

Optimizing insect metabarcoding using replicated mock communities

Elzbieta Iwaszkiewicz-Eggebrecht¹  | Emma Granqvist¹  | Mateusz Buczek²  |
Monika Prus² | Jan Kudlicka³  | Tomas Roslin⁴  | Ayco J. M. Tack⁵  |
Anders F. Andersson⁶  | Andreia Miraldo¹  | Fredrik Ronquist¹  | Piotr Łukasik^{1,2} 

¹Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden; ²Institute of Environmental Sciences, Faculty of Biology, Jagiellonian University, Kraków, Poland; ³Department of Data Science and Analytics, BI Norwegian Business School, Oslo, Norway; ⁴Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden; ⁵Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden and ⁶KTH Royal Institute of Technology, Science for Life Laboratory, Department of Gene Technology, School of Engineering Sciences in Chemistry, Biotechnology and Health, Stockholm, Sweden

Correspondence

Elzbieta Iwaszkiewicz-Eggebrecht
Email: ela.iwaszkiewicz@nrm.se

Funding information

H2020 European Research Council, Grant/Award Number: 856506; Knut och Alice Wallenbergs Stiftelse, Grant/Award Number: KAW 2017.088; Narodowa Agencja Wymiany Akademickiej, Grant/Award Number: PPN/PPO/2018/1/00015; Narodowe Centrum Nauki, Grant/Award Number: 2018/31/B/NZ8/01158; National Institute of General Medical Sciences, Grant/Award Number: R35GM124701; Vetenskapsrådet, Grant/Award Number: 2018-04620, 2018-05973 and 2019-04493

Handling Editor: Douglas Yu

Abstract

1. Metabarcoding (high-throughput sequencing of marker gene amplicons) has emerged as a promising and cost-effective method for characterizing insect community samples. Yet, the methodology varies greatly among studies and its performance has not been systematically evaluated to date. In particular, it is unclear how accurately metabarcoding can resolve species communities in terms of presence-absence, abundance and biomass.
2. Here we use mock community experiments and a simple probabilistic model to evaluate the effect of different DNA extraction protocols on metabarcoding performance. Specifically, we ask four questions: (Q1) How consistent are the recovered community profiles across replicate mock communities?; (Q2) How does the choice of lysis buffer affect the recovery of the original community?; (Q3) How are community estimates affected by differing lysis times and homogenization? and (Q4) Is it possible to obtain adequate species abundance estimates through the use of biological spike-ins?
3. We show that estimates are quite variable across community replicates. In general, a mild lysis protocol is better at reconstructing species lists and approximate counts, while homogenization is better at retrieving biomass composition. Small insects are more likely to be detected in lysates, while some tough species require homogenization to be detected. Results are less consistent across biological replicates for lysates than for homogenates. Some species are associated with strong PCR amplification bias, which complicates the reconstruction of species counts. Yet, with adequate spike-in data, species abundance can be determined

Fredrik Ronquist and Piotr Łukasik equal senior authorship.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

with roughly 40% standard error for homogenates, and with roughly 50% standard error for lysates, under ideal conditions. In the latter case, however, this often requires species-specific reference data, while spike-in data generalize better across species for homogenates.

4. We conclude that a non-destructive, mild lysis approach shows the highest promise for the presence/absence description of the community, while also allowing future morphological or molecular work on the material. However, homogenization protocols perform better for characterizing community composition, in particular in terms of biomass.

KEYWORDS

abundance estimation, homogenization, Malaise trap, metabarcoding, mock communities, non-destructive mild lysis, spike-ins

1 | INTRODUCTION

Insects are likely the most species-rich group of animals on Earth (Stork, 2018) and fulfil a myriad of different ecosystem functions (Jordan et al., 2021). Yet our knowledge about insect diversity in natural ecosystems is still surprisingly scarce – and the key reason for this deficit is their actual diversity. Recent reports on insect abundance and diversity declines (Hallmann et al., 2017; Seibold et al., 2019; van Klink et al., 2020) make monitoring more important than ever. Nonetheless, surveys based on traditional taxonomic work demand enormous time and labour. In one of the most ambitious projects aiming to characterize the insect fauna at the national level—the *Swedish Malaise Trap Project*—sorting 1919 samples into 300 taxonomic fractions took 15 years, despite the high manpower invested (estimated 98K person-hours devoted to sorting; Karlsson et al., 2020). While taxonomic work remains crucial, there is a pressing need for the development of complementary high-throughput and cost-effective methods that enable adequate description of insect community diversity, composition and spatiotemporal dynamics.

In this context, the ever-decreasing costs of high-throughput sequencing (HTS) have made DNA-based methods particularly attractive for large-scale insect monitoring. A number of such methods have been proposed for and tested on bulk invertebrate samples (Kennedy et al., 2020). The method, perhaps, most widely used today is community-level metabarcoding. In this approach, DNA is first extracted from the whole multi-species sample, such as a Malaise trap catch. Afterwards, a short region of a marker gene, typically mitochondrial cytochrome oxidase I (COI), is amplified by PCR with broad-spectrum primers and subsequently sequenced. The resulting collection of marker gene sequences—barcodes—is compared against reference databases, providing information about the identities of species present in the original bulk sample. This rapid and affordable workflow has been used to describe arthropod communities (Cristescu, 2014; Elbrecht et al., 2017; Taberlet et al., 2012) and is increasingly proving useful in ecological studies (Beng et al., 2016;

Liu et al., 2020; Porter et al., 2019). Nevertheless, despite its obvious advantages, metabarcoding is still a relatively young method and many challenges remain. Crucially, it is unclear how accurately metabarcoding can describe species diversity (presence/absence) and community composition (species' relative abundance)—and how its accuracy can be improved.

Research to date has pointed to multiple obstacles to obtaining accurate estimates of both species-specific incidence—adding up to species diversity—and abundance from metabarcoding. Differences in species' biomass and physical structure cause unequal contribution of DNA to the sequenced pool. Added to this, primer biases towards certain taxonomic groups can obscure estimates of species diversity and abundance, as demonstrated through metabarcoding of samples with known composition (Braukmann et al., 2019; Elbrecht et al., 2019; Elbrecht & Leese, 2015; Martoni et al., 2022). One of the solutions proposed to address these challenges is the introduction of known amounts of either tissue or DNA (so-called *spike-ins*; Ji et al., 2020; Luo et al., 2023; Thomas et al., 2016) to facilitate normalizing sequencing results between and within samples.

Many crucial considerations in metabarcoding, however, relate to the early stages of sample processing—before DNA extraction. At present, the most common way of treating bulk samples is to homogenize specimens into an 'insect soup' (Liu et al., 2020; Morinière et al., 2016; Yu et al., 2012). This destructive method results in high DNA yields, but also in the loss of all morphological information—a discouraging outcome for taxonomists. Homogenization can also decrease the odds of detection of smaller and/or rare insects since large or abundant species will contribute much more DNA to the final pool and thereby swamp the DNA of smaller and/or rare species during sequencing. An alternative approach is then to use a non-destructive mild lysis, in which insects soak in the lysis buffer for a short period. The procedure leaves specimens almost intact—as demonstrated in a diverse arthropod sample by Kirse et al. (2023)—allowing future taxonomic work. The mild lysis approach has been developed and applied in freshwater and terrestrial arthropod studies (Batovska et al., 2021; Carew et al., 2018; Martins

et al., 2019; Nielsen et al., 2019), showing promising results in terms of evenness in species representation. However, critical methodological parameters such as the lysis buffer composition or lysis duration and conditions, have not been systematically investigated—despite their likely effects on the relative DNA yield from different species and, thus, on any estimates of community composition.

Arguably, the best way of evaluating methodological aspects of metabarcoding is their application to artificial communities of known composition, also called mock communities (Cristescu, 2014). By controlling specimen numbers and approximate biomass, it is possible to obtain the reference that any metabarcoding results can be compared. Since metabarcoding is sensitive to multiple sources of variation, making sufficient replication is the key to obtaining robust results.

However, many studies to date have used wild-caught insects to construct mock communities (Batovska et al., 2021; Zizka et al., 2019), which precludes proper replication since each sample is unique. Others have based their approach on DNA rather than insect samples and constructed mock communities by blending species-specific DNA extracts or extracts from low-diversity samples into more complex blends (Braukmann et al., 2019; Nielsen et al., 2019). While such approaches might be suitable for testing PCR and sequencing steps of the workflow, they depart significantly from realistic community sample processing procedures. We argue that well-replicated, relevant studies may best be obtained by constructing mock communities from size-standardized, homogeneous populations of different species. Such samples can be prepared in multiple copies, allowing us to evaluate the robustness of our inferences made from metabarcoding.

In this study, we use a set of individual-based mock communities as a powerful tool to systematically compare the effects of methodological choices on inferences from metabarcoding data. To scrutinize our ability to accurately describe insect communities in terms of species presence/absence, abundance and biomass, we use metabarcoding of replicate community samples to generate taxonomically assigned sequence yields, and a simple probabilistic model to analyse the results. We revisit the impact of different methodological solutions in three experiments using mock communities and ask the following questions: (Q1) How consistent are the composition profiles across replicate mock communities?; (Q2) How does the choice of buffer affect community recovery?; (Q3) How are community estimates affected by differing lysis times and homogenization? and (Q4) Is it possible to obtain adequate species abundance estimates through the use of biological spike-ins?

2 | MATERIALS AND METHODS

Our study is built on three separate experiments, all using replicate mock communities constructed from different combinations of up to 25 insect species (Figure 1). In Experiment I, we compare different formulae for lysis buffer and assess if they affect the inferences of species presence. For Experiment II, we select the most promising

buffer and then test a wider range of lysis times applied to larger and more diverse communities. In both Experiments I and II, we also assess how mild lysis and homogenization methods compare and how they affect our ability to accurately reconstruct the true composition of mock communities. Finally, in Experiment III, we use biological spike-ins to assess our ability to recover species abundance from bulk samples subjected to mild lysis, as well as homogenization.

2.1 | Constructing mock communities (Q1)

For constructing replicate insect assemblages of known composition, we used combinations of 31 standardized, reference insect species obtained from biocontrol and pet stores, hobbyists, laboratory cultures and social insect colonies (Table S1). All individuals within a species represented the same developmental stage and were size-standardized. Insects were preserved in 90% ethanol and stored at -20°C until used. Species-specific dry weights were calculated as averages from 10 specimens. Communities were prepared by combining pre-set numbers of individuals from each species in a Falcon tube with ethanol. Every mock community was prepared in at least three identical copies. Each experiment used different community types, as explained below.

2.2 | Experiment I: Assessing the effects of buffer and lysis time/homogenization (Q2, Q3)

To establish the effect of different lysis approaches, we assembled three mock communities—S, M and L—consisting of 8–13 species represented by 11–25 individuals in total (Figure 1, Table S2). Each community was prepared in six replicates: three were lysed with Buffer 1 (Vesterinen et al., 2016; modified from Aljanabi & Martinez, 1997) and the other three with Buffer 2 (arthropod-specific buffer from the Canadian Centre for DNA Barcoding; Ivanova et al., 2006). Communities were incubated in 5 mL of the buffer at 56°C , in a shaking water bath with one negative control per buffer. Fifty μL aliquots were taken after 2, 4 and 6 h of incubation, and then the samples were homogenized using bead beaters. DNA was purified from 20 μL aliquots of the lysate or homogenate. (For buffer recipes and details, see Text S1).

2.3 | Experiment II: Testing a wider range of conditions on more complex mock communities (Q3)

To further explore how mild lysis time influences community estimates and how homogenization and mild lysis compare, we expanded the range of lysis times and increased the size, number and complexity of mock communities. We used five distinct communities, comprising 20–25 species represented by 68–132 individuals. Each mock community was replicated three times (Figure 1b, Table S3). The 15 mock community samples and one

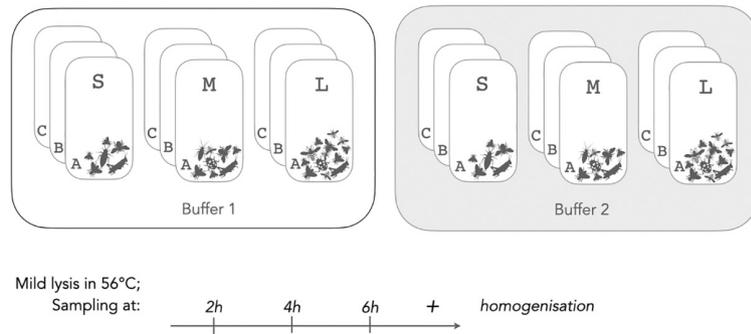
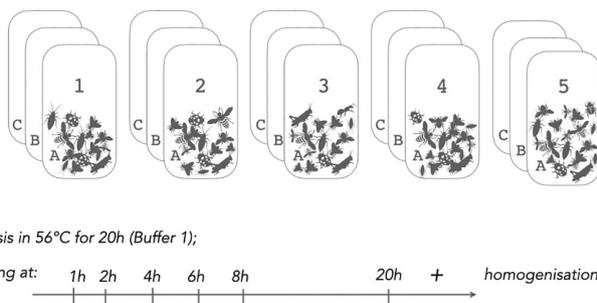
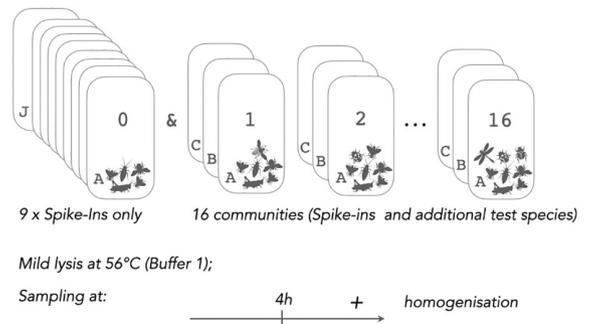
(a) Experiment I: Assessing the effects of buffer and lysis time/homogenization**(b) Experiment II: Testing a wider range of conditions on more complex mock communities****(c) Experiment III: Estimating species abundances**

FIGURE 1 Visual summary of three Experiments that form the basis of the current study. (a) In Experiment I, we tested two lysis buffers on three mock communities of different sizes (S, M and L: up to 25 individuals of 13 species). Communities were prepared in six replicates: three incubated with Buffer 1 and the other three with Buffer 2. We collected lysate aliquots at three time points before homogenizing them. (b) In Experiment II, we made five larger and more complex communities (up to 132 individuals of 25 species), preparing each in three replicates. All were incubated with Buffer 1, and six lysate aliquots were taken, followed by homogenization. (c) In Experiment III, we prepared nine identical communities made of five spike-in species and 16 different mock communities made of the spike-ins and a varying number of seven other species (6–23 individuals per community). Communities, prepared in triplicate, were lysed for 4 h and then homogenized.

negative control were incubated with 20 mL of Buffer 1 at 56°C in a shaking water bath. Two-hundred-and-fifty μL aliquots of lysate were taken at six time points during the incubation (1, 2, 4, 6, 8 and 20 h). Afterwards, each sample was homogenized using a bead beater. DNA was purified using silica-coated magnetic beads from a 225 μL aliquot of each lysate and homogenate. (see details in Text S2).

2.4 | Experiment III: Estimating species abundance with the use of biological spike-ins (Q4)

To explore the utility of biological spike-ins in improving abundance estimations, we assembled 17 mock communities, using up to 12 species. Five species: *Shelfordella lateralis* (Blattodea: Blattellidae), *Drosophila hydei*, *Drosophila yakuba* (both Diptera: Drosophilidae), *Gryllus bimaculatus* and *Grylodes sigillatus* (both Orthoptera:

Gryllidae) were used for calibration and we will henceforth refer to them as 'spike-ins'. Community type '0' comprised only spike-ins and was prepared in nine replicates. The other 16 communities had a variable number of additional test species (1–18 individuals representing up to six species) and were prepared in three copies (Figure 1c, Table S4). Communities and one negative control were subjected to mild lysis in Buffer 1 at 56°C for 4 h and then homogenized using bead beaters. DNA was purified as in Experiment I.

2.5 | Library preparation and bioinformatics

Amplicon libraries were prepared using a two-step PCR approach, as described by Marquina et al. (2021). We amplified a 458-bp region of the COI gene using BF3-BR2 primers (Elbrecht et al., 2019) and sequenced the library pools on Illumina MiSeq using Reagent Kit v3 (2 \times 300 bp reads). (For details see Text S3).

We used *cutadapt* v.3.2 (Martin, 2011) for primer trimming and DADA2 R package for amplicon analysis (Callahan et al., 2016) both in R environment v.4.0 (R Core Team, 2020). Sequences were dereplicated (*derepFastq()*), denoised (*learnErrors()*, *dada()*) and merged (*mergePairs()*) and chimeras removed (*removeBimeraDenovo()*), resulting in ASV (amplicon sequence variants) count tables. Since the makeup of communities was known, ASVs other than those representing verified, error-free species sequences were removed from the count tables.

2.6 | Data visualization and statistical analysis

To statistically analyse differences in inferred community composition depending on buffer and/or treatment (lysis time, homogenization) in Experiments I and II, we performed Permutational Multivariate Analysis of Variance (PERMANOVA) using Bray–Curtis dissimilarity distance for species relative read abundance and Jaccard distance for species presence/absence data with the use of *adonis2()* function of the *VEGAN* package (Oksanen et al., 2020) and *pairwise.adonis()* of the *PAIRWISEADONIS* package (Martinez Arbizu, 2020). To visualize results, we performed nonmetric multidimensional scaling (nMDS) with the *metaMDS()* function of the *VEGAN* package, based on Bray–Curtis distances. Relative read counts from Experiment I were plotted as a heatmap with the use of *PHEATMAP* package (Raivo, 2019) and those from Experiment II as stacked bar plots with the *GGPLOT2* R package (Wickham, 2016).

For Experiment III, we developed a hierarchical Bayesian model that allowed us to estimate the variation across specimens and species in DNA yield and the variation across samples in PCR amplification factors from spike-in data. The model parameters learned were then used to estimate the abundance and biomass of the remaining species. As a basis for the model, let d_{tmij} be the DNA yield, the amount of template DNA extracted from an insect specimen i of species t in sample j of community m and effectively available for PCR amplification. In the simplest case, we model d_{tmij} using a universal gamma distribution

$$d_{tmij} \sim \text{Gamma}(k, \theta),$$

where k is the shape parameter and θ the scale parameter.

We now introduce a PCR factor, c_j , specific to each sample j . This factor represents how many times the DNA template is multiplied in sample j , taking into account amplification, subsampling, sequencing depth and bioinformatic processing steps after DNA extraction. The PCR factor only affects the scale parameter of the gamma distribution, while the number of specimens or biomass affects only the shape parameter. Specifically, given that there are n_{tm} specimens, the read count r_{tmj} (interpreted as a concentration) is distributed as

$$r_{tmj} \sim \text{Gamma}(n_{tm}k, c_j\theta).$$

Given suitable prior probability distributions on k , θ and c , these parameters can be inferred from observed read counts. Specifically, we

used a gamma prior on k and lognormal priors on θ and c , with conjugate normal-inverse-gamma hyperpriors. For abundance estimates, we used a uniform prior to n . We explored variation across species in DNA yield and PCR amplification by introducing species-specific k and θ parameters, respectively.

Models were implemented as probabilistic programs in *Birch* version 2.0.34 (Murray & Schön, 2018), and the inference was performed using sequential Monte Carlo.

Further model and bioinformatic details are available on GitHub: https://github.com/ronquistlab/optimizing_metabarcoding_mock_communities

3 | RESULTS

After primer trimming and quality filtering, the sequencing yielded an average of 7573 (SD +/- 3869) COI reads per sample (Table S5). We recovered the barcode sequences of all species making up mock communities but not in every sample where they were present. Instead, the presence/absence and relative abundance of species varied substantially among mock community types, buffers and treatments (lysis time or homogenization). Furthermore, we observed differences among replicates of the same community. We explore these patterns by answering the methodological questions defined in the Introduction.

3.1 | [Q1] How consistent are the composition profiles across replicate mock communities?

In all Experiments, we observed considerable variation in the relative abundance of species and also in their inferred presence/absence. Certain species were consistently undetected in some of the treatments (patterns explored below, Q3). For others, the rate of false negatives seemed to vary haphazardly from replicate to replicate, especially after mild lysis (Figures 2 and 4). After homogenization estimates from replicates were becoming more similar to one another. The Bray–Curtis distances between replicates were considerably higher among samples treated by mild lysis (0.22 on average in Experiment II) than between homogenized ones (0.12). Overall, inferred community compositions were distinct from one another (Table S9; PERMANOVA, $p_{\text{adjust}} < 0.001$), but the estimates varied from replicate to replicate, despite them coming from the identical mock community.

3.2 | [Q2] How does the choice of lysis buffer affect community reconstruction?

In Experiment I, the choice of buffer significantly affected the reconstruction of species presence/absence patterns in Community S but not others (Table S6; PERMANOVA, $p < 0.05$ for S but $p > 0.05$ for M and L). However, when considering relative read counts

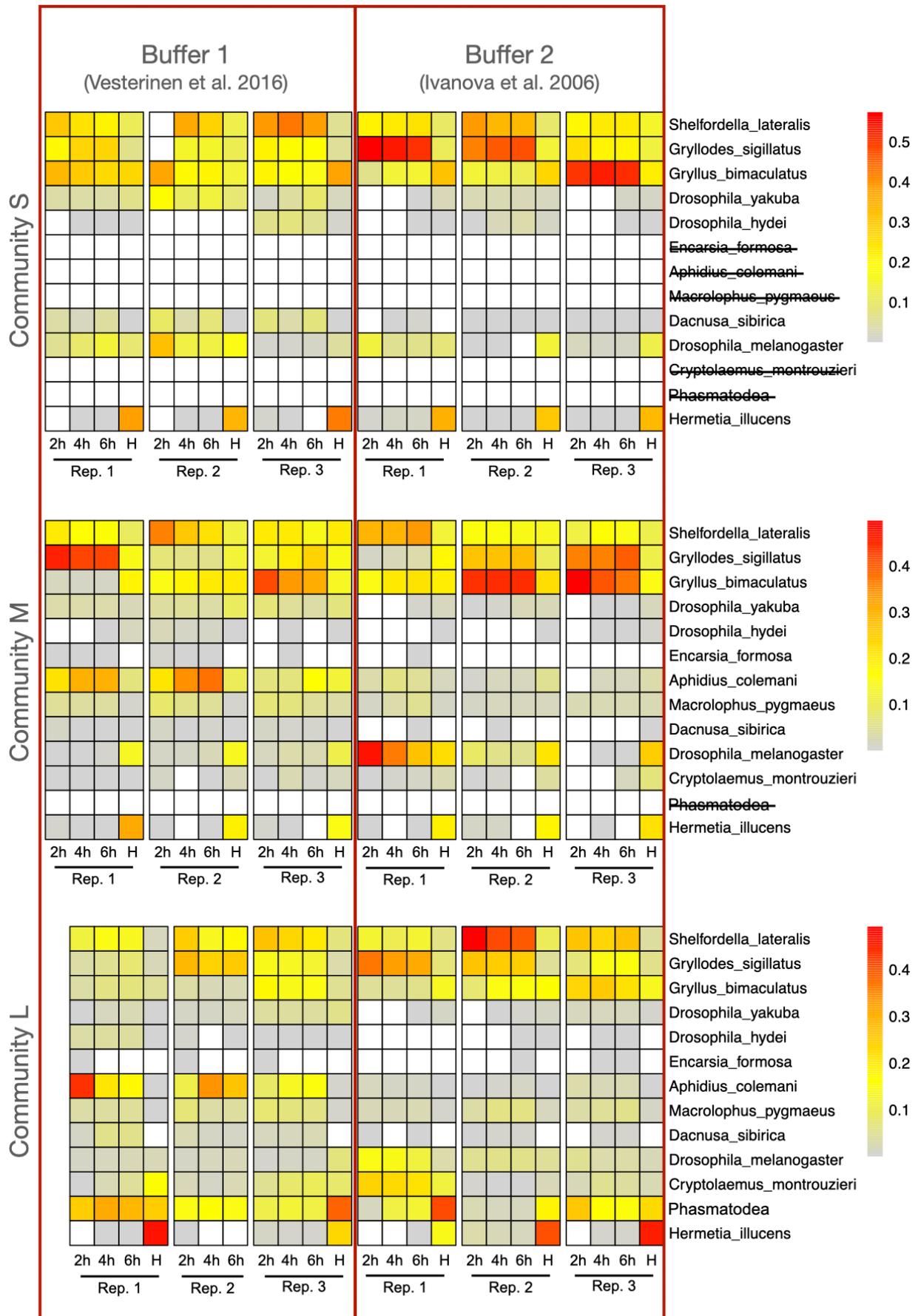


FIGURE 2 Heatmap showing relative abundance of species in Experiment I. Mock communities (three types: S, M, L) were subjected to mild lysis in two buffers (B1 and B2), with aliquots taken after 2, 4 and 6 h, then homogenized (H). We used three biological replicates, identical copies, per community, per buffer (Rep. 1–3). DNA purified from the aliquots was then prepped for metabarcoding. Heatmap colours present relative read counts per experimental species recovered in metabarcoding. White tiles indicate 0 reads. Species absent from a given mock community by design have their name crossed out in the legend; note that there were no false positives.

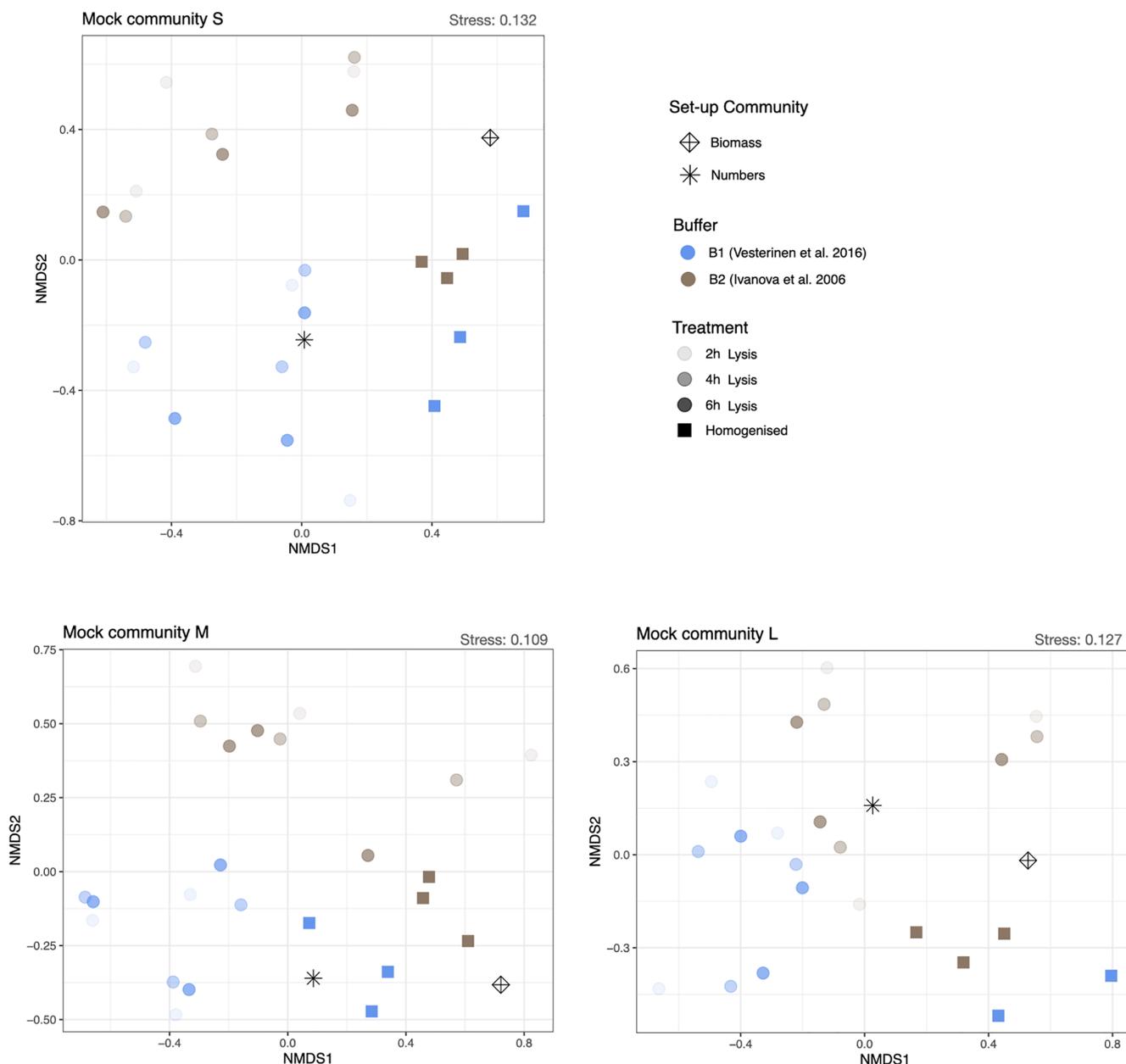


FIGURE 3 Visualization of non-metric multi-dimensional scaling (nMDS) based on metabarcoding results from Experiment I. Mock communities S, M and L are the same as in Figure 2. Different colours represent the two buffers. Homogenized samples are squares while mild lysis is circles with lysis times represented as opacity levels. The original community set-up in terms of the relative number of species and the relative biomass of the species are represented by an asterisk and crossed diamond, respectively. Stress factor for each ordination is given in the top right corner.

we found a strong effect of the buffer used in all Communities (Table S6; PERMANOVA, $p < 0.01$). Overall, digestion with Buffer 1 resulted in substantially more reads from smaller species such as

Aphidius colemani, *Dacnusa sibirica*, *Encarsia formosa* and *Drosophila yakuba*. With Buffer 2, these small species remained undetected in multiple replicates, whereas we scored higher relative read counts

FIGURE 4 Relative species abundance in mock communities from Experiment II. Panels a-e present five mock communities (I–V) and each consists of three sections: (1) two bar plots showing the set-up composition of the given mock community in terms of numbers of individuals of each species (first barplot) and their relative biomass contributions (second barplot); (2) Barplots showing relative read counts obtained through metabarcoding for six lysis times (1 h–20 h) and homogenization for three biological replicates (Rep.1–3); (3) Panel above the bar plots is showing false negatives, where coloured fields indicate species that were not detected in metabarcoding even though they were present in the mock community.

from larger species such as *S. lateralis*, *G. sigillatus* and *G. bimaculatus* (Figure 2).

Multidimensional scaling further corroborated this result (Figure 3). Replicates treated by mild lysis were well separated from homogenized samples along the first axis. While homogenized samples clustered close together regardless of the buffer used, the mild lysis samples were clearly separated between Buffer 1 and 2 along the second nMDS axis (with two individual replicates forming exceptions; see Figure 3).

Given the improved detection of small and rare species conveyed by Buffer 1, we chose this buffer for our subsequent experiments.

3.3 | [Q3] How are community estimates affected by different lysis times and homogenization?

We found no significant differences between samples lysed for different amounts of time ($p_{\text{adjust}} > 0.05$; Tables S7 and S11). However, close inspection of relative read counts revealed that the representation of several species was changing in a predictable manner, either consistently increasing or decreasing with the duration of lysis. For instance, in Experiment II, the relative abundance of *A. colemani* decreased over time, while that of *D. yakuba* was increasing (Figure 4).

Treatments—mild lysis/homogenization—significantly altered the community composition profiles inferred. Species presence/absence differed in all communities from Experiment II (PERMANOVA, $p < 0.01$; Table S10) and in Communities S and L from Experiment I (PERMANOVA, $p < 0.05$; Table S6) but not *M.* When considering relative read counts, treatment had a significant effect in all communities (PERMANOVA $p < 0.01$; Tables S7 and S10).

In more detail, we observed increasing or decreasing relative abundance for many species depending on the treatment (Figure 4). The smallest insects tended to be more reliably detected after mild lysis, yielding few or no reads after homogenization. For instance, the smallest species in our set (the parasitic wasp *E. formosa*) was consistently represented by multiple reads after short lysis (1–4 h) in Communities I and II (Experiment II). However, its proportion of reads decreased rapidly with increasing lysis time and after homogenization, *E. formosa* was no longer detected in any replicate (Figure 4a,b). The same was true for other small insects such as *A. colemani*, *Aphidius ervi* and *Aphidoletes aphidimyza* (Figure 4). For bigger insects (e.g. the lepidopteran *Pieris napi*) and harder-bodied insects (e.g. the mealworm beetle *Tenebrio molitor*), we observed the opposite tendency: species were detectable already after the shortest of lysis times (1 h), but absolute read counts increased

with lysis time and peaked after homogenization (Figure 2). Species with a hard exoskeleton, such as ants *Formica rufa* and *Formica fusca*, released little or no DNA in mild lysis, and thus homogenization was required to detect them (Figure 4a,c). Overall, there was a tendency for intermediate lysis times to result in the highest detected species richness (Figure S1). Homogenization tended to result in lower detected richness, especially in more complex communities. Nonetheless, interspecific variation in the more precise time-dependency of detectability (Figure 4) was so pronounced that it blurred any patterns at the level of overall species richness (Figure S1).

Patterns revealed by nMDS ordination (Figure 5) corroborate different imprints of extraction method on different aspects of community composition. When comparing the composition of mock communities inferred from metabarcoding to their original, known composition, we noted that homogenates resemble the biomass composition of the community, while short lysis samples group closer to their original species abundance composition, with a clear trend as lysis time increases towards homogenization.

3.4 | [Q4] Is it possible to obtain adequate species abundance estimates through the use of biological spike-ins?

Of the five spike-in species used in Experiment III, *D. hydei* amplified poorly and was therefore excluded from further comparisons. Thus, the following results apply to the remaining spike-in species—*S. lateralis*, *G. bimaculatus*, *G. sigillatus* and *D. yakuba*. There was considerable variation among replicates in relative proportions of reads for different spike-ins, especially after the mild lysis (Figure S2).

Using the probabilistic model for lysate data, we found strong evidence in favour of models accommodating species-specific differences in DNA yield over models that did not, while there was no evidence for such a difference in homogenates (Figure 6). Posterior estimates of model parameters illustrate the same results in more detail: there were distinct differences between species in average DNA yield per specimen in lysates but not in homogenates (Figure 7).

Estimates of coefficients of variation (CVs) can be used to pinpoint the level of consistency in read counts (the higher the CV, the lower consistency). Under the simple model that did not accommodate species-specific differences, the estimated CV for specimen numbers was around 0.80 for lysates and around 0.45 for homogenates (Figure 7c,d). In other words, the standard deviation (i.e. standard error, SE) was around 80% vs 45% of the mean, respectively. If we did not use the spike-in data to learn the PCR factor of each

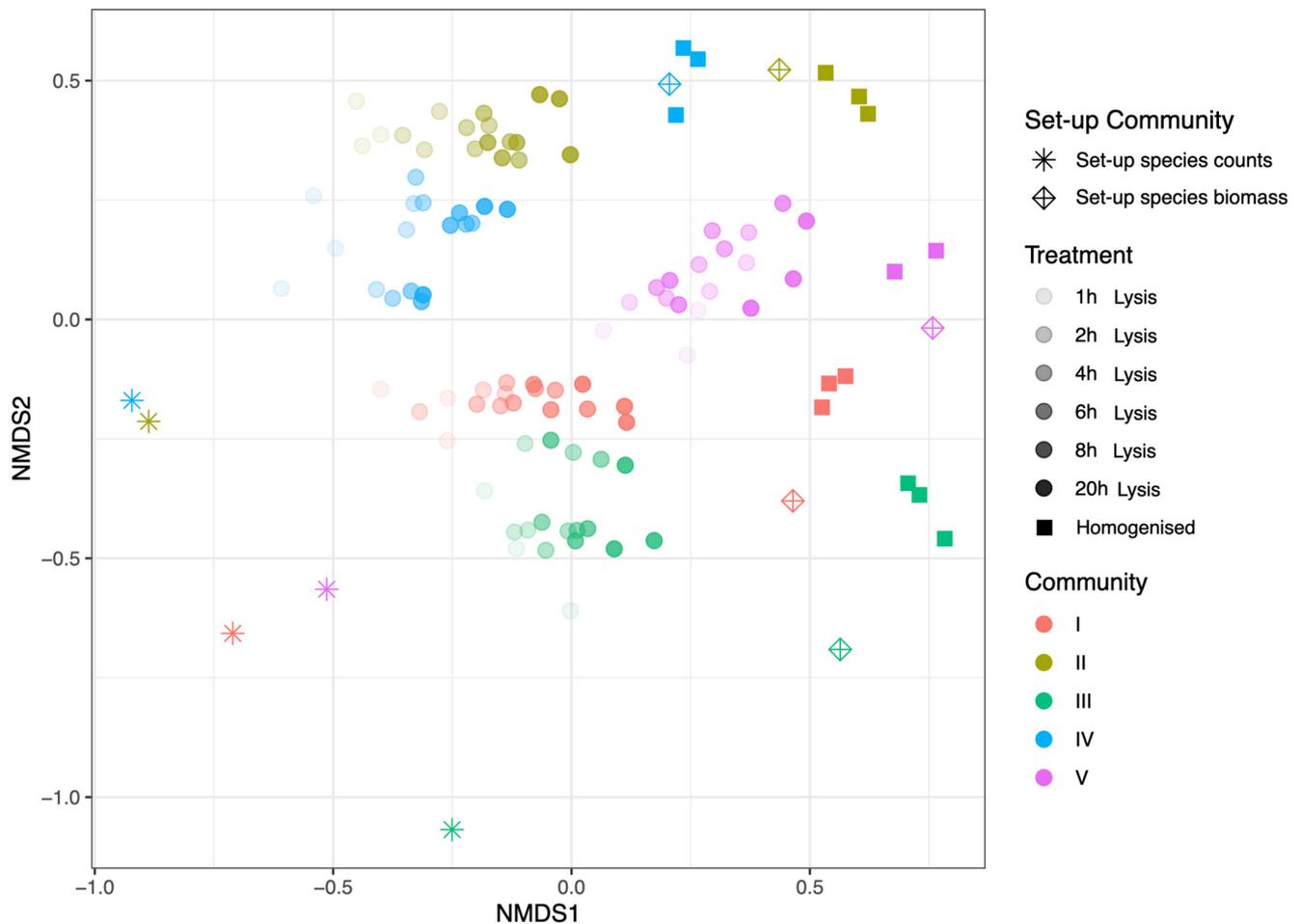


FIGURE 5 nMDS plot summing up COI reads for all samples in Experiment II. Five arthropod mock communities (I–V) in biological triplicates were subjected to mild lysis for 20h, sampled six times throughout the experiment and subsequently homogenized. Lysis time of each sample is represented in the graph as opacity level. Different shapes inform about whether a sample was homogenized (square), lysed (circle) or if it represents the original community set-up in terms of the relative number of individuals contributing to the community (asterisk) or the relative biomass of the species (crossed diamond).

run, the CV increased to around 0.9 for lysates and 0.6 for homogenates (Figure S3c,d; constant PCR factor). When both PCR factor and species-specific differences in DNA yield were accommodated, the CV decreased to 0.5 for three of the four spike-ins in the lysate data, while it remained high for *D. yakuba*, which was also the smallest of the four spike-ins (Figure 7). For the homogenate data, the CV estimates did not change when species-specific differences were accommodated.

When the spike-in data, fitted to a common model for all spike-in species, were used to estimate the abundance of the remaining species, we noticed clear species-specific effects: some species were consistently overestimated while others were consistently underestimated (Figure 8b,d and Figure S5b,d). For lysate data, the estimates of specimen numbers were slightly better than the estimates of biomass, while the opposite was true for homogenate data (Figure 8a,c and Figure S5a,c). The lysate predictions tended to provide close to unbiased estimates of the number of specimens but clearly overestimate biomass, while the homogenate predictions gave approximately unbiased estimates of biomass but slightly underestimated the number of specimens.

4 | DISCUSSION

Our series of experiments provide significant new insights into the methodological choices that determine the reliability of species detection as well as abundance and biomass estimations in the metabarcoding of multi-species community samples. The use of replicate mock communities provided an unprecedented insight into the extent and significance of variation inherent to metabarcoding. Our results contribute to the understanding of trade-offs involved and the limits of DNA metabarcoding broadly. Below, we explore different aspects of the results.

4.1 | Reproducibility of metabarcoding results

Previous studies have reported limited differences in the perception of community composition gained from technical replicates, for example, when using different DNA purification methods (Marquina et al., 2019; Nielsen et al., 2019) or when evaluating taxonomic bias

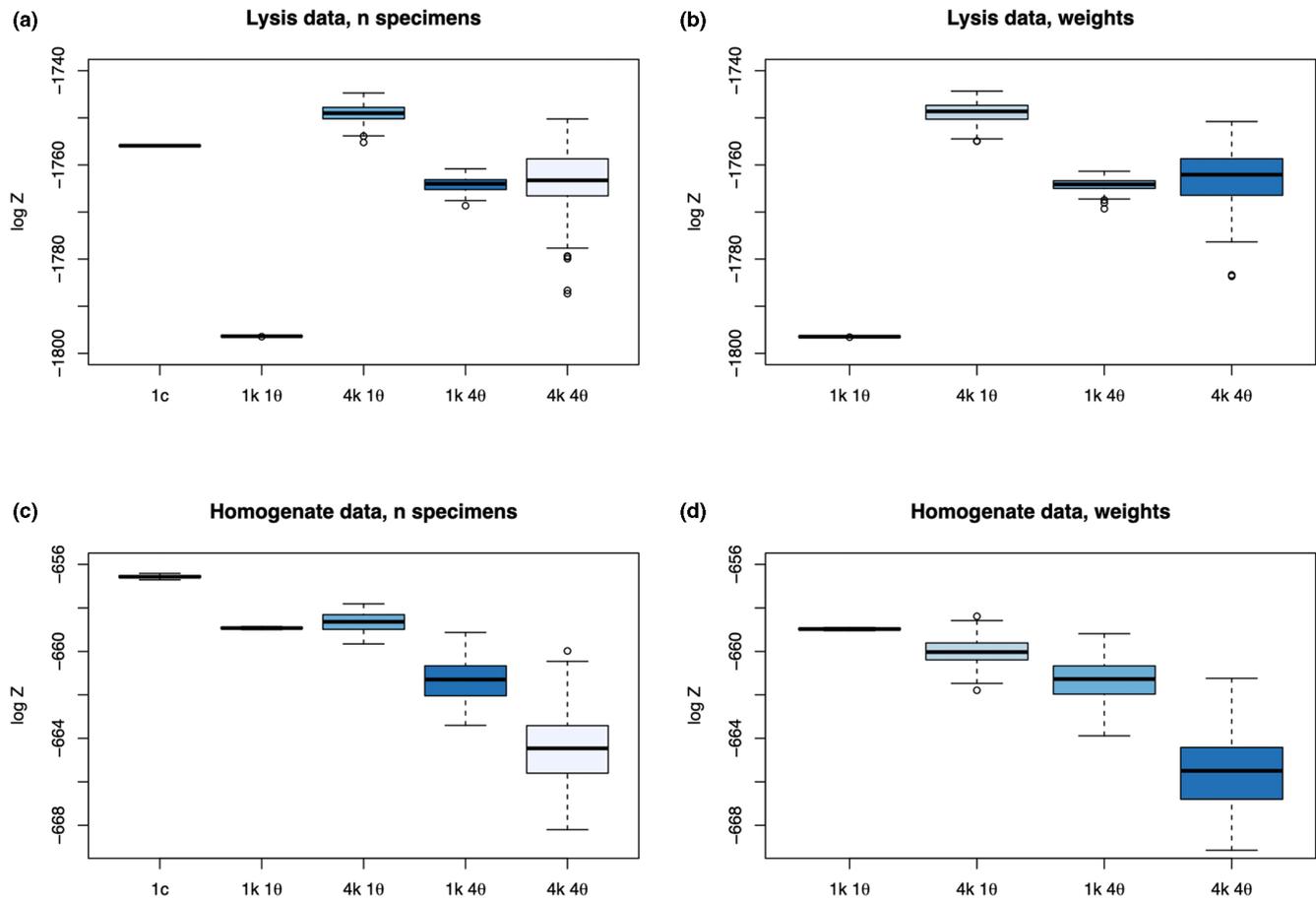


FIGURE 6 Estimated marginal likelihoods ($\log Z$) for four models accommodating different sources of variation in the relative read numbers of the four spike-in species. Results for lysis data are in the top panels (a, b) and for homogenate data in the bottom (c, d). The left panels (a, c) show results for specimen composition, and the right panels (b, d) show results for biomass composition. A difference of five log-likelihood units is considered very strong evidence against the model with lower likelihood (Kass & Raftery, 1995). In lysates, there is strong evidence for species-specific differences in DNA yield (4 k over 1 k models), whereas this is not the case in homogenates. In neither case, there is strong evidence for species-specific PCR amplification biases in these four species (40 over 10 models).

due to PCR primers (Martoni et al., 2022). However, to our knowledge, the current study is the first to examine variation between biological replicates of multi-species communities. Our results reveal a surprising amount of variation among replicates of identical mock communities. It concerns both the relative read abundance and the presence/absence of species. It was particularly strong after mild lysis treatment but also persisted, to a lesser extent, after homogenization. Despite the variation, replicates of the same community were still distinguishable from those of other communities. But we should highlight that our mock communities were explicitly designed to be different.

It seems likely that this variation is largely due to natural variation in DNA yield among individuals of the same species. Recall that in our study, efforts were made to limit variation among individuals; for instance, we used only one life stage per species and excluded individuals that appeared poorly sclerotized, injured, particularly large or small, or otherwise unusual. Thus, one might expect even more variation among real wild-caught insects, where specimens differ in life stage, age, sex, health and physiological

state. Adding to this, how samples are handled, transported and stored, will also contribute to variation in DNA yield of specimens, with the effects likely to vary among species (Marquina et al., 2021). Thus, this variation is inherent in metabarcoding procedures and will clearly limit the precision of any estimates of community composition.

4.2 | Trade-offs between protocols

The different outcomes of our metabarcoding experiments contribute to the understanding of the trade-offs involved. Generally speaking, a milder protocol (short lysis and mild buffer) results in more accurate reflection of the presence/absence of species and community profiles more closely resembling those based on the numbers of specimens per species. The more aggressive the protocol (i.e. buffer with stronger lytic properties, longer lysis, homogenization), the more reflective the results are of the biomass composition of the community, favouring bigger specimens.

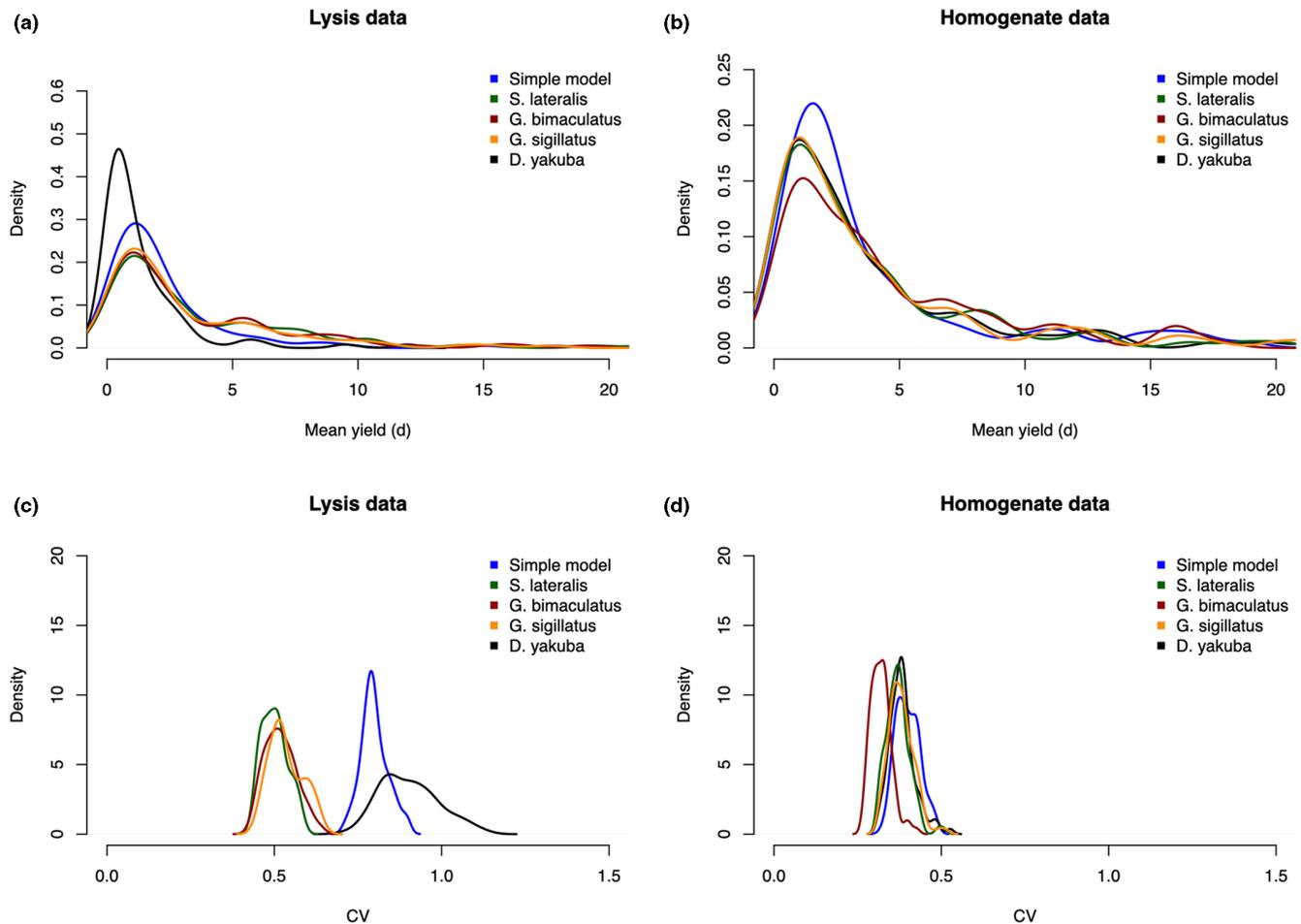


FIGURE 7 Estimated posterior distributions of the mean DNA yield (a, b) and the coefficient of variance (CV) of DNA yield per specimen (c, d). Using two model variants: 'simple model' with no species-specific parameters (blue), and the model that allows the DNA yield of the four spike-in species to be different (remaining colours). The mean DNA yield varies among spike-ins in lysates (a); in particular, the yield of *D. yakuba* is lower than that of the other spike-in species. The DNA yield is more consistent across species in the homogenates (b). The CV was reduced to approximately 0.4 for three of the four species, but not for *D. yakuba*, in lysates (c). In homogenates (d), the CV was estimated to be around 0.45 in the simple model, and accommodating species-specific DNA yields did not improve estimation except for *Gryllus bimaculatus*.

Notably, the volume of lysate and homogenate aliquots used for DNA extraction and library preparation were the same across all treatments. Thus, the shifting representation of species in metabarcoding is reflecting the varying proportion of their DNA present in the starting aliquot. The likely explanation for this is that small insects have a high surface-to-volume ratio and tend to release relatively more DNA initially. With more aggressive treatments the DNA pool becomes more reflective of the insects' volumes. As a result, large species dominate in terms of available DNA and we fail to detect small species. This is demonstrated by the high rate of false negatives for some of the smallest species (*A. aphidimyza*, *D. sibirica*, *E. formosa* and *Tuberculatus annulatus*) in the homogenates of Experiment II. One solution to this issue, however costly, is adding PCR replicates, as they increase the probability of detecting small species (Alberdi et al., 2018).

On top of this, there are species-specific effects, the most obvious of which is the degree of sclerotization. For instance, *Hermetia illucens* larvae yielded little DNA in mild lysis treatments

despite their large size. This may be due to their unusually hard integument (Ståhls et al., 2020). Similarly, the two ants (*F. fusca* and *F. rufa*) required homogenization to be detected at all. Still, their low representation even in homogenates suggests an additional possible PCR bias against them. Certainly, there are numerous other traits that influence how species respond to extraction protocols.

Overall, our findings are in line with the results of Marquina et al. (2023), who found that small and soft species were more difficult to detect after a chemically aggressive treatment (buffer with high SDS, DTT and proteinase K concentrations) than after lysis with the milder buffer, corresponding to Buffer 1 in the current study. They also agree with the results of Marquina et al. (2019), who showed that hard-bodied and large insects are more easily detected in homogenates, while analysis of preservative ethanol—an extremely mild lysis of sorts—is better for the detection of small and soft-bodied insects. However, it remains to be evaluated whether and under what conditions this pattern holds true for real terrestrial arthropod catches (but see Nielsen et al., 2019).

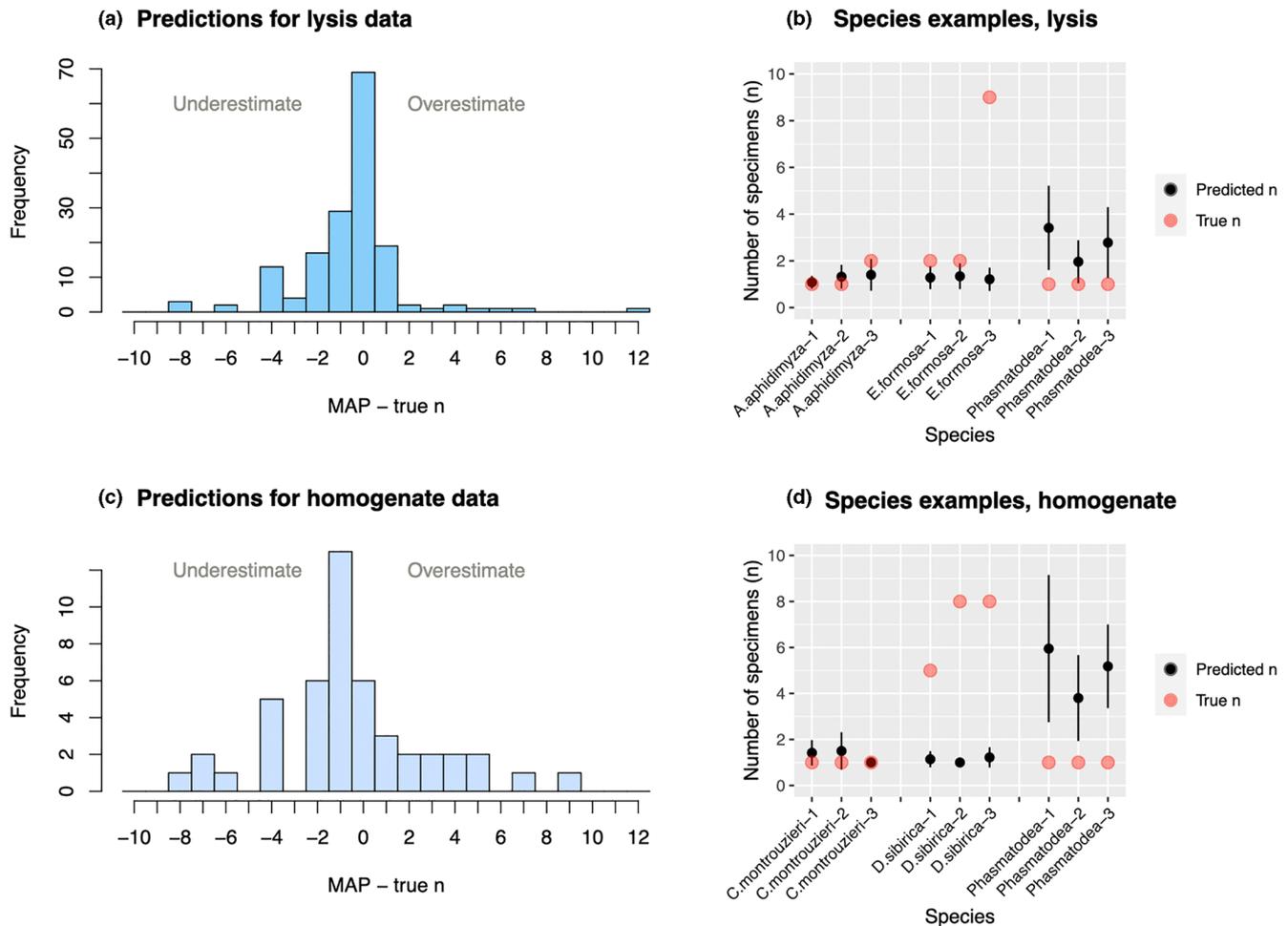


FIGURE 8 Predictions of species abundance in Experiment III for the lysis treatment (a, b) and homogenate treatment (c, d). Histograms (a, c) present specimen numbers (n) predictions for all species and samples combined. The x-axes show the difference between the true value of n and the maximum a posteriori (MAP) estimate. When the difference is zero, the model's MAP estimate is also the true value of n . In (b) and (d), examples of species-specific predictions are shown: three samples each for three species, one species that is well predicted, one species that is underestimated and one species that is overestimated. The black circles mark the mean of the estimates, the bars are standard deviation and the red circles mark the true value.

4.3 | Abundance estimates from metabarcoding data

How well species abundance may be estimated through metabarcoding is currently a hotly debated topic. Early studies suggested that metabarcoding was unsuitable for quantification, but more recent efforts using correction factors or spike-ins of known concentration to calibrate metabarcoding-based estimates have reported encouraging results (Ershova et al., 2021; Luo et al., 2023). Nevertheless, our study offers some of the first quantitative analyses of the accuracy that might be achieved using such approaches.

The use of biological spike-ins is based on the expectation that parameters learned from spike-in data can be applied to the species occurring naturally in the sample, but several sources of variation may influence the results. First, the mean DNA yield per specimen is likely to differ dramatically among species. Second, specimens from a single species can vary substantially in DNA yield. Finally, PCR amplification may also vary considerably among species (Piñol

et al., 2015, 2019). Indeed, in Experiment III, mean DNA yields varied considerably across our biological spike-in species in lysates, making it difficult to calibrate data for species other than those deliberately spiked-in. However, when applying a modelling approach and providing species-specific calibration data, it seems possible to bring down the error of abundance estimates for specimens to a CV of around 50% of the mean under ideal conditions (recall that we size-standardized the spike-in specimens). The mean DNA yield in the homogenate data is more uniform across species, and the CV estimates are lower. Under ideal conditions, one might then be able to reach a CV of 45% when using homogenate data for spike-in species to calibrate biomass estimates for other species.

These results, however, do not apply to species with strong PCR amplification biases such as *D. hydei* and the stick insect (Phasmatodea)—with very low and high PCR factors, respectively. Primer mismatches are usually thought to be the cause of such biases (Piñol et al., 2015, 2019), however, they are an unlikely culprit in this study. We used highly degenerate primers with broad

coverage—BF3-BR2 (Elbrecht et al., 2019)—providing a perfect match (of some of the primer variants) to all experimental species, including *D. hydei*. Additionally, the Qiagen Multiplex polymerase used has relatively little bias when simultaneously amplifying multiple templates in metabarcoding experiments, but factors such as GC contents within the priming site or template may still affect relative amplification efficiency (Nichols et al., 2018). It will be important to explore ways of reducing the bias, perhaps by adjusting the annealing temperature or other PCR reaction parameters. For instance, Yang et al. (2021) demonstrated an improvement in species diversity recovery from mock arthropod DNA mixtures when annealing temperature and number of PCR cycles were decreased. However, it remains to be tested on realistic insect bulk samples.

Overall, our data suggest that there is a natural limit to the precision of metabarcoding-based abundance estimates. Specifically, it appears difficult to bring the standard error of homogenate-based estimates much below 40%–45% of the mean when using biological spike-in references. These limits derive from natural variation between specimens, and stochastic processes in sample preparation. Nonetheless, the statistical method presented here can still be improved to increase the range of situations in which this natural limit can be reached. For instance, the error associated with variable DNA yield from biological spike-ins could possibly be reduced through synthetic spike-ins, such as linearized plasmids with artificial COI targets that are added in pre-defined quantities to homogenates (Palmer et al., 2018; Tourlousse et al., 2017). Alternatively, using individual-specimen barcoding of a subset of lysed or homogenized samples could extend the calibration data to species that occur naturally in the samples. Such an approach would also open up new possibilities for model-free methods by providing a sufficiently large training dataset for machine learning algorithms.

4.4 | Towards the optimal protocol

Given the different advantages and disadvantages of lysis versus homogenization, it will be hard to strike the ideal compromise. Detecting the smallest and rarest species in the sample will always be challenging—particularly for homogenate data, where they constitute a minuscule proportion of the total DNA. One approach, albeit destructive and costly, might be to metabarcode both the lysate and homogenate of the same sample. Such a combination would take advantage of the combined qualitative and quantitative strong suits of the two approaches.

With respect to the specifics of the protocol, our results suggest that intermediate lysis times, between 2 and 8 h, and using our Buffer 1 provide optimal, robust results. 1 h lysis appears to be suboptimal, with increased variation in species representation and lower overall DNA yield. Samples digested for 20 h start to resemble homogenates, with decreased representation of small species.

All in all, our current analysis provides key pointers for the design of an optimal protocol for processing bulk insect samples, with the specifics to be further attuned to local logistics and laboratory setup. We remain optimistic about the prospects of improved statistical modelling to unlock the resulting data's full potential.

AUTHOR CONTRIBUTIONS

Elzbieta Iwaszkiewicz-Eggebrecht, Piotr Łukasik and Fredrik Ronquist conceived the ideas and designed the methodology; Elzbieta Iwaszkiewicz-Eggebrecht and Piotr Łukasik collected the data with the assistance of Mateusz Buczek and Monika Prus; Fredrik Ronquist and Emma Granqvist developed the model and Jan Kudlicka contributed to its implementation in Birch. Elzbieta Iwaszkiewicz-Eggebrecht and Emma Granqvist analysed the data; Elzbieta Iwaszkiewicz-Eggebrecht, Piotr Łukasik and Fredrik Ronquist led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

This project was supported by the Knut and Alice Wallenberg Foundation (KAW 2017.088 to F.R.), Swedish Research Council grant No 2018-04620 (to F.R.), Polish National Agency for Academic Exchange grant PPN/PPO/2018/1/00015 (to P.Ł.), Polish National Science Centre grant 2018/31/B/NZ8/01158 (to P.Ł.) and the National Institute of General Medical Sciences of the NIH under award number R35GM124701 to Brandon S. Cooper (insect culture). T.R. was funded by the European Research Council Synergy Grant 856506 (LIFEPLAN) and a Career Support grant from the Swedish University of Agricultural Sciences. A.J.M.T. was funded by the Swedish Research Council (2019-04493). The data handling was enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC), partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

CONFLICT OF INTEREST STATEMENT

All authors declare that they have no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14073>.

DATA AVAILABILITY STATEMENT

Raw sequence data for this project have been submitted to NCBI's SRA archive under accession number PRJNA885053.

ORCID

Elzbieta Iwaszkiewicz-Eggebrecht  <https://orcid.org/0000-0003-1412-1711>

Emma Granqvist  <https://orcid.org/0000-0002-1513-1674>

Mateusz Buczek  <https://orcid.org/0000-0002-6144-6941>

Jan Kudlicka  <https://orcid.org/0000-0003-3806-4950>

Tomas Roslin  <https://orcid.org/0000-0002-2957-4791>
 Ayco J. M. Tack  <https://orcid.org/0000-0002-3550-1070>
 Anders F. Andersson  <https://orcid.org/0000-0002-3627-6899>
 Andreia Miraldo  <https://orcid.org/0000-0001-6107-006X>
 Fredrik Ronquist  <https://orcid.org/0000-0002-3929-251X>
 Piotr Łukasik  <https://orcid.org/0000-0002-4164-6487>

REFERENCES

- Alberdi, A., Aizpuru, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, 9, 134–147. <https://doi.org/10.1111/2041-210X.12849>
- Aljanabi, S. M., & Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, 25, 4692–4693. <https://doi.org/10.1093/nar/25.22.4692>
- Batovska, J., Piper, A. M., Valenzuela, I., Cunningham, J. P., & Blacket, M. J. (2021). Developing a non-destructive metabarcoding protocol for detection of pest insects in bulk trap catches. *Scientific Reports*, 11, 7946. <https://doi.org/10.1038/s41598-021-85855-6>
- Beng, K. C., Tomlinson, K. W., Shen, X. H., Surget-Groba, Y., Hughes, A. C., Corlett, R. T., & Slik, J. W. F. (2016). The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land-use change in the tropics. *Scientific Reports*, 6, 24965. <https://doi.org/10.1038/srep24965>
- Braukmann, T. W. A., Ivanova, N. V., Prosser, S. W. J., Elbrecht, V., Steinke, D., Ratnasingham, S., Waard, J. R. d., Sones, J. E., Zakharov, E. V., & Hebert, P. D. N. (2019). Metabarcoding a diverse arthropod mock community. *Molecular Ecology Resources*, 19, 711–727. <https://doi.org/10.1111/1755-0998.13008>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13, 581–583. <https://doi.org/10.1038/nmeth.3869>
- Carew, M. E., Coleman, R. A., & Hoffmann, A. A. (2018). Can non-destructive DNA extraction of bulk invertebrate samples be used for metabarcoding? *PeerJ*, 6, e4980. <https://doi.org/10.7717/peerj.4980>
- Cristescu, M. E. (2014). From barcoding single individuals to metabarcoding biological communities: Towards an integrative approach to the study of global biodiversity. *Trends in Ecology & Evolution*, 29, 566–571. <https://doi.org/10.1016/j.tree.2014.08.001>
- Elbrecht, V., Braukmann, T. W. A., Ivanova, N. V., Prosser, S. W. J., Hajibabaei, M., Wright, M., Zakharov, E. V., Hebert, P. D. N., & Steinke, D. (2019). Validation of COI metabarcoding primers for terrestrial arthropods. *PeerJ*, 7, e7745. <https://doi.org/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. <https://doi.org/10.1371/journal.pone.0130324>
- Elbrecht, V., Vamos, E. E., Meissner, K., Aroviita, J., & Leese, F. (2017). Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution*, 8, 1265–1275. <https://doi.org/10.1111/2041-210X.12789>
- Ershova, E. A., Wangenstein, O. S., Descoteaux, R., Barth-Jensen, C., & Præbel, K. (2021). Metabarcoding as a quantitative tool for estimating biodiversity and relative biomass of marine zooplankton. *ICES Journal of Marine Science*, 78, 3342–3355. <https://doi.org/10.1093/icesjms/fsab171>
- Hallmann, C. A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W., Müller, A., Sumser, H., Hören, T., Goulson, D., & Kroon, H. d. (2017). More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLoS ONE*, 12, e0185809. <https://doi.org/10.1371/journal.pone.0185809>
- Ivanova, N. V., Dewaard, J. R., & Hebert, P. D. N. (2006). An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, 6, 998–1002. <https://doi.org/10.1111/j.1471-8286.2006.01428.x>
- Ji, Y., Huotari, T., Roslin, T., Schmidt, N. M., Wang, J., Yu, D. W., & Ovaskainen, O. (2020). SPIKEPIPE: A metagenomic pipeline for the accurate quantification of eukaryotic species occurrences and intraspecific abundance change using DNA barcodes or mitogenomes. *Molecular Ecology Resources*, 20, 256–267. <https://doi.org/10.1111/1755-0998.13057>
- Jordan, A., Patch, H. M., Grozinger, C. M., & Khanna, V. (2021). Economic dependence and vulnerability of United States agricultural sector on insect-mediated pollination service. *Environmental Science & Technology*, 55, 2243–2253. <https://doi.org/10.1021/acs.est.0c04786>
- Karlsson, D., Hartop, E., Forshage, M., Jaschhof, M., & Ronquist, F. (2020). The Swedish Malaise trap project: A 15 year retrospective on a countrywide insect inventory. *Biodiversity Data Journal*, 8, e47255. <https://doi.org/10.3897/BDJ.8.e47255>
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Kennedy, S. R., Prost, S., Overcast, I., Rominger, A. J., Gillespie, R. G., & Krehenwinkel, H. (2020). High-throughput sequencing for community analysis: The promise of DNA barcoding to uncover diversity, relatedness, abundances and interactions in spider communities. *Development Genes and Evolution*, 230, 185–201. <https://doi.org/10.1007/s00427-020-00652-x>
- Kirse, A., Bourlat, S. J., Langen, K., Zapke, B., & Zizka, V. M. A. (2023). Comparison of destructive and nondestructive DNA extraction methods for the metabarcoding of arthropod bulk samples. *Molecular Ecology Resources*, 23, 92–105. <https://doi.org/10.1111/1755-0998.13694>
- Liu, M., Clarke, L. J., Baker, S. C., Jordan, G. J., & Burrige, C. P. (2020). A practical guide to DNA metabarcoding for entomological ecologists. *Ecological Entomology*, 45, 373–385. <https://doi.org/10.1111/een.12831>
- Luo, M., Ji, Y., Warton, D., & Yu, D. W. (2023). Extracting abundance information from DNA-based data. *Molecular Ecology Resources*, 23, 174–189. <https://doi.org/10.1111/1755-0998.13703>
- Marquina, D., Buczek, M., Ronquist, F., & Łukasik, P. (2021). The effect of ethanol concentration on the morphological and molecular preservation of insects for biodiversity studies. *PeerJ*, 9, e10799. <https://doi.org/10.7717/peerj.10799>
- 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. *Molecular Ecology Resources*, 19, 1516–1530. <https://doi.org/10.1111/1755-0998.13071>
- Marquina, D., Roslin, T., Łukasik, P., & Ronquist, F. (2023). Evaluation of non-destructive DNA extraction protocols for insect metabarcoding: Gentler and shorter is better. *Metabarcoding and Metagenomics*, 6, e78871.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, 17, 10–12. <https://doi.org/10.14806/ej.17.1.200>
- Martinez Arbizu, P. (2020). pairwiseAdonis: Pairwise multilevel comparison using Adonis. R package version 0.4.
- Martins, F. M. S., Galhardo, M., Filipe, A. F., Teixeira, A., Pinheiro, P., Paupério, J., Alves, P. C., & Beja, P. (2019). Have the cake and eat it: Optimizing nondestructive DNA metabarcoding of macroinvertebrate samples for freshwater biomonitoring. *Molecular*

- Ecology Resources*, 19, 863–876. <https://doi.org/10.1111/1755-0998.13012>
- Martoni, F., Piper, A. M., Rodoni, B. C., & Blacket, M. J. (2022). Distinguishing bias for non-destructive insect metabarcoding. *PeerJ*, 10, e12981. <https://doi.org/10.7717/peerj.12981>
- Morinière, J., Araujo, B. C. d., Lam, A. W., Hausmann, A., Balke, M., Schmidt, S., Hendrich, L., Doczkal, D., Fartmann, B., Arvidsson, S., & Haszprunar, G. (2016). Species identification in Malaise trap samples by DNA barcoding based on NGS technologies and a scoring matrix. *PLoS ONE*, 11, e0155497. <https://doi.org/10.1371/journal.pone.0155497>
- Murray, L. M., & Schön, T. B. (2018). Automated learning with a probabilistic programming language: Birch. *Annual Reviews in Control*, 46, 29–43. <https://doi.org/10.1016/j.arcontrol.2018.10.013>
- Nichols, R. V., Vollmers, C., Newsom, L. A., Wang, Y., Heintzman, P. D., Leighton, M., Green, R. E., & Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. *Molecular Ecology Resources*, 18, 927–939. <https://doi.org/10.1111/1755-0998.12895>
- Nielsen, M., Gilbert, M. T. P., Pape, T., & Bohmann, K. (2019). A simplified DNA extraction protocol for unsorted bulk arthropod samples that maintains exoskeletal integrity. *Environmental DNA*, 1, 144–154. <https://doi.org/10.1002/edn3.16>
- Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P., O'Hara, R. B., Simpson, G., Solymos, P., Stevens, H., Szoecs, E., & Wagner, H. (2020). Vegan: Community ecology package.
- Palmer, J. M., Jusino, M. A., Banik, M. T., & Lindner, D. L. (2018). Non-biological synthetic spike-in controls and the AMPtk software pipeline improve mycobiome data. *PeerJ*, 6, e4925. <https://doi.org/10.7717/peerj.4925>
- 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. <https://doi.org/10.1111/1755-0998.12355>
- Piñol, J., Senar, M. A., & Symondson, W. O. C. (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. <https://doi.org/10.1111/mec.14776>
- Porter, T. M., Morris, D. M., Basiliko, N., Hajibabaei, M., Doucet, D., Bowman, S., Emilson, E. J. S., Emilson, C. E., Chartrand, D., Wainio-Keizer, K., Séguin, A., & Venier, L. (2019). Variations in terrestrial arthropod DNA metabarcoding methods recovers robust beta diversity but variable richness and site indicators. *Scientific Reports*, 9, 18218. <https://doi.org/10.1038/s41598-019-54532-0>
- R Core Team. (2020). R: A language and environment for statistical computing.
- Raivo, K. (2019). pheatmap: Pretty Heatmaps. R package version 1.0.12. <https://CRAN.R-project.org/package=pheatmap>
- Seibold, S., Gossner, M. M., Simons, N. K., Blüthgen, N., Müller, J., Ambarlı, D., Ammer, C., Bauhus, J., Fischer, M., Habel, J. C., Linsenmair, K. E., Nauss, T., Penone, C., Prati, D., Schall, P., Schulze, E.-D., Vogt, J., Wöllauer, S., & Weisser, W. W. (2019). Arthropod decline in grasslands and forests is associated with landscape-level drivers. *Nature*, 574, 671–674. <https://doi.org/10.1038/s41586-019-1684-3>
- Ståhls, G., Meier, R., Sandrock, C., Hauser, M., Šašić Zorić, L., Laiho, E., Aracil, A., Doderović, J., Badenhorst, R., Unadirekku, P., Mohd Adom, N. A. B., Wein, L., Richards, C., Tomberlin, J. K., Rojo, S., Veselić, S., & Parviainen, T. (2020). The puzzling mitochondrial phylogeography of the black soldier fly (*Hermetia illucens*), the commercially most important insect protein species. *BMC Evolutionary Biology*, 20, 60. <https://doi.org/10.1186/s12862-020-01627-2>
- Stork, N. E. (2018). How many species of insects and other terrestrial arthropods are there on earth? *Annual Review of Entomology*, 63, 31–45. <https://doi.org/10.1146/annurev-ento-020117-043348>
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 21, 2045–2050. <https://doi.org/10.1111/j.1365-294X.2012.05470.x>
- Thomas, A. C., Deagle, B. E., Eveson, J. P., Harsch, C. H., & Trites, A. W. (2016). Quantitative DNA metabarcoding: Improved estimates of species proportional biomass using correction factors derived from control material. *Molecular Ecology Resources*, 16, 714–726. <https://doi.org/10.1111/1755-0998.12490>
- Tourlousse, D. M., Yoshiike, S., Ohashi, A., Matsukura, S., Noda, N., & Sekiguchi, Y. (2017). Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Research*, 45, e23. <https://doi.org/10.1093/nar/gkw984>
- van Klink, R., Bowler, D. E., Gongalsky, K. B., Swengel, A. B., Gentile, A., & Chase, J. M. (2020). Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science*, 368, 417–420. <https://doi.org/10.1126/science.aax9931>
- Vesterinen, E. J., Ruokolainen, L., Wahlberg, N., Peña, C., Roslin, T., Laine, V. N., Vasko, V., Sääksjärvi, I. E., Norrdahl, K., & Lilley, T. M. (2016). What you need is what you eat? Prey selection by the bat *Myotis daubentonii*. *Molecular Ecology*, 25, 1581–1594. <https://doi.org/10.1111/mec.13564>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.
- Yang, C., Bohmann, K., Wang, X., Cai, W., Wales, N., Ding, Z., Gopalakrishnan, S., & Yu, D. W. (2021). Biodiversity soup II: A bulk-sample metabarcoding pipeline emphasizing error reduction. *Methods in Ecology and Evolution*, 12, 1252–1264. <https://doi.org/10.1111/2041-210X.13602>
- Yu, D. W., Ji, Y., Emerson, B. C., 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. <https://doi.org/10.1111/j.2041-210X.2012.00198.x>
- Zizka, V. M. A., Leese, F., Peinert, B., & Geiger, M. F. (2019). DNA metabarcoding from sample fixative as a quick and voucher-preserving biodiversity assessment method. *Genome*, 62, 122–136. <https://doi.org/10.1139/gen-2018-0048>

SUPPORTING INFORMATION

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

Text S1. Experiment I: Assessing the effects of buffer and lysis time/homogenization.

Text S2. Testing a wider range of conditions on more complex mock communities (Q3).

Text S3. Library preparation details.

Figure S1. Species richness detected in Experiment I (a) and Experiment II (b).

Figure S2. Variance and standard deviation in relative count numbers for each spike-in species.

Figure S3. Estimated posterior distributions of mean DNA yield per biomass.

Figure S4. Estimated posterior distribution for the coefficient of variance (CV) of DNA yield per biomass.

Figure S5. Predictions from Experiments III, for the mild lysis and homogenate treatment.

Table S1. Summary of reference insect species.

Table S2. Overview of species used in the mock communities in Experiment I.

Table S3. Overview of species used in the mock communities in Experiment II.

Table S4. Overview of species used in the mock communities in Experiment III.

Table S5. Summary of COI sequencing results - numbers of reads per Experiment.

How to cite this article: Iwaszkiewicz-Eggebrecht, E., Granqvist, E., Buczek, M., Prus, M., Kudlicka, J., Roslin, T., Tack, A. J. M., Andersson, A. F., Miraldo, A., Ronquist, F., & Łukasik, P. (2023). Optimizing insect metabarcoding using replicated mock communities. *Methods in Ecology and Evolution*, 14, 1130–1146. <https://doi.org/10.1111/2041-210X.14073>