

Introduction

The identification of fungi, especially filamentous fungi, has historically been a very difficult task. Due to the amount of experience required to accurately identify filamentous fungi to the species level, it has become acceptable to either identify these organisms to the genus level, or in some cases, simply identify them as “molds.” Over the years, there have been numerous attempts to automate biochemical tests. Since eukaryotic organisms exhibit far less metabolic diversity than prokaryotic organisms, these systems tended not to have the resolution required to differentiate fungi at the species level.

During the manufacture of sterile and non-sterile pharmaceuticals, the presence of molds in the manufacturing area is usually a cause for concern. A well-designed Environmental Monitoring program should detect the presence of molds before they have an opportunity to contaminate the product. However, occasionally, fungal isolates will be recovered from sterility or media fill failures. In these cases, it is extremely important to be able to identify the contaminating organism at the species and possibly the strain level in order to track the origin of the contamination.

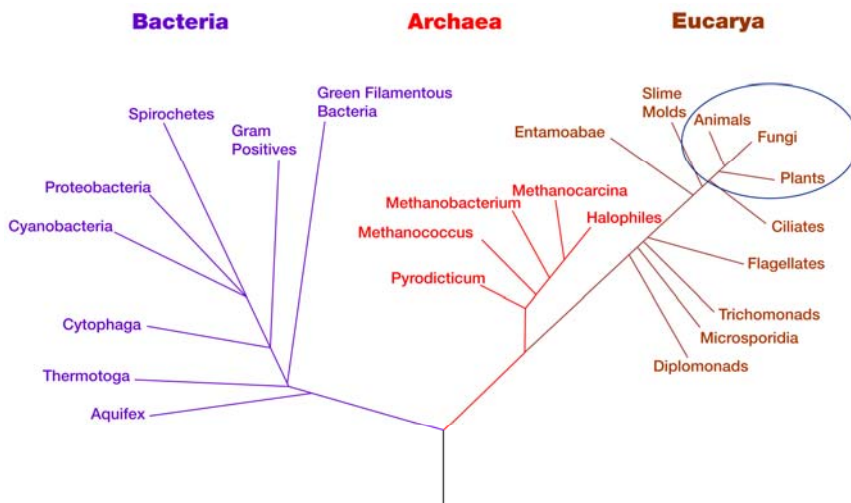
Background and early studies

A recent review of FDA product recalls (Jimenez 2007) illustrates the lack of ability to identify fungal samples to the species, or in many cases, even to the genus level. At the conclusion of the investigations, of the 12 fungal isolates responsible for sterile pharmaceutical product recalls, zero were identified to the species level, 4 to the genus level and 8 were simply designated as “mold” or “yeast”. Similarly, in non-sterile product recalls, 32 products were contaminated with fungi. Of these, 3 were identified to the species level, zero to the genus level and 29 were identified as either “mold” or “yeast”.

While any type of fungal contamination is cause for concern, a species level identification is needed to provide a definitive root cause as part of an investigation. Historically, fungal identification, especially mold identification, has not been held to the same standards as bacterial identification with respect to the investigation of contaminated pharmaceutical products. This has been done primarily because it is recognized that it is difficult to identify fungi to the species level, not because the identity is unimportant. The availability to all pharmaceutical microbiology laboratories of molecular based identification for fungi, both yeasts and molds, have made these limitations virtually disappear.

The use of ribosomal DNA sequences for the purposes of organismal taxonomic classification has been in use for many decades. More recently, ribosomal genes have been used to study the phylogenetic relationships of fungi, with some surprising results.

Phylogenetic Tree of Life



The first came in 1993, when Baldauf and Palmer discovered that fungi are more closely related to animals than they are to plants (Figure 1), as had been assumed to be the case based on morphology and other phenotypic characteristics. Another has been the realization that many species of fungi are synonymous with one another, with many names being used to describe the same organism. For example, *Candida albicans* has 173 synonyms (Kurtzman and Fell), many of which have also been in common usage during different periods of time.

Figure 1: Classification of Fungi

- L. Jimenez. Microbial diversity in Pharmaceutical Product Recalls and Environments. *PDA J Pharm Sci Tech.* 61(5):383-399 (2007).
- S. L. Baldauf and J. D. Palmer. Animals and Fungi Are Each Other's Closest Relatives: Congruent Evidence From Multiple Proteins. *PNAS* 90:11558-11562 (1993).
- C. P. Kurtzman and J. W. Fell (Eds.). *The Yeasts: A Taxonomic Study.* Elsevier, 1988. pp. 476-477.

Recent Advances

Although fungal taxonomists have been using phylogenetic analysis to characterize, classify and re-classify fungi for many years, it is only recently that this approach has gained popularity for the routine identification of fungi in the pharmaceutical manufacturing environment. This has primarily been due to the fact that while the technology existed for performing these analyses, applications which were developed in a compliant manner and are able to be validated in a cGMP laboratory have only recently been made available through products and contract service laboratories. However, now that the technology is available, the expectation is that it will become more widely adopted as an accepted approach over the next several years.

This prediction has been proven true; as the past several years have seen a rapid development of DNA sequencing based fungal identification systems. The introduction of MicroSeq®, a commercial product available from Applied Biosystems, as well as a number of laboratories world-wide using DNA sequencing for fungal identification, has ensured pharmaceutical microbiologists that the technology has the level of support needed to provide high quality and confident and compliant answers.

DNA sequencing provides data for identification that is substantially more accurate and reproducible than relying solely on visual phenotypic characteristics (Figure 2). This is so well understood and accepted that the FDA recommended the use of genetic methods in their 2004 update to the guidance document, "FDA Guidance for Industry. Sterile Drug Products Produced by Aseptic Processing – Current Good Manufacturing Practice." In this document, the FDA states in the section on environmental monitoring, "Genotypic methods have been shown to be more accurate and precise than traditional biochemical and phenotypic techniques. These methods are especially valuable for investigations into failures (e.g., sterility test; media fill contamination)."

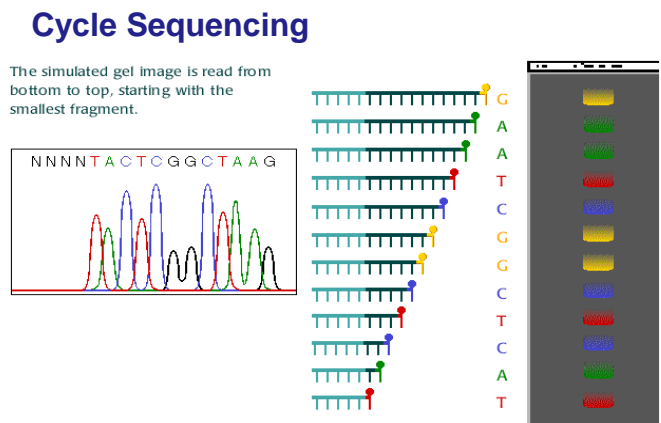


Figure 2: Data resulting from DNA sequencing is not subjective

Also, in the section on sterility testing, the FDA states, "Sterility test isolates should be identified to the species level. Microbiological monitoring data should be reviewed to determine if the organism is also found in the laboratory and production environments, personnel, or product bio-burden. Advanced identification methods (e.g., nucleic-acid based) are valuable for investigational purposes. When comparing results from environmental monitoring and sterility positives, both identifications should be performed using the same methodology."

Limitations with Current Identification Systems

Despite the overwhelming acceptance and support from the scientific community and regulatory agencies, there are still opportunities for improvement with the current commercially available sequence-based identification systems. The first is the coverage of the database of known fungal species. This should not be confused with the size of the database, as a database filled with hundreds, or even thousands, of species not encountered in the pharmaceutical manufacturing environment adds no value to the identification system. This has been pointed out in recent literature by Rozynek, et al. 2004 and Hall, et al. 2003. In both of these publications, the technology is generally well accepted, but the coverage of the database reduces the effectiveness of the technology to provide meaningful identifications.

A second limitation of the commercially available system is the gene target chosen to build the DNA sequence library. This is not a limitation of the technology, but rather a limitation of the application. When choosing an appropriate gene target for phylogenetic analysis, one needs to find a target which undergoes enough genetic mutation for there to be observable differences in the DNA sequences of similar, but different species.

However, the rate of accumulated nucleotide differences should not be so great that truly related species appear to be more dissimilar than they really are. This is the great challenge in choosing the appropriate target, and in many cases it is simply a case of trial and error.

- P. Rozynek, S. Gilges, T. Brüning, and M. Wilhelm. Quality Test of the MicroSeq D2 LSU Fungal Sequencing Kit for the Identification of Fungi. *Int. J Hyg. Environ Health.* 207(3):297-299 (2004).
- L. Hall, S. Wohlfiel, and G. D. Roberts. Experience with the MicroSeq D2 Large-Subunit Ribosomal DNA Sequencing Kit for Identification of Filamentous Fungi Encountered in the Clinical Laboratory. *J Clinical Microbiology* 42(2):622-626 (2003).

Limits of D2 gene resolution

The D2 expansion segment of the Large Subunit (LSU) of the ribosomal gene, as utilized in the MicroSeq® system, typically does a very good job of placing an unknown fungal isolate into the appropriate taxa. However, due to limitations in the resolution of the D2 segment, closely related organisms may have identical, or very similar, DNA sequences.

The D2 expansion segment of the LSU is effective at linking together higher level taxa (Genus, Family, Order, etc.) and can differentiate many species acceptably, but not in all cases. One common example of the inability of D2 to differentiate close but distinct species is *Komagataella pastoris*, *K. phaffii* and *K. pseudopastoris*. *K. pastoris*, formerly known as *Pichia pastoris*, is an organism which is frequently used in the production of biologics. Because of the amount of genetic manipulation required to introduce genes of interest into this species, as well as all of the eventual monitoring with scaling the fermentation up from a starter culture to full production, it is extremely important to ensure the same organism is being used in each step of the process.

Unfortunately, the D2 DNA sequence of *K. pastoris* is very similar to the D2 DNA sequences of 2 other *Komagataella* species, *K. phaffii* and *K. pseudopastoris*. There is only 1 nucleotide difference which separates *K. pastoris* from the other 2 species. This is just one of many examples where the D2 DNA sequence is not able to fully identify an isolate to the species level.

Fungal Identification using Ribosomal Internal Transcribed Spacer (ITS)

More recently, the Internal Transcribed Spacer (ITS) regions of the ribosomal operon have been used for fungal systematics and classification. There are 2 ITS regions in the fungal rRNA operon (Figure 3). The first, ITS1, is found between the 18S and 5.8S rRNA genes. The second, ITS2, is located between the 5.8S and the 28S rRNA genes. The entire rRNA operon is transcribed; however, after transcription, the 2 ITS sequences are excised and are therefore not used for any functional purpose.

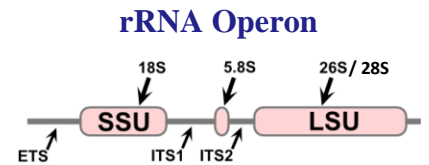


Figure 3: ITS1 and ITS2 genes

Since the ITS sequences are important enough as spacer regions to be maintained by the cell, but not used for any functional purpose, they are allowed to accumulate mutations at a faster rate than the 5.8S, 18S, and 28S rRNA genes. It is this slightly increased rate of accumulated mutations which allows the ITS sequences to provide an improved level of resolution than the D2 sequence. It is generally accepted to sequence the entire stretch of ITS1-5.8S-ITS2 for use in fungal classification. However, for the purposes of routine identification, our laboratory has found that the use of ITS2 alone is usually sufficient for species level identification.

Increased resolution of ITS2 sequence

When we compare the ITS2 DNA sequences of the three *Komagataella* species, we see many more nucleotide differences. There are 16 differences between *K. pastoris* and *K. phaffii* and 18 differences between *K. pastoris* and *K. pseudopastoris*. This increased resolution offers a much more confident identification and allows for small strain to strain variability to still ensure the correct species level identification (Figure 4).

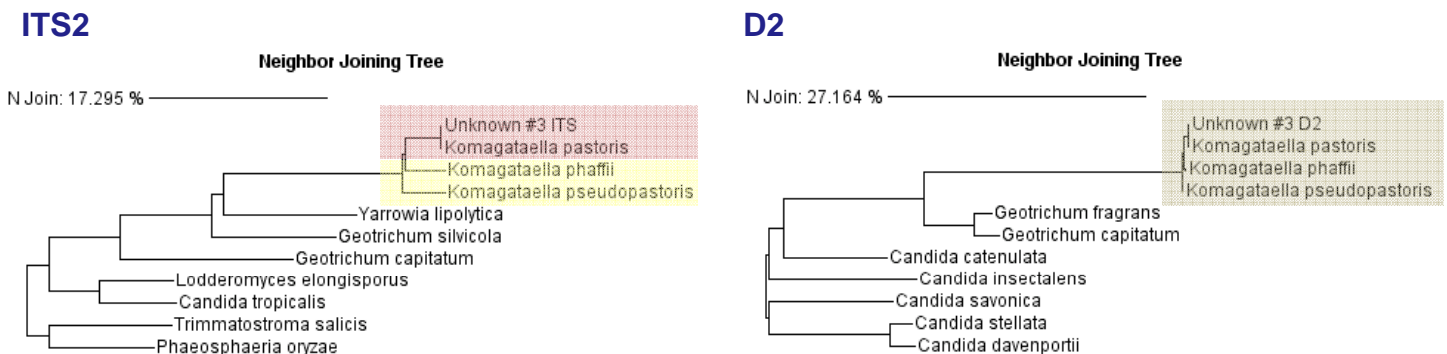


Figure 4: ITS2 and D2 comparison of *Komagataella* spp.

Species level identification of *A. niger* and *A. brasiliensis*

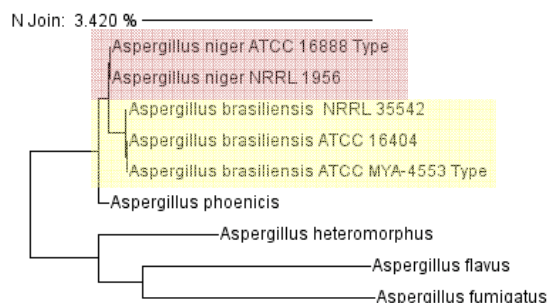


Copyright Dennis Kunkel Microscopy, Inc.

Figure 5: *Aspergillus* sp.

Aspergillus brasiliensis is a newly described species (Varga, et. al, 2007), which was in part created by the transfer of several existing *Aspergillus niger* strains to *A. brasiliensis*. Most significant in the creation of this new species to pharmaceutical microbiologists was the inclusion of *Aspergillus niger* ATCC 16404 (Figure 6). This organism is cited in several USP chapters as a QC organism, including USP <61> “Microbial Limits Test – Enumeration” and USP <71> “Sterility Test”. Because of the number of pharmaceutical companies performing these USP tests, it is very important to be able to correctly identify the QC organism. Unfortunately, this is another example of where virtually all phenotypic tests, as well as D2 DNA sequencing, is unable to differentiate these 2 species from one another. Phenotypically, it has always been difficult to differentiate between *A. niger* strains due to a lack of diversity in morphological features, unstable phenotypic characters and the significant influence of culture conditions on the phenotype (Runyu, et. al. 1995). Furthermore, D2 sequences provide no additional information, as the DNA sequences for all observed *A. niger* and *A. brasiliensis* strains are identical.

ITS2



D2

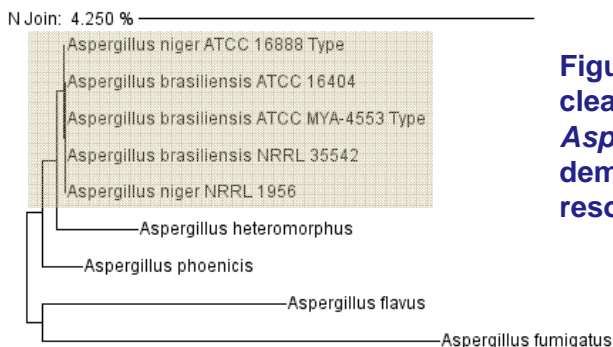


Figure 6: The ITS2 region clearly distinguishes two *Aspergillus* spp., demonstrating greater resolution than D2.

A recent study performed by Houseknecht, et al. from the ATCC looked at the justification for this recent classification. The study concluded that although *A. niger* and *A. brasiliensis* are very similar, there are ways to differentiate isolates of these species. One way is through evaluation of conidia morphology under high magnification with an electron microscope (the differences are not observable using a light microscope). Another way is through ITS DNA sequencing. The entire ITS1-5.8S-ITS2 DNA sequence shows there are 5 differences between the strains of *A. niger* and *A. brasiliensis*. When comparing the ITS2 sequence alone, there is only 1 nucleotide difference between the species (Figure 7); however, this difference has been shown to be extremely reproducible and is therefore considered to be a diagnostic indicator of the species.

ITS2 Sequence Data for Two *Aspergillus* spp.

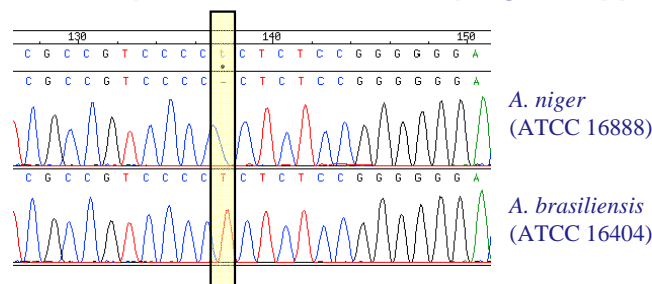


Figure 7: One nucleotide differentiates *A. brasiliensis* from *A. niger*.

Conclusion

Comparative DNA sequencing has greatly enhanced the ability to accurately and reproducibly identify fungi. Fungal taxonomists have been using DNA sequences for many years as a basis for the re-classification of all fungal taxa and have more recently moved to ITS sequencing as the “Gold Standard.” With the ability to use these powerful technologies as part of a comprehensive Environmental Monitoring program, pharmaceutical microbiologists have one more tool with which to ensure product safety. As made clear by the FDA as part of their “Pharmaceutical cGMPs for the 21st Century” and Process Analytical Technology initiative, new rapid and accurate technologies are encouraged to be used by the industry.

- J. Varga, S. Kocsube, B. Toth, J. C. Frisvad, G. Perrone, A. Susca, M. Meijijer, and R. Samson. *Aspergillus brasiliensis* sp. Nov., a Biseriate Black Aspergillus species with World-wide Distribution. *Int. J Sys Evol Microbiology*. 57:1925-1932 (2007).
- E. Rinyu, J. Varga, and L. Ferenczy. Phenotypic and Genotypic Analysis of Variability in *Aspergillus fumigatus*. *J Clinical Microbiology*. 33(10):2567-2575 (1995).
- Houseknecht, J., E. Stamenova, S-O. Suh, B. Beck, M. McKee, and J. Zhou. Reclassification of ATCC® 16404™ and ATCC® 9642™ as *Aspergillus brasiliensis*. *Pharmaceutical Microbiology Forum Newsletter*. 14(10):2-8.