Species identification and strain differentiation of clinical Candida isolates using the DiversiLab system of automated repetitive sequence-based PCR

Mark G. Wise,1 Mimi Healy,1 Kristy Reece,1 Rebecca Smith,2 Dobbie Walton,1 Wendy Dutch,1 Alex Renwick,1 Joe Huong,1 Steve Young,2 Jeffrey Tarrand3 and Dimitrios P. Kontoyiannis4

Correspondence
Dimitrios P. Kontoyiannis
dkontoyi@mdanderson.org

1Bacterial Barcodes Inc., Athens, GA 30602, USA
2TriCore Reference Laboratories, Albuquerque, NM 87102, USA
3Department of Laboratory Medicine, University of Texas M. D. Anderson Cancer Center, Houston, TX 77030, USA
4Department of Infectious Diseases, Infection Control and Employee Health, University of Texas M. D. Anderson Cancer Center, Houston, TX 77030, USA

Received 30 November 2006
Accepted 26 February 2007

The DiversiLab system, which uses repetitive sequence-based PCR (rep-PCR) to genotype micro-organisms, was evaluated as a molecular typing tool for members of the genus Candida. Initially, 41 clinical Candida spp. (7 Candida krusei, 10 Candida parapsilosis, 7 Candida albicans, 10 Candida tropicalis and 7 Candida glabrata), previously identified at the species level by morphological and biochemical analysis, were analysed with the DiversiLab system. Species identification was confirmed by DNA sequence analysis of the contiguous internal transcribed spacer (ITS) region (ITS1–5.8S–ITS2). On the basis of an 80 % similarity threshold, rep-PCR consistently clustered like species and this set of isolates, along with five ATCC reference strains, was used to create a DNA fingerprint library with the DiversiLab software. Subsequently, an additional set of 115 clinical Candida isolates, identified biochemically as C. albicans (n=94), C. glabrata (n=8), C. parapsilosis (n=5), C. tropicalis (n=3), C. krusei (n=3) and Candida lusitaniae (n=2), isolated at a regional reference laboratory, were typed using DiversiLab. One hundred and six of these isolates clustered with members of the Candida library at >80 % similarity and thus could be assigned species identification, and initial calculations showed that identification via rep-PCR fingerprinting was 95 % concordant (101/106) with the biochemical/morphological identification. However, ITS region sequencing of the five discrepant samples, as well as the nine isolates that were <80 % similar to the database samples, showed that nine were misidentified with traditional biochemical/morphological methods. For the misidentified isolates, the sequence-based identification was in agreement with the DiversiLab clustering, yielding an actual correlation of >99 %. As traditional techniques can take several days to provide information about Candida at the genus/species level, genotyping with the DiversiLab system holds promise for more-rapid speciation of members of this genus. This system may also be useful for epidemiological studies such as source tracking that require Candida subspecies discrimination.

INTRODUCTION

Candidiasis and candidaemia have emerged as common infections in debilitated hospitalized patients (Kauffman, 2006). Although Candida albicans remains the most important species causing these infections, non-albicans Candida spp. such as Candida glabrata have emerged as common pathogens, often accounting for more than 50 % of candidiasis cases (Clark & Hajjeh, 2002; Eggimann et al., 2003; Marchetti et al., 2004; Tortorano et al., 2004). Importantly, non-albicans Candida spp. frequently show decreased sensitivity to azoles compared with C. albicans (Imhof et al., 2004; Li et al., 2004; Pfaffer et al., 2002;
Vermitsky & Edlind, 2004) and are associated with high failure rates of azole-based treatment (Tortorano et al., 2004; Viscoli et al., 1999). Traditional automated and biochemical systems of identifying Candida organisms identify them only at the genus/species level (Warren & Hazen, 1999) and the procedures may take several days. Furthermore, some biochemical tests have been reported to misidentify clinical Candida isolates (Dooley et al., 1994). Thus, there is a need for the development of rapid, reliable, diagnostic methods to support the selection of antifungal therapy. Accordingly, several studies reporting the development of rapid-detection methodologies based on real-time PCR (Chau et al., 2004; Hsu et al., 2003; Selvarangan et al., 2003), multiplex PCR (Li et al., 2003; Luo & Mitchell, 2002), fluorescent in situ hybridization with peptide nucleic acid probes (Rigby et al., 2002) and flow cytometry (Page & Kurtzman, 2005) have been reported recently for identification of the clinically important Candida spp.

Similarly, there is an increasing realization that robust molecular subtyping methods for Candida are needed for epidemiological investigations. For example, although the source of candidiasis is endogenous in most patients, nosocomial transmission of Candida spp. has been recognized in intensive care units (Richards et al., 1999). A variety of molecular methods for DNA-based subtyping of Candida spp. have been employed (Lockhart et al., 2005), as no gold standard exists for the genus (Reiss et al., 1998). Current molecular strain typing methods such as PFGE and RFLP are tedious (Fujita & Hashimoto, 2000), whilst others, such as randomly amplified polymorphic DNA, are prone to artefacts (Bart-Delabesse et al., 2001). Additional molecular subtyping methods for Candida spp. include those based on polymorphism in microsatellite markers and loci (Botterel et al., 2001; Dalle et al., 2003; Foulet et al., 2005; Lott et al., 2003; Sampaio et al., 2003; Stephan et al., 2002). Multilocus sequence typing is a highly discriminatory and stable method of unambiguous characterization of C. albicans (Bougnox et al., 2002; Dodgson et al., 2003; Robles et al., 2004; Tavanti et al., 2003). However, the costs associated with this method of sequencing can be prohibitive and, along with the intensive data analysis requirements, may keep such technologies out of reach for many clinical laboratories (Goulding et al., 2000; Olive & Bean, 1999). Overall, few methods alone have the potential for providing both identification and strain-level differentiation in Candida in a time- and cost-effective manner.

Previous reports have described the application of repetitive extragenic palindromic sequence-based PCR (rep-PCR) for characterizing the genotypic relatedness among Candida rugosa isolates (Redkar et al., 1996). The rep-PCR method uses primers that target non-coding repetitive sequences interspersed throughout the bacterial and fungal genome (Koeuth et al., 1995; Stern et al., 1984; Versalovic et al., 1991) and is an established approach for classification of subspecies and strain delineation of bacteria (Versalovic & Luspinki, 2002; Versalovic et al., 1993). Recent studies have demonstrated the potential of a commercially available automated rep-PCR assay system, the DiversiLab system, for rapid identification and strain-level differentiation of Aspergillus (Healy et al., 2004), Fusarium (Healy et al., 2005b), dimorphic fungi (Pounder et al., 2006), Candida (Chau et al., 2004), Zygomycetes (Kontoyiannis et al., 2005) and dermatophytes such as Trichophyton (Pounder et al., 2005). This study explores the use of the DiversiLab system as a genotyping tool for discriminating clinical Candida isolates. Herein, we present a comparison of rep-PCR fingerprinting with sequence-based analysis of the internal transcribed spacer (ITS) region (ITS1–5.8S–ITS2) and assess the utility of the DiversiLab software to create a database of known clinical Candida isolates and assign a species designation to a set of 115 putative Candida isolates.

**METHODS**

Identification of Candida isolates by biochemical and culture characteristics. To construct the initial Candida library, 41 clinical isolates from 5 of the most common clinically encountered Candida species, C. albicans (n=7), C. parapsilosis (n=10), C. glabrata (n=7), C. krusei (n=7) and C. tropicalis (n=10), were characterized in the Clinical Microbiology Laboratory at the M. D. Anderson Cancer Center. Also included were five culture collection reference strains, C. albicans ATCC 10231, C. glabrata ATCC 15126, C. krusei ATCC 200339, C. parapsilosis ATCC 10232 and C. tropicalis ATCC 10610. The test sample isolates, identified initially as C. albicans (n=94), C. glabrata (n=8), C. parapsilosis (n=5), C. tropicalis (n=3), C. krusei (n=3) and C. lusitaniae (n=2), were obtained from TriCore Reference Laboratories. The traditional Candida identification scheme was as follows. Briefly, after positive blood culture, samples were characterized by microscopy. Putative yeast cell-containing samples were tested for germ-tube formation, where a positive test resulted in a C. albicans identification (unless colony morphology was atypical). Germ-tube-negative cultures were resolved using the Vitek yeast biochemical card (YBC; bioMérieux) and the isolate was considered to be identified if the YBC confidence was >85 % and colony morphology was typical. If a yeast showed atypical morphology or the YBC card confidence was low, an API 20C yeast identification system (bioMérieux) and morphological characteristics on cornmeal agar were used to obtain yeast identification. Overall, approximately 2 days and 60 min of technical hands-on time was necessary for completion of this identification procedure.

Culture and DNA extraction. Candida clinical isolates were cultured on trypticase soy agar II with 5 % sheep blood for 24 h at 37 °C. DNA from each isolate was extracted using the UltraClean microbial DNA isolation kit (Mo Bio Laboratories) following the manufacturer’s instructions. DNA was adjusted to approximately 25 ng μl⁻¹ for each sample and the presence of high-molecular-mass DNA was confirmed by agarose gel electrophoresis.

**rep-PCR DNA fingerprinting.** All DNA samples were amplified using the DiversiLab Yeast kit for DNA fingerprinting (Bacterial Barcodes) following the manufacturer’s instructions. Briefly, 50 ng genomic DNA, 2.5 U AmpliTaq polymerase, 2.5 μl 10 X PCR buffer (Applied Biosystems) and 2 μl primer mix were added to the rep-PCR master mix in a total volume of 25 μl per reaction. Thermal cycling parameters were as follows: initial denaturation of 94 °C for 2 min,
Fig. 1. rep-PCR-based dendrogram and virtual gel image fingerprints obtained from 41 clinical *Candida* isolates and 5 ATCC reference strains using the DiversiLab system with the yeast fingerprinting kit. Pearson’s correlation coefficient was used to create a pairwise percentage similarity matrix and the tree was generated using UPGMA. An 80% similarity threshold (vertical line) was chosen for *Candida* species identification. Bar, percentage similarity among strains.
followed by 35 cycles of denaturation at 92 °C for 30 s, annealing at
50 °C for 30 s and extension at 70 °C for 90 s, with a final extension at
70 °C for 3 min. Detection and analysis of rep-PCR products were
implemented using the DiversiLab system (Bacterial Barcodes) and
analysis was performed with the DiversiLab software version 3.1. The
resulting DNA fingerprint patterns were viewed as electropherograms
and the results included a dendrogram with a similarity matrix and a
virtual gel image of the fingerprint for each DNA sample.

**DNA sequencing and analysis.** Species identification of the
*Candida* isolates was confirmed by sequencing the ITS1–5.8S–ITS2
region as described by White et al. (1990). Briefly, the ITS region was
amplified using the ITS1 forward primer 5’-TCCGTAGGTGAAAC-
CTGCGG-3’ and the ITS4 reverse primer 5’-TCCTCCGCTTATGG-
ATATGC-3’. Amplicons were purified using the High Pure PCR
product purification kit (Roche Diagnostics). Bidirectional sequence
data were generated using the BigDye terminator v3.1 cycle
sequencing kit (Applied Biosystems). The products were purified
using Performa DTR gel filtration cartridges (Edge BioSystems) or
ethanol precipitation, and sequenced using an ABI 377 or ABI 3100
sequencer (Applied Biosystems). Contiguous sequences were
assembled and edited using the SEQMAN software program, which is
part of the Lasergene suite of sequence analysis software (DNASTAR).
GenBank/EMBL/DDBJ database searches were conducted for the full-
length sequences using BLAST (Altschul et al., 1990). Sequences were
aligned using the CLUSTALW function of the MEGALIGN program
(DNASTAR). To generate the phylogenetic tree, MEGALIGN uses the
Kimura distance method to create a pairwise genetic distance matrix.
The dendrogram was constructed using the neighbour-joining
method (Saitou & Nei, 1987).

**Fig. 2.** Phylogenetic tree based on the DNA sequence of the ITS1–5.8S–ITS2 region illustrating the relationships among the
41 clinical *Candida* isolates and 5 ATCC reference strains. Sequences were aligned using CLUSTALW and the tree was
constructed from a matrix of pairwise genetic distances using the neighbour-joining algorithm (Saitou & Nei, 1987).
RESULTS AND DISCUSSION

Discrimination of Candida isolates using rep-PCR-based DNA fingerprints and library construction

We obtained rep-PCR-based fingerprint patterns for all of the Candida isolates using the DiversiLab system and yeast DNA fingerprinting kit. For analysis, DiversiLab software used the Pearson correlation coefficient and UPGMA to compare automatically the rep-PCR profiles and create corresponding dendrograms (Healy et al., 2005a). In addition, results included computer-generated virtual gel images and selected demographic fields to aid interpretation of the data. Guidelines have been suggested by the manufacturer and there are several publications for identification of an isolate to the species level (Healy et al., 2004; Pounder et al., 2005). Although there is no universal threshold for species identification via rep-PCR profiles, moulds and fungi can often be identified at the species level if the fingerprints are >80% similar to reference library samples. The 41 clinical Candida isolates from the M. D. Anderson Cancer Center were used to generate an initial Candida library, and the analysis showed that the isolates clustered by species at the 80% threshold (Fig. 1). Six of the seven C. glabrata isolates grouped tightly (approx. 95% similarity) with the C. glabrata reference strain ATCC 15126. The remaining clinical C. glabrata isolate (CL09) was on average approximately 80% similar to members of the main C. glabrata cluster. C. glabrata CL09 appeared to lack two major rep-PCR amplicons present in the other C. glabrata representatives isolates. The ten clinical C. tropicalis strains were part of a single clade at approximately 82% similarity, which included reference strain ATCC 10610. All of the clinical C. krusei isolates clustered at >90% similarity, but C. krusei ATCC 200339 was an outlier with a fingerprint that was only 68% similar to its nearest neighbour in the dataset, C. glabrata CL10. The analysis split the C. albicans strains into two clusters, with C. albicans ATCC 10231 grouping with isolates CL02 and CL06, and the second cluster comprising the five remaining clinical isolates. Finally, the C. parapsilosis strains separated into three clades at the 80% similarity cut-off; C. parapsilosis ATCC 10232 was 95.5% similar to C. parapsilosis CL24.

Sequence analysis of the ITS1–5.8S–ITS2 region

The ITS regions are located between the 18S and 28S rRNA genes, separated by the 5.8S RNA gene. DNA sequence variation in the ITS regions has led to their use in phylogenetic studies of many fungi (Guarro et al., 1999; Turenne et al., 1999). In our study, sequencing of the ITS1–5.8S–ITS2 region from each of the Candida isolates allowed confirmatory identification of the Candida strains via BLAST searches of the GenBank/EMBL/DDJB databases. Fig. 2 illustrates a phylogenetic tree based on the aligned ITS1–5.8S–ITS2 region for the 41 Candida isolates and the 5 ATCC strains. In general, the sequence-based tree bore an overall topology similar to the rep-PCR dendrogram, with the isolates clustered in species-specific groups; however, rep-PCR subdivided the clinical C. albicans and C. parapsilosis strains into separate clades at the 80% similarity threshold. The increased discriminatory ability of rep-PCR is not surprising considering the basis of the two methods. The sequence-based tree was derived solely from variations in the ITS and 5.8S rRNA regions, whereas the rep-PCR DNA profiles represent molecular patterns surveying the entire genome of a given isolate, providing an overview of genomic variation among strains. The rRNA genes can be considered to be molecular chronometers such that sequence variation can be used to infer phylogenetic relationships among organisms. The dendrogram that is created from rep-PCR fingerprints is based on a pattern-matching algorithm that does not necessarily reflect phylogenetic history. This most likely accounts for the fact that the sequence of the C. krusei reference strain ATCC 200339, although only about 72% identical to its closest match, C. krusei CL17, grouped with the other C. krusei isolates in the sequence-based tree, but was an outlier with the rep-PCR analysis, as this strain failed to cluster with the clinical C. krusei isolates.

Identification of Candida isolates by comparison with the Candida library

One hundred and fifteen clinical Candida isolates from a large regional reference laboratory were typed with the DiversiLab system and queried against the Candida library as described above. The isolates were identified previously at the species level by conventional biochemical/morphological tests. Fig. 3 illustrates the results obtained with 94 isolates identified as C. albicans. Eighty-one isolates clustered at >80% with C. albicans library entries. ITS sequencing of the remaining samples was conducted to determine whether the isolates might have been misidentified by the biochemical/morphological methods. One presumptive C. albicans, TCLC-089, grouped within the C. parapsilosis cluster and sequencing analysis via BLAST searches confirmed that the isolate was indeed C. albicans.
Fig. 4. rep-PCR-based dendrogram showing DiversiLab software analysis of eight *C. glabrata*, five *C. parapsilosis*, three *C. tropicalis*, three *C. krusei* and two *C. lusitaniae* isolates (circled) queried against the clinical *Candida* library. For the misclassified sample TCLC-114 (asterisk), ITS sequencing-based identification revealed that the isolate was *S. cerevisiae*. Pearson’s correlation coefficient was used to create a pairwise percentage similarity matrix and the tree was created using UPGMA. An 80% similarity threshold (vertical line) was chosen for *Candida* species identification. Bar, percentage similarity among strains.
parapsilosis. Similarly, isolate TCLC-001 grouped tightly with the C. tropicalis samples in the database and sequencing revealed that the rep-PCR analysis was accurate and that the isolate was indeed C. tropicalis. Also, two putative C. albicans isolates (TCLC-085 and TCLC-086) that clustered with the C. glabrata library entries were confirmed as C. glabrata by sequencing. One sample confirmed by sequencing as C. albicans (TCLC-093) was 81 % similar to C. krusei ATCC 200339; thus, by the 80 % similarity threshold rule-of-thumb, this sample was misclassified by the DiversiLab software.

The remaining eight isolates identified as C. albicans did not group with any library entries at the 80 % cut-off level. Four of these were confirmed as C. albicans by sequence analysis. Therefore, it appears that the Candida library was not of a sufficient scope to be useful for characterizing these particular strains. This is not surprising, as the library was constructed with isolates obtained from a limited patient population (cancer patients) at a single medical facility. The putative C. albicans TCLC-091 isolate was identified as C. parapsilosis by sequencing. Presumptive C. albicans TCLC-092 had a distinctive fingerprint and did not cluster closely with any library entries; sequencing revealed that it was actually Saccharomyces cerevisiae. Finally, the two isolates TCLC-094 and TCLC-095 with unusual rep-PCR profiles were identified as Candida dubliniensis by ITS sequencing and BLAST searches.

The fingerprints of Candida species identified biochemically as non-albicans and a comparison with the Candida library is illustrated in Fig. 4. All query isolates clustered with the appropriate library samples with the exception of one strain (TCLC-114). This isolate, identified as C. glabrata by biochemical methods, was 100 % similar by sequencing to uncultured Saccharomyces isolate A9 (GenBank accession no. AY349457). Table 1 summarizes the species-level identification results among all discrepant samples.

## Conclusions

A simple, automated system of both species-level identification and strain-level differentiation of Candida organisms could be valuable both for accurate treatment decisions and for infection control interventions for these fungal infections. In this study, rep-PCR was able to discriminate between the medically important Candida spp., including C. albicans, C. parapsilosis, C. glabrata, C. krusei and C. tropicalis. The DiversiLab system was also successful at species-level identification of isolates by comparison with a reference library and in fact recognized several strains that were misidentified at the species level by the conventional biochemical/morphological techniques. The availability of large, well-characterized sample sets from dispersed geographical locations to build a comprehensive library of Candida spp. and other yeasts would enhance the ability to classify the samples that were not identified by rep-PCR in this study. In addition, this work suggests that rep-PCR is able to subtype C. albicans and C. parapsilosis species, indicating that this method may be particularly important if specific genotypes are shown to be associated with either increased virulence or unique antifungal-resistance profiles (Foongladda et al., 2004). The DiversiLab system may also be an attractive tool for clinical laboratories where minimizing both the turnaround time and technical hands-on time is paramount.

The most labour-intensive portion of the rep-PCR assay is the DNA extraction, and we have investigated replacing the manual DNA purification procedures with automated systems (data not shown). In particular, for non-albicans

### Table 1. Comparison of identification results for discrepant Candida samples

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Biochemical/morphological identification</th>
<th>rep-PCR-based identification</th>
<th>ITS sequence-based identification via BLAST*</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCLC-001</td>
<td>C. albicans</td>
<td>C. tropicalis</td>
<td>C. tropicalis (DQ680841) (e=1e–137)</td>
</tr>
<tr>
<td>TCLC-083</td>
<td>C. albicans</td>
<td>No identification</td>
<td>C. albicans (AF455524) (e=0)</td>
</tr>
<tr>
<td>TCLC-084</td>
<td>C. albicans</td>
<td>No identification</td>
<td>C. albicans (AF455531) (e=0)</td>
</tr>
<tr>
<td>TCLC-085</td>
<td>C. albicans</td>
<td>C. glabrata</td>
<td>C. glabrata (AY399793) (e=0)</td>
</tr>
<tr>
<td>TCLC-086</td>
<td>C. albicans</td>
<td>C. glabrata</td>
<td>C. glabrata (AY399793) (e=0)</td>
</tr>
<tr>
<td>TCLC-089</td>
<td>C. albicans</td>
<td>C. parapsilosis</td>
<td>C. parapsilosis (DQ683009) (e=0)</td>
</tr>
<tr>
<td>TCLC-090</td>
<td>C. albicans</td>
<td>No identification</td>
<td>C. albicans (AF455524) (e=0)</td>
</tr>
<tr>
<td>TCLC-091</td>
<td>C. albicans</td>
<td>No identification</td>
<td>C. albicans (AY391843) (e=0)</td>
</tr>
<tr>
<td>TCLC-092</td>
<td>C. albicans</td>
<td>No identification†</td>
<td>S. cerevisiae (AB212260) (e=0)</td>
</tr>
<tr>
<td>TCLC-093</td>
<td>C. albicans</td>
<td>C. krusei</td>
<td>C. albicans (AJ551313) (e=0)</td>
</tr>
<tr>
<td>TCLC-094</td>
<td>C. albicans</td>
<td>No identification†</td>
<td>C. dubliniensis (AJ865083) (e=0)</td>
</tr>
<tr>
<td>TCLC-095</td>
<td>C. albicans</td>
<td>No identification†</td>
<td>C. dubliniensis (AJ865083) (e=0)</td>
</tr>
<tr>
<td>TCLC-096</td>
<td>C. albicans</td>
<td>No identification</td>
<td>C. albicans (AF455524) (e=0)</td>
</tr>
<tr>
<td>TCLC-114</td>
<td>C. glabrata</td>
<td>No identification</td>
<td>Uncultured saccharomyces (AY349457) (e=0)</td>
</tr>
</tbody>
</table>

*Highest scoring sequence, GenBank/EMBL/DDBJ accession no. and e value.
†Sample was positively identified by the commercially available DiversiLab yeast library.
Candida spp., which require multiple procedures to be performed to obtain species identification, the automation of rep-PCR would allow rapid genotyping, which could lead to the identification and subspecies differentiation of Candida organisms in hours instead of days. With regard to strain typing, further studies are needed to compare the various methods of molecular typing of sample sets that include demographic data to assess their usefulness for epidemiological studies.

ACKNOWLEDGEMENTS

This work was presented in part at the 44th General Meeting of the Interscience Conference on Antimicrobial Agents and Chemotherapy, Washington, DC, USA, 2004.

REFERENCES


