BIOLOGY LETTERS

royalsocietypublishing.org/journal/rsbl

Opinion piece

Cite this article: Bianchi FM, Gonçalves LT. 2021 Getting science priorities straight: how to increase the reliability of specimen identification? *Biol. Lett.* **17**: 20200874. https://doi.org/10.1098/rsbl.2020.0874

Received: 9 December 2020
Accepted: 8 April 2021

Subject Areas:
taxonomy and systematics

Keywords:
misidentification, public databases, science policy, voucher, taxonomy

Author for correspondence:
Filipe Michels Bianchi
e-mail: bianchi.fm@hotmail.com

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5393787.

THE ROYAL SOCIETY PUBLISHING

Conservation biology

Getting science priorities straight: how to increase the reliability of specimen identification?

Filipe Michels Bianchi[1,2] and Leonardo Tresoldi Gonçalves[3,4]

[1]Laboratório de Entomologia Sistemática, Departamento de Zoologia, [2]Programa de Pós-Graduação em Biologia Animal, and[3]Laboratório de Drosophila, Departamento de Genética, [4]Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

FMB, 0000-0002-5842-8822; LTG, 0000-0002-3593-8424

'We advise the authors to find a native English speaker to proofread the manuscript'. This is a standard feedback journals give to non-native English speakers. Journals are justifiably concerned with grammar but do not show the same rigour about another step crucial to biological research: specimen identification. Surveying the author guidelines of 100 journals, we found that only 6% of them request explicitly citation of the literature used in specimen identification. Authors hamper readers from contesting specimen identification whenever vouchers, identification methods, and taxon concepts are not provided. However, unclear taxonomic procedures violate the basic scientific principle of reproducibility. The scientific community must continuously look for practical alternatives to improve taxonomic identification and taxonomic verification. We argue that voucher pictures are an accessible, cheap and time-effective alternative to mitigate (not abolish) bad taxonomy by exposing preventable misidentifications. Voucher pictures allow scientists to judge specimen identification actively, based on available data. The popularization of high-quality image devices, photo-identification technologies and computer vision algorithms yield accurate scientific photo-documentation, improving taxonomic procedures. Taxonomy is timeless, transversal and essential to most scientific disciplines in biological sciences. It is time to demand rigour in taxonomic identifications.

'We advise the authors to find a native English speaker to proofread the manuscript'. Non-native English speakers commonly receive this feedback during the publication process (although being a native is not the *sine qua non* for academic English proficiency, see [1]). This concern is justified because poor writing may confound readers and overshadow the findings. However, just like words need expert assessment, so does specimen identification. Since species are the fundamental units of biology, accurate specimen identification underpins all biological research. Unlike grammar mistakes which may compromise a sentence's meaning, species misidentification compromises the whole study, weakening scientific integrity. A misidentification triggers a cascade effect in the worst-case scenario, accumulating spurious information around a subject (see examples in [2]).

The 'author guidelines' for most journals describe strict instructions to improve the readability and impact of papers (e.g. English language editing, academic illustration, figure formatting, and graphical abstract design). Surveying submission guidelines and editorial policies of 100 journals from different biological areas and publishers (electronic supplementary material, table S1), we found that most of them (83%) encourage authors to deposit the data from which published results are derived (e.g. trees, scripts, nucleotide sequences) in public databases. Unfortunately, few instructions related to the

taxa *per se* are provided: 32% of the journals require the author to deposit voucher specimens in scientific collections, whereas only 6% request explicitly the literature used to identify the taxa of study. Moreover, Packer *et al.* [3] found that just 50% of the papers complied with vouchering recommendations. This lack of explicit editorial policies raises the question of whether journals are sufficiently committed to providing information on specimen identification.

Concerning the taxonomic treatment of zoological papers, a survey conducted by Monckton *et al.* [4] showed that only 10.7% of papers cited taxonomic identification methods, 6.9% indicated taxon concepts and 29.2% made available vouchers—that is, 70.8% of papers excluded the possibility for double-checking taxonomic identifications permanently. Thus, authors hamper readers from contesting specimen identification whenever vouchers, identification methods and taxon concepts are not provided. Taking a step back, reviewers and editors do not have access to the raw data (i.e. specimens) during the review process for a taxonomic verification. In cases where specimens are vouchered, a researcher may have partial access to the vouchers (e.g. by mail, photographs or visiting the collection where they are deposited), although it demands funding and goodwill. In other words, neither the editor, the reviewers, nor the readers have easy access to the specimens used in the papers to double-check identification.

The lack of supporting information justifying or guaranteeing a careful identification procedure is a usual failure in biological sciences [2] that plays against the basic scientific principle of reproducibility [5]. Thus, suggestions to attenuate the subjectivity of specimen identification and good taxonomic practices have been under debate (e.g. [6–10]). For instance, Meier [11] suggested explicit taxonomic identification methods and taxon concept statements supported by references. Similarly, Bortolus [6] advocated adoption of a 'Taxonomic Verification' section by journals, allowing future taxonomic validation of the specimens under study. These suggestions take little printed-space and can be readily implemented, unless the researcher thinks of specimen identification as an unimportant step—a much bigger problem [5,11]. Besides violating scientific principles, omission to cite taxonomic papers used for specimen identification is also disrespectful to others' intellectual production, an ethical dilemma. Verifying and using the updated literature as a regular procedure for researchers may avoid mistakes beyond the published papers, improving the quality of data deposited in public databases, a key source of information for current research (see also [12,13]).

In the past decades, online databases have expanded and transformed scientists' use of research data [14]. Along with a straightforward process to deposit data and promising sources for users, public databases have their inherent pitfalls. Inadequate taxonomic procedures (e.g. [15–18])—that is, *bad taxonomy sensu* Winston [19]—may now spread misidentifications at a fast rate. For instance, misidentifications and problematic taxonomic meta-data are a recurrent issue in the Global Biodiversity Information Facility (GBIF) [20,21]. Furthermore, in the Barcode of Life Data System (BOLD), misspellings and invalid names may exceed 10% of deposited sequences for a taxon, outnumbering the records with poor-quality sequences and compromising the integrity of databases (see [22]). Published/public misidentifications affect many biological fields, such as conservation [23], invasive species management [24], product traceability [25], and evolutionary biology [26].

Specimen identification requires fundamental expertise since the taxonomic literature may be hostile owing to the technical language and the laboured methodological procedures, even for taxonomists. Thus, the mention of a taxonomist could be expected in papers not strictly taxonomic. But, for example, more than half (62.5%) of studies on community ecology neither have a taxonomist among the authors, nor acknowledge taxonomists for specimen identification or cite taxonomic literature to support taxa identification [2]. It brings into question the quality of the identification of those specimens deposited in public databases owing to the unclear identification process. Again, the end-users are often tried to contest the identification. Most databases provide tools to flag incorrect information and deter bad taxonomy, but this process is possible only when supporting information—such as collection site, voucher pictures or literature used in the identification—is provided.

Similar to how journals request deposition of nucleotide sequences in public databases to enable the reliability and replicability of studies, an additional simple request could improve taxonomic verification: voucher pictures. Images are always powerful allies of taxonomy, and the inclusion of voucher pictures allows a first taxonomic verification in a few mouse-clicks. Researchers should be aware of what (e.g. specimen view, emphasis of structure) maximizes the usefulness of voucher pictures for the taxa under study. As long as the data (e.g. nucleotide sequence, occurrence record) are linked to voucher pictures, the end-user can contest the identification (see [27]) and then opt for using or excluding those data from future analyses.

The popularization of high-quality image devices has changed how science records data, and photographs have been used beyond taxa descriptions in a myriad of different studies. With the current availability of electronic image devices and basic training on specific taxa, accurate scientific photo-documentation can even be conducted in the field by amateur (or nonprofessional) scientists (see [28,29]). One may argue that taking specimen pictures is not feasible for researchers who access large amounts of specimens—but taking scientific photographs is cheaper (most of the times costless) and much less time-consuming than other standard research procedures, such as DNA extraction, preparation for scanning electron microscopy and micro-CT (see examples in [30,31]). Furthermore, time-efficient systems for automatic specimen digitization have been designed to take pictures of thousands of specimens in a few minutes (e.g. [32–34]). Pictures may mitigate (not abolish) bad taxonomy by exposing preventable misidentifications that would not be noted if images were not provided (e.g. BOLD record ASAHE106-12 is identified as *Euschistus tristigmus* (Say) (Hemiptera: Pentatomidae), but it is clearly a non-pentatomid immature [35]).

Presuming that voucher pictures are a panacea for misidentification is a naive position. Photo-identification may be puzzling, sometimes impossible, for reaching lower taxonomic levels (e.g. nematodes [36], insects [37], chiropterans [38]; but see [8]). But another reason to embrace specimen photographs is to develop and popularize photo-identification technologies and computer vision algorithms, already used for diverse taxonomic groups (e.g. [39–42]). Moreover, machine learning and neural networks have improved specimen identification accuracy even using those photos considered of low-quality under human eyes [32]. Undoubtedly, pictures will soon be evaluated by databases' algorithms similarly to how nucleotides sequences are routinely analysed in BLASTn (http://

www.ncbi.nlm.nih.gov/) (e.g. [31]). New possibilities to aid taxonomic identification are welcome; here, we encourage the use of voucher pictures—an overlooked, low-cost and feasible alternative—to lessen the problems mentioned above.

Taxonomy is timeless, transversal and essential to most scientific disciplines in biological sciences and demands immediate strict rigour in taxonomic identifications. Journals should explicitly acknowledge the importance of taxonomic identification and verification, being even tougher than they are on grammar, for instance. We argue that the availability of voucher pictures of specimens used in scientific research would increase reliability and allow identifications to be contested. Our suggestion strengthens the many other practicable alternatives already proposed to cope with this problem (e.g. [9,11,27]). We must continuously look for practical alternatives to improve taxonomic identification and taxonomic verification. Reviewers and editors may encourage authors to make their taxonomic identification clear, replicable and verifiable while pushing forward the proposal of including these directions in the author guidelines of the journals they work for.

Data accessibility. The data are provided in electronic supplementary material.

Authors' contributions. All authors contributed equally to the conception and preparation of the article.

Competing interests. We declare we have no competing interests.

Funding. F.M.B. is supported by a postdoctoral fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (PNPD/CAPES)—Finance code 001; L.T.G. is supported by a doctoral fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Acknowledgements. Dr Christiane Weirauch is acknowledged for thoughtful comments and discussions that improved this paper. We also acknowledge three anonymous reviewers and the Biology Letters' Editorial Board for their extensive efforts to improve this manuscript. We also thank Dr Kaleigh Amanda Russell—from Riverside, California, USA—for reviewing the paper.

References

1. Romero-Olivares AL. 2019 Reviewers, don't be rude to nonnative English speakers. *Science*. (doi:10.1126/science.caredit.aaz7179)
2. Bortolus A. 2008 Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. *Ambio* **37**, 114–118. (doi:10.1579/0044-7447(2008)37[114:ecitbs]2.0.co;2)
3. Packer L, Monckton SK, Onuferko TM, Ferrari RR. 2018 Validating taxonomic identifications in entomological research. *Insect Conserv. Divers.* **11**, 1–12. (doi:10.1111/icad.12284)
4. Monckton SK, Johal S, Packer L. 2020 Inadequate treatment of taxonomic information prevents replicability of most zoological research. *Can. J. Zool.* **98**, 633–642. (doi:10.1139/cjz-2020-0027)
5. Vink CJ, Paquin P, Cruickshank RH. 2012 Taxonomy and irreproducible biological science. *Bioscience* **62**, 451–452. (doi:10.1525/bio.2012.62.5.3)
6. Bortolus A. 2012 Guiding authors to reliably use taxonomic names. *Trends Ecol. Evol.* **27**, 418. (doi:10.1016/j.tree.2012.05.003)
7. Bortolus A. 2012 Good habits come first in science too: a reply to Straka and Starkzomski. *Trends Ecol. Evol.* **27**, 655. (doi:10.1016/j.tree.2012.08.016)
8. Straka JR, Starzomski BM. 2012 Reply to Bortolus: what's in a name? *Trends Ecol. Evol.* **27**, 654. (doi:10.1016/j.tree.2012.08.003)
9. Zeppelini D *et al*. 2021 The dilemma of self-citation in taxonomy. *Nat. Ecol. Evol.* **5**, 2. (doi:10.1038/s41559-020-01359-y)
10. Troudet J, Vignes-Lebbe R, Grandcolas P, Legendre F. 2018 The increasing disconnection of primary biodiversity data from specimens: how does it happen and how to handle it? *Syst. Biol.* **67**, 1110–1119. (doi:10.1093/sysbio/syy044)
11. Meier R. 2017 Citation of taxonomic publications: the why, when, what and what not. *Syst. Entomol.* **42**, 301–304. (doi:10.1111/syen.12215)
12. Pleijel F, Jondelius U, Norlinder E, Nygren A, Oxelman B, Schander C, Sundberg P, Thollesson M. 2008 Phylogenies without roots? A plea for the use of vouchers in molecular phylogenetic studies. *Mol. Phylogenet. Evol.* **48**, 369–371. (doi:10.1016/j.ympev.2008.03.024)
13. Wägele H, Klussmann-Kolb A, Kuhlmann M, Haszprunar G, Lindberg D, Koch A, Wägele JW. 2011 The taxonomist - an endangered race. A practical proposal for its survival. *Front. Zool.* **8**, 25. (doi:10.1186/1742-9994-8-25)
14. Imker HJ. 2018 25 years of molecular biology databases: a study of proliferation, impact, and maintenance. *Front. Res. Metrics Anal.* **3**, 18. (doi:10.3389/frma.2018.00018)
15. Hofstetter V, Buyck B, Eyssartier G, Schnee S, Gindro K. 2019 The unbearable lightness of sequenced-based identification. *Fungal Divers.* **96**, 243–284. (doi:10.1007/s13225-019-00428-3)
16. Valkiūnas G, Atkinson CT, Bensch S, Sehgal RNM, Ricklefs RE. 2008 Parasite misidentifications in GenBank: how to minimize their number? *Trends Parasitol.* **24**, 247–248. (doi:10.1016/j.pt.2008.03.004)
17. Tran PN, Savka MA, Gan HM. 2017 *In-silico* taxonomic classification of 373 genomes reveals species misidentification and new genospecies within the genus *Pseudomonas*. *Front. Microbiol.* **8**, 1296. (doi:10.3389/fmicb.2017.01296)
18. Stavrou AA, Mixão V, Boekhout T, Gabaldón T. 2018 Misidentification of genome assemblies in public databases: the case of *Naumovozyma dairenensis* and proposal of a protocol to correct misidentifications. *Yeast* **35**, 425–429. (doi:10.1002/yea.3303)
19. Winston JE. 1999 *Describing species: practical taxonomic procedure for biologists*. New York, NY: Columbia University Press.
20. Samy G, Chavan V, Ariño AH, Otegui J, Hobern D, Sood R, Robles E. 2013 Content assessment of the primary biodiversity data published through GBIF network: status, challenges and potentials. *Biodivers. Informatics* **8**, 94–172. (doi:10.17161/bi.v8i2.4124)
21. Zizka A *et al*. 2020 No one-size-fits-all solution to clean GBIF. *PeerJ* **8**, e9916. (doi:10.7717/peerj.9916)
22. Bianchi FM, Gonçalves LT. In press. Borrowing the Pentatomomorpha tome from the DNA Barcode library: scanning the overall performance of *cox1* as a tool. *J. Zool. Syst. Evol. Res.* (doi:10.1111/jzs.124_76)
23. Vogel Ely C, Bordignon SA de L, Trevisan R, Boldrini II. 2017 Implications of poor taxonomy in conservation. *J. Nat. Conserv.* **36**, 10–13. (doi:10.1016/j.jnc.2017.01.003)
24. Ng TH, Tan SK, Wong WH, Meier R, Chan S-Y, Tan HH, Yeo DCJ. 2016 Molluscs for sale: assessment of freshwater gastropods and bivalves in the ornamental pet trade. *PLoS ONE* **11**, e0161130. (doi:10.1371/journal.pone.0161130)
25. Almerón-Souza F, Sperb C, Castilho CL, Figueiredo PICC, Gonçalves LT, Machado R, Oliveira LR, Valiati VH, Fagundes NJR. 2018 Molecular identification of shark meat from local markets in southern Brazil based on DNA barcoding: evidence for mislabeling and trade of endangered species. *Front. Genet.* **9**, 138. (doi:10.3389/fgene.2018.00138)
26. Rutherford S, Wilson PG, Rossetto M, Bonser SP. 2015 Phylogenomics of the green ash eucalypts (Myrtaceae): a tale of reticulate evolution and misidentification. *Aust. Syst. Bot.* **28**, 326–354. (doi:10.1071/SB15038)
27. Santos AM, Branco M. 2012 The quality of name-based species records in databases. *Trends Ecol. Evol.* **27**, 6–7. (doi:10.1016/j.tree.2011.10.004)
28. Aristeidou M, Herodotou C, Ballard HL, Young AN, Miller AE, Higgins L, Johnson RF. 2021 Exploring the participation of young citizen scientists in scientific research: the case of iNaturalist. *PLoS ONE* **16**, e0245682. (doi:10.1371/journal.pone.0245682)
29. Wittmann J, Girman D, Crocker D. 2019 Using iNaturalist in a coverboard protocol to measure data quality: suggestions for project design. *Citiz. Sci. Theory Pract.* **4**, 21. (doi:10.5334/cstp.131)

30. Lytle DA *et al*. 2010 Automated processing and identification of benthic invertebrate samples. *J. North Am. Benthol. Soc.* **29**, 867–874. (doi:10.1899/09-080.1)
31. Valan M, Makonyi K, Maki A, Vondráček D, Ronquist F. 2019 Automated taxonomic identification of insects with expert-level accuracy using effective feature transfer from convolutional networks. *Syst. Biol.* **68**, 876–895. (doi:10.1093/sysbio/syz014)
32. Ärje J *et al*. 2020 Automatic image-based identification and biomass estimation of invertebrates. *Methods Ecol. Evol.* **11**, 922–931. (doi:10.1111/2041-210X.13428)
33. Itaki T, Taira Y, Kuwamori N, Maebayashi T, Takeshima S, Toya K. 2020 Automated collection of single species of microfossils using a deep learning–micromanipulator system. *Prog. Earth Planet. Sci.* **7**, 19. (doi:10.1186/s40645-020-00332-4)
34. Gutiérrez-Larruscain D, Santos-Vicente M, Anderberg AA, Rico E, Martínez-Ortega MM. 2018 Phylogeny of the *Inula* group (Asteraceae: Inuleae): evidence from nuclear and plastid genomes and a recircumscription of *Pentanema*. *Taxon* **67**, 149–164. (doi:10.12705/671.10)
35. Yonke TR. 1991 Order Hemiptera. In *Immature insects: volume II* (ed. FW Stehr), pp. 22–65. Dubuque, IA: Kendall/Hunt Publishing Company.
36. Gonçalves LT, Bianchi FM, Deprá M, Calegaro-Marques C. 2021 Barcoding a can of worms: testing *cox1* performance as a DNA barcode of Nematoda. *Genome* **11**, 922–931. (doi:10.1139/gen-2020-0140)
37. Castro-Huertas V, Forero D, Grazia J. 2020 Delicate and diverse: a taxonomic monograph with a phylogenetic analysis of the Neotropical genus *Ghilianella* Spinola (Hemiptera: Reduviidae: Emesinae). *Zootaxa* **4879**, 1–194. (doi:10.11646/zootaxa.4879.1.1)
38. Foley NM, Goodman SM, Whelan CV, Puechmaille SJ, Teeling E. 2017 Towards navigating the Minotaur's labyrinth: cryptic diversity and taxonomic revision within the speciose genus *Hipposideros* (Hipposideridae). *Acta Chiropterol.* **19**, 1–18. (doi:10.3161/15081109ACC2017.19.1.001)
39. Song C, Lin X-L, Wang Q, Wang X-H. 2018 DNA barcodes successfully delimit morphospecies in a superdiverse insect genus. *Zool. Scr.* **47**, 311–324. (doi:10.1111/zsc.12284)
40. Caorsi VZ, Santos RR, Grant T. 2012 Clip or snap? An evaluation of toe-clipping and photo-identification methods for identifying individual southern red-bellied toads, *Melanophryniscus cambaraensis*. *South Am. J. Herpetol.* **7**, 79–84. (doi:10.2994/057.007.0210)
41. Carranza-Rojas J, Goeau H, Bonnet P, Mata-Montero E, Joly A. 2017 Going deeper in the automated identification of herbarium specimens. *BMC Evol. Biol.* **17**, 181. (doi:10.1186/s12862-017-1014-z)
42. Little DP *et al*. 2020 An algorithm competition for automatic species identification from herbarium specimens. *Appl. Plant Sci.* **8**, e11365. (doi:10.1002/aps3.11365)