Curr Genet (1998) 33: 451–459

ORIGINAL PAPER

Lois L. Hoyer · Tracie L. Payne · Maren Bell
Alan M. Myers · Stewart Scherer

Candida albicans ALS3 and insights into the nature of the ALS gene family

Received: 24 December 1997 / 12 March 1998

Abstract The *ALS1* (*ag*glutinin-*l*ike *s*equence) gene of *Candida albicans* encodes a protein similar to alpha-agglutinin, a cell-surface adhesion glycoprotein of *Saccharomyces cerevisiae* (Hoyer et al. 1995). A central domain of a tandemly repeated 108-bp sequence is found in the *ALS1* coding region. This tandem-repeat motif hybridizes to multiple *C. albicans* genomic DNA fragments, indicating the possibility of other *ALS1*-like genes in *C. albicans* (Hoyer et al. 1995). To determine if these fragments constitute a gene family, tandem-repeat-hybridizing genomic fragments were isolated from a fosmid library by PCR screening using primers based on the consensus tandem-repeat sequence of *ALS1* (Hoyer et al. 1995). One group of fosmids, designated *ALS3*, encodes a gene with 81% identity to *ALS1*. The sequences of *ALS1* and *ALS3* are most conserved in the tandem-repeat domain and in the region 5′ of the tandem repeats. Northern-blot analysis using unique probes from the 3′ end of each gene demonstrated that *ALS1* expression varies, depending on which *C. albicans* strain is examined, and that *ALS3* is hyphal-specific. Both genes are found in a variety of *C. albicans* and *C. stellatoidea* strains examined. The predicted Als1p and Als3p exhibit features suggesting that both are cell-surface glycoproteins. Southern blots probed with conserved sequences from the region 5′ of the tandem repeats suggest that other ALS-like sequences are present in the *C. albicans* genome and that the ALS family may be larger than originally estimated.

Key words *Candida albicans* · Gene family · Hyphal-specific · Differential gene regulation

L. L. Hoyer (✉) · T. L. Payne
Department of Veterinary Pathobiology,
University of Illinois at Urbana-Champaign, 2522 VMBSB,
2001 S. Lincoln Avenue, Urbana, IL 61802, USA
e-mail: lhoyer@uiuc.edu
Fax: +1-217-244 7421

L. L. Hoyer · A. M. Myers
Department of Biochemistry and Biophysics,
Iowa State University, Ames, IA 50011, USA

L. L. Hoyer · M. Bell[1] · S. Scherer[2]
Human Genome Center, Lawrence Berkeley Laboratory,
Berkeley, CA 94720, USA

S. Scherer[2]
Department of Microbiology, University of Minnesota,
School of Medicine, Minneapolis, MN 55455, USA

Present addresses:
[1] Department of Molecular and Cellular Biology,
University of California Berkeley, Berkeley, CA 94708, USA
[2] Acacia Biosciences, 4136 Lakeside Drive, Richmond, CA 94806, USA

Communicated by O. C. Yoder

Introduction

The *ALS1* (*ag*glutinin-*l*ike *s*equence) gene of *Candida albicans* was isolated from a screen intended to identify genes transcribed in hyphal forms, but not in yeast forms (Hoyer et al. 1995). Further experiments indicated that *ALS1* is not a hyphal-specific gene, but is transcribed in response to growth-medium components (Hoyer et al. 1995). *ALS1* encodes a protein with similarities to alpha-agglutinin, a cell-surface adhesion glycoprotein of *Saccharomyces cerevisiae* that facilitates cell-cell contact during the mating of haploid yeast (Hauser and Tanner 1989; Lipke et al. 1989). Since, at this time, *C. albicans* has not been observed to undergo meiosis (Odds 1988), it is questionable whether Als1p plays a directly analogous role. Sequence features of Als1p suggest it is a cell-surface glycoprotein. A cell-surface localization, along with similarities to alpha-agglutinin, indicate that Als1p might have an adhesive function, with adhesion to host cells being the most attractive possibility. Adhesion of an ALS-like gene to extracellular matrix proteins and buccal epithelial cells was recently demonstrated lending support to this hypothesis (Gaur and Klotz 1997). Because *Candida* cells that are more adhesive tend to be more pathogenic (Calderone and Braun 1991), it is possible that Als1p may contribute to the pathogenesis of *C. albicans*. The long-term goal of our work is to define the function of Als proteins and their contribution to *C. albicans* pathogenesis.

ALS1 is composed of three domains: a central region consisting of variable numbers of copies of a tandemly repeated 108-bp motif, sequences 5′ of this region, and sequences 3′ of the tandem repeats (Hoyer et al. 1995). The tandem-repeat sequences hybridize with multiple *C. albicans* genomic fragments suggesting the possibility of more *ALS1*-like genes in the *C. albicans* genome. Three to seven tandem-repeat-hybridizing fragments are detected, depending on which *C. albicans* strain is examined (Hoyer et al. 1995). The goal of the present study was to isolate these fragments and determine whether they encode *ALS1*-like genes. This knowledge is essential to formulate the best strategy for testing protein function and its contribution to *C. albicans* pathogenesis. Such studies are often conducted by creating a null mutant for the gene in question and evaluating its phenotype. Because of the diploid nature of *C. albicans*, the creation of null mutants is more difficult than in systems with facile genetics such as *S. cerevisiae* (Scherer and Magee 1990). While advancements in technology have greatly facilitated these experiments (Fonzi and Irwin 1993), the study of gene families is still difficult because of the number of genes that must be disrupted to create a null mutant and because of extensive sequence similarities which complicate the targeting of individual genes. These problems have been encountered previously in the study of the SAP gene family, which encodes secreted aspartyl proteinases (Hube et al. 1997; Sanglard et al. 1997).

In the study reported here, fosmid clones encoding the tandem repeat motif were isolated from a fosmid library by a polymerase-chain-reaction (PCR)-based screening method. Fosmids isolated using this procedure were grouped by the appearance of their PCR products. Subcloning and DNA sequencing of a fosmid from the *ALS3* group revealed a gene with 81% overall identity to *ALS1*. Similarities between *ALS1* and *ALS3* are most striking in the tandem repeat domain and sequences 5′ of that region. Unique probes for *ALS1* and *ALS3* demonstrated that both genes are found in a variety of *C. albicans* and *Candida stellatoidea* strains and that genes in the ALS family are dispersed in the *C. albicans* genome. Northern-blot analysis indicated that ALS genes are differentially regulated and that *ALS3* is hyphal-specific. Additional Southern analysis suggested that the ALS family is larger than originally estimated.

Materials and methods

Media and strains. All standard growth conditions and strains were described previously (Hoyer et al. 1995). Two sources of *C. albicans* strain B311 were used in this study: one, from the SmithKline Beecham culture collection (designated SB B311 in this manuscript), and the other from the American Type Culture Collection (designated B311, ATCC B311, or ATCC 32354). Coding regions for *ALS1* and *ALS3* are identical in these strains, but polymorphisms which alter restriction sites exist in sequences near the coding regions (Hoyer, unpublished observation). *Candida claussenii* and *Candida langeronii* are synonymous with *C. albicans* (Wickes et al. 1992). ATCC 11006 and ATCC 36232 are strains of *C. stellatoidea*.

Library construction and screening. The fosmid library used in this work was a gift from Michael Strathmann. This library consists of 40–50-kb partial *Sau*3AI fragments of genomic DNA from *C. albicans* strain 1161 (Goshorn and Scherer 1989) ligated into a fosmid vector (Kim et al. 1992). Individual clones are seeded in microtiter wells; each clone is uniquely identified by plate, row, and column number. The fosmid library was screened using a PCR protocol. Library clones were pooled by row from each plate; these row pools were grown overnight in LB + Cml medium (per liter: 10 g tryptone, 5 g yeast extract, 5 g sodium chloride, with 20 μg/ml chloramphenicol). DNA was extracted from each pool using an alkaline-lysis protocol (Birnboim and Doly 1979). Each pool was screened by PCR using primers synthesized based on the consensus tandem repeat sequence of *ALS1* (Hoyer et al. 1995). The primers were 5′ TTATCAGAGAGCCACCAAACC 3′ (forward) and 5′ GACCAATATTCAGTAGTAGTGACAGT 3′ (reverse). For each primer, 70% of the nucleotides matched the consensus sequence exactly. Each PCR reaction consisted of 1 μM of each primer, 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 1×Perkin Elmer PCR reaction buffer, 2.5 units of Perkin Elmer *Taq* DNA polymerase, and 10 ng of purified fosmid DNA from each row pool. Thirty five cycles of PCR were performed using the following cycling parameters: 94 °C denaturation for 30 s, 55 °C annealing for 30 s, and 72 °C extension for 90 s. A final extension cycle of 7 min at 72 °C completed the PCR run. PCR products were separated on a 2.5% agarose gel and visualized by ethidium-bromide staining. To identify which well in the positive row pools contained the PCR-positive fosmid, cells from individual wells of each positive row pool were grown overnight in LB + Cml medium and screened by PCR as described above. Positive fosmids were considered to potentially encode ALS genes. Fosmids encoding *ALS1* were identified by PCR using primers based on sequences in the *Xba*I-*Hin*dIII fragment known to be specific for *ALS1* (Hoyer et al. 1995). Forward primer 5′ CTAGTGAACCAACAAATACCAG 3′ and reverse primer 5′ ACATCAGTACCAGAAGAAACAG 3′ were used in PCR reactions as described above. Each reaction included 10 ng of purified fosmid DNA. Thirty cycles of PCR were run with the following parameters: 94 °C denaturation for 30 s, 53 °C annealing for 30 s, and 72 °C extension for 30 s. A final extension cycle of 7 min at 72 °C completed the PCR run. Since fosmids encoding uncharacterized ALS genes were sought, fosmids from which the 325-bp *ALS1*-specific product was amplified were eliminated from further studies.

Southern blots. Genomic Southern blots were performed as previously described (Hoyer et al. 1995) using the DIG Nonradioactive Nucleic Acid Labeling and Detection System (Boehringer Mannheim). Nylon membranes hybridized with several probes in succession were stripped by incubation in 0.2 N NaOH, 0.1% SDS at 65 °C for 1 h. Blots were equilibrated in 2×SSC (300 mM NaCl, 30 mM sodium citrate, pH 7.0) prior to hybridization with another DIG-labeled probe. Several probes in this study were derived from *ALS1* (Hoyer et al. 1995). These included an 875-bp *Xba*I-*Hin*dIII fragment (nucleotides 2611–3486 of the coding region) used as an *ALS1*-specific probe, an 870-bp tandem-repeat-specific *Kpn*I fragment from pLH21, and a 391-bp *Kpn*I-*Hpa*I fragment from the region of *ALS1* 5′ of the tandem repeats (nucleotides 464–855 of the coding region). An *ALS3*-specific probe was synthesized from sequences 3′ of the tandem repeat domain by PCR amplification using the primers 5′ ATGACACCATGTCAAGTTCAGA 3′ (forward) and 5′ CACACCAAATTGGAGGTGATT 3′ (reverse). PCR reactions were as described above and included 10 ng of genomic DNA from *C. albicans* strain 1177 as template. Thirty five cycles of 94 °C denaturation for 30 s, 56 °C annealing for 30 s, and 72 °C extension for 90 s amplified a 1017-bp fragment. This fragment was digested with *Hin*cII to yield a 758-bp probe fragment specific for *ALS3*.

Northern blots. Total RNA for Northern blots was isolated using a phenol extraction procedure (Collart et al. 1993). Twenty five micrograms of total RNA were loaded per lane on 1% formaldehyde agarose gels as previously described (Hoyer et al. 1995) and transferred to nylon membranes (Micron Separations Inc.) using standard protocols (Sambrook et al. 1989). Probes were prepared by random priming using an Oligolabelling Kit (Pharmacia) and [^{32}P]-dCTP

(3000 Ci mmol^{-1}, Amersham). Hybridizations were performed for 16 h at 65°C under the high-stringency conditions described by Monod et al. (1994). Blots were washed twice for 15 min each in 2×SSC, 0.1% (w/v) SDS at 65 °C and exposed to film. A fragment of the *C. albicans TEF1* gene (Sundstrom et al. 1990) was used as a positive control as previously described (Hoyer et al. 1995).

Culture of C. albicans cells for Northern analysis. To investigate the in vitro expression of *ALS1* and *ALS3*, *C. albicans* cells were grown in either YPD (per liter: 10 g yeast extract, 20 g peptone, 20 g dextrose) or RPMI 1640 medium (catalog number 11875-085, Gibco BRL). Strains B311, B792, 1177, 3153 A, SC5314, and WO-1 were included in this analysis. A single *C. albicans* colony from YPD agar was grown for approximately 16 h to stationary phase in YPD liquid at 25 °C. Typically, the density of these cultures was 3–7×10^8 cells/ml, with all strains growing as yeast forms except for SC5314 which typically exhibited 1–5% germ tubes. Cells from stationary phase cultures were used to inoculate fresh YPD (25 °C) and RPMI 1640 (37 °C) cultures at 5×10^6 cells/ml; cultures were incubated with shaking at 200 rpm for 3 h. RPMI 1640-grown cultures exhibited a mixture of yeast and germ tube/hyphal forms at the end of this incubation. Cells of strain WO-1 were white-phase in the YPD culture and formed hyphae upon inoculation into RPMI 1640. Total RNA was harvested from the entire culture and analyzed on Northern blots as described above. Experiments to demonstrate hyphal-specific expression of *ALS3* were similar to those used in the characterization of *C. albicans ECE1* (Birse et al. 1993). *C. albicans* cells from a stationary phase YPD culture, grown as described above, were inoculated at 5×10^6 cells/ml in 10 mM imidazole-HCl (pH 7.0) containing 0.2 mM $MnCl_2$ (Shepherd et al. 1980). Various inducers of hyphal formation were added to the imidazole buffer including 2.5 mM N-acetylglucosamine (NAG) (Shepherd et al. 1980), 4% serum (Shepherd et al. 1980), 2.5 mM L-proline (Dabrowa and Howard 1981), or 2.5 mM glucose plus 2.5 mM glutamine (Glc-Gln) (Shepherd et al. 1980). Cultures were incubated at 25 °C or 37 °C for 3 h and total RNA extracted for Northern analysis. The percentage of cells forming hyphae was noted in each culture. Strains 3153 A and SC5314 were used in this analysis.

CHEF gels. Agarose plugs for CHEF analysis were prepared using minor modifications of a previously described method (Gerring et al. 1991). Modifications included the use of 1.5% low-melting-point agarose and 2.5 mg/ml of proteinase K. Each plug preparation used 2×10^8 cells from an overnight YPD culture. CHEF gels were run as previously described (Wickes et al. 1991a) on a CHEF-DRIII apparatus (Bio Rad). Chromosomes were numbered from largest to smallest with the rDNA-encoding chromosome designated as R (Wickes et al. 1991b).

DNA sequencing and analysis. DNA sequencing was performed by the Iowa State University DNA Sequence and Synthesis Facility using Applied Biosystems sequencers. Templates were prepared by alkaline lysis and PEG precipitation (Kraft et al. 1988); *Taq* cycle sequencing reactions utilized dye-terminator chemistry (Applied Biosystems). The *ALS3* DNA sequence was completed by a combination of sequencing subclones and oligo-walking with custom primers. Custom primers were synthesized on an Applied Biosystems 394 DNA synthesizer. DNA sequences were analyzed using the Genetics Computer Group (GCG) programs (Devereux et al. 1984). CUG codons were translated as serine instead of leucine (Santos and Tuite 1995; White et al. 1995).

Results

Isolation of ALS genes by PCR screening of a fosmid library

A fosmid library constructed from *C. albicans* strain 1161 genomic DNA was screened by PCR using primers synthesized based on the consensus tandem-repeat sequence derived from *ALS1* (Hoyer et al. 1995). Approximately ten genome equivalents were screened yielding 59 positive clones. Examination of PCR products on ethidium bromide-stained agarose gels indicated that patterns of PCR products were variable depending on which fosmid was examined (Fig. 1). This variability was presumed to be due to tandem-repeat-based differences in annealing of the PCR primers at the 55 °C annealing temperature. PCR primers were 70% identical to the consensus tandem-repeat sequence with some individual copies of the tandem repeat sequence matching the primer sequences more closely than others (Hoyer et al. 1995).

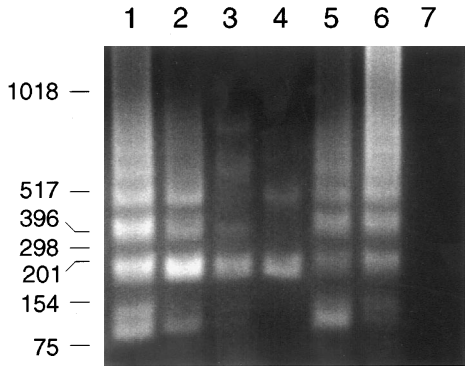

Fig. 1 Representative PCR products from fosmid library screening. Variability of PCR-product patterns observed during fosmid library screening was demonstrated by amplification of 10 ng of purified fosmid or *C. albicans* 1177 genomic DNA with tandem-repeat-specific primers. Size markers (in bp) are shown at left. *Lane 1* fosmid 16G-3 (*ALS1*), *lane 2* fosmid 17D-5 (*ALS3*). Products in *lanes 3–5* represent alleles of *ALS2* and *ALS4* (Hoyer, Payne and Hecht, submitted). *Lane 6 C. albicans* 1177 genomic DNA, *lane 7* No-DNA control

Fosmids were grouped based on the appearance of their PCR products since these potentially represented different ALS genes. Fosmids from which an *ALS1*-specific product could be amplified were identified by PCR with *ALS1*-specific primers. Because the goal of this study was to isolate uncharacterized ALS genes, all *ALS1*-positive fosmids were eliminated from further experiments.

Subcloning and DNA sequencing of *ALS3*

Fosmid 17D-5 was selected as representative of the *ALS3* group. Fosmid 17D-5 DNA was digested with a variety of restriction enzymes, Southern blotted, and probed with an 870-bp *Kpn*I fragment from *ALS1* that contains only tandem-repeat sequences (Hoyer et al. 1995). A 3.8-kb *Bam*HI-*Hin*dIII tandem-repeat-hybridizing fragment of 17D-5 was identified and subcloned into pUC vectors for DNA sequencing (Yanisch-Perron et al. 1985). DNA sequencing from the *Hin*dIII site indicated the presence of a large open reading frame (ORF). The sequence of this ORF was completed by constructing overlapping subclones using unique restriction sites in the region. The resulting ORF was 3357 bp encoding a 1119 amino-acid protein.

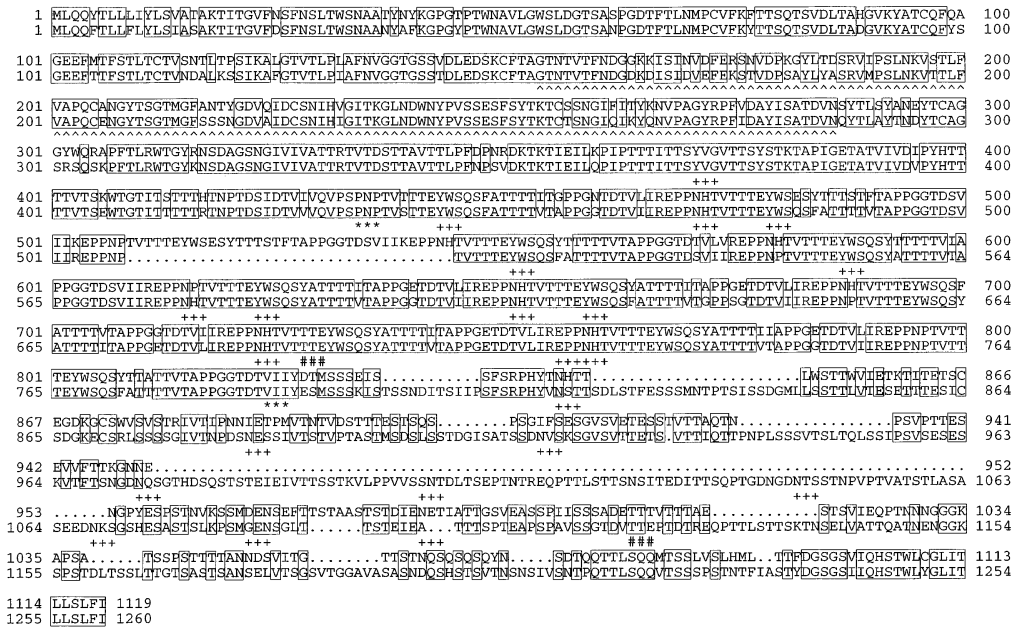

Fig. 2 Optimal alignment of the predicted Als3p (*upper*) and Als1p (*lower*) amino-acid sequences. This alignment was generated with default parameters of the GAP program of GCG software (Genetics Computer Group, Madison, Wis.). Identical amino acids are *boxed*. *Three asterisks*, placed under the lower sequence, mark the beginning and end of the tandem-repeat domain. *Plus signs* are found above (for Als3p) or below (for Als1p) consensus N-glycosylation sites (Asn-X-Ser/Thr where *X* is any residue except proline; Bause 1983). The region corresponding to the *ALS3*-specific probe extends from amino acid 828 to 1076; the start and end of this region are both *overscored by pound signs*. Amino-acid sequences predicted from the nucleotide sequence of the *Kpn*I-*Hpa*I probe are *underscored by carats*. Nucleotide sequences of *ALS1* (accession number L25902) and *ALS3* (accession number U87956) are deposited in the GenBank database

This gene was named *ALS3* because it was originally in the group of fosmids designated *ALS3*. The translational start of *ALS3* was chosen based on alignment to the *ALS1* sequence. The selected ATG was preceded by an AT-rich region and was the only putative start site within 50–100 nucleotides in either direction. Stop codons are located within 50 nucleotides upstream of or downstream from this site in all other reading frames (data not shown).

Comparison between *ALS1* and *ALS3* and their predicted amino-acid sequences

The nucleotide sequences of *ALS1* and *ALS3* were 81% identical; predicted amino-acid sequences were 78% identical and 86% similar. Much of the identity between Als1p and Als3p sequences was found in the tandem repeats (87% identity, 95% similarity) and sequences N-terminal of the repeats (85% identity, 90% similarity; Fig. 2). In contrast, sequences C-terminal of the tandem repeats were only 56% identical and 67% similar. This disparity between the sequences 3′ of the tandem-repeat domain was exploited to design an *ALS3*-specific probe (Fig. 2). This situation is similar to that of *ALS1* where the gene-specific *Xba*I-*Hin*dIII fragment is derived from sequences 3′ of the tandem-repeat domain (Hoyer et al. 1995).

Als1p and Als3p shared many common sequence features. Each protein had hydrophobic N- and C-termini with the characteristics of a signal sequence (von Heijne 1990) and a GPI anchor addition site, respectively (Gerber et al. 1992; Coyne et al. 1993). Both predicted amino-acid sequences were rich in serine and threonine and encoded many consensus N-glycosylation sites (Bause 1983); each of these features was more marked in the tandem-repeat domain and sequences C-terminal of the tandem-repeats. Although the number of amino acids N-terminal of the tandem repeat domain was the same in both proteins (433 amino acids), the region C-terminal of the tandem repeats was only 292 amino acids in Als3p compared to 470 amino acids in Als1p. This smaller C-terminal domain predicted a smaller non-glycosylated protein size for Als3p (120 kDa versus 133 kDa for Als1p) although the size of any Als protein is largely influenced by the number of copies of the tandem-repeat sequence present (Hoyer et al. 1995). Like Als1p, sequence features of Als3p suggested it was a cell-surface glycoprotein.

Transcriptional regulation of *ALS1* and *ALS3*

ALS1 was originally isolated from *C. albicans* strain B311 in a differential hybridization screen intended to isolate genes transcribed in hyphal cells, but not in yeast forms. In this screen, yeast cells were grown in YPD medium and hyphal formation induced by placing YPD-grown cells into RPMI 1640 (Hoyer et al. 1995). Further experiments demonstrated that while *ALS1* is transcribed only in RPMI 1640-grown cells, expression of *ALS1* is not strictly linked to hyphal formation (Hoyer et al. 1995). Experiments were conducted to determine whether *ALS3* expression followed

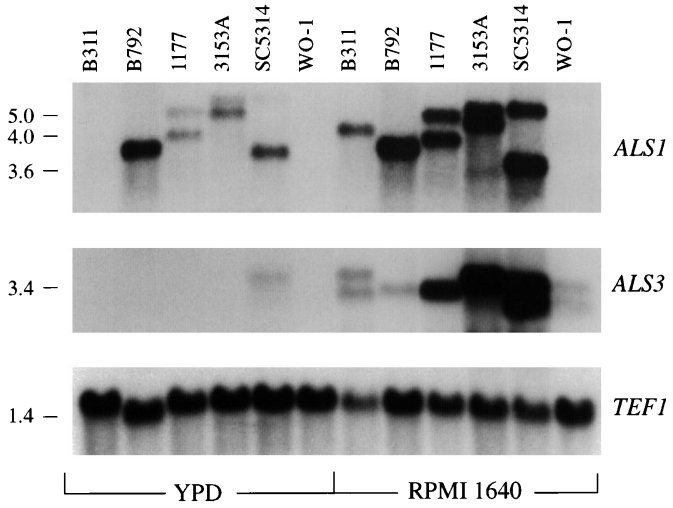

Fig. 3 Northern blot to detect *ALS1*- and *ALS3*-specific messages in different *C. albicans* strains grown in YPD or RPMI 1640. Total RNA from six *C. albicans* strains grown in YPD (left half of blot) or RPMI 1640 (right half of blot) was separated on a formaldehyde-agarose gel. The resulting blot was probed with the *ALS3*-specific fragment (center panel), stripped and re-probed with the *Xba*I-*Hin*dIII *ALS1*-specific fragment (top panel) and then with a fragment of *TEF1* as a control for equal loading of total RNA (Sundstrom et al. 1990). Size markers (in kb) are shown at the left of each blot

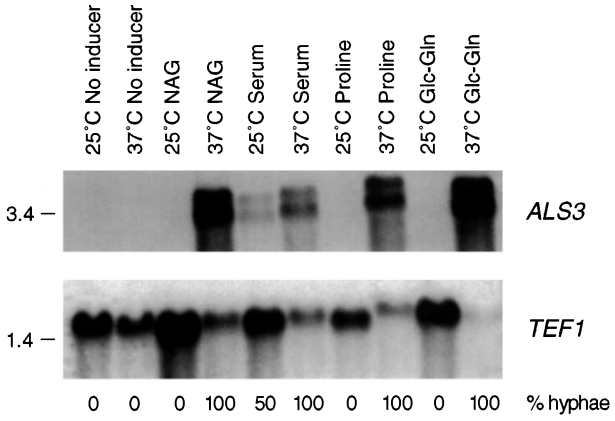

Fig. 4 Northern blot to demonstrate hyphal-specific expression of *ALS3*. *C. albicans* strain SC5314 was incubated in 10 mM imidazole-HCl buffer (pH 7.0) at 25 °C or 37 °C (*lanes 1 and 2*) or in buffer with hyphal-inducing substances added as described in Materials and methods (*lanes 3–10*: *NAG* = N-acetylglucosamine, *Glc-Gln* = glucose and glutamine). The blot was probed with the *ALS3*-specific fragment and then with a fragment derived from *C. albicans TEF1* to control for the loading of RNA (Sundstrom et al. 1990). The percentage of hyphal forms observed in each culture from which RNA was extracted is noted below the blots. Size markers (in kb) are shown at the left of each blot

this same pattern. Six *C. albicans* strains were used in this study since previous experiments indicated that *ALS1* expression varied depending on which *C. albicans* strain was examined (Hoyer, unpublished observations).

ALS1-specific signals were variable in size; some strains showed two distinct *ALS1*-hybridizing species (Fig. 3). The presence of two signals is due to differences in the sizes of *ALS1* alleles in these strains (Hoyer et al. 1995). Size differences between alleles of *ALS1* were shown previously at the DNA level and are due entirely to differences in the number of copies of the tandem-repeat element present in each *ALS1* allele (Hoyer et al. 1995). This same size variability was evident in *ALS3* alleles in certain stains (Fig. 3). PCR experiments using primers to amplify various regions of *ALS3* in each of these strains confirmed that the signals observed in the Northern blot are due to differences in the number of copies of the 108-bp motif in the tandem-repeat domain (Hoyer, unpublished observations).

The Northern blots in Fig. 3 reproduced the original result in strain B311 that was the basis for the isolation of *ALS1* (Hoyer et al. 1995); in strain B311, *ALS1* was transcribed in RPMI 1640-grown cells, but not in YPD-grown cells. Other *C. albicans* strains showed *ALS1*-specific signals in cells grown under both media conditions, although signals appeared to be stronger in RPMI 1640-grown cells (Fig. 3). In the case of strain WO-1, *ALS1*-specific signals were absent in both YPD- and RPMI 1640-grown cells. Examination of WO-1 genomic DNA indicated that the *ALS1* gene was present in this strain, but sequences immediately upstream of *ALS1* in other strains were completely absent from the WO-1 genome (Hoyer and Scherer, unpublished). These observations suggest that *ALS1* may not be transcribed in strain WO-1 or may be under the control of other regulatory mechanisms.

ALS3-specific signals were observed in all lanes from RPMI 1640-grown cells (Fig. 3), but were absent in all strains grown in YPD with the exception of a very faint signal in cells of strain SC5314. Examination of the YPD-grown SC5314 cells indicated that less than 5% grew as hyphal forms while other YPD-grown strains grew solely as yeast forms. In RPMI 1640 medium, cultures grew as mixtures of yeast and hyphal forms. This observation raised the possibility that *ALS3* may be transcribed only in hyphal cells. To explore whether *ALS3* expression was hyphal-specific, cells of *C. albicans* strain SC5314 were grown under a variety of hyphal-inducing conditions (Fig. 4). The percentage of hyphal forms in each culture was recorded. Total RNA extracted from these cells was analyzed on Northern blots probed with the *ALS3*-specific fragment. In each case, *ALS3*-specific signals were detected only in cultures where hyphal forms were found. These data support the conclusion that *ALS3* transcription in strain SC5314 was hyphal-specific. Similar results were obtained with strain 3153A (data not shown).

Chromosomal localization of ALS genes

Southern blots of pulsed-field-electrophoresis-separated *C. albicans* chromosomes were performed to determine whether the ALS gene family was dispersed or clustered in the *C. albicans* genome (Fig. 5). The tandem-repeat-specific probe hybridized to chromosomes 6 and R in strains with a karyotype similar to reference strains (Chu et al.

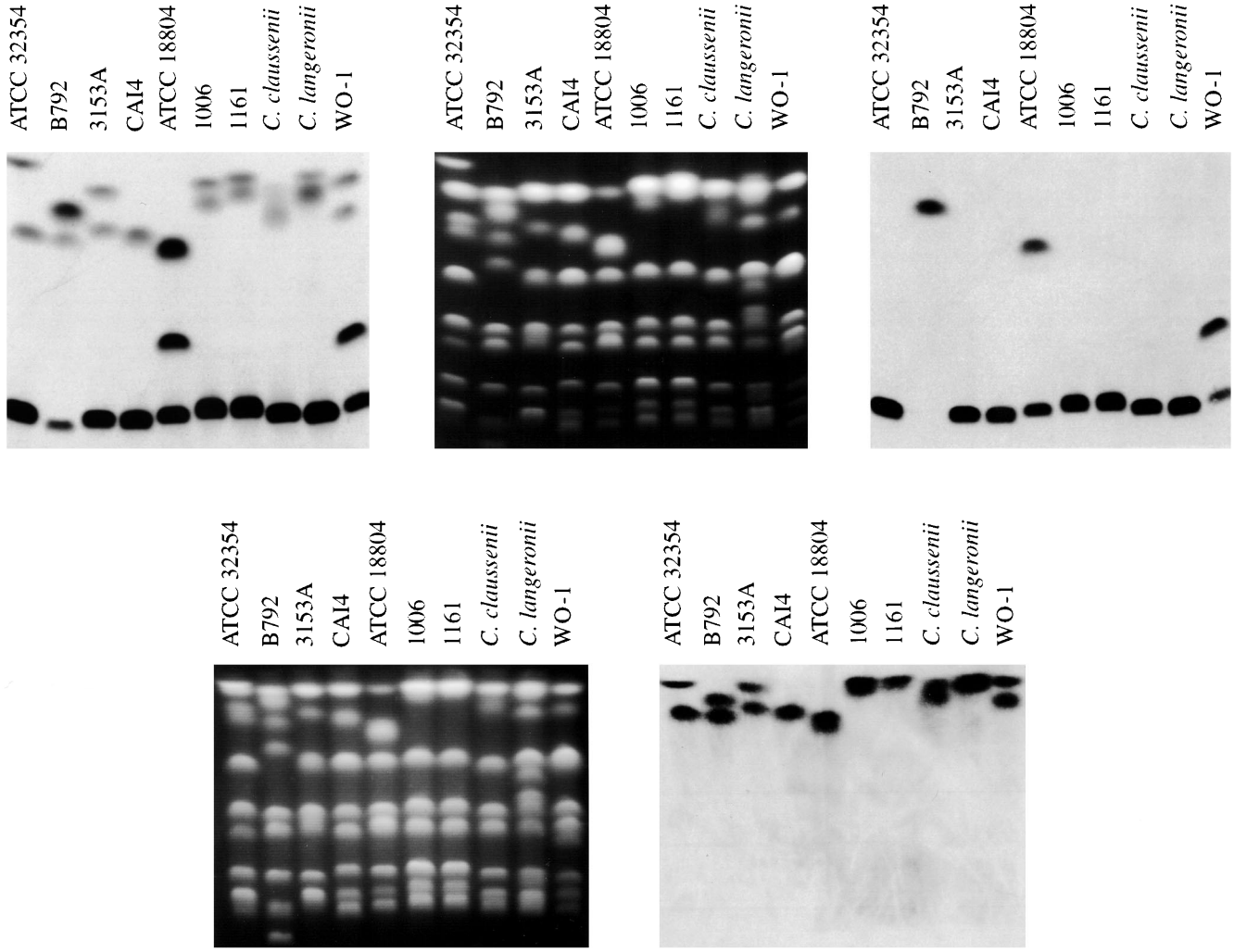

Fig. 5 Chromosomal localization of ALS genes. *C. albicans* chromosomes were separated by pulsed-field gel electrophoresis and visualized by ethidium-bromide staining (top center and bottom left). Gels were Southern blotted and probed with the 870-bp *Kpn*I tandem-repeat-specific fragment (top left), the *Xba*I-*Hin*dIII *ALS1*-specific fragment (top right), or the *ALS3*-specific fragment (bottom right). Blots are matched with the ethidium bromide-stained gel from which they were derived to provide an exact reference to specific chromosomes

1993). Hybridization to other chromosomes or chromosomal fragments was explained by previously described translocation events in strains for which karyotype analysis has been performed, including B792 (Thrash-Bingham and Gorman 1992) and WO-1 (Chu et al. 1993); karyotype analysis has not been performed for strain ATCC 18804. The *Xba*I-*Hin*dIII *ALS1*-specific probe hybridized only to chromosome 6, or fragments derived from chromosome 6, while *ALS3*-specific fragments were only found on chromosome R. Further analysis indicated that *ALS1* was found on *Sfi*I fragment "O" while *ALS3* was located on the "S-U" fragment (Chu et al. 1993; data not shown). These data suggested that the ALS gene family is dispersed in the *C. albicans* genome.

Number of genes in the ALS family

Previous studies demonstrated that *ALS1*-hybridizing fragments are present in genomic DNA from a variety of *C. albicans* and *C. stellatoidea* strains (Hoyer et al. 1995). A similar Southern blot of *Bgl*II-digested genomic DNA indicated that *ALS3*-hybridizing fragments were also present in all strains tested (Fig. 6). Probing the same blot with the tandem-repeat sequences showed that the *ALS1*- and *ALS3*-hybridizing fragments were a subset of those that hybridize with the tandem repeats (Fig. 6). Other tandem-repeat-hybridizing fragments were accounted for by additional fosmid clones isolated using the PCR screening technique (Hoyer, Payne, and Hecht, submitted).

Probing *Bgl*II-digested genomic DNA with the tandem-repeat element showed that the number of repeat-hybridizing fragments was apparently variable across strains (Hoyer et al. 1995, Fig. 6). This variability raised the possibility that each *C. albicans* strain may have a different complement of ALS genes. However, in strains such as WO-1, where fewer distinct tandem-repeat-hybridizing fragments were observed, *ALS1*- and *ALS3*-specific fragments comigrated (Fig. 6). Better separation of fragments greater than 12 kb indicated that there were many tandem-

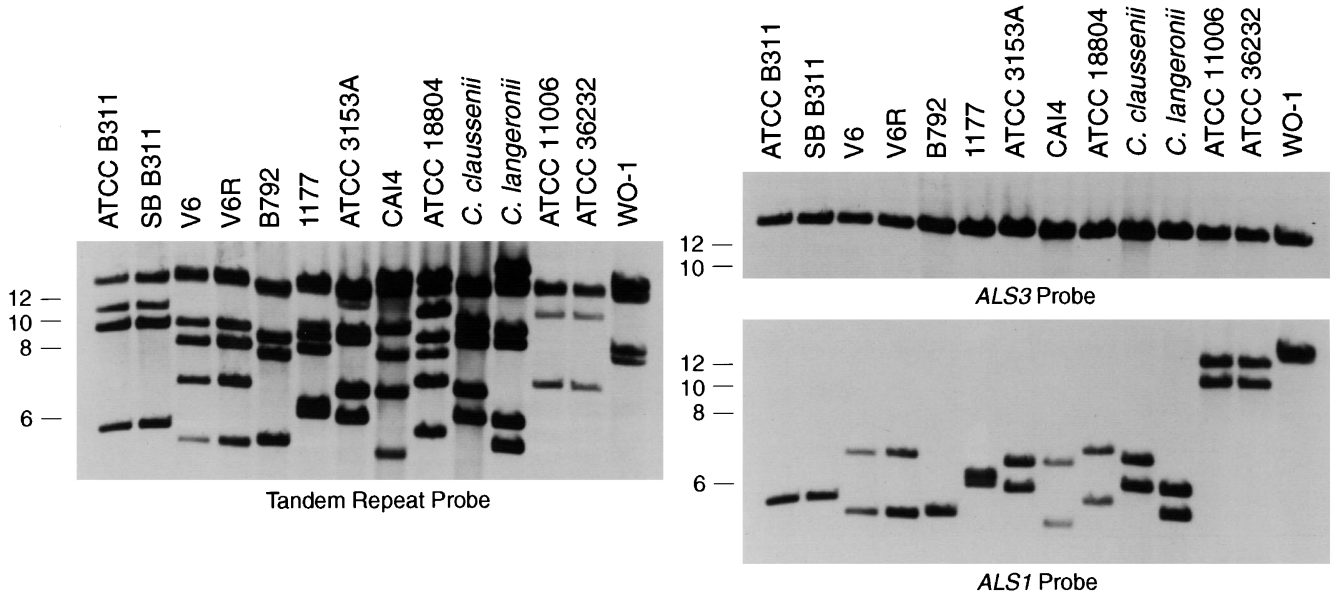

Fig. 6 Southern blots of *Bgl*II-digested genomic DNA from a variety of *C. albicans* and *C. stellatoidea* species. This blot was probed, in succession, with the *Xba*I-*Hin*dIII *ALS1*-specific fragment, the *ALS3*-specific fragment, and the 870-bp *Kpn*I tandem-repeat-specific probe. The blot was stripped between probing with each of these fragments. Size markers (in kb) are shown at the left of each blot

repeat-hybridizing fragments in this size range in a variety of strains and demonstrated that simple counting of fragments was not sufficient to estimate the number of ALS genes present in a particular strain (Fig. 7).

Genes in the ALS family were originally defined by hybridization with the tandem-repeat motif (Hoyer et al. 1995). Comparison of the *ALS1* and *ALS3* sequences showed that the region 5′ of the tandem-repeat domain was also conserved (Fig. 2). In a variety of strains, a *Kpn*I-*Hpa*I fragment derived from this 5′ region of *ALS1* (Fig. 2) hybridized to more genomic *Bgl*II fragments than did the tandem-repeat probe (Fig. 7). Characterization of these fragments will indicate whether they encode additional ALS-like genes.

Discussion

ALS genes are composed of three domains: a central tandem-repeat domain composed of variable numbers of

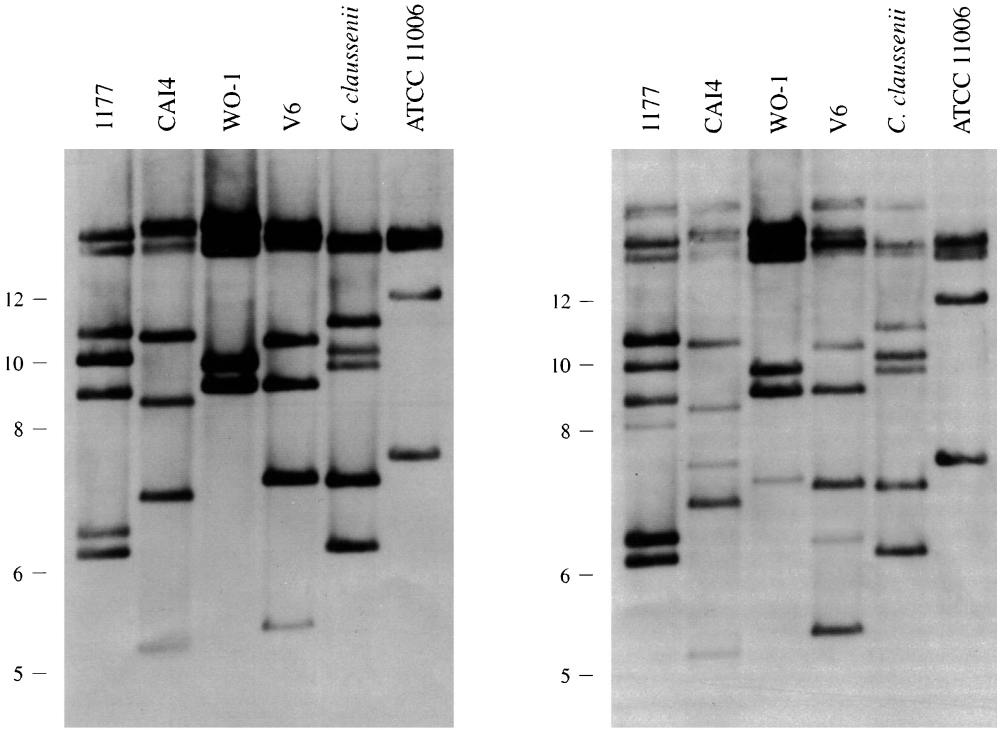

Fig. 7 Southern blots of *Bgl*II-digested genomic DNA. *Bgl*II fragments of genomic DNA from the strains shown were optimally separated in the size range above 12 kb by agarose-gel electrophoresis. The blot was first probed with the 870-bp *Kpn*I tandem-repeat-specific fragment (left), stripped and then probed with the 391-bp *Kpn*I-*Hpa*I fragment from the region 5′ of the tandem-repeat domain in *ALS1* (right). Size markers (in kb) are shown at the left of each blot

head-to-tail copies of the 108-bp sequence, sequences 5′ of this region, and sequences 3′ of the tandem repeats. The tandem-repeat sequences of *ALS3* are very similar to those of *ALS1*. This result was expected since *ALS3*-encoding clones were selected based on similarity to the *ALS1* tandem-repeat sequences. Sequences 5′ of the tandem-repeat domain are also highly conserved between the two genes.

Sequences 3′ of the tandem-repeat domain in each gene are less conserved. These sequences were exploited to develop unique probes for each gene. Although the 3′ sequences are more divergent, they predict proteins with a similar serine/threonine-rich composition. This amino-acid composition is also found in the tandem-repeat region and is consistent with abundant O-glycosylation of proteins (Jentoft 1990). Many consensus N-glycosylation sites are found in the tandem-repeats and the domain C-terminal of the tandem-repeats in Als1p and Als3p (Bause 1983). These observations suggest that the tandem-repeats and the domain C-terminal of the tandem-repeats are heavily N- and O-glycosylated. Abundant glycosylation of these domains is in contrast to the sequences N-terminal of the tandem-repeat domain which are predicted to be relatively free of glycosylation. The presence of a hydrophobic signal peptide (von Heijne 1990) and a GPI anchor addition site (Gerber et al. 1992; Coyne et al. 1993) at the N- and C-terminal ends of both Als1p and Als3p suggests that both are cell-surface glycoproteins. This combination of features (sequence composition, glycosylation pattern, and hydrophobic termini) is found in other *C. albicans* and *S. cerevisiae* proteins known to be cell-surface glycoproteins (summarized by Bailey et al. 1996). Immunohistochemical staining of tissue sections from *C. albicans*-infected mice using an antiserum raised against N-terminal sequences of Als1p has confirmed the cell-surface localization of the Als proteins (Hoyer, Clevenger, Payne, Hecht, Ehrhart and Poulet, submitted).

Experiments to study the in vitro regulation of the ALS genes led to several conclusions. First, regulation of *ALS1* varies depending on which strain of *C. albicans* is examined. The original differential expression of *ALS1* between YPD and RPMI 1640 media is repeatable in strain B311, but, in four other strains studied, *ALS1* is expressed under both growth conditions. The increased abundance of the *ALS1*-specific signal in RPMI-1640-grown cells suggests that *ALS1* expression may be up-regulated under these growth conditions. In strain WO-1, *ALS1* was not expressed in either growth condition. Sequences upstream of *ALS1* in other strains are completely absent from the WO-1 genome suggesting that, in this strain, *ALS1* may not be expressed or may be regulated by a different mechanism. In addition, some strains produce two different-sized *ALS1*-specific messages suggesting that two different sizes of Als1p will be produced. Biochemical analysis of Als1p is in progress to verify this observation. These data indicate the potential for significant variability in the production of Als1p among different *C. albicans* strains.

Two different-sized *ALS3*-specific signals were observed on Northern blots of total RNA from *C. albicans* strains. Like *ALS1*, these signals represented differences in *ALS3* alleles and are due entirely to variation in the number of copies of the tandem-repeat sequence present. The *ALS3*-specific signal was only present in cultures with hyphal cells. Hyphal-specific expression of *ALS3* was confirmed by inducing hyphal formation with a variety of growth conditions in strains SC5314 and 3153A. Therefore, *ALS3* can be added to the list of hyphal-specific genes isolated from *C. albicans*, which includes *ECE1* (Birse et al. 1993), *HWP1* (Staab et al. 1996) and *HYR1* (Bailey et al. 1996). From these data, it is clear that *ALS1* and *ALS3* respond to different regulatory signals. A similar theme of differential regulation of a gene family has been described for the SAP genes of *C. albicans* (Hube et al. 1994). The above data suggest that the profile of Als proteins produced by a given *C. albicans* strain is likely to be highly variable and dependent on growth conditions and morphologic form.

Characterization of *ALS3* indicated that other tandem-repeat-hybridizing genomic fragments encode *ALS1*-like genes and define a gene family. Similar studies of other fosmids isolated by the PCR screening technique revealed ALS-like genes; together, these fosmids account for all of the *ALS1*-tandem-repeat-hybridizing fragments in *C. albicans* strain 1161 (Hoyer, Payne and Hecht, submitted). Comparison of the *ALS1* and *ALS3* sequences indicated that the region 5′ of the tandem-repeat domain is also highly conserved. In genomic Southern blots, a *Kpn*I-*Hpa*I fragment derived from this conserved 5′ region hybridizes to the same fragments that hybridize with the tandem-repeats, and also to additional genomic fragments. This observation suggests that the ALS gene family may be even larger than originally estimated. Characterization of these additional fragments could yield new probes and further expand the ALS family. Characterization of new fosmid clones isolated by direct hybridization with the *Kpn*I-*Hpa*I probe is in progress. At this time, it is difficult to estimate the total number of genes in the ALS family.

Estimating the size of the ALS family is important in order to strategize how best to study the effect of Als proteins on *C. albicans* adhesion and pathogenesis. By comparison to *S. cerevisiae* alpha-agglutinin, it is hypothesized that Als proteins have an adhesive function, with adhesion to host tissues being the most attractive possibility (Lipke et al. 1989; Hoyer et al. 1995). This hypothesis is supported by the work of Gaur and Klotz (1997) who demonstrated that expression of an ALS-like gene in non-adherent *S. cerevisiae* conferred an adherent phenotype to extracellular matrix proteins and buccal epithelial cells. Since adhesion has been positively correlated with pathogenicity (reviewed by Calderone and Braun 1991), Als proteins could make a direct contribution to the disease process. Because of the burgeoning size of the ALS family, constructing true null als mutants in *C. albicans* will be difficult. Therefore, the function of individual Als proteins may be best tested by heterologous expression of ALS genes in *S. cerevisiae*, after considering differences in the genetic code between the organisms (Santos and Tuite 1995; White et al. 1995). However, studies to determine the contribution of Als proteins to pathogenesis must still rely on *C. albicans* con-

structs. The cumulative evidence from these experiments will further define the nature of the ALS gene family, the function of the encoded proteins, and the relevance of the gene family to the adhesion and pathogenesis of *C. albicans*.

Acknowledgements The authors thank Michael Strathmann for the fosmid library which greatly facilitated this study. We also acknowledge the staff of the Iowa State University DNA Sequencing and Synthesis Facility for sequencing the *ALS3* gene. This work was supported by Public Health Service Grant AI23850 (to S. S.) and AI39441 (to L. L. H.), by Cooperative State Research, Education and Extension, U.S. Department of Agriculture, under Project No. ILLU-70-0305 (to L. L. H.), and by the University of Illinois Campus Research Board (to L. L. H.).

References

Bailey DA, Feldman PJF, Bovey M, Gow NAR, Brown AJP (1996) The *Candida albicans HYR1* gene, which is activated in response to hyphal development, belongs to a gene family encoding yeast cell wall proteins. J Bacteriol 178: 5353–5360

Bause E (1983) Structural requirements of N-glycosylation of proteins. Biochem J 209: 331–336

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7: 1513–1523

Birse CE, Irwin MY, Fonzi WA, Sypherd PS (1993) Cloning and characterization of *ECE1*, a gene expressed in association with cell elongation of the dimorphic pathogen *Candida albicans*. Infect Immun 61: 3648–3655

Calderone RA, Braun PC (1991) Adherence and receptor relationships of *Candida albicans*. Microbiol Rev 55: 1–20

Chu W-S, Magee BB, Magee PT (1993) Construction of an *Sfi*I macrorestriction map of the *Candida albicans* genome. J Bacteriol 175: 6637–6651

Collart MA, Oliviero S (1993) Preparation of yeast RNA. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (eds). Current protocols in molecular biology, vol. 2. John Wiley and Sons, New York London, pp 13.12.1–13.12.5

Coyne KE, Crisci A, Lublin DM (1993) Construction of synthetic signals for glycosyl-phosphatidylinositol anchor attachment. J Biol Chem 268: 6689–6693

Dabrowa N, Howard DH (1981) Proline uptake in *Candida albicans*. J Gen Microbiol 127: 391–397

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

Fonzi WA, Irwin MY (1993) Isogenic strain construction and gene mapping in *Candida albicans*. Genetics 134: 717–728

Gaur NK, Klotz SA (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

Gerber LD, Kodukula K, Udenfriend S (1992) Phosphatidylinositol glycan (PI-G)-anchored membrane proteins: amino-acid requirements adjacent to the site of cleavage and PI-G attachment in the COOH-terminal signal peptide. J Biol Chem 267: 12168–12173

Gerring SL, Connelly C, Hieter P (1991) Positional mapping of genes by chromosome blotting and chromosome fragmentation. Methods Enzymol 194: 57–77

Goshorn AK, Scherer S (1989) Genetic analysis of prototrophic natural variants of *Candida albicans*. Genetics 123: 667–673

Hauser K, Tanner W (1989) Purification of the inducible α-agglutinin of *S. cerevisiae* and molecular cloning of the gene. FEBS Lett 225: 290–294

Heijne G von (1990) Protein targeting signals. Curr Opin Cell Biol 2: 604–608

Hoyer LL, Scherer S, Shatzman AR, Livi GP (1995) *Candida albicans* ALS1: domains related to a *Saccharomyces cerevisiae* sexual agglutinin separated by a repeating motif. Mol Microbiol 15: 39–54

Hube B, Monod M, Schofield DA, Brown AJP, Gow NAR (1994) Expression of seven members of the gene family encoding secretory aspartyl proteinases in *Candida albicans*. Mol Microbiol 14: 87–99

Hube B, Sanglard D, Odds FC, Hess D, Monod M, Schafer W, Brown AJ, Gow NA (1997) Disruption of each of the secreted aspartyl proteinase genes *SAP1*, *SAP2*, and *SAP3* of *Candida albicans* attenuates virulence. Infect Immun 65: 3529–3538

Jentoft N (1990) Why are proteins *O*-glycosylated? Trends Biochem Sci 15: 291–295

Kim U-J, Shizuya H, de Jong PJ, Birren B, Simon MI (1992) Stable propagation of cosmid-sized human DNA inserts in an F factor-based vector. Nucleic Acids Res 20: 1083–1085

Kraft R, Tardiff J, Krauter KS, Leinwand LA (1988) Using miniprep plasmid DNA for sequencing double-stranded templates with Sequenase™. Biotechniques 6: 544–547

Lipke PN, Wojciechowicz D, Kurjan J (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

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

Odds FC (1988) *Candida* and candidosis. Balliere Tindall, London

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Sanglard D, Hube B, Monod M, Odds FC, Gow NA (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

Santos MAS, Tuite MF (1995) The CUG codon is decoded in vivo as serine and not leucine in *Candida albicans*. Nucleic Acids Res 23: 1481–1486

Scherer S, Magee PT (1990) Genetics of *Candida albicans*. Microbiol Rev 54: 226–241

Shepherd MG, Yin CY, Ram SP, Sullivan PA (1980) Germ tube induction in *Candida albicans*. Can J Microbiol 26: 21–26

Staab JF, Ferrer CA, Sundstrom P (1996) Developmental expression of a tandemly repeated proline- and glutamine-rich amino-acid motif on hyphal surfaces of *Candida albicans*. J Biol Chem 271: 6298–6305

Sundstrom P, Smith D, Sypherd PS (1990) Sequence analysis and expression of the two genes for elongation factor 1α from the dimorphic yeast *Candida albicans*. J Bacteriol 172: 2036–2045

Thrash-Bingham C, Gorman JA (1992) DNA translocations contribute to chromosome length polymorphisms in *Candida albicans*. Curr Genet 22: 93–100

White TC, Andrews LE, Maltby D, Agabian N (1995) The "universal" leucine codon CTG in the secreted aspartyl proteinase 1 (*SAP1*) gene of *Candida albicans* encodes a serine in vivo. J Bacteriol 177: 2953–2955

Wickes BL, Golin JE, Kwon-Chung KJ (1991a) Chromosomal rearrangement in *Candida stellatoidea* results in a positive effect on phenotype. Infect Immun 59: 1762–1771

Wickes BL, Staudinger J, Magee BB, Kwon-Chung KJ, Magee PT, Scherer S (1991b) 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

Wickes BL, Hicks JB, Merz WG, Kwon-Chung KJ (1992) The molecular analysis of synonymy among medically important yeasts within the genus *Candida*. J Gen Microbiol 138: 901–907

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains; nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103–119