A Physical Map of Chromosome *7* of *Candida albicans*

Hiroji Chibana,* B. B. Magee,* Suzanne Grindle,* Ye Ran,*,[1] Stewart Scherer[†,2] and P. T. Magee*

**Department of Genetics and Cell Biology, University of Minnesota, St. Paul, Minnesota 55108 and [†]Department of Microbiology, University of Minnesota, Minneapolis, Minnesota 55455*

Manuscript received February 2, 1998
Accepted for publication May 20, 1998

ABSTRACT

As part of the ongoing *Candida albicans* Genome Project, we have constructed a complete sequence-tagged site contig map of chromosome *7*, using a library of 3840 clones made in fosmids to promote the stability of repeated DNA. The map was constructed by hybridizing markers to the library, to a blot of the electrophoretic karyotype, and to a blot of the pulsed-field separation of the *Sfi*I restriction fragments of the genome. The map includes 149 fosmids and was constructed using 79 markers, of which 34 were shown to be genes via determination of function or comparison of the DNA sequence to the public databases. Twenty-five of these genes were identified for the first time. The absolute position of several markers was determined using random breakage mapping. Each of the homologues of chromosome *7* is approximately 1 Mb long; the two differ by about 20 kb. Each contains two major repeat sequences, oriented so that they form an inverted repeat separated by 370 kb of unique DNA. The repeated sequence CARE2/Rel2 is a subtelomeric repeat on chromosome *7* and possibly on the other chromosomes as well. Genes located on chromosome *7* in Candida are found on 12 different chromosomes in *Saccharomyces cerevisiae*.

CANDIDA albicans is an opportunistic fungal pathogen that has become a greater and greater medical problem as the number of immunocompromised patients increases due to rising numbers of HIV infections and more aggressive medical procedures, such as organ transplants and cancer chemotherapy. Despite its prevalence as a pathogen, little is known about the characteristics that allow this organism to shift from the commensal to the virulent state. *C. albicans* has a number of special properties that have been proposed to account for the pathogenicity; many of these are of great biological interest, independent of their medical importance. However, no property of *C. albicans* has been shown to be a virulence factor in the sense that it is sufficient to cause disease or that its absence always renders a pathogenic strain avirulent. It seems most likely that virulence will turn out to be a quantitative trait caused by the contributions of a number of genes (Cutler 1991).

In common with a number of pathogenic fungi, *C. albicans* is able to grow either as a yeast or as a filamentous fungus, and the mechanisms by which the transition between these forms (and one intermediate in structure called pseudohyphae, consisting of a string of elongated cells) is controlled are as yet unknown, although genes involved in these transitions have been identified (Kohler and Fink 1996; Braun and Johnson 1997; Leberer *et al.* 1997; Stoldt *et al.* 1997). A second kind of transition is manifested in variable colony morphology. The different distinctive shapes of colonies seem to be related to the propensity to form hyphae or pseudohyphae, but in this case neither the relationship between cell form and colony phenotype nor the regulatory mechanisms are known. There is no obvious environmental control of the phenotypic transition manifested by colony morphology (Soll *et al.* 1991).

At a deeper level, *C. albicans* is biologically interesting because it is usually isolated as a diploid or near diploid, but it has no known sexual cycle (Scherer and Magee 1990). In addition, there is a growing body of evidence that its chromosomes undergo relatively frequent rearrangement, often with an effect on phenotype (Suzuki *et al.* 1989; Wickes *et al.* 1991; Rustchenko *et al.* 1997). Hence, it is an organism concerning which questions about the mechanism of chromosomal translocation, the evolutionary costs and benefits of obligate diploidy, and ways of achieving genetic variability in the absence of meiosis can be addressed.

Genetic manipulation of *C. albicans* is best carried out by molecular approaches: parasexual approaches are possible but cumbersome. It is therefore difficult to establish a genetic map by means of recombination frequencies. However, many genes have been cloned from this organism and the number is increasing rapidly. A physical map would help organize these data and provide much useful information about the genetics of the organism. Furthermore, with the recent achievement

Corresponding author: P. T. Magee, Department of Genetics and Cell Biology, University of Minnesota, 1445 Gortner Ave, St. Paul, MN 55108. E-mail: ptm@biosci.cbs.umn.edu

[1] *Present address:* 2211 Draper Ave, Roseville, MN 55113.

[2] *Present address:* Acacia Biosciences, 4136 Lakeside Drive, Richmond, CA 94806.

of the complete sequence of the genome of the relatively closely related yeast *Saccharomyces cerevisiae* (Goffeau *et al.* 1996), much can be learned by comparison of the genomes of the two organisms.

The complete sequence of the *C. albicans* genome is being determined at Stanford University; present plans are to sequence the equivalent of 1.5 genome equivalents (~25 Mb). This level of coverage will leave many gaps and ambiguous areas near repeated genes. A physical map will allow the assignment of contigs of sequence to chromosomes and will order them with respect to one another.

Finally, the high degree of variation in the electrophoretic karyotype of *C. albicans* clinical isolates may be due to translocation between nonhomologous chromosomes (Chu *et al.* 1993) or to partial aneuploidy (Magee and Magee 1997). A knowledge of the physical structure of the chromosomes is essential to understanding both the origins and the biological implications of this phenomenon.

We therefore set out to develop a physical map of the genome of *C. albicans.* In this article we describe the complete physical map of chromosome *7* from this organism. We used as our starting point the macrorestriction map for the 8-bp-specific restriction enzyme *Sfi*I derived by Chu *et al.* (1993). There are three sites for *Sfi*I on chromosome *7*, yielding four fragments: C, A, F, and G. The *Sfi*I sites between A and F and F and G are located in the repeated sequence RPS (Iwaguchi *et al.* 1992), itself included in a larger repeat. We have named this larger cluster the major repeat sequence (MRS), because it is highly complex; it contains the RPS sequence, which is repeated 40 times in the genome of at least one typical Candida strain (FC18; Iwaguchi *et al.* 1992); and it occurs one or more times on seven of the eight chromosomes (Chindamporn *et al.* 1995). The *Sfi*I restriction site between C and A appears to occur in unique DNA. Our results here describe a sequence-tagged-site map of each of these fragments and describe several new genes identified in the process of this mapping effort. We also describe in detail the location and position of the middle repeat sequences on chromosome *7* and present new information about the structure of this chromosome.

MATERIALS AND METHODS

Strains: The strains we have used are 1006 (Goshorn and Scherer 1989) and 1161. The latter is derived from 1006 by ultraviolet irradiation and selection on 2-deoxygalactose for a homozygous (*gal1*/*gal1*) mutant.

Fosmid library: The library we are using in our physical mapping efforts was constructed by M. Strathmann and consists of 3840 clones in 40 microtiter dishes. We are currently using the first 20 plates of the library, which is equivalent to approximately five haploid *C. albicans* genomes. The clones are a *Sau*3AI partial digestion of strain 1161 DNA inserted into a fosmid vector (Kim *et al.* 1992). This vector, like the better-known cosmids, contains inserts of approximately 40 kpb. The advantage of a fosmid library is that the vector exists in only one or a few copies per bacterial cell; hence, tandemly repeated DNA, of which *C. albicans* contains significant amounts, is relatively stable in such a vector. The clones of the library are numbered by plate, row, and column. Thus, 12G9 is the clone found in the 9th well of the 7th row of the 12th plate in the library. The vector has a T7 polymerase at one side of the cloning site and a (nonfunctional) SP6 promoter at the other. End probes of the inserts are designated by the fosmid number preceded by T or S, depending on which side of the insert they came from.

For probing, filter replicas were made by gridding the library out on five 8 × 11-cm MSI Magnagraph filters (Micron Separations, Inc., Westborough, MA), using a 96-point replicator, with four plates printed on each filter. The filters were then transferred to 15-cm-diameter petri dishes containing Luria broth plus chloramphenicol (20 μg/ml) agar (Sambrook *et al.* 1989) and put at 37° for 24 hr. The filters were then removed and treated by denaturation and neutralization according to the manufacturer's instructions.

Preparation of fosmid DNA for sequencing: *E. coli* cells containing fosmids were cultured in 250 ml of Terrific broth (TB) broth plus 20 μg/ml chloramphenicol (Sambrook *et al.* 1989) at 37° for 20 to 24 hr. After centrifugation, the pellet was rinsed with 25 ml STE (0.1 m NaCl; 10 mm Tris-Cl, pH 8.0; 1 mm EDTA, pH 8.0) or distilled water and resuspended completely in 5 ml of solution I (3.3 mm EDTA, 17 mm glucose, 8.3 mm Tris-HCl, pH 8.0) at room temperature for 10 min. Next 10 ml Solution II (1% SDS, 0.2 N NaOH) was added and the tube quickly inverted three times and put on ice for 10 min. Then 7.5 ml Solution III [3 m KOAc, 11.5% (v/v) glacial acetic acid] was added, and the tube quickly inverted three times more and put on ice for 30 min. The solution was centrifuged at 12,000 × *g* at 4° for 10 min. The supernatant was decanted into a 40-ml tube and mixed with an equal volume of phenol/chloroform. The resulting suspension was centrifuged at 10,000 × *g* for 5 min. The supernatant was removed into a new 40-ml tube, mixed with 13.5 ml isopropanol, and then centrifuged at 16,000 × *g* at 10° for 10 min. The pellet was dissolved in 400 μl TER (10 mm Tris-HCl, pH 8.0; 1 mm EDTA; 100 μg/ml RNase) and incubated at 37° for 30 min. After 40 μl 3 m NaOAc and 800 μl EtOH were added with mixing, the suspension was centrifuged at 16,000 × *g* at room temperature for 10 min. The pellet was dried and resuspended in 420 μl distilled water, and 80 μl 5 m NaCl and 500 μl 13% PEG8000 were added. The resulting suspension was left at 4° overnight, then centrifuged at 16,000 × *g* at room temperature for 10 min. The supernatant was removed, and the pellet was rinsed with 500 μl 70% EtOH. The pellet was resuspended in 39 μl of distilled water. Eleven microliters of the DNA solution was used for each sequencing reaction.

Preparation of fosmid DNA for probing: *E. coli* cells containing fosmids were cultured in 40 ml of TB broth at 37° for 20 to 24 hr. After centrifugation they were resuspended in 4 ml solution I at room temperature for 8 min. Next 10 ml Solution II was added and the tube was quickly mixed by being inverted five times and put on ice for 1 min. Then 6 ml of 7.5 m NH_4OAc was added and the tube quickly mixed and incubated on ice (3 min) as before. The suspension was centrifuged at 12,000 × *g* at room temperature for 10 min. The supernatant was mixed with 11 ml isopropanol in a 40-ml tube, and then centrifuged at 12,000 × *g* at room temperature for 10 min. The pellet was dissolved in 700 μl TER and incubated at room temperature for 30 min. Phenol/chloroform (700 μl) was added; the tube was mixed well and then centrifuged at 12,000 × *g* at room temperature for 15 min.

The upper layer was removed into a new tube and 70 μl 7.5 m NH_4OAc and 700 μl EtOH were added. The suspension was mixed well and quickly centrifuged at 12,000 × *g* at room temperature for 5 min. When dry, the pellet was resuspended in 30 μl TE. Ten and 20 μl of the DNA solution were used for preparation of the T-end probe and the S-end probe, respectively.

Probes: The probes are listed in Table 1 and/or on the *C. albicans* Information Page at http://alces.med.umn.edu/Candida.html/. For the most part they are fragments of cloned genes or fragments which show homology with genes in the public data bases. Although the original fosmid vector has both an SP6 promoter adjacent to the cloning site on one side and a T7 promoter on the other side, sequencing of one of the clones showed that the SP6 site was damaged on the vector used for the library. Therefore, RNA probes were prepared from the T7 end of the insert using the Ambion MAXIscript kit and were labeled with the number of the fosmid preceded by T. The S probes were prepared by a method that takes advantage of the fact that the insert in the fosmid is flanked by *Not*I sites. The *Hin*dIII cloning site in the fosmid vector is on the T7 promoter side of the insert next to that *Not*I site. Hence, comparison of the bands generated by a *Hin*dIII digest to those from a *Not*I-*Hin*dIII double digest identifies a unique band which is a fragment of the insert adjacent to the SP6 promoter. These bands were used as the S probes.

Whole-chromosome probes: Whole chromosomes were separated on a CHEF gel. The areas containing chromosomes *6* and *7* were excised from the gel separately and the pieces of the gel irradiated at 260 nm with a dose of 2 J/m^2. The gel pieces were then placed in a 0.5 N NaOH solution and incubated for 20 min at room temperature with shaking. The NaOH was decanted and the gel pieces washed twice with TE. The DNA was extracted from the gels with phenol-chloroform and labeled as usual.

Generation of the contig map: The contig map was generated by radioactive labeling of one of the probes described above followed by its hybridization to blots of CHEF separations of the chromosomes and the fragments generated by *Sfi*I digestion of the genome (Chu *et al.* 1993) and to filter replicas of the fosmid library. Where necessary to walk to the next contig, probes were prepared from the ends of the inserts of the fosmids located at gaps in the map.

Random breakage mapping: The contig map gives the order of markers, but it does not give physical distances, except as an approximation. Random breakage mapping was therefore used to determine distances between the markers and to measure the size of the chromosome (Game *et al.* 1990). DNA prepared for pulsed-field gels was irradiated in a ^{137}Cs source for times up to 8 hr at 630 rem/min. It was then run under pulsed-field conditions designed to separate bands of the appropriate size. The bands were then blotted and probed to determine the distance of the probe from the end of the chromosome. Figure 2 shows the data from one such experiment that mapped the marker *DFR1*.

Pulsed-field separations: All separations were carried out on a Bio-Rad (Richmond, CA) CHEF DRIII instrument in 0.5× TBE buffer (Sambrook *et al.* 1989). To separate chromosomes, the following conditions were set: 0.9% agarose, 60 to 300 sec, 4.5 V/cm, 120° for 36 hr, followed by 720 to 900 sec, 2.0 V/cm, 106° for 24 hr. To separate *Sfi*I fragments, the conditions were 0.9% agarose, 7 to 100 sec, 4.5 V/cm, 120° for 24 hr followed by 80 to 400 sec, 3.5 V/cm, 120° for 24 hr. All CHEF separations were carried out at 14°.

DNA preparation and labeling: DNA preparation for pulsed-field gels was essentially as described by Chu *et al.* (1993). Preparation of DNA probes for labeling was by the method of Heery *et al.* (1990). The DNA was labeled with [^{32}P]ATP as described by Feinberg and Vogelstein (1983).

RESULTS

Chromosome *7* contig map: The contig map of chromosome *7* was built in five steps:

1. At the beginning of this work 28 DNA markers were used to probe the fosmid library. Of these markers *ARG4*, *DFR 1*, *LEU2*, p212, and p282 have previously been shown to map to chromosome *7* (Chu *et al.* 1993). Ten DNA markers, *ALD1*, *ARS3*, *CHS1*, *CPY1*, *INP53*, *ISP42*, *YOR227*, *RBP1*, *SHM1*, *YDR231*, and 13 DNA markers with no known homology to previously identified genes were assigned to chromosome *7* in this work. The 28 DNA markers identified 118 fosmids. These fosmids provided 81% of the final coverage and fell into nine separate contigs.
2. Whole DNA of chromosome *6* and chromosome *7* isolated as described in materials and methods was labeled and used to probe a pUC18 *Hin*dIII library (S. Grindle, unpublished data) separately. Since the chromosomes of *C. albicans* include repeated DNA, *Hin*dIII clones that hybridized with the whole chromosome *7* probe but not with the whole chromosome *6* probe were chosen. Sequences identified from the *Hin*dIII library identified 10 new fosmids, raising the coverage of the chromosome to 88%; no new contigs were found.
3. Three new fosmids unique to chromosome *7* were identified by probing the fosmid library with chromosome *6* and *7* DNA as with the *Hin*dIII clones. T- and S-end probes from these fosmids in turn hybridized to nine new fosmids, increasing the coverage to 94%, connecting two contigs, and forming a new one as well.
4. The contigs were connected by walking with T and S probes from the ends of the contigs. The walks identified seven new fosmids and connected the nine previous contigs into two. Coverage at this point was 99%.
5. The second half of the library was screened with end probes from the fosmids adjacent to the remaining gap. Two fosmids were identified, each of which contained genes on both sides of the gap.

For the mapping effort, 79 markers were used as gene probes. These included 34 sequences identified as genes either by homology to known genes of the amino acid sequence encoded by the determined DNA sequence or by demonstration of function (Table 1). Twenty had no homology to known genes, and for 24 no sequence data were available. An additional 7 probes containing part of a repeated DNA fragment (MRS or CARE2/Rel2) were used. A total of 149 fosmids were identified by these markers. Eleven of 149 fosmid clones that hybridized with at least one probe of chromosome *7* hybridized with probes of other chromosomes as well. They are likely to be chimeric, containing material from other chromosomes.

TABLE 1

Genes and probes located on chromosome *7* of *Candida albicans*

Gene (function or homology)	Clone	Probe	Vector	*Sfi* I	Source
ACS2 (acetyl-CoA synthetase)	S11G4	Not tested[a]	Fosmid	C	This study
ALD1 (aldehyde dehydrogenase)	1808	4.0 kb E	pUC18	F	This study
ARG4 (argininosuccinate lyase)	1129	PCR	pUC18	G	Hoyer *et al.* (1994)[c]
ARS3 (Autonomous replication site)	pEH7	0.7 kb B-S1	pEMBL18 + tcm1	G	Herreros *et al.* (1992)
CDC34 (ubiquitin conjugating enzyme)	1888	1.5 kb H	pUC18	F	This study
CHS1 (chitin synthase)	pJA16	PCR		G	Au-Young and Robbins (1990)[c]
CPY1 (carboxypeptidase Y)	—	1.6 kb E-H	pGEM7zf	G	Mukhtar *et al.* (1992)
DBP7 (RNA helicase)	p343	3.8 kb H	pUC18	F	This study
DFR1 (dihydrofolate reductase)	p1879	4.9 kb H	pUC18	G	Daly *et al.* (1994)
EPT1 (ethanolamine phosphotransferase)	T2H7	T7 RNA	fosmid	F	This study
INP53 (inositol polyphosphate 5-phosphatase)	1810	4.2 kb E	pUC18	C	This study
ISP42 (mitochondrial import)	1439R	PCR		F	[c]
LEO1 (ORF of *S. cerevisiae*)	T1F12		Fosmid	C	This study
LEU2 (beta-IPM dehydrogenase)	pAL1	2.8 kb E	pUC18	C	L. delCastillo (personal communication)
LOS1 (tRNA splicing)	5H12	Not tested[a]	Fosmid	G	This study
NDH1 (NADH dehydrogenase 49-kd subunit)	1893	1.5 kb E	pUC18	F	This study
ODP2 (dihydrolipoamide S-acetyltransferase)	p54	2.0 kb H	pUC18	F	This study
POL3 (DNA polymerase III)	pTAN2	1.6 kb C	pRS316	G	Nolan and Rosamond (1996)
RBP1 (rapamycin-binding protein)	1819	PCR		F	[c]
SHM1 (serine hydroxymethyl-transferase)	RP1337	1.2 kb E-Sc	pBS	G	R. E. Pearlman, unpublished data
SLA2 (cytoskeleton assembly)	R2H4	2.9 kb H	pUC18	G	This study
SPL1 (tRNA splicing protein)	Not tested[b]				
YCF1 (metal resistance protein)	p306	6.5 kb H	pUC18	F	This study
YDL114 (ORF of *S. cerevisiae*)	p118	11 kb H	pUC18	F	This study
YDR231 (AMP binding protein)	1444R	PCR		F	[c]
YIL1 (ORF of *S. pombe and S. cerevisiae*)	T16A5	T7 RNA	Fosmid	A	This study
YJL54 (ORF of *S. cerevisiae*)	p1879	Not tested[b]	pUC18	G	This study
YJL107 (ORF of *S. cerevisiae*)	G3C11	1.9 kb H	pUC18	F	This study
YJR44 (ORF of *S. cerevisiae*)	R2D7	3.7 kb H	pUC18	G	This study
YML25 (ORF of *S. cerevisiae*)	S9F4	Not tested[a]	Fosmid	A	This study
YMR185 (ORF of *S. cerevisiae*)	S11B11	Not tested[a]	Fosmid	G	This study
YOR227 (ORF of *S. cerevisiae*)	1438R	PCR		F	[c]
YPL12 (ORF of *S. cerevisiae*)	p282	4.0 kb H	pUC18	A	This study
G2E10 (unknown)	G2E10	2.0 kb H	pUC18	G	This study
R2B9 (unknown)	R2B9	4.5 kb H	pUC18	C	This study

(*continued*)

TABLE 1

(Continued)

Gene (function or homology)	Clone	Probe	Vector	*Sfi*I	Source
Repeated sequences					
Ca7 (telomere)	pMM100	0.6 kb P	pBlueSK−	C, G	Sadhu *et al.* (1991)
Care2, Rel2 (subtelomere)	1972	5.5 kb E	pUC18	C, G	This study
HOK (isocitrate dehydrogenase)	pHOK	5.2 kb Sm-C	pBlueSK+	F	Chindamporn *et al.* (1997)
LRT2 (reverse transcriptase)	1972	0.7 kb PCR	pUC18	C	This study
RB2a (unknown)	pRB2	4.1 kb E	pBlueSK+	A,G	Chindamporn *et al.* (1997)
RPS620 (unknown)	pRPS62	2.1 kb P	pUC18	A, F, G	Chibana *et al.* (1994)
T9G4 (unknown)	9G4	T7 RNA	Fosmid	F	This study
820 (unknown)	820	4.0 kb E	pBR322		Scherer and Stevens (1988)
824 (27A; unknown)	824	6.0 kb E	pBR322		Scherer and Stevens (1988)

This table includes only those sequences homologous to known genes or ORFs and probes used for the random-breakage mapping. Their sequence data and other probes used in this work are shown on the Candida home page at http://alces.med.umn.edu/Candida.html. B, *Bam*HI; C, *Cla*I; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; Sc, *Sac*I; Sl, *Sal*I; Sm, *Sma*I.

[a] Since after completion of the contig map, homologous sequence with sequence data from GenBank was found on the fosmids that were assigned on the contig map, it was not used for mapping.

[b] Sequence homologous to SPL1 of *S. cerevisiae* was found in the sequence of Candida LEU2 gene in GenBank entry, and YJL54 was found in an unannotated region of the DFR1 sequence.

[c] The PCR probes were prepared with specific primers which were designed by us according to DNA sequence from GenBank. Sequences of the primers are shown in the web site. ARS sequences that were isolated by Herreros *et al.* (1992) are named ARS3 to distinguish them from other ARS sequences.

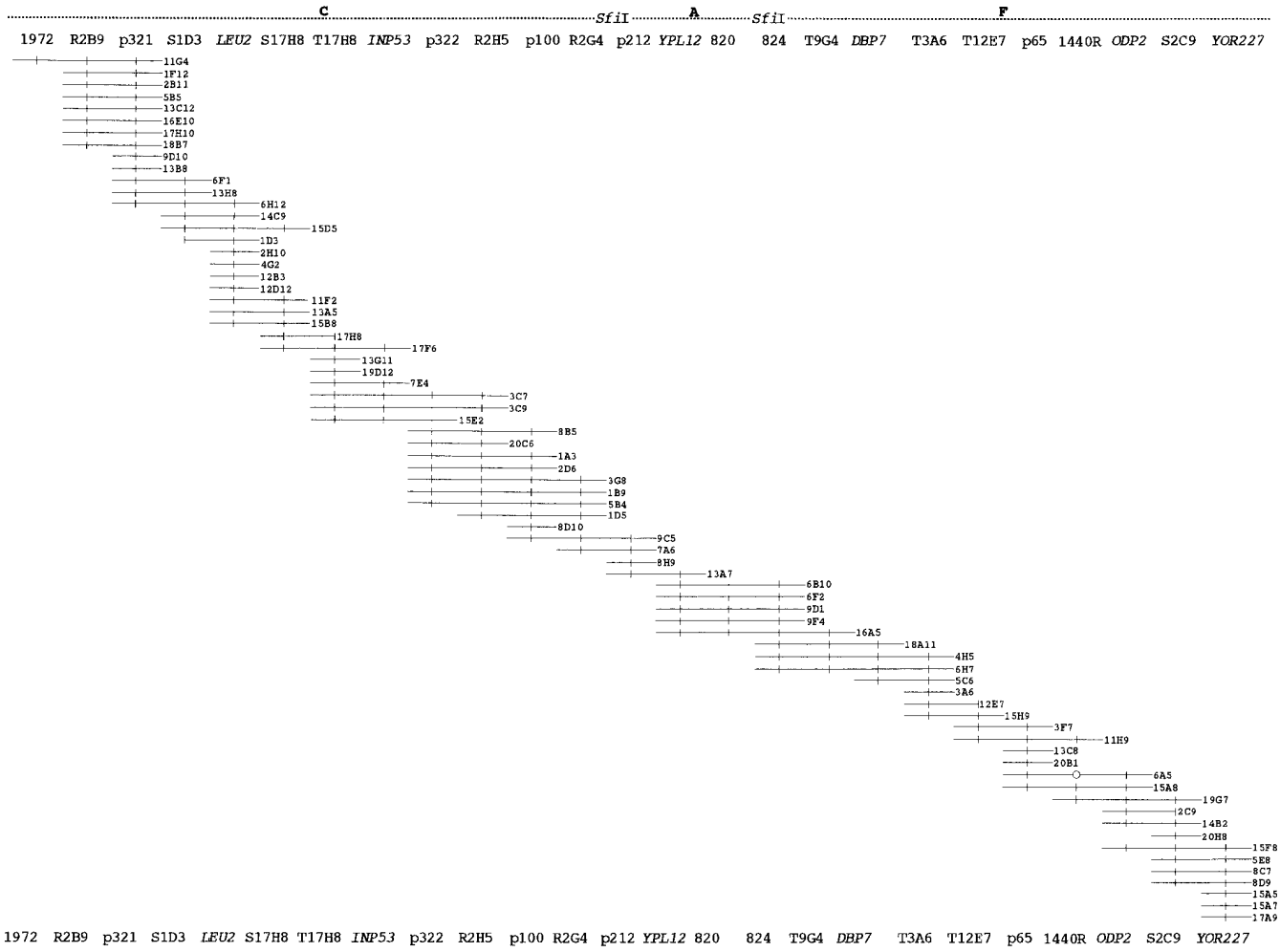

Figure 1.—Contig map of fosmids on chromosome *7*. *Sfi* I fragments are shown on the top by dotted lines. Fosmid clones are presented as horizontal lines. The fosmid clones that hybridized with each probe are marked by vertical lines. Open circles mean that the fosmid clone did not hybridize with the probe. The order of markers pCHR7 and CDC34, 1441R and YDR231, CPY1 and SHM1, and 1950 and ARG4, which are put in parentheses, is not known. The CHS1 and mARS probes are small and are likely to be internal to probe S8B4.

Figure 1 shows the complete contig map of chromosome *7*. Coverage is quite redundant except for the region between markers *YCF1* and T9G4 (the end of the MRS), where coverage is provided by just two fosmids, 13E11 and 8B4, which overlap only at one marker, T8B4. (This marker is the T7 end of the insert of 8B4.) A search of the second half of the library yielded one other fosmid which came from this region. Thus, the library appears to contain a random sampling of the genome.

Spacing of markers on the contig map: The contig map cannot give a reliable estimate of the distance between the markers on the chromosome, because we do not know where on the fosmids the markers are located and the fosmids vary in size in any case. We therefore mapped a series of markers on the chromosome by random breakage mapping. This technique involves induction of double-strand breaks in chromosomes by gamma irradiation. The broken chromosomes are then separated electrophoretically and blotted with a particular gene. If each chromosome molecule sustains on average less than one break, the gene will hybridize with all fragments bigger than the distance from the gene to the farthest telomere. It will not hybridize with any bands smaller than the distance from the gene to the nearest telomere. These distances show up as discontinuities in the blots. To increase our accuracy, we measured the distance of the markers from the *Sfi* I sites on chromosome *7*. Figure 2 shows a typical experiment: the first discontinuity (the end of the very heavy hybridization) occurs at about 270 kb and the second (the end of the intermediate hybridization), at 120 kb. From these data we estimated the location of *DFR1* to be 120 kb from the F-G *Sfi* I site and 270 kb from the telomere on fragment G. Figure 3 shows a map of chromosome *7* indicating the physical distance between a set of selected markers spaced about 100 kb apart. We estimate that the positions of these markers are accurate to less than

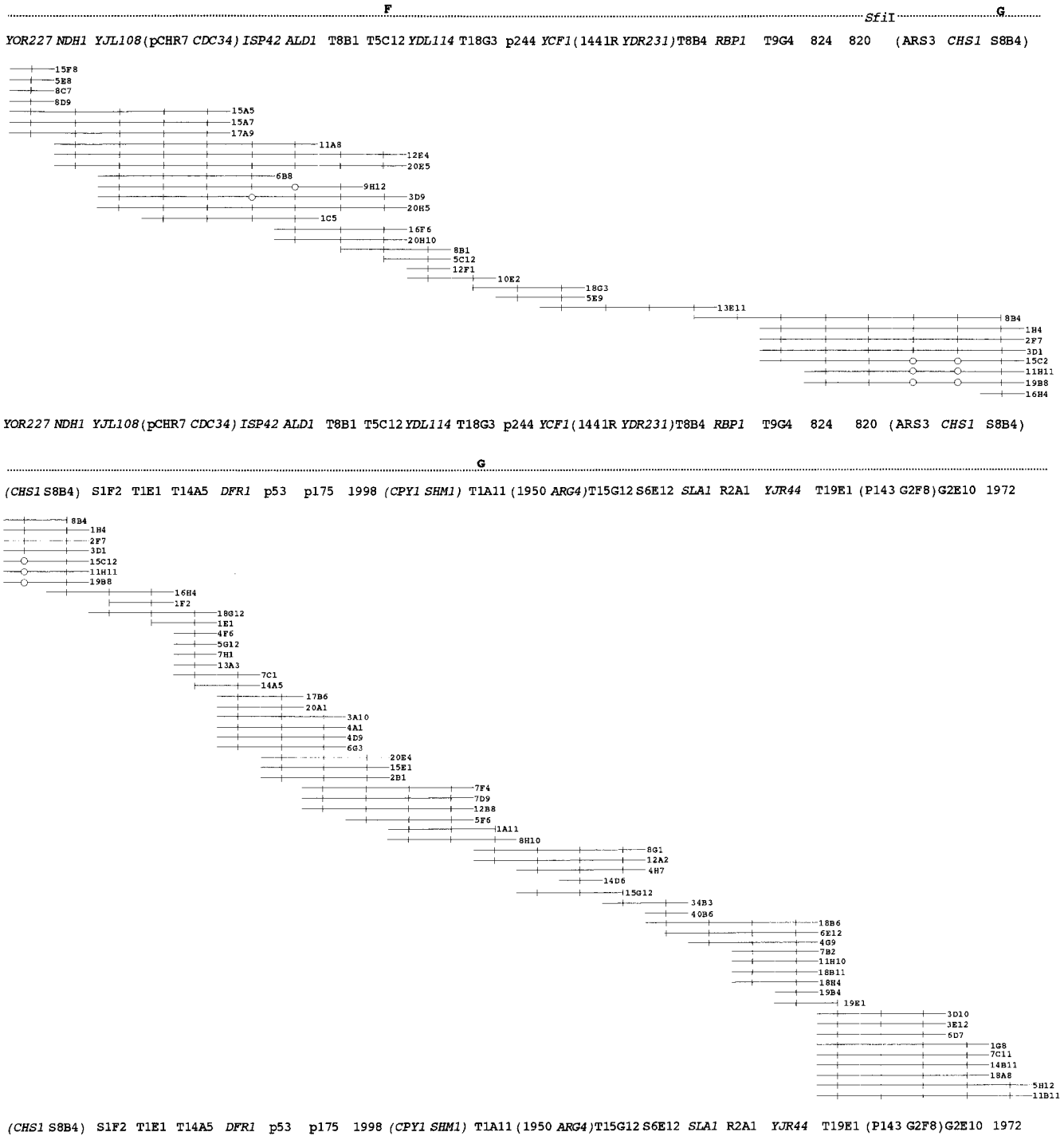

Figure 1.—*Continued.*

±5 kb. Locations of other markers on the chromosome were determined according to their order on the contig map.

Detailed analysis of the sizes of the intact chromosome *7* homologues and of the *Sfi*I fragments of chromosome *7* suggests that the two homologues of chromosome *7* in strain 1006 are 1030 and 1010 kb and that the *Sfi*I fragments of one homologue are C, 170 kb; A, 90 kb; RPS, 24 kb; F, 350 kb; RPS, 2 kb; and G, 390 kb. The *Sfi*I fragments of the other homologue are as follows: C, 180 kb; A, 90 kb; RPS, 2 kb; F, 350 kb; RPS, 2 kb; and G, 390 kb (H. Chibana, unpublished results). The RPS sequences are tandemly repeated in the A-F MRS on one homologue, so that that homologue is 22 kb larger than the other (see below). The homologue whose size is 1030 kb is shown in Figure 3. The assign-

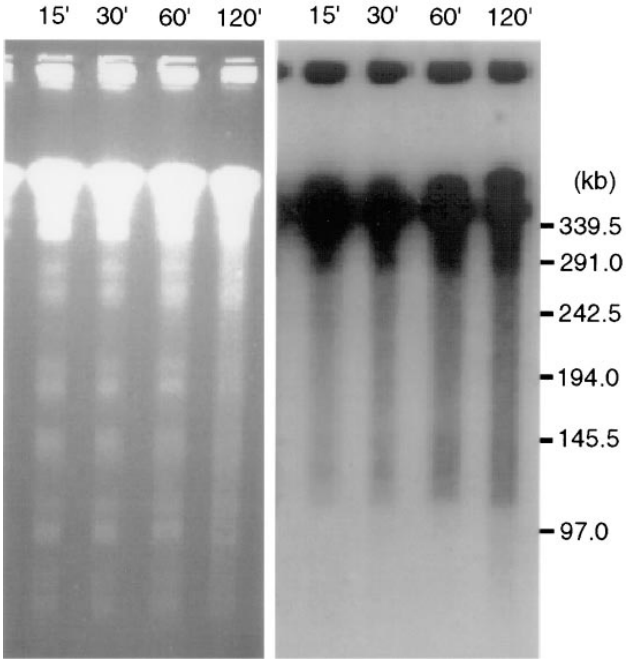

Figure 2.—Random breakage mapping. DNA was digested with *Sfi*I after irradiation in a ^{137}Cs source. The samples were loaded on a 0.9% agarose gel and run with a DRIII CHEF apparatus at 10 sec to 60 sec, 5 V/cm, 36 hr, 120°. A picture of the gel is shown on the left side. Autoradiography of the blot hybridized with a probe from the *DFR1* gene is shown on the right side. The 0 time sample appeared the same as the 15-min sample and is not shown. Lambda ladder DNA was used as a DNA size standard.

ments of fragments of a particular size to each homologue are based on: (1) careful measurement of the *Sfi*I fragment sizes; (2) determination of the MRS size by first using an enzyme, *Xho*I, which cuts outside the repeat, measuring the resulting MRS-containing restriction fragments, and then digesting with *Xho*I and *Sfi*I to determine the size of the fragments between the *Xho*I site and RPS; and (3) calculation of the combination of fragments that best fits the measured size of the homologues.

New and previously identified genes located on chromosome *7*: Table 1 lists the genes localized to chromosome *7*. The two known amino acid biosynthetic genes located on chromosome *7* are *LEU2* (L. delCastillo, personal communication) and *ARG4* (Hoyer *et al.* 1995). *LEU2* is located about 90 kb from the lefthand telomere, and *ARG4* is about 160 kb from the righthand telomere. Other interesting previously identified sequences or genes are *ARS3* (Herreros *et al.* 1992), *CHS1* (Au-Young and Robbins 1990), *CPY1* (Mukhtar *et al.* 1992), *DFR1* (Daly *et al.* 1994), *POL3* (Nolan and Rosamond 1996), and RBP1 (Ferrara *et al.* 1992).

One of the goals of the mapping project is to identify new *C. albicans* sequences corresponding to genes known in other organisms. The present map contains 21 new genes identified by one-pass sequencing. Clones

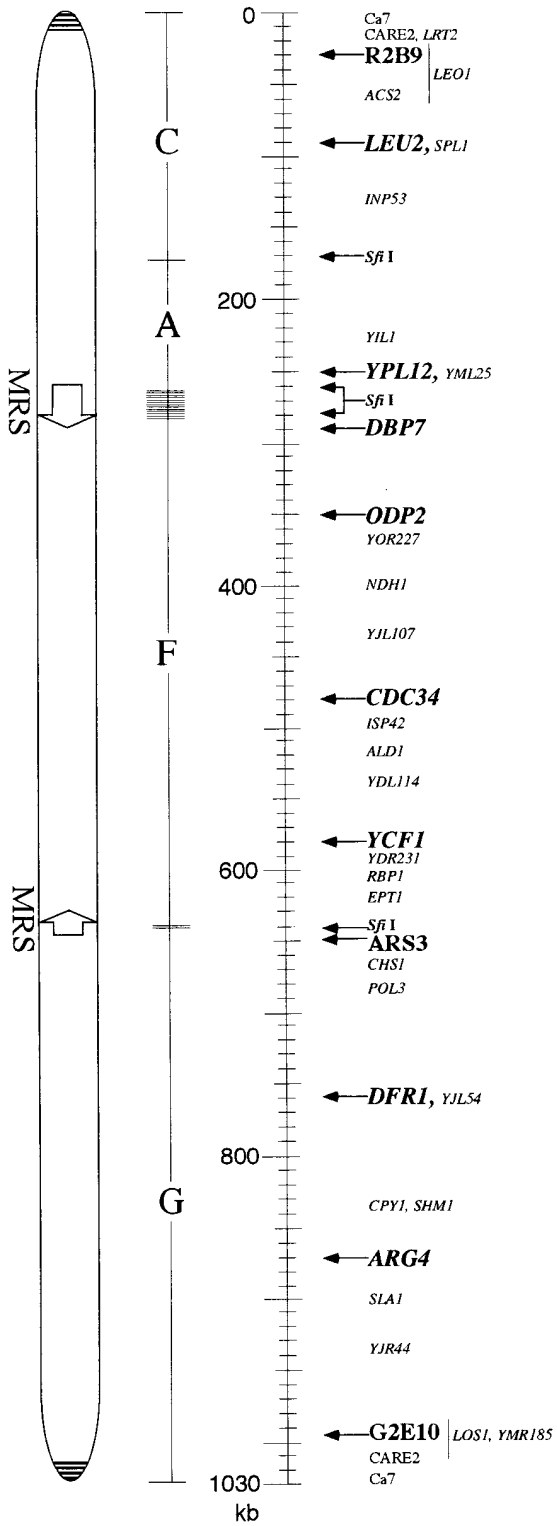

Figure 3.—Physical maps of chromosome *7*. DNA markers shown in boldface letters were determined by random breakage mapping. Their physical locations and *Sfi*I sites on the chromosome are indicated by arrows. The telomere repeated DNA sequence, Ca7, is located on the ends of the chromosome. SPL1 is adjacent to LEU2. Other markers were assigned on the chromosome according to the order of DNA markers on the contig map. LEO1, LOS1, and YMR185 were found on an end of fosmids (Table 1) that were assigned the contig map of chromosome *7*. The directions of the MRSs are shown with arrows.

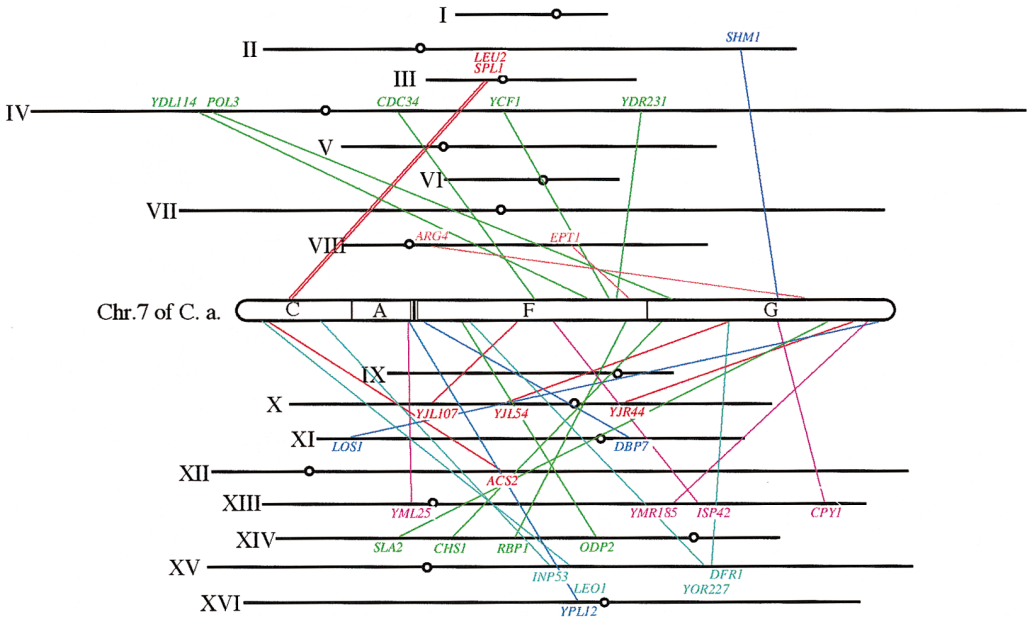

Figure 4.—View of distribution of the genes of chromosome *7* on the genome of *S. cerevisiae*. Chromosomes of *S. cerevisiae* are shown by horizontal lines. Circles indicate the centromeres. Names are those given *C. albicans* genes in Table 1.

were sequenced from each end for about 400–500 bp, and the amino acid sequence specified by any ORFs was compared with the available data bases using blastx (Gish and States 1993). The newly identified genes include such interesting homologues as *ACS2*, *ALD1*, *CDC34*, *DBP7*, *EPT1*, *INP53*, *ISP42*, *LOS1*, *LRT2* (see below), *NDH1*, *ODP2*, *SLA2*, *YCF1*, and *YDR231* (Table 1).

Comparison of genes and gene location with the *S. cerevisiae* map: The completion of the chromosome *7* physical map allows us to compare the distribution of genes located on this molecule in *C. albicans* with their location in the genome of *S. cerevisiae*. Figure 4 and Table 2 show that while the synteny of some pairs of the Saccharomyces genes is conserved, there is no overall indication that significant blocks of genetic material are conserved between these distantly related yeast species. Because *YCF1* and *YDR231* are assigned to the same fosmid, their physical distance is shorter than 40 kb on chromosome *7* of *C. albicans*. On the other hand, their physical distance is about 200 kb on chromosome *IV* of *S. cerevisiae*. *SHM1* and *CPY1*, another pair assigned to one fosmid, are located on different chromosomes in *S. cerevisiae*. *YJL54* is adjacent to *DFR1* on chromosome *7* of *C. albicans*. In *S. cerevisiae YJL54* is located on chromosome *X* and *DFR1* is located on chromosome *XVI*. The one close linkage that seems to be the same is that between *LEU2* and *SPL1*. Hence, these organisms seem likely to have diverged before their present chromosome structures evolved.

There are three identifiable genes found on chromosome *7* that do not have homologues in *S. cerevisiae*. Two of these, *ALD1* and *LRT2*, have highest homology to genes from mammals (human and rat, respectively). A third, *NDH1*, occurs in most fungi but not in baker's yeast.

Telomere-related sequences: Using the telomere-specific probe, Ca7 isolated by Sadhu *et al.* (1991), Chu *et al.* (1993) identified the chromosome *7 Sfi*I fragments (C and G), which contained the telomeres. Because we would not expect to find telomere sequences in our library, we walked to the end of these fragments in an attempt to find subtelomeric sequences that would identify the ends. The clone 1972 turned out to be homologous to sequences at both ends of chromosome *7* (Figures 1 and 3); 1972 contains the repeated se-

TABLE 2

Location of the genes of chromosome *7* on the *S. cerevisiae* genome

Chromosome (kb)	Gene (*Sfi*I fragment, fosmid)
I (230)	
II (813)	SHM1(G)
III (315)	LEU2(C), SPL1(C)
IV (1,520)	CDC34(F,15A5), YCF1(F,13E11), YDR231(F,13E11), YDL114(F,5C12), POL3(G)
V (175)	
VI (270)	
VII (1,091)	
VIII (563)	ARG4(G), EPT1(F)
IX (440)	YIL1(A)
X (745)	YJL54(G), YJR44(G, 18B6), YJL107(F)
XI (666)	DBP7(F), LOS1(G)
XII (1,078,rDNA)	ACS2(C)
XIII(924)	CPY1(G,7F4), ISP42(F,12E4), YMR185(G,11B11), YML25(A)
XIV (784)	ODP2(F, 5A8), RBP1(F,8B4), CHS1(C), SLA2(G)
XV (1,091)	DFR1(G), YOR227(F), INP53(C), LEO1(C)
XVI (948)	YPL12(A)

Genes which are located on the same fosmid are underlined.

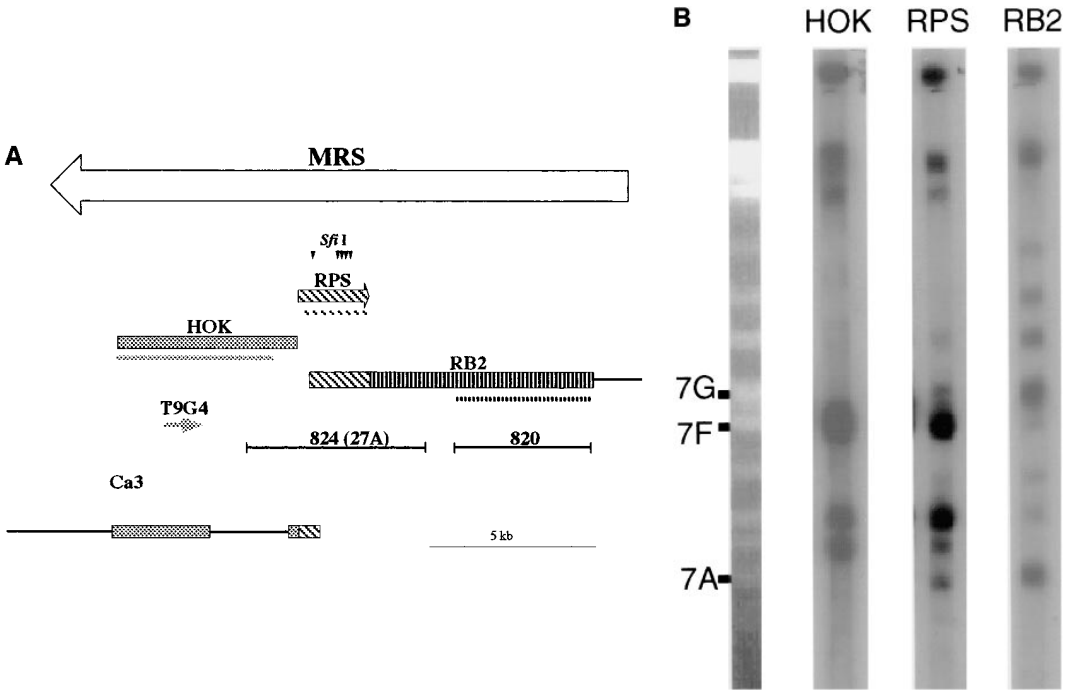

Figure 5.—Analysis of the orientation of the MRSs on chromosome *7*. (A) The elements of the MRS. The open arrow shows the defined direction of the MRS. The shaded bars identify the elements that have been sequenced and their relative positions. The DNA sequences of the lines are unknown. Probe 824 hybridizes with HOK, RPS, and RB2; probe 820 hybridizes that RB2 only. T9G4, one end of the insert of fosmid 9G4, is homologous to HOK. The RPS sequence, containing several *Sfi* I sites, is repeated tandemly; the number of repeats varies among the MRS sites. The probes that were used to identify the various parts of the MRS in B are shown as lines under the various elements. The Ca3 probe is shown for reference; it has been extensively used in molecular epidemiology studies. (B) Hybridization of probes from MRS elements to a pulsed-field separation of genomic *Sfi* I fragments. The fragments from chromosome *7* are indicated at the left; the other positive bands come from other chromosomes.

quence CARE2/Rel2 (Lasker *et al.* 1992) (Thrash-Bingham and Gorman 1993). Clone 1972, isolated from the pUC18-*Hin*dIII library mentioned above, hybridizes to many of the genomic telomeric *Sfi* I fragments and to all of the chromosomes. We therefore surmise that CARE2/Rel2 may be subtelomeric repeats. Clone 1972 is a 5.5-kb sequence that contains, in addition to CARE2, part of a gene, *LRT2*, which bears significant homology to the reverse transcriptases of LINE-type retrotransposons. Such transposons have been located at Drosophila telomeres (Levis *et al.* 1993). *LRT2* or a gene of that family is found near the left telomere of chromosome *7* but is not present on fragment G, which contains the righthand telomere. *LRT2* hybridizes to chromosomes *1*, *4*, and *7*.

The major repeat sequence: The repeat sequence that contains the RPS (Iwaguchi *et al.* 1992) (Chibana *et al.* 1994), HOK, and RB2 regions occurs in part on every *C. albicans* chromosome (Chindamporn *et al.* 1995, 1998). We call this the major repeat sequence (Figure 5A) to distinguish it from less frequent middle repeat sequences such as CARE2. RPS contains multiple *Sfi* I sites, and we estimate that of the 34 *Sfi* I sites (or groups of *Sfi* I sites) that define the large chromosomal fragments in the diploid genome of Candida, the MRS accounts for about 22. The MRS is asymmetric; we have arbitrarily defined it as running from left to right in the order RB2-RPS-HOK. Two MRS sequences were found on the contig map of chromosome *7*, one between *Sfi* I fragments A and F and one between F and G. We have determined the direction of these two on chromosome *7* by probing the *Sfi* I fragments with part of the MRS, HOK. Since this probe hybridizes only to F, the MRSs at the two ends of this fragment must point toward the middle. Fosmid 8B4 reached from the marker *RBP1* on fragment F to *ARS3* on fragment G (Figure 1). This fosmid contained the MRS sequences T9G4, HOK, RPS, 27A, and 820. Of the four *Sfi* I fragments from chromosome *7*, only F hybridizes to HOK (Figure 5B). (*Sfi* I fragment 6C, which runs in the same place as 7C, does hybridize with HOK: that accounts for the band in the C region in Figure 5B.) The HOK probe also hybridized to some fosmids that reacted with the probe *DBP7*, located on the left end of the F fragment, and to some fosmids positive for *YPL12*, located on the right end of the A fragment. These two single-copy genes, then, are near the repeated DNA carrying the HOK sequence. The RB2 probe hybridized to A and to G, and the RPS probe hybridized to A, F, and G (Figure 5B). These results confirm that the two MRS sequences on chromo-

some *7* both point toward the middle of the chromosome, "head to head." This information is important in assessing the role of this sequence in the chromosomal rearrangements, which are often observed in isolates of *C. albicans*. We have also shown that the *Sfi* I site between fragments A and C is not encoded by the MRS, since none of the fosmids covering this site hybridizes with any of the MRS probes.

DISCUSSION

There has been a great deal of interest in molecular genetic approaches to *C. albicans* in recent years. This effort is largely due to the fact that the organism has become an increasingly important human pathogen, and the factors related to its virulence are unknown. Its special biology is thought to be important in its pathogenicity: it can assume a variety of cellular morphologies, and the controls leading to changes in cell shape are not understood. In addition, the organism has no known sexual cycle. Exploration of the *C. albicans* genome and discovery of new genes will no doubt yield important information about its biology and ultimately its pathogenesis, but because no sexual cycle is known for *C. albicans*, there is virtually no genetic map. We therefore decided to create a physical map of the genome to provide a framework for the genetic studies ongoing in a large number of laboratories.

The approach we have chosen is to construct a sequence-tagged-site map, using as tags either random clones or genes characterized by us or others. We have localized these tags to three levels: the chromosomes, identified by CHEF electrophoresis; the restriction fragments generated by the restriction enzyme *Sfi* I, which cuts the diploid genome in 34 places; and a library of *C. albicans* DNA constructed in a fosmid vector. We have now completed the map of the smallest chromosome, *7*. This map has provided significant new information about the Candida genome.

A physical map of the genome of the same strain, 1161, is also being constructed using a restriction enzyme-fingerprinting technique and a cosmid library (Tait *et al.* 1997). A contig including six cosmids surrounding the *HYR1* gene on chromosome *1* has been generated. These efforts should be complementary to ours, allowing us to resolve problem areas that arise in one or the other project.

The strain from which the fosmid library was made was 1161. This strain differs from its wild-type progenitor CBS5736 by one karyotypic change: the two homologues of Chromosome *7* are resolvable by pulsed-field gel electrophoresis in CBS5736; in 1161 they are not. We chose this strain because the homologues of six of the remaining seven chromosomes were not resolvable, suggesting that its homologues did not differ, at least at the level of large blocks of genetic material. The present results suggest but do not prove that the original difference in the chromosome *7* homologues may have been in the number of repeats in the MRS.

General structure of chromosome *7*: Chromosome *7* was earlier shown to contain three *Sfi* I sites, digestion at which yields four fragments, C (190 kb), A (90 kb), F (360 kb), and G (390 kb) (Chu *et al* 1993). This size was corrected to fragments A (90 kb), C (180 kb or 170 kb), F (350 kb), and G (390 kb) by analysis in detail. The two chromosome *7* homologues vary slightly in size, with one containing 1030 kb and the other 1010 kb. There is no evidence of size heterogeneity in the *Sfi* I fragments A, F, and G; the difference is due to the *Sfi* I fragment C and variation in the number of repeats of the RPS sequence in one of the two regions containing the MRS (H. Chibana, unpublished results).

For the most part, the fosmid coverage of chromosome *7* was redundant, but one region was represented on only two fosmids of the 3840 in the library. Although it is possible that this region is difficult to clone in *Escherichia coli* because of the information it contains, there are at least two other possible explanations. One is the probability that some regions will be underrepresented in a finite library. Alternatively, the region may contain several *Sau*3A sites, so that it is underrepresented among the 40-kb fragments in a *Sau*3A partial digest. Such regions are bound to occur as the map grows.

The major repeat sequence: Several kinds of repeated sequences, including 27A and 820 (Scherer and Stevens 1988), Ca3 and Ca7 (Sadhu *et al.* 1991), the *Msp*I fragment (Cutler *et al.* 1988), CARE1 (Lasker *et al.* 1991), CARE2 (Lasker *et al.* 1992), RPS (Iwaguchi *et al.* 1992), Rel-1 and Rel-2 (Thrash-Bingham *et al.* 1993), and HOK and RB2 (Chindamporn *et al.* 1998) have been found in the genome of *C. albicans*. We are beginning to learn more about their functions and locations. Ca7 has been shown to contain telomeric sequences (Sadhu *et al.* 1991). CARE2 and Rel2 are proposed to be subtelomeric DNA in this study. The organization of RPS, HOK, RB2, and Ca3 has been determined by Chindamporn *et al.* (1997), and the relation among these repeated sequences and 27A and 820 is shown in Figure 5A. This large complex containing these repeated DNAs, the MRS, is the point at which the great majority of chromosomal translocations occur in *C. albicans* (B. B. Magee, unpublished data), and variations in the number of repeats of the internal sequence RPS may account for the size differences of resolvable homologues in many strains.

There are two MRSs on chromosome *7*, one between *Sfi*I fragments A and F and one between F and G. The MRS orientation, arbitrarily defined as RB2-RPS-HOK from beginning to end, goes from left to right at the A-F junction and from right to left at the F-G junction; that is, these regions form an inverted repeat, albeit separated by 350 kb. We have determined the orientation of the MRS regions on some other chromosomes

that have been found to be involved in reciprocal translocations with chromosome *7* in other strains, and, in each case, the orientation is such that pairing and a reciprocal recombination event would yield the translocations we have found (Chu *et al.* 1993).

The function of the MRS remains obscure, although the HOK region contains a pseudogene for isocitric dehydrogenase. Every chromosome in *C. albicans* contains at least some homology to this repeat (chromosome *3* has no RPS, but it does contain sequences homologous to RB2), so its role may be important. However, at this time no function has been assigned to the MRS.

Telomere structure: The telomeres of *C. albicans* and several species closely related to it have been shown to be dissimilar to those of most eukaryotes (McEachern and Blackburn 1994). Rather than a simple repeat of 5–8 bp, the telomeres of these fungi have a complicated sequence of 16–25 bp, which is repeated. The terminal 6 bp of this repeat constitute a motif common to this group of organisms. *S. cerevisiae* has a complicated and variable subtelomeric repeat complex that extends about 10 kb inward from the telomere (Louis *et al.* 1994). We have found that a middle-repeat element, CARE2/Rel2, is localized to the end of chromosome *7* and possibly to the ends of other chromosomes as well. Clone 11G4, a fosmid from chromosome *7* that contains the CARE2 repeat, contains a gene homologous to the reverse transcriptases of the LINE-type retrotransposons. This gene, which we have called LRT2, hybridizes to chromosomes *1* and *4* as well as chromosome *7*; hence similar or identical sequences exist on these chromosomes. *Drosophila melanogaster* contains such elements at its subtelomeres (Levis *et al.* 1993); the configuration of the telomere on the *Sfi*I fragment C of chromosome *7*, with three kinds of repeated sequences present, indicates that at least some Candida chromosomes may have a more complex telomere organization than anticipated.

Centromere location: Centromeres in *C. albicans* have resisted isolation. The lack of an adequate genetic or physical map has prevented cloning by walking from closely linked genes. Efforts to identify them by function in *S. cerevisiae* have been unsuccessful (W. Mertz, personal communication), and no one has identified a sequence that confers centromere-like properties on a Candida plasmid.

Although the MRS had been inferred to be a centromere because of its repeated structure (Chibana *et al.* 1994), its localization on virtually every chromosome (Chindamporn *et al.* 1995), and cytological evidence of its association with the spindle apparatus during mitosis (Chibana and Tanaka 1996), there is no evidence that the MRS is the centromere of chromosome *7*. In fact, the presence of two intact copies on each homologue of this chromosome makes it very unlikely that this sequence has a centromere function.

Analysis of a series of translocations in various *C. albicans* strains has suggested that the centromere is found on *Sfi*I fragment F (B. B. Magee, unpublished data), but we found no evidence for or against this location. There are no gaps in the fosmid contig for fragment F or indeed for chromosome *7* itself, suggesting that the centromere can be cloned in *E. coli*, at least in low copy number or in parts (if it is bigger than a single fosmid).

Genes homologous to known genes: The probes we have used to prepare the contig map of chromosome *7* include 34 genes, most of which were identified by partial sequencing followed by a blastx search of the public DNA databases. Of these, 31 were found to have their highest homology to genes of *S. cerevisiae*, and 3 were found to be closest to genes in other organisms. *NDH1* (a subunit of NAD dehydrogenase) and *LRT2* (reverse transcriptase), have no homologue in *S. cerevisiae*, and *ALD1* is more similar to a gene in humans than to the one in yeast. Both *ALD1* and *LRT2* are similar to genes in mammals (humans and rats, respectively). Because Candida is a commensal of these two species, these genes might point to horizontal transmission from the host. The mitochondria of *S. cerevisiae* differ from those of most fungi; the presence of the *NDH1* gene in Candida suggests that its mitochondria resemble those of the majority of fungi, rather than those of Saccharomyces.

Comparison of *S. cerevisiae* and *C. albicans* synteny and linkage: Although there are several cases where groups of genes located on chromosome *7* in Candida are also syntenic in Saccharomyces, the groups are widely dispersed in the latter organism. The only case in which we know that close linkage occurs in both species is between *LEU2* and *SPL1*, which are adjacent in *S. cerevisiae* and *C. albicans*.

Four of the 16 Saccharomyces chromosomes contain no homologues of the genes of chromosome *7* of *C. albicans*. Although no gene from chromosome *7* in Candida was found to have a homologue on chromosome *VII*, one of the largest *S. cerevisiae* chromosomes, it is important to remember that this study includes only 34 of the estimated 6000 to 8000 genes in Candida; hence, it may be surprising that 12 of the 16 chromosomes are represented in Table 2 and Figure 4.

It seems clear that the genomes of Candida and Saccharomyces are not just rearrangements of the same information; *C. albicans* contains several genes that are not found in Saccharomyces, and the converse will very likely turn out to be true when all the genes of Candida are identified. It is certainly true that the MRS, which represents a major structural part of the Candida genome, whatever its function turns out to be, has no analogous component in Saccharomyces. These differences are not surprising, given the evolutionary distance between the two organisms (Kurtzman and Robnett 1997). The overall picture suggests that there has been a great deal of modification of the genome, with genes

lost and gained on both sides since the two yeasts diverged evolutionarily.

The map presented here represents the first detailed analysis of a chromosome of the important pathogen *C. albicans*. The map should allow a variety of experiments to improve our knowledge of the role of chromosome structure and of many genes involved in the biology and pathogenesis of this fungus.

We thank Michael Strathmann for the fosmid library and Marin Bell and Jeff Staudinger for preparing several of the probes used. This work was supported by Grant AI16567 from the National Institute of Allergy and Infectious Diseases.

LITERATURE CITED

Au-Young, J., and P. W. Robbins, 1990 Isolation of a chitin synthase gene (CHS1) from *Candida albicans* by expression in *Saccharomyces cerevisiae*. Mol. Microbiol. **4:** 197–207.

Braun, B. R., and A. D. Johnson, 1997 Control of filament formation in *Candida albicans* by the transcriptional repressor *TUP1* (see comments). Science **277:** 105–109.

Chibana, H., S. Iwaguchi, M. Homma, A. Chindamporn, Y. Nakagawa *et al.*, 1994 Diversity of tandemly repetitive sequences due to short periodic repetitions in the chromosomes of *Candida albicans*. J. Bacteriol. **176:** 3851–3858.

Chibana, H., and K. Tanaka, 1996 Analysis of the cell cycle in the budding yeast Candida *albicans* by positioning of chromosomes by fluorescence in situ hybridization (FISH) with repetitive sequences. Genes Cells **1:** 727–740.

Chindamporn, A., Y. Nakagawa, M. Homma, H. Chibana, M. Doi *et al.*, 1995 Analysis of the chromosomal localization of the repetitive sequences (RPSs) in *Candida albicans*. Microbiology **141:** 469–476.

Chindamporn, A., Y. Nakagawa, I. Mizuguchi, H. Chibana and K. Tanaka, 1998 The repetitive sequences, RPSs, in the chromosomes of *Candida albicans*, are sandwiched by two novel stretches, HOK and RB2, common to each chromosome. Microbiology **144:** 849–857.

Chu, W. S., B. B. Magee and P. T. Magee, 1993 Construction of an SfiI macrorestriction map of the *Candida albicans* genome. J. Bacteriol. **175:** 6637–6651.

Cutler, J., 1991 Putative virulence factors from *Candida albicans*. Annu. Rev. Microbiol. **45:** 187–218.

Cutler, J., P. Glee and H. Horn, 1988 *Candida albicans*- and *Candida stellatoidea*-specific DNA fragment. J. Clin. Microbiol. **26:** 1720–1724.

Daly, S., G. Mastromei, A. Yacoub and R. Lorenzetti, 1994 Sequence of a dihydrofolate reductase-encoding gene from *Candida albicans*. Gene **147:** 115–118.

Ferrara, A., R. Cafferke and G. P. Vivi, 1992 Cloning and sequence analysis of a rapamycin-binding protein-encoding gene (RBP1) from *Candida albicans*. Gene **113:** 125–127.

Feinberg, A. P., and B. Vogelstein, 1983 A technique for radiolabeling DNA restriction endonuclease to high specific activity. Anal. Biochem. **132:** 6–13.

Game, J. C., M. Bell, J. S. King and R. K. Mortimer, 1990 Random-breakage mapping, a rapid method for physically locating an internal sequence with respect to the ends of a DNA molecule. Nucleic Acids Res. **18:** 4453–4461.

Gish, W., and D. J. States, 1993 Identification of protein coding regions by database similarity search. Nat. Genet. **3:** 266–272.

Goffeau, A., B. J. Barrell, H. Bussey, R. W. Davis, B. Dujon *et al.*, 1996 Life with 6000 genes. Science **274:** 546–567.

Goshorn, A. K., and S. Scherer, 1989 Genetic analysis of prototrophic natural variants of *Candida albicans*. Genetics **123:** 667–673.

Heery, D. M., F. Gannon and R. Powell, 1990 A simple method for subcloning DNA fragments from gel slices. Trends Genet. **6:** 173.

Herreros, E., M. I. Garcia-Saez, C. Nombela and M. Sanchez, 1992 A reorganized *Candida albicans* DNA sequence promoting homologous non-integrative genetic transformation. Mol. Microbiol. **6:** 3567–3574.

Hoyer, L. L., B. B. Magee, E. H. Rikkerink and S. Scherer, 1994 The ARG4 gene of *Candida albicans*. Gene **142:** 213–218.

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.

Iwaguchi, S., M. Homma, H. Chibana and K. Tanaka, 1992 Isolation and characterization of a repeated sequence (RPS1) of *Candida albicans*. J. Gen. Microbiol. **138:** 1893–1900.

Kim, U. J., H. Shizuya, P. M. de Jong, B. Birren and M. I. Simon, 1992 Stable propagation of cosmid sized human DNA inserts in an F factor based vector. Nucleic Acids Res. **20:** 1083–1085.

Kohler, J. R., and G. R. Fink, 1996 *Candida albicans* strains heterozygous and homozygous for mutations in mitogen-activated protein kinase signaling components have defects in hyphal development. Proc. Natl. Acad. Sci. USA **93:** 13223–13228.

Kurtzman, C. P., and C. J. Robnet, 1997 Identification of clinically important ascomycetous yeasts based on nucleotide divergence in the 5′ end of the large-subunit (26S) ribosomal DNA gene. J. Clin. Microbiol. **35:** 1216–1223.

Lasker, B. A., L. S. Page, T. J. Lott, G. S. Kobayashi and G. Medoff, 1991 Characterization of CARE-1: *Candida albicans* repetitive element-1. Gene **102:** 45–50.

Lasker, B. A., L. S. Page, T. J. Lott and G. S. Kobayashi, 1992 Isolation, characterization, and sequencing of *Candida albicans* repetitive sequence element 2. Gene **116:** 51–57.

Levis, R. W., R. Ganesan, K. Houtchens, L. A. Tolar and F. Sheen, 1993 Transposons in place of telomeric repeats at a Drosophila telomere. Cell **75:** 1083–1093.

Leberer, E., K. Ziegelbauer, A. Schmidt, D. Harcus, D. Dignard *et al.*, 1997 Virulence and hyphal formation of *Candida albicans* require the Ste20p-like protein kinase CaCla4p. Curr. Biol. **7:** 539–546.

Louis, E. J., E. S. Naumova, A. Lee, G. Naumov and J. E. Haber, 1994 The chromosome end in yeast: its mosaic nature and influence on recombinational dynamics. Genetics **136:** 789–802.

Magee, B. B., and P. T. Magee, 1997 WO-2, a stable aneuploid derivative of WO-1, can switch from white to opaque and form hyphae. Microbiology **143:** 289–295.

McEachern, M. J., and E. H. Blackburn, 1994 A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts. Proc. Natl. Acad. Sci. USA **91:** 3453–3457.

Mukhtar, M., D. A. Logan and N. F. Kaufer, 1992 The carboxypeptidase Y-encoding gene from *Candida albicans* and its transcription during yeast-to-hyphae conversion. Gene **121:** 173–177.

Nolan, T., and J. Rosamond, 1996 Isolation and molecular characterisation of the POL3 gene from *Candida albicans*. Gene **183:** 159–165.

Rustchenko, E. P., D. H. Howard and F. Sherman, 1997 Variation in assimilating functions occurs in spontaneous *Candida albicans* mutants having chromosomal alterations. Microbiology **143:** 1765–1778.

Sadhu, C., M. J. McEachern, E. P. Rustchenko-Bulgac, J. Schmid, D. R. Soll *et al.*, 1991 Telomeric and dispersed repeat sequences in Candida yeasts and their use in strain identification. J. Bacteriol. **173:** 842–850.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Scherer, S., and P. T. Magee, 1990 The genetics of *Candida albicans*. Microbiol. Rev. **54:** 226–241.

Scherer, S., and D. V. Stevens, 1988 A *Candida albicans* dispersed, repeated gene family and its epidemiologic applications. Proc. Natl. Acad. Sci. USA **85:** 1452–1456.

Soll, D. R., J. Anderson and M. Bergen, 1991 The developmental biology of the white-opaque transition in *Candida albicans*, pp. 20–45 in *Molecular Biology of the Candida albicans*, edited by R. Prasad. Springer, Berlin.

Stoldt, V. R., A. Sonneborn, C. E. Leuker and J. F. Ernst, 1997 Efg1p, an essential regulator of morphogenesis of the human pathogen *Candida albicans*, is a member of a conserved class of bHLH proteins regulating morphogenetic processes in fungi. EMBO J. **16:** 1982–1991.

Suzuki, T., I. Kobayashi, T. Kanbe and K. Tanaka, 1989 High frequency variation of colony morphology and chromosome reorganization in the pathogenic yeast *Candida albicans*. J. Gen. Microbiol. **135:** 425–434.

Tait, E., M. C. Simon, S. King, A. J. Brown, N. A. R. Gow *et al.*, 1997 A *Candida albicans* genome project: cosmid contigs, physical mapping, and gene isolation. Fungal Genet. Biol. **21:** 308–314.

Thrash-Bingham, C., and J. A. Gorman, 1993 Identification, characterization and sequence of *Candida albicans* repetitive DNAs Rel-1 and Rel-2. Curr. Genet. **23:** 455–462.

Wickes, B. L., J. E. Golin and K. J. Kwon-Chung, 1991 Chromosomal rearrangement in *Candida Stellatoidea* results in a positive effect on phenotype. Infect. Immun. **59:** 1762–1772.

Communicating editor: A. P. Mitchell