

More than a handful of dirt: sequence-based species description and the role of the ICN (a response to Seifert)

A recent Editorial in *IMA Fungus* (Seifert 2017) is critical of sequence-based species description (Hawksworth *et al.* 2016). The Editorial raises more questions than it answers, concerning the nature of discovery, the minimum evidence that should be required to describe species, and the role of the *International Code of Nomenclature for algae, fungi, and plants* (ICN; McNeill *et al.* 2012) as an enforcer of taxonomic quality. The Editorial trivializes the work of molecular ecologists and paints a bleak picture of the future of taxonomy if sequence-based species description is adopted. Here, I address some of the questions raised in the Editorial and offer a more optimistic vision for the integration of molecular ecology and taxonomy. Responses to quoted elements (in italics) are below:

1. *As stated by Hibbett et al. (2011), "... molecular ecology is clearly the major arena of contemporary species discovery..." rather than conventional taxonomy. My question is whether this process actually discovers new species, or simply indicates that there are new species to be found?*

To "discover" something is to become aware that it exists. It is possible to discover new species through analyses of sequences as well as specimens. It is a different question to ask if sequences (or specimens) provide sufficient information to warrant formal description.

2. *In modern ecology, when you have a substrate in your hand that contains DNA sequences of a thousand species, half of them unknown, have you discovered 500 new species or have you picked up a handful of dirt?*

A molecular ecologist does not simply pick up a handful of dirt. They travel to a sampling locality, collect dirt, record metadata, extract DNA, perform shotgun metagenomics or amplicon sequencing, feed the sequences through a bioinformatics pipeline that compares them to millions of other sequences, make community comparisons using tools such as UniFrac (Loupozone & Knight 2005), and deposit their new sequences in publicly-accessible databases. This is usually where the process ends, without a determination that any new species have been discovered.

How might the research proceed if there was a robust feedback loop between fungal taxonomy and molecular ecology? In that alternate universe, the ecologist would employ a bioinformatics tool to place the new sequences in a phylogenetic tree to seek their closest relatives. The ecologist might then reach out to a taxonomist who could consult relevant monographs

for information about unsequenced taxa, including ecology, and geographic distributions. This newly formed team might request specimens from herbaria or culture collections to obtain sequences to check whether the newly discovered sequences correspond to species that have already been described. At some point, they might decide that some of the new sequences warrant description as new species, but under the current rules of nomenclature they would be unable to validly name their discoveries, because they do not have a physical type specimen. However, if the ecologist had saved a portion of their handful of dirt then that could serve as type material, as demonstrated by the valid publication of *Piromyces cryptodigmaticus* (Kirk 2012; but see Tripp & Lendemer 2012).

3. *Does the act of naming a sequence provide new information that is not already inherent in the sequence itself? I would say not.*

I would say yes. The act of naming communicates the information that someone thinks they have discovered a new species. A DNA sequence is just a DNA sequence, whereas a name embodies a taxonomic hypothesis (including lineage information). One might just as well ask if the act of naming provides new information that is not inherent in a morphological description.

4. *Whether it is one specimen or a hundred, with a specimen in hand it seems clear that you have made a discovery. Does the knowledge that someone else has detected the same DNA in a different handful of dirt really change the picture? There are no characters other than nucleotides, there is no differential ecology or behaviour attributable to the specific unknowns, unless they can be inferred in some way by information inherent in the genetic sequences.*

Yes, the knowledge that someone else has already detected the same DNA represented by the specimen in hand does change the picture. First, it means that the species was already discovered. The taxonomist who has the specimen in hand should be delighted, not threatened, by this knowledge, because the sequence may provide insights that can augment information based on the specimen. Most importantly, the sequence provides information concerning phylogenetic relationships of the species. The phylogeny may make it possible to predict characteristics that are not observable in the specimen, such as physiological attributes. Information about ecology and biogeography can be derived from metadata about the sequence, including sampling locality and source material. For example, if the specimen is a mushroom growing on soil and the sequence comes from a root tip then it may be possible to make an inference about the ecology of the species. With an environmental sequence in hand, the picture has details that might otherwise be invisible.

5. *Is there any conceptual similarity between a species based on one specimen and a species based on a few DNA sequences?*

This is a red herring. The conceptual similarity, to the extent there is one, is of the most general nature, concerning the nature of evidence that justifies formal naming. This question conflates two different issues: (1) whether a species should be described based on a single observation, and (2) whether sequences can serve as the type material. These are not equivalent questions. The answer to one does not automatically inform the answer to the other.

6. *Does a double standard exist, where our historical practise [sic] allows (but is now actively discouraging) what some perceive as*

low quality species descriptions with an old technology, while preventing what some would consider a higher quality of species description using a new technology?

Yes, there is a double standard. A species known only from a single collection can be validly named, even if the type specimen has no observable characters that differentiate it from any other species, has been reduced to dust, or even lost, but a species known by any number of independent environmental sequences, with metadata, cannot be validly named. The absurdity of this double standard will become all the more evident as techniques such as single-cell genomics become more widely used in ecological studies.

7. Are the limitations of what we can determine about a species from a DNA sequence more severe than what we can determine about a species when we have only one specimen? If not, why are so many journals reluctant to allow single species descriptions based on morphology, but lining up to publish controversial papers on DNA defines [sic] taxa that test the limits of the ICN?

See (5) regarding single specimens vs. sequences in species description. The suggestion that “so many” journals are “lining up to publish” papers on sequence-based species description is an exaggeration. Even the most casual perusal of current literature shows that sequence-based species descriptions are extremely rare (perhaps because they are invalid under the ICN), while specimen-based species description is stronger than ever.

8. Quality may not need to be legislated in the ICN, but it still needs to be enforced; there is a strong tendency among mycologists to use the ICN as a quality assurance mechanism. The framers of the ICN have to accept this.

This is a slippery statement. It is not entirely clear what is being recommended, but the implication seems to be that the ICN should accept a role as a taxonomic quality control mechanism, and, one infers, continue to prohibit sequence-based species description. I presume that the author is not trying to suggest that the ICN should add a prohibition on species

description based on single collections, although that would be consistent with a role as a “quality assurance mechanism”.

In my view, the function of the ICN is to dictate the terms for valid publication of names, not to assure taxonomic accuracy. “Quality” in fungal taxonomy is “enforced” by editors and reviewers, and by the scientific community, particularly authors of monographs and other taxonomic compilations. If taxonomists disapprove of sequence-based species, then they should exclude them from their monographs and checklists. However, the rules of nomenclature should not prohibit other workers from formalizing hypotheses based on sequences as valid names with the protection of priority. By the same reasoning, the ICN should not prohibit species description based on a single specimen (or require sequence data).

9. What does it take to raise species description above banality, above trivia that could be extracted by any child or by a machine? Do we want machine taxonomy in fungal biology? From one perspective this seems like a paranoid question and, from another, prescient. If DNA sequences comprise both the description and the type, it is a short step to a pipeline that automatically describes and names the OTUs as species. The question of machine-automated species description is staring us in the face. Surely we should be discussing it?

Yes, this does sound paranoid. I think it is unlikely that we will see a flood of machine-generated names if sequence-based species description is permitted. In any case, even if some rogue bioinformatician decided to flood the literature with spurious names created by an automated pipeline, monographers and other experts would be under no obligation to adopt them (although they would have to deal with them). The rogues would be ostracized, although their h-indices would benefit from all the published criticism (don’t get any ideas, Henrik).

What is more likely, I believe, is that if sequence-based species description was permitted then more taxonomists would begin to screen databases for environmental sequences representing species – both new and previously

discovered – within the groups on which they specialize (they should already be doing this). They would then name only the species for which they think there is adequate evidence. For an example, see the thoughtful (not at all banal or trivial) description of *Hawksworthiomyces sequentia* by De Beer *et al.* (2016).

10. When should we describe species?

The answer to the question posed in the title of the Editorial is that taxonomists should describe species whenever they think that it is warranted based on the evidence at hand. Different taxonomists will have different standards for evidence that justifies species description. One author might insist on several collections and multi-gene phylogenies with Bayesian coalescent analyses. Another may think that a single specimen without molecular data is sufficient. A third may be happy with multiple environmental sequences, backed up by metadata and phylogenetic analyses, as per the proposed Recommendations for sequence-based typification by Hawksworth *et al.* (2016). Currently, the ICN would permit the first two authors to formally name their hypotheses, which would then receive the protection of priority, but the third author would not have that option. This borders on scientific chauvinism, and is inconsistent with the principles of the ICN, which has never dictated the nature of evidence required for species description.

In a different Editorial, James & Seifert (2017) wrote “*After the current trend of documenting the massive biodiversity of unknown fungi subsides, the hard work of finding and describing these unknowns must begin in earnest.*” It is unlikely that research in fungal molecular ecology is going to subside anytime soon. The data are going to keep coming, and with improvements in sequencing technology, they will only become richer and more informative. It would be foolish to wait for molecular ecology to subside before launching major efforts to document fungal diversity with cultures and specimens. There is hard work to be done in all realms of organismal mycology, including specimen- and culture-based taxonomy, molecular systematics, and molecular ecology. The greatest benefits will result if members of traditionally separate

research communities work together, as exemplified by the UNITE consortium. Bioinformatics tools, like the dynamically updated “Top 50 most wanted” list (Nilsson *et al.* 2016), will promote the needed integration (Hibbett *et al.* 2016). Sequence-based species description also has the potential to draw the fungal taxonomic and ecological communities closer together. Each group has much to learn from the other. The leaders of the IMA should promote integrative approaches in mycology, not reinforce historical divisions that maintain disciplinary purity at the cost of progress.

ACKNOWLEDGEMENTS

I am grateful to Henrik Nilsson, Martin Ryberg, and Paul Kirk for comments on this response.

REFERENCES

- De Beer ZW, Marinowitz, 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: 1323–1340.
- Hawksworth DL, Hibbett DS, Kirk PM, Lücking R. (2016) (308–310) Proposals to amend Art. 8 to permit the use of DNA sequences as types of names of fungi. *Taxon* 65: 899–900.
- Hibbett D, Abarenkov K, Kõljalg U, Öpik M, Chai B, Cole J, Wang Q, Crous P, Robert V, Helgason T, Herr JR, Kirk P, Lueschow S, O'Donnell K, Nilsson RH, Oono R, Schoch C, Smyth C, Walker DM, Porras-Alfaro A, Taylor JW, Geiser DM (2016) Sequence-based classification and identification of Fungi. *Mycologia* 108: 1049–1068.
- James TY, Seifert K (2017) Description of *Bifiguratus adelaidae*: The hunt ends for one of the “Top 50 most wanted Fungi”. *Mycologia* 109: 361–362
- Kirk PM (2012) Nomenclatural novelties. *Index Fungorum* 1: 1–1.
- Loupozone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing Microbial Communities. *Applied and Environmental Microbiology* 71: 8228–8235.
- McNeill J, Barrie FR, Buck WR, Demoulin V, Greuter W, *et al.* (2012) *International Code of Nomenclature for algae, fungi, and plants (Melbourne Code) adopted by the Eighteenth International Botanical Congress Melbourne, Australia, July 2011*. [Regnum Vegetabile No. 154.] Königstein: Koeltz Scientific Books.
- Nilsson RH, Wurzbacher C, Bahram M, Coimbra VRM, Larsson E, *et al.* (2016) Top 50 most wanted fungi. *MycologyKeys* 12: 29–40.
- Seifert K (2017) When should we describe species? *IMA Fungus* 8: 37–39.
- Tripp EA, Lendemer JC (2012) Request for binding decisions on the descriptive statements associated with *Mortierella sigyensis* (fungi: *Mortierellaceae*) and *Piromyces cryptodigmaticus* (fungi: *Neocallimastigaceae*). *Taxon* 61: 886–888.

David Hibbett

Biology Department, Clark University,
Worcester, MA 01610, USA
(dhibbett@clarku.edu)

Need for a Web Portal to maintain information on morphological descriptions of all type specimens of fungi

We have some molecular sequence data of different species available in the NCBI database, accessible through internet for molecular taxonomic identification free of cost. In the same way, if we get information through the internet about the morphological description of any particular fungal species on Earth free of cost, it would be great service. If such information were available in a dedicated web portal, then a lot of redundancy and synonymy could be reduced. In fact, if photomicrographs, were also provided then it would be of great help to not only those scientists who are involved in taxonomic research, but also general biologists. This is because the new species are described in various journals and books that are mostly inaccessible. As of now, though some websites give information about individual species, it is often confined to the author's name(s) and the journal volume, year, and page numbers, but does not provide morphological descriptions.

The latest best estimate suggests that there are 2.2–3.8 million species of fungi of which around 120–140 000 have so far been named (Hawksworth & Lücking 2018). There are, however,

perhaps around 260 000 species names in Index Fungorum, indicating that many species have been described several times. One of the reasons for synonymy when naming new species could be the lack of accessibility to information on those already named. Aptroot (1995a) monographed the ascomycete genus *Didymosphaeria*, which has more than 550 species names. Of these names, 100 species had been transferred to other genera, and he accepted only seven species in *Didymosphaeria* – several 100 being synonyms of those seven species (Aptroot 1995b).

Similarly, a world revision of another pyrenocarpous ascomycete genus *Massarina* by Aptroot (1998) led him to accept only 43 species belonging to this genus out of 166 species names attributed to it; many were either found to be synonyms or were redispersed to other new or existing genera. Hence there is a need for a web portal to provide information on morphological description of all type species of fungi known on earth to avoid duplications. At present, some websites (Mycobank, Index Fungorum) provide only bibliographic and nomenclatural

information about most individual species, although since the registration of new fungal names became mandatory from 1 January 2013 the associated diagnosis or description is also available. In addition, MycoBank has a facility for authors to deposit other material, such as illustrations, and Index Fungorum is gradually establishing hyperlinks to original publications where that is permitted.

Including copies of material from books and journals still in copyright is a major constraint, and a way around this needs to be found, perhaps by loading text rather than copies of printed pages. Access to original current journal articles is increasingly difficult due to steep price increases, and many libraries have stopped subscribing to even the core mycological journals, and while e-subscriptions and pay-by-article options exist, they are prohibitively expensive for individuals. This difficulty for accessing information retrieval can result in duplication of work and redundancy when different workers describe the same new species.

Though molecular tools are of great help in solving some identification problems

only around 17 % of the known fungi are isolated and available in fungal genetic resource collections (Hawksworth 2004, 2012). More attempts should be made to isolate and sequence more fungi, but in the case of described species it is necessary to be sure of their identities by also being able to access the original accounts. Until sequences are available for all known fungi, we will continue to rely on morphological features for identification.

This vision of a web portal to access descriptions and photomicrographs of the type specimens of all fungal species needs to be realized. It could perhaps be based on the MyCoPortal model, which links data from 84 institutions worldwide (Miller & Bates 2017) but is not restricted to descriptions and illustrations of type specimens. The

International Mycological Association would be the logical body to consider how to facilitate such an initiative.

This letter is based on a paper presented at the national conference on "Fungal Biology: Recent Trends and Future Prospects" held at the University of Jammu, Jammu, India on 16–17 November 2017.

Aptroot A (1995a) A monograph of *Didymosphaeria* (ascomycetes). *Studies in Mycology* 37: 1–160.

Aptroot A (1995b) Redisposition of some species excluded from *Didymosphaeria* (ascomycetes). *Nova Hedwigia* 60: 325–379.

Aptroot A (1998) A world revision of *Massarina* (Ascomycota). *Nova Hedwigia* 66: 89–162.

Hawksworth DL (2004) Fungal diversity and its implications for genetic resource collections. *Studies in Mycology* 50: 9–18.

Hawksworth DL (2012) Global species numbers of fungi: are tropical studies and molecular approaches contributing to a more robust estimate? *Biodiversity and Conservation* 21: 2425–2433.

Hawksworth DL, Lücking R (2018) Fungal diversity revisited: 2.1 to 3.8 million species. In: *The Fungal Kingdom* (Heitman J, et al., eds): 79–95. Washington DC: American Society for Microbiology.

Miller AN, Bates ST (2017) The Mycology Collections Portal (MyCoPortal). *IMA Fungus* 8: (65)–(66).

V. Venkateswara Sarma

Department of Biotechnology, School of Life Sciences, Pondicherry University, Kalapet, Pondicherry-605014, India (sarmavv@yahoo.com)