JOURNAL OF CLINICAL MICROBIOLOGY, June 1993, p. 1472–1480
0095-1137/93/061472-09$02.00/0

Vol. 31, No. 6

Characterization and Partial Nucleotide Sequence of the DNA Fingerprinting Probe Ca3 of *Candida albicans*

JULIA ANDERSON,[1] THYAGARAJAN SRIKANTHA,[1] BRIAN MORROW,[1] SHELLEY H. MIYASAKI,[2] THEODORE C. WHITE,[2] NINA AGABIAN,[2] JAN SCHMID,[1]† AND DAVID R. SOLL[1]*

Department of Biology, University of Iowa, Iowa City, Iowa 52242-1324,[1] and Intercampus Program in Molecular Parasitology, Department of Stomatology, and Oral AIDS Center, University of California, San Francisco, San Francisco, California 94143[2]

Received 16 November 1992/Accepted 2 March 1993

The moderately repetitive Ca3 fragment of *Candida albicans* has been used as an effective DNA fingerprinting probe in epidemiological studies. *Eco*RI digestion of Ca3 DNA results in seven fragments of 4.2 kb (A), 2.98 kb (B), 2.85 kb (C), 0.77 kb (D_1), 0.77 kb (D_2), 0.38 kb (E), and 0.30 kb (F). Five of these *Eco*RI fragments have been mapped in the 5′-3′ order C B D_1 A D_2. The intact Ca3 probe and the three largest *Eco*RI fragments, A, B, and C, were individually used to probe Southern blots of *Eco*RI-digested DNA of a set of test strains, transverse alternating field electrophoresis-separated chromosomes of strain 3153A, and Northern (RNA) blots of test strain 3153A. Fragments A, B, and C each generate a different Southern blot hybridization pattern with *Eco*RI-digested whole-cell DNA; Ca3 sequences are present in at least five of seven separable chromosomes and a minichromosome of strain 3153A; fragments A, B, and C are distributed differently on chromosomes; and fragments A, B, and C do not cross-hybridize. Ca3 hybridizes to three major transcripts of 2.8, 2.3, and 1.5 kb. Fragment A hybridizes intensely to the 1.5-kb transcript, while fragments B and C both hybridize intensely to the 2.8- and 2.3-kb transcripts. The B fragment, which contains 2,980 bp and contributes to the major portion of the Ca3 pattern, was sequenced. Both direct and inverted repeat sequence motifs were identified. These results provide us with initial insights into the evolution of the Ca3 pattern and the nature of the probe.

Although the yeast *Candida albicans* remains a pervasive pathogen in humans, relatively little is known about the relationship between commensal and infecting strains, the origins of infecting strains, transmission of the organism between individuals, and strain specialization (9, 20, 21). To answer these questions, DNA fingerprinting methods which utilize species-specific, moderately repetitive DNA sequences as probes have been developed (13, 16, 18, 25). Such probes generate relatively complex Southern blot hybridization patterns which differ between most strains but are relatively stable within a single strain over many generations (16, 18). By comparing band positions and intensities between patterns of different strains, a similarity coefficient (S_{AB}) calculated between every pair of strains can be used to generate dendrograms (18). This process has been enhanced by the development of the computer-assisted Dendron system which provides image processing, automated scanning, calculation of similarity coefficients, genesis of dendrograms, and retrospective analyses of large numbers of strains (17, 18, 20, 22).

One such fingerprinting probe, Ca3, is *C. albicans* specific (13, 26), generating a pattern of between 15 and 25 bands of various intensities. Ca3 fingerprinting has proven effective in a number of epidemiological studies involving significant numbers of strains (8, 17, 18, 20, 21, 23–26). Because the Dendron system retains the digitized Ca3 Southern blot hybridization pattern of every *C. albicans* strain analyzed and retrospectively compares all newly analyzed strains with all previously analyzed strains, a data base for epidemiological studies which can be used for the analysis of geographical distribution, transmission, and strain specialization is evolving (17, 19, 22, 24). The data base now contains Ca3 Southern blot hybridization patterns of more than 300 independently isolated strains of *C. albicans*, with information on every host.

Because of the potential value of Ca3 in epidemiological studies, it is important that we learn as much as we can about the evolution of the Ca3 hybridization pattern so that we can weight band differences in the S_{AB} calculation in order to describe more accurately genetic relatedness between strains in dendrograms. A more complete molecular characterization of the Ca3 probe is one prerequisite to this objective. Here, we have used the three largest *Eco*RI fragments of the Ca3 probe, A, B, and C, to identify homologous bands in the total Ca3 hybridization pattern. We have also mapped the distribution of these fragments on individual chromosomes of the originating strain 3153A, sequenced the entire B fragment, and identified two major transcripts which hybridize with the B and C fragments and one which hybridizes with the A fragment. These results provide us with the first insights into the genesis of the Ca3 pattern and the nature of the probe.

MATERIALS AND METHODS

Southern blot hybridization and genesis of dendrograms. Cells from storage cultures maintained on SM agar (10 g of glucose, 10 g of Bacto Peptone, 1.0 g of $MgSO_4 \cdot 7H_2O$, 2.2 g of KH_2PO_4, 1 g of K_2HPO_4, 1 g of yeast extract, 10 g of Bacto Agar, in 1 liter of H_2O) were grown in YPD medium (2% glucose, 2% Bacto Peptone, 1% yeast extract) to late log or early stationary phase at 25°C, and genomic DNA was prepared by the methods of Scherer and Stevens (15). The concentration of DNA in each preparation was measured by ethidium bromide dot quantitation (6). Each DNA prepara-

* Corresponding author.
† Present address: Department of Microbiology and Genetics, Massey University, Palmerston North, New Zealand.

tion was then digested with *Eco*RI, separated on a 0.8% (wt/vol) agarose gel, transferred by capillary blotting to nitrocellulose membrane, and hybridized with nick-translated probes (18). Probes included the intact Ca3 sequence (13, 18, 25), the three major *Eco*RI fragments of Ca3 (A, B, and C), and the three *Xba*I fragments of B (B-1, B-2, and B-3). Three replica gels, each containing the same set of DNA preparations, including the standard strain 3153A and 12 clinical isolates, were simultaneously analyzed with different probes and were stripped and rehybridized with additional probes. The three replica gels exhibited Ca3 hybridization patterns which were nearly indistinguishable.

The method for calculating the similarity coefficient (S_{AB}) between pairs of Southern blot hybridization patterns has been described in detail in a previous report (18). Patterns of the same strain on different gels result in an average S_{AB} of 0.96. Independent, unrelated test isolates result in an average S_{AB} of 0.69 (18), which has been considered a value reflecting unrelatedness. Patterns with no shared molecular size bands result in an S_{AB} of 0.00. An arbitrary threshold for highly related strains has been set at 0.85 by previous studies (24). Dendrograms were generated through the Dendron program (18, 20, 22, 24) by the unweighted pair group method (19).

Mapping of lambda Ca3. Ca3 mapping was done according to the method of Rackwitz et al. (11) by partially digesting Ca3 DNA with *Eco*RI for various timed intervals and then by hybridization in solution to ^{32}P-labeled oligonucleotide probes ON-L and ON-R according to the manufacturers specifications (Promega Biotec, Madison, Wis.). The labeled partial digests were fractionated in a 0.4% agarose gel in 1× TAE (0.04 M Tris-acetate, 0.001 M EDTA) for 48 h at 1.5 V/cm and then autoradiographed. Additional mapping was performed with the restriction enzymes *Dra*III, *Pst*I, and *Sal*I, which had unique sites within each fragment of Ca3. The multiple Southern blots of these digests were hybridized with the radioactive *Eco*RI fragments A, B, C, D_1, and D_2. The common hybridization band(s) resulting from any two Ca3 restriction fragments indicated neighbor relations and the order of the five *Eco*RI fragments of Ca3.

Pulsed-field gel electrophoresis. Chromosomal plugs were made from logarithmically growing suspension cultures of strain 3153A. Cells were treated with a zymolase and novozyme mixture, and the resulting spheroplasts were suspended in 1% SeaPlaque GTG agarose (FMC BioProducts, Rockland, Maine). The plugs were treated for 60 h in an EDTA-proteinase K-sarcosyl mixture at 50°C. Washed plugs were stored at 4°C in a solution of 10 mM Tris-HCl and 1 mM EDTA, pH 8.0. Transverse alternating field electrophoresis (TAFE) was carried out in a Beckman Geneline apparatus with 0.65% LE agarose and 1× TAFE (Tris-acetate) buffer (Beckman Instruments, Inc., Palo Alto, Calif.). The electrophoresis protocol was in five 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; and 80 V for 18 h, with a 10-min switch time. Gels were stained with ethidium bromide and photographed. Chromosomal DNA was then transferred to Zetabind membrane (Cuno, Inc., Meriden, Conn.), and Southern blot hybridization was carried out according to the method of Church and Gilbert (3).

DNA sequencing and computer analysis. The sequence of the 2,980-bp *Eco*RI B fragment of Ca3 was determined by subcloning into pGEM-3Zf(+) plasmid vector. Both orientations of the sequence were derived by the dideoxy-chain termination method (14) with Sequenase version 2.0 (U.S. Biochemicals, Cleveland, Ohio). The entire DNA sequence was achieved by subcloning various smaller restriction fragments into suitable pGEM plasmid vectors. Certain regions of the sequence were derived by using two sets of gene-specific oligonucleotide primers. These primers are: (i) 5-CAAGTTGTGATGGAAAG-3 (1205 to 1221) and its complementary primer 5-CTTTCCATCACAACTTG-3 (1221 to 1205) and (ii) 5-AGCTTGATCTAGGGGCT-3 (1717 to 1733) and its complementary primer 5-AGCCCCTAGATCAA

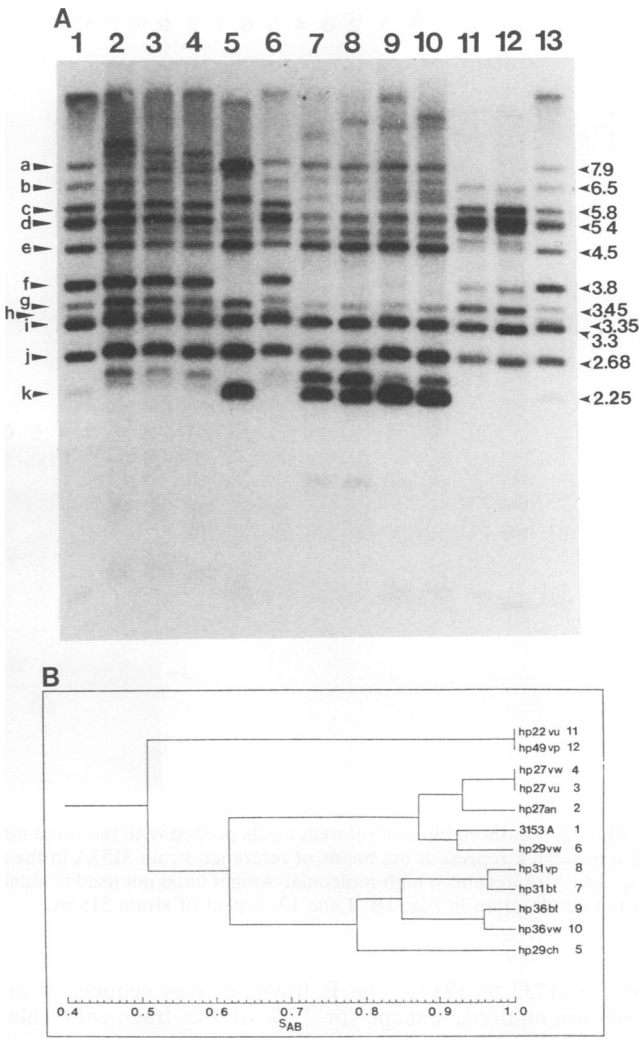

FIG. 1. Southern blot (A) of *Eco*RI-digested DNA of test strains of *C. albicans* hybridized with the moderately repetitive genomic probe Ca3 and a dendrogram (B) based on the similarity coefficients generated by the Dendron program. (A) Southern blot of 12 independent strains. The designations for the 12 independent strains can be found to the right of the dendrogram (B). Lane 13, repeat of strain 3153A also represented in lane 1. The bands of strain 3153A which encompass the size range used in S_{AB} computations between all strains are labeled a through k. The molecular sizes of a through k estimated by comparison to known standards run in parallel are presented to the right of the panel. (B) Dendrogram of the strains in the Southern blot (A) based upon the S_{AB} values computed between all pairs by the Dendron software package. hp22, -27, -29, -31, -36, and -49 represent healthy young women; anatomical origins of the isolates include vu, vulva; vp, vaginal pool; vw, vaginal wall; an, anus; bt, back of tongue; and ch, cheek (buccal epithelium).

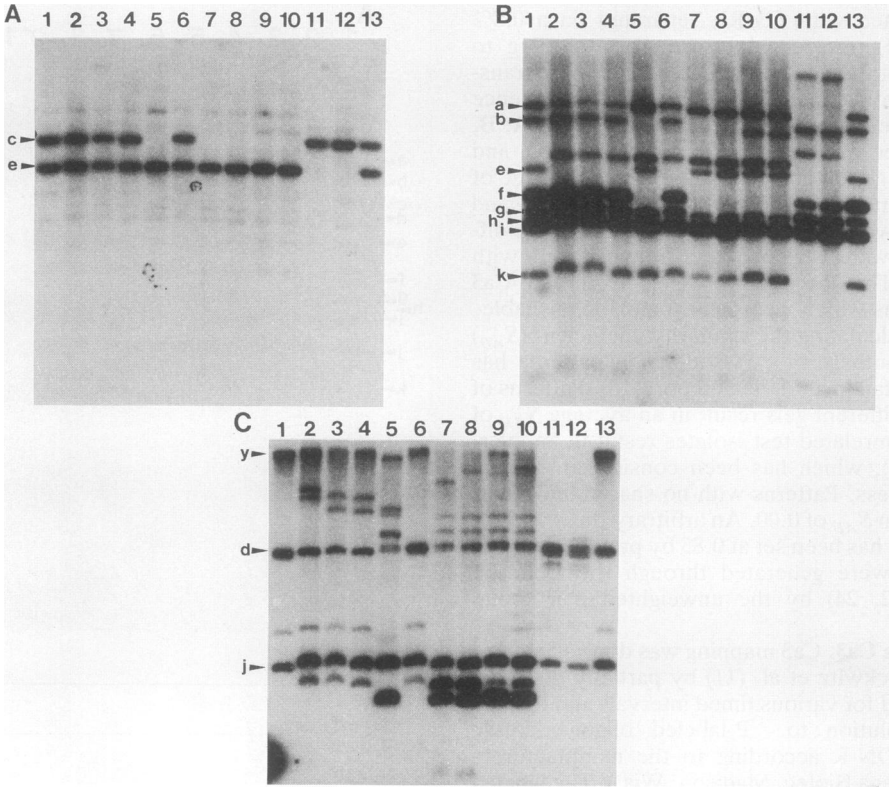

FIG. 2. Southern blots of reference gels probed with the three major *Eco*RI fragments of Ca3. A, fragment A; B, fragment B; C, fragment C. a through k represent the bands of reference strain 3153A in the size range from 2.25 to 7.9 kb (see complete pattern generated by Ca3 in Fig. 1A); y represents a high-molecular-weight band not used in standard computation of S_{AB} values. The strain code can be found to the right of the dendrogram in Fig. 1B. Lane 13, repeat of strain 3153A.

GCT-3 (1733 to 1717). The B fragment was sequenced in both orientations, except for 17% of the fragment. This portion of the sequence was confirmed by using two or more independently isolated plasmid clones. Sequence analyses were performed both with Microgenie Software (Intelligenetics) and with Genetics Computer Group software (University of Wisconsin, Madison, Wis.) (4).

RNA analysis. For Northern (RNA) blot analysis, total cellular RNA from strain 3153A was isolated by the guanidinium isothiocyanate extraction method (2). Fifteen micrograms of RNA was fractionated on a formaldehyde–1.25% agarose gel. Gels were blotted onto a Zetabind nylon membrane and hybridized with radioactively labeled Ca3 as well as with the *Eco*RI fragments A, B, and C by the method of Church and Gilbert (3). Molecular weights were estimated from known molecular weight standards run in parallel.

Nucleotide sequence accession number. The nucleotide sequence reported here for fragment B has been submitted to GenBank and assigned accession number L11068.

RESULTS

The basic Ca3 Southern blot hybridization pattern. When probed with the entire Ca3 fragment, Southern blots of *Eco*RI-digested DNA of *C. albicans* strains exhibit patterns composed of 10 to 20 bands of relatively high intensity and 4 to 6 bands of low intensity. The diversity of patterns obtained from different strains and the similarity of patterns of strains sampled from different body locations of the same individual or from different individuals are demonstrated in the Southern blot in Fig. 1A. These relationships are demonstrated in the dendrogram in Fig. 1B, generated from the similarity coefficients (S_{AB}s) calculated between all pairs in the group of 11 clinical isolates and laboratory strain 3153A in the Southern blot in Fig. 1A.

Southern blot hybridization patterns of the three major *Eco*RI fragments of Ca3. Since the Ca3 sequence is quite large, including approximately 12 kb, we did not know if the entire sequence or only a small portion of the sequence was responsible for the Southern blot hybridization pattern, or what portion of the sequence was repetitive in the genome. We therefore digested the sequence with *Eco*RI in order to examine the hybridization patterns of the individual fragments. Fractionation of the *Eco*RI-digested Ca3 probe on a 1.2% agarose gel resulted in six fragments with the following approximate sizes: A, 4.2 kb; B, 3.2 kb; C, 2.85 kb; D, 0.77 kb; E, 0.38 kb; and F, 0.30 kb. Further restriction fragment analysis (see Materials and Methods) demonstrated two D fragments of 0.77 kb, D_1 and D_2, and ordered the five largest fragments in the sequence C B D_1 A D_2.

Hybridization of *Eco*RI-digested DNA of reference strain 3153A with the entire Ca3 probe generates a pattern which includes bands with approximate molecular sizes of 7.9, 6.5, 5.8, 5.4, 4.5, 3.8, 3.45, 3.35, 3.3, 2.68, and 2.25 kb. These bands have been coded a through k, respectively, in Fig. 1A. *Eco*RI fragments A, B, and C of Ca3 were each used to probe one of the replica gels. In strain 3153A, fragment A hybridized intensely with bands c and e (Fig. 2A); fragment

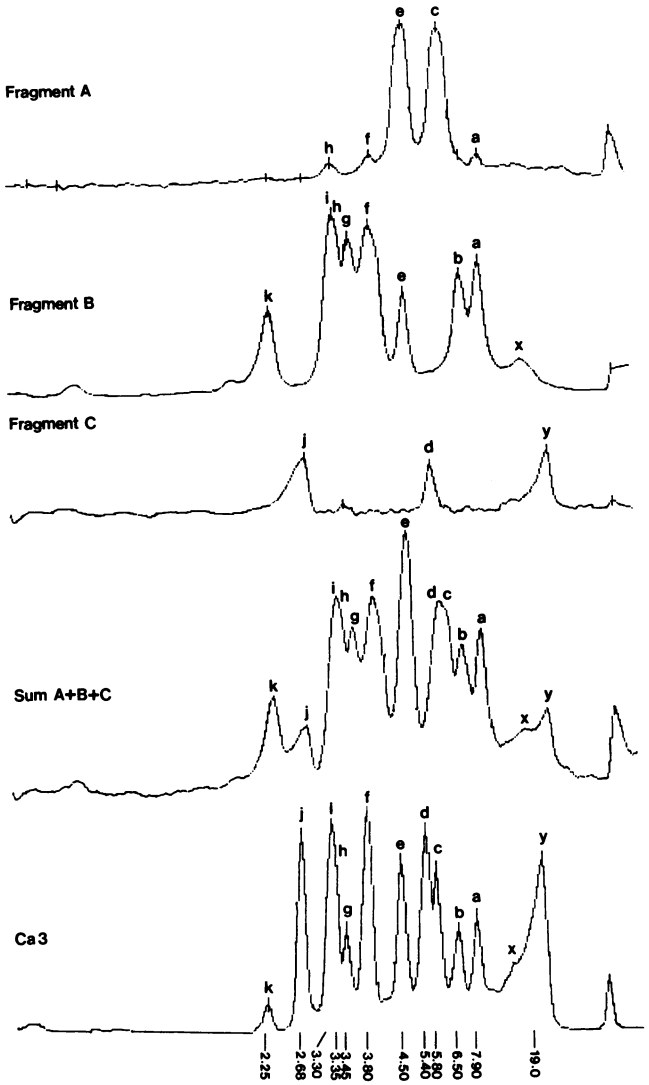

FIG. 3. Pixel density scans of the Southern blot hybridization patterns generated with the *Eco*RI fragments A, B, and C of the Ca3 probe for strain 3153A (lanes 1 in panel A, B, and C of Fig. 2). Band labels are explained in the legends to Fig. 1 and 2. Southern blot hybridization patterns were automatically scanned for pixel density by the Dendron software package for fragments A, B, and C, and summed (Sum A+B+C). The Ca3 hybridization pattern for strain 3153A (lane 1, Fig. 1A) was also scanned and compared with the individual scans and summed scan. Molecular sizes (in kilobases) of the bands are presented at the bottom of the figure.

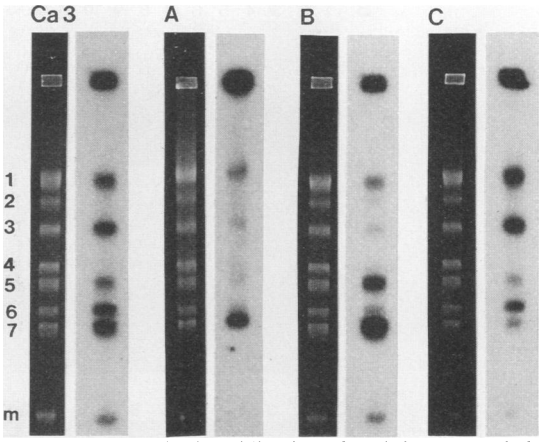

FIG. 4. Southern blot hybridization of TAFE-separated chromosomes of *C. albicans* 3153A with the intact Ca3 probe, and *Eco*RI fragments A, B, and C. Seven chromosomal bands (1 through 7) and a minichromosome (m) were separated in each lane and are visualized by ethidium bromide staining to the left of each blot. Southern blots were exposed for various times to demonstrate the full hybridization pattern.

B hybridized intensely with bands a, b, e, f, g, h, i, and k (Fig. 2B); and fragment C hybridized intensely with bands d and j (Fig. 2C). In addition to the coded bands a through k in the size range from 7.9 to 2.25 kb, fragment C hybridized with a band, coded y, at 19 kb (Fig. 2C). In Fig. 3, scans are plotted of the Southern blot hybridization patterns of *Eco*RI-digested DNA of strain 3153A probed with fragments A, B, or C (lanes 1 in Fig. 2A, B, and C, respectively). The individual scans were compared with a computer-generated summed plot (Sum A+B+C) as well as to a plot of the intact Ca3 pattern. The individual scans not only demonstrate the position of major bands, but also suggest low-intensity bands in the fragment A pattern at a, f, and h, and a high-molecular-weight band (x) in the B pattern at roughly 12.5 kb. The most intense bands in the A-fragment pattern (c and e) clearly differ from the most intense bands in the C-fragment pattern (y, d, and j). Only one major band (e) in the B-fragment pattern, the most complex pattern of the three fragments, correlates with a band in the A-fragment pattern, and no major bands in the B-fragment pattern correlate with major bands in the C-fragment pattern. The summed scans of the A-, B-, and C-fragment hybridization patterns exhibit every band in the intact Ca3 pattern between 2.3 and 19 kb (Fig. 3). Strict additivity in the summed plot (Sum A+B+C) is not expected, since band intensities are not in the linear densitometry range for pixel analysis by the Dendron program. The differences in the banding patterns for A, B, and C suggest that they do not share major homologous sequences; this conclusion is reinforced by the absence of cross-hybridization between the three *Eco*RI fragments of Ca3 (data not shown).

Variability between strains of the Southern blot hybridization patterns generated with the individual Ca3 fragments. (i) Fragment A. The Southern blot hybridization pattern of *Eco*RI-digested DNA of strain 3153A probed with fragment A contained two major bands at 5.8 and 4.5 kb (Fig. 2A and 3). For the additional test strains in the master gel, fragment A hybridization generated three different patterns (Fig. 2A): 5.8- and 4.5-kb bands (hp27an, hp27vu, hp27vw, and hp29vw [lanes 2, 3, 4, and 6, respectively]), as in the case of strain 3153A; a 5.8-kb band alone (hp22vu and hp49vp [lanes 11 and 12, respectively]); or a 4.5-bp band alone (hp29ch, hp31bt, hp31vp, hp36bt, and hp36vw [lanes 5, 7, 8, 9, and 10, respectively]).

(ii) Fragment B. The Southern blot hybridization pattern of *Eco*RI-digested DNA of strain 3153A probed with fragment B contained 8 of the 13 major bands of the intact Ca3 pattern in the molecular size range from 2.25 to 19.0 kb (Fig. 2B; Fig. 3, lane 1). For the additional test strains in the replica gel, fragment B generated a variety of patterns (Fig. 2B). All strains contained the 3.45 (g)- and 3.3 (i)-kb bands. All other bands exhibited variability between strains. In most cases, variability involved the presence or absence of a band at a

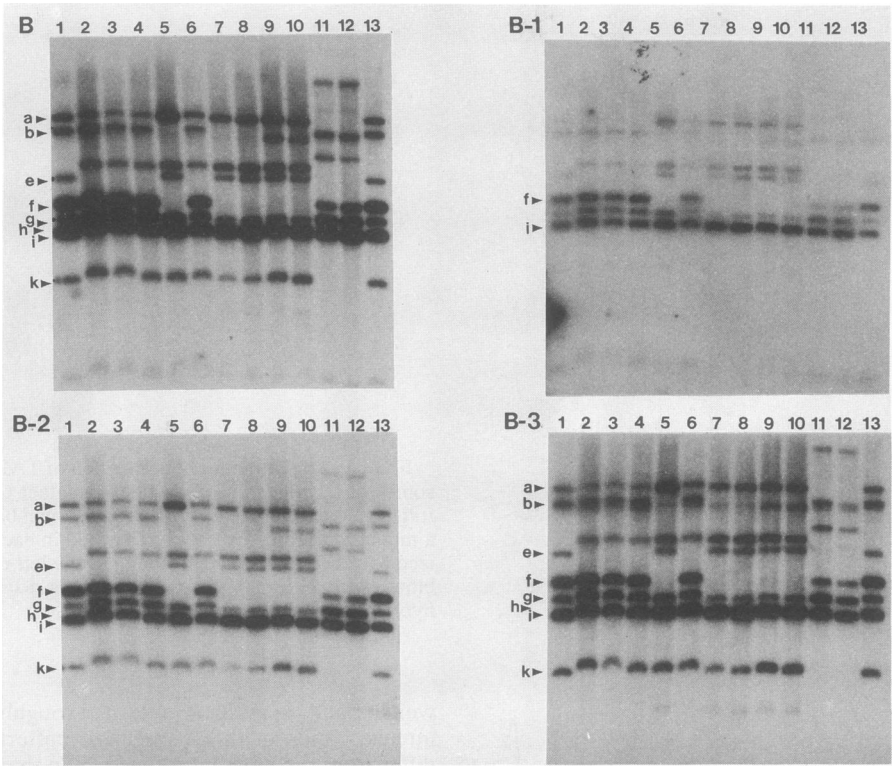

FIG. 5. Southern blots of reference gels probed with the intact fragment B and the three *Xba*I subfragments B-1, B-2, and B-3. a through k represent the bands of reference strains 3153A (see complete pattern generated by Ca3 [Fig. 1A]). The strain code can be found to the right of the dendrogram in Fig. 1B.

particular molecular size position. The most variant patterns were observed in the unusual strains hp22vu (Fig. 2B, lane 11) and hp49vp (Fig. 2B, lane 12) (24). Excluding strains hp22vu and hp49vp, the patterns of all other strains contained the 7.9 (a)- and 2.25 (k)-kb bands (Fig. 2B).

(iii) **Fragment C.** The fragment C pattern of strain 3153A (Fig. 2C, lane 1) contained 3 of the 13 major bands in the molecular size range from 2.25 to 19.0 kb (Fig. 2C and 3). The patterns of all additional strains in the master gel probed with fragment C contained two of these bands, at 5.4 (d) and 2.68 (j) kb (Fig. 2C); except for the unusual strains hp22vu (lane 11) and hp49vp (lane 12), all contained a number of additional bands of high intensity, just below 2.68 kb (j), and several bands between 5.4 kb (d) and 19.0 kb (y), which have been demonstrated to vary in Ca3 Southern blot hybridization patterns between highly similar strains isolated from different body locations of the same individual (24). Examples of this variability are evident both in the Ca3 hybridization patterns (Fig. 1A) and the fragment C hybridization patterns (Fig. 2C) for the pairs hp27an-hp27vu (lanes 2 and 3), hp31bt-hp31vp (lanes 7 and 8), and hp36bt-hp36vw (lanes 9 and 10).

Distribution of the A, B, and C fragments in *C. albicans* chromosomes. TAFE separates the chromosomes of *C. albicans* 3153A into seven major bands (12). Our TAFE gels also discriminated a low-molecular-weight minichromosome previously demonstrated in strain WO-1 (7). The intact lambda Ca3 probe hybridized to chromosomes 1, 3, 5, 6, and 7 and the minichromosome (Fig. 4). Strong hybridization signals were observed with chromosomes 1, 3, 6, and 7, and moderate hybridization was observed with chromosome 5 and the minichromosome. A negligible signal was observed with chromosomes 2 and 4 (Fig. 4). TAFE blots were then probed with each of the three largest *Eco*RI fragments of Ca3. Fragment A hybridized intensely with chromosome 7 and moderately with chromosome 1 (Fig. 4). There was weak to negligible hybridization with chromosomes 2, 3, 4, 5, and 6 and the minichromosome. Fragment B hybridized intensely with chromosomes 5 and 7, moderately with chromosomes 1 and 6 and the minichromosome, and weakly or negligibly with chromosomes 2, 3, and 4, (Fig. 4). Fragment C hybridized intensely with chromosomes 1, 3, and 6, moderately with chromosomes 5 and 7, and negligibly with chromosomes 2 and 4 and the minichromosome (Fig. 4).

Southern blot hybridization patterns of the three *Xba*I fragments of fragment B. Since fragment B generated the hybridization pattern most similar to that of the intact Ca3 probe and generated the highest number of intense bands, we further analyzed the B fragment by digestion with the endonuclease *Xba*I and used the *Xba*I fragments of B to reprobe replica gels of *Eco*RI-digested DNA. The B fragment possessed two *Xba*I sites and generated fragments B-1 (931 bp), B-2 (1,237 bp), and B-3 (812 bp). The hybridization pattern of *Eco*RI-digested DNA of strain 3153A probed with the B fragment included bands a, b, e, f, g, h, i, and k (Fig. 5). The hybridization pattern generated with B-1 included intense bands at f and i, a weak band at g, and very weak bands at a, b, and e (Fig. 5). The patterns generated by B-2 and B-3 included all bands in the pattern generated by the entire B fragment, and at roughly the same relative intensities. In Fig. 6, scans of the 3153A hybridization patterns for the individual *Xba*I fragments of B are plotted, a summed

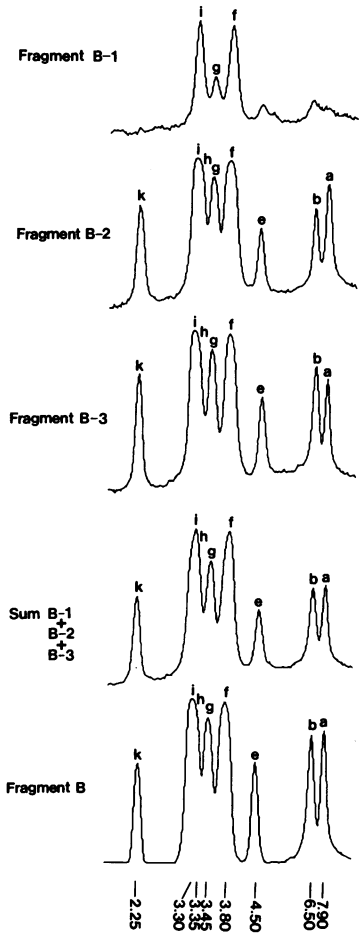

FIG. 6. Pixel density scans of the Southern blot hybridization patterns generated with the *Xba*I subfragments B-1, B-2, and B-3 of fragment B for strain 3153A (lanes 1 in panels B-1, B-2, and B-3, respectively [Fig. 5]). Southern blot hybridization patterns were automatically scanned for pixel density by the Dendron software package. Summed scans and a scan of the B-fragment hybridization pattern are also plotted for comparison. Molecular sizes (in kilobases) of the bands are presented at the bottom.

plot (B-1 + B-2 + B-3) was generated, and the summed plot was compared with the scan of the B-fragment pattern. It is evident in the scans that (i) in the pattern of B-1, band g is reduced and bands a, b, e, h, and k are either greatly reduced or negligible and (ii) the patterns of B-2 and B-3 are similar to that of the intact B fragment (Fig. 6).

The DNA sequence of the B fragment. Because of the significance of the B fragment in the genesis of the Ca3 pattern, the entire fragment was sequenced (Fig. 7). The B fragment is composed of 2,980 bp, and the overall G + C composition of the sequence is 33%. The B fragment contains three putative open reading frames greater than 35 amino acids in length between 831 and 1,034 bp, 2,742 and 2,891 bp and 1,455 and 1,589 bp. Respective lengths were 68, 50, and 45 amino acids. A search indicated no significant amino acid homology with any known protein in the Swiss Protein data base for any of the three open reading frames.

The B fragment contains several direct and inverted repeat sequences. At stringencies between 75 and 92% identity, 18 pairs of direct repeat sequences were identified. Repeats are distributed throughout the B fragment. The two longest direct repeats included a 28-bp repeat with 75% identity, starting at 540 and 1,165 bp of the B sequence (sequences a in Fig. 8), and a 27-bp repeat with 74% identity, starting at 1,559 and 2,850 bp of the B sequence (sequences b in Fig. 8). Four elements varying between 12 and 20 bases were repeated three to four times in the B fragment.

The B fragment also contains a number of inverted repeats. With the same stringency parameters as those for direct repeats, 27 pairs of inverted sequences were identified in the B fragment. The longest pair was 30 bp in length and was located at 1,638 and 2,889 bp along the B fragment (sequences c in Fig. 8). Four pairs of inverted repeats of 10 bp with 100% identity were identified in the B fragment at 1,622 and 1,657, 206 and 1,972, 1,600 and 2,028 and 421 and 2,811 bp. Two novel inverted repeats located at 265 and 1,657 bp of the B fragment were 11 bp long, exhibited 91% identity, and were both flanked at one end by the direct repeat CAT, a characteristic of some transposons.

Northern blot hybridization of the three major Ca3 fragments. The entire Ca3 probe and the three major fragments A, B, and C were individually used to probe Northern blots of whole-cell RNA of strain 3153A (Fig. 8). The entire Ca3 probe hybridized with RNAs of 2.8, 2.3, and 1.5 kb. The A fragment hybridized strongly with RNA at 1.5 kb and very weakly with RNA at 2.8 and 1.5 kb. The B fragment hybridized strongly with RNA at 2.8 and 2.3 kb and very weakly with RNA at 3.2, 1.9, and 1.5 kb. The C fragment hybridized strongly with RNA at 2.8 and 2.3 kb and very weakly with RNA at 1.5 kb.

DISCUSSION

The Ca3 probe was originally cloned from a partial *Eco*RI digest of strain 3153A cellular DNA (13). Complete *Eco*RI digestion of Ca3 divided the probe into five major fragments, which were estimated by comparison with standard DNA markers to be approximately 4.2 kb (fragment A), 3.2 kb (fragment B), 2.85 kb (fragment C), 0.77 kb (fragment D1), and 0.77 kb (fragment D2) in length, and two minor fragments, which were estimated to be approximately 0.38 kb (fragment E) and 0.30 kb (fragment F) in length. The sequence of fragment B provided the more accurate molecular size of 2.98 kb. We have ordered the major fragments as follows: C B D_1 A D_2. The Ca3 probe hybridized with five major chromosomal bands and one minichromosomal band of strain 3153A, separated by TAFE. Two chromosomes of strain 3153A (2 and 4 in the numbering system used in this study) did not hybridize significantly with the intact Ca3 probe. Therefore, these two chromosomes are either devoid of Ca3 sequences or contain very low copy numbers. Chromosomes hybridizing with Ca3 exhibited different levels of intensity, suggesting that they contain different copy numbers of either the entire probe or specific regions of the probe. Our data also suggest that Ca3 does not remain intact among the different repeats. If all repeats were identical, then digestion of whole-cell DNA should result in seven bands. In fact, most strains of *C. albicans* so far examined exhibit between 15 and 25 bands with intensities which do not decrease with decreasing molecular weight, demonstrating that the Ca3 sequence is not complete in some repeats. This conclusion is supported by the hybridization patterns of TAFE-separated chromosomes of strain 3153A probed with the major *Eco*RI fragments of Ca3. In no case did the relative hybridization intensities of a fragment compare exactly with those of the entire Ca3 probe. In the case of fragment A,

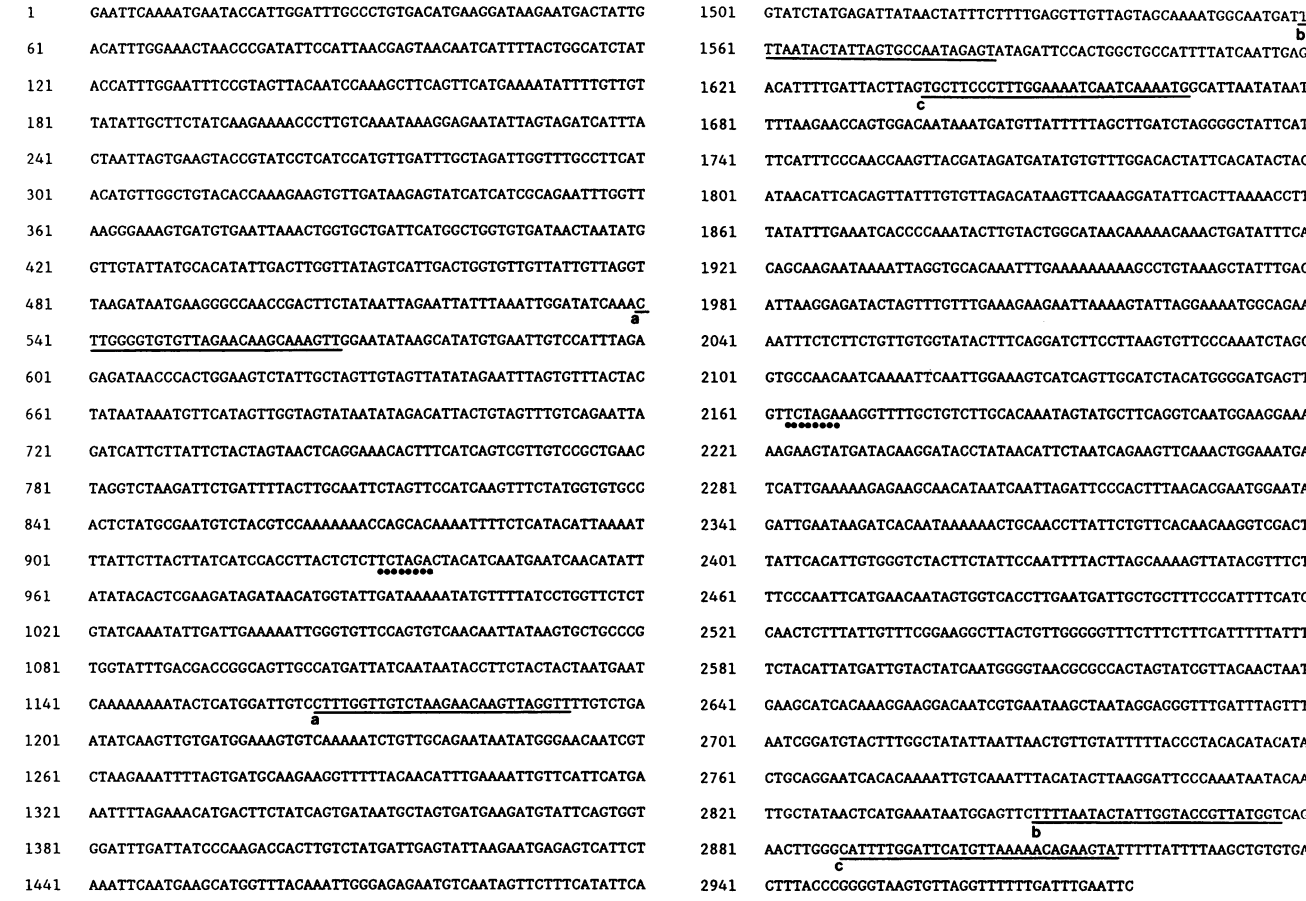

FIG. 7. Sequence of the *Eco*RI fragment B of the Ca3 probe. The sequence contains 2,980 bp. Direct repeat sequences a, at 540 and 1,165 bp (75% identity), and b, at 1,559 and 2,850 bp (74% identity), are underlined. Inverted repeat sequences c, at 1,638 and 2,889 bp (73% identity), are also underlined. The *Xba*I sites are underlined with dots.

there was intense hybridization with chromosome 7 and weak to negligible hybridization to the rest of the chromosomes and minichromosome, suggesting that A may be represented either exclusively on chromosome 7 or in high copy number on 7 and very low copy number on the other chromosomes. The hybridization pattern of A with *Eco*RI-digested DNA indicated a maximum of two band sizes in all strain patterns, suggesting allelic differences. Fragment B hybridized intensely with chromosomes 5 and 7, and fragment C hybridized intensely with 1, 3, and 6. These hybridization patterns suggest that different portions of the Ca3 probe are represented unequally on the chromosomes of strain 3153A.

One would expect all *Eco*RI fragments of the Ca3 probe to have molecular weight counterparts in the pattern of strain 3153A, from which the original Ca3 fragment was cloned. Since the molecular sizes of all fragments (except B) were initially estimated on gels by comparison with known molecular weight standards run in parallel, the estimates cannot be considered very accurate, and one can only approximate the band in the total pattern which correlates with the fragment sequence. In addition, fragments B and C generate hybridization patterns containing several bands of different intensities, indicating differences in redundancy. Since probe Ca3 represents only one of many genomic fragments containing B and/or C, it is not known if the *Eco*RI fragments B and C from Ca3 migrate with any of the intense bands representing redundant fragments of equal size.

Our results demonstrate that fragment B accounts for a significant portion of the Ca3 hybridization pattern. Subfractionation of the B fragment into three *Xba*I and hybridization of each subfragment to *Eco*RI-digested whole-cell DNA of strain 3153A resulted in similar patterns for fragment B and the *Xba*I fragments B-2 and B-3. However, an altered hybridization pattern was obtained for *Xba*I fragment B-1, in which hybridization with bands a, b, and k were dramatically reduced. These results suggest that B-2 and B-3 sequences, but not B-1 sequences, are present in most B repeats.

The three major *Eco*RI fragments of the Ca3 probe also generated different hybridization patterns between strains. Fragment A generated three alternative patterns which included band c alone, band e alone, or a combination of bands c and e. Fragment B, on the other hand, generated a far more complex pattern which varied between strains, but most band positions were relatively invariant and were either present or absent in test strains. The B patterns of all strains contained bands g, h, and i. However, the B patterns contained a, b, or a and b and either possessed or lacked band f. The B hybridization patterns of all but the unusual vaginal strains hp22vu and hp49vp (24) contained a band at e, a band slightly larger than e, or both. In addition, the patterns of strains hp36bt and hp36vw probed with B con-

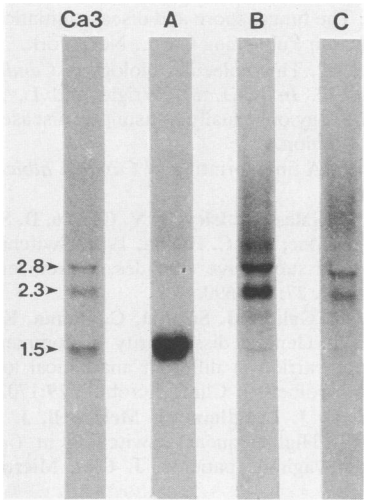

FIG. 8. Northern blots of whole-cell RNA of *C. albicans* 3153A probed with the intact probe Ca3, *Eco*RI fragment A, *Eco*RI fragment B, or *Eco*RI fragment C. The molecular sizes (in kilobases) of only the major hybridization bands are presented to the left.

tained a band slightly smaller than b, and strains hp22vu and hp49vp contained a unique high-molecular-weight band. The C fragment generated patterns with the most variability for a subset of low-intensity, high-molecular-weight bands. These bands have been demonstrated to represent the only differences in Ca3 hybridization patterns between pairs of highly related strains isolated from the oral and vaginal cavities of healthy women, and the differences were previously interpreted to represent genetic divergence between substrains separated in different body locations of the same individual (24).

Because the B fragment contributes more to the Ca3 pattern than the other fragments, it was the first to be sequenced. The sequence contains three open reading frames, but none are homologous to any known sequences in the EMBL nucleic acid data base or to proteins in the Swiss Protein data base. Since the Ca3 fragment is *C. albicans* specific, resulting in weak or negligible hybridization with related species (13, 25, 26), its rate of change must be relatively rapid, which may explain why homology to known genes was not discovered for the B fragment. The B fragment contains a number of direct and inverted repeat sequences. However, the presence of these repeat elements does not result in random rearrangements, since the positions of the majority of bands in the B hybridization pattern are limited in different strains (18).

We have found that there are three major transcripts which hybridize intensely to the Ca3 probe. The smallest of these is 1.5 kb and hybridizes almost exclusively with the A fragment. The two larger transcripts are 2.8 and 2.3 kb, and both hybridize with the B and C fragments but not with the A fragment. Since fragments B and C map as neighbors in the Ca3 probe and since they do not cross-hybridize, the 2.8- and 2.3-kb transcripts most likely represent transcripts of a continuous sequence containing B and C.

This initial characterization of the Ca3 probe has provided us with information on the contribution of *Eco*RI fragments of the probe to the Ca3 hybridization pattern. It has allowed us to identify sequences which are either invariant or which vary between strains. It has provided us with information on the variability between Ca3 sequences in the genome and their distribution on different chromosomes. It is clear that the Ca3 probe is not intact and equally distributed throughout the genome. However, it is also clear that there are not an infinite number of Ca3 hybridization patterns in different strains (18), suggesting that even though there is Ca3 variability within the genome, there must be some prescribed rules for rearrangement. Perhaps the most interesting sequences identified in this study are the high-molecular-weight *Eco*RI fragments which are highly variable between related strains (24) and which are homologous to the *Eco*RI C fragment. Changes in the positions of these bands will be weighted far less than changes in the less variable B bands in a complex computation of the similarity coefficient. We have therefore begun to characterize and sequence the *Eco*RI C fragment in the second phase of the characterization of the Ca3 probe.

ACKNOWLEDGMENTS

We are indebted to H. Vawter-Hugart for help in the genesis of dendrograms and scans of gels and to S. Woodward for photography.

This study was supported by Public Health Service grant AI23922 to D.R.S. and by a grant from the California University-wide AIDS Research Program and the John D. and Catherine T. MacArthur Foundation to N.A. S.H.M. was supported by a USCF Graduate Mentorship Award.

REFERENCES

1. **Bedell, G. W., and D. R. Soll.** 1979. Effects of low concentrations of zinc on the growth and dimorphism of *Candida albicans*: evidence for zinc-resistant and -sensitive pathways for mycelium formation. Infect. Immun. **26:**348–354.
2. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:**156–159.
3. **Church, G. M., and W. Gilbert.** 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:**1991–1995.
4. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the Vax. Nucleic Acids Res. **12:**387–395.
5. **Lee, K. L., H. R. Buckley, and C. C. Campbell.** 1975. An amino acid liquid synthetic medium for the development of mycelial and yeast forms of *Candida albicans*. Sabouraudia **13:**148–153.
6. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
7. **McEachern, M. J., and J. B. Hicks.** 1991. Dosage of the smallest chromosome affects both the yeast-hyphal transition and the white-opaque transition of *Candida albicans* WO-1. J. Bacteriol. **173:**7436–7442.
8. **Miyasaki, S. H., J. B. Hicks, D. Greenspan, I. Polacheck, L. A. McPhail, T. C. White, N. Agabian, and J. Greenspan.** 1992. The identification and tracking of *Candida albicans* isolates from oral lesions in HIV-seropositive individuals. J. Acquired Immune Defic. Syndr. **5:**1039–1046.
9. **Odds, F. C.** 1988. Candida and candidosis: a review and bibliography, 2nd ed. Bailliere Tindall, London.
10. **Queen, C., and L. J. Korn.** 1984. A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids Res. **12:**581–599.
11. **Rackwitz, H. R., G. Zehetner, A. M. Frischauf, and H. Lehrach.** 1984. Rapid restriction mapping of DNA cloned in lambda phage vectors. Gene **30:**195–200.
12. **Sadhu, C., D. Hoekstra, M. J. McEachern, S. I. Reed, and J. Hicks.** 1992. A G-protein subunit from asexual *Candida albicans* functions in the mating signal transduction pathway of *Saccharomyces cerevisiae* and is regulated by the α1-2 repressor. Mol. Cell. Biol. **12:**1977–1985.
13. **Sadhu, C. M., M. J. McEachern, E. P. Rustchenko-Bulgac, J. Schmid, D. R. Soll, and J. B. Hicks.** 1991. Telomeric and repeat

sequences in *Candida* yeasts and their use in strain identification. J. Bacteriol. **173:**842–850.
14. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
15. **Scherer, S., and D. A. Stevens.** 1987. Application of DNA typing methods to epidemiology and taxonomy of *Candida* species. J. Clin. Microbiol. **25:**675–679.
16. **Scherer, S., and D. A. Stevens.** 1988. A *Candida albicans* dispersed, repeated gene family and its epidemiological application. Proc. Natl. Acad. Sci. USA **85:**1452–1456.
17. **Schmid, J., F. C. Odds, M. J. Wiselka, K. G. Nicholson, and D. R. Soll.** 1992. Genetic similarity and maintenance of *Candida albicans* strains from a group of AIDS patients, demonstrated by DNA fingerprinting. J. Clin. Microbiol. **30:**935–941.
18. **Schmid, J., E. Voss, and D. R. Soll.** 1990. Computer-assisted methods for assessing strain relatedness in *Candida albicans* by fingerprinting with the moderately repetitive sequence Ca3. J. Clin. Microbiol. **28:**1236–1243.
19. **Sneath, P. H. A., and R. R. Sokal.** 1973. Numerical taxonomy. The principle and practice of numerical classification. W. H. Freeman & Co., San Francisco.
20. **Soll, D. R.** 1991. Current status of the molecular basis of *Candida* pathogenicity, p. 503–540. *In* G. T. Cole and H. C. Hock (ed.), The fungal spore and disease initiation in plants and animals. Plenum Publishing Corp., New York.
21. **Soll, D. R.** 1992. The molecular biology of *Candida* pathogenesis, p. 131–172. *In* D. J. M. Wright and L. Archard (ed.), Molecular biology of sexually transmitted diseases. Chapman & Hall, Ltd., London.
22. **Soll, D. R.** DNA fingerprinting of *Candida albicans*. J. Mycol. Med., in press.
23. **Soll, D. R., R. Galask, S. Isley, T. V. G. Rao, D. Stone, J. Hicks, J. Schmid, K. Mac, and C. Hanna.** 1989. Switching of *Candida albicans* during successive episodes of recurrent vaginitis. J. Clin. Microbiol. **27:**681–690.
24. **Soll, D. R., R. Galask, J. Schmid, C. Hanna, K. Mac, and B. Morrow.** 1991. Genetic dissimilarity of commensal strains of *Candida* spp. carried in different anatomical locations of the same healthy women. J. Clin. Microbiol. **29:**1702–1710.
25. **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.
26. **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.