

## OPINION

# The omission of critical data in the pursuit of ‘revolutionary’ methods to accelerate the description of species

ALIREZA ZAMANI<sup>1</sup>, VARPU VAHTERA<sup>1</sup>, ILARI EERIKKI SÄ  
ÄKSJÄRVI<sup>1</sup> and MARK D. SCHERZ<sup>2,3\*</sup>

<sup>1</sup>Zoological Museum, Biodiversity Unit, University of Turku, Turku, Finland, <sup>2</sup>Department of Evolutionary Biology, Zoologisches Institut, Technische Universität Braunschweig, Braunschweig, Germany and <sup>3</sup>Department of Herpetology, Zoologische Staatssammlung München (ZSM-SNSB), Munich, Germany

Recently, Meierotto *et al.* (2019) proposed a ‘revolutionary’ protocol for the description of understudied hyperdiverse taxa. The premise of their study was to champion exclusively DNA-barcode-based species descriptions (=diagnoses), which would dramatically increase the rate of description and provide a ‘human-readable record in the literature’ (unlike a Barcode Index Number, BIN; Ratnasingham & Hebert, 2013) that can later be supplemented with additional information.

Species are always delimited against already known species (Linnaeus, 1753, 1758; Mayr, 1992; ICZN, 1999; Naciri & Linder, 2015; Renner, 2016). This was also recognized by Meierotto *et al.* (2019, p. 120): ‘Requirements for the publication of new species include (...) that they be accompanied by either a description or diagnosis which can separate them from any known species with which they are likely to be confused’. However, the latter authors failed to diagnose their 15 new *Zelomorpha* Ashmead, 1900 species from 51 out of 52 previously known species (only the type species was used in the analysis) and their three new *Hemichoma* Enderlein, 1920 species from any of the five previously known species.

This is not the first case of its kind in zoology; Hebert *et al.* (2004) proposed to recognize ten species of skipper butterflies (genus *Astrartes*) based on DNA characters and ecology and some morphological characters, but the species were not formally named until Brower (2010) described them based exclusively on unique mutations in the DNA barcode region. Brower (2010) hailed this method as a flagship example of DNA barcoding’s success in overcoming the ‘taxonomic impediment’ (Brower, 2010). However, it has received extensive criticism (DeSalle *et al.*, 2005; Pons *et al.*, 2006; Rubinoff

*et al.*, 2006; Elias *et al.*, 2007; Dupérré, 2020). Meierotto *et al.* (2019) have taken this approach one step further by immediately assigning names to the lineages. This is, in our opinion, a step too far. Each description includes a lateral habitus image of a single specimen, a short diagnosis based solely on COI barcode nucleotide differences, brief notes on biology, and largely unannotated type specimen information except for brief mention of the locality and host caterpillar of the holotype. There are no morphological descriptions, and as mentioned, the ‘molecular diagnosis’ of their new *Zelomorpha* spp. are compared only with that of the type species, and not to the other 51 already known species of that genus, and in the case of *Hemichoma*, with none of the five species that were already described.

We consider this poor taxonomic practice, and their approach to be fundamentally flawed. We urge innovators to pursue revolutionary new approaches that do not undermine the value of taxonomic expertise or produce sloppy results, but rather seek to draw on the latest methodological advancements to increase the rate of taxonomy without compromising on quality.

Over the last 20 years, there have been many calls for an increase in the rate of taxonomic description (Mora *et al.*, 2011). Few have delivered on that promise. Undoubtedly, the single steepest increase in this rate was made possible by DNA barcoding (Hebert *et al.*, 2003). Yet, in the wake of the genetic revolution, there were already concerns that too much emphasis might be placed on DNA barcoding data alone, leading taxonomists to neglect the importance of other data – integrative approaches, taking the congruence of genetic signals with other datasets, would be required to keep describing biologically meaningful units (Dayrat, 2005; Ebach & Holdrege, 2005; Will *et al.*, 2005).

DNA barcoding is a rapid means to sort specimens into clusters, identify species and discover new ones (when a library of the relevant named species is already available), but does not overcome the bottleneck of the description process itself. Proposals for methods to speed up that process were dubbed ‘turbo’ or ‘fast-track’ taxonomy – an approach that does not

Correspondence: Alireza Zamani, Zoological Museum, Biodiversity Unit, 20014 University of Turku, Turku, Finland. E-mail: zamani.alireza5@gmail.com

\*Present address: Department of Biology, University of Konstanz, Konstanz 78457, Germany

differ fundamentally from previous species descriptions, but relies more heavily on formulaic descriptions of large numbers of new species (Butcher *et al.*, 2012; Riedel *et al.*, 2013). In essence, the approach of Meierotto *et al.* (2019) is simply another one of these turbo-taxonomic approaches except in three key aspects:

### **It ignores almost all of the previously described species in both genera**

Fundamentally, a diagnosis should identify features or combinations of features of a new species that are unique, that is, that allow it to be distinguished from all previously named taxa. It thereby gives a means not only to identify the new species but also to demonstrate that the new taxon is not a synonym of an existing one. The diagnoses of Meierotto *et al.* (2019) are based solely on DNA barcodes, but no barcodes are presented for 51 of the 52 existing species of *Zelomorpha*. Indeed, the existing taxa are summarily ignored, except for a statement that the notes of the second author (M.J. Sharkey) were used to verify that the new species are distinct from the existing taxa, without providing any evidence. As a result, it is impossible, based on the study of Meierotto *et al.* (2019), to assess whether or not their 15 new *Zelomorpha* names are distinct from 51 of the 52 species that were already described. In our opinion, Meierotto *et al.* (2019) have impeded, not enhanced, the taxonomy of these wasps. Ignoring almost all previously described species in a genus is indeed a way to speed up taxonomy – the process of comparison becomes very easy when you neglect practically all existing names – but it also creates chaos. Even the fastest approach to taxonomy will always require consideration of existing names before new ones can be established.

### **It uses purely molecular diagnoses**

Morphologically homogeneous ('cryptic') species are difficult to diagnose from one another, even when substantial differences exist in their DNA barcodes. This can delay taxonomy, because more effort must be invested per species to identify characters that do indeed differ. To overcome this problem, Renner (2016) recently called for more widespread inclusion of DNA sequence data in diagnoses. However, we do not believe that Renner (2016) envisioned the complete replacement of the diagnosis by single nucleotide changes, but rather expansion of concise but comprehensive diagnoses with such information. In some cases, restriction exclusively to genetic markers may be appropriate (e.g. where morphology is highly plastic, or where extremely distinctive genetic lineages are demonstrably cryptic in all other available lines of evidence), but such cases are likely to be the exception rather than the rule. Having complementary lines of evidence, such as morphology, is particularly important when, as is the case in Meierotto *et al.* (2019), only a tiny portion of the available names have DNA sequence data available. The lack of overlap between morphological and genetic data

will further delay the process of clarifying whether or not the new names are synonyms of existing species.

It is also important to note that DNA barcoding relies wholly on mitochondrial markers (usually cytochrome oxidase-I). Mitochondrial trees often disagree with nuclear species trees, especially in taxa where *Wolbachia* may be altering mtDNA introgression (Klopfstein *et al.*, 2016). In these cases, and especially when genetic data are the sole basis of species-level recognition, congruence between nuclear and mitochondrial signal should be tested to better reinforce the species units identified. Moreover, as explained by Dupérré (2020), purely DNA-based descriptions will not only make the identification of millions of historical specimens impossible, it will impair this science in developing countries which house most of the undiscovered portion of biodiversity, due to high costs and lack of staff and technology. Considering the status of taxonomy as a fundamental science, this would drastically affect other related fields of study and, importantly, conservation.

### **It provides a single, low-quality photograph instead of a morphological description**

Renner (2016) also called for more emphasis on diagnosis and not description. With highly descriptive taxonomy, a great deal of time is invested in description of features that are not informative for the distinction of species from one another, which is time that could be spent instead diagnosing substantially more species. Instead, she and others have emphasized the importance of high-resolution photographs as supplements to diagnoses. We agree that detailed high-resolution photographs of specimens can indeed be highly valuable, but we contend that (i) there must be several photographs available, not a single lateral photograph of a single specimen, as provided by Meierotto *et al.* (2019), and (ii) some text highlighting important diagnostic features is valuable to experts, and of paramount importance to non-experts, who must instead play a game of 'spot-the-difference' when such information is lacking. Experts might know the difference between variable and nonvariable characters, whereas such features cannot be distinguished by nonexperts, and it is the purpose of the diagnosis, if not the description, to point such features out. Moreover, we note that the photographs of Meierotto *et al.* (2019) are sometimes blurry and almost all of them cut off the tips of the antennae!

Finally, it has been shown that (e.g. in case of tropical parasitoid wasps) the most time-consuming part of species discovery is field sampling (Sääksjärvi *et al.*, 2004; Hopkins *et al.*, 2019), and the actual description of the species may be written within minutes when material and expertise are already available. To make up the gaps in the existing barcode database, which contains maybe 2% of currently named species worldwide (see <http://www.boldsystems.org/>), far more survey work needs to be undertaken. A comprehensive barcoding database for a given taxon is a prerequisite to contemplating a DNA-only approach akin to that of Meierotto *et al.* (2019), and one that will require substantial further work to assemble.

In the face of the Holocene (=sixth) extinction, taxonomists are racing to describe the 8 million unnamed eukaryotes that lie between the 2.07 million species currently named (Frid & Caswell, 2016), and the estimated ten million extant species (Mora *et al.*, 2011; but see also Larsen *et al.*, 2017 for well-reasoned estimates orders of magnitude higher). Currently, the rate of description is around 18 000 species per year (IISE, 2011), but with species going extinct at a rate 1000 times higher than the natural background rate of extinction, the annual species loss is clearly within or even higher than the rate of new descriptions (Dirzo & Raven, 2003; Mora *et al.*, 2011), and thousands of species will undoubtedly go extinct before they can be described to science (IPBES, 2019). As the current average shelf life of new species between discovery and description is about 21 years (Fontaine *et al.*, 2012), we do indeed need revolutionary new approaches to the discovery and description of new species.

BINs and candidate species numbers (Vieites *et al.*, 2009) already serve a valuable purpose as alphanumeric placeholders to recognize potentially evolutionarily significant diversity before it is taxonomically described. Simply assigning all BINs taxonomic names as Meierotto *et al.* (2019) propose would indeed complete the inventory of life on Earth extremely quickly (at precisely the same pace as the rate of barcoding) – that we do not dispute. But it would also remove the quantitative and qualitative difference between these preliminary identifiers (based on a single DNA marker) and full taxonomic recognition (based on a more comprehensive diagnosis, ideally supported by multiple lines of evidence including genetic data) that lend taxonomy its value. It would supplant taxonomists with technicians, who need to know nothing of the biology of the units with which they are dealing. The purpose of inclusion of molecular data in species descriptions should be to produce more precise taxonomic framework. A species description can be thought of as a hypothesis that can be supported or rejected when more data are obtained. Other researchers must have an opportunity to scientifically evaluate the status of the species in question.

In our eyes, methodological changes to the way species are delineated and described are an important component of increasing the rate of species description, but dismissing the existing literature, and producing ‘descriptions’ that contain almost no information on the morphology of species, its variation, their unique features, their biology, or other aspects, do not constitute a revolution, and cannot be adopted. We note that real revolutions are undoubtedly coming, especially from the fields of machine learning and integrative species delimitation (Solís-Lemus *et al.*, 2015; Favret & Sieracki, 2016), and also that it is possible to produce massive, rapidly assembled taxonomic monographs without compromising on quality (Rakotoarison *et al.*, 2017). But we also want to emphasize that there is no shortcut to nirvana, and a true paradigm shift in taxonomy will come only when there is a revolution in the level of financial investment in taxonomy and the natural history museums that house the described and undescribed reference material of life on Earth (Wheeler, 2020), and when legislature stops acting to prohibit the collecting work of dedicated taxonomists while turning a blind eye to the innumerable organisms destroyed with every hectare of habitat that is lost (Britz *et al.*, 2020).

## Acknowledgements

The authors declare no conflicts of interest.

## Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## References

- Britz, R., Hundsdoerfer, A.K. & Fritz, U. (2020) Funding, training, permits – the three big challenges of taxonomy. *Megataxa*, **1**, 59–52. <https://doi.org/10.11646/megataxa.1.1.10>.
- Brower, A.V.Z. (2010) Alleviating the taxonomic impediment of DNA barcoding and setting a bad precedent: names for ten species of ‘*Astraptes fulgerator*’ (Lepidoptera: Hesperiidae: Eudaminae) with DNA-based diagnoses. *Systematics and Biodiversity*, **8**, 485–491. <https://doi.org/10.1080/14772000.2010.534512>.
- Butcher, B.A., Smith, M.A., Sharkey, M.J. & Quicke, D.L.J. (2012) A turbo-taxonomic study of Thai *Aleiodes* (*Aleiodes*) and *Aleiodes* (*Arcalaiodes*) (hymenoptera: Braconidae: Rogadinae) based largely on COI barcoded specimens, with rapid descriptions of 179 new species. *Zootaxa*, **3457**, 1–232. <https://doi.org/10.11646/zootaxa.3457.1.1>.
- Dayrat, B. (2005) Towards integrative taxonomy. *Biological Journal of the Linnean Society*, **85**, 407–415. <https://doi.org/10.1111/j.1095-8312.2005.00503.x>.
- DeSalle, R., Egan, M.G. & Siddall, M. (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society of London B*, **360**, 1905–1916. <https://doi.org/10.1098/rstb.2005.1722>.
- Dirzo, R. & Raven, P.H. (2003) Global state of biodiversity and loss. *Annual Review of Environment and Resources*, **28**, 137–167. <https://doi.org/10.1146/annurev.energy.28.050302.105532>.
- Dupérré, N. (2020) Old and new challenges in taxonomy: what are taxonomists up against? *Megataxa*, **1**, 59–62. <https://doi.org/10.11646/megataxa.1.1.12>.
- Ebach, M.C. & Holdrege, C. (2005) DNA barcoding is no substitute for taxonomy. *Nature*, **434**, 697. <https://doi.org/10.1038/434697b>.
- Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Brower, A.V.Z., Mallet, J. & Jiggins, C.D. (2007) Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society of London*, **274**, 2881–2889.
- Favret, C. & Sieracki, J.M. (2016) Machine vision automated species identification scaled towards production levels. *Systematic Entomology*, **41**, 133–143. <https://doi.org/10.1111/syen.12146>.
- Fontaine, B., Perrard, A. & Bouchet, P. (2012) 21 years of shelf life between discovery and description of new species. *Current Biology*, **22**, 943–944. <https://doi.org/10.1016/j.cub.2012.10.029>.
- Frid, C.L. & Caswell, B.A. (2016) Does ecological redundancy maintain functioning of marine benthos on centennial to millennial time scales? *Marine Ecology*, **37**, 392–410. <https://doi.org/10.1111/maec.12297>.
- Hebert, P.D.N., Cywinska, A., Ball, S.L. & DeWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B*, **270**, 313–321. <https://doi.org/10.1098/rspb.2002.2218>.
- Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H. & Hallwachs, W. (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 USA*, **101**, 14812–14817. <https://doi.org/10.1073/pnas.0406166101>.

- Hopkins, T., Roininen, H., van Noort, S., Broad, G.R., Kaunisto, K. & Sääksjärvi, I.E. (2019) Extensive sampling and thorough taxonomic assessment of Afrotropical Rhyssinae (Hymenoptera, Ichneumonidae) reveals two new species and demonstrates the limitations of previous sampling efforts. *ZooKeys*, **878**, 33–71. <https://doi.org/10.3897/zookeys.878.37845>.
- ICZN (1999) *International Code of Zoological Nomenclature*, 4th edn. International Trust for Zoological Nomenclature, London.
- IISE (2011) *Retro SOS 2000–2009: A Decade of Species Discovery in Review*. International Institute for Species Exploration, Tempe, Arizona. <http://www.esf.edu/species/>.
- IPBES (2019) *Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (ed. by S. Díaz, J. Settele, E. Brondízio *et al.*), p. 56. IPBES secretariat, Bonn.
- Klopfstein, S., Kropf, C. & Baur, H. (2016) *Wolbachia* endosymbionts distort DNA barcoding in the parasitoid wasp genus *Diplazon* (Hymenoptera: Ichneumonidae). *Zoological Journal of the Linnean Society*, **177**, 541–557. <https://doi.org/10.1111/zoj.12380>.
- Larsen, B.B., Miller, E.C., Rhodes, M.K. & Wiens, J.J. (2017) Inordinate fondness multiplied and redistributed: the number of species on earth and the new pie of life. *The Quarterly Review of Biology*, **92**, 229–265. <https://doi.org/10.1086/693564>.
- Linnaeus, C. (1753) *Species Plantarum*. Salvius, Stockholm.
- Linnaeus, C. (1758) *Systema naturæ per Regna Triâ naturæ, Secundum Classes, Ordines, Genera, Species, Cum Characteribus, Differentiis, Synonymis, Locis*. Salvius, Stockholm.
- Mayr, E. (1992) A local flora and the biological species concept. *American Journal of Botany*, **79**, 222–238.
- Meierotto, S., Sharkey, M.J., Janzen, D.H., Hallwachs, W., Hebert, P.D.N., Chapman, E.G. & Smith, M.A. (2019) A revolutionary protocol to describe understudied hyperdiverse taxa and overcome the taxonomic impediment. *Deutsche Entomologische Zeitschrift*, **66**, 119–145. <https://doi.org/10.3897/dez.66.34683>.
- Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B. & Worm, B. (2011) How many species are there on earth and in the ocean? *PLoS Biology*, **9**, e1001127. <https://doi.org/10.1371/journal.pbio.1001127>.
- Naciri, Y. & Linder, P. (2015) Species delimitation and relationships: the dance of the seven veils. *Taxon*, **64**, 3–16. <https://doi.org/10.12705/641.24>.
- Pons, J., Barraclough, T.G., Gomez-Zurita, J. *et al.* (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–609. <https://doi.org/10.1080/10635150600852011>.
- Rakotoarison, A., Scherz, M.D., Glaw, F. *et al.* (2017) Describing the smaller majority: integrative taxonomy reveals twenty-six new species of tiny microhylid frogs (genus *Stumpffia*) from Madagascar. *Vertebrate Zoology*, **67**, 271–398.
- Ratnasingham, S. & Hebert, P.D.N. (2013) A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. *PLoS One*, **8**, e66213. <http://doi.org/10.1371/journal.pone.0066213>.
- Renner, S.S. (2016) A return to Linnaeus's focus on diagnosis, not description: the use of DNA characters in the formal naming of species. *Systematic Biology*, **65**, 1085–1095. <https://doi.org/10.1093/sysbio/syw032>.
- Riedel, A., Sagata, K., Suhardjono, Y.R., Tänzler, R. & Balke, M. (2013) Integrative taxonomy on the fast track - towards more sustainability in biodiversity research. *Frontiers in Zoology*, **10**, 15. <https://doi.org/10.1186/1742-9994-10-15>.
- Rubinoff, D., Cameron, S. & Will, K. (2006) A genomic perspective on the shortcomings of mitochondrial DNA for barcoding identification. *Journal of Heredity*, **97**, 581–594. <https://doi.org/10.1093/jhered/esl036>.
- Sääksjärvi, I.E., Haataja, S., Neuvonen, S., Gauld, I.D., Jussila, R., Salo, J. & Burgos, A.M. (2004) High local species richness of parasitic wasps (Hymenoptera: Ichneumonidae; Pimplinae and Rhyssinae) from the lowland rainforests of Peruvian Amazonia. *Ecological Entomology*, **29**, 735–743. <https://doi.org/10.1111/j.0307-6946.2004.00656.x>.
- Solís-Lemus, C., Knowles, L.L. & Ané, C. (2015) Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution*, **69**, 492–507. <https://doi.org/10.1111/evo.12582>.
- Vieites, D.R., Wollenberg, K.C., Andreone, F., Köhler, J., Glaw, F. & Vences, M. (2009) Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. *Proceedings of the National Academy of Sciences of the USA*, **106**, 8267–8272. <http://doi.org/10.1073/pnas.0810821106>.
- Wheeler, Q. (2020) A taxonomic renaissance in three acts. *Megataxa*, **1**, 4–8. <https://doi.org/10.11646/megataxa.1.1.2>.
- Will, K.W., Mishler, B.D. & Wheeler, Q.D. (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, **54**, 844–851. <https://doi.org/10.1080/10635150500354878>.

Accepted 7 July 2020