Received: 4 July 2023 | Revised: 24 October 2023 | Accepted: 31 October 2023

DOI: 10.1111/1755-0998.13894

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

Species delimitation with limited sampling: An example from rare trapdoor spider genus *Cyclocosmia* (Mygalomorphae, Halonoproctidae)

Vera Opatova[1] | Kellie Bourguignon[2] | Jason E. Bond[3]

[1]Department of Zoology, Faculty of Sciences, Charles University, Prague 2, Czech Republic

[2]Department of Biological Sciences, Auburn University, Auburn, Alabama, USA

[3]Department of Entomology and Nematology, University of California, Davis, California, USA

Correspondence
Vera Opatova, Department of Zoology, Faculty of Sciences, Charles University, 128 44 Prague 2, Czech Republic.
Email: vera.opatova@gmail.com

Funding information
Charles University Research Centre program, Grant/Award Number: UNCE 204069; Evert and Marion Schlinger Foundation; National Science Foundation, Grant/Award Number: DEB 1937604

Handling Editor: Sean D. Schoville

Abstract

The outcome of species delimitation depends on many factors, including conceptual framework, study design, data availability, methodology employed and subjective decision making. Obtaining sufficient taxon sampling in endangered or rare taxa might be difficult, particularly when non-lethal tissue collection cannot be utilized. The need to avoid overexploitation of the natural populations may thus limit methodological framework available for downstream data analyses and bias the results. We test species boundaries in rare North American trapdoor spider genus *Cyclocosmia* Ausserer (1871) inhabiting the Southern Coastal Plain biodiversity hotspot with the use of genomic data and two multispecies coalescent model methods. We evaluate the performance of each methodology within a limited sampling framework. To mitigate the risk of species over splitting, common in taxa with highly structured populations, we subsequently implement a species validation step via genealogical diversification index (*gdi*), which accounts for both genetic isolation and gene flow. We delimited eight geographically restricted lineages within sampled North American *Cyclocosmia*, suggesting that major river drainages in the region are likely barriers to dispersal. Our results suggest that utilizing BPP in the species discovery step might be a good option for datasets comprising hundreds of loci, but fewer individuals, which may be a common scenario for rare taxa. However, we also show that such results should be validated via *gdi*, in order to avoid over splitting.

KEYWORDS
Bayes factor delimitation, BPP, *gdi*, genomic data, Southern Coastal Plain biodiversity hotspot

1 | INTRODUCTION

Delineating species remains a central topic for both taxonomy and evolutionary research. Each species concept emphasizes different aspects of biological diversity and speciation process; what constitutes a species thus correlates with a framework, in which the species are delimited (Cracraft, 1983; de Queiroz, 2007; Freudenstein et al., 2017; Mallet, 2008; Wang et al., 2020). Defining species as independently evolving lineages (de Queiroz, 1998) is a widely accepted approach; however, the interpretation of what constitutes evolutionary independence may be subjective given the continuous nature of the speciation process (de Queiroz, 2007; Stankowski & Ravinet, 2021). Genetic similarities between closely related sibling taxa are traditionally contributed to the incomplete lineage sorting, or to the homogenizing effect of gene flow that actively hampers the speciation process (Carstens & Knowles, 2007; Leache et al., 2014; Smadja & Butlin, 2011). Nevertheless, there is increasing evidence that speciation in the presence of gene flow is prevalent in nature (Jiao & Yang, 2021; Martin et al., 2013; Morales et al., 2017; Nosil, 2008;

Papadopulos et al., 2011). Although low background levels of gene flow do not significantly affect delimitation outcomes for some methodologies (Zhang et al., 2011), current methodological framework commonly used for phylogenetic inference and species delimitation (Bouckaert et al., 2019; Pons, 2006; Yang, 2015; Zhang et al., 2013) does not explicitly account for the presence of gene flow, introgression or recombination. Consequently, it may yield biased results (Camargo et al., 2012; Carstens et al., 2017; Chan et al., 2020).

Study design (Carstens et al., 2013, 2017), which includes sampling strategy (Chambers & Hillis, 2020; Pentinsaari et al., 2017; Talavera et al., 2013), may influence species delimitation outcomes as much as data type and analytical framework. Single-locus approaches based on mitochondrial data routinely yield inflated lineage/species numbers, particularly in sedentary taxa with allopatric or fragmented population distributions (Hamilton et al., 2014; Opatova & Arnedo, 2014b). Therefore, transitioning towards an implementation of multi-locus, and subsequently genomic data, in coalescent-based analyses appeared to be a viable option to avoid over splitting (Dupuis et al., 2012; Edwards & Knowles, 2014). Unfortunately, the increasing amount of data also reveals finer population structure that can be erroneously interpreted as species-level diversity (Leaché et al., 2019; Sukumaran & Knowles, 2017). Species delimitation should thus be performed in an integrative framework, allowing putative species, delimited from molecular data in the species discovery step, to be validated by other evidence, such as morphology and ecological/behavioural data (Carstens et al., 2013; Edwards & Knowles, 2014). However, in case of, for example, morphologically conserved and rare taxa, molecular data may sometimes constitute the only data type that is easily accessible.

Several approaches have been proposed to guide decision making about species boundaries when delimitation is based solely on molecular data. Applying a threshold of the '10x rule' of intra- versus inter-specific variation (Hebert et al., 2004) to detect the 'barcoding gap' has been shown to be overly simplistic at large phylogenetic scales (Čandek & Kuntner, 2015; Koroiva & Kvist, 2018; Kvist, 2016). Therefore, some authors have argued for establishing a divergence threshold based on well-studied reference groups (e.g. human–chimp divergence for mammals) (Galtier, 2019) that could be applied to the delimited lineages. Additionally, the independence of the putative lineages could be validated by assessing and quantifying the gene flow levels among them (Hey, 2009; Jackson, Morales, et al., 2017). However, such approach is time-consuming, requires significant computational effort and might not be suitable for large datasets comprising many loci and many putative lineages (Jackson, Morales, et al., 2017). Implementation of the genealogical divergence index (*gdi*) (Jackson, Carstens, et al., 2017; Leaché et al., 2019) which accounts for a compounded effect of both genetic isolation and gene flow may represent an elegant solution. The *gdi* is a heuristic metric ranging from 0 (panmixia) to 1 (strong divergence). Low values thus correlate with intra-specific diversity ($gdi < 0.2$), while higher ones indicate ($gdi > 0.7$) species-level divergence (Jackson, Carstens, et al., 2017; Leaché et al., 2019). Although large portion of the spectrum represents an ambiguous outcome, *gdi* was effectively used in species delimitation of vertebrate taxa (Chan et al., 2020; Chan & Grismer, 2019; Jackson, Carstens, et al., 2017; Leaché et al., 2020; Mays et al., 2019). Implementation of the *gdi* estimation might be particularly useful in conservation where accurate species boundaries assessment and individual population differentiation is crucial for developing effective management strategies and protection plans (Coimbra et al., 2021; Devitt et al., 2019; Hosegood et al., 2020; Johnson et al., 2018).

Obtaining sufficient taxon sampling is important in the species delimitation process (Chambers & Hillis, 2020; Pentinsaari et al., 2017; Talavera et al., 2013); however, it might be difficult in endangered or rare taxa. In order to avoid overexploitation, the sampling strategy usually relies on non-lethal tissue collection (Hamm et al., 2010; Machkour M'Rabet et al., 2009; Ožana et al., 2020) or alternatively on completely non-invasive approaches such as analyses of shed hair and faeces (Beja-Pereira et al., 2009; Dufresnes et al., 2019; Waits & Paetkau, 2005). Unfortunately, implementing such sampling methods might be challenging in some invertebrate taxa due to, for example, small body size, cryptic lifestyle and the risk of subsequent wound-related infection. Whole specimen collection is thus often inevitable. For example, dissection of specific body parts or internal structures may be necessary for correct identification (Araujo et al., 2020; Christophoryová et al., 2023; Homma et al., 2020; Kergoat et al., 2015), sufficient DNA yield requires whole body extraction in small organisms or a reference voucher is needed (Català et al., 2021; Johnson et al., 2018). In case of unavoidable whole specimen sampling, the necessity to prevent the overexploitation of the species' natural populations may thus further restrict the number of collected individuals (Hedin, 2015; Jacobs et al., 2018). Conversely, similar situation may also arise due the lack of resources. As a consequence, for example, presence of singletons and insufficient sampling for population size parameter estimations may subsequently limit the methodological framework available for downstream data analyses and bias the results (Bouckaert et al., 2019; Flouri et al., 2018; Huang et al., 2020).

Mygalomorph spiders (tarantula, trap door and funnel web spiders) represent an ancient group of sedentary animals. Their current family-level distribution is linked to major continental-level vicariate events (Godwin et al., 2018; Opatova et al., 2020; Opatova & Arnedo, 2014a) and they typically evolve in allopatry (Bond et al., 2012). Phenotype in mygalomorph spiders is strongly correlated with their behavioural niche, particularly their foraging strategy and web/burrow construction (Wilson et al., 2023). As a consequence, repeated patterns of both morphological stasis and homoplasy are present among distantly related taxa (Hedin et al., 2019; Opatova et al., 2020; Wilson et al., 2023). Morphological homogeneity is particularly common at shallow phylogenetic level, but at the same time, mygalomorphs often possess a remarkably deep intra-specific genetic structuring (Bond et al., 2001). For these particular reasons mygalomorph spiders represent intriguing model in molecular species delimitation research (Bond & Stockman, 2008;

Candia-Ramírez & Francke, 2021; Hamilton et al., 2011; Hedin, 2015; Leavitt et al., 2015; Montes de Oca et al., 2016; Newton et al., 2020; Satler et al., 2013).

The trapdoor spider genus *Cyclocosmia* Ausserer (1871) comprises 10 species with a disjunct distribution, seven inhabit southeastern Asia and three North America (two are known from the Southeastern USA, one from eastern Mexico) (World Spider Catalogue, 2023). The abdomen of the genus is uniquely modified for phragmosis – a defence strategy when a modified body part serves for closing the burrow/nest. It is highly sclerotized and truncated with a disk-shaped plate covered with ribs and grooves and rimmed with thick setae (Gertsch & Platnick, 1975; Zhu et al., 2006). This modification likely evolved as a defence mechanism against parasitic wasps and predators. The abdomen acts like a shield, tightly fitting into the steeply narrowing bottom part of the burrow, where the spider retreats in case of danger (Gertsch & Platnick, 1975; Hunt, 1975). Species descriptions and taxonomy of the genus is primarily based on the morphological differences (number of ribs, groves, setation) found on the truncated abdominal plate (Gertsch & Platnick, 1975; Xu et al., 2017; Zhu et al., 2006).

Two *Cyclocosmia* species inhabit the Southeastern United States, a region recently defined as the Southern Coastal Plain biodiversity hotspot (Noss et al., 2015). *Cyclocosmia truncata* (Hentz 1841) is known from Alabama, central-southern Tennessee and northwestern Georgia; *C. torreya* (Gertsch & Platnick, 1975) is only known from southwestern Georgia and northwestern Florida. The secondary sexual characteristics (male pedipalp and mating clasper, female spermathecae) are homogeneous for these two species; the diagnosis is based on the differences in abdominal truncation. The rim is more protruding in *C. torreya* than in *C. truncata*, and there is generally a difference in the pattern of gaps along the ventral edge (see morphology-based hypothesis *H1* below). Setae projecting off the rim are more numerous in *C. torreya* and the number of ribs on the truncation is slightly higher in *C. truncata* (Gertsch & Platnick, 1975). Both species have affinities towards well-preserved forest habitats, humid ravines and stream banks. Reduced by landscape development, such habitats are now typically highly fragmented with *Cyclocosmia* not particularly abundant even at suitable localities, which many years ago led Gertsch and Platnick (1975) to consider the genus as threatened.

In this article, we aim to assess species boundaries and species diversity of the trapdoor spider genus *Cyclocosmia*, inhabiting the Southern Coastal Plain biodiversity hotspot, based on samples from across most of its geographic range and genomic data. We perform molecular species delimitation with two different methods (BFD, BPP) and then evaluate their performance on our dataset. We also implement the *gdi* metric to validate the species delimitation outcome. Delimiting species in organisms that are rare and occur in low abundance, which is a relatively common phenomenon among arthropods (Coddington et al., 2009), presents certain analytical challenges. We assess various approaches using the *Cyclocosmia* dataset and make recommendations on how to fine tune inferences when based on sparse sampling.

2 | MATERIALS AND METHODS

2.1 | Taxon sampling, molecular protocols and phylogenomic analyses

Twenty-three specimens were used in this study. The genus *Cyclocosmia* was represented by 19 specimens; outgroup taxa comprised four species of the Holanoproctidae genera *Ummidia* Thorell, 1875 and *Hebestatis* Simon, 1903 (Godwin et al., 2018; Opatova et al., 2020). Representatives of the genus *Cyclocosmia*, inhabiting the Southern Coastal Plain biodiversity hotspot, were sampled throughout most of their known geographic range. Fourteen samples morphologically corresponded to *C. truncata* and three to *C. torreya*. Two additional *Cyclocosmia* specimens from Southeast Asia were included in the study. Most of the genomic data was newly generated for this study, additional *Cyclocosmia* sequences and outgroup sequences were obtained from Godwin et al. (2018). Details regarding the sequence data and taxon sampling are provided in Table S1.

Whole genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen), following the manufacturer's guidelines; RNase A was added to the mix following the tissue lysis step in order to ensure complete digestion of residual RNA. Library preparation, enrichment and sequencing were performed at the Center for Anchored Phylogenomics at Florida State University (http://anchoredphylogeny.com/) following the methods described in detail in Lemmon et al. (2012), Hamilton, Lemmon, et al. (2016) and Godwin et al. (2018). Briefly, up to 500 ng of genomic DNA of each sample was sonicated with a Covaris E220 ultrasonicator to a fragment size of ~300–800 bp. Indexed libraries were prepared following Meyer and Kircher (2010) with a modifications for Beckman Coulter Biomek FXp liquid-handling robot (see Hamilton, Lemmon, et al., 2016 for details). Following the blunt-end repair, size selection was performed with SPRI beads (Beckman Coulter, Inc.). Indexed samples were pooled at equal quantities, each pool consisting of 16 samples, and enriched with the AHE Spider Probe kit from Hamilton, Lemmon, et al. (2016) available at the Center for Anchored Phylogenomics at Florida State University. The enriched libraries were then sequenced on PE150 Illumina HiSeq 2500 platform at Florida State University Translational Science Laboratory in the College of Medicine. Orthology was determined among the homologous consensus sequences at each locus following Prum et al. (2015) and Hamilton, Lemmon, et al. (2016). Sequences in each locus were aligned using MAFFT v7.23b (Katoh & Standley, 2013), implementing the E-INS-I algorithm with –genafpair and –maxiterate 1000 flags. Allelic phases were determined for each individual from homologous within-individual sequence clusters at each locus (Lemmon & Lemmon, 2012). The resulting alignment was subsequently trimmed following Hamilton, Lemmon, et al. (2016). In order to have a complete ingroup sequence representation at each locus for downstream species delimitation analyses (see below), only 415 from the 601 recovered loci were retained for the analyses. The quality of the alignment at each locus was scored for accuracy in ALISCORE (Misof & Misof, 2009) and ambiguously

aligned positions were removed with ALICUT (Kück, 2009). The quality of the individual alignments was visually inspected in Geneious 10.1.3 (https://www.geneious.com).

In order to assess the monophyly of both *C. truncata* and *C. torreya*, and reconstruct their internal relationships, the 415 loci were concatenated with FASconCAT (Kück & Longo, 2014), in a supermatrix comprising 19 *Cyclocosmia* terminals and 4 outgroup taxa (hereafter 'Concat' matrix); all individuals were represented by a single allele, each. Partitioning schemes under the GTR+G model were selected based on AICc criterion in PartitionFinder 2 (Lanfear et al., 2016) using the rcluster algorithm (Lanfear et al., 2014) with RAxML implementation (Stamatakis, 2014). Partitioning by codon was not considered in the analyses because ambiguously aligned positions were removed with ALICUT. Maximum likelihood analyses (ML) were conducted in RAxML v 8.2.9 (Stamatakis, 2014), using the 88 partition scheme and an independent GTR+G model assigned to each partition. *Ummidia* samples were used to root the tree. The best ML tree was selected from 1000 iterations, each starting from an independent parsimony-based tree. The node support was inferred from 1000 bootstrap replicates. Bayesian inference (BI) was conducted in ExaBayes v 1.4.1 (Aberer et al., 2014) implementing the same partition scheme and nucleotide substitution model as in the ML analyses. Two independent runs of 4×10^7 generations with four coupled chains each, starting from a parsimony tree with re-sampling every 1000 generations, were run simultaneously. Standard deviation of split frequencies was monitored (<0.01) and the first 25% were discarded as a *burn-in* for the analyses. An extended majority rule consensus tree was obtained with the ExaBayes accompanying program consense (Aberer et al., 2014).

2.2 | Molecular species delimitation

We used two independent molecular species Bayesian framework delimitation approaches under the multispecies coalescent model (MSC) in order to assess species boundaries in sampled North American *Cyclocosmia*. As a species discovery step, we used both Bayes factor delimitation (BFD) approach (Grummer et al., 2014) and joint species delimitation and species tree estimation (A11) approach implemented in the program BPP (Flouri et al., 2018; Yang & Rannala, 2014). Both approaches can be applied to large datasets and neither is limited to species delimitation over one fixed topology in the species discovery step, eliminating potentially biased results. The BFD relies on testing multiple delimitation scenarios based on different topologies (Grummer et al., 2014), while the topology can be directly estimated in BPP analyses (Flouri et al., 2018; Yang & Rannala, 2014). The status of the delimited lineages was subsequently tested by assessing the *gdi* values among sister taxa in the species validation step. In order to fulfil the MCS model assumption and still use singletons in our analyses, sequences of all individuals were phased (see earlier) and therefore represented by two alleles at each locus.

BFD allows for comparison of the marginal likelihoods of the competing models (i.e. species delimitation hypotheses) and their subsequent ranking. Marginal likelihood estimates (MLEs) were obtained via stepping stone (SS) analyses (Xie et al., 2011), an approach known to yield consistently reliable results (Baele et al., 2012; Grummer et al., 2014), implemented in the BEAST v 2 program package (Bouckaert et al., 2019). To overcome the extreme computational demands of the species tree estimation and MLEs impeding the analyses of the full dataset (415 loci) in a time-feasible manner, we created 10 subsets of 50 loci independently selected at random (Morales et al., 2017; O'Neill et al., 2013). The sets comprised all sampled North American *Cyclocosmia* individuals and *Cyclocosmia* sp. from Thailand; all individuals were represented by both alleles of each locus. Additionally, two sets comprising 50 of the most and the least phylogenetically informative loci, respectively, were created. The loci were selected based on the proportion of parsimony informative (pi) sites calculated from the nucleotide alignment of the North American *Cyclocosmia* in Mega 7 (Kumar et al., 2016). The list of randomly selected loci and their respective variability is reported in the Table S2.

Five competing species delimitation hypotheses were tested on all 12 sets of 50 loci. The morphology-based hypothesis (*H1*, three species: *C. torreya*, *C. truncata*, *Cyclocosmia* sp. from Thailand) tested the validity of the current taxonomic scheme and the monophyly of *C. truncata*. Hypotheses 2, 3 and 4 (*H2*, four species; *H3*, five species; *H4*, seven species) were derived from the results of the concatenated analyses (see Figure 1a) and tested whether the clades recovered within *Cyclocosmia* correspond to independent lineages (Figure 2a, Table S3). In these hypotheses, the samples morphologically identified as *C. truncata* were parcelled from two to five groups respectively. In order to test the reliability of the BFD results (Grummer et al., 2014), we also tested an 'over splitting' hypothesis (*H5*, 15 species), where individuals from each sampling locality (or sites within 20 km of each other) were treated as species. This decision was guided by both our knowledge of trapdoor spider genetic background (e.g. each population tends to be delimited as an independent GMYC lineage; Opatova & Arnedo, 2014b; Opatova et al., 2016) and the results of the concatenated analyses (lack of clade geographic overlap). Additionally, a 'reassignment' hypothesis (*H6*, 4 species), where *C. torreya* samples were lumped together with distantly related *C. truncata* samples from the clade I (see results of the concatenated analyses) was also tested.

Both MLEs, obtained with SS analyses for all 12 sets, and species tree analyses were inferred from two independent runs of multispecies coalescent analyses in *BEAST v 2 (Heled & Drummond, 2010). Independent HKY+G substitution models with empirical base frequencies and strict clock models with log normal distribution were defined for each partition. The Yule speciation model was set as a tree prior; diffuse parameters (M=15, S=2) were assigned to the log normal distribution of the birthRate prior. The monophyly of North American *Cyclocosmia* was enforced (see results of the concatenated analyses) in all analyses in order to speed up the computation time. The analyses were run for 2×10^8 generations, retaining samples every 5000 generations. Convergence between the runs and the correct mixing within each run were visualized with TRACER (Rambaut & Drummond, 2009). Individual runs were combined in the BEAST accompanying program LOGCOMBINER, discarding the first

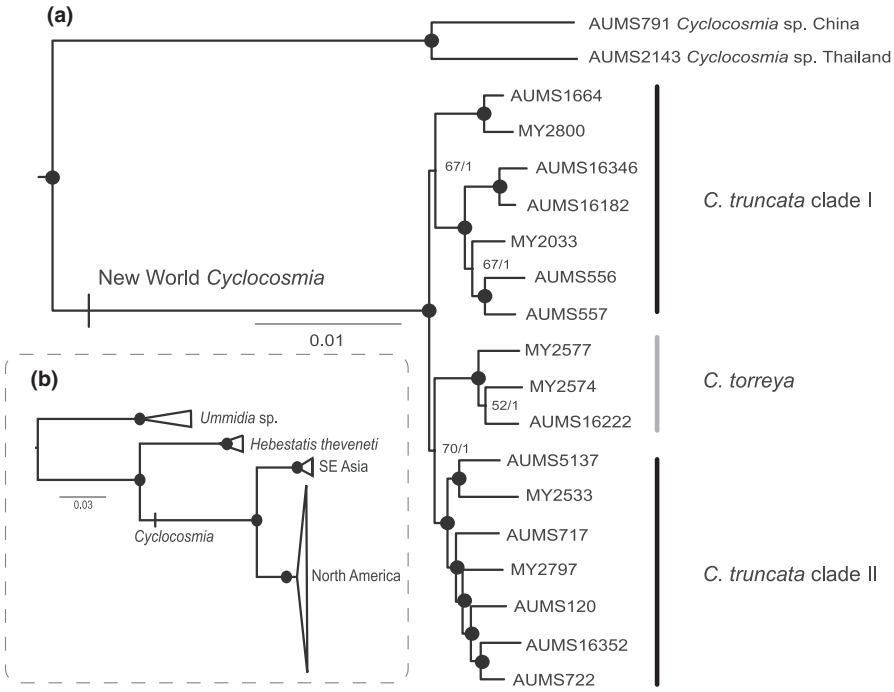

FIGURE 1 Phylogenetic trees of *Cyclocosmia* relationships obtained in concatenated analyses. Topology obtained in the maximum likelihood analyses. Values on nodes denote support values obtained in each approach (left to right): RAxML bootstrap support, ExaBayes Bayesian posterior probabilities (PP), dots denote full support (bootstrap = 100, PP >0.95). (a) Relationships among North American *Cyclocosmia* in detail. (b) Phylogenetic position and reciprocal monophyly of *Cyclocosmia* clades.

20% of the generations of each run as a *burn-in*. A consensus species tree was inferred with TREEANNOTATOR. Stepping stone analyses ($\alpha = 0.3$) were performed in two independent runs of 50 steps, each step running for 1×10^8 generations. The first 5×10^6 generations of samples were set as a pre *burn-in* and were not retained, an additional 20% of the generations were discarded as a *burn-in* for the analyses. The convergence of each step was evaluated through the ESS values ≥100 respectively. The resulting MLEs of both independent SS runs were averaged for subsequent BFD. Bayes factors were calculated as two times the difference of the −lnlikelihood (2lnBf) and interpreted following the methods of Kass and Raftery (1995), i.e. 2lnBf 0–2: 'not worth more than a bare mention', 2lnBf 2–6: 'positive' support, 2lnBf 5–10: 'strong' support, 2lnBf >10: 'decisive' support in distinguishing between competing hypotheses.

To complement the BFD approach, we also implemented rjMCMC algorithm in the program BPP (joint species delimitation and species tree estimation; A11 analysis) (Flouri et al., 2018; Yang & Rannala, 2014). The assignment of the individuals corresponded to the 'over splitting' hypothesis defined for BFD (*H5*, 15 species), where individuals from each sampling locality (or sites 20 km apart) were treated as species. The analyses were conducted independently on all 12 sets of 50 loci and also on the dataset comprising all 415 loci ('set_all'); all individuals were represented by both alleles of each locus. All rooted species trees were assigned equal probabilities (Prior 1; Yang & Rannala, 2014). Within each species tree model, we assigned inverse-gamma prior $\theta \sim$ IG (3, 0.02) for all θ parameters and $\tau \sim$ IG (3, 0.01) for the root ages. The IG (α, β) values for β were empirical, following (Flouri et al., 2018), while $\alpha = 3$ represented a diffuse prior. First 100,000 generations were discarded as a *burn-in* for the analyses, 100,000 samples were collected from the MCMC run, sampling every tenth generation. Three independent runs starting from two different starting topologies: (out, ((sp1, sp2), (((sp3, sp4), (sp5, (sp6, sp7))), ((sp8, (sp9, sp10)), (sp11, (sp12, (sp13, sp14)))))); and (out,((sp8,(sp1,sp2)),(((sp13,sp4),(sp5,(sp6,sp7))),((sp9,sp10),((sp11,sp12),(sp3,sp14)))))) (Figure 2b) were ran for each set of 50 loci in order to assess the convergence of the analyses. Four independent runs were conducted for the 'set_all' dataset.

The resulting delimitation model obtained from the 'set_all' was further tested in the species verification step. Values of *gdi* were calculated for pairs of delimited sister lineages in order to assess the level of their divergence due to the compounded effect genetic isolation and gene flow. The *gdi* was calculated using the following equation: $gdi = 1 - e^{-2\tau/\theta}$. The level of differentiation of lineage A from lineage B was assessed with $2\tau_{AB}/\theta_A$, whereas $2\tau_{AB}/\theta_B$ was used to differentiate B from A (Jackson, Carstens, et al., 2017; Leaché et al., 2019). The *gdi* values were calculated using Excel from the posterior distribution of lineage-specific θ and τ parameter states estimated under the MSC model in the within-model inference approach (A00 analysis) in BPP (Rannala & Yang, 2003) and retained in MCMC runs. The resulting density of *gdi* was visualized in TRACER (Rambaut & Drummond, 2009). The *gdi* values were interpreted as follows: *gdi* <0.2 correspond to intra-specific diversity, *gdi* >0.7 corresponds to species-level divergence, 0.2 < *gdi* < 0.7 represents an ambiguous result (Jackson, Carstens, et al., 2017; Leaché et al., 2019). The settings of inverse-gamma priors and MCMC sampling conditions of the A00 analyses were the same as described earlier. Two independent runs of each analysis were conducted on the 'set_all' dataset, and (out, (((sp1, sp2), (sp3sp4, (sp5, (sp6, sp7)))), (sp8, ((sp9, sp10), (sp12, (sp11, (sp13, sp14)))))), topology in order to check the convergence. Sister terminals with *gdi* <0.7 were merged into one species and the resulting topology was used for subsequent A00 analyses (Figure 3, Figure S2). In cases where one terminal was statistically different from its sister (*gdi* > 0.7), but not vice versa, we maintained the terminals as separate lineages in the downstream analyses (Figure 3, Table 2). The topology and support values of the resulting species delimitation scheme were

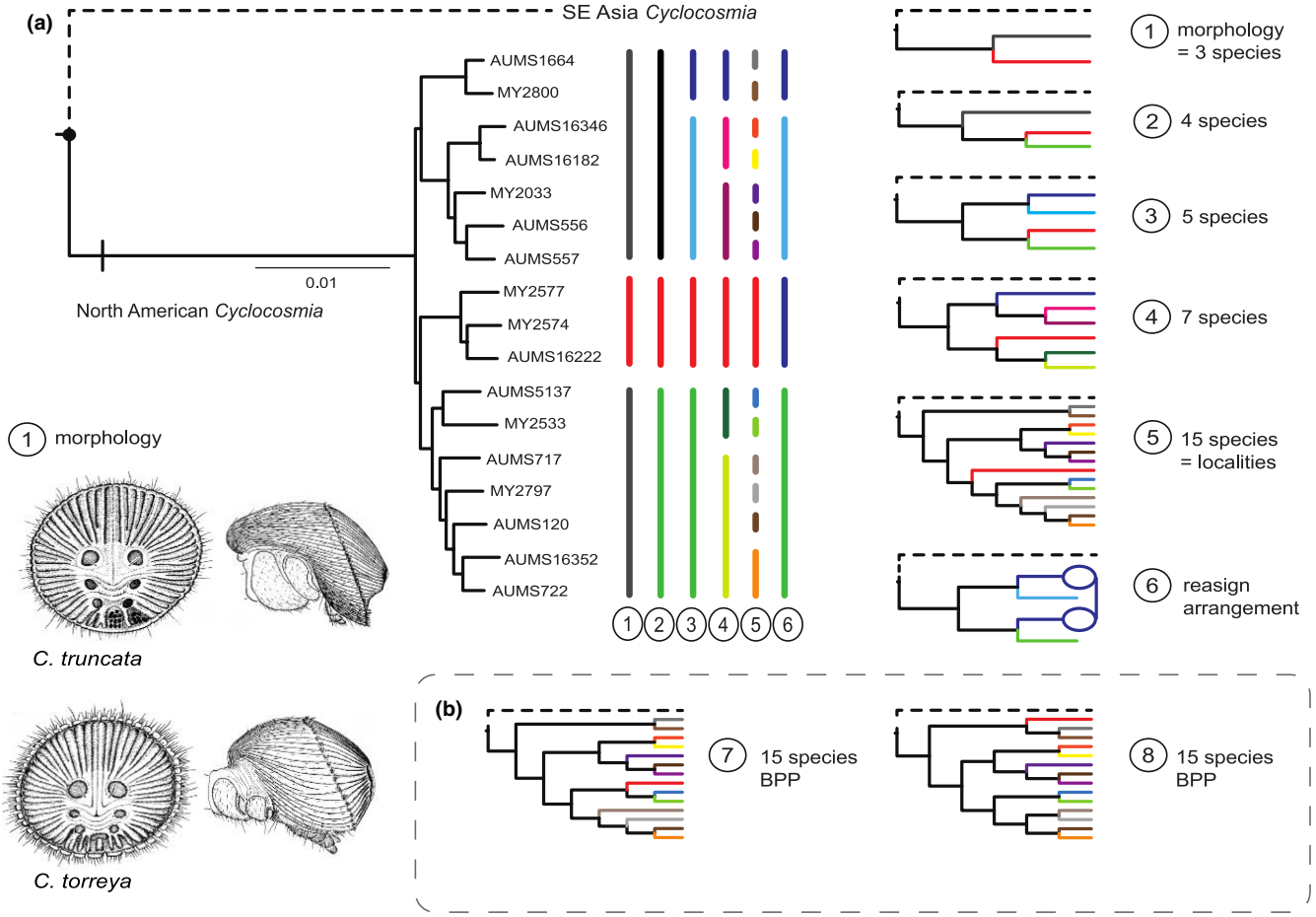

FIGURE 2 Species discovery step delimitation hypotheses and starting topologies. (a) Topology obtained in the concatenated analyses. Colour coding corresponds to the assignments of individuals to different delimitation hypotheses (on the right). 1: Morphology driven hypothesis (adapted from Gertsch & Platnick, 1975), 2–4: hypotheses derived from the structure of *Cyclocosmia* phylogenetic relationships, 5: over split hypothesis (localities assigned as species), 6: reassign arrangement. (b) Starting topologies in BPP A11 analyses. Delimitation scenario and terminal colour coding correspond to over split hypothesis (localities assigned as species, *H5*). 7: Topology derived from the structure of *Cyclocosmia* phylogenetic relationships (*H5*), 8: randomly generated topology for the over split delimitation hypothesis *H5*.

inferred in BPP A01 analyses. To assess the convergence of the runs, we ran the analyses twice from two different starting topologies (Figure S1) and the parameter setting described earlier. Additionally, we also conducted several exploratory analyses, where terminal pairs were merged even when only of the terminals within the pair of sister lineages obtained *gdi* > 0.7 (Figure S2).

All phylogenetic and species delimitation analyses were run on the Hopper Community Cluster at Auburn University.

3 | RESULTS

3.1 | Concatenated analyses

The complete dataset 'Concat' supermatrix comprised 415 loci (131,383 nucleotides) for 23 specimens representing the genus *Cyclocosmia* (19 specimens) and outgroup taxa (Table S1). The alignment possessed complete locus-wise representation. Each individual was represented by a single allele at each locus. The proportion of gaps and missing data was 5.32% in the dataset. The trimming of the alignments was minimal (Portik & Wiens, 2021) in the majority of loci, only 35 loci out of 415 were trimmed more than 30%. The resulting alignment is available from Dryad Digital repository (see Data Availability Statement). Both analyses yielded the same topology (Figure 1a,b). All nodes were fully supported in the BI (PP = 1); ML topology (−ln 384,709.63) received lower supports (bootstrap = 67–70) among North American *Cyclocosmia* clades. The genus *Cyclocosmia* was recovered as monophyletic in all analyses. Two *Cyclocosmia* species from Southeast Asia formed a clade sister to the North American taxa; both clades were recovered with full support in all analyses. The monophyly of *C. truncata* was not recovered in our analyses. Its diversity was divided across two clades (clade I, II) with *C. truncata* clade II as sister to *C. torreya*.

3.2 | Species delimitation – Discovery step

The list of the 50 loci forming each set and the respective variability of each locus derived from the proportion of pi sites within the

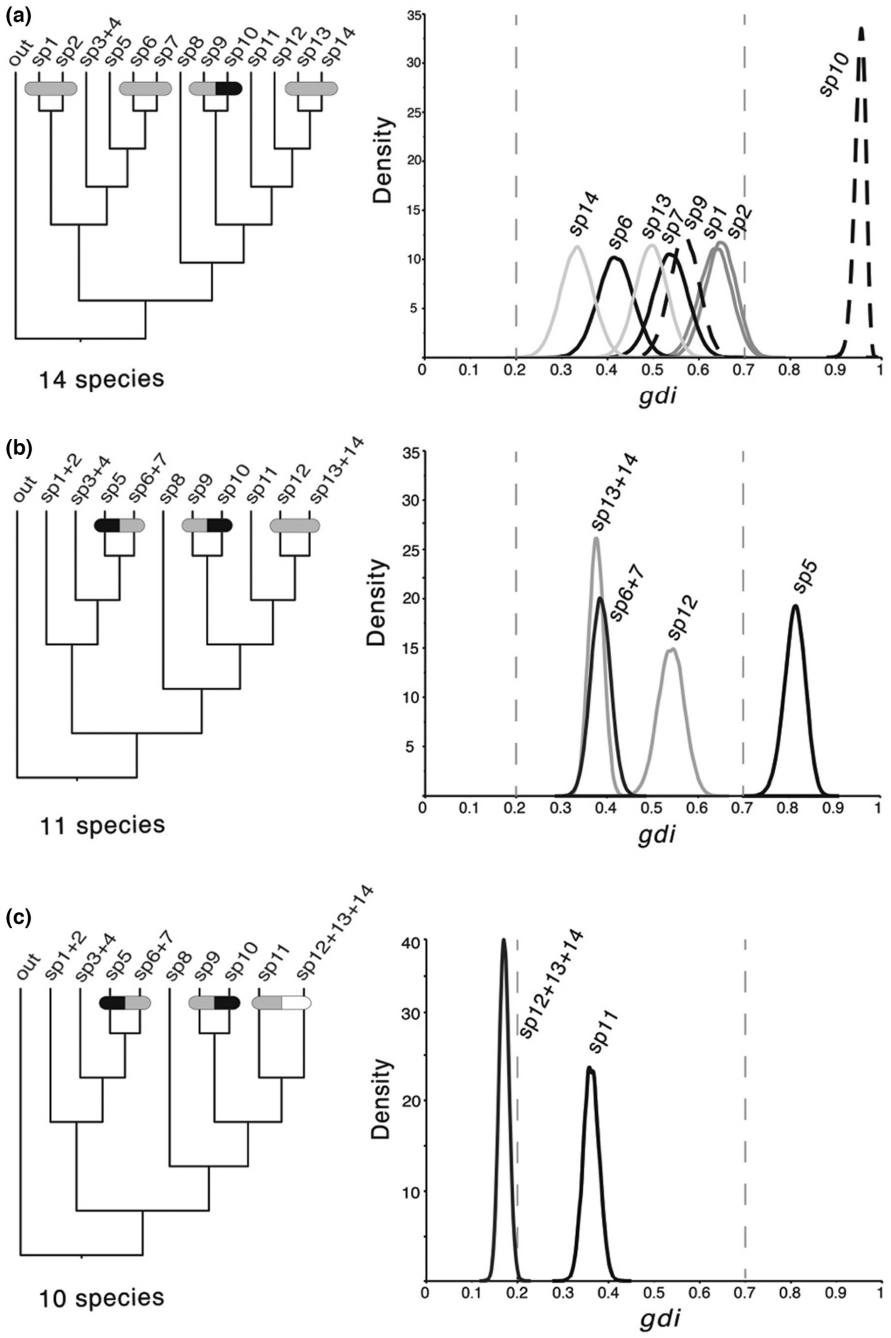

FIGURE 3 Species validation step. Hierarchical process of applying genealogical divergence index (*gdi*) to parameter estimates obtained in A00 analyses in BPP. Tested species trees on the right, posterior probabilities of *gdi* on the left, *gdi* <0.2 indicates a single species (white), 0.2 < *gdi* < 0.7 ambiguous status (grey), *gdi* >0.7 indicates district species (black). Lineages were merged when 0.2 < *gdi* < 0.7 in both tested sister lineages. (a) *gdi* analyses performed on the resulting topology from the species discovery step in BPP (A11 analyses based on 425 loci), (b) *gdi* analyses performed on a merged 14 species topology, (c) *gdi* analyses performed on a merged 11 species topology. The *gdi* values for each validation step are reported in Table 2.

ingroup (North American *Cyclocosmia*) is reported in the Table S2. Ten sets comprised 50 randomly selected loci, and two additional sets represented the most and least variable loci from the pool of 415 loci. MLEs were obtained via two independent SS analyses for all defined species delimitation hypotheses and averaged for subsequent BF calculation. Convergence of most steps of the SS analyses was achieved in all sets of loci and all defined hypotheses with the exception of the 'over splitting' hypothesis (*H5*, 15 species). The lack of convergence in case of the 'over splitting' hypothesis was consistent across all sets of loci and tested hypotheses. On average, less than 10 out of 50 steps of the SS analyses achieved ESS >100 (results not shown). The resulting MLEs average values of the 'over splitting' hypothesis were thus excluded from the subsequent BFD delimitation, despite being the lowest (i.e. the best) (Table 1). The resulting averages of MLEs of the competing species delimitation hypotheses and their ranking according to BFD results (2*ln*BF) in the Table 1. The seven-species delimitation model (hypothesis *H4*) received the best MLEs consistently across the sets. The BF comparison between the favoured model and any other hypotheses/delimitation models yielded values well above 2lnBf >10 in all sets; that is we obtained a 'decisive' support for distinguishing between competing hypotheses.

The species tree topology obtained in *BEAST for the favoured seven-species delimitation model differed among the different sets of loci (Figure S3). The topology shared among the sets 4, 6 and 9 placed *C. torreya* as sister to all the remaining lineages

TABLE 1 Marginal likelihood estimates (MLE), competing species delimitation model ranking and Bayes factor results (2lnBF) proceeding from combined results of two independent runs of steppingstone (SS) analyses. Model receiving the best MLE for each set of loci (in bold) is indicated with 2lnBF = N/A. MLE for 15 species model result from non-converging SS analyses. Set MV: set with 50 most variable loci; Set LV: set with 50 least variable loci.

	Set 1			Set 2			Set 3			Set 4			Set 5			Set 6		
Hypothesis	**MLE**	**Rank**	**2*ln*BF**	**MLE**	**Rank**	**2*ln*BF**	**MLE**	**Rank**	**2*ln*BF**	**MLE**	**Rank**	**2*ln*BF**	**MLE**	**Rank**	**2*ln*BF**	**MLE**	**Rank**	**2*ln*BF**
Morphology	−74,671.6	5	2651	−74,763.4	5	2459	−71,245	5	2866	−71,918.7	5	2816	−75,933.2	5	2537	−73,015	5	2712
4 Species	−74,079.9	4	1468	−74,223.8	4	1380	−70,590.4	4	1557	−71,255.2	3	1489	−75,336.2	4	1343	−72,407.3	4	1496
5 Species	−73,736.8	2	782	−73,898.2	2	729	−70,240.6	2	857	−70,953.9	2	887	−75,040.8	2	753	−72,067.6	2	817
7 Species	**−73,346**	**1**	**N/A**	**−73,533.7**	**1**	**N/A**	**−69,811.9**	**1**	**N/A**	**−70,510.6**	**1**	**N/A**	**−74,664.5**	**1**	**N/A**	**−71,659.1**	**1**	**N/A**
15 Species	−73,243.9	–	–	−73,500.2	–	–	−69,755.5	–	–	−70,457.4	–	–	−74,595.7	–	–	−71,593.4	–	–
Reassign	−74,017.2	3	1342	−74,200.2	3	1333	−70,580.7	3	1538	−71,255.9	4	1491	−75,315.3	3	1302	−72,373.4	3	1429
	Set 7			**Set 8**			**Set 9**			**Set 10**			**Set MV**			**Set LV**		
Hypothesis	**MLE**	**Rank**	**2*ln*BF**	**MLE**	**Rank**	**2*ln*BF**	**MLE**	**Rank**	**2*ln*BF**	**MLE**	**Rank**	**2*ln*BF**	**MLE**	**Rank**	**2*ln*BF**	**MLE**	**Rank**	**2*ln*BF**
Morphology	−70,242.2	5	2041	−78,040.6	5	2112	−76,245.6	5	2755	−70,355.2	5	2022	−90,696.7	5	1763	−66,893.5	5	2528
4 Species	−69,619.9	3	796	−77,350.5	3	732	−75,606.8	3	1477	−69,849.6	4	1010	−90,212.6	4	795	−66,294.2	4	1329
5 Species	−69,316	2	188	−77,284	2	599	−75,327	2	918	−69,570.4	2	452	−90,021.4	2	413	−65,989.5	2	720
7 Species	**−69,221.8**	**1**	**N/A**	**−76,984.4**	**1**	**N/A**	**−74,868.2**	**1**	**N/A**	**−69,344.4**	**1**	**N/A**	**−89,815.1**	**1**	**N/A**	**−65,629.5**	**1**	**N/A**
15 Species	−69,663.1	–	–	−77,393.9	–	–	−74,773.8	–	–	−69,705.8	–	–	−90,163.9	–	–	−65,576.3	–	–
Reassign	−69,714.7	4	986	−77,469.3	4	970	−75,607.2	4	1478	−69,789.1	3	889	−90,164.6	3	699	−66,251.7	3	1244

TABLE 2 The *gdi* values for each pair of lineages and each validation step, values indicating independent status (*gdi* >0.7) marked in bold.

Hypothesis	Species A, B	*gdi* A > B	*gdi* B > A
14 Species	sp1, sp2	0.65	0.64
	sp6, sp7	0.42	0.54
	sp9, sp10	0.57	**0.95**
	sp13, sp14	0.5	0.33
11 Species	sp5, sp6+7	**0.81**	0.39
	sp9, sp10	0.57	**0.95**
	sp12, sp13+14	0.54	0.37
10 Species	sp5, sp6+7	**0.81**	0.39
	sp9, sp10	0.57	**0.95**
	sp11, sp12+13+14	0.37	0.17

morphologically corresponding to *C. truncata*. The same position of *C. torreya* was also recovered in the set 7 and the set of the least variable loci respectively. However, the remaining relationships differed among the topologies. An alternative topology, shared by the sets 3 and 5 is in agreement with the topology recovered in the concatenated analyses.

The results of joint species delimitation and species tree estimation (A11 analysis), starting from two different 15-terminal guide trees (see above), implemented in the program BPP differed among the sets of loci (Table S4). Fourteen species were delimited from eight sets of loci; however, the delimitation model received a consistent support across all runs (PP > 0.95) only in four sets of loci (1, 9, 10 and the set of most variable loci). Thirteen species were delimited from three sets of loci, but without support. Finally, two runs out of three supported an 11-species delimitation model in the set of least variable loci. All four independent runs performed on the full set of 415 loci ('set_all') highly supported (PP >0.99) 14 species delimitation model. The resulting topology of the delimited species was consistent with the topology obtained in the concatenated analyses.

3.3 | Species delimitation – Validation step

Because of the consistency of the results and computational feasibility, the species validation step via *gdi* was performed only on the 14 species delimitation model and the 'set_all' dataset comprising 415 loci. The results of the *gdi* values for each analysis are reported in Table 2. The analyses subsequently reduced the number of terminals (i.e. delimited species) to 11, 10 and finally 9. We encountered in several cases that the *gdi* values supported the independent status of one species in the sister taxa pair, but not vice versa. The topology and support of the final species delimitation scheme (nine species, including the outgroup) obtained in the BPP A01 analyses are depicted in Figure 4a, Figure S1; the topology is consistent with the results of the concatenated analyses. This delimitation model supports a division of *C. truncata* into seven independent lineages (Figure 3).

In our exploratory analyses, where we proceeded to progressively merge all terminals unless they showed values of *gdi* >0.7 reciprocally, we reduced the number of terminals in each round of analyses from 14 to 10, 8, 6, 4, 3 and finally 2. The entire sampled North American *Cyclocosmia* diversity was thus merged into one species (Figure 4a,c, Figure S2).

4 | DISCUSSION

4.1 | Species discovery step

Our delimitation analyses yielded significantly different outcomes, depending both on the size of the analysed datasets and delimitation method. Bayesian factor delimitation (BFD) performed on 10 sets of 50 randomly selected loci and 2 sets of the 50 most and least variable loci delimited 6 North American *Cyclocosmia* lineages in the species discovery step, whereas the A11 analyses performed in BPP delimited up to 13. It is not uncommon for different analytical methods to produce different number of delimited units (Jacobs et al., 2018; Magoga et al., 2021; Petzold & Hassanin, 2020). To an extent, both BFD and BPP may be prone to confounding population structure with species boundaries (i.e. species over splitting) (Jackson, Morales, et al., 2017; Leaché et al., 2019). However, why BFD delimited lower number of species likely stems from a combination of insufficient data sampling and need for predefined hypotheses testing. The results of the two methods are thus not directly comparable. The number of sequences per lineage is particularly influential in the multispecies coalescent delimitation (Huang et al., 2020). In the case of *Cyclocosmia*, the majority of the terminals in the 'over splitting' hypothesis (*H5*, 15 species) were represented by only two alleles that likely did not contain sufficient information for parameter estimation in the *BEAST and SS analyses. The majority of the steps of the SS analyses thus did not converge (Oaks et al., 2019) and the results for this delimitation scenario could not be used in the BFD. The favoured delimitation model for BFD thus artefactually became the seven-species hypothesis (*H4*, Figure 2a), representing the next highest number of species tested after the 'over splitting' hypothesis. These results highlight the potential problems stemming from prior specimen assignment to units, because explicit testing of all the possible delimitation scenarios is likely not feasible for most datasets. Nevertheless, the results of the seven-species hypothesis remained ambiguous, because different sets of loci favoured different topology (Figure S3). Such an outcome suggests that the respective 50 loci sets might contain more information about species delimitation than about the relationships among the delimited taxa (Yang, 2015).

We also encountered inconsistencies in the results of BPP discovery step proceeding from the datasets limited to 50 loci; 13, 12 and 10 North American *Cyclocosmia* species were delimited, respectively, with a varying degree of support (Table S4). The scenario comprising 13 North American species was the most supported hypothesis among the different 50 loci sets and it was also consistently supported in all the analyses based on 415 loci (*set_all*).

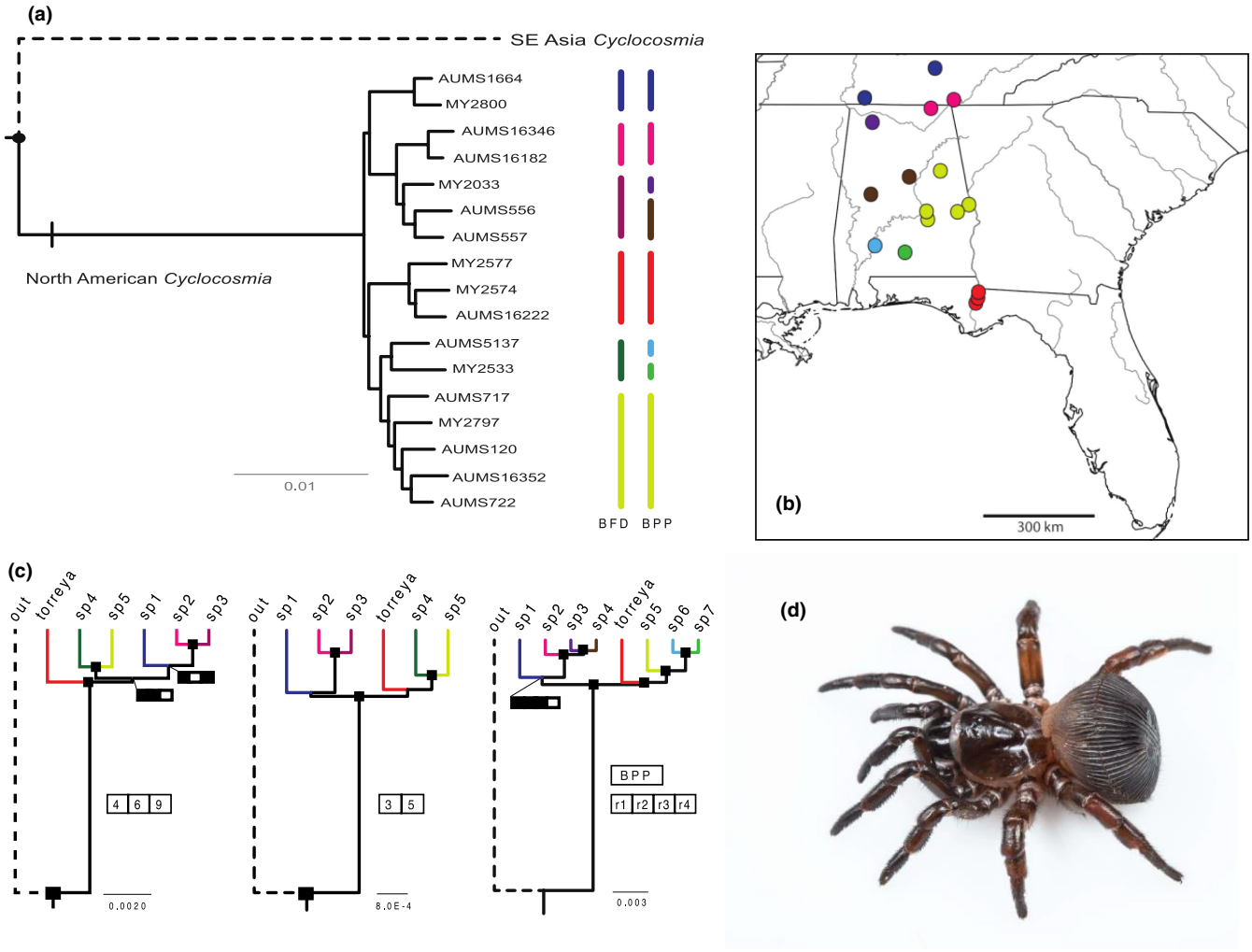

FIGURE 4 Delimitation results obtained in BFD and BPP analyses. (a) Species delimitation results mapped on the topology obtained in the concatenated analyses; left: results of species discovery step, performed on 12 sets of 50 loci in BFD; right: results of the validation step via *gdi* based on estimates obtained in the BPP analyses; colour coding corresponds to the geographic sampling location of the delimited species (b). (c) Left to right: two most commonly inferred topologies in *Beast analyses for the BFD species delimitation *H4* (7 species), topology inferred in four independent runs of the BPP A01 analyses for the final delimitation scenario obtained in the species validation step via *gdi*. (d) Adult female of *Cyclocosmia truncata* sp 5.

Increasing number of loci is known to improve both consistency and support for the delimited species (Huang et al., 2020); however, it may also allow for detection of finer structure within the data and thus potentially lead to over splitting (Leaché et al., 2019; Sukumaran & Knowles, 2017). In case of sampled North American *Cyclocosmia*, the number of delimited species in the discovery step practically equalled the number of sampled localities (13 species vs. 14 localities). Such results are typically obtained from mitochondrial data-based single-locus species delimitations in sedentary taxa. For instance, most localities formed unique GMYC entities in trapdoor spider genera *Titanidiops* (Opatova & Arnedo 2014b) and *Ummidia* despite the fact that the spiderlings of the latter are capable of aerial dispersal via ballooning (Opatova et al., 2016). Similar results are also documented from other mygalomorph lineages (Hamilton et al., 2014), scorpions (Klesser et al., 2021) and pseudoscorpions (Ohira et al., 2018) and generally regarded as inflated in terms of the number of species it delimits. However, it is also possible that these results are a consequence of limited *Cyclocosmia* sampling. Gradual divergence among populations may be missed, creating a false perception of deeper phylogenetic breaks when sampling is sparse in terms of geographic coverage and the number of sampled individuals per locality (Chambers & Hillis, 2020). Nevertheless, taking into account the multi-lineage merging during the species validation step (see below), our results accentuate the species discovery might yield unrealistic results regardless of data type.

4.2 | Species validation step

The *gdi* metric was successfully used for species validation in a variety of animal groups such as reptiles and amphibians (Chan et al., 2020; Chan & Grismer, 2019; Leaché et al., 2020), bats (Jackson, Carstens, et al., 2017) and penguins (Mays et al., 2019)

and our results suggest that it seems to mitigate taxon over splitting also in philopatric organisms such as mygalomorph spiders. When evaluating the resulting *gdi* values, we opted for a conservative approach and only interpreted *gdi* >0.7 as significant; that is, only warranting a species status when that threshold was achieved (Jackson, Carstens, et al., 2017; Leaché et al., 2019). During the species validation process, a pair of putative species was merged on five occasions (Figure 3), resulting thus in delimitation of eight hypothesized North American *Cyclocosmia* species. The *gdi* values did not reach reciprocal significance within a pair of putative species in two instances ('sp9' vs. 'sp10' and 'sp5' vs. 'sp6+7'; Figure 3, Table 2), which likely stems from the shortcomings of *Cyclocosmia* sampling. Such situation has been reported from frogs (Leaché et al., 2019) and penguins (Mays et al., 2019) as a consequence of differences in sample sizes between the compared species pair. Similarly to the previous studies, we considered those species as independent, because merging unilaterally differing lineages may lead to excessive lumping, as revealed by our exploratory analyses (Figure S2).

When sister lineages were merged despite obtaining a *gdi* >0.7, all the North American *Cyclocosmia* lineages within our dataset, including morphologically district *C. torreya*, were merged into a single species. Under this scenario, *C. torreya* would represent 'local' geographic morphological variation. However, such explanation seems unlikely because of the differences also in *Cyclocosmia* phenology. Adult males of *C. torreya* emerge in the spring, whereas the males morphologically corresponding to *C. truncata* appear to be active in the fall (Gertsch & Platnick, 1975). Excessive lumping in this case might be caused by merging two divergent lineages into one and thus biasing subsequent population parameter and coalescent time estimations (Flouri et al., 2018).

4.3 | How many species of *Cyclocosmia* are there?

Although many unrecognized lineages have been detected among mygalomorph spiders by molecular species delimitation, they have been relatively rarely described as species (Hamilton, Hendrixson, et al., 2016). The formal description is usually hampered by conflicting results among the delimitation methods, lack of adult stages of both sexes and apprehension of potential species over splitting (Candia-Ramírez & Francke, 2021; Opatova & Arnedo, 2014b; Rix et al., 2020). The results of our discovery versus validation step show that the risk of over splitting remains in mygalomorph spiders also while using genomic data; however, it can be reduced with a proper delimitation pipeline. The correct selection of θ and τ parameters in BPP analyses is an essential step (Flouri et al., 2018). The values appropriate for 'large populations and shallow divergences' (Flouri et al., 2018; Yang, 2015) might be vastly different from those derived from the empirical data in sedentary taxa (Chan & Grismer, 2019; this study) and thus bias the results towards over splitting. Using empirical values may therefore lead to a better delimitation outcome (Chan & Grismer, 2019).

The two-step process delimited eight species within *Cyclocosmia* inhabiting the Southern Coastal Plain biodiversity hotspot, seven being morphologically homogeneous and consistent with *C. truncata* identification. Ideally, other sources of evidence would be taken into account and the validation would be performed in an integrative framework (Carstens et al., 2013; Edwards & Knowles, 2014). Cytogenetics data (Řezáč et al., 2018; Štundlová et al., 2019), geometric morphometry (Christophoryová et al., 2023; Korba et al., 2022) and species distribution modelling (Newton et al., 2020; Stockman & Bond, 2007) were used for delineating taxa boundaries in morphologically homogeneous arachnids. Unfortunately, our final delimitation hypothesis does not allow statistical validation by the latter two methods, because the majority of the delimited taxa is composed of few individuals and singletons.

Our results suggest that major rivers act as a dispersal barrier for *Cyclocosmia*. *Cyclocosmia* sp. 5, 6, 7 and *C. torreya* inhabit the region south-eastern of the Alabama river, *Cyclocosmia* sp. 1 and 2 occur north of the Tennessee river and *Cyclocosmia* sp. 3 and 4 are confined to the area in between. Large rivers were identified as dispersal barriers in other sedentary spiders such as the Iberian cork oak mygalomorph spider *Macrothele calpeiana* (Arnedo & Ferrández, 2007) and the Chinese primitively segmented trapdoor spider *Ganthela* (Xu et al., 2018), but they also affect species distributions in vertebrates (Barrow et al., 2018; Soltis et al., 2006). The rivers play a particularly important role in the Southern Coastal Plain biodiversity hotspot, their changing course and local flooding created a system of patchy habitats that helped promote diversification in the region (Noss et al., 2015). Similarly to less vagile vertebrate groups, such as frogs and toads (Barrow et al., 2018), the dynamic geologic past of the region also likely promoted diversification in *Cyclocosmia*.

From the molecular species delimitation standpoint, the delimited *Cyclocosmia* lineages fulfil the expectations of how divergent an independent species should be from its closest relatives even under the strictest approach (*gdi* >0.7). Whether the delimited lineages should be recognized and formally described as species is however an inherently subjective question, which should not be expected to be resolved by molecular data analyses (Leaché et al., 2019). In the case of *Cyclocosmia*, species over splitting of *C. truncata* cannot be entirely ruled out even after the *gdi* validation, because of the limited sampling in our study (Leaché et al., 2020). However, based on the non-monophyletic nature of the morphologically defined *C. truncata*, the genus most certainly harbours cryptic species diversity. Our results show that *Cyclocosmia* exhibits strong tendency for short-range endemism combined with morphological stasis (Hedin et al., 2019; Opatova et al., 2020; Wilson et al., 2023). In sedentary taxa, chromosomal rearrangements, likely creating reproduction barriers, may arise even across very short distances (Kotrbová et al., 2016) without being accompanied by morphological differentiation. This is particularly likely if the organisms occupy the same niche (Cerca et al., 2020). The pressure for morphological differentiation also might not be strong in organisms where the mating partner recognition is not based on morphological appearance (Adams & Tsutsui, 2020). Visual cues are involved in courtship behaviour of

spiders with a good eye sight such as jumping spiders and wolf spiders; however, most spider groups rely on chemical communication and specific vibration pattern (Mitoyen et al., 2019).

Even when considering potential over splitting of *C. truncata* caused by limited sampling and some of the geographically proximate independent lineages would rather correspond to populations; these populations would still likely show higher levels of genetic differentiation ($0.2 < gdi < 0.7$) (Jackson, Carstens, et al., 2017; Leaché et al., 2019). In both cases, habitat loss may pose a danger for *Cyclocosmia* diversity in general, either via endangering highly geographically constrained lineages of *C. truncata* or by reducing its intra-specific variability (Amos & Balmford, 2001). Denser sampling, albeit difficult, would be required in order mitigate potential taxa over splitting and ultimately reaching a conclusive decision. Obtaining enough fresh material might not be possible, given the rarity of species. Sequencing of museum specimens, enabled by novel genomic protocols (Wood et al., 2018), combined with geographically targeted sampling guided by the findings in this study, might increase the sample size and help to circumvent this issue.

4.4 | Best practices for genomic data-based species delimitation from small datasets

Limited taxon sampling may be a common impediment for studies tackling the genetic background of endangered or rare taxa (Hedin, 2015; Jacobs et al., 2018). In some cases, it can be avoided by implementation of non-lethal or non-invasive sampling approaches (Dufresnes et al., 2019; Ožana et al., 2020) or analyses of historical museum samples (Derkarabetian et al., 2022; Hedin et al., 2018; Wood et al., 2018). Our study shows that assessing population structure and species boundaries with genomic data might be possible even for small datasets, where obtaining additional samples might be difficult. The methodological pipeline outlined in this article (see Figure S4 for schematic overview) could be particularly helpful for pilot studies aimed at detecting potential cryptic diversity and deeply divergent and isolated populations to ensure preservation of unique lineages and intra-specific diversity in taxa requiring conservation management (Amos & Balmford, 2001). Our results suggest that using BPP for both species discovery and species validation (via *gdi*) is possible in a feasible computational time frame in small dataset (415 loci, 36 terminals). Unlike the BFD approach, it is not prone to bias stemming from the prior assignment of individuals into species and the need of exhaustive testing of all the delimitations scenarios. However, it is important to select appropriate (empirical) values for θ and τ parameters in BPP analyses to avoid over splitting in both steps, particularly when the taxa of interest are characterized by sedentary lifestyle and small population sizes.

5 | CONCLUSIONS

The results of our analyses suggest that the North American trapdoor spider genus *Cyclocosmia* harbours cryptic diversity. *Cyclocosmia truncata*, which comprises seven lineages, was recovered as paraphyletic with respect to *C. torreya*. The delimited lineages have narrow distributions and the major rivers in the region likely represent dispersal barriers. Unfortunately, limited sampling might be inevitable in rare and endangered taxa and some species delimitation approaches thus may not be applicable. The results from *Cyclocosmia* delimitation suggest that utilizing BPP in the species discovery step might be an appropriate option in situations where the delimitation is performed on genomic data and fewer individuals because of the comparatively low computational demand (compared to BFD) and consistency in convergence. However, the results should be validated via *gdi* in order to avoid over splitting.

AUTHOR CONTRIBUTIONS

V.O.: conceptualization, methodology, formal analysis, data curation, writing – original draft, review and editing. K.B.: sample collection, writing – original draft, review and editing. J.E.B.: conceptualization, resources, writing – review and editing.

ACKNOWLEDGEMENTS

We thank R. Godwin, N. Garrison and C. Stephen for the help with the field work. The computational analyses were completed with resources provided by the Auburn University Hopper supercomputing cluster. This project was funded by the National Science Foundation grant DEB 1937604 and the Evert and Marion Schligner Foundation to J.E.B. Additional support was obtained from the Charles University Research Centre program (UNCE 204069) to V.O.

CONFLICT OF INTEREST STATEMENT

The authors assert no conflict of interest.

DATA AVAILABILITY STATEMENT

DNA sequence alignments and raw data are available from the Dryad Digital Repository: Dryad Data Repository (DOI: 10.5061/dryad.p2ngf1vws).

ORCID

Vera Opatova https://orcid.org/0000-0002-4324-1344

REFERENCES

Aberer, A. J., Kobert, K., & Stamatakis, A. (2014). ExaBayes: Massively parallel Bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution*, *31*(10), 2553–2556.

Adams, S. A., & Tsutsui, N. D. (2020). The evolution of species recognition labels in insects. *Philosophical Transactions of the Royal Society B*, *375*(1802), 20190476.

Amos, W., & Balmford, A. (2001). When does conservation genetics matter? *Heredity*, *87*(3), 257–265.

Araujo, A. P. G., Carbayo, F., Riutort, M., & Álvarez-Presas, M. (2020). Five new pseudocryptic land planarian species of *Cratera* (Platyhelminthes: Tricladida) unveiled through integrative taxonomy. *PeerJ*, *8*, e9726.

Arnedo, M. A., & Ferrández, M. A. (2007). Mitochondrial markers reveal deep population subdivision in the European protected spider *Macrothele calpeiana* (Walckenaer, 1805) (Araneae, Hexathelidae). *Conservation Genetics*, *8*(5), 1147–1162.

Ausserer, A. (1871). Beiträge zur Kenntniss der Arachniden-Familie der Territelariae Thorell (Mygalidae Autor). *Verhandlungen der*

Kaiserlich-Königlichen Zoologisch-Botanischen Gesellschaft in Wien, *21*, 117–224.

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A., & Alekseyenko, A. V. (2012). Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, *29*(9), 2157–2167.

Barrow, L. N., Lemmon, A. R., & Lemmon, E. M. (2018). Targeted sampling and target capture: Assessing phylogeographic concordance with genome-wide data. *Systematic Biology*, *67*(6), 979–996.

Beja-Pereira, A., Oliveira, R., Alves, P. C., Schwartz, M. K., & Luikart, G. (2009). Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources*, *9*(5), 1279–1301.

Bond, J. E., Hedin, M. C., Ramirez, M. G., & Opell, B. D. (2001). Deep molecular divergence in the absence of morphological and ecological change in the Californian coastal dune endemic trapdoor spider *Aptostichus simus*. *Molecular Ecology*, *10*(4), 899–910.

Bond, J. E., Hendrixson, B. E., Hamilton, C. A., & Hedin, M. (2012). A reconsideration of the classification of the spider infraorder Mygalomorphae (Arachnida: Araneae) based on three nuclear genes and morphology. *PLoS One*, *7*(6), e38753.

Bond, J. E., & Stockman, A. K. (2008). An integrative method for delimiting cohesion species: Finding the population-species interface in a group of Californian trapdoor spiders with extreme genetic divergence and geographic structuring. *Systematic Biology*, *57*(4), 628–646.

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kuhnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Muller, N. F., Ogilvie, H. A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., … Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *15*(4), e1006650.

Camargo, A., Morando, M., Avila, L. J., & Sites, J. W., Jr. (2012). Species delimitation with ABC and other coalescent-based methods: A test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwinii* complex (Squamata: Liolaemidae). *Evolution: International Journal of Organic Evolution*, *66*(9), 2834–2849.

Čandek, K., & Kuntner, M. (2015). DNA barcoding gap: Reliable species identification over morphological and geographical scales. *Molecular Ecology Resources*, *15*(2), 268–277.

Candia-Ramírez, D. T., & Francke, O. F. (2021). Another stripe on the tiger makes no difference? Unexpected diversity in the widespread tiger tarantula *Davus pentaloris* (Araneae: Theraphosidae: Theraphosinae). *Zoological Journal of the Linnean Society*, *192*(1), 75–104.

Carstens, B. C., & Knowles, L. L. (2007). Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: An example from *Melanoplus* grasshoppers. *Systematic Biology*, *56*(3), 400–411.

Carstens, B. C., Morales, A. E., Jackson, N. D., & O'Meara, B. C. (2017). Objective choice of phylogeographic models. *Molecular Phylogenetics and Evolution*, *116*, 136–140.

Carstens, B. C., Pelletier, T. A., Reid, N. M., & Satler, J. D. (2013). How to fail at species delimitation. *Molecular Ecology*, *22*(17), 4369–4383.

Català, C., Bros, V., Castelltort, X., Santos, X., & Pascual, M. (2021). Deep genetic structure at a small spatial scale in the endangered land snail *Xerocrassa montserratensis*. *Scientific Reports*, *11*(1), 8855.

Cerca, J., Meyer, C., Stateczny, D., Siemon, D., Wegbrod, J., Purschke, G., Dimitrov, D., & Struck, T. H. (2020). Deceleration of morphological evolution in a cryptic species complex and its link to paleontological stasis. *Evolution*, *74*(1), 116–131.

Chambers, E. A., & Hillis, D. M. (2020). The multispecies coalescent over-splits species in the case of geographically widespread taxa. *Systematic Biology*, *69*(1), 184–193.

Chan, K. O., & Grismer, L. L. (2019). To split or not to split? Multilocus phylogeny and molecular species delimitation of southeast Asian toads (family: Bufonidae). *BMC Evolutionary Biology*, *19*(1), 1–12.

Chan, K. O., Hutter, C. R., Wood, P. L., Jr., Grismer, L. L., Das, I., & Brown, R. M. (2020). Gene flow creates a mirage of cryptic species in a southeast Asian spotted stream frog complex. *Molecular Ecology*, *29*(20), 3970–3987.

Christophoryová, J., Krajčovičová, K., Šťáhlavský, F., Španiel, S., & Opatova, V. (2023). Integrative taxonomy approach reveals cryptic diversity within the phoretic Pseudoscorpion genus *Lamprochernes* (Pseudoscorpiones: Chernetidae). *Insects*, *14*(2), 122.

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.

Coimbra, R. T., Winter, S., Kumar, V., Koepfli, K.-P., Gooley, R. M., Dobrynin, P., Fennessy, J., & Janke, A. (2021). Whole-genome analysis of giraffe supports four distinct species. *Current Biology*, *31*(13), 2929–2938. e2925.

Cracraft, J. (1983). Species concepts and speciation analysis. *Ornithology*, *1*, 159–187.

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.

Derkarabetian, S., Paquin, P., Reddell, J., & Hedin, M. (2022). Conservation genomics of federally endangered *Texella* harvester species (Arachnida, Opiliones, Phalangodidae) from cave and karst habitats of central Texas. *Conservation Genetics*, *23*(2), 401–416.

Devitt, T. J., Wright, A. M., Cannatella, D. C., & Hillis, D. M. (2019). Species delimitation in endangered groundwater salamanders: Implications for aquifer management and biodiversity conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(7), 2624–2633.

Dufresnes, C., Remollino, N., Stoffel, C., Manz, R., Weber, J.-M., & Fumagalli, L. (2019). Two decades of non-invasive genetic monitoring of the grey wolves recolonizing the Alps support very limited dog introgression. *Scientific Reports*, *9*(1), 148.

Dupuis, J. R., Roe, A. D., & Sperling, F. A. (2012). Multi-locus species delimitation in closely related animals and fungi: One marker is not enough. *Molecular Ecology*, *21*(18), 4422–4436.

Edwards, D. L., & Knowles, L. L. (2014). Species detection and individual assignment in species delimitation: Can integrative data increase efficacy? *Proceedings of the Biological Sciences*, *281*(1777), 20132765.

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*(10), 2585–2593.

Freudenstein, J. V., Broe, M. B., Folk, R. A., & Sinn, B. T. (2017). Biodiversity and the species concept – Lineages are not enough. *Systematic Biology*, *66*(4), 644–656.

Galtier, N. (2019). Delineating species in the speciation continuum: A proposal. *Evolutionary Applications*, *12*(4), 657–663.

Gertsch, W. J., & Platnick, N. I. (1975). A revision of the trapdoor spider genus *Cyclocosmia* (Araneae, Ctenizidae). American Museum Novitates; no. 2580.

Godwin, R. L., Opatova, V., Garrison, N. L., Hamilton, C. A., & Bond, J. E. (2018). Phylogeny of a cosmopolitan family of morphologically conserved trapdoor spiders (Mygalomorphae, Ctenizidae) using Anchored Hybrid Enrichment, with a description of the family, Halonoproctidae Pocock 1901. *Molecular Phylogenetics and Evolution*, *126*, 303–313.

Grummer, J. A., Bryson, R. W., Jr., & Reeder, T. W. (2014). Species delimitation using Bayes factors: Simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Systematic Biology*, *63*(2), 119–133.

Hamilton, C. A., Formanowicz, D. R., & Bond, J. E. (2011). Species delimitation and phylogeography of *Aphonopelma hentzi* (Araneae, Mygalomorphae, Theraphosidae): Cryptic diversity in North American tarantulas. *PLoS One*, *6*(10), e26207.

Hamilton, C. A., Hendrixson, B. E., & Bond, J. E. (2016). Taxonomic revision of the tarantula genus *Aphonopelma* Pocock, 1901 (Araneae, Mygalomorphae, Theraphosidae) within the United States. *ZooKeys*, *560*, 1–340.

Hamilton, C. A., Hendrixson, B. E., Brewer, M. S., & Bond, J. E. (2014). An evaluation of sampling effects on multiple DNA barcoding methods leads to an integrative approach for delimiting species: A case study of the North American tarantula genus *Aphonopelma* (Araneae, Mygalomorphae, Theraphosidae). *Molecular Phylogenetics and Evolution*, *71*, 79–93.

Hamilton, C. A., Lemmon, A. R., Lemmon, E. M., & Bond, J. E. (2016). Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. *BMC Evolutionary Biology*, *16*(1), 212.

Hamm, C. A., Aggarwal, D., & Landis, D. A. (2010). Evaluating the impact of non-lethal DNA sampling on two butterflies, *Vanessa cardui* and *Satyrodes eurydice*. *Journal of Insect Conservation*, *14*, 11–18.

Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S., & Francis, C. M. (2004). Identification of birds through DNA barcodes. *PLoS Biology*, *2*(10), 1657–1663.

Hedin, M. (2015). High-stakes species delimitation in eyeless cave spiders (*Cicurina*, Dictynidae, Araneae) from Central Texas. *Molecular Ecology*, *24*(2), 346–361.

Hedin, M., Derkarabetian, S., Alfaro, A., Ramírez, M. J., & Bond, J. E. (2019). Phylogenomic analysis and revised classification of atypoid mygalomorph spiders (Araneae, Mygalomorphae), with notes on arachnid ultraconserved element loci. *PeerJ*, *7*, e6864.

Hedin, M., Derkarabetian, S., Blair, J., & Paquin, P. (2018). Sequence capture phylogenomics of eyeless *Cicurina* spiders from Texas caves, with emphasis on US federally-endangered species from Bexar County (Araneae, Hahniidae). *ZooKeys*, *769*, 49–76.

Heled, J., & Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, *27*(3), 570–580.

Hey, J. (2009). Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, *27*(4), 905–920.

Homma, R., Uyeno, D., & Kakui, K. (2020). Integrative taxonomy of *Pseudolepeophtheirus longicauda* (Crustacea: Copepoda: Caligidae) parasitic on *Platichthys stellatus* (Actinopterygii: Pleuronectidae). *Parasitology International*, *78*, 102135.

Hosegood, J., Humble, E., Ogden, R., de Bruyn, M., Creer, S., Stevens, G. M. W., Abudaya, M., Bassos-Hull, K., Bonfil, R., Fernando, D., Foote, A. D., Hipperson, H., Jabado, R. W., Kaden, J., Moazzam, M., Peel, L. R., Pollett, S., Ponzo, A., Poortvliet, M., … Carvalho, G. (2020). Phylogenomics and species delimitation for effective conservation of manta and devil rays. *Molecular Ecology*, *29*(24), 4783–4796.

Huang, J., Flouri, T., & Yang, Z. (2020). A simulation study to examine the information content in phylogenomic data sets under the multispecies coalescent model. *Molecular Biology and Evolution*, *37*(11), 3211–3224.

Hunt, R. H. (1975). Notes on the ecology of *Cyclocosmia truncata* (Araneae, Ctenizidae) in Georgia. *Journal of Arachnology*, *3*(2), 83–86.

Jackson, N. D., Carstens, B. C., Morales, A. E., & O'Meara, B. C. (2017). Species delimitation with gene flow. *Systematic Biology*, *66*(5), 799–812.

Jackson, N. D., Morales, A. E., Carstens, B. C., & O'Meara, B. C. (2017). PHRAPL: Phylogeographic inference using approximate likelihoods. *Systematic Biology*, *66*(6), 1045–1053.

Jacobs, S. J., Kristofferson, C., Uribe-Convers, S., Latvis, M., & Tank, D. C. (2018). Incongruence in molecular species delimitation schemes: What to do when adding more data is difficult. *Molecular Ecology*, *27*(10), 2397–2413.

Jiao, X., & Yang, Z. (2021). Defining species when there is gene flow. *Systematic Biology*, *70*(1), 108–119.

Johnson, N. A., Smith, C. H., Pfeiffer, J. M., Randklev, C. R., Williams, J. D., & Austin, J. D. (2018). Integrative taxonomy resolves taxonomic uncertainty for freshwater mussels being considered for protection under the US Endangered Species Act. *Scientific Reports*, *8*(1), 1–16.

Kass, R., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.

Kergoat, G. J., Toussaint, E. F., Capdevielle-Dulac, C., Clamens, A.-L., Ong'Amo, G., Conlong, D., van Den Berg, J., Cugala, D., Pallangyo, B., Mubenga, O., Chipabika, G., Ndemah, R., Sezolin, M., Bani, G., Molo, R., Ali, A., Calatayud, P. A., Kaiser, L., Silvain, J. F., & Le Ru, B. (2015). Integrative taxonomy reveals six new species related to the Mediterranean corn stalk borer *Sesamia nonagrioides* (Lefèbvre) (Lepidoptera, Noctuidae, Sesamiina). *Zoological Journal of the Linnean Society*, *175*(2), 244–270.

Klesser, R., Husemann, M., Schmitt, T., Sousa, P., Moussi, A., & Habel, J. C. (2021). Molecular biogeography of the Mediterranean *Buthus* species complex (Scorpiones: Buthidae) at its southern Palaearctic margin. *Biological Journal of the Linnean Society*, *133*(1), 166–178.

Korba, J., Opatova, V., Calatayud-Mascarell, A., Enguídanos, A., Bellvert, A., Adrián, S., Sánchez-Vialas, A., & Arnedo, M. A. (2022). Systematics and phylogeography of western Mediterranean tarantulas (Araneae: Theraphosidae). *Zoological Journal of the Linnean Society*, *196*(2), 845–884.

Koroiva, R., & Kvist, S. (2018). Estimating the barcoding gap in a global dataset of cox1 sequences for Odonata: Close, but no cigar. *Mitochondrial DNA Part A DNA Mapping, Sequencing, and Analysis*, *29*(5), 765–771.

Kotrbová, J., Opatova, V., Gardini, G., & Šťáhlavský, F. (2016). Karyotype diversity of pseudoscorpions of the genus *Chthonius* (Pseudoscorpiones, Chthoniidae) in the Alps. *Comparative Cytogenetics*, *10*(3), 325–345.

Kück, P. (2009). *ALICUT: A Perlscript which cuts ALISCORE identified RSS*. Department of Bioinformatics, Zoologisches Forschungsmuseum A. Koenig (ZFMK), Bonn, Germany, Version, 2.

Kück, P., & Longo, G. C. (2014). FASconCAT-G: Extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers in Zoology*, *11*(1), 81.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874.

Kvist, S. (2016). Does a global DNA barcoding gap exist in Annelida? *Mitochondrial DNA Part A DNA Mapping, Sequencing, and Analysis*, *27*(3), 2241–2252.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C., & Stamatakis, A. (2014). Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology*, *14*(1), 82.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2016). PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, *34*(3), 772–773.

Leache, A. D., Harris, R. B., Rannala, B., & Yang, Z. (2014). The influence of gene flow on species tree estimation: A simulation study. *Systematic Biology*, *63*(1), 17–30.

Leaché, A. D., Oaks, J. R., Ofori-Boateng, C., & Fujita, M. K. (2020). Comparative phylogeography of West African amphibians and reptiles. *Evolution*, *74*(4), 716–724.

Leaché, A. D., Zhu, T., Rannala, B., & Yang, Z. (2019). The spectre of too many species. *Systematic Biology*, *68*(1), 168–181.

Leavitt, D. H., Starrett, J., Westphal, M. F., & Hedin, M. (2015). Multilocus sequence data reveal dozens of putative cryptic species in a

radiation of endemic Californian mygalomorph spiders (Araneae, Mygalomorphae, Nemesiidae). *Molecular Phylogenetics and Evolution*, *91*, 56–67.

Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, *61*(5), 727–744.

Lemmon, A. R., & Lemmon, E. M. (2012). High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. *Systematic Biology*, *61*(5), 745–761.

Machkour M'Rabet, S., Hènaut, Y., Dor, A., Pèrez-Lachaud, G., Pèlissier, C., Gers, C., & Legal, L. (2009). ISSR (inter simple sequence repeats) as molecular markers to study genetic diversity in tarantulas (Araneae, Mygalomorphae). *Journal of Arachnology*, *37*(1), 10–14.

Magoga, G., Fontaneto, D., & Montagna, M. (2021). Factors affecting the efficiency of molecular species delimitation in a species-rich insect family. *Molecular Ecology Resources*, *21*(5), 1475–1489.

Mallet, J. (2008). Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society of London B*, *363*, 2971–2986.

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., & Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, *23*(11), 1817–1828.

Mays, H. L., Jr., Oehler, D. A., Morrison, K. W., Morales, A. E., Lycans, A., Perdue, J., Battley, P. F., Cherel, Y., Chilvers, B. L., Crofts, S., Demongin, L., Fry, W. R., Hiscock, J., Kusch, A., Marin, M., Poisbleau, M., Quillfeldt, P., Rey, A. R., Steinfurth, A., Thompson, D. R., & Weakley, L. A. (2019). Phylogeography, population structure, and species delimitation in rockhopper penguins (*Eudyptes chrysocome* and *Eudyptes moseleyi*). *Journal of Heredity*, *110*(7), 801–817.

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc*, *2010*(6), pdb.prot5448.

Misof, B., & Misof, K. (2009). A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: A more objective means of data exclusion. *Systematic Biology*, *58*(1), 21–34.

Mitoyen, C., Quigley, C., & Fusani, L. (2019). Evolution and function of multimodal courtship displays. *Ethology*, *125*(8), 503–515.

Montes de Oca, L., D'Elía, G., & Pérez-Miles, F. (2016). An integrative approach for species delimitation in the spider genus *Grammostola* (Theraphosidae, Mygalomorphae). *Zoologica Scripta*, *45*(3), 322–333.

Morales, A. E., Jackson, N. D., Dewey, T. A., O'Meara, B. C., & Carstens, B. C. (2017). Speciation with gene flow in North American *Myotis* bats. *Systematic Biology*, *66*(3), 440–452.

Newton, L. G., Starrett, J., Hendrixson, B. E., Derkarabetian, S., & Bond, J. E. (2020). Integrative species delimitation reveals cryptic diversity in the southern Appalachian *Antrodiaetus unicolor* (Araneae: Antrodiaetidae) species complex. *Molecular Ecology*, *29*(12), 2269–2287.

Nosil, P. (2008). *Speciation with gene flow could be common*. Wiley Online Library.

Noss, R. F., Platt, W. J., Sorrie, B. A., Weakley, A. S., Means, D. B., Costanza, J., & Peet, R. K. (2015). How global biodiversity hotspots may go unrecognized: Lessons from the North American Coastal Plain. *Diversity and Distributions*, *21*(2), 236–244.

Oaks, J., Cobb, K., Minin, V., & Leaché, A. (2019). Marginal likelihoods in phylogenetics: A review of methods and applications. *Systematic Biology*, *68*(5), 681–697.

Ohira, H., Kaneko, S., Faulks, L., & Tsutsumi, T. (2018). Unexpected species diversity within Japanese *Mundochthonius* pseudoscorpions (Pseudoscorpiones: Chthoniidae) and the necessity for improved species diagnosis revealed by molecular and morphological examination. *Invertebrate Systematics*, *32*(2), 259–277.

O'Neill, E. M., Schwartz, R., Bullock, C. T., Williams, J. S., Shaffer, H. B., Aguilar-Miguel, X., Parra-Olea, G., & Weisrock, D. W. (2013). Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. *Molecular Ecology*, *22*(1), 111–129.

Opatova, V., & Arnedo, M. A. (2014a). From Gondwana to Europe: Inferring the origins of Mediterranean *Macrothele* spiders (Araneae: Hexathelidae) and the limits of the family Hexathelidae. *Invertebrate Systematics*, *28*(4), 361–374.

Opatova, V., & Arnedo, M. A. (2014b). Spiders on a hot volcanic roof: Colonisation pathways and phylogeography of the Canary Islands endemic trap-door spider *Titanidiops canariensis* (Araneae, Idiopidae). *PLoS One*, *9*(12), e115078.

Opatova, V., Bond, J. E., & Arnedo, M. A. (2016). Uncovering the role of the Western Mediterranean tectonics in shaping the diversity and distribution of the trap-door spider genus *Ummidia* (Araneae, Ctenizidae). *Journal of Biogeography*, *43*(10), 1955–1966.

Opatova, V., Hamilton, C. A., Hedin, M., De Oca, L. M., Král, J., & Bond, J. E. (2020). Phylogenetic systematics and evolution of the spider infraorder Mygalomorphae using genomic scale data. *Systematic Biology*, *69*(4), 671–707.

Ožana, S., Pyszko, P., & Dolný, A. (2020). Determination of suitable insect part for non-lethal DNA sampling: Case study of DNA quality and regeneration capability of dragonflies. *Insect Conservation and Diversity*, *13*(4), 319–327.

Papadopulos, A. S., Baker, W. J., Crayn, D., Butlin, R. K., Kynast, R. G., Hutton, I., & Savolainen, V. (2011). Speciation with gene flow on Lord Howe Island. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(32), 13188–13193.

Pentinsaari, M., Vos, R., & Mutanen, M. (2017). Algorithmic single-locus species delimitation: Effects of sampling effort, variation and nonmonophyly in four methods and 1870 species of beetles. *Molecular Ecology Resources*, *17*(3), 393–404.

Petzold, A., & Hassanin, A. (2020). A comparative approach for species delimitation based on multiple methods of multi-locus DNA sequence analysis: A case study of the genus *Giraffa* (Mammalia, Cetartiodactyla). *PLoS One*, *15*(2), e0217956.

Pons, J. (2006). DNA-based identification of preys from non-destructive, total DNA extractions of predators using arthropod universal primers. *Molecular Ecology Notes*, *6*(3), 623–626.

Portik, D. M., & Wiens, J. J. (2021). Do alignment and trimming methods matter for phylogenomic (UCE) analyses? *Systematic Biology*, *70*(3), 440–462.

Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., & Lemmon, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, *526*(7574), 569–573.

Rambaut, A., & Drummond, A. J. (2009). *TRACER v.1.5*. In (Version 1.4). http://tree.bio.ed.ac.uk/software/tracer/.

Rannala, B., & Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, *164*(4), 1645–1656.

Řezáč, M., Arnedo, M. A., Opatova, V., Musilová, J., Řezáčová, V., & Král, J. (2018). Taxonomic revision and insights into the speciation mode of the spider *Dysdera erythrina* species-complex (Araneae: Dysderidae): Sibling species with sympatric distributions. *Invertebrate Systematics*, *32*(1), 10–54.

Rix, M. G., Wilson, J. D., & Harvey, M. S. (2020). First phylogenetic assessment and taxonomic synopsis of the open-holed trapdoor spider genus *Namea* (Mygalomorphae: Anamidae): A highly diverse mygalomorph lineage from Australia's tropical eastern rainforests. *Invertebrate Systematics*, *34*(7), 679–726.

Satler, J. D., Carstens, B. C., & Hedin, M. (2013). Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). *Systematic Biology*, *62*(6), 805–823.

Smadja, C. M., & Butlin, R. K. (2011). A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, *20*(24), 5123–5140.

Soltis, D. E., Morris, A. B., McLachlan, J. S., Manos, P. S., & Soltis, P. S. (2006). Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, *15*(14), 4261–4293.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313.

Stankowski, S., & Ravinet, M. (2021). Defining the speciation continuum. *Evolution*, *75*(6), 1256–1273.

Stockman, A., & Bond, J. E. (2007). Delimiting cohesion species: Extreme population structuring and the role of ecological interchangeability. *Molecular Ecology*, *16*(16), 3374–3392.

Štundlová, J., Šmíd, J., Nguyen, P., & Šťáhlavský, F. (2019). Cryptic diversity and dynamic chromosome evolution in Alpine scorpions (Euscorpiidae: *Euscorpius*). *Molecular Phylogenetics and Evolution*, *134*, 152–163.

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*(7), 1607–1612.

Talavera, G., Dinca, V., & Vila, R. (2013). Factors affecting species delimitations with the GMYC model: Insights from a butterfly survey. *Methods in Ecology and Evolution*, *4*(12), 1101–1110.

Waits, L. P., & Paetkau, D. (2005). Noninvasive genetic sampling tools for wildlife biologists: A review of applications and recommendations for accurate data collection. *The Journal of Wildlife Management*, *69*(4), 1419–1433.

Wang, X., He, Z., Shi, S., & Wu, C.-I. (2020). Genes and speciation: Is it time to abandon the biological species concept? *National Science Review*, *7*(8), 1387–1397.

Wilson, J. D., Bond, J. E., Harvey, M. S., Ramírez, M. J., & Rix, M. G. (2023). Correlation with a limited set of behavioral niches explains the convergence of somatic morphology in mygalomorph spiders. *Ecology and Evolution*, *13*(1), e9706.

Wood, H. M., González, V. L., Lloyd, M., Coddington, J., & Scharff, N. (2018). Next-generation museum genomics: Phylogenetic relationships among palpimanoid spiders using sequence capture techniques (Araneae: Palpimanoidea). *Molecular Phylogenetics and Evolution*, *127*, 907–918.

World Spider Catalog (WSC). (2023). *World Spider Catalog*. Version 24.5. Natural History Museum Bern. http://wsc.nmbe.ch/ https://doi.org/10.24436/2

Xie, W., Lewis, P. O., Fan, Y., Kuo, L., & Chen, M.-H. (2011). Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology*, *60*(2), 150–160.

Xu, X., Kuntner, M., Liu, F., Chen, J., & Li, D. (2018). Formation of rivers and mountains drives diversification of primitively segmented spiders in continental East Asia. *Journal of Biogeography*, *45*(9), 2080–2091.

Xu, X., Xu, C., Li, F., Pham, D. S., & Li, D. (2017). Trapdoor spiders of the genus *Cyclocosmia* Ausserer, 1871 from China and Vietnam (Araneae, Ctenizidae). *ZooKeys*, *643*, 75–85.

Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. *Current Zoology*, *61*(5), 854–865.

Yang, Z., & Rannala, B. (2014). Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution*, *31*(12), 3125–3135.

Zhang, C., Zhang, D.-X., Zhu, T., & Yang, Z. (2011). Evaluation of a Bayesian coalescent method of species delimitation. *Systematic Biology*, *60*(6), 747–761.

Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, *29*(22), 2869–2876.

Zhu, M., Zhang, J., & Zhang, F. (2006). Rare spiders of the genus *Cyclocosmia* (Arachnida: Araneae: Ctenizidae) from tropical and subtropical China. *Raffles Bulletin of Zoology*, *54*, 119–124.

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: Opatova, V., Bourguignon, K., & Bond, J. E. (2023). Species delimitation with limited sampling: An example from rare trapdoor spider genus *Cyclocosmia* (Mygalomorphae, Halonoproctidae). *Molecular Ecology Resources*, *00*, 1–16. https://doi.org/10.1111/1755-0998.13894