Contrasting results of multiple species delimitation approaches cause uncertainty in synecological studies: A case study on Sri Lankan chafers

Article *in* Insect Conservation and Diversity · August 2023

DOI: 10.1111/icad.12684

CITATIONS
0

READS
120

5 authors, including:

U. G. S. L. Ranasinghe
National Institute of Fundamental Studies - Sri Lanka
24 PUBLICATIONS **60** CITATIONS
SEE PROFILE

Jana Thormann
Research Museum Alexander Koenig
11 PUBLICATIONS **98** CITATIONS
SEE PROFILE

Jonas Eberle
Research Museum Alexander Koenig
80 PUBLICATIONS **776** CITATIONS
SEE PROFILE

Suresh P Benjamin
National Institute of Fundamental Studies - Sri Lanka
124 PUBLICATIONS **1,578** CITATIONS
SEE PROFILE

Received: 18 February 2023 | Accepted: 28 July 2023

DOI: 10.1111/icad.12684

ORIGINAL ARTICLE

Insect Conservation and Diversity

Contrasting results of multiple species delimitation approaches cause uncertainty in synecological studies: A case study on Sri Lankan chafers

U. G. S. L. Ranasinghe[1,2] | J. Thormann[1] | S. P. Benjamin[1,2] | A. Bezděk[3] | J. Eberle[1,4] | Dirk Ahrens[1]

[1]Zoological Research Museum A. Koenig, Leibniz Institute for the Analysis of Biodiversity Change (LIB), Bonn, Germany

[2]National Institute of Fundamental Studies, Kandy, Sri Lanka

[3]Biology Centre of the Czech Academy of Sciences, Institute of Entomology, České Budějovice, Czech Republic

[4]Department Environment & Biodiversity, University of Salzburg, Salzburg, Austria

Correspondence
Dirk Ahrens, Zoological Research Museum A. Koenig, Bonn, Leibniz Institute for the Analysis of Biodiversity Change (LIB), Adenauerallee 127, 53113 Bonn, Germany.
Email: d.ahrens@leibniz-lib.de; ahrens.dirk_col@gmx.de

Funding information
Deutscher Akademischer Austauschdienst; National Institute for Fusion Science

Editor: Nusha Keyghobadi and Associate Editor: Yves Basset

Abstract

1. Biodiversity patterns are the sum of multiple overlapping species distributions. Their analysis therefore requires proper species inference. DNA-based species delimitation has become increasingly popular for such assessments, but their robustness is often problematic due to incongruence between multiple delimitation approaches.
2. Here, we explored how contrasting results of different species' delimitations translate into conclusions of synecological studies, exemplified by assemblages of phytophagous scarab beetles in Sri Lanka from different elevations and forest types. Particularly, we compared estimates based on complete assemblages and on cumulated species inventories inferred from individually analysed subclades.
3. The patterns of assemblage similarity were analysed across different spatial scales with reference to morphospecies, molecular operational taxonomic units (MOTUs) and haplotypes.
4. Method-related ambiguity of species (MOTUs) estimates, which also included subclade inferences, severely affected the certainty of apparent biodiversity patterns at most spatial scales. Even more contrasting patterns resulted from individual clade-wise analyses of faunal similarity or even from cumulated species inventories from individual clade-wise species delimitation analysis.
5. In this case study of tropical chafer beetles, haplotypes provided only very little explanatory information, since genetically highly diverse populations widely lacked shared haplotypes. Therefore, searching proper species boundaries should be the ultimate goal of biodiversity assessments to lend an enduring meaning to biodiversity research and its sustainable application.
6. Our study underlines the need for awareness when synecological observations from different studies are integrated, which use different species delimitation methods for their biodiversity assessment, and its potential impact on conservation management.

KEYWORDS
assemblage similarity, DNA-based species delimitation, integrative taxonomy, MOTUs, Scarabaeidae, Sri Lanka

INTRODUCTION

DNA-based species identification has become a standardised and broadly used molecular approach for rapid biodiversity assessments, due to compelling advantage over traditional approaches in terms of speed and automation (Gostel & Kress, 2022; Sun et al., 2016). Their use for biodiversity surveys compared to conventional taxonomy appears immense, including metabarcoding of whole organism communities (Creedy et al., 2022), environmental or extra-organismal DNA (eDNA; Taberlet et al., 2012) and ingested DNA (iDNA; Schnell et al., 2012). All of them use high-throughput sequencing approaches (Leray & Knowlton, 2015), and are often also understood as the 'next-generation' biodiversity assessment (Creedy et al., 2022; Elbrecht & Leese, 2015; Huang et al., 2022; Steinke et al., 2022). This systematic large-scale DNA sequencing of entire communities allows the assessment of molecular diversity as well as the variation in community composition at the species level as well as their infraspecific variation (Baselga et al., 2015; Bush et al., 2019) and its rapid application for conservation management, particularly in highly diverse groups such as insects.

However, the success of DNA-based species inference heavily depends on a distinction between intraspecific and interspecific genetic variation across taxa (Phillips et al., 2019; Phillips et al., 2022) which is often referred to as barcode gap (Meyer & Paulay, 2005; Puillandre et al., 2012; Puillandre et al., 2021; Ratnasingham & Hebert, 2013). For the recognition of this gap, each study requires sufficient sampling effort to capture adequate levels of within-species genetic variation (Eberle et al., 2020; Meyer & Paulay, 2005; Phillips et al., 2022). Independently of the type and number of markers used for the species delimitation, researchers from early on have recognised a certain incongruence between the outcome of DNA-based identification and morphology-based species assignments (e.g. Ahrens et al., 2016; Esselstyn et al., 2012; Fujisawa & Barraclough, 2013; Ratnasingham & Hebert, 2013). This was the reason why (1) the term molecular operational taxonomic units (MOTUs) was introduced, which pragmatically defines groups of individuals by similarity that can, but often do not represent true species (Blaxter et al., 2005; Floyd et al., 2002); and (2) an integrative taxonomy and species delimitation was propagated (e.g., Carstens et al., 2013; Padial et al., 2010; Schlick-Steiner et al., 2010). Such an integrative framework would incorporate multiple lines of evidence but also alternative delimitation approaches and methods (e.g. threshold-based, character-based, tree- and coalescence-based methods) (Fujita et al., 2012; Kapli et al., 2017; Puillandre et al., 2012; Puillandre et al., 2021; Templeton, 2001; Templeton et al., 1992; Will et al., 2005; Zhang et al., 2013). Beyond that, it was found that inherent characteristics of the species and assemblages itself such as fluctuating effective population size (Ahrens et al., 2016; Esselstyn et al., 2012; Fujisawa & Barraclough, 2013) and rareness of species (Coddington et al., 2009; Ahrens et al., 2016; Lim et al., 2012) may have an important impact on the outcome of the species inference in which sampling design (i.e. the extent of taxa studied in a simultaneous species inferences step) becomes a crucial issue (Ahrens et al., 2016; Luo et al., 2018; Zhou et al., 2019).

To bypass some of these difficulties with the accurate DNA-based species delimitation or its incongruence with morphospecies, particularly with mitochondrial DNA (mtDNA) and single marker data (e.g. *COI*), using haplotype data alone has been proposed as an unbiased and even more objective measure for biodiversity (Baselga et al., 2015; García-Lopez et al., 2013; Papadopoulou et al., 2011; Uscanga et al., 2021). A haplotype-based assessment of biodiversity was assumed to be completely independent from species concepts or delimitation methods including their assumptions (Thormann et al., 2016). Such an approach was currently in use in several ecological and metabarcoding studies (Andujar et al., 2022; Gálvez-Reyes et al., 2020; Noguerales et al., 2021). It has been shown to work well for exploring macroecological patterns in poorly known biota (Baselga et al., 2013; Papadopoulou et al., 2011) and predicting large-scale biodiversity patterns by using haplotype diversity as a proxy for genetic and species diversity (Papadopoulou et al., 2011; see also Cruz-Salazar et al., 2021).

Among the different species delimitation methods proposed so far, choosing a particular method has a considerable effect on estimated species entities and thus also on species richness estimates (Ahrens et al., 2016; Eberle et al., 2020; Zhou et al., 2019). With the ongoing employment of integrative approaches using multiple species delimitation methods, it became clear that neither of these always perfectly matches the morphospecies entities (e.g. Ahrens et al., 2013; Ahrens et al., 2016; Dalstein et al., 2019; Eberle et al., 2020; Lukic et al., 2021) nor do they rarely ever match among each other (Ranasinghe et al., 2022b). Although metabarcoding approaches use mainly distance-based clustering algorithms with predefined thresholds for species circumscription (Callahan et al., 2017; Piper et al., 2019), several different pipelines or 'cluster parameter values' are in use that may reveal alternative outcome (e.g. Alberdi et al., 2018; Bailet et al., 2020; Creedy et al., 2019; Potter et al., 2017). The threshold values used for sequence filtering and the number of reads for the identified MOTUs have an effect on the assessment accuracy of data (Edgar & Flyvbjerg, 2015; Meyer et al., 2021; Piper et al., 2019; Piper et al., 2022; Potter et al., 2017). The choice of the clustering threshold at 3% pairwise distance, as applied by the majority of studies (Alberdi et al., 2018; Elbrecht et al., 2017; Yu et al., 2012), or at 2% for similar taxa (Beentjes et al., 2019; Rossini et al., 2016; Smith et al., 2005) can have a significant impact on taxonomic inferences. Each step can potentially introduce its own sources of artefacts and biases which may inflate or deflate sample diversity estimates (Liu et al., 2020; Zinger et al., 2019).

In this study, we investigate the impact of alternative species estimates (MOTUs) on synecological analyses being applied to different

sample entities of spatial scale (regional to local), in the example of diverse tropical scarab chafer assemblages in Sri Lanka. Synecological studies bring together diversity measures at different collection points and integrate them into spatial entities with similar characteristics, such as species number and composition, which is why the methods that examine faunal similarity are applied in ecology as well as in biogeography. Outcome from both is important for providing a robust and stable reference point with biodiversity assessments, particularly for those that have potential impact on decisions of conservation management (e.g. Floren et al., 2020; Ji et al., 2013; Uscanga et al., 2021; van Jaarsveld et al., 1998; Yang et al., 2021).

MATERIALS AND METHODS

Specimen sampling

The impact of alternative MOTU estimates for synecological analyses was investigated using Sri Lankan species assemblages of phytophagous scarab chafers (Coleoptera: Scarabaeidae) (Ahrens et al., 2014). These beetles are polyphagous herbivorous and generally nocturnal, and represented in Sri Lanka by principally three subfamilies: Rutelinae, Melolonthinae and Dynastinae (exceptionally also by Cetoniinae). Sampling of adult beetles was carried out at 15 localities during 2019–2020 (Ranasinghe et al., 2020; Ranasinghe et al., 2022a) which included different forest types (evergreen wet lowland forest, evergreen dry lowland forest, sub-montane forest and montane forest) and elevational zones. Beetles were captured using six UV-light traps per locality. All specimens (for collection details, see Table S1) were preserved in 96% ethanol after collecting. The collected specimens were presorted to morphospecies based on genital morphology.

DNA sequencing

Three to seven individuals of each morphospecies per location were selected for DNA extraction and subsequent sequencing (in total 565 individuals). DNA was extracted from mesothoracic leg and attached muscles using the Qiagen® DNeasy Blood and Tissue Kit (Hilden, Germany) or the Qiagen® BioSprint 96 magnetic bead extractor (Hilden, Germany). Lab work followed the standard protocols of the German Barcode of Life project (Geiger et al., 2016). The primers LCO1490-JJ [5′-CHACWAAYCATAAAGATATYGG-3′] and HCO2198-JJ [5′-AWACTTCVGGRTGVCC AAARAATCA-3′] (Astrin & Stüben, 2008) were used to amplify a 658 bp fragment at the 5′-end of the mitochondrial gene cytochrome c oxidase subunit 1. Polymerase chain reactions (PCRs) of 90 samples were performed using the QIAGEN® Multiplex PCR Kit. The amplification products were subsequently checked by electrophoresis on a 1.5% agarose gel containing GelRed®. Successfully amplified DNA fragments were purified using Illustra™ ExoProStar™ Enzymatic PCR and Sequencing Clean-Up Kit. Forward and reverse strands were sequenced by Macrogen Europe (Macrogen, Amsterdam, the Netherland; www.macrogen.com). PCRs for 475 samples were done in 96-well plates. Unpurified PCR products were subsequently sent for purification and bidirectional Sanger sequencing to BGI Tech Solutions (Hongkong, China). Sequences were assembled, edited and aligned using Geneious R7 (version 7.1.9, Biomatters Ltd.). All data are deposited in BOLD (projects: SCOIB, PLEUBARC) and GenBank (accession numbers MW698204 – MW698469 (Sericini; Ranasinghe et al., 2022a) and OP270279 – OP270470 (other taxa; see Table S1)).

Phylogenetic analysis

Maximum likelihood (ML; Felsenstein, 1973) searches were performed in IQ-TREE version 1.6.12 (Nguyen et al., 2015) under the (GTR + F + I + G4) model of nucleotide substitution that was inferred as the best fit model by ModelFinder (Kalyaanamoorthy et al., 2017). A total of 1000 ultrafast bootstrap (Hoang et al., 2018) replicates were done to assess branch supports. Then, the tree search was repeated 10 times with the above parameters and the tree with the highest likelihood was selected for further analysis.

Species delimitation

DNA-based species delimitation was performed using three different approaches: (1) the multi-rate Poisson tree processes (mPTP) model (Zhang et al., 2013). mPTP was performed on its dedicated web server (https://mptp.h-its.org/#/tree; accessed on 21 December 2021). It is an improved method of PTP which does not require user-defined parameters (Kapli et al., 2017). Using a Markov chain Monte Carlo (MCMC) sampling approach, it computes support values for each clade, which can be used to assess the confidence of the ML delimitation. The IQ-TREE result from previous phylogenetic analysis was used as input for all PTP analysis. (2) Statistical parsimony analysis (Templeton, 2001; Templeton et al., 1992) was performed as implemented in TCS v. 1.21 (Clement et al., 2000). The procedure partitions the sequence data into clusters, that is, subgroups (or networks) of closely related haplotypes connected by changes with a <95% probability to be non-homoplastic. Resulting networks have been found to be largely congruent with morphospecies at the 95% threshold (Ahrens et al., 2007; Meier et al., 2006) and are considered here as MOTUs. (3) Assemble Species by Automatic Partitioning (ASAP) (Puillandre et al., 2021) which was conducted using the ASAP web server (https://bioinfo.mnhn.fr/abi/public/asap/) (Puillandre et al., 2021) and the distance matrix generated by IQ-TREE. ASAP divides species partitions based on pairwise genetic distances. ASAP also computes a probability of panmixia (*p*-val), a relative gap width metric (W) and ranked results by the ASAP-score: the lower the score, the better the partition (Puillandre et al., 2021). The number of MOTUs predicted by ASAP's best score was selected and compared with other methods. Furthermore, we employed clustering algorithms similar to those used in metabarcoding approaches, to explore the reliability of this critical step in current metabarcoding analysis pipelines

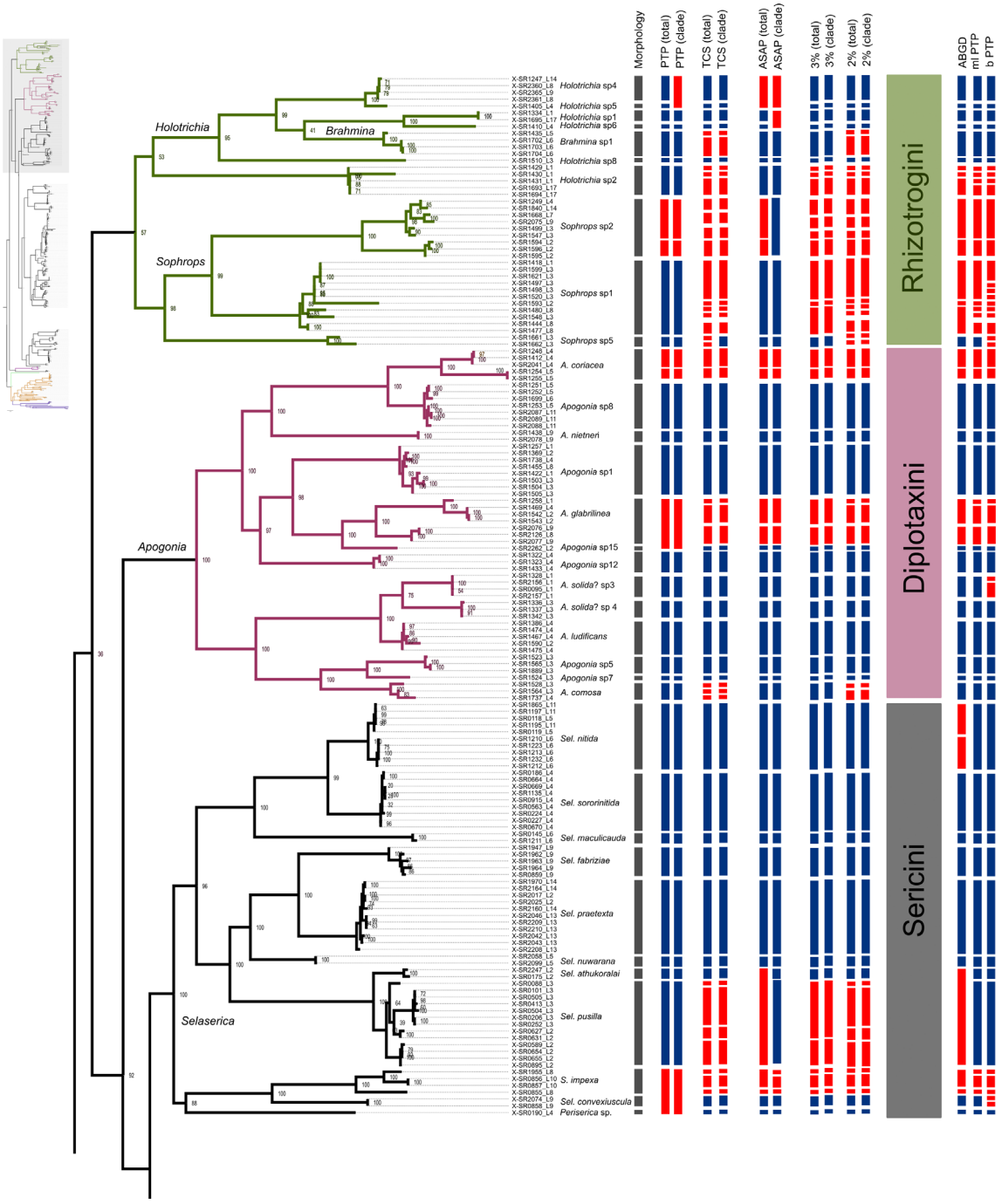

FIGURE 1 Maximum likelihood tree with information about morphospecies assignments, sampling locations, results of species delimitations (mPTP, TCS, ASAP, 3%, 2% clustering). Blue boxes indicate agreement between molecular species delimitation method and morphospecies assignment, red boxes indicate disagreement. Results from subclade analyses are shown in a separate column indicated by 'clade'. Ultrafast bootstrap supports (%) >50 are shown next to the branches.

(Macher et al., 2018). Distance-based clustering was done with the R-package (v. 4.1.2) SPIDER (v. 1.5.0; Brown et al., 2012). A threshold of 3% was applied to the pairwise distance matrix of all specimens obtained from MEGA X (p-distance) (Kumar et al., 2018).

Since it is known that also the tree depth, that is, the phylogenetic extent of sampling, may impact species delimitation analysis (Ahrens et al., 2016), we reanalysed the current dataset for its four principal monophyletic lineages as evident from the maximum likelihood tree obtained with IQ-TREE: clade 1: Rhizotrogini (+ Leucopholini); clade 2: *Apogonia* spp. (Diplotaxini); clade 3: Sericini; clade 4: Rutelinae + Dynastinae + Cetoniinae. Clade 1–3 formally comprises the subfamily Melolonthinae. The latter, however, in most current molecular phylogenies does not result monophyletic (Ahrens et al., 2014; McKenna et al., 2019). On these four subclades, the same delimitation methods

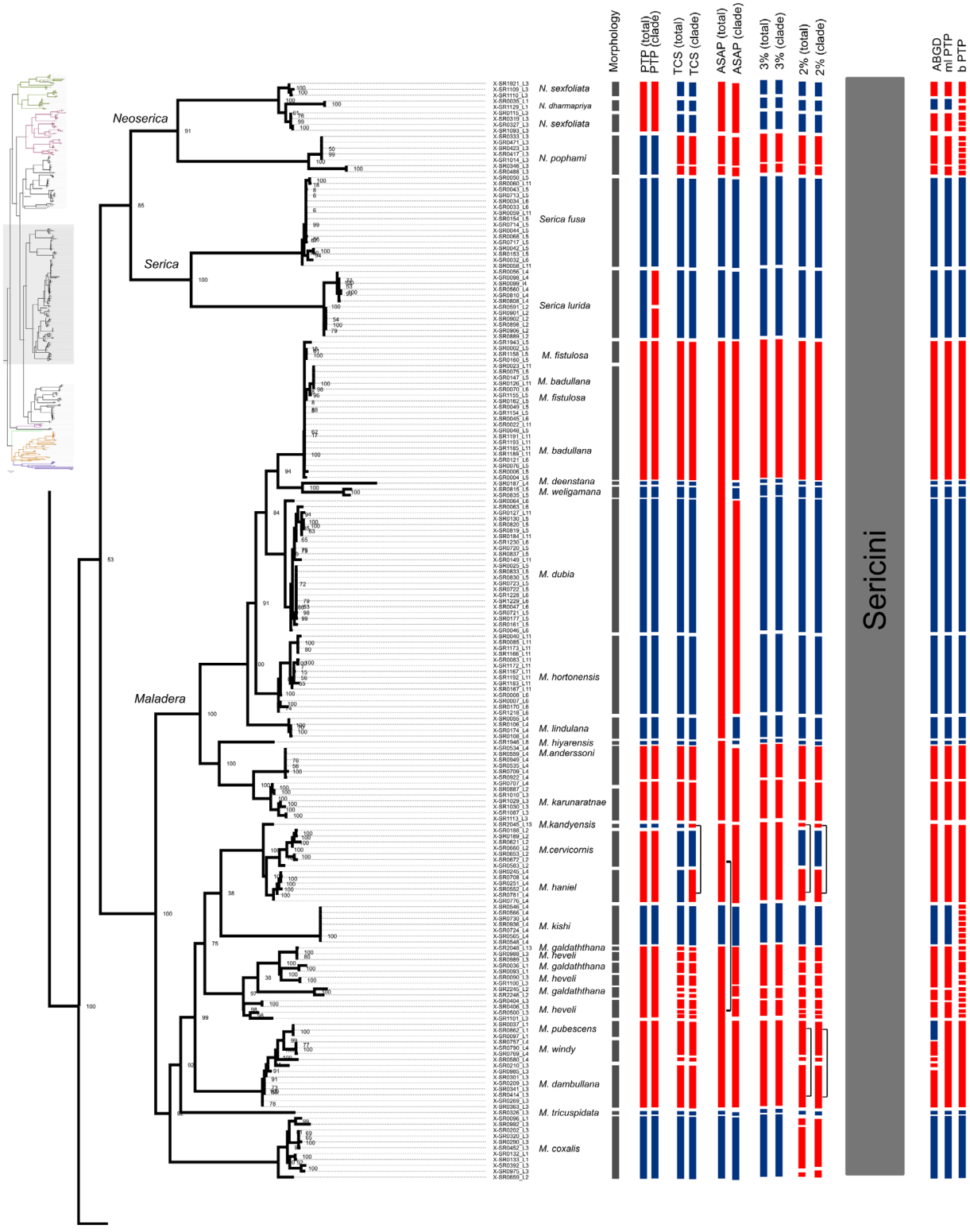

FIGURE 1 (Continued)

as described above were applied (i.e. mPTP, TCS, ASAP and 3% clustering).

The accuracy of DNA-based methods with prior morphospecies assignment was assessed by the match ratio (Ahrens et al., 2016): $2 \times N_{match}/(N_{mol} + N_{morph})$, where N_{match} is the number of exact matches of morphospecies (all individuals) with MOTUs of different delimitation methods, N_{mol} is the number of MOTUs that resulted from different delimitation methods and N_{morph} is the number of morphospecies (Table S2). All morphospecies were mapped onto the terminals of the maximum likelihood tree and MOTUs obtained from

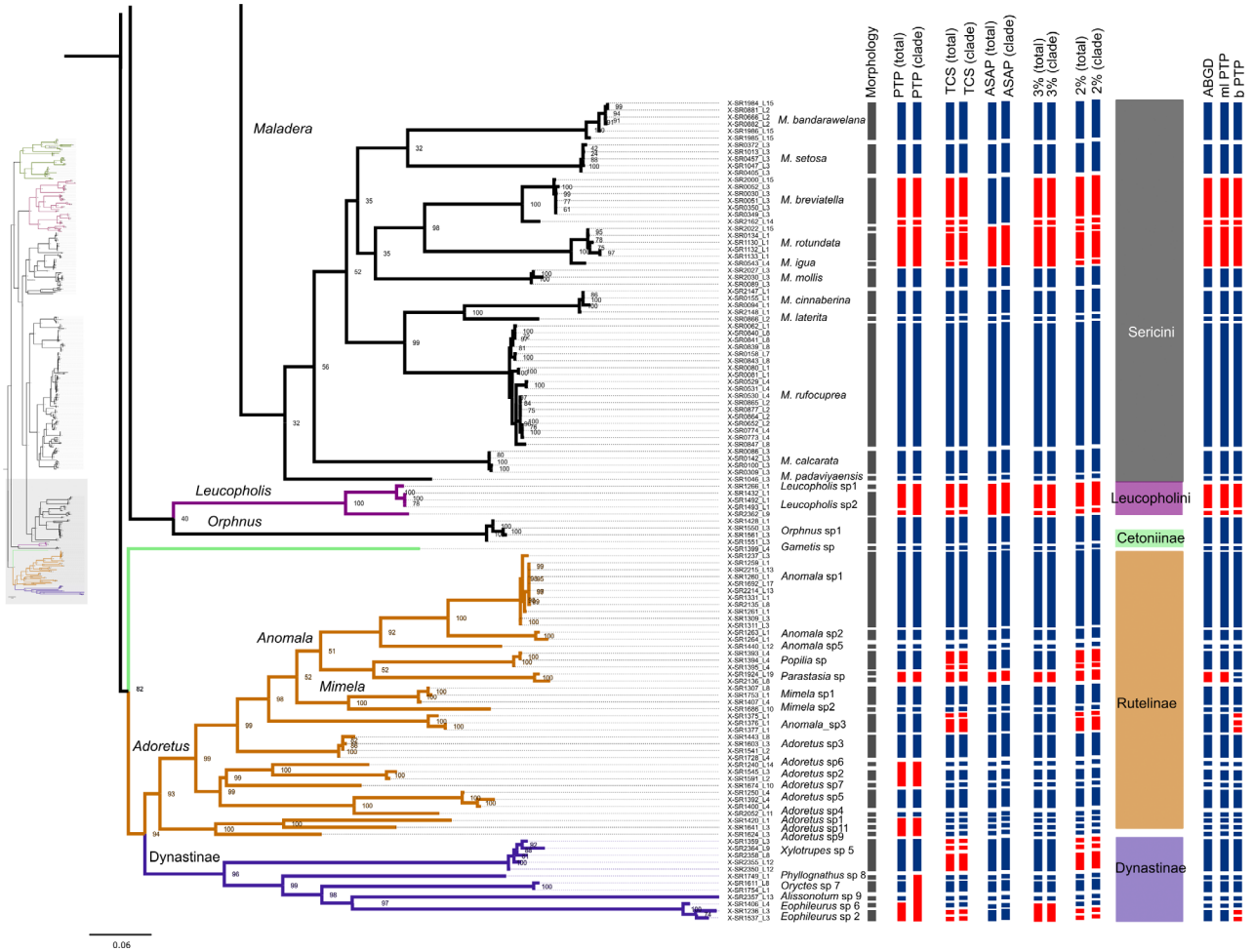

FIGURE 1 (Continued)

different species delimitation methods, including subclade analysis (Figure 2). Furthermore, the match ratios for all pairs of delimitation methods were calculated analogously as explained above and compared in a similarity matrix. Subsequently, the matrix was transformed into a distance matrix and a principal coordinate analysis (PCoA) was performed in PAST v.3.25 (Hammer et al., 2001) to visualise the similarity of outcome between the different methods (Ranasinghe et al., 2022a). The same was done for species inventories resulting from individual subclade analyses and cumulated inventories from individually performed species delimitation analyses on subclades, in which entities of each subclade delimitation were taken again together for the entire assemblage, to explore whether species delimitation on subclades alone affected the overall outcome of delimitation analyses.

In addition, a few more alternative species delimitation approaches were run for the full assemblage data, for which, however, we did not perform synecological analyses, but we compared only the outcome of species delimitation. Poisson tree process modelling was performed using the maximum likelihood implementation (hereafter mlPTP; Zhang et al., 2013) with a single Poisson distribution, as well as the Bayesian implementation (bPTP), which adds Bayesian support (pp) values for putative species to branches in the input tree. Automatic Barcode Gap Discovery (ABGD) was conducted using the ABGD web server (https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html) using the distance matrix generated through IQ-TREE analysis with default parameters (i.e. a relative gap width of 1 and 50 steps, $p_{min} = 0.001$, $p_{max} = 0.1$, Nb bins for distance distribution = 20) (Puillandre et al., 2012). And finally, we performed again distance-based clustering, but with a 2% threshold (performed again in SPIDER). Then match ratio was calculated as above.

Synecological analysis

The analysis of species diversity and assemblage composition was performed for different spatial levels (forest type, elevation, locality) using morphospecies and the different MOTUs, namely haplotypes, mPTP clusters, TCS networks, ASAP clusters as well as 3% and 2% clusters. Species and MOTUs composition of each of these spatial entities was assessed for the entire assemblage, for individual subclades and for cumulative MOTUs of subclades. Forest types included four entities: (a) evergreen wet lowland forests, (b) evergreen dry

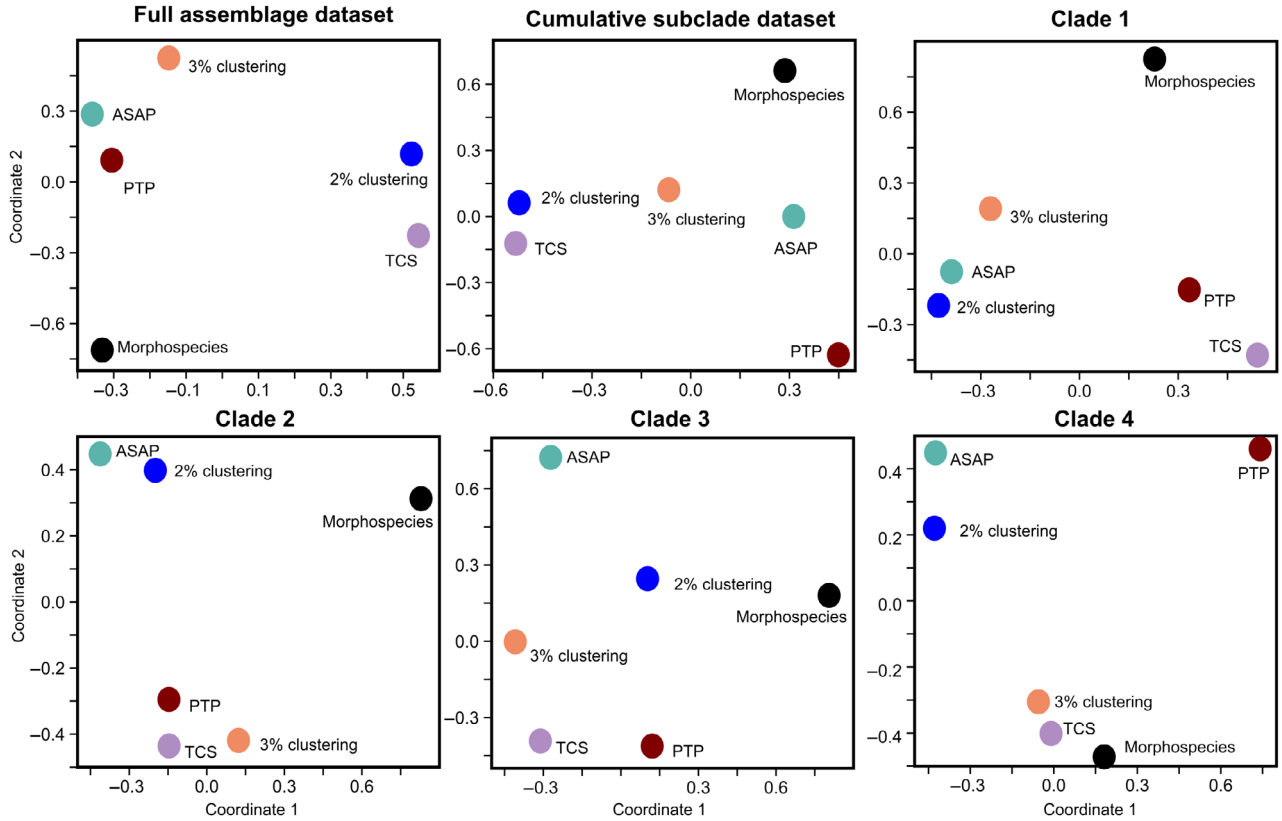

FIGURE 2 Principal coordinate analyses (PCoA) of results of different species delimitation methods based on pairwise match ratios for the total assemblage, individual subclades and cumulative subclade analyses. Clade 1: Rhizotrogini (+ Leucopholini); clade 2: *Apogonia* spp. (Diplotaxini); clade 3: Sericini; clade 4: Rutelinae + Dynastinae + Cetoniinae.

lowland forests, (c) sub-montane forests and (d) montane forests. Elevational zones (EZs) included five entities: EZ1: 0–500 m, EZ2: 501–1000 m, EZ3: 1001–1500 m, EZ4: 1501–2000 m and EZ5: 2001–2500 m. Localities included all 15 individual sampling localities. The dissimilarity in species/MOTU composition of entities at each spatial level was measured using the Jaccard index (i.e. presence-absence data) as implemented in PAST v. 3.25 (Hammer et al., 2001). Results were visualised using non-metric multidimensional scaling (NMDS) and single linkage clustering (based on Jaccard index) at each spatial level.

Endemicity (unique vs. shared occurrences) for entities at each spatial level was calculated for the morphospecies and MOTUs for the entire assemblage, single subclades and cumulated species inventories from individually performed analyses of species delimitation for subclades.

RESULTS

Of 4901 sampled specimens, 565 individuals were sequenced, of which 458 (81%) specimens representing 101 morphospecies were sequenced successfully and included 332 distinct haplotypes. Of the 101 morphospecies, 27 were singletons (26.7% of all species), that is, species represented by one specimen. Sixteen morphospecies (i.e. 15.8% of all species) had infraspecific distances larger than 3% (Figure S5). Thirty-three morphospecies (or 12 haplotypes out of 332) were represented from more than one locality, but no one was found in more than half of all localities. The resulting maximum likelihood tree (Figure 1) showed general agreement with subfamily and genus level classification (Ahrens et al., 2014; Dietz, Seidel, et al., 2023; McKenna et al., 2019). Monophyletic clades resulted for all tribes (Diplotaxini, Sericini, Rhizotrogini and Leucopholini) and most subfamilies (Rutelinae and Dynastinae), the latter two formed a monophyletic sister clade. Melolonthinae was not monophyletic.

Species delimitation: Full dataset versus subclade datasets

The different species delimitation methods (PTP, TCS, ASAP, 3% and 2% clustering) resulted in different numbers of MOTUs (Table S2). We found relatively limited congruence between MOTUs and morphospecies as well as among the different DNA-based delimitation approaches (Figure 1). None of the employed species delimitation methods identically inferred species partitions, neither did they compared to the prior morphospecies assignments. The total number of MOTUs varied from 82 to 129, compared to the morphospecies count of 101. These contradictions arise from

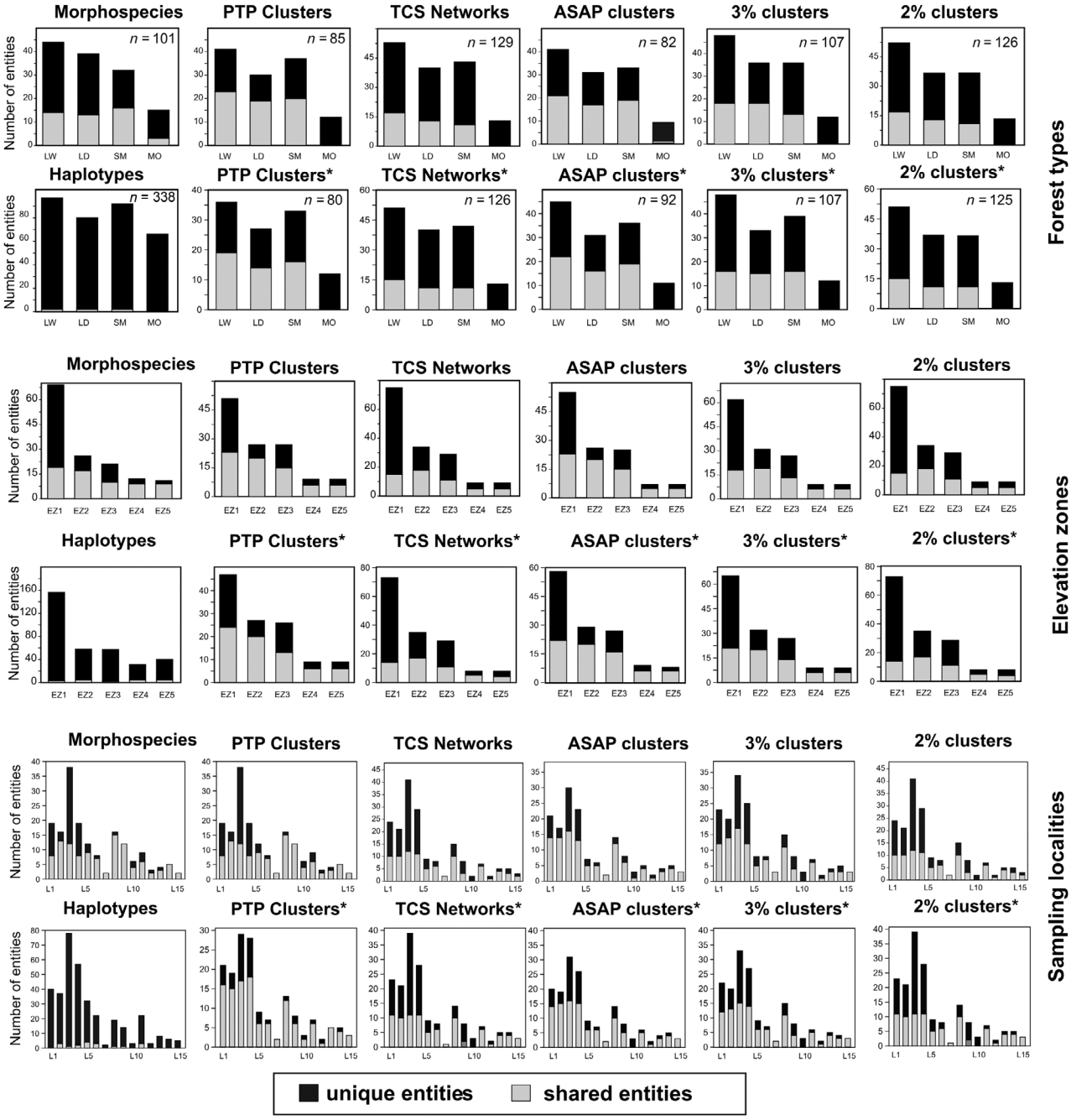

FIGURE 3 Number of 'species entities' reported for morphospecies, haplotypes, PTP-clusters, TCS networks and distance clusters from the total assemblage and from cumulative subclade analyses (indicated by asterisk) for forest types, elevation zones and sampling localities. LW: wet lowland; LD: dry lowland; SM: sub-montane; MO: montane. EZ1: 0–500 m, EZ2: 501–1000 m, EZ3: 1001–1500 m, EZ4: 1501–2000 m, EZ5; 2001–2500 m. L1–L15 sampling localities.

splitting (the individuals of one morphospecies are assigned to two or more different MOTUs) or lumping phenomena (individuals of two or more different morphospecies are fused into one MOTU) (Figure 1). Only 37 MOTUs were obtained from all methods and also perfectly matched morphospecies. Eighty-three morphospecies assignments entirely matched with MOTUs of at least one delimitation method. Out of the 27 singleton morphospecies, 14 were also 'molecular singletons' for all delimitation methods, that is, they were the unique representatives of a MOTU, while remaining 13 singletons were lumped with other specimens into one MOTU.

Compared to the analyses of the full dataset, separate delimitation analyses on individual subclades showed minor differences (ASAP more splitting; PTP and TCS more lumping; 3%, 2% clustering with no differences). The match ratio was higher in few cases (clade 4: ASAP; clade 3: TCS; clades 2 and 4: 3%, 2% clustering) for delimitation analyses on individual subclades (Table S2). The match

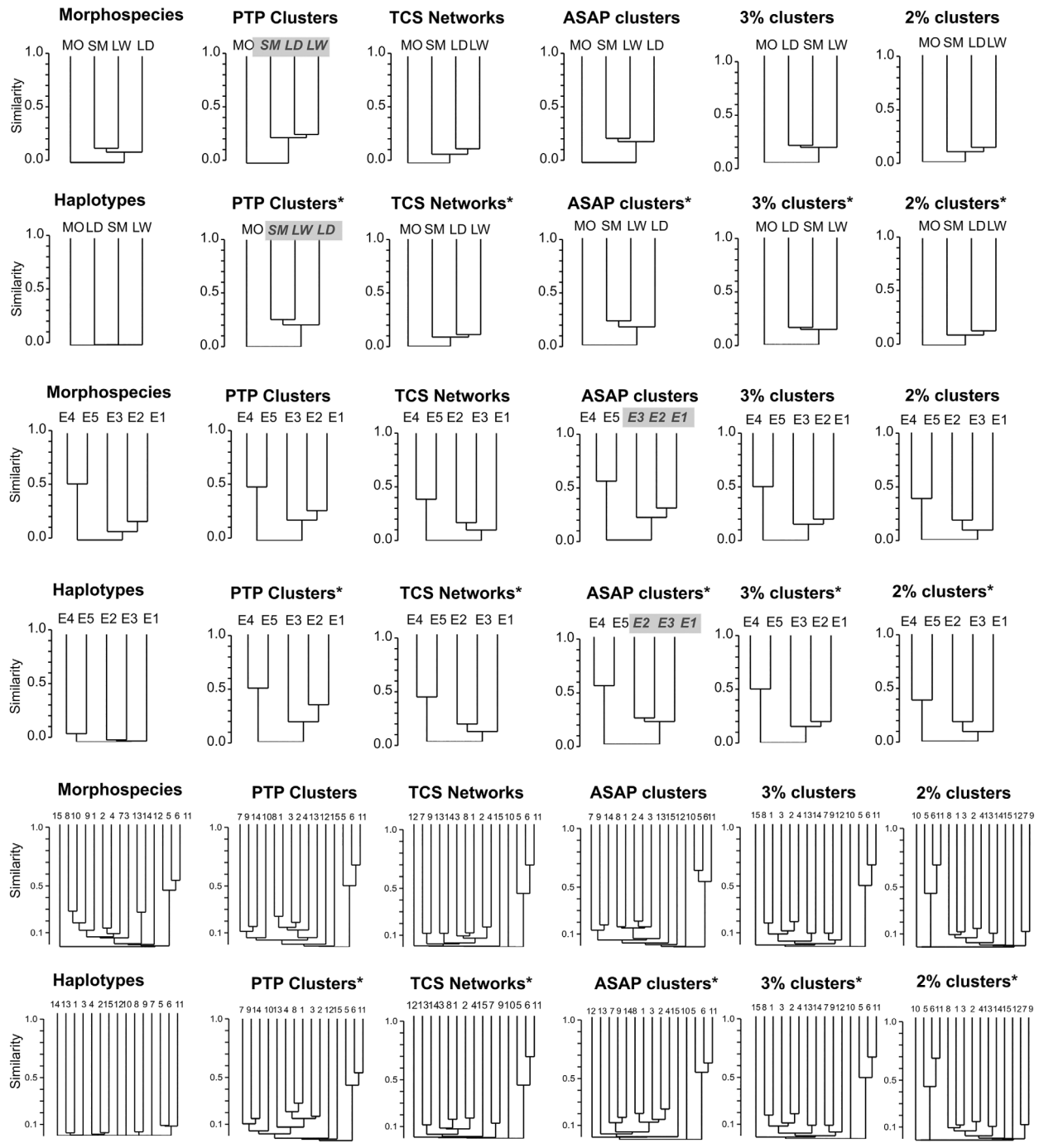

FIGURE 4 Clustering analysis based on the Jaccard index among morphospecies, haplotypes and MOTUs that result from delimitation of the total assemblage and from cumulative subclade analyses (indicated by asterisk) for forest types, elevation zones and sampling localities. Observed differences between total versus cumulative analyses are shown in bold italics. LW: wet lowland; LD: dry lowland; SM: sub-montane; MO: montane. EZ1: 0–500 m, EZ2: 501–1000 m, EZ3: 1001–1500 m, EZ4: 1501–2000 m, EZ5; 2001–2500 m. L1–L15 sampling localities.

ratio compared to the morphospecies assignments was lower for cumulative subclade analyses for ASAP, PTP, TCS, while for 3%, 2% it remained the same (Table S2). The number of summed MOTUs varied from 80 to 126 compared to 82 to 129 for the analysis of the full dataset.

The PCoA ordination based on the similarity of pairwise match ratios of the different delimitation methods, also in relation to their match with morphospecies revealed contrasting patterns between the differently partitioned analyses (Figure 2); that is, the full assemblage dataset, the four individual subclades and for cumulative subclade analyses. None of the DNA-based methods conformed with prior morphospecies matches. Most importantly, the ordination patterns for the full assemblage dataset and cumulative subclade dataset are highly contrasting.

Species richness and assemblage similarity

The operational specimen entities (haplotypes, morphospecies, MOTUs) that were gained for the spatial partitions were similar between the full assemblage and the cumulative subclade analyses within each species delimitation approach, but differed strongly between the different delimitation approaches, also in comparison to the morphospecies (Figure 3, see Figure S2 for subclades). The numbers of unique and shared operational taxonomic entities for the spatial partitions were inconsistent between delimitations. All MOTUs (except one MOTU in ASAP) that occurred in montane forest were not found in other spatial partitions. The most striking result here was that almost all haplotypes were unique to each spatial entity and were restricted to either one forest type, elevation zone or locality, except only few cases. Twelve haplotypes were shared among localities. This occurred in only 10 morphospecies, in each case with single species that shared haplotypes at a maximum of three locations or a single species that shared haplotypes within two forest types (three individual cases) or adjacent elevation zones (10 individual cases). In contrast to exclusive haplotypes, morphospecies generally occurred at more than one locality and also at higher level spatial zones. Some, however, also occurred exclusively at a single entity. Shared morphospecies were observed in 37 cases with a maximum of six localities; 21 cases for forest types; and 30 cases for elevation zones, and no species was found at all elevation zones.

The assemblage's compositional similarity (Jaccard index) was assessed for morphospecies, haplotypes and for MOTUs from the total dataset and cumulative subclade analysis. Outcome of the species delimitation approaches resulted in considerably different species compositions, leading to considerable variation in clustering of spatial partitions (Figure 4; see Figure S3 for subclades). For the clustering of haplotype composition there were of course no alterations between full assemblage dataset or subclades (Figure 4; see Figure S3 for subclades). However, levels of haplotype sharing was by two magnitudes lower than that of MOTUs or of morphospecies (Table S3), which is why relations between spatial entities remain often unresolved, or very weakly connected. Contrasting patterns of assembly composition were obtained between the full assemblage dataset versus the cumulative subclade analyses with PTP (forest types) and ASAP (elevation zone). Otherwise, very similar patterns were observed for the respective pairs of taxon sampling and delimitation strategies, despite generally differing total numbers of MOTUs. Match of compositional similarity with morphospecies pattern was rarely found, however. NMDS based on species compositional similarity (Jaccard index) showed similar results, although relations of entities based on haplotypes (i.e., their NMDS 'ecospace') were based simply on divergences rather than on shared entities (Figure S1; see Figure S4 for subclades).

DISCUSSION

Conflicts among different DNA-based species delimitation approaches are common (Ahrens et al., 2016; Zhou et al., 2019). However, comparative studies to detect such conflicts and to show how this impacts synecological analysis in biodiversity assessments are rare. This is particularly true for studies which investigate beyond species diversity also the similarity of species composition of the assessed probes or sites. Our study showed how contrasting results of species delimitation translate into patterns and conclusions of synecology, exemplified by assemblages of phytophagous scarab beetles in Sri Lanka.

The congruence between the different delimitation methods was rather moderate as number of MOTUs varied from 82 to 129 (also 80–126 for cumulative analysis). TCS, 2%, 3% clustering produced higher species numbers (overestimation), while ASAP and mPTP produced lower numbers (underestimation) than prior morphospecies sorting. Consequently, patterns obtained by morphospecies, DNA-based delimitation methods and haplotypes were strongly contrasting each other in both total diversity and similarity patterns across different spatial scales and species turnover among assemblages (Figure 3). Method-related ambiguity of DNA-based species estimates affected severely the certainty of biodiversity patterns at different spatial scales such as different elevations, forest types or sampling localities. We also demonstrated that differences in assemblage similarity patterns across different spatial scales emerged when DNA-based species estimates were based on the entire assemblage versus cumulatively composed assemblage with MOTUs delimited from subclade datasets.

Subject to variation in delimitation results were especially allopatric, slightly divergent genetic clusters which represent diverging populations or recently separated lineages that may have recently speciated or are still undergoing genetic differentiation (or speciation). Since key diagnostic morphological characters such as the genital organs in most insects are under strong selection, they can evolve even more rapidly than mitochondrial DNA and often result in clearly definable morphospecies while species are not yet sorted by slowly recombining and possibly more slowly diverging mitochondrial haplotypes (Dietz, Eberle, et al., 2023; Eberle et al., 2016; Gibbs, 2017; Thompson, 1998). Additionally, introgression may add which may result in incompatible gene and species trees as well as incongruence between morphospecies and units delimited using mtDNA. However, introgression is presumably more common in closely related species (Dalstein et al., 2019) and may further confound species delimitation and specimen identification using mtDNA barcodes (Gibbs, 2017; Pentinsaari et al., 2014). In the end, these issues may lead to taxonomic incongruence between delimitation approaches and thus to contrasting resulting synecological patterns.

Ecological studies using metabarcoding protocols for their rapid biodiversity assessment may potentially exaggerate this problem. As many DNA barcoding approaches, they generally also use a single gene, such as *COI* as the marker of choice (Deagle et al., 2014) and process homogenised ('souped') mass-collected specimens from large-scale trapping. The resulting huge number of sequences from mass-PCR and high-throughput sequencing (Yu et al., 2012), as all other DNA-based species inference approaches, suffer the same abovementioned problems. Each one of their highly automatised steps can potentially introduce its own sources of artefacts and biases

which may inflate or deflate sample diversity estimates (Liu et al., 2020; Zinger et al., 2019). A majority of metabarcoding studies uses a 3% threshold clustering (Alberdi et al., 2018; Elbrecht et al., 2017; Yu et al., 2012), while alternative clustering thresholds with 2%, which are also used quite often, may exaggerate the effect of oversplitting. Lineage-specific differences in the amount of interspecific divergence between species will lead to either overestimation or underestimation, depending on the cluster settings (Beentjes et al., 2019) which in consequence bias synecological analyses.

That the choice of the 'clustering threshold' during species inference leads to different number of MOTUs (Smith et al., 2005) was also confirmed here (Figure 2). To date it is widely accepted that there is no universal genetic distance threshold value (or 'barcode gap') to group species (Meier et al., 2008; Phillips et al., 2022; Wiemers & Fiedler, 2007). Using a fixed genetic threshold to distinguish taxa with different evolutionary histories may overestimate or underestimate species diversity (Meyer & Paulay, 2005). However, there is growing evidence that the overlap between mean intra- and interspecific genetic distances is considerably greater with larger proportions of closely related taxa, particularly also due to quite commonly occurring introgression (Ballard & Whitlock, 2004; Edelman & Mallet, 2021; Meyer & Paulay, 2005; Sun et al., 2016), but also with increased geographic sampling (Bergsten et al., 2012). Although, new software arises constantly and hardware improves rapidly in order to recover taxonomic information from a wider range of taxa by de novo OTU-picking pipelines (Elbrecht & Leese, 2015; Krehenwinkel et al., 2017; Yu et al., 2012), general problems in species delimitation caused by the fluctuating effective population size, given by the nature of each species (Ahrens et al., 2016), might remain.

In our study, we noted highest levels of intraspecific divergence at 10.5% (Figure S5), while in most species intraspecific divergences were rather low. In the few cases of deep coalescence (e.g. in *Sophrops* sp2, *Apogonia glabrilinea*, *A. coriacea*, *Maladera galdaththana*, *M. heveli*), more than one of the DNA-based species delimitation methods split these morphospecies into different MOTUs. Such species with genetically well-differentiated populations occur in time and space, possibly due to limited gene flow or by low dispersal between populations, for example, due to climatic fluctuation during the Pleistocene in geographically highly structured areas (Arora et al., 2010; Beck et al., 2017; Dincă et al., 2021; French et al., 2022; Lukic et al., 2021; Murria et al., 2017; Voris, 2000). Nevertheless, the nature of maternal inheritance of mtDNA and its very low recombination rate also affects and partly causes the patterns of deep coalescence in mtDNA (Ballard & Whitlock, 2004). This is one major reason encountered for inconsistencies in results of species delimitation using mtDNA compared to morphospecies or nuclear DNA data, which may affect also sympatric populations (Lukic et al., 2021).

Because of these problems, some studies used haplotypes as proxies for genetic diversity in the framework of a haplotype-based macroecology (Baselga et al., 2013; Papadopoulou et al., 2011). They demonstrated the utility of haplotype data for exploring macroecological patterns in poorly known biota and predicting large-scale biodiversity patterns based on genetic inventories of local samples (Arribas et al., 2020; Papadopoulou et al., 2011). Metabarcoding-based haplotyping used this approach as it allows access to the intraspecific diversity and facilitated enhanced biodiversity monitoring (Shum & Palumbi, 2021; Sigsgaard et al., 2020). However, the general correlation between species and genetic diversity has been also questioned (e.g. Hahn et al., 2017; Reisch & Schmid, 2018), since often species diversity mainly depended on habitat conditions, whereas genetic diversity is significantly affected by habitat fragmentation. This would imply that the applicability of a haplotype-based macroecology would be less universal than thought.

Our results showed that the dissimilarity of assemblage composition between haplotypes at diverse spatial levels (Table S3) was extremely high. Almost all haplotypes were unique at any spatial entity, except a few rare cases. Therefore, clustering of haplotype composition (based on shared characteristics, i.e. haplotypes) became almost meaningless due to missing of shared haplotypes. In consequence, basis for the comparative analysis (i.e. PCoA, clustering) was the amount of divergence alone resulting often in a dendrogram-based presentation of the outcome in unresolved polytomies of entities (see Figure 4, Figure S3): this makes haplotypes a poor proxy for compositional comparison of species diversity, at least in hyperdiverse and ancient tropical habitats (Barros et al., 2020; Cruz-Salazar et al., 2021; Rodríguez et al., 2015), for which shared entities are highly informative and should be considered. Haplotypes are simply products of divergence, species often carry more than a single haplotype (or can be even heteroplasmic: i.e. the presence of multiple haplotypes within an individual) (Gibbs, 2017). In contrast to that, species integrate different patterns by dispersal, inheritance, introgression and recombination, all may occur at different intensity in time and space (Epp et al., 2018; Vellend & Geber, 2005), having thus a more complex, integrative and significative meaning than simple diversity, that is, the number of entities (e.g. MOTUs). Moreover, haplotype-derived patterns might be confounded if mtDNA groupings or genealogy does not reflect true species limits or the species' geographic extent (Murria et al., 2017; Paradis, 2018; Salinas-Ivanenko & Múrria, 2021).

CONCLUSION

DNA-based species identification is increasingly used and popular for biodiversity assessments. However, due to incongruent outcomes from various species delimitation methods, particularly when researchers follow an integrative taxonomic approach (Carstens et al., 2013), provide certain potential for ambiguities for synecological studies. Our study showed that such method-related ambiguity of DNA-based species estimates severely affected the patterns of faunal similarity and the certainty of apparent biodiversity patterns. Even more contrasting patterns may result from individual clade-wise analyses of faunal similarity or even from cumulated species inventories from separately performed, individual, clade-wise species delimitation analyses. These potential contrasts revealed here with real-life, empirical data underline the need for awareness when synecological

observations from different studies are integrated, particularly if these used different species delimitation methods or protocols (e.g. clustering thresholds) for their biodiversity assessment. Alternatively, the same problems may occur, if uniform clustering thresholds do not fit with lineage specific species divergence. In this case, MOTUs barely reflect species. Limitations in a purely haplotype-based biodiversity assessment found here reveal, why searching proper species boundaries should be the ultimate goal of biodiversity assessment: to place trust in species delimitations and to give an enduring (and sustainable) meaning to biodiversity research and its sustainable and universal application (Carstens et al., 2013). Integrative and multilayered analyses are possibly the most obvious solution. In this context, morphospecies, especially due to their globally complete and enormous reference system based on morphology (Ahrens et al., 2021) and their good chances of a universal and well-expandable implementation in approaches using artificial intelligence (Borowiec et al., 2022; Klasen et al., 2022; Van Klink et al., 2022), remain a valid and highly valuable proxy for biodiversity research (Padial et al., 2010; Schlick-Steiner et al., 2010; Yeates et al., 2011).

AUTHOR CONTRIBUTIONS

U.G.S.L. Ranasinghe: Conceptualization; methodology; formal analysis; validation; investigation; visualization; project administration; writing – original draft; writing – review and editing; data curation; funding acquisition. **J. Thormann:** Formal analysis; data curation; writing – review and editing; resources. **S.P. Benjamin:** Resources; supervision; writing – review and editing; funding acquisition; investigation. **A. Bezděk:** Writing – review and editing; validation. **J. Eberle:** Writing – review and editing; investigation; conceptualization; supervision. **D. Ahrens:** Conceptualization; methodology; data curation; resources; project administration; writing – review and editing; writing – original draft; funding acquisition; visualization; supervision; validation; investigation.

ACKNOWLEDGEMENTS

Fieldwork for this study was supported by institutional funding of the ZFMK, the Alexander Koenig Stiftung and by the German Academic Exchange Service (DAAD). S.R. was funded by the DAAD; S.P.B. was funded by the NIFS. We thank N. Athukorala (NIFS), S.R. family members, P.L. Dharmapriya, B. Ranasinghe and T.N. Dharmpriya for helping in the fieldwork. Furthermore, fieldwork was supported by C. Bohacz (ZFMK), D. Bopearachchi, A. Sathkunanathan, M. Tharmarajan, C. Jayatissa (all NIFS) and Supipi Wijesundara (UoP). Professor S. Wijesundara and C. Lekamge made available facilities in the NIFS Arboretum. Also, thanks to M. de Silva and Sampath for access of the facilities in the Hiyare Conservation Center and to all wildlife rangers (Nuwara Eliya, Hiyare, Horton Plains), regional forest officers (Kandy, Knuckles, Nuwara Eliya, Kottawa, Kanneliya) who helped us to conduct fieldwork. For providing research and collection permits to S.P.B, we thank the Department of Wildlife Conservation, Sri Lanka (permit no: WL/3/2/61/18), the Department of Forest Conservation, Sri Lanka (permit no: R&E/RES/NFSRCM/2019-01 & R&E/RES/NFSRCM/EXTENSION/2020) and the Divisional Forest Office, Kandy, Sri Lanka (permit no: K/G/01/06/03); Galle (M/0/03/2019). Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data (DNA sequences) are available on NCBI and BOLD databases: GenBank accessions (MW698204-MW698469; OP270279 – OP270470; see Table S1); BOLD databases (projects: SCOIB, PLEUBARC).

ORCID

U. G. S. L. Ranasinghe https://orcid.org/0000-0001-6455-8079
S. P. Benjamin https://orcid.org/0000-0003-4666-0330
A. Bezděk https://orcid.org/0000-0002-8748-3281
J. Eberle https://orcid.org/0000-0003-2519-0640
Dirk Ahrens https://orcid.org/0000-0003-3524-7153

REFERENCES

Ahrens, D., Ahyong, S.T., Ballerio, A., Barclay, M.V.L., Eberle, J., Espeland, M. et al. (2021) Is it time to describe new species without diagnoses? – A comment on Sharkey et al. (2021). *Zootaxa*, 5027(2), 151–159.

Ahrens, D., Fabrizi, S., Šipek, P. & Lago, P. (2013) Integrative analysis of DNA phylogeography and morphology of the European rose chafer (*Cetonia aurata*) to infer species taxonomy and patterns of postglacial colonisation in Europe. *Molecular Phylogenetics and Evolution*, 69, 83–94.

Ahrens, D., Fujisawa, T., Krammer, H.J., Eberle, J., Fabrizi, S. & Vogler, A.P. (2016) Rarity and incomplete sampling in DNA based species delimitation. *Systematic Biology*, 65, 478–494.

Ahrens, D., Monaghan, M.T. & Vogler, A.P. (2007) DNA-based taxonomy for associating adults and larvae in multi-species assemblages of chafers (coleoptera, Scarabaeidae). *Molecular Phylogenetics and Evolution*, 44(1), 436–449.

Ahrens, D., Schwarzer, J. & Vogler, A.P. (2014) The evolution of scarab beetles tracks the sequential rise of angiosperms and mammals. *Proceedings of the Royal Society B*, 281, 20141470.

Alberdi, A., Aizpurua, O., Gilbert, M.T.P. & Bohmann, K. (2018) Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology & Evolution*, 9, 134–147.

Andujar, C., Arribas, P., López, H., Arjona, Y., Perez-Delgado, A., Oromi, P. et al. (2022) Community assembly and metaphylogeography of soil biodiversity: insights from haplotype-level community DNA metabarcoding within an oceanic Island. *Molecular Ecology*, 31, 4078–4094. Available from: https://doi.org/10.1111/mec.16560

Arora, N., Nater, A., van Schaik, C.P., Willems, E.P., van Noordwijk, M.A., Goossens, B. et al. (2010) Effects of Pleistocene glaciations and rivers on the population structure of *Bornean orangutans* (*Pongo pygmaeus*). *Proceedings of the National Academy of Sciences of the United States of America*, 107(50), 21376–21381.

Arribas, P., Andujar, C., Salces-Castellano, A., Emerson, B.C. & Vogler, A.P. (2020) The limited spatial scale of dispersal in soil arthropods revealed with whole-community haplotype-level metabarcoding. *Molecular Ecology*, 30, 48–61.

Astrin, J.J. & Stüben, P.E. (2008) Phylogeny in cryptic weevils: molecules, morphology and new genera of western Palaearctic Cryptorhynchinae (coleoptera: Curculionidae). *Invertebrate Systematics*, 22(5), 503–522.

Bailet, B., Apothéloz-Perret-Gentil, L., Baričević, A., Chonova, T., Franc, A., Frigerio, J. et al. (2020) Diatom DNA metabarcoding for ecological assessment: comparison among bioinformatics pipelines used in six European countries reveals the need for standardization. *Science of the Total Environment*, 745, 140948.

Ballard, J.W. & Whitlock, M.C. (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, 13, 729–744.

Barros, M.J.F., Silva-Arias, G.A., Segatto, A.L.A., Reck-Kortmann, M., Fregonezi, J.N., Diniz-Filho, J.A.F. et al. (2020) Phylogenetic niche conservatism and plant diversification in south American subtropical grasslands along multiple climatic dimensions. *Genetics and Molecular Biology*, 43(2), e20180291.

Baselga, A., Fujisawa, T., Crampton-Platt, A., Bergsten, J., Foster, P.G., Monaghan, M.T. et al. (2013) Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels. *Nature Communications*, 4, 1892.

Baselga, A., Gómez-Rodríguez, C. & Vogler, A.P. (2015) Multi-hierarchical macroecology at species and genetic levels to discern neutral and non-neutral processes. *Global Ecology and Biogeography*, 24, 873–882.

Beck, S.V., Carvalho, G.R., Barlow, A., Ruber, L., Tan, H.H., Nugroho, E. et al. (2017) Plio- Pleistocene phylogeography of the southeast Asian blue panchax killifish, *Aplocheilus panchax*. *PloS One*, 12(7), e0179557.

Beentjes, K.K., Speksnijder, A.G.C.L., Schilthuizen, M., Hoogeveen, M., Pastoor, R. & van der Hoorn, B.B. (2019) Increased performance of DNA metabarcoding of macroinvertebrates by taxonomic sorting. *PLoS One*, 14(12), e0226527.

Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M.T., Balke, M. et al. (2012) The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology*, 61, 851–869.

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R. et al. (2005) Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462), 1935–1943.

Borowiec, M.L., Dikow, R.B., Frandsen, P.B., McKeeken, A., Valentini, G. & White, A.E. (2022) Deep learning as a tool for ecology and evolution. *Methods in Ecology and Evolution*, 13, 1640–1660.

Brown, S.D.J., Collins, R.A., Boyer, S., Lefort, M.C., Malumbres-Olarte, J., Vink, C.J. et al. (2012) SPIDER: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources*, 12, 562–565.

Bush, A., Compson, Z.G., Monk, W.A., Porter, T.M., Steeves, R., Emilson, E. et al. (2019) Studying ecosystems with DNA metabarcoding: lessons from biomonitoring of aquatic macroinvertebrates. *Frontiers in Ecology and Evolution*, 7, 434.

Callahan, B., McMurdie, P. & Holmes, S. (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *International Society for Microbial Ecology Journal*, 11, 2639–2643.

Carstens, B.C., Pelletier, T.A., Reid, N.M. & Satler, J.D. (2013) How to fail at species delimitation. *Molecular Ecology*, 22, 4369–4383.

Clement, M., Posada, D. & Crandall, K.A. (2000) TCS: A computer program to estimate gene genealogies. *Molecular Ecology*, 9, 1657–1660.

Coddington, J.A., Agnarsson, I., Miller, J.A., Kuntner, M. & Hormiga, G. (2009) Undersampling bias: the null hypothesis for singleton species in tropical arthropod surveys. *Journal of Animal Ecology*, 78, 573–584.

Creedy, T.J., Andújar, C., Meramveliotakis, E., Noguerales, V., Overcast, I., Papadopoulou, A. et al. (2022) Coming of age for COI metabarcoding of whole organism community DNA: towards bioinformatic harmonisation. *Molecular Ecology Resources*, 22, 847–861.

Creedy, T.J., Ng, W.S. & Vogler, A.P. (2019) Toward accurate species-level metabarcoding of arthropod communities from the tropical forest canopy. *Ecology and Evolution*, 9(6), 3105–3116.

Cruz-Salazar, B., Ruiz-Montoya, L., Ramírez-Marcial, N. & García-Bautista, M. (2021) Relationship between genetic variation and diversity of tree species in tropical forests in the El Ocote biosphere reserve, Chiapas, Mexico. *Tropical Conservation Science*, 14, 1–14.

Dalstein, V., Eberle, J., Fabrizi, S., Etzbauer, C. & Ahrens, D. (2019) COI-based species delimitation in Indochinese *Tetraserica* chafers reveal hybridisation despite strong divergence in male copulation organs. *Organisms Diversity and Evolution*, 19, 277–286.

Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F. & Taberlet, P. (2014) DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, 10, 20140562.

Dietz, L., Eberle, J., Kukowka, S., Mayer, C., Bohacz, C., Baur, H. et al. (2023) Standardized nuclear markers improve and homogenize species delimitation in Metazoa. *Methods in Ecology and Evolution*, 4, 543–555.

Dietz, L., Seidel, M., Eberle, J., Misof, B., Pacheco, T.L., Podsiadlowski, L. et al. (2023) A transcriptome-based phylogeny of Scarabaeoidea confirms the sister group relationship of dung beetles and phytophagous pleurostict scarabs (Coleoptera). *Systematic Entomology*, 1–15. Available from: https://doi.org/10.1111/syen.12602

Dincă, V., Dapporto, L., Somervuo, P., Vodă, R., Cuvelier, S., Gascoigne-Pees, M. et al. (2021) High resolution DNA barcode library for European butterflies reveals continental patterns of mitochondrial genetic diversity. *Communications Biology*, 4(1), 315.

Eberle, J., Ahrens, D., Mayer, C., Niehuis, O. & Misof, B. (2020) A plea for standardized nuclear markers in metazoan DNA taxonomy. *Trends in Ecology & Evolution*, 35(4), 336–345.

Eberle, J., Warnock, R.C.M. & Ahrens, D. (2016) Bayesian species delimitation in *Pleophylla* chafers (coleoptera) – the importance of prior choice and morphology. *BMC Evolutionary Biology*, 16, 94.

Edelman, N.B. & Mallet, J. (2021) Prevalence and adaptive impact of introgression. *Annual Review of Genetics*, 55, 265–283.

Edgar, R.C. & Flyvbjerg, H. (2015) Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics*, 31, 3476–3482.

Elbrecht, V. & Leese, F. (2015) Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PLoS One*, 10, e0130324.

Elbrecht, V., Vamos, E.E., Meissner, K., Aroviita, J. & Leese, F. (2017) Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution*, 8, 1265–1275.

Epp, L.S., Kruse, S., Kath, N.J., Stoof-Leichsenring, K.R., Tiedemann, R., Pestryakova, L.A. et al. (2018) Temporal and spatial patterns of mitochondrial haplotype and species distributions in Siberian larches inferred from ancient environmental DNA and modeling. *Scientific Reports*, 8, 17436.

Esselstyn, J., Evans, B., Sedlock, J., Anwarali Khan, F. & Heaney, L. (2012) Single-locus species delimitation: a test of the mixed yule-coalescent model, with an empirical application to Philippine round-leaf bats. *Proceedings of the Royal Society B: Biological Sciences*, 279, 3678–3686.

Felsenstein, J. (1973) Maximum likelihood and minimum-step methods for estimating evolutionary trees from data on discrete characters. *Systematic Zoology*, 22, 240–249.

Floren, A., von Rintelen, T., Hebert, P.D.N., de Araujo, B.C., Schmidt, S., Balke, M. et al. (2020) Integrative ecological and molecular analysis indicate high diversity and strict elevational separation of canopy beetles in tropical mountain forests. *Scientific Reports*, 10, 16677.

Floyd, R., Abebe, E., Papert, A. & Blaxter, M. (2002) Molecular barcodes for soil nematode identification. *Molecular Ecology*, 11(4), 839–850.

French, C.M., Bertola, L.D., Carnaval, A.C., Economo, E.P., Kass, J.M., Lohman, D.J. et al. (2022) Global determinants of the distribution of

insect genetic diversity. *bioRxiv* https://doi.org/10.1101/2022.02.09.479762

Fujisawa, T. & Barraclough, T.G. (2013) Delimiting species using single-locus data and the generalized mixed yule coalescent approach: a revised method and evaluation on simulated data sets. *Systematic Biology*, 62, 707–724.

Fujita, M.K., Leaché, A.D., Burbrink, F.T., McGuire, J.A. & Moritz, C. (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution*, 27(9), 480–488.

Gálvez-Reyes, N., Arribas, P., Andujar, C., Emerson, B., Pinero, D. & Mastretta-Yanes, A. (2021) Dispersal limitations and long-term persistence drive differentiation from haplotypes to communities within a tropical sky-island: Evidence from community metabarcoding. *Molecular Ecology*, 30, 6611–6626. Available from: https://doi.org/10.1111/mec.16195

García-Lopez, A., Mico, E., Murria, C., Galante, E. & Vogler, A.P. (2013) Beta diversity at multiple hierarchical levels: explaining the high diversity of scarab beetles in tropical montane forests. *Journal of Biogeography*, 40, 2134–2145.

Geiger, M.F., Moriniere, J., Hausmann, A., Haszprunar, G., Wägele, W., Hebert, P.D.N. et al. (2016) Testing the global malaise trap program – how well does the current barcode reference library identify flying insects in Germany? *Biodiversity Data Journal*, 4, e10671.

Gibbs, J. (2017) DNA barcoding a nightmare taxon: assessing barcode index numbers and barcode gaps for sweat bees. *Genome*, 61(1), 21–31.

Gostel, M.R. & Kress, W.J. (2022) The expanding role of DNA barcodes: indispensable tools for ecology, evolution, and conservation. *Diversity*, 14, 213.

Hahn, C.Z., Michalski, S.G., Fischer, M. & Durka, W. (2017) Genetic diversity and differentiation follow secondary succession in a multi-species study on woody plants from subtropical China. *Journal of Plant Ecology*, 10, 213–221.

Hammer, O., Harper, D.A.T. & Ryan, P.D. (2001) Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, 4, 1–9.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Vinh, L.S. (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35, 518–522.

Huang, J., Miao, X., Wang, Q., Menzel, F., Tang, P., Yang, D. et al. (2022) Metabarcoding reveals massive species diversity of Diptera in a subtropical ecosystem. *Ecology and Evolution*, 12, e8535.

Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A. et al. (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, 16, 1245–1257.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. & Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14, 587–589.

Kapli, T., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A. et al. (2017) Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics*, 33(11), 1630–1638.

Klasen, M., Ahrens, D., Eberle, J. & Steinhage, V. (2022) Image-based automated species identification: can virtual data augmentation overcome problems of insufficient sampling? *Systematic Biology*, 71(2), 320–333.

Krehenwinkel, H., Wolf, M., Lim, J.Y., Rominger, A.J., Simison, W.B. & Gillespie, R.G. (2017) Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Scientific Reports*, 7(1), 1–12.

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35, 1547–1549.

Leray, M. & Knowlton, N. (2015) DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, 112(7), 2076–2081.

Lim, G., Balke, M. & Meier, R. (2012) Determining species boundaries in a world full of rarity: singletons, species delimitation methods. *Systematic Biology*, 61, 165–169.

Liu, M., Clarke, L.J., Baker, S.C., Jordan, G.J. & Burridge, C.P. (2020) A practical guide to DNA metabarcoding for entomological ecologists. *Ecological Entomology*, 45, 373–385.

Lukic, D., Eberle, J., Thormann, J., Holzschuh, C. & Ahrens, D. (2021) Excluding spatial sampling bias does not eliminate over-splitting in DNA-based species delimitation analyses. *Ecology and Evolution*, 11, 10327–10337.

Luo, A., Ling, C., Ho, S.Y.W. & Zhu, C.D. (2018) Comparison of methods for molecular species delimitation across a range of speciation scenarios. *Systematic Biology*, 67, 830–846.

Macher, J.N., Vivancos, A., Piggott, J.J., Centeno, F.C., Matthaei, C.D. & Leese, F. (2018) Comparison of environmental DNA and bulk-sample metabarcoding using highly degenerate cytochrome c oxidase I primers. *Molecular Ecology Resources*, 18(6), 1456–1468.

McKenna, D.D., Shin, S., Ahrens, D., Balke, M., Beza-Beza, C., Clarke, D.J. et al. (2019) The evolution and genomic basis of beetle diversity. *Proceedings of the National Academy of Sciences*, 116, 24729–24737.

Meier, R., Shiyang, K., Vaidya, G. & Ng, P.K.L. (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, 55(5), 715–728.

Meier, R., Zhang, G. & Ali, F. (2008) The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. *Systematic Biology*, 57, 5809–5813.

Meyer, A., Boyer, F., Valentini, A., Bonin, A., Ficetola, G.F., Beisel, J. et al. (2021) Morphological vs. DNA metabarcoding approaches for the evaluation of stream ecological status with benthic invertebrates: testing different combinations of markers and strategies of data filtering. *Molecular Ecology*, 30, 3203–3220.

Meyer, C.P. & Paulay, G. (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, 3, 2229–2238.

Murria, C., Bonada, N., Vellend, M., Zamora-Munoz, C., Alba-Tercedor, J., Elisa Sainz-Cantero, C. et al. (2017) Local environment rather than past climate determines community composition of mountain stream macroinvertebrates across Europe. *Molecular Ecology*, 26(21), 6085–6099.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., & Minh, B.Q. (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32 (1), 268–274. https://doi.org/10.1093/molbev/msu300

Noguerales, V., Meramveliotakis, E., Castro-Insua, A., Andújar, C., Arribas, P., Creedy, T.J. et al. (2021) Community metabarcoding reveals the relative role of environmental filtering and spatial processes in metacommunity dynamics of soil microarthropods across a mosaic of montane forests. *Molecular Ecology*, 00, 1–19.

Padial, J.M., Miralles, A., De la Riva, I. & Vences, M. (2010) The integrative future of taxonomy. *Frontiers in Zoology*, 7, 16.

Papadopoulou, A., Anastasiou, I., Spagopoulou, F., Stalimerou, M., Terzopoulou, S., Legakis, A. et al. (2011) Testing the species–genetic diversity correlation in the Aegean archipelago: toward a haplotype-based macroecology? *The American Naturalist*, 178(2), 241–255.

Paradis, E. (2018) Analysis of haplotype networks: the randomized minimum spanning tree method. *Methods in Ecology and Evolution*, 9, 1308–1317.

Pentinsaari, M., Mutanen, M. & Kaila, L. (2014) Cryptic diversity and signs of mitochondrial introgression in the *Agrilus viridis* species complex (coleoptera: Buprestidae). *European Journal of Entomology*, 111(4), 475–486.

Phillips, J.D., Gillis, D.J. & Hanner, R.H. (2019) Incomplete estimates of genetic diversity within species: implications for DNA barcoding. *Ecology and Evolution*, 9, 2996–3010.

Phillips, J.D., Gillis, D.J. & Hanner, R.H. (2022) Lack of statistical rigor in DNA barcoding likely invalidates the presence of a true species' barcode gap. *Frontiers in Ecology and Evolution*, 10, 859099.

Piper, A.M., Batovska, J., Cogan, N.O.I., Weiss, J., Cunningham, J.P., Rodoni, B.C. et al. (2019) Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. *Gigascience*, 8, giz092.

Piper, A.M., Cunningham, J.P., Cogan, N.O.I. & Blacket, M.J. (2022) DNA metabarcoding enables high-throughput detection of spotted wing *drosophila* (*Drosophila suzukii*) within unsorted trap catches. *Frontiers in Ecology and Evolution*, 10, 822648.

Potter, C., Tang, C.Q., Fonseca, V., Lallias, D., Gaspar, J.M., Thomas, K. et al. (2017) De novo species delimitation in metabarcoding datasets using ecology and phylogeny. *Peer Journal*, 5, e3121v1.

Puillandre, N., Brouillet, S. & Achaz, G. (2021) ASAP: assemble species by automatic partitioning. *Molecular Ecology Resources*, 21, 609–620.

Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. (2012) ABGD, automatic barcode gap discovery for primary species delimitation. *Molecular Ecology*, 21, 1864–1877.

Ranasinghe, S., Eberle, J., Bohacz, C., Benjamin, S. & Ahrens, D. (2020) New species of Sericini from Sri Lanka (Coleopera, Scarabaeidae). *European Journal of Taxonomy*, 621, 1–20.

Ranasinghe, S., Eberle, J., Athukorala, N., Benjamin, S. & Ahrens, D. (2022a) New species of Sericini from Sri Lanka (Coleopera, Scarabaeidae) II. *European Journal of Taxonomy*, 821, 57–101.

Ranasinghe, U.G.S.L., Eberle, J., Thormann, J., Bohacz, C., Benjamin, S. & Ahrens, D. (2022b) Multiple species delimitation approaches with COI barcodes poorly fit each other and with morphospecies – an integrative taxonomy case of Sri Lankan Sericini chafers (coleoptera: Scarabaeidae). *Ecology and Evolution*, 12, e8942.

Ratnasingham, S. & Hebert, P.D.N. (2013) A DNA-based registry for all animal species: the barcode index number (BIN) system. *PLoS One*, 8, e66213.

Reisch, C. & Schmid, C. (2018) Species and genetic diversity are not congruent in fragmented dry grasslands. *Ecology and Evolution*, 12, 664–671.

Rodríguez, A., Börner, M., Pabijan, M., Gehara, M., Haddad, C.F.B. & Vences, M. (2015) Genetic divergence in tropical anurans: deeper phylogeographic structure in forest specialists and in topographically complex regions. *Evolutionary Ecology*, 29, 765–785.

Rossini, B.C., Oliveira, C.A.M., de Melo, F.A.G., Bertaco, V.D., de Astarloa, J.M.D., Rosso, J.J. et al. (2016) Highlighting *Astyanax* species diversity through DNA barcoding. *PLoS One*, 11, e0167203.

Salinas-Ivanenko, S. & Múrria, C. (2021) Macroecological trend of increasing values of intraspecific genetic diversity and population structure from temperate to tropical streams. *Global Ecology and Biogeography*, 30, 1685–1697.

Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E. & Crozier, R.H. (2010) Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual Review of Entomology*, 55, 421–438.

Schnell, I.B., Thomsen, P.F., Wilkinson, N., Rasmussen, M., Jensen, L.R.D., Willerslev, E. et al. (2012) Screening mammal biodiversity using DNA from leeches. *Current Biology*, 22(8), R262–R263.

Shum, P. & Palumbi, S.R. (2021) Testing small-scale ecological gradients and intraspecific differentiation for hundreds of kelp forest species using haplotypes from metabarcoding. *Molecular Ecology*, 30, 3355–3373.

Sigsgaard, E.E., Jensen, M.R., Winkelmann, I.E., Møller, P.R., Hansen, M.M. & Thomsen, P.F. (2020) Population-level inferences from environmental DNA— current status and future perspectives. *Evolutionary Applications*, 13, 245–262.

Smith, M.A., Fisher, B.L. & Hebert, P.D.N. (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360, 1825–1834.

Steinke, D., de Waard, S.L., Sones, J.E., Ivanova, N.V., Prosser, S.W.J., Perez, K. et al. (2022) Message in a bottle - metabarcoding enables biodiversity comparisons across ecoregions. *GigaScience*, 11, giac040.

Sun, S. et al. (2016) DNA barcoding reveal patterns of species diversity among northwestern Pacific molluscs. *Scientific Reports*, 6, 33367.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 8, 2045–2050.

Templeton, A.R. (2001) Using phylogeographic analyses of gene trees to test species status and processes. *Molecular Ecology*, 10, 779–791.

Templeton, A.R., Crandall, K.A. & Sing, C.F. (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, 132(2), 619–633.

Thompson, J.N. (1998) Rapid evolution as an ecological process. *Trends Ecology and Evolution*, 13(8), 329–332.

Thormann, B., Ahrens, D., Armijos, D.M., Peters, M.K., Wagner, T. & Wägele, J.W. (2016) Exploring the leaf beetle fauna (Coleoptera: Chrysomelidae) of an Ecuadorian mountain forest using DNA barcoding. *PlosOne*, 11(2), e0148268.

Uscanga, A., López, H., Piñero, D., Emerson, B.C. & Mastretta-Yanes, A. (2021) Evaluating species origins within tropical sky-islands arthropod communities. *Journal of Biogeography*, 48, 2199–2210.

van Jaarsveld, A.S., Freitag, S., Chown, S.L., Muller, C., Koch, S., Hull, H. et al. (1998) Biodiversity assessment and conservation strategies. *Science*, 279, 2106–2108.

Van Klink, R., August, T., Bas, Y., Bodesheim, P., Bonn, A., Fossøy, F. et al. (2022) Emerging technologies revolutionise insect ecology and monitoring. *Trends in Ecology and Evolution*, 37(10), 872–885.

Vellend, M. & Geber, M.A. (2005) Connections between species diversity and genetic diversity: species diversity and genetic diversity. *Ecology Letters*, 8(7), 767–781.

Voris, H.K. (2000) Maps of Pleistocene Sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography*, 27, 1153–1167.

Wiemers, M. & Fiedler, K. (2007) Does the DNA barcoding gap exist? – a case study in blue butterflies (lepidoptera: Lycaenidae). *Frontiers in Zoology*, 4, 8.

Will, K.W., Mishler, B.D. & Wheeler, Q.D. (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, 54(5), 844–851.

Yang, J., Zhang, X., Jin, X., Seymour, M., Richter, C., Logares, R. et al. (2021) Recent advances in environmental DNA-based biodiversity assessment and conservation. *Diversity and Distribution*, 27, 1876–1879.

Yeates, D.K., Seago, A., Nelson, A., Cameron, S.L., Joseph, L. & Trueman, J.W.H. (2011) Integrative taxonomy, or iterative taxonomy? *Systematic Entomology*, 36, 209–217.

Yu, D.W., Ji, Y., Emerson, B.C., Wang, X., Ye, C., Yang, C. et al. (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, 3, 613–623.

Zhang, J., Kapli, P., Pavlidis, P. & Stamatakis, A. (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, 29, 2869–2876.

Zhou, Z., Guo, H., Han, L., Chai, J., Che, X. & Shi, F. (2019) Singleton molecular species delimitation based on COI-5P barcode sequences revealed high cryptic/undescribed diversity for Chinese katydids (orthoptera: Tettigoniidae). *BMC Ecology and Evolution*, 19, 79.

Zinger, L., Bonin, A., Alsos, I.G., Bálint, M., Bik, H., Boyer, F. et al. (2019) DNA metabarcoding—need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology*, 28, 1857–1862.

SUPPORTING INFORMATION

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

Figure S1. NMDS analysis (non-metric multidimensional scaling: Jaccard index) for morphospecies, haplotypes and MOTUs that resulted from delimitation of the total assemblage and from cumulative subclade analyses (indicated by asterisk) for forest types, elevation zones and sampling localities.

Figure S2. Number of putative species entities reported in forest types, elevation zones and sampling localities as morphospecies, haplotypes, PTP-clusters, TCS networks, distance clusters for subclades.

Figure S3. Clustering analysis (Jaccard index) based on presence/absence for forest types, elevation zones and sampling localities for morphospecies, haplotypes and MOTUs for four subclade analysis. LW: wet lowland; LD: dry lowland; SM: sub-montane; MO: montane. EZ1: 0–500 m, EZ2: 501–1000 m, EZ3: 1001–1500 m, EZ4: 1501–2000 m, EZ5; 2001–2500 m. L1–L15 sampling localities.

Figure S4. NMDS (non-metric multidimensional scaling: Jaccard index) based on presence/absence of species. Analyses were done for forest types, elevation zones and sampling localities and for morphospecies, haplotypes and MOTUs and four subclade analysis. LW: wet lowland; LD: dry lowland; SM: sub-montane; MO: montane. EZ1: 0–500 m, EZ2: 501–1000 m, EZ3: 1001–1500 m, EZ4: 1501–2000 m, EZ5; 2001–2500 m. L1–L15 sampling localities.

Figure S5. Frequency of intra-and interspecific distances of the phytophagous scarab chafer data from Sri Lanka

Table S1. Sample details: voucher number, species identification, locality id (Sri Lanka), barcode index number (BIN) assignments and GenBank accession numbers.

Table S2. Number of MOTUs, number of matches between MOTUs and morphospecies (in parenthesis) and match ratios (Ahrens et al., 2016) of DNA-based species delimitation methods. Numbers are given for the total dataset, individual clades and cumulative subclades. Match ratio $= 2 \times$ Nmatch / (Nmol + Nmorph).

Table S3. Species similarity among different forest types and elevation zones among morphospecies, haplotypes and MOTUs.

LW: wet lowland; LD: dry lowland; SM: sub-montane; MO: montane. EZ1: 0–500 m, EZ2: 501–1000 m, EZ3: 1001–1500 m, EZ4: 1501–2000 m, EZ5: 2001–2500 m.

How to cite this article: Ranasinghe, U.G.S.L., Thormann, J., Benjamin, S.P., Bezděk, A., Eberle, J. & Ahrens, D. (2023) Contrasting results of multiple species delimitation approaches cause uncertainty in synecological studies: A case study on Sri Lankan chafers. *Insect Conservation and Diversity*, 1–16. Available from: https://doi.org/10.1111/icad.12684