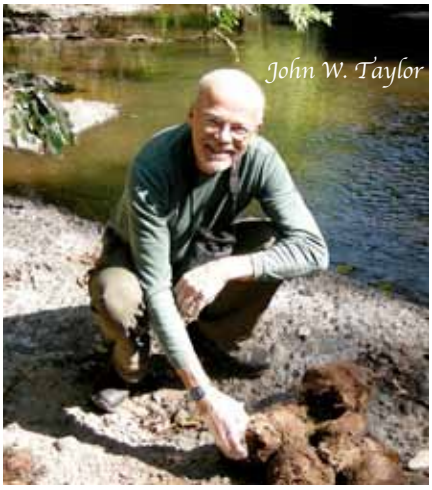
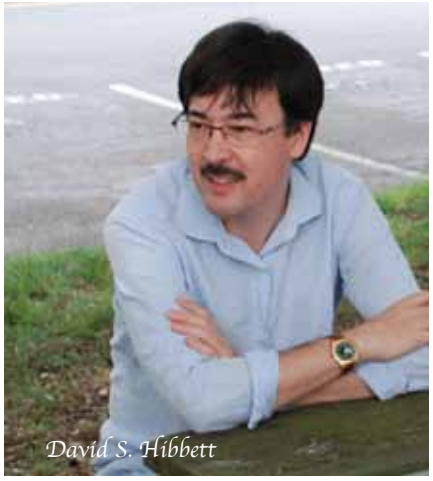
TOWARD SEQUENCE-BASED CLASSIFICATION OF FUNGAL SPECIES

The November issue of *Molecular Ecology* leads off with a wonderful News and Views piece from the architects of the UNITE project (http://uite.ut.ee) and a diverse, international group of collaborators. The article, "Towards a unified paradigm for sequence-based identification of fungi" (Kõljalg *et al.* 2013), reports on a January 2013 workshop focused on advancing ITS-based identification of fungal communities using the latest tools from the UNITE consortium. For more than a decade, UNITE has worked to provide a curated database of fungal ITS sequences that can be used to identify new sequences, whether from cultures, specimens, or environmental samples, as well as sophisticated informatics tools for sequence-based identification. UNITE addresses two daunting problems associated with the International Nucleotide Sequence Databases (INSD - GenBank, EMBL and DDJB), that is, the lack of community-curation and the presence of many unnamed DNA sequences. Their efforts are to be welcomed by anyone wanting to identify fungi using DNA sequences.

The example that forms the core of the article shows how the UNITE resources can be used to cluster ITS sequences into "species hypotheses" (SH) based on interspecific similarities of between 97 and 99 % (corresponding to "gaps" of 1–3 % between SH clusters). For each SH, a consensus ITS sequence is constructed and a reference sequence is designated (by default this is the sequence closest to the consensus, but expert users can designate a different reference sequence, such as that derived from a type specimen). Kõljalg and colleagues report that reference sequences have been designated for nearly 2000 of the nearly 53 000 SHs in the UNITE database. With the new platform for delimiting ITS-based SHs, the UNITE consortium has produced a valuable resource for mycology. However, is this resource broad enough to address the problem of fungi known only from DNA sequences obtained from environmental samples? And, can it satisfy the diversity of opinion found among mycologists?

Species recognition is not a simple matter, and one size does not fit all, as acknowledged by Kõljalg and colleagues. UNITE aims for a level of taxonomic precision midway between that accepted by ecologists and that needed by population genomicists. At its most inclusive, the UNITE approach equates 97 % similarity with conspecificity, but UNITE's more exclusive percentages, up to 99 %, will inflate species numbers (Amend *et al.* 2010). At the Genome Wide Association end of the spectrum, where the taxonomic unit is a freely interbreeding population, an ITS limit of 100 % would be far too inclusive. Its acceptance as the fungal BarCode not withstanding, ITS is plagued by intragenomic variation, absence from some clades, and lack of resolution in others (Schoch *et al.* 2012). Although efforts to find an alternative to ITS that is effective in all branches of the fungal kingdom have yet to succeed, other regions have been shown to be superior for species identification in particular clades, or have simply been more extensively sampled than ITS. For example, in arbuscular mycorrhizal fungi, the 18S nuclear rRNA gene has been developed as a molecular marker for "Virtual Taxa" in the MaarjAM database (Öpik *et al.* 2010). This, and other examples that could be cited, should not be taken as a criticism of the tools and approach of UNITE, which currently offer a state-of-the-art, flexible system for ITS-based taxon delimitation (that could be extended to other genes or even "phenomic" data). Rather, they illustrate that the optimal criteria and approaches for sequence-based taxon delimitation are likely to vary across clades of fungi, and will be determined by the preferences of the communities specializing in particular groups. All methods have a fixed lifespan. Barcoding using ITS, or any other single-gene marker, may be eclipsed by genomics when, in the not too distant future, it proves easier to first sequence the genomes of a group of fungi and then choose the best regions for systematics from among all possible candidates.


John W. Taylor


David S. Hibbett

The UNITE approach is tailored to sequences tied to cultures or specimens, but fungal ecologists face the problem of novel, fungal ITS sequences from environmental samples (Hibbett *et al*. 2011, Hibbett & Taylor 2013). Here the current level of automation at UNITE, which requires operator input to identify species clusters, may not be sufficient to enable ecologists to put names on the thousands of SHs that are found in a typical ecological study. There also is additional labor involved in linking metadata to the sequences, data that currently are manually extracted from publications by UNITE.

Then there is the sociological problem of one system satisfying all mycologists. When registration of new species was proposed in the new *International Code of Nomenclature*, it proved impossible to settle on just one site. Even the term SH is different from the more generally understood molecular-OTU (mOTU) or an alternative, Environmental Nucleic Acid Sequence (ENAS). UNITE is wonderful because it is curated, but if mycologists want to reach all of biology, the same level of community curation will be needed for the much larger and far better known, INSDs (Bidartondo *et al*. 2008). This tension is seen throughout science; on one end of the rope are the visionary groups, UNITE being a prime example, that "just do it," and show the rest of us how to make a curated database and how to semi-automatically name species. On the other end is the larger community of biologists that would like to find all sequences and all metadata in one location and who find it difficult enough to submit sequences to GenBank and metadata to MycoBank. This tug-of-war seems to be inevitable and eternal; the best that we can hope for may be a series of UNITE-like groups, all operating on a playground that has just enough rules to allow for communication among the groups but not so many as to stifle innovation.

For the reasons given above, we would argue against the adoption of any single approach as the sole method for delimiting sequence-based taxonomic entities that are recorded in community taxonomic databases such as MycoBank, Index Fungorum, or the NCBI taxonomy (although unified resources certainly are important for ecological studies). Rather, we suggest that a pluralistic approach to sequence-based taxonomy is needed to allow different ideas to compete in the taxonomic marketplace. In the short run, such an open approach has the potential to create confusion, particularly if different groups of researchers use different methods to delimit taxa within the same clades. To minimize such conflicts, while allowing different ideas to flourish and be tested, it will be necessary for the taxonomic community to engage in a broad, open conversation about sequence-based classification. An inclusive meta-category for sequence-based taxa in public databases that is non-restrictive with regard to genes, similarity cut-offs or other criteria, would make sequence based taxonomy concrete and promote discovery of best practices.

The next step is to broaden the conversation, and that is what will happen this coming year in a series of meetings that will address the challenges and opportunities for sequence-based fungal classification. The first meetings will take place in April at the Centraalbureau voor Schimmelculture's symposia and workshops in Utrecht and Amsterdam (http://www.cbs.knaw.nl/index.php/meetings/396-2014-cbs-spring-symposium-the-genera-of-fungi), then in June at the Mycological Society of America meetings in East Lansing, Michigan (http://msaconference.msafungi.org/) and finally in August at the International Mycological Congress in Bangkok (http://www.imc10.com/2014/home.html). The discussions are certain to be lively. Let us hope that they are at least as productive as the UNITE workshop.

Acknowledgment
We thank Urmas Kõljalg and Henrik Nilsson for helpful comments.

Amend AS, Seifert KA, Samson R, Bruns TD (2010) Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. *Proceedings of the National Academy of Sciences, USA* **107**: 13748–13753.

Bidartondo MI, Bruns TD, Blackwell M, Edwards I, Taylor AFS [and 252 others] (2008) Preserving accuracy in GenBank. *Science* **319**: 1616.

Hibbett DS, Ohman A, Glotzer D, Nuhn M, Kirk P, Nilsson RH(2011) Progress in molecular and morphological taxon discovery in *Fungi* and options for formal classification of environmental sequences. *Fungal Biology Reviews* **25**: 38–47.

Hibbett DS, Taylor JW (2013) Fungal systematics: is a new age of enlightenment at hand? *Nature Reviews Microbiology* **11**: 129–133.

Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS [and 37 others] (2013) Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology* **22**: 5271–5277.

Öpik M, Vanatoa A, Vanatoa E, Moora M, Davison J, Kalwij JM, Reier Ü, Zobel M (2010) The online database Maarj*AM* reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (*Glomeromycota*). *New Phytologist* **188**: 223–241.

Palma-Guerrero J, Hall CR, Kowbel D, Welch J, Taylor JW, Brem RB, Glass L (2013) Genome wideassociation identifies novel loci involved in fungal communication. *Plos Genetics* **9**: DOI: 10.1371/journal.pgen.1003669

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL [and 151 others] (2012) Fungal barcoding: nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*. *Proceedings of the National Academy of Sciences, USA* **109**: 6241–6246.

John W. Taylor and David S. Hibbett
(jtaylor@berkeley.edu; dhibbett@clarku.edu)