Received: 23 March 2022 | Revised: 6 June 2022 | Accepted: 14 June 2022

DOI: 10.1111/jbi.14463

COMMENTARY

Journal of Biogeography WILEY

Taxonomic uncertainty and the challenge of estimating global species richness

Juliana Stropp[1] | **Richard James Ladle[2,3]** | **Thaise Emilio[4]** | **Thainá Lessa[3]** | **Joaquín Hortal[1,5]**

[1]Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain

[2]Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO-InBIO), Universidade do Porto, Vairão, Portugal

[3]Institute of Biological and Health Sciences, Federal University of Alagoas, Maceió, Alagoas, Brazil

[4]Programa Nacional de Pós-Doutorado (PNPD), Programa de Pós-graduação em Ecologia, Universidade Estadual de Campinas (UNICAMP), Campinas, Brazil

[5]Departamento de Ecologia, Instituto de Ciências Biológicas, Universidade Federal de Goiás (UFG), Goiânia, Brazil

Correspondence
Juliana Stropp, Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (MNCN-CSIC), José Gutiérrez Abascal 2, 28006 Madrid, Spain.
Email: juliana.stropp@gmail.com

Funding information
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award Number: 001; Horizon 2020 Framework Programme, Grant/Award Number: 843234; Instituto Nacional de Ciência e Tecnologia em Ecologia, Evolução e Conservação da Biodiversidade, Grant/Award Number: FAPEG 201810267000023 and MCTIC/CNPq 465610/2014-5; European Union's Horizon 2020, Grant/Award Number: 854248

Handling Editor: Alexander Zizka

Keywords: global species richness, macroecology, species identification, taxonomic uncertainty, taxonomy, tree species

It is a biogeographical trope that after several centuries of exploration, our knowledge of the world's biodiversity is still staggeringly incomplete (Meyer et al., 2016), even for well-studied groups such as trees (Keppel et al., 2021). Thus, when a study publishes a new estimate of the number of known (and unknown) tree species on Earth, it often gains global media attention. Recently, a new study suggested that there were approximately 73,000 tree species worldwide and as much as 14% of the world's tree flora still remains unknown to science (Gatti et al., 2022). But how accurate are such estimates?

Assessments of the global number of tree species vary between 45,000 (table 1 in Fine & Ree, 2006) and 100,000 (Oldfield et al., 1998). Estimates of tree species richness in the Amazonian rainforest range from 14,000 (Cardoso et al., 2017) to over 15,000 species (ter Steege et al., 2020). The disparity in the estimates triggered a scientific debate about the most appropriate data and methods to identify the total number of tree species (Cardoso et al., 2017; Chao & Colwell, 2017; ter Steege et al., 2020). Much progress has been made in understanding the sensitivity of estimates to different mathematical models (e.g. parametric or non-parametric estimators), and the completeness of biodiversity inventories. Surprisingly, there has been little progress in quantifying uncertainty associated with underlying data, specifically related to changes in taxonomy and errors in species identification. Here, we highlight two reasons why understanding taxonomic uncertainty is crucial for estimating the number of known and unknown species in the world.

First, taxonomy does not advance homogeneously across taxa or regions. Until 2015, 15,000 plant species were described for the Brazilian Atlantic forest but only 12,000 for the Brazilian Amazon, a region that is nearly four times larger than the Atlantic forest. In fact, the Brazilian Amazon lags 65 years behind the Atlantic forest in terms of species description (Hopkins, 2019). Moreover, taxonomic effort and reclassification do not affect taxa equally. For several well-studied taxa, such as palms, taxonomic lumping has reduced the number of valid species over time. For example, the 27 species that currently comprise the single palm genus, *Attalea*, were described as 16 different genera and 167 species (!) over the past 200 years (Henderson, 2020) (Figure 1). For other taxa, such as the Neotropical genus *Protium*, taxonomic splitting has increased the number of species (Damasco et al., 2021). Thus, it is questionable whether taxonomic lumping and splitting is a 'zero-sum game'; specifically, at the

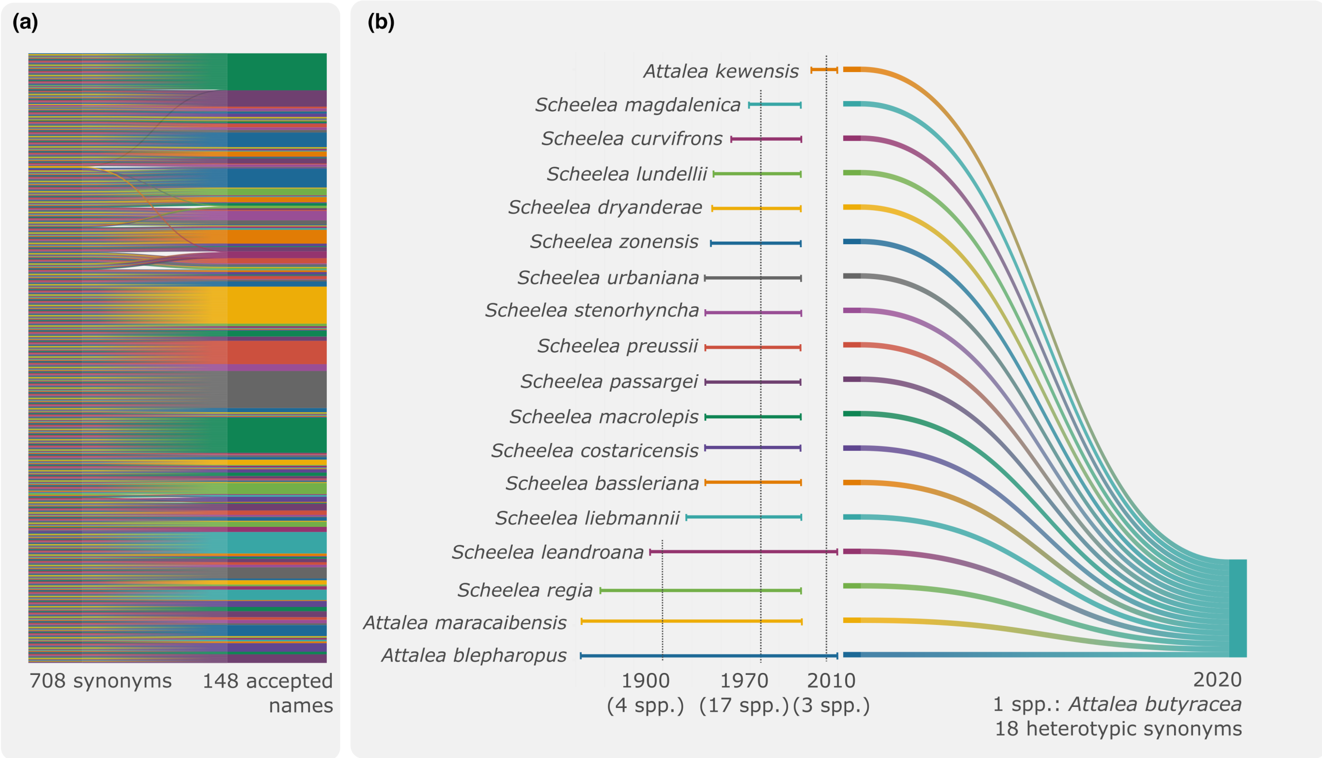

FIGURE 1 Taxonomic change in palms (Arecaceae; Palmae). (a) Shows the taxonomic reclassification of 708 synonyms into 148 accepted names of palm species that occur in Amazonia; each colour represents a species name. (b) Depicts the link between heterotypic synonyms and the accepted name in a timeline of taxonomic lumping for one palm species, *Attalea butyracea*; horizontal coloured lines mark the year of description of 18 heterotypic synonyms and the year of synonymization, with each colour representing a current synonym; vertical lines indicate the counts of accepted species (in parentheses) in a given year; curved lines represent the lumping of 18 heterotypic synonyms into *A. butyracea*. Data sources: For (a), the list of accepted names of palms that occur in Amazonia was extracted from Cardoso et al. (2017) and ter Steege et al. (2019), whereas the list of synonyms was obtained from Govaerts et al. (2022); the list of heterotypic synonyms shown in (b) was obtained from Henderson (2020).

level of individual genera or families. Taxonomic change will continue to impact species numbers, particularly for poorly studied taxa and regions. Given the different stages of taxonomic knowledge for the various plant families, it is difficult to establish robust estimates of species richness, especially when such estimates are based on asymptotic values of extrapolated species accumulation curves or non-parametric species richness estimators (as e.g. in Gatti et al., 2022).

Second, species misidentification is common across taxa and prevalent in temperate and tropical forest plots (Dexter et al., 2010). Yet, the absolute number of misidentified trees is particularly high in floristically diverse tropical forests as many tree families lack recent taxonomic studies that provide, e.g. identification keys. This problem is reflected by tropical herbaria that tend to contain a large number of misidentified vouchers, which are used to match unidentified tree samples (Goodwin et al., 2015). By contrast, temperate forests have lower diversity and a longer taxonomic history and thus their tree flora is less prone to misidentification. Species counts are affected by misidentification in at least two ways. A single species can be incorrectly identified as several species, which inflates the count of "rare" species. Conversely, several rare species can be incorrectly identified as one single species, which inflates the count of "common" species (Dexter et al., 2010). Both ways of misidentification cause spurious variation in the species composition between sampling units and may affect estimates of species richness as calculated, for example, by the Chao2 adjusted estimator (Chao, 1987). This estimator can be adequate for predicting regional species richness from heterogeneous data sources (Hortal et al., 2006). However, to minimise a possible overestimation of species richness, it requires that all unique species in a sample have the same mean detection probability (Chao & Colwell, 2017). Several studies, such as Gatti et al. (2022), partly account for this by estimating true unique species, but this approach assumes that the probability of errors in the recording of unique species is similar across continents—an unrealistic assumption given the large regional discrepancies in the resources and capacity to accurately identify tree species.

A single, universally accepted estimate of the number of species on Earth may never be achieved as it would require each species being unambiguously described, named and allocated to a unique branch in the tree of life based on a single species concept; an out-of-reach goal for taxonomy. Understanding how taxonomic uncertainty affects estimates of species richness continues to be a major challenge that requires collaboration across disciplines. This opens a research avenue with at least three fronts. First, it is necessary to document the taxonomic history of species, that is, describing temporal trends and understanding the drivers of new discoveries, taxonomic lumping, splitting and re-assignments. Second, such historical taxonomic data need to be incorporated into mathematical models of species richness, thereby capturing variations in species numbers that result purely from progress in taxonomy (e.g. Alroy, 2002; Edie et al., 2017). Third, insight from taxonomic history can be used to forecast how future taxonomic explorations, including new discoveries and reclassifications, may reshape the current pattern of species richness (Edie et al., 2017). Tackling these challenges requires a network of taxonomists, macroecologists and data scientists collaborating across multiple taxa in different continents.

Expeditions to remote forests are surely needed to unveil the diversity of the world's tree flora, but uncovering the number of known and unknown tree species also requires bridging the gap between taxonomy and macroecology. In the age of advanced data mining and statistical modelling, reconstructing centuries of taxonomic discoveries may be as rewarding as new field expeditions.

ACKNOWLEDGEMENTS

This work was funded by the European Union's Horizon 2020 research and innovation programme under a Marie Skłodowska-Curie Action led by JS (grant agreement #843234; project TAXON-TIME). RJL is supported by the European Union's Horizon 2020 research and innovation programme (grant agreement #854248). TE and TL are funded by the "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior"—Brazil (CAPES)—Finance Code 001. Additional support has been provided by the "Instituto Nacional de Ciência e Tecnologia em Ecologia, Evolução e Conservação da Biodiversidade" (INCT-EECBio) (MCTIC/CNPq 465610/2014-5; FAPEG 201810267000023). No permits were needed to carry out this research.

DATA AVAILABILITY STATEMENT

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

ORCID

Juliana Stropp https://orcid.org/0000-0002-2831-4066

REFERENCES

Alroy, J. (2002). How many named species are valid? *Proceedings of the National Academy of Sciences USA*, *99*(6), 3706–3711. https://doi.org/10.1073/pnas.062691099

Cardoso, D., Särkinen, T., Alexander, S., Amorim, A. M., Bittrich, V., Celis, M., Daly, D. C., Fiaschi, P., Funk, V. A., & Giacomin, L. L. (2017). Amazon plant diversity revealed by a taxonomically verified species list. *Proceedings of the National Academy of Sciences USA*, *114*(40), 10695–10700. https://doi.org/10.1073/pnas.1706756114

Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, *43*(4), 783–791.

Chao, A., & Colwell, R. K. (2017). Thirty years of progeny from Chao's inequality: Estimating and comparing richness with incidence data and incomplete sampling. *SORT-Statistics and Operations Research Transactions*, *41*, 3–54.

Damasco, G., Baraloto, C., Vicentini, A., Daly, D. C., Baldwin, B. G., & Fine, P. V. A. (2021). Revisiting the hyperdominance of Neotropical tree species under a taxonomic, functional and evolutionary perspective. *Scientific Reports*, *11*(1), 9585. https://doi.org/10.1038/s41598-021-88417-y

Dexter, K. G., Pennington, T. D., & Cunningham, C. W. (2010). Using DNA to assess errors in tropical tree identifications: How often are ecologists wrong and when does it matter? *Ecological Monographs*, *80*(2), 267–286.

Edie, S. M., Smits, P. D., & Jablonski, D. (2017). Probabilistic models of species discovery and biodiversity comparisons. *Proceedings of the National Academy of Sciences USA*, *114*(14), 3666–3671. https://doi.org/10.1073/pnas.1616355114

Fine, P. V. A., & Ree, R. H. (2006). Evidence for a time-integrated species-area effect on the latitudinal gradient in tree diversity. *The American Naturalist*, *168*(6), 796–804. https://doi.org/10.1086/508635

Gatti, R. C., Reich, P. B., Gamarra, J. G. P., Crowther, T., Hui, C., Morera, A., Bastin, J.-F., De-Miguel, S., Nabuurs, G.-J., & Svenning, J.-C. (2022). The number of tree species on Earth. *Proceedings of the National Academy of Sciences USA*, *119*(6), e2115329119. https://doi.org/10.1073/pnas.2115329119

Goodwin, Z. A., Harris, D. J., Filer, D., Wood, J. R. I., & Scotland, R. W. (2015). Widespread mistaken identity in tropical plant collections. *Current Biology*, *25*(22), R1066–R1067. https://doi.org/10.1016/j.cub.2015.10.002

Govaerts, R., Dransfield, J., Zona, S., Hodel, D.R., & Henderson, A. (2022). *World checklist of Arecaceae. Facilitated by the Royal Botanic Gardens, Kew*. http://wcsp.science.kew.org/

Henderson, A. J. (2020). A revision of *Attalea* (Arecaceae, Arecoideae, Cocoseae, Attaleinae). *Phytotaxa*, *444*, 1–76. https://doi.org/10.11646/phytotaxa.444.1.1

Hopkins, M. J. G. (2019). Are we close to knowing the plant diversity of the Amazon? *Anais Da Academia Brasileira de Ciencias*, *91*, e20190396. https://doi.org/10.1590/0001-3765201920190396

Hortal, J., Borges, P. V. A., & Gaspar, C. (2006). Evaluating the performance of species richness estimators: Sensitivity to sample grain size. *Journal of Animal Ecology*, *75*(1), 274–287. https://doi.org/10.1111/j.1365-2656.2006.01048.x

Keppel, G., Craven, D., Weigelt, P., Smith, S. A., van der Sande, M. T., Sandel, B., Levin, S. C., Kreft, H., & Knight, T. M. (2021). Synthesizing tree biodiversity data to understand global patterns and processes of vegetation. *Journal of Vegetation Science*, *32*(3), e13021. https://doi.org/10.1111/jvs.13021

Meyer, C., Weigelt, P., & Kreft, H. (2016). Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters*, *19*(8), 992–1006. https://doi.org/10.1111/ele.12624

Oldfield, S., Lusty, C., & MacKinven, A. (1998). *The world list of threatened trees*. World Conservation Press.

ter Steege, H., Mota de Oliveira, S., Pitman, N. C. A., Sabatier, D., Antonelli, A., Guevara Andino, J. E., Aymard, G. A., & Salomão, R. P. (2019). Towards a dynamic list of Amazonian tree species. *Scientific Reports*, *9*(1), 3501. https://doi.org/10.1038/s41598-019-40101-y

ter Steege, H., Prado, P. I., de Lima, R. A. F., Pos, E., de Souza Coelho, L., de Andrade Lima Filho, D., Salomão, R. P., Amaral, I. L., de Almeida Matos, F. D., Castilho, C. V., Phillips, O. L., Guevara, J. E., de Jesus Veiga Carim, M., Cárdenas López, D., Magnusson, W. E., Wittmann, F., Martins, M. P., Sabatier, D., Irume, M. V., … Pickavance, G. (2020). Biased-corrected richness estimates for the Amazonian tree flora. *Scientific Reports*, *10*(1), 10130. https://doi.org/10.1038/s41598-020-66686-3

How to cite this article: Stropp, J., Ladle, R. J., Emilio, T., Lessa, T., & Hortal, J. (2022). Taxonomic uncertainty and the challenge of estimating global species richness. *Journal of Biogeography*, *49*, 1654–1656. https://doi.org/10.1111/jbi.14463