FUNGAL BIOLOGY 120 (2016) 125–136

British Mycological Society promoting fungal science

journal homepage: www.elsevier.com/locate/funbio

DNA barcoding of fungi causing infections in humans and animals

Laszlo IRINYI[a], *Michaela LACKNER*[b], *G. Sybren DE HOOG*[c], *Wieland MEYER*[a,*]

[a]*Molecular Mycology Research Laboratory, Centre for Infectious Diseases and Microbiology, Sydney Medical School – Westmead Hospital, Marie Bashir Institute for Infectious Diseases and Biosecurity, The University of Sydney, Westmead Millennium Institute, Sydney, NSW, Australia*

[b]*Division of Hygiene and Medical Microbiology, Medical University of Innsbruck, Innsbruck, 6020, Austria*

[c]*CBS-KNAW Fungal Biodiversity Centre, Utrecht, 3508 AD, The Netherlands*

ARTICLE INFO

Article history:
Received 3 February 2015
Received in revised form
12 April 2015
Accepted 17 April 2015
Available online 30 April 2015
Corresponding Editor:
Sybren de Hoog

Keywords:
DNA barcoding
Fungal identification
Intraspecies/interspecies genetic diversity
ITS region
Pathogenic fungi
Reference ITS database

ABSTRACT

Correct species identification is becoming increasingly important in clinical diagnostics. Till now, many mycological laboratories rely on conventional phenotypic identification. But this is slow and strongly operator-dependent. Therefore, to improve the quality of pathogen identification, rapid, reliable, and objective identification methods are essential. One of the most encouraging approaches is molecular barcoding using the internal transcribed spacer (ITS) of the rDNA, which is rapid, easily achievable, accurate, and applicable directly from clinical specimens. It relies on the comparison of a single ITS sequence with a curated reference database. The International Society for Human and Animal Mycology (ISHAM) working group for DNA barcoding has recently established such a database, focusing on the majority of human and animal pathogenic fungi (ISHAM-ITS, freely accessible at http://www.isham.org/ or directly from http://its.mycologylab.org). For some fungi the use of secondary barcodes may be necessary.

Introduction

In microorganisms such as bacteria and fungi species delimitation is challenging and often ambiguous because of unclear sexuality when compared e.g. to higher animals. Morphology of fungi may be very simple (yeasts) or very complex (filamentous fungi), but in any case identification with phenotypic data requires well-trained experts. Moreover, complex life cycles, such as yeast–mycelial transitions or synanamorphs of certain fungi, may impede correct morphological identification (Seifert & Samuels 2000). Besides the fact that fungi compose one of the largest kingdoms, many of them are still yet to be described (Mora *et al.* 2011). Numerous fungi are not culturable, making species identification often nearly impossible (Nilsson *et al.* 2009; Begerow *et al.* 2010). Accurate identification of fungal pathogens continues to be a difficult task in

* *Corresponding author*. Molecular Mycology Research Laboratory, MBI, CIDM, Level 4, Room O.4.04, 176 Hawkesbury Road, Westmead, NSW 2145, Australia. Tel.: +61 2 86273430; fax: +61 2 98915317.
E-mail address: wieland.meyer@sydney.edu.au (W. Meyer).

Abbreviations; BOLD, Barcode of Life Data System.; ISHAM, International Society for Human and Animal Mycology.; RTL, RefSeq Targeted Loci

http://dx.doi.org/10.1016/j.funbio.2015.04.007
1878-6146/

Terminology used:

Barcode	single diagnostic marker showing variation that is universally usable for species distinction.
Barcoding gap	difference between inter- and intraspecific genetic distances in a group of organisms.
Diagnostic marker	character present in all members of the species enabling confident distinction from other species.
Geneological concordance	concordance of multiple gene genealogies as displayed in phylogenetic trees of independent markers.
Haplotype	the unity of genes that a progeny inherits from its parent(s).
Haplotype diversity	relative haplotype frequency in a species.
Nucleotide diversity	average number of nucleotide differences per site.
Polyphasic classification	taxonomy using a multitude of characters, of both phenotypic and genotypic nature.
Polymorphism	more than one allele occurring on a single locus within an alignment.
Population	group of organisms differing from members of other groups within the same biological species, i.e. potentially interbreeding with these members yielding viable progeny.
Segregating site	position in an alignment which is polymorphic.

the clinical laboratory. It is important to define the etiologic agent of a disease to detect novel agents of disease and to predict intrinsic resistance to antifungal agents. Molecular information has widely contributed to the taxonomy, classification, phylogenetic inference, species delimitation and identification of fungi (James *et al.* 2006; Hibbett *et al.* 2007; Kõljalg *et al.* 2013). In the future, advances in molecular-based technologies are expected to drastically improve species identification and discovery of fungal species.

The process of barcode development: from phylogeny to identification

Barcoding requires the availability of a sound taxonomic system and thus is no replacement for comprehensive taxonomic and phylogenetic analysis (Ebach & Holdrege 2005; Will *et al.* 2005). Taxonomy is the science of species delimitation, while barcoding is the selection of traits for identification (Fig 1). Today, taxonomic studies require phylogenetic information to position the fungus in the fungal kingdom. A representative sample is deposited as type specimen as an anchor of the species for future reference. Additional individuals are studied with a multitude of techniques, of both phenotypic and genotypic nature. All characters in such a polyphasic study are combined to recognize clusters of mutually similar individuals. Group members that cluster in a single group should show genealogical concordance and then are judged to belong to the same taxonomic entity. The recognized species clusters may have different degrees of heterogeneity; strains of a single cluster may be invariant for that gene, or deviate from each other significantly. Cluster-recognition via concordance is relative, and there is no cut-off level of difference where strains can no longer be accepted as belonging to the same species. Consistently different populations are proven to belong to a single species by positive mating results (Fig 2). However, their offspring may have lower fitness than progeny from parents within the same population. Thus, even if mating is technically possible *in vitro* – which is often not the case – species are still delimited with some difficulties and species borderlines may not be sharp. Acknowledging these limitations, diagnostic markers have to be selected that are present in all members of the species and with which the species can be distinguished confidently from other species. A barcode is a single diagnostic marker that is universally usable for that purpose. Validation of the marker, i.e. determining its reproducibility, sensitivity, and specificity is subsequently needed. Only with this full understanding of the species other characteristics, including antifungal susceptibility and virulence, can be meaningfully investigated.

Obviously, if an unknown specimen does not provide a close match (defined below) to an existing record in the barcode database, the barcode sequence does not entitle automatically the unknown specimen as a new species. It merely states that this specimen could be a species not present in the database being used and it should be selected for a more detailed large-scale taxonomic analysis. DNA barcoding could help species discovery, especially in poorly studied taxonomic groups. Due to the high taxonomic diversity of medically relevant fungal taxa, different genetic loci are commonly used including the D1/D2 domain of the 28S rRNA large subunit (LSU) in *Candida* (Kurtzman & Robnett 1997; Fell *et al.* 2000; Scorzetti *et al.* 2002), internal transcribed spacer (ITS) of the ribosomal DNA (rDNA) in *Exophiala* (Rainer and de Hoog 2006), β-tubulin (Balajee *et al.* 2007) and calmodulin (Samson *et al.* 2014) in *Aspergillus*, or multilocus sequencing in *Cryptococcus* (Ngamskulrungroj *et al.* 2009). Phylogenetic studies mostly use multi-gene approaches selecting those markers which were proven to be the most suitable to the targeted group of species. These loci are carefully selected by the mycological community to optimize the construction of a strongly supported phylogenetic tree. The amplification of these genetic markers often requires tailored PCR conditions and species- or genus-specific primers. In contrast, barcodes need to be standardized for an entire eukaryotic kingdom and easily amplifiable by universal primers using standard laboratory conditions. Given the fact that standard Sanger sequencing is the most commonly applied sequencing technique so far, the barcodes need to be easily sequenced without any major modification. It is important to note that species identification should differ from resolving phylogenies at higher taxonomic levels which requires a different strategy for the selection of loci and algorithms. To study phylogenetic relationships, single copy, slowly evolving genetic markers (e.g. protein coding genes) are more suitable