Copyright © 2001 by the Genetics Society of America

Characterization of Agglutinin-like Sequence Genes From Non-*albicans* Candida and Phylogenetic Analysis of the ALS Family

Lois L. Hoyer,* Ruth Fundyga,[†] Jennifer E. Hecht,* Johan C. Kapteyn,[‡] Frans M. Klis[‡] and Jonathan Arnold[†]

**Department of Veterinary Pathobiology, University of Illinois, Urbana, Illinois 61802, [†]Department of Genetics, University of Georgia, Athens, Georgia 30602 and [‡]Swammerdam Institute for Life Sciences, BioCentrum Amsterdam, University of Amersterdam, 1098 SM Amsterdam, The Netherlands*

Manuscript received September 4, 2000
Accepted for publication January 18, 2001

ABSTRACT

The ALS (*a*gglutinin-*l*ike *s*equence) gene family of *Candida albicans* encodes cell-surface glycoproteins implicated in adhesion of the organism to host surfaces. Southern blot analysis with ALS-specific probes suggested the presence of ALS gene families in *C. dubliniensis* and *C. tropicalis*; three partial ALS genes were isolated from each organism. Northern blot analysis demonstrated that mechanisms governing expression of ALS genes in *C. albicans* and *C. dubliniensis* are different. Western blots with an anti-Als serum showed that cross-reactive proteins are linked by β1,6-glucan in the cell wall of each non-*albicans* Candida, suggesting similar cell wall architecture and conserved processing of Als proteins in these organisms. Although an ALS family is present in each organism, phylogenetic analysis of the *C. albicans*, *C. dubliniensis*, and *C. tropicalis* ALS genes indicated that, within each species, sequence diversification is extensive and unique ALS sequences have arisen. Phylogenetic analysis of the ALS and SAP (*s*ecreted *a*spartyl *p*roteinase) families show that the ALS family is younger than the SAP family. ALS genes in *C. albicans*, *C. dubliniensis*, and *C. tropicalis* tend to be located on chromosomes that also encode genes from the SAP family, yet the two families have unexpectedly different evolutionary histories. Homologous recombination between the tandem repeat sequences present in ALS genes could explain the different histories for co-localized genes in a predominantly clonal organism like *C. albicans*.

CANDIDA albicans is an opportunistic pathogenic fungus that causes mucocutaneous and disseminated forms of disease. Two well-characterized gene families of *C. albicans* are believed to produce proteins that function in pathogenesis. The first characterized family, the SAP family, encodes secreted aspartyl proteinases (Hube *et al.* 1998). Disruption of SAP genes or inhibition of SAP gene products reduces pathogenicity of *C. albicans*, providing evidence for the role of aspartyl proteinases in the disease process (Sanglard *et al.* 1997; Hoegl *et al.* 1998; Borg-Von Zepelin *et al.* 1999; Cassone *et al.* 1999; Gruber *et al.* 1999; Schaller *et al.* 1999). The second large gene family in *C. albicans* is called ALS (*a*gglutinin-*l*ike *s*equence) due to the resemblance of domains of its encoded proteins to α-agglutinin, a cell-surface adhesion glycoprotein in *Saccharomyces cerevisiae* (Lipke *et al.* 1989; Hoyer *et al.* 1995). Presently, eight genes in the ALS family have been reported in the literature, although a small number of additional genes are found in the *C. albicans* genome (Hoyer *et al.* 1995, 1998a,b; Gaur and Klotz 1997; Hoyer and Hecht 2000, 2001). ALS genes conform to a basic three-domain structure that includes a relatively conserved 5′ domain of 1299 to 1308 nucleotides (433 to 436 amino acids), a central domain of variable length consisting entirely of a tandemly repeated 108-bp motif, and a 3′ domain of variable length and sequence that encodes a serine-threonine-rich protein (Hoyer *et al.* 1998b). Heterologous expression of ALS genes in *S. cerevisiae* confers an adherence phenotype on the organism, suggesting Als proteins function in adhesion to host surfaces, a property that is positively correlated with Candida pathogenesis (Calderone and Braun 1991; Gaur and Klotz 1997; Fu *et al.* 1998). In addition to the potential for Als proteins to function in pathogenesis, ALS genes are differentially expressed under a variety of conditions that include morphological form, growth medium composition, growth phase, and strain of *C. albicans*, similar to the SAP family (Hube *et al.* 1994; Hoyer *et al.* 1998a,b). Association of Als proteins with pathogenicity mechanisms and differential expression of ALS genes suggest that, similar to the SAPs, the ALS family is important in *C. albicans* pathogenesis.

If these gene families play an important role in *C. albicans* pathogenesis, it is possible that they also contribute to the pathogenicity of clinically relevant non-*albicans* Candida. Previous studies identified SAP genes in other Candida species including *C. dubliniensis*, *C. tropi-*

Corresponding author: Lois L. Hoyer, 2522 VMBSB, 2001 S. Lincoln Ave., Urbana, IL 61802. E-mail: lhoyer@uiuc.edu

calis, *C. parapsilosis*, and *C. guilliermondii* (Monod *et al.* 1994; Gilfillan *et al.* 1998). These organisms are among non-*albicans* Candida species that are isolated with increasing frequency from clinical specimens (Wingard *et al.* 1979; Wingard 1995; Fridkin and Jarvis 1996; Van't Wout 1996; Hoppe *et al.* 1997; Kunova *et al.* 1997; Sullivan and Coleman 1997; Weinberger *et al.* 1997; Darwazah *et al.* 1999; Rangel-Frausto *et al.* 1999). These observations led us to question whether an ALS gene family was also present in clinically important non-*albicans* Candida.

In this study, we present evidence on the DNA, RNA, and protein level that ALS genes exist as a family in *C. dubliniensis* and *C. tropicalis*. We isolate multiple ALS gene sequences from each organism using a PCR-based strategy and demonstrate that, although the basic structure of ALS genes is likely to be conserved in these organisms, there is little conservation of individual gene sequences across the different species. Using ALS and SAP gene probes, we also demonstrate that ALS and SAP genes are co-localized on the same chromosomes in each organism. Data from these studies demonstrate conservation of basic cell wall architecture between *C. albicans* and the non-*albicans* species and highlight significant differences in ALS gene expression patterns between the two most closely related organisms, *C. albicans* and *C. dubliniensis*. Finally, ALS gene sequence data from *C. albicans*, *C. dubliniensis*, and *C. tropicalis* are used to present a phylogenetic analysis of the ALS family. The data presented here indicate that the ALS family is younger than the SAP family. The presence of genes on the same chromosome with different evolutionary histories is expected under sexual recombination and provides indirect evidence that *C. albicans* has mated throughout its evolutionary past (Barton and Wilson 1996; Hull *et al.* 2000; Magee and Magee 2000).

MATERIALS AND METHODS

Candida strains: Multiple strains of each organism were used in initial Southern blotting studies to detect ALS genes. As studies progressed, two strains of each organism were chosen as representative of results and used in the figures in this article. Strains listed here include all those used in the study. *C. albicans* strain SC5314 was a gift from W. A. Fonzi; strain B311 was purchased from the American Type Culture Collection (ATCC; Manassas, VA), and strain 1177 was a gift from Stewart Scherer. *C. tropicalis* strains CAPG3 and T60700 were a gift from Patricia Kammeyer; strains 13803, 201380, and 201381 were purchased from ATCC. *C. dubliniensis* strains CD36 (type strain), CM1, and 16F were provided by David Coleman; strain LY261 was a gift from Richard Barton. *C. parapsilosis* strain SB was a gift from Carrie Frey; Patricia Kammeyer provided strains 44 and X36406. The identity of the *C. tropicalis* and *C. parapsilosis* strains was verified using either the API 20C AUX or API 32C system. Cellular morphology of each organism was examined following growth on corn meal-Tween agar plates (Remel, Lenexa, KS) and matched descriptions provided in standard sources (Larone 1995; Sullivan *et al.* 1995). Strains were maintained as glycerol stocks at −80° and streaked on YPD agar plates as needed.

ALS gene probes: All methods for making ALS gene-specific probes were published previously (Hoyer *et al.* 1995, 1998a,b; Hoyer and Hecht 2000, 2001); Table 1 summarizes these probes. To date, eight ALS genes were reported in the literature (Hoyer *et al.* 1995, 1998a,b; Gaur and Klotz 1997; Hoyer and Hecht 2000, 2001). The sequences of *ALS3* and *ALS8* are essentially identical and are detected by the same probe (Hoyer and Hecht 2000). To avoid redundancy, *ALS8* was omitted from certain figures in this article.

Nucleic acid gels and blotting: Protocols for genomic DNA extraction, running contour-clamped homogeneous electrical field (CHEF) gels, and Southern blotting were described previously (Hoyer *et al.* 1995, 1998a,b). All Southern blots were performed with the digoxigenin nonradioactive nucleic acid labeling and detection system (Roche Molecular Biochemicals, Indianapolis). Separation of *C. albicans* total RNA on formaldehyde gels and subsequent Northern blotting were described; detection of specific messages utilized radiolabeled DNA fragments (Hoyer *et al.* 1995, 1998a,b). Hybridization conditions for individual blots are included in the figure legends. ALS cross-hybridizing fragments were detected in Southern blots of *C. parapsilosis* genomic DNA after 40° hybridization and washing at 50° in 0.5× SSC/0.1% SDS.

Growth of Candida for Northern blot analysis: A single colony each of *C. dubliniensis* CD36 and *C. tropicalis* 13803 was inoculated into separate flasks of YPD (yeast extract, peptone, dextrose) medium and grown overnight (~16 hr) at 30° and 200 rpm shaking. Cells from each culture were counted and inoculated into a variety of growth media at a density of 5×10^6 cells/ml. Growth media included fresh YPD, RPMI 1640 (catalog no. 11875-085; Life Technologies, Rockville, MD), and Lee medium (Lee *et al.* 1975) adjusted to pH 4.5, 5.5, 6.5, or 7.5. Cultures were grown for various lengths of time ranging from 2 to 8 hr. Cells were harvested, washed in pyrocarbonic acid diethyl ester-treated sterile water, flash-frozen in an ethanol-dry ice bath, and stored at −80° until RNA was extracted.

Cell wall fractionation and analysis of cell wall proteins: Cells for cell wall fractionation were grown using the same conditions as for Northern analyses (see above). Methods for cell wall fractionation and protein analysis were previously described (Kapteyn *et al.* 2000). In brief, Als proteins were released by β1,6-glucanase digestion of isolated, SDS-extracted cell walls, separated by electrophoresis, and electrophoretically transferred onto polyvinylidene difluoride membranes. Membranes were treated for 30 min with 50 mm periodic acid, 100 mm sodium acetate (pH 4.5) to abolish any cross-reactivity of the serum to *N*- and *O*-linked glycan. Als proteins were visualized by treating the membranes with a polyclonal anti-Als antiserum raised by immunization of a New Zealand White rabbit with the purified N-terminal domain of *C. albicans* Als5p (Hoyer and Hecht 2001). A serum dilution of 1:5000 in phosphate-buffered saline (PBS), containing 5% (w/v) nonfat milk powder, was used. Binding of the anti-Als antiserum was assessed with goat anti-rabbit IgG peroxidase (Pierce Chemical Co.) at a dilution of 1:10,000 in PBS/5% (w/v) milk powder. The blots were developed using enhanced chemiluminescence (ECL) Western blotting detection reagents (Amersham Pharmacia Biotech). The anti-Als serum did not show any signal on a Western blot of *S. cerevisiae* β1,6-glucanase-released cell wall proteins (J. C. Kapteyn, unpublished data).

PCR amplification of ALS gene fragments from *C. dubliniensis* and *C. tropicalis*: Nucleotide sequences from the 5′ domain of *ALS1* through *ALS7* were aligned using the PILEUP program of the GCG sequence analysis package (Devereux *et al.* 1984) and regions of conserved sequence were defined.

These regions were used to design consensus oligonucleotide primers where degenerate bases were included in positions of ambiguity. The resulting primers were 5′ GCH ART SCN GGD GAY ACA TTY AYR TT 3′ (forward) and 5′ GGM AYA TCA AYR AHA ASA GTW GCW GTK YCH CC 3′ (reverse). PCR reactions including these primers used genomic DNA from *C. dubliniensis* strain CD36 and from *C. tropicalis* strain ATCC 13803, a 52° annealing temperature, and *Pfu* polymerase (Stratagene, La Jolla, CA). Each strain produced the predicted PCR product of ~1 kb. The products were cloned into pCRBlunt (Invitrogen, Carlsbad, CA) and transformed into *Escherichia coli* TOP10 (Invitrogen). Plasmid DNA from the resulting clones was analyzed by DNA sequencing. Open reading frames that resembled ALS sequences were given the accession nos. AF201685 (*ALSD1*), AF202529 (*ALSD2*), AF202530 (*ALSD3*), and AF201686 (*ALST1*). A second forward PCR primer was designed to amplify additional ALS-like sequences from *C. tropicalis* genomic DNA. The primer (5′ GCH GGT TAT CGW CCW TTT DTK GA 3′) was paired with the reverse primer above; amplification, cloning, and DNA sequencing followed the previous methods. DNA sequences isolated using this procedure were assigned accession nos. AF211865 (*ALST2*) and AF211866 (*ALST3*). All *C. dubliniensis* and *C. tropicalis* DNA sequences were translated with the alternate yeast genetic code tables because, like *C. albicans*, these species decode CUG as serine instead of leucine (Sugita and Nakase 1999).

***C. tropicalis* SAP gene probe:** Four SAP gene sequences from *C. tropicalis* have been reported in the GenBank database (accession nos. X61438, AF115320, AF115321, and AF115322). Coding regions from these sequences were aligned using the PILEUP program of GCG and consensus regions were identified. Primers were made to these regions using degenerate bases in positions of ambiguity. The resulting primers were 5′ GTT DTB RTW GAY ACY GGW TCH TCY GAT 3′ (forward) and 3′ CCD GTA TAY TTR GCA TKR TCA AYV CC 3′ (reverse). A consensus SAP gene probe of ~460 nucleotides was amplified from genomic DNA of strain ATCC 13803. This fragment was purified from an agarose gel and labeled by random priming using the Genius nonisotopic system (Roche Molecular Biochemicals). The resulting probe was hybridized to Southern blots at 65° and washed in 0.5× SSC/0.1% SDS at the same temperature.

Phylogeny analysis: The predicted amino acid sequences for *C. albicans*, *C. dubliniensis*, and *C. tropicalis* Als proteins were aligned using the PILEUP program of GCG software (Wisconsin Package Version 10, Genetics Computer Group, Madison, WI). While the full sequence of each gene is not known, the missing sequences are repetitive regions and would be excluded from phylogenetic analyses were they known (Orti *et al.* 1997). Three maximum parsimony trees were constructed using PAUP in GCG, with all characters assigned equal weights, branches added stepwise, and bootstrap values computed. Bootstrap values were assigned from heuristic searches for topologies found with exhaustive searches. Heuristic and exhaustive searches produced the same topology.

The first tree was found by an exhaustive search and included seven *C. albicans* sequences, three *C. dubliniensis* sequences, and *ALST1* from *C. tropicalis* and spanned amino acids 22 through 357. In this first tree, *ALST2* and *ALST3* were omitted to take advantage of the longer sequence available for the remaining genes. The second tree was found by an exhaustive search of the seven *C. albicans* amino acid sequences. Only the N-terminal and C-terminal domains were used in this analysis since the tandem repeat sequences may be phylogenetically misleading (Orti *et al.* 1997). The third tree was found by a branch and bound search and included seven sequences from *C. albicans*, three from *C. dubliniensis*, and three from *C. tropicalis*, which spanned amino acids 231 to 357. The first and third trees were rooted with *ALST1* because previous rDNA studies have placed *C. tropicalis* outside of *C. albicans* and *C. dubliniensis* (Barns *et al.* 1991; Gilfillan *et al.* 1998). The *C. albicans*-only tree was rooted with *ALS7*, as determined by the first tree. The maximum parsimony tree of SAP sequences included those studied by Monod *et al.* (1998) and the three additional SAP sequences reported in GenBank (listed above). SAP sequences were aligned using PILEUP and a heuristic search conducted using *S. cerevisiae YAP3* as the root (Barns *et al.* 1991). In addition, the distance-based tree building method UPGMA was used to cross-validate inferred topology of the trees generated (Hillis *et al.* 1996).

RESULTS

Detection of ALS gene sequences in non-*albicans* Candida species by Southern blotting: Southern blots of genomic DNA from several non-*albicans* Candida species were hybridized with various ALS-specific fragments (Table 1). Because they hybridize to multiple ALS genes in *C. albicans*, fragments derived from the *ALS1* and *ALS5* tandem repeat domains were used first (Hoyer *et al.* 1995, 1998a,b; Table 1). At high stringency, the *ALS1* repeats and *ALS5* repeats probes largely differentiate between subfamilies of the ALS genes in *C. albicans*: the *ALS1* repeats probe hybridizes to *ALS1*, *ALS2*, *ALS3*, *ALS4*, and *ALS8* while the *ALS5* repeats probe minimally recognizes *ALS5*, *ALS6*, and *ALS7* (Hoyer *et al.* 1998b; Hoyer and Hecht 2000). Each probe recognized multiple genomic fragments in strains of *C. dubliniensis* and *C. tropicalis*, although the hybridization signals in *C. tropicalis* were weaker (Figure 1). Decreasing the hybridization stringency increased the intensity of the *C. tropicalis* signals (data not shown); however, efforts were made to screen at higher stringencies to avoid potentially misleading nonspecific results. These initial results indicated the presence of ALS-like tandem repeat fragments in the genomes of *C. dubliniensis* and *C. tropicalis*.

The 5′ domain of *C. albicans* ALS genes is conserved, showing 55–90% identity among known sequences (Hoyer and Hecht 2000). To determine if ALS 5′ domain sequences were also present in non-*albicans* Candida, the genomic Southern blot described above was stripped and reprobed with a *Kpn*I-*Hpa*I fragment derived from the 5′ domain of *ALS1* (Table 1). This probe recognizes multiple fragments in the *C. albicans* genome that are largely the same as the fragments that hybridize with the *ALS1* tandem repeat probe (Hoyer *et al.* 1998a; Figure 1). These results suggested that the 5′ domain and tandem repeat domain were found on the same genomic fragment in many cases. A similar result was achieved for *C. dubliniensis* DNA, but signals were not observed at higher stringency for *C. tropicalis* DNA (Figure 1).

In addition to a conserved 5′ domain followed by a domain of tandem repeats, *C. albicans* ALS genes encode a 3′ domain sequence that is variable in length and

TABLE 1

Hybridization probes derived from *C. albicans* ALS sequences

Probe name	Probe type	*C. albicans* gene(s) detected	GenBank accession no.	ALS domain detected	5′ coordinate or forward primer[a]	3′ coordinate or reverse primer[a]	Reference
ALS1 repeats	Restriction fragment	*ALS1*, *ALS2*, *ALS3* *ALS4*, *ALS8*	L25902	Tandem repeats	nt 1378	nt 2247	Hoyer *et al.* (1995)
ALS5 repeats	PCR product	*ALS5*, *ALS6*, *ALS7* and possibly others	AF189016	Tandem repeats	5′ GGT ACA AGT TCC ACT GCC AAA 3′	5′ AAG ACA GTT CTT CCA ATG GAT CA 3′	Hoyer *et al.* (1998b)
*Kpn*1-*Hpa*I	Restriction fragment	*ALS1*, *ALS2*, *ALS3* *ALS4*, *ALS5*, *ALS8*, and possibly others	L25902	5′ domain	nt 460	nt 858	Hoyer *et al.* (1995)
*Xba*I-*Hin*dIII	Restriction fragment	*ALS1*	L25902	3′ domain	nt 2611	nt 3491	Hoyer *et al.* (1995)
ALS2/ALS4	PCR product	*ALS2*, *ALS4*	AF024581	3′ domain	5′ TCC GAG TCC ATT CCA GTA CTA A 3′	5′ GTT ACA GCA TCA CTA GAA GGA ATA TC 3′	Hoyer *et al.* (1998b)
ALS3	PCR product	*ALS3*, *ALS8*	U87956	3′ domain	5′ ATG ACA CCA TGT CAA GTT CAG A 3′	5′ GTT GAC TCA ATG TAG TTT GTT G 3′	Hoyer *et al.* (1998a)
ALS5/ALS6	PCR product	*ALS5*, *ALS6*	AF068866	3′ domain	5′ CAA TTC CTT CTC CAC CTA CTT CAA C 3′	5′ GCA GCA CTG TCT CCA TTC ATC 3′	Hoyer and Hecht (2001)
ALS7 VASES repeats	Restriction fragment	*ALS7*	AF201684	3′ domain	nt 4099	nt 4576	Hoyer and Hecht (2000)

[a] Nucleotide (nt) numbers marking restriction sites within the sequence or PCR primers required to amplify the probes are shown.

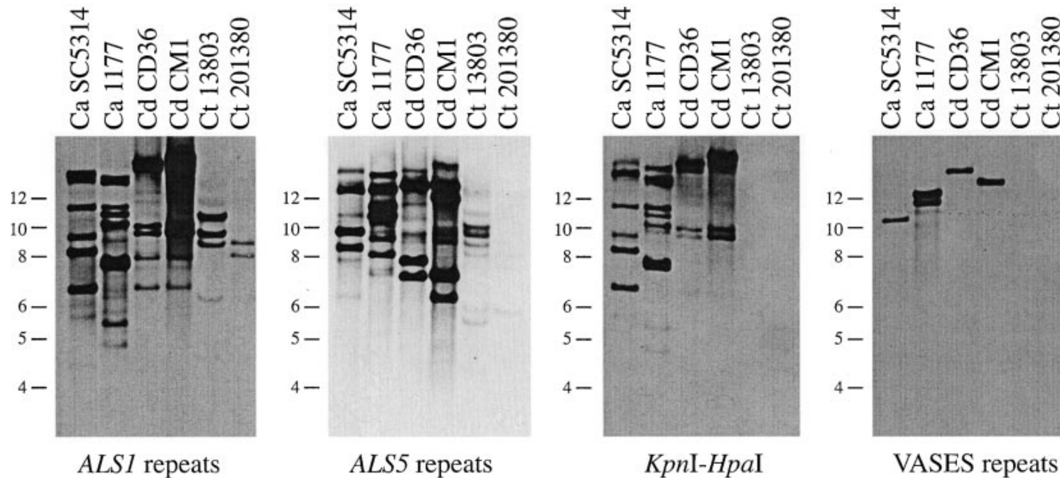

Figure 1.—Southern blots of *Bgl*II-digested genomic DNA from *C. albicans* (Ca), *C. dubliniensis* (Cd), and *C. tropicalis* (Ct) strains. The Southern blot was probed with the indicated fragments derived from *C. albicans* ALS genes. The blots were hybridized at 50° and washed at 60° in 0.5× SSC/0.1% SDS. Molecular size markers (in kilobases) are indicated at the left of each blot.

sequence among the known genes (Hoyer *et al.* 1998b; Hoyer and Hecht 2000). To determine whether ALS genes in non-*albicans* species were homologous to certain *C. albicans* ALS genes, we hybridized blots of genomic DNA with fragments derived from the 3′ end of *ALS1*, *ALS3*, *ALS2/ALS4*, *ALS5/ALS6*, and *ALS7* (Table 1). Even at lowered stringencies, no 3′-domain-derived probes gave any signal on blots of *C. dubliniensis* or *C. tropicalis* DNA, with the exception of *ALS7* (Figure 1). Results presented here suggest that there are a similar number of ALS genes in *C. albicans* and *C. dubliniensis*. While ALS genes are likely to be present in *C. tropicalis*, they are likely to be fewer in number and less related in sequence to *C. albicans* ALS genes. Finally, the juxtaposition of the 5′ domain and tandem repeat domain of the ALS fragments in *C. dubliniensis* suggests that these genes have a similar three-domain structure as ALS genes in *C. albicans*, but have unique 3′ sequences.

Northern blotting with ALS-specific probes: Northern blot analysis was pursued to confirm that the ALS-hybridizing sequences detected on Southern blots encoded expressed genes. *C. dubliniensis* CD36 and *C. tropicalis* ATCC 13803 cells for RNA extraction were grown under a variety of conditions as described in materials and methods above. For many growth conditions, *C. dubliniensis* showed multiple bands that cross-hybridized with the *ALS1* tandem repeats probe (Figure 2). Hybridization of *C. tropicalis* RNA with the same probe failed to show strong signals with the exception of a high-molecular-weight band observed for RPMI-grown cells (data not shown). Lack of signals on *C. tropicalis* Northern blots may be due to difficulties in specific detection with *C. albicans*-derived sequences noted on Southern blots above and suggested that other means were needed to demonstrate a gene family in this organism.

Cell wall analysis of *C. dubliniensis* and *C. tropicalis*: For cell wall analysis, *C. dubliniensis* CD36 and *C. tropicalis* 13803 were grown in YPD and RPMI media, respectively, since these conditions yielded the best signals on Northern blots described above. Western blot analysis of SDS-PAGE-separated, β1,6-glucanase-released cell wall proteins with an antiserum raised against the N-terminal domain of Als5p revealed diffuse bands at ∼470 kD (Figure 3). These apparent molecular sizes were similar to those observed in the analysis of Als proteins in the *C. albicans* cell wall (Kapteyn *et al.* 2000). These data suggested that, similar to *C. albicans*, both *C. dubliniensis* and *C. tropicalis* had Als proteins, which were incorporated in their cell wall through linkage to β1,6-glucan.

Isolation of non-*albicans* Candida ALS sequences by PCR with consensus primers: Southern and Western blot data presented above suggested that DNA encoding

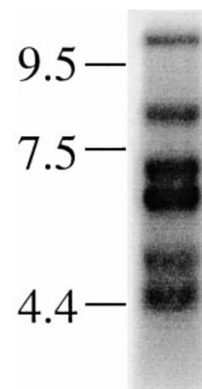

Figure 2.—Northern blot of *C. dubliniensis* total RNA probed with the *ALS1* tandem repeats fragment. RNA was isolated from strain CD36 grown overnight in YPD medium at 30° and 200 rpm shaking. Identical signals were also observed for CD36 grown in fresh YPD medium (30°) and RPMI medium (37°) for 2 and 5 hr and in Lee medium (37°) at four different pH values for 8 hr. Molecular size markers (in kilobases) are shown at the left of the blot.

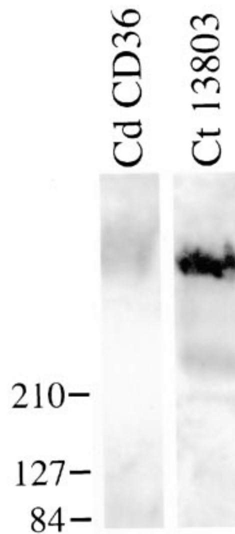

Figure 3.—Western blot of *C. dubliniensis* CD36 and *C. tropicalis* 13803 cell wall extracts with an anti-Als serum.

the 5′ end of ALS genes was conserved across the three species studied. Alignment of the 5′ domains of all known *C. albicans* ALS genes showed regions of sequence identity that could be used to design consensus PCR primers. Two forward primers and one reverse primer were selected from the aligned sequences. The combination of the first forward primer with the reverse primer predicted a PCR product of ~1 kb; using the second forward primer predicted a 370-bp product. Amplification of genomic DNA from *C. dubliniensis* CD36 and *C. tropicalis* 13803 using the first primer set yielded PCR products of the expected size. Cloning of these products and DNA sequencing of selected clones revealed three distinct *C. dubliniensis* clones and one *C. tropicalis* clone with an open reading frame similar to the 5′ end of *C. albicans* ALS genes. Because the newly isolated gene fragments did not directly correspond to known *C. albicans* ALS genes, nomenclature for the genes followed that in use for SAP genes in non-*albicans* Candida (Monod *et al.* 1998): *C. dubliniensis* genes were designated *ALSD1*, *ALSD2*, and *ALSD3*; the *C. tropicalis* sequence was named *ALST1*. Despite sequencing many clones from *C. tropicalis*, no additional ALS-like coding regions were isolated. Cloning and sequencing of fragments isolated from amplification of *C. tropicalis* DNA with the second PCR primer pair revealed two new open reading frames (*ALST2* and *ALST3*) that resembled ALS genes. Alignment of amino acid sequences from *C. albicans*, *C. dubliniensis*, and *C. tropicalis* Als proteins showed regions of conservation present in each; of particular note were the eight Cys residues, which were conserved in every sequence with the exception of the last Cys residue, which was missing from Alst2p (Figure 4). Comparison of the new ALS sequence fragments to corresponding regions of *C. albicans* ALS genes showed a lower degree of identity between *C. tropicalis* and *C. albicans* sequences (53 to 63% at the nucleotide level and 42 to 59% at the amino acid level), consistent with the weak hybridization signals observed in Northern and Southern blotting. Finding multiple ALS-like coding regions in *C. dubliniensis* and *C. tropicalis* suggested that ALS genes existed as a family in these organisms.

Chromosomal co-localization of ALS and SAP family sequences in *C. dubliniensis* and *C. tropicalis*: Isolation of three *C. dubliniensis* and three *C. tropicalis* ALS sequences suggested the presence of an ALS gene family in each organism. The presence of SAP-like DNA sequences in *C. dubliniensis* was shown by cross-hybridization on genomic Southern blots and on CHEF gels (Gilfillan *et al.* 1998). A SAP family in *C. tropicalis* was postulated from genomic Southern blots (Monod *et al.* 1994) and substantiated by the presence of multiple *C. tropicalis* SAP gene sequences in the GenBank database. Previous work in *C. albicans* showed that ALS and SAP genes are located mainly on chromosomes 3, 6, and R (Monod *et al.* 1994, 1998; Hoyer *et al.* 1998a; Hoyer and Hecht 2000). To determine if this conservation of localization was also true for *C. dubliniensis* and *C. tropicalis*, Southern blots of CHEF-separated chromosomes were probed with SAP and ALS sequences.

Separation of *C. dubliniensis* chromosomes on a CHEF agarose gel showed a wide variability in karyotype between strains CD36 and CM1 (Figure 5, left). The presence of multiple *C. dubliniensis* strains with a karyotype similar to *C. albicans* has been demonstrated (Gilfillan *et al.* 1998), although wide variation in karyotype for commonly studied *C. dubliniensis* isolates has also been shown (Magee *et al.* 1999). The *ALS1* repeats and *ALS5* repeats fragments hybridized to *C. dubliniensis* chromosomes the size of 3 and 6 and, in CM1, fragments that were likely derived from these chromosomes (Figure 5). These results matched data presented for *C. dubliniensis* where SAP probes hybridized to chromosomes the size of 3 and 6 and a fragment of similar size to R (Gilfillan *et al.* 1998). Lack of signals for chromosome R with ALS probes indicated either that ALS sequences are not found on this chromosome or that ALS sequences present on chromosome R are sufficiently dissimilar in sequence that they cannot be detected with *C. albicans*-derived probes by high-stringency Southern hybridization.

Limited references are available for the karyotype of *C. tropicalis*, but published information and experimentation with CHEF running conditions indicated that *C. tropicalis* chromosomes were separable with the same running conditions used for separating the largest *C. albicans* chromosomes (Mahrous *et al.* 1992; Figure 5). Using these conditions, seven distinct chromosomal bands were separated and numbered from 1 to 7 (largest to smallest). Hybridization of *C. tropicalis* chromosomes with ALS sequences was done at high stringency to avoid potentially nonspecific hybridization and misleading signals. With this procedure the same two chromosomes

```
  Alsd1p  ~~~~~~~~~~ ~~~~~~~~~~ ~ANAGDTFTL IMPCVFKFIT TQTSVDLTAN GVKYATCTFH AGEDFTAFSS MSCVVNNGLT SNIKAFGTVR IPISFNVGGT GSSVNLQDSK
   Als6p  YGGPGYPTWT AVLGWSLDGT LASPGDTFTL VMPCVFKFIT TQTSVDLTAN GVKYATCTFH AGEDFTTFSS MSCVVNNGLS SNIRAFGTVR LPISFNVGGT GSSVNIQDSK
   Als5p  FKGPGYPTWN AVLGWSLDGT SANPGDTFIL NMPCVFKFTA SQKSVDLTAD GVKYATCQFY SGEEFTTFST LTCTVNDALK SSIKAFGTVT LPIAFNVGGT GSSTDLEDSK
   Als1p  FKGPGYPTWN AVLGWSLDGT SANPGDTFTL NMPCVFKYTT SQTSVDLTAD GVKYATCQFY SGEEFTTFST LTCTVNDALK SSIKAFGTVT LPIAFNVGGT GSSTDLEDSK
   Als3p  YKGPGTPTWN AVLGWSLDGT SASPGDTFTL NMPCVFKFTT SQTSVDLTAH GVKYATCQFQ AGEEFMTFST LTCTVSNTLT PSIKALGTVT LPLAFNVGGT GSSVDLEDSK
  Alsd2p  ~~~~~~~~~~ ~~~~~~~~~~ ~ASPGDTFTL IMPCVFKFIT TQTSVDLTAN GVKYATCSFY SGEEFTTFSS LTCTVNSALT SSVKAFGTVT LPISFNVGGT GSSVDLEDSK
 Als2-1p  YKGPNRPTWN AVLGWSLDGT SANPGDTFTL NMPCVFKFIT DQTSVDLTAE GVKYATCQFY SGEEFTTFSS LKCTVSNTLT SSIKALGTVT LPISFNVGGT GSSVDLESSQ
  Alsd3p  ~~~~~~~~~~ ~~~~~~~~~~ ~ANAGDTFTL NMPCVFKFIT DQTSVDLVAD GRTYATCDLY SGEEFTTFSS LKCTVSNALN SQTKALGTVT LPLSFNIGGS GSDVDITSSQ
 Als4-1p  YRGPATPTWT AVIGWSLDGA TASAGDTFTL DMPCVFKFIT DQTSIDLVAD GRTYATCNLN SAEEFTTFSS VSCTVTTTMT ADTKAIGTVT LPFSFSVGGS GSDVDLANSQ
  Alst1p  ~~GISMKRVA VSPANVNVSP AANAGDTFTL IMPCVFKFTT SETSIDLTVG SKSYATCNFN AGEHFTTFSS LSCTVTQSVP DNTNAYGTIT VPLAFNVGGS GRDVDLTDAK
   Als7p  YEEISTLTAN AQLEWALDGT IASPGDTFTL VMPCVYKFMT YETSVQLTAN SIAYATCDFD AGEDTKSFSS LKCTVTDELT EDTSVFGSVI LPIAFNVGGS GSKSTITDSK
    Cons  .......... .......... .A..GDTF.L .MPCV.K... ...S..L... ...YATC... ..E....FS. ..C.V..... ......G... .P..F..GG. G.........

  Alsd1p  CFTAGTNSVT FTDGDHKISI PVDFPKTPES SSGLIKYSRV IPTLDKLSSL AVASQCTAGY KSGVLGFSAT KNDVTIECSN VHVGITNGLN SWNMPVSSDS FSYTKTCTSS
   Als6p  CFTAGTNTVT FTDGDHKIST TVNFPKTPQS SSSLVYFARV IPSLDKLSSL VVASQCTAGY ASGVLGFSAT KDDVTIDCST IHVGITNGLN SWNMPVSSES FSYTKTCTPN
   Als5p  CFTAGINTVT FNDGSKKLSI AVNFEKSTVD RSGYLTTSRF MPSLNKIATL YVAPQCENGY TSGTMGFSTS YGDVAIDCSN VHIGISKGVN DWNHPVTSES FSYTKSCSSF
   Als1p  CFTAGTNTVT FNDGDKDISI DVEFEKSTVD PSAYLYASRV MPSLNKVTTL FVAPQCENGY TSGTMGFSSS NGDVAIDCSN IHIGITKGLN DWNYPVSSES FSYTKTCTSN
   Als3p  CFTAGTNTVT FNDGGKKISI NVDFERSNVD PKGYLTDSRV IPSLNKVSTL FVAPQCANGY TSGTMGFANT YGDVQIDCSN IHVGITKGLN DWNYPVSSES FSYTKTCSSN
  Alsd2p  CFTAGTNTVT FTDGDNKVST TVDFEKSTVD STGYLTSSRL MPSLNKVTSL FVAPQCARGY TSGTIGFSSS NGGVSFDCSN VHVGITNGVN DWNFPVSSES FSYTKTCSST
 Als2-1p  CFKAGTNTVT FNDGDKKISI DVDFEKTNED ASGYFIASRL IPSINKVSIT YVAPQCANGY TSGAMGFIVL TGDTTIDCSN VHVGITKGLN DWNFPVSSDS LSYNKTCSST
  Alsd3p  CFKEGTNTVT FNDGDTTFST TANFQRSDVN ANDRILLSRI LPSLAKSVTI FIPPRCASGY SSGTMGFSTA GTDAIIDCST VHAGISNGLN DWNYPIESKS FSYTTSCSSK
 Als4-1p  CFTAGINTVT FNDGDTSIST TVDFEKSTVA SSDRILLSRI LPSLSQAVNL FLPQECANGY TSGTMGFSTA GTGATIDCST VHVGISNGLN DWNYPISSES FSYTKTCTST
  Alst1p  CFTTGDNTVT FSDGDKSFST TANFEGAGTL NDDY.ESSRL IPSLGKTDAL LVAPLCSNGY KSGTIGFSST TKGFSIDCNN IQAGITSQLN AWGFPTDSQS FSYTTQCTTT
   Als7p  CFSSGYNTVT FFDGNNQLST TANFLPRREL AFGLVVSQRL SMSLDTMTNF VMSTPCFMGY QSGKLGFTSN DDDFEIDCSS IHVGITNEIN DWSMPVSSVP FDHTIRCTSR
    Cons  CF..G.N.VT F.DG....S. ...F...... ........R. .......... .....C..GY .SG..GF... .......C.. ...GI....N .W..P..S.. ......C...

  Alsd1p  SFIITYENVP AGYRPFIDTY VKKTSTTSTG FNLNYTNSYV CTDGKKGNDP LIYFWTS.YT NSDAGSDGAV VIVTTKTVTD STTAITTLPF DPTVDKTKTI EVLEPIPTTT
   Als6p  SFIITYENVP AGYRPFIDSY VKKSATATNG FNLNYTNIYN CMDGKKGNDP LIYFWTS.YT NSDAGSNGAA VVVTTRTVTD STTAITTLPF DPTVDKTKTI EVIEPIPTTT
   Als5p  GISITYQNVP AGYRPFIDAY I..SPSDNNQ YQLSYKNDYT CVDDYWQHAP FTLKWTG.YK NSDAGSNGIV IVATTRTVTD STTAVTTLPF NPSVDKTKTI EILQPIPTTT
   Als1p  GIQIKYQNVP AGYRPFIDAY I..SATDVNQ YTLAYTNDYT CAGSRSQSKP FTLRWTG.YK NSDAGSNGIV IVATTRTVTD STTAVTTLPF NPSVDKTKTI EILQPIPTTT
   Als3p  GIFITYKNVP AGYRPFVDAY I..SATDVNS YTLSYANEYT CAGGYWQRAP FTLRWTG.YR NSDAGSNGIV IVATTRTVTD STTAVTTLPF DPNRDKTKTI EILKPIPTTT
  Alsd2p  GITITYRNVP AGYRPFIDAY I..TASDVNS YTLSYTNDYT CVGGSVQHKP FTLRWSG.YK NGEAGSNGIV IVATTRTVTD STTAVTTLPF NSDVDKTKTI EILQPIPTTT
 Als2-1p  GISITYENVP AGYRPFFDVY T..SVSGQNR .QLRYTNDYA CVGSSLQSKP FNLRLRG.YN NSEANSNGFV IVATTRTVTD STTAVTTLPF NPSVDKTKTI EILQPIPTTT
  Alsd3p  GVSVTYQNVP AGYRPFVDAY I..SA..LTS YTMQYTNQYT CVGARPVDAS FSYNWLG.YD NAEAGSRGIT IVVTTSTVTD STTAVTTLPF NSDVDRTKTI AVLQPIPTTT
 Als4-1p  SVLVTFQNVP AGYRPFVDAY I..SATRVSS YTMQYTNIYA CVGAASVDDS FTHTWRG.YS NSQAGSNGIT IVVTTRTVTD STTAVTTLPF NSDTDKTKTI EILQPIPTTT
  Alst3p  ~~~~~~~~~~ AGYRPFFAAL VQAPSSD... YAIQYTAKYR CEGSVQRDDS QKISWAG.YT NSDPDSNGAV VVLTTITGTQ SNTIVTTLPF NPTADHTKTI EVIVPIPTVT
  Alst1p  SYSITFSTIP KGLRPFIDAY IKAPTST... YPMTYTYKYV CSDGKSYNGN TKLNWSG.YV NSDADSEGME IVVATTTGTG STTGVTTLPF DKTKDKTKTI QVIEPIPTTT
  Alst2p  ~~~~~~~~~~ AGYRPFVDVL FSHTASDI.. FTMLYTNEYV GADGVTYDAS MKKAWKS.YQ DSLPSGDGAI IIVTTRTGTQ STTAVSTLPY DPEIDLTKTI EVLVPIPTTT
   Als7p  ALYIEFKTIP AGYRPFVDAI VQIPTTE..P FFVKYTNEFA CVNGIYTSIP FTSFFSQPIL YDEALAIGAD LVRTTSTVIG SITRTTTLPF ISRLQKTKTI LVLEPIPTTT
    Cons  .........P .G.RPF.... .......... ....Y..... .......... .......... .......G.. ....T.T... S.T...TLP. ......TKTI ....PIPT.T

  Alsd1p  ITTSYVGVST SFFTKTATIG ETATLIIDV~
   Als6p  ITTSYVGIST SLSTKTATIG GTATVVVDVP
   Als5p  ITTSYVGVTT SYSTKTAPIG ETATLIVDVP
   Als1p  ITTSYVGVTT SYSTKTAPIG ETATVIVDVP
   Als3p  ITTSYVGVTT SYSTKTAPIG ETATVIVDIP
  Alsd2p  ITTLYVGVTT SYSTKTAPIG DTATLFIDI~
 Als2-1p  ITTSYVGVTT SYSTKTAPIG ETATVIVDVP
  Alsd3p  ITTSYVGVTT SYRTQTVPIG ETATLFI~~~
 Als4-1p  ITTSYVGVTT SYSTKTAPIG ETATVIVDVP
  Alst3p  TTTSYIGVTT SYTTITGTIG DTATLVIDM~
  Alst1p  VTTSYLGVTT SFSTITATIG ETATLVI~~~
  Alst2p  TTTSYLGVST YYSTITATIG DTATLVIDM~
   Als7p  VTTSHHGFDT WYYTKKATIG DTATVFIDVP
    Cons  .TT...G..T ...T....IG .TAT......
```

FIGURE 4.—Amino acid sequence alignment of predicted Als proteins corresponding to the PCR-amplified region. Amino acid sequences of Als proteins from *C. albicans* were aligned with those predicted from the PCR-amplified *C. dubliniensis* and *C. tropicalis* sequences. A consensus sequence is provided. The positions of conserved and semiconserved Cys residues are double-underlined in the consensus sequence. Amplification of the original clones with the Pfu proofreading polymerase and double-stranded sequencing of each fragment suggested that lack of the last Cys residue in Alst2p was not due to a PCR-induced or DNA sequencing error. Because their gene fragments were amplified with the second primer pair and yielded a shorter product, amino acid sequences of Alst2p and Alst3p do not begin until the third sequence block. Gaps in the alignment are denoted by periods; a tilde (~) is used to indicate sequence information that lies outside of the PCR-amplified region.

were detected with each probe; however, hybridization signals were weak (Figure 5). The main question we sought to answer with this experiment was whether SAP and ALS sequences hybridized to the same chromosome. Since no SAP chromosomal localization data have been published for *C. tropicalis*, we constructed a consensus SAP probe by PCR using degenerate oligonucleotide primers designed from alignment of the *C. tropicalis* SAP sequences available in GenBank (see MATERIALS AND METHODS). Validation that this probe recognized *C. tropicalis* SAP genes was done by Southern blot of *Eco*RI-digested genomic DNA (Figure 6A). A similar blot probed with a consensus SAP oligo was reported by MONOD *et al.* (1994). The fragments observed in our blot were all present in the published blot, demonstrating that our probe detected multiple *C. tropicalis* SAP genes. Hybridization of our SAP probe to CHEF-separated *C. tropicalis* chromosomes detected four chromosomal bands (Figure 6B). Hybridization of each *C. tropicalis* ALS gene to *C. tropicalis* chromosomes showed that ALS genes were located only on chromosomes where SAP genes are also found (Figure 6B). Interestingly, the *ALST3* sequence hybridized to two chromosomes. Whether these chromosomes are homologous or whether there is more than one *ALST3*-like gene in *C. tropicalis* remains to be determined. Although characterization of the gene families in *C. tropicalis* is not complete, these initial data support the conclusion that ALS and SAP genes are found mainly on the same chromosomes in a variety of Candida species.

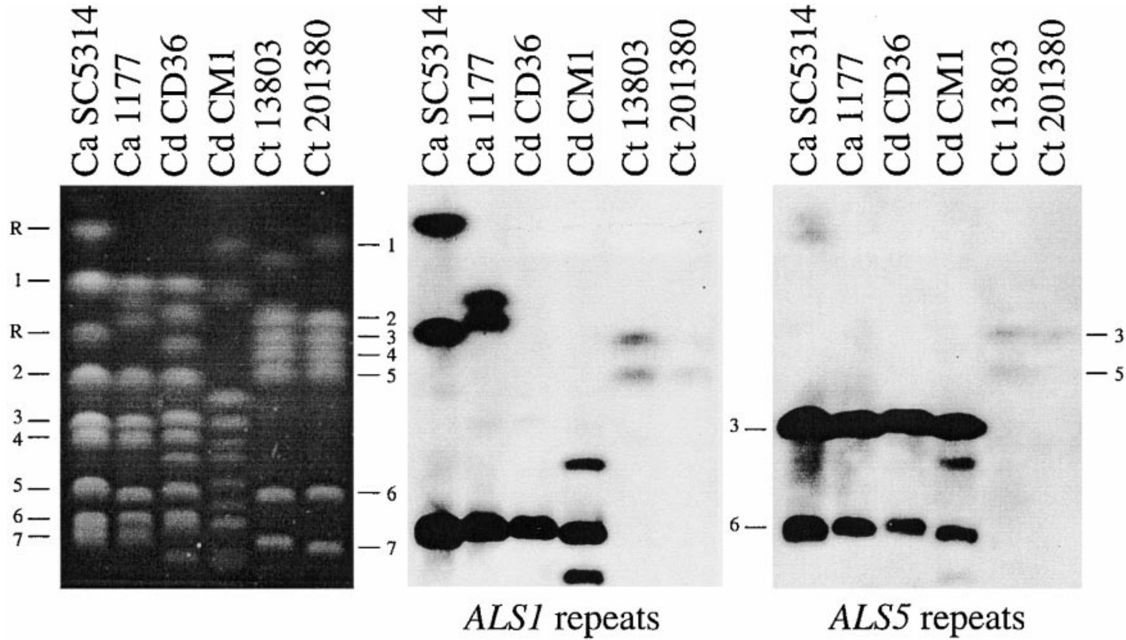

FIGURE 5.—Chromosomal localization of ALS genes in *C. albicans* (Ca), *C. dubliniensis* (Cd), and *C. tropicalis* (Ct) strains as defined by hybridization with *C. albicans* ALS repeats probes. Chromosomes of two strains each of *C. albicans*, *C. dubliniensis*, and *C. tropicalis* were separated on a CHEF gel and stained with ethidium bromide (left). Subsequently, the gel was Southern blotted and probed with the *ALS1* repeats (middle) and *ALS5* repeats (right). Blots were hybridized at 65° and washed in 0.5× SSC/0.1% SDS at the same temperature. *C. albicans* chromosomes were numbered as previously indicated (WICKES *et al.* 1991). A numbering system has not been defined for *C. dubliniensis* chromosomes; *C. tropicalis* chromosomes were numbered from 1 to 7 (largest to smallest). The smudge at the upper left corner of the *ALS5* repeats blot was not aligned with a single lane and is interpreted as a blotting artifact.

Molecular evolution of the ALS family: Phylogeny analysis of the ALS family was conducted to determine the oldest gene (most basal lineage) in the ALS family, to compare the rate of evolution of the ALS family to that of the SAP family, and to understand how the history of the ALS family compares to the history of *C. albicans*. To accomplish these goals, we constructed three maximum parsimony trees from amino acid se-

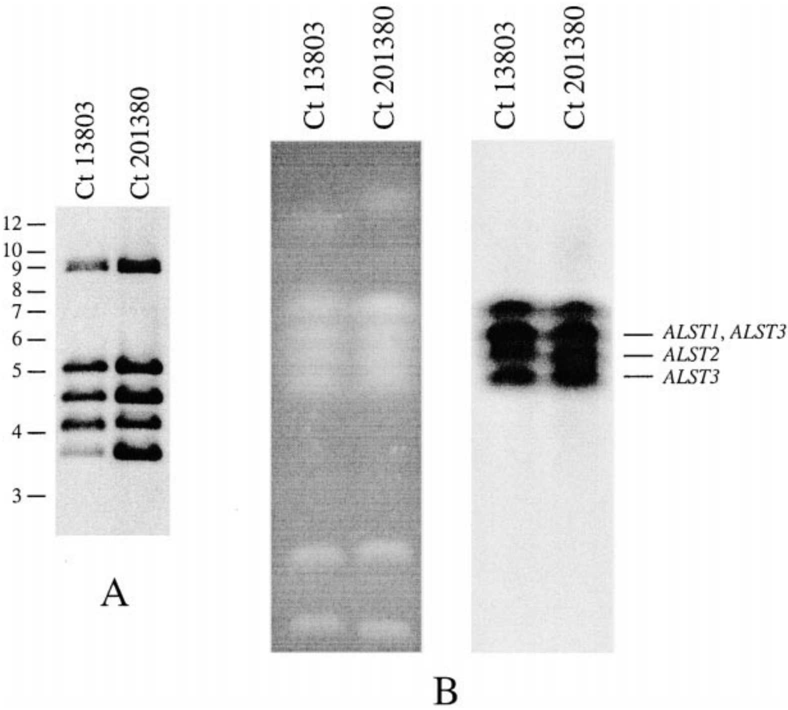

FIGURE 6.—Co-localization of ALS and SAP genes on *C. tropicalis* chromosomes. The PCR-amplified SAP consensus probe was hybridized to Southern-blotted, *Eco*RI-digested genomic DNA from two *C. tropicalis* strains (A) and also to a Southern blot of CHEF-separated chromosomes (B, right side). The chromosomes to which each *C. tropicalis* ALS gene fragment hybridized are indicated at the right of the blot. All blots were hybridized at 65° and washed in 0.5× SSC/0.1% SDS at the same temperature. Molecular size markers (in kilobases) are shown at the left of the genomic Southern blot in A.

quences. UPGMA trees were also constructed with similar results (data not shown). The first tree was constructed to better resolve the *C. albicans* and *C. dubliniensis* family structure, specifying *ALST1* as the root (Figure 7A). This phylogram was based on 336 characters of which 95 were constant and 156 were informative. The tree length was 708 with a consistency index of 0.8107 and a homoplasy index of 0.1893, including uninformative characters. *C. tropicalis* acted as an outgroup in the otherwise unrooted parsimony tree and placed *ALS7* as the most basal lineage within *C. albicans*; *ALS6* appeared to be the second most basal. The *ALS4* and *ALS6* ancestors existed before the *C. albicans* and *C. dubliniensis* split as each appeared to have a sister gene in *C. dubliniensis*. Because nearly all of the ALS genes from *C. albicans* have been characterized (Hoyer and Hecht 2000), we concluded that *ALSD2* probably arose within *C. dubliniensis* as it lacked a corresponding *C. albicans* gene. Concluding that *ALS1*, *ALS2*, *ALS3*, and *ALS5* arose within *C. albicans* after the split from *C. dubliniensis* assumed that none of the other ALS sequences that were likely to exist in *C. dubliniensis* group with these genes.

A second maximum parsimony tree was constructed solely from *C. albicans* sequences to determine the relative age of the genes in the family (Figure 7B). In this phylogram, there were 1016 characters of which 239 were constant and 410 were informative. The tree length was 1533, the consistency index was 0.9328, and the homoplasy index was 0.0672. This phylogram specified *ALS7* as the root, based on results from the first tree. This analysis confirmed the conclusion that *ALS6* was the second most-basal *C. albicans* lineage and that *ALS2* and *ALS4* were the youngest, most rapidly evolving genes in the family.

The third phylogram, with all sequences from *C. albicans*, *C. dubliniensis*, and *C. tropicalis*, was based on 127 total characters of which 33 were constant and 63 were informative (Figure 7C). Tree length was 340 with a consistency index of 0.76 and a homoplasy index of 0.24, including uninformative characters. Surprisingly, even though the SAP and ALS sequences examined co-localized in all three species, none of the known *C. tropicalis* ALS sequences grouped with *C. albicans* sequences as occurred for the SAP family (see below). This result was unexpected since genes located on the same chromosome in predominantly clonal organisms are expected to have the same evolutionary history (Barton and Wilson 1996). As observed in the third tree (Figure 7C), the pairs of *ALS6* with *ALSD1* and *ALS4* with *ALSD3* indicated that family structure was detected in a close relative of *C. albicans*.

A maximum parsimony phylogram of all reported SAP sequences was found by a heuristic search in order to include additional *C. tropicalis* sequences not found in the most recently published SAP tree (Monod *et al.* 1998, Figure 7D). This new tree compared favorably with that reported by Monod *et al.* (1998). The SAP tree was based on 703 characters of which 140 were constant and 318 were informative. Tree length was 2408 with a consistency index of 0.75 and a homoplasy index of 0.25. The tree was rooted with the *S. cerevisiae* sequence *YAP3*. Unlike the ALS trees, the SAP sequences from *C. tropicalis* did not form a basal group separate from *C. albicans*. The *C. tropicalis* sequences instead grouped more distally, after branching off of the *C. albicans SAP9* and *SAP7* sequences.

ALS genes in other Candida species: Because of the emerging similarities between the ALS and SAP gene families, species in which SAP genes have been documented are obvious ones to examine for the presence of ALS genes. An example of such an organism is *C. parapsilosis* (Monod *et al.* 1994). Low-stringency hybridization of *Bgl*II-digested *C. parapsilosis* genomic DNA with the *ALS1* repeats fragment showed a single cross-hybridizing fragment of ∼16 kb (data not shown). PCR amplification of *C. parapsilosis* DNA using both sets of ALS consensus primers yielded fragments of the predicted length; however, DNA sequencing of cloned fragments yielded sequences that did not resemble known ALS genes. Additional experimentation is required to define the nature of the *C. parapsilosis* cross-hybridizing fragment and to determine if this organism encodes ALS genes. Other species of Candida remain to be tested.

DISCUSSION

DNA, RNA, and protein evidence reported here demonstrate that ALS gene families are found in *C. dubliniensis* and *C. tropicalis*. PCR screening procedures yielded the sequences of three ALS genes from each organism. These sequences revealed that the ALS family in the non-*albicans* species is not identical to that in *C. albicans*. Chromosomal analysis of each organism indicated that the ALS and SAP gene sequences are largely co-localized. Phylogenetic analysis of the ALS family suggests that the ALS family has a different evolutionary history from the SAP family as expected for an organism with an evolutionary history of mating (Hull *et al.* 2000; Magee and Magee 2000).

Comparative biology of Candida species: Analysis of the ALS family in *C. dubliniensis* and *C. tropicalis* yielded new insights about two different biological processes in these organisms. First, release of Als proteins from the cell walls of *C. dubliniensis* and *C. tropicalis* with β1,6-glucanase suggests that Als proteins in the non-*albicans* species encode the correct signals for cell wall localization and that the basic wall structure of these organisms is similar to that of *S. cerevisiae* and *C. albicans* (Kapteyn *et al.* 1999, 2000; Smits *et al.* 1999). Second, studies of ALS genes in *C. dubliniensis* suggest differences in regulation of the gene family and in production of cell wall proteins. The multiplicity of similarly expressed ALS-hybridizing messages on Northerns of *C. dubliniensis* total RNA contrast sharply with the appearance of similar ALS Northerns for *C. albicans* in two ways: a seemingly increased number of expressed genes in *C. dubliniensis* and the apparent constitutive nature of the gene expression (Hoyer *et al.* 1995, 1998a,b).

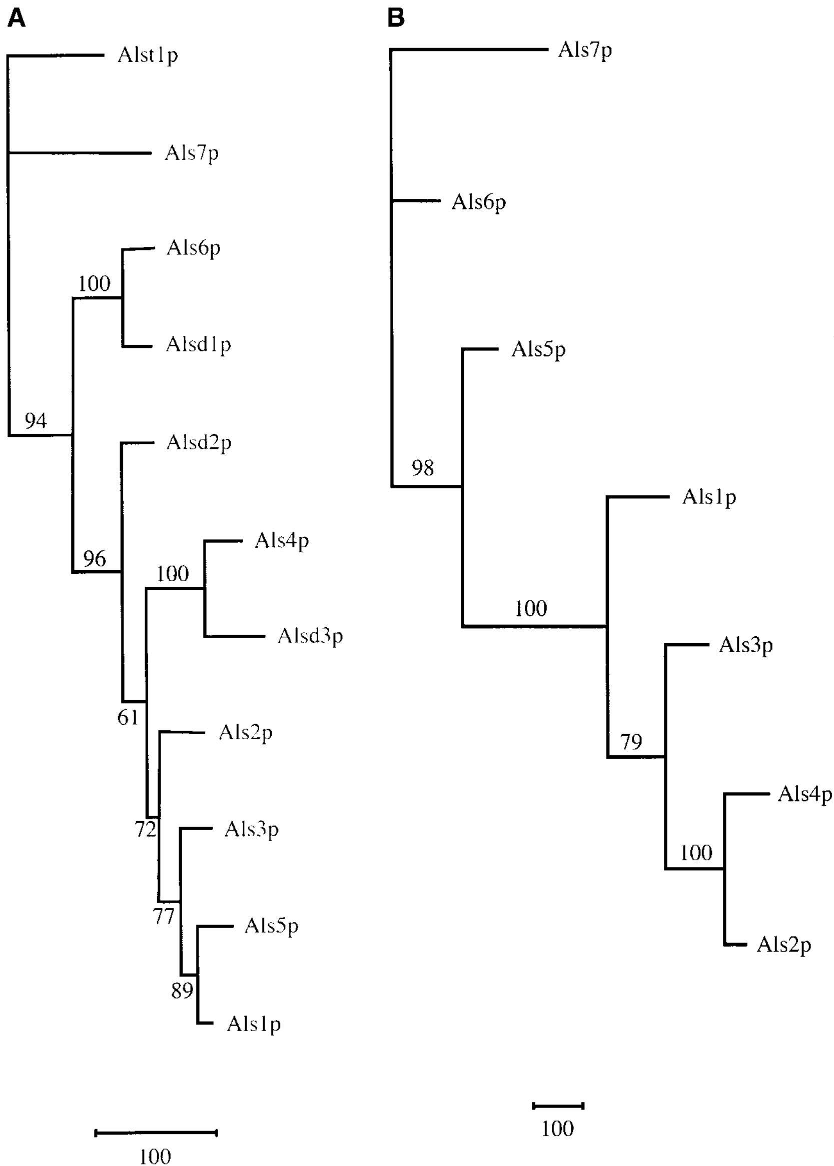

Figure 7.—Phylogenetic trees. Maximum parsimony trees were constructed from Als and Sap amino acid sequences. Bars indicate the branch length that corresponds to 10 or 100 substitutions per 100 amino acids. Bootstrap values represent the percentage of times the topology was generated in 1000 replicates.

Understanding the phenotypic effect of this altered expression pattern requires additional analysis.

Phylogenetic relationship between ALS and SAP gene families: The phylogenetic analysis of the ALS and SAP families focused on three main questions: (i) Which *C. albicans* ALS gene is the most basal and therefore likely to be the most ancestral form?, (ii) how does the rate of evolution compare between the ALS and SAP gene families?, and (iii) how does the evolutionary history of the ALS family compare with that of *C. albicans*? Phylogenetic reconstruction by maximum parsimony indicated that *ALS7* is the most basal lineage of the *C. albicans* ALS family when the trees are rooted with *C. tropicalis* sequences; *ALS6* is the second most-basal lineage in this analysis. Assuming a constant molecular clock for all ALS genes, this result implied that *ALS7* is the oldest gene. However, *ALS7* possesses a unique composition within the ALS family (Hoyer and Hecht 2000), so it potentially evolved under a different clock. Since we believe that few *C. albicans* ALS sequences remain uncharacterized, we expect these results to be robust to the addition of sequences from other species.

Evidence for a younger ALS family was provided by examination of the ALS and SAP phylogenetic trees, which showed that all of the ALS sequences from *C. tropicalis* formed a basal group, while the SAP sequences from *C. tropicalis* branched off more distal nodes. One *C. tropicalis* SAP sequence even grouped as a sister with a *C. albicans* SAP sequence. If a family arose before the *C. albicans*/*C. tropicalis* split, the descendant species would each receive a copy of the gene and we would expect to see grouping by related family members across species as we do with SAP. If a family arose after the *C. albicans*/*C. tropicalis* split, we would see each species group sepa-

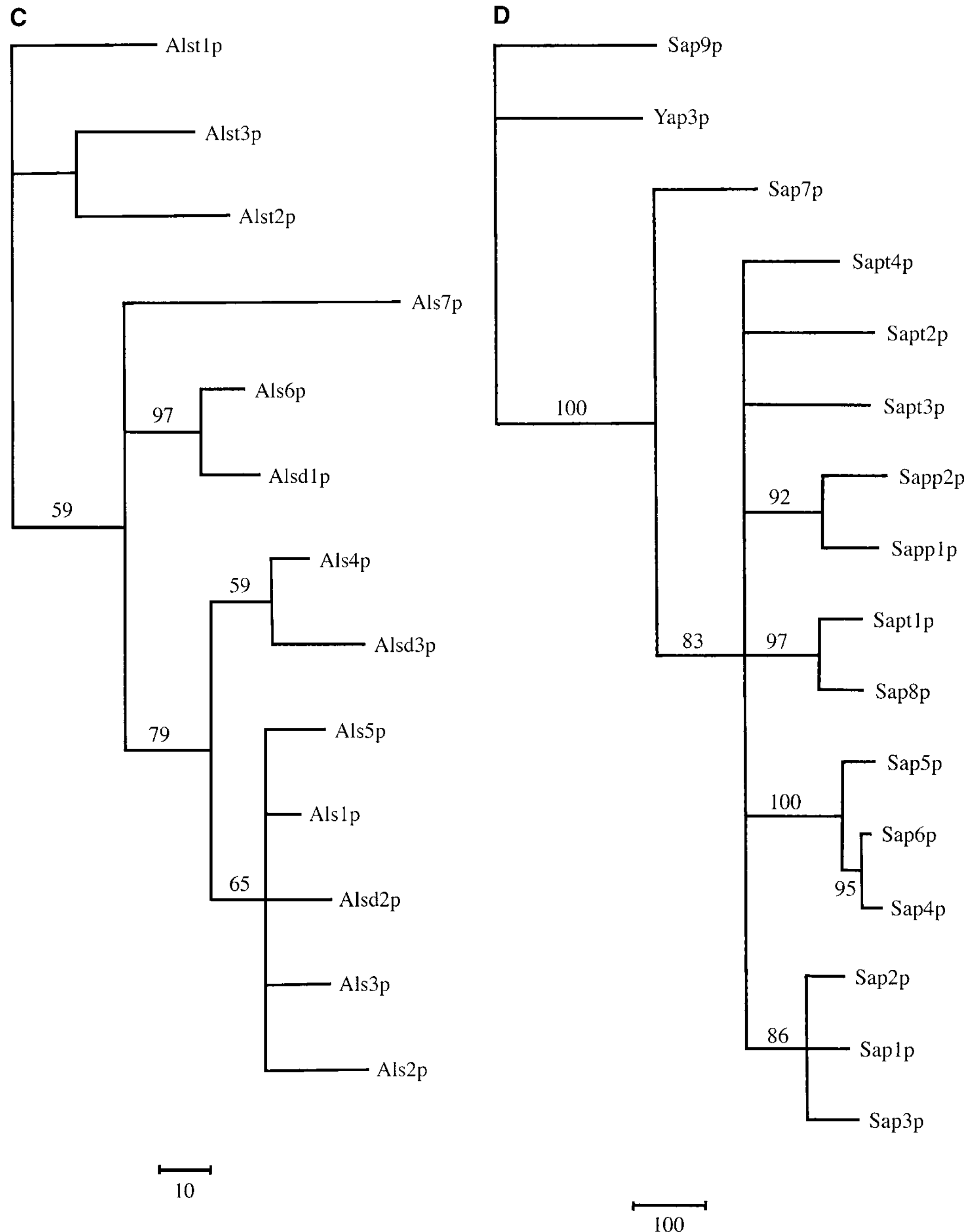

FIGURE 7.—*Continued.*

rately, as occurs with ALS. Families that arise during a split exhibit a blend of these two patterns.

An alternative explanation for our molecular phylogeny data is concerted evolution between ALS genes, in which repeats have a homogenizing effect and all gene copies within a species become the same or very similar over time (ZIMMER *et al.* 1980; HUGHES 1999). Using this explanation, we would argue that the *C. tropicalis* sequences group together because of sequence homogenization within *C. tropicalis*, rather than because of ALS gene diversification after the *C. tropicalis* and *C. dubliniensis/C. albicans* split. Hence, the history of the gene family and the history of the species would not be the same. In this event, we would expect species to group separately and not in the pairs that occur for *ALS4/ALSD3* and *ALS6/ALSD1* (Figure 7). Because few additional ALS or SAP genes are likely to be found in *C. albicans*, it is unlikely the basal branches will change and we conclude that ALS is a younger family than SAP.

Given that ALS is a younger family, it is remarkable that the number of genes is comparable to that of the SAP family. It is possible that the repeats found in ALS genes provide a mechanism for the rearrangement and amplification of the family (BIERNE and MICHEL 1994;

Parniske and Jones 1999). Homologous recombination between repeats could explain why the ALS and SAP families have such different histories.

Why are Candida gene families present? The preservation and expansion of the Candida ALS and SAP families could suggest that gene families lend a selective advantage to the organism. Several possibilities exist to explain their presence. First, gene families may exist because of the need for multiple specificities of the same general function. For example, the various Sap proteins may digest specific proteins with varying degrees of efficiency while different Als proteins may allow Candida to adhere to a variety of host surfaces. Conversely, specific proteins in each family may have redundant function that provides backup function in case one protein in the family is compromised. One example of redundancy may be found in the hypha-specific genes in both the ALS and SAP families (Hube *et al.* 1994; Hoyer *et al.* 1999a; Hoyer and Hecht 2000).

Other potential explanations for the presence of gene families include the possibility that a particular gene dosage is required upon confer a specific phenotype on the cell. Although the necessary probes to detect all ALS and SAP genes in non-*albicans* Candida may not be defined, a positive correlation exists between the number of SAP and ALS genes and the frequency with which a given Candida species is isolated from clinical specimen: *C. albicans* has the most genes from each family and the gene number appears to decrease in *C. tropicalis* and further decrease in *C. parapsilosis* (de Viragh *et al.* 1993). Perhaps the presence of additional proteins in each family signals greater colonization or pathogenic potential for that species.

Finally, the presence of two large gene families might suggest that products of each family have synergistic effects. These effects could occur between proteins of the same family or between proteins of the different families. Future research will clarify these possibilities and further define the role of gene families in Candida biology and pathogenesis.

We thank Patricia Kammeyer, David Coleman, and Richard Barton for Candida isolates. This work was supported by U.S. Public Health Service Grant AI39441, National Science Foundation Grant MCB-9630910, the Netherlands Technology foundation (STW) and the Earth Life Sciences Foundation (ALW). R.F. was supported in part by a Training Grant in Molecular and Cellular Mycology (T32-AI07373) from the National Institutes of Health.

LITERATURE CITED

Barns, S. M., D. J. Lane, M. L. Sogin, C. Bibeau and W. G. Weisburg, 1991 Evolutionary relationships among pathogenic *Candida* species and relatives. J. Bacteriol. **173:** 2250–2255.

Barton, N. H., and I. Wilson, 1996 Genealogies and geography, pp. 23–56 in *New Uses for New Phylogenies*, edited by P. H. Harvey, A. J. L. Brown, J. M. Smith and S. Nee. Oxford University Press, New York.

Bierne, H., and B. Michel, 1994 When replication forks stop. Mol. Microbiol. **13:** 17–23.

Borg-Von Zepelin, M., I. Meyer, R. Thomassen, R. Wurzner, D. Sanglard *et al.*, 1999 HIV-protease inhibitors reduce cell adherence of *Candida albicans* strains by inhibition of yeast secreted aspartic proteases. J. Invest. Dermatol. **113:** 747–751.

Calderone, R. A., and P. C. Braun, 1991 Adherence and receptor relationships of *Candida albicans*. Microbiol. Rev. **55:** 1–20.

Cassone, A., F. De Bernardis, A. Torosantucci, E. Tacconelli, M. Tumbarello *et al.*, 1999 In vitro and in vivo anticandidal activity of human immunodeficiency virus protease inhibitors. J. Infect. Dis. **180:** 448–453.

Darwazah, A., G. Berg and B. Faris, 1999 *Candida parapsilosis*: an unusual organism causing prosthetic heart valve infective endocarditis. J. Infect. **38:** 130–131.

de Viragh, P. A., D. Sanglard, G. Togni, R. Falchetto and M. Monod, 1993 Cloning and sequencing of two *Candida parapsilosis* genes encoding acid proteases. J. Gen. Microbiol. **139:** 335–342.

Devereux, J., P. Haeberli and O. Smithies, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

Fridkin, S. K., and W. R. Jarvis, 1996 Epidemiology of nosocomial fungal infections. Clin. Microbiol. Rev. **9:** 499–511.

Fu, Y., G. Rieg, W. A. Fonzi, P. H. Belander, J. E. Edwards, Jr. *et al.*, 1998 Expression of the *Candida albicans* gene *ALS1* in *Saccharomyces cerevisiae* induces adherence to endothelial and epithelial cells. Infect. Immun. **66:** 1783–1786.

Gaur, N. K., and S. A. Klotz, 1997 Expression, cloning, and characterization of a *Candida albicans* gene, *ALA1*, that confers adherence properties upon *Saccharomyces cerevisiae* for extracellular matrix proteins. Infect. Immun. **65:** 5289–5294.

Gilfillan, G. D., D. J. Sullivan, K. Haynes, T. Parkinson, D. C. Coleman *et al.*, 1998 *Candida dubliniensis*: phylogeny and putative virulence factors. Microbiology **144:** 829–838.

Gruber, A., C. Speth, E. Lukasser-Vogl, R. Zangerle, M. Borg-Von Zepelin *et al.*, 1999 Human immunodeficiency virus type 1 protease inhibitor attenuates *Candida albicans* virulence properties in vitro. Immunopharmacology **41:** 227–234.

Hillis, D. M., C. Moritz and B. K. Mable, 1996 *Molecular Systematics*, Ed. 2. Sinauer Associates, Sunderland, MA.

Hoegl, L., E. Thoma-Greber, M. Rocken and H. C. Korting, 1998 HIV protease inhibitors influence the prevalence of oral candidosis in HIV-infected patients: a 2-year study. Mycoses **41:** 321–325.

Hoppe, J. E., M. Klausner, T. Klingebiel and D. Niethammer, 1997 Retrospective analysis of yeast colonization and infections in paediatric bone marrow transplant recipients. Mycoses **40:** 47–54.

Hoyer, L. L., and J. E. Hecht, 2000 The *ALS6* and *ALS7* genes of *Candida albicans*. Yeast **16:** 847–855.

Hoyer, L. L., and J. E. Hecht, 2001 The *ALS5* gene of *Candida albicans* and analysis of the Als5p N-terminal domain. Yeast **18:** 49–60.

Hoyer, L. L., S. Scherer, A. R. Shatzman and G. P. Livi, 1995 *Candida albicans* ALS1: domains related to a *Saccharomyces cerevisiae* sexual agglutinin separated by a repeating motif. Mol. Microbiol. **15:** 39–54.

Hoyer, L. L., T. L. Payne, M. Bell, A. M. Myers and S. Scherer, 1998a *Candida albicans ALS3* and insights into the nature of the ALS gene family. Curr. Genet. **33:** 451–459.

Hoyer, L. L., T. L. Payne and J. E. Hecht, 1998b Identification of *Candida albicans ALS2* and *ALS4* and localization of Als proteins to the fungal cell surface. J. Bacteriol. **180:** 5334–5343.

Hube, B., M. Monod, D. A. Schofield, A. J. Brown and N. A. Gow, 1994 Expression of seven members of the gene family encoding secretory aspartyl proteinases in *Candida albicans*. Mol. Microbiol. **14:** 87–99.

Hube, B., R. Ruchel, M. Monod, D. Sanglard and F. C. Odds, 1998 Functional aspects of secreted *Candida* proteinases. Adv. Exp. Med. Biol. **436:** 339–344.

Hughes, A. L., 1999 Concerted evolution of exons and introns in the MHC-linked tenascin-X genes of mammals. Mol. Biol. Evol. **16:** 1558–1567.

Hull, C. M., R. M. Raisner and A. D. Johnson, 2000 Evidence for mating of the "asexual" yeast *Candida albicans* in a mammalian host. Science **289:** 307–310.

Kapteyn, J. C., H. Van Den Ende and F. M. Klis, 1999 The contribu-

tion of cell wall proteins to the organization of the yeast cell wall. Biochim. Biophys. Acta **1426:** 373–383.

Kapteyn, J. C., L. L. Hoyer, J. E. Hecht, W. H. Muller, A. Andel *et al.*, 2000 The cell wall architecture of *Candida albicans* wild-type cells and cell-wall-defective mutants. Mol. Microbiol. **35:** 601–611.

Kunova, A., J. Trupl, A. Demitrovicova, Z. Jesenska, S. Grausova *et al.*, 1997 Eight-year surveillance of non-*albicans Candida* spp. in an oncology department prior to and after fluconazole had been introduced into antifungal prophylaxis. Microb. Drug Resist. **3:** 283–287.

Larone, D. H., 1995 *Medically Important Fungi*. American Society for Microbiology, Washington, DC.

Lee, K. L., H. R. Buckley and C. C. Campbell, 1975 An amino acid liquid synthetic medium for the development of mycelial and yeast forms of *Candida albicans*. Sabouraudia **13:** 148–153.

Lipke, P. N., D. Wojciechowicz and J. Kurjan, 1989 *AGα1* is the structural gene for the *Saccharomyces cerevisiae* α-agglutinin, a cell surface glycoprotein involved in cell-cell interactions during mating. Mol. Cell. Biol. **9:** 3155–3165.

Magee, B. B., and P. T. Magee, 2000 Induction of mating in *Candida albicans* by construction of MTLa and MTLalpha strains. Science **289:** 310–313.

Magee, B. B., M. Sanchez and P. T. Magee, 1999 Karyotypic analysis of the recently identified oral pathogen *Candida dubliniensis*, Abstract C52. ASM Conference on *Candida* and Candidiasis, Charleston, SC, American Society for Microbiology.

Mahrous, M., A. D. Sawant, W. R. Pruitt, T. Lott, S. A. Meyer *et al.*, 1992 DNA relatedness, karyotyping and gene probing of *Candida tropicalis*, *Candida albicans* and its synonyms *Candida stellatoidea* and *Candida claussenii*. Eur. J. Epidemiol. **8:** 444–451.

Monod, M., G. Togni, B. Hube and D. Sanglard, 1994 Multiplicity of genes encoding secreted aspartic proteinases in *Candida* species. Mol. Microbiol. **13:** 357–368.

Monod, M., B. Hube, D. Hess and D. Sanglard, 1998 Differential regulation of *SAP8* and *SAP9*, which encode two new members of the secreted aspartic proteinase family in *Candida albicans*. Microbiology **144:** 2731–2737.

Orti, G., D. E. Pearse and J. C. Avise, 1997 Phylogenetic assessment of length variation at a microsatellite locus. Proc. Natl. Acad. Sci. USA **94:** 10745–10749.

Parniske, M., and J. D. Jones, 1999 Recombination between diverged clusters of the tomato Cf-9 plant disease resistance gene family. Proc. Natl. Acad. Sci. USA **96:** 5850–5855.

Rangel-Frausto, M. S., T. Wiblin, H. M. Blumberg, L. Saiman, J. Patterson *et al.*, 1999 National epidemiology of mycoses survey (NEMIS): variations in rates of bloodstream infections due to *Candida* species in seven surgical intensive care units and six neonatal intensive care units. Clin. Infect. Dis. **29:** 253–258.

Sanglard, D., B. Hube, M. Monod, F. C. Odds and N. A. Gow, 1997 A triple deletion of the secreted aspartyl proteinase genes *SAP4*, *SAP5*, and *SAP6* of *Candida albicans* causes attenuated virulence. Infect. Immun. **65:** 3539–3546.

Schaller, M., H. C. Korting, W. Schafer, J. Bastert, W. C. Chen *et al.*, 1999 Secreted aspartic proteinase (Sap) activity contributes to tissue damage in a model of human oral candidosis. Mol. Microbiol. **34:** 169–180.

Smits, G. J., J. C. Kapteyn, H. Van Den Ende and F. M. Klis, 1999 Cell wall dynamics in yeast. Curr. Opin. Microbiol. **2:** 348–352.

Sugita, T., and T. Nakase, 1999 Non-universal usage of the leucine CUG codon and the molecular phylogeny of the genus *Candida*. Syst. Appl. Microbiol. **22:** 79–86.

Sullivan, D. J., and D. Coleman, 1997 *Candida dubliniensis*: an emerging opportunistic pathogen. Curr. Top. Med. Mycol. **8:** 15–25.

Sullivan, D. J., T. J. Westerneng, K. A. Haynes, D. E. Bennett and D. C. Coleman, 1995 *Candida dubliniensis* sp. nov.: phenotypic and molecular characterization of a novel species associated with oral candidosis in HIV-infected individuals. Microbiology **141:** 1507–1521.

Weinberger, M., T. Sacks, J. Sulkes, M. Shapiro and I. Polacheck, 1997 Increasing fungal isolation from clinical specimens: experience in a university hospital over a decade. J. Hosp. Infect. **35:** 185–195.

Wickes, B., J. Staudinger, B. B. Magee, K. J. Kwon-Chung, P. T. Magee *et al.*, 1991 Physical and genetic mapping of *Candida albicans*: several genes previously assigned to chromosome 1 map to chromosome R, the rDNA-containing linkage group. Infect. Immun. **59:** 2480–2484.

Wingard, J. R., 1995 Importance of *Candida* species other than *C. albicans* as pathogens in oncology patients. Clin. Infect. Dis. **20:** 115–125.

Wingard, J. R., W. G. Merz and R. Saral, 1979 *Candida tropicalis*: a major pathogen in immunocompromised patients. Ann. Intern. Med. **91:** 539–543.

Van't Wout, J. W., 1996 Fluconazole treatment of candidal infections caused by non-*albicans Candida* species. Eur. J. Clin. Microbiol. Infect. Dis. **15:** 238–242.

Zimmer, E. A., S. L. Martin, S. M. Beverly, Y. W. Kan and A. C. Wilson, 1980 Rapid duplication and loss of genes coding for the alpha chains of hemoglobin. Proc. Natl. Acad. Sci. USA **77:** 2158–2162.

Communicating editor: M. E. Zolan