

(308–310) Proposals to permit DNA sequence data to serve as types of names of fungi

David L. Hawksworth,¹ David S. Hibbett,² Paul M. Kirk³ & Robert Lücking⁴

1 *Departamento de Biología Vegetal II, Facultad de Farmacia, Universidad Complutense de Madrid, Plaza Ramón y Cajal, Madrid 28040, Spain; Department of Life Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, U.K.; Comparative Plant and Fungal Biology, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, U.K.*

2 *Biology Department, Clark University, 950 Main Street, Worcester Massachusetts 01610, U.S.A.*

3 *State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China; Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, U.K.*

4 *Botanischer Garten und Botanisches Museum Berlin, Freie Universität Berlin, Königin-Luise-Straße 6–8, 14195 Berlin, Germany*

Author for correspondence: David L. Hawksworth, d.hawksworth@kew.org; d.hawksworth@nhm.ac.uk

DOI <http://dx.doi.org/10.12705/654.31>

Numerous novel fungi are being discovered among molecular sequences recovered from environmental samples, from the ranks of phylum (Jones & al. in *Nature* 474: 200–203. 2011) and class (Rosling & al. in *Science* 333: 876–879. 2011) down to species (e.g., Hinchcliff & al. in *Proc. Natl. Acad. Sci. U.S.A.* 112: 12764–12769. 2015; Grantham & al. in *PLoS ONE* 19: e122105. 2015; Nilsson & al. in *Mycology* 12: 29–40. 2016). In 2012, there were 43,290 sequences not named to species in GenBank (<http://www.ncbi.nlm.nih.gov/genbank>), of which 11,429 were not even named to genus (Schoch in *Biodivers. & Conservation* 21: 2425–2433. 2012). The problem is most dramatic in the case of voucherless sequences obtained from next-generation sequencing of environmental samples, such as SFF (Roche 454), Illumina Native, Illumina SRF, AB SOLiD Native, and AB SOLiD SRF (O'Brien & al. in *Appl. Environ. Microbiol.* 71: 5544–5550. 2005; Taylor & al. in *Molec. Ecol. Resources* 8: 742–752. 2008; Buée & al. in *New Phytol.* 184: 449–456. 2009; Lumini & al. in *Environ. Microbiol.* 12: 2165–2179. 2010; Unterseher & al. in *Molec. Ecol.* 20: 275–285. 2011; Dai & al. in *Canad. J. Microbiol.* 58: 81–92. 2012; McGuire & al. in *Microbial Ecol.* 63: 804–812. 2012; Shokralla & al. in *Molec. Ecol.* 21: 1794–1805. 2012). In 2014, the Sequence Read Archive SRA (<http://www.ncbi.nlm.nih.gov/sra>) had amassed over 12 million sequence reads of the fungal ITS barcoding locus, from only about 700 biosamples corresponding to fewer than 50 studies

performed over just five years, compared to fewer than 1 million voucher-based fungal ITS sequences accumulated in GenBank over 25 years (Lücking in *IMC10 Book of Abstracts: O 8.6.1, Abstract ID ABS0123*; http://www.fabinet.up.ac.za/newsitem/112-IMC10_eBook_of_Abstracts.pdf). This corresponds to a 100:1 ratio in sequence data generation, a ratio likely to further increase. SRA sequence reads assigned to the recently described genus *Archaeorhizomyces* Rosling & T.Y. James, with two formally recognized species (Rosling & al. in *Science* 333: 876–879. 2011; Menkis & al. in *Fungal Biol.* 118: 943–955. 2014) suggest the existence of hundreds of undescribed taxa in this clade (Lücking, l.c.; Smith & Lücking, unpub. data), and other clades of ecologically cryptic fungi appear to show similar patterns (Jones & al., l.c.). These taxa require scientific names in order to facilitate communication about them. Under the current *Code* (McNeill & al. in *Regnum Veg.* 154. 2012), such lineages cannot be formally named in the absence of any physical material attributable to a given sequence (either dried specimens or cultures preserved in a metabolically inactive state) or illustration that can serve as the holotype. This is contrary to the objective of the *Code*, which aims to provide a stable system of applying names (Pre. 1) to all algae, fungi, and plants where they are required.

How this issue should be handled under the *Code* has become an increasing concern among mycologists (Hibbett & al. in *Fungal Biol.*

Rev. 25: 38–47. 2011; Hawksworth & al. in IMA Fungus 2: 105–112. 2011; Hibbett in Science 351: 1150–1151. 2016). The issue was mentioned repeatedly during the 10th International Mycological Congress (IMC10) in Bangkok in 2014 and specifically addressed in presentations (e.g., Lücking, l.c.), but time did not permit it to be discussed in the Nomenclature Sessions held during that Congress; nevertheless, 44% of Congress members voting supported the concept of naming of such taxa (Redhead & al. in IMA Fungus 5: 449–462. 2014). We suspect that percentage would now be greater in view of the papers that have since appeared, and continue to appear, showing the scale of the problem.

We do not consider it an option to let this issue drift. Authors are already free to use any characters, including molecular sequences, in diagnoses, but are precluded from allocating names to environmental sequences obtained through voucherless sequencing techniques by the inability to designate a physical holotype corresponding to particular sequence data. The current *Code* is, therefore, failing to meet the needs of the mycological community.

DNA sequence data have already been used in rare cases as sole diagnostic characters, even if a physical environmental specimen from which the sequence was obtained was available as type, to overcome the current nomenclatural constraints (Kirk in Index Fung. 1: 1. 2012). This approach is not ideal, as recovery of the sequence from the material in the long term, i.e., validation of the diagnostic characters, cannot be guaranteed. However, the same problem already exists with other ephemeral characters, such as the oil bodies in *Hepaticae* (von Konrat & al. in PhytoKeys 8: 13–36. 2012), which are not technically excluded by the *Code* to serve as a diagnostic feature.

The *Code* does not prohibit the use of any category of characters for the separation of taxa, that being a matter of taxonomy and not of nomenclature; thus DNA sequence data as a sole diagnostic character are acceptable. The situation encountered with voucherless environmental sequence data therefore needs to be addressed. In order to remedy this, we propose that DNA sequence data alone should be permissible as types for fungi when no physical specimen is available for technical reasons.

(308) Insert a new paragraph after Art. 8.5 as follows:

“8.6. In fungi, when DNA sequence data corresponding to a new taxon have been detected, but no physical specimen has been found

to serve as the type of the name of the new taxon (Art. 8.1–8.4), the type may be composed of DNA sequence data deposited in a public repository.”

(309) Add a new Recommendation 8C:

“8C.1. When the type is composed only of DNA sequence data (Art. 8.6), the new taxon should be described with reference to a published phylogenetic analysis; both the phylogenetic tree and the DNA sequence alignment that was used to create the phylogenetic tree should be deposited in a publicly accessible repository.”

“8C.2. A new taxon typified only by DNA sequence data should be represented by multiple sequences obtained in independent studies, of which one is designated as the holotype.”

“8C.3. DNA sequence data used for typification should be drawn from the molecular regions that are appropriate for delimiting species, based on prevailing best practices as determined by the relevant taxonomic communities.”

(310) Amend Art. 9.1 as follows:

“9.1. A holotype of a name of a species or infraspecific taxon is the one specimen, **or sequence (Art. 8.6)**, or illustration (but see Art. 40.4) used by the author, or designated by the author as the nomenclatural type. As long as the holotype is extant, it fixes the application of the name concerned (but see Art. 9.15).”

Should this suite of proposals pass, the Editorial Committee will need to consider making small changes in Art. 40, in particular in Art. 40.2, 40.3, 40.4, and perhaps 40.5, to ensure consistency with them, as well as adding a “but see” reference in Art. 8.1. In addition, the Nomenclature Section may consider it desirable to add “algae” to these proposals, in line with the special provisions for cultures in Art. 8.4.

Acknowledgements

We are most grateful to Nicholas J. Turland and John H. Wiersema for their assistance in finalizing these proposals. This contribution was prepared while D.L.H. was in receipt of funds from Spanish Ministerio de Economía y Competitividad project CGL2014-55542-P. R.L. is indebted to Brian Eugene Smith (Field Museum, Chicago) for analytical work on SRA sequence read data for the *Archaeorhizomyces* project.