


## Forum

### Heuristics, species, and the analysis of systematic data

Tom Wells <sup>1,\*</sup>  
Tom Carruthers,<sup>2</sup> and  
Robert W. Scotland<sup>1</sup>



**Disagreements over how to define species potentially render them incomparable, yet biologists routinely count and compare species. This ‘species problem’ persists despite the wealth of data and methods available to contemporary systematists. A heuristic approach to species provides a consistent yet flexible means of selecting, assessing, and integrating different biological data.**

#### Persistent confusion stemming from the species problem

Unlike elements in chemistry, and particles in physics, biology’s best-known unit – the species – remains ill-defined and disputed. Although it continues to be used routinely by researchers from across the field of biology, there are at least 30 competing definitions – termed ‘species concepts’ – in existence [1]. Ongoing discussion [2] about the limitations of one of the most famous species definitions – the biological species concept (BSC) – highlights the continued difficulty of resolving this ‘species problem’. Its emphasis on reproductive isolation (RI) means the BSC remains notoriously difficult to test empirically and cannot be applied to asexual organisms including prokaryotes [3]. Worse, even its proponents admit that notions of RI are undermined by evidence of widespread hybridization between otherwise distinct species, as well as by the protracted and variable nature of speciation itself [2,4]. Many biologists

therefore choose to use other definitions that emphasise criteria such as mate recognition, ecological niche, or monophyly [1,5]. But incompatibility among these competing definitions means that this pluralism in relation to species concepts can lead to differing conclusions about species delimitations and counts [5,6]. This ‘species problem’ has ramifications beyond taxonomy because an indefinable unit is a potentially incomparable one. Knowing which species an organism belongs to is integral to studying and understanding both that organism and the wider world, but if different researchers are using inconsistent definitions of the term species, the comparability of their results might be brought into question [6]. Since species form the basic unit in research as diverse as understanding speciation, estimating biodiversity, modelling species distributions, and conducting conservation assessments, inconsistency in their delimitation impacts biologists from a wide range of disciplines such as systematics, evolutionary biology, ecology, and conservation [4,6,7].

#### The promise and limitations of integrative taxonomy

Attempts to overcome the species problem have distinguished theoretical ideas of what species are from operational criteria useful for identifying them [6]. By focusing on what different species share in common, this has allowed species to be treated theoretically as either ‘diverging lineages’ [5,8] or ‘cohesive clusters’ [3,9]. These opposing views might be characterised as flip sides of the same coin, in that both endeavour to provide relatively stable and comparable definitions by promoting the integration of the different types of data that represent operational criteria for identifying species [6]. If species are treated as diverging phylogenetic lineages for example [5,8], then different ‘contingent properties’ such as RI, morphological similarity, or ecological niche simply represent different criteria for identifying

those lineages [5]. RI and other contingent properties thus remain important facets of the process of speciation, but should no longer be treated as definitional of species themselves [2].

An increase in the volume, type, and precision of data and methods available to contemporary researchers for lineage identification should in theory result in more accurate and consistent species delimitation [8]. Unfortunately, treating species as lineages has instead led to excessive focus on the most empirical and readily available data and methods for assessing lineage divergence – molecular phylogenetics and genomics [4,9]. This has resulted in the neglect of other important aspects of species – in particular their morphology and ecological role – which increases the risk of mistaking temporarily isolated populations for species [4,9]. Switching the focus from lineage divergence to the cohesive forces binding individuals into species more explicitly requires the integration of different forms of data [3,9] (an approach often referred to as ‘integrative taxonomy’ [10]), but it still necessitates a decision about exactly how much evidence of cohesion is enough or too little to recognise a species [6]. Meanwhile, both lineage- and cohesion-based approaches raise questions about how to treat conflicting or incongruent data [10].

The chief challenge to integrative taxonomy continues to be the heterogeneous nature of speciation, which precedes under a variety of different processes, giving rise to the inconsistent manifestation of different ‘contingent properties’ for its recognition [2,5]. This means that incongruence in different types of systematic data is to be expected, and should not necessarily rule out species status [10]. Simply combining multiple forms of data or methods is not therefore guaranteed to resolve the species problem on its own, and deciding what properties to assess, how to do so, and

which to prioritise in the event of incongruence remains an unresolved issue.

### A heuristic framework provides specific guidelines on integrating data

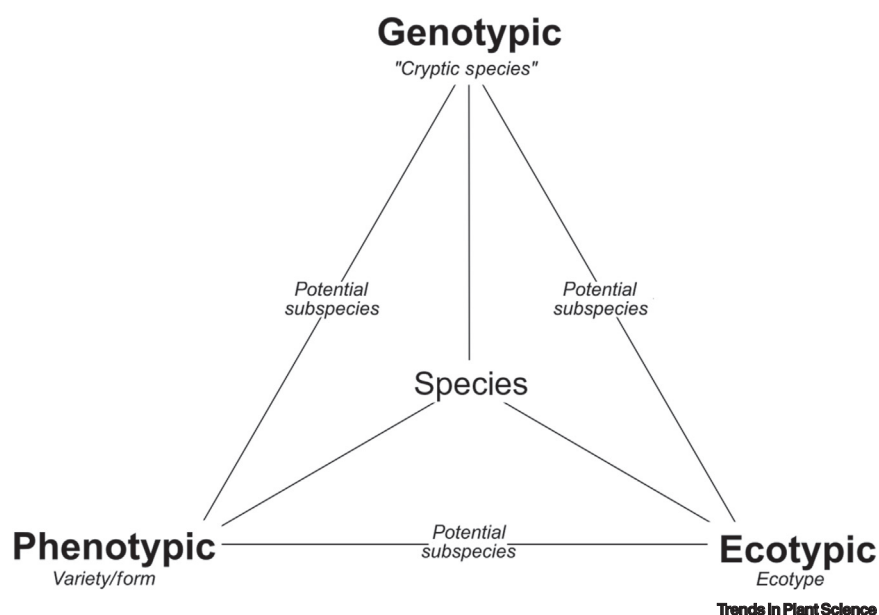
Although increasingly thought of as specific algorithms or protocols, heuristics are any methods or approaches that allow us to make difficult but important decisions despite uncertainty arising from limited methods or incomplete data, while ensuring those decisions are as accurate as possible. In this sense, Wells *et al.* [6] showed that species delimitation inherently involves the use of heuristics – for example, through the use of selected diagnostic characters – but also that species themselves are routinely used in a heuristic manner to investigate broader patterns of biodiversity and the processes underpinning them. As a result, what makes species an important unit of biological research is not simply a ‘privileged reality’ based on a unique form of evolution distinct from other taxa, as criticised by Mallet [2]. Instead, it is the utility of species in providing a means of quantifying and assessing complex biological phenomena otherwise beyond human comprehension; offering for example, a framework for investigating the mechanisms driving evolution by natural selection, as well as a quantifiable notion of biodiversity, and a means of measuring ecological change.

Consideration of this heuristic application of species as a means of understanding Earth’s biodiversity can therefore inform both operational ways of accurately and consistently delimiting species, and ideas about their theoretical nature [6]. Rather than being an impediment to the formation of a coherent species definition [10], understanding the means of identifying and delimiting species is in fact necessarily entwined with theoretical notions of what they are. The different ways in which species are used heuristically are united in

their need for species to be lineages or clusters of closely related individuals, with a shared ecological role derived from shared phenotypic adaptations [6]. As such, species should be identified by properties indicative of those three things – genotypic, phenotypic, and ecotypic cohesion (Figure 1).

When faced with a decision about species delimitation, this heuristic definition can be used to select, assess, and integrate potentially relevant data (Figure 1 and Box 1). For example, monophyly in relation to an established DNA barcode marker represents clear evidence of genotypic cohesion, consistent for instance with prolonged geographic isolation in Andean valleys [11]. It cannot however provide information about either phenotypic or ecotypic cohesion, and this is also the case for analyses based on the Multi-species Coalescent [4]. In this sense, if phylogenetic structure is incongruent with observations of morphological traits

or ecological observations, it is not on its own sufficient to overrule species hypotheses, or to delimit species [12]. Perhaps most importantly, morphological characters should not simply be treated as synonymous with evidence of phenotypic cohesion. First, because the shared possession of a morphological character can also represent evidence of the genotypic cohesion underpinning that character. And second, because many morphological characters have clearly adaptive functions, providing evidence for ecotypic cohesion [6]. A particular floral trait (evidence of phenotypic cohesion) might therefore be safely taken as an indication of common ancestry (evidence of genotypic cohesion) and a likely adaption to a specific mode of pollination (evidence of ecotypic cohesion). Similarly, distribution data could provide evidence of ecotypic cohesion if it correlates with a particular habitat, while a contiguous distribution can often reasonably be taken as evidence of genotypic cohesion. This is of course why species



**Figure 1. Species recognition ideally requires evidence of genotypic, phenotypic, and ecotypic cohesion.** The properties used in species recognition should ideally provide evidence of genotypic, phenotypic, and ecotypic cohesion. Properties indicative of one form of cohesion may indicate a variety, ecotype, or ‘cryptic’ lineage, two together could perhaps lead to recognition as a subspecies, but all three are ideally required for a species. Some properties can be simultaneously indicative of more than one type of cohesion (see Box 1 in the main text).

### Box 1. Integrating data for species recognition

Contemporary researchers often call for integrated taxonomic analysis, without necessarily providing specific guidelines for what types of data to combine, how to do so, or how to deal with discordance. A heuristic approach to species provides an explicit framework for what types of properties should be used in species recognition – those that provide evidence of genotypic, phenotypic, and ecotypic cohesion. Examples of data that might be used to infer these properties are listed in Table I. Evidence of all three forms of cohesion is required for species recognition (see also Figure 1 in the main text). Some individual properties can provide evidence of all three types of cohesion, since for example, shared possession of putatively adaptive morphological traits results from genotypic cohesion, and enables accommodation of the same ecological pressures.

Table I. Integrating data for species recognition<sup>a</sup>

Contingent Property	Example	Type of cohesion		
		Genotypic	Phenotypic	Ecotypic
Monophyly	DNA barcode	•	–	–
	Fixed morphological trait	•	•	◦
Population structure	Multispecies coalescent	•	–	–
Shared habitat	Distribution data	•	–	•
	Edaphic data	–	◦	•
	Climatic data	–	◦	•
Mutualism (e.g., pollinator)	Floral structure/reward	•	•	•
	Field observations	•	–	•
Adaptive trait	Character with clear function	•	•	•
Not apparently adaptive trait	No clear function	•	•	–

<sup>a</sup>Key: • indicate evidence of cohesion, ◦ indicate potential evidence of cohesion, – indicate a lack of evidence.

delimitation was perfectly feasible to those like Darwin with no knowledge or means of directly assessing the mechanisms of genetics.

What is more, different characters, data, and properties will be of greater or lesser importance depending on the taxonomic group, mode of speciation, or research priorities in question; and yet species can still remain united in possessing evidence of genotypic, phenotypic, and ecotypic cohesion. By explicitly promoting the integration of data on genotype, phenotype, and ecotype, a heuristic approach therefore avoids the risk of recognising ephemeral variability at the rank of species [4,6,9] (Figure 1), and renders species more comparable despite their inherent variability. This heuristic approach therefore frees researchers from the need for explicit thresholds that have traditionally limited the utility of concept-based approaches to species delimitation.

### Integrating genotypic, phenotypic, and ecotypic data generates more accurate and consistent species

The integrated analysis of data indicative of genotypic, phenotypic, and ecotypic cohesion within a ‘species complex’ previously treated as one of the most common tree species in Amazonia – *Protium heptaphyllum* Marchand [13] – demonstrates how this approach facilitates species delimitation while simultaneously improving our understanding of biodiversity. Damasco *et al.* [13] showed that *P. heptaphyllum* contained eight distinct lineages worthy of recognition as separate species, challenging its previous status as an example of a single ‘hyperdominant’ species. Crucially, all but two of these lineages had previously been described as varieties or subspecies, implying that their distinctiveness had been recognised in the past, but additional phylogenetic, morphological, functional trait, and spatial analyses revealed they warranted species

status. Integrated analysis that considers the heuristic nature and role of species as a tool for understanding tropical plant diversity thus leads to a more accurate recognition of the diversity of the tree flora of Amazonia, with fundamental implications for conservation priorities and interventions [13].

### Concluding remarks

A heuristic definition might seem to imply that every species delimitation requires vast and detailed data. This would inevitably impede rather than accelerate the process of discovery at a time when the slow pace of taxonomy [12,14] is in stark contrast to accelerated rates of plant extinction [15]. In fact, since the majority of the operational criteria generally treated as species concepts are all different ways of identifying the same form of lineage cohesion [3,5], most of the time simple combinations of readily diagnosable traits continue to provide sufficient evidence for species recognition [6,12] (Box 1). This enables us to make rapid and consistent definitions, while tailoring our approach to the taxa in question. If we want to accurately and efficiently describe the Earth’s plant diversity before it is too late [12,14,15], a heuristic approach to species is a good place to start.

### Acknowledgments

This work was supported by a Biotechnology and Biological Sciences Research Council (BBSRC) research grant (T001445/1 to R.S.). T.W. is also funded by the BBSRC as part of the Interdisciplinary Biosciences Doctoral Training Partnership at the University of Oxford. We would like to thank the handling editor Dr Susanne C. Brink and three anonymous reviewers for their constructive and insightful comments on earlier drafts. We would also like to thank Alex K. Monro, David J. Harris, and Alex Twyford for discussion and Laura Guillardin Calvo for the initial idea for Figure 1.

### Declaration of interests

No interests are declared.

<sup>1</sup>Department of Plant Sciences, University of Oxford, Oxford, UK  
<sup>2</sup>Royal Botanic Gardens, Kew, Richmond, UK

\*Correspondence:

tom.wells@plants.ox.ac.uk (T. Wells).

<https://doi.org/10.1016/j.tplants.2022.03.013>

© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## References

1. Zachos, F.E. (2016) An annotated list of species concepts. In *Species Concepts in Biology* (Zachos, F.E., ed.), pp. 77–96, Springer
2. Mallet, J. (2020) Alternative views of biological species: reproductively isolated units or genotypic clusters? *Natl. Sci. Rev.* 7, 1400–1401
3. Templeton, A.R. (1989) The meaning of species and speciation. In *Speciation and its Consequences* (Endler, J.A. and Otte, D., eds), pp. 3–27, Sinauer Associates
4. Sukumaran, J. and Knowles, L.L. (2017) Multispecies coalescent delimits structure, not species. *Proc. Natl. Acad. Sci. U. S. A.* 114, 1607–1611
5. de Queiroz, K. (2007) Species concepts and species delimitation. *Syst. Biol.* 56, 879–886
6. Wells, T. *et al.* (2021) Species as a heuristic: reconciling theory and practice. *Syst. Biol.* Published online October 21, 2021. <https://doi.org/10.1093/sysbio/syab087>
7. Baker, T.R. *et al.* (2017) Maximising synergy among tropical plant systematists, ecologists, and evolutionary biologists. *Trends Ecol. Evol.* 32, 258–267
8. Padial, J.M. and De la Riva, I. (2020) A paradigm shift in our view of species drives current trends in biological classification. *Biol. Rev.* 96, 731–751
9. Freudenstein, J.V. *et al.* (2017) Biodiversity and the species concept – lineages are not Enough. *Syst. Biol.* 66, 644–656
10. Padial, J.M. and De La Riva, I. (2010) A response to recent proposals for integrative taxonomy. *Biol. J. Linn. Soc.* 101, 747–756
11. Pennington, R.T. *et al.* (2010) Contrasting plant diversification histories within the Andean biodiversity hotspot. *Proc. Natl. Acad. Sci. U. S. A.* 107, 13783–13787
12. Muñoz-Rodríguez, P. *et al.* (2019) A taxonomic monograph of *Ipomoea* integrated across phylogenetic scales. *Nat. Plants* 5, 1136–1144
13. Damasco, G. *et al.* (2021) Revisiting the hyperdominance of Neotropical tree species under a taxonomic, functional and evolutionary perspective. *Sci. Rep.* 11, 1–11
14. Goodwin, Z.A. *et al.* (2020) How long does it take to discover a species? *Syst. Biodivers.* 18, 784–793
15. Humphreys, A.M. *et al.* (2019) Global dataset shows geography and life form predict modern plant extinction and rediscovery. *Nat. Ecol. Evol.* 3