

# How to know unknown fungi: the role of a herbarium

Patrick M. Brock<sup>1,2,3</sup>, Heidi Döring<sup>1</sup> and Martin I. Bidartondo<sup>1,2</sup>

<sup>1</sup>Royal Botanic Gardens, Kew TW9 3DS, UK; <sup>2</sup>Imperial College London, London SW7 2AZ, UK; <sup>3</sup>Natural History Museum, London SW7 5BD, UK

## Summary

Author for correspondence:

Martin I. Bidartondo

Tel: +44 208 332 5379

Fax: +44 208 332 5310

Email: m.bidartondo@imperial.ac.uk

Received: 11 August 2008

Accepted: 22 October 2008

*New Phytologist* (2009) **181**: 719–724

doi: 10.1111/j.1469-8137.2008.02703.x

**Key words:** BLAST, DNA barcoding, ecology, evolution, nuclear ribosomal internal transcribed spacer.

- The development of a universal approach to the identification of fungi from the environment is impeded by the limited number and narrow phylogenetic range of the named internal transcribed spacer DNA sequences available on GenBank. The goal here was to assess the potential impact of systematic DNA sequencing from a fungal herbarium collection.
- DNA sequences were generated from a diverse set of 279 specimens deposited at the fungal herbarium of the Royal Botanic Gardens at Kew (UK) and bioinformatic analyses were used to study their overlap with the public database.
- It is estimated that *c.* 70% of the herbarium taxonomic diversity is not yet represented in GenBank and that a further *c.* 10% of our sequences match solely to 'environmental samples' or fungi otherwise unidentified.
- Here it is shown that the unsampled diversity residing in fungal herbaria can substantially enlarge the coverage of GenBank's fully identified sequence pool to ameliorate the problem of environmental unknowns and to aid in the detection of truly novel fungi by molecular data.

## Introduction

Enormously evolutionarily and functionally diverse, yet difficult to identify based on morphology alone, the fungi present a great challenge to taxonomists and ecologists alike (Hawksworth, 2001). A universal molecular approach to their identification (White *et al.*, 1990; Gardes *et al.*, 1991; Henrion *et al.*, 1992; Druzhinina *et al.*, 2005; Kõljalg *et al.*, 2005) through what are now referred to as DNA barcodes (Herbert *et al.*, 2003; Seifert & Crous, 2008) uses the nuclear ribosomal internal transcribed spacer region (ITS) because of its practicality and wide taxonomic applicability (Gardes & Bruns, 1993; Nilsson, 2007). A targeted, more accurate molecular approach to species recognition has also been developed (Taylor *et al.*, 2000, 2006). Environmental studies are generating increasingly vast numbers of ITS sequences with no accompanying morphological characters (O'Brien *et al.*, 2005; 'environmental unknowns'), which are typically identified – or not as the case may be – by querying GenBank (Benson *et al.*, 2008) for matching sequences using BLAST (Altschul *et al.*, 1990).

Globally, it is estimated that there are between 1.5 million (based on extrapolations from the fungi–plant ratio in Britain; Hawksworth, 2001) and 3.5 million (based on environmental ITS surveys; O'Brien *et al.*, 2005) species of fungi. Less than 5% have been described (Hawksworth & Mueller, 2005) and only 17% of these are thought to be cultivable (Hawksworth,

1991). Only when every described species is represented on GenBank will researchers be able to definitively recognize when they detect a 'known' species vs a truly novel 'unknown' species using molecular techniques. Populating the fungal ITS dataset on GenBank with named, specimen-referenced sequences should also lower the unfortunately high rate of fungal species re-description (*c.* 2.5 : 1; Hawksworth, 1992).

The accumulation and propagation of naming errors associated with GenBank's fungal ITS sequences have been discussed elsewhere (Vilgalys, 2003; Holst-Jensen *et al.*, 2004; Nilsson *et al.*, 2006; Bidartondo *et al.*, 2008); however, the taxonomic coverage of the dataset is a deeper problem. Here we emphasize the fundamental role fungal herbaria can play in addressing it by showing that there is considerable fungal ITS diversity in the Kew mycology herbarium not represented on GenBank.

## Materials and Methods

A total of 509 specimens representing 215 species in 49 genera from 22 Basidiomycota families and nine Ascomycota families were sampled from the *c.* 800 000 specimen-rich Kew fungal herbarium, emphasizing British gasteroid and hypogeous fungi using floristic accounts by Pegler *et al.* (1993, 1995) and genera of grassland fungi. This sample of predominantly uncultured macrofungi comprised 38 plant pathogenic, 164 mycorrhizal and 307 saprophytic fungi,

including four type specimens. One to five specimens from different counties were chosen per species. When multiple specimens were available from a species, the youngest and largest were selected. Destructive sampling from Kew's fungal herbarium specimens is permitted by the curator on a case-by-case basis following guidelines aimed at safeguarding the collection for future users. DNA extraction and polymerase chain reaction (PCR) were carried out following Gardes & Bruns (1993), using the fungus-specific ITS1F primer and the eukaryotic ITS4 primer to amplify the nuclear ribosomal internal transcribed spacer region. The extraction was modified for silica emulsion binding and purification (Gene-Clean; Q-Biogene, Irvine, CA, USA) and the hot-start enzyme JumpStart (Sigma, St Louis, MO, USA) was used to catalyse the PCR with 2 min at 94°C, followed by 35 cycles of 30 s at 94°C, 30 s at 53°C, 45 s + 5 s per cycle at 72°C, and finishing with 5 min at 72°C. The PCR products were purified with QIAquick (Qiagen Inc., Valencia, CA, USA), sequenced bidirectionally using the PCR primers and BigDye 3.1 on an ABI3730 (Applied Biosystems, Foster City, CA, USA) and edited in SEQUENCHER 4.5 (Gene Codes, Ann Arbor, MI, USA). Sequences with a read length of < 350 bp obtained from direct sequencing of PCR products were excluded from further analysis and classified as 'failed'. PCR products that could not be sequenced directly were cloned using TOPO TA (Invitrogen, Carlsbad, CA, USA). All DNA sequences were submitted to BLAST and used to query the nucleotide collection using default settings. Those sequences that consistently matched common contaminants (e.g. *Penicillium*) were discarded and also classified as 'failed'.

Overlap between the herbarium and GenBank in terms of matching sequences was initially assessed in two ways. First, using the widely used  $\geq 97\%$  sequence similarity cut-off point for fungal species delimitation (Izzo *et al.*, 2005; O'Brien *et al.*, 2005; Morris *et al.*, 2008; Ryberg *et al.*, 2008; Walker *et al.*, 2008), we counted the number of matches between our sample and GenBank using BLAST. The top hit, the next fully identified hit and the next insufficiently identified hit were recorded. 'Fully identified' was defined as named to the species level without ambiguity. Thus, any GenBank DNA sequence accession that did not include any of the following in its title ([titl] field) was defined as fully identified: *cf.*, *sp.*, *uncultured*, *unidentified* and *fungus*. Accessions that did include any of these were therefore defined as insufficiently identified. Although the inclusion of *fungus* incorrectly placed some fully identified sequences into the insufficiently identified pool, it retrieved > 600 insufficiently identified sequences that would have otherwise remained in the fully identified pool. The first hit on the BLAST results list was assumed to represent the best match and the next fully identified and insufficiently identified matches were used to check for consistency. If these did not appear correct, the first 100 BLAST hits were examined in detail. This would highlight cases of synonymy and incorrectly identified sequences on the

database. If two names were found to be synonymous using Index Fungorum (<http://www.indexfungorum.org>), this was counted as a correct match. If it was obvious that a DNA sequence on the database was incorrectly identified (e.g. if a basidiomycete sequence from our sample produced a top hit to an ascomycete with the subsequent 99 hits to closely related basidiomycetes including conspecifics) it was ignored as top hit. When situations such as these occurred, which they did rarely, the next apparently correct hit was selected for analysis. Additionally, if alignment coverage of the three selected hits appeared too low (approx. < 70%), the alignments of the top 100 hits were examined in detail. If only a small proportion or none of the variable spacer regions (ITS1 and ITS2) were covered by the results, despite percentage matches of  $\geq 97\%$  to the conserved 5.8S rRNA gene region, the sequence was classified as nonmatching.

Second, to simulate the identification of unnamed environmental samples, a decision was made for each of our query DNA sequences as to whether the top 100 BLAST results and their alignments provided enough evidence to assign it a genus name. The judgement made for each sequence during this part of the analysis was made on a case-by-case basis. Our basic criterion was that if the top four named hits were of the same genus and had sequence similarity and alignment coverage of approx. > 70%, this counted as strong enough evidence to assign the DNA sequence a genus name during the simulation. However, all 100 results were taken into account, and we paid close attention to the relationship between the proportion of the spacer regions covered by each result and their relatedness to the query sequence. This became particularly relevant when dealing with genera with very few GenBank representatives. If, for example, a query DNA sequence from an under-represented group had two or three significant spacer matches to the same genus and 97 subsequent matches to other genera only in the 5.8S conserved gene region, this was considered sufficient evidence to assign a genus name to the query sequence during the simulation.

Furthermore, to estimate the overlap between our sample and the database in terms of taxonomic units, we used BLASTclust to calculate the number of  $\geq 97\%$  sequence clusters – a proxy for species – that (1) included only herbarium sequences, (2) included only GenBank sequences and (3) included both herbarium and GenBank sequences (see the Supporting Information, Methods S1).

To complement the analyses we examined changes in the number of fungal ITS sequences on GenBank and the proportion that were annotated with references to voucher specimens from 1998 to 2008. Sequences were downloaded with the search strings utilized for the BLASTclust analysis (Methods S1), however, the string ('"1980"[PDAT] : "2008/04/14"[PDAT]') was edited accordingly to download the data for the years 1998–2007 individually and ('"voucher" [ALL]') was added to retrieve the specimen-referenced subsets of the fully and insufficiently identified sequences.

## Results

A total of 279 specimens, representing 157 species in 40 genera from 19 families of Basidiomycota and eight families of Ascomycota, yielded uncontaminated DNA sequences (GenBank Accessions EU784164–EU784445; Table S1). We estimate that these include 22 plant pathogenic, 66 mycorrhizal and 191 saprophytic fungi. Direct sequencing of PCR products produced 263 sequences from the 509 specimens sampled, after 54 sequences were discarded because of insufficient quality. However, 21 of these 263 originated from contaminant fungi. A total of 190 PCR products were cloned and they produced 71 DNA sequences, of which 10 were from contaminants and 61 were incorporated into the analysis. This added further samples to the dataset and improved the quality of sequences obtained via direct PCR product sequencing, respectively (see Table S1 for specimen details).

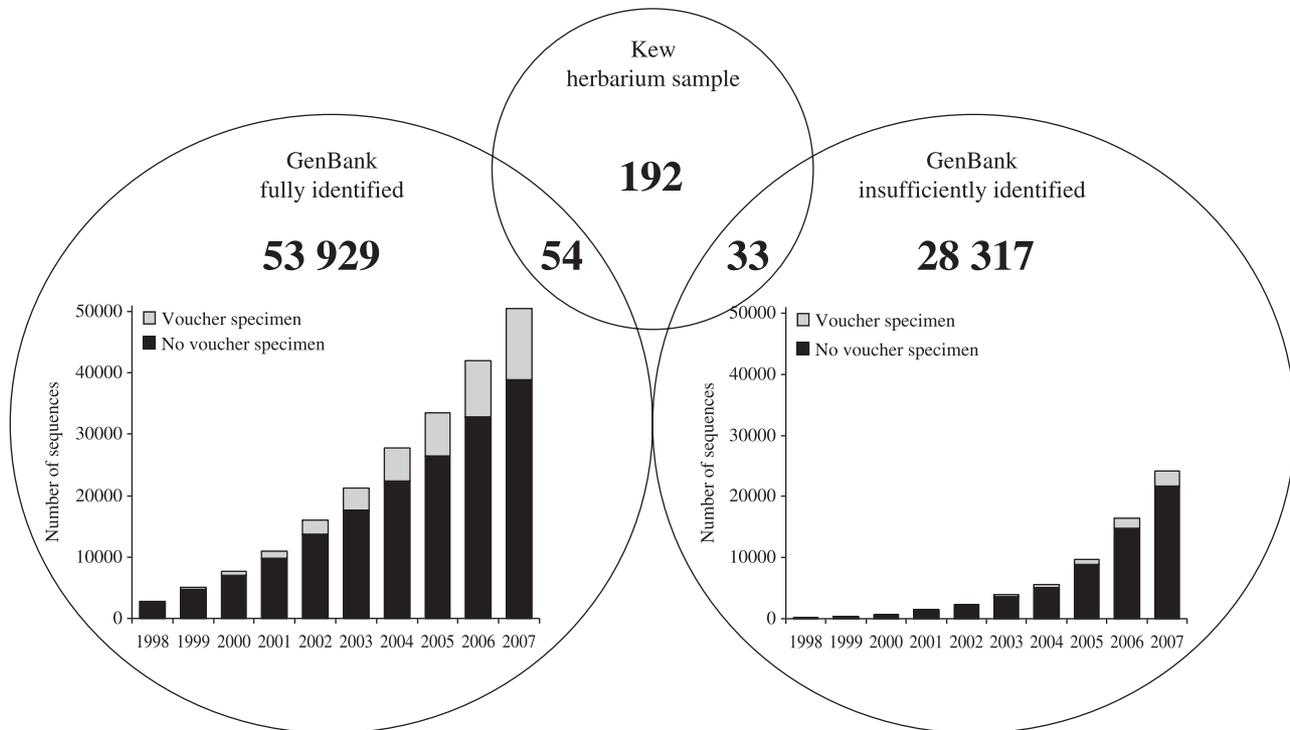
Although some of the oldest specimens in our sample were successful (55 yr old in one case), the age of specimens had a strong effect on the success of DNA sequencing (two-sample  $t$ -test,  $df = 507$ ,  $P < 0.0001$ ; Fig. S1). Older specimens were also significantly more likely to be contaminated (two-sample  $t$ -test,  $df = 306$ ,  $P < 0.0001$ ). An examination of other factors that might affect the success of DNA sequencing revealed no significant bias. All 12 orders (not including *Diehliomyces microsporus*, Ascomycota *incertae sedis*) represented by our sample included specimens that were both successful and failed ('failed' hereinafter including contaminants). Orders represented by large numbers of specimens (e.g. *Agaricales* 153 successful, 78 failed) did not display a conclusive bias. A similar pattern emerged at the family and genus level. Although there were some families and genera in which all or none of the specimens were successful, these were all groups containing no more than four specimens (except in the case of the Helvellaceae, in which all seven failed). Different taxonomic groups in our sample also had different specimen age distributions as a result of collecting bias. The Eurotiales (*Elaphomyces*), for example, were represented mostly by older specimens, as were some groups of Pezizales (particularly *Tuber*). Finally, specimens were categorized as either being gasteroid (Alexopoulos *et al.*, 1996) or open (i.e. exposed hymenium). We might expect fungi in which spores mature inside the sporocarp to be more likely to yield DNA sequences, their spores being enclosed and protected during the collection, drying and storage processes. However, this did not appear to be the case: of the successfully sequenced specimens 61% were open and 39% were gasteroid and of the failed specimens 55% were open and 45% were gasteroid.

Sixty-nine per cent of our DNA sequences did not match any GenBank accessions with  $\geq 97\%$  sequence similarity, 19% matched GenBank accessions named to the species level (henceforth 'fully identified') and 12% matched GenBank accessions which did not have species names (henceforth 'insufficiently identified'; Fig. 1). The 54 sequences that

matched fully identified accessions comprised 34 species in 21 genera and were dominated by no particular group. However, 21 of these matches were to congeneric but incorrect species (Table S2). These mismatches may be caused by species complexes such as in *Armillaria* (Cox *et al.*, 2006) but could also be the effects of inconsistencies of the BLAST algorithm (Koski & Golding, 2001) and misidentifications. The 33 sequences that matched insufficiently identified accessions were from 28 species in 13 genera and were dominated by *Tuber*. During the naming simulation at the genus level, 68% of our sequences (189) could not be confidently assigned to a genus, while just 32% (90) could be. This simulation was only able to assign a name to three more sequences than the  $\geq 97\%$  species-level match could. Congeneric sequences were not always grouped together by this method: only some *Hymenogaster* sequences, for example, could be confidently assigned to a genus.

We estimated the number of species names in GenBank's fully identified sequence pool by downloading the GenBank title for all 53 983 sequences ([titl] field; 14 April 2008) and removing all but the genus and species names. The number of unique genus–epithet combinations was then counted, and it resulted in a total of 14 223 fungal names. This estimate does not take into account synonymy and variations in spelling. Even using the most conservative estimate for the total number of fungal species that have been described (74 000; Hawksworth & Mueller, 2005), this suggests that a mere 19% of the known fungal species are represented by fully identified ITS sequences on GenBank.

The results of the BLASTclust analysis performed to estimate taxonomic overlap between our herbarium sample and the sequence pools in GenBank (Methods S1) were similar to the BLAST analyses estimating overlap based on individual sequence matches. In addition, this evaluation revealed that the 157 morphologically identified species in our sample formed 173 clusters and that the estimated 14 223 species in GenBank's fully identified sequence pool formed 37 427 clusters (Fig. S2). Although misidentifications may contribute, it is likely that there are more clusters than species names owing to taxonomic problems, such as cryptic species and poorly defined morphologically variable species complexes, because within-species ITS variation exceeds 3% in some fungal lineages, and due to introns. Eighty-four of the 157 morphological species in our sample formed fully congruent molecular clusters and 73 were either subdivided or grouped with other species of the same genus (Table S3). Within the last 3–5 yr the number of insufficiently identified sequences deposited in GenBank has rapidly increased largely owing to 'environmental unknowns' most of which are not accompanied by voucher information (Fig. 1). The BLASTclust analysis showed that just 1.9% (691) of the clusters that included DNA sequences from GenBank's fully identified sequence pool also included sequences from GenBank's insufficiently identified sequence pool, whereas 15.6% (27) of the clusters that included sequences



**Fig. 1** Results of the BLAST analysis performed on 14 April 2008, showing the number of sequences in each category (i.e. DNA sequence pool) and their overlaps based on  $\geq 97\%$  DNA sequence similarity. The plots show how the numbers of fully (53 983; 66%) and insufficiently (28 350; 34%) identified fungal nuclear ribosomal internal transcribed spacer sequences available on GenBank, and the proportions of these associated with voucher specimens (tinted), have changed since 1998.

from our herbarium sample also included insufficiently identified sequences from GenBank.

## Discussion

We have shown that there is considerable fungal ITS diversity in our herbarium sample which is not represented on GenBank and that the exploitation of the herbarium as a genetic resource has the potential to substantially enlarge the taxonomic coverage of GenBank's fully identified specimen-referenced DNA sequence pool. Based on one of the most thorough molecular fungal surveys to date, O'Brien *et al.* (2005) hypothesized that most unidentifiable ITS sequences belong to known fungi which are not yet represented by ITS data on GenBank. Our results corroborate this view. In fact, 30 of our sequences (including 17 *Hygrocybe* species from Britain) have their top BLAST matches to 'uncultured basidiomycete' sequences generated by O'Brien *et al.* (2005) from North American forests.

The 192 herbarium specimen DNA sequences that did not match any accessions in the database comprised 95 morphological species from 28 genera and were dominated by *Hygrocybe* (76 sequences) and *Geastrum* (32 sequences). There was nothing on the database that matched any of our *Geastrum* sequences with  $\geq 97\%$  similarity, and this was also the case with 14 other genera. Despite being among the most eye-catching of macro-

fungi, both *Hygrocybe* (waxcaps) and *Geastrum* (earthstars) are examples of 'orphan' genera which have not been targeted by molecular phylogenetic studies so far, or for which the ITS has not been favoured.

At present, ecologists are often forced to name environmental fungal samples only to the genus level because of a lack of sufficient reference DNA sequences in the public database. A genus naming simulation may therefore be a more appropriate way of testing the database as a tool for molecular identification. Even using such a relaxed matching criterion it appears again that *c.* 70% of the diversity of the fungal herbarium sample is not represented on GenBank. Conservatively, this means that Kew alone may have > 4500 additional species, collected during the last 10 yr and readily suitable for DNA sequencing, not yet represented on the public database.

Undoubtedly, because our sample was not random, our estimates may be imprecise. A truly random sample would require fully databasing the entire herbarium first. However, it will remain more efficient, and more valuable from evolutionary and ecological perspectives, to select genera – and not just individual species or specimens – that are (1) well represented in the herbarium by recent collections and (2) covered by up-to-date regional checklist, floristic or monographic treatments that have been conducted in conjunction with revisions of the material in the herbarium. Sampling macrofungi for DNA barcoding from well-annotated herbarium specimens requires

only minimal training and can largely be performed by students and volunteers; in our study, sampling from specimens was by far the most time-consuming step. Along with targeting taxa unrepresented or poorly represented in GenBank, focusing on regional mycotas will be a promising strategy for many herbaria to maximize success rates, because of the increased likelihood of holding numerous recent collections. More accurate estimates of worldwide fungal diversity and better species delimitations will also result from this approach. Given the mismatches between morphological species and molecular clusters found in this study, we encourage the sampling of multiple specimens from each species. This helps to identify potentially misidentified specimens, but more importantly, it highlights species complexes, cryptic species and yet undescribed taxa that are in need of revision by mycologists.

Only 1.4% of all clusters which included GenBank accessions grouped both fully and insufficiently identified DNA sequences (Methods S1; Fig. S2). This lack of taxonomic overlap between GenBank's two sequence pools may arise because two distinct types of mycological study contribute to GenBank: environmental surveys, which deposit diverse sets of unnamed DNA sequences, and phylogenetic studies, which deposit related sets of named sequences. Our data show that fungal herbaria are instrumental in informatively linking the taxonomic breadth and metadata associated with the former (cf. Ryberg *et al.*, 2008) with the specimens and detailed taxonomic data of the latter (cf. Kausserud *et al.*, 2008), because the herbarium sample analysed here overlaps with both the pool of identified and insufficiently identified sequences in the database. Using the ITS diversity in fungal herbaria will not only allow us to build a better reference database for identifying 'unknowns' from the environment, it will allow us to discover yet undescribed fungal species to investigate fungal diversity in depth. It will also enable researchers to examine the morphological features related to their unvouchered environmental samples.

## Acknowledgements

We thank the Royal Society for a small research grant to M.I.B. ('Closing the sequence gap: fungal DNA barcoding at Kew'), Anja Bus, Richard Waterman and Begoña Aguirre-Hudson for help with sampling and databasing, Doug Chesters for help with the analysis, and Ian Alexander and three anonymous reviewers for comments on the manuscript.

## References

- Alexopoulos CJ, Mims CW, Blackwell M. 1996. *Introductory mycology*, 4th edn. Hoboken, NJ, USA: John Wiley & Sons Inc.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2008. GenBank. *Nucleic Acids Research Database Issue* 36: D25–D30.
- Bidartondo MI, Bruns TD, Blackwell M, Edwards I, Taylor AFS, Horton T, Zhang N, Kõljalg U, May G, Kuyper TW *et al.* 2008. Preserving accuracy in GenBank. *Science* 319: 1616.
- Cox KD, Scherm H, Riley MB. 2006. Characterization of *Armillaria* spp. from peach orchards in the southeastern United States using fatty acid methyl ester profiling. *Mycological Research* 110: 414–422.
- Druzhinina IS, Kopchinskiy AG, Komón M, Bissett J, Szakacs G, Kubicek CP. 2005. An oligonucleotide barcode for species identification in *Trichoderma* and *Hypocrea*. *Fungal Genetics and Biology* 42: 813–828.
- Gardes M, Bruns TD. 1993. ITS primers with enhanced specificity for basidiomycetes: application to the identification of mycorrhizae and rusts. *Molecular Ecology* 2: 113–118.
- Gardes M, Fortin J, White TJ, Bruns TD, Taylor JW. 1991. Identification of indigenous and introduced symbiotic fungi in mycorrhizae by amplification of the internal transcribed spacer. *Canadian Journal of Botany* 69: 180–190.
- Hawksworth DL. 1991. The fungal dimension of biodiversity: magnitude, significance and conservation. *Mycological Research* 95: 641–655.
- Hawksworth DL. 1992. The need for a more effective biological nomenclature for the 21st century. *Botanical Journal of the Linnean Society* 109: 543–567.
- Hawksworth DL. 2001. The magnitude of fungal diversity: the 1.5 million species estimate revised. *Mycological Research* 105: 1422–1432.
- Hawksworth DL, Mueller GM. 2005. Fungal communities: their diversity and distribution. In: Dighton J, White JF, Oudemans P, eds. *The fungal community: its organisation and role in the ecosystem*. Boca Raton, FL, USA: CRC Press, 27–37.
- Henrion B, Le Tacon F, Martin F. 1992. Rapid identification of genetic variation of ectomycorrhizal fungi by amplification of ribosomal RNA genes. *New Phytologist* 122: 289–298.
- Herbert PDN, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, B* 270: 313–321.
- Holst-Jensen A, Vrålstad T, Schumacher T. 2004. On reliability. *New Phytologist* 161: 11–13.
- Izzo A, Agbowo J, Bruns TD. 2005. Detection of plot-level changes in ectomycorrhizal communities across years in an old-growth mixed-conifer forest. *New Phytologist* 166: 619–630.
- Kausserud H, Stige LC, Vik JO, Økland RH, Høiland K, Stenseth NC. 2008. Mushroom fruiting and climate change. *Proceedings of the National Academy of Sciences, USA* 105: 3811–3814.
- Kõljalg U, Larsson K-H, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjøller R, Larsson E *et al.* 2005. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist* 166: 1063–1068.
- Koski LB, Golding GB. 2001. The closest BLAST hit is often not the nearest neighbour. *Journal of Molecular Evolution* 52: 540–542.
- Morris HM, Smith ME, Rizzo DM, Rejmánek M, Beldsoe CS. 2008. Contrasting ectomycorrhizal fungal communities on the roots of co-occurring oaks (*Quercus* spp.) in a California woodland. *New Phytologist* 178: 167–176.
- Nilsson RH. 2007. *Fungal taxonomy and systematics in the digital era with an outlook on the cantharelloid clade (Basidiomycota)*. PhD thesis, Göteborg University, Göteborg, Sweden.
- Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson K-H, Kõljalg U. 2006. Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *PLoS ONE* 1: e59.
- O'Brien HE, Parrent JL, Jackson JA, Moncalvo J-M, Vilgalys R. 2005. Fungal community analysis by large-scale sequencing of environmental samples. *Applied and Environmental Microbiology* 71: 5544–5550.
- Pegler DN, Læssøe T, Spooner BM. 1995. *British puffballs, earthstars and stinkhorns; an account of the British gasteroid fungi*. Richmond, UK: Kew Publishing.
- Pegler DN, Spooner BM, Young TWK. 1993. *British truffles; a revision of the British hypogeous fungi*. Richmond, UK: Kew Publishing.

- 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: 50–64.
- Seifert KA, Crous PW. 2008. The All-Fungi DNA barcoding campaign (FunBOL). *Persoonia* 20: 106.
- Taylor JW, Jacobson DK, Kroken S, Kasuga T, Geisler DM, Hibbett DS, Fisher MC. 2000. Phylogenetic species recognition and species concepts in fungi. *Fungal Genetics and Biology* 31: 21–32.
- Taylor JW, Turner E, Townsend JP, Dettman JR, Jacobson D. 2006. Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences* 361: 1947–1963.
- Vilgalys R. 2003. Taxonomic misidentification in public DNA databases. *New Phytologist* 160: 4–5.
- Walker JF, Miller OK Jr, Horton JL. 2008. Seasonal dynamics of ectomycorrhizal fungus assemblages on oak seedlings in the southeastern Appalachian Mountains. *Mycorrhiza* 18: 123–132.
- White TJ, Bruns TD, Lee S, Taylor JW. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenies. In: Innis MA, Gelfand DH, Sninsky JJ, White T. eds. *PCR protocols: a guide to methods and applications*. San Diego, CA, USA: Academic Press, 135–322.

## Supporting Information

Additional supporting information may be found in the online version of this article.

**Fig. S1** Relationship between success in obtaining nuclear ribosomal internal transcribed spacer DNA sequence data and the age of specimens.

**Fig. S2** Results of the BLASTclust analysis, showing the number of clusters or taxonomic units – a proxy for species – in each category (i.e. DNA sequence pool) and their overlaps based on  $\geq 97\%$  DNA sequence similarity.

**Table S1** Specimens which yielded internal transcribed spacer DNA sequences used in the analysis

**Table S2** Summary of the 21 ( $\geq 97\%$ ) BLAST matches which were to the wrong species

**Table S3** Output from the BLASTclust analysis showing only the herbarium sequences

**Methods S1** BLASTclust analysis.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.



## About *New Phytologist*

- *New Phytologist* is owned by a non-profit-making **charitable trust** dedicated to the promotion of plant science, facilitating projects from symposia to open access for our Tansley reviews. Complete information is available at [www.newphytologist.org](http://www.newphytologist.org).
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as-ready' via *Early View* – our average submission to decision time is just 29 days. Online-only colour is **free**, and essential print colour costs will be met if necessary. We also provide 25 offprints as well as a PDF for each article.
- For online summaries and ToC alerts, go to the website and click on 'Journal online'. You can take out a **personal subscription** to the journal for a fraction of the institutional price. Rates start at £139 in Europe/\$259 in the USA & Canada for the online edition (click on 'Subscribe' at the website).
- If you have any questions, do get in touch with Central Office ([newphytol@lancaster.ac.uk](mailto:newphytol@lancaster.ac.uk); tel +44 1524 594691) or, for a local contact in North America, the US Office ([newphytol@ornl.gov](mailto:newphytol@ornl.gov); tel +1 865 576 5261).