PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

CC BY

Review

Cite this article: Meier R, Hartop E, Pylatiuk C, Srivathsan A. 2024 Towards holistic insect monitoring: species discovery, description, identification and traits for all insects. *Phil. Trans. R. Soc. B* **379**: 20230120. https://doi.org/10.1098/rstb.2023.0120

Received: 7 October 2023
Accepted: 25 January 2024

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

Subject Areas:
taxonomy and systematics, environmental science

Keywords:
insect decline, megabarcoding, integrative taxonomy ecological traits, natural history

Author for correspondence:
Rudolf Meier
e-mail: rudolf.meier@hu-berlin.de

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.7151304.

THE ROYAL SOCIETY PUBLISHING

Towards holistic insect monitoring: species discovery, description, identification and traits for all insects

Rudolf Meier[1,3], Emily Hartop[1,2], Christian Pylatiuk[4] and Amrita Srivathsan[1]

[1]Center for Integrative Biodiversity Discovery, Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Invalidenstraße 43, 10115 Berlin, Germany
[2]Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology, Trondheim, NO-7491, Norway
[3]Institute of Biology, Humboldt University, 10115 Berlin, Germany
[4]Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany

RM, 0000-0002-4452-2885; EH, 0000-0002-8598-9844; CP, 0000-0002-3507-7134

Holistic insect monitoring needs scalable techniques to overcome taxon biases, determine species abundances, and gather functional traits for all species. This requires that we address taxonomic impediments and the paucity of data on abundance, biomass and functional traits. We here outline how these data deficiencies could be addressed at scale. The workflow starts with large-scale barcoding (megabarcoding) of all specimens from mass samples obtained at biomonitoring sites. The barcodes are then used to group the specimens into molecular operational taxonomic units that are subsequently tested/validated as species with a second data source (e.g. morphology). New species are described using barcodes, images and short diagnoses, and abundance data are collected for both new and described species. The specimen images used for species discovery then become the raw material for training artificial intelligence identification algorithms and collecting trait data such as body size, biomass and feeding modes. Additional trait data can be obtained from vouchers by using genomic tools developed by molecular ecologists. Applying this pipeline to a few samples per site will lead to greatly improved insect monitoring regardless of whether the species composition of a sample is determined with images, metabarcoding or megabarcoding.

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

1. Introduction

Insects deliver many ecosystem services such as pollination, nutrient recycling, seed dispersal, etc. They also transmit pathogens, prey on other species and are an important food source [1–3]. Given the high species diversity and biomass of insects, it is inconceivable that any ecosystem can be understood and monitored without comprehension of the ecology of its insect communities. This is why biomonitoring should not stop with determining species richness, but also resolve taxonomic issues and include an assessment of abundance and functional diversity [4,5]. However, how can we obtain all this information? We here outline how to go from mass samples collected with standardized traps (e.g. Malaise traps) to generating data for functional analysis. The workflow starts with individually barcoding all specimens in a mass sample ('megabarcoding' [6]), grouping them into molecular operational taxonomic units (MOTUs), revising the MOTU boundaries to obtain species limits, training artificial intelligence (AI) identification algorithms for common species, and using vouchers to collect trait and natural history data at scale. Once this pipeline has been applied to a few samples at each monitoring site, all subsequent monitoring becomes easier,

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

because AI tools can be used to identify at least some of the common species. In addition, species lists and abundances become more meaningful, because the samples can be analysed at species-level and use trait data for functional analysis.

Given the enormous number of insect species on our planet, the vision outlined above does come with many challenges. Before we outline the workflow in more depth, we would thus like to mitigate some typical concerns, while simultaneously stressing specific advantages. At present, only very few, charismatic species (e.g. some species of birds, butterflies, trees, etc.) are really monitored at a global scale [7]. By contrast, insect biomonitoring is conducted locally at a limited number of sites. Monitoring only becomes 'global' through the analysis of trends across local sites. Given this approach to global monitoring it is important that local sites are analysed in detail. This means that we need methods that go well beyond generating species lists. Such data should include abundance, trait, and natural history information for key species. Fortunately, gathering meaningful trait and natural history information for locally dominant species is made distinctly easier by typical patterns of species abundances [8]: most specimens at a site belong to a moderate number of species. These species then become a priority for collecting trait and natural history data. For example, analysis of Malaise trap data from eight countries in Srivathsan *et al*. [9] shows that on average, 39% of the specimens in a trap belong to the 20 most abundant species. Depending on sample, biogeographic region and habitat, the number of species contributing 50% of all specimens ranged from 7 to 195 (median = 53) (figure 1; electronic supplementary material). This suggests that many functional insights into insect communities can be gained by collecting trait data for a manageable number of species. However, this requires that biologists let go of their usual habit of studying only particular taxa and/or rare species. Instead, we propose that the focus should be on abundant species. Such a shift to abundance-based research priorities will be desirable not only for holistic insect biomonitoring, but also for addressing taxonomic impediments. Srivathsan *et al*. [9] also showed that more than half the specimens and species in Malaise trap samples belong to only 20 family level insect taxa [9]. Since most of these taxa are neglected by taxonomists [9], the setting of taxonomic priorities according to abundance will automatically shift attention to taxa that are currently understudied.

The need for the holistic study of bulk samples also becomes apparent when we consider current approaches. Morphospecies sorting and cherry-picking specimens from bulk samples have been largely replaced with bulk sequencing techniques using metabarcoding and metagenomics [10,11]. These techniques were developed to allow entire samples to be characterized at once and en masse [11–13]. However, all metabarcoding comes with the downside of losing the association between DNA barcode and specimen. In addition, many popular metabarcoding techniques are destructive, requiring the homogenization of samples into 'insect soups' before sequencing, leaving no voucher specimens to validate results [14,19]. This is problematic, as the barcode region targeted by metabarcoding is only a few hundred base pairs of DNA long. For 10–20% of taxa, this information has repeatedly been shown to be insufficient for delimiting species [15,16]. This problem also remains when non-destructive methods for metabarcoding are used (e.g. mild lysis, extraction of DNA from preservatives [17,18]). A further complication of metabarcoding is that different DNA extraction methods and the use of different primers yield different results [18,19]. Thus, method development is still ongoing to ensure the comparability of the presence–absence and approximate abundance data obtained from bulk sequencing of whole samples across studies [11,17,20].

2. Towards holistic insect monitoring

Each insect sample is like a library documenting the environmental conditions in the habitat where the sample was collected. To adequately extract this information, we must resolve species limits and gather data on the natural history of these species. We thus argue that metabarcoding should be complemented with techniques that generate barcode databases for validated species, reveal how common these species are, and how they interact with the environment. The data obtained from such a workflow will increase the value of species lists regardless of whether they were collected in the past or will be collected in the future.

Our proposal is that each site chosen for biomonitoring should be covered by an in-depth analysis of a few samples obtained with the kind of traps that will also be used for routine biomonitoring. These samples should be analysed by applying megabarcoding [6] combined with semi-automatic imaging of all specimens. In addition to establishing accessible vouchers, such an approach will reveal species richness and abundances. The resulting vouchers are then available for collecting important trait data. For example, data obtained with genome skimming or low-coverage genome sequencing can be mined for ecological information (see below), but also used to construct phylogenetic relationships [21], determine genetic diversity [22], population connectivity [23] and population size [24] (for further reading, see Theissinger *et al*. [25]). Inferences about the environmental history of a monitoring site will thus cover long time periods ranging from the condition at the time of collecting over recent years (e.g. population connectivity [23]) to thousands of years ago (e.g. community assembly [26]) This information can then also enrich the analysis of samples processed with metabarcoding, because the trait data would be associated with specific species names and the species would be identifiable via barcodes. As outlined below, this is eminently feasible, because we already know how to get barcodes and images at scale and metagenomics or whole genome sequencing can yield much of the remaining information [27].

3. From megabarcoding to species

Bulk insect samples generally consist of a relatively small number of species in high abundance and a very large number of rare species (figures 1 and 2). Many abundant species belong to 'dark taxa', which Hartop *et al*. [28] defined as high-diversity clades (greater than 1000 species) that are poorly known (fewer than 10% of species described) [28]. What makes such taxa understudied is a combination of high abundance and diversity, usually accompanied by small size, poorly developed taxonomy and a reliance on microscopic characters for species identification [28,29]. Dark taxa are arguably the biggest obstacle for a holistic monitoring of insect communities. They are the main reason why, until recently, Malaise or pitfall trap samples were very rarely fully processed.

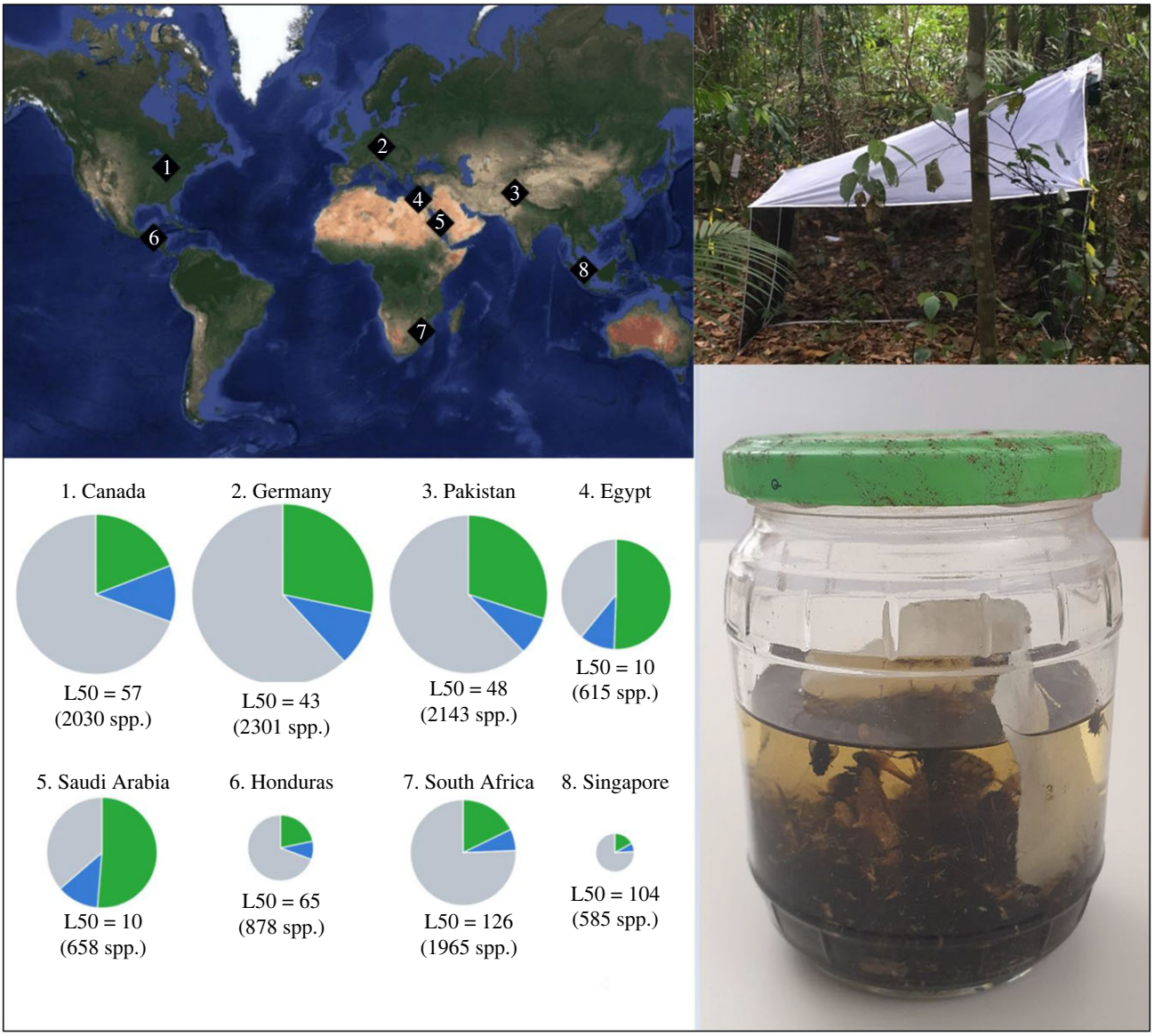


Figure 1. Proportion of specimens in Malaise traps belonging to the top 10 (green), next 10 (blue) and remaining (grey) MOTUs in Malaise traps characterised in Srivathsan *et al*. [9]. Pie charts are scaled to abundance. L50 values represent the number of MOTUs that cover 50% of the specimens. Map made using ggmap in R (Google maps, satellite, 2023 NASA).

Such dark taxa can be resolved at scale when bulk samples are processed by megabarcoding (figure 2; [6,9,30–33]). Such individual barcoding of all specimens is now feasible owing to the plummeting cost, high efficiency and rapidly increasing accuracy of third generation sequencing [34]. Megabarcoding associates every specimen in a sample with a DNA barcode and relies on use of non-destructive techniques to preserve morphological features of the specimens [31]. It thus generates a large number of vouchers that allow for downstream taxonomic and genomic research. One important use of these vouchers is species delimitation and description based on multiple types of data (integrative taxonomy) [35,36]. Recently, large-scale integrative taxonomy (LIT) was proposed as a systematic way to generate integrative species hypotheses [28]. Preliminary hypotheses are generated based on a data source such as DNA barcodes that is inexpensive and easy to obtain at scale. The hypotheses are then validated or revised using a second type of data obtained for a smaller number of specimens that was specifically chosen to test the preliminary hypotheses (e.g. based on haplotype divergence [28]).

LIT has been successfully tested twice. The first test [28] was based on approximately 18 000 specimens shown to contain 365 species of Phoridae (Diptera), although the specimens were initially only assigned to 315 molecular clusters. Barcodes were obtained for all specimens, but the targeted morphological checks only required revisiting just over 5% of specimens. A second successful validation of LIT was recently undertaken by Meier *et al*. [37] for 1456 specimens of Mycetophilidae (Diptera) which were shown to contain 120 species based on a LIT analysis. Compared to the first test based on phorid flies, a larger proportion of specimens had to be checked for MOTU validation because studying only one specimen for each of the 120 species already meant inspecting greater than 8% of all specimens in the sample ($n = 1456$). Still, LIT was again found to be effective at resolving conflict between MOTUs and morphology by studying the morphology of only a small subset of all specimens [37,38].

In the above applications of LIT, DNA barcodes were the data source used to generate preliminary species hypotheses and morphology was the second data source used to validate these hypotheses. However, LIT could also include other types of data. For example, automated imaging combined with AI algorithms for unsupervised learning may eventually advance sufficiently for new solutions: the first data source could be image data, while the validation of preliminary

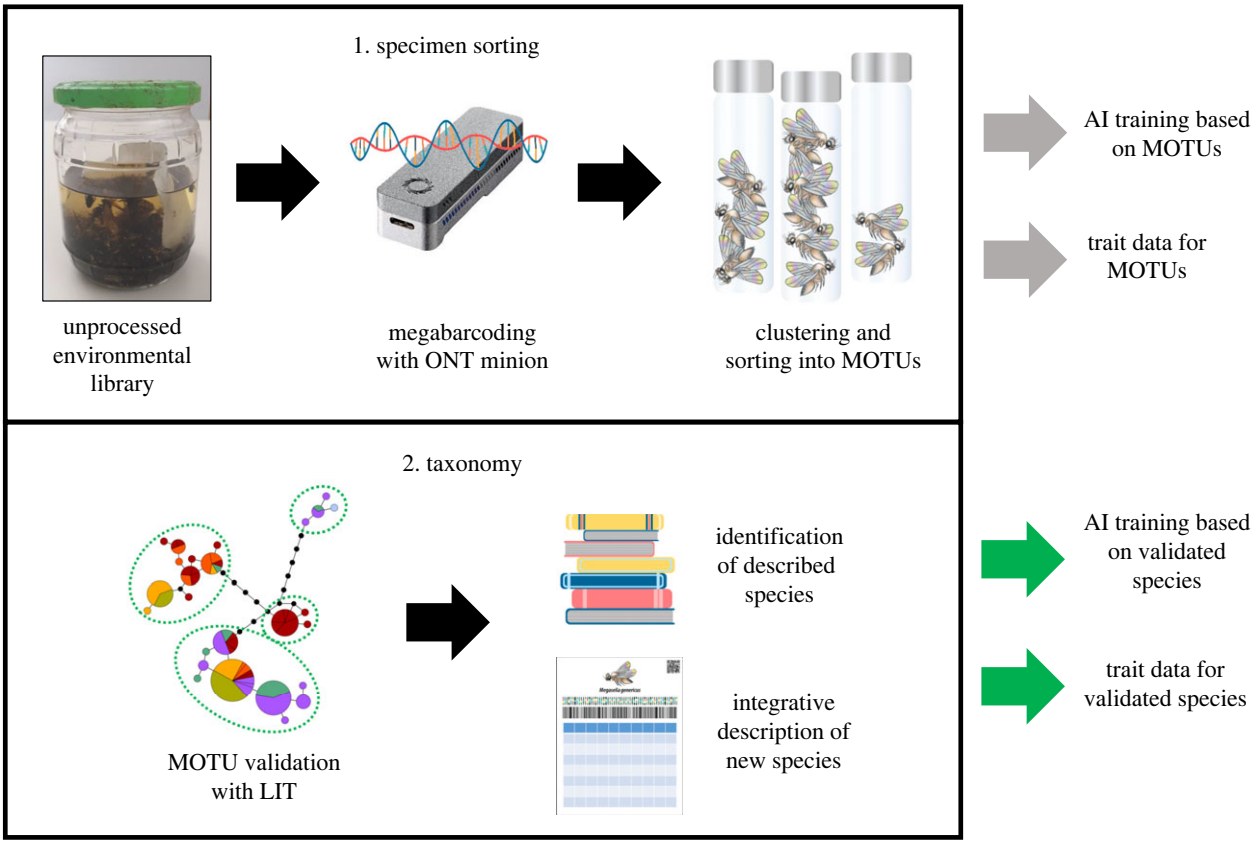


Figure 2. From bulk insect sample to species. Sorting involves megabarcoding and sorting specimens into MOTUs (ONT, Oxford Nanopore Technologies). Taxonomic research includes MOTU validation with LIT followed by either identification or description. AI training and trait data should be collected for validated species, but processing at MOTU-level can yield approximate data.

species hypothesis could be barcodes or morphological data from genitalia. Since automated imaging could yield thousands of images at low cost, this would be in line with the basic principle of LIT, i.e. that the data used for sorting should be cheap.

Delimiting all species found in a sample precedes one of the most difficult work phases in biodiversity science—identification or description. The task of distinguishing between species that have already been described and those that need description is often complicated by crumbling type material, muddled public databases, and insufficiently detailed legacy descriptions in the historical literature [36,39]. Resolving legacy issues in taxonomy is time-consuming, but this time only needs to be invested once per taxon and region. For example, once the described species of fungus gnats (Diptera: Mycetophilidae) in Southeast Asia had been evaluated based on descriptions and types, Amorim *et al*. [38] were able to describe 115 of the 120 species of one set of Malaise trap samples from Singapore as new to science. When Meier *et al*., analysed a second set of samples, more than 90% of the specimens belonged to the species described based on the first set samples. Indeed, the barcodes suggested the presence of fewer than 20 additional new species [37]. This implies that a few iterations of taxonomic revision of a taxon at any one site will be sufficient for allowing most species to be assigned to a species. The use of barcode and morphological data will also mean that the species are robustly delimited and can be identified with either barcodes or morphological features [28,38,40]. Associated high quality images will enable future-proofing the descriptions.

4. Robotics, machine learning and biomass

Processing several bulk samples per site using specimen-based approaches requires the use of robotics. In fact, we argue that the ultimate goal should be to develop AI identification algorithms, so that specimens can be largely identified based on images. Robots such as the DiversityScanner [41] have already been deployed to accelerate specimen transfer from bulk samples to 96-well microplates. This robot detects, photographs and sorts insects in preparation for megabarcoding [41]. Concomitant with the automation of barcoding, the DiversityScanner is programmed to generate stacked images for each specimen. These images (and associated specimens) are then grouped into putative species using barcode data. Even now, the DiversityScanner can already identify common families without sequencing, because it uses a trained convolutional neural network (CNN) for the most abundant insect families [41]. The next logical step is to train identification algorithms to recognize common species, first using MOTUs as approximations of species and later species after validation with LIT. Common species will benefit first, but even species that are rare in individual samples can eventually be covered by CNNs because specimens for rare species from many samples can be pooled until there are a sufficiently large number of images for training the algorithms.

At present, the extent to which CNNs will be able to identify specimens to species remains unclear [42]. However, this is probably a matter of what kinds of images (e.g. orientation, body part) are provided for training and identification. Importantly, the

quality, variety, and orientation of images needed for CNN training may vary across taxa (see approaches used for nematodes: [43]). However, it is likely that the need for barcoding of specimens belonging to common species will dramatically decrease. Instead, such species will be identified based on images, then counted, and finally transferred into species-specific vials. Only unknown specimens will subsequently have to be picked up, imaged, and barcoded. This yields a continuous feedback loop that slowly whittles away the unknown biodiversity, and facilitates finding rare species that require additional sources of data, or expert attention. DNA barcoding may eventually be used only to validate identification generated by CNNs or for the identification of body parts that have lost diagnosable features.

The next step in holistic biomonitoring is semi-automatic collection of biomass information. Some tools are already available, but their use is still limited, because they either require identified specimens [44] or the biomass of three-dimensional insects is estimated based on two-dimensional images of whole samples [45] or individual specimens [41]. It appears to us that the next logical step is determining biomass based on three-dimensional models. They could initially be produced for common species by modelling several specimens covering the species' size range. Afterwards, one could establish the relationship between three-dimensional volumes and two-dimensional measurements so that routine biomonitoring could continue to rely on two-dimensional imaging. Once AI identification tools are available and conversion factors are known, one could spread an insect sample across a large table-sized tray and use cameras to identify and measure all specimens in almost real-time. When combined with information on what the species does in its environment, this will allow us to resolve changes in not only the size distribution and abundance of insects over time, but also to relate such changes to ecosystem stability. Such advances will get us one step closer to reading a bulk insect sample like we would read the books in an environmental library.

5. Large-scale natural history data

Holistic insect biomonitoring requires an understanding of the functional diversity of insect communities detected at particular sites [46,47]. Characterization of functional diversity ranges from using classifications of species into functional groups to the quantification of functional traits for species [48–50]. Some trait information can be obtained from the natural history literature [51], but this literature is in decline [52] and tends to cover mostly charismatic species. We thus argue that collecting trait data has to be greatly expanded and more targeted on species that are particularly common and/or contribute much of the biomass in different trophic groups. Semi-automatic imaging [41] immediately provides information on morphological traits such as body size, biomass, hairiness, eye number and size [4]. The vouchers generated by megabarcoding can also be used to obtain other life-history traits such sex ratios, time of reproduction, egg or clutch size. Beyond morphological traits, metagenomics (via whole genome sequencing) of vouchers will arguably offer the most promising approach for collecting additional ecological insights. This technique can simultaneously detect many species interactions across a broad range of taxa without amplification-related biases [53]. Indeed, metagenomics is particularly suitable when there is a need to rapidly characterize the feeding ecology, host genetics, parasites and microbes of a particular species [54]. Note that the same low-coverage genomic data also enables reconstructing phylogenetic relationships and estimating the genetic diversity within species [27,55].

However, metagenomics remains expensive and metabarcoding will retain much of its importance. For example, prey detection via metabarcoding of the gut content of predators using targeted primers is now well established [56,57], as is the sequencing of pollen found on insects [58–60]. Indeed, pollen can sometimes simply be shaken off insects and used directly as DNA template [58–60]. Given that much of the biomass of holometabolous insects is accumulated by larvae, dietary metabarcoding is particularly needed for immature stages [61]. This requires that the larvae and adults are first associated via megabarcoding [62]. Even tools that were initially developed only for vertebrate detection via invertebrate derived DNA (iDNA) can now also be used for revealing species interactions between insects and vertebrates [63–67], and plants [68]. The study of iDNA is particularly scalable and cost-effective when applied to insect faeces or regurgitates, as neither substrate requires DNA extraction [69]. Other cost-effective techniques are the analysis of fresh or archived plants for arthropod DNA [70,71] and the analysis of spider webs for prey DNA [72,73]. Occasionally, the interpretation of the signal can be difficult, but tools such as metabolite screens can even resolve whether the detected iDNA originated from carrion or dung [74,75].

For monitoring environmental health and ecosystem functioning, the study of microbiomes [76,77] is becoming integral to understanding of insect ecology [78], but the analysis of microbiomes also allows for the detection of antimicrobial resistance among the microorganisms in a habitat and sometimes it can even resolve whether the resistance genes were acquired by the insects when feeding on contaminated food (as has been shown for cockroaches, houseflies, ants, mosquitoes: [79]). Unfortunately, it is currently unknown to what extent bacteria with such resistance genes are prevalent in common insects, but in-depth study of vouchers for common species can answer these questions. Note, that screens of host genomes and microbiomes are also important for discovering new insect-derived antimicrobial substances [80] and can reveal mutations related to pesticide resistance, thus predicting the susceptibility of insect populations and contributing to selecting appropriate control measures [81–83].

6. Conclusion

To achieve holistic insect monitoring, we must overcome taxon biases and the lack of trait data for common species. This can be done by paying more attention to resolving species limits and making species identifiable through a variety of user-friendly tools. We predict that image-based identification will eventually turn out to be the tool of choice for many species. In the meantime, DNA barcoding using third generation sequencing is a viable and valuable intermediate solution, because it also yields the barcodes needed for the analysis of environmental DNA with metabarcoding and metagenomics. Following megabarcoding, the next logical step in holistic biomonitoring is collecting functional traits at scale. Body size, biomass, feeding habits, seasonality, clutch

and egg size, and many other traits are low-hanging fruits, but many additional traits can be obtained through shallow sequencing of whole vouchers or body parts. By using these methods for characterizing the most common species in a few samples per site, a transformative shift in the quality of insect monitoring will be feasible.

Data accessibility. The data used for this submission is available from Nature: https://www.nature.com/articles/s41559-023-02066-0 [84]. Supplementary material is available online [85].

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

Authors' contributions. R.M.: conceptualization, project administration, writing—original draft, writing—review and editing; E.H.: conceptualization, visualization, writing—original draft, writing—review and editing; C.P.: conceptualization, project administration, writing—review and editing; A.S.: conceptualization, data curation, formal analysis, visualization, writing—original draft, writing—review and editing.

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

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

Funding. We received no funding for this study.

Acknowledgements. We thank the numerous scientists and laboratory members of the former Evolutionary Biology Laboratory at the National University of Singapore and the Museum für Naturkunde's Centre for Integrative Biodiversity Discovery for contributions to developing megabarcoding and contributing stimulating discussions regarding future insect monitoring.

References

1. Noriega JA *et al.* 2018 Research trends in ecosystem services provided by insects. *Basic Appl. Ecol.* **26**, 8–23. (doi:10.1016/j.baae.2017.09.006)
2. Losey JE, Vauhan M. 2006 The economic value of ecological services provided by insects. *Bioscience* **56**, 311–323.
3. Debinski DM. 2023 Insects in grassland ecosystems. In *Rangeland wildlife ecology and conservation* (eds LB McNew, DK Dahlgren, JL Beck), pp. 897–929. Cham, Switzerland: Springer.
4. Moretti M *et al.* 2017 Handbook of protocols for standardized measurement of terrestrial invertebrate functional traits. *Funct. Ecol.* **31**, 558–567. (doi:10.1111/1365-2435.12776)
5. Greenop A, Woodcock BA, Outhwaite CL, Carvell C, Pywell RF, Mancini F, Edwards FK, Johnson AC, Isaac NJB. 2021 Patterns of invertebrate functional diversity highlight the vulnerability of ecosystem services over a 45-year period. *Curr. Biol.* **31**, 4627–4634.e3. (doi:10.1016/j.cub.2021.07.080)
6. Chua PYS, Bourlat SJ, Ferguson C, Korlevic P, Zhao L, Ekrem T, Meier R, Lawniczak MKN. 2023 Future of DNA-based insect monitoring. *Trends Genet.* **39**, P531–P544. (doi:10.1016/j.tig.2023.02.012)
7. Schmeller DS *et al.* 2017 Building capacity in biodiversity monitoring at the global scale. *Biodivers. Conserv.* **26**, 2765–2790. (doi:10.1007/s10531-017-1388-7)
8. Callaghan CT, Borda-de-Água L, van Klink R, Rozzi R, Pereira HM. 2023 Unveiling global species abundance distributions. *Nat. Ecol. Evol.* **7**, 1600–1609. (doi:10.1038/s41559-023-02173-y)
9. Srivathsan A *et al.* 2023 Convergence of dominance and neglect in flying insect diversity. *Nat. Ecol. Evol.* **7**, 1012–1021. (doi:10.1038/s41559-023-02066-0)
10. Crampton-Platt A, Yu DW, Zhou X, Vogler AP. 2016 Mitochondrial metagenomics: letting the genes out of the bottle. *Gigascience* **5**, 15. (doi:10.1186/s13742-016-0120-y)
11. Buchner D *et al*. 2023 German-wide Malaise trap metabarcoding estimates over 33,000 insect species. *BioRxiv*. (doi:10.1101/2023.05.04.539402)
12. Elbrecht V, Leese F. 2017 Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. *Front. Environ. Sci.* **5**, 11. (doi:10.3389/fenvs.2017.00011)
13. Cristescu ME. 2014 From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends Ecol. Evol.* **29**, 566–571. (doi:10.1016/j.tree.2014.08.001)
14. Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z. 2012 Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods Ecol. Evol.* **3**, 613–623. (doi:10.1111/j.2041-210X.2012.00198.x)
15. Bergsten J *et al*. 2012 The effect of geographical scale of sampling on DNA barcoding. *System. Biol.* **61**, 851–869. (http://dx.doi.org/10.1093/sysbio/sys037)
16. Meier R, Blaimer BB, Buenaventura E, Hartop E, von Rintelen T, Srivathsan A, Yeo D. 2021 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. (doi:10.1111/cla.12489)
17. Iwaszkiewicz-Eggebrecht E *et al.* 2023 Optimizing insect metabarcoding using replicated mock communities. *Methods Ecol. Evol.* **14**, 1130–1146. (doi:10.1111/2041-210X.14073)
18. Marquina D, Esparza-Salas R, Roslin T, Ronquist F. 2019 Establishing arthropod community composition using metabarcoding: surprising inconsistencies between soil samples and preservative ethanol and homogenate from Malaise trap catches. *Mol. Ecol. Resour.* **19**, 1516–1530. (doi:10.1111/1755-0998.13071)
19. Elbrecht V, Braukmann TWA, Ivanova NV, Prosser SWJ, Hajibabaei M, Wright M, Zakharov EV, Hebert PDN, Steinke D. 2019 Validation of COI metabarcoding primers for terrestrial arthropods. *PeerJ* **7**, e7745. (doi:10.7717/peerj.7745)
20. Luo M, Ji Y, Warton D, Yu DW. 2023 Extracting abundance information from DNA-based data. *Mol. Ecol. Resour.* **23**, 174–189. (doi:10.1111/1755-0998.13703)
21. Magnussen T, Johnsen A, Kjærandsen J, Struck TH, Søli GEE. 2022 Molecular phylogeny of Allodia (Diptera: Mycetophilidae) constructed using genome skimming. *Syst. Entomol.* **47**, 267–281. (doi:10.1111/syen.12529)
22. Foote AD *et al.* 2016 Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nat. Commun.* **7**, 11693. (doi:10.1038/ncomms11693)
23. Lou RN, Jacobs A, Wilder AP, Therkildsen NO. 2021 A beginner's guide to low-coverage whole genome sequencing for population genomics. *Mol. Ecol.* **30**, 5966–5993. (doi:10.1111/mec.16077)
24. Fournier R, Tsangalidou Z, Reich D, Palamara PF. 2023 Haplotype-based inference of recent effective population size in modern and ancient DNA samples. *Nat. Commun.* **14**, 7945. (doi:10.1038/s41467-023-43522-6)
25. Theissinger K *et al.* 2023 How genomics can help biodiversity conservation. *Trends Genet.* **39**, 545–559. (doi:10.1016/j.tig.2023.01.005)
26. Mérot C *et al.* 2021 Locally adaptive inversions modulate genetic variation at different geographic scales in a seaweed fly. *Mol. Biol. Evol.* **38**, 3953–3971. (doi:10.1093/molbev/msab143)
27. Srivathsan A, Nagarajan N, Meier R. 2019 Boosting natural history research via metagenomic clean-up of crowdsourced feces. *PLoS Biol.* **17**, e3000517. (doi:10.1371/journal.pbio.3000517)
28. Hartop E, Srivathsan A, Ronquist F, Meier R. 2022 Towards large-scale integrative taxonomy (LIT): resolving the data conundrum for dark taxa. *Syst. Biol.* **71**, 1404–1422. (doi:10.1093/sysbio/syac033)

29. Bickel D. 2009 Why Hilara is not amusing: the problem of open-ended taxa and the limits of taxonomic knowledge. In *Diptera diversity: status, challenges and tools* (eds T Pape, R Meier), pp. 279–301. Leiden, The Netherlands: Brill.
30. DeWaard JR *et al.* 2019 Expedited assessment of terrestrial arthropod diversity by coupling Malaise traps with DNA barcoding. *Genome* **62**, 85–95. (doi:10.1139/gen-2018-0093)
31. Wang WY, Srivathsan A, Foo M, Yamane SK, Meier R. 2018 Sorting specimen-rich invertebrate samples with cost-effective NGS barcodes: validating a reverse workflow for specimen processing. *Mol. Ecol. Resour.* **18**, 490–501. (doi:10.1111/1755-0998.12751)
32. Srivathsan A, Lee L, Katoh K, Hartop E, Kutty SN, Wong J, Yeo D, Meier R. 2021 ONTbarcoder and MinION barcodes aid biodiversity discovery and identification by everyone, for everyone. *BMC Biol.* **19**, 1–21. (doi:10.1186/s12915-021-01141-x)
33. Meier R, Wong W, Srivathsan A, Foo M. 2016 $1 DNA barcodes for reconstructing complex phenomes and finding rare species in specimen-rich samples. *Cladistics* **32**, 100–110. (doi:10.1111/cla.12115)
34. Srivathsan A, Feng V, Suárez D, Emerson B, Meier R. 2023 ONTbarcoder 2.0: rapid species discovery and identification with real-time barcoding facilitated by Oxford Nanopore R10.4. *Cladistics* **40**, 192–203. (doi:10.1111/cla.12566)
35. Puillandre N, Lambert A, Brouilette S, Achaz G. 2012 ABGD, Automatic barcode gap discovery for primary species delimitation. *Mol. Ecol.* **21**, 1864–1877. (doi:10.1111/j.1365-294X.2011.05239.x)
36. Dayrat B. 2005 Towards integrative taxonomy. *Biol. J. Linn. Soc.* **85**, 407–415. (doi:10.1111/j.1095-8312.2005.00503.x)
37. Meier R, Srivathsan M, Oliveira SS, Balbi MIPA, Ang Y, Yeo D, Amorim D de S. 2023 'Dark taxonomy': a new protocol for overcoming the taxonomic impediments for dark taxa and broadening the taxon base for biodiversity assessment. *BioRxiv*. (doi:10.1101/2023.08.31.555664)
38. Amorim D de S, Oliveira SS, Balbi MIPA, Ang Y, Yeo D, Srivathsan A, Meier R. 2023 An integrative taxonomic treatment of the Mycetophilidae (Diptera: Bibionomorpha) from Singapore reveals 115 new species on 730 km^2. *BioRxiv*. (doi:10.1101/2023.09.02.555672)
39. Godfray HCJ. 2002 Challenges for taxonomy. *Nature* **417**, 17–19. (doi:10.1038/417017a)
40. Puillandre N, Modica MV, Zhang Y, Sirovich L, Boisselier M-C, Cruaud C, Holford M, Samadi S. 2012 Large-scale species delimitation method for hyperdiverse groups. *Mol. Ecol.* **21**, 2671–2691. (doi:10.1111/j.1365-294X.2012.05559.x)
41. Wührl L, Pylatiuk C, Giersch M, Lapp F, von Rintelen T, Balke M, Schmidt S, Cerretti P, Meier R. 2021 DiversityScanner: robotic handling of small invertebrates with machine learning methods. *Mol. Ecol. Resour.* **22**, 1626–1638. (doi:10.1111/1755-0998.13567)
42. Valan M, Makonyi K, Maki A, Vondráček D, Ronquist F. 2019 Automated taxonomic identification of insects with expert-level accuracy using effective feature transfer from convolutional networks. *Syst. Biol.* **68**, 876–895. (doi:10.1093/sysbio/syz014)
43. Filgueiras CC, Kim Y, Wickings KG, El Borai F, Duncan LW, Willett DS. 2023 The smart soil organism detector: an instrument and machine learning pipeline for soil species identification. *Biosens. Bioelectron.* **221**, 114417. (doi:10.1016/j.bios.2022.114417)
44. Ärje J *et al.* 2020 Automatic image-based identification and biomass estimation of invertebrates. *Methods Ecol. Evol.* **11**, 922–931. (doi:10.1111/2041-210X.13428)
45. Schneider S, Taylor GW, Kremer SC, Burgess P, McGroarty J, Mitsui K, Zhuang A, deWaard JR, Fryxell JM. 2022 Bulk arthropod abundance, biomass and diversity estimation using deep learning for computer vision. *Methods Ecol. Evol.* **13**, 346–357. (doi:10.1111/2041-210X.13769)
46. Díaz S, Cabido M. 2001 Vive la différence: plant functional diversity matters to ecosystem processes. *Trends Ecol. Evol.* **16**, 646–655. (doi:10.1016/S0169-5347(01)02283-2)
47. Barabás G, Parent C, Kraemer A, Van de Perre F, De Laender F. 2022 The evolution of trait variance creates a tension between species diversity and functional diversity. *Nat. Commun.* **13**, 2521. (doi:10.1038/s41467-022-30090-4)
48. McGill B, Enquist B, Weiher E, Westoby M. 2006 Rebuilding community ecology from functional traits. *Trends Ecol. Evol.* **21**, 178–185. (doi:10.1016/j.tree.2006.02.002)
49. Dawson SK *et al.* 2021 The traits of 'trait ecologists': an analysis of the use of trait and functional trait terminology. *Ecol. Evol.* **11**, 16 434–16 445. (doi:10.1002/ece3.8321)
50. Mouchet MA, Villéger S, Mason NWH, Mouillot D. 2010 Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. *Funct. Ecol.* **24**, 867–876. (doi:10.1111/j.1365-2435.2010.01695.x)
51. Drager KI, Rivera MD, Gibson JC, Ruzi SA, Hanisch PE, Achury R, Suarez AV. 2023 Testing the predictive value of functional traits in diverse ant communities. *Ecol. Evol.* **13**, e10000. (doi:10.1002/ece3.10000)
52. Tewksbury JJ *et al.* 2014 Natural history's place in science and society. *Bioscience* **64**, 300–310. (doi:10.1093/biosci/biu032)
53. Paula DP *et al.* 2016 Uncovering trophic interactions in arthropod predators through DNA shotgun-sequencing of gut contents. *PLoS ONE* **11**, e0161841. (doi:10.1371/journal.pone.0161841)
54. Srivathsan A, Ang A, Vogler AP, Meier R. 2016 Fecal metagenomics for the simultaneous assessment of diet, parasites, and population genetics of an understudied primate. *Front. Zool.* **13**, 17. (doi:10.1186/s12983-016-0150-4)
55. Ang A, Roesma DI, Nijman V, Meier R, Srivathsan A, Rizaldi. 2020 Faecal DNA to the rescue: shotgun sequencing of non-invasive samples reveals two subspecies of Southeast Asian primates to be critically endangered species. *Sci. Rep.* **10**, 9396. (doi:10.1038/s41598-020-66007-8)
56. Masonick P, Hernandez M, Weirauch C. 2019 No guts, no glory: gut content metabarcoding unveils the diet of a flower-associated coastal sage scrub predator. *Ecosphere* **10**, e02712. (doi:10.1002/ecs2.2712)
57. Toju H, Baba YG. 2018 DNA metabarcoding of spiders, insects, and springtails for exploring potential linkage between above- and below-ground food webs. *Zool. Lett.* **4**, 4. (doi:10.1186/s40851-018-0088-9)
58. Baksay S, Andalo C, Galop D, Burrus M, Escaravage N, Pornon A. 2022 Using metabarcoding to investigate the strength of plant-pollinator interactions from surveys of visits to DNA sequences. *Front. Ecol. Evol.* **10**, 735588. (doi:10.3389/fevo.2022.735588)
59. Arstingstall KA, DeBano SJ, Li X, Wooster DE, Rowland MM, Burrows S, Frost K. 2023 Investigating the use of pollen DNA metabarcoding to quantify bee foraging and effects of threshold selection. *PLoS ONE* **18**, e0282715. (doi:10.1371/journal.pone.0282715)
60. Suchan T, Talavera G, Sáez L, Ronikier M, Vila R. 2019 Pollen metabarcoding as a tool for tracking long-distance insect migrations. *Mol. Ecol. Resour.* **19**, 149–162. (doi:10.1111/1755-0998.12948)
61. McPherson C, Avanesyan A, Lamp WO. 2022 Diverse host plants of the first instars of the invasive *Lycorma delicatula*: insights from eDNA metabarcoding. *Insects* **13**, 534. (doi:10.3390/insects13060534)
62. Yeo D, Puniamoorthy J, Ngiam RWJ, Meier R. 2018 Towards holomorphology in entomology: rapid and cost-effective adult-larva matching using NGS barcodes. *Syst. Entomol.* **43**, 678–691. (doi:10.1111/syen.12296)
63. Calvignac-Spencer S, Merkel K, Kutzner N, Kühl H, Boesch C, Kappeler PM, Metzger S, Schubert G, Leendertz FH. 2013 Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Mol. Ecol.* **22**, 915–924. (doi:10.1111/mec.12183)

64. Lee PS, Sing KW, Wilson JJ. 2015 Reading mammal diversity from flies: the persistence period of amplifiable mammal mtDNA in blowfly guts (*Chrysomya megacephala*) and a new DNA mini-barcode target. *PLoS ONE* **10**, e0123871. (doi:10.1371/journal.pone.0123871)
65. Lee PS, Gan HM, Clements GR, Wilson JJ. 2016 Field calibration of blowfly-derived DNA against traditional methods for assessing mammal diversity in tropical forests1. *Genome* **59**, 1008–1022. (doi:10.1139/gen-2015-0193)
66. Drinkwater R, Williamson J, Clare EL, Chung AYC, Rossiter SJ, Slade E. 2021 Dung beetles as samplers of mammals in Malaysian Borneo-a test of high throughput metabarcoding of iDNA. *PeerJ* **9**, e11897. (doi:10.7717/peerj.11897)
67. Frolov AV, Akhmetova LA, Vishnevskaya MS, Kiriukhin BA, Montreuil O, Lopes F, Tarasov SI. 2023 Amplicon metagenomics of dung beetles (Coleoptera, Scarabaeidae, Scarabaeinae) as a proxy for lemur (Primates, Lemuroidea) studies in Madagascar. *Zookeys* **1181**, 29–39. (doi:10.3897/zookeys.1181.107496)
68. Nboyine JA, Boyer S, Saville DJ, Wratten SD. 2019 Identifying plant DNA in the faeces of a generalist insect pest to inform trap cropping strategy. *Agron Sustain. Dev.* **39**, 57. (doi:10.1007/s13593-019-0603-1)
69. Srivathsan A, Loh RK, Ong EJ, Lee L, Ang Y, Kutty SN, Meier R. 2022 Network analysis with either Illumina or MinION reveals that detecting vertebrate species requires metabarcoding of iDNA from a diverse fly community. *Mol. Ecol.* **32**, 6418–6435. (doi:10.1111/mec.16767)
70. Yoneya K, Ushio M, Miki T. 2023 Non-destructive collection and metabarcoding of arthropod environmental DNA remained on a terrestrial plant. *Sci. Rep.* **13**, 7125. (doi:10.1038/s41598-023-32862-4)
71. Krehenwinkel H *et al.* 2022 Environmental DNA from archived leaves reveals widespread temporal turnover and biotic homogenization in forest arthropod communities. *Elife* **11**, e78521. (doi:10.7554/eLife.78521)
72. Xu CCY, Yen IJ, Bowman D, Turner CR. 2015 Spider web DNA: a new spin on noninvasive genetics of predator and prey. *PLoS ONE* **10**, e0142503. (doi:10.1371/journal.pone.0142503)
73. Grabarczyk EE *et al.* 2023 DNA metabarcoding analysis of three material types to reveal Joro spider (*Trichonephila clavata*) trophic interactions and web capture. *Front. Ecol. Evol.* **11**, 1177446. (doi:10.3389/fevo.2023.1177446)
74. Owings CG, Gilhooly WP, Picard CJ. 2021 Blow fly stable isotopes reveal larval diet: a case study in community level anthropogenic effects. *PLoS ONE* **16**, e0249422. (doi:10.1371/journal.pone.0249422)
75. Owings CG *et al.* 2019 Female blow flies as vertebrate resource indicators. *Sci. Rep.* **9**, 10594. (doi:10.1038/s41598-019-46758-9)
76. Engel P, Moran NA. 2013 The gut microbiota of insects – diversity in structure and function. *FEMS Microbiol. Rev.* **37**, 699–735. (doi:10.1111/1574-6976.12025)
77. Jing T-Z, Qi F-H, Wang Z-Y. 2020 Most dominant roles of insect gut bacteria: digestion, detoxification, or essential nutrient provision? *Microbiome* **8**, 38. (doi:10.1186/s40168-020-00823-y)
78. Jang S, Kikuchi Y. 2020 Impact of the insect gut microbiota on ecology, evolution, and industry. *Curr. Opin. Insect Sci.* **41**, 33–39. (doi:10.1016/j.cois.2020.06.004)
79. Gwenzi W, Chaukura N, Muisa-Zikali N, Teta C, Musvuugwa T, Rzymski P, Abia ALK. 2021 Insects, rodents, and pets as reservoirs, vectors, and sentinels of antimicrobial resistance. *Antibiotics* **10**, 68. (doi:10.3390/antibiotics10010068)
80. Chevrette MG *et al.* 2019 The antimicrobial potential of *Streptomyces* from insect microbiomes. *Nat. Commun.* **10**, 516. (doi:10.1038/s41467-019-08438-0)
81. Pasteur N, Raymond M. 1996 Insecticide resistance genes in mosquitoes: their mutations, migration, and selection in field populations. *J. Heredity* **87**, 444–449. (doi:10.1093/oxfordjournals.jhered.a023035)
82. Hsu J-C, Chien T-Y, Hu C-C, Chen M-JM, Wu W-J, Feng H-T, Haymer DS, Chen C-Y. 2012 Discovery of genes related to insecticide resistance in *Bactrocera dorsalis* by functional genomic analysis of a de novo assembled transcriptome. *PLoS ONE* **7**, e40950. (doi:10.1371/journal.pone.0040950)
83. Oakeshott JG, Home I, Sutherland TD, Russell RJ. 2003 The genomics of insecticide resistance. *Genome Biol.* **4**, 202. (doi:10.1186/gb-2003-4-1-202)
84. Srivathsan A *et al.* 2023 Convergence of dominance and neglect in flying insect diversity. *Nat. Ecol. Evol.* **7**, 1012–1021. (doi:10.1038/s41559-023-02066-0)
85. Meier R, Hartop E, Pylatiuk C, Srivathsan A. 2024 Towards holistic insect monitoring: species discovery, description, identification, and traits for all insects. Figshare. (doi:10.6084/m9.figshare.c.7151304)