A Candida albicans dispersed, repeated gene family and its epidemiologic applications
(gene conversion/mobile element/orthogonal field alternating gel electrophoresis/DNA polymorphism)

STEWART SCHEerer*† AND DAVID A. STEVEnS‡

*Department of Microbiology, University of Minnesota School of Medicine, Minneapolis, MN 55455; and ‡Division of Infectious Diseases, Department of Medicine, Santa Clara Valley Medical Center and Institute for Medical Research, San Jose and Stanford University School of Medicine, Stanford, CA 94305

Communicated by John Abelson, November 9, 1987

ABSTRACT Candida albicans causes a wide variety of infections but can readily be isolated from the skin and mucosa of healthy individuals. To enable high-resolution epidemiologic studies on this common pathogen, a species-specific DNA probe has been isolated from its genome. There are ~10 copies of the sequence dispersed among the chromosome-sized DNA molecules resolved by pulsed-field electrophoresis. New DNA polymorphisms in this gene family arise at high rates. As a consequence, this probe will readily distinguish strains from different patients in the same hospital and from various sites in individual patients. The DNA polymorphisms detected by using this probe are largely due to internal changes in members of the family rather than movement to new genomic locations. This suggests recombination or gene conversion rather than transposition as the mechanism producing the observed variation.

Epidemiologic studies of microorganisms rely on precise and reproducible methods for their identification. Many of the tests in current use measure phenotypes with the assumption that the underlying genetic differences that produce the phenotypic changes are the same in different isolates. With the advent of recombinant DNA technology it has become possible to replace many of these phenotypic tests with ones that directly examine the genomes of the organisms of interest (1–4). The highest resolution in studies of this type using molecular probes can be achieved by examining segments of the genome that are changing the most rapidly.

We have taken this approach in studies of the ubiquitous eukaryotic pathogen Candida albicans. The spectrum of disease caused by Candida ranges from superficial infections of the skin and mucosa to severe systemic disease in individuals with impaired immunity (5, 6). Recently, Candida has emerged as a common source of hospital-acquired infections. Because Candida can be readily isolated from the skin and mucosa of healthy individuals, the availability of precise methods for typing of strains is especially important in studies of transmission (7–9).

An initial step in DNA-based methods for typing Candida species and strains is direct examination of the fluorescence pattern obtained after restriction digests of total genomic DNA and electrophoresis. These DNA fingerprints distinguish Candida species and many C. albicans subtypes (10, 11). One drawback of this approach is that only a limited number of sites in the genome of a given species can be scored for differences, primarily the highly repeated ribosomal DNA and the mitochondrial DNA (10, 12). To examine many sites in the genome simultaneously, a dispersed, repeated DNA segment from the C. albicans genome has been isolated. This sequence family produces new DNA polymorphisms at rates that can be measured in the laboratory. Therefore, it should be possible not only to resolve closely related strains but to approximate the number of generations that clinical isolates have been separated. In addition, the availability of DNA probes for highly polymorphic loci dispersed throughout the genome should facilitate the development of a combined physical and genetic map of the organism.

High frequencies of DNA polymorphisms in populations are due to a number of mechanisms. Often, the polymorphisms are associated with movement of transposable elements (13–15). The gene family described here has many of the features common to a variety of eukaryotic transposable elements. These include moderate repetition and dispersion throughout the genome, production of new DNA polymorphisms at high rates, and species specificity. However, the DNA polymorphisms detected by this probe are generally due to changes of internal sequences in copies of the element. Mechanisms capable of producing the observed changes are presented in the Discussion.

MATERIALS AND METHODS

C. albicans strains 609 and 616 (lanes A and B of Fig. 1) and Candida tropicalis strain 615 (lane C of Fig. 1) are recent clinical isolates from the University of California Los Angeles Medical Center. Strain 609 is a blood isolate, strain 616 derives from an abdominal abscess, and strain 615 was cultured from urine. All cloned DNA segments described in this work are derived from strain 616. Strains 609 and 615 were used as the DNA probes to isolate the species-specific clones.

Growth of Candida strains and small-scale DNA isolations have been described (11). Larger-scale Candida DNA preparations suitable for molecular cloning were made as follows. Exponentially growing cells in rich medium were washed with 1 M sorbitol. Protoplasts were prepared by resuspending the cells 10-fold concentrated in 1 M sorbitol/50 mM potassium phosphate, pH 7.5/14 mM 2-mercaptoethanol with 100 μg of zymolyase 100T per ml (Miles) and incubating at 30°C for 30 min. The protoplasts were pelleted and resuspended in 50 mM Na₂EDTA/0.2% sodium dodecyl sulfate (NaDodSO₄) with 100 μg of proteinase K per ml. This was incubated at 50°C for 3 hr. The DNA was extracted three times with phenol/chloroform (1:1) and precipitated with 2 vol of ethanol. The DNA was resuspended in 10 mM Tris-HCl, pH 7.5/1/1 mM Na₂EDTA (TE buffer) with 10 μg of RNase A per ml. After overnight incubation at 4°C the DNA was precipitated with 2 vol of 2-propanol and resuspended in TE buffer.

Agarose gel electrophoresis and Southern transfers were performed by standard procedures (12). After transfer to nitrocellulose, hybridization was done in 50% formamide/0.9 M NaCl/50 mM sodium phosphate, pH 7.4/5 mM

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†To whom reprint requests should be addressed.
EDTA/0.3% NaDodSO4/100 µg of denatured salmon DNA per ml with 10⁶ dpm of nick-translated probe per ml (16). Pulsed-field electrophoresis was done in an orthogonal field alternating gel electrophoresis apparatus (17) with the following parameters: 125 V, 5-min pulse time, 11°C buffer, 72-hr run time. The buffer was 50 mM Tris borate/2.5 mM Na₂EDTA.

Libraries of EcoRI fragments from strain 616 in λ vector 1149 (18) were prepared as described by Davis et al. (16). The phage were packaged in vitro and plated on an hsdR hflA Escherichia coli strain to select recombinant phage. Plate lysates were prepared that represent a pool of ~750,000 independent recombinants. Approximately 85% of the phage, including the species-specific clones 23A, 24A, and 27A (Fig. 1), contained a single EcoRI insert.

A library of Sau3A partial-digestion products [15–20 kilobases (kb)] from strain 616 was prepared by using vector EMBL3 (18). Products of the in vitro packaging reaction were screened directly with 32P-labeled plasmid 701 (an internal EcoRI–HindIII fragment of clone 27A in pBR322). Among the isolates recovered from this screen were two phage that contained an EcoRI fragment identical to that in clone 27A.

RESULTS

Isolation of Species-Specific DNA Probes. One approach for the isolation of mobile DNA segments is the examination of closely related species for differences in their genomes (19). Given the small size of the Candida genome, it is possible to screen cloned segments in phage λ directly by using total genomic DNA labeled by nick-translation. The vector NM1149 was selected because its vigorous growth would result in enhanced signals in plaque filter hybridization and recombinant phage can be selected on an hflA E. coli strain to remove rejoined vector (18). Using this vector, a library of about 100 genome equivalents of C. albicans EcoRI fragments was constructed. Approximately 3000 recombinant phage from this library were plated at a density such that individual plaques were clearly resolved and duplicate plaque filters were prepared (8). DNA from C. albicans and C. tropicalis were labeled by nick-translation to 2 × 10⁶ dpm/µg and hybridized to the replica filters overnight at 37°C (see Materials and Methods).

After washing and autoradiography, all plaques were found to label with the C. albicans probe. Only plaques that labeled strongly with the C. albicans labeled with the C. tropicalis probe. Five plaques were identified that labeled strongly with the C. albicans (self) probe and weakly or not at all with the C. tropicalis probe. Phage from these plaques were plaque-purified, rescreened, plaque-purified again, and grown to prepare DNA. Labeled DNA from each phage was hybridized to identical Southern transfers of EcoRI digests of DNA from two unrelated C. albicans and one C. tropicalis isolate (Fig. 1). Two of these clones had been identified previously (unpublished results). They are likely to be a spacer segment of the rRNA-encoding DNA (rDNA) repeat that is not present in C. tropicalis DNA. A third clone (24A) was repeated highly in C. albicans, gave similar patterns with the two C. albicans strains, and was present at reduced copy number in C. tropicalis. The final two clones differed greatly in their hybridization patterns with the two C. albicans strains and did not hybridize at all with the C. tropicalis DNA.

One of this final pair (27A) was chosen for further study. Labeled DNA of this clone showed no hybridization to a variety of C. tropicalis and C. parapsilosis strains and distinguishable hybridization patterns to EcoRI digests of 30 C. albicans strains representing various phenotypic groups and subgroups (not shown).

FIG. 1. Hybridization of cloned segments to C. albicans and C. tropicalis DNA. Clones representative of the three types described in the text are shown. Each filter has EcoRI digests of two C. albicans DNAs (lanes A and B) and one C. tropicalis DNA (lane C). Hybridization was overnight at 37°C. The first panel shows a dispersed, repeated clone (24A) present at reduced copy number in C. tropicalis. In panel II, a clone (23A) likely to be rRNA-encoding DNA (rDNA) spacer is shown. Weak hybridization to other rDNA fragments and to partial-digestion products from the highly repeated rDNA cluster is seen. The final panel is the species-specific probe (27A) used in the epidemiologic studies.

To demonstrate that the hybridizing fragments are dispersed throughout the C. albicans genome, a subclone of 27A in pBR322 (701 in Materials and Methods) was hybridized to the chromosome-sized DNA molecules of C. albicans (Fig. 2) separated by pulsed-field electrophoresis (17). All but one of the molecules resolved by this technique were

FIG. 2. Sequences homologous to the mobile DNA probe are found on the molecules resolved by pulsed-field electrophoresis in C. albicans. Candida DNA was prepared by the method used for Saccharomyces DNA (20, 21). The fluorescent pattern is shown at the left in panel I with Saccharomyces cerevisiae size standards. The intensity of the fluorescence of the bands does not appear to be proportional to molecular weight. Splitting of homologous pairs and unresolved chromosomes are the likely explanations of this result. In the hybridization pattern in panel II, all size classes of Candida DNA molecules but one are labeled. None of the Saccharomyces chromosomes that remain on the gel label with the probe. No smaller Candida molecules have been seen in shorter runs. The smallest Candida chromosomes are estimated to be 1000 kb. The largest chromosomes are significantly larger than Saccharomyces chromosome XII.
labeled with the probe. Given the absence of labeling of one of the bands in the fluorescence pattern, it is unlikely that this family of repeated sequences is involved in chromosome structure (for example, the telomere repeats).

**Clone 27A Will Detect DNA Polymorphisms That Arise in the Laboratory.** To estimate the rate at which the DNA polymorphisms associated with this family of sequences arise, two experiments were done. One was a serial passage four different *C. albicans* strains 30 times on rich medium (≈300 generations) and examine six single colonies obtained from the final passage of each strain. DNA from 1 of these 24 colonies showed a change in the Southern pattern. Other eukaryotic mobile elements produce DNA sequence polymorphisms at detectable rates (14, 15).

In a second experiment, 10 spontaneous *C. albicans* 5-fluorocytosine-resistant mutants (22) were examined in similar fashion. One of these mutants showed a change in its hybridization pattern with this probe (Fig. 3). It is unknown whether the changes in hybridization pattern of this family are due to the mutation recovered by selection; however, the result underscores the sensitivity this type of probe can give to the detection of strain differences. Similar changes are seen when spontaneous cadmium-resistant mutants are examined by using this probe (S. Grindle and S.S., unpublished observations). The loss of a hybridizing fragment need not be due to a transposition event. Resistance to 5-fluorocytosine is generally due to a recessive mutation (23). The loss could arise from the elimination (by mitotic recombination) or of a heterozygous copy of the element linked to the locus conferring drug resistance when the new mutation was made homozygous.

**Use of Probe 27A to Resolve Epidemiologically Related Strains.** Isolates from different body sites from several patients were examined by using the mobile DNA probe. In five of six individuals, some differences could be detected in the different body-site isolates. One of the five patients appeared to be infected by two unrelated *C. albicans*, as judged by DNA fingerprints (11). The four isolates from the same patient with both DNA typing techniques do not agree with those from phenotyping meth-
ods (7, 8). Some similarity in hybridization pattern is seen among strains with clearly different fingerprint patterns.

Characterization of the DNA Sequence Polymorphisms. To determine the nature of the DNA polymorphisms revealed in the experiments described above, additional physical characterization of the gene family was performed. A library of Sau3A partial-digestion products of C. albicans cloned in vector EMBL3 was screened with a plasmid subclone of 27A. Among the positive clones, two overlapping clones were identified that contained an EcoRI fragment identical in size and restriction map to 27A (not shown). Fig. 6 shows a map of this region of the genome. Hybridization of radioactive total genomic DNA to Southern blots of restriction fragments of these clones revealed that most of the cloned region was repeated DNA. Additional copies of the element were obtained in the same screen. The terminal fragments from the cloned region surrounding 27A did not label a different clone recovered from the EMBL3 library that extends a corresponding distance from the portions of the clone homologous to 27A. This suggests that the ends of the repeated segment lie within the cloned region shown in Fig. 6.

Additional mapping studies (not shown) indicate that one end of the repeated region lies within the boundaries of probe 1. The EcoRI fragments to the left of probe 1 label single bands in Southern blots to EcoRI-digested genomic DNA. A HindIII fragment spanning the left end of probe 1 labels many bands in the same digest, indicating that the repeated region ends near the left end of probe 1. Similar mapping studies place probe 3 near the right end but within the repeated segment. The size of the repeated region is at least 12 kb. The terminal probes did not label other restriction fragments from the cloned region, indicating an absence of long terminal repeats in the cloned region.

Clone 24A (Fig. 1) is strongly homologous to probe 1 of Fig. 6. It has a different restriction map than clones 137 and 146 and probably derives from another copy of the element. Hybridization of clone 24A to other Candida species might derive from the DNA adjacent to the element in that probe.

Hybridization of probes from terminal regions of the element to DNA from epidemiologically related strains is not effective in revealing DNA polymorphisms. Three pairs of isolates, each pair having indistinguishable DNA fingerprint, were examined. Southern blots of EcoRI-digested DNA from all six strains were hybridized with each of the three probes shown in Fig. 6. Strains with different DNA fingerprints often will be distinguished by using probes 1 and 3. Strains that cannot be resolved by either DNA fingerprinting or the end probes will often reveal many differences with probe 2. Results obtained with one pair are shown in Fig. 7. Differences in band intensity might be due to changes in the number of copies at particular loci on homologous chromosomes or substitution of sequences with less homology to the probe. Larger copies of the element than those described here might exist and substitute larger regions of nonhomologous DNA. If the changes in band intensity in panel III of Fig. 7 were due to deletions, it would imply a hot spot for DNA rearrangements internal to the element. If all copies of the element are similar in size and organization, only the largest one or two labeled fragments shown in Figs. 1 and 3–5 could extend into the flanking low-copy DNA.

**DISCUSSION**

We have isolated a repeated DNA segment that provides a species-specific probe for C. albicans and will distinguish all but the most closely related strains. DNA polymorphisms associated with the gene family arise at rates that can be measured in the laboratory. The polymorphisms are primarily internal changes in the gene family and not due to movement to new genomic locations.

The great sensitivity that highly mobile DNA probes can give to epidemiologic studies may necessitate some change in the notion of the identity of strains. Clearly, the strains examined in this study are changing at measurable rates. A more useful measure may be an estimate for the number of generations that separate two strains. It should be possible to examine the spread of Candida infections within an individual by techniques analogous to those used to study development with genetic mosaics. The rates of change observed by using probe 27A are sufficiently low that reversion to prior configurations would be unlikely. A number of questions in Candida epidemiology will be open to more detailed investigation by this approach. These include recurrent vaginitis and the association of particular types of infections with particular strains. Comparison must be made of the rates at which DNA polymorphisms arise in animals and the rates at which they arise in laboratory media. Clearly, different probes vary in their efficacy in identifying polymorphisms in the genome.

The techniques used here to isolate the mobile gene family can be extended with little modification to screen for C. albicans single-copy DNA segments that are absent in other C. albicans strains. Additional copies of the gene family described here might exist that are not detected with internal probes such as 27A. As it is not yet clear how much of the observed polymorphism is due to heterozygosity, it is uncertain whether the end probes are detecting the same number of copies of the element as the central probe.
Continuing structural characterization of the gene family will resolve these issues. Analysis of the transcription of the family and its DNA sequence should provide clues to its biological role.

Examination of cloned fragments from corresponding genomic locations in different strains indicates the presence of different copies of the element at the site rather than the existence of empty sites (S.S. and S. Grindle, unpublished observations). This observation, combined with the results presented here, points to some mechanism other than transposition as the origin of the rapid sequence alterations.

Other mechanisms that might have produced the observed changes include mutational hot spots, recombination or gene conversion, and genetic exchange between strains. Movement of repeated DNA sequences to different chromosomes by gene conversion has been observed in Saccharomyces for transposable and nontransposable sequences (24, 25). Detailed examination of all of the copies of the family in closely related strains should permit the determination of which mechanisms are operating. If a gene conversion mechanism is at work, it should be possible to identify the donor sequences for the recombinational events.

We have detected high frequencies of new DNA polymorphisms with unselected passage and with spontaneous mutations. The frequency at which polymorphisms are seen when single colonies of spontaneous mutants are examined is much greater than that seen in unselected serial passage. It is not yet known which of these two frequencies more closely reflects the in vivo rate. As much of the polymorphism observed with this gene family is due to internal changes and the same frequency is seen with two different drug selections, it is unlikely that the polymorphisms are in fact the selected mutations. More likely would be that the selective agents are inducers of recombination or that the observed changes are linked to the newly selected mutations.

C. albicans is known to have at least two morphologic switching systems that result in high frequencies of spontaneous phenotypic changes (26, 27). No correlation between changes in this mobile gene family and the known phenotypic systems has been seen. The association of spontaneous mutations with either movement of this gene family to new sites or the production of linked DNA polymorphisms would provide a major tool in the development of the Candida genetic system.

J. Goodman was the source of Candida strains used to initiate this work. S. Grindle provided technical assistance. We thank R. W. Davis for advice on the separation of large DNA molecules and A. T. Haase for helpful suggestions on this manuscript. Tim Leonard and Colleen O’Neill provided assistance on the preparation of graphics and text. This work was supported by funds from the University of Minnesota and Grant AI23850 from the National Institutes of Health.