COMMENTARY

New light on names and naming of dark taxa

Martin Ryberg[1], R. Henrik Nilsson[2,3]

I *Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden* **2** *Department of Biological and Environmental Sciences, University of Gothenburg, Box 463, 405 30 Göteborg, Sweden* **3** *Gothenburg Global Biodiversity Centre, Box 461, 405 30 Göteborg, Sweden*

Corresponding author: *Martin Ryberg* (martin.ryberg@ebc.uu.se)

Abstract

A growing proportion of fungal species and lineages are known only from sequence data and cannot be linked to any physical specimen or resolved taxonomic name. Such fungi are often referred to as "dark taxa" or "dark matter fungi". As they lack a taxonomic identity in the form of a name, they are regularly ignored in many important contexts, for example in legalisation and species counts. It is therefore very urgent to find a system to also deal with these fungi. Here, issues relating to the taxonomy and nomenclature of dark taxa are discussed and a number of questions that the mycological community needs to consider before deciding on what system/s to implement are highlighted.

Keywords

Introduction

The first step in wisdom is to know the things themselves; this notion consists in having a true idea of the objects; objects are distinguished and known by classifying them methodically and giving them appropriate names. Therefore, classification and name-giving will be the foundation of our science.

Linnaeus (1735)

Public DNA sequence databases abound with fungal entries that defy all attempts at taxonomic identification. These poorly understood lineages are known from more or less all imaginable substrates and environments, including soil, wood and water but also spacecraft, tumuli and residential areas (e.g. Nilsson et al. 2016). They have been referred to as "dark taxa" or "dark matter fungi" (Page 2016; Grossart et al. 2016) and most of them likely represent undescribed taxa (Tedersoo and Smith 2017), while a limited proportion presumably represents described, but never before sequenced, taxa (cf. Nagy et al. 2011). The number of dark taxa is growing rapidly in the wake of the increasing use of sequence-based approaches to characterisation of biodiversity. Today, few researchers would contest that dark taxa merit both scientific and societal interest, and there is a growing need to classify and communicate these taxa and to record and accumulate data on them. Biodiversity data collections and legislation, in sharp contrast, are largely centred on names of species and higher groups. Being nameless, the dark taxa are not easily incorporated into many of these contexts and are consequently often omitted. They are, for example, usually left out of species counts of areas and therefore from decisions on nature conservation. This unsatisfactory situation where much of our resources and infrastructure cannot properly accommodate dark taxa has spurred a debate in the life sciences on how they should be handled (Samyn and De Clerck 2012; Patterson 2014).

The mycological community has been aware of the problem of dark taxa for a long time (Nilsson et al. 2005; Porter et al. 2008; Ryberg et al. 2008; Hibbett et al. 2009) and several solutions have been proposed. The UNITE database for molecular identification of fungi, for instance, has presented a system where ITS sequence clusters are presented as species hypotheses tagged with unique digital object identifiers (DOIs) to facilitate unambiguous communication (Kõljalg et al. 2016), while Hawksworth et al. (2016) suggested that it should be possible to publish valid names under the International Code of Nomenclature for algae, fungi and plants (ICN) using a DNA sequence, instead of a physical specimen, as type. However, there is no consensus on how the dark taxa should be handled and mycology clearly faces a number of extremely difficult questions in the very near future. These questions will dictate how we refer to, name and to some extent identify, fungi – in short, how mycology is done. The pivotal nature of the pending decisions suggests that all perspectives and points of view must be brought to the surface and vetted thoroughly. Yet it is felt, when following the debate, that this has not been the case so far. In this opinion piece, the authors wish to clarify several overlooked and perhaps obscured aspects of dark taxa and their naming. While a stand will not be taken regarding whether DNA sequences should suffice as species types, it is hoped that several matters that should be resolved will be identified before such a decision can be made. This commentary's contribution is meant for mycologists at large, because dark taxa concern mycology at large. Through longstanding work with difficult-to-identify fungal lineages from soil and other environmental samples, the authors can attest to the frequency and the widespread nature of dark taxa across the fungal tree of life. Disregarding these un-named taxa for the

simple reason that they lack a formal name would be a severe, costly mistake – and one that has already impeded mycology for far too long. This is an issue whose resolution can no longer be postponed.

Nomenclature and taxonomy

In the context of dark taxa, the distinction between taxonomy (delimiting and characterising taxa) and nomenclature (naming and, to some extent, diagnosing taxa) must be stressed. While these are connected and commonly discussed together, blurring their distinction, they are fundamentally different (de Queiroz 2006). The ICN only governs taxon names and the process of naming taxa; it is by design essentially silent on the processes of characterising and delimiting taxa. A name can therefore be the correct name for a taxon according to the code, even if the taxon delimitation/characterisation with which it was published does not suit the taxon. In analogy, a taxon description can be suitable for a taxon even if the name with which it is published is not the correct name or even a validly published name.

According to the ICN, only a validly published name can be considered for the correct name of a taxon (the name that should be used for the taxon). Two of the requirements for a name to be validly published are that it is published with a diagnosis or description and that a type is designated. The diagnosis can be based on molecular characters (Tripp and Lendemer 2014; Sheikh et al. 2017), while the type of a species name needs to be a physical specimen or exceptionally, an illustration of a specimen. Names of taxa of higher rank have a taxon at the immediate lower rank as their type and thus also refer back to a type specimen in the end. The type indicates what taxon a name refers to, and all validly published names whose type is included in the taxon should be considered for that taxon. Determining which name goes where is sometimes made more complicated by the fact that types may be difficult to get hold of or may even be missing altogether for some names. In other cases, the type may not manifest all characters needed to clarify to which taxon it belongs. When the type is missing or lacking important characters, new or additional types can be designated (neo- and epitypes, respectively; Ariyawansa et al. 2014). The basic principle to establish the correct name for a taxon is that the first validly published name should be used, although there are exceptions to this rule. Given that all validly published names must be considered for a taxon, an inflation in names that are difficult to interpret or apply may hamper taxonomy more than an inflation in species descriptions without valid names.

Since the description of taxa is only controlled by the regular scientific principles for publishing, taxa may be described and diagnosed based solely on molecular characters. Such taxa can also be given a name (De Beer et al. 2016), although such names cannot be considered validly published unless there is a specimen to serve as type. This is what the Hawksworth et al. (2016) proposal seeks to amend. The issue with dark taxa is thus not that they cannot be described, but that they cannot be given formal names for communication.

What are names for?

Names are used for communicating objects and concepts. Unless an object or concept is very straightforward indeed, the lack of a name is a major obstacle in its communication and may – implicitly or not – be taken to mean that its communication is not necessary to begin with. This is a general societal issue, but it certainly pertains to mycology as well. Many newly described taxa have, in fact, been represented in DNA sequence databases for 10 years or more before they piqued somebody's interest or were possible to typify according to the ICN (see examples in Nilsson et al. 2016). Upon closer inspection, several of these taxa were found to be both ubiquitous and of significant taxonomic and ecological interest (Rosling et al. 2011; James and Seifert 2017). There is, thus, data to suggest both that the lack of names for dark taxa have retarded progress on their study and that there is ample reason to study and communicate dark taxa in the first place.

Communication of taxa includes aspects such as incorporation into biodiversity datasets, sequence repositories and legislation but also regular scientific and societal communication. What type of name to use will depend on the identity of the communicating parties. For computers, accession numbers such as DOIs will suffice for communication. DOIs are, however, less suitable for human communication (e.g. http://dx.doi.org/10.15156/BIO/SH004915.07FU versus *Vishniacozyma victoriae*). For scientists, latinised binomials may work well, while society at large may prefer vernacular names.

For efficient communication, it is important to consider that one taxon should have only one name and that any name should refer to only one taxon. One of the major purposes of the ICN is to ensure and uphold these relations, while vernacular names are not governed by and, indeed, often violate such rules. The use of parallel naming systems does little to facilitate unambiguous naming in biological systematics. For instance, one and the same name can be the correct or valid name for different taxa under the ICN and the International Code for Zoological nomenclature at the same time, e.g. *Erica* in Ericaceae (Viridiplantae) and Arachnida (Metazoa). For ambiregnal taxa, different names may be the correct/valid name under different nomenclature codes. For instance, a dinoflagellate genus was named *Phalacroma* under the ICN, a name that subsequently was found to be already occupied under the zoological code. Thus, the name *Prodinophysis* was introduced for the same genus (Taylor et al. 1987). In the case where different names are used for the same taxon, databases can link the different uses of names across the systems.

Delimitation of taxa

Descriptions of taxa may be based on different sets of characters, for example sexual or asexual reproductive structures, physiological parameters or DNA bases. It may therefore be difficult to tell whether a taxon, described based on one type of character, is the same as a taxon described from another character type or set of characters. This is the basis behind both the former dual system of naming for "Eumycota" vis-à-vis "Deuteromycota" and the situation which is now faced with dark taxa. In the case of

dark taxa, it is not immediately clear how to correlate a species delimited from environmental sequence data to, say, a range of physiological parameters quantified in the lab or a handful of morphological traits gleaned from microscopy studies of soil samples. Obtaining such additional data and mapping them to individual species will not be straightforward from heterogeneous, mixed-species substrates such as soil and water, but emerging single-cell techniques (e.g. Castelle et al. 2015) offer promise in this regard. As overlapping character sets gradually become available, improved understanding of the underlying taxon will follow piece by piece and the correct name can eventually be assigned according to the nomenclature code. In the context of "Eumycota" and "Deuteromycota", molecular data are often used to link the teleomorph and anamorph stages of species, thus resolving the issue (e.g. Piątek et al. 2017). This is also the reason why sequence data from type specimens are very valuable to sort out nomenclature and DNA barcoding issues and to bring dark taxa under the realm of taxonomy by providing them with a name (Robbertse et al. 2017; Torres-Cruz et al. 2017).

As taxon delimitation and naming are two different things, another complication in what a name refers to is that different taxonomists may advocate different circumscriptions of taxa while the name itself is determined by the ICN. In these cases, the same name may be the correct name for different taxa, with little overlap in the underlying organisms. Furthermore, a name can be correct for some taxon or a synonym of another name depending on the specific taxonomy. If it is required that taxa be monophyletic, then changes in the taxonomy should be expected due to changes in the understanding of evolutionary relationships. Even if this stabilises with time as better estimates of evolutionary relationships are obtained, there may still be conflicts as to what clades are considered as taxa and at what taxonomic level. For example, Hibbett et al. (2007) treat Monoblepharidomycetes as a class in Chytridiomycota while Powell and Letcher (2014) classify it in the monotypic phylum Monoblepharidomycota. Similarly, the small genus *Entorrhiza* is variously recognised as a basidiomycete lineage or as a separate phylum depending on what resource is turned to (Bauer et al. 2015).

The species level is often viewed as a separate evolutionary lineage of special standing (Mayden 1997; but see, e.g. Baum 2009), but there will still be disagreement on how species are delimited. Any species delimitation will always be a hypothesis and different lines of evidence may disagree as to which hypothesis is best supported. Although molecular data provide significant explanatory power in systematics and taxonomy, their use is not devoid of complications. Clustering of sequences into operational taxonomic units, for instance, depends on the choice of clustering algorithm and the parameter settings used (e.g. selection of sequence similarity cut-off) as well as the choice of genetic marker and the individual sequences to be clustered. Thus, equating a sequence-derived operational taxonomic unit with a species is problematic (Schoch et al. 2012; Ryberg 2015). A sequence-derived operational taxonomic unit may, nevertheless, be a species hypothesis.

Without a reference to which taxonomy is employed, what is referred to by a name is more or less ambiguous. The UNITE species hypothesis system provides an unambiguous way to refer to sets of sequences at approximately the species level and additions and removals of sequences to those species hypotheses can be traced back in time

(e.g. https://unite.ut.ee/bl_forw_sh.php?sh_name=SH181628.07FU). However, this approach is limited to sequences included in the underlying dataset, the given set of hypotheses and to taxa represented by ITS sequence data in the first place. Changes between taxonomies are a part of the progress in understanding nature and only scientific advances, together with a dialogue amongst scientists to arrive at a consensus, can resolve this problem. However, without some sort of names for communication such progress seems difficult.

Outlook

The number of dark taxa increases with more or less every new metabarcoding study, but the pace at which these taxa are formalised is many orders of magnitude slower (James and Seifert 2017). This hints at an untenable situation and it is becoming increasingly clear that a system to handle dark taxa in the context of taxonomy, nomenclature and biodiversity at large is needed. The authors plea for the adoption of dark taxa into regular mycology and argue for an expedient establishment of a system or approach to handle dark taxa in mycology and elsewhere. When constructing such a system, many urgent questions present themselves. What are being communicated: sequence clusters, taxa at an undefined level or taxa as recognised by the ICN? Who are communicating: is it computers, scientists, the society at large or any combination of these? If a system with computer-facilitated communication on sequence clusters is wanted, UNITE already fills this role. If a system with a flexible set of taxonomic hypotheses that consider more than just ITS sequence distances is also wanted, something more is needed. For such a system, the impact on biodiversity research should also be considered. Will it encourage and/or deter research and, if so, what kind of research? Should it, unlike the traditional nomenclature codes, encourage best practices in taxonomy? If so, will it achieve these best practices or will it engender poor practices and increased confusion? There is clearly a risk that allowing sequences as types will inflate the number of (rogue?) names and serve to hamper taxonomy in the end (cf. Seifert 2018). As sequences are fundamentally different from a physical specimen, consideration should also be given as to how well sequences will serve as types. It is true that they are perfectly well amenable to digitalisation and that they are easy to share and compare. At the same time, they contain little additional information if further taxonomic resolution is needed and may increase the need for epitypes. Finally, it should be asked if the aim is a separate, DNA-based system for the dark taxa or an integrated system including other characters and taxa too. If the aim is comparable units and not the creation of disjunct taxonomic systems, one system is certainly recommended.

These questions are urgent, because dark taxa permeate mycology and the fungal tree of life, and *ad hoc* names are being used to communicate them without any system to ensure stability of those names (De Beer et al. 2016; Tedersoo and Smith 2017). The answers to the above questions are not immediately clear and they may furthermore differ depending on personal perspectives. A heated debate is therefore expected, perhaps without any consensus at the end. Whatever system is implemented, there needs to be an active discussion in the mycological community as to what confidence should

be required for the named taxa and whether a specific system should be implemented to safeguard these quality measures (cf. Tedersoo et al. 2015).

Seifert (2018) asks what is the point of a name when there is no additional information attached to it. At the same time, he gives a species the name "the brain fungus" (not valid according to the ICN) to be able to talk about it and makes a plea for more information on it. However, without a stable and precise name as an identifier, it will be difficult to accumulate precise knowledge about exactly this species. A name is not an end to our understanding of a taxon, but a means and a beginning. If mycology is to be the study of all fungi and not just the perhaps < 10 % which can be readily observed, then dark taxa should be welcomed into the light.

Acknowledgements

References

Ariyawansa HA, Hawksworth DL, Hyde KD, Jones EG, Maharachchikumbura SS, Manamgoda DS, Thambugala KM, Udayanga D, Camporesi E, Daranagama A, Jayawardena R (2014) Epitypification and neotypification: guidelines with appropriate and inappropriate examples. Fungal Diversity 69(1): 57–91. https://doi.org/10.1007/s13225-014-0315-4

Bauer R, Garnica S, Oberwinkler F, Riess K, Weiß M, Begerow D (2015) Entorrhizomycota: a new fungal phylum reveals new perspectives on the evolution of fungi. PLoS One 10(7): e0128183. https://doi.org/10.1371/journal.pone.0128183

Baum DA (2009) Species as ranked taxa. Systematic Biology 58: 74–86. https://doi.org/10.1093/sysbio/syp011

Castelle CJ, Wrighton KC, Thomas BC, Hug LA, Brown CT, Wilkins MJ, Frischkorn KR, Tringe SG, Singh A, Markillie LM, Taylor RC (2015) Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. Current Biology 25(6): 690–701. https://doi.org/10.1016/j.cub.2015.01.014

De Beer ZW, Marincowitz S, Duong TA, Kim JJ, Rodrigues A, Wingfield MJ (2016) *Hawksworthiomyces* gen. nov. (Ophiostomatales), illustrates the urgency for a decision on how to name novel taxa known only from environmental nucleic acid sequences (ENAS). Fungal Biology 120(11): 1323–1340. https://doi.org/10.1016/j.funbio.2016.07.004

De Queiroz K (2006) The PhyloCode and the distinction between taxonomy and nomenclature. Systematic Biology 55(1): 160–162. https://doi.org/10.1080/10635150500431221

Grossart HP, Wurzbacher C, James TY, Kagami M (2016) Discovery of dark matter fungi in aquatic ecosystems demands a reappraisal of the phylogeny and ecology of zoosporic fungi. Fungal Ecology 19: 28–38. https://doi.org/10.1016/j.funeco.2015.06.004

Hawksworth DL, Hibbett DS, Kirk PM, Lücking R (2016) (308–310) Proposals to permit DNA sequence data to serve as types of names of fungi. Taxon 65(4): 899–900. https://doi.org/10.12705/654.31

Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE, Huhndorf S, James T, Kirk PM, Lücking R, Lumbsch HT (2007) A higher-level phylogenetic classification of the Fungi. Mycological research 111(5): 509–547. https://doi.org/10.1016/j.mycres.2007.03.004

Hibbett DS, Ohman A, Kirk PM (2009) Fungal ecology catches fire. New Phytologist 184(2): 279–82. https://doi.org/10.1111/j.1469-8137.2009.03042.x

James TY, Seifert KA (2017) Description of *Bifiguratus adelaidae*: The hunt ends for one of the "Top 50 Most Wanted Fungi". Mycologia 109(3): 361–362. https://doi.org/10.1080/00275514.2017.1372667

Kõljalg U, Tedersoo L, Nilsson RH, Abarenkov K (2016) Digital identifiers for fungal species. Science 352(6290): 1182–1183. https://doi.org/10.1126/science.aaf7115

Linnaeus C (1735) Systema Naturæ: Facsimile of the First Edition, translation Engel-Ledeboer MSH, Engel H (1964) B. De Draaf, Nieuwkoop.

Mayden RL (1997) A hierarchy of species concepts: the denouement in the saga of the species problem. In: Claridge MF, Dawah HA, Wilson MR (Eds) Species the unit of biodiversity, vol 54 the Systematic Association Special volume series. Chapman and Hall, Cornwall, 381–423.

Nagy LG, Petkovits T, Kovács GM, Voigt K, Vágvölgyi C, Papp T (2011) Where is the unseen fungal diversity hidden? A study of *Mortierella* reveals a large contribution of reference collections to the identification of fungal environmental sequences. New Phytologist 191(3): 789–94. https://doi.org/10.1111/j.1469-8137.2011.03707.x

Nilsson RH, Kristiansson E, Ryberg M, Larsson KH (2005) Approaching the taxonomic affiliation of unidentified sequences in public databases–an example from the mycorrhizal fungi. BMC Bioinformatics 6: 178. https://doi.org/10.1186/1471-2105-6-178

Nilsson RH, Wurzbacher C, Bahram M, Coimbra VR, Larsson E, Tedersoo L, Eriksson J, Duarte C, Svantesson S, Sánchez-García M, Ryberg M (2016) Top 50 most wanted fungi. MycoKeys 12: 29. https://doi.org/10.3897/mycokeys.12.7553

Page RD (2016) DNA barcoding and taxonomy: dark taxa and dark texts. Philosophical Transactions of the Royal Society B 371(1702): 20150334. https://doi.org/10.1098/rstb.2015.0334

Patterson DJ (2014) Helping protists to find their place in a big data world. Acta Protozoologica 53(1): 115.

Piątek M, Lutz M, Sousa FM, Santos AR, Félix CR, Landell MF, Gomes FC, Rosa CA (2017) *Pattersoniomyces tillandsiae* gen. et comb. nov.: linking sexual and asexual morphs of the only known smut fungus associated with Bromeliaceae. Organisms Diversity & Evolution 17(3): 531–543. https://doi.org/10.1007/s13127-017-0340-8

Porter TM, Schadt CW, Rizvi L, Martin AP, Schmidt SK, Scott-Denton L, Vilgalys R, Moncalvo J-M (2008) Widespread occurrence and phylogenetic placement of a soil clone group adds a prominent new branch to the fungal tree of life. Molecular Phylogenetics and Evolution 46(2): 635–644. https://doi.org/10.1016/j.ympev.2007.10.002

Powell MJ, Letcher PM (2014) Chytridiomycota, monoblepharidomycota, and neocallimastigomycota. In: McLaughlin D, Spatafora JW (Eds) Mycota VII, part A, Systematics and evolution. Springer, Berlin, 141–175.

Robbertse B, Strope PK, Chaverri P, Gazis R, Ciufo S, Domrachev M, Schoch CL (2017) Improving taxonomic accuracy for fungi in public sequence databases: applying 'one name one species' in well-defined genera with *Trichoderma/Hypocrea* as a test case. https://doi.org/10.1093/database/bax072

Rosling A, Cox F, Cruz-Martinez K, Ihrmark K, Grelet GA, Lindahl BD, Menkis A, James TY (2011) Archaeorhizomycetes: unearthing an ancient class of ubiquitous soil fungi. Science 333(6044): 876–879. https://doi.org/10.1126/science.1206958

Ryberg M, Nilsson RH, Kristiansson E, Töpel M, Jacobsson S, Larsson E (2008) Mining metadata from unidentified ITS sequences in GenBank: a case study in *Inocybe* (Basidiomycota). BMC Evolutionary Biology 8: 50. https://doi.org/10.1186/1471-2148-8-50

Ryberg M (2015) Molecular operational taxonomic units as approximations of species in the light of evolutionary models and empirical data from Fungi. Molecular Ecology 24(23): 5770–5777. https://doi.org/10.1111/mec.13444

Samyn Y, De Clerck O (2012) No name, no game. European journal of taxonomy 10: 1–3. https://doi.org/10.5852/ejt.2012.10

Seifert KA (2018) When should we describe species? IMA fungus 8(2): 37–39.

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Bolchacova E, Voigt K, Crous PW, Miller AN (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proceedings of the National Academy of Sciences 109(16): 6241–6246. https://doi.org/10.1073/pnas.1117018109

Sheikh S, Thulin M, Cavender JC, Escalante R, Kawakami SI, Lado C, Landolt JC, Nanjundiah V, Queller DC, Strassmann JE, Spiegel FW (2017) A new classification of the dictyostelids. Protist 169(1): 1–28. https://doi.org/10.1016/j.protis.2017.11.001

Taylor FJ, Sarjeant WA, Fensome RA, Williams GL (1987) Standardisation of nomenclature in flagellate groups treated by both the Botanical and Zoological Codes of Nomenclature. Systematic zoology 36(1): 79–85. https://doi.org/10.2307/2413311

Tedersoo L, Smith ME (2017) Ectomycorrhizal Fungal Lineages: Detection of Four New Groups and Notes on Consistent Recognition of Ectomycorrhizal Taxa in High-Throughput Sequencing Studies. In Tedersoo L (Ed.) Biogeography of Mycorrhizal Symbiosis. Springer International Publishing, Cham, 125–142. https://doi.org/10.1007/978-3-319-56363-3_6

Tedersoo L, Ramirez KS, Nilsson RH, Kaljuvee A, Kõljalg U, Abarenkov K (2015) Standardizing metadata and taxonomic identification in metabarcoding studies. GigaScience 4: 34. https://doi.org/10.1186/s13742-015-0074-5

Torres-Cruz TJ, Billingsley Tobias TL, Almatruk M, Hesse CN, Kuske CR, Desirò A, Benucci GM, Bonito G, Stajich JE, Dunlap C, Arnold AE (2017) *Bifiguratus adelaidae*, gen. et sp. nov., a new member of Mucoromycotina in endophytic and soil-dwelling habitats. Mycologia 109(3): 363–78. https://doi.org/10.1080/00275514.2017.1364958

Tripp EA, Lendemer JC (2014) Sleepless nights: When you can't find anything to use but molecules to describe new taxa. Taxon 63(5): 969–971. https://doi.org/10.12705/635.1