Received: 22 June 2022 | Accepted: 10 October 2022

DOI: 10.1111/2041-210X.14041

Methods in Ecology and Evolution

RESEARCH ARTICLE

Standardized nuclear markers improve and homogenize species delimitation in Metazoa

Lars Dietz[1] | Jonas Eberle[1,2] | Christoph Mayer[1] | Sandra Kukowka[1] | Claudia Bohacz[1] | Hannes Baur[3] | Marianne Espeland[1] | Bernhard A. Huber[1] | Carl Hutter[4] | Ximo Mengual[1] | Ralph S. Peters[1] | Miguel Vences[5] | Thomas Wesener[1] | Keith Willmott[6] | Bernhard Misof[1,7] | Oliver Niehuis[8] | Dirk Ahrens[1]

[1]Museum A. Koenig, Leibniz Institute for the Analysis of Biodiversity Change (LIB), Bonn, Germany; [2]Paris-Lodron-University, Salzburg, Austria; [3]Naturhistorisches Museum Bern/Institute of Ecology and Evolution, University of Bern, Bern, Switzerland; [4]Museum of Natural Sciences and Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana, USA; [5]Technische Universität Braunschweig, Braunschweig, Germany; [6]Florida Museum of Natural History, University of Florida, Gainesville, Florida, USA; [7]Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany and [8]Abt. Evolutionsbiologie und Ökologie, Institut für Biologie I, Albert-Ludwigs-Universität Freiburg, Freiburg im Breisgau, Germany

Correspondence
Dirk Ahrens
Email: ahrens.dirk_col@gmx.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: AH 175/3-1, AH 175/6-1, AH 175/6-2, MA 3684/5-2, MI 649/18-1, NI 1387/6-1 and NI 1387/7-1; National Science Foundation grant NSF, Grant/Award Number: DEB-1256742

Handling Editor: Douglas Yu

Abstract

1. Species are the fundamental units of life and evolution. Their recognition is essential for science and society. Molecular methods have been increasingly used for the identification of animal species, despite several challenges.
2. Here, we explore with genomic data from nine animal lineages a set of nuclear markers, namely metazoan-level universal single-copy orthologs (metazoan USCOs), for their use in species delimitation. Our data sets include arthropods and vertebrates. We use various data assembly strategies and use coalescent-based species inference as well as population admixture analyses and phenetic methods.
3. We demonstrate that metazoan USCOs distinguish well closely related morphospecies and consistently outperform classical mitochondrial DNA barcoding in discriminating closely related species in different animal taxa, as judged by comparison with morphospecies delimitations. USCOs overcome the general shortcomings of mitochondrial DNA barcodes, and due to standardization across Metazoa, also those of other approaches. They accurately assign samples not only to lower but also to higher taxonomic levels.
4. Metazoan USCOs provide a powerful and unifying framework for DNA-based species delimitation and taxonomy in animals and their employment could result in a more efficient use of research data and resources.

KEYWORDS
animals, barcoding, DNA taxonomy, metazoan USCOs, species delimitation

1 | INTRODUCTION

Defining and classifying groups of organisms faces new opportunities in terms of reproducibility, automation, and robustness due to major innovations in morphological and genomic analytical methods as well as continuously increasing computational power (Gaston & O'Neill, 2004; Lemmon & Lemmon, 2013; MacLeod et al., 2010; Mayer et al., 2016; Rannala & Yang, 2008). Rapid and correct identification of species is paramount since they are the fundamental entities of biodiversity and evolution. DNA-based approaches revolutionized the possibilities in biology (Blaxter et al., 2022). In particular, they are able to resolve questions that were formerly intractable or unfeasible through morphology-based approaches (Godfray, 2007; Hebert et al., 2003). During the past 20 years, DNA barcoding has increased the quality and reproducibility of species delimitation and identification and enabled a rapid assessment and monitoring of biodiversity (Taberlet et al., 2012; Yu et al., 2012). Its hallmark is the capability to standardize and automate species recognition by using a specific single and easily amplified gene fragment. In animals, the most widely used DNA barcoding marker has been the mitochondrial protein-coding gene cytochrome oxidase subunit 1 (*COI*) (Hebert et al., 2003). Beyond that, DNA barcoding paved the way for direct inference of species boundaries from unknown samples (Pons et al., 2006). However, species delimitation and identification based on information from a single mitochondrial gene are prone to errors due to extrachromosomal inheritance, incomplete lineage sorting, sex-biased dispersal, asymmetrical introgression, or *Wolbachia*-mediated genetic sweeps (Ballard & Whitlock, 2004; Funk & Omland, 2003). As a consequence, results from delimiting species by means of barcoding are not always congruent with those obtained from analysing morphology or other data. Integrative taxonomic approaches have therefore been proposed to overcome these problems by complementing barcode-based species hypotheses with additional evidence (Carstens et al., 2013; Padial et al., 2010; Schlick-Steiner et al., 2010; Yeates et al., 2011).

Recent species delimitation approaches have considerably improved in accuracy by taking advantage of the phylogenetic information contained in multiple nuclear markers (Dowton et al., 2014; Knowles & Carstens, 2007; Prebus, 2021). Simulations have demonstrated increasing accuracy and robustness with the use of more genes (Yang & Rannala, 2010, 2014). Previously, investigating a small number of loci has been a compromise between costs and the accuracy of the inferred results (Eberle et al., 2020; Edwards & Knowles, 2014). However, progress in sequencing technologies promises a continuous decrease of DNA sequencing costs. In consequence, whole genome and transcriptome sequencing (e.g. Harvey et al., 2016; Misof et al., 2014; Niehuis et al., 2012) or DNA target enrichment approaches would enable sequencing thousands of single copy target loci from an organism's genome, even with degraded DNA, to resolve questions that cannot be answered with only a limited number of loci (Gnirke et al., 2009, Faircloth et al., 2012; Lemmon et al., 2012; Lemmon & Lemmon, 2013; Hancock-Hanser et al., 2013; Mayer et al., 2016).

Besides *COI*, different markers have been used for metazoan DNA taxonomy: nuclear ribosomal RNA genes (rDNA, Chen et al., 2017; Krehenwinkel et al., 2019; Lebonah et al., 2014), restriction site associated DNA sequences (RADseq; Baird et al., 2008; Herrera & Shank, 2016; Pante et al., 2015), and ultra-conserved elements (UCE; Bejerano et al., 2004; Faircloth et al., 2012), the latter typically including more variable flanking regions (Gueuning et al., 2020; Ješovnik et al., 2017; Prebus, 2021; Zarza et al., 2018). However, they can hardly be applied universally across animals, either because of insufficient infraspecific variation or a lack of homologous loci between distantly related taxa (Eberle et al., 2020; Pierce, 2019). Therefore, metazoan-level universal single-copy orthologs (metazoan USCOs) have been proposed as a core set of nuclear-encoded protein-coding genes for species delimitation in Metazoa (Eberle et al., 2020). USCOs are under strong selection for occurring only in single copy within a genome (Feron & Waterhouse, 2022; Waterhouse et al., 2011, 2013). In Metazoa, 978 USCOs were recognized based on a representative selection of 65 high-quality genomes (Simão et al., 2015). A prerequisite for a gene to be a USCO is that it is present as single copy in at least 90% of these genomes (Simão et al., 2015).

USCOs have primarily been utilized for assessing the completeness and quality of sequenced genomes and transcriptomes (Simão et al., 2015). However, they have also been useful for phylogenetic studies (Fernández et al., 2018; Suvorov et al., 2021; Waterhouse et al., 2018; Zhang et al., 2019) and as markers to establish a universally applicable genomic species identification and delimitation procedure and to allow inferring the wider systematic placement (e.g. class, order, or family) of an unknown sample (Eberle et al., 2020). Since the number of single-copy genes increases with increasing relatedness of the species under consideration, sets of USCOs are larger for lower systematic levels (Eberle et al., 2020).

Here we demonstrate the suitability of metazoan USCOs for species identification within Metazoa using empirical data and show the impact of sequence assembly protocols on data yield. In nine study cases, we sequenced metazoan USCOs of different genera from various arthropod and vertebrate groups that include morphologically well-recognizable and closely related species which cannot always be distinguished by *COI* barcodes.

2 | MATERIALS AND METHODS

2.1 | Samples and case studies

We mined metazoan USCOs of selected species of seven genera representing six major groups of Arthropoda (Araneae: *Stygopholcus*; Coleoptera: *Pleophylla*; Diptera: *Sphaerophoria*; Hymenoptera: *Chrysis* and *Pteromalus*; Lepidoptera: *Taygetis*; Myriapoda: *Lithobius*) and of two genera of Amphibia (Anura: *Discoglossus*, *Rana*; Table S1). Specimens were chosen to include (i) well-studied examples of closely related species being genetically divergent but including also conspecifics, to provide controls for the accuracy of species

delimitation; and (ii) particularly challenging cases with (mitochondrial) introgression or morphologically cryptic lineages. Whenever possible, at least four or five samples were used per species, representing different populations within the species' range.

2.2 | Data generation and assembly

USCO data were produced using target DNA enrichment (Gnirke et al., 2009; Mayer et al., 2016). Bait design based on individual exon alignments and lab procedures are extensively described in the Supporting Information (SI). The major objective was maximizing the number of USCOs found per studied taxon, but also inferring the robustness of species delimitation given different properties of the assembly methods. We considered seven different data assembly approaches (Figure 1), which may impact data yield and species delimitation of closely related taxa. Assembly techniques (A1–A7) used partly new and partly published pipelines and orthology verification methods (for extensive details, see SI). Of these, A1 and A2 generated diploid consensus sequences based on direct mapping to a reference, that is, heterozygous sites were represented by ambiguity codes. The other approaches generated (pseudo)haploid contigs, where each site was represented by the most commonly found nucleotide. Approaches A4, A6, and A7 did not distinguish between coding and non-coding sequences, while the other approaches automatically exclude non-coding data.

In addition, we generated reduced datasets for all approaches in which alignment positions having a gap in at least one individual were removed. Furthermore, for A1 and A2, we generated datasets with all positions removed which either had a gap or ambiguity (i.e. heterozygous position) in at least one individual.

2.3 | Phylogenetic analyses

For all assemblies (A1–A7), phylogenetic analyses were performed in two ways: (1) We conducted multi-species coalescent analyses with *ASTRAL III* v. 5.6.1 (Zhang et al., 2018) based on individual gene trees that were produced with maximum-likelihood analyses on separate nucleotide alignments for each locus using *IQ-TREE* v. 1.6.3 (Nguyen et al., 2015). (2) Alignments of all recovered genes (Table S3) were concatenated into a single dataset for each study case (and assembly), for which tree searches were then conducted with *IQ-TREE*. In both approaches, tree searches were performed for full and reduced USCO datasets (for details, see SI).

Finally, we combined the results of A3 for all study cases into a single super-alignment, including the reference sequences that were used for bait design (Table S4). Phylogenetic analysis of the

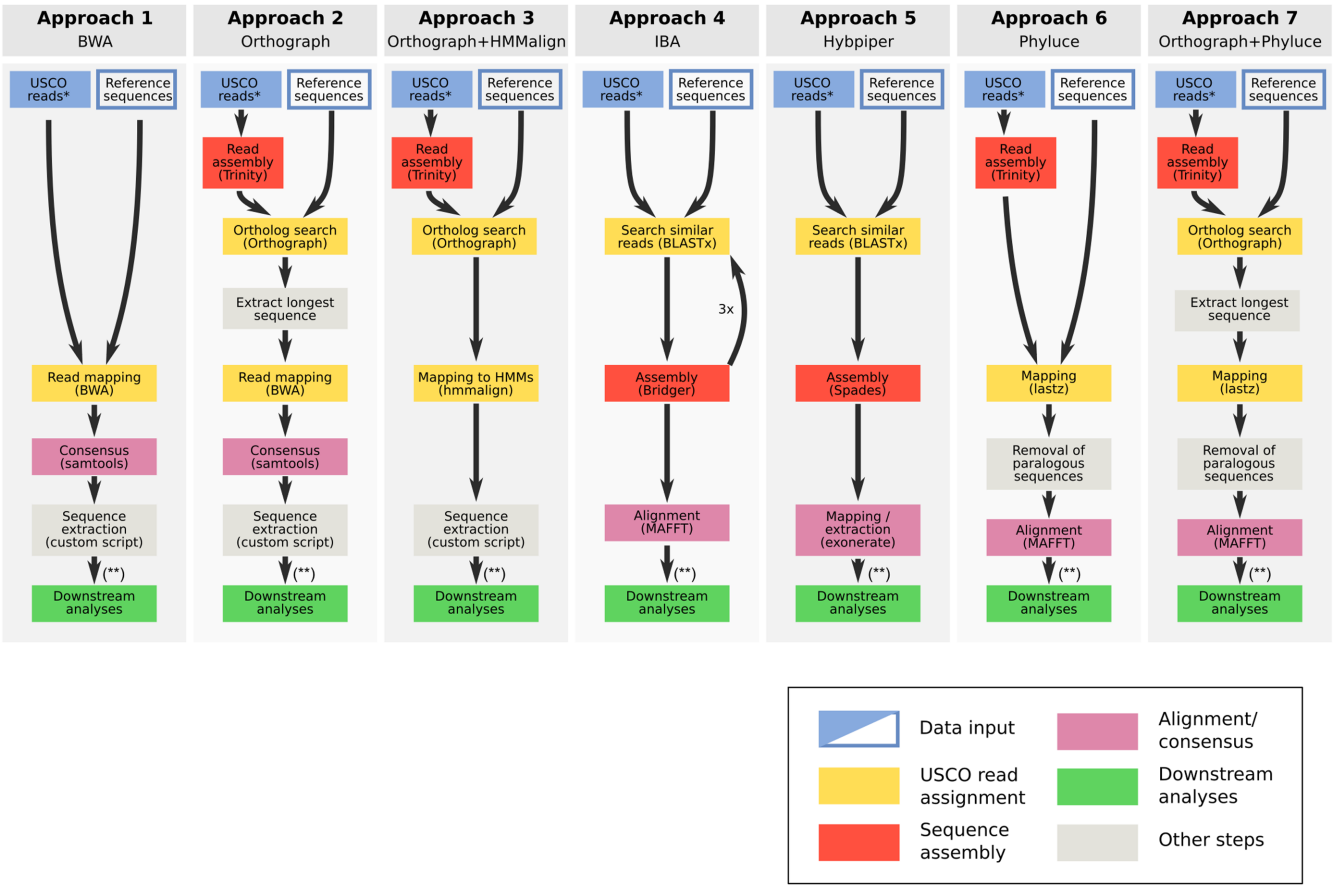

FIGURE 1 Data assembly workflows used for the seven different approaches (A1–A7) (for details, see Supporting Infomartion).

non-partitioned concatenated alignment was performed for nucleotides and amino acid sequences.

2.4 | Analysis of sequence variation and species delimitation

For analyses of sequence variation on resulting USCOs, single nucleotide polymorphisms (SNPs) were extracted from A2 datasets using *SNP sites* (Page et al., 2016), excluding low-quality bases. With these SNPs, we performed phenetic clustering analyses with *STRUCTURE* v. 2.3.4 (Pritchard et al., 2000) using different numbers of assumed populations as well as non-metric multidimensional scaling (NMDS) analyses with PAST v. 4.03 (Hammer et al., 2001; for details, see Supporting Information).

Under the general lineage concept (De Queiroz, 1998), some important criteria for species entities (monophyly, absence of genetic intermediates, and diagnosability) result from evolutionary patterns that become evident from a range of species delimitation approaches. Metazoan USCOs of species that represent independently evolving meta-populations (De Queiroz, 2007; Freudenstein et al., 2017) should fit a species tree with gene tree distributions described by the multi-species coalescent model (Rannala & Yang, 2003). Consequently, we applied various implementations of the multi-species coalescent model to each study case to delimit species using parametric (Boukaert et al., 2014; Jones, 2017; Yang & Rannala, 2010, 2014) and nonparametric methods (Fujisawa et al., 2016).

We analysed full and reduced metazoan USCOs with the program *BPP* v. 4.1.4, testing the validity of species-level groupings (Flouri et al., 2018; analysis details: see Supporting Information). Optimal priors for the analysis, that is, real population size (theta) and divergence time (tau), were unknown for the study cases. Since a determination of priors from the data itself prior to the analysis would make species inference circular, as priors would be inferred for a priori assumed species-level groupings, we used nine predefined prior combinations of theta and tau (Figure 5; Figures S9–S14). For each prior combination, the final criterion for defining a species split was the median posterior probability, computed from five independent runs of each analysis. Species splits were accepted as valid if the posterior probability was higher than 0.9. At least in the median of the five independent analyses, the ESS always exceeded 200, that is, the runs reached stationarity. Consistency across runs was considered as an indicator of MCMC convergence (Flouri et al., 2018).

As nonparametric method, we used the trinomial distribution of triplets model (*tr2*; Fujisawa et al., 2016) to infer species boundaries based on topological variations in gene trees. Analyses were run on full and reduced datasets from all seven assembly approaches.

Finally, tree inferences and results of species delimitation with USCOs were compared with those from *COI* sequence data of the same specimens. We applied various de novo species delimitation approaches to the sequences generated with Sanger sequencing or with DNA target enrichment and Illumina sequencing (for details, see Supporting Information).

3 | RESULTS

3.1 | Data recovery success

We obtained nucleotide sequences by DNA target enrichment and Illumina sequencing comprising an average of 2,291,310 reads (SD 1,858,696; Table S2), which assembled into up to 950 metazoan USCO loci with more than 0.5 million base pairs per specimen (Table S3). The success of data recovery varied considerably between assembly approaches (A1–A7; Figures 1, Figure S2) and study cases (Figure 2, Figure S1; Tables S3 and S5). Best performing assembly approaches, in terms of number of recovered loci and base pairs, yielded at least 700 metazoan USCOs and more than 200,000 base pairs per case study (see Supporting Information). Here, most metazoan USCOs were recovered in majority of the study cases and frequently in all or almost all individuals, with considerable overlap of alignment regions (Figure 2a–c). Sometimes metazoan USCOs were recovered in only a few (1–3) individuals, in some assembly approaches more frequently than in others (Figure S1); these were excluded to avoid an excess of missing data (for details, see Supporting Information).

3.2 | Phylogenetic analyses

Species inference under the general lineage concept of species (De Queiroz, 1998, 2007; Yang & Rannala, 2010) with metazoan USCOs relies on accurate phylogenetic hypotheses. In the phylogenies inferred using coalescent-based species tree approaches and maximum likelihood analyses of concatenated data, 98% of the morphospecies in the study cases resulted as being monophyletic (Figure 3; Figures S2–S7). All trees had robust and well-resolved interspecific topologies that widely agreed among tree reconstruction methods and assembly approaches (A1–A7). Infraspecific relationships sometimes varied among different assemblies. Morphospecies were almost always (except in two cases) recovered as monophyletic in USCO-based trees but sometimes not in our *COI* benchmarking analyses (13 non-monophyletic cases, see below) (Figure 3, Figure S5; Table 1). Just one case showed a disagreement between metazoan USCO tree topology and morphology-based taxonomy: the beetle *Pleophylla fasciatipennis* was split into two separate clades that are not sister to each other, possibly reflecting the presence of true cryptic species. Thorough *IQ-TREE* analyses (parameter: -m MFP+MERGE, repeated 50 times) on concatenated and partitioned data resulted in only slight changes of topology for poorly resolved infraspecific nodes compared with the 'explorative' runs, showing

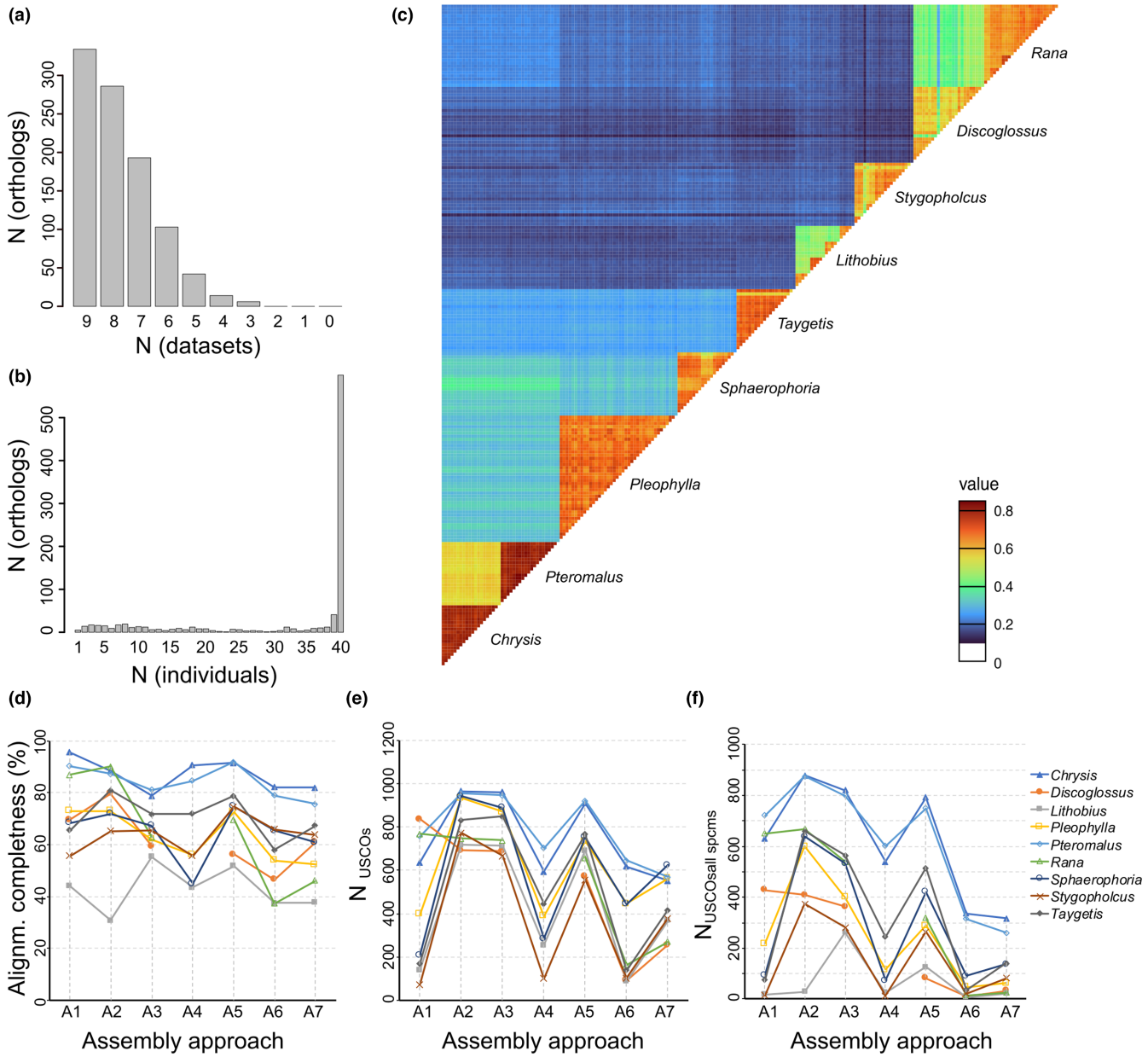

FIGURE 2 Evidence for the universal applicability of metazoan USCOs as taxonomic markers. (a) Number of orthologs shared exclusively by N datasets among the nine study cases in assembly approach A2. (b) Distribution of orthologs over the number of specimens within one exemplary case study (*Pleophylla*) in assembly approach A2 (see Figure S1 for a complete overview of all taxa and assemblies). (c) Proportion of pairwise sequence overlaps in the concatenated alignment (A3). (d) Completeness of concatenated alignments (percentage of non-missing data), (e) number of metazoan USCOs and (f) number of metazoan USCOs present in all specimens of each case study in each assembly approach.

that the tree topologies did not depend greatly on model parameters neither at infraspecific branches nor at species-level branches.

A concatenated super-alignment of all target groups combined with the reference taxa (Table S4) with 260,233 amino acids or 780,699 nucleotides in 978 metazoan USCOs produced a meaningful tree (Figure 3a; Figure S8): all genera and higher-level systematic groups (e.g. orders), as well as most of the morphospecies, were monophyletic. Internal relationships were well resolved and recovered topologies reflected widely accepted phylogenetic relationships among the major lineages.

3.3 | Analysis of sequence variation and species delimitation

Single-nucleotide polymorphisms (SNPs) extracted from the metazoan USCOs obtained with the best-performing approach A2 (see Supporting Information) varied considerably in number among the study cases due to different numbers of recovered USCOs, ranging from 1617 (*Stygopholcus*) to 17,364 SNPs (*Pleophylla*) (Table S6). NMDS on SNPs (Figure 4) showed nearly all morphospecies as discrete clusters. In the diphyletic *Pleophylla fasciatipennis*, the two

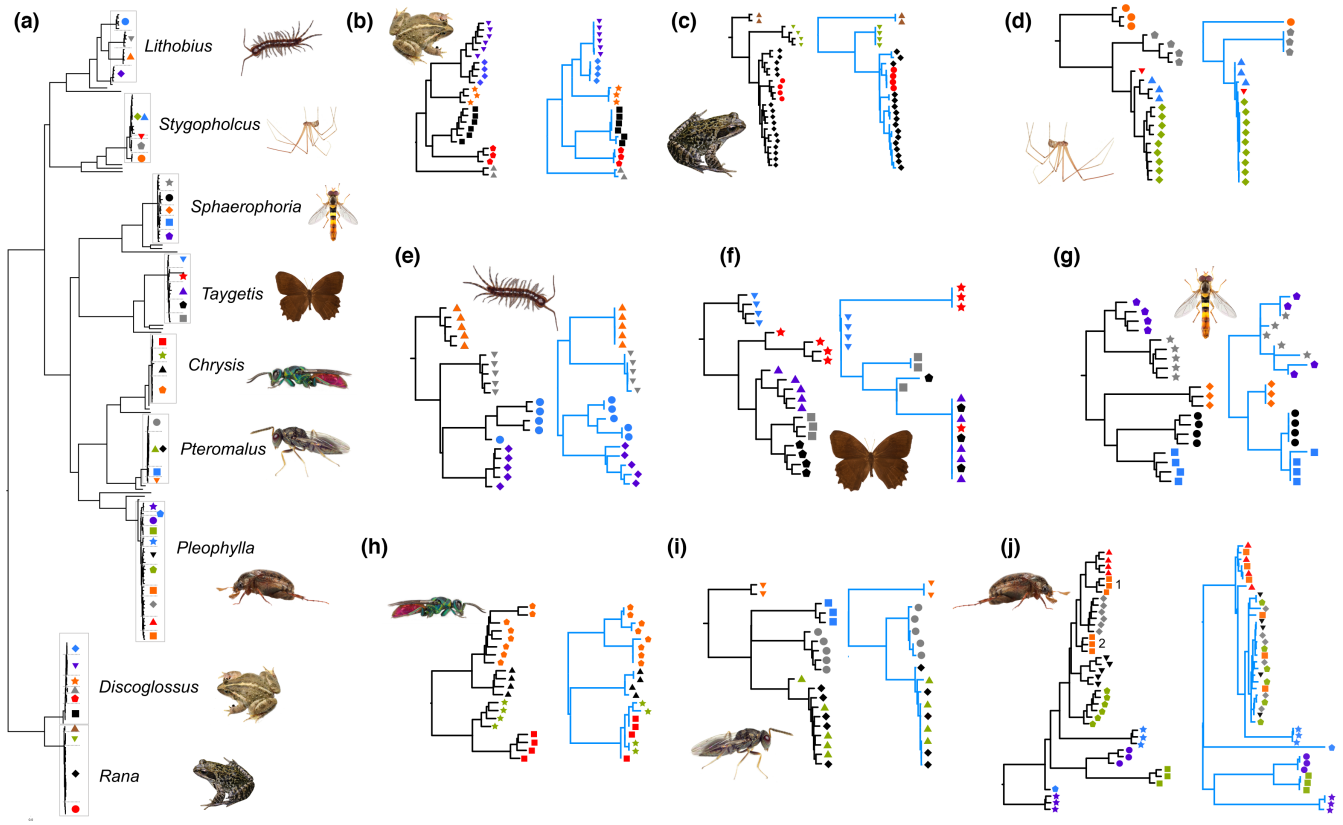

FIGURE 3 Phylogenetic resolution of metazoan USCOs. (a) Tree computed with maximum likelihood from concatenated metazoan USCOs for all study cases (A3; all nucleotides); branches of reference taxa (Supporting Information) are unlabeled. (b–j) metazoan USCO trees (black) obtained from ASTRAL analysis of metazoan USCOs (A2) of individual study cases compared with the *COI* benchmarking tree (blue). Study cases: *Discoglossus* (b), *Rana* (c), *Stygopholcus* (d), *Lithobius* (e), *Taygetis* (f), *Sphaerophoria* (g), *Chrysis* (h), *Pteromalus* (i) and *Pleophylla* (j). Morphospecies indicated by coloured symbols. *Pt. brachygaster* sp2 (blue square) did not yield sequences for *COI*.

TABLE 1 Number of morphospecies entities inferred as monophyletic from USCO and COI datasets, in comparison with the total number of morphospecies involved (based on A2, see Figure 5)

Taxon case	N_{morph}	USCOs	*COI*
Chrysis	4	4	2
Discoglossus	6	6	6
Lithobius	4	4	4
Pleophylla	10	9	5
Pteromalus	5	3	2[a]
Rana	4	3	3
Sphaerophoria	5	5	3
Stygopholcus	5	5	4
Taygetis	5	5	1

[a]Samples of one morphospecies could not be amplified successfully (*P. brachygaster* sp2).

distinct clades were also recovered as separate clusters. However, the visibility of this outcome in the plot is dependent on scaling since the genetic divergence between the included species sometimes differed strikingly within a dataset (Figure 4i); species sometimes formed distinct clusters although these were concealed in the plots due to dense packing. The separation between closely related species became clearer after removing more distantly related taxa from NMDS analysis. In one case, individuals of a clearly monophyletic morphospecies (*Lithobius crassipes*) did not form a distinct cluster due to low data recovery in the group and strong infraspecific divergence (Figure 4d).

Population admixture analyses with *STRUCTURE* (Pritchard et al., 2000) confirmed the monophyletic morphospecies in most cases, and further sub-splitting was not evident (Figure 4). Considering only results with the highest likelihood score, individuals were always assigned to the clusters corresponding to morphospecies with at least 90% probability. In all study cases, increasing K beyond the number of included morphospecies did not result in further species splits. However, the analysis had to be repeated multiple times for each K as in some runs the MCMC got trapped in a local maximum within which some morphospecies remained indistinguishable.

Admixture between species was generally only found if the species were very closely related to each other. In the case of *Pleophylla fasciatipennis*, where the morphospecies was not recovered as monophyletic, no admixture between the two different clades was detected, suggesting that the morphological similarity between those lineages cannot be explained by hybridization (see Supporting Information).

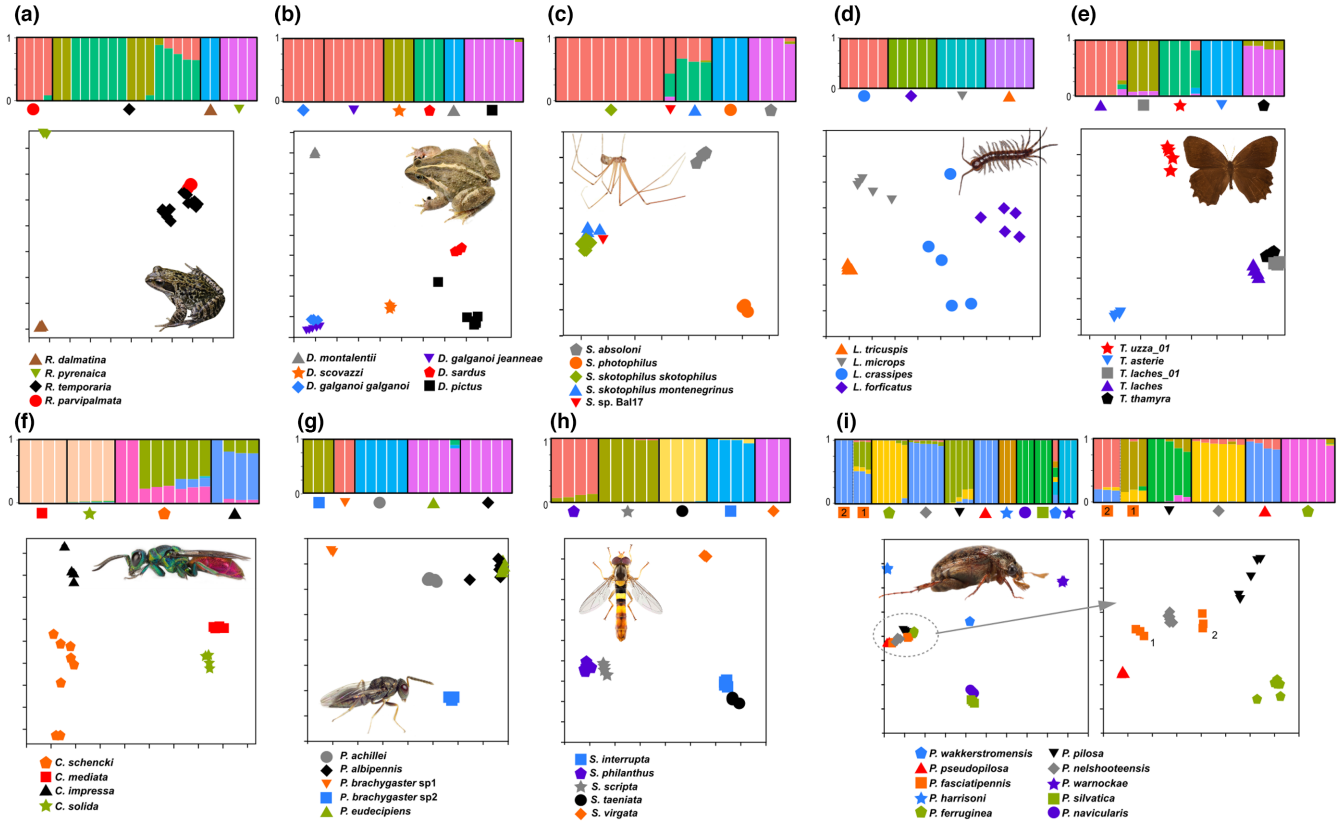

FIGURE 4 Discriminative power of metazoan USCOs. (a–i) probabilities of cluster assignment from *STRUCTURE* analysis (above) and plots from NMDS (below) based on SNPs derived from metazoan USCO data. Study cases: *Rana* (a), *Discoglossus* (b), *Stygopholcus* (c), *Lithobius* (d), *Taygetis* (e), *Chrysis* (f), *Pteromalus* (g), *Sphaerophoria* (h) and *Pleophylla* (i). (i) Shows the results of analyses of all species on the left, and a subset on the right (*P. pilosa* group; encircled in left NMDS plot). Morphospecies indicated by coloured symbols.

The overall outcome of multi-species coalescent analyses with *BPP* (Flouri et al., 2018) and *tr2* (Fujisawa et al., 2016) suggested that many morphospecies should be split into additional entities. Only in one case two morphospecies were lumped (*Pteromalus eudecipiens* and *P. albipennis*), probably due to the taxonomic misinterpretation of one of the morphospecies.

Over-splitting in comparison with morphospecies (and/or geographically separated populations) was sometimes observed with the full USCO data set at all levels of infraspecific nodes (even in syntopic specimens) (Figures S12–S14). However, after excluding sites containing missing or ambiguous (heterozygote) data, over-splitting at the infraspecific level was reduced, although most geographically separated populations within a morphospecies were still proposed as separate species (Figure 5; Figures S9–S11 and S15–S23). All *tr2* analyses similarly resulted in over-splitting of several morphospecies, many of the additional splits were consistent with those of *BPP*, but not all. No consistent tendency towards (more or less) splitting relative to *BPP* was observed (Figures S9–S14).

The genealogical divergence index (gdi) (Jackson et al., 2017) proved to not be a suitable general proxy to evaluate species status and to reject over-splitting in connection with *BPP* analyses using full metazoan USCO data (Figures S12–S14). For many of the splits, the index value fell between the established inter- and infraspecific gdi thresholds ($0.2 < gdi < 0.7$), assuming gradual values considering the degree of divergence of the examined lineages. Many of the lineages split with intermediate gdi values were not only well differentiated morphologically and genetically, but also occurred syntopically (Table S7). While some species, especially those with long branches in the phylogenetic trees, such as *Pleophylla harrisoni* and *Stygopholcus photophilus*, were clearly supported as distinct ($gdi > 0.7$) and in other cases splits within morphospecies would clearly be rejected ($gdi < 0.2$, e.g. within *Pleophylla nelshoogteensis*, *P. pseudopilosa* and *Sphaerophoria scripta*), for most groupings the gdi was between 0.2 and 0.7, indicating an ambiguous species status.

With reduced metazoan USCO data, obtained after excluding missing or ambiguous nucleotides, gdi values at nodes matching morphospecies boundaries were all well above the established inter- and infraspecific gdi threshold (i.e. 0.7; Jackson et al., 2017), with very few exceptions in *Sphaerophoria philanthus* and *Taygetis laches* (Figure 5). Gdi values were yet distinctly above the lower threshold (i.e. 0.2) in the few available infraspecific nodes indicating ambiguity in the use of gdi.

3.4 | COI benchmarking

In only two of the nine study cases (*Discoglossus*, *Lithobius*), the ML tree based on *COI* data recovered all morphospecies as monophyletic

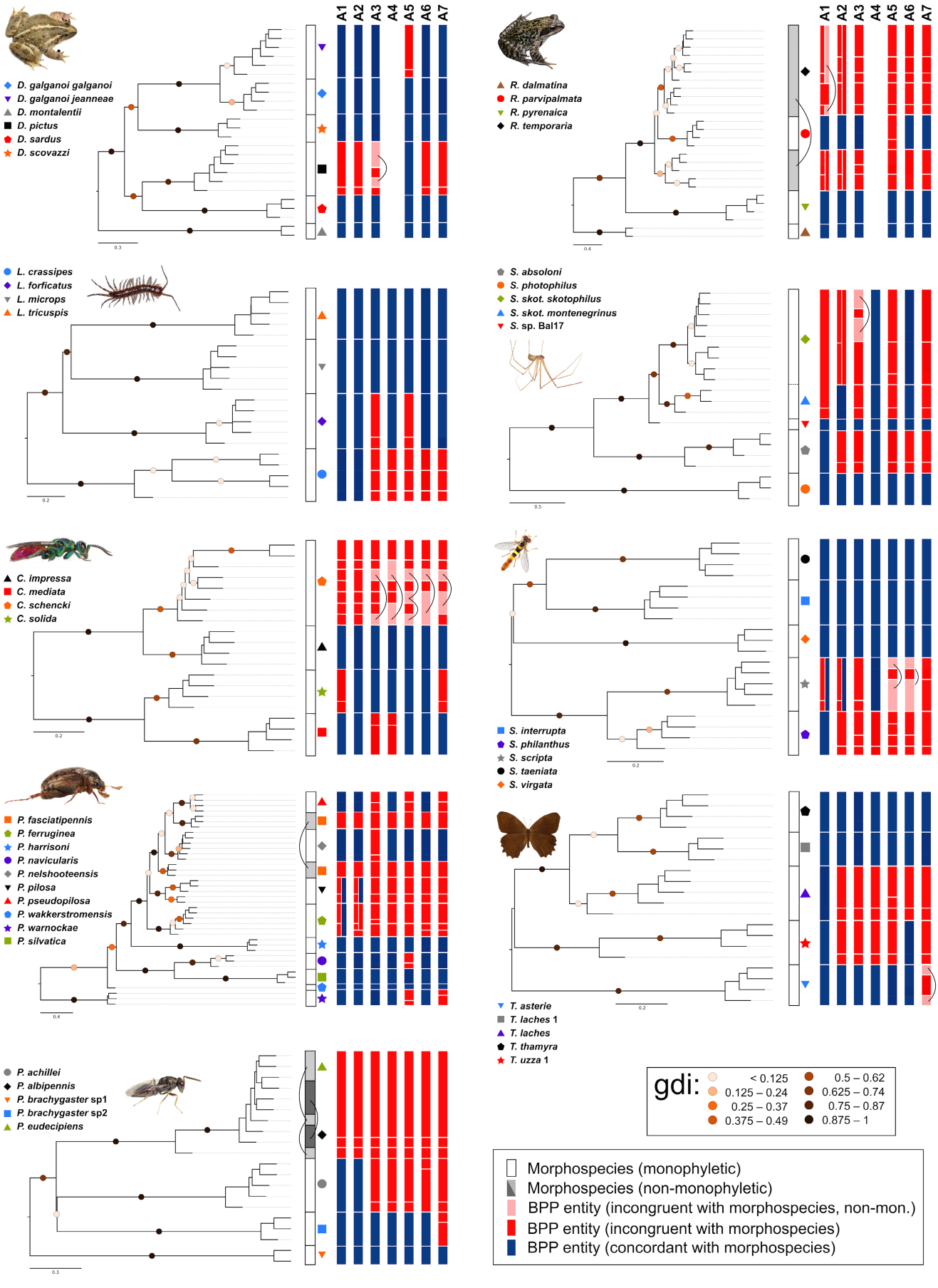

FIGURE 5 Results of *BPP* species delimitation analyses based on reduced metazoan USCO data* showing each case study and assembly approach (A1–A7) mapped onto ASTRAL trees (inferred from the full dataset obtained from A2). Visualized *BPP* entities (coloured rectangles in columns) are based on the median of posterior probabilities of all five replicates and all nine prior combinations. Morphospecies entities are indicated by the rectangle in the first column to the right of each tree. Gdi values are mapped onto branches. Species entities from different assembly approaches may not be monophyletic in this tree because alternative assembly approaches may result in differing guide tree topologies). * Sites with gaps were removed. For the right part of the divided columns of A1 and A2, also sites with ambiguous data were removed.

groups (Figure 3; Figure S30). In one additional case (*Pteromalus*), two potential morphospecies, which were not separated from the metazoan USCO data, also remained indistinguishable. *COI* performed worst in *Pleophylla*, where five morphospecies that are easily distinguished by morphology (Eberle et al., 2016, 2017) could not be resolved as monophyletic groups (Table 1). While three *Taygetis* morphospecies had identical haplotypes, in *Lithobius*, two morphospecies showed very deep coalescence, leading to problems with species delimitation.

Species delimitation analyses with *COI* showed a tendency for over-splitting in most of the study cases (Figure S30). Results of the different species delimitation methods were partly quite divergent within each study case, particularly in *Taygetis* and *Sphaerophoria*, in which for the same *COI* data some methods produced just one MOTU (parsimony network analysis; Templeton et al., 1992), others up to 17 and 16, respectively (*bPTP*; Zhang et al., 2013). The most consistent results were obtained in *Discoglossus*, where results of *ABGD* (Puillandre et al., 2012) had a 100% correspondence between MOTUs and morphospecies.

4 | DISCUSSION

Our results demonstrate that metazoan USCOs are a quite powerful universal marker system for species delimitation using nuclear genomic data. USCOs allowed us to distinguish closely related species even in study cases in which taxa were indistinguishable and phylogenetically unresolved by *COI* sequences alone (Table 1). Besides the striking fit between metazoan USCO phylogenies and morphospecies, our results demonstrate the potential to resolve open taxonomic questions and evaluate detected and potentially unrecognized, truly cryptic species (e.g. *Pleophylla*, *Discoglossus*). Results of phylogenetic analyses including taxa of all study cases indicate that metazoan USCOs also differentiate between higher taxa at the level of orders or classes, being compatible with the current hypotheses of within-arthropod relationships (Misof et al., 2014; Figures 3a, Figure S8).

Consequently, our empirical evidence confirmed (a) sufficient overlap in the recovery of metazoan USCOs between different groups of Metazoa to allow comparison and large-scale analyses of even distantly related species and groups such as arthropod and tetrapod species (for higher level assignment), (b) robust results based on clearly specified wet lab and bioinformatic protocols that recover data from a high proportion of samples, (c) sufficient phylogenetic resolution to separate closely related species, and (d) agreement of resulting groupings with morphospecies (and/or alternative evidence for robust species hypotheses such as independently published hybrid zone analyses).

The phylogenetic trees from metazoan USCO were robust with respect to differences in alignment completeness due to varying data yield: different assembly approaches as well as the use of different tree reconstruction methods (concatenation-based vs. coalescent analyses) resulted in very similar trees (Figures S2–S7). Even in cases with a higher amount of missing data, resulting from a non-optimal data recovery (caused by more distant assembly reference taxa), metazoan USCOs yielded enough information to generate mostly well-resolved (at species level) and reliable phylogenies. Phylogenetic analyses with reduced data (ambiguous and/or gaps omitted) showed very similar tree topologies (Figures S21–S26), except for three assembly approaches in *Chrysis* and one in *Discoglossus* (Figure S26) in which monophyly of one morphospecies failed to be recovered.

The choice of which assembly strategy (Figure 1) might be best suited will depend on the availability of potential reference data for a certain study group (for bait design and data assembly) as well as on the scope of the study to be performed: For larger-scale phylogenetic analyses or the higher-level identification of an unknown sample, approach A3 seemed most suitable, for species-level analyses of a narrow case study A2 is preferable (see SI). The versatility to assemble USCOs from raw reads in various ways is one of the major advantages for their universality to address different systematic levels and reference data situations successfully.

USCOs, as protein-coding genes, can be analysed on both the transcriptional (nucleotide) and translational (amino acid) level and thus allow a more reliable assessment of homology. Since amino acid sequences are typically more conserved than the underlying coding nucleotides, using USCOs permits the inclusion of more diverged orthologous DNA sequences. These qualities strengthen the premise of metazoan USCOs as a universal marker system in DNA taxonomy, even to assign unknown samples (with lacking species-specific reference data) to higher systematic categories, such as genera, tribes, families, and orders, which is typically not reliable with single gene markers.

Over-splitting with multispecies coalescent species delimitation approaches is yet a problem in species delimitation and is also seen in USCO datasets. This phenomenon is well known even if only a few markers are used (Barley et al., 2018; Chambers & Hillis, 2020; Eberle et al., 2019; Fujisawa, 2018; Sukumaran & Knowles, 2017) and is caused by currently used speciation model implementations (Sukumaran et al., 2021). Here, in contrast to

BPP analyses, the removal of gaps led to even more infraspecific splits in *tr2* analyses, probably because highly incomplete gene alignments were removed that caused incongruent gene tree topologies and thus likely exaggerated coalescence. Overcoming the over-splitting is the major task for current and future research in the field but is likely to be resolved in future implementations of these species delimitation methods by incorporating for example an extended speciation process model for species delimitation (Sukumaran et al., 2021). The main difficulty to date, particularly in young species or poorly dispersing species, is to distinguish population structure from speciation (Barley et al., 2018; Chambers & Hillis, 2020; Eberle et al., 2019; Sukumaran et al., 2021; Sukumaran & Knowles, 2017), a problem that seems to be more serious rather than alleviated in datasets with large numbers of loci, because population structure becomes more clearly visible with more data (Domingos et al., 2017; Prebus, 2021). This has been demonstrated with *tr2* (Fujisawa, 2018) and is also expected for *BPP* (Leaché et al., 2019). Our analyses based on the full USCO data confirmed this, as even closely related syntopic specimens were often split into different species. The genealogical divergence of populations/species as expressed by the genealogical divergence index (gdi) (Jackson et al., 2017) also turned out to be of limited use here, as gdi values for many nodes subject to splitting were found to be in the range (0.2–0.7) in which species status is uncertain. Although phenetic clustering algorithms (*STRUCTURE* and NMDS) were less prone to over-splitting, we suggest that this could partly be caused by a lower resolution, as only high-quality SNPs were included in these approaches rather than whole alignments. This should be tested in further studies with extended sampling of specimens, and also with groups of different dispersal capabilities.

So far, the good success of the rather conserved metazoan USCO markers to resolve shallow lineages alleviates the need to exploit introns adjacent to USCOs (theoretically available in the raw data) for species delineation purposes. This is advantageous, because the nucleotide sequences of introns are more difficult to align, especially if the simultaneously examined taxa are phylogenetically relatively distant from each other. Note that, even in cases where reads from different exons of a gene cannot be assembled into a single contig due to intervening introns, approaches such as A2 and A3 still concatenate the resulting exonic sequences into a single coding sequence for each gene based on the reference.

To our knowledge, metazoan USCOs represent the only universal multi-marker system that is applicable to the DNA taxonomy of all metazoans (Eberle et al., 2020). They performed here successfully for both arthropods and vertebrates and, at the same time, are ready to overcome all problems of *COI* barcoding (e.g. introgression of heterospecific DNA and *Wolbachia*-mediated genetic sweeps; Eberle et al., 2020). However, approaches based on a single marker or few genes (e.g. *COI* barcoding) will continue yet to be cheaper, quicker to implement and bioinformatically less demanding in the foreseeable future, particularly when analysing large numbers of individuals. Therefore, we suggest USCOs as the first choice supplementary marker instead of replacing single-gene data entirely. Whenever genomic multi-marker analyses are used and needed to resolve taxonomic problems, we recommend sequencing and studying metazoan USCOs instead of or in addition to markers that are narrowly taxon- or digestion reaction specific. Due to their universality, USCOs are perfectly suitable for the accumulation of databases in which data can be reused and assembled on both higher and lower systematic levels without losing discriminative power. Although there might be currently quantitative restrictions due to higher sequencing costs (ca 80 $ per sample), USCOs could be sequenced for smaller subsets of specimens, particularly when barcoding results contradict other evidence such as morphology.

Finally, we expect that USCOs can be extracted from complete whole-genome data. It remains to be tested how this will perform when the annotation has to be carried out de novo, as annotation errors may lead to incomplete datasets. In our data, highly resolved and robust phylogenetic trees were also obtained for datasets that were less than fully complete.

However, bait design for sequencing USCOs via target enrichment still requires reference genomes or transcriptomes from relatively closely related taxa, which may be a problem for organisms hard to collect. However, we expect that such data will become increasingly available for progressively more animal taxa.

Given the current tendency of over-splitting with the multispecies coalescent delimitation methods, our results also underline the steady need for a critical evaluation, integration and additional refinement of the results of species delimitation (Carstens et al., 2013) with additional evidence and methods (Campillo et al., 2020; Dufresnes et al., 2021; Hausdorf & Hennig, 2020), also when inferring species boundaries with genomic datasets. Limitations of current species delimitation approaches (Sukumaran & Knowles, 2017) and the nature of species and speciation (Ahrens et al., 2016) continue to urge for integrative approaches in species delimitation (Padial et al., 2010) and emphasize the continued need to improve protocols and algorithms and to implement models for species delimitation. The accurate application of these will eventually require exact knowledge of the speciation circumstances (e.g. geography, hybrid zones, host information) and sufficient sampling depth, taxonomically and geographically. In this context, metazoan USCOs are excellently suited to significantly and sustainably complement cross-taxon hypothesis testing and substantially increase the accuracy and comparability of DNA markers.

AUTHOR CONTRIBUTIONS

Dirk Ahrens, Oliver Niehuis, Bernhard Misof, and Christoph Mayer designed the study and acquired funding; Dirk Ahrens, Carl Hutter, Miguel Vences, Oliver Niehuis, Bernhard Misof, Christoph Mayer, and Jonas Eberle conceptualized and supervised data collection and sequencing; Lars Dietz performed USCO data assembly and analyses; Dirk Ahrens, Hannes Baur, Marianne Espeland, Bernhard A. Huber, Carl Hutter, Ximo Mengual, Ralph S. Peters, Thomas Wesener, Keith Willmott and Oliver Niehuis provided and identified specimens; Christoph Mayer, Jonas Eberle, and Claudia

Bohacz assembled and analysed the *COI* data; Lars Dietz, Jonas Eberle, Christoph Mayer, Oliver Niehuis and Dirk Ahrens wrote the original draft of the manuscript; all authors reviewed and edited the manuscript.

ACKNOWLEDGEMENTS

Amphibian samples and data were contributed by Pierre-Andre Crochet, David Donaire, and Pedro Galán. We also thank Tomas Flouri, Michael Hofreiter, Loïs Rancilhac, and Ziheng Yang for fruitful discussions, as well as Karen Meusemann, Alexander Donath and Lars Podsiadlowski for their help to access 1Kite and ZFMK graduate school data. Furthermore, we are grateful to Frank Burbrink, Bryan Carstens, Brant Faircloth, and two anonymous reviewers for helpful comments on the manuscript. The study was funded by the following grants: German Research Foundation grant AH 175/3-1, AH 175/6-1, AH 175/6-2 (D.A.), MA 3684/5-2 (C.M.), MI 649/18-1 (B.M.), NI 1387/6-1, NI 1387/7-1 (O.N.) and National Science Foundation grant NSF, DEB-1256742 (K.W.).

CONFLICT OF INTEREST

The authors declare no competing interest.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.14041.

DATA AVAILABILITY STATEMENT

Specimen data including references to collection permits are given in Table S1. Raw sequence data are deposited in NCBI (Table S1). Additional Supplementary Files (protocols, bait files, code, trees, alignments, etc.) are available on Dryad Digital Repository https://doi.org/10.5061/dryad.hhmgqnkg5; (Dietz et al., 2022).

ORCID

Lars Dietz https://orcid.org/0000-0001-6469-381X
Jonas Eberle https://orcid.org/0000-0003-2519-0640
Christoph Mayer https://orcid.org/0000-0001-5104-6621
Claudia Bohacz https://orcid.org/0000-0003-4889-4341
Marianne Espeland https://orcid.org/0000-0002-6800-4783
Carl Hutter https://orcid.org/0000-0001-6381-6339
Ximo Mengual https://orcid.org/0000-0002-6185-9404
Ralph S. Peters https://orcid.org/0000-0001-7784-9203
Miguel Vences https://orcid.org/0000-0003-0747-0817
Thomas Wesener https://orcid.org/0000-0002-2028-3541
Keith Willmott https://orcid.org/0000-0002-9228-0219
Bernhard Misof https://orcid.org/0000-0003-4175-6798
Oliver Niehuis https://orcid.org/0000-0003-4253-1849
Dirk Ahrens https://orcid.org/0000-0003-3524-7153

REFERENCES

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.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, *3*, e337.

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

Barley, A. J., Brown, J. M., & Thomson, R. C. (2018). Impact of model violations on the inference of species boundaries under the multispecies coalescent. *Systematic Biology*, *67*, 269–284.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., & Haussler, D. (2004). Ultraconserved elements in the human genome. *Science*, *304*, 1321–1325.

Blaxter, M., Archibald, J. M., Childers, A. K., Coddington, J. A., Crandall, K. A., Di Palma, F., Durbin, R., Edwards, S. V., Graves, J. A. M., Hackett, K. J., Hall, N., Jarvis, E. D., Johnson, R. N., Karlsson, E. K., Kress, W. J., Kuraku, S., Lawniczak, M. K. N., Lindblad-Toh, K., Lopez, J. V., … Lewin, H. A. (2022). Why sequence all eukaryotes? *Proceedings of the National Academy of Sciences of the United States of America*, *119*(4), e2115636118. https://doi.org/10.1073/pnas.2115636118

Boukaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*, 1–6.

Campillo, L. C., Barley, A. J., & Thomson, R. C. (2020). Model-based species delimitation: Are coalescent species reproductively isolated? *Systematic Biology*, *69*, 708–721.

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

Chambers, E. A., & Hillis, D. M. (2020). The multispecies coalescent over-splits species in the case of geographically widespread taxa. *Systematic Biology*, *69*, 184–193.

Chen, C.-S., Huang, C.-T., & Hseu, R.-S. (2017). Evidence for two types of nrDNA existing in Chinese medicinal fungus *Ophiocordyceps sinensis*. *AIMS Genetics*, *4*, 192–201.

De Queiroz, K. (1998). The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations. In D. J. Howard & S. H. Berlocher (Eds.), *Endless forms: Species and speciation* (pp. 57–75). Oxford University Press.

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*, 879–886.

Dietz, L., Eberle, J., Mayer, C., Kukowka, S., Bohacz, C., Baur, H., Espeland, M., Huber, B. A., Hutter, C., Mengual, X., Peters, R. S., Vences, M., Wesener, T., Willmott, K., Misof, B., Niehuis, O., & Ahrens, D. (2022). Standardized nuclear markers improve and homogenize species delimitation in Metazoa. *Dryad Digital Repository*. https://doi.org/10.5061/dryad.hhmgqnkg5

Domingos, F. M. C. B., Colli, G. R., Lemmon, A., Lemmond, E. M., & Beheregaray, L. B. (2017). In the shadows: Phylogenomics and coalescent species delimitation unveil cryptic diversity in a Cerrado endemic lizard (Squamata: Tropidurus). *Molecular Phylogenetics and Evolution*, *107*, 455–465.

Dowton, M., Meiklejohn, K., Cameron, S. L., & Wallman, J. (2014). A preliminary framework for DNA barcoding, incorporating the multispecies coalescent. *Systematic Biology*, *63*, 639–644.

Dufresnes, C., Brelsford, A., Jeffries, D. L., Mazepa, G., Suchan, T., Canestrelli, D., Nicieza, A., Fumagalli, L., Dubey, S., Martínez-Solano, I., Litvinchuk, S. N., Vences, M., Perrin, N., & Crochet, P.-A. (2021). Mass of genes rather than master genes underlie the genomic architecture of amphibian speciation. *Proceedings of the National Academy of Sciences of the United States of America*, *118*, e2103963118.

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*, 336–345.

Eberle, J., Beckett, M., Özgül-Siemund, A., Frings, J., Fabrizi, S., & Ahrens, D. (2017). Afromontane forests hide 19 new species of an ancient chafer lineage (Coleoptera: Scarabaeidae: Melolonthinae): *Pleophylla* Erichson, 1847 – Phylogeny and taxonomic revision. *Zoological Journal of the Linnean Society*, *180*, 321–353.

Eberle, J., Fabrizi, S., Bazzato, E., Rossi, M., Columba, M. S., Cillo, D., Uliana, M., Sparacio, I., Sabatinelli, G., Warnock, R. C. M., Carpaneto, G. M., & Ahrens, D. (2019). Sex-biased dispersal obscures species boundaries in integrative species delimitation approaches. *Systematic Biology*, *68*, 441–459.

Eberle, J., Warnock, R. C. M., & Ahrens, D. (2016). Bayesian species delimitation in *Pleophylla* chafers (Coleoptera) – The importance of prior choice and morphology. *BMC Ecology and Evolution*, *16*, 94.

Edwards, D. L., & Knowles, L. L. (2014). Species detection and individual assignment in species delimitation: Can integrative data increase efficacy? *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20132765.

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, *61*, 717–726.

Fernández, R., Kallal, R. J., Dimitrov, D., Ballesteros, J. A., Arnedo, M. A., Giribet, G., & Hormiga, G. (2018). Phylogenomics, diversification dynamics, and comparative transcriptomics across the spider tree of life. *Current Biology*, *28*, 1489–1497.e5.

Feron, R., & Waterhouse, R. M. (2022). Assessing species coverage and assembly quality of rapidly accumulating sequenced genomes. *GigaScience*, *11*, giac006. https://doi.org/10.1093/gigascience/giac006

Flouri, T., Jiao, X., Rannala, B., & Yang, Z. (2018). Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Molecular Biology and Evolution*, *35*, 2585–2593.

Freudenstein, J. V., Broe, M. B., Folk, R., & Sinn, B. T. (2017). Biodiversity and the species concept - lineages are not enough. *Systematic Biology*, *66*, 644–656.

Fujisawa, T. (2018). How tr2-delimitation over-splits structured populations. https://tmfujis.wordpress.com/2018/03/11/how-tr2-delimitation-over-splits-structured-populations/

Fujisawa, T., Aswad, A., & Barraclough, T. G. (2016). A rapid and scalable method for multilocus species delimitation using Bayesian model comparison and rooted triplets. *Systematic Biology*, *65*, 759–771.

Funk, D. J., & Omland, K. E. (2003). Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, *34*, 397–423.

Gaston, K. J., & O'Neill, M. A. (2004). Automated species identification: Why not? *Philosophical Transactions of the Royal Society B*, *359*, 655–667.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S., & Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, *27*, 182–189.

Godfray, H. C. J. (2007). Linnaeus in the information age. *Nature*, *446*, 259–260.

Gueuning, M., Frey, J. E., & Praz, C. (2020). Ultraconserved yet informative for species delimitation: Ultraconserved elements resolve long-standing systematic enigma in Central European bees. *Molecular Ecology*, *29*, 4203–4220.

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

Hancock-Hanser, B. L., Frey, A., Leslie, M. S., Dutton, P. H., Archer, F. I., & Morin, P. A. (2013). Targeted multiplex next-generation sequencing: Advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. *Molecular Ecology Resources*, *13*, 254–268.

Harvey, M. G., Tilston Smith, B., Glenn, T. C., Faircloth, B. C., & Brumfield, R. T. (2016). Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology*, *65*, 910–924.

Hausdorf, B., & Hennig, C. (2020). Species delimitation and geography. *Molecular Ecology Resources*, *20*, 950–960.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, *270*, 313–321.

Herrera, S., & Shank, T. M. (2016). RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. *Molecular Phylogenetics and Evolution*, *100*, 70–79.

Jackson, N., Carstens, B., Morales, A., & O'Meara, B. C. (2017). Species delimitation with gene flow. *Systematic Biology*, *66*, 799–812.

Ješovnik, A., Sosa-Calvo, J., Lloyd, M. W., Branstetter, M. G., Fernández, F., & Schultz, T. R. (2017). Phylogenomic species delimitation and host-symbiont coevolution in the fungus-farming ant genus *Sericomyrmex* Mayr (hymenoptera: Formicidae): Ultraconserved elements (UCEs) resolve a recent a recent radiation. *Systematic Entomology*, *42*, 523–542.

Jones, G. (2017). Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *Journal of Mathematical Biology*, *74*, 447–467.

Knowles, L. L., & Carstens, B. C. (2007). Delimiting species without monophyletic gene trees. *Systematic Biology*, *56*, 887–895.

Krehenwinkel, H., Pomerantz, A. F., Henderson, J. B., Kennedy, S. R., Lim, J. Y., Swamy, V., Shoobridge, J. D., Graham, N. R., Patel, N. H., Gillespie, R. G., & Prost, S. (2019). Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. *Gigascience*, *8*, 1–16.

Leaché, A. D., Zhu, T., Rannala, B., & Yang, Z. (2019). The spectre of too many species. *Systematic Biology*, *68*, 168–181.

Lebonah, D. E., Dileep, A. O., Chandrasekhar, K. B., Sreevani, S., Sreedevi, B., & Kumari, J. P. (2014). DNA barcoding on bacteria: A review. *AIMS Genetics*, *5*, 1–23.

Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, *61*, 727–744.

Lemmon, E. M., & Lemmon, A. R. (2013). High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 99–121.

MacLeod, N., Benfield, M., & Culverhouse, P. (2010). Time to automate identification. *Nature*, *467*, 154–155.

Mayer, C., Sann, M., Donath, A., Meixner, M., Podsiadlowski, L., Peters, R. S., Petersen, M., Meusemann, K., Liere, K., Wägele, J. W., Misof, B., Bleidorn, C., Ohl, M., & Niehuis, O. (2016). BaitFisher: A software package for multispecies target DNA enrichment probe design. *Molecular Biology and Evolution*, *33*, 1875–1886.

Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., Frandsen, P. B., Ware, J., Flouri, T., Beutel, R. G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A. J., Aspöck, U., Aspöck, H., Bartel, D., … Zhou, X. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science*, *346*, 763–767.

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*, 268–274.

Niehuis, O., Hartig, G., Grath, S., Pohl, H., Lehmann, J., Tafer, H., Donath, A., Krauss, V., Eisenhardt, C., Hertel, J., Petersen, M., Mayer, C., Meusemann, K., Peters, R. S., Stadler, P. F., Beutel, R. G., Bornberg-Bauer, E., McKenna, D. D., & Misof, B. (2012). Genomic

and morphological evidence converge to resolve the enigma of Strepsiptera. *Current Biology*, *22*, 1309–1313.
Padial, J. M., Miralles, A., De la Riva, I., & Vences, M. (2010). The integrative future of taxonomy. *Frontiers in Zoology*, *7*, 16.
Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A., & Harris, S. R. (2016). SNP-sites: Rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics*, *2*(4), e000056.
Pante, E., Abdelkrim, J., Viricel, A., Gey, D., France, S. C., Boisselier, M. C., & Samadi, S. (2015). Use of RAD sequencing for delimiting species. *Heredity*, *114*, 450–459.
Pierce, M. P. (2019). Filling in the gaps: Adopting ultraconserved elements alongside COI to strengthen metabarcoding studies. *Frontiers in Ecology and Evolution*, *7*(469), 1–6.
Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D., & Vogler, A. P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, *55*, 595–609.
Prebus, M. M. (2021). Phylogenomic species delimitation in the ants of the *Temnothorax salvini* group (Hymenoptera: Formicidae): An integrative approach. *Systematic Entomology*, *46*, 307–326.
Pritchard, J. K., Stephens, M., & Donnelly, P. J. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.
Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, automatic barcode gap discovery for primary species delimitation. *Molecular Ecology*, *21*, 1864–1877.
Rannala, B., & Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, *164*, 1645–1656.
Rannala, B., & Yang, Z. (2008). Phylogenetic inference using whole genomes. *Annual Review of Genomics and Human Genetics*, *9*, 217–231.
Schlick-Steiner, B. C., Steiner, F. M., Seifert, B., Stauffer, C., Christian, E., & Crozierm, R. H. (2010). Integrative taxonomy: A multisource approach to exploring biodiversity. *Annual Review of Entomology*, *55*, 421–438.
Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*, 3210–3212.
Sukumaran, J., Holder, M. T., & Knowles, L. L. (2021). Incorporating the speciation process into species delimitation. *PLoS Computational Biology*, *17*(5), e1008924.
Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 1607–1612.
Suvorov, A., Kim, B. Y., Wang, J., Armstrong, E. E., Peede, D., D'Agostino, E. R. R., Price, D. K., Waddell, P. J., Lang, M., Courtier-Orgogozo, V., David, J. R., Petrov, D., Matute, D. R., Schrider, D. R., & Comeault, A. A. (2021). Widespread introgression across a phylogeny of 155 Drosophila genomes. *Current Biology*, *32*, 111–123.e5.
Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*, 2045–2050.
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. 3. Cladogram estimation. *Genetics*, *132*, 619–633.
Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, *35*, 543–548.
Waterhouse, R. M., Tegenfeldt, F., Li, J., Zdobnov, E. M., & Kriventseva, E. V. (2013). OrthoDB: A hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Research*, *41*, D358–D365.
Waterhouse, R. M., Zdobnov, E. M., & Kriventseva, E. V. (2011). Correlating traits of gene retention, sequence divergence, duplicability and essentiality in vertebrates, arthropods, and fungi. *Genome Biology and Evolution*, *3*, 75–86.
Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 9264–9269.
Yang, Z., & Rannala, B. (2014). Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution*, *31*, 3125–3135.
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., & Ding, Z. (2012). Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, *3*, 613–623.
Zarza, E., Connors, E. M., Maley, J. M., Tsai, W. L. E., Heimes, P., Kaplan, M., & McCormack, J. E. (2018). Combining ultraconserved elements and mtDNA data to uncover lineage diversity in a Mexican highland frog (Sarcohyla; Hylidae). *PeerJ*, *6*, e6045.
Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*, 153.
Zhang, F., Ding, Y., Zhu, C. D., Zhou, X., Orr, M. C., Scheu, S., & Luan, Y. X. (2019). Phylogenomics from low-coverage whole-genome sequencing. *Methods in Ecology and Evolution*, *10*, 507–517.
Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, *29*, 2869–2876.

SUPPORTING INFORMATION

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

How to cite this article: Dietz, L., Eberle, J., Mayer, C., Kukowka, S., Bohacz, C., Baur, H., Espeland, M., Huber, B. A., Hutter, C., Mengual, X., Peters, R. S., Vences, M., Wesener, T., Willmott, K., Misof, B., Niehuis, O., & Ahrens, D. (2023). Standardized nuclear markers improve and homogenize species delimitation in Metazoa. *Methods in Ecology and Evolution*, *14*, 543–555. https://doi.org/10.1111/2041-210X.14041