PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

CC BY

Opinion piece

Cite this article: Iwaszkiewicz-Eggebrecht E, Zizka V, Lynggaard C. 2024 Three steps towards comparability and standardization among molecular methods for characterizing insect communities. *Phil. Trans. R. Soc. B* **379**: 20230118.
https://doi.org/10.1098/rstb.2023.0118

Received: 29 September 2023
Accepted: 10 December 2023

One contribution of 23 to a theme issue 'Towards a toolkit for global insect biodiversity monitoring'.

Subject Areas:
genetics, ecology

Keywords:
comparability, metadata, COI synthetic spike-in, metabarcoding, internal standard, multiplex

Author for correspondence:
Ela Iwaszkiewicz-Eggebrecht
e-mail: ela.iwaszkiewicz@nrm.se

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.7159013.

THE ROYAL SOCIETY PUBLISHING

Three steps towards comparability and standardization among molecular methods for characterizing insect communities

Ela Iwaszkiewicz-Eggebrecht[1], Vera Zizka[2] and Christina Lynggaard[3]

[1]Bioinformatics and Genetics Department, Swedish Museum of Natural History, PO Box 50007, Stockholm, 104 05, Sweden
[2]Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig Bonn, 53113, Germany
[3]Section for Molecular Ecology & Evolution, Globe Institute, Faculty of Health and Medical Sciences, University of Copenhagen, 1353 Copenhagen, Denmark

EI-E, 0000-0003-1412-1711

Molecular methods are currently some of the best-suited technologies for implementation in insect monitoring. However, the field is developing rapidly and lacks agreement on methodology or community standards. To apply DNA-based methods in large-scale monitoring, and to gain insight across commensurate data, we need easy-to-implement standards that improve data comparability. Here, we provide three recommendations for how to improve and harmonize efforts in biodiversity assessment and monitoring via metabarcoding: (i) we should adopt the use of synthetic spike-ins, which will act as positive controls and internal standards; (ii) we should consider using several markers through a multiplex polymerase chain reaction (PCR) approach; and (iii) we should commit to the publication and transparency of all protocol-associated metadata in a standardized fashion. For (i), we provide a ready-to-use recipe for synthetic cytochrome *c* oxidase spike-ins, which enable between-sample comparisons. For (ii), we propose two gene regions for the implementation of multiplex PCR approaches, thereby achieving a more comprehensive community description. For (iii), we offer guidelines for transparent and unified reporting of field, wet-laboratory and dry-laboratory procedures, as a key to making comparisons between studies. Together, we feel that these three advances will result in joint quality and calibration standards rather than the current laboratory-specific proof of concepts.

This article is part of the theme issue 'Towards a toolkit for global insect biodiversity monitoring'.

1. Introduction

Insects represent one of the most diverse and ecologically significant groups of animals on Earth [1]. Accurate and efficient identification of insect species underpins numerous scientific disciplines, including ecology, evolutionary biology, agriculture and medical entomology [2–4]. While traditional morphological techniques have historically dominated insect taxonomic studies, these methods are limited by the availability of skilled taxonomists, the existence of cryptic species, and the degradation of morphological features in fragmented or preserved specimens [5]. DNA-based molecular methods, such as DNA barcoding and metabarcoding, play a crucial role in expediting taxonomic identification owing to their easy implementation and high discriminatory power [6,7]. While DNA barcoding is primarily employed for individual species identification [8], it has also recently been adapted for high-throughput specimen-based barcoding [9]. However, currently, DNA metabarcoding stands as the predominant method

© 2024 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

used for community-level characterization [10]. In metabarcoding, DNA is isolated directly from mixtures of different specimens and species (bulk samples) or the samples' fixative [11]. DNA traces shed in the environment by organisms can also be extracted from environmental substrates (e.g. soil, water, air), yielding so-called environmental (e)DNA metabarcoding [12]. Subsequently, taxonomically informative DNA regions—barcodes—are mass-amplified and sequenced in parallel using high-throughput technologies (HTS). By bioinformatical analysis of sequences and by comparisons to reference databases, we may thus identify a broad spectrum of species present in those complex samples.

The field of metabarcoding is rapidly and continuously developing, but there is little agreement on basic methodological choices and standard operating procedures. In this context, the standardization and harmonization of technical protocols and data reporting is a pressing issue [13,14]. Only by such harmonization may we ensure the reproducibility of research, large-scale data synthesis and the transfer into applied biodiversity monitoring [15,16]. Recent work by Arribas *et al*. [17] and Chua *et al*. [18] highlighted the fact that current metabarcoding procedures vary tremendously from one study to another. In response, these authors called for efforts to develop standard procedures. Nonetheless, achieving protocol standardization is challenging owing to the fast-evolving and diverse nature of the field. Since current approaches are marked by high variability of techniques and sample types, it would require a great effort to arrive at a universally agreed method for the metabarcoding of insect diversity. Neither may such streamlining be wholly desirable, as the constant development of new approaches is a key motor for driving the field forwards.

Nonetheless, the current state of the art is also an obstacle to the advancement of global insight into the insect fauna through DNA metabarcoding. Most importantly, it hampers comprehensive analyses across studies. Since different methods come with different biases and different resolution, we lack a common benchmark for gauging study-specific patterns against each other. For other fields of analytical science, the current state of metabarcoding would be inconceivable. Consider analytical chemistry. Here, all calibration is based on comparison to a set of samples of known contents—and in support of one's findings, one would naturally report on the precision of one's measures. In metabarcoding, such calibration is typically made against a study-specific sample (or 'mock community'; [19–21]) of a composition purpose-invented by the laboratory in question. In the end, we face a high probability of comparing apples to oranges.

We are then confronted with a situation where biodiversity measurements are urgently needed and new methods are constantly being developed, but joint standards seem out of reach. As an alternative to rigorous standardization which implements the usage of strictly frameworked and compulsory protocols, we feel that the preferred route may be another one: that we should aim for at least a minimum set of recommendations which may in themselves promote data comparability. Therefore, standardization here refers to the integration of quality and reporting standards in metabarcoding protocols and not the streamlining of protocols themselves. In this regard, we discuss: (i) universal internal standards for normalization as well as quality assessment and provide an easy-to-use synthetic spike-in protocol that can be used in metabarcoding assays; (ii) the need for consent on a group-specific marker gene, but also the implementation of multiplex polymerase chain reaction (PCR) approaches, thereby achieving more comprehensive taxonomic coverage; and (iii) the necessity of minimum metadata requirements, and for open-access availability (figure 1).

Overall, these considerations will improve the comparability and reproducibility of insect metabarcoding surveys across samples and experiments, while circumventing the need for strictly enforced standardized guidelines. We are confident that an increased comparability of research will enhance large-scale meta-analysis of data and the integration of molecular approaches in applied surveys.

2. Universal internal standards

To arrive at a benchmark for calibration, an increasing number of sequencing studies (i.e. genomics, transcriptomics) rely on the use of so-called 'spike-ins' [22–24], also known as internal standards (ISD, [25]). In the context of metabarcoding, this term refers to the deliberate addition of a precise quantity of a known DNA sequence to each experimental sample. The integration of spike-ins into the metabarcoding workflow serves a dual purpose (figure 2). Firstly, they act as sample-specific positive controls, enhancing the evaluation of data quality. Moreover, spike-ins play a pivotal role in decreasing variation that occurs during molecular processing and sequencing. By providing a consistent reference point against which the abundance of a sample's DNA sequences can be gauged, they effectively correct for biases or technical variations introduced during the experimental workflow [25,26]. Bringing all samples to an approximately uniform scale allows, in principle, for meaningful comparisons between samples and experiments. Additionally, this implementation of standards holds the potential to transition from assessing relative abundances to quantifying absolute numbers, thereby overcoming the major limitation of metabarcoding datasets [26–31].

Spike-ins have been successfully applied and proved useful in multiple microbial community studies (reviewed in [25]). For instance, Lin *et al*. [32] spiked-in seawater samples with genomic DNA from yeast and thermophilic bacterium. The authors concluded that spike-ins allowed them to quantitatively compare microbial communities across different samples, as well as to accurately establish phytoplankton chloroplast 16S and genomic 18S rRNA gene abundances. Tkacz and collaborators [30] developed synthetic spike-ins for 16S, 18S and ITS markers, and added them to environmental soil samples. They demonstrated an improved absolute quantification of the microbial community. Despite the utility of spike-ins, they have not gained much use in animal metabarcoding studies, where the dominating marker is a fragment of the cytochrome *c* oxidase gene (COI). So far, only a relatively small number of studies have used COI spike-ins. For instance, Ji *et al*. [27] used a specified mixture of COI barcode amplicon DNA of three insect species to spike-in mock communities as well as standardized environmental samples. They showed that spike-ins performed well as internal standards and improved correction for stochasticity introduced during molecular processing and sequencing. Luo *et al*. [26] used the same three spike-ins as Ji *et al.* [27], adding them to a serially diluted mixture of DNA extracts from various arthropod species (called 'mock soup'), thus demonstrating that spike-ins helped recovering

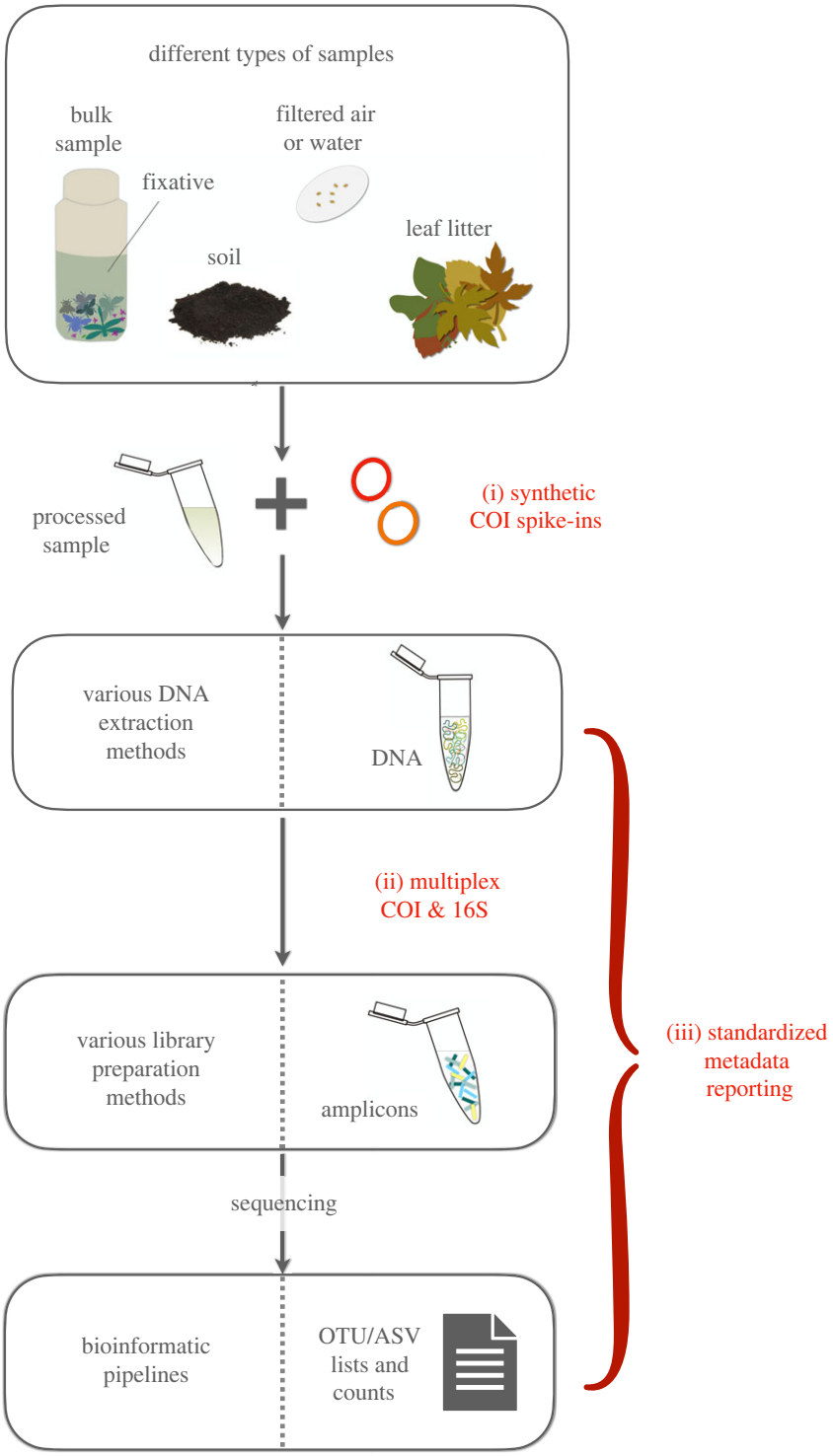


Figure 1. Schematic metabarcoding workflow with the proposed attempts of harmonization highlighted in red. While the overall procedure is variable (different substrates, DNA extraction and libraries preparation methods, bioinformatic pipelines), three proposed approaches for improved standardization can be adopted: (i) integration of artificial spike-ins, (ii) consistent marker gene amplification during PCR and consideration of multiplex approaches, and (iii) standardized reporting of metadata. OTU, operational taxonomic unit; ASV, amplicon sequence variant.

within-species abundance change between dilutions of the 'mock soups'. Marquina *et al*. [33] found COI ISD to be a useful tool in assessing the preservation of DNA after long-term storage in a range of ethanol concentrations. They added a spike-in standard right before the DNA amplification step and then compared the ratio of insect amplicon copy numbers to this spike-in.

When considering internal standards for use in insect biodiversity surveys, one should distinguish three types: (i) biological spike-ins—in which insect specimens are added to every sample in the same numbers [21]; (ii) DNA spike-in—in which pre-amplified insect marker DNA is added in a specific amount [26,27]; and (iii) synthetic spike-in—in which artificial DNA molecules, designed *in silico* and synthesized in a laboratory, are added to the sample. For ease of identification, these molecules are designed to lack similarity to any sequence in public databases [24,31].

While each of the spike-in types has its advantages and limitations, there are a few common aspects to consider when searching for universal spike-in for insect studies. Firstly, the principle of internal standard demands that the same amount of the spike-in material is added to each sample. DNA and synthetic spike-ins allow, in principle, for precise measurements of the input material,

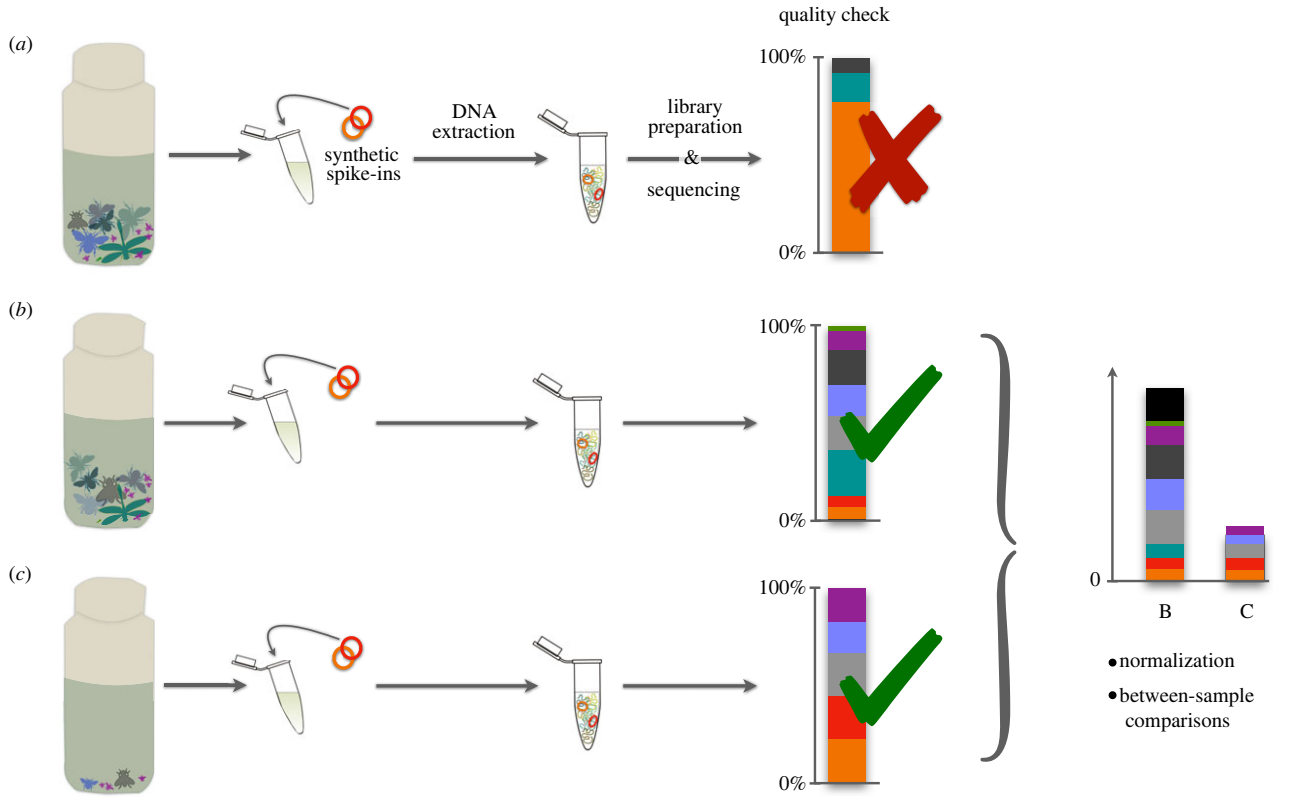


Figure 2. Spike-ins improve quality checks and facilitate normalization. Synthetic spike-ins aid in quality assessment, with deviations in the number and relative proportions of spikes indicating procedural issues (*a*). When spike-ins yield expected proportions of reads (*b*) and (*c*), they can be used for normalization, correcting for amplification and dilutions, and between-sample comparisons.

while biological spike-ins do not permit full control of the amount of input DNA owing to biological differences between spike-in individuals [21]. Secondly, an efficient spike-in must be readily available in almost infinite amounts, both at present and in the future. As we aspire to employ long-term terrestrial insect surveys over decades, we must make sure that the same internal standard will be available in the future. Long-term supply of biological spike-ins is difficult to assure and maintaining populations is costly and time-consuming. DNA and synthetic spike-ins can be amplified into infinity; however, the DNA spike-ins have a limited source (actual insects) and demand careful maintenance. In an event of DNA degradation or material damage, the source can be lost and difficult (or impossible) to recreate. Synthetic spike-ins are free from this problem, as they can be resynthesized anywhere at any moment using commercial services and following standard protocols. Last but not the least, spike-in DNA sequence must be easily distinguished from the sample material being studied. This implies that the species chosen as a source of biological or DNA spike-ins must be carefully selected to ensure they are not present in the region where the study samples come from. This fact precludes the universal, worldwide use of biological or DNA spike-ins. Synthetic spike-ins appear particularly attractive as a solution to this issue, as they are by design different from any biological sequence in the databases and can be added to any sample irrespective of their origin and composition. Taking all this into consideration, and consistent with the conclusions of Harrison *et al*. [25], synthetic spike-ins are the most suitable choice for DNA-based community studies. Nevertheless, to the best of our knowledge, there are currently no universal synthetic spike-ins in use for COI-based metabarcoding biodiversity surveys.

Here, we address this gap by proposing two synthetic COI spike-ins, named *Callio-synth* and *tp53-synth*. *Callio-synth* has already been described and used in the DNA degradation experiment by Marquina *et al*. [33], but for clarity we will summarize the design in brief here. As the starting point for designing both spike-ins, we used a sequence of a bluebottle fly (*Calliphora vomitoria*). We preserved the primer binding sites that match the primers BF3-BR2 [34] and fwhF2-fwhR2n [35] as well as a flanking region around each primer site (±6 bp). In the case of *Callio-synth*, the remainder of the sequence was replaced by a random DNA sequence that does not resemble any known sequence deposited in GenBank, while keeping GC content similar to the original *C. vomitoria* sequence. For the other spike-in, we replaced sequences between preserved primer-binding sites with short fragments of the *tp53* gene (*tp53-synth*). Since the *tp53* sequences are very short and used in the context of COI metabarcoding, we do not risk any confusion when analysing results. Both spike-in sequences (471 bp total length) can be easily synthesized by a commercial laboratory anywhere in the world. They should then be inserted into standard plasmids (also commercially available) and transformed into monoclonal bacteria—a stable system for long-term storage and amplification. Spike-ins can be extracted from the bacterial culture whenever needed by commonly used kits. The concentration of spike-in DNA must be quantified and the same number of spike-in copies should be added to every sample before DNA extraction. Subsequently spike-ins are co-extracted, co-amplified and sequenced together with the DNA of the insect sample. A step-by-step protocol providing spike-in sequences and detailed information on design and how to proceed with bacterial cultures and quantifications is available following the link: https://dx.doi.org/10.17504/protocols.io.14egn33ryl5d/v2 (also available as the electronic supplementary material, file S1). More information on how to design a synthetic spike-in is summarized in the review by Harrison *et al*. [25].

As the use of COI synthetic spike-ins is in its nascent stage, many questions naturally remain. As a first question, we may ask whether the use of just two spike-ins is enough? The comprehensive review by Harrison *et al*. [25] suggests that at least three distinct spike-ins should be added to each sample and that evaluating ratios of spike-in reads should become part of the quality control. As a second question, we may consider how much spike-in should be added? The exact amount will depend on the nature of the sample, i.e. whether it has high DNA yields (e.g. homogenized tissue samples) or just trace amounts (i.e. eDNA). The challenge is to have enough reads to make full use of spike-ins while not losing too much sequencing bandwidth on them. The literature review conducted by Harrison *et al*. [25] suggests that one should aim for spike-ins to constitute 1–3% of the total DNA in a sample. Naturally, it is not an easy goal to achieve when processing large numbers of diverse samples. However, dividing samples into distinct size categories, based on insect biomass or the quantity of input DNA, presents a feasible approach for incorporation of suitable levels of spike-ins. In our proposed protocol, developed for a large collection of homogenized Malaise trap samples (originating from the Insect Biome Atlas project, insectbiomeatlas.org), we add approximately five million copies of each spike-in to samples. While this exact procedure can only be recommended for samples of similar type, the rationale and methodology outlined in the protocol can serve as a template or starting point for other studies. Undoubtedly, standards regarding the exact amount of spike-in DNA appropriate for different kinds of samples remain to be developed and agreed upon by the community. A third question relates to how the presence of spike-ins influences community reconstruction of the samples? It is known that presence of one DNA sequence can affect the recovery of other sequences present in the samples [36]. Perhaps the addition of synthetic spike-ins could influence our ability to recover certain COI barcodes? While this aspect remains to be studied, a simple null model is to assume that the same effect will hold across all samples and studies that use them.

The proposed spike-ins are designed to work in experiments targeting two different primer pairs (or combination of those primers). Of these, *BF3-BR2* produces a 418 bp long fragment of the Folmer region and is commonly used in Illumina-based studies, whereas *fwhF2-fwhR2n* allows for the amplification of a shorter fragment, only 205 bp long, which is particularly useful when dealing with degraded DNA. If a different pair of primers is desired, or the study focuses on a full-length barcode, then new synthetic spike-ins will be necessary. The same holds true if a single reaction targets different markers to improve taxonomic resolution (multiplex approaches—see next section). Here synthetic spike-ins for all target fragments should be integrated in the process. We do not claim to have generated the final set of synthetic spike-ins, but merely encourage the community to discuss, test and develop synthetic spike-ins, to thereby bring the whole field into a more rigorously quantitative and standardized future.

We propose that synthetic COI spike-ins should become the standard and be included in all insect surveys, as they are a powerful tool improving the reliability and accuracy of COI metabarcoding. However, we also highlight that the synthetic spike-ins put forward in this paper are merely a starting point. For future development, it will be important to consider expanding the length of the spike-in to cover the whole Folmer region, to design additional spike-ins and to test them in different combinations and concentrations in order to find the optimal standard spike-in mixture for universal use.

3. Targeting multiple gene regions

In both DNA barcoding and metabarcoding, it is possible to target and amplify different mitochondrial gene regions (such as COI, and the ribosomal RNA genes 16S, 12S, 18S and 28S) through PCR. Given that these molecular methods rely on databases to assign taxonomy to the sequences obtained, the completeness of the database is crucial for accurate assignment [37]). At present, the database for COI is more complete than that of other regions [38]. Therefore, if using other regions, the taxonomic resolution will be influenced by the database incompleteness. As a further advantage, the COI region presents enough variability between species and sufficiently little variation within species to allow for accurate species-level taxonomic assignment [17]. All this has led researchers to recommend targeting the COI region in insect community metabarcoding (e.g. [39]). Nonetheless, at lower taxonomic resolution, other loci such as 16S rRNA can provide a better overview of the insect community present in the sample. In some cases its use has provided data complementary data to that obtained with COI [40,41]. Because of this, and together with primer bias in the efficiency of PCR amplification [42,43], some researchers recommend and have used more than one primer set [44–47]. Nevertheless, this can be labour intensive and expensive, as it increases the amount of reagents needed for PCR reactions and sequencing if each primer is used in a separate reaction. Therefore, to keep the costs down and to allow comparability between studies, there is consensus on the use of the COI region as the first option [38].

In a multiplex approach, several primer sets are used in the same amplification reaction, which thereby produces multilocus data [6,48]. If the primers are chosen to generate amplicons of different lengths, it is possible to differentiate which sequences belong to which primer [49]. However, multiplexing PCR can present technical challenges especially in mixed samples where the amount of the template DNA differs across taxa. Owing to differences in the signal from these taxa, primer concentration in the reaction mix must be compensated based on their amplification efficiency [50]. In addition, other calibrations should be performed to balance the reaction, such as optimizing thermocycling conditions and making sure there are no cross primer dimers (see [48,51] for more detailed information). The sensitivity and specificity of the assay can be assessed by the use of spike-ins as described in the previous section.

Multiplexing has been successfully used with species-specific primers. For example, it has been used to target and amplify the 16S rRNA and COI region of several species of edible insects [52]. Multiplex has also been implemented using general primers targeting a wider taxonomic spectrum. For example, de Kerdrel and collaborators [53] used multiplex PCR to target four nuclear DNA regions (28SrDNA, V1-2 and V6-7 regions of 18SrDNA and histone H3) in bulk arthropod samples. In evidence of its efficiency, multiplexing has been used to amplify spider DNA through a mitochondrial multiplex assay simultaneously targeting the COI, 12S and 16S regions and a nuclear multiplex assay targeting the D6 region of the 28S gene, the V1-V2 and the V6-V7 regions of the 18S, the ribosomal ITS2 and histone H3 [48]. In addition, the COI, 16S, 18S and 28S genes have been targeted to study the

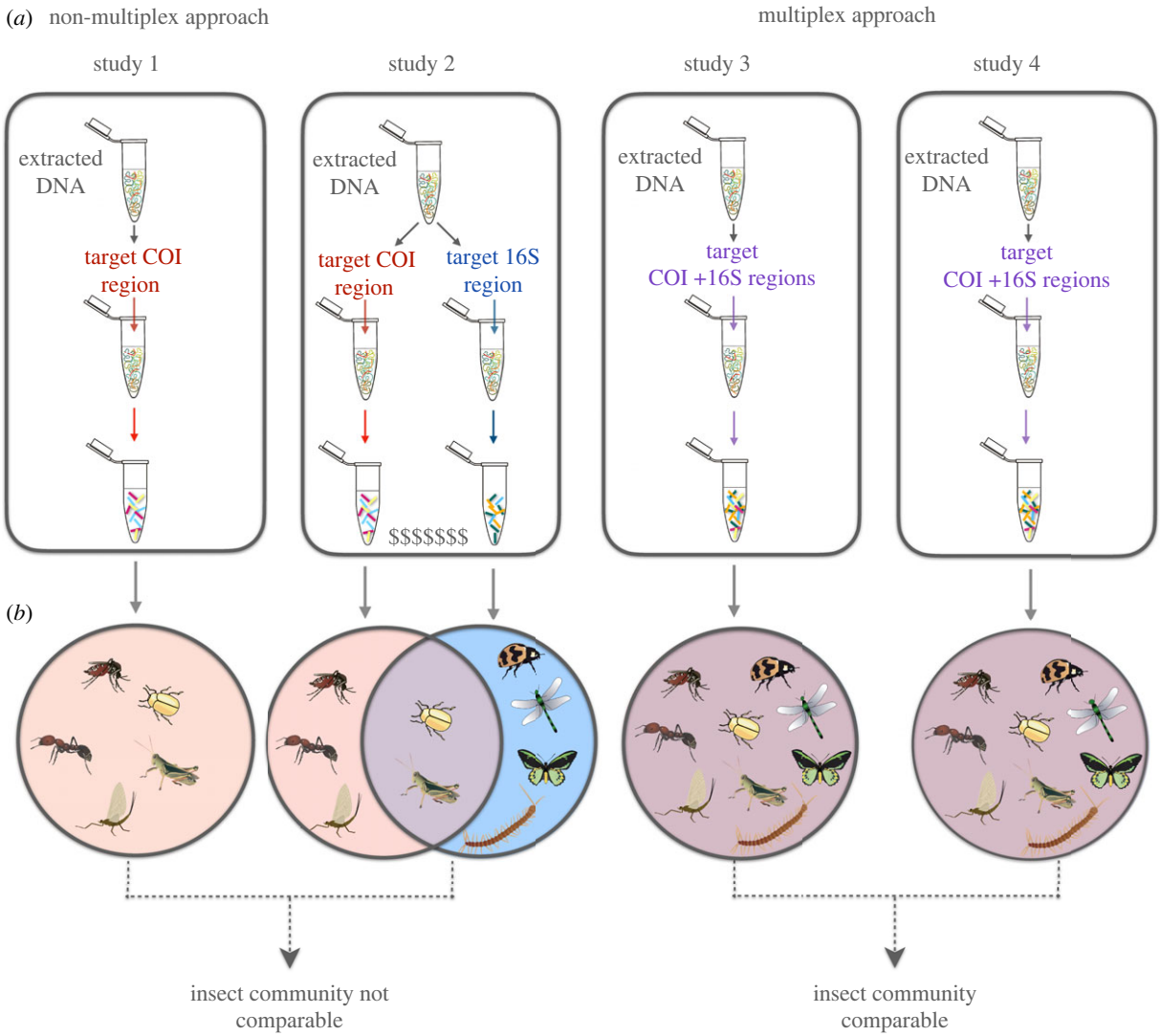


Figure 3. The choice of the gene region targeted in metabarcoding studies will impact the insect community detected and can hinder data comparability across studies. (*a*) Targeting several gene regions using a multiplex approach can be time and cost-effective and (*b*) targeting the same regions can further allow data comparability between different studies. Animal images in the figure obtained from the Integration and Application Network, University of Maryland Center for Environmental Science (ian.umces.edu/symbols/).

diet of spiders [54], and the COI, 12S and 18S genes for pest insect identification [55]. Batuecas *et al*. [49] used two primers targeting different regions within COI. These studies show that the use of a multiplex approach for arthropod identification is successful, but they also illustrate the lack of consensus about which loci to target.

Although the multiplex method has been applied in several studies, it is still less frequently used than the non-multiplex approach. While Krehenwinkel *et al*. [56] targeted different regions within the COI gene to identify insect DNA in tea bags, we are not aware of other studies implementing multiplex PCR for targeting insect DNA in environmental samples (eDNA). This seems a missed opportunity, as it could accelerate the application of high-throughput surveillance [55], especially if also applied to new sources of insect DNA such as air [47,57], rain water [58] and other novel DNA substrates (reviewed in [18]).

Therefore, we suggest that the multiplex approach could be routinely used across different research groups, projects and sample types. The use of several loci can improve the characterization of the community [40] and support phylogenetic analyses [48]. Further, if multiple laboratories were to target the same loci in their multiplex PCR, then this would allow comparability across laboratories (figure 3). In conclusion, we consider that owing to the relative completeness of the reference database and the possibility of obtaining high taxonomic resolution of the sequenced data, it is important to keep targeting the COI gene as a standard solution for metabarcoding. In addition to COI, another primer set targeting a different region should be added in a standard multiplex assay. Beyond COI, the regions most commonly used (also in a multiplex setting) are 18S and 16S. While the 18S region is generally used for a broader taxonomic overview including metazoans [59], the 16S region has been proved to outperform 18S when targeting insect DNA, especially in the detection of dipterans [59]. Further, the 16S region has been found to provide data complementary to that gained from the COI gene [40]. Therefore, we suggest that 16S should be included in the multiplex assay, together with the COI gene. However, we also need to point out that reference databases for this locus should be improved to allow for better taxonomic assignments. Naturally, we encourage the use of other loci, too, but offer COI and 16S as a solid starting point. Finally, we suggest that the use of multiplex PCR will prove useful beyond insects in bulk samples, and should be explored further when targeting small traces of insect DNA in environmental samples.

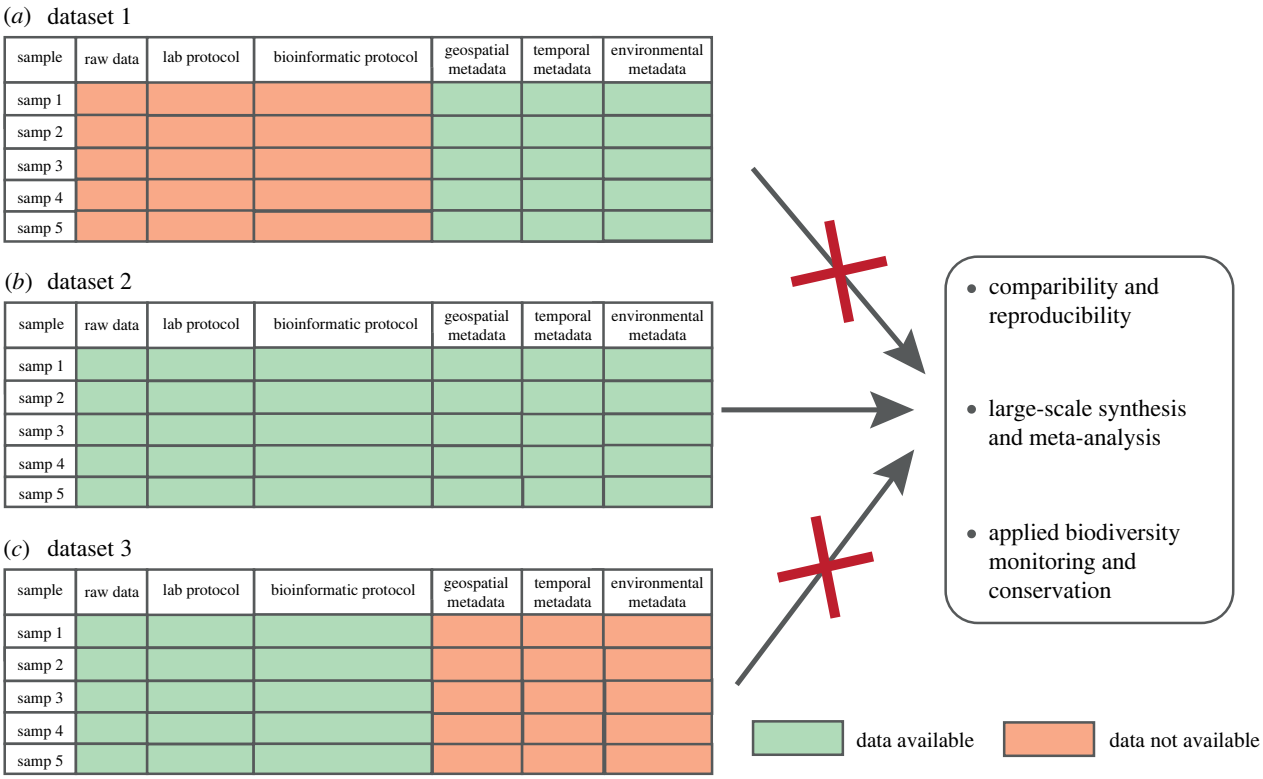


Figure 4. Complete and well reported metadata (*b*) strengthens comparability of different metabarcoding studies and enables meta-analysis. Incomplete and non-published documentation of metadata from metabarcoding community analysis (*a*) and (*c*) prevents comparability of datasets, hampering meta-analysis and large-scale synthesis as well as the integration in applied monitoring schemes and conservation. *Raw data*: link to archive (such as NCBI Sequence Read Archive or EBI-ENA), where raw sequence data are stored also including sample description; *lab protocol*: laboratory protocol information based on the MIMARKS reporting standards (see main text, [72]) and extended in the electronic supplementary material, table S2—laboratory protocol); *bioinformatic protocol:* minimum reporting standards of main programs and parameter settings for bioinformatic sequence analysis as summarized in the electronic supplementary material, table S2—bioinformatic analyses; *geospatial, temporal and environmental metadata:* all sample associated metadata as proposed in MIMARKS standards and associated environments (e.g. through GEOME platform, see main text).

4. Standardization of metadata reports

Molecular biodiversity research yields enormous quantities of community-level data, comprising raw sequence material but also associated technical protocols and environmental metadata [17,60]. Accessibility of raw sequences and technical information is a requisite for retrospective data analysis, long-term studies and global syntheses, and so is public access to environmental and ecological metadata [20,61–63]. While centralized publication of raw sequence data is widely adapted in metabarcoding surveys through platforms such as the National Center for Biotechnology Information (NCBI) Sequence Read Archive [64] or the European Bioinformatics Institute Nucleotide Archive (EBI-ENA—also linked with MGnify as a platform for comparative analysis for microbiome data) [62,65,66], the associated metadata are often scarce, inconsistent and/or non-standardized—or remain completely unreported [67–70]. Here, current tools and standards for data deposition remain of little help, as they tend to focus exclusively either on genomic or ecological attributes [62,71]. The unavailability of technical and ecological metadata hampers transparency and reproducibility of research, preventing data exchange and reuse and subsequently the integration of data into large scale meta-analysis (figure 4). This is a major concern, as it restricts access to metabarcoding data by governmental conservation organizations and large-scale monitoring initiatives, which rely on standardized data collection and reporting guidelines [61,68,73,74]. While huge variability in laboratory protocols, bioinformatic pipelines and sample types impedes strict technical standardizations (see Introduction), there is clearly a need for improved reporting standards [17,62,72]. This necessity for streamlined reporting of the metadata behind molecular surveys has been discussed by different consortia and publications (e.g. [17,61,62,67,75]). Nonetheless, given the continuing challenges in making an actual change, we here want to stress the salient points and the key items needed to ensure the true replicability and large-scale synthesis of data.

To remedy the state of the art, several authors have argued for the use of data sheets with at least minimum metadata to be associated with published HTS data [17,62,76] with a focus on eDNA approaches [68,75,77]. However, for large-scale analysis of insects, different amplicon-based methods can be used, and metadata storage should not exclusively focus on eDNA standards. Checklists for minimum information to improve mechanisms of metadata capture and exchange have been initiated by the Genomics Standards Consortium. Those were initially designed for genomic and metagenomic approaches targeting microorganisms, with the attempt to record minimum technology-specific, geospatial and temporal metadata (see [72]). Extant checklists (e.g. 'minimum information about a marker gene sequence (MIMARKS) (https://genomicsstandardsconsortium.github.io/mixs/, [60,61,72] can be universally applied to any marker gene at levels from single individuals to complex communities and incorporates standards developed by the Consortium for the Barcode of Life (CBOL). Yilmaz *et al*. [72] identified the minimum information to be associated with the sequence data, including geospatial and technical details (e.g. sample collection method and device, sample size, target gene, PCR condition, integrated negative and positive controls). Requirements for minimum information are available and differ slightly for different environmental samples (e.g. water, soil, sediment, air). In brief, rather than reinventing the wheel, we might thus adopt

the checklists introduced for minimum sequence information as a first standard for the publication of DNA metabarcoding metadata and reporting. This is also the goal of Genomic Observatories Metadatabases (GEOME, https://geome-db.org), which is an open access platform permanently linking molecular raw data (Sequence Read Archive) with associated sample metadata including technical protocols and environmental information [71].

However, while previous checklists for metadata do include detailed information from sampling to sequencing, the information required about bioinformatic processing and taxonomic assignment of sequences is limited (including steps such as sequence quality check, chimera identification, etc.), especially if no precast pipeline is used ([78] see [79] for detailed review of existing pipelines). Therefore, we here introduce further minimum information about sequence analysis that ensures reproducibility and should be considered for publication in metabarcoding studies in table format (electronic supplementary material, file S2). Notably, such a solution can also reduce the textual portion in the material and methods section. In particular, we solicit data on the programmes and detailed parameter settings used for the different steps of bioinformatic analysis. In addition, the table provided (electronic supplementary material, file S2) includes minimum information about sample replication, integrated positive and negative controls and how those are used in downstream data quality improvement.

While the need for standard open-access metadata publication along with metabarcoding datasets has been repeatedly raised [17,62,69,76], documentation remains poor. By pointing to extant minimum requested checklists and associated platforms, we emphasize the continuing need for open-access publication of environmental and technical data, and for adequate information about bioinformatic analysis. We believe that the availability of geospatial and environmental metadata, and the standardized reporting of the protocols applied, forms the very basis for cross-laboratory meta-analysis of metabarcoding data and the successful transfer of methods into applied monitoring schemes.

5. Moving towards comparability and standardization

In the light of the increasing volume of insect DNA data being generated globally, there is undeniable evidence of the successful application of molecular methods in yielding valuable insights. Nevertheless, the question regarding how to create comprehensive insect monitoring schemes still remains, as each study uses different protocols, calibrates the data in different ways and reports the data using different formats. We suggest that rather than pursuing strict standardization of protocols, the adoption of a small number of easy-to-implement standards is crucial. To this end, we propose (i) synthetic spike-ins as internal standards, (ii) emphasis on a standard marker gene but consider additional fragments through multiplex PCR, and (iii) commit to the publication and transparency of all protocol-associated metadata.

To unlock the full potential of molecular methods for characterizing insect communities, global standardization and cross-study comparisons are imperative. We believe that the three suggested advances will help in achieving uniform quality and calibration across different studies, thereby paving the way for large-scale data meta-analysis.

Data accessibility. The protocol detailing use of synthetic spike-ins is published on protocols.io (doi:10.17504/protocols.io.14egn33ryl5d/v2): https://www.protocols.io/view/synthetic-coi-spike-ins-for-use-in-metabarcoding-b-14egn33ryl5d/v2 [80].

Supplementary material is available online [81].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. E.I.-E.: conceptualization, supervision, visualization, writing—original draft, writing—review and editing; V.Z.: conceptualization, visualization, writing—original draft, writing—review and editing; C.L.: conceptualization, visualization, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. E.I.-E. was supported by the Knut and Alice Wallenberg Foundation (grant no. KAW 2017.088). C.L. was further supported by a research grant no. (grant no. VIL41390) from VILLUM FONDEN. V.Z. is funded by the German Federal Ministry of Education and Research (grant no. 01UT2101A).

Acknowledgements. We are thankful to Tomas Roslin from the Swedish University of Agricultural Sciences (SLU) for his help in shaping the ideas and the manuscript. We also thank Niklas Noll from the Leibniz Institute for the Analysis of Biodiversity Change (LIB) for helpful comments to table the electronic supplementary material, file S2. We are grateful to Piotr Łukasik from Jagiellonian University (Poland) for pioneering the use of synthetic COI spike-ins. He also provided helpful feedback on the first version of the manuscript. The spike-ins were beautifully produced by Monika Prus-Frankowska and Anna Michalik from Jagiellonian University and Monika helped immensely in putting together the protocol.

References

1. Grimaldi D, Engel M. 2005 *Evolution of the insects*. Cambridge, UK: Cambridge University Press.
2. Eggleton P. 2020 The state of the World's insects. *Annu. Rev. Environ. Resour.* **45**, 61–82. (doi:10.1146/annurev-environ-012420-050035)
3. Jankielsohn A. 2018 The importance of insects in agricultural ecosystems. *Adv. Entomol.* **6**, 62–73. (doi:10.4236/ae.2018.62006)
4. Scudder GGE. 2017 The importance of insects. In *Insect biodiversity* (eds RG Foottit, PH Adler), pp. 9–43. Chichester, UK: John Wiley & Sons, Ltd.
5. Kim KC. 2017 Taxonomy and management of insect biodiversity. In *Insect biodiversity* (eds RG Foottit, PH Adler), pp. 767–782. Chichester, UK: John Wiley & Sons, Ltd.
6. Compson ZG, McClenaghan B, Singer GAC, Fahner NA, Hajibabaei M. 2020 Metabarcoding from microbes to mammals: comprehensive bioassessment on a global scale. *Front. Ecol. Evol.* **8**, 581835. (doi:10.3389/fevo.2020.581835)
7. Ge Y, Xia C, Wang J, Zhang X, Ma X, Zhou Q. 2021 The efficacy of DNA barcoding in the classification, genetic differentiation, and biodiversity assessment of benthic macroinvertebrates. *Ecol. Evol.* **11**, 5669–5681. (doi:10.1002/ece3.7470)
8. Hebert PDN, Cywinska A, Ball SL, de Waard JR. 2003 Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B* **270**, 313–321. (doi:10.1098/rspb.2002.2218)
9. Hartop E, Srivathsan A, Ronquist F, Meier R. 2022 Towards large-scale integrative taxonomy (LIT): resolving the data conundrum for dark taxa. *Syst. Biol.* **71**, 1404–1422. (doi:10.1093/sysbio/syac033)

10. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. 2012 Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21**, 2045–2050. (doi:10.1111/j.1365-294X.2012.05470.x)
11. Kirse A, Bourlat SJ, Langen K, Zapke B, Zizka VMA. 2023 Comparison of destructive and nondestructive DNA extraction methods for the metabarcoding of arthropod bulk samples. *Mol. Ecol. Resour.* **23**, 92–105. (doi:10.1111/1755-0998.13694)
12. Lacoursière-Roussel A, Deiner K. 2021 Environmental DNA is not the tool by itself. *J. Fish Biol.* **98**, 383–386. (doi:10.1111/jfb.14177)
13. Blackman RC *et al.* 2019 Advancing the use of molecular methods for routine freshwater macroinvertebrate biomonitoring – the need for calibration experiments. *Metabarcoding Metagenomics* **3**, 49–57. (doi:10.3897/mbmg.3.34735)
14. Loeza-Quintana T, Abbott CL, Heath DD, Bernatchez L, Hanner RH. 2020 Pathway to increase standards and competency of eDNA Surveys (PISCeS)—advancing collaboration and standardization efforts in the field of eDNA. *Environ. DNA* **2**, 255–260. (doi:10.1002/edn3.112)
15. Blancher P *et al.* 2022 A strategy for successful integration of DNA-based methods in aquatic monitoring. *Metabarcoding Metagenomics* **6**, e85652. (doi:10.3897/mbmg.6.85652)
16. Mergen P, Meissner K, Hering D, Leese F, Bouchez A, Weigand A, Kueckmann S. 2018 DNAqua-Net or how to navigate on the stormy waters of standards and legislations. *Biodiv. Inform. Sci. Standards* **2**, e25953. (doi:10.3897/biss.2.25953)
17. Arribas P *et al.* 2022 Toward global integration of biodiversity big data: a harmonized metabarcode data generation module for terrestrial arthropods. *GigaScience* **11**, giac065. (doi:10.1093/gigascience/giac065)
18. Chua PYS, Bourlat SJ, Ferguson C, Korlevic P, Zhao L, Ekrem T, Meier R, Lawniczak MKN. 2023 Future of DNA-based insect monitoring. *Trends Genet.* **39**, 531–544. (doi:10.1016/j.tig.2023.02.012)
19. Braukmann TWA *et al.* 2019 Metabarcoding a diverse arthropod mock community. *Mol. Ecol. Resour.* **19**, 711–727. (doi:10.1111/1755-0998.13008)
20. Bruce K *et al.* 2021 A practical guide to DNA-based methods for biodiversity assessment. *Adv. Books* **1**, e68634. (doi:10.3897/ab.e68634)
21. Iwaszkiewicz-Eggebrecht E *et al.* 2023 Optimizing insect metabarcoding using replicated mock communities. *Methods Ecol. Evol.* **14**, 1130–1146. (doi:10.1111/2041-210X.14073)
22. Chen K, Hu Z, Xia Z, Zhao D, Li W, Tyler JK. 2016 The overlooked fact: fundamental need for spike-in control for virtually all genome-wide analyses. *Mol. Cell. Biol.* **36**, 662–667. (doi:10.1128/MCB.00970-14)
23. Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, Gingeras TR, Oliver B. 2011 Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* **21**, 1543–1551. (doi:10.1101/gr.121095.111)
24. Palmer JM, Jusino MA, Banik MT, Lindner DL. 2018 Non-biological synthetic spike-in controls and the AMPtk software pipeline improve mycobiome data. *PeerJ* **6**, e4925. (doi:10.7717/peerj.4925)
25. Harrison JG, John Calder W, Shuman B, Alex Buerkle C. 2021 The quest for absolute abundance: the use of internal standards for DNA-based community ecology. *Mol. Ecol. Resour.* **21**, 30–43. (doi:10.1111/1755-0998.13247)
26. Luo M, Ji Y, Warton D, Yu DW. 2023 Extracting abundance information from DNA-based data. *Mol. Ecol. Resour.* **23**, 174–189. (doi:10.1111/1755-0998.13703)
27. Ji Y, Huotari T, Roslin T, Schmidt NM, Wang J, Yu DW, Ovaskainen O. 2020 SPIKEPIPE: a metagenomic pipeline for the accurate quantification of eukaryotic species occurrences and intraspecific abundance change using DNA barcodes or mitogenomes. *Mol. Ecol. Resour.* **20**, 256–267. (doi:10.1111/1755-0998.13057)
28. Shelton AO *et al.* 2023 Toward quantitative metabarcoding. *Ecology* **104**, e3906. (doi:10.1002/ecy.3906)
29. Sickel W, Zizka V, Scherges A, Bourlat SJ, Dieker P. 2023 Abundance estimation with DNA metabarcoding – recent advancements for terrestrial arthropods. *ARPHA Prepr.* **4**, e109709. (doi:10.3897/arphapreprints.e109709)
30. Tkacz A, Hortala M, Poole PS. 2018 Absolute quantitation of microbiota abundance in environmental samples. *Microbiome* **6**, 110. (doi:10.1186/s40168-018-0491-7)
31. Tourlousse DM, Yoshiike S, Ohashi A, Matsukura S, Noda N, Sekiguchi Y. 2017 Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Res.* **45**, e23. (doi:10.1093/nar/gkw984)
32. Lin Y, Gifford S, Ducklow H, Schofield O, Cassar N. 2019 Towards quantitative microbiome community profiling using internal standards. *Appl. Environ. Microbiol.* **85**, e02634-18. (doi:10.1128/AEM.02634-18)
33. Marquina D, Buczek M, Ronquist F, Łukasik P. 2021 The effect of ethanol concentration on the morphological and molecular preservation of insects for biodiversity studies. *PeerJ* **9**, e10799. (doi:10.7717/peerj.10799)
34. Elbrecht V, Braukmann TWA, Ivanova NV, Prosser SWJ, Hajibabaei M, Wright M, Zakharov EV, Hebert PDN, Steinke D. 2019 Validation of COI metabarcoding primers for terrestrial arthropods. *PeerJ* **7**, e7745. (doi:10.7717/peerj.7745)
35. Vamos EE, Elbrecht V, Leese F. 2017 Short COI markers for freshwater macroinvertebrate metabarcoding. *PeerJ* **5**, e3037v2. (doi:10.7287/peerj.preprints.3037v2)
36. Kelly RP, Shelton AO, Gallego R. 2019 Understanding PCR processes to draw meaningful conclusions from environmental DNA studies. *Sci. Rep.* **9**, 12133. (doi:10.1038/s41598-019-48546-x)
37. Li R, Ratnasingham S, Zarubiieva I, Somervuo P, Taylor GW. 2024 PROTAX-GPU: a scalable probabilistic taxonomic classification system for DNA barcodes. *Phil. Trans. R. Soc. B* **379**, 20230124. (doi:1098/rstb.2023.0124)
38. Andújar C, Arribas P, Yu DW, Vogler AP, Emerson BC. 2018 Why the COI barcode should be the community DNA metabarcode for the metazoa. *Mol. Ecol.* **27**, 3968–3975. (doi:10.1111/mec.14844)
39. Elbrecht V, Taberlet P, Dejean T, Valentini A, Usseglio-Polatera P, Beisel J, Coissac E, Boyer F, Leese F. 2016 Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. *PeerJ* **4**, e1966. (doi:10.7717/peerj.1966)
40. Marquina D, Esparza-Salas R, Roslin T, Ronquist F. 2019 Establishing arthropod community composition using metabarcoding: surprising inconsistencies between soil samples and preservative ethanol and homogenate from Malaise trap catches. *Mol. Ecol. Resour.* **19**, 1516–1530. (doi:10.1111/1755-0998.13071)
41. Sato JJ, Ohtsuki Y, Nishiura N, Mouri K. 2022 DNA metabarcoding dietary analyses of the wood mouse *Apodemus speciosus* on Innoshima Island, Japan, and implications for primer choice. *Mammal Res.* **67**, 109–122. (doi:10.1007/s13364-021-00601-7)
42. Alberdi A, Aizpurua O, Gilbert MTP, Bohmann K. 2018 Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods Ecol. Evol.* **9**, 134–147. (doi:10.1111/2041-210X.12849)
43. Krehenwinkel H, Wolf M, Lim JY, Rominger AJ, Simison WB, Gillespie RG. 2017 Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Sci. Rep.* **7**, 17668. (doi:10.1038/s41598-017-17333-x)
44. Hajibabaei M, Porter TM, Wright M, Rudar J. 2019 COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. *PLoS ONE* **14**, e0220953. (doi:10.1371/journal.pone.0220953)
45. 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. *Mol. Ecol. Resour.* **15**, 819–830. (doi:10.1111/1755-0998.12355)
46. Thomsen PF, Sigsgaard EE. 2019 Environmental DNA metabarcoding of wild flowers reveals diverse communities of terrestrial arthropods. *Ecol. Evol.* **9**, 1665–1679. (doi:10.1002/ece3.4809)

47. Roger F, Ghanavi HR, Danielsson N, Wahlberg N, Löndahl J, Pettersson LB, Andersson GKS, Boke Olén N, Clough Y. 2022 Airborne environmental DNA metabarcoding for the monitoring of terrestrial insects—a proof of concept from the field. *Environ. DNA* **4**, 790–807. (doi:10.1002/edn3.290)
48. Krehenwinkel H, Kennedy SR, Rueda A, Lam A, Gillespie RG. 2018 Scaling up DNA barcoding – primer sets for simple and cost efficient arthropod systematics by multiplex PCR and Illumina amplicon sequencing. *Methods Ecol. Evol.* **9**, 2181–2193. (doi:10.1111/2041-210X.13064)
49. Batuecas I, Alomar O, Castañe C, Piñol J, Boyer S, Gallardo-Montoya L, Agustí N. 2022 Development of a multiprimer metabarcoding approach to understanding trophic interactions in agroecosystems. *Insect Sci.* **29**, 1195–1210. (doi:10.1111/1744-7917.12992)
50. Markoulatos P, Siafakas N, Moncany M. 2002 Multiplex polymerase chain reaction: a practical approach. *J. Clin. Lab. Anal.* **16**, 47–51. (doi:10.1002/jcla.2058)
51. Sint D, Raso L, Traugott M. 2012 Advances in multiplex PCR: balancing primer efficiencies and improving detection success. *Methods Ecol. Evol.* **3**, 898–905. (doi:10.1111/j.2041-210X.2012.00215.x)
52. Tramuta C, Gallina S, Bellio A, Bianchi DM, Chiesa F, Rubiola S, Romano A, Decastelli L. 2018 A set of multiplex polymerase chain reactions for genomic detection of nine edible insect species in foods. *J. Insect Sci.* **18**, 3. (doi:10.1093/jisesa/iey087)
53. de Kerdrel GA, Andersen JC, Kennedy SR, Gillespie R, Krehenwinkel H. 2020 Rapid and cost-effective generation of single specimen multilocus barcoding data from whole arthropod communities by multiple levels of multiplexing. *Sci. Rep.* **10**, 78. (doi:10.1038/s41598-019-54927-z)
54. Krehenwinkel H, Kennedy SR, Adams SA, Stephenson GT, Roy K, Gillespie RG. 2019 Multiplex PCR targeting lineage-specific SNPs: a highly efficient and simple approach to block out predator sequences in molecular gut content analysis. *Methods Ecol. Evol.* **10**, 982–993. (doi:10.1111/2041-210X.13183)
55. Batovska J, Piper AM, Valenzuela I, Cunningham JP, Blacket MJ. 2021 Developing a non-destructive metabarcoding protocol for detection of pest insects in bulk trap catches. *Sci. Rep.* **11**, 7946. (doi:10.1038/s41598-021-85855-6)
56. Krehenwinkel H, Weber S, Künzel S, Kennedy SR. 2022 The bug in a teacup—monitoring arthropod–plant associations with environmental DNA from dried plant material. *Biol. Lett.* **18**, 20220091. (doi:10.1098/rsbl.2022.0091)
57. Pumkaeo P, Takahashi J, Iwahashi H. 2021 Detection and monitoring of insect traces in bioaerosols. *PeerJ* **9**, e10862. (doi:10.7717/peerj.10862)
58. Macher T-H, Schütz R, Hörren T, Beermann AJ, Leese F. 2023 It's raining species: rainwash eDNA metabarcoding as a minimally invasive method to assess tree canopy invertebrate diversity. *Environ. DNA* **5**, 3–11. (doi:10.1002/edn3.372)
59. Ficetola GF, Boyer F, Valentini A, Bonin A, Meyer A, Dejean T, Gaboriaud C, Usseglio-Polatera P, Taberlet P. 2021 Comparison of markers for the monitoring of freshwater benthic biodiversity through DNA metabarcoding. *Mol. Ecol.* **30**, 3189–3202. (doi:10.1111/mec.15632)
60. Field D *et al.* 2008 The minimum information about a genome sequence (MIGS) specification. *Nat. Biotechnol.* **26**, 541–547. (doi:10.1038/nbt1360)
61. Crandall ED *et al.* 2022 Metadata preservation and stewardship for genomic data is possible, but must happen now. *bioRxiv* (doi:10.1101/2022.09.12.507034)
62. Tedersoo L, Ramirez KS, Nilsson RH, Kaljuvee A, Kõljalg U, Abarenkov K. 2015 Standardizing metadata and taxonomic identification in metabarcoding studies. *GigaScience* **4**, s13742-015. (doi:10.1186/s13742-015-0074-5)
63. Thalinger B, Deiner K, Harper LR, Rees HC, Blackman RC, Sint D, Traugott M, Goldberg CS, Bruce K. 2021 A validation scale to determine the readiness of environmental DNA assays for routine species monitoring. *Environ. DNA* **3**, 823–836. (doi:10.1002/edn3.189)
64. Katz K, Shutov O, Lapoint R, Kimelman M, Brister JR, O'Sullivan C. 2022 The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Res.* **50**, D387–D390. (doi:10.1093/nar/gkab1053)
65. Balk MA *et al.* 2022 A solution to the challenges of interdisciplinary aggregation and use of specimen-level trait data. *iScience* **25**, 105101. (doi:10.1016/j.isci.2022.105101)
66. Richardson L *et al.* 2023 MGnify: the microbiome sequence data analysis resource in 2023. *Nucleic Acids Res.* **51**, D753–D759. (doi:10.1093/nar/gkac1080)
67. Harris MA, Slippers B, Kemler M, Greve M. 2023 Opportunities for diversified usage of metabarcoding data for fungal biogeography through increased metadata quality. *Fungal Biol. Rev.* **46**, 100329. (doi:10.1016/j.fbr.2023.100329)
68. Nicholson A *et al.* 2020 An analysis of metadata reporting in freshwater environmental DNA research calls for the development of best practice guidelines. *Environ. DNA* **2**, 343–349. (doi:10.1002/edn3.81)
69. Rimet F *et al.* 2021 Metadata standards and practical guidelines for specimen and DNA curation when building barcode reference libraries for aquatic life. *Metabarcoding Metagenomics* **5**, e58056. (doi:10.3897/mbmg.5.58056)
70. Taylor CF *et al.* 2008 Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nat. Biotechnol.* **26**, 889–896. (doi:10.1038/nbt.1411)
71. Riginos C *et al.* 2020 Building a global genomics observatory: using GEOME (the Genomic Observatories Metadatabase) to expedite and improve deposition and retrieval of genetic data and metadata for biodiversity research. *Mol. Ecol. Resour.* **20**, 1458–1469. (doi:10.1111/1755-0998.13269)
72. Yilmaz P *et al.* 2011 Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nat. Biotechnol.* **29**, 415–420. (doi:10.1038/nbt.1823)
73. Deck J, Gaither MR, Ewing R, Bird CE, Davies N, Meyer C, Riginos C, Toonen RJ, Crandall ED. 2017 The Genomic Observatories Metadatabase (GeOMe): A new repository for field and sampling event metadata associated with genetic samples. *PLoS Biol.* **15**, e2002925. (doi:10.1371/journal.pbio.2002925)
74. Hering D *et al.* 2018 Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. *Water Res.* **138**, 192–205. (doi:10.1016/j.watres.2018.03.003)
75. Shea MM, Kuppermann J, Rogers MP, Smith DS, Edwards P, Boehm AB. 2023 Systematic review of marine environmental DNA metabarcoding studies: toward best practices for data usability and accessibility. *PeerJ* **11**, e14993. (doi:10.7717/peerj.14993)
76. Davies N *et al.* 2021 Internet of Samples (iSamples): toward an interdisciplinary cyberinfrastructure for material samples. *GigaScience* **10**, giab028. (doi:10.1093/gigascience/giab028)
77. Kimble M, Allers S, Campbell K, Chen C, Jackson LM, King BL, Silverbrand S, York G, Beard K. 2022 medna-metadata: an open-source data management system for tracking environmental DNA samples and metadata. *Bioinformatics* **38**, 4589–4597. (doi:10.1093/bioinformatics/btac556)
78. Creedy TJ *et al.* 2022 Coming of age for COI metabarcoding of whole organism community DNA: towards bioinformatic harmonisation. *Mol. Ecol. Resour.* **22**, 847–861. (doi:10.1111/1755-0998.13502)
79. Hakimzadeh A *et al.* In press. A pile of pipelines: an overview of the bioinformatics software for metabarcoding data analyses. *Mol. Ecol. Resour*. (doi:10.1111/1755-0998.13847)
80. Iwaszkiewicz-Eggebrecht E, Zizka V, Lynggaard C. 2024 Three steps towards comparability and standardization among molecular methods for characterizing insect communities. protocols.io. (https://www.protocols.io/view/synthetic-coi-spike-ins-for-use-in-metabarcoding-b-14egn33ryl5d/v2)
81. Iwaszkiewicz-Eggebrecht E, Zizka V, Lynggaard C. 2024 Three steps towards comparability and standardization among molecular methods for characterizing insect communities. Figshare. (doi:10.6084/m9.figshare.c.7159013)