Molecular Identification of Fungi: Rationale, Philosophical Concerns, and the UNITE Database

R.H. Nilsson[*,1,2,§], K. Abarenkov[3,§], K.-H. Larsson[4,§] and U. Kõljalg[2]

[1]*Department of Plant and Environmental Sciences, University of Gothenburg, Box 461, 405 30 Göteborg, Sweden*

[2]*Department of Botany, Institute of Ecology and Earth Sciences, University of Tartu, 40 Lai St., 51005 Tartu, Estonia*

[3]*Natural History Museum, University of Tartu. 46 Vanemuise Str., 51014 Tartu, Estonia*

[4]*The Mycological Herbarium, Natural History Museum, University of Oslo, P.O. Box 1172, Blindern, 0318 Oslo, Norway*

Abstract: Fungi form a ubiquitous group of largely inconspicuous organisms that play key ecological roles in terrestrial nutrient cycling. The typically subterranean or otherwise unnoticeable nature of fungal life has left mycology with a partial understanding of fungal biology, and progress is hampered by factors such as the difficulty to delimit species and individuals of fungi and the sparsity of discriminatory morphological and physiological characters for reliable species identification. These and other complications have paved the way for DNA sequence data as an important source of information in mycology, and there are now twenty years' worth of fungal sequence data available for scientific purposes. However, issues of data reliability and generality impede the use of publicly available fungal DNA sequences. The UNITE database for molecular identification of fungi (http://unite.ut.ee) was built as a response to the difficulties facing anyone seeking to identify environmental samples of fungi to species level using molecular data and the major international sequence databases. The present study describes the UNITE database and examines the role of UNITE in the light of emerging sequencing technologies, notably massively parallel ("454") pyrosequencing. Environmental sampling of fungi is discussed from a taxonomic perspective.

Keywords: Fungi, molecular identification, environmental sampling, taxonomy, ITS.

INTRODUCTION

The access to DNA sequence data has reshaped much of the life sciences, and there are few, if any, branches of biology that have not benefited from molecular data. This holds true also for the study of the diversification of life itself - evolutionary and systematic biology. Indeed, sequence data have proved a detailed source of information highly relevant to endeavours such as taxonomic classification, evolutionary studies of the origin of organism groups and their geographical distribution, and the detection and identification of species in various habitats and substrates [1-4]. The latter field - molecular identification - has seen particular progress in relation to sequence data in that specimens and substrates previously unamenable to precise biological examination - such as non-fertile plant parts, soil, and wood - have become available for detailed scientific scrutiny. There are countless examples of the value of molecular identification in linking specimens and ecological properties to species names in systematics and taxonomy, ecology, and conservation biology [5-7]. Agriculture, forest management, and the food industry are other areas of application; here the identification of pathogens, pests, and contaminations are often in focus [8-10].

The heterogeneous uses and applications of molecular identification are reflected in the different kinds of results each respective study is designed to generate. While simple lists of species names form the end product in many cases, other efforts will have very specific requirements on the auxiliary data associated with each species name, such as detailed information on the collection site, soil type, and any host species. Other studies are concerned with the relative composition of biological communities and so may not bother with species names in the first place. These parties will to various extents see the need for, or benefits of, depositing their data in public sequence databases and will do so with varying levels of commitment. As a result these repositories accumulate very nonuniform entries at different stages along the trajectory to rich and reliable annotations. For certain genes and groups of organisms the integrity of the data is severely compromised, often without immediate means through which to tell satisfactory data from the unsatisfactory [11-13]. These inconsistencies are regularly subsumed into the data and results of other studies and thus percolate through the research community, eventually reaching those who may not be in a position to examine them critically [14].

When it comes to molecular identification, fungi form an example of a group of organisms where technology has outpaced taxonomy. The inconspicuous nature of fungal life coupled with the lack of reliable morphological characters for identification and discrimination among species have precluded a detailed understanding of the fungal kingdom,

*Address correspondence to this author at the Department of Plant and Environmental Sciences, University of Gothenburg, Box 461, 405 30 Göteborg, Sweden; Tel: +46-31-786 2623; Fax: +46-31-786 2560; E-mail: henrik.nilsson@dpes.gu.se

§Equal contribution.

and DNA sequence data have therefore been formative for our current view of fungi [15, 16]. The internal transcribed spacer (ITS) region of the nuclear ribosomal repeat unit has become the primary genetic marker for molecular identification and other species-level pursuits in many groups of fungi [17]. Yet less than 1% of the hypothesized number of extant fungal species - 1.5 million - have been sequenced for the ITS region [18, 19]. As a consequence, most large-scale sequencing efforts of fungi recover a significant number of species that cannot be identified to species or even genus level [20]. The ease with which the ITS region can be sequenced - and the scale at which such efforts can be launched today - have left nearly half of the 165,000 fungal ITS sequences in the International Nucleotide Sequence Databases (INSD: GenBank/EMBL/ DDBJ [21]) without a full species name and as much as another 10% with incorrect names [22]. Anyone seeking to identify newly generated fungal ITS sequences to species level using the international sequence databases therefore regularly needs to address non-trivial issues such as contradictory or competing species names, the presence of large numbers of relevant but unidentified sequences, and the lack of similar sequences altogether. In addition, the use of emerging sequencing technologies, where the sequence read length and quality may not always match those of traditional Sanger sequencing, adds a layer of technical complexity to such searches [23, 24].

THE UNITE DATABASE - TECHNOLOGY AND APPLICATIONS

Conceived by a North European team of researchers with a common interest in mycology, the UNITE database for molecular identification of fungi (http://unite.ut.ee) [25-26] grew from the incipient need for prompt yet satisfactory analysis of newly generated fungal DNA sequences in the light of properly identified, and richly annotated, reference sequences. The making of UNITE was characterized by the need to develop new - and modify existing - software solutions to fill the gap between what software infrastructure was able to provide at the time and the demands imposed by structured large-scale sequencing efforts of fungi. Several commercial and freely available database systems - with and without a biological connotation - were evaluated for use as platform for the database. All were rejected, however, for lacking the flexibility and extendibility deemed necessary for the task. Instead the open source MySQL database (http://www.mysql.com) was chosen to form the mainstay of UNITE, and the interfaces and search features around it were built *ab initio*. UNITE was initially accessed through a relatively simple web interface and Perl / PHP scripts, but as the complexity of the data and relationships that needed to be represented increased, the decision was made to orient the development towards a unifying, workbench-type environment for management of environmental samples of fungi. The number of tables was increased to about 160, and much effort was expended on streamlining the retrieval and deposition of large quantities of sequences and their metadata. Similarly mechanisms to compare and correlate data of different origins were developed [26].

UNITE currently holds 2,968 reference sequences from 1,120 fungal species in 155 genera, primarily of ectomycorrhizal fungi from North Europe (Fig. **1**); the

taxonomic and geographic scopes have however recently been widened to include all true fungi on a global basis. To maintain the taxonomic integrity of UNITE, the submission of reference sequences is a restricted process. To qualify as a reference sequence, an entry must stem from a thoroughly annotated voucher specimen residing in a public herbarium (or culture collection) and must be submitted by a person with a documented experience of the fungal group in question. All new reference sequences are subjected to peer review prior to inclusion. In addition to the reference sequences, UNITE hosts public sequences that, although being of sufficient length and quality, do not meet the full requirements to serve as reference sequences; these are provided as environmental samples or otherwise insufficiently identified sequences. These entries, too, have an important role to play in that they may be used to link species to, e.g., nutritional modes, substrates, or geographical regions [27, 28].

The primary genetic marker targeted by UNITE is the ITS region, which together with the D1/D2 region of the neighbouring nuclear large subunit (nLSU/28S) gene forms a prime candidate for the role as the standardized region for molecular identification of fungi [6, 29]. The ITS region does not work equally well for all groups of fungi, however [30, 31], and UNITE welcomes sequences from other genes and markers relevant to molecular identification provided that the sequences are annotated with primer and amplification details. At present most of the search and analysis features of UNITE are however tailored towards the ITS region. Newly generated ITS sequences can be queried for taxonomic affiliation in several ways at the UNITE entry page (http://unite.ut.ee). In addition to the sequence similarity search suite BLAST [32] for single sequences or in batch mode for larger datasets, UNITE offers additional search functions drawing from phylogenetic analysis and hidden Markov models [33, 34]. The user can choose to include the fungal ITS sequences in INSD in these searches; UNITE reference sequences are highlighted in the output to mark their distinctness from non-reference sequences. The database can be queried for species names to show sequences, morphological descriptions, illustrations, and collection data for that species (Fig. **1**). This information often proves helpful in the sequence identification step, and UNITE encourages deposition of supplementary data and files such as relevant photos and illustrations along with sequences.

UNITE also offers a workbench for sequence management, PlutoF, which ultimately is intended as a comprehensive solution to environmental sequencing efforts targeting fungi. A system login is required to access the more advanced features of PlutoF. Upon logging in the user can upload and manage projects and sequence sets and prepare various batch analyses, such as comparing sequences from different plots of some given study for differences in taxonomic composition (Supplementary Material 1). A module to handle herbarium specimens and cultures is available, and there is support for projecting the distribution of species and specimens onto maps of regional or global scope. Several tools to assist in quality control of sequence data have been devised, including a chimera checker for the ITS region, a utility to reorient sequences that are inadvertently given in the reverse complementary

Fig. (1). A) Fruiting body of the ectomycorrhizal (plant symbiotic) fungus *Thelephora palmata*. **B)** The ectomycorrhizal root-tip mycelia of a closely related species, seen here ensheathing the root tips of Scots pine (*Pinus sylvestris*). The morphology of ectomycorrhizal root tips rarely offers sufficient detail to allow identification of the fungus to species level, and there is similarly little hope of tracing the mycelium back to any associated, nearby fruiting body, which may or may not exist. Sequence data thus often represent one's only chance of a satisfactory species-level identification of the fungal component of mycorrhiza. UNITE was initially conceived for molecular identification of ectomycorrhizal root tips, but its taxonomic scope has recently been extended to include all true fungi on a global basis.

orientation, and a tool to extract the highly variable ITS1 and ITS2 subregions from the neighbouring and very conserved genes of the ITS region [35-37]. Incremental annotation of INSD entries is supported, giving the user the opportunity to tag unidentified or incorrectly identified, chimeric, or otherwise problematic INSD sequences with relevant information or taxonomic annotations. A software pipeline for analysis of massively parallel ("454") pyrosequencing [38] ITS datasets was likewise developed and recently introduced in PlutoF [39]. Environmental sequencing draws from many branches of the biological sciences, however, and the workbench is actively developed to encompass all facets of the field. UNITE co-operates with the Fungal Environmental Sampling and Informatics Network [40] to co-ordinate a joint resource infrastructure and to avoid duplicate development. Initiatives such as the specimen image database MorphBank [41] and the authoritative nomenclatural resource MycoBank [42] are also likely to be of importance to UNITE.

TECHNOLOGICAL ADVANCES AND THEIR TAXO-NOMIC RAMIFICATIONS

At the onset of the UNITE initiative, the process of obtaining DNA sequences was a relatively slow and at best semi-automated undertaking. The last few years have however witnessed enormous leaps in the fields of DNA amplification and sequencing. Emerging sequencing technologies such as massively parallel pyrosequencing have the capacity to generate hundreds of thousands of sequences - though of limited length - in the course of a day [43]. Needless to say, the possibilities offered by these technologies in terms of characterization of organisms at the community level seem endless. And yet it is not difficult to identify potential problems that may surface as a result of their use. With that amount of output, automated processing and interpretation of the results are likely to prove the only

feasible way forward. Several studies have however pointed at the substantial difficulties inherent to assigning a broad selection of fungal sequences to species level in a uniform yet satisfactory way [44-46]. To rely on static threshold values of sequence similarity to demarcate distinct species is a practice that hardly can be recommended other than for the reason that equally rapid alternative approaches are largely lacking. The phylogenetic identification tools of UNITE do perhaps represent a step in the direction of a more analytical approach to the taxonomic assignment of fungal ITS sequences, but they currently process only one or a handful of sequences at the time and require at least cursory manual interpretation. To streamline these tools to handle multiple query sequences and to interpret the results generically will be given high priority in the UNITE development process. That these needs are not unique to UNITE is demonstrated by an increasing array of software development efforts targeting similar problems [47-49]. Even so, due to their speed, sequence clustering and similarity searches are likely to remain popular tools in community-level sequencing efforts. It therefore seems inevitable that such efforts, for the time being, be associated with species names that in part are approximate. A good practice may be to refrain from providing full species names other than in very clear-cut cases. As much as one would like to see all sequences in one's dataset and downstream publications resolved to species level, a partial but correct name - such as the generic name only - is much preferable to a full but incorrect species name.

Given that species delimitation is handled in a satisfactory way, environmental sequencing efforts are in an unprecedented position to fuel and advance taxonomic research by indicating the presence of new, undescribed species and by providing depth and dimension to intraspecific and geographic sampling. Similarly the emerging technologies should prove powerful tools in the

pursuit of new or complementary genes for species delimitation and systematics. One can however also think of ways in which taxonomy could be adversely affected by the new developments. Taxonomy is a depauperate and undermanned discipline that may find it difficult to compete for funding with large international sequencing efforts employing bleeding-edge technology. Environmental sequencing projects also have the capacity to flood mycology with unidentified sequences lacking the auxiliary data needed to be of use in any other project than the very study in which they were generated. The comparability of biodiversity results obtained under different spatiotemporal conditions is at a clear risk of being compromised if species for which no Latin names are available would be referred to by different *ad hoc* names - such as "*Cortinarius* clade 3" *contra* "*Cortinarius* cf. *rubellus*" for the same taxon - in every study in which they were recovered [cf. 50]. To establish correspondence among the underlying taxa for such names across different published scientific studies would require significant effort - if it could be done at all - and is likely to prove enough of an obstacle that any interesting properties reported in any one of those studies will go unnoticed by others. In a way not altogether different from how species are conceptualized under phylogenetic nomenclature [51-52], UNITE is looking to support the provision of temporary operational names of the accession number type for such unidentified - and unidentifiable - clusters of hypothetically conspecific sequences until the data is there to warrant formal description of the species. Such operational names would serve as handles to which properties of the underlying hypothetical species could be attached and through which direct and precise communication of results could be achieved.

COMMUNITY-LEVEL SEQUENCE DATA: CONCLU-DING REMARKS

It is often said that technology (or the cost of employing the desired technology) forms the bottleneck for what mycological research questions that can be pursued. Now, at least in some respects, the reverse is true: now we can sequence an entire community of fungi at a cost lower than that of sequencing 1,000 specimens ten years ago. We have yet to grasp the full implications of the emerging sequencing technologies for mycology, however. The few pyrosequencing efforts of fungi published to date all testify to the great diversity of fungi - a diversity that is not easily fitted into the known species, genera, and orders of fungi [e.g., 53-55]. Indeed, the highly controversial question of whether the *International Code of Botanical Nomenclature* [56] should be modified to allow species descriptions based on sequence data alone is likely to resurface in the wake of environmental sequencing efforts. It may however not be to the purpose to allow such descriptions at this stage since we still have ITS data from less than 1% of the estimated number of fungal species. At present we are not in a position to tell species that do not form fruiting bodies or other appreciable morphological structures at all - and that hence for all practical purposes can be detected only through DNA sequencing - from species that *do* form fruiting bodies but for which no reference sequences are available. Most herbaria and other collections of fungal specimens remain largely unsampled for ITS - or other reference - sequences,

and there is evidence to suggest that a considerable proportion of the presently unidentified ITS sequences in INSD could be identified to species level if the herbaria worldwide were explored for ITS sequences from suitable reference specimens [57]. To secure funding for such elementary research is however likely to prove a nontrivial undertaking.

Compounding the problem, less than a quarter of the fungal species described between 2000 and 2009 were accompanied by sequence data [20]. Thus, if the *Botanical Code* indeed is to be modified, it may at this stage be better to mandate the supply of a relevant DNA sequence with each new species description (as applicable) than to allow species descriptions based on sequence data only. If the mycological community deems this approach to be too controversial, then other means of stimulating the growth of the reference sequence corpus must be explored. The way it is now, anyone interested in the true face of the fungal diversity at any particular site may well be better off spending 75% of the money sequencing random high-quality herbarium (or culture collection) specimens from previously unsequenced species and only 25% of the money sequencing environmental samples from that site than, as often is the case now, diverting all resources towards the site in question. That is clearly below what can be accepted given the full span of mycological research questions the emerging sequencing technologies have the capacity to resolve.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

The authors have no competing interests to declare.

SUPPLEMENTARY MATERIAL

This article is also accompanied with supplementary material and it can be viewed at publisher's website.

REFERENCES

[1]   J. Felsenstein, *Inferring phylogenies*. Sunderland, MA: Sinauer Associates, 2004.
[2]   B. D. Shenoy, R. Jeewon, and K. D. Hyde, "Impact of DNA sequence-data on the taxonomy of anamorphic fungi", *Fungal Divers.*, vol. 26, pp. 1-54, 2007.
[3]   J. E. Stajich, M. L. Berbee, M. Blackwell, D. S. Hibbett, T. Y. James, J. W. Spatafora, and J. W. Taylor, "The Fungi", *Curr. Biol.*, vol. 19, pp. R840-R845, 2009.
[4]   R. H. Nilsson, M. Ryberg, E. Sjökvist, and K. Abarenkov, "Rethinking taxon sampling in the light of environmental sequencing", *Cladistics*, vol. 27, pp. 197-203, 2011.
[5]   A. F. S. Taylor, "Recent advances in our understanding of fungal ecology", *Coolia*, vol. 52, pp. 197-212, 2008.
[6]   D. Begerow, R. H. Nilsson, M. Unterseher, and W. Maier, "Current state and perspectives of fungal DNA barcoding and rapid identification procedures", *Appl. Microbiol. Biotechnol.*, vol. 87, pp. 99-108, 2010.
[7]   E. Naro-Maciel, B. Reid, N. N. FitzSimmons, M. Le, R. DeSalle, and G. Amato, "DNA barcodes for globally threatened marine

turtles: a registry approach to documenting biodiversity", *Mol. Ecol. Res.*, vol. 10, pp. 252-263, 2010.

[8]   E. G. Brockerhoff, A. M. Liebhold, and H. Jactel, "The ecology of forest insect invasions and advances in their management", *Can. J. For. Res.*, vol. 36, pp. 263-268, 2006.

[9]   E. H.-K. Wong and R. H. Hanner, "DNA barcoding detects market substitution in North American seafood", *Food Res. Int.*, vol. 41, pp. 828-837, 2008.

[10] R. Floyd, J. Lima, J. deWaard, L. Humble, and R. Hanner, "Common goals: policy implications of DNA barcoding as a protocol for identification of arthropod pests", *Biol. Invasions*, vol. 12, pp. 2947-2954, 2010.

[11] M. I. Bidartondo, "Preserving accuracy in GenBank", *Science*, vol. 319, pp. 1616, 2008.

[12] R. Christen, "Global sequencing: a review of current molecular data and new methods available to assess microbial diversity", *Microbes. Environ.*, vol. 23, pp. 253-268, 2008.

[13] S. Kang, M. A. Mansfield, B. Park, and 7 co-authors, "The promise and pitfalls of sequence-based identification of plant-pathogenic fungi and oomycetes", *Phytopathology*, vol. 100, pp. 732-737, 2010.

[14] W. R. Gilks, B. Audit, D. D. Angelis, S. Tsoka, and C. A. Ouzonis, "Modelling the percolation of annotation errors in a database of protein sequences", *Bioinformatics*, vol. 18, pp. 1641-1649, 2002.

[15] M. Blackwell, D. S. Hibbett, J. W. Taylor, and J. W. Spatafora, "Research coordination networks: a phylogeny for kingdom *Fungi* (Deep Hypha)", *Mycologia*, vol. 98, pp. 829-837, 2006.

[16] T. Y. James, F. Kauff, C. L. Schoch, and 70 co-authors, "Reconstructing the early evolution of *Fungi* using a six-gene phylogeny", *Nature*, vol. 443, pp. 818-822, 2006.

[17] K. A. Seifert, "Progress towards DNA barcoding of fungi", *Mol. Ecol. Res.*, vol. 9, pp. 83-89, 2009.

[18] D. L. Hawksworth, "The magnitude of fungal diversity: the 1.5 million species estimate revisited", *Mycol. Res.*, vol. 105, pp. 1422-1432, 2001.

[19] M. Ryberg, E. Kristiansson, E. Sjökvist, and R. H. Nilsson, "An outlook on the fungal internal transcribed spacer sequences in GenBank and the introduction of a web-based tool for the exploration of fungal diversity", *New Phytol.*, vol. 181, pp. 471-477, 2009.

[20] D. S. Hibbett, A. Ohman, and P. M. Kirk, "Fungal ecology catches fire", *New Phytol.*, vol. 184, pp. 279-282, 2009.

[21] D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers, "GenBank", *Nucleic Acids Res.*, vol. D38, pp. D46-D51, 2010.

[22] R. H. Nilsson, M. Ryberg, E. Kristiansson, K. Abarenkov, K.-H. Larsson, and U. Kõljalg, "Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective", *PLoS One*, vol. 1, art. e59, 2006.

[23] S. M. Huse, J. A. Huber, H. G. Morrison, M. L. Sogin, and D. M. Welch, "Accuracy and quality of massively parallel DNA pyrosequencing", *Genome. Biol.*, vol. 8, R143, 2007.

[24] J. M. Rothberg and J. H. Leamon, "The development and impact of 454 sequencing", *Nat. Biotechnol.*, vol. 26, pp. 1117-1124, 2008.

[25] U. Kõljalg, K.-H. Larsson, K. Abarenkov, and 13 co-authors, "UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi", *New Phytol.*, vol. 166, pp. 1063-1068, 2005.

[26] K. Abarenkov, R. H. Nilsson, K.-H. Larsson, and 15 co-authors, "The UNITE database for molecular identification of fungi - recent updates and future perspectives", *New Phytol.*, vol. 186, pp. 281-285, 2010.

[27] K. G. Peay, P. G. Kennedy, and T. D. Bruns, "Fungal community ecology: a hybrid beast with a molecular master", *BioScience*, vol. 58, pp. 799-810, 2008.

[28] M. Ryberg, R. H. Nilsson, E. Kristiansson, M. Töpel, S. Jacobsson, and E. Larsson, "Mining metadata from unidentified ITS sequences in GenBank: a case study in *Inocybe* (*Basidiomycota*)", *BMC Evol. Biol.*, vol. 8, art. 50, 2008.

[29] U. Eberhardt, "A constructive step towards selecting a DNA barcode for fungi", *New Phytol.*, vol. 187, pp. 265-268, 2010.

[30] E. Bellemain, T. Carlsen, C. Brochmann, E. Coissac, P. Taberlet, and H. Kauserud, "ITS as an environmental DNA barcode for fungi: An in silico approach reveals potential PCR biases", *BMC Microbiol.*, vol. 10, art. 189, 2010.

[31] H. Stockinger, M. Krüger, and A. Schüßler, "DNA barcoding of arbuscular mycorrhizal fungi", *New Phytol.*, vol. 187, pp. 461-474, 2010.

[32] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.,* vol. 25, pp. 3389-3402, 1997.

[33] S. R. Eddy, "Profile hidden Markov models", *Bioinformatics*, vol. 14, pp. 755-763, 1998.

[34] R. H. Nilsson, K.-H. Larsson, and B. Ursing, "galaxie - CGI scripts for sequence identification through automated phylogenetic analysis", *Bioinformatics,* vol. 20, pp. 1447-1452, 2004.

[35] R. H. Nilsson, K. Abarenkov, V. Veldre, S. Nylinder, P. De Wit, S. Brosche, J. F. Alfredsson, M. Ryberg, and E. Kristiansson, "An open source chimera checker for the fungal ITS region", *Mol. Ecol. Resour.,* vol. 10, pp. 1076-1081, 2010.

[36] R. H. Nilsson, V. Veldre, M. Hartmann, and 7 co-authors, "An open source software package for automated extraction of ITS1 and ITS2 from fungal ITS sequences for use in high-throughput community assays and molecular ecology", *Fungal. Ecol.*, vol. 3, pp. 284-287, 2010.

[37] R. H. Nilsson, V. Veldre, Z. Wang, and 10 co-authors, "A note on the incidence of reverse complementary fungal ITS sequences in the public sequence databases and a software program for their detection and reorientation", *Mycoscience*, in press, DOI: 10.1007/s 10267-010-0086-z.

[38] M. Margulies, M. Egholm, W. E. Altman, and 53 co-authors, "Genome sequencing in microfabricated high-density picolitre reactors", *Nature*, vol. 437, pp. 376-380, 2005.

[39] L. Tedersoo, R. H. Nilsson, K. Abarenkov, and 7 co-authors, "454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases", *New Phytol.*, vol. 188, pp. 291-301, 2010.

[40] T. D. Bruns, A. E. Arnold, and K. W. Hughes, "Fungal networks made of humans: UNITE, FESIN, and frontiers in fungal ecology", *New Phytol.*, vol. 177, pp. 586-588, 2008.

[41] Morphbank (http://www.morphbank.net, 04 September 2010). Florida State University, Department of Scientific Computing, Tallahassee, FL 32306-4026 USA.

[42] P. W. Crous, W. Gams, J. A. Stalpers, V. Robert, and G. Stegehuis, "MycoBank: an online initiative to launch mycology into the 21st century", *Stud. Mycol.*, vol. 50, pp. 19-22, 2004.

[43] J. Shendure and H. Ji, "Next-generation DNA sequencing", *Nat. Biotechnol.*, vol. 26, pp. 1135-1145, 2008.

[44] R. H. Nilsson, E. Kristiansson, M. Ryberg, N. Hallenberg, and K.-H. Larsson, "Intraspecific ITS variability in the kingdom *Fungi* as expressed in the international sequence databases and its implications for molecular species identification", *Evol. Bioinform. Online*, vol. 4, pp. 193-201, 2008.

[45] K. W. Hughes, R. H. Petersen, and E. B. Lickey, "Using heterozygosity to estimate a percentage DNA sequence similarity for environmental species delimitation across basidiomycete fungi", *New Phytol.*, vol. 182, pp. 795-798, 2009.

[46] J. M. U'Ren, J. W. Dalling, R. E. Gallery, D. R. Maddison, E. C. Davies, C. M. Gibson, and A. E. Arnold, "Diversity and evolutionary origins of fungi associated with seeds of a neotropical pioneer tree: a case study for analysing fungal environmental samples", *Mycol. Res.*, vol. 113, pp. 432-449, 2009.

[47] D. H. Huson, A. F. Auch, J. Qi, and S. C. Schuster, "MEGAN analysis of metagenomic data", *Genome Res.*, vol. 17, pp. 377-386, 2007.

[48] S. Götz, J. M. García-Gómez, J. Terol, and 7 co-authors, "High-throughput functional annotation and data mining with the Blast2GO suite", *Nucleic Acids Res.*, vol. 36, pp. 3420-3435, 2008.

[49] S. Kumar, M. Nei, J. Dudley, and K. Tamura, "MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences", *Brief. Bioinform.*, vol. 9, pp. 299-306, 2008.

[50] T. R. Horton, A. E. Arnold, and T. D. Bruns, "FESIN workshops at ESA - the mycelial network grows", *Mycorrhiza*, vol. 19, pp. 283-285, 2009.

[51] M. Härlin, "Phylogenetic approaches to nomenclature: a comparison based on a nemertean case study", *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 266, pp. 2201-2207, 1998.

[52] P. D. Cantino, H. N. Bryant, K. de Quieroz, M. J. Donoghue, T. Eriksson, D. M. Hillis, and M. S. Y. Lee, "Species names in

phylogenetic nomenclature", *Syst. Biol.*, vol. 48, pp. 790-807, 1999.

[53]   M. Buée, M. Reich, C. Murat, E. Morin, R. H. Nilsson, S. Uroz, and F. Martin, "454 pyrosequencing analyses of forest soils reveal an unexpected high fungal diversity", *New Phytol.*, vol. 184, pp. 449-456, 2009.

[54]   A. Jumpponen and K. L. Jones, "Massively parallel 454-sequencing indicates hyperdiverse fungal communities in temperate *Quercus macrocarpa* phyllosphere", *New Phytol.*, vol. 184, pp. 438-448, 2009.

[55]   M. A. Ghannoum, R. J. Jurevic, P. K. Mukherjee, F. Cui, M. Sikaroodi, A. Naqvi, and P. M. Gillevet, "Characterization of the oral fungal microbiome (mycobiome) in healthy individuals", *PLoS Pathog.*, vol. 6, art. 1, 2010.

[56]   J. McNeill, F. R. Barrie, H. M. Burdet *et al.*, Ed., *International Code of Botanical Nomenclature (Vienna Code) adopted by the Seventeenth International Botanical Congress Vienna, Austria, July 2005*. Ruggell: Gantner Verlag.

[57]   P. M. Brock, H. Döring, and M. I. Bidartondo, "How to know unknown fungi: the role of a herbarium", *New Phytol.*, vol. 181, pp. 719-724, 2009.