Zootaxa 5027 (2): 151–159
https://www.mapress.com/j/zt/

Article

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

https://doi.org/10.11646/zootaxa.5027.2.1
http://zoobank.org/urn:lsid:zoobank.org:pub:1A2457B4-C496-4CEB-9D95-F87B822DC619

Is it time to describe new species without diagnoses? —A comment on Sharkey *et al.* (2021)

DIRK AHRENS[1*], SHANE T. AHYONG[2], ALBERTO BALLERIO[3], MAXWELL V. L. BARCLAY[4], JONAS EBERLE[5], MARIANNE ESPELAND[1,9], BERNHARD A. HUBER[1,10], XIMO MENGUAL[1,11], THAYNARA L. PACHECO[1,6], RALPH S. PETERS[1,12], BJÖRN RULIK[1,13], FERNANDO VAZ-DE-MELLO[7], THOMAS WESENER[1,14] & FRANK-THORSTEN KRELL[8]

[1]*Zoological Research Museum Alexander Koenig, Bonn, Germany*
[2]*Australian Museum Research Institute, Sydney, Australia, and School of Biological, Earth and Environmental Sciences, University of New South Wales, Australia*
Shane.Ahyong@Australian.Museum; https://orcid.org/0000-0002-2820-4158
[3]*Viale Venezia 45, Brescia, Italy*
alberto.ballerio.bs@aballerio.it; https://orcid.org/0000-0001-9772-976X
[4]*Department of Life Sciences, Natural History Museum, London, UK*
m.barclay@nhm.ac.uk; https://orcid.org/0000-0003-4989-2014
[5]*Paris-Lodron-University, Hellbrunner Straße 34, 5020 Salzburg, Austria*
eberle.jonas@gmail.com; https://orcid.org/0000-0003-2519-0640
[6]*Museu Nacional—Universidade Federal do Rio de Janeiro, Brazil*
thay.lpacheco@gmail.com; https://orcid.org/0000-0001-9503-7751
[7]*Universidade Federal de Mato Grosso, Cuiaba, Brazil*
vazdemello@gmail.com; https://orcid.org/0000-0001-9697-320X
[8]*Department of Zoology, Denver Museum of Nature & Science, Colorado, U.S.A.*
Frank.Krell@dmns.org; https://orcid.org/0000-0001-6065-0812
[9] *M.Espeland@leibniz-zfmk.de; https://orcid.org/0000-0002-6800-4783*
[10] *B.Huber@leibniz-zfmk.de; https://orcid.org/0000-0002-7566-5424*
[11] *X.Mengual@leibniz-zfmk.de; https://orcid.org/0000-0002-6185-9404*
[12] *ralph_peters@hotmail.com, R.Peters@leibniz-zfmk.de; https://orcid.org/0000-0001-7784-9203*
[13] *B.Rulik@leibniz-zfmk.de; https://orcid.org/0000-0002-2808-9305*
[14] *T.Wesener@leibniz-zfmk.de; https://orcid.org/0000-0002-2028-3541*
[*]*Corresponding author. D.Ahrens@leibniz-zfmk.de, ahrens.dirk_col@gmx.de; https://orcid.org/0000-0003-3524-7153*

Abstract

New methods in taxonomy and systematics can influence the overall practice of formally naming and describing biodiversity. DNA barcoding has been controversial since its emergence, but now, large scale species descriptions exclusively based on barcodes have created what can be called a 'new quality of performance'. Its limitations are discussed from different perspectives: nomenclature, general pragmatism, and problems of DNA-based species delimitation in the light of the central aim of achieving a robust and stable nomenclature of organisms, essential for all applications of biodiversity research. This issue needs to be addressed to prevent restraining the progress of taxonomy and its ability to contribute to modern science.

Key words: Taxonomy, species delimitation, *COI* barcoding, nomenclature rules

Finding new species: Hypotheses, methods, tools

Taxonomy implies two distinct processes of knowledge acquisition: 1) referencing specimens to an existing system of taxa and therefore to their names adopted according to the principle of priority and with reference to type specimens, in the context of the Linnaean nomenclature (ICZN 1999); and 2) the delimitation of species units. The first step is a procedure that combines archival activity, i.e., considering the previous publication record and the

Accepted by A. Minelli: 20 Aug. 2021; published: 31 Aug. 2021

type specimens preserved in natural history collections, and a systematic classification effort based on a scientific hypothesis of phylogenetic relationships. The latter involves testing a scientific hypothesis about how to delineate a boundary between a species taxon and the most closely related putative taxon (once more with reference to type specimens if a new species is to be distinguished from a known one). The functionality of this reference system, on the basis of which over 1.5 million. animal species have been described to date (IUCN 2020), has historically been guaranteed, despite the consistent use of a large number of different species concepts (Stankowski & Ravinet 2021), by the existence of one prevailing reference trait complex for all names, i.e., morphology, and is expressed by the fixation (and preservation for future additional study) of a type specimen (current nomenclature rules imply this but do not state it explicitly).

Species delimitation methods, along with various species concepts, continue to evolve and improve with new technology, more robust data, and increasingly sophisticated algorithms (de Queiroz 2007). Morphology-based species delimitations (historically performed by trained taxonomists and being sometimes perceived as 'subjective') are being refined by mechanized ('objective') methodologies and new lines of evidence, such as behavior, natural history, morphometric data, genetic data, etc.

The Code and the DNA barcode-based species

In a recent paper, Sharkey *et al.* (2021) described 403 hymenopteran species new to science in what they call a "minimalist revision". Their species descriptions are mostly composed of a nucleotide sequence of the mitochondrial protein-coding gene cytochrome *c* oxidase subunit 1 (*COI* or the DNA barcode, for animals at least) (accompanied by a simple neighbor-joining tree) and one photograph, rarely more, including collection data and often host data of the preserved type specimens.

While taxonomists have long debated standards for species descriptions, that discussion has recently mainly focused on the principle of type specimen preservation (e.g., Amorim *et al.* 2016; Ceríaco *et al.* 2016; Krell and Marshall 2017) for which the International Code of Zoological Nomenclature (ICZN 1999, 2017) provides rather clear rules and recommendations. In contrast to this, ICZN recommendations regarding the quality of data used for the delimitation of new taxa are rather general and allow for a wider margin of interpretation: "An author, when drawing up the description of a new nominal taxon, should include comparisons with appropriate related taxa in order to assist later identification of the taxon..." (ICZN 1999: 125 [Appendix B]). This is in line with recent demands for more hypothesis-driven research in modern taxonomy (Vences 2020), in which diagnoses should play a major role, to reduce the image of taxonomy as purely descriptive. The majority of new species "descriptions" by Sharkey *et al.* (2021) fail to comply with such a recommendation. The user of such descriptions has to search the "diagnosis", either reexamining the type specimens or by reanalysis of the *COI* barcode data. This by itself does not threaten the nomenclatural availability of the names proposed, since Appendix B (ICZN 1999) is a mere recommendation, but their minimalist approach challenges the criteria of availability of newly established names set by the Code. In fact, Article 13.1.1 (ICZN 1999) states that a criterion of a name's availability is the presence of a "description or definition that states in words characters that are purported to differentiate the taxon". Both "description" and "definition" are also defined in the Code Glossary (an integral mandatory part of the Code, see also Art. 89 of the Code). According to the Glossary a "definition" is "a statement in words that purports to give those characters which, in combination, uniquely distinguish a taxon" and a "description" is "a statement in words of taxonomic characters of a specimen or a taxon". It is debatable whether a DNA sequence alone can fulfil the requirement of being "a statement in words". It could be argued that the letters forming the sequence are the initials of the nucleotide bases, and therefore the DNA sequence is nothing but a list of characters (even if abbreviated, based on a universal convention); others will argue that the sequence does not fulfil the requirements of Art. 13.1.1 and therefore deem those names unavailable. This minimalism could therefore become a source of "limbo names", i.e., names which may or may not be considered available, according to subjective evaluations by different users of taxonomy. The Code of Zoological Nomenclature needs to become more explicit and should discourage minimalist approaches in a way that such divergent interpretations are no longer possible.

Standardizing new species' discovery

In the past two decades, DNA barcoding has increased the quality and reproducibility of species' characterization and enabled rapid assessments of biodiversity (e.g., Taberlet *et al.* 2012). A major advantage is the ability to standardize and automate species recognition by using a single gene fragment with standardized protocols (e.g., Carstens *et al.* 2013). DNA sequences such as *COI* barcodes include all the possibly informative characters that they have. The 5'-half of the *COI* gene has been the most widely used marker in animals (Blaxter 2016). DNA-based species delimitation also made way for the direct inference of species boundaries from unknown samples (Pons *et al.* 2006), and allowed association of different life stages with each other (Ahrens *et al.* 2007). At an earlier stage of the DNA barcoding campaign, Hebert *et al.* (2004) stated that their results "illustrate the value of DNA barcoding, especially when coupled with traditional taxonomic tools, in disclosing hidden diversity". Others have enthusiastically suggested completely replacing traditional taxonomy by a "DNA taxonomy" (Tautz *et al.* 2003). While in these early years no species were formally described and named according to the Zoological Code exclusively using barcodes (Hebert *et al.* 2004), this happened later (e.g., Brower 2010; Meierotto *et al.* 2019; Sharkey *et al.* 2021). In some cases, diagnoses were made specific by highlighting diagnostic nucleotide positions, and in others, such as Meierotto *et al.* (2019), delimitation was based on threshold clustering (2%) as implemented in the Barcode of Life Index Number (BIN; Ratnasingham and Hebert 2013). In Sharkey *et al.* (2021), most of the descriptions solely authored by Sharkey himself lack any species diagnostics (same applies to Doerder 2019); instead, pure nucleotide data (in the case of Doerder 2019 only NCBI accession numbers) are provided along with neighbor joining trees. These trees, however, do not show qualitative or diagnostic characters but only reflect genetic distances within an assemblage of taxa of (poorly defined) limited geographical scope.

Molecular characters, alone or supplemented with other evidence such as ecological or life history data, have been used in the past for describing new species of different organisms that are morphologically extremely similar and the taxonomic distinctiveness of which is not easily diagnosed morphologically (e.g., Brower 2010). Most of these works have also assessed morphological variation, thus providing a link to the existing diagnostic reference system based on morphology (Burns *et al.* 2008; Halt *et al.* 2009; Pfeiler *et al.* 2011). Renner (2016) discussed the use of DNA characters in the formal naming of species based on previous works that used DNA marker information for species diagnosis (of which the majority provided a link to morphological traits in the descriptions). She suggested incorporating DNA barcodes into species diagnoses as a recommendation in all codes of organismal nomenclature.

Zamani *et al.* (2021) already raised critical arguments against an exclusively DNA-based approach of species description (Meierotto *et al.* 2019), especially regarding the ignorance of previously described species, the pure use of molecular diagnoses, the use of insufficient photographic documentation instead of a "morphological description", and the huge gaps in the existing barcode database, which includes maybe 2% of currently named species worldwide (see http://www.boldsystems.org/). These issues alone would be sufficient to make *COI* barcodes unsuitable as stand-alone reference system for the delineation of new species. Future discussions might also involve environmental DNA as already being the case in fungal taxonomy (e.g., Hongsanan *et al.* 2018; Wu *et al.* 2019).

The approach chosen by Sharkey *et al.* (2021a) is problematic, since it provides neither DNA-based nor morphology-based species diagnoses, i.e., it does not say *how* one species differs from other related ones. This, however, is precisely hypothesis testing in taxonomy, and therefore represents a bedrock of science. Neglecting the recommendations of the ICZN (see above), Sharkey *et al.* (2021a) name species in this limited way for the first time on a large scale. In a reply to Zamani *et al.* (2021), Sharkey *et al.* (2021b) again justified this approach without giving a better rationale than before. If this sets a precedent, we might soon have to cope with two parallel taxonomies, in which long established (morphology-defined) names might become *nomina dubia* (e.g., Pfeiler & Nazario-Yepiz 2020) since reference systems (morphology vs *COI* sequences) for the taxa are not directly and immediately relatable. Such parallel taxonomies that consider the names of the other faction doubtful would be extremely detrimental for taxonomy, a result that most supporters of the current barcoding community would neither like to see nor stand for. Moreover, such a situation would paralyze cautious taxonomists working with morphology because reference systems are not directly compatible and also exclude taxonomists (Dupérré 2020) who do not have access to appropriate funding or resources. The BioScan consortium (Pennisi 2019), for example, seems as yet unconnected to the taxonomic community and its needs (Pinheiro *et al.* 2019).

Problems with COI barcode taxonomy

Working outside the existing reference system for established taxonomic names would not be the only flaw of a stand-alone-*COI* barcode taxonomy. In the past 15 years, empirical studies of DNA- and integrative taxonomy (e.g., Carstens *et al.* 2013) have demonstrated that DNA taxonomy is far more complex and complicated than the oversimplified assumption that the "barcode gap" will distinguish species (e.g., Meyer & Paulay 2005; Wiemers & Fiedler 2007). This assumption is only the tip of an iceberg of problems. Despite its apparent simplicity and increasing low-cost, *COI* is rather unsuitable for inferring species boundaries due to many issues: species delimitation and identification based on information from a single mitochondrial gene is prone to errors due to extrachromosomal inheritance and accordingly reduced rate of gene flow, little recombination, incomplete lineage sorting, sex-biased dispersal, asymmetrical introgression, and/or *Wolbachia*-mediated genetic sweeps (Funk & Omland 2003; Ballard & Whitlock 2004; Petit & Excoffier 2009; Klopfstein *et al.* 2016). Furthermore, coalescence times are three to four times faster than in nuclear markers (Birky *et al.* 1989; Palumbi *et al.* 2001). Thus, divergence in mtDNA may be a result of speciation, or it may not if it is not correlated with evidence from nuclear/genomic DNA. There is certainly some practical convenience in examining environmental samples using a single marker as opposed to multiple markers. Also, costs are currently still lower compared to nuclear multi-marker approaches. Its inaccuracy is due to widely distributed mitochondrial paraphyly of species, which increases with geographical upscaling and more extensive sampling of the analyzed data (Bergsten *et al.* 2012), common sex-biased dispersal and thus biased patterns of *COI* divergence as well as ancestral polymorphism leading to false "cryptic" divergence (e.g., Ahrens *et al.* 2013; Eberle *et al.* 2019). The *COI*-inferred species number may in such cases exceed the true species number by up to ten times (Eberle *et al.* 2019). This level of inaccuracy is too high to be acceptable for a robust and stable species delimitation and nomenclature. Furthermore, the outcome of species delimitation analyses is affected by (unbalanced) sampling, effective population size (of each species), fluctuation of effective population size (within a group of taxa whose species limits are to infer), the depth of phylogenetic sampling, and last but not least, by the choice of sampling or the method of species delimitation (Esselstyn *et al.* 2012; Fujisawa & Barraclough 2013; Ahrens *et al.* 2016; Meier *et al.* 2021). In short, results of a DNA barcode-based taxonomy are not as objective and stable as widely claimed. Meier *et al.* (2021) also underlined the instability of BINs, on which species entities of Sharkey *et al.* (2021a) were based, and that their circumscription is founded partly on proprietary and unpublished data. To uncritically derive conclusions based on DNA data is problematic (Carstens *et al.* 2013; Meier *et al.* 2021); even multi-gene nuclear DNA data may result in over-splitting of true species, despite applying current state-of-the-art species delimitation methods, like those based on the multispecies coalescent model (e.g., Sukuruman & Knowles 2017).

Integrative taxonomy approaches

As most of these problems have long been recognized, taxonomists have developed integrative taxonomy approaches incorporating both molecular and morphological data (Padial *et al.* 2010; Schlick-Steiner *et al.* 2010; Yeates *et al.* 2011; Eberle *et al.* 2016). While there are many excellent examples of good practice of "turbo taxonomy" using such integrative approaches (e.g., Riedel *et al.* 2013a,b; Srivathsan *et al.* 2019), a further step forward was achieved by the establishment of integrative taxonomic approaches using iterative analyses (e.g., Solís-Lemus *et al.* 2015). Contrary to the claim that these methods require multiple molecular markers (Sharkey *et al.* 2021a), this approach can also be used with DNA barcodes and trait measurements (although it benefits from additional markers). Further approaches that are more scalable are being developed (Derkarabetian *et al.* 2019). Such methods combine different sources of evidence such as morphology, multiple nuclear genes, etc. in a single species delimitation analysis, which helps to overcome major philosophical problems in taxonomy through the application of alternative species concepts (Conix 2018). However, most of the highly informative markers of the multi-gene studies have problems in common, either a lack of universality (for a universal reference system such as *COI* barcodes) or a lack of resolution (Eberle *et al.* 2020). This is why *COI* data should and will have a crucial role in species identification and biodiversity monitoring. Recently, however, nuclear markers have been identified that are much more widely applicable and show even better discriminating power than mitochondrial DNA (Eberle *et al.* 2020; Dietz *et al.* 2021). With the ever decreasing sequencing costs of genomic data and increasing computational power, these markers might soon be ready to complement *COI* barcodes in the context of a taxonomy in which data for species discovery are inclusive rather than exclusive, and in which the major goal will be a satisfactory upgrade of reference data sets (Miralles *et al.* 2020), including morphology, of the entirety of global diversity.

Parallel taxonomies?

Sharkey *et al.* (2021a,b) in reply to Zamani *et al.* (2021), are well aware of some of the problems we raise here, but the answers they offer are a combination of postponing the solution or ignoring the problem: parallel taxonomies will persist until other taxonomists can close the gap between the two mutually exclusive reference systems by subsequently documenting the morphology of Sharkey's type specimens. Species exclusively based on *COI* barcodes would dictate a temporary taxonomy as the result of the current opportunities, relying on an outdated methodology (Meier *et al.* 2021) and determined by financial constraints. Future taxonomists, likely using emerging and more appropriate genomic data, would have to correct the created taxonomic inflation (Vences 2020) by linking *COI*-based species to those so far only known by morphology by undertaking morphological analysis afresh – a non-trivial task. Taxonomists or ecologists who are not in a position to produce barcodes, cannot test and use the data from Sharkey's species descriptions. Collections with samples suitable for DNA extraction have become more common in the last decades, but often still cover a too limited number of species and geographical areas to allow conclusions about biodiversity in time and space. In contrast, natural history collections still harbor millions of unidentified insect specimens from all over the world (in addition to the type material). To obtain comprehensive biodiversity data in time and space and not only from appropriately preserved recent material, these collections will have to continue to be examined based on morphology.

Another question concerns the fate of the diagnostic morphological evidence that was elaborated in the background of Sharkey's study (Sharkey *et al.* 2021a) and that was used to verify the barcoding databases and to assign the species to the respective genera in order to properly name them. If there is no published record, not even a brief one, of morphological characters in the species diagnosis, then it remains unclear how integration with barcode data, as mentioned by Sharkey *et al*. (2021a), could be implemented. The knowledge of how to do this is likely to be lost once the specialist who examined the original type specimens (as outlined by Sharkey *et al.* 2021a) is gone.

Morphology-based taxonomy can also be rapid in species discovery, if at least moderate funding is available. We acknowledge the need for innovative and even unorthodox approaches in largely unknown, megadiverse taxa (Meierotto *et al.* 2019; Vences 2020), but we consider the "minimalist" approach of Sharkey *et al.* (2021a) to be more harmful than useful. What is the advantage of naming BIN clusters or molecular operational taxonomic units if it is unclear whether they are species or not (e.g., Brower 2010; Blaxter 2016; Vences 2020; Meier *et al.* 2021), and if they are not integrated into the existing reference system? What is the criterion for applying one reference system (morphology) or another (*COI* sequences), and who makes this decision? We contend that a temporary and parallel taxonomy exclusively based on *COI* barcodes is not a constructive way to face the taxonomic impediment or to guarantee nomenclatural stability in the 21st century. Examples of integrative "turbo taxonomy" (see above) have shown that we can do better. The use of molecular data for species assessments and discovery is only progressive in an integrated framework where new evidence is compatible with existing knowledge, and where competing hypotheses can be tested against each other in a framework of established, widely accepted and clearly specified scientific rules.

A way ahead

Adapting scientific processes and hypotheses to the most economic or available methodology and technical feasibility is problematic. Massive innovation from new technologies, such as barcoding, deep learning, or information science, however, should not bring us to a decline in our understanding of species, and in consequence their naming (e.g., Kennedy *et al.* 2005; Garnett *et al.* 2020; Cellinese *et al.* 2021). This would be extremely detrimental to taxonomy as a science. Last but not least, deep learning and automated image recognition (Høye *et al.* 2021; Gerovichev *et al.* 2021) provide great potential even in small insects to employ morphology in taxonomy and rapid automatized biodiversity assessment, as well as for fast and automatized morphological trait extraction (e.g., Klasen *et al.* 2020). Thus, future biodiversity research will rely on all available diagnostic data (morphology and DNA) for species identification and delimitation. Therefore, the future of taxonomy will embrace integrative rather than exclusive approaches.

Acknowledgements

We are thankful to Sven Kullander (Stockholm, Sweden) and Alessandro Minelli (Padova, Italy) for the discussion and helpful comments on the manuscript.

References

Ahrens, D., Monaghan, M.T. & Vogler, A.P. (2007) DNA-based taxonomy for associating adults and larvae in multi-species assemblages of chafers (Coleoptera: Scarabaeidae). *Molecular Phylogeneny and Evolution*, 44, 436–449.
https://doi.org/10.1016/j.ympev.2007.02.024

Ahrens, D., Fabrizi, S., Sipek, P. & Lago, P. (2013) Integrative analysis of DNA phylogeography and morphology of the European rose chafer (*Cetonia aurata*) to infer species taxonomy and patterns of postglacial colonisation in Europe. *Molecular Phylogeneny and Evolution*, 69, 83–94.
https://doi.org/10.1016/j.ympev.2013.05.016

Ahrens, D., Fujisawa, T., Krammer, H.-J., Eberle, J., Fabrizi, S. & Vogler, A.P. (2016) Rarity and incomplete sampling in DNA-based species delimitation. *Systematic Biology*, 65, 478–494.
https://doi.org/10.1093/sysbio/syw002

Amorim, D.S., Santos, C.M.D., Krell, F.-T., Dubois, A., Nihei, S.S., Oliveira, O.M.P., Pont, A., Song, H., Verdade, V.K., Fachin, D.A., Klassa, B., Lamas, C.J.E., Oliveira, S.S., Carvalho, C.J.B. de, Mello-Patiu, C.A., Hajdu, E., Couri, M.S., Silva, V.C., Capellari, R.S., Falaschi, R.L., Feitosa, R.M., Prendini, L., Pombal Jr., J.P., Fernández, F., Rocha, R.M., Lattke, J.E., Caramaschi, U., Duarte, M., Marques, A.C., Reis, R.E., Kurina, O., Takiya, D.M., Tavares. M., Fernandes, D.S., Franco, F.L., Cuezzo, F., Paulson, D., Guénard, B., Schlick-Steiner, B.C., Arthofer, W., Steiner, F.M., Fisher, B.L., Johnson, R.A., Delsinne, T.D., Donoso, D.A., Mulieri, P.R., Patitucci, L.D., Carpenter, J.M., Herman, L. & Grimaldi, D. (2016) Timeless standards for species delimitation. *Zootaxa*, 4137 (1), 121–128.
https://doi.org/10.11646/zootaxa.4137.1.9

Ballard, J.W. & Whitlock, M.C. (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, 13, 729–744.
https://doi.org/10.1046/j.1365-294X.2003.02063.x

Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M.T., Balke, M., Hendrich, L., Geijer, J., Herrmann, J., Foster, G.N., Ribera, I., Nilsson, A.N., Barraclough, T.G. & Vogler, A.P. (2012) The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology*, 61, 851–869.
https://doi.org/10.1093/sysbio/sys037

Birky, C.W., Fuerst, P. & Maruyama, T. (1989) Organelle gene diversity under migration, mutation and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics*, 121, 613–627.
https://doi.org/10.1093/genetics/121.3.613

Blaxter, M. (2016) Imagining Sisyphus happy: DNA barcoding and the unnamed majority. *Philosophical Transactions of the Royal Society B*, 371, 20150329.
https://doi.org/10.1098/rstb.2015.0329

Brower, A.V.Z. (2006) Problems with DNA barcodes for species delimitation: 'ten species' of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). *Systematics and Biodiversity*, 4 (2), 127–132.
https://doi.org/10.1017/S147720000500191X

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

Burns, J.M., Janzen, D.H., Hajibabaei, M., Hallwachs, W. & Hebert, P.D.N. (2008) DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservacion Guanacaste, Costa Rica. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 6350–6355.
https://doi.org/10.1073/pnas.0712181105

Carstens, B.C., Pelletier, T.A., Reid, N.M. & Satler, J.D. (2013) How to fail at species delimitation. *Molecular Ecology*, 22, 4369–4383.
https://doi.org/10.1111/mec.12413

Cellinese, N., Conix, S. & Lapp, H. (2021) Phyloreferences: Tree-Native, Reproducible, and Machine-Interpretable Taxon Concepts. Available from; https://doi.org/10.32942/osf.io/57yjs (accessed 5 March 2021)

Ceríaco, L.M.P., Gutiérrez, E.E. & Dubois, A. (2016) Photography-based taxonomy is inadequate, unnecessary, and potentially harmful for biological sciences. *Zootaxa*, 4196 (3), 435–445.
https://doi.org/10.11646/zootaxa.4196.3.9

Conix, S. (2018) Integrative taxonomy and the operationalization of evolutionary independence. *European Journal for Philosophy of Science*, 8, 587–603.
https://doi.org/10.1007/s13194-018-0202-z

De Queiroz, K. (2007) Species concepts and species delimitation. *Systematic Biology*, 56 (6), 879–886.
https://doi.org/10.1080/10635150701701083

Derkarabetian, S., Castillo, S., Koo, P.K., Ovchinnikov, S. & Hedin, M. (2019) A demonstration of unsupervised machine learning in species delimitation. *Molecular Phylogeneny and Evolution*, 139, 106562.
https://doi.org/10.1016/j.ympev.2019.106562

Dietz, L., Eberle, J., Mayer, C., Kukowka, S., Bohacz, C., Baur, H., Espeland, M., Huber, B.A., Hutter, C., Mengual, X., Peters, R.S., Vences, M., Wesener, T., Willmott, K., Misof, B., Niehuis, O. & Ahrens, D. (2021) Standardized nuclear markers advance metazoan taxonomy. *bioRxiv*. [published online]
https://doi.org/10.1101/2021.05.07.443120

Dupérré, N. (2020) Old and new challenges in taxonomy: what are taxonomists up against? *Megataxa*, 1 (1), 59–62.
https://doi.org/10.11646/megataxa.1.1.12

Doerder, F.P. (2019) Barcodes reveal 48 new species of *Tetrahymena, Dexiostoma*, and *Glaucoma*: phylogeny, ecology, and biogeography of new and established species. *Journal of Eukaryotic Microbiology*, 66, 182–208.
https://doi.org/10.1111/jeu.12642

Eberle, J., Warnock, R.C.M. & Ahrens, D. (2016) Bayesian species delimitation in *Pleophylla* chafers (Coleoptera)—the importance of prior choice and morphology. *BMC Ecology and Evolution*, 16, 94.
https://doi.org/10.1186/s12862-016-0659-3

Eberle, J., Fabrizi, S., Bazzato, E., Rossi, M., Stella Columba, M., Cillo, D., Uliana, M., Sparacio, I., Sabatinelli, G., Warnock, R.C.M., Carpaneto, G.M. & Ahrens, D. (2019) Sex-biased dispersal obscures species boundaries in integrative species delimitation approaches. *Systematic Biology*, 68, 441–459.
https://doi.org/10.1093/sysbio/syy072

Eberle, J., Ahrens, D., Mayer, C., Niehuis, O. & Misof, B. (2020) A plea for implementing a standardized set of nuclear markers in DNA taxonomy. *Trends in Ecology and Evolution*, 35 (4), 336–345.
https://doi.org/10.1016/j.tree.2019.12.003

Esselstyn, J.A., Evans, B.J., Sedlock, J.L., Khan, F.A.A. & Heaney, L.R. (2012) Single-locus species delimitation: a test of the mixed Yule–coalescent model, with an empirical application to Philippine round-leaf bats. *Proceedings of the Royal Society B*, 279, 3678–3686.
https://doi.org/10.1098/rspb.2012.0705

Fujisawa, T. & Barraclough, T. (2013) Delimiting species using the Generalized Mixed Yule Coalescent (GMYC) approach: a revised method and evaluation on simulated datasets. *Systematic Biology*, 62, 707–724.
https://doi.org/10.1093/sysbio/syt033

Funk, D.J. & Omland, K.E. (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, 34, 397–423.
https://doi.org/10.1146/annurev.ecolsys.34.011802.132421

Garnett, S.T., Christidis, L., Conix, S., Costello, M.J., Zachos, F.E., Bánki, O.S., Bao, Y., Barik, S.K., Buckeridge, J.S., Hobern, D., Lien, A., Montgomery, N., Nikolaeva, S., Pyle, R.L., Thomson, S.A., van Dijk, P.P., Whalen, A., Zhang, Z.-Q. &Thiele, K.R. (2020) Principles for creating a single authoritative list of the world's species. *PLoS Biology*, 18 (7), e3000736.
https://doi.org/10.1371/journal.pbio.3000736

Gerovichev, A., Sadeh, A., Winter, V., Bar-Massada, A., Keasar, T. & Keasar, C. (2021) High throughput data acquisition and deep learning for insect ecoinformatics. *Frontiers in Ecology and Evolution*, 9, 1–11.
https://doi.org/10.3389/fevo.2021.600931

Halt, M.N., Kupriyanova, E.K., Cooper, S.J.B. & Rouse, G.W. (2009) Naming species with no morphological indicators: species status of *Galeolaria caespitosa* (Annelida: Serpulidae) inferred from nuclear and mitochondrial gene sequences and morphology. *Invertebrate Systematics,* 23, 205–222.
https://doi.org/10.1071/IS09003

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 United States of America*, 101 (41), 14812–14817.
https://doi.org/10.1073/pnas.0406166101

Høye, T.T., Ärje, J., Bjerge, K., Hansen, O.L.P., Iosifidis, A., Leese, F., Mann, H.M.R., Meissner, K., Melvad, C. & Raitoharju, J. (2021) Deep learning and computer vision will transform entomology. *Proceedings of the National Academy of Sciences of the United States of America*, 118, e2002545117.
https://doi.org/10.1073/pnas.2002545117

Hongsanan, S., Jeewon, R., Purahong, W., Xie, N., Liu, J.-K., Jayawardena, R.S., Ekanayaka, A.H., Dissanayake, A., Raspé, O., Hyde, K.D., Stadler, M. & Peršoh, D. (2018) Can we use environmental DNA as holotypes? *Fungal Diversity*, 92, 1–30.

https://doi.org/10.1007/s13225-018-0404-x

ICZN (1999) *International Code of Zoological Nomenclature. 4th Edition*. International Trust for Zoological Nomenclature, London, XXIX + 306 pp.

ICZN (2017) Declaration 45—Addition of Recommendations to Article 73 and of the term "specimen, preserved" to the Glossary. *The Bulletin of Zoological Nomenclature*, 73, 96–97. https://doi.org/10.21805/bzn.v73i2.a2

IUCN (2020) The IUCN Red List of Threatened Species. Version 2020-3. Summary statistics. Available from: https://www.iucnredlist.org (accessed 11 February 2021)

Kennedy, J.B., Kukla, R. & Paterson, T. (2005) Scientific names are ambiguous as identifiers for biological taxa: their context and definition are required for accurate data integration. *In*: Ludäscher, B. & Raschid, L. (Eds.), *Data integration in the life sciences. DILS 2005. Lecture Notes in Computer Science. Vol. 3615*. Springer, Berlin and Heidelberg, pp. 80–95. https://doi.org/10.1007/11530084_8

Klasen, M., Ahrens, D., Eberle, J. & Steinhage, V. (2021) Image-based automated species identification: Can virtual data augmentation overcome problems of insufficient sampling? *Systematic Biology.* [published online] https://doi.org/10.1093/sysbio/syab048

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

Krell, F.-T. & Marshall, S.A. (2017) New species described from photographs: Yes? No? Sometimes? A fierce debate and a new Declaration of the ICZN. *Insect Systematics and Diversity*, 1, 3–19. https://doi.org/10.1093/isd/ixx004

Meyer, C.P. & Paulay, G. (2005) DNA Barcoding: error rates based on comprehensive sampling. *PLoS Biology*, 3 (12), e422.

Meier, R., Blaimer, B., Buenaventura, E., Hartop, E., von Rintelen, T., Srivathsan, A. & Yeo, D. (2021) 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. *bioRxiv*. [published online] https://doi.org/10.1101/2021.04.28.441626

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 (2), 119–145. https://doi.org/10.3897/dez.66.34683

Miralles, A., Bruy, T., Wolcott, K., Scherz, M.D., Begerow, D., Beszteri, B., Bonkowski, M., Felden, J., Gemeinholzer, B., Glaw, F., Glöckner, F.O., Hawlitschek, O., Kostadinov, I., Nattkemper, T.W., Printzen, C., Renz, J., Rybalka, N., Stadler, M., Weibulat, T., Wilke, T., Renner, S.S. & Vences, M. (2020) Repositories for taxonomic data: Where we are and what is missing. *Systematic Biology*, 69, 1231–1253. https://doi.org/10.1093/sysbio/syaa026

Padial, J.M., Miralles, A., De la Riva, I. & Vences, M. (2010) The integrative future of taxonomy. *Frontiers in Zoology*, 7, 16. https://doi.org/10.1186/1742-9994-7-16

Palumbi, S.R., Cipriano, F. & Hare, M.P. (2001) Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. *Evolution*, 55 (5), 859–868. https://doi.org/10.1554/0014-3820(2001)055[0859:PNGCFM]2.0.CO;2

Pennisi, E. (2019) DNA barcodes jump-start search for new species. *Science*, 364 (6444), 920–921. https://doi.org/10.1126/science.364.6444.920

Petit, R.J. & Excoffier, L. (2009) Gene flow and species delimitation. *Trends in Ecology and Evolution*, 24, 386–393. https://doi.org/10.1016/j.tree.2009.02.011

Pfeiler, E., Van der Heiden, A.M., Ruboyianes, R.S. & Watts, T. (2011) *Albula gilberti*, a new species of bonefish (Albuliformes: Albulidae) from the eastern Pacific, and a description of adults of the parapatric *A. esuncula*. *Zootaxa*, 3088 (1), 1–14. https://doi.org/10.11646/zootaxa.3088.1.1

Pfeiler, E. & Nazario-Yepiz, N.O. (2020) DNA-based taxonomy and potential suppression of long-established names: the case of *Telegonus fulgerator* (Lepidoptera: Hesperiidae), *Systematics and Biodiversity*, 18, 338–346. https://doi.org/10.1080/14772000.2020.1758825

Pinheiro, H. T., Moreau, C. S., Daly, M. & Rocha, L.A. (2019) Will DNA barcoding meet taxonomic needs? *Science*, 365 (6456), 873–874. https://doi.org/10.1126/science.aay7174

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W.D. & Vogler, A.P. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, 55, 595–609. https://doi.org/10.1080/10635150600852011

Ratnasingham, S. & Hebert, P.D.N. (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, 7, 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

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. (2013a) 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

Riedel, A., Sagata, K., Surbakti, S., Tänzler, R. & Balke, M. (2013b) One hundred and one new species of *Trigonopterus* weevils from New Guinea. *ZooKeys*, 280, 1–150.

Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E. & Crozier, R.H. (2010) Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual Review of Entomology*, 55, 421–438.
https://doi.org/10.1146/annurev-ento-112408-085432

Sharkey, M.J., Janzen, D.H., Hallwachs, W., Chapman, E.G., Smith, M.A., Dapkey, T., Brown, A., Ratnasingham, S., Naik, S., Manjunath, R., Perez, K., Milton, M., Hebert, P., Shaw, S.R., Kittel, R.N., Solis, M.A., Metz, M.A., Goldstein, P.Z., Brown, J.W., Quicke, D.L.J., van Achterberg, C., Brown, B.V. & Burns, J.M. (2021a) 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.figure403

Sharkey, M., Brown, B., Baker, A. & Mutanen, M. (2021b) Response to Zamani *et al.* (2020): The omission of critical data in the pursuit of "revolutionary" methods to accelerate the description of species. *ZooKeys*, 1033, 191–201.
https://doi.org/10.3897/zookeys.1033.66186

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

Srivathsan, A., Hartop, E., Puniamoorthy, J., Lee, W.T., Kutty, S.N., Kurina, O. & Meier, R. (2019) Rapid, large-scale species discovery in hyperdiverse taxa using 1D MinION sequencing. *BMC Biology*, 17, 96.
https://doi.org/10.1186/s12915-019-0706-9

Stankowski, S. & Ravinet, M. (2021) Quantifying the use of species concepts. *Current Biology*, 31, R428–R429.
https://doi.org/10.1016/j.cub.2021.03.060

Sukumaran, J. & Knowles, L.L. (2017) Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 1607–1612.
https://doi.org/10.1073/pnas.1607921114

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 21, 2045–2050.
https://doi.org/10.1111/j.1365-294X.2012.05470.x

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H. & Vogler, A.P. (2003) A plea for DNA taxonomy. *Trends in Ecology and Evolution*, 18, 70–74.
https://doi.org/10.1016/S0169-5347(02)00041-1

Vences, M. (2020) The promise of next-generation taxonomy. *Megataxa*, 1 (1), 35–38.
https://doi.org/10.11646/megataxa.1.1.6

Wiemers, M. & Fiedler, K. (2007) Does the DNA barcoding gap exist?—a case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology*, 4, 8.
https://doi.org/10.1186/1742-9994-4-8

Wu, B., Hussain, M., Zhang, W., Stadler, M., Liu, X. & Xiang, M. (2019) Current insights into fungal species diversity and perspective on naming the environmental DNA sequences of fungi. *Mycology*, 10 (3), 127–140.
https://doi.org/10.1080/21501203.2019.1614106

Yeates, D.K., Seago, A., Nelson, A., Cameron, S.L., Joseph, L. & Trueman, J.W.H. (2011) Integrative taxonomy, or iterative taxonomy? *Systematic Entomology*, 36, 209–217.
https://doi.org/10.1111/j.1365-3113.2010.00558.x

Zamani, A., Vahtera, V., Sääksjärvi, I.E. & Scherz, M.D. (2021) The omission of critical data in the pursuit of 'revolutionary' methods to accelerate the description of species. *Systematic Entomology*, 46, 1–4.
https://doi.org/10.1111/syen.12444