Molecular Ecology

WILEY

MOLECULAR ECOLOGY

OPINION OPEN ACCESS

Delimiting Species—Prospects and Challenges for DNA Barcoding

Brent C. Emerson

Island Ecology and Evolution Research Group, Institute of Natural Products and Agrobiology (IPNA-CSIC), Santa Cruz de Tenerife, Spain

Correspondence: Brent C. Emerson (bemerson@ipna.csic.es)

Received: 2 December 2024 | **Revised:** 3 January 2025 | **Accepted:** 23 January 2025

Handling Editor: Loren Rieseberg

Funding: This work was also supported by Biodiversity Genomics Europe (Grant no. 101059492), funded by Horizon Europe under the Biodiversity, Circular Economy and Environment call (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 22.00173 and 24.00054; and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. This work was supported by project PID2020-116788GB-I00, funded by MCIN/AEI/10.13039/501100011033, Spain.

Keywords: arthropod | DNA Barcoding | insect | taxonomy

ABSTRACT

Discovering, describing and cataloguing global species diversity remains a fundamental challenge both for biodiversity research and for the management and conservation of biodiversity. Among animals, the challenge is particularly acute within the arthropods, which comprise approximately 85% of all described animals, with approximately 1 million described species. The true number of arthropod species is estimated to be in excess of 10 million species. This estimate is likely to be revised upward in the light of global DNA barcode sequencing initiatives that are cataloguing unprecedented levels of cryptic or overlooked diversity. The scale of diversity that is being recovered with barcode sequencing places further strain on a taxonomic system confronted by ever-limited global taxonomic capacity to verify and describe new species. It is predicted that the number of novel operational taxonomic units delimited by barcode sequencing is likely to eclipse the number of species described by Linnean taxonomy by as early as 2029. Unless addressed, this may see an increasing proportion of arthropod species falling outside of protective legislative frameworks as a consequence of their lack of formal description. Confronted with this challenge, there is increasing, but controversial, acceptance of species delimitation and species description based on barcode sequence clustering thresholds. In response to the evolving controversy surrounding this issue, it is both timely and important to identify and clarify prospects and challenges for DNA barcoding, with a specific focus on species delimitation to address important shortfalls and impediments in biodiversity research.

1 | Introduction

Species are the fundamental unit for measuring, quantifying and comparing biodiversity. At the same time, defining what a species is remains an ongoing source of discussion and debate (Coyne and Orr 2004). Because of this debate, it is important that, within any discipline of science where species are being delimited, a working definition is employed. The classical typological definition of species, rooted in the development of the taxonomic nomenclatural system of Linnaeus, serves as a framework within which falsifiable species hypotheses can be further tested and evaluated. Species for which further evaluation reveals them to be indistinct from each other may be merged (synonymised). Alternatively, new species may be erected from within existing species when distinct subgroups are found. Centuries of morphology-based species descriptions from the classical typological taxonomic approach are the bedrock of biodiversity research, which in turn may provide

new data that lead to taxonomic revision. Within this context, molecular data, in particular DNA barcode sequencing (the use of short standardised sequences that differ between species), are revealing underappreciated levels of diversity that one may typically associate with species-level divergences (Hebert et al. 2003). For many taxonomic groups, this is rewriting our understanding of their true species richness and diversity (e.g., Cicconardi, Fanciulli, and Emerson 2013; Fouquet et al. 2007; Hebert et al. 2004; Young et al. 2019). We are now generating DNA barcode data and identifying what are likely to be new species at a rate that far exceeds our ability to describe those species in the classical way. Approaches based on DNA barcode sequences are thus providing a path forward to address, at scale, what has become known as the Linnaean shortfall—the discrepancy between formally described species and the number of species that actually exist (Brown and Lomolino 1998). However, this increased rate of discovery is a double-edged sword, as it greatly exceeds global taxonomic capacity to verify and describe new species. This phenomenon is referred to as the 'taxonomic impediment', a term that can be traced back to the Convention on Biological Diversity signed at the Rio de Janeiro Earth Summit in 1992 (Khuroo et al. 2007).

The taxonomic impediment has important legal and administrative implications. Conservation laws such as the US Endangered Species Act and the European Union Habitat Directive rely on the formal classification of species for protection. Thus, if species are not formally described, they may not be recognised within such legal frameworks. At the global level, treaties and conventions often require formal species description. Without formal description, species will remain invisible within the legislative framework of the Convention on Biological Diversity (CBD) and will unlikely be recognised in biodiversity monitoring or conservation efforts that are under international law. Many national and international regulations concerning invasive species, such as the US Lacey Act and the European Union Invasive Alien Species Regulation, are based on taxonomic identification. If an invasive species is not formally described, it may be difficult to apply existing laws and regulations designed to target such species. Thus, as the scale of species discovery from barcode sequencing initiatives increases, so too does the need for their formal species description, with the suggestion that this can be efficiently integrated within barcoding initiatives (Meierotto et al. 2019; Sharkey et al. 2021).

Twenty years ago, when the practise of DNA barcoding was in its infancy, Moritz and Cicero (2004) put forward an assessment of the promise and pitfalls of DNA barcoding as a proposed tool for (i) assigning unknown individuals to species and (ii) enhancing the discovery of new species (Hebert et al. 2003; Stoeckle 2003). They drew attention to criticisms that single gene sequences should be the primary identifier for species and pointed out that the DNA barcoding community was moving toward embedding any large-scale sequence database within the existing systematics framework, including voucher specimens and the integration of molecular with morphological characters. Moritz and Cicero (2004) also drew attention to the contention surrounding DNA-led species discovery and the use of some level of mtDNA divergence as a yardstick for species boundaries.

Much change has occurred since the critique of Moritz and Cicero (2004). The Barcode of Life Data System (BOLD) was launched in 2005, and now provides a data storage and analysis platform that is a global hub for diverse research projects that use or generate barcode sequence data. Barcode sequencing itself is transitioning from classical Sanger sequencing technology to high-throughput sequencing platforms (e.g., Hebert et al. 2025; Srivathsan et al. 2024, 2019). This can reduce both logistical and financial limitations for generating barcode sequence data, and is expected to see a step change in data generation as a function of more data being generated per researcher and more researchers generating data. It is thus both timely and important to consider the prospects and challenges for DNA barcoding, particularly in the context of the Linnaean shortfall and the taxonomic impediment.

1.1 | DNA Barcoding, the Barcode of Life Data System, and the Era of High-Throughput Sequencing

DNA barcoding is a tool for specimen identification using short sequences of DNA that can be broadly amplified across a taxonomic group of interest and that will typically be expected to be different between species. In the case that species can be distinguished from each other by a barcode sequence and that reference barcode sequences are available for those species, then unknown individuals can be identified by sequencing their DNA barcode. BOLD is a cloud-based data storage and analysis platform developed at the Centre for Biodiversity Genomics in Canada and devoted to DNA barcoding. BOLD accepts sequences from more than 150 genetic markers from animals, plants, fungi and protists. Among these, the most represented group is the Arthropoda, with more than 16,000,000 barcode sequences representing more than 376,000 described species and many other unknown, unidentified or undescribed species. For simplicity, I focus attention on the Arthropoda and the 658 base pair region of the cytochrome oxidase I gene (COI) of the mitochondrial DNA (mtDNA) genome. This region is commonly referred to as the 'Folmer region' (Folmer et al. 1994) and is almost universally employed across animals.

The emergence of DNA barcoding as a standardisable genetic tool for the taxonomic assignment of unknown individuals to a reference sequence library has been followed by massive uptake and implementation. A seminal paper by Hebert et al. (2003) alone has been cited more than 10,500 times (Scopus). Although the logistics and cost of DNA sequencing were, until recently, limiting factors for sequencing at scale, this is no longer the case. High-throughput sequencing platforms, such as those of Oxford Nanopore Technologies and Pacific Biosciences, reduce the time needed for post-sequencing quality control through massively parallelised read processing (e.g., Runnel et al. 2022; Srivathsan et al. 2019). In the case of Oxford Nanopore Technologies, the traditional barrier to in-house sequencing has also been broken, with their affordable MinION sequencing platform, for which the cost of barcode sequencing can be reduced by up to two orders of magnitude compared to traditional Sanger sequencing (Srivathsan et al. 2019). Thus, sequencing at scale is now substantially more accessible than ever before, which is likely to see a dramatic increase in the production of barcode sequences. How barcode sequences

are collated, curated and ultimately assigned taxonomic identity is an evolving procedure, central to which is the refined single linkage (RESL) clustering algorithm within BOLD (Ratnasingham and Hebert 2013), which summarises barcode sequence variation into operational taxonomic units (OTUs).

1.2 | The Barcode Index Number (BIN) System and Species Delineation

Coalescent theory (Kingman 1982; Wakeley 2008) predicts that under a model of isolation without gene flow and in the absence of selection, both populations within species and closely related species will ultimately become reciprocally monophyletic and thus diagnosable from each other. This simple principle underpins many algorithmic approaches to infer species boundaries from DNA sequence data (e.g., Hao, Jiang, and Chen 2011; Jones, Ghoorah, and Blaxter 2011; Pons et al. 2006; Puillandre et al. 2012). Among these, the RESL algorithm within BOLD (Ratnasingham and Hebert 2013) essentially clusters sequences below a 2.2% divergence threshold into OTUs that are assigned a unique BIN. Ratnasingham and Hebert (2013) provide further detail of how the RESL algorithm works, although specific detail is lacking due to the proprietary and patented nature of the algorithm (Hebert and Ratnasingham 2016), which has received criticism (Meier et al. 2022). OTUs are a convenient estimate of independent evolutionary lineages that may be appropriate for comparative community ecological analyses, as is done in many meta-barcoding studies. However, studies that seek to conduct taxonomic revisions and describe new species are less justified in using algorithmic inferences of OTUs (Carstens et al. 2013). Indeed, it is recognised that even with multi-locus data, coalescent approaches diagnose genetic structure, and it is not possible to distinguish structure associated with population isolation from that associated with species boundaries (Sukumaran and Knowles 2017). Despite these concerns, there is an increasing acceptance of BINs as indicative of species boundaries. In describing the BIN system, Ratnasingham and Hebert (2013) suggest that when a typological species is divided into two or more BINs, this most likely involves overlooked species. Put another way, the suggestion is that BINs probably represent species. Indeed, BINs are now commonly referred to as proxies for species (e.g., deWaard et al. 2019; Pentinsaari et al. 2020; Seymour et al. 2024) and are being used as a framework for describing species (e.g., Meierotto et al. 2019; Sharkey et al. 2021). Given the emerging debate surrounding this (e.g., Ahrens et al. 2021; Engel et al. 2021; Fernandez-Triana 2022; Meier et al. 2022; Zamani et al. 2022), it is timely to evaluate the theoretical and empirical basis of the BIN system.

1.3 | Speciation: Process, Concepts and Thresholds

Views about species and speciation are highly variable (Stankowski and Ravinet 2021b), with conceptual understanding ranging from speciation as a discrete event—a boundary marking a transition from populations to species, through to speciation as a process (Stankowski and Ravinet 2021a). Despite the central importance of species delimitation for barcoding, it is often unclear which species concept different practitioners adhere to, which is a problem for the interpretation of results and decision-making (Stankowski and Ravinet 2021b). Conceptually, the BIN system aligns with the phylogenetic species concept but differs fundamentally through the implementation of a threshold. A threshold for speciation assumes that the time for speciation to complete, as measured in mtDNA mutations, is effectively the same across species. While convenient, a universal mtDNA divergence threshold for speciation does not align with speciation theory and has been subject to much debate (Collins and Cruickshank 2013).

In their critique of inferences from barcoding, Moritz and Cicero (2004) pointed out the lack of appropriate data to understand levels of mtDNA variation within and between species. In addition to this, it can be added that we still have a rather limited understanding of how mtDNA diagnosability of species is impacted by hybridisation. Mitonuclear incompatibilities should select against hybrids and limit the exchange of mtDNA variants between species (Burton and Barreto 2012). However, under certain conditions, both direct selection and indirect selection are also expected to favour mtDNA introgression between arthropod species that will lead to non-diagnosability through mtDNA paraphyly, and such paraphyly is expected to be a non-trivial phenomenon among arthropod species (Noguerales and Emerson 2024). Moritz and Cicero (2004) note that for a robust estimation of variation within species, sampling needs to be geographically representative, and appropriate data sets are now emerging that meet this criterion. An analysis of geometrid moths of the Mediterranean region found that 7% of species partitioned into multiple OTUs at a local scale, but this increased to 17% at the more inclusive Mediterranean scale (Hausmann 2011). In a geographically more inclusive analysis of 459 European Lepidoptera species, 19% of species present intraspecific divergences that exceed the BIN threshold (Dincă et al. 2021). Hendrich et al. (2015) have sequenced the barcode region for 15,948 individuals from 3514 species of beetle, sampled from Germany and neighbouring countries, whereas Astrin et al. (2016) sequenced 3339 individuals from 561 species of spider across the same region. These two datasets also provide for an assessment of the extent to which OTUs align with taxonomically described species. To conservatively exclude species with low intraspecific divergences due to limited geographic or genetic sampling, I have filtered their data to only include species with eight or more barcode sequences, sampled across at least three countries in the case of beetles or from at least three geographically disjunct regions of Germany in the case of spiders. Maximum intraspecific divergences exceed the BIN threshold for 19% of beetle species and 20% of spider species. It is worth noting that many of these species have ranges across Europe that exceed the sampling of Hendrich et al. (2015) and Astrin et al. (2016). Thus, many other species may also exceed the BIN threshold with a more geographically representative sampling.

Ratnasingham and Hebert (2013) acknowledge that typological species are more likely to be partitioned into different BINs as the geographic scale of sampling increases. However, they suggest that much of this segregation into multiple BINs is likely to represent reproductively isolated lineages that are not currently recognised as different species. This explicitly links BINs to the Biological Species Concept (BSC), providing the opportunity for direct testing when such BIN-based OTUs can be sampled in sympatry.

1.4 | Parameterising mtDNA Thresholds With Cosegregation Patterns in Sympatry

When divergent mtDNA lineages occur in allopatry, their association with biological species can only be tested with controlled crosses, which are often impractical. However, when in sympatry, their significance with regard to biological species can be indirectly evaluated with segregation patterns for morphological or niche data or directly tested with multi-locus nuclear genomic data (Figure 1). Although not explicitly presented in a sympatric context, Sharkey et al. (2021) contextualised some of the 403 BIN-based OTUs they found within the Ichneumonoidea using niche data in the form of larval hosts from rearing experiments. Emerson et al. (2017) were able to support species inferences for sympatric spider lineages using segregation patterns for vertical forest stratification and phenology. However, the richest source of co-segregating data is the nuclear genome.

It has been suggested that the tactical incorporation of one or more nuclear DNA markers in BOLD would facilitate the detection of different biological species that cluster as a single OTU because of mtDNA introgression (Ratnasingham and Hebert 2013). However, the stochasticity of the coalescent, together with different rates of coalescence for haploid uniparentally inherited markers compared to diploid bi-parentally inherited markers, will likely lead to uncertainty or even erroneous conclusions when the species split time is short relative to the effective population size. Moving beyond single loci to genotypes derived from genome-wide SNP arrays (e.g., 100s or 1000s of unlinked SNPs) would provide a statistically more robust approach of even broader utility, capable of not only (i) differentiating biological species of very recent origin but also (ii) determining the significance of divergent mtDNA lineages with regard to the BSC when they can be sampled in sympatry (Figure 1). It is important to emphasise that SNP genotyping would be a tactical tool providing genomic depth to resolve species boundaries within scenarios (i) and (ii), thus complementing barcode sequencing initiatives that can more efficiently provide individual data at scale.

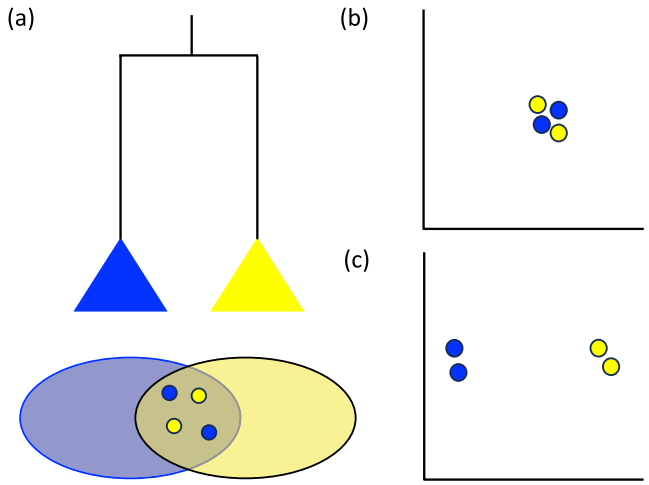

FIGURE 1 | Direct testing of association between mtDNA lineages and biological species. (a) Two mtDNA lineages are represented above graphically, with coloured triangles representing mtDNA haplotype variation within each. The corresponding geographic ranges of each lineage are represented by ellipses below, showing areas of allopatry for each species, and an area of sympatry. As few as two individuals from each mtDNA lineage in sympatry are needed to test for biological species using multi-locus nuclear data. (b) A single biological species is inferred by principal component analysis (PCA) of nuclear genomic variation from four individuals in (a), with no differentiation or cosegregation of nuclear genotypes with mtDNA variation. (c) PCA analysis reveals each mtDNA lineage to represent a different biological species through cosegregation of nuclear genotypes with mtDNA lineages.

Cost-effective techniques can now generate statistically powerful genome-wide arrays of single nucleotide polymorphisms (SNPs) within non-model organisms (e.g., Kess et al. 2016; Peterson et al. 2012), where as few as two individuals from each mtDNA lineage in sympatry may be used to test for biological species (Figure 1). Cosegregation of nuclear genotype variation with mtDNA lineages in sympatry reveals the boundaries of biological species (e.g., Pérez-Delgado et al. 2022), whereas a lack of segregation reveals no barriers to gene flow associated with mtDNA lineages (e.g., Giska, Sechi, and Babik 2015). Such analyses do not need to be at scale, nor do they require a standardised set of genomic markers across all species. The purpose is to characterise cosegregation across a sufficiently informative number of comparisons for the estimation of thresholds (Figure 2).

1.5 | Embracing the Speciation Continuum With a Dual Threshold Approach

Strategic sampling of sympatric mtDNA-derived OTUs for nuclear genotyping provides a pathway to empirically parameterise upper and lower barcode divergence thresholds, between which the uncertainty of time for speciation can be accounted for (Figure 2). In the absence of other taxonomically informative data, divergent mtDNA lineages in allopatry that fall between such lower and upper thresholds would thus be representative of the 'grey zone of speciation' (Roux et al. 2016), within which isolating barriers causing reproductive isolation (RI) are such that $0 \leq RI \leq 1$, with RI confidently assumed to be 0 below the lower threshold and 1 above the upper threshold.

After filtering the barcode data sets of Hendrich et al. (2015) and Astrin et al. (2016) as described above, Jiménez-Garcia et al. (2025) have suggested that divergences below the 99% confidence interval for average maximum intraspecific divergence within species (1.62% and 1.67% for species of beetle and spider respectively) are likely to be representative of single biological species (excluding the possibility of mtDNA introgression). Thus, restricting the analysis of sympatric mtDNA lineages to those that are divergent above these values should focus sampling within the grey zone of speciation. Alternatively, the 2.2% BIN threshold itself could be used to define the lower limits of the sampling range. However, the sampling concerns that underlie this estimate (e.g., Meier et al. 2022; Moritz and Cicero 2004) suggest that taxonomically focused and geographically representative estimates are more desirable. As such, sampling sympatric barcode lineages

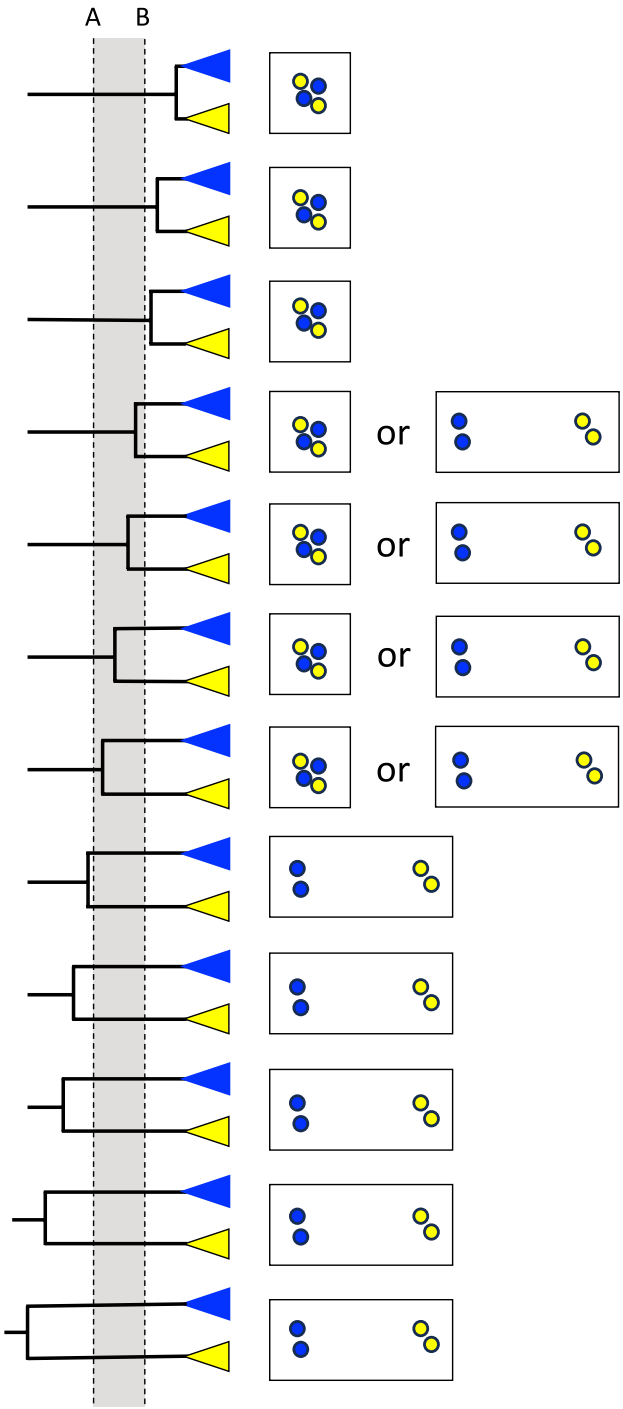

FIGURE 2 | Parameterising upper and lower barcode divergence thresholds, above which mtDNA divergences can reliably be inferred to represent different species and below which they can reliably be inferred to represent single biological species. Phylogenies represent mtDNA lineages of increasing divergence from top to bottom, within which coloured triangles represent mtDNA haplotype variation that coalesces into two monophyletic lineages. Insets beside each phylogeny represent the range of nuclear genomic segregation outcomes (see Figure 1) encountered across independent comparisons with a similar divergence. Dashed line B represents the divergence threshold below which mtDNA variation consistently corresponds to a single biological species. Dashed line A represents the divergence threshold above which mtDNA variation consistently corresponds to different biological species. The grey band between A and B represents the 'grey zone of speciation' (Roux et al. 2016).

that encompass the 2.2% BIN threshold has the potential to corroborate or not its generality.

As pointed out by Moritz and Cicero (2004), using a unique value of mtDNA divergence as a yardstick for species boundaries ignores the low precision with which coalescence of mtDNA predicts phylogenetic divergence at nuclear genes. It also ignores inherent variation in time to speciation that occurs across species. This incompatibility between the BIN threshold and species delimitation may be addressed by transitioning away from a binary classification, where individuals are either categorised as belonging to the same or different species, to a system that also incorporates classification uncertainty. Within such a system, BINs may be assigned the epithet *stet*. (Sigovini, Keppel, and Tagliapietra 2016) to reflect their uncertain species status (Figure 3), also indicating their priority for more detailed taxonomic assessment. This may ultimately provide a more informative framework for species discovery and taxonomy.

1.6 | A BOLD Framework to Tackle the Linnean Shortfall and the Taxonomic Impediment

A transversal value of BOLD is that it can democratise taxonomy. Centuries of specialist taxonomic knowledge can be digitised for global access when specimens with reliable species-level taxonomic assignment are barcode sequenced. This contributes to reducing the taxonomic impediment by providing non-specialists with a mechanism for species identification through barcode reference sequences. However, for groups that have seen only limited taxonomic progress in relation to global estimates of their species richness, barcode initiatives are likely to be sequencing species that are yet to be described, for which no barcode reference sequences exist. Indeed, most species-level diversity is concentrated in taxonomically poorly known groups that are now referred to as 'dark taxa' (Meier et al. 2022). OTUs that are taxonomically anonymous at the species level still hold great value for biodiversity measurement, monitoring and management at scale through the application of mega- and meta-barcoding approaches (e.g., Hartop et al. 2024; Noguerales et al. 2023). As such, the BIN system provides an unparalleled resource for comparability and contextualisation across independent biodiversity assessments, independent of the level of taxonomic assignment to individual BINs. However, how we then reconcile individual BINs with taxonomy has become a contentious issue.

The acceptance of BIN-based OTUs as proxies for species and thus a framework for describing species is becoming normalised (e.g., deWaard et al. 2019; Meierotto et al. 2019; Pentinsaari et al. 2020; Seymour et al. 2024; Sharkey et al. 2021). Meierotto et al. (2019) and Sharkey et al. (2021) provide compelling arguments for why developing species descriptions within a barcode sequence framework is needed for dark taxa. There are too many species, there are too few taxonomists and there is too little time. The gap between species inferences from DNA barcoding and their formal description will only widen with time if we do not come up with alternatives to the classical taxonomic approach. However, this attractively simple single threshold solution to the Linnaean shortfall and the taxonomic impediment has been met with concerns (e.g., Ahrens et al. 2021; Brower and

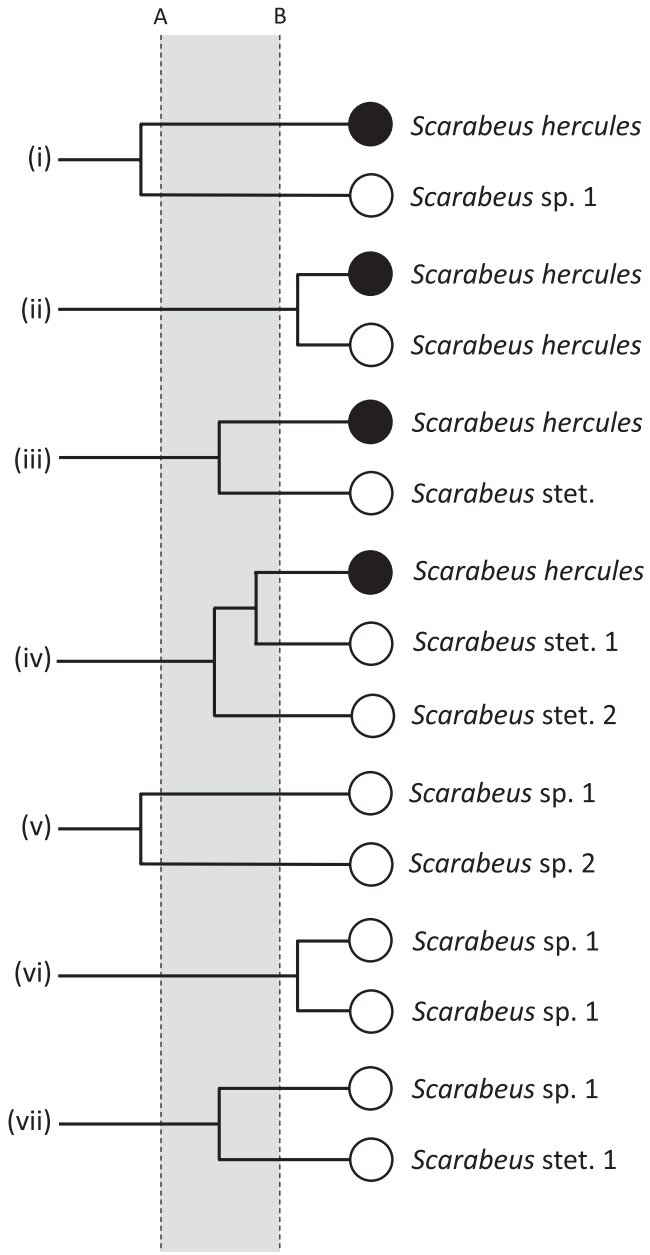

FIGURE 3 | Taxonomic classification of BINs to account for uncertain species status. A and B represent upper and lower divergence thresholds, between which the grey area represents the 'grey zone of speciation' (see Figure 2). Circles represent BINs, where filled circles represent taxonomically described species, using *Scarabaeus hercules* for illustrative purposes. Open circles represent BINs that do not correspond to existing taxonomy. (i) BIN is assigned species status. (ii) BIN is assigned to the described species. (iii) BIN assigned as *species stetit*, recognising that additional data are needed for taxonomic assignment. (iv) Each BIN is assigned as a unique *species stetit*. (v) Both BINs are assigned species status. (vi) Both BINs are assigned to the same species. (vii) One BIN is assigned species status, with the second BIN assigned as *species stetit*, whereby *Scarabaeus* sp. 1 is assigned to the first of the two BINs to be characterised.

DeSalle 2024; Engel et al. 2021; Fernandez-Triana 2022; Meier et al. 2022; Zamani et al. 2022), indicative of a new iteration of the centuries-old debate where the 'species problem' is the centre (Stankowski and Ravinet 2021b).

In their critique of Sharkey et al. (2021), two of the three concerns of Meier et al. (2022) centre around the reliability of species delimitation with a single locus and single threshold. Both concerns can potentially be mitigated by incorporating classification uncertainty, and assigning species an epithet, rather than new species-level taxonomy, to novel mtDNA lineages that are divergent from other lineages within the thresholds of speciation uncertainty (Figure 3). The third concern of Meier et al. (2022) is the description of species where the only diagnostic character is a barcode sequence (or sequences). However, in the case of truly cryptic biological species, these can only be diagnosed based on molecular data (e.g., Cicconardi, Fanciulli, and Emerson 2013; Pérez-Delgado et al. 2022). Thus, at some point, it would seem inevitable that DNA-based diagnoses of species will need to be incorporated within the taxonomic framework. What is needed is a broader discussion and agreement about how this can be best achieved. This is particularly relevant and urgent, given that the number of BINs registered in BOLD is predicted to exceed the number of Linnaean species by 2029 (P. Hebert, personal communication, October 8, 2024).

2 | Conclusions

All fields of biology are dynamic and evolve in response to new knowledge and tools, and taxonomy is no exception. Classical taxonomy and molecular approaches have come together over recent decades to give rise to the field of integrative taxonomy. However, recent initiatives that substantially shift the balance of weight toward molecular data for taxonomic decision-making are proving divisive, and there is a risk of this debate becoming further polarised. Despite differences of opinion about how best to interpret and integrate barcode sequence data into taxonomy, both sides of the debate have a common and thus unifying goal—to overcome the Linnaean shortfall and the taxonomic impediment. Progress toward this goal will need broad participation and discussion across fields such as taxonomy, evolutionary biology, speciation theory and population genetics, together with the barcoding community. Stakeholder representation from government agencies and institutions that rely upon, manage or legislate with regard to biodiversity data will also be important for a fuller contextualisation of the importance of formal description. Taxonomy is at an interesting crossroads that needs to be carefully navigated to maintain a common and unified pathway. The challenge now is to arrive at a consensus decision across different disciplines and stakeholders that both contribute to and rely upon taxonomy.

Acknowledgements

The author is grateful to Ben Price, Sean Stankowski and Mark Ravinet who provided comments on an earlier version of the manuscript. This work was supported by Biodiversity Genomics Europe (Grant no. 101059492), funded by Horizon Europe under the Biodiversity, Circular Economy and Environment call (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 22.00173 and 24.00054; and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. This work was also supported by project PID2020-116788GB-I00, funded by MCIN/AEI/10.13039/501100011033, Spain.

Conflicts of Interest

The author declares no conflicts of interest.

Data Availability Statement

There is no associated data.

References

Ahrens, D., S. T. Ahyong, A. Ballerio, et al. 2021. "Is It Time to Describe New Species Without Diagnoses?—A Comment on Sharkey Et Al. (2021)." *Zootaxa* 5027, no. 2: 151–159. https://doi.org/10.11646/zootaxa.5027.2.1.

Astrin, J. J., H. Höfer, J. Spelda, et al. 2016. "Towards a DNA Barcode Reference Database for Spiders and Harvestmen of Germany." *PLoS One* 11, no. 9: e0162624. https://doi.org/10.1371/journal.pone.0162624.

Brower, A. V. Z., and R. DeSalle. 2024. "DNA Barcodes in Taxonomic Descriptions." In *DNA Barcoding. Methods in Molecular Biology*, edited by R. DeSalle, vol. 2724, 105–115. New York, NY: Humana.

Brown, J. H., and M. V. Lomolino. 1998. *Biogeography*. Sunderland, Massachusetts: Sinauer Associates Publishers.

Burton, R. S., and F. S. A. Barreto. 2012. "A Disproportionate Role for mtDNA in Dobzhansky–Muller Incompatibilities?" *Molecular Ecology* 21: 4942–4957. https://doi.org/10.1111/mec.12006.

Carstens, B. C., T. A. Pelletier, N. M. Reid, and J. D. Satler. 2013. "How to Fail at Species Delimitation." *Molecular Ecology* 22, no. 17: 4369–4383. https://doi.org/10.1111/mec.12413.

Cicconardi, F., P. P. Fanciulli, and B. C. Emerson. 2013. "Collembola, the Biological Species Concept, and the Underestimation of Global Species Richness." *Molecular Ecology* 22: 5382–5396.

Collins, R. A., and R. H. Cruickshank. 2013. "The Seven Deadly Sins of DNA Barcoding." *Molecular Ecology Resources* 13, no. 6: 969–975. https://doi.org/10.1111/1755-0998.12046.

Coyne, J. A., and H. A. Orr. 2004. *Speciation*. Sunderland, Massachusetts: Sinauer.

deWaard, J. R., V. Levesque-Beaudin, S. L. DeWaard, et al. 2019. "Expedited Assessment of Terrestrial Arthropod Diversity by Coupling Malaise Traps With DNA Barcoding." *Genome Research* 62, no. 3: 85–95. https://doi.org/10.1139/gen-2018-0093.

Dincă, V., L. Dapporto, P. Somervuo, et al. 2021. "High Resolution DNA Barcode Library for European Butterflies Reveals Continental Patterns of Mitochondrial Genetic Diversity." *Communications Biology* 4, no. 1: 315. https://doi.org/10.1038/s42003-021-01834-7.

Emerson, B. C., J. Casquet, H. Lopez, et al. 2017. "A Combined Field Survey and Molecular Identification Protocol for Comparing Forest Arthropod Biodiversity Across Spatial Scales." *Molecular Ecology Resources* 17, no. 4: 694–707. https://doi.org/10.1111/1755-0998.12617.

Engel, M. S., L. M. P. Ceríaco, G. M. Daniel, et al. 2021. "The Taxonomic Impediment: A Shortage of Taxonomists, Not the Lack of Technical Approaches." *Zoological Journal of the Linnean Society* 193, no. 2: 381–387. https://doi.org/10.1093/zoolinnean/zlab072.

Fernandez-Triana, J. L. 2022. "Turbo Taxonomy Approaches: Lessons From the Past and Recommendations for the Future Based on the Experience With Braconidae (Hymenoptera) Parasitoid Wasps." *ZooKeys* 2022, no. 1087: 199–220. https://doi.org/10.3897/zookeys.1087.76720.

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. "DNA Primers for Amplification of Mitochondrial Cytochrome c Oxidase Subunit I From Diverse Metazoan Invertebrates." *Molecular Marine Biology and Biotechnology* 3: 294–299.

Fouquet, A., A. Gilles, M. Vences, C. Marty, M. Blanc, and N. J. Gemmell. 2007. "Underestimation of Species Richness in Neotropical Frogs Revealed by mtDNA Analyses." *PLoS One* 2, no. 10: e1109. https://doi.org/10.1371/journal.pone.0001109.

Giska, I., P. Sechi, and W. Babik. 2015. "Deeply Divergent Sympatric Mitochondrial Lineages of the Earthworm *Lumbricus rubellus* Are Not Reproductively Isolated." *BMC Evolutionary Biology* 15: 217. https://doi.org/10.1186/s12862-015-0488-9.

Hao, X., R. Jiang, and T. Chen. 2011. "Clustering 16S rRNA for OTU Prediction: A Method of Unsupervised Bayesian Clustering." *Bioinformatics* 27, no. 5: 611–618. https://doi.org/10.1093/bioinformatics/btq725.

Hartop, E., L. Lee, A. Srivathsan, et al. 2024. "Resolving biology's Dark Matter: Species Richness, Spatiotemporal Distribution, and Community Composition of a Dark Taxon." *BMC Biology* 22, no. 1: 215. https://doi.org/10.1186/s12915-024-02010-z.

Hausmann, A. 2011. "An Integrative Taxonomic Approach to Resolving Some Difficult Questions in the Larentiinae of the Mediterranean Region." *Mitteilungen der Münchner Entomologischen Gesellschaft* 101: 73–97.

Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. DeWaard. 2003. "Biological Identifications Through DNA Barcodes." *Proceedings of the Royal Society of London B* 270: 313–321.

Hebert, P. D. N., R. Floyd, S. Jafarpour, and S. W. J. Prosser. 2025. "Barcoding 100K Specimens: In a Single Nanopore Run." *Molecular Ecology Resources* 25: e14028. https://doi.org/10.1111/1755-0998.14028.

Hebert, P. D. N., E. H. Penton, J. M. Burns, D. H. Janzen, and W. Hallwachs. 2004. "Ten Species in One: DNA Barcoding Reveals Cryptic Species in the Neotropical Skipper Butterfly *Astraptes fulgerator*." *Proceedings of the National Academy of Sciences of the United States of America* 101, no. 41: 14812–14817. https://doi.org/10.1073/pnas.0406166101.

Hebert, P. D. N., and S. Ratnasingham. 2016. "US Patent No. 20,160,102,958,2016."

Hendrich, L., J. Morinière, G. Haszprunar, et al. 2015. "A Comprehensive DNA Barcode Database for Central European Beetles With a Focus on Germany: Adding More Than 3500 Identified Species to BOLD." *Molecular Ecology Resources* 15: 795–818.

Jiménez-Garcia, E., D. Suárez Ramos, C. Andújar, H. López, and B. C. Emerson. 2025. "Molecular Delimitation of Evolutionary Significant Units Reveals Hidden Geographic Drivers of Extinction Risk Within Island Arthropods." *Diversity and Distributions* 31, no. 1: e13966.

Jones, M., A. Ghoorah, and M. Blaxter. 2011. "JMOTU and Taxonerator: Turning DNA Barcode Sequences Into Annotated Operational Taxonomic Units." *PLoS One* 6, no. 4: e19259. https://doi.org/10.1371/journal.pone.0019259.

Kess, T., J. Gross, F. Harper, and E. G. Boulding. 2016. "Low-Cost ddRAD Method of SNP Discovery and Genotyping Applied to the Periwinklr *Littorina saxatilis*." *Journal of Molluscan Studies* 82, no. 1: 104–109. https://doi.org/10.1093/mollus/eyv042.

Khuroo, A. A., G. H. Dar, Z. S. Khan, and A. H. Malik. 2007. "Exploring an Inherent Interface Between Taxonomy and Biodiversity: Current Problems and Future Challenges." *Journal for Nature Conservation* 15, no. 4: 256–261. https://doi.org/10.1016/j.jnc.2007.07.003.

Kingman, J. F. C. 1982. "The Coalescent." *Stochastic Processes and their Applications* 13, no. 3: 235–248.

Meier, R., B. B. Blaimer, E. Buenaventura, et al. 2022. "A Re-Analysis of the Data in Sharkey Et Al.'s (2021) Minimalist Revision Reveals That BINs Do Not Deserve Names, but BOLD Systems Needs a Stronger Commitment to Open Science." *Cladistics* 38, no. 2: 264–275. https://doi.org/10.1111/cla.12489.

Meierotto, S., M. J. Sharkey, D. H. Janzen, et al. 2019. "A Revolutionary Protocol to Describe Understudied Hyperdiverse Taxa and Overcome the Taxonomic Impediment." *Deutsche Entomologische Zeitschrift* 66, no. 2: 119–145. https://doi.org/10.3897/dez.66.34683.

Moritz, C., and C. Cicero. 2004. "DNA Barcoding: Promise and Pitfalls." *PLoS Biology* 2, no. 10: e354. https://doi.org/10.1371/journal.pbio.0020354.

Noguerales, V., and B. C. Emerson. 2024. "Arthropod mtDNA Paraphyly: A Case Study of Introgressive Origin." *Journal of Evolutionary Biology* 10: voae152. https://doi.org/10.1093/jeb/voae152.

Noguerales, V., E. Meramveliotakis, A. Castro-Insua, et al. 2023. "Community Metabarcoding Reveals the Relative Role of Environmental Filtering and Spatial Processes in Metacommunity Dynamics of Soil Microarthropods Across a Mosaic of Montane Forests." *Molecular Ecology* 32, no. 23: 6110–6128. https://doi.org/10.1111/mec.16275.

Pentinsaari, M., G. A. Blagoev, I. D. Hogg, et al. 2020. "A DNA Barcoding Survey of an Arctic Arthropod Community: Implications for Future Monitoring." *Insects* 11, no. 1: 46. https://doi.org/10.3390/insects11010046.

Pérez-Delgado, A. J., P. Arribas, C. Hernando, et al. 2022. "Hidden Island Endemic Species and Their Implications for Cryptic Speciation Within Soil Arthropods." *Journal of Biogeography* 49, no. 7: 1367–1380. https://doi.org/10.1111/jbi.14388.

Peterson, B. K., J. N. Weber, E. H. Kay, and H. E. Hoekstra. 2012. "Double Digest RADseq: An Inexpensive Method for *de Novo* SNP Doscovery and Genotyping in Model and Non-model Species." *PLoS One* 7, no. 5: e37135. https://doi.org/10.1371/journal.pone.0037135.

Pons, J., T. G. Barraclough, J. Gomez-Zurita, et al. 2006. "Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects." *Systematic Biology* 55, no. 4: 595–609. https://doi.org/10.1080/10635150600852011.

Puillandre, N., A. Lambert, S. Brouillet, and G. Achaz. 2012. "ABGD, Automatic Barcode Gap Discovery for Primary Species Delimitation." *Molecular Ecology* 21, no. 8: 1864–1877. https://doi.org/10.1111/j.1365-294X.2011.05239.x.

Ratnasingham, S., and P. D. N. Hebert. 2013. "A DNA-Based Registry for all Animal Species: The Barcode Index Number (BIN) System." *PLoS One* 8, no. 7: e66213. https://doi.org/10.1371/journal.pone.0066213.

Roux, C., C. Fraïsse, J. Romiguier, Y. Anciaux, N. Galtier, and N. Bierne. 2016. "Shedding Light on the Grey Zone of Speciation Along a Continuum of Genomic Divergence." *PLoS Biology* 14, no. 12: e2000234. https://doi.org/10.1371/journal.pbio.2000234.

Runnel, K., K. Abarenkov, O. Copoț, et al. 2022. "DNA Barcoding of Fungal Specimens Using PacBio Long-Read High-Throughput Sequencing." *Molecular Ecology Resources* 22, no. 8: 2871–2879. https://doi.org/10.1111/1755-0998.13663.

Seymour, M., T. Roslin, J. R. deWaard, et al. 2024. "Global Arthropod Beta-Diversity Is Spatially and Temporally Structured by Latitude." *Communications Biology* 7, no. 1: 552. https://doi.org/10.1038/s42003-024-06199-1.

Sharkey, M. J., D. H. Janzen, W. Hallwachs, et al. 2021. "Minimalist Revision and Description of 403 New Species in 11 Subfamilies of Costa Rican Braconid Parasitoid Wasps, Including Host Records for 219 Species." *ZooKeys* 1013: 1–665. https://doi.org/10.3897/zookeys.1013.55600.

Sigovini, M., E. Keppel, and D. Tagliapietra. 2016. "Open Nomenclature in the Biodiversity Era." *Methods in Ecology and Evolution* 7, no. 10: 1217–1225. https://doi.org/10.1111/2041-210X.12594.

Srivathsan, A., V. Feng, D. Suárez, B. Emerson, and R. Meier. 2024. "ONTbarcoder 2.0: Rapid Species Discovery and Identification With Real-Time Barcoding Facilitated by Oxford Nanopore R10.4." *Cladistics* 40, no. 2: 192–203. https://doi.org/10.1111/cla.12566.

Srivathsan, A., E. Hartop, J. Puniamoorthy, et al. 2019. "Rapid, Large-Scale Species Discovery in Hyperdiverse Taxa Using 1D MinION Sequencing." *BMC Biology* 17: 1–20. https://doi.org/10.1186/s12915-019-0706-9.

Stankowski, S., and M. Ravinet. 2021a. "Defining the Speciation Continuum." *Evolution* 75, no. 6: 1256–1273. https://doi.org/10.1111/evo.14215.

Stankowski, S., and M. Ravinet. 2021b. "Quantifying the Use of Species Concepts." *Current Biology* 31, no. 9: R428–R429. https://doi.org/10.1016/j.cub.2021.03.060.

Stoeckle, M. 2003. "Taxonomy, DNA, and the Bar Code of Life." *BioScience* 53, no. 9: 796–797. https://doi.org/10.1641/0006-3568(2003)053[0796:TDATBC]2.0.CO;2.

Sukumaran, J., and L. L. Knowles. 2017. "Multispecies Coalescent Delimits Structure, Not Species." *Proceedings of the National Academy of Sciences of the United States of America* 114, no. 7: 1607–1611. https://doi.org/10.1073/pnas.1607921114.

Wakeley, J. 2008. *Coalescent Theory: An Introduction*. Greenwood Village, CO, United States: Roberts & Company Publishers.

Young, M. R., H. C. Proctor, J. R. deWaard, and P. D. N. Hebert. 2019. "DNA Barcodes Expose Unexpected Diversity in Canadian Mites." *Molecular Ecology* 28, no. 24: 5347–5359. https://doi.org/10.1111/mec.15292.

Zamani, A., Z. F. Fric, H. F. Gante, et al. 2022. "DNA Barcodes on Their Own Are Not Enough to Describe a Species." *Systematic Entomology* 47, no. 3: 385–389. https://doi.org/10.1111/syen.12538.