## COMMENTARY

# Urgent need for authentic (derived from type or typified material) ITS sequence database for all fungi

*Rahul Sharma*

The identification of fungal species has come a long way from the time when PCR was not in existence in mycological laboratories. Today it is difficult to find any mycological laboratory without a PCR machine. There are several reasons for the trend in using molecular methods, especially for identification of fungal species. Among them, the foremost is the accuracy and authenticity of using the DNA sequence, because it seldom changes (due to mutations) in contrast to morphology which does not remain constant and changes with environment. There are several examples where one is able to identify a non-sporulating strain to species level using the DNA sequence of rDNA, which otherwise would have been impossible due to lack of spores or fruiting structures useful in species identification. Due to the added advantage that we are able to deduce the phylogenetic relationships of fungi using DNA sequence data, molecular identification has not only gained support but has become an essential component in description of novel species.

Identification of fungi based on the sequence of the ITS region of nrDNA depends on sequence comparison with orthologous sequences from other species. The National Center of Biotechnology Information (NCBI) along with European Molecular Biology Laboratory (EMBL) and DNA Database Japan (DDBJ) have a huge database which has ITS sequences from around 20,000 species. However, there are a large number of sequences that are redundant apart from those that are incorrect (wrongly identified) in the database. The lack of sequences from a considerable number of fungi makes the database-based identification incomplete and on several occasions misleading. This is because if the BLASTn search shows less than 95% similarity, one is prompted to assume that the test fungus is definitely a new species. However, actually the test fungus might belong to a known fungal species whose sequence is not deposited in GenBank or any other database; hence no high similarity match. When one goes through the recent papers published in mycological journals, one finds several papers on phylogenetic studies, many of which comprise newly derived sequences of those fungi whose sequences are already known albeit from other strains. These publications no doubt enhance the understanding of the phylogenetic relationships of members of families and orders or even higher taxa along with revealing intra-specific variability. But we feel it is at the cost of several fungi (> 60% of all known fungi) whose rDNA has never been sequenced even once. Sequencing efforts worldwide need to be concentrated on those fungi that lack even a single ITS sequence in GenBank. Once all the known fungi have at least one ITS rDNA locus sequenced, detection of novel forms will be rapid and authentic. This will greatly enhance the utility of the public databases as the reference centre for identification of fungi using rDNA. Although in recent years, the sequencing costs have come down considerably which has led to increase in the number of newly sequenced fungi, this has not helped increase the sequencing of the remaining fungi. It has led to increasing number of projects that are involved in higher-level phylogenetic study or whole genome sequencing of fungi. As pointed out by Herendeen for plants, most of the biodiversity to be discovered is in the tropics, but most of the standing expertise is in North America and Europe. He adds 'If we're really worried about documenting and understanding diversity, we need to have more people trained in these various countries where a lot of this biodiversity is endemic.' Working on development of a complete ITS database by enriching GenBank and other databases is the best strategy to uncover the global fungal diversity. Type material of all described fungi is held in two forms: either as living cultures (in culture collections) or herbarium specimens (in herbaria). Of all the known fungi that have been cultured so far or for which living culture exists, more than 80% is held in the culture collections of Europe or North America. Several culture collections have over several decades (or a century) acquired fungal cultures from across the globe, which includes developing countries like India. Procurement of cultures for taxonomic study (sequencing the ITS region) by the developing world from the collections in Europe or North America is a costly affair. An example is cited later for one of the fungus studied by us, which clearly reveals a potential problem faced by mycologist in the developing world.

Most of the novel taxa of fungi that are being described require ITS sequence information along with morphological data; however, a large number of them still lack sequence data. Also, to fill the gap of the missing ITS sequences of fungi in GenBank, mycologists around the world should sequence the ITS region from type or typified material that is present in their respective culture collections or herbaria. And when the only type material of a fungus is lost, mycologists of the respective countries should make new collections, identify them, typify them and also generate the ITS sequence. Although sequence of the ITS region does not differentiate all fungal species, it certainly gives a head start to the identification and indicates its possible phylogenetic relatives. For example, species of two very common genera, viz. *Aspergillus* or *Penicillium* have very high level of sequence similarity in the ITS region, but the sequence of this region at least gives an indication of its close relatives which can be resolved by sequencing a second or third genetic locus for species resolution. Likewise species of *Sclerotium* are highly divergent in their ITS region within species unlike the former two genera. The probable explanation for this difference in the sequence conservation in the ITS region is because speciation in the former two genera is a recent event, while *Sclerotium* seems to be an evolutionarily older genus and therefore the ITS region has accumulated a large number of changes (mutation) during the course of time. A similar story is seen in anthropophilic dermatophytes that have recently evolved and are believed to have diverged from zoophiles[1–3].

Therefore, the ITS sequence may not be applied to identify all fungi but its utility in species identification cannot be undermined because this is the single locus which is most represented in GenBank compared to other regions like beta-tubulin, chitin synthase, RBP1, etc. and hence it has been selected as the locus for the barcoding of fungi (CBOL). Several articles appear that question the use of one gene to deduce the phylogeny of the whole organism. Many advocate that more than one or a combination of genes is needed (multigene) to correctly represent the true phylogeny of fungi. In our view the ideal approach will be to identify genes that are responsible for morphology of spores or fruiting structures. For example, morphology-based identification relies mainly on differences in the shape and size of spores (conidia, ascospores, basidiospores, etc.). There is always a conflict between mycologist who study only morphology (morpho-taxonomists) and those who only study DNA (phylo-taxonomists). However, there is increasing evidence (from the published literature) that both groups are learning the intricacies of the other approach, i.e. the morphologists are incorporating molecular methods and phylogeneticists are incorporating morphology in their study.

It would be interesting if genes are known that are responsible for septation in fungi; then species distinction on the basis of the number of septations can be correlated with genetically. Similar case with genes that are responsible for colour or shape of spores or fruit bodies in ascomycetes and basidiomycetes since species delimitation in several genera is based on differences in such features. Although as in other organisms like humans and animals, for one trait there are generally more than one gene involved, there might be a combination of genes that results in a particular phenotypic trait in fungi as well. Therefore, sequencing all of those genes would be a difficult task for species identification. Hence, the ITS region of nrDNA happens to be the appropriate candidate locus for sequencing all fungi. In fact, the phylogenetic analysis based on sequence comparison of rDNA has been questioned by many as it represents gene phylogeny and not organismal phylogeny. Here we would like to point out that using the ITS sequence for taxonomic purpose is a different issue and utilizing it for phylogeny derivation is different. ITS or other regions should be used for an indication of the supposed phylogeny of a taxa and not to represent true phylogeny. Hence for classification purpose or identification it is quite appropriate to use the ITS sequence as it allows us to compare practically all taxa based on one character, since there is no fungus which lacks the ITS region altogether. Although one may not find sufficient differences to distinguish all fungal species, comparison of the ITS sequence certainly brings us close to a selected few genera/species from the 100,000 odd ones for comparison. And, because rDNA evolution in most cases corresponds to the differences in the species, it is taken as a good parameter albeit an indirect one for species distinction. As ribosome is universal and is present in all organisms, DNA sequences from its different regions (that are greatly conserved) are compared across a wide range of fungi for deducing phylogenetic relationships of higher taxa. The ITS regions that span the 5.8S gene on both sides are the lesser conserved regions of the rDNA. The sequence conservancy in the SSU, LSU and 5.8S regions is helpful for proper alignment of the ITS sequences among genera or species, as these conserved regions are used as standard reference points thereby revealing the insertion and deletion events within the spacers.

Species concept in fungi is not uniform as in plants and animals where mostly biological species exist, i.e. members of a species which are sexually interbreeding belong to one species. Because many fungi reproduce asexually, it is not uncommon to designate species purely on morphological grounds. More recently, the situation is getting further complicated when we distinguish species based on phylogenetic grounds. The so-called cryptic species that are recognized using molecular methods are actually those forms that are in the process of speciation. This means that morphologically they are similar and belong to one species but genetically they have differences, which makes us believe that they are different species. As they are in the process of speciation, they have acquired differences only in certain genes which we have analysed but still have to diverge in those genes which are responsible for morphology; hence there are no differences in morphology. The question now arises – whether it is appropriate to name those genetic entities as species which have still not fully diverged but are in the process of divergence? Morphologists have long considered such forms at subspecies rank (including variety). What is the need of the hour is to put more efforts to uncover the remaining fungal diversity than to study the already known diversity by hair splitting with multiple genes. If one considers the number of taxa sequenced (rDNA) for the first time (whose sequence is not in GenBank) in comparison to the number of total taxa sequenced, the ratio is very small. This point was also raised by Korf in his excellent analysis of Wheelar's paper, where he advocated that more efforts should be focused on unravelling the unknown diversity rather than repeated phylogenetic studies[4].

After the Biodiversity Convention several countries have restricted the transfer of biological material to other countries, including India, due to possible commercial exploitation of bioresources. The worst hit by this scenario are the taxonomists who need to examine materials available in other culture collections or herbaria the world over when they encounter a possibly new form. It is difficult to explain the situation to the administrators who are in charge of the Biodiversity Authority in different countries even after several requests from taxonomic groups. The better alternative is to have rDNA (preferably ITS or LSU-D1, D2 regions) of all the valid fungal species sequenced and made available in GenBank. The usefulness of GenBank/EMBL/DDBJ databases will then be considerably enhanced and the science of taxonomy will be more globalized. This will allow decentralization of the science of taxonomy which currently seems to be slowly getting restricted to big culture collections and those who have large repositories and have access to almost all recognized taxa of a group. Culture collections or herbaria around the world which hold type or typified fungal materials should join hands in this effort to form a consortium to generate rDNA sequences of all valid taxa and making it available to the scientific community across the globe. This will greatly speed up the inventorying of unknown diversity of fungi across the globe. The world's major culture collections or herbaria are situated in countries which have very little diversity to explore, whereas countries with immense diversity have less

resources or even access to large collections, for making taxonomic comparisons. We were in dilemma when we wanted to compare a newly recovered isolate of fungus *Gliocephalotrichum*. The fungus was known from India before, but the culture of the previous isolation was deposited by the original authors at IMI (now CABI), and none of the Indian culture collections has the culture. The story was similar to what Abdul Kalam quoted in his book *Ignited Minds* about gyros whose raw material was originally from India and the finished product was refused by another country due to some reasons. When we wanted to obtain the culture of *Gliocephalotrichum* from CABI for comparison, it was costing us Rs 12,000. The irony was that the original source of the culture was India and when we needed it for comparison we had to pay a huge amount to obtain the same. From an Indian culture collection, we can obtain a fungal culture for anything between Rs 200 and 500. Therefore obtaining at least an indigenous fungal strain from an Indian culture collection seems economically feasible. Several authors have raised concern over the dwindling status of the science of taxonomy and the meagre staff positions and funding it is getting in recent years, but very little is being done to prevent the situation from getting worse. The foremost reason for this is that we are not creating a second line of fungal taxonomists in laboratories that were known to have expertise on a particular group of fungi. The situation in India is at least evident and we do feel that it is a global phenomenon. The change from morphology-based taxonomy to purely molecular-based taxonomy within a few years is one of the main reasons for this situation. Several schools of taxonomy that were involved in fungal diversity exploration and which regularly described new species have stopped because mycology journals now no longer accept papers that do not have molecular data. Unlike many branches it is difficult to train taxonomists (morphologist) overnight; it often requires more than a decade for a person to acquire sufficient expertise on a group. However, it is equally difficult for a morphotaxonomist to get insight into the theoretical and practical aspects of molecular techniques. The only plausible solution to this difficult situation is the collaborative efforts needed to uncover the fungal diversity. However, it is equally important that we do not stop our efforts in encouraging students to study morphology and training them to observe morphological features. In India, several schools of taxonomy have died because of lack of a second generation of trained taxonomists, e.g. Osmania University – hyphomycetes and ascomycetes; Gorakhpur University – Cercosporiod fungi; University of Jabalpur – hypho and ascomycetes; University of Delhi – ascomycetes; CAS Botany, Chennai – Hyphomycetes and macromycetes, etc. Nonetheless, many institutions have made efforts to keep taxonomic science alive by recruiting and training a new generation, e.g. Agharkar Research Institute – lichen, mushroom and ascomycetes; Panjabi University Patiala – resupinate basidiomycetes; BSI – mushrooms and hyphomycetes, etc.

1. Ohst, T. *et al*., *J. Clin. Microbiol*., 2004, **42**, 4444–4448.
2. Sharma, R. *et al*., *J. Med. Microbiol*., 2007, **56**, 1377–1385.
3. Sharma, R. and Gräser, V., In *Molecular Detection of Human Fungal Pathogens* (ed. Dongau Liu), CRC Press, Taylor & Francis Group, Boca Raton, USA, 2011, pp. 285–298.
4. Konf, R. P., *Mycotaxon*, 2005, **93**, 407–415.

ACKNOWLEDGEMENT. The Department of Science and Technology, New Delhi, is greatly acknowledged for funding National Facility at ARI.

*Rahul Sharma is in the National Fungal Culture Collection of India, Agharkar Research Institute, G.G. Agarkar Road, Pune 411 004, India.*
*e-mail: rsharma@aripune.org*