Phylogenetic analysis and rapid identification of *Candida dubliniensis* based on analysis of *ACT1* intron and exon sequences

Samantha M. Donnelly, Derek J. Sullivan, Diarmuid B. Shanley and David C. Coleman

The phylogenetic position of *Candida dubliniensis* has previously been established on the basis of the sequence of rRNA genes. In order to confirm the relationship between *C. dubliniensis* and other yeast species, particularly *Candida albicans*, using non-rRNA gene sequences the *ACT1* gene was chosen for analysis. Three overlapping fragments that together span the entire *C. dubliniensis* *ACT1* gene (CdACT1) were amplified from a recombinant phage library using PCR. These were cloned and used to determine the contiguous sequence of the gene. Analysis of the sequence data revealed the presence of a 1131 bp ORF interrupted by a single 632 bp intron at the 5’ extremity of the gene. Comparison of the CdACT1 sequence with the *C. albicans* homologue (CaACT1) revealed that although the exons are 97-98% identical the introns are only 83-4% identical. Phylogenetic trees generated using *ACT1* exon and intron sequences from a range of yeast species unequivocally confirmed the phylogenetic position of *C. dubliniensis* as a unique taxon within the genus *Candida*. Analysis of the *ACT1*-associated intron sequences from 10 epidemiologically unrelated *C. dubliniensis* isolates from disparate geographical locations showed a very low level of intraspecies sequence variation. In order to develop an accurate and rapid method to identify *C. dubliniensis* from primary isolation plates the significant divergence between the *C. dubliniensis* and *C. albicans* *ACT1* intron sequences was exploited by designing *C. dubliniensis*-specific PCR primers. Using a rapid boiling method to produce template DNA directly from colonies from primary isolation plates in 10 min, these primers were used in a blind test with 122 isolates of *C. dubliniensis*, 53 isolates of *C. albicans*, 10 isolates of *C. stellatoidea* and representative isolates of other clinically relevant *Candida* and other yeast species. Only the *C. dubliniensis* isolates yielded the *C. dubliniensis*-specific 288 bp amplimer. Use of this technique on colonies suspected to be *C. dubliniensis* allows their correct identification as *C. dubliniensis* in as little as 4 h.

**Keywords:** *ACT1*, phylogenetics, *Candida dubliniensis*, PCR identification

**INTRODUCTION**

*Candida dubliniensis* is a yeast species first described in 1995 (Sullivan *et al.*, 1995). Although *C. dubliniensis* is phenotypically similar to *C. albicans*, the two species differ significantly at the genetic level. In particular, phylogenetic analysis of large- and small-subunit rRNA gene sequences provided the basis for the designation of *C. dubliniensis* as a separate species (Sullivan *et al.*, 1995, 1997; Gilfillan *et al.*, 1998). The first isolates identified as *C. dubliniensis* were recovered from the oral cavities of Irish human immunodeficiency virus (HIV)-infected individuals. However, over the last three
years there have been increasing numbers of reports of the recovery of *C. dubliniensis* isolates by laboratories throughout the world, including Europe, North and South America, and Australia (Coleman et al., 1997b; Sullivan et al., 1997; Sullivan & Coleman, 1998; Salkin et al., 1998; Kirkpatrick et al., 1998). Although the majority of these isolates have been primarily associated with oral candidosis in HIV-infected and AIDS patients (Coleman et al., 1997a), *C. dubliniensis* isolates have also been recovered from the oral cavities, gastrointestinal tracts and vaginas of HIV-negative individuals (Moran et al., 1997, 1998; Sullivan & Coleman, 1998; Odds et al., 1998). There have also been reports of *C. dubliniensis* isolates associated with systemic disease (Pinjon et al., 1998; Meis et al., 1999). The majority of clinical isolates of *C. dubliniensis* have been shown to be susceptible to commonly used antifungal drugs, including fluconazole (Moran et al., 1997, 1998; Kirkpatrick et al., 1998; Pfaller et al., 1999). However, fluconazole resistance has been detected in clinical isolates (Moran et al., 1997, 1998; Kirkpatrick et al., 1998; Pfaller et al., 1999) and isolates of *C. dubliniensis* susceptible to fluconazole can be readily induced to produce fluconazole-resistant derivatives following exposure to the drug in vitro (Moran et al., 1997, 1998).

The prevalence of *C. dubliniensis* in the oral cavities of HIV-infected individuals and AIDS patients and reports of its association with disease in other body sites warrant in-depth epidemiological analysis. However, these investigations have been hampered by the lack of a simple, reliable method capable of unequivocally differentiating between *C. dubliniensis* and *C. albicans* in the clinical laboratory. Indeed, since *C. dubliniensis* and *C. albicans* share the ability to produce germ tubes and chlamydo-spores, features previously used for the definitive identification of *C. albicans*, it is likely that many isolates of *C. dubliniensis* have been misidentified as *C. albicans*. Investigations of our own collection of stored oral *Candida* isolates, originally identified as *C. albicans*, have shown that 1-8% of isolates recovered from asymptomatic normal healthy individuals and 16.5% of isolates recovered from HIV-infected individuals were in fact *C. dubliniensis* (Coleman et al., 1997a). In a similar study, Odds et al. (1998) have recently shown that approximately 2% of a stored archival culture collection of 2500 yeast isolates, originally identified as *C. albicans*, was *C. dubliniensis*. They found that the prevalence of *C. dubliniensis* was significantly higher among HIV-infected individuals than among HIV-negative individuals (Odds et al., 1998). Although first described in 1995, the earliest known *C. dubliniensis* isolates were recovered in the 1950s, thus predating the HIV pandemic. One of these strains, NCPF 3108, was recovered in the UK in 1957 and was originally deposited in the British National Collection for Pathogenic Fungi as *C. stellatoidea* (Sullivan et al., 1995), while another strain, CBS 2747, which was recovered in the Netherlands in 1952, was originally deposited in the Centraal Bureau fur Schimmelcultures as *C. albicans* (Meis et al., 1999).

A variety of tests have been developed to discriminate between *C. dubliniensis* and *C. albicans* based upon phenotypic characteristics. These include carbohydrate assimilation profiles and colonial coloration on differential media such as CHROMagar Candida and methyl blue-Sabouraud agar (Sullivan et al., 1995, 1997; Coleman et al., 1997a; Schoofs et al., 1997). However, some of these assays have been shown to be unreliable in some instances and should only be used for the presumptive identification of *C. dubliniensis* from clinical specimens (Schoofs et al., 1997; Sullivan & Coleman, 1998; Kirkpatrick et al., 1998). The accuracy of *C. dubliniensis* isolate identification based on carbohydrate assimilation profiles has been improved by the recent inclusion of the assimilation profiles of some *C. dubliniensis* strains in the databases of commercially available yeast identification systems, including the API ID 32C and 20C AUX systems. It has been reported recently that *C. dubliniensis* and *C. albicans* can be distinguished on the basis of differential growth at 45 °C, with isolates of the former species unable to grow at this temperature (Pinjon et al., 1998). However, in a recent study a significant number of *C. albicans* isolates were found to be unable to grow at this temperature (Kirkpatrick et al., 1998). Currently, the most reliable tests available to differentiate between these species are based on molecular techniques such as DNA fingerprinting with repetitive-sequence-containing probes, randomly amplified polymorphic DNA (RAPD) analysis and pulsed-field gel electrophoresis (Sullivan et al., 1995), but these are not suitable for the analysis of large sample numbers in routine diagnostic laboratories. However, since the differences between *C. dubliniensis* and *C. albicans* are most pronounced at the genetic level such differences should provide the basis for a specific and rapid identification test. One molecular technique with the required degree of specificity and ease of use is the polymerase chain reaction (PCR). This technology is increasingly available in diagnostic laboratories and due to its speed, reproducibility and high sample volume throughput is ideally suited for application to large numbers of clinical isolates.

The phylogenetic position of *C. dubliniensis* in relation to other yeast species has been established on the basis of the comparison of small- and large-subunit rRNA gene sequences (Sullivan et al., 1995, 1997; Gilfillan et al., 1998). In the present study we sought to confirm these phylogenetic relationships using sequences of non-rRNA gene origin. It was also hoped that these sequence data would lead to the identification of *C. dubliniensis*-specific nucleotide sequences that could be exploited in the design of a rapid PCR-based identification test. To achieve these goals the ACT1 gene of *C. dubliniensis* was chosen for analysis. ACT1 encodes actin, a protein that is abundant in all eukaryotic cells, where it is the major component of cytoplasmic microfilaments. Due to structural constraints the amino acid sequence of actin proteins from different eukaryotic species is highly conserved (Korn et al., 1978; Hightower et al., 1986; Pollard et al., 1990; Hennessey et al., 1993; Welch et al., 1993).
C. dubliniensis ACT1 gene

1994). Since C. albicans and C. dubliniensis are very closely related it was anticipated that the ACT1 genes of these species would be very similar. Results presented in this study for C. dubliniensis and in a previous study for C. albicans (Losberger & Ernst, 1989) showed that both ACT1 genes contain a single class IV intron and it was anticipated that these intron sequences would be subject to less evolutionary conservation than the actin-protein-coding exons. Therefore we decided to investigate whether the exons and introns of C. albicans and C. dubliniensis would be sufficiently divergent to allow an accurate determination of the phylogenetic relationship between the two species and to allow the design of C. dubliniensis-specific primers suitable for rapid and specific identification of this species in the clinical laboratory using a rapid template DNA preparation procedure.

METHODS

Candida strains and culture media. All C. dubliniensis strains were isolated by this laboratory or received from other laboratories and identified using the molecular and phenotypic methods described by Sullivan et al. (1995). All Candida strains and isolates were routinely grown on Potato Dextrose Agar (PDA, Oxoid) at pH 5.6 for 48 h at 37°C. For liquid culture, Candida strains and isolates were grown at 37°C in Yeast Peptone Dextrose Broth (YPD) in an orbital incubator (Gallenkamp) set at 150 r.p.m.

Bacterial strains and culture media. Escherichia coli DH5α was used as the host strain for phagemid pBluescript II KS( +/−) (Stratagene) and was maintained on Luria–Bertani (LB) agar, supplemented with 100 μg ampicillin ml⁻¹ to maintain plasmids where appropriate. For liquid culture, strains harbouring plasmids were grown at 37 °C in LB broth containing 100 μg ampicillin ml⁻¹ in an orbital incubator set at 150 r.p.m. Transformation of E. coli DH5α and identification of transformants containing recombinant plasmids were carried out by standard protocols (Sambrook et al., 1989). E. coli LE 392 and its P2 phage lysogenic derivative (P2 392) were used for propagating the bacteriophage λ cloning vector EMBL3 and its recombinant derivatives. These strains were grown and maintained on LB agar supplemented with 10 mM MgSO₄ and 0.2% (w/v) maltose. Organisms for phage infection were grown in LB broth containing 10 mM MgSO₄ and 0.2% (w/v) maltose (Sambrook et al., 1989).

Chemicals, enzymes, radioisotopes and oligonucleotides. Analytical-grade or molecular-biology-grade chemicals were purchased from Sigma-Aldrich, BDH or Boehringer Mannheim. Enzymes were purchased from Promega or Boehringer Mannheim and used according to the manufacturer’s instructions. [α-³²P]dATP (3000 Ci mmol⁻¹; 110 TBq mmol⁻¹) was purchased from Amersham. Custom-synthesized oligonucleotides were purchased from Genosys Biotechnologies (Europe).

DNA extraction procedures. Plasmid DNA for restriction endonuclease digestion and Southern hybridization was prepared by the alkaline lysis method described by Sambrook et al. (1989). Plasmid DNA for sequencing was prepared using a Quantum Prep Plasmid Miniprep kit (Bio-Rad). Total cellular DNA from Candida isolates was prepared as described by Gallagher et al. (1992). High-molecular-mass total cellular DNA from C. dubliniensis for the construction of a genomic library was isolated by the method described by Bennett et al. (1998). Candida template DNA for use in PCR experiments with the C. dubliniensis-specific oligonucleotide primer pair DUBF/DUBR (Table 1) was prepared as follows. A single colony from a culture grown for 48 h at 37°C on PDA or CHROMagar Candida medium (CHROMagar Candida, Paris, France) was suspended in 50 μl sterile distilled water. Cell suspensions were boiled for 10 min and the lysed cells subjected to a clearing spin for 5 min at 20000 g. Template DNA contained in 25 μl supernatant was used for PCR amplification.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Nucleotide coordinates</th>
<th>RE site†</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. albicans</td>
<td>5’- CGGAATTCCTTTAGAACCATTATCTCCGAT-3’</td>
<td>-49 to −30</td>
<td>EcoRI</td>
</tr>
<tr>
<td></td>
<td>5’- GGTCTAGAAATAATTGTCGACA-3’</td>
<td>126 to 145</td>
<td>XbaI</td>
</tr>
<tr>
<td>ACTF</td>
<td>5’- CGGAATTCATGGACGGTTGATGTT-3’</td>
<td>-1 to 17</td>
<td>EcoRI</td>
</tr>
<tr>
<td>ACTR</td>
<td>5’- CGGAATTCATGGACTAGAGATCTCCGAC-3’</td>
<td>1746 to 1767</td>
<td>EcoRI</td>
</tr>
<tr>
<td>3’F</td>
<td>5’- CGTAATCTAGATATTAGCTCCACCCAC-3’</td>
<td>1641 to 1660</td>
<td>EcoRI</td>
</tr>
<tr>
<td>3’R</td>
<td>5’- GGTCTAGAAATATTCTCCGAC-3’</td>
<td>1792 to 1811</td>
<td>XbaI</td>
</tr>
<tr>
<td>INF</td>
<td>5’- CGGAATTCATGACGCTTCGCACCCG-3’</td>
<td>-1 to 17</td>
<td>EcoRI</td>
</tr>
<tr>
<td>INTR</td>
<td>5’- CGGAATTCATGACGCTTCGCACCCG-3’</td>
<td>724 to 739</td>
<td>EcoRI</td>
</tr>
<tr>
<td>C. dubliniensis</td>
<td>DUBF</td>
<td>5’- GTATTTCTGGAGGCTTCCGAT-3’</td>
<td>251 to 270</td>
</tr>
<tr>
<td></td>
<td>DUBR</td>
<td>5’- GTGATTCTGGACACTAAGTC-3’</td>
<td>519 to 538</td>
</tr>
</tbody>
</table>

* Primers were complementary to ACT1 gene sequences as follows: CaACT1, accession no. X16377 (Losberger & Ernst, 1989); and CdACT1, accession no. AJ236897 (this study). Nucleotide coordinates shown are numbered in the 5’ to 3’ direction with the first base of the translation start codon being +1.

†Restriction endonuclease recognition sequence included within the primer sequence (underlined).
Large-scale E. coli phage lysates were prepared according to the plate method of Sambrook et al. (1989) and recombinant phage DNA was purified from phage preparations using a Wizard Lambda Prep kit (Promega).

**Construction of a C. dubliniensis CD36 genomic DNA library.** The DNA library was constructed by ligating Sau3A-generated partial digest products of C. dubliniensis CD36 chromosomal DNA >10 kb in size with BamHI-generated pre-prepared λ bacteriophage replacement vector EMBL3 arms (Promega) followed by packaging in vitro into pre-prepared phage heads and tails (Promega) according to the manufacturer's instructions. Previous studies have shown that DNA fragments ranging in size from 9 to 23 kb can be cloned into the EMBL3 vector (Frischauf et al., 1983). The packaged recombinant phage particles were propagated on the E. coli lysogenic strain P2 392. A recombinant library containing 2.0 × 10^8 p.f.u. was obtained.

Recombinant phages were propagated on E. coli LE 392 to yield ~1000 p.f.u. per 90 mm Petri plate and were transferred from the plaques onto nitrocellulose membrane filters (Schleicher & Schuell) by overlaying the plaques with the filters, which were then screened by plaque hybridization (Sambrook et al., 1989) using α-32P-labelled DNA probes.

**DNA hybridization.** Probe DNA used in screening the C. dubliniensis genomic DNA library and in Southern hybridization experiments was labelled with α-32PdATP (3000 Ci mmol⁻¹; 110 TBq mmol⁻¹) by random priming with a random hexanucleotide primer labelling kit (Prime-a-Gene, Promega). All hybridizations were performed under conditions of high stringency (65 °C), as described by Sambrook et al. (1989).

**PCR isolation of ACT1-associated introns.** Amplification of ACT1-associated intron sequences from *Candida* strains was performed in 100 μl reaction mixtures containing 100 pmol each of the forward and reverse primers, INTF/INTR (Table 1), 250 μM deoxythymidine triphosphates, 2.0 mM MgSO₄, 20 mM Tris/HC1 (pH 8.8 at 25 °C), 10 mM KCl, 10 mM (NH₄)₂SO₄, 0.1% (v/v) Triton X-100, 1 U Vent DNA polymerase (New England Biolabs) and 500 ng template DNA. PCR reactions were performed in a DNA Thermal Cycler (Perkin Elmer Cetus) with 30 cycles of 1 min at 94 °C, 1 min at 52 °C and 1 min at 72 °C, followed by 10 min. The amplimers were purified using the Wizard PCR Prep DNA Purification system (Promega) and cloned into pBluescript using standard procedures (Sambrook et al., 1989).

**DNA sequencing.** This was performed by the dye terminator method of Sanger et al. (1977) using an automated Applied Biosystems 370A DNA sequencer and dye-labelled terminators (ABI Prism Dye Terminator Cycle Sequencing Ready Reaction kit, Applied Biosystems). Searches of the GenBank database for nucleotide sequence similarities were performed using the BLAST family of computer programs (Altschul et al., 1990). Nucleotide sequence alignments were carried out using the CLUSTAL W sequence analysis program (Thompson et al., 1994).

**PCR identification of C. dubliniensis.** PCR identification of C. dubliniensis using the C. dubliniensis-specific primer pair DUBF/DUBR (Table 1) was carried out in a 50 μl final volume containing 10 pmol each of the forward and reverse primers, 2.5 mM MgCl₂, 10 mM Tris/HC1 (pH 9.0 at 25 °C), 10 mM KCl, 0.1% (v/v) Triton X-100, 2.5 U Tag DNA polymerase...
(Promega) and 25 μl template-DNA-containing cell supernatant (prepared as described above). Each reaction mixture also contained 10 pmol each of the universal fungal primers RNAF/RNAR (Fell, 1993), which amplify approximately 610 bp from all fungal large-subunit rRNA genes and were used as an internal positive control. Cycling conditions consisted of 6 min at 95 °C followed by 30 cycles of 30 s at 94 °C, 30 s at 58 °C, 30 s at 72 °C, followed by 72 °C for 10 min. Amplification products were separated by electrophoresis through 2.0% (w/v) agarose gels containing 0.5 μg ethidium bromide ml⁻¹ and were visualized on a UV transilluminator.

RESULTS

Isolation and sequence analysis of the C. dubliniensis ACT1 gene

A C. dubliniensis CD36 genomic library, constructed in bacteriophage λ EMBL3, was screened by plaque hybridization with a radioactively labelled probe consisting of the entire C. albicans ACT1 (CaACT1) gene cloned into pBR322 (p1002, a gift from B. Magee, University of Minnesota). Approximately 10000 recombinant plaques were screened, from which 10 p1002-reactive plaques were detected. The plaque with the strongest signal was selected and subjected to single-plaque purification. DNA was extracted from the recombinant phage, termed λCDACT1, and restriction endonuclease digestion analysis and Southern hybridization analysis showed that λCDACT1 contained a DNA insert of approximately 15 kb. Attempts to subclone smaller hybridizing fragments from the cloned insert DNA of λCDACT1 into the vector phagemid pBluescript, to facilitate the sequencing of the C. dubliniensis ACT1 (CdACT1) gene, failed. A number of p1002-reactive recombinant plasmids were obtained. However, upon further investigation these were shown to contain both ACT1 and EMBL3 vector homologous sequences, and no recombinant plasmids harbouring only ACT1-homologous DNA were obtained. It was concluded that the ACT1-homologous insert DNA from λCDACT1 was

Table 2. Yeast species and strains used in the phylogenetic analysis of C. dubliniensis

<table>
<thead>
<tr>
<th>Yeast strain*</th>
<th>ACT1 intron sequence†</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C. albicans</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>132A</td>
<td>This study</td>
<td>Gallagher et al. (1992)</td>
</tr>
<tr>
<td>179B</td>
<td>This study</td>
<td>Gallagher et al. (1992)</td>
</tr>
<tr>
<td>ATCC 10123</td>
<td>X16377</td>
<td>Losberger &amp; Ernst (1989)</td>
</tr>
<tr>
<td><strong>C. dubliniensis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD36 (Ireland)</td>
<td>AJ236897; this study</td>
<td>Sullivan et al. (1995)</td>
</tr>
<tr>
<td>CD91 (Ireland)</td>
<td>This study</td>
<td>This study</td>
</tr>
<tr>
<td>CD70 (UK)</td>
<td>This study</td>
<td>Sullivan et al. (1997)</td>
</tr>
<tr>
<td>NCPF 3108 (UK)</td>
<td>This study</td>
<td>Sullivan et al. (1995)</td>
</tr>
<tr>
<td>CD93 (Finland)</td>
<td>This study</td>
<td>This study</td>
</tr>
<tr>
<td>94191 (Spain)</td>
<td>This study</td>
<td>Pinjon et al. (1998)</td>
</tr>
<tr>
<td>P2 (Switzerland)</td>
<td>This study</td>
<td>Boerlin et al. (1995)</td>
</tr>
<tr>
<td>CD71 (Argentina)</td>
<td>This study</td>
<td>Sullivan et al. (1997)</td>
</tr>
<tr>
<td>CM2 (Australia)</td>
<td>This study</td>
<td>Sullivan et al. (1995)</td>
</tr>
<tr>
<td>CD92 (Canada)</td>
<td>This study</td>
<td>This study</td>
</tr>
<tr>
<td><strong>C. glabrata</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCC 90876</td>
<td>AF069746</td>
<td>Unpublished data submitted to GenBank</td>
</tr>
<tr>
<td><strong>C. stellatoidea</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCC 11006</td>
<td>AJ237919; this study</td>
<td>Kwon-Chung et al. (1989)</td>
</tr>
<tr>
<td>303530</td>
<td>This study</td>
<td>bioMerieux‡</td>
</tr>
<tr>
<td>303531</td>
<td>This study</td>
<td>bioMerieux‡</td>
</tr>
<tr>
<td><strong>C. tropicalis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCPF 3111</td>
<td>AJ237918; this study</td>
<td>NCPF catalogue</td>
</tr>
<tr>
<td><strong>K. lactis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J7</td>
<td>M25826</td>
<td>Deshler et al. (1989)</td>
</tr>
<tr>
<td><strong>S. cerevisiae</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A364A</td>
<td>L00026</td>
<td>Gallwitz &amp; Sures (1980)</td>
</tr>
</tbody>
</table>

* Abbreviations: ATCC, American Type Culture Collection, (Manassas, VA, USA); NCPF, National Collection of Pathogenic Fungi, Bristol, UK. The country of origin of the C. dubliniensis isolates is shown in parentheses.
† Accession numbers are for the EMBL/GenBank nucleotide sequence databases.
‡ From the culture collection of bioMerieux, St Louis, MO, USA.
unstable when cloned in pBluescript. Therefore it was
decided to amplify CdACT1 from the recombinant
phage by PCR using a mixture containing Taq polymer-
ase and the proof-reading polymerase Pwo (Expand
high-fidelity PCR system, Boehringer) and three primer
sets homologous to regions of the CaACT1 gene and
flanking sequences, including 5'F/5'R, ACTF/ACTR
and 3'F/3'R (Table 1, Fig. 1). The three amplimers,
containing overlapping sequences, obtained following
PCR with these primers were cloned into pBluescript to
to yield recombinant plasmids p5, pACT and p3, respect-
ively (Fig. 1). The insert DNA cloned in p5 and p3 was
sequenced fully in both directions using universal
primers, while the insert DNA cloned in pACT was
sequenced fully by primer walking. These three over-
lapping sequences yielded a contiguous sequence of
1827 bp revealing an ORF of 1131 bp interrupted by a
single 632 bp intron at the 5' end (Fig. 1). The overall
nucleotide sequence identity between this ORF
(CdACT1) and the CaACT1 gene was 90.6%. This
divergence is mainly due to differences between the
intron sequences, which are 83.4% identical (Fig. 2),
while the spliced coding sequences, which are identical
in length in both species, are 97.9% identical. The
differences between the exon sequences correspond to
a total of 24 base changes. However, only one of these
base substitutions [A→G, at position 660, numbering
the sequences in the 5'-3' direction from the first base
(+1) of the translation start codon of CdACT1], results
in a change in the predicted amino acid sequence, a
conservative substitution from isoleucine to valine. At
632 bp the CdACT1 intron is 25 bp shorter than the
corresponding sequence in CaACT1; however, it is
situated in exactly the same position at the 5' end of
the gene and is recognizable by the presence of yeast intron
consensus sequences (Fig. 2). These are the 5' consensus
sequence GTATG, the 3' consensus sequence YAG, and
the branchpoint sequence TACTAAC located near the
3' end of the intron which is essential for efficient
splicing (Mount, 1982; Langford et al., 1984; Leer et al.,
1984; Molenaar et al., 1984; Teem et al., 1984).
In order to determine the level of intraspecies intron
sequence conservation in epidemiologically unrelated
isolates from geographically divergent parts of the world
the introns of nine additional C. dubliniensis isolates,
two additional C. albicans isolates (132A and 179B) and
three C. stellatoidea isolates (Table 2) were amplified
using the primer set INTF/INTR, which were com-
plementary to CaACT1 sequences flanking the intron
(Table 1). The intron sequences of the 10 C. dubliniensis
isolates tested, including CD36 (Table 2), were found to
be very highly conserved, with only one or two base
changes found within each isolate. Similar intraspecies
sequence conservation was observed with the C. albicans
and C. stellatoidea strains studied.

**Phylogenetic analysis based on ACT1 sequences**

The ACT1 gene has been used extensively to infer
interspecies relationships across broad evolutionary
distances (Zakut et al., 1982; Mertins & Gallwitz, 1987;
Wildeman et al., 1988; Fletcher et al., 1994; Cox et al.,
1995; Wery et al., 1996). This part of the study was
undertaken to confirm the phylogenetic position of C.
dubliniensis in relation to other yeast species using
ACT1 sequences. This is the first time that the phylo-
genome of C. dubliniensis has been investigated using non-
rRNA gene sequences. Since the ACT1 gene of many
yeast species contains highly conserved (i.e. exon) and
less well-conserved (i.e. intron) sequences, these regions
were compared separately. Firstly, the ACT1 spliced
coding sequences of C. albicans, C. dubliniensis, C.
glabrata, Klyveromyces lactis and Saccharomyces cere-
sisae, obtained in this study or from the databases
(Table 3), were compared using the CLUSTAL W sequence
alignment software package. Secondly, the ACT1-
associated intron sequences from selected strains of C.
albicans, C. dubliniensis, C. stellatoidea, C. tropicalis,
C. glabrata and K. lactis (Table 4) were obtained either
from GenBank or following amplification using the
INTF/INTR primer set (Table 1) and also compared using
CLUSTAL W. An evolutionary distance matrix for
each group of sequences was generated incorporating
corrections for multiple base changes according to the
These data indicated that the C. dubliniensis coding and
intron sequences differ from the corresponding C.
albicans sequences by 2-1% and 16-6%, respectively.
Evolutionary trees constructed using the neighbour-
joining method of Saitou & Nei (1987) based on these
data are shown in Fig. 3. These trees and the bootstrap
values determined for each node unequivocally con-
irmed the unique species designation of C. dubliniensis
and its phylogenetic position in relation to the other
yeast species, including C. albicans. In addition, these
data also confirm that C. albicans and C. stellatoidea are

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C. albicans</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. dubliniensis</td>
<td>2-1</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>12-9</td>
<td>12-6</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>C. glabrata</td>
<td>13-2</td>
<td>12-3</td>
<td>9-8</td>
<td>–</td>
</tr>
<tr>
<td>K. lactis</td>
<td>13-2</td>
<td>12-6</td>
<td>10-1</td>
<td>7-8</td>
</tr>
</tbody>
</table>

Table 3. Genetic distance matrix based on comparison of ACT1 gene coding sequences

Values correspond to percentages of difference corrected for multiple base changes by the method of Jukes & Cantor (1969). The ACT1 gene coding sequences used were as follows: C. albicans ATCC 10123 (X16377; Losberger & Ernst, 1989); C. dubliniensis CD36 (AJ236897; this study); S. cerevisiae A364A (L00026; Gallwitz & Sures, 1980); C. glabrata NCPF 90876 (AF069746; O. Kurzai and others, unpublished data) and K. lactis J7 (M25826; Deshler et al., 1989). Neither the C. tropicalis nor the C. stellatoidea ACT1 coding sequences are currently available in the databases, so they could not be compared with the sequences of the other yeast species used to construct the matrix.
**Table 4.** Genetic distance matrix based on comparison of ACT1-associated intron sequences

Values correspond to percentages of difference corrected for multiple base changes by the method of Jukes & Cantor (1969). The intron sequences used were as follows: *C. albicans* ATCC 10123 (X16377; Losberger & Ernst, 1989); *C. stellatoidea* ATCC 11006 (AJ237919, this study); *C. dubliniensis* CD36 (AJ236897; this study); *C. tropicalis* NCPF 3111 (AJ237918, this study); *C. glabrata* NCPF 90876 (AF069746; unpublished data submitted to GenBank) and *K. lactis* J7 (M25826; Deshler et al., 1989). The sequence of the *S. cerevisiae* ACT1-associated intron (L00026; Gallwitz & Sures, 1980) was not included in the construction of the matrix because it was only 308 bp in length, significantly shorter than the intron sequences of the other yeasts studied, and so valid genetic distance determinations with this sequence and the others used to construct the matrix could not be made.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. albicans</em></td>
<td>~</td>
<td>0.2</td>
<td>~</td>
<td>16.6</td>
<td>43.4</td>
<td>54.8</td>
</tr>
<tr>
<td><em>C. stellatoidea</em></td>
<td>0.2</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>58.1</td>
</tr>
<tr>
<td><em>C. dubliniensis</em></td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td><em>C. tropicalis</em></td>
<td>16.6</td>
<td>16.6</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td><em>C. glabrata</em></td>
<td>43.4</td>
<td>43.5</td>
<td>47.1</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td><em>K. lactis</em></td>
<td>54.8</td>
<td>55.0</td>
<td>57.1</td>
<td>54.0</td>
<td>61.4</td>
<td>63.1</td>
</tr>
</tbody>
</table>

**Fig. 3.** Unrooted phylogenetic neighbour-joining trees generated from the alignment of the ACT1-exon (a) and -intron (b) sequences of *C. dubliniensis* and other yeast species. Numbers at the nodes were generated by bootstrap analysis (Felsenstein, 1985) and represent the percentage of times the arrangement occurred in 1000 randomly generated trees. The sequences used to construct the trees are indicated in the legends to Tables 3 and 4.

so closely related as to be considered a single species (Kamiyama et al., 1989; Sullivan et al., 1995).

**PCR-based identification of *C. dubliniensis***

Because of the many phenotypic similarities shared by *C. albicans* and *C. dubliniensis* it is not easy to discriminate between isolates of these species in the clinical laboratory. However, examination of an alignment of the ACT1-associated intron sequences of these two species (Fig. 2) and the observation that they differ by 16.6% (Table 4) suggested that this significant sequence divergence could be exploited as a means to identify *C. dubliniensis* accurately and rapidly in combination with a rapid template DNA preparation procedure. PCR primers specific for the *C. dubliniensis* intron (DUBF/DUBR; Table 1, Fig. 1) were synthesized and used to amplify a DNA fragment of 288 bp from *C. dubliniensis* template DNA obtained by boiling cells from a single 48 h colony suspended in 50 μl water for 10 min. PCR reactions also contained the fungal universal primers RNAF/RNAR (Fell, 1993), which amplify a product of approximately 610 bp from the fungal large-subunit rRNA gene and serve as an internal positive control. While all fungal species should produce a product of approximately 610 bp with the RNAF/RNAR primers, only *C. dubliniensis* isolates should yield the 288 bp amplimer with the DUBF/DUBR primer set. The *C. dubliniensis*-specific primer pair DUBF/DUBR was tested in a blind trial using template DNA...
Table 5. Yeast species used in PCR identification experiments with the C. dubliniensis-specific primers DUBF/DUBR

<table>
<thead>
<tr>
<th>Species</th>
<th>No. of isolates</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. albicans</td>
<td>53</td>
<td>This study; Pinion et al. (1998); Jabra-Rzik et al. (1999)</td>
</tr>
<tr>
<td>C. dubliniensis</td>
<td>122</td>
<td>This study; Sullivan et al. (1995); Sullivan et al. (1997); Coleman et al. (1997a); Moran et al. (1997, 1998); Pinion et al. (1998); Jabra-Rzik et al. (1999); Pujol et al. (1997)</td>
</tr>
<tr>
<td>C. glabrata</td>
<td>1</td>
<td>Haynes &amp; Westerneng (1996)</td>
</tr>
<tr>
<td>C. kefyr</td>
<td>1</td>
<td>NCPF 3234</td>
</tr>
<tr>
<td>C. krusei</td>
<td>1</td>
<td>Haynes &amp; Westerneng (1996)</td>
</tr>
<tr>
<td>C. norvegensis</td>
<td>1</td>
<td>NCPF 3860</td>
</tr>
<tr>
<td>C. parapsilosis</td>
<td>4</td>
<td>This study</td>
</tr>
<tr>
<td>C. sake</td>
<td>1</td>
<td>NCPF 8360</td>
</tr>
<tr>
<td>C. stellatoidea</td>
<td>1</td>
<td>ATCC 11006</td>
</tr>
<tr>
<td>C. tropicalis</td>
<td>1</td>
<td>This study</td>
</tr>
<tr>
<td>T. beigeli</td>
<td>1</td>
<td>NCPF 3857</td>
</tr>
</tbody>
</table>

One hundred and fourteen of the C. dubliniensis isolates were recovered from oral specimens, five were recovered from faecal specimens and one each was recovered from a vaginal, sputum and a post-mortem lung specimen. The isolates were recovered from different countries as follows: Argentina, 1 isolate; Australia, 2; Belgium, 5; Canada, 6; France, 4; Germany, 4; Greece, 1; Ireland, 48; Scandinavia, 4; Spain, 5; Switzerland, 4; UK, 17; USA, 21.

![Agarose gel electrophoresis of PCR-amplified DNA products generated using the C. dubliniensis-specific primers DUBF/DUBR (Table 1, Fig. 2) and the universal fungal primers RNAWRNAR (Fell, 1993) with template DNA from yeast strains. The profiles shown correspond to: the C. dubliniensis type strain CD36 (lane 1); C. albicans (lane 2); C. glabrata (lane 3); C. kefyr (lane 4); C. krusei (lane 5); C. norvegensis (lane 6); C. sake (lane 7); C. stellatoidea (lane 8); C. tropicalis (lane 9); Trichosporon beigeli (lane 10); C. dubliniensis American isolate (lane 11); C. dubliniensis Argentinian isolate (lane 12); C. dubliniensis Australian isolate (lane 13); C. dubliniensis Canadian isolate (lane 14); C. dubliniensis French isolate (lane 15); C. dubliniensis German isolate (lane 16); C. dubliniensis UK isolate (lane 17). A negative control in which no template DNA was used in the PCR reaction was also included (lane 18). The 288 bp C. dubliniensis-specific amplimer generated by the DUBF/DUBR primers is present in lane 1 and lanes 11–17. Lanes marked M contain 100 bp size reference markers.](image)

Fig. 4. Agarose gel electrophoresis of PCR-amplified DNA products generated using the C. dubliniensis-specific primers DUBF/DUBR (Table 1, Fig. 2) and the universal fungal primers RNAWRNAR (Fell, 1993) with template DNA from yeast strains. The profiles shown correspond to: the C. dubliniensis type strain CD36 (lane 1); C. albicans (lane 2); C. glabrata (lane 3); C. kefyr (lane 4); C. krusei (lane 5); C. norvegensis (lane 6); C. sake (lane 7); C. stellatoidea (lane 8); C. tropicalis (lane 9); Trichosporon beigeli (lane 10); C. dubliniensis American isolate (lane 11); C. dubliniensis Argentinian isolate (lane 12); C. dubliniensis Australian isolate (lane 13); C. dubliniensis Canadian isolate (lane 14); C. dubliniensis French isolate (lane 15); C. dubliniensis German isolate (lane 16); C. dubliniensis UK isolate (lane 17). A negative control in which no template DNA was used in the PCR reaction was also included (lane 18). The 288 bp C. dubliniensis-specific amplimer generated by the DUBF/DUBR primers is present in lane 1 and lanes 11–17. Lanes marked M contain 100 bp size reference markers.

from the yeast isolates listed in Table 5 as follows: C. albicans (n = 53), C. dubliniensis (n = 122), C. glabrata (n = 1), C. kefyr (n = 1), C. krusei (n = 1), C. norvegensis (n = 1), C. parapsilosis (n = 4), C. sake (n = 1), C. stellatoidea (n = 10), C. tropicalis (n = 1) and Trichosporon beigeli (n = 1). All 196 yeast isolates yielded an amplimer of approximately 610 bp, but only the C. dubliniensis isolates yielded the 288 bp amplimer. Fig. 4 shows examples of the PCR amplimers obtained with representative strains belonging to a variety of different yeast species, including epidemiologically unrelated C. dubliniensis isolates from disparate geographical locations. Use of this PCR test in conjunction with the rapid template DNA preparation procedure used here means that a C. dubliniensis isolate can be identified unequivocally in less than 4 h.
**DISCUSSION**

Because *C. dubliniensis* was only described recently it is important to further investigate and confirm its taxonomic and phylogenetic relationship to other medically important yeast species, especially the closely related *C. albicans*. The close relationship between *C. dubliniensis* and *C. albicans* has resulted in difficulties in developing rapid and reliable identification techniques capable of definitively discriminating between the two species. As a result, in-depth epidemiological studies on the prevalence of this organism have been hampered. There is an urgent requirement for a test which can rapidly and definitively identify *C. dubliniensis* directly following primary culture from clinical specimens. Such a test must be simple to use, inexpensive, easily accessible to clinical diagnostic laboratories and suitable for application to large numbers of samples.

The objectives of this study were to investigate the phylogenetic relationship of *C. dubliniensis* to other yeast species using non-rRNA gene sequences, and to develop a rapid identification technique for this organism. To achieve these objectives the ACT1 gene was chosen for investigation, firstly because it is ideal for inferring phylogenetic relationships due to its high degree of sequence conservation in all eukaryotes, and secondly because it is unusual among yeast genes in that it is interrupted by an intron in most yeasts studied. Cloning and gene sequence analysis revealed that the overall similarity between the *C. dubliniensis* and *C. albicans* ACT1 genes is 90.6%. Comparison of the spliced coding sequences of the two species revealed that the exon sequences are 97.9% identical. This level of homology is comparable to the percentage sequence identity between the two species reported previously for the V3 variable region of the large-subunit rRNA gene (97.52–97.75% ; Sullivan et al., 1995, 1997) and the small-subunit rRNA gene (98.6% ; Gilfillan et al., 1998).

The predicted *C. dubliniensis* ACT1 protein sequence was identical to that of *C. albicans*, apart from a single conservative substitution. A phylogenetic tree generated from nucleotide comparisons of *ACT1* coding sequences from *C. dubliniensis* and a variety of yeast species showed that *C. dubliniensis* was grouped separately from *C. albicans* and the other yeast species in 100% of trees generated (Fig. 3a). These studies represent the first phylogenetic investigation of *C. dubliniensis* based on non-rRNA gene sequences, and they unequivocally confirm its unique position as a separate taxon within the genus *Candida* as determined previously by comparative rRNA gene sequence analysis (Sullivan et al., 1995, 1997; Gilfillan et al., 1998). In contrast to the highly conserved nature of the *C. dubliniensis* and *C. albicans* ACT1 exon sequences there was considerable divergence (16.6%) between the ACT1-associated introns of the two species. When these and the ACT1-associated intron sequences from a number of other yeast species were used to generate a second phylogenetic tree (Fig. 3b) the unique position of *C. dubliniensis* as a separate taxon within the genus *Candida* was affirmed. These results also confirmed that *C. dubliniensis* is most closely related to *C. albicans*. In addition, the *C. albicans* and *C. stellatoidea* ACT1-associated introns were found to differ by one basepair substitution, corresponding to a 0.2% sequence divergence. These findings provided further evidence that *C. albicans* and *C. stellatoidea* should be considered as the same species. This situation is analogous to that found between *S. cerevisiae* and *S. carlsbergensis*, where the ACT1-associated introns differ by one basepair deletion and one basepair substitution, and it is accepted that these two organisms are in fact the same species (Nellen et al., 1981). The *C. tropicalis* intron sequence differs from that of *C. albicans* by 43.4%, confirming that it is more distantly related to *C. albicans* than *C. dubliniensis* (Table 4, Fig. 3b). All of these findings indicate that the ACT1-associated intron sequences are not subject to the same level of evolutionary constraint as the ACT1 coding sequences.

The ACT1 genes of fungal species, in general, are noteworthy because of the presence of introns (Gallwitz & Sures, 1980; Fidel et al., 1988; Wildeman et al., 1988; Deshler et al., 1989; Losberger & Ernst, 1989, 1994; Fletcher et al., 1994; Cox et al., 1995; Matheucci et al., 1995, 1996). At present, most known introns can be assigned unambiguously to one of four classes, depending on the intron structure and location (Krainer & Maniatis, 1988). ACT1-associated introns belong to class IV, which are nuclear pre-mRNA introns. The *C. dubliniensis* ACT1-associated intron is located at the 5' end of the gene, where it interrupts the fourth codon. The ACT1 genes of *C. albicans*, *C. glabrata*, *S. cerevisiae* and *K. lactis* all contain introns located at this codon (Losberger & Ernst, 1989; O. Kurzai and others, unpublished data submitted directly to GenBank – accession no. AF069746); Gallwitz & Sures, 1980; Deshler et al., 1989). This position is conserved amongst fungi, as all fungal actin genes that contain an intron do so at the third, fourth or fifth codon. Three conserved sequence elements have also been identified in the nuclear pre-mRNA introns of yeasts, at the 5' and 3' splice sites and at a site within the intron near the 3' splice site, known as the branchpoint sequence. All three conserved elements have been shown to be important for the accurate and efficient splicing of introns in *S. cerevisiae* (Langford et al., 1984; Leer et al., 1984; Molenar et al., 1984; Teem et al., 1984; Mount, 1982). The *C. dubliniensis*, *C. albicans*, *C. stellatoidea* and *C. tropicalis* ACT1-associated introns possess all three conserved elements, namely GTATG (5' consensus), TAG (3' consensus) and TACTAAC (branchpoint), (this study, Fig. 2; Losberger & Ernst, 1989). These sequences are also present in *C. glabrata* and *K. lactis* although the 3' consensus sequence is CAG (Deshler et al., 1989; see GenBank accession no. AF069746 for the *C. glabrata* ACT1-intron sequence).

One striking feature of the *C. dubliniensis* introns was that they showed little intraspecies variation, even among isolates from geographically divergent locations. The small changes which were recorded consisted of single base changes, some of which were shared by more
than one strain, and deletions which occurred at the end of poly(T) and poly(A) stretches. Introns containing these deletions were sequenced on separate occasions using different preparations of template DNA to rule out the possibility of sequencing or amplification artefacts. We concluded that these deletions are genuine and are probably the result of slipped-strand miscalching during replication. Similar intraspecies sequence conservation was observed with the ACT1-associated introns from C. albicans and C. stellatoidea. Boucher et al. (1996) made similar findings with their analysis of the group I self-splicing intron present in the large-subunit rRNA gene, in which the intron is present in a similar location in C. albicans, C. stellatoidea and C. dubliniensis. Again there was no significant intraspecies variation in the intron sequence. Furthermore, the C. albicans self-splicing intron and that of C. stellatoidea showed a high degree of homology, differing only by three single basepair substitutions. They also found that the homology between the C. albicans and C. dubliniensis group I introns (CaLSU and CdLSU, respectively) was quite high except for two regions of divergence contained in two stem-loop regions, both of which are much longer in C. dubliniensis than in C. albicans. These two regions lie outside the catalytic core, and although they are predicted to have a more complex secondary structure than those of C. albicans they do not affect the self-splicing ability of the intron. Our analysis of the C. dubliniensis and C. albicans ACT1-associated introns showed that although identical conserved elements are present in both species, nucleotide differences accounting for a 16.6% sequence divergence were dispersed throughout the length of the intron (Fig. 2). With group I introns, conservation of the nucleotide sequence may be important as it dictates the secondary structure of the intron and therefore its self-splicing ability. However, with group IV introns, such as the C. dubliniensis and C. albicans ACT1-associated introns, the splicing event is mediated by the spliceosome and although maintenance of the three conserved elements is important for splicing there do not appear to be any other constraints upon conservation of the nucleotide sequence. This may explain why divergence between the C. dubliniensis and C. albicans ACT1-associated introns sequences is dispersed throughout the intron.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Wellcome Trust (no. 047204). We thank Dr B. Magee, University of Minnesota, for providing the C. albicans ACT1-encoding plasmid p1002. We thank our colleagues who sent us isolates of Candida dubliniensis, including Aristeia Velegriki, National University of Athens, Greece; Markus Ruhnke, Vichkik Klinikum der Humboldt Universitat, Berlin, Germany; Luc Giasson, Laval University, Quebec, Canada; Jose Ponton, Universidad del Pais Vasco, Bilbao, Spain; Elizabeth Johnson, Public Health Laboratory Service, Mycology Reference Laboratory, Bristol, UK; Frank Odds, Jansen Research Foundation, Beerse, Belgium; Lakshman Samaranayake, University of Hong Kong; Mary Ann Jabra-Rizk, Johns Hopkins University, Baltimore, USA; Jean-Marie Bastide, Universite de Montpellier, France; Fiona Mulcahy, St. James’s Hospital, Dublin, Ireland; and Christine McCreary, Dublin Dental Hospital, Dublin, Ireland.

REFERENCES

albicans: presence of homologous sequences in Candida species other than C. albicans. Microbiology 144, 55–72.


Morgan, G. P., Sanglard, D., Donnelly, S. M., Shanley, D. B.,


Nellen, W., Donath, C., Moos, M. & Gallwitz, D. (1981). The nucleotide sequences of the actin genes from Saccharomyces carlsbergensis and Saccharomyces cerevisiae are identical except for the introns. J Mol Appl Genet 1, 239–244.


Received 9 April 1999; revised 26 May 1999; accepted 27 May 1999.