Comparative genomics using *Candida albicans* DNA microarrays reveals absence and divergence of virulence associated genes in *Candida dubliniensis*

Gary Moran,¹ Cheryl Stokes,¹ Sascha Thewes,² Bernhard Hube,² David C. Coleman,¹ and Derek Sullivan¹

Author for correspondence: Derek Sullivan. Tel: + 353 1 6127275, Fax: +353 1 6127295
e-mail: derek.sullivan@dental.tcd.ie

¹ Microbiology Research Unit, Department of Oral Surgery, Oral Medicine and Oral Pathology, School of Dental Science, University of Dublin, Trinity College, Dublin 2, Republic of Ireland

² Robert Koch Institut, Nordufer 20, 13353 Berlin, Germany

**Running title:** Comparative genomics of *Candida albicans* and *Candida dubliniensis*

**Subject Category:** Genes and Genomes

**Keywords:** *Candida albicans*, *Candida dubliniensis*, microarrays, comparative genomics
SUMMARY

*Candida dubliniensis* is a pathogenic yeast species closely related to *Candida albicans*. However, it is less frequently associated with human disease and displays reduced virulence in animal models of infection. We have used comparative genomic hybridisation (CGH) in order to discover why *C. dubliniensis* is apparently less virulent than *C. albicans*. In these experiments we compared the genomes of the two species by co-hybridising *C. albicans* microarrays with fluorescently labeled *C. albicans* and *C. dubliniensis* genomic DNA. We found that *C. dubliniensis* genomic DNA hybridised reproducibly to 96% percent of *C. albicans* gene-specific sequences indicating a significant degree of nucleotide sequence homology (>60%) in these sequences. The remaining 4% of sequences (representing 234 genes) gave *C. albicans/C. dubliniensis* normalised fluorescent signal ratios indicative of significant sequence divergence (<60% homology) or absence in *C. dubliniensis*. We identified sequence divergence in several genes (confirmed by Southern Blot analysis and sequencing analysis of PCR products) with putative virulence functions including the gene encoding the hypha-specific human transglutaminase substrate Hwp1p. Poor hybridisation of *C. dubliniensis* genomic DNA to the secreted aspartyl proteinase encoding gene SAP5 array sequences also led us to determine that SAP5 was absent in *C. dubliniensis* and that this species possesses only one gene homologous to SAP4 and SAP6 of *C. albicans*. In addition, divergence and absence of sequences in several gene families was identified including a family of HYR1-like GPI-anchored proteins, a family of genes homologous to a putative transcriptional activator (CTA2) and several ALS genes. This study has confirmed the close relatedness of *C. albicans* and *C. dubliniensis* and has identified a subset of unique *C. albicans* genes that may contribute to the increased prevalence and virulence of this species.

INTRODUCTION

*Candida dubliniensis* is associated with oral candidosis in the HIV-infected population in which this species was first identified in 1995 (Jabra-Rizk et al., 2001; Sullivan et al., 1995; Sullivan et al., 2004). Recently, several studies have also identified *C. dubliniensis* as a cause of oral disease in diabetic and cancer patients (Sebti et al., 2001; Willis et al., 2000). However, the closely related species *C. albicans* appears to be more successful than *C. dubliniensis* as a commensal of the human oral cavity in healthy individuals, as
determined by standard oral swab sampling methods (Coleman et al., 1997). In addition, the incidence of *C. dubliniensis* isolation from blood cultures is extremely low compared to *C. albicans* (Kibbler et al., 2003; Pfaller et al., 2004). In a recent study of *Candida* spp. recovered from blood cultures in 6 sentinel hospitals in England and Wales between 1997 and 1999, *C. dubliniensis* was isolated from only 2% of samples compared to 65% for *C. albicans*. Two studies have also demonstrated that *C. dubliniensis* is less virulent than *C. albicans* in a murine model of systemic candidosis (Gilfillan et al., 1998; Vilela et al., 2002). The reason for the apparent difference in virulence between the two species is unknown as they are phenotypically very similar and seem to share many of the traits traditionally associated with virulence in *C. albicans*. In particular both species have the ability to form true hyphae, to adhere to human epithelium and to produce secreted aspartyl proteinases (Gilfillan et al., 1998; Hannula et al., 2000; Vilela et al., 2002). However, *C. dubliniensis* does not form hyphae as rapidly as *C. albicans* in response to shifts in pH/temperature or when incubated in serum (Gilfillan et al., 1998). In contrast, when cultured on Staib agar or Pal’s agar *C. dubliniensis* forms abundant hyphae, pseudohyphae and chlamydospores, whereas *C. albicans* remains in the yeast phase (Al Mosaid et al., 2003; Al Mosaid et al., 2001). *Candida dubliniensis* also seems to be more sensitive to environmental stress such as elevated temperature and NaCl concentration (Alves et al., 2002; Pinjon et al., 1998).

Comparative genomic hybridisation (CGH) studies with DNA microarrays provide a rapid and cost effective method to obtain informative data about unsequenced genomes and has been used extensively to compare gene content in prokaryotic and eukaryotic microorganisms (Daran-Lapujade et al., 2003; Dong et al., 2001; Murray et al., 2001). The completion of the *C. albicans* genome project and the availability of *C. albicans* DNA microarrays now enables genomes of different strains of *C. albicans* and closely related species such as *C. dubliniensis* to be compared. In the present study, CGH was performed between *C. albicans* and *C. dubliniensis* using *C. albicans* DNA microarrays in order to identify genomic differences that might account for the difference in virulence between *C. albicans* and *C. dubliniensis*. This approach was deemed feasible as all *C. dubliniensis* genes analysed to date share greater than 90% identity at the nucleotide sequence level with the orthologous *C. albicans* genes. Total genomic DNA from *C. albicans* and *C. dubliniensis* was co-
hybridised to *C. albicans* DNA microarrays and the relative hybridisation efficiency of *C. dubliniensis* and *C. albicans* DNA to each gene-specific spot was compared. This approach allowed us to identify the presence of thousands of *C. albicans* homologous genes in *C. dubliniensis* without the need for sequence analysis and has guided us towards genes which are highly divergent or even absent from *C. dubliniensis*. We anticipate that this collection of *C. albicans*-specific sequences may contain genes that contribute to the observed differences in virulence and epidemiology between these two organisms.

**METHODS**

*Candida* strains and culture conditions

*Candida albicans* strain SC5314 was used as a control in all comparative genomic hybridisation experiments using Eurogentec *C. albicans* DNA microarrays. *Candida dubliniensis* strains used in this study included the *C. dubliniensis* type strain CD36 (American Type Culture Collection reference ATCCMYA-178, British National Collection of Pathogenic Fungi reference NCPF3949), which is a representative of *C. dubliniensis* Cd25 fingerprint group I (genotype 1) and *C. dubliniensis* strain CD514, a strain representative of Cd25 fingerprint group II (genotype 3) (Gee *et al.*, 2002). Strains were routinely grown on Potato Dextrose Agar (PDA; Oxoid) medium, pH 5.6, at 37°C. For liquid culture, cells were grown in yeast extract-peptone-dextrose (YEPD) broth, also at 37°C (Gallagher *et al.*, 1992).

Chemicals, enzymes and radioisotopes

All chemicals used were of molecular biology grade and were purchased from Sigma-Aldrich. Molecular biology enzymes and kits were purchased from Promega or New England Biolabs unless otherwise indicated. Cy5 and Cy3 dUTP were purchased from Amersham Biosciences Europe. Supplies of [α-P\(^{32}\)]dATP (6000 Ci/mmol\(^{-1}\), 220 TBq/mmol\(^{-1}\)) were purchased from NEN Life Sciences.

DNA Microarrays

*Candida albicans* DNA microarrays used in this study were constructed by Eurogentec based on the Galar Fungail consortium’s annotation of the *C. albicans* SC5314 genome sequence in the CandidaDB database
This annotation was produced based on the genome sequence released by the Stanford Genome Technology Center. Each glass slide microarray contained sequences corresponding to 6039 ORFs (98% of annotated genes) that were approximately 300 bp in length and spotted in duplicate.

Genomic DNA preparation

High molecular weight total genomic DNA was recovered from Candida strains by organic extraction following digestion of the cell wall with Zymolyase 20T (Seikagaku Corp.) and proteinase K treatment (Roche diagnostics) as described by Gallagher et al. (Gallagher et al., 1992).

Genomic DNA labeling and microarray hybridisation

For DNA labeling experiments with Cy5 dUTP and Cy3 dUTP, total genomic DNA (2 µg) was first fragmented by either restriction endonuclease digestion or sonication. For restriction endonuclease digests, two 1 µg aliquots of DNA were separately digested with Tru1I or RsaI (Fermentas). These digests were then heat inactivated, extracted once with a mixture phenol:chloroform:isoamyl alcohol (25:24:1) and ethanol precipitated. The two separate aliquots of digested DNA were combined to give a mixture of Tru1I and RsaI fragments (50 to 4,000 bp) for labeling. Alternatively, separate DNA samples were prepared for labeling by sonication using a Sonoplus HD70 sonicator (Bandelin Electronic) at 75% power for 30 cycles producing DNA fragments ranging from 500 to 5,000 bp.

Each labeling reaction was carried out with 2 µg of either sheared or digested genomic DNA using the RadPrime random priming labeling system (Invitrogen) incorporating Cy5 dUTP into C. albicans SC5314 genomic DNA and Cy3 dUTP into C. dubliniensis DNA fragments. After labeling, reaction products were purified with Nucleospin PCR clean up columns (Macherey-Nagel) and concentrated to a final volume < 5 µl with a Microcon YM-30 column (Millipore). Cy5-labeled and Cy3-labeled reactions were mixed together in DIG EasyHyb buffer (Roche Diagnostics) to a final volume of 60 µl for hybridisation. The mixture was denatured at 98 °C for 5 min then chilled on ice. Microarray slides (Eurogentec) were placed in a
hybridisation chamber (Corning), covered with a glass LifterSlip (Erie Scientific Company) and the labeling reaction was carefully applied at the edges of the slide. The chamber was sealed and incubated in the dark at 42 °C for 16-18 h. Slides were washed at high stringency at room temperature as follows: (i) 5 min in 1 x SSC, 0.03 % (w/v) SDS, (ii) 5 min in 0.2 x SSC and (iii) 5 min in 0.05 x SSC. Following washing, slides were dried thoroughly by centrifugation at low speed for 5 min in a 50 ml disposable plastic tube (Greiner Bio-One) and scanned immediately.

Each of the hybridisations performed using digested DNA and sheared DNA were performed on two separate occasions. One additional hybridisation was also performed between sheared Cy5-labeled *C. albicans* DNA and sheared Cy3-labeled *C. dubliniensis* CD514 DNA.

**Data analysis**

DNA microarray slides were scanned with the GenePix 4000B scanner (Axon Instruments). Data were extracted from scanned images using the GenePix Pro 4.1.1.4 software package (Axon Instruments). Data normalisation and subsequent analysis was carried out with the GeneSpring 6.1 software package (Silicon Genetics). Hybridisation data from each DNA ‘spot’ on the slide was only included for analysis if the control (*C. albicans*) channel signal was above local background plus 2 standard deviations (2SD). Signal intensities in both channels were background corrected. Measurements were normalised across the whole chip by dividing each measurement by the median of all measurements taken for that chip. A normalised fluorescence ratio value was determined for each spot by dividing the *C. albicans* control channel normalised signal by the *C. dubliniensis* normalised signal values. The log₂ value of each ratio was determined and the log₂ ratios of duplicate spots were averaged. The significance of normalised ratios < 1 was determined in replicate experiments using the Student’s *t* test. The raw data has been submitted in a MIAME compliant format to the ArrayExpress database at the European Bioinformatics Institute.

The relationship between log₂ ratio values and nucleotide sequence homology was determined by linear regression analysis using Prism 4.0 (GraphPad Software). For this analysis, nucleotide sequence homology between the array printed *C. albicans* sequences (~ 300 bp) and the corresponding region of available
homologous *C. dubliniensis* sequences was determined using DNA Strider 3.1 software. Sequences used in this analysis included the available *C. dubliniensis* gene sequences from GenBank and PCR-amplified sequences described here (Table 1). Sequences were included for analysis only when a minimum of 100 bp of uninterrupted sequence could be aligned. Gaps of over 50 bp in length were excluded from homology calculations. On the basis of this analysis, we chose genes with normalised ratios < 0.5 (p value < 0.05) in replicate experiments with both sheared and digested genomic DNA for further study (see results).

Larger sequence alignments (> 500 bp) described in the results were carried out using the CLUSTAL W software package (Higgins & Sharp, 1988).

**Southern hybridisation**

Southern hybridisation analysis was carried out as described previously (Moran *et al.*, 1998; Southern, 1975) using DNA sequences labeled with [α-32P]dATP by random primer labeling (Prime-a-Gene system; Promega) or using DIG-labeled probes incorporating DIG-11-dUTP (Roche Diagnostics) during PCR amplification as described by the manufacturers. In all instances, post hybridisation washes were performed at reduced stringency (60 °C with 0.5 x SSC, 0.1% [w/v] SDS) unless otherwise indicated.

**PCR amplification of *C. dubliniensis* genome sequences**

PCR amplification from *C. dubliniensis* genomic DNA template was carried out as described previously (Moran *et al.*, 1998; Moran *et al.*, 2002). Oligonucleotide primers used in this study were synthesised by Sigma-Genosys (Table 2). PCR amplified DNA fragments were sequenced where indicated using the dideoxy chain termination method by Lark Technologies (Saffron Walden, United Kingdom)
RESULTS

Comparative genomic microarray hybridisation

To label *Candida* chromosomal DNA efficiently by random priming with Cy3 or Cy5 dUTP, it was first necessary to fragment the chromosomal DNA. Two DNA fragmentation methods (sonication and restriction endonuclease digestion) were compared in order to determine if either labeling method introduced artifacts into the microarray results. The hybridisation efficiency of *Candida albicans* SC5314 genomic DNA prepared by either sonication or restriction endonuclease digestion to Eurogentec *C. albicans* microarrays was found to be comparable. In replicate experiments involving *C. albicans* genomic DNA labeled following restriction endonuclease digestion, 5912 duplicate spots (98.4%) gave *C. albicans* signals 2SD above background. Similarly, using genomic DNA labeled following random shearing by sonication, 5892 (98%) duplicate spots were included for analysis. Using the same criteria, *C. dubliniensis* genomic DNA prepared by either sonication or restriction endonuclease digestion hybridised to at least 95% of gene-specific spots in replicate hybridisations.

Relative hybridisation efficiency of co-hybridised Cy5-labeled *C. albicans* and Cy3-labeled *C. dubliniensis* genomic DNA to the microarrays was assessed by determining a normalised ratio of *C. albicans* and *C. dubliniensis* signal intensities (Cy5/Cy3 normalised ratios). In order to investigate whether a relationship existed between the strength of hybridisation of *C. dubliniensis* DNA to the array and the degree of nucleotide homology between the corresponding *C. albicans* and *C. dubliniensis* sequences, we plotted log2 ratio values versus percent nucleotide sequence homology. We determined the nucleotide sequence homology between the *C. albicans* probe sequences present on the array and the corresponding sequences of 11 *C. dubliniensis* genes sequences available in GenBank (Table 1). We also attempted to PCR amplify sequences using *C. albicans*-specific oligonucleotide primers (Table 2) from *C. dubliniensis* corresponding to 35 genes which hybridised poorly with *C. albicans* genomic DNA (normalised ratios ranging from 0.17 to 0.47) on the microarray. At low stringency conditions, 10 of these PCR primer sets yielded PCR amplification products with sequences homologous to the corresponding *C. albicans* gene (nucleotide sequence homology range 59%-80%, Table 1). We plotted the percent nucleotide sequence homology
between the 11 GenBank and 10 PCR amplified sequences and their *C. albicans* homologue versus the log$_2$ ratio values from the 21 DNA spots representing these genes on the array (Fig. 1). Linear regression analysis was used to generate a best fit-line from the data set which demonstrated a relationship between nucleotide sequence homology and log$_2$ ratio ($r^2 = 0.81$, $p < 0.0001$). The 11 most homologous nucleotide sequences (83.6% to 98.8% identity), including housekeeping genes such as *ACT1*, *URA3* and *ERG11*, all had normalised ratio values > 0.55. The remaining 10 genes formed a second group with intermediate sequence homologies of 59% – 80%. These 10 sequences possessed normalised ratios ranging from 0.17 to 0.47.

Based on this analysis we categorised genes into three homology groups based on normalised ratio values; (I) a high homology group (normalised ratio > 0.5; > 80% nucleotide sequence homology), (II) a medium homology group (normalised ratio 0.25 to 0.5; 60-80% nucleotide sequence homology) and (III) sequences that possessed low homology or were possibly absent in *C. dubliniensis* (normalised ratio < 0.25; < 60% nucleotide sequence homology).

In total, 751 sequence-specific spots exhibited reduced normalised ratios below 0.5 ($p < 0.05$) in replicate array experiments with both digested and sheared genomic DNA. From this group, 500 sequences were classified as likely to possess intermediate nucleotide sequence homology (60-80%). The remaining 251 sequences (representing 4.25% of the spots analysed) gave normalised ratios < 0.25 and were predicted to possess low nucleotide sequence homology (< 60%), or were possibly absent in *C. dubliniensis*.

**Categorisation of divergent genes**

We decided to examine the group of sequences predicted to possess low nucleotide sequence homology in more detail as these genes are most likely to be functionally different or possibly even absent in *C. dubliniensis*. This group of 251 sequences were found to correspond to 234 genes (Table 3), as 17 genes were found to be represented on the array by two different duplicate spots. However, 124 (53%) of these were hypothetical genes with no homology to genes of known function. Within this group 38 genes could be classified as conserved hypothetical due to significant homology to other hypothetical genes in GenBank or
CandidaDB (Table 4). Nineteen sequences (8.1%) were found to have homology to genes encoding C. albicans retrotransposon elements, including transposases and reverse transcriptases.

(i) **Putative transcriptional regulators.** A group of 21 genes were identified to possess homology to putative transcriptional regulators. Seven of these regulators had strong homology to genes encoding transcriptional activators in *S. cerevisiae* with Zn-finger DNA binding motifs. A further 9 corresponded to a family of genes encoding proteins with homology to a putative *C. albicans* transcriptional activator, CTA2 (Kaiser *et al.*, 1999). All 9 CTA2-like genes included for analysis exhibited normalised ratios < 0.25. A CLUSTAL W-generated alignment of the nucleotide sequences of CTA21, CTA22, CTA25 and CTA26 from *C. albicans* revealed that these sequences were at least 89% identical. A PCR primer pair (CTA2F/CTA26R, Table 2) homologous to conserved sequences in these ORFs did not yield amplimers from *C. dubliniensis* genomic DNA template at primer annealing temperatures of 50 °C. Southern hybridisation analysis with a CTA2 probe at least 90% homologous to 6 *C. albicans* CTA2-like sequences annotated in the CandidaDB database revealed multiple hybridising fragments in EcoRI- or HindIII-digested *C. albicans* DNA (Fig. 2). Hybridisation of the same sequence to *C. dubliniensis* genomic DNA digested with EcoRI or HindIII at reduced stringency did not reveal any hybridising fragments (Fig. 2). These findings are in agreement with the array data that this gene family are significantly divergent (i.e. share low nucleotide sequence homology) or are absent in *C. dubliniensis*.

(ii) **Putative membrane transporters.** Seven genes with strong homology to membrane transporters were also found to hybridise poorly with *C. dubliniensis* genomic DNA including the oligopeptide transporter encoding gene *OPT1*, the choline transporters *HNM3* and *HNM4*, the uracil permease *FUR4* and the allantoin permease *DAL52*. The absence of homologous sequences for this group of genes was confirmed in *C. dubliniensis* by Southern hybridisation analysis (Fig. 3).

(iii) A **leucine-rich repeat family of proteins.** A large family of genes encoding proteins with leucine-rich repeats (termed IFA family, Pasteur CandidaDB) in *C. albicans* were also identified. Of 27 IFA sequences included on the array, 20 gave normalised ratios < 0.25 following hybridisation of *C. dubliniensis* genomic DNA. Four IFA sequences gave intermediate ratios (0.25 to 0.5) and only three sequences (*IFA3, IFA20* and *IFA21*) gave ratios above 0.5. A CLUSTAL W-generated alignment of the *IFA1, IFA2, IFA4* and *IFA5* ORF
nucleotide sequences (ranging in size from 1,731 to 2079 bp) revealed that most homology was present within the first 800 bp of the 5' region of the ORF family members. A PCR primer pair (IFA1F/IFA1R, Table 2) based on conserved IFA sequences in this region were designed and allowed amplification of a DNA fragment from \textit{C. dubliniensis} genomic DNA with 70% homology to IFA8, in keeping with its ratio value between 0.25 and 0.5.

\textbf{(iv) Genes encoding GPI-anchored proteins.} Genes encoding GPI-anchored proteins were also identified in this analysis, including the hypha-specific protein \textit{HYR1} (Bailey \textit{et al.}, 1996). Two sequences representing the 3' end and an internal fragment of the \textit{HYR1} gene respectively were present on the array. Both sequences yielded normalised ratios < 0.25 and low stringency Southern blot analysis of \textit{C. dubliniensis} genomic DNA with PCR amplified \textit{C. albicans} \textit{HYR1} gene sequences (nucleotides 95 to 1183, the domain bearing most homology to other GPI-anchored protein encoding genes in \textit{C. albicans}) did not identify any homologous sequence in \textit{C. dubliniensis} (Fig. 4a). Several genes encoding \textit{HYR1}-related proteins (termed the IFF gene family, CandidaDB database) were also present on the array. Of the 9 IFF sequences included for analysis (IFF2 to IFF11) only IFF5 and IFF11 gave ratios above 0.5. Sequence alignments of \textit{C. albicans} IFF genes generated with CLUSTAL W identified IFF1 as the most likely ancestral gene based on its homology to other members, particularly in the 5' region. As sequences homologous to this region of IFF1 were not included on the Eurogentec array, we amplified the 5' region of IFF1 from \textit{C. albicans} with the primer pair IFF1F/IFF1R (Table 2) and hybridised it to \textit{C. dubliniensis} genomic DNA to reveal a single hybridising band (Fig. 4b). We used this primer pair in PCR amplification reactions using \textit{C. dubliniensis} template DNA and successfully amplified a 750 bp region with 86% homology to \textit{C. albicans} IFF1.

Other sequences with characterised gene products or functions that could be inferred from homology searches included genes involved in biotin synthesis (\textit{BIO3}, \textit{BIO4}) and several unrelated genes encoding metabolic enzymes.
**Putative virulence factors**

We searched the data set of genes with normalised ratios < 0.5 to identify genes which have previously been associated with *C. albicans* virulence.

(i) **Genes encoding putative adhesins.** Of the 8 sequences with homology to members of the *ALS* gene family of GPI-anchored proteins (encoding putative adhesins) included for analysis, all gave normalised ratios < 0.5, with sequences specific for *ALS1, ALS5, ALS6* and *ALS7* yielding normalised ratios < 0.25 (Hoyer, 2001). Spots homologous to *ALS2, ALS3* and *ALS9* were excluded from our analysis due to poor hybridisation with *C. albicans* genomic DNA.

We also investigated whether sequences encoding another group of GPI-anchored proteins were present, namely those related to *HWP1*, encoding a hyphal adhesin and related sequences (*RBT1* and *IPF14331*) (Braun et al., 2000; Staab et al., 1999). *HWP1* has been associated with virulence in *C. albicans* by mediating adhesion to epithelial cells (Staab et al., 1999). *HWP1* and *RBT1* both yielded normalised ratios < 0.5 (0.48 and 0.37 respectively) in all experiments with *C. dubliniensis* CD36 genomic DNA. In order to identify a homologue of *HWP1*, a primer set designed based on the *C. albicans* *HWP1* sequence (*HWP1F/HWP1R*, Table 2) was used to amplify a 1.3 kb region of *C. dubliniensis* genomic DNA. The putative 5' upstream region was also amplified with primers designed based on the sequence of the corresponding *C. albicans* region (APL6F and HWPR2, Table 2). An ORF of 1,266 bp with homology to *C. albicans* *HWP1* was identified in these sequences. However the ORF shared only 49% identity with the nucleotide sequence of *C. albicans* *HWP1* due to the presence of several large deletions within the coding sequence (GenBank Accession No: AJ632273). The overlapping 5' region amplified from *C. dubliniensis* contained upstream sequences homologous to the *C. albicans* *APL6* gene. This synteny between *HWP1* and *APL6* is conserved in *C. albicans*, and provides further evidence that this gene is a *C. dubliniensis* *HWP1* homologue. However, the predicted protein encoded by the *C. dubliniensis* gene was 421 amino acids in length, 213 residues shorter than the *C. albicans* 634 amino acid protein. The first 50 residues of each protein were highly homologous, both containing the KR signature of the KEX2 cleavage site (Fig 5a). However, the remainder of the N-terminal half of CdHwp1p contained several large deletions compared to the *C. albicans* protein, including most of the region rich in proline, glutamine and aspartate residues (Fig.
5a)(Sundstrum, 2002). Two of these deletions (of 89 bp and 119 bp, respectively) spanned the region homologous to the microarray probe and were likely to be responsible for the low signal detected with C. dubliniensis genomic DNA. Further deletions were found in the serine-threonine rich region, however the o–site for GPI-anchor addition is conserved.

Similarly, a C. dubliniensis sequence PCR-amplified using the primers RBTF2/RBTR2 (Table 2) had homology to RBT1 (termed CdRBT1) and was also found to contain deletions of 52 bp and 82 bp, respectively. Southern blot analysis was performed to determine whether single or multiple HWP1 and RBT1 homologues could be detected in C. dubliniensis. Hybridisation of the C. dubliniensis HWP1 amplified sequence to EcoRI digested C. albicans genomic DNA revealed a single hybridising fragment of 4 kb (Fig. 5b). In C. dubliniensis genomic DNA, a strongly hybridising fragment of 9 kb was detected and a second weak hybridising fragment of 5 kb in EcoRI-digested DNA (Fig. 5b). This second fragment was identical in size to the fragment detected in Southern blots of C. dubliniensis DNA with sequences corresponding to CaRBT1 (Fig. 5c) indicating that this second hybridising fragment was likely to correspond to CdRBT1, and the most closely related gene to CdHWP1 in C. dubliniensis.

(ii) Secreted aspartyl proteinases. Sequences homologous to the 10 C. albicans secreted aspartyl proteinase (SAP) encoding genes (SAP1 to SAP10) were included on the arrays. All of the SAP genes with the exception of SAP4, SAP5 and SAP6 gave normalised ratios > 0.6. SAP5 gave an average ratio of 0.4 (among genes with intermediate homology) in all C. dubliniensis CD36 experiments. We probed the C. dubliniensis genome for homologues of SAP4-6 using PCR primers homologous to conserved regions of these genes (SAP4-6F/SAP4-6R, Table 2). Amplification using C. dubliniensis genomic DNA as template yielded a PCR product of 750 bp that shared 86% identity with SAP4 and SAP6. Using an inverse PCR strategy (primers InvSAPF/InvSAPR, Table 2) we amplified flanking sequences from C. dublineinsis genomic DNA to obtain the complete ORF (Accession no: AJ634382). This C. dubliniensis gene was found to lie upstream of the C. dubliniensis homologue of the C. albicans SAP1 gene and was equally homologous to SAP4 and SAP6 (~85%). This ORF was designated CdSAP4 as the synteny at this locus with SAP1 is identical to that at the SAP4 locus in C. albicans. We used this C. dubliniensis SAP4 gene as a probe in Southern blots with C. albicans and C. dubliniensis genomic DNA in order to identify SAP5 and SAP6 homologues in C.
The SAP4-6 genes in *C. albicans* are highly homologous (89% nucleotide sequence identity) so we anticipated that the *C. dubliniensis* SAP4 gene should hybridise strongly to any *C. dubliniensis* SAP5 or SAP6 homologues. Indeed, the *C. dubliniensis* SAP4 gene hybridised to four separate *Kpn*I fragments in *C. albicans* genomic DNA that correspond to SAP5, SAP6 and two alleles of SAP4 that could be differentiated on the basis of a restriction fragment length polymorphism (Fig. 6). However, hybridisation of the *C. dubliniensis* SAP4 gene to *C. dubliniensis* genomic DNA digested with *Kpn*I, *Hind*III and several restriction endonucleases that do not cleave within the *CdSAP4* ORF (*Bgl*II, *Spe*I, *Sal*I, *Xba*I) revealed only one significantly hybridising band in *C. dubliniensis* genomic DNA (Fig. 6). Furthermore, hybridisation of the *C. albicans* SAP5 and SAP6 genes to *C. dubliniensis* DNA resulted in hybridisation to the same restriction fragment harboring the *C. dubliniensis* SAP4 gene (data not shown). These findings were confirmed by Southern blot analysis on 8 epidemiologically unrelated isolates of *C. albicans* and *C. dubliniensis*, respectively (data not shown).

Hybridisation of a second *C. dubliniensis* strain to microarrays

In order to confirm the above data obtained with *C. dubliniensis* CD36 and to investigate the levels of intraspecies variation between unrelated *C. dubliniensis* strains, we hybridised genomic DNA from a second *C. dubliniensis* isolate, CD514, to these arrays. We chose this strain as it has been shown to be genetically unrelated to *C. dubliniensis* CD36 based on its DNA fingerprint pattern obtained with the *C. dubliniensis* fingerprint probe Cd25. Sheared genomic DNA from *C. dubliniensis* CD514 was co-hybridised with *C. albicans* SC5314 DNA to arrays. We compared the data set from CD514 with that generated from CD36 in order to identify genes unique to each strain. Only three additional genes were discovered that hybridised significantly to CD514 DNA (ratio > 0.59) that were deemed absent in CD36 (normalised ratio < 0.2, p value < 0.035). These genes were IPF4450 and IPF17652.3 with homology to an integrase and a reverse transcriptase, respectively and the oligopeptide transporter encoding gene *OPT1*. The presence of the *OPT1* sequence in CD514 and its absence in CD36 was confirmed by Southern hybridisation with the *C. albicans OPT1* sequence (Fig. 2). Conversely, only one sequence encoding *GIT1* (glycerophosphoinositol transporter)
was identified which failed to hybridise with CD514 DNA (ratio 0.112, p-value 0.008) and gave significant signals with CD36 DNA (ratio 1.1).

DISCUSSION

Phylogenetic analysis of rRNA sequences has confirmed that *C. dubliniensis* and *C. albicans* are the two most closely related *Candida* species of clinical importance in humans (Sullivan et al., 1995; Sullivan et al., 2004). However, whereas *C. albicans* is the most significant yeast pathogen responsible for superficial and deep seated infections, *C. dubliniensis* is of lesser clinical importance in mucosal infections in non-HIV-infected patients, and in the case of bloodstream infection is relatively insignificant (Kibbler et al., 2003; Meis et al., 2000). This apparent lower virulence of *C. dubliniensis* is also evident in data from animal model infection studies. However, the exact reasons why *C. albicans* is more virulent are not clear. In this study we have utilised recently available *C. albicans* whole genome DNA microarrays to investigate and identify genomic differences between *C. albicans* and *C. dubliniensis* that could account, at least in part, for the enhanced virulence potential of *C. albicans* relative to *C. dubliniensis* and for the differences in epidemiology between the two species.

The findings presented in this study obtained by CGH reinforce the phylogenetic data that originally inferred the close relatedness of the two organisms (Gilfillan et al., 1998; Sullivan et al., 1995). Our data show that only 4.25% of *C. albicans* sequences analysed in our studies (normalised ratio < 0.25) were likely to be absent or highly divergent (< 60% homologous at the nucleotide sequence level) in *C. dubliniensis*. That the vast majority of *C. albicans* genes are highly conserved in *C. dubliniensis* indicates that the two species have probably only diverged relatively recently and thus are likely to inhabit similar environments in the human body. Thus only a small subset of *C. albicans* genes seem to be unique to this species and are likely to be important contributory factors to the greater success of *C. albicans* as a commensal on human mucosal epithelium and as a pathogen in compromised hosts.
Of the 234 *C. albicans* genes identified predicted to have < 60% homology at the nucleotide sequence level or even possibly be absent in *C. dubliniensis*, 124 were hypothetical genes of unknown function (Table 3). However, 38 of these hypothetical ORFs were conserved with homology to genes in *Saccharomyces cerevisiae*, *Aspergillus nidulans* or paralogous sequences in the *C. albicans* genome. Of the 110 genes with a confirmed or hypothetical function, few were identified that corresponded to housekeeping genes involved in central metabolism, cell structure, or molecular biosynthesis. However, several transporter-encoding genes involved in nutrient uptake were identified as being absent, including two of the four genes encoding choline permeases in *C. albicans* (*HNM3*, *HNM4*) a uracil permease (*FUR4*) and an allantoin permease (*DAL52*). The only confirmed intrastrain difference between the two Cd25 fingerprint group *C. dubliniensis* strains analysed here was the presence of sequences homologous to the oligopeptide transporter *OPT1* in the Cd25 fingerprint group II strain CD514 which was absent in the fingerprint group I isolate CD36 (Lubkowitz *et al.*, 1997). It is not known whether absence of these genes could affect the ability of *C. dubliniensis* to grow relative to *C. albicans in vivo*, as for example our data suggest other genes encoding choline (*HNM2*) and allantoin permeases (*DAL51*) are likely to be present in *C. dubliniensis*. *Candida dubliniensis* also seems to be missing sequences involved in the biosynthesis of biotin (*BIO3* encoding DAPA aminotransferase and *BIO4* encoding dethiobiotin synthetase). Although biotin is required for growth, *C. albicans* and *C. dubliniensis* probably acquire sufficient biotin from exogenous sources in the oral cavity, most likely from commensal bacteria (Phalip *et al.*, 1999).

Twenty-two sequences corresponded to genes present in retrotransposons of *C. dubliniensis*, indicating that since their divergence the genomes of the two species may have acquired different mobile genetic elements.

Ten sequences homologous to genes encoding various GPI-anchored proteins were identified in our analysis as being absent or of low homology in *C. dubliniensis* by a combination of array hybridisation data, PCR analysis and Southern blot analysis (Sundstrum, 2002). Poor *C. dubliniensis* hybridisation signals were detected from sequences homologous to the *C. albicans* hyphal specific *HYRI* gene (average ratio 0.13), and no homologous gene was identified in *C. dubliniensis* following Southern hybridisation with conserved *C.
albicans HYRI sequences (Bailey et al., 1996). Sequences corresponding to several HYRI-related GPI-anchored proteins in C. albicans also exhibited poor hybridisation signals with C. dubliniensis genomic DNA (IFF family genes). Although specific functions have not been assigned to proteins encoded by these genes, their location on the cell surface indicates possible roles in maintaining cell wall integrity, environmental signaling or adhesion to host surfaces. Interestingly, subsequence analysis (Southern blotting) with sequences homologous to the C. albicans IFF1 gene (absent from Eurogentec microarrays) identified a homologous gene in C. dubliniensis for which sequences were later identified by PCR. These data suggest that at least one IFF-like gene is present in the C. dubliniensis genome. This may represent an ancestral IFF-related gene, however additional IFF-related genes may be present in the C. dubliniensis genome but may be difficult to detect by CGH as they could have diverged more extensively than essential housekeeping genes with greater sequence based constraints on protein function. A similar conclusion could be reached with regard to sequences homologous to genes encoding proteins of the α-agglutinin-like ALS family of adhesins. The ALS probes on the Eurogentec arrays used in this study consist of sequences from the 3′ region of the ORFs, which within the ALS family are the least conserved regions (Hoyer et al., 2001). By Southern hybridisation analysis Hoyer et al. noted that the 3′ regions of the C. albicans ALS genes are poorly conserved in C. dubliniensis (Hoyer et al., 2001). We observed low hybridisation ratios (< 0.25) for several members of this family including ALS1, ALS5, ALS6 and ALS7. However, Hoyer et al. identified partial 5′ nucleotide sequences for three ALS homologues in C. dubliniensis (ALSD1, ALSD2 and ALSD3). Their study demonstrated that ALSD1 is closely related to ALS6 and ALSD3 is closely related to ALS4. The present study confirms the findings of Hoyer et al. that the 3′ regions of the ALS genes are poorly conserved in C. dubliniensis, but does not provide further evidence for the existence of other C. dubliniensis ALS homologues. Since the microarray DNA spots correspond to 300-400 bp regions of each gene, our data reflect differences present in these regions only. As the CGH data obtained for the ALS gene family demonstrates, data indicating the absence or divergence of a particular gene requires confirmation as these regions may encompass non-conserved regions of the gene. Conversely, there may be divergent regions in many genes that remain undetected as they lie outside the regions compared here. Similarly, minor genetic differences (e.g. point mutations) and differences in non-translated regions that cannot be detected using
these methods could also influence virulence and epidemiology. In addition, phenotype can also be influenced by post-transcriptional events unrelated to DNA sequence.

Low hybridisation ratios were also observed for the HWP1 gene and the related sequences RBT1 and IPF14331 (Braun & Johnson, 1997; Staab et al., 1999; Sundstrum, 2002). The HWP1 encoded protein has been identified as a hypha-specific substrate for host transglutaminases involved in covalent adhesion to host cells. However, the functions of the other two gene products are as yet uncharacterised. In this study we identified the C. dubliniensis HWP1 homologue. The C. dubliniensis HWP1 gene hybridised poorly to the C. albicans array HWP1 sequences due to the presence of large deletions in the C. dubliniensis ORF. The predicted translated protein encoded by CdHWP1 contains several large deletions compared to the C. albicans protein (Fig. 5a). These deletions lie within the N-terminal glutamine- and proline-rich repeat domain containing the transglutaminase substrate activity and the internal serine and threonine-rich domain. It will be of interest to determine whether the transglutaminase substrate activity of the C. dubliniensis homologue is affected by the presence of deletions in glutamine rich regions of the N-terminus. A defect in the ability of C. dubliniensis to form stable attachments to oral epithelium may partly explain its reduced prevalence in the oral cavities of healthy individuals and patients with oral disease.

One of the most intensely studied virulence attributes of C. albicans is the ability to secrete aspartyl proteinases, encoded by 10 separate genes (Naglik et al., 2003). Sequences from all 10 SAP genes were present on the array. Sequences from only one of these genes, SAP5, gave consistently low signals from C. dubliniensis hybridising DNA. SAP5 is a member of the SAP4-6 subfamily of proteinases, which are all highly homologous at the nucleotide sequence level and preferentially expressed by hyphae (Hube et al., 1994). In our efforts to determine if SAP5 was present in C. dubliniensis, we identified a gene most homologous to SAP4 and SAP6, which we have designated CdSAP4, as the ORF was located upstream of CdSAPI, identical to the synteny observed in C. albicans (Miyasaki et al., 1994). Southern hybridisation analysis with this CdSAP4 sequence revealed that it could hybridise to multiple fragments of C. albicans restriction endonuclease-digested DNA corresponding to sequences of SAP4, SAP5 and SAP6. Such cross-
hybridisation is likely to be responsible for the strong signal detected from spots representing SAP6 on the C. albicans microarray. However, the CdsAP4 sequence consistently hybridised to only one single band (between ~2 and 10 kb) in Southern hybridisation experiments with C. dubliniensis genomic DNA. These data indicate that only one gene with strong homology to the SAP4-6 subfamily exists in C. dubliniensis. Attempts to identify the corresponding genomic loci of putative SAP5 and SAP6 genes in C. dubliniensis by low stringency PCR were unsuccessful (data not shown). Together, the SAP4-6 subfamily has been shown to play an important role in the establishment of C. albicans systemic infections in mice and SAP6 has been shown to be the most important gene within this family in the establishment of murine intraperitoneal infection (Felk et al., 2002; Sanglard et al., 1997). As C. dubliniensis only possesses one gene with homology to SAP4-6, C. dubliniensis might be expected to be less able than C. albicans to establish systemic infection. In vivo virulence studies and epidemiological data support this hypothesis, as C. dubliniensis is less virulent than C. albicans in a murine systemic model of infection and the incidence of recovery of this organism from human blood cultures is extremely low (Gilfillan et al., 1998; Kibbler et al., 2003). The absence of these hypha-specific proteinases in C. dubliniensis may affect the ability of its hyphae to penetrate host tissues, acquire nutrients or evade killing by macrophages. We are currently testing the role of C. dubliniensis Saps in infection models.

Differences in gene regulation have not been explored in any great detail in C. dubliniensis to date. The array CGH data indicates the presence of genes homologous to many of the transcription factors involved in regulating hypha formation in C. albicans (e.g. EFG1, CPH1, TUP1). However, poor hybridisation signals were detected from several other genes encoding putative transcriptional regulators that could affect regulatory circuits in C. dubliniensis. Seven of these sequences had homology to genes encoding proteins with Zn-finger DNA binding motifs in S. cerevisiae. Another group of genes with homology to the putative C. albicans transcriptional activator encoding gene CTA2 were also identified. CTA2 (GenBank ID AJ006637) was identified by Kaiser et al. in a one-hybrid screen in S. cerevisiae for C. albicans proteins with transcriptional activating properties (Kaiser et al., 1999). A family of possibly up to 10 genes with strong homology to CTA2 has been identified in C. albicans. Twelve sequences homologous to these genes
were included in our analysis and all gave normalised ratios < 0.25. Southern hybridisation also failed to identify any sequences with significant homology to these genes in *C. dubliniensis*. Although the function of these proteins has yet to be confirmed, the absence or divergence of a large family of transcriptional activators in *C. dubliniensis* could have major implications for the growth and virulence of this fungus.

In the present study, *C. albicans* DNA microarrays facilitated a whole genome comparison between *C. albicans* and its close relative *C. dubliniensis* in the absence of significant amounts of available *C. dubliniensis* sequence information. Our experiments have revealed the absence and divergence of several genes and gene families in *C. dubliniensis*. These include putative virulence factors and many genes specific or preferentially expressed in the hyphal phase such as *SAP5, SAP6, HWP1* and *HYR1*. *Candida dubliniensis* is generally less efficient than *C. albicans* at forming hyphae in response to serum and the absence of these hypha-regulated genes may also indicate that *C. dubliniensis* hyphae are less specialised as virulence promoting structures (Gilfillan *et al.*, 1998). We have endeavoured to confirm the absence or divergence of genes directly involved in virulence (e.g. *HWP1, SAP5*), however conclusive confirmation of this data will have to await the completion of the *C. dubliniensis* genome sequencing project. At present, this data set represents a framework for further investigations into genetic and phenotypic differences between *C. albicans* and *C. dubliniensis*. 

ACKNOWLEDGEMENTS

This study was supported by the Microbiology Research Unit, Dublin Dental School and Hospital.
Table 1. Percentage nucleotide sequence homology of C. dubliniensis genes to corresponding C. albicans Eurogentec microarray sequences

<table>
<thead>
<tr>
<th>Gene</th>
<th>% homology</th>
<th>Normalised ratio^</th>
<th>GenBank Accession no.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT1</td>
<td>98.8</td>
<td>2.19</td>
<td>AJ236897</td>
<td>(Donnelly et al., 1999)</td>
</tr>
<tr>
<td>URA3</td>
<td>93.0</td>
<td>2.05</td>
<td>AJ302032</td>
<td>(Staib et al., 2001)</td>
</tr>
<tr>
<td>MDRI</td>
<td>92.0</td>
<td>1.22</td>
<td>AJ227752</td>
<td>(Moran et al., 1998)</td>
</tr>
<tr>
<td>CDR1</td>
<td>91.9</td>
<td>1.16</td>
<td>AJ439073</td>
<td>(Moran et al., 2002)</td>
</tr>
<tr>
<td>CDR2</td>
<td>91.0</td>
<td>1.69</td>
<td>AJ439075</td>
<td>(Moran et al., 2002)</td>
</tr>
<tr>
<td>ERG3</td>
<td>90.1</td>
<td>0.92</td>
<td>AJ421248</td>
<td>(Pinjon et al., 2003)</td>
</tr>
<tr>
<td>ERG11</td>
<td>90.4</td>
<td>1.64</td>
<td>AY034876</td>
<td>(Perea et al., 2002)</td>
</tr>
<tr>
<td>PHR2</td>
<td>90.5</td>
<td>1.40</td>
<td>AF184908</td>
<td>(Kurzai et al., 1999)</td>
</tr>
<tr>
<td>PHR1</td>
<td>88.0</td>
<td>1.67</td>
<td>AF184907</td>
<td>(Kurzai et al., 1999)</td>
</tr>
<tr>
<td>SAP4</td>
<td>86.0</td>
<td>0.56</td>
<td>AJ634382</td>
<td>This study</td>
</tr>
<tr>
<td>SAP2</td>
<td>83.6</td>
<td>1.14</td>
<td>AJ634672</td>
<td>This study</td>
</tr>
<tr>
<td>CZF1</td>
<td>80.0</td>
<td>0.47</td>
<td>AJ634475</td>
<td>This study</td>
</tr>
<tr>
<td>IPF8147</td>
<td>79.6</td>
<td>0.29</td>
<td>AJ634476</td>
<td>This study</td>
</tr>
<tr>
<td>IPF1873</td>
<td>75.0</td>
<td>0.28</td>
<td>AJ634664</td>
<td>This study</td>
</tr>
<tr>
<td>RVS161</td>
<td>75.0</td>
<td>0.32</td>
<td>AJ634665</td>
<td>This study</td>
</tr>
<tr>
<td>IPF16104</td>
<td>74.0</td>
<td>0.24</td>
<td>AJ634666</td>
<td>This study</td>
</tr>
<tr>
<td>IFA8</td>
<td>70.4</td>
<td>0.37</td>
<td>AJ634667</td>
<td>This study</td>
</tr>
<tr>
<td>IPF9787</td>
<td>70.0</td>
<td>0.35</td>
<td>AJ634668</td>
<td>This study</td>
</tr>
<tr>
<td>IPF3448</td>
<td>69.6</td>
<td>0.42</td>
<td>AJ634669</td>
<td>This study</td>
</tr>
<tr>
<td>SSN6</td>
<td>68.0</td>
<td>0.38</td>
<td>AJ634670</td>
<td>This study</td>
</tr>
<tr>
<td>IPF2057</td>
<td>59.0</td>
<td>0.17</td>
<td>AJ634671</td>
<td>This study</td>
</tr>
</tbody>
</table>

^The average ratio of normalised fluorescence values of C. albicans and C. dubliniensis hybridising DNA at each gene specific spot
Table 2. Sequences of oligonucleotide primers used in this study.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5’ – 3’)</th>
<th>nucleotide coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRH12F</td>
<td>ACTGGAAATGGGAAACTGAAC</td>
<td>+1159 to 1178</td>
</tr>
<tr>
<td>CRH12R</td>
<td>CACAAACATCTGAGAAAGATG</td>
<td>+1476 to 1457</td>
</tr>
<tr>
<td>1760R</td>
<td>CAGAACTGTGTTAGACAC</td>
<td>+712 to +731</td>
</tr>
<tr>
<td>1760R</td>
<td>TCTATATCCATATCCAGATTCC</td>
<td>+1121 to +1102</td>
</tr>
<tr>
<td>11560F</td>
<td>CAATCTAAAACCTCCATAGG</td>
<td>+451 to +470</td>
</tr>
<tr>
<td>11560R</td>
<td>AGGGAAATTACTAACTGACTG</td>
<td>+821 to +812</td>
</tr>
<tr>
<td>8147F</td>
<td>CCCTACTACCTACATACG</td>
<td>+405 to +424</td>
</tr>
<tr>
<td>8147R</td>
<td>AATTAATTCCTCAAGATAC</td>
<td>+769 to +750</td>
</tr>
<tr>
<td>CHS5F</td>
<td>AGGAAACAGATATTGTGAG</td>
<td>+1055 to +1074</td>
</tr>
<tr>
<td>CHS5R</td>
<td>TCTGTAGATGTGGCTACAG</td>
<td>+1476 to +1457</td>
</tr>
<tr>
<td>IPF3448F</td>
<td>GGAGATTGTGGAGACCAAG</td>
<td>+492 to +511</td>
</tr>
<tr>
<td>IPF3448R</td>
<td>TGGCATTGTTCTCCACCAAC</td>
<td>+868 to +849</td>
</tr>
<tr>
<td>BIO3F</td>
<td>CAAGGAAATGCTGTATCTG</td>
<td>+791 to +809</td>
</tr>
<tr>
<td>BIO3R</td>
<td>ACCAAACACTCAATCGAC</td>
<td>+1187 to +1168</td>
</tr>
<tr>
<td>HIK1.3F</td>
<td>AAAAGTCTAACCAATTCAG</td>
<td>+443 to +462</td>
</tr>
<tr>
<td>HIK1.3R</td>
<td>TTCAGCATTGTGATGATCG</td>
<td>+843 to +824</td>
</tr>
<tr>
<td>HIK1.5F</td>
<td>AAAACGTTAGCCGTCAAAGC</td>
<td>+1753 to +1772</td>
</tr>
<tr>
<td>HIK1.5R</td>
<td>TCTAATATTAGATGGGAC</td>
<td>+2202 to +2183</td>
</tr>
<tr>
<td>HIT1F</td>
<td>TGTCTCAATATGTGTTAGC</td>
<td>+49 to +58</td>
</tr>
<tr>
<td>HIT1R</td>
<td>ATTCAATATGCAAGACATC</td>
<td>+444 to +425</td>
</tr>
<tr>
<td>HNM4F</td>
<td>TAAGTGGAGAACCATAATGG</td>
<td>+887 to +906</td>
</tr>
<tr>
<td>HNM4R</td>
<td>AAAGCACACCCAAAAGGACAG</td>
<td>+1350 to +1331</td>
</tr>
<tr>
<td>CZF1F</td>
<td>ATTCACACCTTGTATTCTG</td>
<td>+657 to +676</td>
</tr>
<tr>
<td>CZF1R</td>
<td>TTTGCTATCTCTCCTGATCC</td>
<td>+1087 to +1068</td>
</tr>
<tr>
<td>8627F</td>
<td>AATCAACACCTGTAGTAATCG</td>
<td>+1249 to +1268</td>
</tr>
<tr>
<td>8627R</td>
<td>TATGGATAAAATTACCATGAG</td>
<td>+1671 to +1652</td>
</tr>
<tr>
<td>15920F</td>
<td>CTACCAAACTAAACACAG</td>
<td>+1433 to +1452</td>
</tr>
<tr>
<td>15920R</td>
<td>ACTTTGGTGAATGAGTTAGG</td>
<td>+1875 to +1856</td>
</tr>
<tr>
<td>2057F</td>
<td>ACTACTGTCCTGCTGCTAC</td>
<td>+967 to +986</td>
</tr>
<tr>
<td>2057R</td>
<td>CCGTGTATAGCAGAAGTGC</td>
<td>+1385 to +1366</td>
</tr>
<tr>
<td>2971F</td>
<td>AATGGTTAAACAAAGGAGATC</td>
<td>+927 to +956</td>
</tr>
<tr>
<td>2971R</td>
<td>TATCTTTTCTCTGTTCAAAC</td>
<td>+1372 to +1352</td>
</tr>
<tr>
<td>DAL52F</td>
<td>TGTGGTAGAGTATATTCCC</td>
<td>+1119 to +1138</td>
</tr>
<tr>
<td>DAL52R</td>
<td>CATCTTCACTATCGTAGCT</td>
<td>+1558 to +1539</td>
</tr>
<tr>
<td>FUR4F</td>
<td>AGGGTTCATCTGCTATGAT</td>
<td>+1281 to +1300</td>
</tr>
<tr>
<td>FUR4R</td>
<td>GTACCATCTGACTTTCAT</td>
<td>+1662 to +1643</td>
</tr>
<tr>
<td>HNM3F</td>
<td>ATTTGACTGTTCTGTTTGG</td>
<td>+976 to +995</td>
</tr>
<tr>
<td>HNM3R</td>
<td>GATACATATACTGTATTGGT</td>
<td>+1409 to +1390</td>
</tr>
<tr>
<td>BIO4F</td>
<td>TGGACGCAATACTGGAACAG</td>
<td>+91 to +110</td>
</tr>
<tr>
<td>BIO4R</td>
<td>CACGGTCTCTAATGGCTCT</td>
<td>+467 to +448</td>
</tr>
<tr>
<td>OPT1F</td>
<td>GTGTAAGTTTTCTCTCATAAGT</td>
<td>+1843 to +1862</td>
</tr>
<tr>
<td>OPT1R</td>
<td>AGTCGGTGTTGTAAGACT</td>
<td>+2239 to +2220</td>
</tr>
<tr>
<td>IFF2F</td>
<td>ATGTGTTGCTGTAATGGCT</td>
<td>+3340 to +3359</td>
</tr>
<tr>
<td>IFF2R</td>
<td>CAAAGAAAACAAACACATTTG</td>
<td>+3728 to +3747</td>
</tr>
<tr>
<td>4805F</td>
<td>AAAACGTTACTATGTTTGGAG</td>
<td>+4069 to +4088</td>
</tr>
<tr>
<td>4805R</td>
<td>ATTTCAATATCCTAACATACC</td>
<td>+4418 to +4399</td>
</tr>
<tr>
<td>IPF257F</td>
<td>TACCAACCTATGGTTCAAC</td>
<td>+58 to +78</td>
</tr>
</tbody>
</table>
IPF257R  AATGACTGATTTGATCTAGTA  +402 to +383
1873F   GAATTGCTAAAGCAGATCGG  +433 to +452
1873R   AGATGACTGTTGTTAAATCAG  +886 to +869
RSV161F ACAATCGCTCTACTCGAATG  +246 to +265
RSV161R CAATCTGTTGGAATCTCGT  +690 to +671
RRN3F   CGCCGACATTCCAGGATCTGC  +1248 to +1267
RRN3R   CTATATATCGTCTCTACTCT  +1671 to +1651
13135F  TAATGGTTTGTGATGCAATG  +218 to +238
13135R  GATATGATCTGTTTGAATCGG  +590 to +571
9787F   AAGAACAGTTGGATTCAGAG  +989 to +1008
9787R   AATTGATCATTTCTTGGACG  +1349 to +1350
SSN6F   ACCAGTTAACCAACCTCCTT  +2817 to +2836
SSN6R   TCATGATTCTTTCTATCTCTC  +3233 to +3116
16104F  AACTCTAAAGATTTGAGT  +1741 to +1760
16104R  TCATACAGATTTCTACTTG  +2172 to +2153
IFA1F   GCGGACTCAGGATTATGATGC  +1279 to +1297
IFA1R   GCGGACTCAGGATTATGATGC  +1969 to +1951
HYR1F   GAATCTGTTGGAATCTCGT  +117 to +117
HYR1R   GAATCTGTTGGAATCTCGT  +1200 to +1183
IFF1F   GCGGACTCAGGATTATGATGC  +1 to +20
IFF1R   GCGGACTCAGGATTATGATGC  +844 to +826
SAP4-6F  GCGGACTCAGGATTATGATGC  +14 to +32
SAP4-6R  GCGGACTCAGGATTATGATGC  +745 to +763
InvSAPF GCGGACTCAGGATTATGATGC  +40 to +21
InvSAPR GCGGACTCAGGATTATGATGC  +623 to +642
CdSAP4F  GCGGACTCAGGATTATGATGC  -226 to -207
CdSAP4R  GCGGACTCAGGATTATGATGC  +1463 to +1444
RBTF2   GCGGACTCAGGATTATGATGC  +694 to +714
RBTR2   GCGGACTCAGGATTATGATGC  +1410 to +1390
HWPR2   GCGGACTCAGGATTATGATGC  -14 to +5
HWPR2   GCGGACTCAGGATTATGATGC  +1908 to +1889
APL6F   GCGGACTCAGGATTATGATGC  +223 to +210
CTA26F  GCGGACTCAGGATTATGATGC  +1 to +20
CTA26R  GCGGACTCAGGATTATGATGC  +781 to +751

* Primer names refer to gene annotations in CandidaDB (http://genolist.pasteur.fr/CandidaDB/)
✝ Nucleotide coordinates are given for each gene where +1 refers to the first base of the ATG start codon. All coordinates are for C. albicans genes with the exception of InvSAPF/R and CdSAP4/R which refer to the coordinates of the CdSAP4 gene and HWPR3 which refer to the CdHW1 gene.
Table 3. Functional categories of *C. albicans* genes predicted to be of low nucleotide sequence homology or absent in *C. dubliniensis* (normalised ratio < 0.25).

<table>
<thead>
<tr>
<th>Functional Category</th>
<th>Number of genes</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothetical genes</td>
<td>124</td>
<td>53.0%</td>
</tr>
<tr>
<td>Putative transcriptional regulator</td>
<td>21</td>
<td>9.0%</td>
</tr>
<tr>
<td>Retrotransposon elements</td>
<td>19</td>
<td>8.1%</td>
</tr>
<tr>
<td>Leucine-rich repeat family (<em>IFA</em>)</td>
<td>19</td>
<td>8.1%</td>
</tr>
<tr>
<td>GPI-anchored proteins</td>
<td>10</td>
<td>4.3%</td>
</tr>
<tr>
<td>Cell metabolism/biosynthesis</td>
<td>8</td>
<td>3.4%</td>
</tr>
<tr>
<td>Transporters</td>
<td>7</td>
<td>3.0%</td>
</tr>
<tr>
<td>Protein processing/modification</td>
<td>7</td>
<td>3.0%</td>
</tr>
<tr>
<td>Cell division and mating</td>
<td>5</td>
<td>2.1%</td>
</tr>
<tr>
<td>mRNA Processing</td>
<td>4</td>
<td>1.7%</td>
</tr>
<tr>
<td>Chromatin/DNA binding</td>
<td>3</td>
<td>1.3%</td>
</tr>
<tr>
<td>Cytoskeletal</td>
<td>3</td>
<td>1.3%</td>
</tr>
<tr>
<td>Morphogenesis related</td>
<td>2</td>
<td>0.8%</td>
</tr>
<tr>
<td>Mitochondrial</td>
<td>2</td>
<td>0.8%</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>234</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>
Table 4. *C. albicans* SC5314 genes predicted to be of low homology (< 60% nucleotide sequence identity) or absent in *C. dubliniensis* CD36

<table>
<thead>
<tr>
<th>Functional category</th>
<th>Putative or known function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unknown function (124)</strong></td>
<td></td>
</tr>
<tr>
<td>IPF14519</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF3468</td>
<td>Homology to IPF708</td>
</tr>
<tr>
<td>IPF2960.3f/IPF2960.5f</td>
<td>Contains DEAD helicase box</td>
</tr>
<tr>
<td>IPF17640</td>
<td>Homology to IPF15492</td>
</tr>
<tr>
<td>IPF417.3f</td>
<td>Homology to Sc YBR075w</td>
</tr>
<tr>
<td>IPF9794</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF7010.3</td>
<td>Homology to IPF324.3</td>
</tr>
<tr>
<td>IPF17417</td>
<td>Homology to IPF15492</td>
</tr>
<tr>
<td>IPF708</td>
<td>Homology to Sc YBR075w</td>
</tr>
<tr>
<td>IPF6387.3</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF12498.3f/IPF12498.53f</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF13810.3</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF5661</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF2702</td>
<td>Homology to Sc YBR074w</td>
</tr>
<tr>
<td>IPF17661</td>
<td>Homology to ScYBR075w</td>
</tr>
<tr>
<td>IPF17272</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF13072</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF16173.3f</td>
<td>Homology to IPA5</td>
</tr>
<tr>
<td>IPF3105</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF17488.3f/IPF17488.5f</td>
<td>Homology to IPF13810</td>
</tr>
<tr>
<td>IPF7940</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF6266</td>
<td>Homology to Sc YPR036w</td>
</tr>
<tr>
<td>IPF11508</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF14254</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF19766</td>
<td>No homology detected, dubious ORF</td>
</tr>
<tr>
<td>IPF11506</td>
<td>Homology to IPF17417</td>
</tr>
<tr>
<td>IPF15772</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF19377</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF14706</td>
<td>Homology to IPF13135</td>
</tr>
<tr>
<td>IPF10280</td>
<td>Homology to IPF3748</td>
</tr>
<tr>
<td>IPF7804.5f</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF19720.3eoc</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF4504</td>
<td>Homology to <em>Aspergillus nidulans</em> AN3284.2</td>
</tr>
<tr>
<td>IPF2815</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF6488</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF17131</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF9655</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF2754</td>
<td>Homology to Sc YER181c</td>
</tr>
<tr>
<td>IPF9401</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF4751</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF6325</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF13290</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF9400</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF17727.3/IPF17727</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF17322.3f</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF2195</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF11936.3f</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF17794</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF474</td>
<td>No homology detected</td>
</tr>
<tr>
<td>IPF2617</td>
<td>No homology detected</td>
</tr>
</tbody>
</table>
IPF17991  No homology detected
IPF12093  No homology detected
IPF14587.3  No homology detected
IPF20134  No homology detected
IPF11051  No homology detected
IPF12399  No homology detected
IPF324.3  Homology to IPF7010.3
IPF3444.5f  No homology detected
IPF635  No homology detected
IFB1  No homology detected
IPF18833  No homology detected
IPF10231.exon  No homology detected
IPF9211.5f  No homology detected
IPF15506  Homology to Sc YGR025w
IPF19812  No homology detected
IPF5373  No homology detected
IPF13724  Local homology to Sc Rsa1p
IPF8642  Homology to IPF10761
IPF243  No homology detected
IPF3301  No homology detected
IPF5730  Homology to Sc YNL211c
IPF7578  Homology to Aspergillus nidulans AN4487.2
IPF10168.3  No homology detected
IPF14618  No homology detected
IPF9057  No homology detected
IPF3233  No homology detected
IPF7539  No homology detected
IPF5978  No homology detected
IPF5217  No homology detected
IPF931  Homology to Sc YDR124w
IPF4880  No homology detected
IPF609  No homology detected
IPF3416  Local homology to Sc Sap30p, histone deacetylase
IPF15255  Local homology to Sc YEL007w
IPF13135  Homology to IPF13070
IPF2057  No homology detected
IPF14107  No homology detected
IPF11756  No homology detected
IPF16988  No homology detected
IPF14081  No homology detected
IPF15824  Local homology to Sc YKR023w
IPF7338  Homology to IPF13810
IPF12  No homology detected
IPF15335  No homology detected
IPF8741.5f  No homology detected
IPF16057  No homology detected
IPF8627  No homology detected
IPF1709  Homology to Sc Typp15 and A. nidulans ANO175.2
IPF16368.3f  No homology detected
IPF8942  Local homology to Sc Rim2p
IPF7848  No homology detected
IPF1742.3f.eoc  No homology detected
IPF19554.3f  Local homology to A. nidulans AN2129.2 and COP9 signal transduction domain
IPF13231  No homology detected
IPF19807  No homology detected
IPF16231  No homology detected
IPF17542  No homology detected
IPF2082  No homology detected
IPF2062  Local homology to C. albicans Cyr1p
IPF19542.5f
IPF7644
IPF15601
IPF5453
IPF17483
IPF9013
IPF14827
IPF3733
IPF11118
IPF9325
IPF6070
IPF13613
IPF19731
IPF2150
IPF10761
IPF6235
IPF17652.3
POL.3
POL0
Tca5a
IPF2535
IPF19295.3f
IPF4450
IPF13885.repeat/IPF13885
IPF14825
IPF9315
IPF4708
IPF16067
IPF9191.3f
IPF8612
IPF16104

**Retrotransposon encoded sequences (19)**

- **Zorro1b.5f/Zorro1b.3f**: Reverse transcriptase
- **Zorro2a.3f**: Reverse transcriptase
- **Zorro2b.3f/Zorro2b.5f**: Reverse transcriptase
- **Cirt**: Transposase
- **Cirt1a**: Transposase
- **Cirt2**: Transposase
- **Cirt3**: Transposase
- **Cirt4a**: Transposase
- **Cirt4b**: Transposase
- **IPF6235**: *Candida albicans* Tca2 retrotransposon
- **IPF17652.3**: Reverse transcriptase
- **POL.3**: Pol polyprotein, reverse transcriptase
- **POL0**: Pol part of pCal retrotransposon
- **Tca5a**: Polyprotein of Tca5 retrotransposon
- **IPF2535**: Homology to *Tca5* polyprotein (pol) gene
- **IPF19295.3f**: Homology to pol polyprotein *Arabidopsis thaliana*
- **IPF4450**: Polypeptide
- **IPF13885.repeat/IPF13885**: Homology to gag
- **IPF14825**: Homology to reverse transcriptases

**Putative transcriptional regulators (21)**

- **CTA20.exon2**: Homology to *C. albicans* CTA2
- **CTA21**: Homology to *C. albicans* CTA2
- **CTA22**: Homology to *C. albicans* CTA2
- **CTA24.1.exon1/CTA24.1.exon2**: Homology to *C. albicans* CTA2
- **CTA24.3/CTA24**: Homology to *C. albicans* CTA2
- **CTA2.5.3f/CTA25**: Homology to *C. albicans* CTA2
- **CTA26**: Homology to *C. albicans* CTA2
- **CTA27**: Homology to *C. albicans* CTA2
- **CTA29.exon2/CTA29.exon1**: Homology to *C. albicans* CTA2
- **SPT7**: Homology to Sc SPT7 transcription factor
- **RRN3**: Required for transcription of rDNA by RNA Polymerase I
- **IPF9315**: Homology to Sc HAP3 activator
- **IPF4708**: Involved in transcriptional elongation
- **IPF13021**: Zn finger protein, GAL4 domain, homology to Sc HAP1
- **IPF14255**: Zn finger protein, GAL4 domain
- **IPF15920**: Zn finger protein
- **IPF10533.exon1/IPF10533.exon2**: Zn finger protein, homology to AFLR in *A. nidulans*
- **IPF16067**: Zn finger protein, GAL4 domain, homologous to Sc HAL9
- **IPF9191.3f**: Zn finger protein, GAL4 domain, homologous to Sc HAL9
- **IPF8612**: Zn finger protein, homology to Sc *MGA1* activator
- **IPF16104**: Homology to Sc YJR119c, pfam matches to transcription factor domains
### Leucine-rich repeat family (19)
- IFA1: Leucine-rich repeat protein
- IFA2: Leucine-rich repeat protein
- IFA4: Leucine-rich repeat protein
- IFA5: Leucine-rich repeat protein
- IFA6: Leucine-rich repeat protein
- IFA7: Leucine-rich repeat protein
- IFA9: Leucine-rich repeat protein
- IFA10: Leucine-rich repeat protein
- IFA11: Leucine-rich repeat protein
- IFA12: Leucine-rich repeat protein
- IFA13: Leucine-rich repeat protein
- IFA15: Leucine-rich repeat protein
- IFA17.5f/IFA17.3f: Leucine-rich repeat protein
- IFA18.3: Leucine-rich repeat protein
- IFA19: Leucine-rich repeat protein
- IFA22: Leucine-rich repeat protein
- IFA24.3/IFA24.3: Leucine-rich repeat protein
- IFA25: Leucine-rich repeat protein
- IPF3540: *C. albicans* IFA family homologue

### GPI-Anchored (10)
- ALS1.3eoc: GPI-anchored protein, putative adhesin
- ALS5: GPI-anchored protein, putative adhesin
- ALS6: GPI-anchored protein, putative adhesin
- ALS7: GPI-anchored protein, putative adhesin
- ALS11.3f: GPI-anchored protein, putative adhesin
- HYR1.53/HYR1: GPI-anchored protein
- IFF2: GPI-anchored protein
- IFF4: GPI-anchored protein
- IFF8: GPI-anchored protein
- CRH12: GPI-anchored protein

### Central metabolism/biosynthesis (8)
- IPF19538: Putative isocitrate dehydrogenase
- IPF5239: Putative aldose reductase
- PPX1: Putative exopolyphosphatase
- BIO4: Dethiobiotin synthetase
- IPF4940: Putative isoamyl acetate esterase
- IFD2: Putative aryl alcohol dehydrogenase
- ADH3: Alcohol dehydrogenase
- CHS5: Chitin biosynthesis protein

### Protein trafficking/modification (7)
- IPF4710: Homology to *Sc* VTAl
- CTM1: Homology to *Sc* CTM1, cytochrome c methyltransferase
- IPF4195: Sc Ulp2p involved in ubiquitin protein degradation
- IPF6812: Homology to YLR224w, ubiquitin catabolism
- UBR11.3: Sc UBR1 homolog, ubiquitin metabolism
- PBN1: Homology to protease B
- IPF2997: Homology to *Sc* Reg1p, Protein phosphatase

### Putative membrane transporters (7)
- OPT1: Oligopeptide transporter
- DAL52: Putative allantoin permease
- IPF11550.3f/IPF11560.5f: Homology to *Ca*+ transporting ATPases
- FUR4: Uracil permease
- HNM3: Choline permease
- HNM4: Choline permease
- IPF1992: Homology to *Sc* AZR1, drug efflux pump
### mRNA processing (4)
- **IPF4706**: Homology to Sc Upf3p, nonsense mRNA decay
- **IPF1911**: Homology to Sc Syf2p, mRNA splicing
- **IPF6444**: Homology to Sc Tgs1p, RNA methyltransferase
- **CUS1**: Spliceosome associated protein

### Cytoskeletal (3)
- **IPF11222**: Homology to Dynein light chain proteins
- **IPF4032**: Homology to Sc Spc110p for microtubule component
- **YKE2.3**: Actin binding protein

### Chromatin/DNA binding (3)
- **IPF10490**: Homology to Sc ESC1
- **IPF4805**: Homology to Sc NF11/SIZ1
- **IPF670**: Homology to AHC1, histone acetyltransferase component

### Morphogenesis related (2)
- **IPF13247**: Homology to ECE1
- **IPF946**: EFG1 dependant transcript EDT1

### Mitochondrial proteins (2)
- **IPF5224**: Homology to Sc mitochondrial protein YHR083w
- **IPF11802**: Homology to Sc mitochondrial protein YDR332w

### Mating and cell division (5)
- **KAR5**: Nuclear fusion protein
- **IPF2589**: Homology to Sc Sog2p
- **IPF2971**: Homology to Sc BUR2 and *S. pombe* cyclin c homologue
- **IPF1760.3f/IPF1760.3f**: Homology to endochitinases
- **IPF1759.53f**: Homology to endochitinases

---

*Gene identifiers refer to those in the CandidaDB database (http://genolist.pasteur.fr/CandidaDB/)
†Gene functions as assigned in the CandidaDB database, except where significant homology was independently detected by searches of GenBank or *Saccharomyces* Genome Database (SGD). Sc indicates homology to *S. cerevisiae* genes*
REFERENCES


**Fig. 1.** Standard curve used to determine the relationship between percent nucleotide sequence homology of *C. albicans* SC5314 and *C. dubliensis* CD36 sequences and normalised fluorescence ratio. GenBank sequences for 11 *C. dubliensis* genes of known homology and 10 novel PCR amplified sequences were included. The average of the Log2 ratio values for each gene was plotted against percent nucleotide sequence homology. Linear regression analysis was used to predict the best fitting line. The slope value (y) and the coefficient of variance ($R^2$) were calculated using Prism 4.0.
Fig. 2. Southern hybridisation analysis of *C. albicans* and *C. dubliniensis* DNA with a DIG-11-dUTP labeled probe homologous to nucleotides +1 to +781 of CTA26. Lanes 1 and 3 contain *C. albicans* genomic DNA digested with *EcoRI* and *HindIII*, respectively. Lanes 2 and 4 contain *C. dubliniensis* CD36 genomic DNA digested with *EcoRI* and *HindIII*, respectively. Molecular size markers in kilobases (kb) are indicated on the left. Washes were performed at reduced stringency (60 °C in 0.5 x SSC).
Fig. 3

Southern hybridisation analysis of *C. albicans* and *C. dubliniensis* DNA with [α-P³²]dATP labeled probes corresponding to *C. albicans* microarray sequences of the genes (a) OPT1, (b) HNM3, (c) HNM4 and (d) FUR4. Each blot contains *Hind*III-digested genomic DNA from *C. albicans* SC5314 (lane 1), *C. dubliniensis* CD36 (lane 2) and *C. dubliniensis* CD514 (lane 3). Molecular size markers in kilobases (kb) are indicated on the left. Washes were performed at reduced stringency (60 °C in 0.5 x SSC).
Fig. 4

Southern hybridisation analysis of *C. albicans* and *C. dubliniensis* genomic DNA with sequences corresponding to *C. albicans* GPI-anchored protein encoding genes. Panel (a) was hybridised with an [α-\text{P}^32]dATP labeled probe corresponding to nucleotides +122 to +1027 of the *C. albicans HYR1* gene. Panel (b) was hybridised with an [α-\text{P}^32]dATP labeled probe corresponding to nucleotides +1 to +844 of *IFF1*. Lanes 1 and 2 in both panels contain EcoRI digested genomic DNA from *C. albicans* and *C. dubliniensis* respectively. Lanes 3 and 4 contain HindIII-digested genomic DNA from *C. albicans* and *C. dubliniensis* respectively. Molecular size markers in kilobases (kb) are indicated on the left. Washes were performed at reduced stringency (60 °C in 0.5 x SSC).
**Fig. 5**

(a) Diagram illustrating regions of homology to *C. albicans* CaHwp1p and the extent of deletions in the predicted CdHwp1p protein sequence. The upper rectangular box represents the CaHwp1p protein and shows the position of the KEX2 cleavage site (arrow), the recombinant rHwp1p domain (shaded area) shown to possess transglutaminase substrate activity (Sundstrum, 2002), the serine-threonine rich region (Ser-Thr rich) and the carboxy terminal ω-site. The lower boxes represent the homologous regions of the predicted *C. dublinsiens* CdHwp1p protein. The numbers below indicate the positions of the homologous *C. dublinsiens*
protein domains relative to the corresponding \textit{C. albicans} amino acid residues. (b) and (c) Southern hybridisation analysis of \textit{C. albicans} and \textit{C. dubliniensis} genomic DNA with sequences corresponding to (b) \textit{HWPI} and (c) \textit{RBT1}. DNA in (a) was hybridised with an [\(\alpha\)-\(P^{32}\)]dATP labeled probe corresponding to the entire \textit{C. dubliniensis HWPI} ORF. DNA in (b) was hybridised with an [\(\alpha\)-\(P^{32}\)]dATP labeled probe corresponding to nucleotides +694 to +1410 of \textit{RBT1} amplified from \textit{C. albicans} genomic DNA. Lanes 1 and 2 in both panels contain \textit{EcoRI}-digested genomic DNA from \textit{C. albicans} and \textit{C. dubliniensis}, respectively. Molecular size markers in kilobases (kb) are indicated on the left. Washes were performed at reduced stringency (60 °C in 0.5 x SSC).
Fig. 6  Southern hybridisation analysis of *C. albicans* and *C. dubliniensis* genomic DNA with sequences corresponding to the *C. dubliniensis* CdSAP4 gene. The blot was hybridised with an [α-32P]dATP labeled probe of the entire CdSAP4 ORF. Lane 1 contains *Kpn*I-digested genomic DNA from *C. albicans* SC5314. The markers on the left side of the panel indicate the predicted positions of the SAP4 (two alleles), SAP5 and SAP6 genes in SC5314. Lanes 2 to 7 contain genomic DNA from *C. dubliniensis* digested with *Kpn*I, *Bgl*II, *Hind*III, *Spe*I, *Sal*I and *Xba*I as indicated. Molecular size markers in kilobases (kb) are indicated on the right. Washes were performed at reduced stringency (60 °C in 0.5 x SSC).
**Fig. 1.** Standard curve used to determine the relationship between percent nucleotide sequence homology of *C. albicans* SC5314 and *C. dubliniensis* CD36 sequences and normalised fluorescence ratio. GenBank sequences for 11 *C. dubliniensis* genes of known homology and 10 novel PCR amplified sequences were included. The average of the Log₂ ratio values (Table 1) for each gene was plotted against percent nucleotide sequence homology. Linear regression analysis was used to predict the best fitting line. The coefficient of variance (R²) was calculated using Prism 4.0.
Fig. 2. Southern hybridisation analysis of *C. albicans* and *C. dubliniensis* DNA with a DIG-11-dUTP labeled probe homologous to nucleotides +1 to +781 of *CTA26*. Lanes 1 and 3 contain *C. albicans* genomic DNA digested with *Eco*RI and *Hind*III, respectively. Lanes 2 and 4 contain *C. dubliniensis* CD36 genomic DNA digested with *Eco*RI and *Hind*III, respectively. Molecular size markers in kilobases (kb) are indicated on the left. Washes were performed at reduced stringency (60 °C in 0.5 x SSC).
Fig. 3

Southern hybridisation analysis of *C. albicans* and *C. dubliniensis* DNA with [α-³²P]dATP-labeled probes corresponding to the complete ORF sequences of the *C. albicans* genes *OPT1*, *FUR4*, *HNM3* and *HNM4*. The ORFs were amplified from *C. albicans* genomic DNA with the primer sets OPTA/B, FUR4A/B, HNM3A/B and HNM4A/B (Table 2). Each blot contains EcoRI-digested genomic DNA from *C. albicans* SC5314, *C. dubliniensis* CD36 and *C. dubliniensis* CD514. Molecular size markers in kilobases (kb) are indicated on the left. Washes were performed at reduced stringency (60 °C in 0.5 x SSC).
**Fig. 4**

Southern hybridisation analysis of C. albicans and C. dubliniensis genomic DNA with sequences corresponding to highly conserved regions of C. albicans GPI-anchored protein encoding genes. Panel (a) was hybridised with an [α-32P]dATP-labeled probe corresponding to nucleotides +95 to +1183 of the C. albicans HYR1 gene. Panel (b) was hybridised with an [α-32P]dATP labeled probe corresponding to nucleotides +1 to +844 of IFF1. Lanes 1 and 2 in both panels contain EcoRI digested genomic DNA from C. albicans and C. dubliniensis respectively. Lanes 3 and 4 contain HindIII-digested genomic DNA from C. albicans and C. dubliniensis respectively. Molecular size markers in kilobases (kb) are indicated on the left. Washes were performed at low stringency (60 °C in 0.5 x SSC).
**Fig. 5**

(a) Diagram illustrating regions of homology to *C. albicans* CaHwp1p and the extent of deletions in the predicted CdHwp1p protein sequence. The upper rectangular box represents the CaHwp1p protein and shows the position of the KEX2 cleavage site (arrow), the recombinant rHwp1p domain (shaded area) shown to possess transglutaminase substrate activity (Sundstrum, 2002), the serine-threonine rich region (Ser-Thr rich) and the carboxy terminal ω-site. The lower boxes represent the homologous regions of the predicted *C. dubliniensis* CdHwp1p protein. The numbers below indicate the positions of the homologous *C. dubliniensis* residues.

(b) and (c) Gel blots of *C. dubliniensis* strains SC5314 and CD36. The blots show the expression of *CdHWP1* and *CaRBT1* genes.

*Fig. 5.* (a) Diagram illustrating regions of homology to *C. albicans* CaHwp1p and the extent of deletions in the predicted CdHwp1p protein sequence. The upper rectangular box represents the CaHwp1p protein and shows the position of the KEX2 cleavage site (arrow), the recombinant rHwp1p domain (shaded area) shown to possess transglutaminase substrate activity (Sundstrum, 2002), the serine-threonine rich region (Ser-Thr rich) and the carboxy terminal ω-site. The lower boxes represent the homologous regions of the predicted *C. dubliniensis* CdHwp1p protein. The numbers below indicate the positions of the homologous *C. dubliniensis* residues.
protein domains relative to the corresponding *C. albicans* amino acid residues. (b) and (c) Southern hybridisation analysis of *C. albicans* and *C. dubliniensis* genomic DNA with sequences corresponding to (b) *HWP1* and (c) *RBT1*. DNA in (a) was hybridised with an [α-32P]dATP-labeled probe corresponding to the entire *C. dubliniensis HWP1* ORF. DNA in (b) was hybridised with an [α-32P]dATP-labeled probe corresponding to nucleotides +694 to +1410 of *RBT1* amplified from *C. albicans* genomic DNA. Lanes 1 and 2 in both panels contain *Eco*RI-digested genomic DNA from *C. albicans* and *C. dubliniensis*, respectively. Molecular size markers in kilobases (kb) are indicated on the left. Washes were performed at reduced stringency (60 °C in 0.5 x SSC).
**Fig. 6** Southern hybridisation analysis of *C. albicans* and *C. dubliniensis* genomic DNA with sequences corresponding to the *C. dubliniensis* CdSAP4 gene. The blot was hybridised with an [α-P³²]dATP labeled probe of the entire CdSAP4 ORF. Lane 1 contains *Kpn*I-digested genomic DNA from *C. albicans* SC5314. The markers on the left side of the panel indicate the predicted positions of the SAP4, SAP5 and SAP6 genes in SC5314. Lanes 2 to 7 contain genomic DNA from *C. dubliniensis* digested with *Kpn*I, *Bgl*II, *Hind*III, *Spe*I, *Sal*I and *Xba*I as indicated. Molecular size markers in kilobases (kb) are indicated on the right. Washes were performed at reduced stringency (60 °C in 0.5 x SSC).