

Species level identification of moulds

and yeasts has long been considered a challenging area of microbiology and consequently, in contrast to bacteria, fungal isolates obtained from Environmental Monitoring (EM) programmes are often identified to the genus rather than the species level.

As yeasts and moulds can cause serious problems if they find their way into pharmaceutical products, more detailed identification can provide important information that helps manufacturers track the origin of isolates and reduce the risk of contamination. So what are the difficulties when it comes to species level identification of yeasts and moulds?

Bacteria are commonly identified by sequencing of the 16S ribosomal DNA, and the approaches that have been developed for genotypic identification of yeasts and moulds also involve sequencing sections of ribosomal DNA. Fungal ribosomes have a large and small subunit. The ribosomal RNA operon, that is the DNA that codes for ribosomal RNA, has three rRNA sequences and two internal transcribed spacer regions: ITS1 and ITS2.

Two distinct approaches have emerged for genetic identification of fungi: sequencing of the D2 region of the large subunit ribosomal DNA (D2 LSU) and sequencing of either one or both of the internal transcribed spacer regions (ITS).

D2 LSU sequencing is probably the most widely used approach for mould and yeast identification at present.

Similarly to 16S rDNA sequencing for bacteria, sequence data obtained can be analysed using a validated commercial database that has been built using reference strains, as well as the publically available but non validated sources, such as the EMBL-EBI (European Molecular Biology Laboratory – European Bioinformatics Institute) database. This is a very active area of research and the amount of data available is continually expanding as researchers upload new results.

ITS sequencing is also used for the identification of moulds and yeasts, but in contrast to D2 LSU sequencing, at present it generally relies more heavily on the use of data obtained from unvalidated public databases.

Identification strategy

When sequencing fungal isolates, NCIMB always uses a validated D2 LSU database in the first instance, as the use of a validated database gives the most reliable result. However, some families and genera are known to be difficult to identify to

HOW TO APPROACH THE IDENTIFICATION OF YEASTS AND MOULDS

Moulds and yeasts can cause serious problems for pharmaceutical, personal care and food product manufacturers. Accurate fungal identification is part of any Environmental Monitoring programme, but species level ID is not always straightforward. NCIMB's Identification Service Manager, Vikki Mitchell, looks at where the difficulties are often encountered



species level using D2 sequencing and in my own experience, while it is usually possible to obtain a species level identification of yeasts using D2 LSU sequencing, we sometimes cannot get that level of identification for moulds.

In cases where species level identification cannot be obtained using D2 LSU sequences, it is often possible to obtain species level identification using ITS sequencing. Generally, there is a higher level of differentiation between ITS

Detailed identification of yeasts and moulds can help manufacturers track the origin of isolates and reduce the risks of contamination

sequences. Unlike ribosomal DNA, ITS sequences have no functional role, and consequently have accumulated a greater level of mutation, which aids identification.

A good example of moulds that are difficult to identify to species level using D2LSU sequencing comes from the genus *Penicillium*. I have found *Penicillium camemberti*, *Penicillium clavigerum*,

Penicillium commune, *Penicillium corylophilum* and *Penicillium crustosum* have all matched a single isolate sequence at 100% similarity.

However, when ITS sequencing has subsequently been undertaken, a species level result has been achieved – in one specific example the isolate was identified as *Penicillium crustosum* with a 100% match.

The above example is an illustration of where D2 LSU sequencing cannot differentiate between different species of the same genera, but in some cases D2 LSU sequencing alone cannot distinguish between different, but closely related, genera.

For example, I have found several species of *Cladosporium* and *Mycosphaerella* have all matched to the same isolate sequence. *Cladosporium* is a large genus that has been reported to be the most common fungal component isolated from air, and is therefore quite commonly found in the course of EM programmes.

Again, we have found ITS sequencing to be successful in providing a species level identification – one isolate which matched to *Cladosporium cladosporioides*, *Cladosporium herbarum*, *Cladosporium oxysporum*, *Mycosphaerella aronici* and *Mycosphaerella tassiana*, was identified as *Cladosporium cladosporioides* when ITS was used.

In another example, we found that it was not possible to differentiate between an even larger group of closely related genera. Searching the non-validated EMBL database for D2 LSU sequences failed to distinguish between *Saccharothecium sepincola*, *Mycosphaerella sojae*, *Pithomyces chartarum*, *Pleospora gaeumannii*, *Leptosphaerulina saccharicola*, *Heterophoma adonidis*, *Nothophoma quercina* and *Leptosphaerulina australis*.

These eight species, from seven different genera, all matched a single D2 LSU isolate sequence at 100% similarity. In this case however, ITS sequencing did not result in a species level match either – the isolate matched to *Leptosphaerulina saccharicola*, *Leptosphaerulina australis*, *Leptosphaerulina chartarum* and *Leptosphaerulina trifolii*.

It was, however, successful in narrowing the identification down to *Leptosphaerulina* species rather than several different genera – a substantial improvement on the initial result obtained.

As many fungi do match very well to the validated D2 LSU database, at



present, we recommend using D2 LSU sequencing in the first instance. If a match is not found using the validated database, we would analyse the results against the non-validated EMBL database, before considering whether to follow up with ITS sequencing.

It is recommended that any relevant published papers should be referred to for additional supporting information when using unvalidated databases for identification purposes.

How far to go

The decision on whether to progress to ITS sequencing if a species or genus level identification cannot be obtained using D2 LSU sequencing is really dependent on the individual circumstances and whether family, genus or species level identification is required.

For example, it may be requested by customers investigating excursions from normal populations or contamination issues. While we cannot guarantee that ITS sequencing will always provide a sequence level match where it has not been obtained using D2 LSU, the examples above illustrate that it has been successful in doing so with some quite commonly found isolates, and where a species level match has not been found it

Development of genotypic techniques has greatly improved the accuracy, speed and reliability of yeast and mould identification

has still given an improved result.

In conclusion, the development of genotypic techniques has greatly improved the accuracy, speed and reliability of yeast and mould identification but this is an area which is still developing rapidly. Validated databases of ITS sequences are being developed and in future, as the availability of validated sequences improves, ITS may become more commonly used method for identification of yeasts and moulds. At the present time, however combining both the D2 LSU and ITS techniques has been found to offer a good approach to the identification of yeasts and moulds.

CONTACT

**Vikki Mitchell, Identification Services
Manager
NCIMB Ltd
Ferguson Building
Craibstone Estate Bucksburn
Aberdeen AB21 9YA UK
T +44 1224 711100
enquiries@ncimb.com**