
On the Unreliability of Published DNA Sequences

Author(s): Paul D. Bridge, Peter J. Roberts, Brian M. Spooner and Gita Panchal

Source: *The New Phytologist*, Oct., 2003, Vol. 160, No. 1 (Oct., 2003), pp. 43–48

Published by: Wiley on behalf of the New Phytologist Trust

Stable URL: <https://www.jstor.org/stable/1514203>

REFERENCES

Linked references are available on JSTOR for this article:

https://www.jstor.org/stable/1514203?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

New Phytologist Trust and Wiley are collaborating with JSTOR to digitize, preserve and extend access to *The New Phytologist*

On the unreliability of published DNA sequences

Paul D. Bridge^{1,2,3}, Peter J. Roberts², Brian M. Spooner² and Gita Panchal¹

¹School of Biological and Chemical Sciences, Birkbeck College, Malet Street, London WC1E 7HX, UK; ²Mycology Section, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, UK; ³Present address, British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Rd, Cambridge CB3 0ET, UK

Summary

Author for correspondence:

Paul Bridge

Tel: +44 1223 221 567

Fax: +44 1223 221 259

Email: pdb@bas.ac.uk

Received: 17 March 2003

Accepted: 16 June 2003

doi: 10.1046/j.1469-8137.2003.00861.x

- Here, the reliability of published fungal nucleic acid sequences is tested by the critical re-evaluation of 206 named sequences obtained from public-access databases.
- Sequences from the ribosomal RNA (rRNA) gene cluster were examined as these are commonly used to establish fungal phylogeny and evolution, and are also increasingly employed in the identification of fungi from nonculture based studies.
- Fifty-one rRNA internal transcribed spacer (ITS) sequences were obtained for species of *Amanita*, 55 ITS sequences were obtained for species of *Phoma* and 100 rRNA small subunit sequences were obtained from representative genera of the order Helotiales. In each case, the fungal group was selected partly on the basis of sequences deposited by three or more laboratories in order to avoid sample bias. The results suggest that up to 20% of the sequences available for each group may be unreliable, and this proportion is supported by additional informal observations.

Key words: DNA sequences, fungi, systematics, reliability, ribosomal RNA, databases.

© *New Phytologist* (2003) **160**: 43–48

Introduction

Nucleic acid sequences deposited in public-access databases are widely used by systematic and evolutionary biologists to construct systematic frameworks and model evolutionary pathways. They are also increasingly used for routine identification of cryptic organisms (e.g. fungi found in soil samples) by simple sequence matching. Ideally, public-access databases should provide a working archive of available sequences, forming a valuable resource, analogous to herbarium and culture collections, for current and future researchers. Recent publications have stressed that the continuing generation of such data is crucial to the development of systematics (Tautz *et al.*, 2002; Blaxter, 2003). It is therefore of primary importance that the sequences deposited in reference databases are accurate and reliable. Recently, a few reports have questioned the identity of some fungal data (Deckert *et al.*, 2002; Crous, 2002) and highlighted the accumulation of chimeric bacterial 16S sequences in public access databases (Hugenholtz & Huber, 2003).

The publicly available DNA sequences for fungi are an example of a very incomplete data set. Estimates for the total number of fungal species vary, but a generally accepted figure is 80 000 described to date, with possibly as many as 1.4

million awaiting discovery (Kirk *et al.*, 2001). However, at the time of writing, only 13 764 species of fungi have any DNA sequences deposited in GenBank (<http://www.ncbi.nlm.nih.gov/taxonomy/txstat.cgi>).

Fungi also provide a good example of organisms where sequence information can be difficult to interpret. There are several reasons for this. First, different levels of variation may occur in the same DNA region in different taxa, resulting in problems in making generalised comparisons between taxa (Bridge, 2002); second, most fungi occur in close association with other organisms (including other fungi), providing ample opportunity for cross-contamination; and third, many fungi are known only from environmental collections, making it difficult to confirm sequence results.

Fungal systematics and phylogeny have generally been based on a restricted number of genomic regions, with the conserved ribosomal RNA (rRNA) subunit genes being used almost universally for familial and higher-level studies. Similarly, the variable spacers between and within the subunits have been widely applied for comparisons within genera or species where the very highly conserved gene sequences may not provide sufficient discrimination to unequivocally identify organisms (Bruns *et al.*, 1991; Bainbridge, 1994).

The identity of some fungal DNA sequences deposited in public databases has recently been contested (Deckert *et al.*, 2002; Crous, 2002) and, as a result, there is a need to determine if such reports reveal a widespread phenomenon. In this study we show that up to 20% of publicly available, taxonomically important DNA sequences for three randomly chosen groups of fungi may be incorrectly named, chimeric, of poor quality or too incomplete for reliable comparison.

Methods

Selection criteria

In order to test the taxonomic integrity of the available data, three data sets of sequences used for fungal systematics and phylogeny were constructed. These data sets were obtained by collecting together particular DNA sequences held in the Fungi subset of the EMBL sequence database.

Fungal data sets were selected so as to contain sequences deposited at different times, from different laboratories worldwide in order avoid reflecting the working practice or standards of a single laboratory. The major criteria for selecting the genomic regions and fungal groups were to examine different sequences that have routinely been used for systematic studies, and fungal groups where there were sufficient independent data from named material to allow comparisons to be made. One-hundred small subunit (SSU) sequences were available for 50 genera of the Helotiales, 55 internal transcribed spacer (ITS) sequences were available for species of the genus *Phoma*, and 51 ITS sequences were available for 31 species of *Amanita*.

These are relatively large data sets for fungal sequences, but all are incomplete in that 88 genera are currently placed in Helotiales s.l., some 200 species or forms have been described in *Phoma* and some 200 species have been described in *Amanita*.

Sequence retrieval and comparison

Sequences (Table 1) were obtained by repeated key word searching in EMBL with the Sequence Retrieval Service. The identity of the sequences was initially tested by comparison with the 49 nearest matches in the EMBL Fungi database as disclosed by the FASTA facility (Pearson & Lipman, 1988). Sequences were considered identified if they showed greater than 90% similarity with those of at least three closely related fungi. In the event of ambiguous results, sequences were further checked with the WU-BLASTN facility. Where these fungi were closely related to the species being considered, the sequence was scored as identified. Conversely, where the best matches were to fungi unrelated to the sequence under consideration, sequences were provisionally scored as misidentified or equivocal before examination of alignments. All close matches were aligned in CLUSTALW (Higgins *et al.*, 1994), and alignments were checked through JalView (<http://www.ebi.ac.uk/jalview/index.html>). All comparisons were performed interactively through the European Bioinformatics Institute web site (<http://www.ebi.ac.uk>).

Results and Discussion

Fifty-five ITS sequences were obtained for species of *Phoma*, a genus of anamorphic ascomycetes including some important

Table 1 Accession numbers of sequences used

	Accession number
<i>Phoma</i> sequences used	AY183371, AY183372, AY206462, AY207331, AY207336, AY210335, AY210347, AF046013, AF046016, AF046017, AF046018, AF046019, AF046020, AF046021, AF046022, AF079775, AF079892, AF079893, AF079894, AF079895, AF126816, AF126819, AF149931, AF149932, AF149933, AF149934, AF218789, AF218792, AF268182, AF268183, AF268184, AF268185, AF268186, AF268187, AF268188, AF268189, AF268190, AF268191, AF268192, AF272552, AF272553, AF272554, AF423119, AY089978, AY131199, AY131200, AY160204, AJ301701, AJ301711, AJ310556, AJ310557, U19274, L38711, AJ310558, AJ310559
<i>Amanita</i> sequences used	AB015676, AB015677, AB015678, AB015679, AB015680, AB015681, AB015682, AB015683, AB015684, AB015685, AB015686, AB015687, AB015688, AB015689, AB015690, AB015691, AB015692, AB015693, AB015694, AB015695, AB015696, AB015697, AB015698, AB015699, AB015700, AB015701, AB015702, AF085428, AF085483, AF085484, AF085485, AF085486, AF085487, AF085488, AF085489, AF085490, AF085491, AF085492, AF085493, AF085494, AF335440, AF438557R1, AF438558R, AF438559R, AF438560R, AF438561R, AF438562R, AF438563R, AF438593R, Z54294, AJ308097
<i>Helotiales</i> sequences used	AF106015, AF107343, AF113715, AF113716, AF203456, AF203459, AF203460, AF203461, AF203462, AF203464, AF203466, AF203467, AF222515, AF222532, AF222533, AF281072, AF292087, AF292088, AF292089, AF292091, AF292092, AF292093, AJ224362, L37536, U45444, U45445, U46031, U67427, U67428, U67429, U67430, U67431, U67432, U67433, U67434, Z30239, Z30240, Z73711, Z73712, Z73713, Z81372, Z81375, Z81382, Z81383, Z81384, Z81385, Z81389, Z81390, Z81393, Z81395, Z81398, Z81400, Z81401, Z81376, Z81377, Z73714, Z81378, AF292090, Z81379, Z81381, Z73715, Z73716, Z73717, Z73718, Z73719, Y14211, Y14210, AF010505, Z73722, Z73723, Z73724, Z73726, Z73727, Z73729, Z73730, Z73731, AJ226069, AJ226070, Z81386, Z73733, Z73734, Z73735, Z81392, Z81397, L37541, X69850, Z73738, Z81399, U45438, Z81380, Z81388.

¹R indicates sequences deposited as 3'–5' and converted to reverse complement for analysis.

Table 2 Identification of 100 rRNA small subunit (SSU) sequences deposited in the EMBL database as members of the *Helotiales*, 51 rRNA internal transcribed spacer (ITS) sequences deposited as from species of *Amanita* and 55 rRNA ITS sequences from *Phoma*

	<i>Helotiales</i>	<i>Amanita</i>	<i>Phoma</i>
Number of sequences	100	51	55
Misidentified	7 + 8 ¹	8	3 + 3 ²
Dubious/chimeric	4	5	3
Too short to align	5	5	14
Unpublished	21	11	33
Unvouchered (in public collections)	30	19	30

¹See specific comments in the text. ²Includes three sequences consisting of only ITS1 and 5.8 s regions.

crop and livestock pathogens (see Table 1). Of these, 14 were incomplete, consisting of either the ITS1 and 5.8 s gene or the ITS2 and 5.8 s gene only. One of these sequences gave only moderate matches (83–90%) to other anamorphic ascomycetes, and two others gave high matches (96–99%) to unidentified Dothideomycetes and species of *Paraphaeosphaeria*. It therefore seems unlikely that these three sequences are correctly identified. None of these 14 incomplete sequences were used in subsequent alignments because of the data reduction required to obtain sequences of comparable length.

The remaining 41 sequences were aligned through CLUSTAL-W and an average distance tree (Fig. 1) was produced. Three of the sequences had previously been identified as either ‘cf *Phoma*’ or ‘*Phoma*-like coelomycetes’. These three sequences were 98.8% similar to each other and came closest to a single record of *Guignardia philoprina* (98.8%) or to *Pezicula* and *Neofabraea* species (87–92%). The single *Guignardia* sequence (AB041243) is likely to be a misidentification as it is significantly different from the other 68 complete ITS sequences available for this genus, and showed close homology with *Pezicula* and *Neofabraea* species and *Cryptosporiopsis* anamorphs. It seems likely that the three *Phoma*-like organisms are actually anamorphs of *Pezicula* or *Neofabraea* species. An additional *Phoma* sequence gave its 49 best matches to *Alternaria* species, and two further sequences gave best matches to *Phomopsis* (99.8%). *Phomopsis* species are morphologically similar to *Phoma*, and it would seem likely that these were simply misidentified. Sequences received as *Phoma epicoccina* gave their best matches to *Epicocum nigrum* (100%), and this supports the proposed synonymy of these species (Arenal *et al.*, 2000). Two identical sequences were labelled as *Phoma wasa-biae*. This species is of uncertain affiliation; it is not listed in the Index of Fungi and it was transferred to the genus *Plenodomus* by Reddy *et al.* (1998), partly on the basis of rRNA sequences, although the genus *Plenodomus* has also been considered as a section of the genus *Phoma* by van der Aa *et al.* (1990). The sequence database details available list the strain sequenced as ‘Yokogi’ in both instances, and it is possible that both are derived from the same original material. The sequences showed 92–95% homology with those from *Leptosphaeria* and 89% similarity with those from *Phoma tracheiphila*, a species of anamorphic *Leptosphaeria*. Given

these concerns it is not possible to make a definitive statement regarding the identity of this sequence (Table 2).

Eight of the 51 sequences of *Amanita*, a genus of agarics, including several lethally poisonous toadstools, were considered misidentified. Seven of these gave consistent close matches to other larger fungi and one matched a different species of *Amanita*. All the incorrect *Amanita* sequences had originally been obtained from cultures, and this may be of particular significance since *Amanita* isolates generally grow very slowly in culture while the suggested true taxa were all fungi that grow relatively rapidly.

A further five *Amanita* sequences, all from a single study, gave equivocal identifications in that they showed good matches to each other, but poor matches to sequences from other laboratories for any agaric. None of the species concerned had been sequenced by other laboratories, so this finding is difficult to interpret because of the small proportion of fungal species for which information is available. In *Amanita* species, variable rates of molecular divergence have been reported within nuclear rRNA sequences (Moncalvo *et al.*, 2000). While this may have some effect on the use of ITS sequences within this genus, the lack of close matches to other agarics would suggest that this is unlikely.

The remaining 35 *Amanita* sequences, which were assumed to be identified correctly, could be aligned over most of the ITS1/5.8 s/ITS2 region, but five sequences were only of sections of ITS1 or ITS2 regions and could not be adequately compared.

Seven of the 100 SSU sequences from the *Helotiales*, an order of ascomycetous ‘cup-fungi’, were considered misidentified. Six of these matched with existing unrelated taxa, and one failed to identify satisfactorily with any fungal SSU sequence. Three of the seven incorrect *Helotiales* sequences were described from cultures. Many of the fungi in this order are closely associated with higher plants, and the suggested correct identifications were generally fast-growing fungi that could be reasonably expected to occur on the hosts.

Thirteen additional *Helotiales* sequences gave equivocal identifications. Sequences of eight of the nine representatives of the genus *Myriosclerotinia* showed good matches to other sequences obtained in the same study from the same laboratory, but either no match, or very low matches, to other *Helotiales*. This could suggest that this genus is unrelated to others

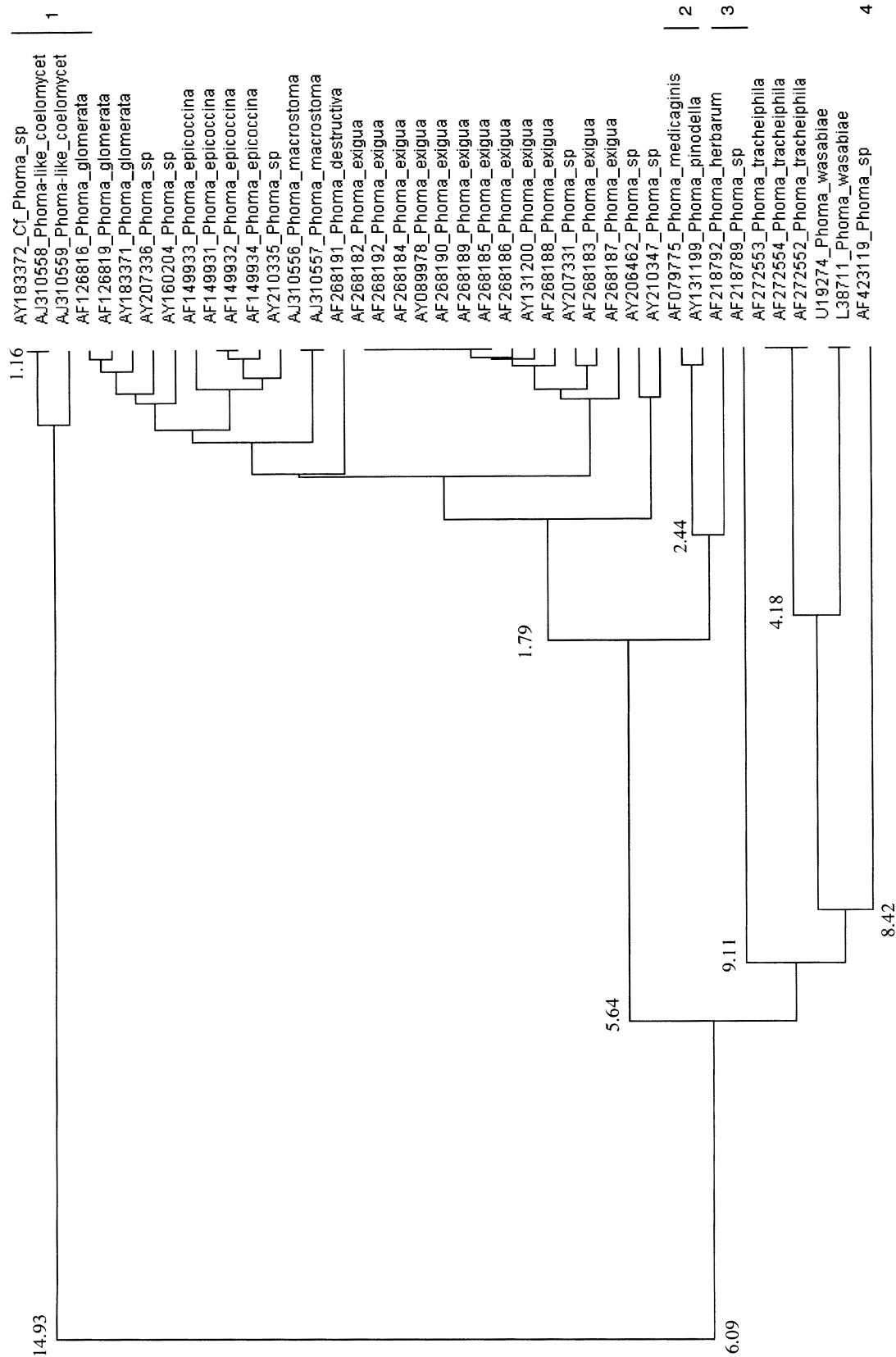


Fig. 1 Average distance tree of 41 aligned internal transcribed spacer (ITS)1/5.8 s/ITS2 sequences from species of *Phoma* (distances of major branches shown). 1, Sequences identified as *Peziza*/*Neofabraea* anamorphs; 2, sequences identified as *Phomopsis*; 3, two sequences that show closest matches to mixed *Leptosphaeria*/*Parphaeosphaeria* species; 4, sequence identified as *Alternaria*.

currently placed in the order, but this is not supported by classical morphology. The ninth sequence from this genus was obtained in a different study, and identified closely with the genus *Monilinia*. This is the expected result as *Monilinia* and *Myriosclerotinia* are morphologically similar and are placed together in the Sclerotiniaceae. This finding therefore supports the suggestion that the eight ambiguous sequences have not been correctly identified. All of these sequences were obtained from cultures and so could again be the result of faster-growing contaminants in the original samples. The five remaining equivocal sequences all contained large runs of 20 or more unidentified bases, generally towards the centre of the sequence. None of these sequences gave acceptable matches to any known fungi when considered as single molecules. When the sequence was subdivided, and matches were attempted for the sequences before or after the ambiguous regions, either the 5' or the 3' region generally matched with other members of the *Helotiales*, and the remaining fragments matched with a variety of fungal groups. One explanation for this could be that the sequencing reactions were performed on mixed DNA species, resulting in chimeric sequences, as already reported for comparable bacterial sequences (Hugenholtz & Huber, 2003).

The complete rRNA SSU in *Helotiales* is approximately 1600–1700 bp, and the database sequences varied in length from 483 bp to 1700+ bp. As a result, only 75 of the 80 presumed correct sequences were of sufficient length, and showed enough overlap, to allow a central region of approximately 500 bp to be aligned.

Of the 206 sequences obtained for all three groups, only 141 had been included in published studies and only 127 included reference to material available in either national herbaria or culture collections. This last figure may be artificially low, as some published sequences may be linked to reference collections in the publication, but this information might not have been deposited in the database.

The results presented here are summarized in Table 2 and show that there is a significant number of fungal rRNA sequences, currently deposited in databases, that may be incorrectly named. This raises questions as to what extent these sequences can be used as reliable reference sources without prior checking of the sequence, and in some cases the associated literature.

Whether this finding is representative of the integrity of the database as a whole is open to question, but it is probable that it is. The three groups used in this study were selected on the criteria that (1) they included sequences from different geographic regions, (2) that they included sequences generated independently from three or more laboratories, and (3) they included sufficient named organisms to allow interpretative comparisons. However, it was noticeable that some single, clearly unrelated sequences, were included in the best matches for correctly identified sequences during the FASTA/BLAST searches. When these individual sequences were further checked they were also clearly misidentified, as was noted above for the

single *Guignardia* sequence. This strongly suggests that there are misidentified sequences for many genera other than those studied here.

The overall figures found here (32% of sequences unpublished, around 38% not linked to traceable material and around 20% misidentified or dubious) appear remarkably consistent between the three groups, and similar figures have been obtained in a further study on the ascomycete genus *Hypoxylon* (results not shown). In this study, 41 sequences out of 206 (around 20%) appeared to be misidentified, dubious or chimeric. Twenty-seven of these sequences were deposited in the period 1996–99, whereas only 14 were deposited in the following 3-yr period (2000–02). However, as sequences were deposited unevenly over these periods, this does not necessarily imply that there are more problems with older sequences. In *Phoma*, only 23 sequences had been submitted by the end of 1999, of which only two are placed in the misidentified/dubious categories, whereas 32 sequences have been deposited since 1999, of which seven can be considered in the misidentified/dubious category.

The generation of incorrect sequences would appear to happen for several different reasons. The simplest of these is misidentification of the original material, but others may result from poor isolation techniques. In nature, fungal species are frequently associated with other organisms, including other fungi, as parasites and symbiotic partners, or occur on plant material that supports a variety of fungi. As a result, direct culture from fresh collections can result in the growth of one of these associated fungi, as opposed to the intended species. Similarly, in dried specimens on woody substrates, the DNA of nontarget endophytic fungi may be better protected than that of the target fungi. These situations can easily result in the extraction or amplification of incorrect or chimeric DNA sequences.

There remains a significant problem in verifying DNA from fungal species due to the very incomplete nature of the reference data available. The 13 764 species for which any sequence information is available constitute around 10–15% of the named fungal species, and less than 1% of the estimated total number of species that may exist (as noted above). As a result, initial comparisons of new sequences will often show either low matches with existing sequences, or low matches with a diverse range of organisms because of background similarity from conserved sequences. This is complicated by systematic uncertainty, as illustrated by the ascomycetous species *Endomyces scopularum*. This is a mycoparasite occurring on other fungi and in the past the genus has been taxonomically associated with the Ophiostomatales, although this has since been discounted (Kirk *et al.*, 2001). There are three SSU sequences from isolates labelled as *E. scopularum* in the EMBL database, and one of these, AF267227, shows greater similarity to sequences from *Ophiostoma* species than it does to the other two (Suh *et al.*, 2001). However, the taxonomic uncertainty in this group makes it impossible to say if this is a

misidentified isolate, if the other two are incorrect or if the species concept is polyphyletic

A major concern is the relatively small proportion of sequences deposited for which a specimen or culture is readily available for examination. The deposition, in a public-access collection, of material used to derive sequences has been recommended on a number of occasions (Agerer *et al.*, 2000), and this would provide a mechanism to allow some verification. This should also be extended to include the deposition of both original collections and any cultures derived from them. Deposition of material is, however, not possible for the increasing numbers of sequences obtained from clone libraries from environmental studies. Wherever possible such sequences will need to be compared with existing material, but it must also be accepted that this will not always be possible.

A further concern identified here is the deposition of strain or specimen details that are easily traceable. In the case of *Phoma wasabiae* it would appear that the sequence from the 'Yokogi' strain has been deposited with two different accession numbers. However, sequences from the same isolate can also be deposited under different strain numbers. One example of this is the two ITS1/5.8 s/ITS2 sequences held in EMBL for *Hypoxylon atroroseum*. One is described as ATCC 76081 from a major culture collection, and the second is described as from strain H6A. This latter sequence was included in a published study, where the strain list identifies H6A as ATCC 76081 (Sanchez-Ballesteros *et al.*, 2000). While both sequences are examples of the designation of traceable material, further access to the literature is required to determine that they are both sequences from the same isolate. This leads to a redundancy in the database that is not evident from the data itself, particularly in this case where the two sequences differ by one indel, one transversion and one unresolved base. This may seem surprisingly large for sequencing errors over approximately 800 bases, but this level of difference is reflected in other instances of repeated sequencing of a single fungal isolate.

This study has shown that while published sequences can be used subsequently, a degree of comparison, and in some cases recourse to other literature, may be necessary before the sequences can be accurately interpreted. In this work, sequences of defined taxonomic groups were used, and the numbers considered were relatively small. However, there are some potentially significant implications for the interpretation of sequence data from environmental studies, particularly those where large numbers of unrelated sequences may be obtained. DNA sequences are already playing an important role in systematics and routine identifications, and this will undoubtedly increase in the future. It is essential, therefore, that the sequences deposited should be accurate, or at least not misleading. Many herbaria and culture collections already have protocols relating to the identification, verification and deposition of material, often supervised by specialists employed within the collections. These practices ensure that the material and the associated information is reliable and, in the event of

any query, is available for re-examination and further research. It would be helpful if public-access databases, who rarely employ specialist systematists, require depositors to follow similar protocols to ensure that the information they release is as accurate as possible, properly annotated and (wherever possible) based on material available for future examination and verification.

References

- Agerer R, Ammirati J, Blanz P, Courtecuisse R, Desjardin DE, Gams W, Hallenberg N, Halling R, Hawksworth DL, Horak E, Korf RP, Mueller GM, Oberwinkler F, Rambold G, Summerbell RC, Triebel D, Watling R. 2000. Always deposit vouchers. *Mycological Research* 104: 642–644.
- Arenal F, Platas G, Monte E, Pelaez F. 2000. ITS sequencing supports *Epicoccum nigrum* and *Phoma epicoccina* being the same biological species. *Mycological Research* 104: 301–303.
- Bainbridge BW. 1994. Modern approaches to the taxonomy of *Aspergillus*. In: Powell KA, Renwick A, Peberdy JF, eds. *The genus Aspergillus*, New York, NY, USA: Plenum Press, 291–301.
- Blaxter M. 2003. Counting angels with DNA. *Nature* 421: 122–124.
- Bridges PD. 2002. The history and application of molecular mycology. *Mycologist* 16: 90–99.
- Bruns TD, White TJ, Taylor JW. 1991. Fungal molecular systematics. *Annual Reviews in Ecology and Systematics* 22: 525–564.
- Crous P. 2002. Adhering to good cultural practice. *Mycological Research* 106: 1378–1379.
- Deckert RJ, Hsiang T, Peterson RL. 2002. Genetic relationships of endophytic *Lophodermium nitens* isolates from needles of *Pinus strobus*. *Mycological Research* 106: 305–313.
- Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673–4680.
- Hugenholtz P, Huber T. 2003. Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. *International Journal of Systematic and Evolutionary Microbiology* 53: 289–293.
- Kirk PM, Cannon PF, David JC, Stalpers JA. 2001. *Dictionary of the fungi*, 9th edn. Wallingford, UK: CABI Publishing.
- Moncalvo JM, Drehermel D, Vilgalys R. 2000. Variation in modes and rates of evolution in nuclear and mitochondrial ribosomal DNA in the mushroom genus *Amanita* (Agaricales, Basidiomycota): phylogenetic implications. *Molecular Phylogenetics and Evolution* 16: 48–63.
- Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences, USA* 85: 2444–2448.
- Reddy PV, Patel R, White JF. 1998. Phylogenetic and developmental evidence supporting reclassification of cruciferous pathogens *Phoma lingam* and *P. wasabiae* in genus *Plenodomus*. *Canadian Journal of Botany* 76: 1916–1922.
- Sanchez-Ballesteros J, Gonzalez V, Salazar O, Acero J, Portal MA, Julian M, Rubio V, Bills GF, Polishook JD, Platas G, Mochales S, Pelaez F. 2000. Phylogenetic study of *Hypoxylon* and related genera based on ribosomal ITS sequences. *Mycologia* 92: 964–977.
- Suh S-O, Kurtzman CP, Blackwell M. 2001. The status of *Endomyces scopularum* – a filamentous fungus and two yeasts. *Mycologia* 93: 317–322.
- Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP. 2002. DNA points the way ahead in taxonomy. *Nature* 418: 479.
- van der Aa HA, Noordeloos ME, de Gruyter J. 1990. Species concepts in selected genera of Coelomycetes. *Studies in Mycology* 32: 3–19.