Accurate identification of *Candida parapsilosis* (*sensu lato*) using mitochondrial DNA and real-time PCR

Ana Carolina R. Souza¹, Renata C. Ferreira¹,²,³, Sarah S. Gonçalves¹, Guillermo Quindós⁴, Elena Eraso⁴, Fernando C. Bizerra¹, Marcelo R. S. Briones²,³ and Arnaldo L. Colombo²,#

¹Laboratório Especial de Micologia, Disciplina de Infectologia, Universidade Federal de São Paulo, Rua Pedro de Toledo 669, 5 andar, CEP 04039-032, São Paulo, SP, Brasil
²Departamento de Microbiologia, Imunologia e Parasitologia, Universidade Federal de São Paulo, Rua Botucatu 862, ECB 3 andar, CEP 04023-062, São Paulo, SP, Brasil
³Laboratório de Genômica Evolutiva e Biocomplexidade, Universidade Federal de São Paulo, Rua Pedro de Toledo 669, 4 andar, CEP 04039-032, São Paulo, SP, Brasil
⁴Laboratorio de Micología Médica, Departamento de Inmunología, Microbiología y Parasitología, Facultad de Medicina y Odontología, Universidad del País Vasco/Euskal Herriko Unibertsitatea, Bilbao, Spain

Running title: *C. parapsilosis* species identification by TaqMan PCR

#Corresponding author:

Arnaldo L. Colombo

Laboratório Especial de Micologia
Universidade Federal de São Paulo,
Rua Botucatu 740, CEP 04023-062, São Paulo, SP, Brasil
Phone and Fax: 55 11 5083-0806
EMAIL: arnaldocolombo@gmail.com
Abstract

*Candida parapsilosis* is the second most frequently isolated *Candida* species from blood cultures in South America and some European countries, such as Spain. Since 2005, this species has been considered a complex of 3 closely related species: *C. parapsilosis*, *Candida metapsilosis* and *Candida orthopsilosis*. Here, we described a real-time TaqMan-MGB PCR assay, using mitochondrial DNA (mtDNA) as target, which readily distinguishes these 3 species. We first used comparative genomics to locate syntenic regions between these 3 mitochondrial genomes and then selected *NADH5* as target for the real-time PCR assay. Probes were designed to include a combination of different SNPs that are able to differentiate each species within *C. parapsilosis* complex. This new methodology was first tested using mtDNA and then genomic DNA from 4 reference and 5 clinical strains. For assay validation, a total of 96 clinical isolates and 4 American Type Culture Collection (ATCC) previously identified by *ITS* rDNA (internal transcribed spacers of rDNA) sequencing were tested. Real-time PCR using genomic DNA was able to differentiate the 3 species with 100% accuracy. No amplification was observed when DNA from other species was used as templates. We observed 100% congruence with *ITS* rDNA sequencing identification, including 30 strains used in blind testing. This novel method allows a quick and accurate intra-complex identification of *C. parapsilosis* and saves time compared with sequencing, which so far has been considered the “gold standard” for yeast identification. In addition, this assay provides a useful tool for epidemiological and clinical studies of these emergent species.
Introduction

Candida albicans remains the most prevalent species in human superficial and invasive Candida infections, although there is concern over the increasing rates of non-C. albicans infections worldwide (2, 18). C. parapsilosis is a relevant pathogen primarily in South America, where it causes from 19% to 38% of all episodes of candidemia (6, 25, 26). Similarly, data by Almirante et al. (2005) and Canton et al. (2011) show that C. parapsilosis causes from 15% to 23% of all hematogenous candidiasis cases in Spanish hospitals (1, 4).

Recently, the genetically heterogeneous taxon C. parapsilosis was reclassified into 3 species: Candida parapsilosis (sensu stricto), Candida orthopsilosis and Candida metapsilosis (30). Although the epidemiological and clinical differences caused by these Candida species have not yet been fully determined, several in vitro studies have demonstrated biological differences between them, including the expression of virulence factors, susceptibility to antifungal agents and geographical distribution (3, 10, 12, 13, 22, 24).

Currently, species identification within the C. parapsilosis complex is based on DNA techniques such as RAPD, RFLP of the SADH locus and sequencing of the internal transcribed spacer, ITS (13, 20, 31). These methods are time consuming, labor intensive and, with the exception of DNA sequencing, may have accuracy and reproducibility limitations (23, 27).

Because mutation rates in the mitochondrial genome are higher than in the nuclear genome, it is possible to obtain sufficient resolution to distinguish close
phylogenetic relationships (5). Indeed, our group has shown that mitochondrial DNA (mtDNA) sequences can be readily used to discriminate Candida glabrata isolates. Analysis of the cytochrome c oxidase subunit 2 (COX2) enables typing of C. glabrata and clustering of strains associated with their geographical origins (28).

In the present study, we describe a specific, fast, sensitive and accurate real-time PCR method based on mtDNA-specific target detection that is capable of discriminating all 3 species within the C. parapsilosis complex.

Material and Methods

Microorganisms: A total of 100 C. parapsilosis (sensu lato) strains were used to validate our assay: (i) four reference strains obtained from the American Type Culture Collection (ATCC), C. parapsilosis (sensu stricto) ATCC 22019 and ATCC 90018, C. orthopsilosis ATCC 96141 and C. metapsilosis ATCC 96143; (ii) a panel of 66 clinical isolates obtained from different body sites and geographical regions of Brazil, previously identified at the species level by sequencing of the ITS region (part of this collection had already been deposited in GenBank (NCBI) (13); and (iii) a group of 30 well-characterized C. parapsilosis (sensu lato) clinical isolates that were kindly provided in a blind fashion by Dr. Quindós (Bilbao, Spain) (Table S1) (24). In addition, the reference strains C. albicans (ATCC 90029), Candida tropicalis (ATCC 750), C. glabrata (ATCC 90030), Candida lusitaniae (ATCC 66035), Candida krusei (ATCC 6258) and Lodderomyces elongisporus (CBS 2605) were also tested as controls. Viability, purity and phenotypic identification of the samples were performed as previously described (13).
Microorganism identification using amplification and sequencing of the ITS region of rDNA: The genomic DNA was extracted from single colonies using glass beads (19). Amplification and sequencing were performed with the universal primers ITS1 and ITS4 (Table 1) (32). PCR and sequencing were performed as previously described by our group (14). The sequences generated in this study have been deposited in the GenBank (NCBI) database (http://www.ncbi.nlm.nih.gov/genbank).

In silico selection of the mtDNA target gene for species identification: Complete mitochondrial genome sequences from 3 Candida species were retrieved from GenBank (NCBI): C. parapsilosis CBS7157 (access number-X74411), C. parapsilosis CLIB214 (DQ026513), C. orthopsilosis MCO456 (AY962590), C. orthopsilosis MCO471 (DQ026513), C. metapsilosis MCO448 (NC006971) and C. metapsilosis PL448 (AY391853). We generated a whole genome alignment using the MAUVE program, Progressive Mauve algorithm (7) to verify the synteny. Selected sequences were aligned using Seaview v.4 (16).

Extraction and amplification of yeast mitochondrial DNA for verification of the SNP presence on the target gene: Mitochondrial DNA was isolated as described elsewhere with a yield of 100 ng/µl (8). A 620bp fragment of mitochondrial NADH5 was amplified and sequenced with specific primers (Table 1). PCR (25µl) were performed using the MasterMix protocol (Promega, Madison, WI). The conditions of PCR cycles were: 3 min at 95 °C, followed by 40 cycles at
95 °C for 45 s, 52 °C for 45 s, and 1 min at 72 °C, followed by a final 5 min, 72 °C, extension.

PCR products were purified using Amicon Ultra-0.5 kit (Millipore Massachusetts, United States) and sequenced in an ABI3100 automated sequencer (Applied Biosystems). Phred-Phrap-Consed were used for the assembly and finishing of high quality sequences (Phred scores>40) (9, 15). Species were identified by Basic Local Alignment Search Tool (BLAST), with e-value <10^-5 as cutoff.

Identification of C. parapsilosis (sensu stricto), C. orthopsilosis and C. metapsilosis strains by Real-Time PCR

Primers and probes design: Primers and specific probes for NADH5 gene were designed using the Primer Express Software (Applied Biosystems). Sequences and characteristics of primers and probes are shown in table 1.

Real-Time PCR assay: Real-time PCR was performed in a StepOnePlus Real-Time PCR System (Applied Biosystems). The reaction mix contained 12.5 μl of TaqMan Universal PCR Master Mix (Applied Biosystems), 400 nM of each primer, 250 nM of probe and 160 ng of genomic DNA (or 40 ng of mtDNA) from the different tested strains in a final volume of 25 μl. A single-color simplex assay was used, in which each probe was added to different reaction tubes. Real-time PCR conditions were as follows: initial denaturation at 50 °C for 2 min and 95 °C for 10 min; 40 cycles of 95 °C for 15 s and 51 °C for 30 s; and a final elongation step of 1 min at 60 °C. Fluorescence was measured for 1 min during the elongation step.
The assay sensitivity was estimated by comparing the real-time identification results with the identification by ITS sequencing. To ensure the specificity of the assay, we used genomic DNA from *L. elongisporus* and different *Candida* species (*C. albicans*, *C. tropicalis*, *C. glabrata*, *C. lusitaniae* and *C. krusei*) as templates in the reaction.

**Results**

**Molecular Identification by sequencing of the ITS region of all clinical isolates (66 Brazilian and 30 Spanish strains):** The strains were initially identified as *C. parapsilosis* (*sensu lato*) on the basis of their micromorphology characteristics and biochemical profiles using ID32C System (bioMérieux, Marcy l’Etoile, France) (data not shown). ITS sequencing was used as “gold standard” for the identification at the species level of all clinical strains tested. The amplicons lengths of the ITS regions were 593bp to 612bp and sequences generated were used in BLAST searches (http://www.ncbi.nlm.nih.gov) to confirm the preliminary identifications. BLAST analysis with the ITS sequences were capable of identifying all reference and clinical strains as follows: 50 isolates were identified as *C. parapsilosis* (*sensu stricto*); 40, as *C. orthopsilosis*; and 10, as *C. metapsilosis*.

**In silico analysis:** *In silico* analysis showed that the mitochondrial genomes of *C. parapsilosis* (*sensu stricto*), *C. orthopsilosis* and *C. metapsilosis* are syntenic. The criterion for selecting the target gene was to find regions with different single-nucleotide polymorphism (SNP) profiles that were capable of discriminating the 3 species and that were flanked by conserved sequences in the same region of the
chromosome. On the basis of the alignments from the 6 mitochondrial whole-genome sequences, we selected NADH5 as the target for real-time PCR. The pairwise uncorrected distances between three species for complete sequence were: 5.10% for C. parapsilosis (sensu stricto) and C. metapsilosis; 4.85% for C. parapsilosis (sensu stricto) and C. orthopsilosis and 5.16% for C. metapsilosis and C. orthopsilosis. SNPs identified in silico were confirmed by sequencing 620bp of NADH5 of the 4 reference strains and 6 clinical isolates (LEMI6492, LEMI6814, LEMI7685, LEMI8521, LEMI7787, LEMI7518) (Figure 1). The partial sequence of C. parapsilosis (sensu stricto) and C. metapsilosis showed 100% sequence similarity with each of the control strains. Strain LEMI7518 identified as C. orthopsilosis has seven nucleotide differences to the reference strain, one inside the probe region (Figure 1).

**Molecular Identification by Real-time PCR:** Primers were designed to be specific for the C. parapsilosis species complex and to anneal to a highly conserved region of the NADH5. Three different specific probes (TaqMan-MGB) targeting a region of the NADH5 were designed for the 3 species within the C. parapsilosis complex. Initially, the ability of each probe to specifically identify its target was assessed using mtDNA as template from 4 reference strains and 5 clinical isolates (LEMI7294, LEMI6492, LEMI3523, LEMI3494, LEMI8521). Different annealing temperatures, probes and mtDNA concentrations were tested to ensure specificity and good reproducibility (data not shown) of the PCR assay, and the best performance was obtained using 51 °C, 250 nM and 40 ng/reaction, respectively. The assay was determined to be species specific because: (i) all strains tested
were correctly identified; (ii) non-specific amplification was not observed; and, (iii) no cross-species probe signal was observed.

To show that we obtain identical results using either purified mtDNA or total cellular DNA we compared the specificity and sensitivity of the method using total DNA from reference strains as template. We used the same annealing temperature and probe concentrations established for the mtDNA assays. Different genomic DNA concentrations were tested (data not shown), and a concentration of 160 ng/reaction was selected. Similar to the mtDNA analysis, the probes were very specific at the species level, and all strains tested were correctly identified.

After the standardization of the real-time PCR assay, we applied this method to analyze the genomic DNA of all the clinical isolates from Brazil included in this study (n= 66). By real-time PCR, we identified 37 as *C. parapsilosis* (*sensu stricto*), 24 as *C. orthopsilosis* and 5 as *C. metapsilosis*. The results generated by real-time PCR showed 100% concordance with the results obtained by *ITS* sequencing (Table 2). In addition, all isolates that were sent to our lab in a blind fashion by Dr. Quindós (n= 30) were correctly identified as the following: 21 *C. parapsilosis* (*sensu stricto*), 5 *C. orthopsilosis* and 4 *C. metapsilosis*. The Spanish strains were originally identified in Dr. Quindós’ lab by *ITS* sequencing (24).

**Discussion**

In the present study, we described a rapid and accurate method using mtDNA and TaqMan technology for identification of the 3 species within the *C. parapsilosis* complex. We decided to use mtDNA because it has been recognized
as a valuable target for evaluating close phylogenetic relationships because of its higher mutation rate compared with nuclear DNA (5, 28, 33). To explore the mtDNA potential for species identification, we first performed an in silico analysis to check for possible polymorphic regions capable of differentiating the 3 target species within the *C. parapsilosis* complex. After checking the whole mtDNA sequence of the 3 species, we found polymorphic regions within the *NADH5* gene, which were chosen as targets for real time PCR. The pairwise distances of *NADH5* homologs are well balanced between the three pair-groups and are higher than those found for *ITS* and *COX3* in the *C. parapsilosis* complex and for *COX2* in *C. glabrata* (28,30).

The probes were designed to include a combination of different SNPs able to differentiate each species within the *C. parapsilosis* complex (Figure 1). The large adenine and thymine content in mtDNA as a whole and, consequently, in the selected region impairs the design of probes with high annealing temperatures. This problem was solved using the TaqMan Minor Groove Binder (MGB), which enhances the melting temperature of probes.

Despite all progress obtaining specific probes based on mtDNA, we considered the protocol for its extraction to be time consuming and labor intensive (8). Therefore, we showed that during the genomic DNA extraction from yeast cells, some mtDNA is also obtained, making it possible to run the PCR assays on extracted total DNA and still correctly identify all strains (Table 2). Consequently, the new assay becomes more suitable to clinical laboratories.
In the present study, we tested 100 *C. parapsilosis* (*sensu lato*) strains and a control group of 6 other species (Table 2). Because all *C. parapsilosis* (*sensu lato*) were correctly identified, we have established that the primer and probe sets exhibited 100% specificity, with no background fluorescence within the species of the complex. No amplification was observed when DNA from other species were used as the template, including *L. elongisporus*, which is biochemically indistinguishable from the *C. parapsilosis* species complex (21). Finally, we observed 100% concordance with the ITS sequencing identification, including the thirty strains used for blind testing.

Recently, some authors proposed new identification assays to distinguish these closely related species. In 2011, Hays *et al.* described a real-time PCR assay using melting curve analysis on a portion of the *SADH* gene. This method provides a rapid identification; however, the assay may exhibit low reproducibility because it is possible to observe variations in the melting curve when different thermocyclers are used (17). Garcia-Effron *et al.* (2011) proposed a new real-time PCR assay that can differentiate these 3 species using molecular beacons technology and the *ITS* region as the target. This assay was highly specific, exhibited good reproducibility and showed that the *C. parapsilosis* species complex can be identified quickly, thus demonstrating its suitability for clinical applications (11). Santos *et al.* (2011) developed an assay using matrix-assisted laser desorption/ionization time-of-flight intact cell mass spectrometry (MALDI-TOF ICMS) for identification of some emerging pathogenic *Candida* species. These authors have shown that this method identifies the 3 species within the *C.*
parapsilosis complex, however, a quality-controlled database must be available to provide an accurate analysis (29). In all methods described above, only a limited number of C. metapsilosis (1, 6 and 2, respectively) were tested by the authors (11, 17, 29). In our analysis, we were able to validate our strategy by testing 50 strains of C. parapsilosis (sensu stricto), 40 strains of C. orthopsilosis and 10 strains of C. metapsilosis, including strains from 2 different countries.

In conclusion, we found an unexpected high variability in NADH5, a gene that encodes a respiratory chain protein and therefore, expected to be under strong negative selection. Nevertheless, we described a level of variability whose analysis reveals that it can be exploited for epidemiological purposes. Our data shows that the variability described in our study is sufficient to unequivocally distinguish the three species of the C. parapsilosis complex. We were able to validate a real-time PCR assay using mtDNA as template. This new identification method allows the C. parapsilosis species complex to be identified quickly and accurately and saves time when compared with conventional sequencing, which has been considered the gold standard for yeast identification. In addition, this assay provides a useful tool for epidemiological and clinical studies of these emergent species.

Acknowledgments

This work was supported in part by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil (grant 2007/08575-1), and the Conselho Nacional de Pesquisas Científicas e Tecnológicas (CNPq), Brazil (grant 308011/2010-4). A.C.R.S. received a master fellowship from FAPESP
(2010/02752-1). R.C.F. and F.C.B. received a postdoctoral fellowship from FAPESP (grants 2009/01230-4 and 2010/17179-5). S.S.G. received a postdoctoral fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil (PNPD 23038.007393/2011-11). A. L. C received grants from FAPESP and CNPq. M.R.S.B. received grants from FAPESP, CNPq, and the International Program of the Howard Hughes Medical Institute. We thank Dr. Quindós for sending strains for blind testing.

References


Figure 1: Local alignment of the \textit{NADH5} gene. Primer regions are shown in boxes (forward primer from 223 to 246 bp; reverse from 284 to 307 bp), and the probe region (from 251 to 276 bp) is in the dashed box. The SNP between the \textit{C. orthopsilosis} strains located in the probe region is at position 270 bp. \textit{Cp}: \textit{C. parapsilosis} (sensu stricto); \textit{Cm}: \textit{C. metapsilosis} and \textit{Co}: \textit{C. orthopsilosis}.
TABLE 1. Oligonucleotide primers and probes sequences used in this study.

<table>
<thead>
<tr>
<th>Oligonucleotide*</th>
<th>Sequence 5´ to 3´</th>
<th>Modification</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>5´ end</td>
<td>3´ end</td>
<td>5´ end</td>
<td>3´ end</td>
</tr>
<tr>
<td>ITS1**</td>
<td>TCCGTAGGTGAACCTGCGG</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>ITS4**</td>
<td>GCATACTCAAAGCAGGA</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>NADH5 F</td>
<td>GTGCTGGTTCTTATTCATA</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>NADH5 R</td>
<td>GCCAGAGTCTACCATACCTAC</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>RT - NADH5 F</td>
<td>CATATACTGAAAATAATACGATG</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>RT - NADH5 R</td>
<td>CATTATCTCTTAAATCCAATAA</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>Cp NADH5</td>
<td>TTCTATGATTATAGTTATCTAT</td>
<td>FAM</td>
<td>MGB</td>
</tr>
<tr>
<td>Co NADH5</td>
<td>TACCTAGGAATATTATAGTAGTATAT</td>
<td>VIC</td>
<td>MGB</td>
</tr>
<tr>
<td>Cm NADH5</td>
<td>TACCTAGGTATTATATATATATTAT</td>
<td>NED</td>
<td>MGB</td>
</tr>
</tbody>
</table>

*The letters F and R in the primer names describe the orientation of the primer 5´ to 3´. F: Forward (sense) or R: Reverse (antisense).  
** White et al. (17).

<table>
<thead>
<tr>
<th>Strains (nº tested)</th>
<th>Species Identified by Real-Time PCR assay (No. of isolates tested)</th>
<th>% agreement*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C. parapsilosis (sensu stricto)</td>
<td>C. orthopsilosis</td>
</tr>
<tr>
<td>Reference strains (4)</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>C. parapsilosis (sensu stricto) (2)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C. orthopsilosis (1)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C. metapsilosis (1)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Clinical isolates (66)</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>C. parapsilosis (sensu stricto) (27)</td>
<td>0</td>
<td>34</td>
</tr>
<tr>
<td>C. orthopsilosis (34)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C. metapsilosis (5)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Blinded strains (30)</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>C. parapsilosis (sensu stricto) (21)</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>C. orthopsilosis (5)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C. metapsilosis (4)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Control strains (5)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C. albicans (1)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C. tropicalis (1)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C. krusei (1)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C. glabrata (1)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C. lusitaniae (1)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>L. elongisporus (1)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Agreement analysis was performed between the Real-Time PCR assay described in the current study and the "gold standard" technique (ITS sequencing)

**N/A: not applicable