German-wide Malaise trap metabarcoding estimates over 33,000 insect species

Dominik Buchner[1], Arne Beermann[1,11], Thomas Hörren[2], Julian Enss[2], Mark Frenzel[3], Yuanheng Li[1], Jörg Müller[4,5], Steffen U. Pauls[6,7,8], Martin Sorg[2], LTER-D consortium, Peter Haase [9,10,11$], Florian Leese[1,11$]

1 Aquatic Ecosystem Research, University of Duisburg Essen, Universitätsstraße 5, 45141 Essen, Germany

2 Entomological Society Krefeld, Magdeburger Straße 38-40, 47800 Krefeld, Germany

3 Helmholtz Centre for Environmental Research – UFZ, Department of Community Ecology, Theodor-Lieser-Str. 4, 06120 Halle, Germany

4 Field Station Fabrikschleichach, Department of Animal Ecology and Tropical Biology, Julius-Maximilians-Universität Würzburg, 97070 Würzburg, Germany

5 Bavarian Forest National Park, 94481 Grafenau, Germany

6 Senckenberg Research Institute and Natural History Museum Frankfurt, Senckenberganlage 25, 60325 Frankfurt am Main, Germany

7 LOEWE Centre for Translational Biodiversity Genomics, Senckenberganlage 25, 60325 Frankfurt am Main, Germany

8 Institute for Insect Biotechnology, Justus-Liebig-University Gießen, Ludwigstraße 23, 35390 Gießen, Germany

9 Faculty of Biology, University of Duisburg Essen, Universitätsstraße 5, 45141 Essen, Germany

10 Senckenberg Research Institute and Natural History Museum Frankfurt, Clamecystraße 12, 63571 Gelnhausen, Germany

11 Centre for Water and Environmental Research, Universitätsstraße 2, 45141 Essen, Germany

$ equally contributing senior authors

Abstract

Insects are unique in terms of their high species diversity and deliver key ecological functions. Despite this importance, we know little about true species numbers and biodiversity trends for most insect orders. A key limitation is a lack of time, funding and taxonomic expertise needed to identify the huge number of often small species, of which many remain undescribed as “dark taxa”. We here present a holistic, scalable and affordable insect diversity monitoring approach using DNA metabarcoding of the mitochondrial COI gene. Using data from the German Malaise trap program we analyzed 1,815 samples obtained from 2019 and 2020 through 75 traps across Germany. We uncovered 10,803 plausible insect species of which the majority (83.9%) was represented by a single OTU. We estimated another 22,500 potential insect species, which lack reference data or represent undescribed species. The overall number of >33,000 insect species reported here is almost as high as the total number of insect species known for Germany (~35,500). As Malaise traps capture only a certain fraction of insect species, we argue that many species recorded here are unknown from Germany or new to science in general. Our approach uses robotics and replicated analysis with on average 1.4 M sequence reads per sample for less than 50 € including supply, labor and maintenance. Our up-scalable analysis of 141 samples within two weeks with one person provides a blueprint for large-scale insect monitoring in almost real time.

Introduction

Global biodiversity and associated ecosystem services are changing rapidly in the Anthropocene (Pereira et al. 2012). A key organism group in this respect are insects, because they provide essential ecosystem processes and services, such as pollination and decomposition (Cardoso et al. 2020). Accordingly, intact insect communities bring immense ecological and economic benefits to humanity (Losey and Vaughan 2006). Yet, despite their importance, we still know little about the ecology and distribution of insect species, trends in insect diversity and cannot even present the overall number of extant insect species. To date, about 1 million insect species have been described, but studies estimate that about 5.5 M species (Stork 2018) or even more exist (IPBES 2019). The vast majority of insect species remains unknown. Even for extremely well-studied and generally species-poor countries with long entomological history, like Sweden or Germany, recent studies report large amounts of new insect species. From 165,000 identified specimens, caught with Malaise traps in Sweden, Karlsson et al. (2020) reported almost 700 insect species new to science and ~1,300 new to Sweden. In an analysis of about 62,000 specimens from German Malaise traps, Chimeno et al. (2022) estimated about 2,000 new dipteran species, raising the total number of so far 33,341 known species from Germany (Klausnitzer 2005) to approximately 35,500. These examples show how shallow our current understanding of insect diversity really is.

A key constraint to closing this knowledge gap is the low and globally declining number of taxonomic experts. Malaise traps often collect over 10,000 specimens over a 2-week period (Karlsson et al. 2020). Many of these specimens are small and identification even difficult for experts, resulting in much greater times required for identification than collection. While national and international networks, such as the Global Malaise Trap program (Geiger et al. 2016), BioScan (Hobern 2021), LifePlan (www.helsinki.fi/en/projects/lifeplan) or the Swedish Malaise Trap program (Karlsson et al. 2020) have now initiated large-scale samplings of insects using standardized traps and associated analysis procedures (Hallmann et al. 2017), classical taxonomic quantitative analysis of these samples is not feasible at the rate of collection. Therefore, novel methods are needed to efficiently and reliably assess insect species diversity (van Klink et al. 2020, Chua et al. 2023). Two key innovations for insect biodiversity monitoring are high-throughput image-based recognition and DNA metabarcoding. Image data have the advantage that the specimens are assessed digitally and that the physical specimens remain intact. This means morphological features as well as sample abundance and biomass data can additionally be inferred retrospectively (Besson et al. 2022). But image-based insect species diversity assessments are still in early phases of development and currently do not allow for the technical throughput and taxonomic resolution needed. DNA metabarcoding, on the other hand, is largely qualitative and requires destructive sampling of the specimens. However, following more than a decade of research and trial applications, the technology has reached a high technical readiness level. International projects have created an accessible platform for insect barcode reference data (Ratnasingham and Hebert 2007). Suitable primer pairs (Braukmann et al. 2019, Elbrecht et al. 2019) and a range of suitable protocols for field data collection, lab processing and data analysis exist (Buchner et al. 2021c, Montgomery et al. 2021). The undisputed strength of DNA

metabarcoding is the rapid delivery of taxonomically highly resolved taxa lists. Despite these known strengths, DNA metabarcoding has not been implemented as part of nation-wide, long-term and high-resolution insect-biodiversity monitoring programs. One aspect for this is that the technique's throughput is still limited by significant manual labor, which increases processing time and associated costs. However, incomplete and partly inconsistent reference databases remain a key challenge, limiting the accuracy and quantity of species-level assignments and thus the scope of species lists produced by DNA metabarcoding.

While the possibility to use DNA metabarcoding to assess insect species diversity has greatly improved in many regions and for many taxonomic groups, problems remain most obvious for taxonomically difficult and species-rich groups like Diptera and Hymenoptera. The unknown (=undescribed) species in these poorly known groups are often referred to as "dark taxa" (Hartop 2021). However, in the absence of reference barcodes for species, or even the lack of formal species descriptions for many "dark taxa", DNA metabarcoding can use genetic distance thresholds to approximate entities that roughly reflect species diversity such as Operational Taxonomic Units (OTUs), or Barcode Index Numbers (BINs, Ratnasingham and Hebert 2013). Even in the absence of species names, these distance-based entities can approximate species numbers and they have been applied in ecological and ecotoxicological research for many years (Sturmbauer et al. 1999, Hoppeler et al. 2016, Beermann et al. 2018). Interestingly, no study has hitherto quantified the number of possible "dark taxa" from DNA metabarcoding at larger scale. Other larger-scale insect DNA-metabarcoding studies in Germany either have a regional (Geiger et al. 2016, Uhler et al. 2021) or taxonomic focus (Chimeno et al. 2022). These studies thus cannot benchmark the suitability of metabarcoding as a method for large-scale insect biodiversity monitoring in terms of time and costs (Karlsson et al. 2020).

Our study fills this gap by analyzing a large-scale sampling from the German Malaise trap monitoring program (Welti et al. 2022) and consists of two major elements. 1) We present and benchmark a reproducible DNA metabarcoding workflow that can be applied at large scale and provide related cost estimates. We combine the workflow with a multi-level species validation procedure to quantify the validity of species records. The underlying dataset consists of 1,815 Malaise trap samples collected from 2019 and 2020 from 75 individual traps. 2) Knowing that the data come from one of the taxonomically best-accessed countries of the world, we then assessed, how many of the currently estimated 35,500 German insect species we find. By calibrating genetically defined OTUs to known morphospecies, we subsequently estimate the number of species per insect order to quantify the number of "dark taxa" in the data set to extrapolate the known and unknown insect diversity of the German fauna. Based on the integrated analysis of molecular ecologists, traditional taxonomists and biodiversity researchers, we derive recommendations for future insect biodiversity monitoring.

Material and Methods

Sampling

Sampling was conducted as part of the German Malaise trap program (Welti et al. 2022, https://www.ufz.de/lter-d/index.php?de=46285). This program comprises 31 sites across Germany. The sites are mainly situated in Long-Term Ecological Research (LTER, Mirtl et al. 2018) sites as well as in national parks and biosphere reserves. At each site 1-6 Malaise traps are operated since early 2019 to explore and monitor the biomass and species composition of flying insects in a wide range of habitats. In this study, we used a total of 1,815 samples from 63 Malaise traps in 2019 and 75 Malaise traps in 2020. The locations of the traps are shown in Figure 1. Traps were emptied every two weeks from the beginning of April until the end of October in both years (approx. 15 samples per year). Insects were caught in 80% Ethanol and biomass was measured following Welti et al. (2022). Samples were stored in 96% Ethanol protected from light until genetic analysis.

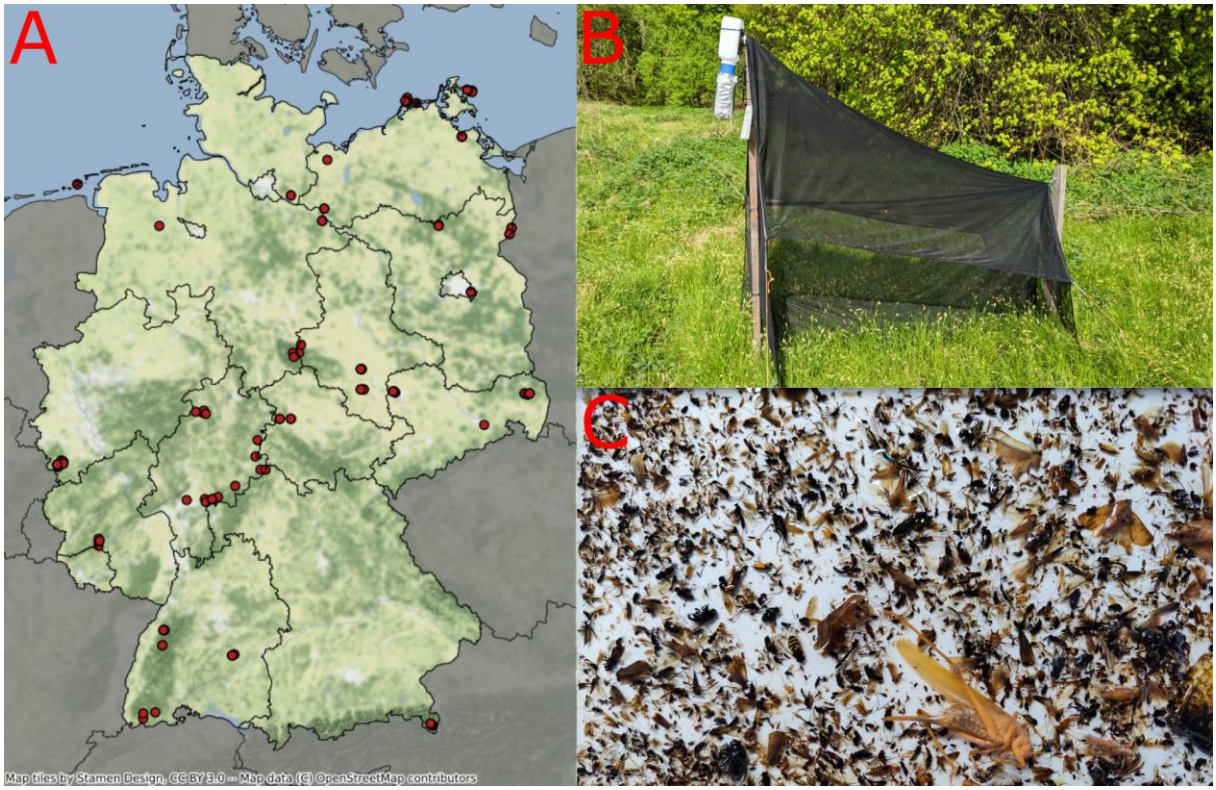

Figure 1: A: Locations of the Malaise traps used in this study**, B:** Exemplary picture of one of the traps**, C:** Exemplary picture of a sample.

Laboratory processing

Before DNA extraction, all samples were divided into two size classes (small ≤ 4 mm; large > 4 mm) by stirring the samples at 750 RPM in ethanol on top of a perforated plate sieve (4 mm diameter, Supplementary Figure S1). The smaller individuals sink to the bottom section while larger individuals stay on top of the sieve.

The size fractions were following the protocol described in Buchner et al. (2021a), with the modification that the homogenization time was reduced to 30 s. The two size

fractions were then pooled in a 1:5 ratio (large: small, 200 µL: 800 µL) as suggested by Elbrecht et al. (2021) to increase taxon recovery of smaller size insects.

DNA Extraction

Generally, the laboratory steps followed the workflow outlined in Buchner et al. (2021b). All protocols used in this study are available as step-by-step protocols in the protocols.io repository (Buchner 2022b). All subsequent steps were completed on a Biomek FX^P liquid handling workstation (Beckman Coulter, Brea, CA, USA). After sample lysis (Buchner 2022e) each sample was replicated once in separate plates and kept separate for the entire library preparation to control for possible cross-contamination. DNA was extracted using a magnetic bead protocol (Buchner 2022a). Extraction success was verified on a 1% agarose gel. For all samples that did not amplify, the extraction was repeated with silica spin columns (Buchner 2022a).

Sequencing library preparation

The PCR for the metabarcoding library was done in a two-step PCR protocol (Zizka et al. 2019) targeting a 205 bp fragment (Vamos et al. 2017) of the cytochrome oxidase I subunit (COI). Samples were amplified in the 1st PCR with the Qiagen Multiplex Plus Kit (Qiagen, Hilden, Germany) with a final concentration of 1x Multiplex Mastermix, 200 nM of each primer (fwh2F, fwhR2n, Vamos et al. 2017) and 1 µL of DNA filled up to a final volume of 10 µL with PCR-grade water. The amplification protocol was 5 min of initial denaturation at 95 °C, 20 cycles of 30 s denaturation at 95 °C, 90 s of annealing at 58 °C and 30 s of extension at 72 °C finished by a final elongation step of 10 min at 68 °C. Each of the PCR plates used in the first step was tagged with a unique combination of inline tags. Additionally, the primers contain a universal binding site for the 2nd step PCR primer to anneal (Supplementary. Table S1).

PCR products were purified using a bead-based protocol and a ratio of 0.8x to remove remaining primers and potential primer dimers (Buchner 2022c).

In the 2nd PCR samples were amplified with a final concentration of 1x Multiplex Mastermix, 100 nM of each primer (Supplementary Table S1), 1x Corralload Loading Dye, and 2 µL of cleaned-up 1st PCR product. The amplification protocol was 5 min of initial denaturation at 95 °C, 25 cycles of 30 s denaturation at 95 °C, 90 s of annealing at 61 °C and 30 s of extension at 72 °C finished by a final elongation step of 10 min at 68 °C. PCR success was visualized on a 1% agarose gel.

To achieve a similar sequencing depth, the PCR products were normalized to equal concentrations. Normalization was achieved with a bead-based protocol and a ratio of 0.7x (Buchner 2022d) and an elution volume of 50 µL. The whole volume of the normalized samples was then pooled into the final libraries. Libraries were then concentrated using a silica spin-column protocol (Buchner 2022a). Library concentrations were quantified on a Fragment Analyzer (High Sensitivity NGS Fragment Analysis Kit; Advanced Analytical, Ankeny, USA). The libraries were then sequenced using the HiSeq X platform with a paired-end (2x150 bp, 15 lanes) kit at Macrogen Europe and using the MiSeq V2 platform (2x150 bp, 1 lane) at CeGaT (Tübingen, Germany).

Bioinformatics

Raw data of the sequencing run were delivered demultiplexed by index reads. Since no differences were detected between the sequencing runs, all of them were pooled before subsequent analysis. Further demultiplexing of the inline tags was done with the python module "demultiplexer" (v1.1.0, https://github.com/DominikBuchner/demultiplexer). Reads were then further processed with the APSCALE pipeline (v1.4.0, https://github.com/DominikBuchner/apscale, Buchner et al. 2022) with default settings. First, paired-end reads were merged using vsearch (v2.21.1, Rognes et al. 2016). Primer sequences were trimmed using cutadapt (v3.5, Martin 2011). Only reads with a length of 205 bp (±10) and a maximum expected error of 1 were retained. Before OTU clustering using a similarity threshold of 97% reads were dereplicated excluding reads with an abundance below 4. For subsequent mapping of reads to OTUs all reads, including singletons, were used. The OTU table was filtered for erroneous OTUs with the LULU algorithm (Frøslev et al. 2017) as implemented in APSCALE.

Taxonomic assignment was carried out using BOLDigger (v1.5.4, https://github.com/DominikBuchner/BOLDigger, Buchner and Leese 2020). The best hit was determined with the BOLDigger method and the API verification method. This resulted in a raw OTU table (Supplementary Table S2) that was used in subsequent analysis.

Data filtering

To control for possible contamination that happened in the laboratory workflow, the technical replicates of each sample, as well as the negative controls, were merged by summing up the reads if reads were present in both replicates. After that, the maximum number of reads per OTU present in all of the negative controls was subtracted from the respective OTU (Supplementary Table S3). All OTUs were analyzed for stop-codons, and any OTUs that contained stop-codons were removed. We analyzed two datasets: 1) OTUs that were be assigned at species level and 2) OTUs that were assigned to insects. To further clean dataset 1, all OTUs with identical species assignments were merged by summing up their reads. Retrieved species names containing numbers or punctuation marks were also removed (e.g., incomplete database records). This resulted in a final species list (Supplementary Table S4) for all samples that was used for plausibility checking.

Plausibility checking of taxonomic assignment

To check the resulting species list for plausibility, three different approaches were used: a) all species were checked for plausibility by taxonomic experts from the Entomological Society Krefeld, Germany, b) for all detected species we checked whether there are GBIF records within a 200 km radius around the given trap. To do so, a polygon was drawn around all trap locations with occurrences of the respective species, and the border of this polygon was then widened by 200 km. Records were extracted using the "rgbif" package (Chamberlain et al. 2022). c) all species were checked against the German Barcode of Life database which was downloaded with a custom python script (Supplementary Script S1).

A record was accepted as plausible if two out of the three plausibility checks were true. This resulted in a final species list (Supplementary Table S5) that was used for the final analysis.

Statistical analysis

To test if sequencing depth was sufficient across all samples we performed a read-based rarefaction analysis for all samples with a custom python script. Reads were randomly drawn without replacement from a sample in 0.1% steps (50 iterations per step) of the total read count. To the resulting rarefaction curve, a Michaelis-Menten equation was fitted and the slope at maximum richness derived. The derivation of the resulting function at maximum richness was computed to determine the slope of the function at its maximum. If by doubling the sequencing depth the recovered richness increased by less than 5%, sequencing depth was considered sufficient (Supplementary Figure S2, Supplementary Table S6).

To estimate total insect OTU richness as well as total insect species in the dataset the iChao2 (Chiu et al. 2014) estimator was computed across the whole dataset by randomly drawing subsets from the dataset without replacement.

All figures were created using the python package “seaborn” (Waskom et al. 2020) with custom python scripts.

Time and cost estimations

To estimate the time and costs needed for a nationwide insect diversity assessment via metabarcoding we used vendor list prices as well as runtimes on the robot plus estimates for laboratory set up before and after the individual working steps. Costs include all labware and chemicals needed to complete the respective step of the protocol. Incubation times are not included in the time estimates since they can be used to process other samples. The per sample price is based on 1,815 samples including replicates and negative controls resulting in a total number of 4,149 individual reactions spread across 44 96-well PCR plates. A total of 10 blenders were used for homogenization which are included in the cost estimate. For the sequencing, cost estimates are based on the current costs of 1100 € for 110 Gb output using the NovaSeq 6000 S4 at a commercial service provider (Macrogen Europe) and a sequencing depth of at least 1.5 Mio reads per sample (~ 817 Gb). Labour costs were estimated at 60.35 € per hour, which corresponds to the hourly wage of a permanent scientist with the highest level of experience (TV-L E13, level 6) in Germany. The costs for the use of the pipetting robot were estimated based on linear depreciation and a total expected life-time of 10 years. Annual maintenance costs of 40,000 € incl. taxa were assumed, resulting in total depreciation and maintenance costs of approximately 4,000 € for this this study.

Results

Sequencing results

Overall, we sequenced 1,815 insect samples (775 for 2019 and 1,040 for 2020; Table 1). Based on all samples, all sequencing runs combined yielded 3,999,082,169

demultiplexed read pairs. Average read number per sample in the final read table was 1,401,469 reads (SD ± 618,631). Sequencing depth was sufficient for all samples with a mean increase of richness of 0.34% by doubling the sequencing depth (0% – 2.96%).

Sequencing yielded a total of 52,981 raw OTUs of which 50,087 OTUs were assigned to insects in 2019 and 2020. For 15,042 out of the 50,087 OTUs we were able to assign in total 11,776 insect species names, and 10,803 of these were considered plausible, leaving a total of 35,045 insect OTUs that could not be assigned to species names, either because they lack species-level reference data or are hitherto undiscovered species, i.e., dark taxa (Table 1). This leads to the high discrepancy between reported OTUs and species. While in general several OTUs can be assigned to the same species, for the majority of our plausible species (9,061 = 83.9%), only one OTU per species was recorded (Supplementary Table S7).

Table 1: Summary statistics for the dataset divided by years of sampling. A species was accepted as plausible if 2 out of the 3 plausibility checks were true.

	2019 (n=775)	**2020 (n=1,040)**	**Σ (n=1,815)**
Raw OTUs	41,189	50,166	52,981
Insect OTUs	39,367	47,652	50,087
Insect species	9,728	11,183	11,776
Plausible according to:			
Expert validation	9,132	10,475	11,030
GBIF validation	7,969	8,867	9,254
GBOL validation	8,777	10,056	10,574
Plausible	**8,983**	**10,276**	**10,803**

Species richness estimation

For a total of 50,087 observed OTUs the iChao2 estimator suggested 4,725 additional undetected OTUs (+9.4%), while it suggested 934 additional undetected plausible species for the 10,803 observed species (+8.6%). Most of the OTUs were found in both sampling years (36,932 = 73.7%) with 2020 having more exclusive insect OTUs (10,720) compared to 2019 (2,435). The same is true for the plausible insect species. The majority of species were found in both years (8,456 = 78.3%) with the year 2020 having a higher number of exclusive species (1,820) than 2019 (527).

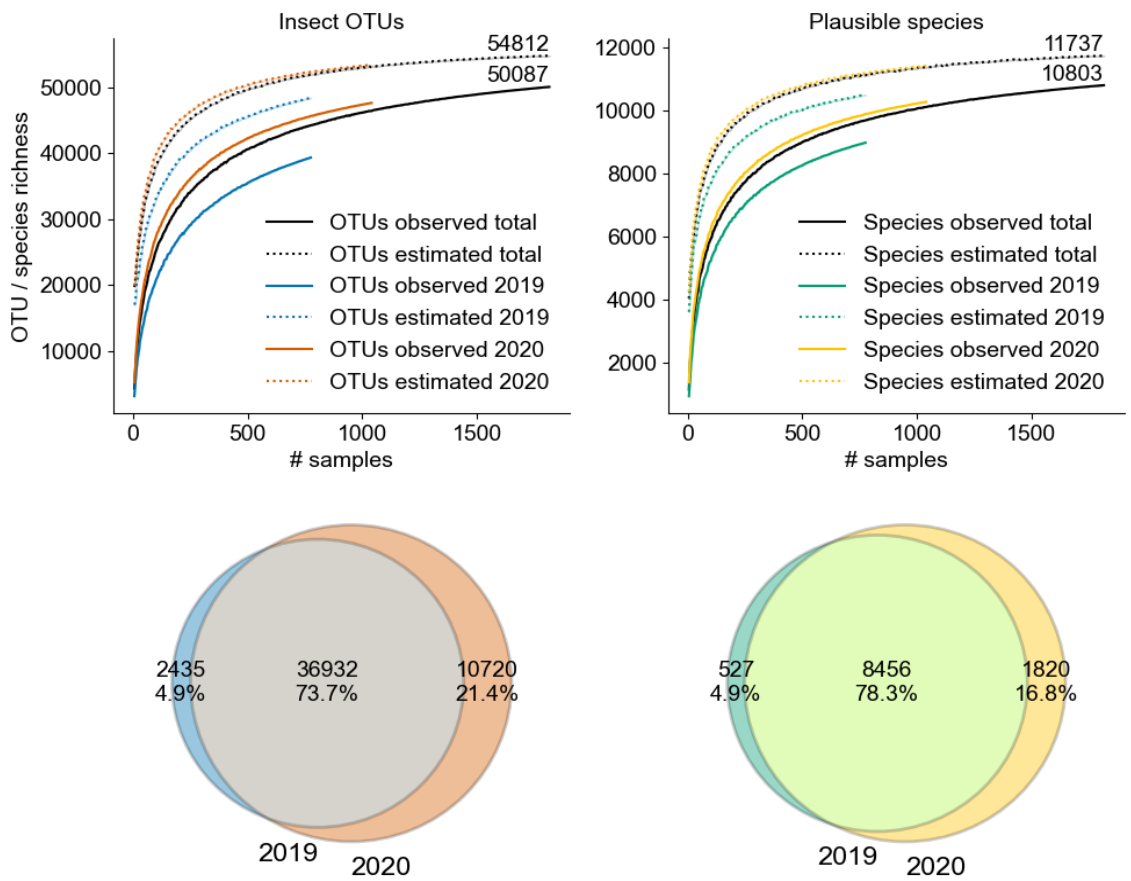

Figure 2: Top row: Insect OTU and plausible insect species richness rarefaction curves for observed (straight lines) and estimated richness (dotted lines). Both sampling years (colored lines) as well as total richness (black lines) are shown. **Bottom row:** Absolute and relative overlap between the two years of the sampling campaign for insect OTUs (left circles) and plausible insect species (right circles).

OTU distribution across insect orders and unknown taxa

All of the four most diverse insect orders in Germany (Diptera, Hymenoptera, Lepidoptera, and Coleoptera) are well represented in the dataset (Figure 3, top row) with Diptera and Hymenoptera being the most common on both the OTU (22,732 = 45.4% and 12,823 = 25.6%) and species level (3,851 = 35.6% and 2,370 = 21.9%). We detected 10,803 insect species, which represent 33% of the 32,257 insect species known for the orders found in this study (Klausnitzer 2005). Those 10,803 species represent 83% of the 13,076 barcoded species from Germany (Supplementary Table S8).

The percentage of OTUs with species-level assignment differs among those groups and is highest for the Lepidoptera (75%) followed by Coleoptera (59%). For OTUs of Diptera and Hymenoptera, less than a fifth was assigned at species-level (17% and 18%, respectively). The lowest percentage of OTUs assigned to species-level is in Orthoptera (<1%) (Figure 3, bottom left panel).

The mean number of OTUs per plausible insect species lies between 1 and 1.5 for most of the groups, while Orthoptera (~4 OTUs per species) and Zygentoma (~2 OTUs per species) are notable exceptions (Figure 3, bottom right panel).

Applying the inferred correction factors per insect order to the ~43.500 OTUs assigned to order level (excluding orthopterans) and subtracting plausible species from the corrected counts we estimate another ~ 22.500 (Supplementary Table S9) insect species in our data set that either a) could not yet be assigned at species-level with the present reference data base, b) have previously not been recorded for Germany, or c) represent new species. For the order Diptera we estimate ~4,000 species new to science, while for all other orders missing reference sequences are hindering species assignment. For this estimate all OTUs belonging to the order Orthoptera were removed to not artificially inflate the number.

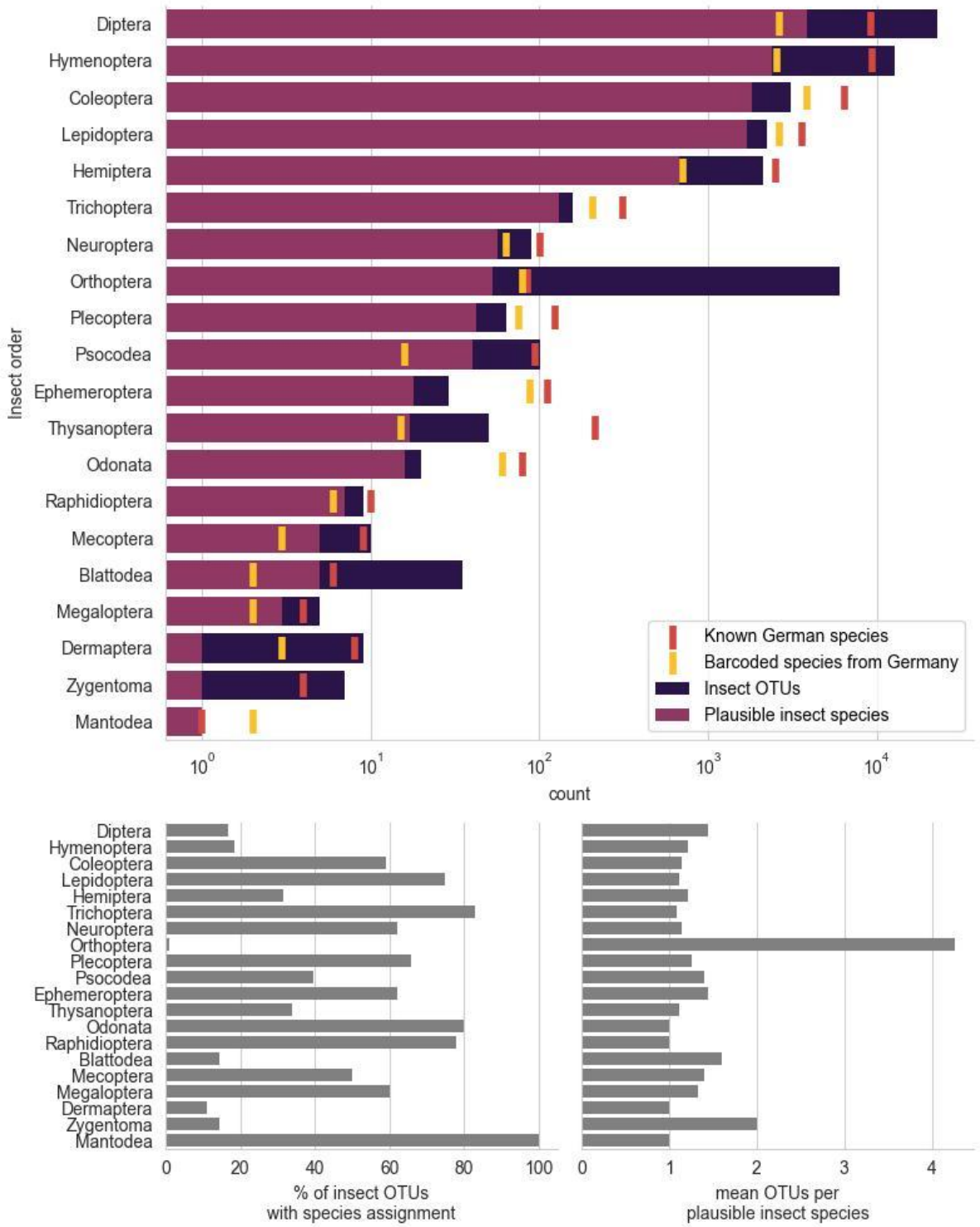

Figure 3: Top row: The number of plausible insect species per order in this study compared to the total number of OTUs per group. Vertical lines show the number of known insect species per order in Germany (Klausnitzer 2005) and the number of barcoded species sampled in Germany in the BOLD database (data release 14th of April, 2023). **Bottom left panel**: The percentage of insect OTUs with species-level assignment. **Bottom right panel:** The mean number of OTUs that were grouped per plausible insect species per OTU with species level assignment.

Time & cost estimation

The costs of the laboratory workflow for replicated samples, including all needed labware, chemicals, sequencing and personnel, were estimated to be roughly 88,000€, leading to costs of 48.50 € per sample (Table 2). All laboratory costs excluding labor make up 11.92 € (26%) of the total costs. Labour costs are the main driver of total costs with 34.28 € (74%) per sample. The sequencing step incurs the highest laboratory cost (4.85 € per sample), followed by enzymatic steps such as PCR (1.65 € per sample) and sample lysis (1.54 € per sample). The complete list of all materials and estimated times can be found in Supplementary Table S9.

The total time needed to process all 1,815 samples was roughly 1,100 working hours or 27.5 weeks of full-time work (equivalent to 141 samples within 2 weeks, one person and one robot). Most time was needed for the size-sorting and the homogenization of the samples. All steps after sample homogenization can be completed in 6 weeks of full-time work.

Table 2: Cost estimates for the presented workflow: Cost estimates are based on vendor list prices in 2022 where available. Time estimates are either hands-on time in the laboratory or mean runtimes of the robotic protocols. Costs and time per sample are based on 1,815 samples. Time total time and total costs are based on the respective replicates in the step. Labor costs were estimated at 60.35 € per hour.

Processing step	Material costs per sample [€]	Total material costs [€]	Labor costs per sample [€]	Time per sample [h:min:s]	Total time [h:min:s]	Total costs [€]
Size-sorting	-	-	15.09	0:15:00	453:45:00	27,383.81
Homogenization	0.21	379.90	11.5	0:11:26	345:45:00	21,252.45
Sample lysis	1.54	2,795.64	1.91	0:01:54	57:37:30	6,264.26
DNA extraction	1.02	1,845.55	1.09	0:01:05	33:00:00	3,823.27
QA DNA extraction	0.04	78.48	0.74	0:00:44	22:00:00	1,417.24
1st PCR	1.65	2,994.43	0.25	0:00:15	07:20:00	3,450.83
PCR cleanup	0.50	898.46	0.74	0:00:44	22:00:00	2,237.22
2nd PCR	1.65	2,994.43	0.37	0:00:22	11:00:00	3,663.81
Normalization	0.47	844.47	0.74	0:00:44	22:00:00	2,183.23
Pooling	<0.01	0.45	0.13	0:00:08	04:00:00	243.86
QA libraries	<0.01	10.86	0.13	0:00:08	04:00:00	254.27
Sequencing	4.85	8,800.00	-	-	-	8,800.00
Bioinformatic analysis	-	-	1.59	0:01:35	48:00:00	2,890.51
Robot depreciation and maintenance	-	-	-	-	-	4,000
Total	11.92	21,642.67	34.28	34:04	1030:27:00	**87,864.76**

Discussion

A metabarcoding workflow to upscale insect biodiversity monitoring

While several DNA metabarcoding workflows have been published, (e.g. Braukmann et al. 2019, Hardulak et al. 2020, Hausmann et al. 2020), our DNA metabarcoding approach differs from current DNA metabarcoding frameworks in important aspects that are key to increase throughput on the one hand and to decrease costs and time on the other: i) homogenization of the samples was done in preservative liquid to avoid time-consuming drying of the samples (Buchner et al. 2021a), ii) all samples were replicated before DNA extraction, which is regarded a key quality assurance and

quality control step, iii) all laboratory steps except for the agarose gels were performed on an automated liquid handler (Buchner et al. 2021c), iv) a mean sequencing depth of ~1.4 M reads per sample, and v) all protocols from wet-lab techniques until the bioinformatic analysis are published as open-source protocols (Buchner 2022b) or programs (Buchner and Leese 2020, Buchner et al. 2022) and are therefore transparent and fully reproducible to support open science and open data.

To date, the costs of metabarcoding protocols are often considered prohibitively high for large-scale monitoring programs (Borrell et al. 2017, Montgomery et al. 2021). Cost estimates for comparable protocols often exceed 200 € per sample and even range up to 400 € per sample (Ji et al. 2013, Elbrecht et al. 2017, Aylagas et al. 2018). Compared to others, the here presented workflow reduces these costs considerably (down to 48.50 €) by automating almost all crucial laboratory steps and presenting self-made solutions instead of expensive kits. The wet-lab processing costs without labor costs are below 12 € per sample including replication and sequencing costs. Since the presented pricing is based on vendor list prices, the costs could be further reduced by ordering chemicals and consumables needed in bulk or further downscaling of steps that involve enzymes when possible (Buchner et al. 2021b). What is even more important, is the reduced time needed to fully process one sample. We estimated about half an hour of processing time per sample with the majority of the time needed for size sorting and homogenization leading to personnel costs of about 34 € per sample. These costs could be further reduced by automating size sorting and homogenization of samples, which should be a focal point of future development. However, our cost analysis did not include any costs related to the sampling in the field. Based on experience from the German Malaise trap program, costs for Malaise trap installation and processing (e.g., April – October with a 2-week interval, resulting in 15 samples) are estimated to a total of 6 working hours (here: approx. 360 €) plus travel time (depending on distance to trap location) and material (here: approx. 400 €). This results in additional costs of roughly 65 € per sample, which is higher comparted to the lab costs.

Insect diversity in Germany

Without considering unassigned OTUs, we identified 10.803 insect species. And even with a greater sampling effort at our sites, this number is expected to increase by only 934 additional species that were missed due to insufficient sampling effort. The 10,803 detected species are already equivalent to almost one third of all insect species known to Germany and to ~83.1% of all barcoded insects from Germany. In Hymenoptera and Hemiptera, for example, we recovered close to 100% of the barcoded species in Germany. Still, our dataset very likely also contains many species, that were previously not found or not yet barcoded in Germany, but in adjacent countries. This would also explain, that for e.g. Diptera we find about 50% more species than there are reference sequences from Germany. Using our plausibility check system, only ~ 9% of the initially reported species were found to be implausible. This highlights the reliability of the DNA metabarcoding protocol and using a focus that includes also species found in adjacent countries. At the same time, the plausibility check also highlights the need to further improve reference databases. Since the checking by taxonomic experts is a time-consuming and potentially expensive process, our results suggest that checking the reported species via multiple publicly available databases, like GBIF (Telenius 2011)

and the Entomofauna Germanica (Klausnitzer 2005), might be sufficient to confirm the plausibility of reported species. While being reproducible and easily automatable the checking reported a lower number of plausible insect species than the expert judgment but maybe biased regarding poorly studied groups, e.g. Diptera, Hymenoptera, or specific groups of beetles. Also, the interpretation of taxa lists from all insects requires extremely broad expert knowledge. Our results support the need to further expand the synergies of taxonomists and molecular ecologists working together. An important task for the future remains the completion of interregional reference sequence databases and the generation of reference sequences of unknown species (Chua et al. 2023).

Assessing the dark taxa biodiversity

A key finding of our study is the approximately 5 times higher number of OTUs than plausible species. While in our study 1 OTU represents 1 species in the majority of cases for the plausible species there are some notable exceptions. One explanation for this could be that the sequences represent true biological signals but cannot be assigned to species-level because of the lack of reference sequences for described species. In addition, many of these OTUs can represent dark taxa. Karlsson et al. (2020) highlighted this, and especially the megadiverse Hymenoptera and Diptera comprise a significant proportion of undescribed taxa. In this context Parr et al. (2012) rightfully note that the generation of sequences is much faster than the formal description of new species leading to an enormous lag time between the deposition of sequences in GenBank or other repositories, new species recognition and their formal description. While it is difficult to precisely quantify the number of OTUs attributed to either known species without a reference barcode or to new species, we can roughly quantify the ratio using OTUs. OTUs with a fixed threshold of 3% sequence identity do not perfectly represent species, as taxa with a higher evolutionary rates may be split into multiple OTUs (Mahé et al. 2014, Müller and Nebel 2021). However, the same also happens with other approaches like BINs (Meier et al. 2022). Our mean number of OTUs per plausible insect species of 1-1.5 suggests that this is generally not the case for our marker/study group combination except in Orthoptera. The majority of plausible species (83.9%) are represented by a single OTU. We, therefore, conclude that in this case one OTU is a well-suited proxy for one species and we may have a large portion of yet-to-be-described species in the OTU dataset. This agrees with results reported from the Swedish Malaise Trap Program, where most of the observed 700 species new to science (Karlsson et al. 2020) belong to Diptera and Hymenoptera (Ronquist et al. 2020). Obviously, this is an overall stark underrepresentation of putatively undiscovered species in Germany. Especially insects living in habitats types not well sampled with Malaise traps (e.g. tree canopies) are likely to be a trove of undiscovered species diversity, even in countries like Germany and Sweden with a relatively species-poor fauna and a long taxonomic tradition. DNA Metabarcoding can help discover the extent of this unknown diversity. For the order of Orthoptera, however, OTUs do not adequately represent species. Orthopterans are fairly large and well-described for Germany with 84 known species (Reinhardt et al. 2005). Of those, we found 53 in our study, but also additional ~6,000 unassigned orthopteran OTUs. Bensasson et al. (2000, 2001) already reported high rates of mitochondrial pseudogenes (“Numts”) in orthopterans, potentially leading to inflated orthopteran estimates based on OTUs. Although we checked all OTUs for stop-codons, we might have still missed a considerable portion of Numts since not all must contain a stop codon.

A possible solution to both, incomplete reference databases as well as the identification of large intraspecific distances and Numts may be image-based approaches in combination with high-throughput single specimen barcoding and machine learning (Hartop et al. 2022, Wührl et al. 2022). While metabarcoding will outcompete those methods in time and costs for large-scale biomonitoring, those methods present a scalable solution to the described problems. At the same time, we need to acknowledge and counteract the fact that globally taxonomic expertise, particularly in megadiverse and understudied taxa, is vital to moving beyond OTUs in biodiversity research. Particularly, if we want to link biodiversity observations with ecological patterns and processes, expert knowledge of species' taxonomy and ecology are key. This includes among others insect trait information, which is currently largely lacking. Large-scale correlative analysis can provide information on trends and some ecosystem processes, yet many insect species rely on specific microhabitat conditions that currently cannot be modelled at scale.

Conclusions & Outlook

We present a reliable, scalable, fast and cost-effective DNA metabarcoding workflow that can be integrated into large scale or long-term insect biodiversity monitoring. With over 33,000 insect species we report almost the number of species that is known from whole Germany. While the approach does not deliver abundance data and relies on sample homogenization, thereby making it impossible to retroactively study specimens and their morphology, this amount of samples with millions of specimens could realistically never be analyzed with traditional manual sorting and determination by expert taxonomists. For the first time, the approach allows for almost real-time data generation as the presented workflow allows to completely process 141 samples with many ten thousands of specimens with a single person in 2 weeks including sequencing and bioinformatic analysis, so that even monitoring of hundreds of traps across a whole country like Germany is possible. For the dipterans we even report more species than are known from Germany. This shows that in these ''dark taxa" new species exist that need to be described and characterized. Thus, while we present a scalable biodiversity monitoring solution, we highlight the need to improve taxonomic and ecological characterization of insect diversity and to complete reference libraries. Only through these combined steps can we reliably monitor biodiversity change at large scales and provide data to support conservation actions.

Acknowledgements

This project received funding from the Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz of the German federal State of Hesse (Center for Translational Biodiversity Genomics; LOEWE-TBG) and the Hessische Landesamt für Naturschutz, Umwelt und Geologie (HLNUG).

References

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

Beermann AJ, Zizka VMA, Elbrecht V, Baranov V, Leese F (2018) DNA metabarcoding reveals the complex and hidden responses of chironomids to multiple stressors. Environmental Sciences Europe 30: 26. https://doi.org/10.1186/s12302-018-0157-x

Bensasson D, Zhang D-X, Hewitt GM (2000) Frequent Assimilation of Mitochondrial DNA by Grasshopper Nuclear Genomes. Molecular Biology and Evolution 17: 406–415. https://doi.org/10.1093/oxfordjournals.molbev.a026320

Bensasson D, Zhang D-X, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecology & Evolution 16: 314–321. https://doi.org/10.1016/S0169-5347(01)02151-6

Besson M, Alison J, Bjerge K, Gorochowski TE, Høye TT, Jucker T, Mann HM, Clements CF (2022) Towards the fully automated monitoring of ecological communities. Ecology Letters 25: 2753–2775.

Borrell YJ, Miralles L, Huu HD, Mohammed-Geba K, Garcia-Vazquez E (2017) DNA in a bottle—Rapid metabarcoding survey for early alerts of invasive species in ports. PLOS ONE 12: e0183347. https://doi.org/10.1371/journal.pone.0183347

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

Buchner D (2022a) Guanidine-based DNA extraction with silica-coated beads or silica spin columns. https://doi.org/10.17504/protocols.io.eq2ly73mmlx9/v2

Buchner D (2022b) Invertebrate bulk sample metabarcoding protocol collection. https://doi.org/10.17504/protocols.io.j8nlkw4n6l5r/v4

Buchner D (2022c) PCR cleanup and size selection with magnetic beads. https://doi.org/10.17504/protocols.io.36wgqj45xvk5/v3

Buchner D (2022d) PCR normalization and size selection with magnetic beads. https://doi.org/10.17504/protocols.io.q26g7y859gwz/v3

Buchner D (2022e) Sample preparation and lysis of homogenized malaise trap samples. https://doi.org/10.17504/protocols.io.dm6gpjrmjgzp/v1

Buchner D, Leese F (2020) BOLDigger – a Python package to identify and organise sequences with the Barcode of Life Data systems. Metabarcoding and Metagenomics 4: e53535. https://doi.org/10.3897/mbmg.4.53535

Buchner D, Haase P, Leese F (2021a) Wet grinding of invertebrate bulk samples – a scalable and cost-efficient protocol for metabarcoding and metagenomics. Metabarcoding and Metagenomics 5: e67533. https://doi.org/10.3897/mbmg.5.67533

Buchner D, Macher T-H, Leese F (2022) APSCALE: advanced pipeline for simple yet comprehensive analyses of DNA metabarcoding data. Bioinformatics: btac588. https://doi.org/10.1093/bioinformatics/btac588

Buchner D, Beermann AJ, Leese F, Weiss M (2021b) Cooking small and large portions of “biodiversity-soup”: Miniaturized DNA metabarcoding PCRs perform as good as large-volume PCRs. Ecology and Evolution. https://doi.org/10.1002/ece3.7753

Buchner D, Macher T-H, Beermann AJ, Werner M-T, Leese F (2021c) Standardized high-throughput biomonitoring using DNA metabarcoding: Strategies for the adoption of automated liquid handlers. Environmental Science and Ecotechnology 8: 100122. https://doi.org/10.1016/j.ese.2021.100122

Cardoso P, Barton PS, Birkhofer K, Chichorro F, Deacon C, Fartmann T, Fukushima CS, Gaigher R, Habel JC, Hallmann CA, Hill MJ, Hochkirch A, Kwak ML, Mammola S, Ari Noriega J, Orfinger AB, Pedraza F, Pryke JS, Roque FO, Settele J, Simaika JP, Stork NE, Suhling F, Vorster C, Samways MJ (2020) Scientists’ warning to humanity on insect extinctions. Biological Conservation 242: 108426. https://doi.org/10.1016/j.biocon.2020.108426

Chamberlain S, Oldoni D, Waller J (2022) rgbif: Interface to the Global Biodiversity Information Facility API. https://doi.org/10.5281/zenodo.6023735

Chimeno C, Hausmann A, Schmidt S, Raupach MJ, Doczkal D, Baranov V, Hübner J, Höcherl A, Albrecht R, Jaschhof M, Haszprunar G, Hebert PDN (2022) Peering into the Darkness: DNA Barcoding Reveals Surprisingly High Diversity of Unknown Species of Diptera (Insecta) in Germany. Insects 13: 82. https://doi.org/10.3390/insects13010082

Chiu C-H, Wang Y-T, Walther BA, Chao A (2014) An improved nonparametric lower bound of species richness via a modified good–turing frequency formula. Biometrics 70: 671–682. https://doi.org/10.1111/biom.12200

Chua PYS, Bourlat SJ, Ferguson C, Korlevic P, Zhao L, Ekrem T, Meier R, Lawniczak MKN (2023) Future of DNA-based insect monitoring. Trends in Genetics. https://doi.org/10.1016/j.tig.2023.02.012

Elbrecht V, Vamos EE, 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

Elbrecht V, Braukmann TWA, Ivanova NV, Prosser SWJ, Hajibabaei M, Wright M, Zakharov EV, Hebert PDN, Steinke D (2019) Validation of COI metabarcoding primers for terrestrial arthropods. PeerJ 7: e7745. https://doi.org/10.7717/peerj.7745

Elbrecht V, Bourlat SJ, Hörren T, Lindner A, Mordente A, Noll NW, Schäffler L, Sorg M, Zizka VMA (2021) Pooling size sorted Malaise trap fractions to maximize taxon recovery with metabarcoding. PeerJ 9: e12177. https://doi.org/10.7717/peerj.12177

Frøslev TG, Kjøller R, Bruun HH, Ejrnæs R, Brunbjerg AK, Pietroni C, Hansen AJ (2017) Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nature Communications 8: 1188. https://doi.org/10.1038/s41467-017-01312-x

Geiger MF, Moriniere J, Hausmann A, Haszprunar G, Wägele W, Hebert PDN, Rulik B (2016) Testing the Global Malaise Trap Program – How well does the current barcode reference library identify flying insects in Germany? Biodiversity Data Journal: e10671. https://doi.org/10.3897/BDJ.4.e10671

Hallmann CA, Sorg M, Jongejans E, Siepel H, Hofland N, Schwan H, Stenmans W, Müller A, Sumser H, Hörren T, Goulson D, Kroon H de (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

Hardulak LA, Morinière J, Hausmann A, Hendrich L, Schmidt S, Doczkal D, Müller J, Hebert PDN, Haszprunar G (2020) DNA metabarcoding for biodiversity monitoring in a national park: Screening for invasive and pest species. Molecular Ecology Resources 20: 1542–1557. https://doi.org/10.1111/1755-0998.13212

Hartop E (2021) A multi-faceted approach to a “dark taxon”: The hyperdiverse and poorly known scuttle flies (Diptera: Phoridae). Available from: http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-192276 (April 12, 2023).

Hartop E, Srivathsan A, Ronquist F, Meier R (2022) Towards Large-Scale Integrative Taxonomy (LIT): Resolving the Data Conundrum for Dark Taxa. Systematic Biology 71: 1404–1422. https://doi.org/10.1093/sysbio/syac033

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

Hobern D (2021) BIOSCAN: DNA barcoding to accelerate taxonomy and biogeography for conservation and sustainability. Genome 64: 161–164. https://doi.org/10.1139/gen-2020-0009

Hoppeler F, Tachamo Shah RD, Shah DN, Jähnig SC, Tonkin JD, Sharma S, Pauls SU (2016) Environmental and spatial characterisation of an unknown fauna using DNA sequencing – an example with Himalayan Hydropsychidae (Insecta: Trichoptera). Freshwater Biology 61: 1905–1920. https://doi.org/10.1111/fwb.12824

IPBES (2019) Summary for policymakers of the global assessment report on biodiversity and ecosystem services. Zenodo https://doi.org/10.5281/zenodo.3553579

Ji Y, Ashton L, Pedley SM, Edwards DP, Tang Y, Nakamura A, Kitching R, Dolman PM, Woodcock P, Edwards FA, Larsen TH, Hsu WW, Benedick S, Hamer KC, Wilcove DS, Bruce C, Wang X, Levi T, Lott M, Emerson BC, Yu DW (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecology Letters 16: 1245–1257. https://doi.org/10.1111/ele.12162

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

Klausnitzer B (2005) Die Insektenfauna Deutschlands (“Entomofauna Germanica”) - ein Gesamtüberblick. 37: 87–97.

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

Losey JE, Vaughan M (2006) The Economic Value of Ecological Services Provided by Insects. BioScience 56: 311–323. https://doi.org/10.1641/0006-3568(2006)56[311:TEVOES]2.0.CO;2

Mahé F, Rognes T, Quince C, Vargas C de, Dunthorn M (2014) Swarm: robust and fast clustering method for amplicon-based studies. PeerJ 2: e593. https://doi.org/10.7717/peerj.593

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

Meier R, Blaimer BB, Buenaventura E, Hartop E, von Rintelen T, Srivathsan A, Yeo D (2022) A re-analysis of the data in Sharkey et al.’s (2021) minimalist revision reveals that BINs do not deserve names, but BOLD Systems needs a stronger commitment to open science. Cladistics 38: 264–275.

Mirtl M, T. Borer E, Djukic I, Forsius M, Haubold H, Hugo W, Jourdan J, Lindenmayer D, McDowell WH, Muraoka H, Orenstein DE, Pauw JC, Peterseil J, Shibata H, Wohner C, Yu X, Haase P (2018) Genesis, goals and achievements of Long-Term Ecological Research at the global scale: A critical review of ILTER and future directions. Science of The Total Environment 626: 1439–1462. https://doi.org/10.1016/j.scitotenv.2017.12.001

Montgomery GA, Belitz MW, Guralnick RP, Tingley MW (2021) Standards and Best Practices for Monitoring and Benchmarking Insects. Frontiers in Ecology and Evolution 8. Available from: https://www.frontiersin.org/articles/10.3389/fevo.2020.579193 (March 17, 2023).

Müller R, Nebel M (2021) On the use of sequence-quality information in OTU clustering. PeerJ 9: e11717. https://doi.org/10.7717/peerj.11717

Parr CS, Guralnick R, Cellinese N, Page RDM (2012) Evolutionary informatics: unifying knowledge about the diversity of life. Trends in Ecology & Evolution 27: 94–103. https://doi.org/10.1016/j.tree.2011.11.001

Pereira HM, Navarro LM, Martins IS (2012) Global Biodiversity Change: The Bad, the Good, and the Unknown. Annual Review of Environment and Resources 37: 25–50. https://doi.org/10.1146/annurev-environ-042911-093511

Ratnasingham S, Hebert PDN (2007) bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7: 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Ratnasingham S, Hebert PDN (2013) A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. PLOS ONE 8: e66213. https://doi.org/10.1371/journal.pone.0066213

Reinhardt K, Köhler G, Maas S, Detzel P (2005) Low dispersal ability and habitat specificity promote extinctions in rare but not in widespread species: the Orthoptera of Germany. Ecography 28: 593–602. https://doi.org/10.1111/j.2005.0906-7590.04285.x

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ 4: e2584. https://doi.org/10.7717/peerj.2584

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

Stork NE (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

Sturmbauer C, Opadiya GB, Niederstätter H, Riedmann A, Dallinger R (1999) Mitochondrial DNA reveals cryptic oligochaete species differing in cadmium resistance. Molecular Biology and Evolution 16: 967–974. https://doi.org/10.1093/oxfordjournals.molbev.a026186

Telenius A (2011) Biodiversity information goes public: GBIF at your service. Nordic Journal of Botany 29: 378–381. https://doi.org/10.1111/j.1756-1051.2011.01167.x

Uhler J, Redlich S, Zhang J, Hothorn T, Tobisch C, Ewald J, Thorn S, Seibold S, Mitesser O, Morinière J, Bozicevic V, Benjamin CS, Englmeier J, Fricke U, Ganuza C, Haensel M, Riebl R, Rojas-Botero S, Rummler T, Uphus L, Schmidt S, Steffan-Dewenter I, Müller J (2021) Relationship of insect biomass and

richness with land use along a climate gradient. Nature Communications 12: 5946. https://doi.org/10.1038/s41467-021-26181-3

Vamos E, Elbrecht V, Leese F (2017) Short COI markers for freshwater macroinvertebrate metabarcoding. Metabarcoding and Metagenomics 1: e14625. https://doi.org/10.3897/mbmg.1.14625

Waskom M, Botvinnik O, Gelbart M, Ostblom J, Hobson P, Lukauskas S, Gemperline DC, Augspurger T, Halchenko Y, Warmenhoven J, Cole JB, De Ruiter J, Vanderplas J, Hoyer S, Pye C, Miles A, Swain C, Meyer K, Martin M, Bachant PQ, Kunter G, Villalba S, Brian, Fitzgerald C, Evans CG, Williams ML, O'Kane D, Yarkoni T, Brunner T (2020) seaborn: Statistical data visualization. Astrophysics Source Code Library: ascl:2012.015.

Welti EAR, Zajicek P, Frenzel M, Ayasse M, Bornholdt T, Buse J, Classen A, Dziock F, Engelmann RA, Englmeier J, Fellendorf M, Förschler MI, Fricke U, Ganuza C, Hippke M, Hoenselaar G, Kaus-Thiel A, Kerner J, Kilian D, Mandery K, Marten A, Monaghan MT, Morkel C, Müller J, Puffpaff S, Redlich S, Richter R, Rojas-Botero S, Scharnweber T, Scheiffarth G, Yáñez PS, Schumann R, Seibold S, Steffan-Dewenter I, Stoll S, Tobisch C, Twietmeyer S, Uhler J, Vogt J, Weis D, Weisser WW, Wilmking M, Haase P (2022) Temperature drives variation in flying insect biomass across a German malaise trap network. Insect Conservation and Diversity 15: 168–180. https://doi.org/10.1111/icad.12555

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

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