

RESOURCE ARTICLE OPEN ACCESS

Balancing Barcoding and Genomics: gDNA Quality in Insect Vouchers After HotSHOT DNA Extraction

Vivian Feng^{1,2}  | Caroline Høegh-Guldberg^{3,4} | Rudolf Meier^{1,2}  | Aleš Buček³ 

¹Center for Integrative Biodiversity Discovery, Leibniz Institute for Evolution and Biodiversity Science, Museum für Naturkunde, Berlin, Germany | ²Humboldt-Universität Zu Berlin, Institut für Biologie, Berlin, Germany | ³Biology Centre of the Czech Academy of Sciences, Institute of Entomology, České Budějovice, Czech Republic | ⁴Department of Zoology, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

Correspondence: Aleš Buček (ales.bucek@entu.cas.cz) | Rudolf Meier (rudolf.meier@hu-berlin.de)

Received: 16 October 2025 | **Revised:** 19 December 2025 | **Accepted:** 20 January 2026

Keywords: barcoding | genomic DNA quality | genomics | HotSHOT | malaise trap

ABSTRACT

DNA extraction with an alkaline buffer system called ‘HotSHOT’ is widely used for barcoding because it is rapid, inexpensive, and voucher preserving, but it remains unclear whether sufficient genomic DNA (gDNA) remains in small vouchers for downstream use in genomics. We here evaluate gDNA quality and quantity before and after HotSHOT treatment of 11 insect families representing six orders. Some specimens were flash frozen immediately after collection, while others were kept for 1 week at tropical temperatures in ethanol to mimic Malaise trap conditions. Encouragingly, we show that gDNA of sufficiently high quality and quantity for genomic sequencing remained in specimens treated with HotSHOT. We also show that DNA integrity was strongly influenced by field storage, with specimens exposed to Malaise trap conditions showing such pronounced degradation that the standard HotSHOT treatment no longer significantly altered DNA quality. For control material, HotSHOT treatments involving longer exposure to high temperature led to smaller fragment lengths, with the effect apparently being influenced by the degree of specimen sclerotization. Our results thus suggest that optimised HotSHOT treatments, together with carefully controlled pre-extraction storage, preserve voucher gDNA of sufficient quality for downstream genomic analyses with both short-read and possibly even some long-read sequencing technologies. We provide protocol selection guidelines that improve voucher gDNA preservation in HotSHOT-treated samples. This is particularly important for many species which are only known from one or few specimens discovered during barcoding projects.

1 | Introduction

Bulk sampling methods such as Malaise, pitfall and light traps routinely collect thousands of specimens belonging to hundreds of species in a single event (Srivathsan et al. 2023). Such samples can be thought of as potential metagenomic libraries in which genetic diversity is largely partitioned by species because intraspecific variability is generally much smaller than interspecific variability. Because many species are singletons and only a few dominate in numbers (see Figure 1 in Meier et al. 2024), performing full DNA extractions and sequencing

for every individual is neither efficient nor desirable for answering most biological questions. It concentrates effort and budget on repeatedly sequencing common species and it creates a ‘two-voucher problem’ in that every specimen generates not only a physical voucher but also a separate DNA extract vial that must be archived, tracked and curated. This duplicates storage needs, inflates consumables and freezer costs, complicates chain-of-custody and can compromise the morphological integrity of specimens when aggressive DNA extraction techniques are used. Many scientists thus use a different workflow that starts with generating specimen-specific

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.

© 2026 The Author(s). *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

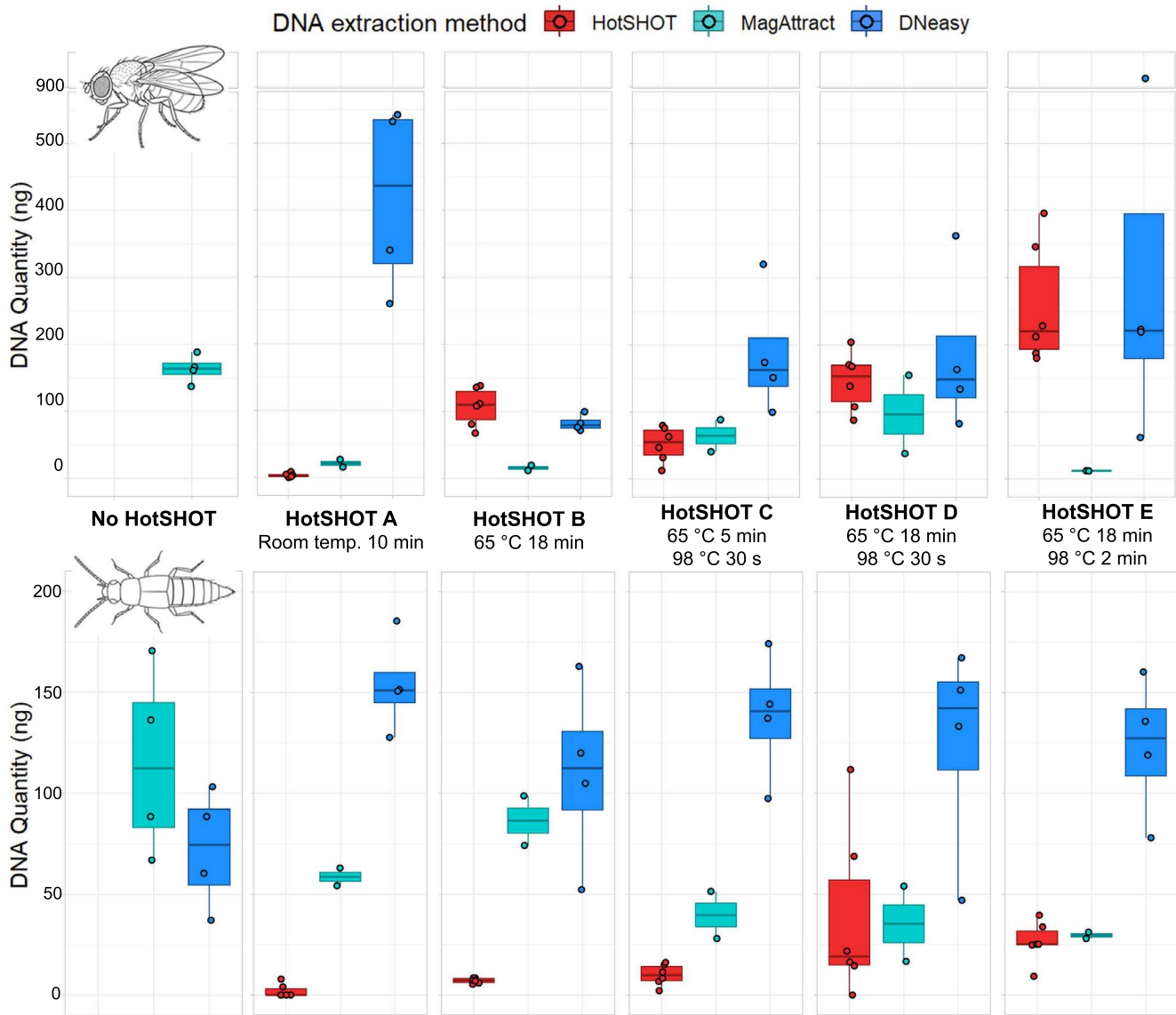


FIGURE 1 | Increasing intensity of HotSHOT treatment increases DNA quantity in HotSHOT extract but does not affect DNA quantity in subsequent genomic extractions (with MagAttract and DNeasy kits) of the HotSHOT-treated specimens.

barcodes that are then used for rapid sorting to approximately species level. In the next step, only selected specimens are subjected for targeted downstream analyses to answer specific questions. Such a 'reverse workflow' (Wang et al. 2018), where barcoding precedes morphological or genomic exploration of select specimens, has been demonstrated to be technically and financially feasible due to the wide availability of third-generation sequencing platforms (Hebert et al. 2018; Srivathsan et al. 2018, 2019, 2021, 2024). Such workflows are thus also considered central to accelerating biodiversity discovery at scale to address 'entomological dark matter' (Hartop et al. 2024; Chua et al. 2023; Brydegaard et al. 2024; Meier et al. 2024; Meier, Srivathsan, et al. 2025).

Data produced during this two-step workflow can directly or indirectly address all seven biodiversity shortfalls highlighted in Hortal et al. (2015). The Linnean shortfall of undescribed species is reduced because barcodes can be used for

species discovery in at scale and preparing them for integrative taxonomy (Hartop et al. 2022; Amorim et al. 2025; Meier, Srivathsan, et al. 2025). The Wallacean shortfall, concerning incomplete distribution knowledge, and the Prestonian shortfall, concerning missing abundance data, are also addressed because each barcoded specimen is counted and linked to a collection locality. Unfortunately, the remaining biodiversity shortfalls in Hortal et al. (2015) require more expensive genomic data (Luikart et al. 2003). The Darwinian shortfall, incomplete phylogenies, can be tackled by obtaining low-coverage genomes for barcoded vouchers as long as they yield enough phylogenetically informative markers (e.g., Call et al. 2023). The Raunkiaeran shortfall on traits can be reduced by combining morphological information with genomic data on functions such as metabolic pathways. The Hutchinsonian shortfall on abiotic tolerances can be approached by examining population genomic variation for signals of adaptation, while the Eltonian shortfall on ecological interactions can

occasionally already be addressed with the DNA obtained in the first step (Quintana et al. 2022; Srivathsan et al. 2022) but usually requires community-level genomic datasets.

A critical element for the feasibility of this two-step pipeline is that vouchers retain genomic DNA (gDNA) of sufficient quality and quantity after barcoding. This explains the popularity of techniques that just 'leaches' a small amount of DNA from the specimen. Several methods have been proposed (e.g., Margam et al. 2010; Wong et al. 2014; Thongjued et al. 2019; Wang et al. 2018), but many labs have recently started to use Hot Sodium Hydroxide and Tris (HotSHOT) because it is rapid, inexpensive, and preserves vouchers (Truett et al. 2000). HotSHOT extraction skips DNA purification and only requires two inexpensive buffers, one for leaching DNA and one for neutralisation (Srivathsan et al. 2023; Fikáček et al. 2024). Neither needs to be kept cold. Unfortunately, however, it remains unclear how much DNA remains in the specimens after HotSHOT extraction. Work on microbiomes after HotSHOT leaching has provided mixed signals (Andriienko et al. 2024): the overall DNA yield was reduced and the fragment length distribution shifted toward smaller DNA fragments. Yet, it did not appear to bias the reconstructed bacterial communities when assessed with 16S sequencing. This suggests that HotSHOT may not alter DNA quality and quantity to such an extent that it would prevent meaningful downstream genomic analyses based on short amplicon sequencing.

Here, we explicitly tested whether insect vouchers treated with HotSHOT retain gDNA of sufficient integrity for low-coverage genome sequencing. We carried out two experiments. We quantified gDNA of fresh control specimens for two species whose DNA was extracted with and without exposure to HotSHOT. In the second experiment, we asked whether DNA degradation caused by HotSHOT is more limiting than the degradation experienced while specimens are in a Malaise trap for 1 week in the tropics. To test this, we exposed a taxonomically diverse set of specimens to conditions typical of Malaise traps pitched in a tropical country. We then compared gDNA quality and quantity of specimens with and without exposure to HotSHOT. Our results serve as empirical guidelines for firstly assessing and then maximising the quantity and quality of gDNA remaining in the specimen post-extraction.

2 | Materials and Methods

2.1 | Control: Impact of HotSHOT on DNA Quality and Quantity in Control Specimens

Fresh *Drosophila melanogaster* (Diptera: Drosophilidae) from a laboratory colony and freshly collected *Myllaena dubia* (Coleoptera: Staphylinidae) from the Slavkov Forest, Czech Republic (50.054N, 12.859E) were flash frozen in liquid nitrogen and stored in absolute ethanol at -80°C prior to further analysis. The gDNA of control specimens was extracted immediately without HotSHOT (Truett et al. 2000) exposure, thus providing the baseline for expected gDNA quality and quantity. The remaining specimens were exposed to different HotSHOT treatments whereby the reagents were prepared following Truett et al. (2000) and kept consistent across all experiments.

The alkaline lysis reagent (pH 12) consists of 25 mM NaOH and 0.2 mM disodium EDTA and the neutralising reagent (pH 5) of 40 mM Tris-HCl. In the first treatment, specimens were placed in alkaline lysis reagent at room temperature for 10 min before the addition of an equal volume of neutralising reagent and subsequent gDNA extraction ('A'). The next treatment applied the standard HotSHOT treatment (65°C for 18 min plus 98°C for 2 min; 'E'), which is widely used in high throughput barcoding workflows, but we also tested three additional HotSHOT treatments with shorter exposure: treatment 'B', 65°C for 18 min; treatment 'C', 65°C for 5 min plus 98°C for 30 s; and treatment 'D', 65°C for 18 min plus 98°C for 30 s. To evaluate the effect of HotSHOT exposure, the gDNA of all specimens was extracted with either the Qiagen DNeasy Blood and Tissue Kit (non-destructive extraction, final elution into 25 μL of water; $n = 4$ biological replicates) or Qiagen MagAttract HMW DNA Kit (liquid nitrogen-frozen individuals disrupted with mortar and pestle, final elution into 20 μL of water; $n = 2-4$ biological replicates). The MagAttract Kit-based DNA extraction followed modified manufacturer's protocol with following modifications: the specimen frozen by liquid nitrogen was pulverised with plastic pestle in 1.5 mL plastic tube and 187 μL of mastermix (consisting of 100 μL PBS buffer, 10 μL of Proteinase K, 2 μL of RNase A and 75 μL of Buffer AL) was added to the still frozen pulverised sample and was incubated at room temperature for 2 h with gentle inversion of the tube every 30 min. All subsequent incubation steps except for the last elution step were performed without shaking at room temperature. The Blood and Tissue Kit-based DNA extraction followed modified manufacturer's protocol with following modifications: the specimen was submerged in 180 μL of ATL buffer and 20 μL Proteinase K mix, vortexed briefly for 15 s and incubated overnight at 56°C without shaking. This protocol did not include physical sample disintegration and homogenisation, and the extracted specimens are typically intact with occasional breakage of some appendages during the procedure.

2.2 | Malaise Trap: Impact of HotSHOT on DNA Quality and Quantity of Exposed Specimens

We next asked whether typical delays between a specimen falling into a Malaise trap collecting jar and its gDNA being extracted have a strong impact on DNA quality before and after HotSHOT treatment. To test this, specimens from nine families representing five insect orders and spanning a wide range of body sizes were collected at Kent Ridge Park, Singapore (1.287N, 103.789 E) and stored in a Malaise trap jar filled with 70% ethanol (Diptera: Sepsidae; Hemiptera: Alydidae, Cicadellidae, Pyrrhocoridae; Hymenoptera: Apidae; Lepidoptera: Lycaenidae; Orthoptera: Gryllidae). The jar was left outdoors in a partially shaded environment for 1 week to mimic tropical field conditions. After this period, half the specimens for each family were extracted directly with the Qiagen DNeasy Blood and Tissue Kit (non-destructive extraction, final elution into 50 μL of buffer AE; $n = 2-3$ biological replicates), while the other half were first subjected to the standard HotSHOT treatment ('E': 65°C for 18 min, 98°C for 2 min; alkaline lysis reagent volume 15–20 μL) and then extracted with the same kit (non-destructive extraction, final elution into 50 μL of buffer AE; $n = 2-3$ biological replicates). Entire insects were processed for most families, while only mid legs were used for the larger specimens of Apidae and Gryllidae.

2.3 | DNA Quality Assessment With Qubit and TapeStation

The DNA concentration of each extract (from HotSHOT, DNeasy and MagAttract) was measured with the Qubit fluorometer. Purified gDNA extracts (from DNeasy and MagAttract) were assessed with the Genomic DNA ScreenTape (Agilent) assay for DNA integrity, given by the DNA Integrity Number (DIN). Fragment length distributions were obtained by classifying the most common fragment sizes of each sample, given in tables under each electropherogram, into one of five size classes (more than 10 kbp, 5–10 kbp, 1–5 kbp, 300 bp–1 kbp, less than 300 bp). Estimated quantities of each fragment size group were calculated by multiplying integrated areas provided in the same table with sample DNA concentrations from Qubit. Both DINs and electropherograms were obtained via the TapeStation Analysis Software (versions A.02.01 and 5.1).

2.4 | Amplification

Because varying intensity and duration of HotSHOT treatments can have a measurable impact on the quality and concentration of the HotSHOT extract (hsDNA), we also tested for the fresh specimens whether the hsDNA still contained sufficient template DNA to yield reliable PCR products. We attempted to amplify the Folmer region (658 bp) of the COI gene from the hsDNA using indexed LCO1490 and HCO2198 primers (Folmer et al. 1994). The primers had 9 bp indices attached at the 5' end, sourced from Srivathsan et al. (2024), allowing for subsequent demultiplexing. PCR reactions consisted of 10 µL hsDNA, 9 µL PCR water, 1 µL forward and reverse primer, and 4 µL HOT FIREPol Blend Master Mix (04–25–00120–10, Solis BioDyne). Cycling conditions were: 94°C for 15 min, followed by five cycles (94°C for 30 s, 47°C for 40 s, 72°C for 1 min), 30 cycles (94°C for 30 s, 52°C for 40 s, 72°C for 1 min), and a final extension of 72°C for 7 min. Amplicons were visualised by electrophoresis on a 2% agarose gel.

2.5 | Statistical Analyses

All analyses were performed in R version 4.4.2 (R Core Team 2024) using RStudio version 2023.06.0 (Posit Team 2023). DNA yield and integrity were analysed with analysis of variance (ANOVA) models fitted using the function `aov()`. Post hoc multiple comparisons among HotSHOT treatments were carried out with Tukey's HSD test using the function `TukeyHSD()`. To summarise groupings, compact letter displays were generated with the function `multcompLetters4()` from the `multcompView` package (Graves et al. 2024), where groups sharing a letter are not significantly different. For the Malaise trap sample, DIN values were compared using Welch's two-sample *t*-test with the function `t-test()`. Figures were produced with the `ggplot2` package (Wickham 2016). Boxplots were drawn with `geom_boxplot()`, where the central line indicates the median, the box spans the interquartile range (IQR) and whiskers extend no further than 1.5 × IQR. Individual observations were overlaid as jittered points with `geom_jitter()`. For line plots, group means were added with `stat_summary(fun = mean, geom = point)` and connected with `stat_summary(fun = mean, geom = line)`. Error

bars were calculated as \pm standard error of the mean using `stat_summary(fun = mean_se, geom = errorbar)`.

3 | Results

3.1 | Impact of HotSHOT on DNA Quality and Quantity in Control Specimens

DNA extraction with HotSHOT includes a high temperature step (95°C–98°C). The original recommendation of exposure for 30 min is followed by most studies, but shorter exposure has also been shown to be effective for insects (see Table S1). Therefore, we tested one HotSHOT treatment without and four with exposures to high temperature. We find that the treatments influenced the amount of DNA in HotSHOT extract (hsDNA) of control specimens. The quantity of hsDNA was strongly dependent on the temperature protocol of HotSHOT. hsDNA yield increased with the temperature exposure during HotSHOT, with the standard treatment “E” producing the largest quantities (Figure 1) as tested in a two-way ANOVA of a linear model modelling taxon-specific differences (Amount of hsDNA ~ Treatment × Taxon). It revealed significant effects of treatment ($F_{4,50} = 26.27$, $p < 0.001$), taxon ($F_{1,50} = 102.93$, $p < 0.001$) and the interaction between treatment and taxon ($F_{4,50} = 16.93$, $p < 0.001$). Surprisingly, the amount of genomic DNA (gDNA) extractable from the HotSHOT-treated specimen for further genomic work was not significantly reduced when increasing high-temperature exposure during HotSHOT treatments (Figure 1) as tested in a two-way ANOVA (Amount of gDNA ~ Treatment × Taxon; $F_{5,60} = 1.24$, $p = 0.303$) nor the interaction ($F_{5,60} = 0.962$, $p = 0.449$) and the only significant effect was the variable “Taxon” with the amount of gDNA in the specimens being significantly higher in *D. melanogaster* than *M. dubia* ($F_{1,60} = 5.39$, $p = 0.024$). The DNeasy extraction method also produced significantly higher amounts of gDNA than the MagAttract extraction method (Amount of gDNA ~ Treatment × DNA extraction method; $F_{1,60} = 12.7$, $p < 0.001$), and the interaction between treatment and extraction method was significant ($F_{5,60} = 2.40$, $p = 0.048$) but treatment alone was not significant ($F_{5,60} = 1.43$, $p = 0.226$).

Across all treatments, *Drosophila melanogaster* consistently produced higher yields than the more heavily sclerotized *Myllaena dubia*. However, yield alone is not a reliable indicator of usable DNA, because treatments that maximise yield may also produce the most degraded extracts. This effect was confirmed in our experiments. Untreated control specimens had the highest gDNA quality as quantified by DIN, while even a brief 10-min room temperature exposure to HotSHOT caused a measurable decline in quality (Figure 2). DIN decreased further under all HotSHOT incubations, with the most widely adopted treatment ‘E’ producing the lowest DIN. These results were statistically significant with a one-way ANOVA showing that treatment had a significant effect on DIN (DIN ~ Treatment; $F_{5,66} = 15.69$, $p < 0.001$) and a post hoc Tukey HSD test showing significantly higher DIN for the control (group ‘a’) compared to all other treatments (groups ‘b’ and ‘c’) except treatment ‘B’ (group ‘ab’) and significantly lower DIN for treatment ‘E’ (group ‘c’) compared to all other treatments (Figure 2). Less sclerotised *D. melanogaster* had significantly higher DINs than more sclerotized *M. dubia* across treatments (DIN ~ Treatment × Taxon; $F_{1,60} = 5.59$, $p = 0.021$).

except for treatment ‘E’ where they exhibited a large decline in DIN (Figure 3) which suggests greater resilience to high temperature treatments for more sclerotized taxa. In the same two-way ANOVA, both treatment ($F_{5,60}=19.17$, $p<0.001$) as well as the interaction between treatment and taxon ($F_{5,60}=3.01$, $p=0.017$) were significant.

Size distributions of gDNA fragments provided a complementary perspective (Figure 4). Untreated control specimens contained substantial amounts of high molecular weight gDNA, with fragments longer than 10 kbp being abundant. Specimens

with HotSHOT treatments ‘B’, ‘C’ and ‘D’ showed a similar profile of gDNA fragments, though with greater variability in the amount of fragments longer than 10 kbp. Under the standard HotSHOT treatment ‘E’, gDNA fragments larger than 10 kbp mostly disappeared and most gDNA fragments varied in length between 1 and 5 kbp. Note, however, that despite strong differences in yield and quality, PCR amplification of the COI barcode was robust across most HotSHOT treatments (Figure 5) with amplification being successful for both taxa under all extraction treatments. However, band intensity was reduced for *M. dubia* in treatments lacking a 98°C step.

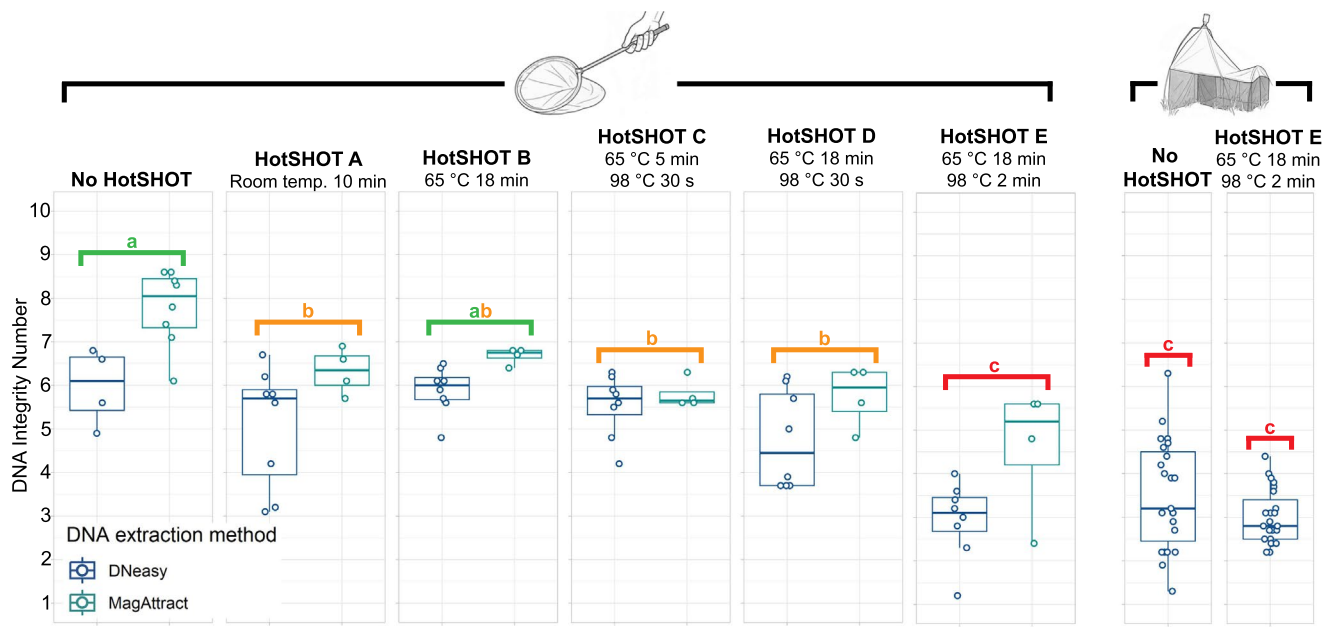


FIGURE 2 | HotSHOT treatment caused a measurable decline in DNA Integrity Number (DIN) in control samples, with untreated specimens having significantly higher DIN than all treatments except “B”. HotSHOT treatment did not cause a significant decline in DIN in the Malaise trap sample. Groups sharing a letter above the boxplot are not significantly different from each other.

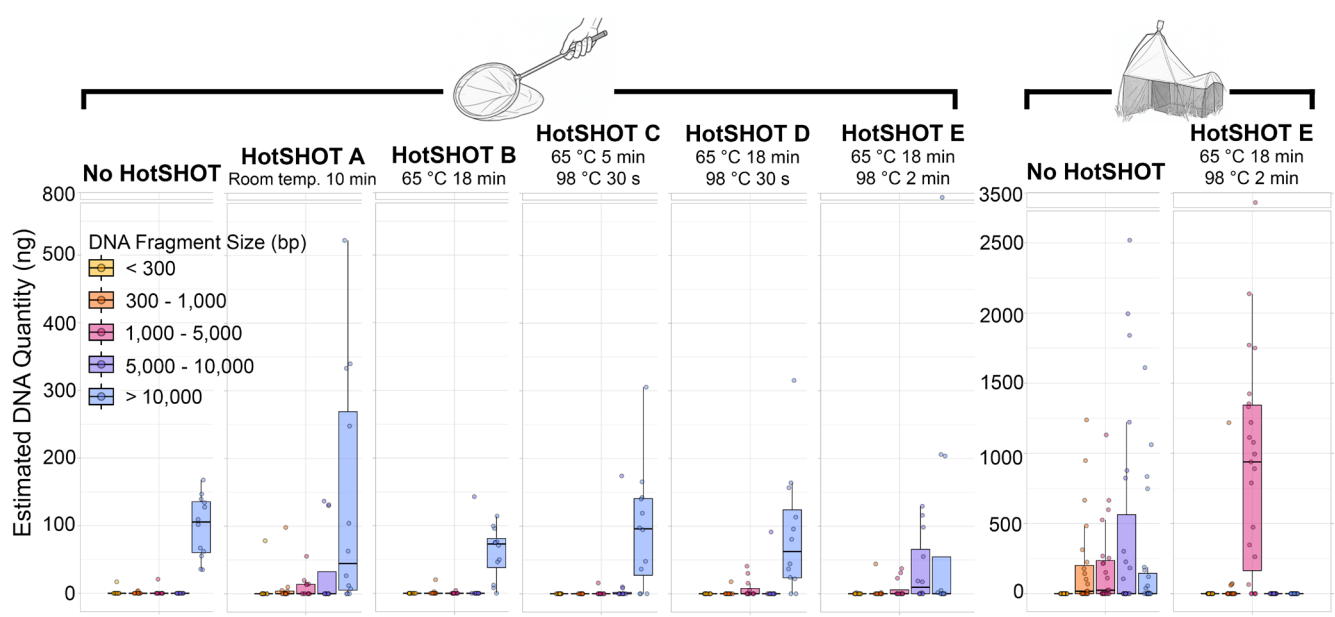


FIGURE 3 | Less sclerotized taxa (e.g., *D. melanogaster*) have a stronger response than more sclerotized taxa (e.g., *M. dubia*) to the most intense HotSHOT treatment (“E”), while taxon did not have a significant effect in the Malaise trap sample.

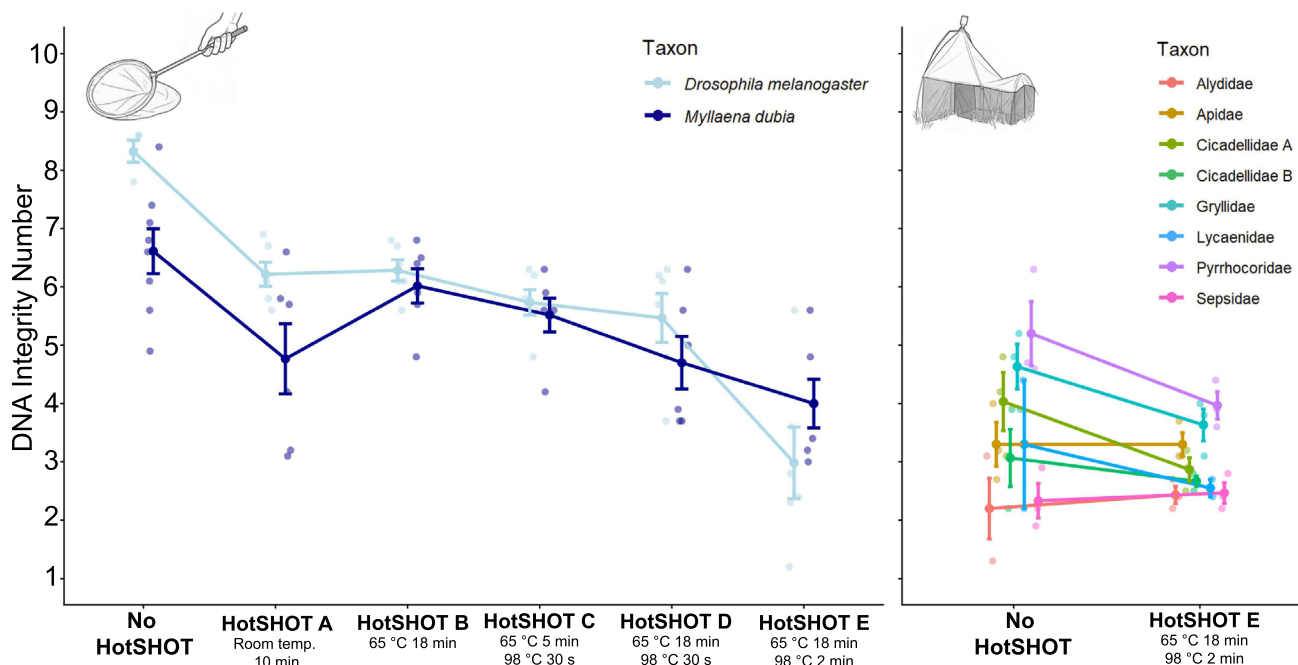


FIGURE 4 | Only HotSHOT treatment ‘E’ reduced the amount of fragments over 10 kbp in gDNA of control specimens obtained with both extraction kits. Using a Malaise trap results in more fragmented gDNA, and the combination of a Malaise trap and HotSHOT treatment “E” eliminated all remaining long fragments.

3.2 | Impact of HotSHOT on DNA Quality and Quantity of Malaise Trap Specimens

Overall, the Malaise trap experiment showed that extended exposure of specimens to field conditions was a major driver of DNA degradation, effectively being as destructive as the harshest HotSHOT treatment (roughly equivalent to treatment ‘E’ in experiment 1; Figure 2), thus leaving less scope for further decline after applying HotSHOT treatment ‘E’ because most gDNA was already fragmented to 300bp–5 kbp size range and fragments above 10 kbp were rare (Figure 4). Additional HotSHOT treatment shifted these distributions further, with treatment ‘E’ eliminating nearly all remaining long fragments and making the 1–5 kbp size class dominant (Figure 4). Accordingly, the mean DIN value for Malaise trap specimens after HotSHOT treatment declined only moderately from 3.26 to 3.00, a difference that was not significant (Welch’s two-sample *t*-test, $32.31 = 1.76$, $p = 0.088$), while experiment 1 yielded a significant decline between control specimens (DIN = 7.18) and specimens with HotSHOT treatment ‘E’ (DIN = 3.49). When we used a linear model for the taxa in the Malaise trap experiment ($\text{DIN} \sim \text{Treatment} \times \text{Taxon}$), we found in a two-way ANOVA that both treatment ($F_{1,30} = 7.13$, $p = 0.012$) and taxon ($F_{7,30} = 8.91$, $p < 0.001$) had a significant effect on DIN, but the interaction between treatment and taxon was not significant ($F_{7,30} = 1.25$, $p = 0.309$). This implied that the treatment effects were consistent across taxa (Figure 3).

4 | Discussion

Our results show that HotSHOT-treated specimens retain sufficient gDNA for downstream genomic analysis. Voucher gDNA quantities (11.78 ng–6000 ng) are well within the input

requirements of commonly used Illumina library preparation kits (e.g., Illumina DNA Prep: 1–500 ng input, QIAseq FX DNA Library Kit: 20 pg–1 µg input, NEBNext Ultra II FS DNA Library Prep Kit: 100 pg–500 ng input). In some cases, the recovered amounts also fell within the nominal range for long-read sequencing kits such as the Oxford Nanopore Ligation Sequencing Kit V14, but whether DNA exposed to HotSHOT is truly suitable for generating high-quality long-read data will require testing. Taken together, our results suggest that a two-step workflow in which HotSHOT is first used to obtain DNA barcodes for sorting into molecular operational taxonomic units (mOTUs) leaves enough gDNA of sufficient quality for genomics as long as the data are generated with short-read technologies. It thus also appears realistic to use the workflow to address not only the Linnean, Wallacean, and Prestonian but also the Darwinian, Raunkiaeran, Hutchinsonian, and Eltonian shortfalls as envisioned by recent reviews of the use of genomics in biodiversity science (Chua et al. 2023; Meier, Lawniczak, and Srivathsan 2025).

Our data confirm that gDNA integrity in insect specimens is strongly dependent on preservation and storage history. The gDNA of Malaise trap samples exposed to tropical field conditions already has heavily fragmented DNA ranging from 300 bp to 5 kbp. Additional HotSHOT treatment thus inflicts limited further damage to DNA in the > 5 kbp size range category. Similar patterns are well documented in museomics, where sequencing success and recoverable fragment length overall decline with specimen age and depend strongly on preservation method, yet low-coverage genomic sequencing has become feasible even for material that is decades to more than a century old (e.g., Strutzenberger et al. 2012; Prosser et al. 2016; Gilbert et al. 2007; Tin et al. 2014). The data suggest that a few weeks of exposure in a Malaise trap in tropical temperatures mirrors

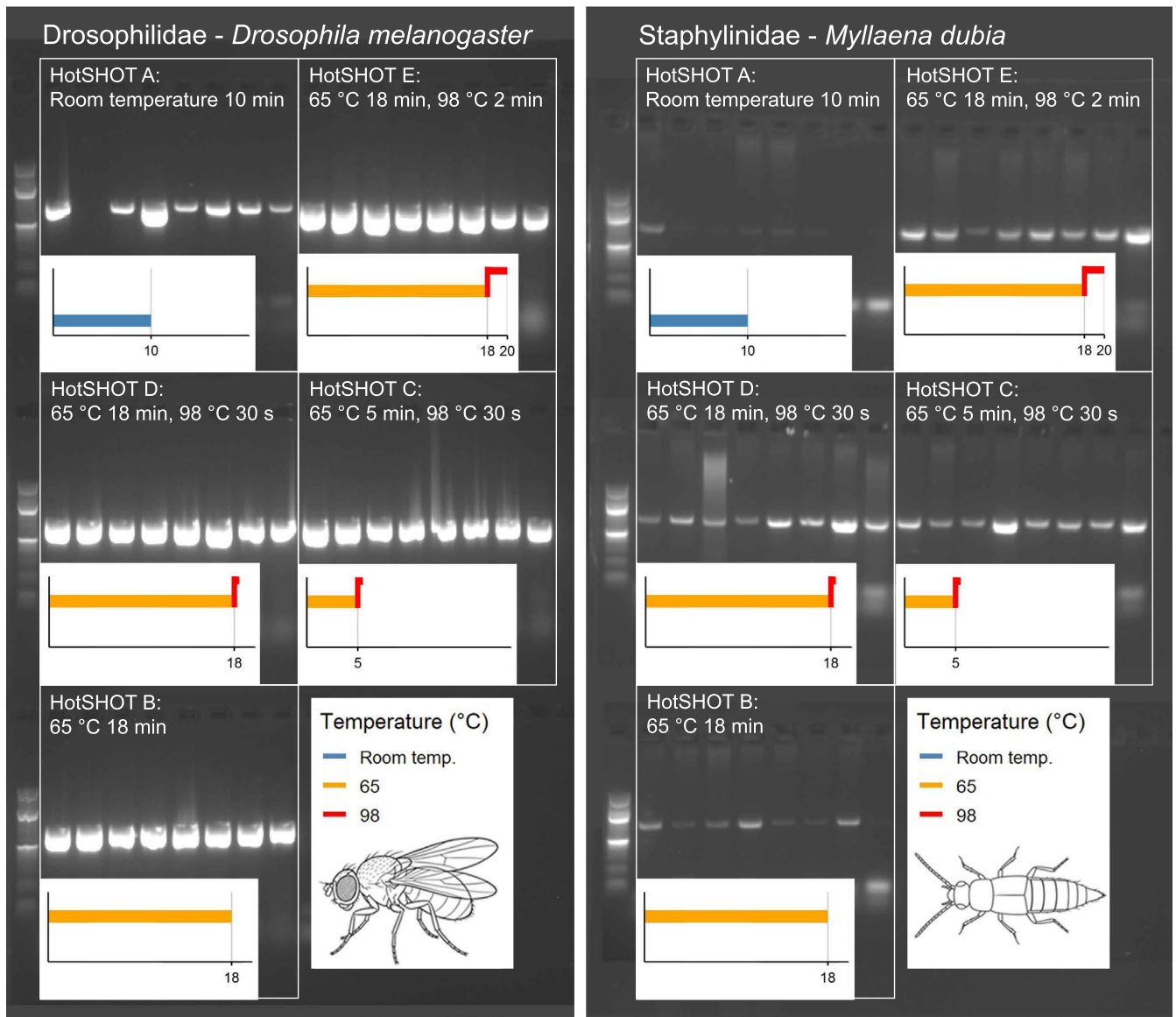


FIGURE 5 | 658bp COI amplification success was high (100% in both taxa except 'A') across all HotSHOT treatments used. Note the fainter gel electrophoresis bands for *M. dubia* specimens treated without any 98 °C step.

the gradual DNA degradation over considerably longer time observed in specimens from natural history museums. However, as in museomics, this level of degradation does not preclude short-read based genomic analysis, but it does severely limit the usability of such samples for long read-based analyses, such as *de novo* genome assembly.

Since the level of DNA degradation limits the sequencing options, two-step workflows that aim to maximise information from bulk biodiversity samples should minimise damage in the first DNA leaching step. Despite the usage of a range of temperature regimes in the published literature (see Table S1), no attempt to optimise the temperature exposure during the HotSHOT treatment was made previously. However, our results show that it matters: the harshest but also most popular treatment released the most DNA but at the cost of severely reduced DNA integrity, while milder treatments such as 'B' yielded sufficient DNA for barcoding while preserving longer fragments useful for genomic work. Comparative studies of extraction

methods also indicate that HotSHOT typically yields very low concentrations and purity compared to column-based protocols (Holmquist et al. 2025; Martincová and Aghová 2020), which is consistent with our finding that the amount of DNA removed by HotSHOT is so small that it does not preclude obtaining significant amounts of DNA in subsequent genomic extractions (see Figure 1). Even the most aggressive treatment that maximises DNA yield in HotSHOT leads only to a modest and not significant reduction in DNA remaining in the specimen. This bodes well for obtaining genomic data from voucher specimens that had been barcoded, provided that they have been stored appropriately. Storage conditions must balance the preservation of morphological and molecular integrity, as ethanol concentrations that are too high make specimens brittle and difficult to manipulate for morphological study, whereas concentrations that are too low compromise DNA quality (Marquina et al. 2021). Optimal storage therefore depends on specimen size and anticipated future uses. For large specimens, we recommend dissecting a small tissue sample (preferably without cuticle) for

long-term DNA storage under optimal conditions (frozen and/or in 96% ethanol), while preserving the remaining voucher in 70%–80% ethanol.

Our results show a significant interaction between taxon and treatment in experiment 1, with *Drosophila* yielding more and higher-integrity DNA than *Myllaena* under mild HotSHOT conditions. These differences cannot be attributed to a single factor, since extraction success is likely determined by multiple specimen properties, including the total amount of DNA available and its accessibility, for which sclerotization may be a proxy. A universal HotSHOT protocol will therefore not be optimal for all specimens. Sorting trap catches by factors such as sclerotization would make sense, but it would also be expensive unless automation is embraced. Robotic sorting systems such as the DiversityScanner, which combines high-throughput handling with convolutional neural network-based family-level classification (Wühl et al. 2022), could provide the technical foundation for automating these choices. This would include the development of predictive AI models that assign specimens to optimal treatments based on traits that explain variation in DNA yield. Such models should also consider the amount of genomic DNA remaining in the specimen to balance barcoding efficiency with preservation for downstream genomic analyses.

Overall, post-HotSHOT genomic DNA yield was apparently influenced by technical and biological factors, with substantial inherent variability limiting predictability at the level of individual specimens. Genomic DNA yield was consistently higher for non-destructive extraction with the Qiagen DNeasy Blood and Tissue kit than for the magnetic bead-based Qiagen MagAttract kit, despite the latter involving complete specimen pulverisation (Figure 1). This may indicate that the MagAttract protocol would require more optimization. Beyond this extraction method effect, we detected a modest species effect that was independent of the extraction kit: increasing heat exposure during HotSHOT led to higher post-extraction gDNA yields in *Drosophila*, whereas *Myllaena* showed little or no systematic response. However, this species-level signal was weak relative to the pronounced intraspecific variability, with gDNA yields varying widely among biological replicates and overlapping broadly between taxa. These effects may be due to specimen habitus or sclerotisation, with their impact potentially being modulated by incidental breakage increasing the ease with which DNA can be obtained.

In summary, sorting specimens into putative MOTUs based on barcodes directly addresses three major biodiversity shortfalls, namely the Linnean, Wallacean, and Prestonian shortfalls, while the remaining Darwinian, Raunkiaeran, Hutchinsonian, and Eltonian shortfalls can be addressed by genomic sequencing of vouchers. The latter could be further optimised by reducing the damage to DNA incurred in the field and by tailoring HotSHOT treatments to what taxon is being extracted. Shorter exposure of specimens to field conditions and automation provide realistic paths toward such individualised treatments. Despite the modest sample size and single field condition tested, we believe our results provide actionable evidence that HotSHOT is well suited for a two-step approach to studying bulk biodiversity samples holistically.

Author Contributions

A.B. and R.M. conceived the study. V.F. and C.H.-G. carried out specimen collection and laboratory work. V.F. performed data analyses. V.F. and R.M. wrote the manuscript with input from all authors. All authors read and approved the final version of the manuscript.

Acknowledgements

A.B. and C.H.-G. were supported by the Czech Science Foundation grant Junior STAR No. 23-08010M. C.H.-G. was also supported by grant GA JU 04-115/2025/P provided by the University of South Bohemia.

Funding

This work was supported by Grantová Agentura České Republiky, 23-08010M.

Disclosure

Benefit-Sharing Statement: Benefits from this research accrue from the sharing of our data and results on public databases as described above.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in Figshare at <https://doi.org/10.6084/m9.figshare.30265897>, reference number 10.6084/m9.figshare.30265897 (Feng et al. 2025).

References

- Amorim, D. d. S., S. S. Oliveira, M. I. P. A. Balbi, et al. 2025. "An Integrative Taxonomic Treatment of the Mycetophilidae (Diptera: Bibionomorpha) From Singapore Reveals 115 New Species on 730 km²." *bioRxiv, Preprint*. <https://doi.org/10.1101/2023.09.02.555672>.
- Andriienko, V., M. Buczek, R. Meier, A. Srivathsan, P. Łukasik, and M. R. Kolasa. 2024. "Implementing High-Throughput Insect Barcoding in Microbiome Studies: Impact of Non-Destructive DNA Extraction on Microbiome Reconstruction." *PeerJ* 12: e18025. <https://doi.org/10.7717/peerj.18025>.
- Brydegaard, M., R. D. Pedales, V. Feng, et al. 2024. "Towards Global Insect Biomonitoring With Frugal Methods." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 379: 20230103. <https://doi.org/10.1098/rstb.2023.0103>.
- Call, E., V. G. Twort, M. Espeland, and N. Wahlberg. 2023. "One Method to Sequence Them All? Comparison Between Whole-Genome Sequencing and Target Enrichment of Museum Specimens From the Moth Families Epicopeiidae and Uraniidae." *bioRxiv, Preprint*. <https://doi.org/10.1101/2023.08.21.553699>.
- Chua, P. Y. S., S. J. Bourlat, C. Ferguson, et al. 2023. "Future of DNA-Based Insect Monitoring." *Trends in Genetics* 39, no. 7: 531–541. <https://doi.org/10.1016/j.tig.2023.02.012>.
- Feng, V., C. Hoegh-Guldberg, R. Meier, and A. Buček. 2025. Datasets for "Balancing Barcoding and Genomics: gDNA Quality in Insect Vouchers After HotSHOT DNA Extraction." [dataset]. figshare. Version 4. <https://doi.org/10.6084/m9.figshare.30265897.v4>.
- Fikáček, M., F.-S. Hu, M.-H. Le, and J.-P. Huang. 2024. "Can Immature Stages Be Ignored in Studies of Forest Leaf Litter Arthropod Diversity? A Test Using Oxford Nanopore DNA Barcoding." *Insect Conservation and Diversity* 17, no. 1: 31–50. <https://doi.org/10.1111/icad.12702>.

- Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. "DNA Primers for Amplification of Mitochondrial Cytochrome c Oxidase Subunit I From Diverse Metazoan Invertebrates." *Molecular Marine Biology and Biotechnology* 3, no. 5: 294–299.
- Gilbert, M. T. P., W. Moore, L. Melchior, and M. Worobey. 2007. "DNA Extraction From Dry Museum Beetles Without Conferring External Morphological Damage." *PLoS One* 2, no. 3: e272. <https://doi.org/10.1371/journal.pone.0000272>.
- Graves, S., H. Piepho, and L. Selzer. 2024. "MultcompView: Visualizations of Paired Comparisons. R package version 0.1–10." <https://github.com/liselzer/multcompview>.
- Hartop, E., L. Lee, A. Srivathsan, et al. 2024. "Resolving Biology's Dark Matter: Species Richness, Spatiotemporal Distribution, and Community Composition of a Dark Taxon." *BMC Biology* 22: 215. <https://doi.org/10.1186/s12915-024-02010-z>.
- Hartop, E., A. Srivathsan, F. Ronquist, and R. Meier. 2022. "Towards Large-Scale Integrative Taxonomy (LIT): Resolving the Data Conundrum for Dark Taxa." *Systematic Biology* 71, no. 6: 1404–1422. <https://doi.org/10.1093/sysbio/syac033>.
- Hebert, P. D. N., T. W. A. Braukmann, S. W. J. Prosser, et al. 2018. "A Sequel to Sanger: Amplicon Sequencing That Scales." *BMC Genomics* 19, no. 1: 219. <https://doi.org/10.1186/s12864-018-4611-3>.
- Holmquist, A., H. Tavris, G. Kim, L. Esposito, B. Fisher, and A. Lam. 2025. "Towards Large-Scale Museomics Projects: A Cost-Effective and High-Throughput Extraction Method for Obtaining Historical DNA From Museum Insect Specimens." *Molecular Ecology Resources* 25: e14117. <https://doi.org/10.1111/1755-0998.14117>.
- Hortal, J., F. de Bello, J. A. F. Diniz-Filho, T. M. Lewinsohn, J. M. Lobo, and R. J. Ladle. 2015. "Seven Shortfalls That Beset Large-Scale Knowledge of Biodiversity." *Annual Review of Ecology, Evolution, and Systematics* 46: 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>.
- Luikart, G., P. England, D. Tallmon, S. Jordan, and P. Taberlet. 2003. "The Power and Promise of Population Genomics: From Genotyping to Genome Typing." *Nature Reviews Genetics* 4: 981–994. <https://doi.org/10.1038/nrg1226>.
- Margam, V. M., E. W. Gachomo, J. H. Shukle, O. O. Ariyo, M. J. Seufferheld, and S. O. Kotchoni. 2010. "A Simplified Arthropod Genomic-DNA Extraction Protocol for Polymerase Chain Reaction (PCR)-based Specimen Identification Through Barcoding." *Molecular Biology Reports* 37: 3631–3635. <https://doi.org/10.1007/s11033-010-0014-5>.
- Marquina, D., M. Buczek, F. Ronquist, and P. Łukasik. 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>.
- Martincová, I., and T. Aghová. 2020. "Comparison of 12 DNA Extraction Kits for Vertebrate Samples." *Animal Biodiversity and Conservation* 43, no. 1: 67–77. <https://doi.org/10.32800/abc.2020.43.0067>.
- Meier, R., E. Hartop, C. Pylatiuk, and A. Srivathsan. 2024. "Towards Holistic Insect Monitoring: Species Discovery, Description, Identification and Traits for All Insects." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 379: 20230120. <https://doi.org/10.1098/rstb.2023.0120>.
- Meier, R., M. K. N. Lawniczak, and A. Srivathsan. 2025. "Illuminating Entomological Dark Matter With DNA Barcodes in an Era of Insect Decline, Deep Learning, and Genomics." *Annual Review of Entomology* 70: 185–204. <https://doi.org/10.1146/annurev-ento-040124-014001>.
- Meier, R., A. Srivathsan, S. S. Oliveira, et al. 2025. "Dark taxonomy: A New Protocol for Overcoming the Taxonomic Impediments for Dark Taxa and Broadening the Taxon Base for Biodiversity Assessment." *Cladistics* 41: 1–16. <https://doi.org/10.1111/cla.12609>.
- Posit Team. 2023. "RStudio: Integrated Development Environment for R." Posit Software, PBC. <http://www.posit.co>.
- Prosser, S. W. J., J. R. deWaard, S. E. Miller, and P. D. N. Hebert. 2016. "DNA Barcodes From Century-Old Type Specimens Using Next-Generation Sequencing." *Molecular Ecology Resources* 16, no. 2: 487–497. <https://doi.org/10.1111/1755-0998.12474>.
- Quintana, M., L. de-León, J. Cubero, and F. Siverio. 2022. "Assessment of Psyllid Handling and DNA Extraction Methods in the Detection of 'Candidatus Liberibacter Solanacearum' by qPCR." *Microorganisms* 10: 1104. <https://doi.org/10.3390/microorganisms10061104>.
- R Core Team. 2024. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing. <https://www.R-project.org>.
- Srivathsan, A., Y. Ang, J. M. Heraty, et al. 2023. "Convergence of Dominance and Neglect in Flying Insect Diversity." *Nature Ecology & Evolution* 7: 1012–1021. <https://doi.org/10.1038/s41559-023-02066-0>.
- Srivathsan, A., B. Baloglu, W. Wang, et al. 2018. "A MinION-Based Pipeline for Fast and Cost-Effective DNA Barcoding." *Molecular Ecology Resources* 18: 1035–1049. <https://doi.org/10.1111/1755-0998.12890>.
- Srivathsan, A., V. Feng, D. Suárez, B. Emerson, and R. Meier. 2024. "ONTbarcoder 2.0: Rapid Species Discovery and Identification With Real-Time Barcoding Facilitated by Oxford Nanopore R10.4." *Cladistics* 40: 192–203. <https://doi.org/10.1111/cla.12566>.
- Srivathsan, A., E. Hartop, J. Puniamoorthy, et al. 2019. "Rapid, Large-Scale Species Discovery in Hyperdiverse Taxa Using 1D MinION Sequencing." *BMC Biology* 17: 96. <https://doi.org/10.1186/s12915-019-0706-9>.
- Srivathsan, A., L. Lee, K. Katoh, et al. 2021. "ONTbarcoder and MinION Barcodes Aid Biodiversity Discovery and Identification by Everyone, for Everyone." *BMC Biology* 19: 217. <https://doi.org/10.1186/s12915-021-01141-x>.
- Srivathsan, A., R. K. Loh, E. J. Ong, et al. 2022. "Network Analysis With Either Illumina or MinION Reveals That Detecting Vertebrate Species Requires Metabarcoding of iDNA From a Diverse Fly Community." *Molecular Ecology* 31: 1–18. <https://doi.org/10.1111/mec.16767>.
- Strutzenberger, P., G. Brehm, and K. Fiedler. 2012. "DNA Barcoding-Based Species Delimitation Increases Species Count of Eois (Geometridae) Moths in a Well-Studied Tropical Mountain Forest." *BMC Evolutionary Biology* 12: 82. <https://doi.org/10.1186/1471-2148-12-82>.
- Thongjued, K., W. Chotigeat, S. Bumrungsri, P. Thanakiatkrai, and T. Kitpipit. 2019. "A New Cost-Effective and Fast Direct PCR Protocol for Insects Based on PBS Buffer." *Molecular Ecology Resources* 19, no. 3: 691–701. <https://doi.org/10.1111/1755-0998.12998>.
- Tin, M. M. Y., E. P. Economo, and A. S. Mikheyev. 2014. "Sequencing Degraded DNA From Non-Destructively Sampled Museum Specimens for RAD-Tagging and Low-Coverage Shotgun Phylogenetics." *PLoS One* 9, no. 5: e96793. <https://doi.org/10.1371/journal.pone.0096793>.
- Truett, G. E., P. Heeger, R. L. Mynatt, A. A. Truett, J. A. Walker, and M. L. Warman. 2000. "Preparation of PCR-Quality Mouse Genomic DNA With Hot Sodium Hydroxide and Tris (HotSHOT)." *Biotechniques* 29, no. 1: 52–54. <https://doi.org/10.2144/00291bm09>.
- Wang, W. Y., A. Srivathsan, M. Foo, S. K. Yamane, and R. Meier. 2018. "Sorting Specimen-Rich Invertebrate Samples With Cost-Effective NGS Barcodes: Validating a Reverse Workflow for Specimen Processing." *Molecular Ecology Resources* 18: 490–501. <https://doi.org/10.1111/1755-0998.12751>.
- Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wong, W. H., Y. C. Tay, J. Puniamoorthy, M. Balke, P. S. Cranston, and R. Meier. 2014. "'Direct PCR' Optimization Yields a Rapid, Cost-Effective, Nondestructive and Efficient Method for Obtaining DNA

Barcodes Without DNA Extraction.” *Molecular Ecology Resources* 14, no. 6: 1271–1280. <https://doi.org/10.1111/1755-0998.12275>.

Wührl, L., C. Pylatiuk, M. Giersch, et al. 2022. “DiversityScanner: Robotic Handling of Small Invertebrates With Machine Learning Methods.” *Molecular Ecology Resources* 22, no. 4: 1626–1638. <https://doi.org/10.1111/1755-0998.13567>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** Overview of peer-reviewed studies published from 2006 to 2026 employing the HotSHOT DNA extraction protocol. For each reference, the column ‘HotSHOT protocol used’ indicates whether the original protocol (Standard: 95°C for 30min) or a modified version was applied. This classification is based on the methodological descriptions provided by the authors, which are reproduced verbatim in the ‘Verbatim protocol’ column. **Dataset: S1** Supporting Information.