JOURNAL OF BACTERIOLOGY, Jan. 1991, p. 842–850
0021-9193/91/020842-09$02.00/0

Vol. 173, No. 2

Telomeric and Dispersed Repeat Sequences in *Candida* Yeasts and Their Use in Strain Identification

CHANCHAL SADHU,[1]† MICHAEL J. McEACHERN,[1] ELENA P. RUSTCHENKO-BULGAC,[1]‡ JAN SCHMID,[2] DAVID R. SOLL,[2] AND JAMES B. HICKS[1]†*

Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, California 92037,[1] and Department of Biology, University of Iowa, Iowa City, Iowa 52242[2]

Received 20 June 1990/Accepted 19 October 1990

Several different repetitive DNA sequences have been isolated from the pathogenic yeast *Candida albicans*. These include two families of large dispersed repeat sequences (Ca3, Ca24) and a short (23-bp) tandemly repeated element (Ca7) associated with *C. albicans* telomeres. In addition, a large subtelomeric repeat (WOL17) has been cloned. DNA fragments containing the telomeric repeats are highly variable among different *C. albicans* strains. We have shown that the Ca3 repeat is relatively more stable and is suitable for use as a species-specific and strain-specific probe for *C. albicans*.

Candida albicans and related species are responsible for the majority of oral and vaginal yeast infections (1a, 26) and continue to increase in frequency as a major cause of systemic infections in immunocompromised hosts (1, 4, 5, 25, 30). Because *C. albicans* and related species lack sexual cycles (46), genetic studies have been difficult. Analyses of mitotic recombination of auxotrophic markers (47, 48), DNA content (32, 42), and complexity (32) and chromosome organization by orthogonal field alternating gel electrophoresis (21) indicate that *C. albicans* possesses a diploid genome organized into about eight chromosome pairs (16, 20), with a DNA content and complexity similar to those of bakers' yeast, *Saccharomyces cerevisiae* (17).

The lack of a sexual phase in *C. albicans* and related species presumably prevents opportunities for recombination between strains, leading to a direct clonal relationship in the descendants. This characteristic may be useful in clinical analyses of colonization and epidemiological studies, since it allows direct determination of strain relatedness by employing genomic markers. Distinguishing between species and strains is of paramount importance in such studies, since it has been demonstrated that more than one *Candida* species and more than one strain of a single species can colonize or infect the same individual (9, 23, 27, 28, 44). In addition, multiple "switch phenotypes" (growth forms with distinctive cell types and colony morphologies [40, 41]) of the same strain can be found at the same site of infection (43). Sugar assimilation and biotyping methods are normally accurate in distinguishing species and, in many cases, strains (15, 29, 31) but can provide false distinctions between switch phenotypes of the same strain (1). Serological typing methods can discriminate between only two *C. albicans* subtypes (8). Even electrophoretic karyotyping can be misleading, since subclones of single strains can exhibit significant changes in chromosome size or structure over relatively short periods of culture (36, 45). Clearly, a combination of methods is necessary to assess strain relatedness and distinguish between closely related species.

* Corresponding author.
† Present address: ICOS Corporation, 22021 20th St., Bothell, WA 98021.
‡ Present address: University of Rochester Medical School, Rochester, NY 14642.

Repetitive DNA sequences provide useful markers for strain identification and also provide keys to understanding chromosome structure and organization. Scherer and Stevens (38) reported the isolation of a dispersed middle repetitive DNA sequence from *C. albicans* and showed that it could be used for fingerprinting strains of that species. We have also generated species-specific repetitive DNA probes that have been used in analysis of clinical isolates from patients suffering from acute vaginal candidiasis (43) and systemic candidiasis (44) as well as in species identification. We present here an initial characterization of several of these repetitive DNA elements. These elements (with the exception of a ribosomal DNA [rDNA] probe, Ca5) are dispersed throughout the *C. albicans* genome. The element Ca3 appears to have undergone internal rearrangement in a strain-specific manner and is suitable as a strain-specific probe. The elements Ca7 and WOL17 are highly variable and are located near the chromosome ends.

MATERIALS AND METHODS

Strains and growth conditions. The sources of yeast strains used in this work are indicated in Table 1. Strains were grown in YEPD medium (39) (either in liquid or on agar plates) at 30°C, and DNA was isolated by standard procedures used with *S. cerevisiae* as described by Sherman et al. (39), except that in the yeast DNA minipreparation procedure three phenol extractions were required to make the DNA clean enough to cut with restriction enzymes.

Isolation of healthy human mouth strains. Healthy mouth strains were isolated by scrubbing the inner cheeks of healthy individuals with a sterile Culturette swab (American Scientific Products, McGaw Park, Ill.), rigorously swirling the swab in 0.5 ml of sterile water, and plating on nutrient agar. Single colonies were cloned and initially typed as *C. albicans* by examining the pattern of sugar assimilation with a commercial kit obtained from Analytab Products (Plainview, N.Y.).

Library construction and cloning. A library of *C. albicans* 3153a DNA was constructed by inserting a partial *Eco*RI digest of DNA into the bacteriophage vector λgt10 (11). Phage arms and packaging extracts were obtained from Stratagene Cloning Systems (La Jolla, Calif.). After in vitro

TABLE 1. Yeast strains used in this study

Strain	Species[a]	Source or reference
3153a	*C. albicans*	40
WO-1	*C. albicans*	41
32032	*C. albicans*	ATCC 32032
32033	*C. albicans*	ATCC 32033
32077	*C. stellatoidea*	ATCC 32077
18814	*C. claussenii*	ATCC 18814
B4201	*C. albicans*	14
B4365	*C. stellatoidea*	ATCC 20408
B4252	*C. stellatoidea*	ATCC 11006
HMH2, 4, 5, 6, 8, and 9[b]	*C. albicans*	This study
CTAT	*C. tropicalis*	ATCC 34139
CT1	*C. tropicalis*	D. R. Soll (41a)
28707	*C. tropicalis*	ATCC 28707
CK	*C. krusei*	ATCC 34135
CG	*C. glabrata*	ATCC 34138
RLK	*S. cerevisiae*	13

[a] *C. stellatoidea* and *C. claussenii* are closely related or identical to *C. albicans* and are considered as being *C. albicans* for the purpose of this study.
[b] Healthy human mouth strains.

packaging, the library was amplified by plating 10^5 plaques and harvesting with 10 ml of λ storage buffer (22).

The method used to isolate repetitive DNA-containing clones was based on differences in hybridization signal among members of the clone library that had been transferred to nitrocellulose filters (7). Unfractionated 3153a DNA was labeled with ^{32}P by nick translation (22) or with a random priming kit (Amersham Corp., Arlington Heights, Ill.) and used as probe. The sizes of the DNA strands used for probing were 200 to 400 bases. In accordance to previous studies of DNA complexity in *C. albicans* (32), approximately 10% of the plaques in the library showed strong hybridization, indicating that the cloned DNA segment was present in multiple copies in the whole-cell DNA probe. The strongly hybridizing plaques were screened by using labeled rDNA from *S. cerevisiae* and cloned mitochondrial DNA of *C. albicans* obtained from W. S. Riggsby (49). As expected, ribosomal sequences make up the majority of the strongly hybridizing clones. Likewise, fragments containing tRNA genes were identified by using a probe made from low-molecular-weight (4S) RNA.

In addition to the *C. albicans* 3153a library, we constructed two libraries of the DNA from strain WO-1. One of the libraries was in the plasmid vector, pBluescript (Stratagene Cloning Systems, La Jolla, Calif.) and the other is in the bacteriophage vector EMBL-3. For the plasmid library, WO-1 DNA was digested to completion with *Sau*3AI, end filled with G and A, and ligated to pBluescript DNA that was cut with *Sal*I and end filled with T and C. After transformation to *Escherichia coli* DH5α, DNA minipreparations were made from 75 individual colonies and separately hybridized to *C. albicans* genomic DNA blots. Clone 17 from the white/opaque library (WOL17) was chosen for further characterization. For the bacteriophage library, WO-1 DNA was partially digested with *Sau*3AI and ligated with the *Bam*HI-cut arms of the vector EMBL-3. This library was used to isolate larger segments of WO-1 DNA containing the WOL17 sequences.

Hybridizations for Southern blots and library screening were performed at 65°C in 4× SSC (1× SSC is 0.15 M sodium chloride plus 0.015 M sodium citrate [pH 6.8]) containing 10 μg of denatured salmon sperm DNA per ml and 0.2% sodium lauryl sarcosinate. After hybridization, the blots were washed at 65°C in 0.5× SSC containing 0.2% sodium lauryl sarcosinate.

Bacterial and phage genetic manipulations were performed as outlined by Maniatis et al. (22). Restriction enzymes and other supplies for manipulating nucleic acids were obtained from standard commercial sources.

Ca7 subcloning and sequencing. The Ca7 repeats were subcloned in both orientations into the plasmid Bluescript (Stratagene) as a 1,057-bp *Sau*3A fragment to create plasmids pMM100 and pMM101. These plasmids and deletion derivatives made from them were used to sequence the Ca7 repeats. Sequencing was carried out by the dideoxy method (37) with reagents from the Sequenase Kit manufactured by U.S. Biochemical Corp. (Cleveland, Ohio).

Pulsed field gel electrophoresis. Transverse alternating-field electrophotoresis was carried out on a Beckman (Palo Alto, Calif.) Geneline apparatus with the Tris-acetate buffer recommended by Beckman on 0.6% agarose (Bio-Rad chromosomal grade). Electrophoresis was in six stages as follows: 100 V for 6 h with a 1-min switch time, 100 V for 12 h with a 2-min switch time, 100 V for 16 h with a 4-min switch time, 100 V for 20 h with a 7-min switch time, 80 V for 18 h with a 10-min switch time, and 80 V for 18 h with a 13-min switch time.

Nomenclature. To avoid unnecessary complexity in referring to the novel genetic elements described here, we have named the elements after the accession numbers of the clones from which they were identified. That is, the names λCa10, λCa3, etc., refer to our laboratory catalog numbers for the recombinant phage or plasmid isolated in the screen for repeats. For the elements that we have characterized, we propose the corresponding names Ca3, Ca7, and Ca24. We have also used Ca3 and Ca7 to designate hybridization probes derived from the original clones.

RESULTS

Hybridization of the *C. albicans* repeat elements to genomic DNA. After secondary screening of the *C. albicans* λgt10 library, DNA minipreparations were made from each clone of interest; these DNAs were labeled and used to probe blots of whole-cell DNA and digested with restriction enzymes to assess copy number and pattern of hybridization.

Each set of tracks in Fig. 1 consists of *Eco*RI restriction digests of DNA from two *C. albicans* strains probed with labeled DNA from hybrid bacteriophage clones identified as either single-copy or repeated sequences by the procedure described above. Clones designated λCa12 and λCa10 (Fig. 1A and B) are recombinant phage that were originally picked from among those giving weak hybridization signals in the initial screen. Because the library was constructed from a partial *Eco*RI digest of genomic DNA, each clone contains multiple *Eco*RI fragments, five in λCa12 and three in λCa10. The positions where the actual *Candida* restriction fragments from the phage clones would migrate are indicated next to each panel. As expected, for unique single-copy DNA, each restriction fragment present in λCa12 and λCa10 hybridizes to a single corresponding band in the genomic digest. A restriction polymorphism can be seen in Fig. 1B; in strain 3153a the band corresponding to the second-largest band in the clone is relatively weak, and a new band (also weak) appears at a higher molecular weight. This is the pattern expected if the polymorphism occurred in one of the two alleles on homologous chromosomes in this diploid organism.

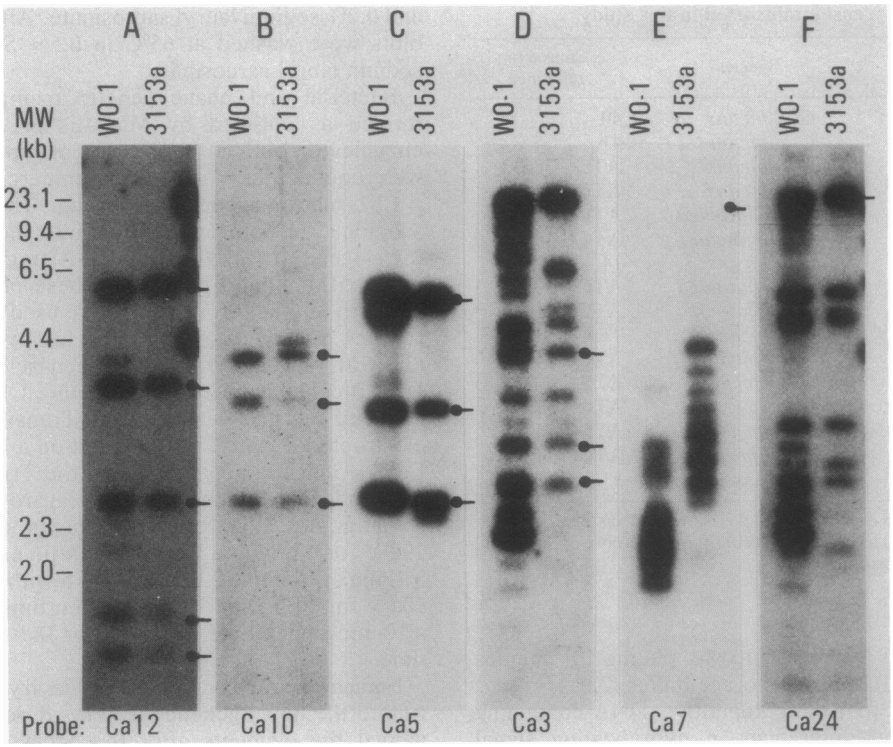

FIG. 1. Hybridization patterns of different *C. albicans*-derived probes to *Eco*RI-digested DNA from *C. albicans* WO-1 and 3153a. The positions of the *Eco*RI fragments of the respective clones are indicated.

The polymorphism shown by λCa10 is a notable exception to the overall similarity of the restriction patterns of the two strains. The conserved pattern of DNA hybridization in this and other digests probed with several other single-copy clones (data not shown) indicates that the two strains are closely related and that the majority of the unique DNA component has not diverged substantially during their separate microevolutionary histories. The repeated DNA component has undergone more extensive change, as shown below.

The tracks in Fig. 1C, D, E, and F have been probed with clones picked as repeated sequences by virtue of their strong initial hybridization signal. Probe λCa5 (Fig. 1C) contains three fragments of the tandemly repeated rRNA cistrons. The pattern is consistent with a tandemly repeated array of a sequence approximately 12.1 kbp in length and is in substantial agreement with results of Magee et al. (19) for a different set of *C. albicans* strains. λCa5 hybridizes to one or two of the largest *C. albicans* chromosomal bands seen in tranverse alternating-field electrophoresis (Fig. 2C). This is consistent with the rDNA being present on one pair of homologous chromosomes. *C. albicans* typically contains eight pairs of chromosomes, the largest two of which have only recently been separated from one another (16). Our recent observations (M. McEachern, data not shown) have shown that the rDNA chromosomes of *C. albicans* fluctuate in size at a very rapid rate from one colony to another in the same strain. This presumably reflects the effects of unequal crossing over between rDNA repeats. Also noteworthy in Fig. 2C is the fact that *C. tropicalis* produces at least eight chromosome bands, of which two of the largest hybridize to the λCa5 probe. A polymorphism, originally noted by Rikkerink et al. (33), appears in the top two bands in the digest of strain B9 (Fig. 1C). A weaker band appears below the 2.6-kbp band in strain 3153a. This probably represents a polymorphism that is present in only a fraction of the copies.

In contrast to the tandem-repeat pattern of Ca5, probes Ca3, Ca7, and Ca24 yielded hybridization patterns (Fig. 1D, E, and F, respectively) characteristic of dispersed repeat sequences.

Clones λCa3 and λCa24 (Fig. 1D and F) were two of eight independent recombinant phage that showed similar multiband hybridization patterns. Four of these phages, including λCa3, contained the three *Eco*RI restriction fragments (indicated next to the tracks in Fig. 1D) of 4.3, 3.3, and 3.1 kbp plus two additional fragments of 0.8 and 0.3 kbp (run off of the gel shown in Fig. 1D). Corresponding fragments were found in the genomes of both 3153a and WO-1. The other four clones, including λCa24, contained a single *Eco*RI fragment of approximately 12.4 kbp and hybridized with a subset of the genomic bands shown in Fig. 1D, including three bands at 2.3 to 3.0 kbp in strain WO-1 as well as the expected band of 12.5 kpb (Fig. 1F).

The Southern blots of Fig. 1 and other blots (data not shown) clearly show that λCa24 is a member of a family of repetitive elements that is distinguishable from the Ca3 family of repeats. Curiously, however, the internal 3.0-kb *Eco*RI fragment of Ca3 hybridizes to Ca24, indicating that the two repeat families share some sequences in common. Additional restriction and hybridization results (data not shown) indicate that Ca24 is of a size comparable to that of Ca3 but must be made up of mostly non-Ca3 sequences.

Transverse alternating-field electrophoretic analysis (Fig. 2B) confirms that the Ca3 element is dispersed among most of the chromosomes of *C. albicans*. Only the homologs of the fifth smallest pair of chromosomes (chromosome 3 in the nomenclature of Magee et al. [20]) appear to lack any copy of the Ca3 element, something that appears to be the case with

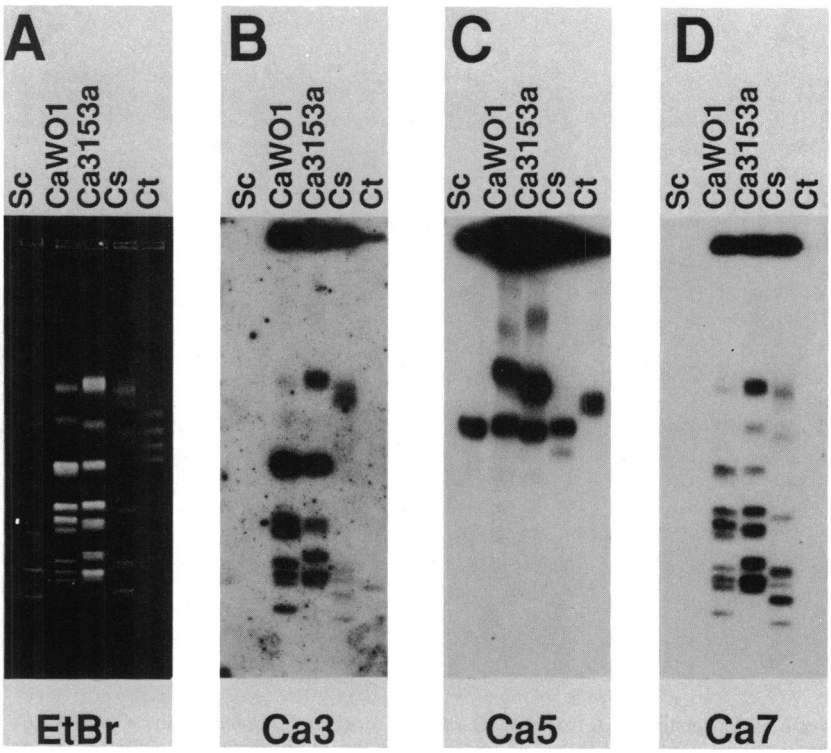

FIG. 2. (A) Ethidium bromide (EtBr)-stained pattern of *S. cerevisiae* RLK (Sc), *C. albicans* WO-1 (CaWO-1), *C. albicans* 3153a (Ca3153a), *C. stellatoidea* (Cs), and *C. tropicalis* (Ct) chromosomes separated by transverse alternating-field electrophoresis. (B, C, and D) Hybridization patterns of the chromosome blot probed with Ca3, Ca5, and the Ca7 repeats, respectively.

C. albicans strains that we have examined. Curiously, Ca3 hybridizes to one of the chromosomal bands of *Candida tropicalis*. From the restriction patterns and chromosome hybridization patterns, λCa3 appears to represent the same family of repeated sequences that was reported by Scherer and Stevens (38).

Ca7 repeat. The hybridization pattern of λCa7 shown in Fig. 1E is striking in that it is made up of many bands, the sizes of which are clustered in a tight distribution around a modal size characteristic of each of the two strains. Furthermore, in contrast to the strong similarity between strains WO-1 and 3153a shown in Fig. 1 with the other four probes, the λCa7 patterns for the two strains seem unrelated. This observation was extended by comparing the hybridization pattern with those of seven other *C. albicans* strains. Each tested strain shows a pattern quite distinct from any other (Fig. 3A). These findings indicate that λCa7 represents a rapidly changing segment of DNA.

Hybridization of a subcloned fragment of λCa7 to *Candida* chromosomes (Fig. 2D) demonstrates that this sequence is present on every chromosome of every *C. albicans* strain examined but is absent from *S. cerevisiae*, *Candida krusei*, and *Candida glabrata* (Fig. 3A). Furthermore, the hybridization intensities are approximately equal among the different chromosomes. These results indicate that the Ca7 repetitive element is not only dispersed in the genome but also uniformly dispersed among all of the chromosomes.

Examination of the λCa7 clone demonstrated that it did not have the expected insert of *Candida Eco*RI fragments but rather that it contained a large duplication of λ sequences fused to *Candida* sequences without the benefit of an *Eco*RI site, apparently as the result of an aberrant ligation event. The stretch of *C. albicans* DNA immediately adjacent to the duplicated λ sequences proved to exhibit nearly the complete hybridization pattern produced by the intact Ca7 phage. Sequencing this region from a 1.1-kb *Sau*3AI fragment revealed the presence of 16 tandem copies of a 23-bp repeat, the sequence of which is shown in Fig. 4C. The repeat shares a 4-bp homology with the λ sequences at the point of fusion. Such a rare event would explain why the Ca7 repeat sequence, which occurs as many as 30 times in the genome, occurred only once in our survey of cloned repetitive DNA. On the basis of arguments described below, we believe that this genomic fragment represents the last *Eco*RI restriction fragment on the end of a *Candida* chromosome and that the repeated element, which we have designated Ca7, represent sequences characteristic of the *C. albicans* telomere.

Ca7 is located at the ends of chromosomes. The highly variable and often smeared banding pattern of Ca7 (Fig. 1E and 3A) bears a striking resemblance to the banding pattern observed for telomeres in *S. cerevisiae* when probed with poly(G-T) probes (18). This feature, combined with the fact that the Ca7 repeat resides on every *Candida* chromosome and along with the observation that *S. cerevisiae* telomeres bear a specific set of repeated sequences, led us to the notion that Ca7 may reside near the ends of *Candida* chromosomes and may be part of the telomere structure. We therefore applied the test for terminal location devised by DeLange and Borst (6) to verify the terminal location of certain expression-linked variable surface antigen genes in *Trypanosoma brucei*.

In the DeLange and Borst method, high-molecular-weight chromosomal DNA is subjected to limited digestion with the exonuclease *Bal*31. DNA sequences located at or near natural chromosome ends should be more sensitive to diges-

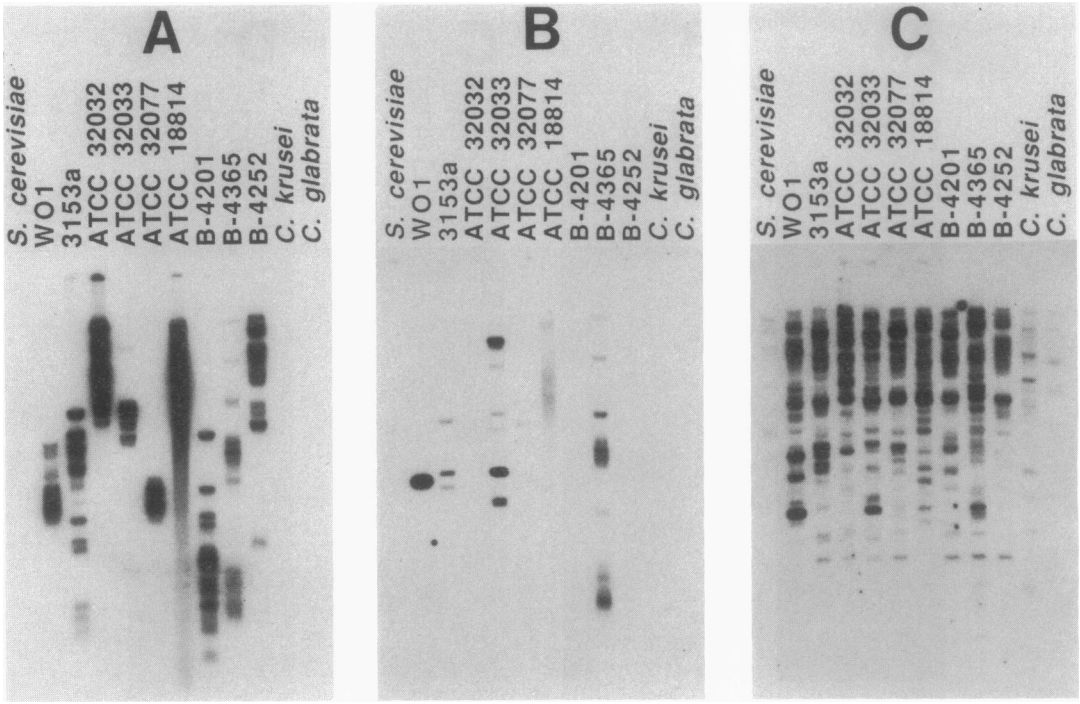

FIG. 3. Autoradiograms of *Eco*RI-digested DNA from nine *C. albicans* or *C. stellatoidea* strains, *C. krusei*, *C. glabrata*, and *S. cerevisiae* probed with a plasmid carrying a *Sau*3A fragment from Ca7 that contains the 23-bp repeats (A), WOL17 (B), or λWOL17.1 (C).

tion than sequences located elsewhere on the chromosomes. In our experiment, we compared the *Bal* 31 sensitivity of Ca7 sequences with those of bands from Ca3 and a single-copy gene, *CAG1* (Fig. 4A and B). Each gel track of Fig. 4 contains DNA from *C. albicans* WO-1 digested with *Bal* 31 for the designated time and then completely digested with *Eco*RI restriction enzyme. The blot was first probed with Ca7 and *CAG1* and autoradiographed (Fig. 4B). The same blot was then boiled in water to remove the ^{32}P-labeled probes and probed with ^{32}P-labeled λCa3 (Fig. 4A). The hybridization patterns of Ca3 and *CAG1* are not altered even after 120 min of *Bal* 31 digestion. In contrast, the fragments probed by Ca7 in Fig. 4B exhibit sensitivity to *Bal* 31. The entire population of fragments shows a decrease in size by 45 min of digestion. By measuring relative to marker bands on the ethidium bromide-stained gel (data not shown), we estimate that by 100 min, approximately 600 bp has been digested from each of the prominent bands. Considering the size of the Ca7 repeat we have cloned (approximately 400 bp), it seems that the Ca7 sequence begins no more than 200 bp from the ends of the chromosomes.

Isolation of a second telomere-associated sequence. During the course of screening a variety of random *Sau*3AI fragments cloned from strain WO-1, we identified another DNA fragment associated with *C. albicans* telomeres. This was originally detected by hybridizing a random *C. albicans* DNA clone, designated WOL17, with *Eco*RI-digested genomic DNAs from nine *C. albicans* strains (Fig. 3B). It is evident from Fig. 3B that both the hybridization pattern and the number of copies of the WOL17 sequence vary enormously from strain to strain; three of the nine tested strains lack the element entirely. Although four of the remaining *C. albicans* strains show hybridization to one or more discreet bands, two strains, ATCC 18814 and B-4365, show hybridization to smears in the same places as those that appear when the Ca7 repeats are used as a probe of the same digests (Fig. 3A). This suggested that the WOL17 sequence was telomere linked in at least these two strains. The strain from which the WOL17 sequence was isolated, WO-1, displays a single intense *Eco*RI band that hybridizes to the WOL17 sequence. Like the Ca7 sequence, the WOL17 sequence hybridized to each of the chromosomes of WO-1.

To determine whether the WOL17 sequence was telomere associated in WO-1, we performed a Southern blot with WO-1 DNA digested with a number of 6-base-recognizing restriction enzymes. The resulting filter was first probed with WOL17 and, after exposure to film, was boiled and reprobed with Ca7 (data not shown). The observed patterns for the two probes typically proved to be identical, indicating that WOL17 is in fact telomere linked. Two of the 12 tested enzymes, *Eco*RI (Fig. 3A and B) and *Pvu*II, did not produce the same pattern with the two probes. However, these 2 enzymes, unlike the 10 other enzymes tested, produce relatively short fragments (<3 kb) that hybridize to Ca7 and presumably cut in between the WOL17 sequence and the Ca7 repeats.

DNA sequence analysis of the 144-bp *Sau*3AI WOL17 fragment has shown that it does not contain any short repeats like those seen in Ca7, nor does it contain any open reading frames. In a *Bal* 31 digestion experiment, we demonstrated that the WOL17 sequence was not as sensitive to digestion as the Ca7 repeats, indicating that, at least in strain WO-1, WOL17 is located further from the telomere.

We have used WOL17 DNA as a probe of the EMBL-3 library of *C. albicans* WO-1 DNA to isolate a number of 15- to 18-kb subtelomeric stretches of DNA from different *Candida* telomeres. DNA samples from 11 hybridizing phage clones were purified and analyzed by restriction digestion; 9 of the clones proved to be clearly unique (data not shown). None of the clones hybridizes to the Ca7 repeats. When one

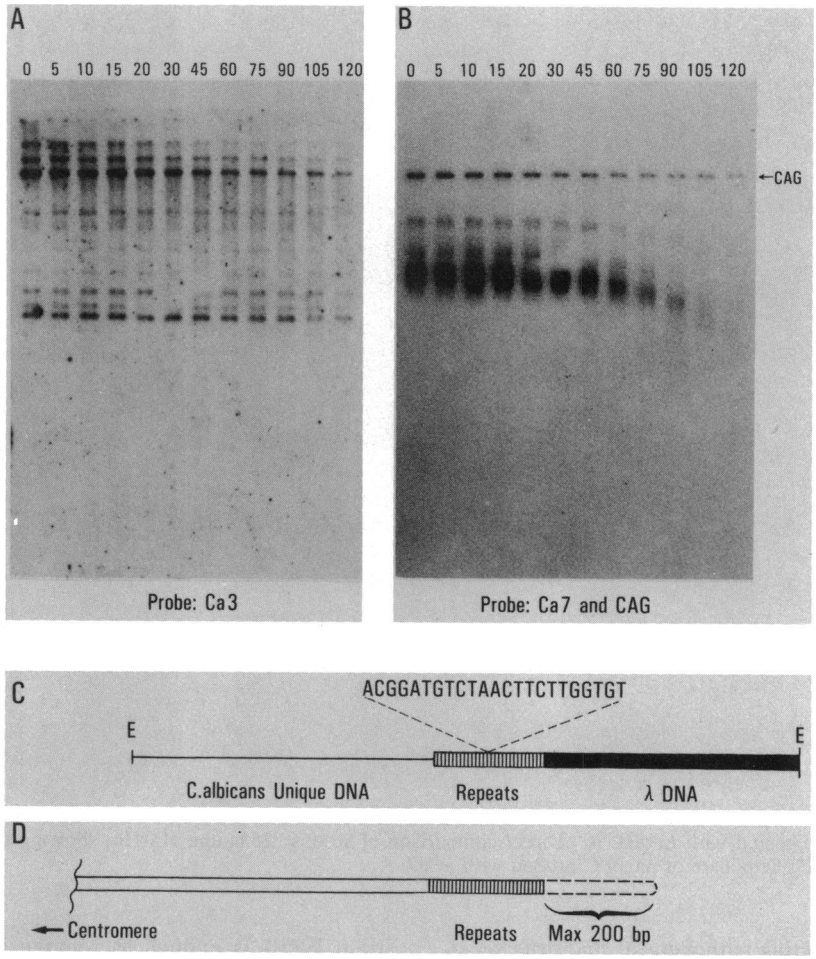

FIG. 4. *Bal* 31 sensitivity of Ca3 and Ca7 elements. *C. albicans* WO-1 DNA isolated from white or opaque cells was digested with *Bal* 31 for the indicated times (in minutes) and then digested with *Eco*RI. The DNA samples were probed with Ca3 (A) or Ca7 and *CAG1* (B) after gel electrophoresis and blotting. (C) Schematic representation of the λCa7 clone. The DNA sequence of the repeat element is also shown. (D) Diagram showing the relative position of the Ca7 repeats on a *C. albicans* chromosome.

of the clones (λWOL17.1), chosen arbitrarily, was radiolabeled and hybridized to a Southern blot containing digests of the rest of the clones, hybridization was observed to most of the *Candida*-derived DNA fragments of each of the other clones. This result indicates that even though there are restriction site polymorphisms, quite large subtelomeric regions may be conserved among different WO-1 chromosomes. Hybridization of this phage clone (λWOL17.1) to *Eco*RI-digested DNA of a number of *Candida* strains (Fig. 3C) revealed that all tested *C. albicans* strains show similar but distinct multiband patterns. This suggests that there may be appreciable conservation of at least some subtelomeric sequences among different *C. albicans* strains. Some hybridization is also detected to DNA from *C. krusei*, *C. glabrata*, and *S. cerevisiae*, although the significance of this is not known.

Use of Ca3 in strain identification. In previous studies (43, 44), we have shown that Ca3 and Ca7 fingerprints can be used to distinguish individual strains in infected individuals as well as to assess strain relatedness. Preliminary results (Fig. 5) allow some conclusions concerning their relative stability and utility as markers. We have surveyed a series of strains isolated from the oral mucosa of healthy human subjects. Figure 5A and B contain a panel of *Eco*RI-digested DNA from 3153a and strains isolated from six different healthy subjects. The set of exposures in Fig. 5 shows several properties of the Ca3 repeat. First, each mouth strain displays a distinct banding pattern made up of both band positions and relative intensities. Comparison of Fig. 5A and B shows that the lower half of the gel, containing mainly single-copy junction fragments, is nearly identical in all seven strains, whereas the heavy bands corresponding to the repeated internal fragments of Ca3 show different patterns in each strain. We infer from this result that internal rearrangement is more common in Ca3 than is transposition to new sites in the genome. In addition, it is noteworthy that the patterns are not infinitely variable but seem to be made up of interchangeable parts. In other words, there appears to be a limited repertoire for the variability of Ca3. In fact, the Ca3 fingerprint of one of the clinical stains, HMH6c, appears to be identical to that of 3153a, a strain that has been in the laboratory for over 10 years. Confirmation of this identity is shown in Fig. 5C and D, in which the two strains have been run and compared separately. The appearance of the 3153a fingerprint in a clinical isolate leads to the notion that *C. albicans* may be classified into related subgroups by using Ca3-related probes, raising the possibility that eventually these subgroups might show useful correlations with genetic

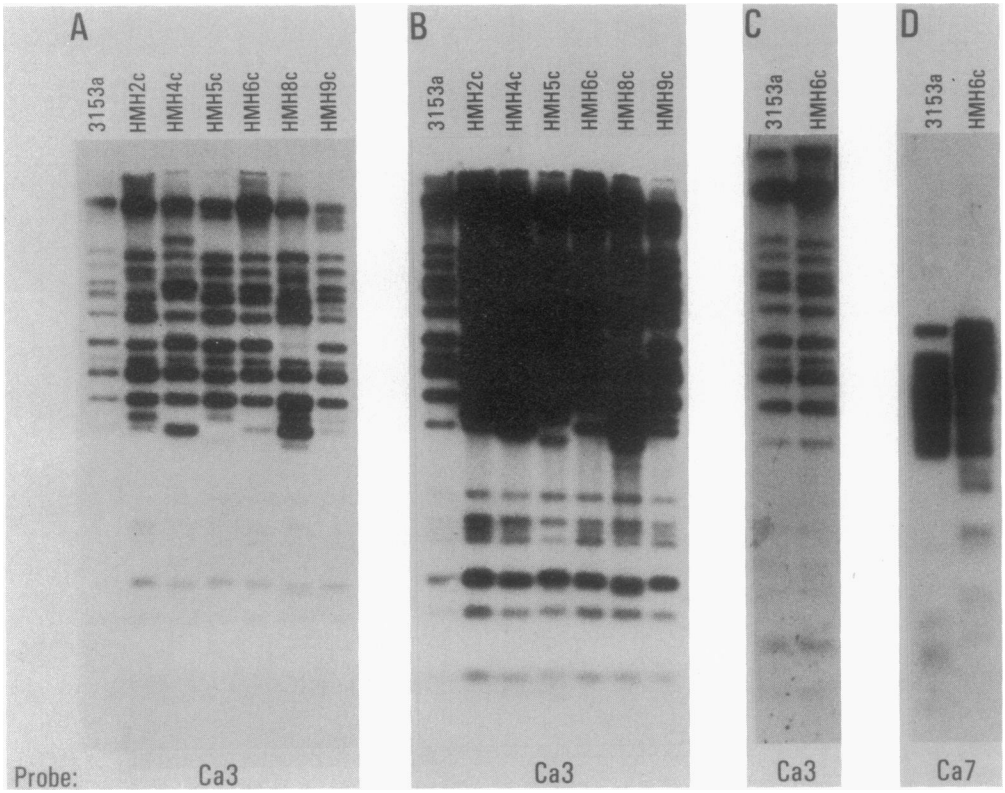

FIG. 5. (A and B) Different exposures of autoradiograms from a Southern blot of DNA from 3153a and six *C. albicans* strains isolated from healthy human mouths and digested with *Eco*RI. (C) Direct comparison of strains 3153a and HMH6, showing the identity in Ca3 pattern between these two strains. (D) Duplicate of panel C probed with λCa7.

differences in factors affecting pathogenesis and drug sensitivity.

DISCUSSION

In this report we have identified middle repetitive elements from the pathogenic yeast *C. albicans*. The elements that we have characterized the fullest are a telomere-associated tandem repeat (Ca7) and two related families of elements (Ca3 and Ca24) found in at least 20 different sites in the genome.

It is too early to tell whether the dispersed repetitive element designated Ca3 is similar in structure to a typical eucaryotic transposon, such as the *Drosophila* element copia (35) or *S. cerevisiae* TY1 (34), having terminal repeats flanking a relatively conserved core region. Some evidence, however, suggests that the Ca3 element has seldom moved during the time since *C. albicans* diverged from other yeast species. For example, in contrast to TY1 (34), the restriction map of the core region of Ca3 appears to be more variable than that of the junction fragments, and it appears to share at least one segment with a repeat family of different structure that we have designated Ca24. Our data are consistent with those of Scherer and Stevens (38) and indicate that Ca3 undergoes internal rearrangement more frequently than it hops to other sites in the genome.

In contrast to Ca3 and Ca24, the Ca7 repeat is made up of tandem copies of a simple 23-bp sequence. It is found on all *C. albicans* chromosomes, but in a specific location at or near the telomere. *Bal* 31 digestion indicates that removing 500 to 600 bp of DNA from the ends of chromosomes from strain WO-1 is enough to remove all of the Ca7 repeats. Given that the stretch of repeats in the Ca7 phage is ~370 bp in size, this suggests that there can be no more than ~250 bp of sequences separating the Ca7 repeats from the ends of chromosomes. The Ca7 repeats must therefore either be the telomeres themselves or else be a type of subtelomeric repeat situated extremely near the telomere.

The size and sequence of the Ca7 repeat are dissimilar to those of the chromosomal telomeric repeats that have thus far been described. Known telomeric repeats are somewhat variable, but all are shorter in size than the Ca7 sequence. Examples of sequences of telomeric repeats include C_4A_4 (from *Tetrahymena* sp.), C_3TA_2 (from vertebrates), and $C_{2\text{-}3}A\ (CA)_{1\text{-}3}$ (from *Saccharomyces* sp.) (see reference 50 for a review). This indicates that if the Ca7 repeats are, in fact, the *Candida* telomeres, they are not typical chromosomal telomeres.

The alternative explanation of the Ca7 sequences is that they represent not telomeric repeats but subtelomeric repeats. Subtelomeric repetitive elements appear to be very common, having been identified in many species, including *C. albicans*, as demonstrated in this study by the WOL17 sequence and its neighboring sequences. In *S. cerevisiae*, where subtelomeric repeats have been most studied, there are two large subtelomeric repeats, X and Y′ (2, 3). Y′ is a highly conserved 6.7-kb sequence, and X is a less well conserved element that varies from 0.3 to 3.7 kb. The Y′ element bears some resemblance to Ca7 in that it has been shown to contain a number of tandem 36-bp repeats (10).

With the available data it is impossible to distinguish whether the Ca7 repeats are telomeric or subtelomeric. The

lack of similarity to known chromosomal telomeres argues for a subtelomeric location, but there are reasons to not yet rule out that the Ca7 repeats are the actual telomeres. For example, the Ca7 repeats are present on every chromosome of every *C. albicans* isolate that we have examined. This is expected for a telomeric sequence but is often untrue for subtelomeric sequences, including the X and Y elements of *S. cerevisiae* (12, 51) and the WOL17 sequence of *C. albicans* (this work). Also consistent with the possibility that the Ca7 repeats are the *C. albicans* telomeres is the fact that labeled *Saccharomyces* and *Tetrahymena* telomeric sequences do not hybridize to *Candida* telomeric fragments (unpublished observations).

Interestingly, there is precedent for telomeres having a structure similar to that of the Ca7 repeats. The mitochrondrial telomeres of six *Tetrahymena* species have been examined and found to be composed of tandemly repeated sequences from 31 to 53 bp in size with no conserved sequence among them and also with no sequence similarity to the known chromosomal telomeres (24). It has been proposed that these telomeres maintain themselves via unequal crossing over rather than by utilizing telomerase-catalyzed de novo DNA synthesis, as is thought to maintain the chromosomal telomeres that have been studied in detail.

The data presented here indicate that middle-repetitive DNA sequences can be readily isolated from medically important yeast species and that these sequences can be employed to identify distinctive chromosomal structures, to distinguish between species, and to assess strain relatedness within a species.

ACKNOWLEDGMENTS

This work was supported by two National Institutes of Health grants, AI25213 (to J.B.H.) and AI23911 (to D.R.S.), a grant from the National Institute of Allergy and Infectious Diseases, and a research fellowship (DRG-963) from the Damon Runyon-Walter Winchell Cancer Fund (to M.J.M.).

REFERENCES

1. **Anderson, J., and D. R. Soll.** Unpublished observations.
1a. **Bodey, G. P., and V. Fainstein (ed.).** 1985. Candidiasis. Raven Press, New York.
2. **Chan, C. S. M., and B.-K. Tye.** 1983. Organization of DNA sequences and replication origins at yeast telomeres. Cell **33:** 563–573.
3. **Chan, C. S. M., and B.-K. Tye.** 1983. A family of *Saccharomyces cerevisiae* repetitive autonomously replicating sequences that have very similar genomic environments. J. Mol. Biol. **168:**505–523.
4. **Chandler, F. W.** 1985. Pathology of the mycoses in patients with the acquired immunodeficiency syndrome (AIDS). Curr. Top. Med. Mycol. **1:**1–23.
5. **Degregorio, M. W., W. M. F. Lee, C. A. Linker, R. A. Jacobs, and C. A. Ries.** 1982. Fungal infections in patients with acute leukemia. Am. J. Med. **73:**542–548.
6. **DeLange, T., and P. Borst.** 1982. Genomic environment of the expression-linked extra copies of genes for surface antigens of *Trypanosoma brucei* resembles the end of a chromosome. Nature (London) **299:**451–453.
7. **Gusella, J. F., C. Keys, A. Varsanyi-Breiner, F.-T. Kao, C. Jones, T. T. Puck, and D. Housman.** 1980. Isolation and localization of DNA segments from specific human chromosomes. Proc. Natl. Acad. Sci. USA **77:**2829–2833.
8. **Hansenlever, H. R., and W. O. Mitchell.** 1961. Antigenic studies of *Candida*. I. Observation of two antigenic graphs in *Candida albicans*. J. Bacteriol. **82:**570–573.
9. **Hopfer, R. L., V. Fainstein, M. P. Luna, and G. P. Bodey.** 1981. Disseminated candidiasis caused by four different *Candida* species. Arch. Pathol. Lab. Med. **105:**454–455.
10. **Horowitz, H., and J. E. Haber.** 1984. Subtelomeric regions of yeast chromosomes contain a 36 base-pair tandemly repeated sequence. Nucleic Acids Res. **12:**7105–7121.
11. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1984. Constructing and screening cDNA libraries in λgt10 and λgt11, p. 49–78. *In* D. Glover (ed.), DNA cloning: a practical approach, vol. 1. IRL Press, Oxford.
12. **Jager, D., and P. Philippsen.** 1989. Many yeast chromosomes lack the telomere-specific y′ sequence. Mol. Cell. Biol. **9:**5754–5757.
13. **Keil, R. L., and G. S. Roeder.** 1984. *cis*-acting, recombination-stimulating activity in a fragment of the ribosomal DNA of *S. cerevisiae*. Cell **39:**377–386.
14. **Kwon-Chung, K. J., W. S. Riggsby, R. A. Uphoff, J. B. Hicks, W. L. Whelan, E. Reiss, B. B. Magee, and B. L. Wicks.** 1989. Genetic differences between type I and type II *Candida stellatoidea*. Infect. Immun. **57:**527–532.
15. **Land, G. A., B. A. Harrison, K. L. Hulme, B. H. Cooper, and J. C. Byrd.** 1979. Evaluation of the new API 20C strip for yeast identification against a conventional method. J. Clin. Microbiol. **10:**357–364.
16. **Lasker, B. A., G. F. Carle, G. S. Kobayashi, and G. Medoff.** 1989. Comparison of the separation of *Candida albicans* chromosome-sized DNA by pulsed-field gel electrophoresis techniques. Nucleic Acids Res. **17:**3783–3793.
17. **Lauer, G. D., T. M. Roberts, and L. C. Klotz.** 1977. Determination of the nuclear DNA content of *Saccharomyces cerevisiae* and implications for the organization of DNA in yeast chromosomes. J. Mol. Biol. **114:**507–526.
18. **Lustig, A. J., and T. D. Petes.** 1986. Identification of yeast mutants with altered telomere structure. Proc. Natl. Acad. Sci. USA **83:**1398–1402.
19. **Magee, B. B., T. M. D'Souza, and P. T. Magee.** 1987. Strain and species identification by restriction fragment length polymorphisms in the ribosomal DNA repeat of *Candida* species. J. Bacteriol. **169:**1639–1643.
20. **Magee, B. B., Y. Koltin, J. A. Gorman, and P. T. Magee.** 1988. Assignment of cloned genes to the seven electrophoretically separated *Candida albicans* chromosomes. Mol. Cell. Biol. **8:**4721–4726.
21. **Magee, B. B., and P. T. Magee.** 1987. Electrophoretic karyotypes and chromosome numbers in *Candida* species. J. Gen. Microbiol. **133:**425–430.
22. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **McCreight, M. C., D. W. Warnock, and M. V. Martin.** 1985. Resistogram typing of *Candida albicans* isolates from oral and cutaneous sites in irradiated patients. Sabouraudia **23:**403–406.
24. **Morin, G. B., and T. R. Cech.** 1988. Mitochondiral telomeres: surprising diversity of repeated telomeric DNA sequences among six species of *Tetrahymena*. Cell **52:**367–374.
25. **Myerowitz, R. L., G. J. Pazin, and C. M. Allen.** 1977. Disseminated candidiasis: changes in incidence, underlying diseases and pathology. Am. J. Clin. Pathol. **68:**29–38.
26. **Odds, F. C.** 1988. Candida and candidosis. Balliere Tindall, Baltimore.
27. **Odds, F. C.** 1982. Genital candidiasis. Clin. Exp. Dermatol. **7:**345–354.
28. **Odds, F. C.** 1987. Candida infections: an overview. Crit. Rev. Microbiol. **15:**1–5.
29. **Odds, F. C., and A. B. Abbott.** 1980. A simple system for the presumptive identification of *Candida albicans* and differentiation of strains within the species. Sabouraudia **18:**302–317.
30. **Parker, J. C., J. J. McCloskey, and K. A. Knauer.** 1976. Pathobiologic features of human candidiasis: a common deep mycosis of the brain, heart and kidney in the altered host. Am. J. Clin. Pathol. **65:**991–1000.
31. **Polonelli, L., C. Archibusacci, M. Sestito, and G. Morace.** 1983. Killer system: a simple method for differentiating *Candida albicans* strains. J. Clin. Microbiol. **17:**774–780.
32. **Riggsby, W. S., L. J. Torres-Bauza, J. W. Wills, and T. M. Tomes.** 1982. DNA content, kinetic complexity and the ploidy

question in *Candida albicans*. Mol. Cell. Biol. **2**:853–862.
33. **Rikkerink, E. H. A., B. B. Magee, and P. T. Magee.** 1988. Opaque-white phenotype transition: a programmed morphological transition in *Candida albicans*. J. Bacteriol. **170**:895–899.
34. **Roeder, G. S., and G. R. Fink.** 1983. Transposable elements in yeast, p. 299–324. *In* J. A. Shapiro (ed.), Mobile genetic elements. Academic Press, Inc., New York.
35. **Rubin, G. M.** 1983. Dispersed repetitive DNAs in Drosophila, p. 329–362. *In* J. A. Shapiro (ed.), Mobile DNA elements. Academic Press, Inc., New York.
36. **Rustchenko-Bulgac, E. P., F. Sherman, and J. B. Hicks.** 1990. Chromosomal rearrangements associated with morphological mutants provide a means for genetic variation of *Candida albicans*. J. Bacteriol. **172**:1276–1283.
37. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
38. **Scherer, S., and D. A. Stevens.** 1988. A *Candida albicans* dispersed, repeated gene family and its epidemiologic applications. Proc. Natl. Acad. Sci. USA **85**:1452–1456.
39. **Sherman, F., G. R. Fink, and J. B. Hicks.** 1987. Methods in yeast genetics: a laboratory course manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
40. **Slutsky, B., J. Buffo, and D. R. Soll.** 1985. High frequency switching of colony morphology in *Candida albicans*. Science **230**:666–669.
41. **Slutsky, B., M. Staebell, J. Anderson, L. Risen, M. Pfaller, and D. R. Soll.** 1987. "White-opaque transition": a second high-frequency switching system in *Candida albicans*. J. Bacteriol. **169**:189–197.
41a. **Soll, D. R.** Unpublished data.
42. **Soll, D. R., G. W. Bedell, J. Thiel, and M. Brummel.** 1981. The dependency of nuclear division on volume in the dimorphic yeast *Candida albicans*. Exp. Cell Res. **133**:55–62.
43. **Soll, D. R., C. J. Langtimm, J. McDowell, J. Hicks, and R. Galask.** 1987. High-frequency switching in *Candida* strains isolated from vaginitis patients. J. Clin. Microbiol. **25**:1611–1622.
44. **Soll, D. R., M. Staebell, C. Langtimm, M. Pfaller, J. Hicks, and T. V. G. Rao.** 1988. Multiple *Candida* strains in the course of a single systemic infection. J. Clin. Microbiol. **26**:1448–1459.
45. **Suzuki, T., I. Kobayashi, T. Kanbe, and K. T. Tanaka.** 1989. High-frequency variation of colony morphology and chromosome reorganization in the pathogenic yeast *Candida albicans*. J. Gen. Microbiol. **135**:425–434.
46. **Whelan, W.** 1987. The genetics of medically important fungi. Crit. Rev. Microbiol. **14**:99–170.
47. **Whelan, W. L., R. M. Partridge, and P. T. Magee.** 1980. Heterozygosity and segregation in *Candida albicans*. Mol. Gen. Genet. **180**:107–113.
48. **Whelan, W. L., and D. R. Soll.** 1982. Mitotic recombination in *Candida albicans*: recessive lethal alleles linked to a gene required for methionine biosynthesis. Mol. Gen. Genet. **187**:477–485.
49. **Wills, J. W., B. A. Lasker, K. Sirotkin, and W. S. Riggsby.** 1984. Repetitive DNA of *Candida albicans*: nuclear and mitochondrial components. J. Bacteriol. **157**:918–924.
50. **Zakian, V. A.** 1989. Structure and function of telomeres. Annu. Rev. Genet. **23**:579–604.
51. **Zakian, V. A., and H. M. Blanton.** 1988. Distribution of telomere-associated sequences on natural chromosomes in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **8**:2257–2260.