ARPHA Preprints

PENSOFT

PREPRINT

Author-formatted, not peer-reviewed document posted on 18/07/2023

DOI: https://doi.org/10.3897/arphapreprints.e109709

Abundance estimation with DNA metabarcoding – recent advancements for terrestrial arthropods

Wiebke Sickel, Vera Zizka, Alice Scherges, Sarah J. Bourlat, Petra Dieker

ARPHA Preprints *Author-formatted, not peer-reviewed document posted on 18/07/2023*. DOI: https://doi.org/10.3897/arphapreprints.e109709

Abundance estimation with DNA metabarcoding – recent advancements for terrestrial arthropods

Wiebke Sickel[a]*, Vera Zizka[b], Alice Scherges[b], Sarah J. Bourlat[b], Petra Dieker[a]

[a]Thünen Institute of Biodiversity, Bundesallee 65, 38116 Braunschweig, Germany

[b]Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig Bonn, Adenauerallee 127, 53113 Bonn, Germany

*__Corresponding Author:__ Wiebke Sickel (wiebke.sickel@thuenen.de)

Running Title: Abundance estimation with DNA metabarcoding

Word count: 4722

Abstract: 172

Main body text (without references): 4550

Abstract

Biodiversity is declining at alarming rates worldwide and large-scale monitoring is urgently needed to understand changes and their drivers. While classical taxonomic identification of species is time and labor intensive, the combination with DNA-based methods could upscale monitoring activities to achieve larger spatial coverage and increased sampling effort. However, challenges remain for DNA-based methods when species counts and/or biomass estimates are required. Several methodological advancements exist to improve the potential of DNA metabarcoding for abundance analysis, which however need further evaluation. Here, we discuss laboratory, as well as some bioinformatic adjustments to DNA metabarcoding workflows regarding their potential to achieve species abundance estimation from arthropod community samples. Our review includes pre-laboratory processing methods such as specimen photography, laboratory methods such as the use of spike-in DNA as an internal standard and bioinformatic advancements like correction factors. We conclude that specimen photography coupled with DNA metabarcoding currently promises the greatest potential to achieve species counts and biomass estimates, but that approaches such as spike-ins and correction factors are promising methods to pursue further.

Key words: biodiversity monitoring, COI, abundance, metabarcoding, insects, spike-ins.

Introduction

Biodiversity is declining at alarming rates worldwide (Díaz et al. 2020). The startling observation of a 75% decline in flying insect biomass in German nature reserves over 25 years (Hallmann et al. 2017) triggered an earthquake in society and politics and raised awareness for arthropod declines, which have since been further documented (Lister and Garcia 2018; Seibold et al. 2019; Simmons et al. 2019; van Klink et al. 2020). Subsequently, numerous initiatives have been launched or reinforced on global to European and regional scales to assess arthropod diversity and also define guidelines for applied, large-scale biodiversity monitoring schemes (Seibold et al. 2019; Ronquist et al. 2020; Potts et al. 2021). Monitoring programs are frequently limited in spatial coverage and sampling effort due to the morpho-taxonomical analysis of specimens, which is costly and time-consuming (Yu et al. 2012) and further limited by taxonomic impediment (Fernandes et al. 2019; Watts et al. 2019; Darby et al. 2020). Thus, in order to meet the increased demand for arthropod diversity assessments, traditional morpho-taxonomy approaches need to be combined with other methods (Pawlowski et al. 2018; Compson et al. 2020).

DNA-based approaches are promising to overcome the above mentioned shortcomings in arthropod diversity surveys and monitoring (Porter and Hajibabaei (2018); Zinger et al. (2020); Box 1: Glossary; Supplementary file 1: Background information), where DNA metabarcoding in particular enables high sample throughput (Elbrecht and Steinke 2018; de Kerdrel et al. 2020), automation of laboratory and bioinformatic processes (Krehenwinkel et al. 2017a; Buchner et al. 2021; Buchner et al. 2023) and is widely accepted as a time- and cost-effective approach for large-scale biodiversity assessments

(Piper et al. 2019; Watts et al. 2019). Molecular methods further have the potential to resolve cryptic species (Sow et al. 2019) and intraspecific genetic diversity (Elbrecht et al. 2018) and open up the possibility to include degraded and non-invasively collected material, e.g. feces (Andriollo et al. 2019), or plant material in biodiversity surveys, which yields high potential for trophic interaction and food web analysis.

However, implementation in policy-mandated monitoring programs is still hampered (Altermatt 2021; Meissner 2021). Reasons for the limited application include general skepticism among taxonomists, missing expertise and infrastructure within state monitoring agencies, a lack of standardized molecular protocols (Dickie et al. 2018; Pawlowski et al. 2018; Zinger et al. 2019; Compson et al. 2020; Creedy et al. 2021), as well as incomplete reference databases (Watts et al. 2019; van der Heyde et al. 2020; Zenker et al. 2020) and the destruction of specimens for DNA extraction (Zizka et al. 2019), although non-destructive approaches are gaining ground (Castalanelli et al. 2010; Carew et al. 2018; Zenker et al. 2020; Batovska et al. 2021; Kirse et al. 2023). The most important criticism concerns the limitation to assess species counts and biomass (here summarized into abundance; see Box 1: Glossary), which is essential in standardized monitoring and ecological analysis, but still remains one of the greatest challenges for high-throughput DNA-based approaches (Compson et al. 2020).

Several factors within the metabarcoding workflow affect extraction of abundance data (Pawlowski et al. 2018; Zinger et al. 2019). Firstly, sample properties such as complexity seem to affect abundance information (Piñol et al. 2019), including e.g. variation in sample biomass as well as across and within species (Elbrecht and Leese 2015; Elbrecht et al. 2017; Braukmann et al. 2019). Read numbers are further affected by variations in

marker gene copy numbers (Krehenwinkel et al. 2017b). Secondly, methodological biases skew abundance estimations. During DNA extraction, a protocol-dependent taxonomic bias can be introduced due to variations in species size and morphology, causing differences in isolated DNA yields (Krehenwinkel et al. 2017a; Pornon et al. 2017; Matos-Maraví et al. 2019; Iwaszkiewicz-Eggebrecht et al. 2022). Several subsampling steps in the metabarcoding laboratory workflow can introduce stochastic processes affecting read counts (Leray and Knowlton 2017; Shirazi et al. 2021; Zizka et al. 2022). Arguably, the strongest bias is caused by taxon-specific differences in primer binding efficiency (Piñol et al. 2015; Krehenwinkel et al. 2017a; Krehenwinkel et al. 2017b; Pawlowski et al. 2018). The magnitude of primer bias depends on the number of mismatches between primer and target sequence, especially towards the 3‘-end of the primer (Piñol et al. 2019). Apart from primer choice, additional PCR bias can be caused by variable GC content in the target genetic marker (Nichols et al. 2018), amplicon length (Krehenwinkel et al. 2017b) or the occurrence of pseudogenes (Andujar et al. 2021). Thirdly, post-laboratory steps in the bioinformatic processing of sequencing data can skew final read distribution (Frøslev et al. 2017; Alberdi et al. 2018; Matos-Maraví et al. 2019; Darby et al. 2020; Creedy et al. 2021).

A meta-analysis targeting 22 DNA metabarcoding studies revealed a weak relationship between biomass and generated read counts, with a large degree of uncertainty (Lamb et al. 2019). The studies included in Lamb et al. (2019) used different protocols and a wide range of target organisms and sample types, which somewhat hampers overall comparability, but does emphasize that raw read counts are not suitable to infer abundance estimates.

A variety of different approaches have emerged recently, that can help improve abundance estimates from metabarcoding data, including correction factors, spike-ins, primer optimization or multi-locus metabarcoding (e.g. Richardson et al. (2015); Krehenwinkel et al. (2017b); Richardson et al. (2019); Darby et al. (2020); Luo et al. (2022); also see Box 1: Glossary; Supplementary file 1: Background information). However, these advances have so far not been compared systematically for complex arthropod samples.

Here, we review potential methods that can improve abundance estimation in arthropod whole organism community (woc) samples. Considering the variety of approaches and applications, we aim to formulate general recommendations for DNA metabarcoding workflows in arthropod monitoring. In addition, we explore approaches from metabarcoding studies targeting e.g. aquatic samples and have so far not been applied to terrestrial arthropods and their trophic interactions.

Methods

Collection of relevant literature and assessment of methodological approaches

We collected relevant literature in two steps: First, we performed an online literature search in Google Scholar and EBSCO Discovery Service on 20 April 2020, which was repeated on 17 January 2022. We used the keywords [(quant*) AND (insect) AND (metabarcod*) AND (DNA)] and included only peer reviewed publications in English. Although the search term specifically targeted insects, we use the more general term ‘terrestrial arthropods’ throughout the text. Secondly, some publications were added to the list based on the authors’ expertise.

We included studies that applied DNA metabarcoding to terrestrial arthropods as target organisms and/or in relation to their tropic interactions within ecosystems (e.g. pollination and food web studies), as these topics are strongly connected and play an important role in monitoring schemes (e.g. ecosystem services of pollination or natural pest control). With these criteria, woc and tissue samples were included covering also pollen, gut contents and feces as well eDNA metabarcoding approaches, such as extraction from soil and sample fixative. We excluded studies that applied individual-based DNA barcoding and NGS barcoding, PCR-free approaches as well as long-read sequencing methodologies, as we wanted to focus on metabarcoding specifically. PCR-free approaches are, however, briefly discussed in an outlook section.

Based on 113 publications matching our search criteria (Supplementary file 2: Literature review), we extracted information on article type, study type, sample type, species group, methods and parameters (Table 1; Supplementary file 3: Categories assessed in the literature review). We examined these methods regarding their applicability to study types (species richness assessments, pollen analysis, food web studies) and to sample types (woc samples, pollen, eDNA and gut contents/feces). The overall suitability was assessed based on whether certain abundance metrics (species counts, relative abundance, biomass) were achievable, whilst also considering the extent of additional equipment, cost and labor (Supplementary file 4: Evaluation of methodological approaches). These considerations are based on the available literature.

Results

Description and assessment of methods

Reviewing the literature, we identified three main methods to estimate species abundance with metabarcoding (Table 1): (i) semi-quantitative metrics (Fig. 1), (ii) approaches that can potentially reduce read abundance biases (Fig. 2), and (iii) the combination of DNA (meta-)barcoding with other methodological approaches (Fig. 3), which we present in more detail in the following sections (also see Supplementary file 1: Background information). This review focuses on studies including developments associated with the laboratory workflow. For a critical assessment of missing standards in bioinformatics we refer to (Creedy et al. 2021). Since many metabarcoding studies refer to relative abundances, whilst monitoring aims to determine species counts, we make a clear distinction of these terms throughout this manuscript by referring to ‘species counts‘ (absolute number of individuals belonging to the same species), ’relative abundance‘ (proportion of a species within a sample) and ’biomass‘ (weight of individuals belonging to the same species), respectively. We use the more general term ’abundance‘ as a summary term for the three (see Box 1: Glossary).

1) Semi-quantitative metrics

DNA metabarcoding is comprehensively used to assess presence/absence from complex sample mixtures. While some ecological questions, including biodiversity measures, can be answered with presence/absence data (e.g. alpha diversity), abundance information is needed for complex network analysis as e.g. food web structures or plant-pollinator

interactions. There are different approaches to conduct semi-quantitative analysis of DNA metabarcoding data (Fig. 1A). In diet analyses, frequency or percentage of occurrence (FOO/POO; Fig. 1B) are often applied (Deagle et al. (2019), but see Cuff et al. (2022)). In bipartite networks, link strength (Fig. 1C) is a meaningful quantitative metric for plant-pollinator or prey-predator networks (Thomsen and Sigsgaard 2019; Cuff et al. 2022). Alternatively, relative read abundance (RRA) summarized over biological replicates is often used (Fig. 1D), especially for pollen samples (Kratschmer et al. 2019; Wilson et al. 2021). Read counts, RRA as well as derived metrics, such as log- or rank-transformed or rarefied read abundance are applied to obtain community composition for different sample types (pollen, feces, gut and woc samples; Hope et al. (2014); Hawkins et al. (2015); Richardson et al. (2015); Krehenwinkel et al. (2018); Macías-Hernández et al. (2018); Marquina et al. (2019)). The use of any of the above-mentioned metrics to assess abundances is easy to achieve from metabarcoding data, but species counts or biomass cannot be assessed.

2) Approaches that reduce read bias

Correction factors. Mock community experiments have shown a positive correlation of read counts per species with genomic template DNA concentration in pollen and woc samples (Gueuning et al. 2019; Baksay et al. 2020), while other studies revealed a PCR bias introduced through taxon- and marker-specific primer efficiency (Krehenwinkel et al. 2017b; Bell et al. 2019; Braukmann et al. 2019; Darby et al. 2020). Since these biases are strongly affected by primer binding efficiencies (Piñol et al. 2019), they are assumed to be predictable (Krehenwinkel et al. 2017b). Thus, correcting read counts using species-

specific correction factors can improve metabarcoding-derived abundance estimates (Krehenwinkel et al. (2017b); Darby et al. (2020); Fig. 2A and 2B). Such correction factors can be obtained using mock communities (Krehenwinkel et al. (2017b), Fig. 2A) or iterative algorithms (Darby et al. (2013); Darby et al. (2020), Table 1, Fig. 2B, Supplementary file 1: Background information). In order to derive correction factors using mock communities, artificial community samples of defined composition are processed alongside unknown samples. However, the derived correction factors can only be applied to species that are present both in environmental and artificial community samples, which is a strong limitation for hyperdiverse woc arthropod samples such as Malaise trap catches that contain many unknown taxa. Phylogenetic inference methods, as described for microbial analyses (Goberna and Verdu 2016; McLaren et al. 2019), are a possibility to extend correction factors to closely-related taxa, but this remains to be tested for arthropods, pollen and gut/fecal samples. To obtain correction factors, mock communities have so far only been used in combination with woc arthropod samples (Krehenwinkel et al. 2017b), but this approach could also be transferred to pollen samples, as processing mock communities alongside such samples is common (Bell et al. 2019; Baksay et al. 2020; Swenson and Gemeinholzer 2021). Species-specific correction factors obtained from mock community samples are helpful to reduce read abundance biases, however, some sources of bias still exist, e.g. related to the evenness of a community sample (Piñol et al. 2019), copy number variations of the target gene (Krehenwinkel et al. 2017b) or differences in DNA quality between specimens used for mock community samples versus field collected samples (Krehenwinkel et al. 2018).

Correction factors can also be calculated using an iterative algorithm which mitigates data skews due to copy number variations of the target gene (Darby et al. (2013; 2020); Fig. 2B, Supplementary file 1: Background information). It uses randomly generated correction factors for each species and compares predicted specimen counts with counts obtained from morphological identifications. The correction factors are then iteratively adjusted until predicted and actual counts converge (Darby et al. 2013).

The algorithm can only be applied to samples with high concordance between morphological and DNA-based taxonomies, but it is a promising approach, as the predicted species counts were highly correlated with actual count data (Darby et al. 2013; Darby et al. 2020). It requires high-quality material and specimens to be identified morphologically (Darby et al. 2013; Darby et al. 2020) and thus can only be used for woc and tissue samples (see Supplementary file 1: Background information for more limitations). Time and cost of the overall analysis increases, as a reference set of morphologically identified species is required, but this could be worth it in the case of repeat monitoring of sites with known species composition, or for the monitoring of known arthropod pests.

Spike-ins. Spike-ins (Fig. 2C) may also be referred to as internal standards (ISDs; Harrison et al. (2021)). Here, a defined amount of reference material DNA is added to each sample, which allows read count correction and thereby improves abundance estimation (Luo et al. 2022). The reference DNA can be added as tissue (Darby et al. 2020), genomic DNA, pre-amplified DNA (Ji et al. 2020), plasmids (Luo et al. 2022) or synthetic DNA (Palmer et al. 2018). Synthetic DNA fragments should include primer

binding sites and need to be added to the reference database (Tkacz et al. 2018; Luo et al. 2022). Spike-ins are added to the samples in a standardized manner, e.g. a defined mass of reference DNA (ng) per defined volume of lysis buffer (µl; Ji et al. (2020); Luo et al. (2022)). It is recommended to add the spike-in prior to DNA extraction, so that it is co-extracted, co-amplified and co-sequenced along with the sample DNA and therefore underlies the same methodological biases. Since all samples receive the same amount of spike-in, they should theoretically return the same spike-in read counts. However, sample complexity affects read numbers (Piñol et al. 2019) and thus different samples will return different read numbers for the spike-in (Luo et al. 2022). Read correction can be achieved by dividing the number of reads assigned to amplicon sequence variants (ASVs) by the number of reads assigned to the spike-in, resulting in significant improvement in within-species abundance across samples (Ji et al. 2020; Luo et al. 2022).

The use of spike-ins is not restricted by sample type, but comes with a low increase in effort and costs, because the spiking of samples is an additional, albeit minimal, step in the laboratory workflow, which has to then be integrated in the bioinformatic workflow. It should be noted that spike-in correction does not correct for biases across species within samples. It has been proposed that species-specific correction factors obtained from mock communities (see previous section) or unique molecular identifiers (UMIs, see outlook section) can be used to correct for within-sample across species biases (Ji et al. 2020; Luo et al. 2022). Spike-in correction is a straightforward and powerful approach with high potential to improve abundance estimations via DNA metabarcoding.

Primer optimization. A variety of studies have shown that primer design is an essential part determining the success of DNA metabarcoding studies, both in terms of taxon recovery and read abundance biases (Esnaola et al. 2018; Jusino et al. 2018; Lafage et al. 2019; Pedro et al. 2020). Primers used in DNA metabarcoding need to be universal and the fragment length needs to be suitable for the sequencing platform of choice, whilst allowing for species-level identification (Meusnier et al. 2008). Over and under amplification of different lineages of arthropods (Krehenwinkel et al. 2017b; Darby et al. 2020) as well as certain plant species dominating pollen samples (Bell et al. 2019; Baksay et al. 2020) have been reported and should be minimized as much as possible. Thus, primer design, including validation and evaluation, is a labor and time consuming and ongoing task (Elbrecht et al. 2019).

Multi-locus metabarcoding. Different genetic markers suffer from different taxonomic biases and thus some studies employ several different loci for the same organismal group, which is referred to as multi-marker (Adamowicz et al. 2019) or multi-locus metabarcoding (Batovska et al. 2021). Multi-locus metabarcoding has been applied to woc and tissue samples (Marquina et al. 2019; Giebner et al. 2020), pollen (Richardson et al. 2015; Bell et al. 2019; Richardson et al. 2019), fecal samples and gut contents (Swift et al. 2018; Krehenwinkel et al. 2019; Gil et al. 2020) as well as soil and even eDNA samples (Ritter et al. 2019; Thomsen and Sigsgaard 2019).

Locus-specific biases can be mitigated by using rank order abundance or median-based proportional abundance summarized over all loci, as has been demonstrated in pollen DNA metabarcoding (Richardson et al. 2015; Richardson et al. 2019). The locus-specific

PCRs are often performed separately (Richardson et al. 2015; Swift et al. 2018; Richardson et al. 2019; Baksay et al. 2020; Darby et al. 2020), which increases time and cost for sample processing. Multi-locus metabarcoding can be performed in multiplexed reactions (de Kerdrel et al. 2020; Batovska et al. 2021) to improve time and cost efficiency. However, this may introduce additional read abundance skews, possibly due to PCR competition between loci (Batovska et al. 2021). During analysis, data from different markers need to be analyzed separately (Thomsen and Sigsgaard 2019), which increases time for analysis. It should be emphasized that different markers usually yield discordant taxon lists (Alberdi et al. 2018; da Silva et al. 2019), e.g. because of incomplete reference databases for markers other than *COI* (Andujar et al. 2018). Such discordant taxa lists allow a broader taxon coverage, but it also means that data from different markers are complementary (Kirse et al. 2021), complicating data analysis. For pollen samples, some evidence exists that chloroplast markers (e.g. *trnL*) are more suitable for assessing relative abundances than ribosomal markers (e.g. ITS2; Richardson et al. (2019); Baksay et al. (2020)), and these differences need to be carefully considered. In the case of discordant taxa lists, abundance estimates (e.g. rank-based) can only be determined for taxa identified by more than one marker (Richardson et al. 2015; Richardson et al. 2019).

3) Combining DNA metabarcoding with other methods

Some studies combine DNA metabarcoding with other methodologies (Fig. 3). One example is morphological analysis of gut content remains, pollen grains or arthropod specimens (Keller et al. 2015; Darby et al. 2020; Gil et al. 2020), others are weighing bulk

samples (Hausmann et al. 2020), using flow cytometry of pollen (Baksay et al. 2020) or other forms of PCR (Schneider et al. 2016; Tedersoo et al. 2019). In such cases, DNA metabarcoding may be used to obtain a comprehensive species list of the detected taxa, whilst abundance estimates (e.g. species counts, biomass, DNA copy number) are obtained with the other methodology. The choice of additional methodology determines the sample types that can be used, for example, combining metabarcoding with quantitative PCR (qPCR; Schneider et al. (2016)) or digital droplet PCR (ddPCR; Tedersoo et al. (2019)) can be performed on all sample types. For other methodologies, for example weighing, woc samples are required (Hausmann et al. 2020).

One noteworthy approach of method combination is the photographic documentation of specimens from woc samples before analyzing them with DNA metabarcoding. This combined approach enables individual counts, body size measurements and thereby biomass estimation (Gueuning et al. 2019). As specimens are handled individually (Wührl et al. 2022), the use of body parts for DNA extraction, instead of full specimens, is furthermore facilitated (Gueuning et al. 2019; Darby et al. 2020), keeping voucher specimens mostly intact. Specimen photography further allows documentation of specimens for future reference as well as incorporating a pre-sorting strategy (Elbrecht et al. 2020). Whilst handling of individual specimens is exceptionally time- and labor-intensive, automated solutions can improve time-efficiency (Ärje et al. 2020; Wührl et al. 2022). In combination with machine learning approaches, the automated screening of high-resolution pictures of arthropod woc samples for abundance estimation is emerging and would facilitate large-scale assessments, e.g. for monitoring schemes (Høye et al.

2021). While these approaches are still in development, the vision of completely automated protocols, incorporating image recognition before molecular sample processing, exists (Høye et al. 2021; Besson et al. 2022; van Klink et al. 2022; Wührl et al. 2022).

Discussion

General conclusions and recommendations for arthropod monitoring and related questions

The available literature has revealed that the majority of (terrestrial) arthropod DNA metabarcoding studies do not sufficiently address the potential to infer species counts and/or biomass estimates (Supplemental file 2). In terms of pollen analysis, research exists that discusses abundance estimation via DNA metabarcoding, but with inconsistent results (Keller et al. 2015; Kraaijeveld et al. 2015; Richardson et al. 2015; Bell et al. 2019; Richardson et al. 2019; Baksay et al. 2020). In contrast to this, DNA metabarcoding has received considerably more attention in the aquatic sector in recent years (Elbrecht and Leese 2015; Elbrecht et al. 2017; Beentjes et al. 2019; Hoshino et al. 2021). Existing policies, like the EU Water Framework Directive (Directive 2000/60/EC) and the Marine Strategy Framework Directive (MSFD; Directive 2008/56/EC), legally require routine monitoring of aquatic environments. As a consequence, standards for sampling, processing and reporting already exist (Haase et al. (2004), but see Birk et al. (2012)), as well as DNA-based indicators (Aylagas et al. 2014). Especially the DNAqua-Net COST Action (Leese et al. 2016; Leese et al. 2018) has published many advancements regarding the suitability and integration of (e)DNA metabarcoding in biomonitoring

(Pawlowski et al. 2018; Buchner et al. 2019), as well as resources to facilitate standardization and quality control for DNA-based monitoring (Bruce et al. (2021); Bruce and Keskin (2021); Vasselon et al. (2021), DNAqua-Hub; accessed 24 May 2022). This work has a high potential to be transferred into terrestrial arthropod monitoring. However, this transfer could be hampered by the lack of data on diversity and distributions of hyperdiverse arthropods.

Additionally, the collected literature focused on approaches that apply to the sample processing stage of metabarcoding workflows. The effect of bioinformatics and data analysis strategies on abundance and biomass estimations is strongly underrepresented (Supplemental File 2: Literature collection). Although a variety of non-harmonized bioinformatic tools and pipelines exists (Creedy et al. 2021), a more detailed discussion on the bioinformatics and data analysis side of this topic, however, is outside the scope of this review, but future research needs to address this.

As expected, there is a variety of adjustments attempting to improve abundance estimation via DNA metabarcoding (Supplemental file 2: Literature collection). It remains, however, difficult to find a ‘one-size-fits-all’ approach to assessing individual counts and biomass from DNA metabarcoding, partly because different approaches are applicable only to certain sample types or because recent advancements still do not translate to individual counts and/or biomass estimates.

Overall suitability of DNA metabarcoding approaches for species counts and biomass estimation for terrestrial arthropod monitoring

373 Currently, the most promising approach is to combine DNA metabarcoding with specimen
374 photography, which would ideally be automated (Ärje et al. 2020; Wührl et al. 2022). In
375 addition, promising avenues such as correction factors and spike-ins should be further
376 developed (Darby et al. 2013; Krehenwinkel et al. 2017b; Darby et al. 2020; Ji et al. 2020;
377 Luo et al. 2022). Specimen photography coupled with automatic image recognition
378 facilitates body size measurements to achieve biomass estimates as well as species
379 counts. Combining the approaches of Darby et al. (2020), Gueuning et al. (2019) and de
380 Kerdrel et al. (2020) seems especially promising, as recombining specimens to 'pseudo-
381 community' samples allows cost-efficient mixed-species DNA (meta-)barcoding. We
382 would like to point out that this strategy is not the same as NGS barcoding (see Box 1;
383 Wang et al. (2018); Srivathsan et al. (2021)), since individual specimens or parts of them
384 are combined to mixed-species samples (Gueuning et al. 2019; de Kerdrel et al. 2020).
385 Thus, samples are processed following a metabarcoding workflow, but obtained barcodes
386 can be traced back to specimens (de Kerdrel et al. 2020). We argue that despite the
387 increase in processing time and associated costs, (automated) specimen photography is
388 a simple and effective way to achieve considerable improvement in taxon recovery,
389 species counts and biomass estimates (Fig. 4). This approach is limited to woc samples,
390 although a similar approach may potentially be applied to pollen samples, for example by
391 flow cytometry (Baksay et al. 2020; Dunker et al. 2020). Theoretically, these approaches
392 could be combined to achieve count and biomass data, although the above-mentioned
393 studies did not comment on this potential.
394

Regardless of application or sample type, general recommendations for every metabarcoding workflow are to use appropriate positive controls, i.e. mock communities (Ji et al. 2020), as well as negative controls, biological and technical replicates (Alberdi et al. 2018; Elbrecht and Steinke 2018; Liu et al. 2019; Zinger et al. 2019; Yang et al. 2020) and consider multi-locus metabarcoding. Each of these steps can improve taxon detection and the correlation between relative read abundances and input DNA mass (Richardson et al. 2019; Ritter et al. 2019; Thomsen and Sigsgaard 2019; Ji et al. 2020). Associated increases in costs and labor are justified by the improvement in the generated data, although budget limitations may deem technical replicates unfeasible. With optimized metabarcoding and bioinformatic workflows, more robust relative abundance and biomass estimates are thus potentially achievable in the foreseeable future. However, species counts cannot be obtained, as other sources of bias still exist. We therefore recommend considering additional approaches discussed further down.

For eDNA, obtaining count data is extremely difficult. Since eDNA dynamics (Barnes and Turner 2016; Compson et al. 2020) are affected by various uncontrollable factors prior to sampling, analysis of abundance information is further impeded. Thus, presence/absence and derived FOO/POO data from replicates currently seem to be the best option, although promising approaches exist that will move towards more informative data obtainable from eDNA. For example, combining species detections with information about the cellular and molecular state of eDNA (e.g. intra- versus extra-cellular eDNA, genetic region, fragment size) is expected to improve the abundance estimation, as demonstrated in water samples (Jo et al. 2021). Other options for eDNA-based monitoring are: an overall

418 experimental design and sampling strategy that allows indirect counts, developing and
419 applying novel metrics (e.g. the ‘eDNA index’; Kelly et al. (2019)) or coupling
420 presence/absence data with site-occupancy models (van Strien et al. 2010; van Strien et
421 al. 2013). We argue that eDNA approaches are worth considering for arthropod
422 monitoring, as they are non-invasive (Andriollo et al. 2019; Thomsen and Sigsgaard 2019;
423 Pumkaeo et al. 2021; Roger et al. 2022), which is especially important for protected and
424 endangered species.
425 Additionally, (e)DNA-based analyses open up new avenues that move away from
426 traditional species counts or biomass estimates. On such avenue to pursue further is
427 more sensitive detection rates of parasitism and invasive species (Sow et al. 2019; Young
428 et al. 2021). Furthermore, genetic tools facilitate the analysis of intraspecific diversity
429 (Elbrecht et al. 2018; Arribas et al. 2021; Shum and Palumbi 2021; Weitemier et al. 2021),
430 which is greatly underappreciated in arthropod monitoring schemes. Thus, (e)DNA
431 metabarcoding deserves to be incorporated in such schemes at least as a complementary
432 approach to morpho-taxonomy.
433

434 *Outlook: Further molecular approaches for the estimation of species abundances*
435 In the following, we explore selected approaches from the wider literature that were not
436 within the scope of the present review. However, there is high potential for the
437 implementation in monitoring programs in the future. Novel data analysis pipelines are
438 constantly being developed and some focus on integrating uncertainties associated with
439 the dynamics of DNA in the environment (Barnes and Turner 2016; Compson et al. 2020).
440 One such example, a tracer model, has successfully been applied to estimate the

441 abundance of target fish species (Fukaya et al. 2020). Another example, an ‘eDNA index’,
442 which is a double-transformation of read-counts, holds potential to assess abundance
443 trends across time and space (Kelly et al. 2019). Additionally, species occupancy models
444 can detect false negatives (Compson et al. 2020) and Bayesian hierarchical models can
445 integrate primer choice and other parameters of the metabarcoding workflow (Doi et al.
446 2019; Compson et al. 2020), which would allow correcting read count-derived abundance
447 estimates. Lastly, the application of half-life corrections has successfully been applied to
448 infer relative frequencies of prey items based on metabarcoding data (Uiterwaal and
449 DeLong 2020).
450
451 When grouping sequencing reads as ASVs instead of molecular operational taxonomic
452 units, DNA metabarcoding can potentially deliver conservative species counts in the
453 sense of ‘minimum census estimates’, similar to those obtained from non-invasive
454 sampling of hair and feces (Frantz et al. 2004; Miotto et al. 2007). In this case, the
455 evolutionary rate of the chosen marker would have to be considered (Wang et al. 2016),
456 as it may affect the recovery of ASVs per species and consequently the obtained
457 minimum census estimates. Furthermore, the potential to recover signals of intraspecific
458 diversity via ASVs has recently been demonstrated (Elbrecht et al. 2018; Shum and
459 Palumbi 2021; Weitemier et al. 2021).
460 Another promising approach is to further refine the qSeq protocol (Hoshino and Inagaki
461 2017; Hoshino et al. 2021) and similar workflows employing unique molecular identifiers
462 (UMIs; Luo et al. (2022)). Here, a single-primer extension is included in the workflow
463 before performing PCR. During this step, each DNA fragment is labelled with a random

tag and the number of random tags per ASV can be used to accurately infer starting copy numbers of each recovered sequence in the original sample. This allows simultaneous species identification and inference of relative abundances from eDNA and woc samples (Hoshino and Inagaki 2017; Hoshino et al. 2021; Luo et al. 2022). Unique molecular identifiers have also been applied in detecting rare allele variants and mutations and have been reported as being especially useful for read error corrections (Jabara et al. 2011; Kinde et al. 2011; Kivioja et al. 2012; Fields et al. 2020).

There is an urgent need to shift away from a purely morpho-taxonomic approach and related indicators for long-term arthropod monitoring, towards an integrative framework, in which morphological and molecular biological methodologies are applied in parallel. This requires the development and implementation of novel proxies and indicators to indirectly assess species abundance based on genetic data. One possible approach is to apply Hill numbers to DNA-based and morpho-taxonomic assessments alike, as this improves comparability and they can even be applied to (phylo-)genetic data (Alberdi and Gilbert 2019). Since, biomass and DNA mass of a taxon are correlated (Elbrecht et al. 2017), the amount of genomic DNA per taxon could serve as a proxy for abundance. This could be assessed by combining metabarcoding with qPCR or ddPCR, although these usually focus on specific target species (Schneider et al. 2016; Tedersoo et al. 2019), but also via the use UMIs (see above).

PCR-free methods represent a further alternative (Garrido-Sanz et al. 2020; Ji et al. 2020; Cordier et al. 2021). The advantage of these approaches is that no amplification step is

conducted and therefore, the complete mitochondrial or nuclear DNA is sequenced and analyzed. As PCR amplification is omitted, mito- and metagenomic approaches are associated with more reliable abundance estimations. However, both approaches depend on the accessibility of whole mitochondrial or nuclear genomes of target taxa in order to assign generated reads to the species of origin (Schmidt et al. 2022). So far, whole-genome reference databases exist only for a limited number of species (Formenti et al. 2022; Lewin et al. 2022). In addition, sequencing effort for mito- or metagenomic approaches is much higher than in metabarcoding studies, limiting the current application of those approaches in large-scale arthropod monitoring through higher costs, computing power and data storage requirements. Further, bias introduced through extraction and variable gene copy number still exist in those approaches.

Concluding remarks

Even though there are many details to consider when applying DNA metabarcoding to arthropod monitoring, pollen and food web analyses, we were able to make some general recommendations. Generally, DNA metabarcoding should always be optimized for maximum taxon recovery and minimal amplification biases. The processing of adequate positive and negative controls is essential. Incorporating appropriate biological and technical replicates reduces the impact of certain methodological biases.

Methodologies employed need to be time and cost efficient and ensure as little disturbance to ecosystems as possible despite regular sampling. However, current arthropod monitoring efforts rely on morpho-taxonomy, which is time-consuming and requires adequate taxonomic expertise. As a result, monitoring schemes are patchy and

largely limited. DNA metabarcoding as a rapid tool to obtain species occurrences is a very promising method for large-scale monitoring activities, especially when species counts are not required. When combining DNA metabarcoding with specimen photography and body size measurements, species counts and biomass can also be assessed.

Going forward, creating new DNA-based metrics to report (relative) abundances based on genetic units rather than processing individual specimens offers new innovations addressing the most central questions in arthropod monitoring, as these rarely require absolute measures of abundance. Detecting and assessing trends in monitoring relates more to within- and between-sample comparisons taken across spatial and temporal scales. This has already successfully been done with metabarcoding (Hope et al. 2014; Danner et al. 2017; Baroja et al. 2019; Moran et al. 2019; Steinke et al. 2022). Additionally, DNA metabarcoding facilitates the assessment of ecosystem services in a time- and cost-efficient manner, via processing pollen and food web analyses.

There are still many challenges to face until metabarcoding data can deliver robust abundance estimations. Currently, sorting and individual handling of specimens from woc samples is unavoidable to obtain such data. However, it is important to apply both classical morpho-taxonomy and molecular biological approaches in parallel, which will allow the management and analysis of the large amounts of data generated by monitoring programs in a timely and cost-effective manner. Thus, despite its limitations, DNA metabarcoding can and should be incorporated as an additional tool in routine arthropod monitoring to increase sample sizes and cover a broader range of taxonomic groups.

Abbreviations

ASV – amplicon sequence variant

ddPCR – digital droplet PCR

eDNA – environmental DNA

FOO / POO – frequency of occurrence / percent of occurrence

NGS / HTS – next generation sequencing / high-throughput sequencing

qPCR – quantitative PCR

RRA – relative read abundance

UMI – unique molecular identifier

woc samples – whole organism community samples

Author contributions

WS and PD devised the study. WS performed the literature review and drafted the first version of the manuscript. WS, PD, VZ and SJB were substantially involved in subsequent drafts. AS created the figures. All authors agreed to the final version of the manuscript.

Declaration of competing interest

None.

Acknowledgements

The presented study is part of the joint project *Monitoring of biodiversity in agricultural landscapes* (MonViA) that has been funded by the German Federal Ministry of Food and Agriculture.

Supplementary files

Supplementary file 1: Background information. Background information on DNA metabarcoding and methodological adjustments to improve abundance estimates

Supplementary file 2: Literature collection. Relevant publications were identified in two steps: 1) via an online literature search in Google Scholar and EBSCO Discovery Service; keywords: ((quant*) AND (insect) AND (metabarcod*) AND (DNA)), including only peer reviewed publications in English, results were screened for suitability based on title and abstract; 2) addition of publications based on the authors' expertise; included studies that applied metabarcoding with arthropods and/or in relation to their trophic interactions within ecosystems (e.g. pollination, food web studies); excluded topics were: individual-based DNA barcoding and NGS barcoding, long-read sequencing methodology, mito-/metagenomics and genome skimming.

Supplementary file 3: Categories assessed in the literature review. For each category, possible parameters are given, together with examples and more detailed explanation where appropriate and necessary

Supplementary file 4: Evaluation of methodological approaches

References

Adamowicz SJ, Boatwright JS, Chain F, Fisher BL, Hogg ID, Leese F, Lijtmaer DA, Mwale M, Naaum AM, Pochon X, Steinke D, Wilson JJ, Wood S, Xu J, Xu S, Zhou X, van der Bank M (2019) Trends in DNA barcoding and metabarcoding. Genome 62: v-viii. doi:10.1139/gen-2019-0054.

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

Alberdi A, Gilbert MTP (2019) A guide to the application of Hill numbers to DNA-based diversity analyses. Mol Ecol Resour 19: 804-817. doi:10.1111/1755-0998.13014.

Altermatt F (2021) The challenges and opportunities for implementation of eDNA biomonitoring in riverine systems. ARPHA Conference Abstracts 4: doi:10.3897/aca.4.e65521.

Andriollo T, Gillet F, Michaux JR, Ruedi M (2019) The menu varies with metabarcoding practices: A case study with the bat Plecotus auritus. PLOS ONE 14: e0219135-e0219135. doi:10.1371/journal.pone.0219135.

Andujar 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.

Andujar C, Creedy TJ, Arribas P, Lopez H, Salces-Castellano A, Perez-Delgado AJ, Vogler AP, Emerson BC (2021) Validated removal of nuclear pseudogenes and sequencing artefacts from mitochondrial metabarcode data. Mol Ecol Resour 21: 1772-1787. doi:10.1111/1755-0998.13337.

Ärje J, Melvad C, Jeppesen MR, Madsen SA, Raitoharju J, Rasmussen MS, Iosifidis A, Tirronen V, Gabbouj M, Meissner K, Høye TT, Kotze DJ (2020) Automatic image-based identification and biomass estimation of invertebrates. Methods in Ecology and Evolution 11: 922-931. doi:10.1111/2041-210x.13428.

Arribas P, Andujar C, Salces-Castellano A, Emerson BC, Vogler AP (2021) The limited spatial scale of dispersal in soil arthropods revealed with whole-community haplotype-level metabarcoding. Mol Ecol 30: 48-61. doi:10.1111/mec.15591.

Aylagas E, Borja A, Rodriguez-Ezpeleta N (2014) Environmental status assessment using DNA metabarcoding: towards a genetics based Marine Biotic Index (gAMBI). PLOS ONE 9: e90529. doi:10.1371/journal.pone.0090529.

Baksay S, Pornon A, Burrus M, Mariette J, Andalo C, Escaravage N (2020) Experimental quantification of pollen with DNA metabarcoding using ITS1 and trnL. Scientific Reports 10: 4202-4202. doi:10.1038/s41598-020-61198-6.

Barnes MA, Turner CR (2016) The ecology of environmental DNA and implications for conservation genetics. Conservation Genetics 17: 1-17. doi:10.1007/s10592-015-0775-4.

Baroja U, Garin I, Aihartza J, Arrizabalaga-Escudero A, Vallejo N, Aldasoro M, Goiti U (2019) Pest consumption in a vineyard system by the lesser horseshoe bat (Rhinolophus hipposideros). PLOS ONE 14: e0219265-e0219265. doi:10.1371/journal.pone.0219265.

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.

Beentjes KK, Speksnijder AGCL, Schilthuizen M, Hoogeveen M, Pastoor R, van der Hoorn BB (2019) Increased performance of DNA metabarcoding of macroinvertebrates by taxonomic sorting. PLOS ONE 14: e0226527-e0226527. doi:10.1371/journal.pone.0226527.

Bell KL, Burgess KS, Botsch JC, Dobbs EK, Read TD, Brosi BJ (2019) Quantitative and qualitative assessment of pollen DNA metabarcoding using constructed species mixtures. Molecular Ecology 28: 431-455. doi:10.1111/mec.14840.

Besson M, Alison J, Bjerge K, Gorochowski TE, Hoye TT, Jucker T, Mann HMR, Clements CF (2022) Towards the fully automated monitoring of ecological communities. Ecol Lett 25: 2753-2775. doi:10.1111/ele.14123.

Birk S, Bonne W, Borja A, Brucet S, Courrat A, Poikane S, Solimini A, van de Bund W, Zampoukas N, Hering D (2012) Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. Ecological Indicators 18: 31-41. doi:10.1016/j.ecolind.2011.10.009.

Braukmann TWA, Ivanova NV, Prosser SWJ, Elbrecht V, Steinke D, Ratnasingham S, de Waard JR, Sones JE, Zakharov EV, Hebert PDN (2019) Metabarcoding a diverse arthropod mock community. Mol Ecol Resour 19: 711-727. doi:10.1111/1755-0998.13008.

Bruce K, Blackman R, Bourlat SJ, Hellström AM, Bakker J, Bista I, Bohmann K, Bouchez A, Brys R, Clark K, Elbrecht V, Fazi S, Fonseca V, Hänfling B, Leese F, Mächler E, Mahon AR, Meissner K, Panksep K, Pawlowski J, Schmidt Yáñez P, Seymour M, Thalinger B, Valentini A, Woodcock P, Traugott M, Vasselon V, Deiner K (2021) A practical guide to DNA-based methods for biodiversity assessment. Advanced Books: doi:https://doi.org/10.3897/ab.e68634.

Bruce K, Keskin E (2021) Field and Lab Protocols: Achievements of the DNAqua-Net Working Group 3. ARPHA Conference Abstracts 4: doi:10.3897/aca.4.e65306.

Buchner D, Beermann A, Hörren T, Enss J, Frenzel M, Li Y, Müller J, Pauls SU, Sorg M, Haase P, Leese F (2023) German-wide Malaise trap metabarcoding estimates over 33,000 insect species. BioRxiv: doi:10.1101/2023.05.04.539402.

Buchner D, Beermann AJ, Laini A, Rolauffs P, Vitecek S, Hering D, Leese F (2019) Analysis of 13,312 benthic invertebrate samples from German streams reveals minor deviations in ecological status class between abundance and presence/absence data. PLOS ONE 14: e0226547. doi:10.1371/journal.pone.0226547.

Buchner D, Macher TH, Beermann AJ, Werner MT, Leese F (2021) Standardized high-throughput biomonitoring using DNA metabarcoding: Strategies for the adoption of automated liquid handlers. Environ Sci Ecotechnol 8: 100122. doi:10.1016/j.ese.2021.100122.

Carew ME, Coleman RA, Hoffmann AA (2018) Can non-destructive DNA extraction of bulk invertebrate samples be used for metabarcoding? PeerJ 6: e4980. doi:10.7717/peerj.4980.

Castalanelli MA, Severtson DL, Brumley CJ, Szito A, Foottit RG, Grimm M, Munyard KG, David M. (2010) A rapid non-destructive DNA extraction method for insects and other arthropods. Journal of Asia-Pacific Entomology 13: 243-248. doi:10.1016/j.aspen.2010.04.003.

Compson ZG, McClenaghan B, Singer GAC, Fahner NA, Hajibabaei M (2020) Metabarcoding From Microbes to Mammals: Comprehensive Bioassessment on a Global Scale. Frontiers in Ecology and Evolution 8: doi:10.3389/fevo.2020.581835.

Cordier T, Alonso-Saez L, Apotheloz-Perret-Gentil L, Aylagas E, Bohan DA, Bouchez A, Chariton A, Creer S, Fruhe L, Keck F, Keeley N, Laroche O, Leese F, Pochon X, Stoeck T, Pawlowski J, Lanzen A (2021) Ecosystems monitoring powered by environmental genomics: A review of current strategies with an implementation roadmap. Mol Ecol 30: 2937-2958. doi:10.1111/mec.15472.

Creedy T, Andujar C, Meramveliotakis E, Noguerales V, Overcast I, Papadopoulou A, Morlon H, Vogler A, Emerson B, Arribas P (2021) Coming of age for COI metabarcoding of whole

organism community DNA: towards bioinformatic harmonisation. Authorea Preprints: under review. doi:10.22541/au.162141276.61766048/v1.

Cuff JP, Windsor FM, Tercel MPTG, Kitson JJN, Evans DM (2022) Overcoming the pitfalls of merging dietary metabarcoding into ecological networks. Methods in Ecology and Evolution: doi:10.1111/2041-210x.13796.

da Silva LP, Mata VA, Lopes PB, Pereira P, Jarman SN, Lopes RJ, Beja P (2019) Advancing the integration of multi-marker metabarcoding data in dietary analysis of trophic generalists. Mol Ecol Resour 19: 1420-1432. doi:10.1111/1755-0998.13060.

Danner N, Keller A, Härtel S, Steffan-Dewenter I (2017) Honey bee foraging ecology: Season but not landscape diversity shapes the amount and diversity of collected pollen. PLOS ONE 12: e0183716-e0183716. doi:10.1371/journal.pone.0183716.

Darby B, Bryant R, Keller A, Jochim M, Moe J, Schreiner Z, Pratt C, Euliss NH, Park M, Simmons R, Otto C (2020) Molecular sequencing and morphological identification reveal similar patterns in native bee communities across public and private grasslands of eastern North Dakota. PLOS ONE 15: e0227918-e0227918. doi:10.1371/journal.pone.0227918.

Darby BJ, Todd TC, Herman MA (2013) High-throughput amplicon sequencing of rRNA genes requires a copy number correction to accurately reflect the effects of management practices on soil nematode community structure. Mol Ecol 22: 5456-5471. doi:10.1111/mec.12480.

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. Scientific Reports 10: 78-78. doi:10.1038/s41598-019-54927-z.

Deagle BE, Thomas AC, McInnes JC, Clarke LJ, Vesterinen EJ, Clare EL, Kartzinel TR, Eveson JP (2019) Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? Molecular Ecology 28: 391-406. doi:10.1111/mec.14734.

Díaz S, Settele J, Brondízio E, Ngo HT, Guèze M, Agard J, Arneth A, Balvanera P, Brauman K, Butchart S, Chan K, Garibaldi L, Ichii K, Liu J, Subramanian SM, Midgley G, Miloslavich P, Molnár Z, Obura D, Pfaff A, Polasky S, Purvis A, Razzaque J, Reyers B, Chowdhury RR, Shin Y-J, Visseren-Hamakers I, Willis K, Zayas C (2020) Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. pp.

Dickie IA, Boyer S, Buckley HL, Duncan RP, Gardner PP, Hogg ID, Holdaway RJ, Lear G, Makiola A, Morales SE, Powell JR, Weaver L (2018) Towards robust and repeatable sampling methods in eDNA-based studies. Mol Ecol Resour: doi:10.1111/1755-0998.12907.

Doi H, Fukaya K, Oka SI, Sato K, Kondoh M, Miya M (2019) Evaluation of detection probabilities at the water-filtering and initial PCR steps in environmental DNA metabarcoding using a multispecies site occupancy model. Sci Rep 9: 3581. doi:10.1038/s41598-019-40233-1.

Dunker S, Motivans E, Rakosy D, Boho D, Mäder P, Hornick T, Knight TM (2020) Pollen analysis using multispectral imaging flow cytometry and deep learning. New Phytol 229: 593-606. doi:10.1111/nph.16882.

Elbrecht V, Bourlat SJ, Hörren T, Lindner A, Mordente A, Noll NW, Sorg M, Zizka VMA (2020) Pooling size sorted malaise trap fractions to maximise taxon recovery with metabarcoding. BioRxiv: doi:10.1101/2020.06.09.118950.

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-e7745. doi:10.7717/peerj.7745.

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: e0130324-e0130324. doi:10.1371/journal.pone.0130324.

Elbrecht V, Peinert B, Leese F (2017) Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. Ecology and Evolution 7: 6918-6926. doi:10.1002/ece3.3192.

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

Elbrecht V, Vamos EE, Steinke D, Leese F (2018) Estimating intraspecific genetic diversity from community DNA metabarcoding data. PeerJ 6: e4644. doi:10.7717/peerj.4644.

Esnaola A, Arrizabalaga-Escudero A, González-Esteban J, Elosegi A, Aihartza J (2018) Determining diet from faeces: Selection of metabarcoding primers for the insectivore Pyrenean desman (Galemys pyrenaicus). PLOS ONE 13: e0208986-e0208986. doi:10.1371/journal.pone.0208986.

Fernandes K, van der Heyde M, Coghlan M, Wardell-Johnson G, Bunce M, Harris R, Nevill P (2019) Invertebrate DNA metabarcoding reveals changes in communities across mine site restoration chronosequences. Restoration Ecology 27: 1177-1186. doi:10.1111/rec.12976.

Fields B, Moeskjær S, Friman V-P, Andersen SU, Young JPW (2020) MAUI-seq: Metabarcoding using amplicons with unique molecular identifiers to improve error correction. BioRxiv: doi:10.1101/538587.

Formenti G, Theissinger K, Fernandes C, Bista I, Bombarely A, Bleidorn C, Ciofi C, Crottini A, Godoy JA, Hoglund J, Malukiewicz J, Mouton A, Oomen RA, Paez S, Palsboll PJ, Pampoulie C, Ruiz-Lopez MJ, Svardal H, Theofanopoulou C, de Vries J, Waldvogel AM, Zhang G, Mazzoni CJ, Jarvis ED, Balint M, European Reference Genome Atlas C (2022) The era of reference genomes in conservation genomics. Trends Ecol Evol 37: 197-202. doi:10.1016/j.tree.2021.11.008.

Frantz A, Schaul M, Pope LC, Fack F, Schley L, Muller CP, Roper TM (2004) Estimating population size by genotyping remotely plucked hair: the Eurasian badger. Journal of Applied Ecology 41: 985-995. doi:10.1111/j.0021-8901.2004.00951.x.

Frøslev TG, Kjøller R, Bruun HH, Ejrnæs R, Brunbjerg AK, Pietroni C, Hansen AJ (2017) Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nature Communications 8: doi:10.1038/s41467-017-01312-x.

Fukaya K, Murakami H, Yoon S, Minami K, Osada Y, Yamamoto S, Masuda R, Kasai A, Miyashita K, Minamoto T, Kondoh M (2020) Estimating fish population abundance by integrating quantitative data on environmental DNA and hydrodynamic modelling. Mol Ecol: doi:10.1111/mec.15530.

Garrido-Sanz L, Senar MÀ, Piñol J (2020) Estimation of the relative abundance of species in artificial mixtures of insects using low-coverage shotgun metagenomics. Metabarcoding and Metagenomics 4: doi:10.3897/mbmg.4.48281.

Giebner H, Langen K, Bourlat SJ, Kukowka S, Mayer C, Astrin JJ, Misof B, Fonseca VG (2020) Comparing diversity levels in environmental samples: DNA sequence capture and metabarcoding approaches using 18S and COI genes. Mol Ecol Resour 20: 1333-1345. doi:10.1111/1755-0998.13201.

Gil V, Pinho CJ, Aguiar CAS, Jardim C, Rebelo R, Vasconcelos R (2020) Questioning the proverb 'more haste, less speed': classic versus metabarcoding approaches for the diet study of a remote island endemic gecko. PeerJ 8: e8084-e8084. doi:10.7717/peerj.8084.

Goberna M, Verdu M (2016) Predicting microbial traits with phylogenies. ISME J 10: 959-967. doi:10.1038/ismej.2015.171.

Gueuning M, Ganser D, Blaser S, Albrecht M, Knop E, Praz C, Frey JE (2019) Evaluating next-generation sequencing (NGS) methods for routine monitoring of wild bees: Metabarcoding, mitogenomics or NGS barcoding. Molecular Ecology Resources 19: 847-862. doi:10.1111/1755-0998.13013.

Haase P, Lohse SP, S., Schindehütte KS, A., Rolauffs P, Hering D (2004) Assessing streams in Germany with benthic invertebrates: development of a practical standardised protocol for macroinvertebrate sampling and sorting. Limnologica 34: 349-365.

Hallmann CA, Sorg M, Jongejans E, Siepel H, Hofland N, Schwan H, Stenmans W, Müller A, Sumser H, Hörren T, Goulson D, de Kroon H (2017) More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLOS ONE 12: e0185809. doi:10.1371/journal.pone.0185809.

Harrison JG, Randolph GD, Buerkle CA (2021) Characterizing microbiomes via sequencing of marker loci: techniques to improve throughput, account for crosscontamination, and reduce cost. MSystems 6: e00294-00221. doi:https://doi.org/10.1128/mSystems.00294-21.

Hausmann A, Segerer AH, Greifenstein T, Knubben J, Morinière J, Bozicevic V, Doczkal D, Günter A, Ulrich W, Habel JC (2020) Toward a standardized quantitative and qualitative insect monitoring scheme. Ecology and Evolution 10: 4009-4020. doi:10.1002/ece3.6166.

Hawkins J, de Vere N, Griffith A, Ford CR, Allainguillaume J, Hegarty MJ, Baillie L, Adams-Groom B (2015) Using DNA Metabarcoding to Identify the Floral Composition of Honey: A New Tool for Investigating Honey Bee Foraging Preferences. PLOS ONE 10: e0134735. doi:10.1371/journal.pone.0134735.

Hope PR, Bohmann K, Gilbert MTP, Zepeda-Mendoza ML, 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: 39-39. doi:10.1186/1742-9994-11-39.

Hoshino T, Inagaki F (2017) Application of Stochastic Labeling with Random-Sequence Barcodes for Simultaneous Quantification and Sequencing of Environmental 16S rRNA Genes. PLOS ONE 12: e0169431. doi:10.1371/journal.pone.0169431.

Hoshino T, Nakao R, Doi H, Minamoto T (2021) Simultaneous absolute quantification and sequencing of fish environmental DNA in a mesocosm by quantitative sequencing technique. Scientific Reports 11: 4372. doi:10.1038/s41598-021-83318-6.

Høye TT, Ärje J, Bjerge K, Hansen OLP, Iosifidis A, Leese F, Mann HMR, Meissner K, Melvad C, Raitoharju J (2021) Deep learning and computer vision will transform entomology. Proc Natl Acad Sci U S A 118: doi:10.1073/pnas.2002545117.

Iwaszkiewicz-Eggebrecht E, Granqvist E, Buczek M, Prus M, Roslin T, Tack AJM, Andersson AF, Miraldo A, Ronquist F, Łukasik P (2022) Optimizing insect metabarcoding using replicated mock communities. BioRxiv: doi:10.1101/2022.06.20.496906.

Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R (2011) Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci U S A 108: 20166-20171. doi:10.1073/pnas.1110064108.

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.

Jo T, Takao K, Minamoto T (2021) Linking the state of environmental DNA to its application for biomonitoring and stock assessment: Targeting mitochondrial/nuclear genes, and different DNA fragment lengths and particle sizes. Environmental DNA: doi:10.1002/edn3.253.

Jusino MA, Banik MT, Palmer JM, Wray AK, Xiao L, Pelton E, Barber JR, Kawahara AY, Gratton C, Peery MZ, Lindner DL (2018) An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. Molecular Ecology Resources 19: 176-190. doi:10.1111/1755-0998.12951.

Keller A, Danner N, Grimmer G, Ankenbrand M, von der Ohe K, von der Ohe W, Rost S, Härtel S, Steffan-Dewenter I (2015) Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. Plant Biology 17: 558-566. doi:10.1111/plb.12251.

Kelly RP, Shelton AO, Gallego R (2019) Understanding PCR processes to draw meaningful conclusions from environmental DNA studies. Scientific Reports 9: 12133-12133. doi:10.1038/s41598-019-48546-x.

Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B (2011) Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci U S A 108: 9530-9535. doi:10.1073/pnas.1105422108.

Kirse A, Bourlat SJ, Langen K, Fonseca VG (2021) Unearthing the Potential of Soil eDNA Metabarcoding—Towards Best Practice Advice for Invertebrate Biodiversity Assessment. Frontiers in Ecology and Evolution 9: doi:10.3389/fevo.2021.630560.

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.

Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, Taipale J (2012) Counting absolute numbers of molecules using unique molecular identifiers. Nature Methods 9: 72-74.

Kraaijeveld K, de Weger LA, Ventayol Garcia M, Buermans H, Frank J, Hiemstra PS, den Dunnen JT (2015) Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. Mol Ecol Resour 15: 8-16. doi:10.1111/1755-0998.12288.

Kratschmer S, Petrović B, Curto M, Meimberg H, Pachinger B (2019) Pollen availability for the Horned mason bee (Osmia cornuta) in regions of different land use and landscape structures. Ecological Entomology: doi:10.1111/een.12823.

Krehenwinkel H, Fong M, Kennedy S, Huang EG, Noriyuki S, Cayetano L, Gillespie R (2018) The effect of DNA degradation bias in passive sampling devices on metabarcoding studies of arthropod communities and their associated microbiota. PLOS ONE 13: e0189188-e0189188. doi:10.1371/journal.pone.0189188.

Krehenwinkel H, Kennedy S, Pekár S, Gillespie RG (2017a) A cost-efficient and simple protocol to enrich prey DNA from extractions of predatory arthropods for large-scale gut content analysis by Illumina sequencing. Methods in Ecology and Evolution 8: 126-134. doi:10.1111/2041-210X.12647.

Krehenwinkel H, Pomerantz A, Henderson JB, Kennedy SR, Lim JY, Swamy V, Shoobridge JD, Graham N, Patel NH, Gillespie RG, Prost S (2019) Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. GigaScience 8: 1-16. doi:10.1093/gigascience/giz006.

Krehenwinkel H, Wolf M, Lim JY, Rominger AJ, Simison WB, Gillespie RG (2017b) Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Scientific Reports 7: 17668-17668. doi:10.1038/s41598-017-17333-x.

Lafage D, Elbrecht V, Cuff JP, Steinke D, Hambäck PA, Erlandsson A (2019) A new primer for metabarcoding of spider gut contents. Environmental DNA 2: 234-243. doi:10.1002/edn3.62.

873 Lamb PD, Hunter E, Pinnegar JK, Creer S, Davies RG, Taylor MI (2019) How quantitative is
874 metabarcoding: A meta-analytical approach. Mol Ecol 28: 420-430.
875 doi:10.1111/mec.14920.
876 Leese F, Altermatt F, Bouchez A, Ekrem T, Hering D, Meissner K, Mergen P, Pawlowski J,
877 Piggott J, Rimet F, Steinke D, Taberlet P, Weigand A, Abarenkov K, Beja P, Bervoets L,
878 Björnsdóttir S, Boets P, Boggero A, Bones A, Borja Á, Bruce K, Bursić V, Carlsson J,
879 Čiampor F, Čiamporová-Zatovičová Z, Coissac E, Costa F, Costache M, Creer S,
880 Csabai Z, Deiner K, DelValls Á, Drakare S, Duarte S, Eleršek T, Fazi S, Fišer C, Flot J-
881 F, Fonseca V, Fontaneto D, Grabowski M, Graf W, Guðbrandsson J, Hellström M,
882 Hershkovitz Y, Hollingsworth P, Japoshvili B, Jones J, Kahlert M, Kalamujic Stroil B,
883 Kasapidis P, Kelly M, Kelly-Quinn M, Keskin E, Kõljalg U, Ljubešić Z, Maček I, Mächler
884 E, Mahon A, Marečková M, Mejdandzic M, Mircheva G, Montagna M, Moritz C, Mulk V,
885 Naumoski A, Navodaru I, Padisák J, Pálsson S, Panksep K, Penev L, Petrusek A,
886 Pfannkuchen M, Primmer C, Rinkevich B, Rotter A, Schmidt-Kloiber A, Segurado P,
887 Speksnijder A, Stoev P, Strand M, Šulčius S, Sundberg P, Traugott M, Tsigenopoulos C,
888 Turon X, Valentini A, van der Hoorn B, Várbíró G, Vasquez Hadjilyra M, Viguri J,
889 Vitonytė I, Vogler A, Vrålstad T, Wägele W, Wenne R, Winding A, Woodward G, Zegura
890 B, Zimmermann J (2016) DNAqua-Net: Developing new genetic tools for bioassessment
891 and monitoring of aquatic ecosystems in Europe. Research Ideas and Outcomes 2:
892 doi:10.3897/rio.2.e11321.
893 Leese F, Bouchez A, Abarenkov K, Altermatt F, Borja Á, Bruce K, Ekrem T, Čiampor F,
894 Čiamporová-Zaťovičová Z, Costa FO, Duarte S, Elbrecht V, Fontaneto D, Franc A,
895 Geiger MF, Hering D, Kahlert M, Kalamujić Stroil B, Kelly M, Keskin E, Liska I, Mergen
896 P, Meissner K, Pawlowski J, Penev L, Reyjol Y, Rotter A, Steinke D, van der Wal B,
897 Vitecek S, Zimmermann J, Weigand AM (2018) Why We Need Sustainable Networks
898 Bridging Countries, Disciplines, Cultures and Generations for Aquatic Biomonitoring 2.0:
899 A Perspective Derived From the DNAqua-Net COST Action. In: Next Generation
900 Biomonitoring: Part 1. 63-99. doi:10.1016/bs.aecr.2018.01.001.
901 Leray M, Knowlton N (2017) Random sampling causes the low reproducibility of rare eukaryotic
902 OTUs in Illumina COI metabarcoding. PeerJ 5: e3006-e3006. doi:10.7717/peerj.3006.
903 Lewin HA, Richards S, Lieberman Aiden E, Allende ML, Archibald JM, Balint M, Barker KB,
904 Baumgartner B, Belov K, Bertorelle G, Blaxter ML, Cai J, Caperello ND, Carlson K,
905 Castilla-Rubio JC, Chaw SM, Chen L, Childers AK, Coddington JA, Conde DA,
906 Corominas M, Crandall KA, Crawford AJ, DiPalma F, Durbin R, Ebenezer TE, Edwards
907 SV, Fedrigo O, Flicek P, Formenti G, Gibbs RA, Gilbert MTP, Goldstein MM, Graves JM,
908 Greely HT, Grigoriev IV, Hackett KJ, Hall N, Haussler D, Helgen KM, Hogg CJ, Isobe S,
909 Jakobsen KS, Janke A, Jarvis ED, Johnson WE, Jones SJM, Karlsson EK, Kersey PJ,
910 Kim JH, Kress WJ, Kuraku S, Lawniczak MKN, Leebens-Mack JH, Li X, Lindblad-Toh K,
911 Liu X, Lopez JV, Marques-Bonet T, Mazard S, Mazet JAK, Mazzoni CJ, Myers EW,
912 O'Neill RJ, Paez S, Park H, Robinson GE, Roquet C, Ryder OA, Sabir JSM, Shaffer HB,
913 Shank TM, Sherkow JS, Soltis PS, Tang B, Tedersoo L, Uliano-Silva M, Wang K, Wei X,
914 Wetzer R, Wilson JL, Xu X, Yang H, Yoder AD, Zhang G (2022) The Earth BioGenome
915 Project 2020: Starting the clock. Proc Natl Acad Sci U S A 119:
916 doi:10.1073/pnas.2115635118.
917 Lister BC, Garcia A (2018) Climate-driven declines in arthropod abundance restructure a
918 rainforest food web. Proc Natl Acad Sci U S A 115: E10397-E10406.
919 doi:10.1073/pnas.1722477115.
920 Liu M, Clarke LJ, Baker SC, Jordan GJ, Burridge CP (2019) A practical guide to DNA
921 metabarcoding for entomological ecologists. Ecological Entomology:
922 doi:10.1111/een.12831.

Luo M, Ji Y, Yu DW (2022) Extracting abundance information from DNA-based data. BioRxiv: doi:10.1101/2022.01.06.475221.

Macías-Hernández N, Athey K, Tonzo V, Wangensteen OS, Arnedo M, Harwood JD (2018) Molecular gut content analysis of different spider body parts. PLOS ONE 13: e0196589-e0196589. doi:10.1371/journal.pone.0196589.

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.

Matos-Maraví P, Ritter CD, Barnes CJ, Nielsen M, Olsson U, Wahlberg N, Marquina D, Sääksjärvi I, Antonelli A (2019) Biodiversity seen through the perspective of insects: 10 simple rules on methodological choices and experimental design for genomic studies. PeerJ 7: e6727-e6727. doi:10.7717/peerj.6727.

McLaren MR, Willis AD, Callahan BJ (2019) Consistent and correctable bias in metagenomic sequencing experiments. Elife 8: doi:10.7554/eLife.46923.

Meissner K (2021) The pathway of molecular methods from research to routine use. ARPHA Conference Abstracts 4: doi:10.3897/aca.4.e65556.

Meusnier I, Singer GA, Landry JF, Hickey DA, Hebert PD, Hajibabaei M (2008) A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9: 214. doi:10.1186/1471-2164-9-214.

Miotto RA, Rodrigues FP, Ciocheti G, Galetti Jr PM (2007) Determination of the Minimum Population Size of Pumas (Puma concolor) Through Fecal DNA Analysis in Two Protected Cerrado Areas in the Brazilian Southeast. Biotropica 39: 647-654. doi:10.1111/j.1744-7429.2007.00315.x.

Moran AJ, Prosser SWJ, Moran JA (2019) DNA metabarcoding allows non-invasive identification of arthropod prey provisioned to nestling Rufous hummingbirds (Selasphorus rufus). PeerJ 7: e6596-e6596. doi:10.7717/peerj.6596.

Nichols RV, Vollmers C, Newsom LA, Wang Y, Heintzman PD, Leighton M, Green RE, Shapiro B (2018) Minimizing polymerase biases in metabarcoding. Molecular Ecology Resources 18: 927-939. doi:10.1111/1755-0998.12895.

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.

Pawlowski J, Kelly-Quinn M, Altermatt F, Apotheloz-Perret-Gentil L, Beja P, Boggero A, Borja A, Bouchez A, Cordier T, Domaizon I, Feio MJ, Filipe AF, Fornaroli R, Graf W, Herder J, van der Hoorn B, Iwan Jones J, Sagova-Mareckova M, Moritz C, Barquin J, Piggott JJ, Pinna M, Rimet F, Rinkevich B, Sousa-Santos C, Specchia V, Trobajo R, Vasselon V, Vitecek S, Zimmerman J, Weigand A, Leese F, Kahlert M (2018) The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. Sci Total Environ 637-638: 1295-1310. doi:10.1016/j.scitotenv.2018.05.002.

Pedro PM, Amorim J, Rojas MVR, Sa IL, Galardo AKR, Santos Neto NF, Pires de Carvalho D, Nabas Ribeiro KA, Razzolini MTP, Sallum MAM (2020) Culicidae-centric metabarcoding through targeted use of D2 ribosomal DNA primers. PeerJ 8: e9057. doi:10.7717/peerj.9057.

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: 819-830. doi:10.1111/1755-0998.12355.

Piñol J, Senar MA, Symondson WOC (2019) The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. Molecular Ecology 28: 407-419. doi:10.1111/mec.14776.

Pornon A, Andalo C, Burrus M, Escaravage N (2017) DNA metabarcoding data unveils invisible pollination networks. Sci Rep 7: 16828. doi:10.1038/s41598-017-16785-5.

Porter TM, Hajibabaei M (2018) Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. Mol Ecol 27: 313-338. doi:10.1111/mec.14478.

Potts SG, Dauber J, Hochkirch A, Oteman B, Roy DB, Ahmé K, Biesmeijer K, Breeze TD, Carvell C, Ferreira C, FitzPatrick Ú, Isaac NJB, Kuussaari M, Ljubomirov T, Maes J, Ngo H, Pardo A, Polce C, Quaranta M, Settele J, Sorg M, Stefanescu C, Vujić A (2021) Proposal for EU Pollinator Monitoring Scheme. JRC Technical Report. Publications Office of the European Union, Luxembourg, pp. doi:10.2760/881843.

Pumkaeo P, Takahashi J, Iwahashi H (2021) Detection and monitoring of insect traces in bioaerosols. PeerJ 9: e10862. doi:10.7717/peerj.10862.

Richardson RT, Curtis HR, Matcham EG, Lin CH, Suresh S, Sponsler DB, Hearon LE, Johnson RM (2019) Quantitative multi-locus metabarcoding and waggle dance interpretation reveal honey bee spring foraging patterns in Midwest agroecosystems. Molecular Ecology 28: 686-697. doi:10.1111/mec.14975.

Richardson RT, Lin CH, Quijia JO, Riusech NS, Goodell K, Johnson RM (2015) Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding approach. Applications in Plant Sciences 3: 1500043-1500043. doi:10.3732/apps.1500043.

Ritter CD, Häggqvist S, Karlsson D, Sääksjärvi I, Muasya AM, Nilsson RH, Antonelli A (2019) Biodiversity assessments in the 21st century: the potential of insect traps to complement environmental samples for estimating eukaryotic and prokaryotic diversity using high-throughput DNA metabarcoding. Genome 62: 147-159.

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. Environmental DNA: doi:10.1002/edn3.290.

Ronquist F, Forshage M, Haggqvist S, Karlsson D, Hovmoller R, Bergsten J, Holston K, Britton T, Abenius J, Andersson B, Buhl PN, Coulianos CC, Fjellberg A, Gertsson CA, Hellqvist S, Jaschhof M, Kjaerandsen J, Klopfstein S, Kobro S, Liston A, Meier R, Pollet M, Riedel M, Rohacek J, Schuppenhauer M, Stigenberg J, Struwe I, Taeger A, Ulefors SO, Varga O, Withers P, Gardenfors U (2020) Completing Linnaeus's inventory of the Swedish insect fauna: Only 5,000 species left? PLOS ONE 15: e0228561. doi:10.1371/journal.pone.0228561.

Schmidt A, Schneider C, Decker P, Hohberg K, Rombke J, Lehmitz R, Balint M (2022) Shotgun metagenomics of soil invertebrate communities reflects taxonomy, biomass, and reference genome properties. Ecol Evol 12: e8991. doi:10.1002/ece3.8991.

Schneider J, Valentini A, Dejean T, Montarsi F, Taberlet P, Glaizot O, Fumagalli L (2016) Detection of invasive mosquito vectors using environmental DNA (eDNA) from water samples. PLOS ONE 11: e0162493-e0162493. doi:10.1371/journal.pone.0162493.

Seibold S, Gossner MM, Simons NK, Blüthgen N, Muller J, Ambarli D, Ammer C, Bauhus J, Fischer M, Habel JC, Linsenmair KE, Nauss T, Penone C, Prati D, Schall P, Schulze ED, Vogt J, Wollauer S, Weisser WW (2019) Arthropod decline in grasslands and forests is associated with landscape-level drivers. Nature 574: 671-674. doi:10.1038/s41586-019-1684-3.

Shirazi S, Meyer RS, Shapiro B (2021) Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding. Ecol Evol 11: 15766-15779. doi:10.1002/ece3.8239.

Shum P, Palumbi SR (2021) Testing small-scale ecological gradients and intraspecific differentiation for hundreds of kelp forest species using haplotypes from metabarcoding. Mol Ecol 30: 3355-3373. doi:10.1111/mec.15851.

Simmons BI, Balmford A, Bladon AJ, Christie AP, De Palma A, Dicks LV, Gallego-Zamorano J, Johnston A, Martin PA, Purvis A, Rocha R, Wauchope HS, Wordley CFR, Worthington TA, Finch T (2019) Worldwide insect declines: An important message, but interpret with caution. Ecol Evol 9: 3678-3680. doi:10.1002/ece3.5153.

Sow A, Brévault T, Benoit L, Chapuis M-p, Galan M, Coeur d'acier A, Delvare G, Sembène M, Haran J (2019) Deciphering host-parasitoid interactions and parasitism rates of crop pests using DNA metabarcoding. Scientific Reports 9: 3646-3646. doi:10.1038/s41598-019-40243-z.

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

Steinke D, deWaard SL, Sones JE, Ivanova NV, Prosser SWJ, Perez K, Braukmann TWA, Milton M, Zakharov EV, deWaard JR, Ratnasingham S, Hebert PDN (2022) Message in a Bottle-Metabarcoding enables biodiversity comparisons across ecoregions. GigaScience 11: doi:10.1093/gigascience/giac040.

Swenson SJ, Gemeinholzer B (2021) Testing the effect of pollen exine rupture on metabarcoding with Illumina sequencing. PLOS ONE 16: e0245611. doi:10.1371/journal.pone.0245611.

Swift JF, Lance RF, Guan X, Britzke ER, Lindsay DL, Edwards CE (2018) Multifaceted DNA metabarcoding: Validation of a noninvasive, next-generation approach to studying bat populations. Evolutionary Applications 11: 1120-1138. doi:10.1111/eva.12644.

Tedersoo L, Drenkhan R, Anslan S, Morales-Rodriguez C, Cleary M (2019) High-throughput identification and diagnostics of pathogens and pests: Overview and practical recommendations. Molecular Ecology Resources 19: 47-76. doi:10.1111/1755-0998.12959.

Thomsen PF, Sigsgaard EE (2019) Environmental DNA metabarcoding of wild flowers reveals diverse communities of terrestrial arthropods. Ecology and Evolution 9: 1665-1679. doi:10.1002/ece3.4809.

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.

Uiterwaal SF, DeLong JP (2020) Using patterns in prey DNA digestion rates to quantify predator diets. Mol Ecol Resour 20: 1723-1732. doi:10.1111/1755-0998.13231.

van der Heyde M, Bunce M, Wardell-Johnson G, Fernandes K, White NE, Nevill P (2020) Testing multiple substrates for terrestrial biodiversity monitoring using environmental DNA metabarcoding. Molecular Ecology Resources 00: 1-14. doi:10.1111/1755-0998.13148.

van Klink R, Bowler DE, Gongalsky KB, Chase JM (2022) Long-term abundance trends of insect taxa are only weakly correlated. Biol Lett 18: 20210554. doi:10.1098/rsbl.2021.0554.

van Klink R, Bowler DE, Gongalsky KB, Swengel AB, Gentile A, Chase JM (2020) Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. Science 368: 417-420. doi:10.1126/science.aax9931.

van Strien AJ, Termaat T, Groenendijk D, Mensing V, Kéry M (2010) Site-occupancy models may offer new opportunities for dragonfly monitoring based on daily species lists. Basic and Applied Ecology 11: 495-503. doi:10.1016/j.baae.2010.05.003.

van Strien AJ, van Swaay CAM, Termaat T, Devictor V (2013) Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. Journal of Applied Ecology 50: 1450-1458. doi:10.1111/1365-2664.12158.

Vasselon V, Ács É, Almeida S, Andree K, Apothéloz-Perret-Gentil L, Bailet B, Baricevic A, Beentjes K, Bettig J, Bouchez A, Capelli C, Chardon C, Duleba M, Elersek T, Genthon

C, Hurtz M, Jacas L, Kahlert M, Kelly M, Lewis M, Macher J, Mauri F, Moletta-Denat M, Mortágua A, Pawlowski J, Pérez Burillo J, Pfannkuchen M, Pilgrim E, Pissaridou P, Porter J, Rimet F, Stanic K, Tapolczai K, Theroux S, Trobajo R, van der Hoorn B, Vasquez Hadjilyra M, Walsh K, Wanless D, Warren J, Zimmermann J, Zupančič M (2021) The Fellowship of the Ring Test: DNAqua-Net WG2 initiative to compare diatom metabarcoding protocols used in routine freshwater biomonitoring for standardisation. ARPHA Conference Abstracts 4: doi:10.3897/aca.4.e65142.

Wang J, Santiago E, Caballero A (2016) Prediction and estimation of effective population size. Heredity (Edinb) 117: 193-206. doi:10.1038/hdy.2016.43.

Wang WY, Srivathsan A, Foo M, Yamane SK, Meier R (2018) Sorting specimen-rich invertebrate samples with cost-effective NGS barcodes: Validating a reverse workflow for specimen processing. Mol Ecol Resour 18: 490-501. doi:10.1111/1755-0998.12751.

Watts C, Dopheide A, Holdaway R, Davis C, Wood J, Thornburrow D, Dickie IA (2019) DNA metabarcoding as a tool for invertebrate community monitoring: a case study comparison with conventional techniques. Austral Entomology 58: 675-686. doi:10.1111/aen.12384.

Weitemier K, Penaluna BE, Hauck LL, Longway LJ, Garcia T, Cronn R (2021) Estimating the genetic diversity of Pacific salmon and trout using multigene eDNA metabarcoding. Mol Ecol 30: 4970-4990. doi:10.1111/mec.15811.

Wilson RS, Keller A, Shapcott A, Leonhardt SD, Sickel W, Hardwick JL, Heard TA, Kaluza BF, Wallace HM (2021) Many small rather than few large sources identified in long-term bee pollen diets in agroecosystems. Agriculture, Ecosystems & Environment 310: doi:10.1016/j.agee.2020.107296.

Wührl L, Pylatiuk C, Giersch M, Lapp F, von Rintelen T, Balke M, Schmidt S, Cerretti P, Meier R (2022) DiversityScanner: Robotic handling of small invertebrates with machine learning methods. Mol Ecol Resour 22: 1626-1638. doi:10.1111/1755-0998.13567.

Yang C, Bohmann K, Wang X, Cai W, Wales N, Ding Z, Gopalakrishnan S, Yu DW (2020) Biodiversity Soup II: A bulk-sample metabarcoding pipeline emphasizing error reduction. BioRxiv: doi:10.1101/2020.07.07.187666.

Young RG, Milian-Garcia Y, Yu J, Bullas-Appleton E, Hanner RH (2021) Biosurveillance for invasive insect pest species using an environmental DNA metabarcoding approach and a high salt trap collection fluid. Ecol Evol 11: 1558-1569. doi:10.1002/ece3.7113.

Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution 3: 613-623. doi:10.1111/j.2041-210X.2012.00198.x.

Zenker MM, Specht A, Fonseca VG (2020) Assessing insect biodiversity with automatic light traps in Brazil: Pearls and pitfalls of metabarcoding samples in preservative ethanol. Ecology and Evolution 10: 2352-2366. doi:10.1002/ece3.6042.

Zinger L, Bonin A, Alsos IG, Balint M, Bik H, Boyer F, Chariton AA, Creer S, Coissac E, Deagle BE, De Barba M, Dickie IA, Dumbrell AJ, Ficetola GF, Fierer N, Fumagalli L, Gilbert MTP, Jarman S, Jumpponen A, Kauserud H, Orlando L, Pansu J, Pawlowski J, Tedersoo L, Thomsen PF, Willerslev E, Taberlet P (2019) DNA metabarcoding-Need for robust experimental designs to draw sound ecological conclusions. Mol Ecol 28: 1857-1862. doi:10.1111/mec.15060.

Zinger L, Donald J, Brosse S, Gonzalez MA, Iribar A, Leroy C, Murienne J, Orivel J, Schimann H, Taberlet P, Lopes CM (2020) Advances and prospects of environmental DNA in neotropical rainforests. In: Dumbrell AJ, Turner EC, Fayle TM (Eds) Advances in ecological research. 331-373.

Zizka VMA, Geiger MF, Hörren T, Kirse A, Noll NW, Schäffler L, Scherges AM, Sorg M (2022) Recommendations for tissue homogenisation and extraction in DNA metabarcoding of Malaise trap samples. doi:10.1101/2022.01.25.477667.

1127 Zizka VMA, Leese F, Peinert B, Geiger MF (2019) DNA metabarcoding from sample fixative as
1128 a quick and voucher-preserving biodiversity assessment method (1). Genome 62: 122-
1129 136. doi:10.1139/gen-2018-0048.
1130

Figures and tables

Figure 1: Semi-quantitative metrics. (A) ASV table as the outcome of a DNA metabarcoding experiment, rows are samples, columns are ASVs, numbers are raw read counts. From the ASV table, semi-quantitative metrics can be derived, e.g. frequency and percentage of occurrence, bipartite networks and relative read abundance; (B) Frequency and percentage of occurrence derived from ASV table, frequency of occurrence simplifies the ASV table into presence/absence data, indicated by presence or absence of a rectangle, when summarising this over all samples, percentage of occurrence can be an informative metric for abundance in a system (right); (C) Bipartite networks derived from the ASV table, samples and ASVs are nodes, edges indicate presence/absence of the ASVs per sample (left), when summarising this over all samples, link strength can be an informative metric for abundance in a system (right); (D) Relative read abundance derived from ASV table, relative read abundance for individual samples is determined by dividing raw read counts of individual ASVs by total read count per sample (left), when summarising this over all samples, mean relative read abundance can be an informative metric for abundance in a system (right); abbreviations: S – Sample, ASV – Amplicon sequence variant, RRA – relative read abundance; ASVs are colour coded and refer to ASVs from (A), artwork: Alice Scherges.

Figure 2: Reducing read abundance biases. (A) Processing mock communities (bottle) with defined composition allows determining taxon-specific correction factors, that can be applied to correct relative read abundance of samples with unknown composition, indicated by a red line. Correction factors can only be determined for taxa included in the

mock community; (B) Correction factors can be determined using iterative algorithms and a guess-and-test approach based on a morphological reference data set (not shown). The correction factors can be applied to correct relative read abundance of samples with unknown composition, indicated by a red line. Correction factors can only be determined for samples, that show a good agreement in terms of taxa detected between the reference and the DNA metabarcoding data set; (C) Adding spike-ins, e.g. a defined amount of genomic DNA, to all samples and co-amplifying and co-sequencing the reference material allows correcting raw read counts by simply dividing read counts assigned to taxa (blue and brown bars) by read counts assigned to the spike-in (red bars); abbreviations: RRA – relative read abundance, S – sample, artwork: Alice Scherges.

Figure 3: Combination of methods. Samples or sample subsets (bottle) can be processed and analysed by a variety of methodological approaches in parallel to achieve complementary datasets, examples are, starting from the top and moving clock-wise, specimen photography (camera icon), sample weighing or body size measurements for biomass estimation (scales icon), quantitative PCR (amplification curves), DNA metabarcoding (ASV table), digital droplet PCR (event number ~ amplitude graph), morphological analyses (binocular icon), abbreviations: qPCR – quantitative PCR, ASV – amplicon sequence variant, ddPCR – digital droplet PCR, artwork: Alice Scherges.

Figure 4: **Recommended workflow for biodiversity assessments with bulk samples and DNA metabarcoding that obtains count and biomass data with species level taxonomic identifications.** (A) Specimens from a bulk sample (bottle) are first

processed individually. (B) Processing includes specimen photography (camera), specimen counts (abacus), body size measurements (caliper) and biomass estimation (scales). Ideally, this is done automatically (green robot icon) and involves automatic image recognition to achieve preliminary taxa identifications on broad taxonomic scales. (C) Specimens are then re-combined to a community sample, a spike-in is added and DNA is extracted (microcentrifuge tube). (D) DNA metabarcoding delivers species level identifications and raw read counts (ASV table), which are corrected via the spike-in. (E) Image data is combined to a taxon list containing count, size and biomass data (taxon list). (F) Image data and DNA metabarcoding data are combined using machine learning approaches (data assembly, orange robot icon) to obtain a data set that contains information on species level identities, along with count data and biomass estimates (taxa bubbles), abbreviations: ASV – amplicon sequence variant, artwork: Alice Scherges.

Table 1: Overview of methodological approaches discussed in this publication.

Box 1: Glossary

Figure 1: Semi-quantitative metrics.

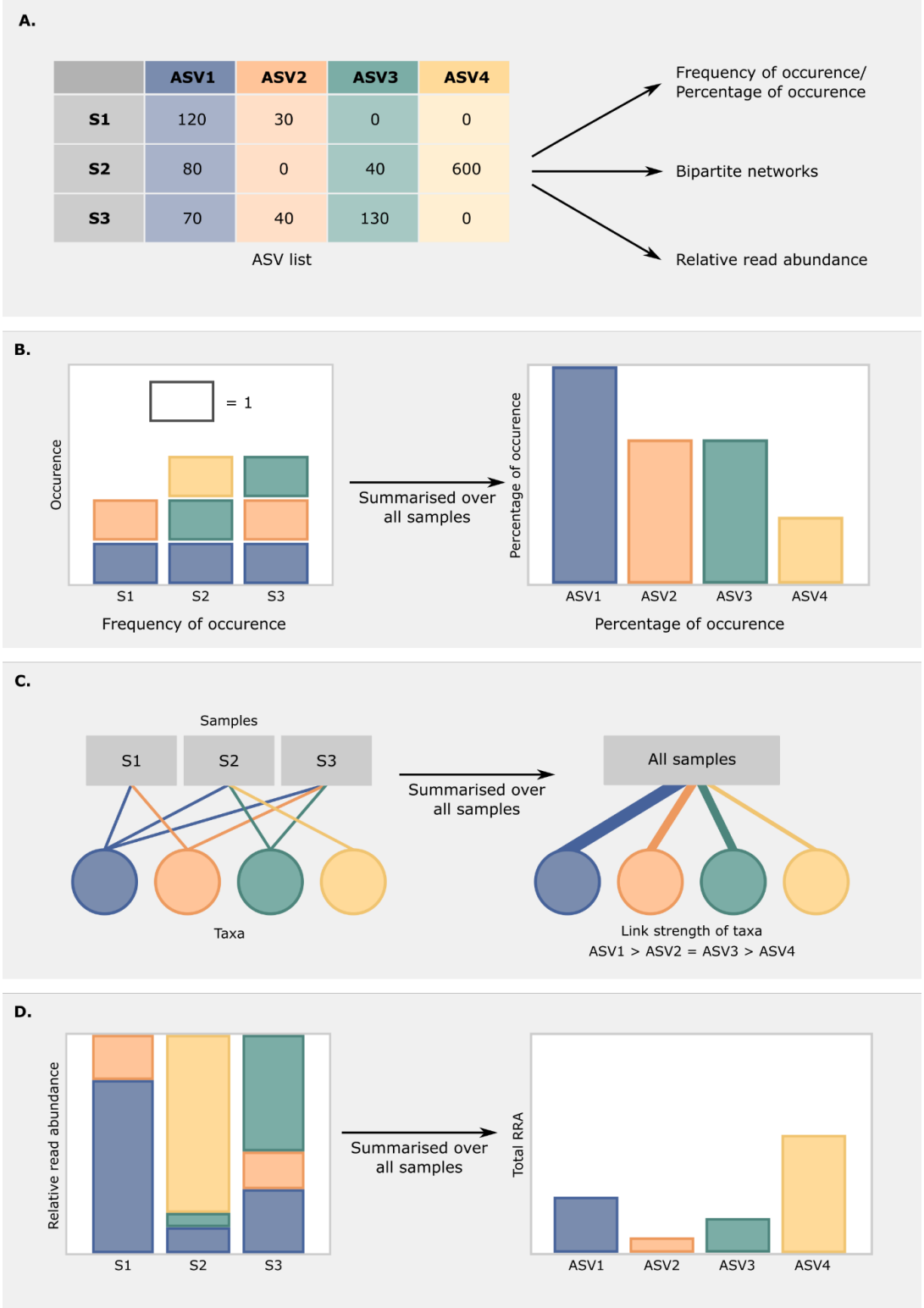


Figure 2: Reducing read abundance biases.

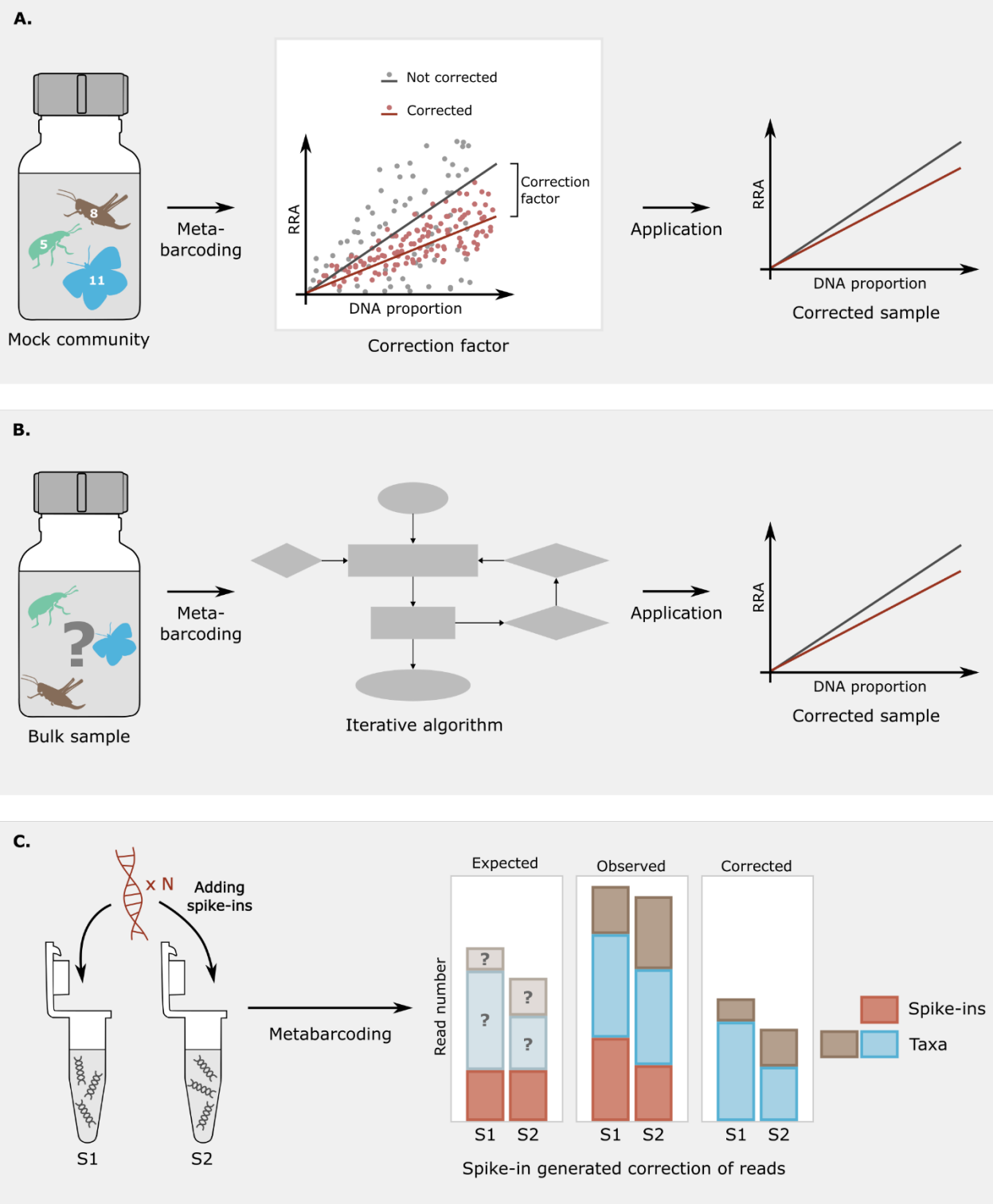
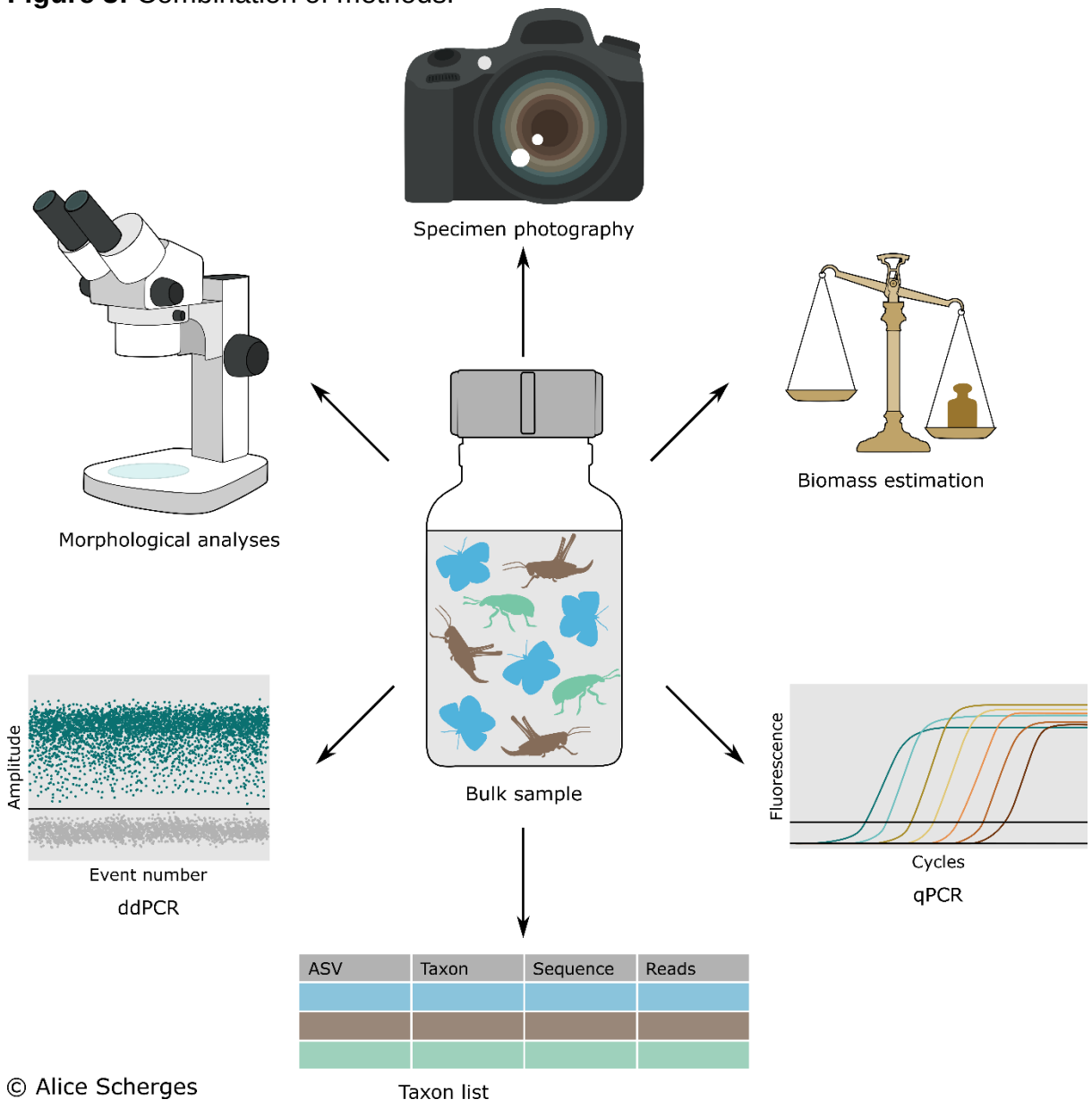
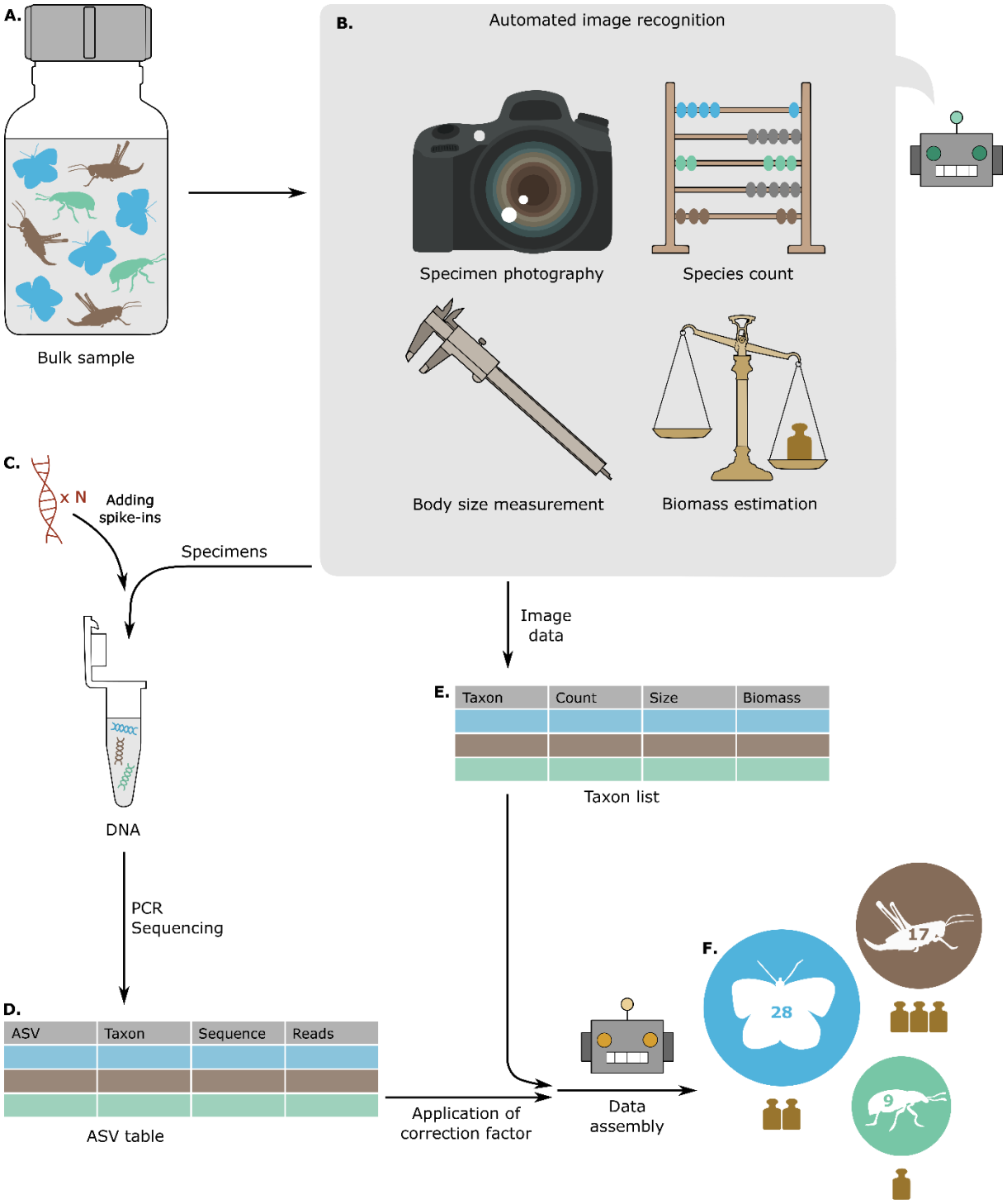


Figure 3: Combination of methods.



© Alice Scherges

1202 **Figure 4**: Recommended workflow for biodiversity assessments with bulk samples and
1203 DNA metabarcoding that obtains count and biomass data with species level taxonomic
1204 identifications.



1205 © Alice Scherges

Table 1: Overview of methodological approaches discussed in this publication.

Category	Approach	Sample types	Quantitative?
Semi-quantitative metrics	FOO/POO	all	semi-quantitative
	RRA	all	semi-quantitative
	rarefaction	all	semi-quantitative
	transformation	all	semi-quantitative
Reducing read abundance biases	correction factors via algorithm	woc samples	yes; virtual species counts
	correction factors via mock communities	all	yes
	spike-ins	all	yes; relative abundance
	primer optimisation	all	no
	multi-locus metabarcoding	all	yes
Combination of methods	general	all; depending on approach	yes; depending on approach
	photography and body measurements of single specimens	woc samples	yes; species counts, biomass

Box1: Glossary, only terms that have not been defined in (Porter and Hajibabaei 2018) are explained in more detail

Abundance	in this review: general term to refer to different measures of species abundance, including species counts, relative abundance, biomass
Biomass	weight of individuals belonging to the same species
Correction factors	taxon-specific correction of read numbers assigned to that taxon; obtained via mock community sequencing or iterative algorithms (e.g. Darby et al. 2020; Darby et al. 2013; Krehenwinkel et al. 2017b)
DNA metabarcoding	performing DNA barcoding on mixed-species samples within one single PCR and coupled with high throughput sequencing ('mixed template PCR' in (Porter and Hajibabaei 2018))
eDNA	see (Porter and Hajibabaei 2018); environmental DNA
ASV	see (Porter and Hajibabaei 2018); amplicon sequence variant; alternative terms: exact sequence variant (ESV), zOTU (zero-radius OTU; Callahan et al. 2017; Edgar 2016)
FOO / POO	frequency of occurrence/percent of occurrence: taxon occurrences are combined across a number of individual samples (biological replicates) for a population-wide estimate; FOO – number of samples; POO – percentage of samples reported (Deagle et al. 2019)
Species counts	absolute number of individuals belonging to the same species

Metagenomics, mitogenomics / mito-metagenomics	see (Porter and Hajibabaei 2018)
Mock community	artificial communities of known species composition / DNA concentration or similar; often used as a positive control (e.g. Ji et al. 2020; Krehenwinkel et al. 2017b)
Multi-locus metabarcoding	DNA metabarcoding based on the (separate) amplification and sequencing of different genetic markers for the same organismal group
NGS / HTS	next-generation sequencing, modern sequencing platforms that generate millions of DNA sequences in a single run (e.g. Illumina desktop sequencers), alternative term: high-throughput sequencing
NGS barcoding	performing DNA barcoding on individual specimens, but exploiting the time- and cost-effectiveness of high-throughput sequencing platforms, rather than relying on traditional Sanger sequencing (e.g. Wang et al. 2018)
Rarefaction	sample-wise random sub-sampling of reads to even sequencing depth across all samples
Relative abundance	Proportion of a species within a sample
RRA	relative read abundance; taxon-wise number of reads per sample divided by total number of reads per sample
Spike-ins / internal	fixed amount of reference material added to each separate

standards	sample (e.g. Ji et al. 2020; Saitoh et al. 2016)
Transformation	application of (mathematical) function to sequence reads, e.g. log-based transformations or rank-based transformations
woc samples	whole organism community samples, mixed species samples, obtained from mass-sampling traps, e.g. Malaise trap

1210