Database establishment for the secondary fungal DNA barcode translational elongation factor 1α (TEF1α)

Wieland Meyer, Laszlo Irinyi, Minh Thuy Vi Hoang, Vincent Robert, Dea Garcia-Hermoso, Marie Desnos-Ollivier, Chompoonek Yurayart, Chi-Ching Tsang, Chun-Yi Lee, Patrick Woo, et al.

NRC Research Press

ARTICLE

Database establishment for the secondary fungal DNA barcode *translational elongation factor 1α* (*TEF1α*)[1]

Wieland Meyer, Laszlo Irinyi, Minh Thuy Vi Hoang, Vincent Robert, Dea Garcia-Hermoso, Marie Desnos-Ollivier, Chompoonek Yurayart, Chi-Ching Tsang, Chun-Yi Lee, Patrick C.Y. Woo, Ivan Mikhailovich Pchelin, Silke Uhrlaß, Pietro Nenoff, Ariya Chindamporn, Sharon Chen, Paul D.N. Hebert, Tania C. Sorrell, and the ISHAM barcoding of pathogenic fungi working group[2]

Abstract: With new or emerging fungal infections, human and animal fungal pathogens are a growing threat worldwide. Current diagnostic tools are slow, non-specific at the species and subspecies levels, and require specific morphological expertise to accurately identify pathogens from pure cultures. DNA barcodes are easily amplified, universal, short species-specific DNA sequences, which enable rapid identification by comparison with a well-curated reference sequence collection. The primary fungal DNA barcode, ITS region, was introduced in 2012 and is now routinely used in diagnostic laboratories. However, the ITS region only accurately identifies around 75% of all medically relevant fungal species, which has prompted the development of a secondary barcode to increase the resolution power and suitability of DNA barcoding for fungal disease diagnostics. The translational elongation factor 1α (*TEF1α*) was selected in 2015 as a secondary fungal DNA barcode, but it has not been implemented into practice, due to the absence of a reference database. Here, we have established a quality-controlled reference database for the secondary barcode that together with the ISHAM-ITS database, forms the ISHAM barcode database, available online at http://its.mycologylab.org/. We encourage the mycology community for active contributions.

Key words: fungal DNA barcoding, secondary fungal DNA barcode database, translational elongation factor 1α.

Résumé : En raison d'infections fongiques nouvelles ou émergentes, les champignons pathogènes humains et animaux constituent une menace croissante mondialement. Les outils diagnostiques actuels sont lents, non-discriminants au niveau spécifique ou sous-spécifique, et ils nécessitent une expertise morphologique pour identifier avec justesse les agents pathogènes en cultures pures. Les codes à barres de l'ADN sont des séquences

Received 9 May 2018. Accepted 5 November 2018.

Corresponding Editor: Jianping Xu.

W. Meyer, L. Irinyi, M.T.V. Hoang, and T.C. Sorrell. Molecular Mycology Research Laboratory, Centre for Infectious Diseases and Microbiology, Faculty of Medicine and Health, Sydney School of Medicine, Westmead Clinical School, Marie Bashir Institute for Infectious Diseases and Biosecurity, The University of Sydney, Westmead Hospital (Research and Education Network), Westmead Institute for Medical Research, Westmead, NSW, Australia.
V. Robert. Westerdijk Fungal Biodiversity Institute, Utrecht, the Netherlands.
D. Garcia-Hermoso and M. Desnos-Ollivier. Institut Pasteur, National Reference Center for Invasive Mycoses and Antifungals (NRCMA), Molecular Mycology Unit, CNRS UMR2000, Paris, France.
C. Yurayart. Mycology Unit, Department of Microbiology, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand; Department of Veterinary Microbiology and Immunology, Faculty of Veterinary Medicine, Kasetsart University, Bangkok, Thailand.
C.-C. Tsang, C.-Y. Lee, and P.C.Y. Woo. Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong.
I.M. Pchelin. Laboratory of Molecular Genetic Microbiology, Kashkin Research Institute of Medical Mycology, I.I. Mechnikov North-Western State Medical University, St Petersburg, Russia.
S. Uhrlaß and P. Nenoff. Laboratory of Medical Microbiology, Partnership Dr. C. Krueger & Prof. Dr. P. Nenoff, Roetha OT Moelbis, Germany.
A. Chindamporn. Mycology Unit, Department of Microbiology, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand.
S. Chen. Molecular Mycology Research Laboratory, Centre for Infectious Diseases and Microbiology, Faculty of Medicine and Health, Sydney School of Medicine, Westmead Clinical School, Marie Bashir Institute for Infectious Diseases and Biosecurity, The University of Sydney, Westmead Hospital (Research and Education Network), Westmead Institute for Medical Research, Westmead, NSW, Australia; Centre for Infectious Diseases and Microbiology Laboratory Services, ICPMR, Westmead Hospital, Westmead, NSW, Australia.
P.D.N. Hebert. Department of Integrative Biology and Director of the Biodiversity Institute of Ontario at the University of Guelph, Guelph, ON, Canada.

Corresponding author: Wieland Meyer (email: wieland.meyer@sydney.edu.au).

[1]This paper is part of a special issue entitled "Trends in DNA Barcoding and Metabarcoding".

[2]See Acknowledgments for complete list of working group members.

d'ADN courtes, spécifiques, faciles à amplifier et universelles, ce qui rend possible une identification rapide par simple comparaison avec une collection de séquences de référence bien constituée. Le principal code à barres de l'ADN chez les champignons est la région ITS, laquelle a été employée initialement en 2012 et est maintenant employée de manière routinière dans les laboratoires diagnostiques. Cependant, la région ITS ne permet d'identifier avec justesse qu'environ 75 % de tous les spécimens fongiques d'intérêt médical, ce qui a créé la nécessité de développer un code à barres secondaire de manière à augmenter la résolution et la pertinence du codage à barres de l'ADN pour le diagnostic des maladies fongiques. Le facteur d'élongation de la traduction 1α (*TEF1α*) a été choisi en 2015 comme code à barres secondaire chez les champignons, mais n'a pas été tellement employé en raison de l'absence d'une base de données de séquences de référence. Dans ce travail, les auteurs ont établi une telle base de référence dont la qualité a été vérifiée pour le code à barres secondaire qui, conjointement à la base de données ISHAM-ITS, forme la base de données « ISHAM barcode database », disponible en ligne à http://its.mycologylab.org/. Les auteurs encouragent la communauté des mycologues à y contribuer activement. L'utilisation d'un système de codage à barres double facilitera l'identification juste de tous les champignons pathogènes d'intérêt clinique.

Mots-clés : codage à barres de l'ADN fongique, base de données d'un code à barres fongique secondaire, facteur d'élongation de la traduction 1α.

Introduction

The identification and delimitation of fungal species are often challenging using traditional methods, either because of the lack of distinct morphological and biochemical characters or the absence of sexual reproduction. Complex life cycles, such as yeast–mycelial transitions or the existence of synanamorphs of certain species further hamper correct morphological identification (Seifert and Samuels 2000). Conventional fungal identification methods using culture and microscopic analyses are often insensitive, slow (7–14 days), and heavily dependent on the level of mycological expertise (Alexander and Pfaller 2006; Balajee et al. 2007). Further morphological and biochemical traits are subjective, error prone, and frequently result in misidentifications (Sangoi et al. 2009). In addition, the fact that numerous fungal species cannot grow under laboratory conditions precludes their identification (Begerow et al. 2010; Nilsson et al. 2009). Phenotypic traits also fail to differentiate closely related species or species complexes with near-identical morphological characters but distinguishable genetic traits, such as, e.g., the members of the genera *Fusarium*, *Aspergillus*, and *Scedosporium* (Balajee et al. 2005; Gilgado et al. 2005; O'Donnell 2000).

Rapid and accurate identification is essential in clinical diagnostics, outbreak investigation, or other epidemiological studies, as well as in agricultural, quarantine, and industrial settings. Culture and "expert-free" methods capable of identifying fungi directly from biological specimens are needed. The early diagnosis of pathogens is critical in fungal disease management, public health, animal welfare, and plant protection, for quarantine purposes and in other industrial applications. To counter these difficulties in identification, many molecular techniques have been developed over the last 20 years. These include PCR-RFLP analysis (Dendis et al. 2003; Diguta et al. 2011), RAPD (Baires-Varguez et al. 2007; Brandt et al. 1998), hybridisation with genus/species specific DNA or RNA probes (Lindsley et al. 2001; Sandhu et al. 1995), PCR fingerprinting (Hierro et al. 2004; Lieckfeldt et al. 1993), species-specific PCR assays (Kulik et al. 2004; Martin et al. 2000), real time PCR (Bergman et al. 2007; Klingspor and Jalal 2006), and, increasingly, DNA sequencing (Pryce et al. 2003; Romanelli et al. 2010). Sequence-based identification methods have proven to be more accurate than conventional methods in diagnostic clinical mycology (Balajee et al. 2007; Ciardo et al. 2006). Developments in molecular-based technologies are expected to improve further and simplify species or even strain identification, making it faster and more accurate in the future.

Among the applied molecular techniques, DNA barcoding is one of the most promising and efficient methods, as it enables rapid identification of species and recognition of cryptic species across all kingdoms of eukaryotic life. The concept was first proposed by Hebert and colleagues in 2003 (Hebert et al. 2003*a*). Barcodes are standardized, easily amplified, universal short DNA sequences (500–800 bp), which exhibit a high divergence at species level and allow rapid identification by comparison with a reference sequence collection of accurately identified species (Hajibabaei et al. 2007). Characteristics of barcodes necessary for their routine use include the following: ease of attainment from any specimen, irrespective of morphology or stage of the fungal life cycle; high taxonomic coverage; and a high resolution power (Hebert et al. 2003*a*). Ideally, barcodes must be unique to a single species, and to be constant within each species to ensure consistency of identification (Hebert et al. 2003*a*, 2003*b*; Letourneau et al. 2010). Additionally, interspecies variation should exceed the intraspecies variation, generating a "break" in the distribution of distances that is referred to as the barcoding gap (Meyer and Paulay 2005). This gap is generally evaluated and represented by the Kimura 2-parameter distance model (Kimura 1980).

Numerous genetic loci have been evaluated as a barcode for fungi, with varying success rates. Protein-coding genes show high taxonomic resolution and identification success in Ascomycota. The most commonly used

genes are *RPB1* (Crespo et al. 2007; Hofstetter et al. 2007; McLaughlin et al. 2009; O'Donnell et al. 2013; Tanabe et al. 2005), *β-tubulin* (Frisvad and Samson 2004), and partial *translational elongation factor 1α* (*TEF1α*) (James et al. 2006; O'Donnell et al. 2010; Schoch et al. 2009). The D1/D2 region of the large subunit (LSU) of the rDNA gene cluster has been used successfully for species identification and strain recognition in yeasts (Fell et al. 2000; Kurtzman and Robnett 1998; Scorzetti et al. 2002). However, the applicability of these genetic loci as a potential barcode is hampered by the lack of standardization of the amplification process, differences in primer selection, their species/genera-dependent resolution power, and the lack of quality-controlled reference databases.

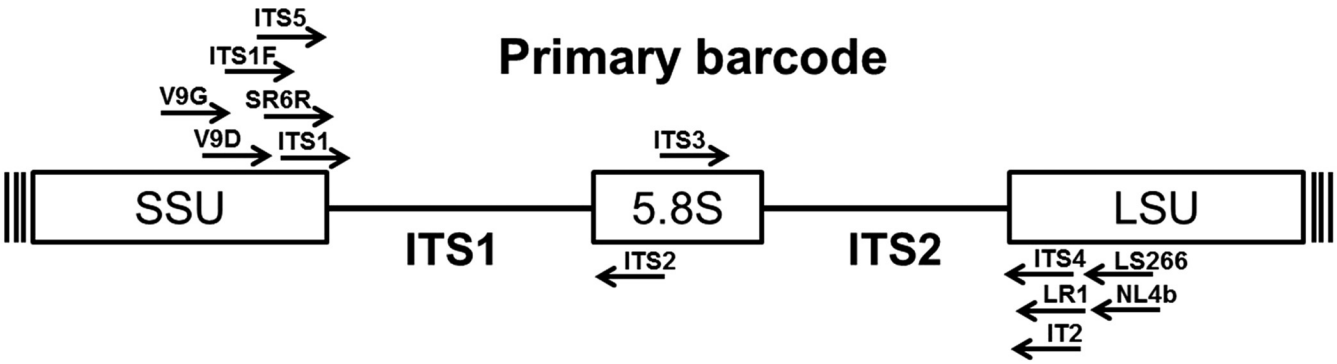
The primary fungal barcode

The internal transcribed spacer (ITS) region is comprised of three units, two non-coding, variable, ITS1 and ITS2 regions, separated by the highly conserved 5.8S gene, which is located between the 18S (small subunit (SSU)) and 28S (LSU) genes in the nrDNA repeat unit (White et al. 1990). The ITS region has long been used for species identification and phylogenetic studies in fungi (Bruns et al. 1991; Gardes and Bruns 1993; Hillis and Dixon 1991; Leaw et al. 2006; Meyer and Gams 2003). Its widespread popularity relies on several facts. There are many universal primers available to bind to DNA of most of the fungal taxa, with the most commonly used being the ITS1, ITS1F, ITS2, ITS3, ITS4, and ITS5 (Gardes and Bruns 1993; White et al. 1990). To avoid cross reactivity with plant or animal DNA, fungus-specific primers were also designed, such as SR6R and LR1 (Vilgalys and Hester 1990), V9D, V9G, and LS266 (Gerrits van den Ende and de Hoog 1999), IT2 (Beguin et al. 2012), ITS1F (Gardes and Bruns 1993), and NL4b (O'Donnell 1993) (Fig. 1). Multiple copies of the ITS are present in the genome which increases significantly the amplification efficiency, even from samples where the initial amount of DNA is low, such as environmental or clinical specimens (Vilgalys and Gonzalez 1990). The relatively short length of the ITS region allows easy amplification and subsequent Sanger sequencing (Seifert 2009). Finally, the ITS region has proven to have a good resolution power leading to species discrimination in most fungal taxa. A significant drawback is its limited resolution in closely related species, or species complexes containing cryptic species (Nilsson et al. 2008). Recognising the many advantages and relatively few limitations of the ITS region for use in identification across the fungal kingdom, it was proposed as a standard fungal DNA barcode in 2012 by Schoch and colleagues (Schoch et al. 2012) — although there were strong criticisms of this proposal by some mycologists (Kiss 2012). The main shortcoming of barcoding for microorganisms, not limited to fungi, has been the lack of adequate and validated reference libraries of DNA barcodes (Seifert 2009). In response to this, the ISHAM-ITS reference database for human and animal pathogenic fungi was established in 2015 (Irinyi et al. 2015). Currently (as of 21 September 2018) this database is in global use and contains 4200 complete ITS sequences, representing 645 fungal species. The ITS locus proved to be sufficient for correct identification of most medically relevant fungal species (Irinyi et al. 2015). However, an exact cut-off point for identification could not be defined, in agreement with other studies (Nilsson et al. 2008), due to a high number of polymorphisms (up to 2.5% of the ITS region) in certain species. For these taxa, it is essential to consider alternative genetic regions to obtain superior resolution power and separation. Correct identification of these taxa cannot be guaranteed by using only ITS sequences, as has been documented in the global project "Assembling the Fungal Tree of Life" (https://aftol.umn.edu) (James et al. 2006). This is particularly relevant to species complexes and sibling species delineation. Preferably, an alternative barcode should contain a single, probably protein-coding gene sequence with high interspecies and low intraspecies divergence and high taxonomic coverage, hence meeting the requirements of a standard barcode, namely, amplified by universal primers with high PCR efficiencies.

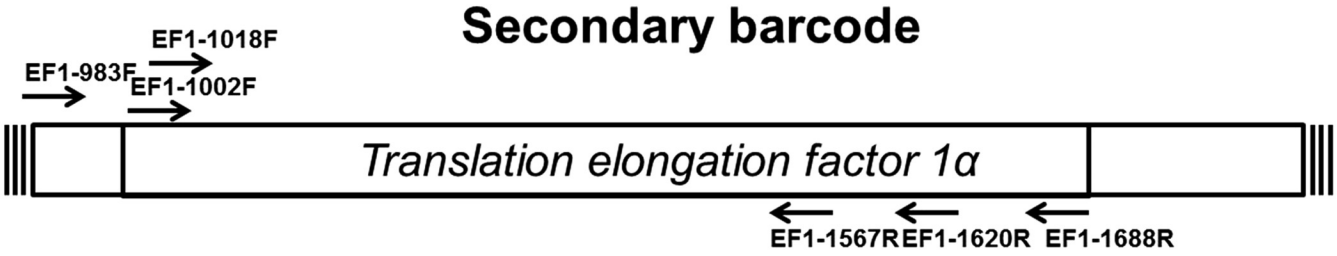
The secondary fungal barcode

Despite the advantages of the ITS locus, such as robust PCR amplification and a high taxonomic coverage, it can only accurately identify around 75% of all medically relevant fungal species using a cut-off of 98.5% sequence identity (Irinyi et al. 2015). It has also been shown that the resolution power of ITS at a higher taxonomic level is inferior to that of many protein coding genes (Nilsson et al. 2006; Seifert 2009). To enhance the selectivity of fungal DNA barcoding and enable the correct identification of all fungal species, the most pragmatic solution was to establish a secondary fungal barcode, recognising that it would be difficult to identify only one DNA region that fulfilled all requirements of a barcode (Stielow et al. 2015). Over the years, many loci have been tested as alternatives to the ITS region but no consensus had been reached among mycologists mainly due to the lack of standardization and the non-availability of universal primers (Frisvad and Samson 2004; McLaughlin et al. 2009; O'Donnell et al. 2010). Whole genome sequencing (WGS) has recently made it possible to carry out full genome comparisons of evolutionarily distinct fungi to identify potential secondary barcode markers and design universal primers for robust PCR amplification (Robert et al. 2011). The proposal was to identify genetic markers with high interspecies and low intraspecies sequence divergence, which accurately reflect higher-level taxonomic affiliations. In a recent study, 14 (partially) universal primer pairs targeting eight genetic markers were tested across more than 1500 species (1931 strains or specimens) to select the most optimal secondary barcode

Fig. 1. Schematic structure of the primary (ITS) and secondary (*TEF1α*) fungal DNA barcode regions indicating universal primers for their amplification.



Primer name	Sequences	References
ITS1:	(5' TCCGTAGGTGAACCTGCGG 3')	(White et al., 1990)
ITS2:	(5' GCTGCGTTCTTCATCGATGC 3')	(White et al., 1990)
ITS3:	(5' GCATCGATGAAGAACGCAGC 3')	(White et al., 1990)
ITS4:	(5' TCCTCCGCTTATTGATATGC 3')	(White et al., 1990)
ITS5:	(5' GGAAGTAAAAGTCGTAACAAGG 3')	(White et al., 1990)
ITS1F:	(5' CTTGGTCATTTAGAGGAAGTAA 3')	(Gardes and Bruns, 1993)
IT2:	(5' CCTCCGCTTATTGATATGCTTAGG 3')	(Beguin et al., 2012)
SR6R:	(5' AAGTATAAGTCGTAACAAGG 3')	(Vilgalys and Hester, 1990)
LR1:	(5' GGTTGGTTTCTTTCCT 3')	(Vilgalys and Hester, 1990)
V9D:	(5' TTAAGTCCCTGCCCTTTGTA 3')	(Gerrits van den Ende and de Hoog, 1999)
V9G:	(5' TACGTCCCTGCCCTTTGTA 3')	(Gerrits van den Ende and de Hoog, 1999)
LS266:	(5' GCATTCCCAAACAACTCGACTC 3')	(Gerrits van den Ende and de Hoog, 1999)
NL4b:	(5' GGATTCTCACCCTCTATGAC 3')	(O'Donnell, 1993)



Primer name	Sequences	References
EF1-1018F (Al33F):	(5' GAYTTCATCAAGAACATGAT 3')	(Stielow et al., 2015)
EF1-1620R (Al33R):	(5' GACGTTGAADCCRACRTTGTC 3')	(Stielow et al., 2015)
EF1-1002F (Al34F):	(5' TTCATCAAGAACATGAT 3')	(Stielow et al., 2015)
EF1-1688R (Al34R):	(5' GCTATCATCACAATGGACGTTCTTGGAG 3')	(Stielow et al., 2015)
EF1-983F:	(5' GCYCCYGGHCAYCGTGAYTTYAT 3')	(Rehner and Buckley, 2005)
EF1-1567R:	(5' ACHGTRCCRATACCACCRATCTT 3')	(Rehner and Buckley, 2005)

marker (Stielow et al. 2015). As a result of this study the *TEF1α* gene proved to be the most promising candidate, and as such it was proposed as the universal secondary fungal DNA barcode, based on its universal taxon applicability and the availability of universal primers, such as EF1-1018F (Al33F)/EF1-1620R (Al33R) or EF1-1002F (Al34F)/EF1-1688R (Al34R) (Fig. 1).

The aims of this publication are as follows:

- To announce the establishment of a quality controlled reference database for the secondary DNA barcode (*TEF1α*) of human and animal pathogenic fungi as a tool to fill the current gap in the diagnosis of mycoses.
- To make a call to the global medical mycology community for further sequence contributions to enable full coverage of all medically important fungal species.

Materials and methods

Cultures

In this study 908 strains, representing 186 human and animal pathogenic fungal species, were used to generate quality controlled *TEF1α* sequences.

DNA extraction

DNA was isolated and purified from cultures using either previously described methods (Ferrer et al. 2001)

or the Quick-DNA Fungal/Bacterial Kit (D6007, Zymo Research), according to the manufacturer's instructions. Laboratories contributing to the current study carried out the DNA extractions by using the methods routinely used in their laboratories.

PCR amplification of the secondary barcode

All PCR reactions were performed in a final volume of 50 μL, made of 5 μL dNTP mix, 5 μL 10 × reaction buffer, 4 μL 10 ng/μL forward and reverse primers, 3 μL 50 mM $MgCl_2$, 15 μL 10 ng/μL genomic DNA, 0.5 μL Taq polymerase, and nuclease-free water up to 50 μL. Partial amplification of the secondary fungal DNA barcode, *TEF1α* region, was done using two sets of primers: Al33F (5′-GAYTTCATCAAGAACATGAT-′3) and Al33R (5′-GACGTT GAADCCRACRTTGTC-′3) or Al34F (5′-TTCATCAAGAAC ATGAT-′3) and Al34R (5′-GCTATCATCACAATGGACGTT CTTGGAG-′3) with the following amplification protocol: 5 min initial denaturing at 94 °C, followed by 40 cycles of 50 min at 94 °C, 50 s annealing at 48 °C, 50 s at 72 °C, and 7 min final extension at 72 °C (Stielow et al. 2015). First, the primer set Al33F-Al33R was used to amplify the *TEF1α* region, and if there was no amplification achieved the second primer set Al34F-Al34R was applied. PCR products were subjected to electrophoresis in a 1.5% agarose gel containing EtBr and visualised using UV illumination. Successfully amplified PCR products were sent for commercial sequencing, e.g., to Macrogen Inc., Korea, in both forward and reverse directions.

Data analysis

Bidirectional sequences were assembled and edited using Sequencher® ver. 5.3. (Gene Codes Corporation, Ann Arbor, MI USA). Sequences were manually checked, and ambiguous bases were corrected based on the forward and reverse trace files, considering the PHRED scores received. The sequences for each taxon were aligned with the program CLUSTALW (Thompson et al. 1994), which is part of the software MEGA ver. 5.2.2 (Tamura et al. 2011). Resulting multiple alignments were then checked visually and edited when needed. For further analyses, the sequences were truncated at conserved sites to obtain equal 3′- and 5′-endings. The intraspecies diversity was estimated by calculating the average nucleotide diversity (π) within species, which had sequences from more than three strains, giving the proportion of nucleotide differences in all haplotypes in the studied sample using the software DnaSP ver. 5.10.01 (Librado and Rozas 2009).

Results and discussion

Establishment of the secondary fungal barcode database

With the *TEF1α* locus being selected as secondary barcode, a dedicated quality-controlled database covering all medically relevant fungal species was established to complement the ISHAM-ITS database.

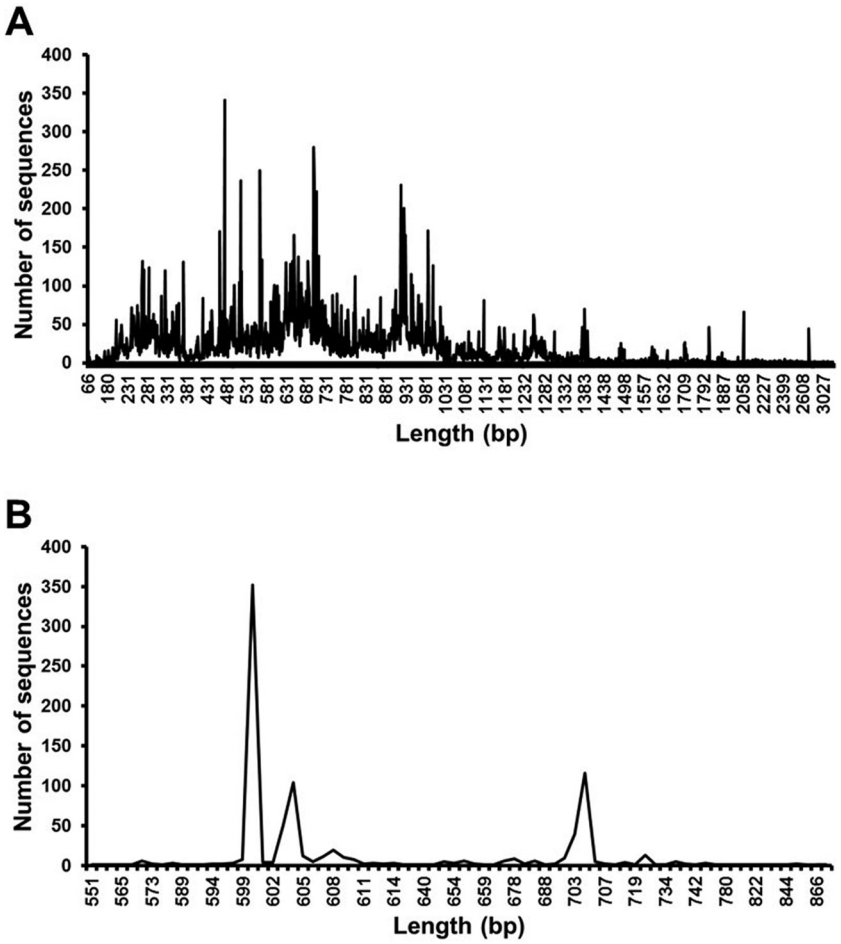
Currently there are 37 773 *TEF1α* sequences available in GenBank (as of 15 March 2018), representing 6283 taxa, of which 670 are only classified to the genus level. However, there are only 6539 sequences from 269 clinically relevant species in 108 fungal genera. The majority of them belong to the genera *Fusarium* (2647), *Cryptococcus* (694), and *Sporothrix* (154). Of these, 114 species are represented by less than two sequences. The length of the *TEF1α* sequences in GenBank ranges from 66 to ~3000 bp (Fig. 2A), indicating that some of the sequences might be too short for correct species identification. In addition, these sequences are not quality controlled and have not been verified at species level, which may lead to an incorrect identification.

The secondary fungal DNA barcode database was established using only quality-controlled *TEF1α* sequences obtained from taxonomically verified fungal cultures. The newly established database contains 908 quality-controlled *TEF1α* sequences, representing 186 pathogenic fungal species, was launched at the recent 7th International DNA Barcoding Conference at the Kruger National Park in South Africa (20–24 November 2017). Sequences from the genera *Candida*, *Cryptococcus*, and *Scedosporium* are most abundant in the new database. In total, 143 species are represented by three strains or less. The length of the partial *TEF1α* sequences in the database ranges from 550–900 bp (Fig. 2B). In a few cases, shorter amplification products of the *TEF1α* were obtained (e.g., for *Trichosporon* spp.).

Overall, the PCR success rate was very good, e.g., in the Molecular Mycology Research Laboratory, Westmead Institute for Medical Research, Westmead, NSW, Australia, 270 *TEF1α* sequences were generated. From those, 220 were produced using the Al33F-Al33R primer set and 50 were amplified with the Al34F-Al34R primer set. There was no trend in amplification success for different fungal species; however, all strains for the species *Aspergillus niger*, *Candida albicans*, *Candida dublinsiensis*, *Kluyveromyces marxianus*, and *Pichia kudriavzevii* required the Al34f-A134R primer set for amplification. However, in some cases no amplification products for some strains of *Cladosporium* spp., *Rhodotorula* spp., or *Trichosporon* spp. were obtained. In those cases, modifications of the amplification conditions, like touchdown PCR, may be appropriate.

Comparison of the nucleotide diversity (π) and number of polymorphic sites (S) of 43 fungal species, with more than three strains per species in the new secondary fungal DNA barcode database (Fig. 3), showed that the *TEF1α* locus is less diverse than the ITS locus for most of the species, with the intraspecies variation being below 1.5%. The pilot data reported herein showed a further reduction in the intraspecies diversity using the *TEF1α* locus compared with the ITS region for 12 fungal species, including *Candida albicans*, *C. glabrata*, *C. metapsilosis*, *C. pararugosa*, *C. tropicalis*, *Clavispora lusitaniae*, *C. gattii*,

Fig. 2. Length distribution of secondary barcode (*TEF1α*) sequences in (A) GenBank and (B) ISHAM barcode database. Only *TEF1α* sequences longer than 550 bp were included.



Histoplasma capsulatum, *Kodamaea ohmeri*, *Scedosporium apiospermum*, *S. aurantiacum*, and *Yarrowia lipolytica*. The combination of the increased number of fungal species, which had less than 1.5% intraspecies variation, with the further reduction in intraspecies variation compared to the ITS confirms the secondary DNA barcode locus (*TEF1α*) as a more discriminatory marker. Further studies, including more strains and covering all human and animal pathogenic fungi, are needed to enable specific indications for the use of the dual fungal DNA barcode.

As a result of this study, the ITS and the *TEF1α* databases where combined together to form the new ISHAM barcode database, which currently contains 4200 ITS and 908 *TEF1α* sequences. The ISHAM barcode database is available online at the following websites: http://its.mycologylab.org/ or the ISHAM website (http://isham.org). This new database allows single locus or polyphasic identification of human and animal pathogenic fungi based on sequence alignments of either the ITS or the *TEF1α* locus sequences, or both, against the reference sequences maintained in the ISHAM barcode database.

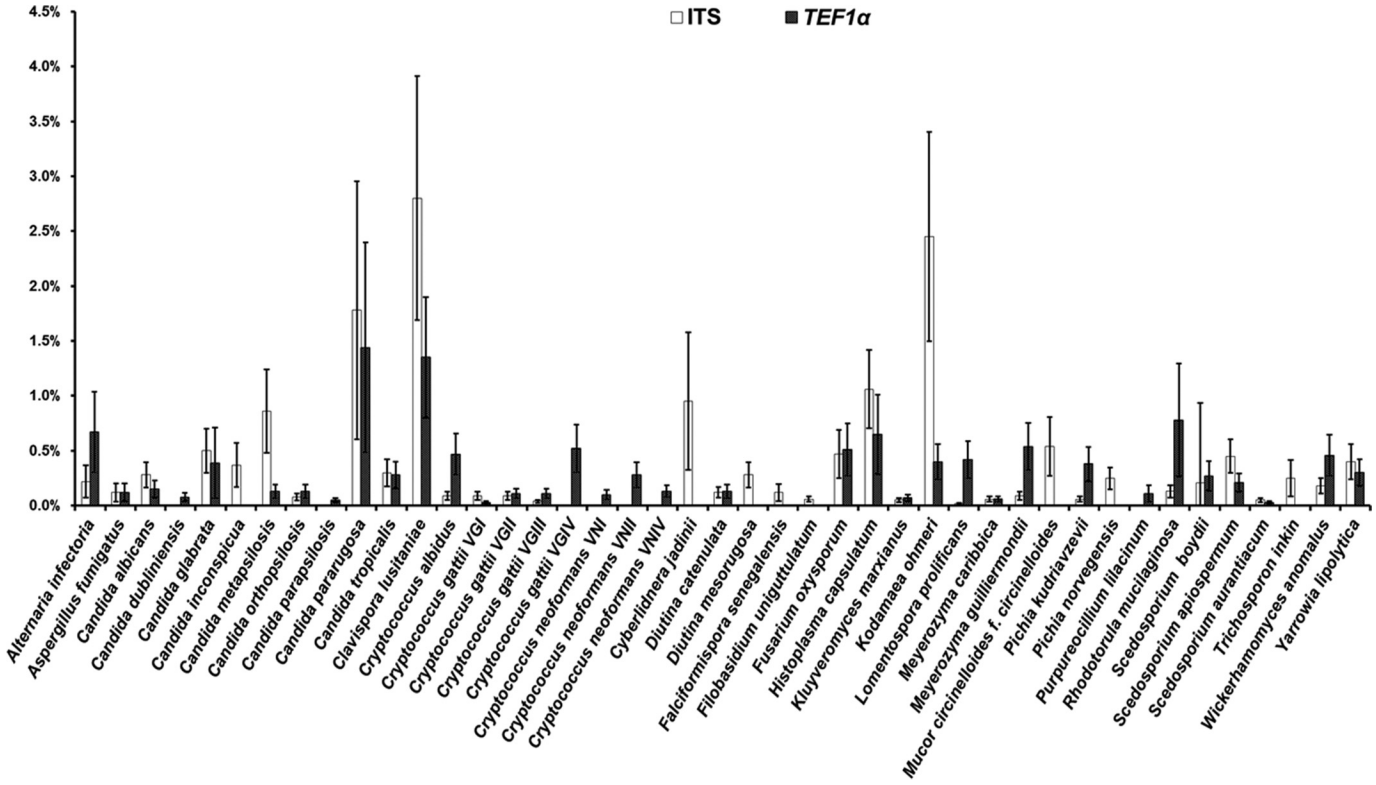
Call for data submission

The database is open for further sequence submissions with the goal of eventually containing sequences from multiple strains of all medically relevant species. Data obtained from taxonomically validated strains can be submitted directly into the database or via contacting the curators of the database: DNA-barcode.database@sydney.edu.au. After validation of the taxonomic metadata and verification of the submitted sequences by the database curator(s), they will be added to the database.

Value of the dual barcoding system for fungal identification

In summary, our data confirm that most medically relevant fungal species can be identified from their ITS sequence, confirming its status as a primary universal fungal DNA barcode. However, some species may not be distinguishable using the ITS region. This can be the case for the following two reasons: either the taxa are not adequately represented in the database, or the ITS region is unable to distinguish between biologically consistent groups. Under these circumstances, the application of a dual DNA barcoding system that incorporates a secondary barcode, *TEF1α*, supported by a quality-controlled reference database, increases the ability to accurate identification of all clinically important fungal pathogens. The herein established secondary fungal DNA barcode

Fig. 3. Intraspecies variation for species, which are represented by more than three strains, in the ITS regions (white bars) — primary DNA barcode — compared with the translation elongation factor 1α (black bars) — secondary DNA barcode.



database offers the basis for a comprehensive database of the fungal kingdom.

Acknowledgements

The authors thank Jeffrey M. Loch (U.S. Geological Survey, National Wildlife Health Center, Madison, USA) for submitting sequences and all reference centres and clinical microbiology laboratories that kindly contributed strains to this study. The study was supported by NH&MRC grants APP1031952 and APP1121936 to W.M., V.R., S.C., and T.C.S., and a REN grant from Western Sydney Local Health Research & Education Network to W.M., T.C.S., S.C., L.I. T.C.S. is a Sydney Medical Foundation Fellow.

Members of the ISHAM barcoding of pathogenic fungi working group include the following: Sharon Chen, Catriona Halliday, Laszlo Irinyi, Wieland Meyer, Tania C. Sorrell (University of Sydney/Westmead Hospital, Sydney, Australia); Ian Arthur (PathWest Laboratory Medicine WA, Nedlands, Australia); Maria Luiza Moretti (University of Campinas, Campinas, Brazil); Célia Maria de Almeida Soares (Universidade Federal de Goiás, Goiânia, Brazil); Mauro de Medeiros Muniz, Rosely Maria Zancopé-Oliveira (Instituto de Pesquisa Clínica Evandro Chagas (IPEC), Fundação Oswaldo Cruz (FIOCRUZ), Rio de Janeiro, Brazil); Analy Salles de Azevedo Melo, Arnaldo L. Colombo, Angela Satie Nishikaku (Universidade Federal de São Paulo, São Paulo, Brazil); Marijke Hendrickx, Dirk Stubbe (BCCM/IHEM, Scientific Institute of Public Health, Brussels, Belgium); Anne-Cécile Normand, Renaud Piarroux, Stéphane Ranque (Aix-Marseille University, Marseille, France); Françoise Dromer, Dea Garcia-Hermoso, Marie Desnos-Ollivier (Institut Pasteur, Paris, France); Michael Arabatzis, Aristea Velegraki (National Kapodistrian University of Athens, Athens, Greece); Gianluigi Cardinali (Università degli Studi di Perugia, Perugia, Italy); Laura Rosio Castañón, Maria Lucia Taylor, Conchita Toriello (Universidad Nacional Autónoma de México, Ciudad de México, México); Sybren de Hoog, Vincent Robert (Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands); Célia Pais (University of Minho, Braga, Portugal); Wilhelm de Beer (FABI, Pretoria, South Africa); Marieka Gryzenhout (University of the Free State, Bloemfontein, South Africa); Josep Guarro, Jose F. Cano-Lira (Universitat Rovira i Virgili, Reus, Spain); Ariya Chindamporn (Chulalongkorn University, Bangkok, Thailand); Chompoonek Yurayart (Kasetsart University, Bangkok, Thailand); Barbara Robbertse, Conrad Schoch (NIH, Bethesda, MD, USA).

References

Alexander, B.D., and Pfaller, M.A. 2006. Contemporary tools for the diagnosis and management of invasive mycoses. Clin. Infect. Dis. **43**: S15–S27. doi:10.1086/504491.

Baires-Varguez, L., Cruz-García, A., Villa-Tanaka, L., Sánchez-García, S., Gaitán-Cepeda, L.A., Sánchez-Vargas, L.O., et al. 2007. Comparison of a randomly amplified polymorphic DNA (RAPD) analysis and ATB ID 32C system for identification of clinical isolates of different *Candida* species. Rev. Iberoam.

Micol. **24**(2): 148–151. doi:10.1016/S1130-1406(07)70031-1. PMID: 17604435.

Balajee, S.A., Gribskov, J.L., Hanley, E., Nickle, D., and Marr, K.A. 2005. *Aspergillus lentulus* sp. nov., a new sibling species of *A. fumigatus*. Eukaryotic Cell, **4**(3): 625–632. doi:10.1128/ec.4.3.625-632.2005.

Balajee, S.A., Sigler, L., and Brandt, M.E. 2007. DNA and the classical way: identification of medically important molds in the 21st century. Med. Mycol. **45**(6): 475–490. doi:10.1080/13693780701449425. PMID:17710617.

Begerow, D., Nilsson, H., Unterseher, M., and Maier, W. 2010. Current state and perspectives of fungal DNA barcoding and rapid identification procedures. Appl. Microbiol. Biotechnol. **87**(1): 99–108. doi:10.1007/s00253-010-2585-4. PMID:20405123.

Beguin, H., Pyck, N., Hendrickx, M., Planard, C., Stubbe, D., and Detandt, M. 2012. The taxonomic status of *Trichophyton quinckeanum* and *T. interdigitale* revisited: a multigene phylogenetic approach. Med. Mycol. **50**(8): 871–882. doi:10.3109/13693786.2012.684153. PMID:22587727.

Bergman, A., Fernandez, V., Holmström, K.O., Claesson, B.E., and Enroth, H. 2007. Rapid identification of pathogenic yeast isolates by real-time PCR and two-dimensional melting-point analysis. Eur. J. Clin. Microbiol. Infect. Dis. **26**(11): 813–818. doi:10.1007/s10096-007-0369-2. PMID:17680284.

Brandt, M.E., Padhye, A.A., Mayer, L.W., and Holloway, B.P. 1998. Utility of random amplified polymorphic DNA PCR and TaqMan automated detection in molecular identification of *Aspergillus fumigatus*. J. Clin. Microbiol. **36**(7): 2057–2062. PMID:9650962.

Bruns, T.D., White, T.J., and Taylor, J.W. 1991. Fungal molecular systematics. Annu. Rev. Ecol. Syst. **22**(1): 525–564. doi:10.1146/annurev.es.22.110191.002521.

Ciardo, D.E., Schar, G., Bottger, E.C., Altwegg, M., and Bosshard, P.P. 2006. Internal transcribed spacer sequencing versus biochemical profiling for identification of medically important yeasts. J. Clin. Microbiol. **44**(1): 77–84. doi:10.1128/JCM.44.1.77-84.2006. PMID:16390952.

Crespo, A., Lumbsch, H.T., Mattsson, J.E., Blanco, O., Divakar, P.K., Articus, K., et al. 2007. Testing morphology-based hypotheses of phylogenetic relationships in Parmeliaceae (Ascomycota) using three ribosomal markers and the nuclear *RPB1* gene. Mol. Phylogenet. Evol. **44**(2): 812–824. doi:10.1016/j.ympev.2006.11.029. PMID:17276700.

Dendis, M., Horvath, R., Michalek, J., Ruzicka, F., Grijalva, M., Bartos, M., and Benedik, J. 2003. PCR-RFLP detection and species identification of fungal pathogens in patients with febrile neutropenia. Clin. Microbiol. Infect. **9**(12): 1191–1202. doi:10.1111/j.1469-0691.2003.00719.x. PMID:14686984.

Diguta, C.F., Vincent, B., Guilloux-Benatier, M., Alexandre, H., and Rousseaux, S. 2011. PCR ITS-RFLP: a useful method for identifying filamentous fungi isolates on grapes. Food Microbiol. **28**(6): 1145–1154. doi:10.1016/j.fm.2011.03.006. PMID: 21645813.

Fell, J.W., Boekhout, T., Fonseca, A., Scorzetti, G., and Statzell-Tallman, A. 2000. Biodiversity and systematics of basidiomycetous yeasts as determined by large-subunit rDNA D1/D2 domain sequence analysis. Int. J. Syst. Evol. Microbiol. **50**(3): 1351–1371. doi:10.1099/00207713-50-3-1351. PMID:10843082.

Ferrer, C., Colom, F., Frases, S., Mulet, E., Abad, J., and Alio, J. 2001. Detection and identification of fungal pathogens by PCR and by ITS2 and 5.8S ribosomal DNA typing in ocular infections. J. Clin. Microbiol. **39**: 2873–2879. doi:10.1128/JCM.39.8.2873-2879.2001. PMID:11474006.

Frisvad, J.C., and Samson, R.A. 2004. Polyphasic taxonomy of *Penicillium* subgenus *Penicillium*— a guide to identification of food and air-borne terverticillate penicillia and their mycotoxins. Stud. Mycol. **49**: 1–174.

Gardes, M., and Bruns, T.D. 1993. ITS primers with enhanced specificity for *Basidiomycetes* — application to the identification of mycorrhizae and rusts. Mol. Ecol. **2**(2): 113–118. doi:10.1111/j.1365-294X.1993.tb00005.x. PMID:8180733.

Gerrits van den Ende, A.H.G., and de Hoog, G.S. 1999. Variability and molecular diagnostics of the neurotropic species *Cladophialophora bantiana*. Stud. Mycol. **43**: 151–162.

Gilgado, F., Cano, J., Gené, J., and Guarro, J. 2005. Molecular phylogeny of the *Pseudallescheria boydii* species complex: proposal of two new species. J. Clin. Microbiol. **43**(10): 4930–4942. doi:10.1128/JCM.43.10.4930-4942.2005. PMID:16207945.

Hajibabaei, M., Singer, G.A., Hebert, P.D., and Hickey, D.A. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet. **23**(4): 167–172. doi:10.1016/j.tig.2007.02.001. PMID:17316886.

Hebert, P.D., Cywinska, A., Ball, S.L., and deWaard, J.R. 2003*a*. Biological identifications through DNA barcodes. Proc. Biol. Sci. **270**(1512): 313–321. doi:10.1098/rspb.2002.2218. PMID: 12614582.

Hebert, P.D., Ratnasingham, S., and deWaard, J.R. 2003*b*. Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. Proc. R. Soc. B Biol. Sci. **270**(S1): S96–S99. doi:10.1098/rsbl.2003.0025.

Hierro, N., González, A., Mas, A., and Guillamón, J.M. 2004. New PCR-based methods for yeast identification. J. Appl. Microbiol. **97**(4): 792–801. doi:10.1111/j.1365-2672.2004.02369.x. PMID: 15357729.

Hillis, D.M., and Dixon, M.T. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Q. Rev. Biol. **66**(4): 411–453. doi:10.1086/417338. PMID:1784710.

Hofstetter, V., Miadlikowska, J., Kauff, F., and Lutzoni, F. 2007. Phylogenetic comparison of protein-coding versus ribosomal RNA-coding sequence data: a case study of the Lecanoromycetes (Ascomycota). Mol. Phylogenet. Evol. **44**(1): 412–426. doi:10.1016/j.ympev.2006.10.016. PMID:17207641.

Irinyi, L., Serena, C., Garcia-Hermoso, D., Arabatzis, M., Desnos-Ollivier, M., Vu, D., et al. 2015. International Society of Human and Animal Mycology (ISHAM)-ITS reference DNA barcoding database-the quality controlled standard tool for routine identification of human and animal pathogenic fungi. Med. Mycol. **53**(4): 313–337. doi:10.1093/mmy/myv008. PMID:25802363.

James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., et al. 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature, **443**(7113): 818–822. doi:10.1038/nature05110. PMID:17051209.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**(2): 111–120. doi:10.1007/BF01731581. PMID:7463489.

Kiss, L. 2012. Limits of nuclear ribosomal DNA internal transcribed spacer (ITS) sequences as species barcodes for Fungi. Proc. Natl. Acad. Sci. U.S.A. **109**(27): E1811. doi:10.1073/pnas.1207143109. PMID:22715287.

Klingspor, L., and Jalal, S. 2006. Molecular detection and identification of *Candida* and *Aspergillus* spp. from clinical samples using real-time PCR. Clin. Microbiol. Infect. **12**(8): 745–753. doi:10.1111/j.1469-0691.2006.01498.x. PMID:16842569.

Kulik, T., Fordoński, G., Pszczółkowska, A., Płodzień, K., and Łapiński, M. 2004. Development of PCR assay based on ITS2 rDNA polymorphism for the detection and differentiation of *sporotrichioides*. FEMS Microbiol. Lett. **239**(1): 181–186. doi:10.1016/j.femsle.2004.08.037. PMID:15451117.

Kurtzman, C., and Robnett, C. 1998. Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. Antonie van Leeuwenhoek, **73**(4): 331–371. doi:10.1023/A:1001761008817. PMID:9850420.

Leaw, S.N., Chang, H.C., Sun, H.F., Barton, R., Bouchara, J.-P.,

and Chang, T.C. 2006. Identification of medically important yeast species by sequence analysis of the internal transcribed spacer regions. J. Clin. Microbiol. **44**(3): 693–699. doi:10.1128/JCM.44.3.693-699.2006. PMID:16517841.

Letourneau, A., Seena, S., Marvanová, L., and Bärlocher, F. 2010. Potential use of barcoding to identify aquatic hyphomycetes. Fungal Divers. **40**(1): 51–64. doi:10.1007/s13225-009-0006-8.

Librado, P., and Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics, **25**(11): 1451–1452. doi:10.1093/bioinformatics/btp187. PMID:19346325.

Lieckfeldt, E., Meyer, W., and Börner, T. 1993. Rapid identification and differentiation of yeasts by DNA and PCR fingerprinting. J. Basic Microbiol. **33**(6): 413–425. doi:10.1002/jobm.3620330609. PMID:8271158.

Lindsley, M.D., Hurst, S.F., Iqbal, N.J., and Morrison, C.J. 2001. Rapid identification of dimorphic and yeast-like fungal pathogens using specific DNA probes. J. Clin. Microbiol. **39**(10): 3505–3511. doi:10.1128/JCM.39.10.3505-3511.2001. PMID:11574564.

Martin, C., Roberts, D., van Der Weide, M., Rossau, R., Jannes, G., Smith, T., and Maher, M. 2000. Development of a PCR-based line probe assay for identification of fungal pathogens. J. Clin. Microbiol. **38**(10): 3735–3742. PMID:11015393.

McLaughlin, D.J., Hibbett, D.S., Lutzoni, F., Spatafora, J.W., and Vilgalys, R. 2009. The search for the fungal tree of life. Trends Microbiol. **17**(11): 488–497. doi:10.1016/j.tim.2009.08.001.

Meyer, C.P., and Paulay, G. 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol. **3**(12): e422. doi:10.1371/journal.pbio.0030422. PMID:16336051.

Meyer, W., and Gams, W. 2003. Delimitation of *Umbelopsis* (*Mucorales*, *Umbelopsidaceae* fam. nov.) based on ITS sequence and RFLP data. Mycol. Res. **107**(3): 339–350. PMID:12825503.

Nilsson, R., Ryberg, M., Kristiansson, E., Abarenkov, K., Larsson, K., and Kõljalg, U. 2006. Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. PLoS ONE, **1**(1): e59. doi:10.1371/journal.pone.0000059. PMID:17183689.

Nilsson, R.H., Kristiansson, E., Ryberg, M., Hallenberg, N., and Larsson, K.H. 2008. Intraspecific ITS variability in the kingdom *Fungi* as expressed in the international sequence databases and its implications for molecular species identification. Evol. Bioinform. Online, **4**: 193–201. PMID:19204817.

Nilsson, R.H., Ryberg, M., Abarenkov, K., Sjokvist, E., and Kristiansson, E. 2009. The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. FEMS Microbiol. Lett. **296**(1): 97–101. doi:10.1111/j.1574-6968.2009.01618.x. PMID:19459974.

O'Donnell, K. 1993. *Fusarium* and its near relatives. *In* The fungal holomorph: mitotic, meiotic and pleomorphic speciation in fungal systematics. *Edited by* D.R. Reynolds and J.W. Taylor. CAB International, Wallingford, U.K. pp. 225–233.

O'Donnell, K. 2000. Molecular phylogeny of the *Nectria haematococca-solani* species complex. Mycologia, **92**(5): 919–938. doi:10.2307/3761588.

O'Donnell, K., Sutton, D.A., Rinaldi, M.G., Sarver, B.A., Balajee, S.A., Schroers, H.J., et al. 2010. Internet-accessible DNA sequence database for identifying fusaria from human and animal infections. J. Clin. Microbiol. **48**(10): 3708–3718. doi:10.1128/JCM.00989-10. PMID:20686083.

O'Donnell, K., Rooney, A.P., Proctor, R.H., Brown, D.W., McCormick, S.P., Ward, T.J., et al. 2013. Phylogenetic analyses of *RPB1* and *RPB2* support a middle Cretaceous origin for a clade comprising all agriculturally and medically important fusaria. Fungal Genet. Biol. **52**: 20–31. doi:10.1016/j.fgb.2012.12.004. PMID:23357352.

Pryce, T.M., Palladino, S., Kay, I.D., and Coombs, G.W. 2003. Rapid identification of fungi by sequencing the ITS1 and ITS2 regions using an automated capillary electrophoresis system. Med. Mycol. **41**(5): 369–381. doi:10.1080/13693780310001600435. PMID:14653513.

Rehner, S.A., and Buckley, E. 2005. A *Beauveria* phylogeny inferred from nuclear ITS and EF1-a sequences: evidence for cryptic diversification and links to *Cordyceps* teleomorphs. Mycologia, **97**: 84–98. doi:10.1080/15572536.2006.11832842. PMID:16389960.

Robert, V., Szöke, S., Eberhardt, U., Cardinali, G., Seifert, K.A., Lévesque, C.A., et al. 2011. The quest for a general and reliable fungal DNA barcode. Open Appl. Inform. J. **5**(Suppl. 1-M6): 45–61. doi:10.2174/1874136301105010045.

Romanelli, A.M., Sutton, D.A., Thompson, E.H., Rinaldi, M.G., and Wickes, B.L. 2010. Sequence-based identification of filamentous basidiomycetous fungi from clinical specimens: a cautionary note. J. Clin. Microbiol. **48**(3): 741–752. doi:10.1128/JCM.01948-09.

Sandhu, G.S., Kline, B.C., Stockman, L., and Roberts, G.D. 1995. Molecular probes for diagnosis of fungal infections. J. Clin. Microbiol. **33**(11): 2913–2919. PMID:8576345.

Sangoi, A.R., Rogers, W.M., Longacre, T.A., Montoya, J.G., Baron, E.J., and Banaei, N. 2009. Challenges and pitfalls of morphologic identification of fungal infections in histologic and cytologic specimens: a ten-year retrospective review at a single institution. Am. J. Clin. Pathol. **131**(3): 364–375. doi:10.1309/AJCP99OOOZSNISCZ. PMID:19228642.

Schoch, C.L., Sung, G.H., Lopez-Giráldez, F., Townsend, J.P., Miadlikowska, J., Hofstetter, V., et al. 2009. The Ascomycota tree of life: a phylum-wide phylogeny clarifies the origin and evolution of fundamental reproductive and ecological traits. Syst. Biol. **58**(2): 224–239. doi:10.1093/sysbio/syp020. PMID:20525580.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., and Chen, W. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc. Natl. Acad. Sci. U.S.A. **109**(16): 6241–6246. doi:10.1073/pnas.1117018109. PMID:22454494.

Scorzetti, G., Fell, J.W., Fonseca, A., and Statzell-Tallman, A. 2002. Systematics of basidiomycetous yeasts: a comparison of large subunit D1/D2 and internal transcribed spacer rDNA regions. FEMS Yeast Res. **2**(4): 495–517. doi:10.1111/j.1567-1364.2002.tb00117.x. PMID:12702266.

Seifert, K.A. 2009. Progress towards DNA barcoding of fungi. Mol. Ecol. Resour. **9**(S1): 83–89. doi:10.1111/j.1755-0998.2009.02635.x. PMID:21564968.

Seifert, K.A., and Samuels, G.J. 2000. How should we look at anamorphs? Stud. Mycol. **45**: 5–18.

Stielow, J.B., Lévesque, C.A., Seifert, K.A., Meyer, W., Iriny, L., Smits, D., et al. 2015. One fungus, which genes? Development and assessment of universal primers for potential secondary fungal DNA barcodes. Persoonia, **35**: 242–263. doi:10.3767/003158515X689135. PMID:26823635.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**(10): 2731–2739. doi:10.1093/molbev/msr121. PMID:21546353.

Tanabe, Y., Watanabe, M.M., and Sugiyama, J. 2005. Evolutionary relationships among basal fungi (Chytridiomycota and Zygomycota): insights from molecular phylogenetics. J. Gen. Appl. Microbiol. **51**(5): 267–276. doi:10.2323/jgam.51.267. PMID:16314681.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**(22): 4673–4680. doi:10.1093/nar/22.22.4673. PMID:7984417.

Vilgalys, R., and Gonzalez, D. 1990. Organization of ribosomal

DNA in the basidiomycete *Thanatephorus praticola*. Curr. Genet. **18**(3): 277–280. doi:10.1007/BF00318394. PMID:2249259.

Vilgalys, R., and Hester, M. 1990. Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several *Cryptococcus* species. J. Bacteriol. **172**(8): 4238–4246. doi:10.1128/jb.172.8.4238-4246.1990. PMID:2376561.

White, T.J., Bruns, T., Lee, S., and Taylor, J.W. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In* PCR protocols: a guide to methods and applications. *Edited by* M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White. Academic Press, New York. pp. 315–322.