Received: 18 December 2023 | Revised: 23 January 2024 | Accepted: 5 February 2024

DOI: 10.1111/1755-0998.13938

MOLECULAR ECOLOGY RESOURCES WILEY

PERSPECTIVE

Delimiting the rare, endangered and actively speciating

Kristen M. Martinet | **Luke J. Harmon**

Department of Biological Sciences, University of Idaho, Moscow, Idaho, USA

Correspondence

Luke J. Harmon, Department of Biological Sciences, University of Idaho, 875 Perimeter Drive, Moscow, ID 83844, USA.
Email: lukeh@uidaho.edu

Handling Editor: Joanna Kelley

Species delimitation is a contentious topic. The genomics revolution initially brought hope that identifying and classifying species would be easier through better methods and more data, but genomics has also brought complexity and controversy to delimitation. One solution can be to collect a larger sample of individuals at a finer geographic scale. But what if taxa are rare and collecting more samples is difficult or detrimental to the organisms at hand? In this issue of *Molecular Ecology Resources*, Opatova et al. (2023) tackle the ambiguity of species delimitation in rare and endangered trapdoor spiders (genus *Cyclocosmia*). The authors propose a framework for delimiting species when samples are hard to come by, such as in these rare and cryptic spiders. The authors combine extensive genomic sampling with statistical approaches that consider both the genetic distinctiveness of each population of spiders and how much gene flow occurs between these populations. Their proposed taxonomy balances two opposing signals, structure and gene flow, to count eight lineages of *Cyclocosmia*, and to point the way for future taxonomic studies of the rare or difficult to obtain.

KEYWORDS
genomics/proteomics, speciation, species delimitation, spiders

It once seemed possible that speciation was a rare occurrence. Perhaps, we thought, most organisms on the Earth were a part of a distinct species and could be neatly tucked into this or that box. The genomics revolution has shattered this possibility. It is now clear that most, if not all, species include substantial variation from one place to the next, and that variation frequently blurs the boundary between one species and two (De Queiroz, 1998). This discovery has elevated species delimitation—where scientists use statistical approaches to identify species and classify individuals within them—to the forefront of modern biodiversity science. However, species delimitation is a contentious topic (Carstens et al., 2013). Species concepts are hard-fought, methodological innovations are rapid but tax the capabilities of our computers, and biological discoveries—especially those concerning the frequency of gene flow across porous species—continually disrupt the field. Furthermore, results can be incredibly important, both economically and socially, deepening the taxonomic drama (Dufresnes et al., 2023). One common call is for "more data," with the hope that additional sampling can resolve this issue. However, for the rare species that we might be most passionate about, such as the trapdoor spider genus *Cyclocosmia* considered by Opatova et al. in this issue, sample sizes are inherently limited. These trapdoor spiders have a large abdominal disk that they can use to plug the hole of their burrow to hide from predators. These spiders are also extremely rare, and every collected individual reduces the precariously small populations of these spectacular animals.

Using minimal field sampling, Opatova et al. instead inputs extensive genomic sampling along with empirically informed priors into a statistical approach to species delimitation. They start with a sophisticated clustering algorithm, Bayesian Phylogenetics and Phylogeography (BPP; Yang, 2015). BPP identifies how individual spiders are sorted into putative species. Although there are a range of computational methods for delimiting species, BPP may be the most commonly used today. One reason for this is that the results of BPP are less influenced by the effects of confounding factors like geography and migration than competing methods (Zhang et al., 2014). Under the right conditions, the method also demonstrates minimal overestimation and underestimation of

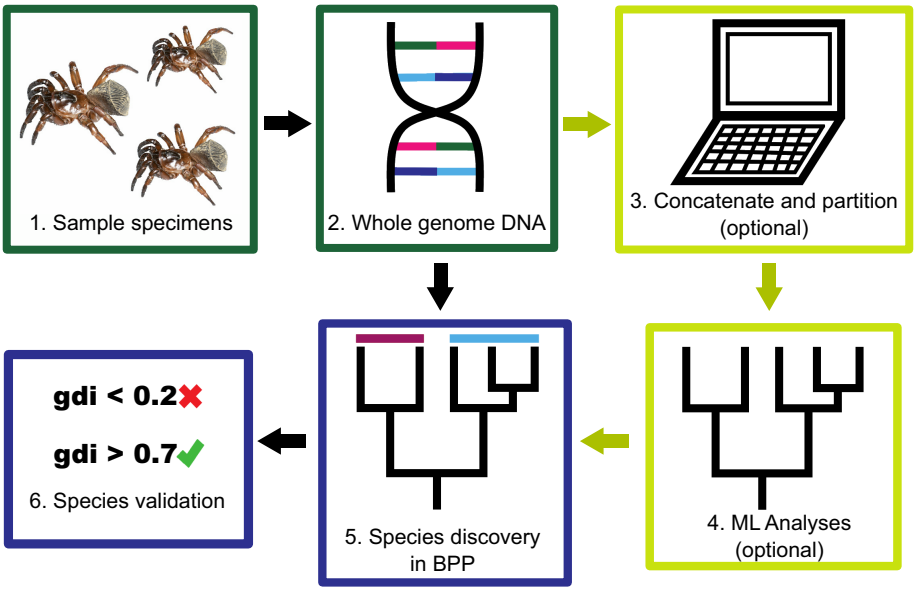

FIGURE 1 A simplified visualisation of the framework used by Opatova et al. for delimiting *Cyclocosmia* species. In Step 1, specimens are sampled from the environment. In Step 2, whole genome DNA is extracted from samples (see section 2.1 in Opatova et al., 2023 for specific sequencing details for *Cyclocosmia*). The loci from Step 2, if aligned, may be directly used with BPP for the species discovery step (Step 5), but optional Steps 3 and 4 help make the BPP analyses run faster. In Step 3, sequences are aligned, alignment quality is assessed in ALISCORE, ambiguously aligned positions are removed with ALICUT, and finally the remaining loci are concatenated with FASconCAT and partitioned in PartitionFinder 2. In Step 4, maximum likelihood (ML) analyses are conducted in RAxML to generate ML trees using the partition scheme with the most support from Step 3. In Step 5, the full set of loci from Step 2 is input into BPP for use in a Bayesian framework delimitation approach called joint species delimitation and species-tree estimation (A11). The A11 analysis is conducted with rjMCMC in BPP. Finally, in Step 6, the genealogical divergence index is used to verify the final delimitation model given for the dataset containing all loci. If the *gdi* value is less than 0.2, the two lineages have intraspecific diversity. If the *gdi* value is greater than 0.7, the two lineages have species-level divergence. Any value in between is too ambiguous to make a claim. Photo of *Cyclocosmia truncata* taken by Jason Bond and uploaded on Wikimedia Commons under the Creative Commons Attribution 3.0 Unported licence.

species counts (Luo et al., 2018). However, BPP may be susceptible to oversplitting, where species are too finely divided based on relatively minor genomic variation across space. This is particularly likely when there is gene flow, since BPP assumes that populations, once split, have evolved completely independently from one another.

To guard against oversplitting in the face of gene flow, Opatova et al. use a heuristic approach called the genealogical divergence index (*gdi*). The *gdi* combines information on both divergence and gene flow (Jackson et al., 2017), making it a perfect partner for delimitation analyses using BPP (Leaché et al., 2019). The *gdi* considers the probability that any two alleles within one putative species are each others' closest evolutionary relative compared to an allele from another putative species. Imagine that we select a gene at random from the genome, and within that gene we consider three different alleles: two within one focal putative species and one from another putative species. The *gdi* represents the probability that the two alleles from the focal species are the most closely related of the three. This property, called monophyly, shares a close connection to traditional species delimitation methods—and allows the investigators to collapse potential species that are only weakly divergent from one another. If the probability represented by the *gdi* is high, most of the variation within the focal species is monophyletic. While the *gdi* alone is not useful for delimiting species, it allows researchers to explore the accuracy of delimitation models when gene flow is present (Jackson et al., 2017). Opatova et al. use both BPP and *gdi* together in a workflow to determine the number of lineages present in their dataset (Figure 1)—which turns out to be eight.

Previous research has suggested that combining a model-based method like BPP with the *gdi* is a reliable species delimitation method in datasets with a large number of individuals (Leaché et al., 2019). Opatova et al. demonstrate that this method can be used reliably with datasets that include a small number of individuals as well, opening up new possibilities for identifying species when sampling is limited. Trapdoor spiders serve as an ideal test system for the robustness of this framework because of their intriguing combination of morphological similarity and deep molecular divergence (Bond et al., 2001). This study also hints towards the future of taxonomy in a speciating world, where all species—even the cryptic and rare—are made up of a complex web of populations that span the spectrum between one species and many (Rosenblum et al., 2012). Adding additional biological processes like gene flow to our species delimitation models may help us more accurately describe this spectrum of speciation.

CONFLICT OF INTEREST STATEMENT

The authors assert no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Luke J. Harmon https://orcid.org/0000-0002-4985-5750

REFERENCES

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*, 899–910.

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

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

Dufresnes, C., Poyarkov, N., & Jablonski, D. (2023). Acknowledging more biodiversity without more species. *Proceedings of the National Academy of Sciences of the United States of America*, *120*, e2302424120.

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

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

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

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*, 24, e13894. https://doi.org/10.1111/1755-0998.13894

Rosenblum, E. B., Sarver, B. A., Brown, J. W., Des Roches, S., Hardwick, K. M., Hether, T. D., Eastman, J. M., Pennell, M. W., & Harmon, L. J. (2012). Goldilocks meets Santa Rosalia: An ephemeral speciation model explains patterns of diversification across time scales. *Evolutionary Biology*, *39*, 255–261.

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

Zhang, C., Rannala, B., & Yang, Z. (2014). Bayesian species delimitation can Be robust to guide-tree inference errors. *Systematic Biology*, *63*, 993–1004.

How to cite this article: Martinet, K. M., & Harmon, L. J. (2024). Delimiting the rare, endangered and actively speciating. *Molecular Ecology Resources*, *24*, e13938. https://doi.org/10.1111/1755-0998.13938