Microevolutionary changes in *Candida albicans* identified by the complex Ca3 fingerprinting probe involve insertions and deletions of the full-length repetitive sequence RPS at specific genomic sites

Claude Pujol, Sophie Joly, Bridgid Nolan, Thyagarajan Srikantha and David R. Soll

Author for correspondence: David R. Soll. Tel: +1 319 335 1117. Fax: +1 319 335 2772.
e-mail: david-soll@uiowa.edu

Department of Biological Sciences, 440 BB, University of Iowa, Iowa City, IA 52242, USA

The 11 kb complex DNA fingerprinting probe Ca3 is effective both in cluster analyses of *Candida albicans* isolates and in identifying microevolutionary changes in the size of hypervariable genomic fragments. A 2.6 kb *EcoRI* fragment of Ca3, the C fragment, retains the capacity to identify these microevolutionary changes, and when the C fragment is cleaved with *SacI*, the capacity is retained exclusively by a 1 kb subfragment, C1, which contains a partial RPS repeat element. The microevolutionary changes identified by Ca3, therefore, may involve reorganization of RPS elements dispersed throughout the genome. To test this possibility, hypervariable fragments from several strains of *C. albicans* were sequenced and compared. The results demonstrate that the microevolutionary changes identified by Ca3 are due to the insertion and deletion of full-length tandem RPS elements at specific genomic sites dispersed throughout the *C. albicans* genome. The RPS elements at these dispersed sites are bordered by the same upstream and downstream sequences. The frequency of recombination was estimated to be one recombination per 1000 cell divisions by following RPS reorganization *in vitro*. The results are inconsistent with unequal recombination between homologous or heterologous chromosomes, but consistent with intrachromosomal recombination. Two alternative models of intrachromosomal recombination are proposed: unequal sister-chromatid exchange and slipped misalignment at the replication fork.

Keywords: *Candida albicans*, RPS repetitive units, microevolution, Ca3 fingerprinting

INTRODUCTION

The complex *Candida albicans*-specific DNA fingerprinting probe Ca3 (Soll *et al.*, 1987; Sadhu *et al.*, 1991; Anderson *et al.*, 1993) has been used in a broad range of epidemiological studies (Soll *et al.*, 1987, 1989, 1991; Schmid *et al.*, 1990, 1993, 1995; Odds *et al.*, 1990; Soll, 1993; Hellstein *et al.*, 1993; Schroepel *et al.*,

1994; Lockhart *et al.*, 1995, 1996; Kleinegger *et al.*, 1996; Pfaller *et al.*, 1998; Marco *et al.*, 1999; White *et al.*, 1997; Lockhart *et al.*, 1999). Its effectiveness has been verified in a comparison with multilocus enzyme electrophoresis and random amplified polymorphic DNA analysis (Pujol *et al.*, 1997). When used to probe Southern blots of *EcoRI*-digested DNA, Ca3 identifies over 20 bands that include monomorphic, moderately variable and hypervariable bands (Schmid *et al.*, 1990). While the entire pattern has been used to assess the relatedness of isolates, the subset of hypervariable bands has been used to monitor the microevolution of strains at sites of infection or carriage (Soll *et al.*, 1991; Schmid

The GenBank accession numbers for the sequences reported in this paper are AF121119, AF121120, AF121121, AF121122, AF121123, AF121124 and AF121125.

et al., 1993; Schroepel *et al.*, 1994; Lockhart *et al.*, 1995, 1996).

The Ca3 probe is composed of seven *EcoRI* fragments: A, B, C, D1, D2, E and F (Anderson *et al.*, 1993). The C fragment alone identifies hypervariable sequences in Southern blots of *EcoRI*-digested genomic DNA (Anderson *et al.*, 1993), and for that reason the C fragment has been used effectively to monitor microevolution in colonizing strains (Schroepel *et al.*, 1994; Lockhart *et al.*, 1995, 1996). Recently, we demonstrated (Lockhart *et al.*, 1995) that the C fragment contains a portion of the *C. albicans* repetitive element RPS (Iwaguchi *et al.*, 1992). RPS elements are present on seven of the eight *C. albicans* chromosomes (Iwaguchi *et al.*, 1992; Chindamporn *et al.*, 1995), and those RPS elements sequenced to date contain three to four tandem repeats, each approximately 172 bp in length, referred to as the 'alt' element (Iwaguchi *et al.*, 1992; Chibana *et al.*, 1994). The 29 bp at the end of each alt element represent a common palindromic sequence referred to as COM29. Because of structural similarities between COM29 and both the λ attachment site and the bacterial DNA inversion/cross-over sequence (Ehrlich, 1989), Iwaguchi *et al.* (1992) have suggested that COM29 sequences may represent specific recombination sites in the *C. albicans* genome. The possibility therefore exists that the microevolutionary changes that are identified in clonal populations by the C1 fragment of the Ca3 probe (Lockhart *et al.*, 1995) are due to genomic reorganizations involving RPS elements. Southern analysis and DNA sequencing experiments described here demonstrate that high-frequency changes in the size of genomic sequences that hybridize to the C fragment of Ca3 involve the insertion and deletion of full-length RPS elements, providing the first estimate of the rate of RPS reorganization, and lead to two alternative models for intrachromosomal RPS recombination.

METHODS

***C. albicans* isolates.** Yeast isolates were collected in two family practice clinics in Ann Arbor, MI, USA, from women suffering from vaginal infections (Lockhart *et al.*, 1996). Samples were cultured initially on Sabouraud's agar and the species identified by sugar-assimilation patterns using the commercially available API 20C kit (Analytab Products). Isolates identified as *C. albicans* were transferred to brain-heart-infusion agar slants and fingerprinted. Recurrence was defined as two or more infections in a six-month period. Each recurrence patient (RP) was distinguished by a number (e.g. RP39), and each isolate from that patient was distinguished by a second number. Therefore isolates 1 and 2 from patient RP39 would be identified as RP39.1 and RP39.2, respectively.

DNA fingerprinting and Southern-blot analysis. Cells from stored agar slants were plated on YPD agar and harvested after 2 d. DNA was extracted and then digested with one or more selected restriction enzymes following the recommendations of the manufacturer (Promega). Digested DNA (3 μ g per lane) was electrophoresed through a 0.6% (w/v) or 0.8% (w/v) agarose gel overnight at 35 V. The gel was stained with ethidium bromide to assess loading and separation, then the DNA was transferred by capillary blotting to a nylon Hybond-

N+ membrane (Amersham). For sequential hybridization, the membrane was first prehybridized for 7 h at 65 °C with 10 mg denatured calf thymus DNA ml⁻¹ and then hybridized overnight at 65 °C with the first random-primer-labelled ([³²P]dCTP) probe. These hybridization steps were performed in a solution of 5 $\times$ SSPE (1 $\times$ SSPE contains 10 mM NaH₂PO₄, pH 7.5, 10 mM EDTA and 0.18 M NaCl) containing 5% (w/v) dextran sulfate and 0.3% (w/v) SDS. The membrane was washed at 45 °C with a solution of 2 $\times$ SSPE containing 0.2% (w/v) SDS and exposed to XAR-S film (Eastman Kodak) with a Cronex Lightning-Plus intensifying screen (Du Pont). The blot was then stripped of the first probe by incubating it in 0.4 M NaOH for 30 min at 45 °C, then incubated for 15 min in a solution of 1 $\times$ SSPE containing 0.1% (w/v) SDS and 0.2 M Tris/HCl, pH 7.5. The blot was then hybridized with the second random primer-labelled probe and again exposed to XAR-S film. This procedure was repeated for subsequent hybridizations.

Cloning hypervariable band fragments. Subgenomic libraries were constructed for *C. albicans* isolates RP5.1, RP10.3, RP39.1 and RP39.4. To accomplish this, total genomic DNA of each strain was digested with *EcoRI* and fractionated in a sucrose gradient. Fractions were analysed by Southern blot hybridization with the C1 subfragment of the Ca3 probe (Lockhart *et al.*, 1995), and those containing hypervariable band fragments were selected. Subgenomic libraries were constructed from the selected fractions in phage λ EMBL3 (Promega) between *EcoRI* sites, according to established protocols (Sambrook *et al.*, 1989). Libraries were plated at a density of 10⁴ recombinant phage per plate and screened with the C1 probe (Church & Gilbert, 1984). The *EcoRI* fragments of interest were then subcloned into the pGEM-7Zf(+) plasmid vector (Promega). When cloning fragments into *E. coli* JM109, we found that the larger the insert, the less stable the clone. Therefore, the integrity of each clone was tested by Southern analysis after each passage in *E. coli*. Cloned fragments and the genomic DNA of the isolate used to obtain each clone were digested with *EcoRI*, and Southern blots of these preparations were probed with C1. This provided a direct comparison between each clone and the original genomic fragment. In cases in which the clone showed evidence of recombination, the preparation was abandoned and prepared again.

Nucleotide sequence of hypervariable band fragments. Plasmid DNA was isolated using a CsCl-ethidium bromide protocol (Sambrook *et al.*, 1989). Prior to sequencing, each clone was compared to the original fragment by Southern analysis to ensure it had maintained its integrity. The nucleotide sequence of each plasmid insert was determined in both directions with the ABI model 373A Auto Sequence system (Perkin Elmer/Applied Biosystems) using the PCR cycle-sequencing protocol and fluorescent dye terminator dideoxynucleotides (Perkin Elmer/Applied Biosystems). Homology and alignment of nucleotide sequences to gene databases were performed with MacDNASIS Pro v3.6 software (Hitachi Software Engineering). The 5' and 3' ends of each clone were sequenced using customized synthetic primers. Both ends of the clones were sequenced until a region located in the centre of the fragments and displaying RPS sequence was reached. To sequence the RPS-containing regions of clones RP39.4–9 kb, RP39.1–13.7 kb and RP39.1–6.5 kb, a sequential series of deletions were prepared with the Erase-a-Base System kit (Promega). H3, a *HindIII*/*EcoRI* subfragment of RP39.1–13.7 kb, and RPS39, a *Csp45I* subfragment of RP39.1–13.7 kb, were both subcloned in the pGEM-7Zf(+) plasmid vector for further analysis.

RESULTS

Microevolutionary changes are identified by the C1 subfragment of Ca3

The Ca3 probe is approximately 11 kb in length (Sadhu *et al.*, 1991) and is composed of seven *Eco*RI fragments (Fig. 1a; Anderson *et al.*, 1993). Although the major portion of the Ca3 pattern is constant between isolates

from a colonizing strain, hypervariable high-molecular-mass fragments are identified by Ca3 that can differ, and these changes have been used to assess microevolution within a strain (Schroeppel *et al.*, 1994; Lockhart *et al.*, 1995, 1996). Examples of the microevolutionary changes discriminated by Ca3 are presented in Fig. 2(a) for collections of infecting isolates from two recurrent vaginitis patients, RP39 and RP5. The C fragment of

Fig. 1. Physical map showing the relationship between the Ca3 probe (a) (Soll *et al.*, 1987; Sadhu *et al.*, 1991), the hypervariable fragments analysed in the present study (b), HOK (c) (Chindamporn *et al.*, 1998), and the RPS cluster (Iwaguchi *et al.*, 1992). Known DNA sequences are represented by boxes with thick borders. Relative positions have been assigned by DNA homology. The physical map of the Ca3 probe has been adapted from Anderson *et al.* (1993) and Lockhart *et al.* (1995). The RPS region of the hypervariable fragments analysed here contained different numbers of full-length RPS tandem repeats, demarcated by vertical dashed lines. E, *Eco*RI; Sm, *Sma*; Sa, *Sac*; N, *Nsi*; H, *Hind*III; S, *Sal*. The model for hypervariable fragments (b) represents the consensus structure observed for fragments RP5.1–18.2 kb, RP5.1–7.5 kb, RP10.3–18.2 kb, RP39.1–13.7 kb, and RP39.4–9 kb. The vertically dashed *Eco*RI site located on the right site of the RPS region of the hypervariable fragments was observed only for clone RP5.1–7.5 kb and represented the 3' extremity of that clone. The remaining cloned hypervariable fragments possessed identical 3' *Eco*RI sites as represented in the figure.

Fig. 2. Southern-blot hybridization patterns of *Eco*RI-digested DNA of isolates from patients RP39 and RP45 probed with Ca3 (a), the C1 fragment of Ca3 (b) and RPS39 (c). DNA from reference strain 3153A was run in the first lane of the gel for normalization and unwarping. Molecular sizes (kb) are shown to the left of the hybridization patterns. White crosses indicate hypervariable bands that show differences among isolates from two recurrent vaginitis patients, RP39 and RP45.

Fig. 3. Demonstration that the C1 subfragment, but not the C2 subfragment, of the C fragment of Ca3 identifies the hypervariable high-molecular-mass bands of the Ca3 pattern. *SacI* was shown to cleave the *EcoRI* C fragment of the Ca3 probe into subfragments C1 and C2. Total genomic DNA of isolates from three patients, RP5, RP10 and RP39, were digested with either *EcoRI* (E) or *EcoRI/SacI* (+S) and probed with either C1 (a, b, c) or C2 (d, e, f). Molecular sizes (kb) are shown to the left of each panel.

the Ca3 probe (Fig. 1a) exclusively identifies the high-molecular-mass hypervariable fragments (Anderson *et al.*, 1993). The C fragment can be reduced further to two subfragments, C1 and C2, by *SacI* digestion (Lockhart *et al.*, 1995) (Fig. 1a). The C1 fragment retains the capacity to identify the hypervariable changes identified by the C fragment of the Ca3 probe (Lockhart *et al.*, 1995) (Fig. 2b). In the experimental results presented in Fig. 3, it is demonstrated that the C2 portion of the C fragment does not retain this capacity. If microevolutionary

changes are restricted to C1-containing fragments, then Southern blots of DNA digested either with *EcoRI* alone or with *EcoRI* plus *SacI* should both show differences between isolates from a single patient when probed with C1, but only Southern blots of DNA digested with *EcoRI* should show differences when probed with C2, since *SacI* removes C1 sequences from C2 sequences. This was exactly the result obtained in each of the three sets of isolates (Fig. 3), demonstrating that the microevolutionary changes identified by the Ca3 probe were

due exclusively to the reorganization of genomic sequences identified by the C1 subfragment.

The sequence of the C1 subfragment revealed that it contained a partial repetitive RPS element (Lockhart *et al.*, 1995). When the Southern blot that had been probed with Ca3 and fragment C1 in Fig. 2(a) and (b), respectively, was reprobed with the cloned RPS element RPS39 (Fig. 2c), all of the hypervariable fragments identified by Ca3 and C1 were also identified by RPS39. This result suggested that the microevolutionary changes identified by Ca3, the C fragment and the C1 subfragment involved the reorganization of RPS elements in hypervariable genomic fragments.

Hypervariable genomic fragments contain similar non-RPS 3' and 5' ends

The results obtained from the experiment described in Fig. 3 suggested not only that the hypervariability of high-molecular-mass fragments was due to the reorganization of sequences that hybridize to the RPS-containing C1 subfragment, but also that these hypervariable genomic fragments contained similar upstream C2 regions. To assess directly sequence homologies between hypervariable genomic fragments, several were cloned and sequenced from the 5' and 3' ends until an RPS sequence was penetrated. Those cloned included fragment RP5.1–18.2 kb and RP5.1–7.5 kb (both from isolate RP5.1), fragment RP10.3–18.2 kb (from isolate RP10.3), fragment RP39.1–13.7 kb (from isolate RP39.1) and fragment RP39.4–9 kb (from isolate RP39.4). The 3' terminal 2.74 kb sequences of all clones except RP5.1–7.5 kb were 100% identical. RP5.1–7.5 kb was truncated at the 3' end due to a transversion from A to T. This resulted in the genesis of an *EcoRI* restriction site unique to RP5.1–7.5 kb, 864 bp upstream from the consensus 3' *EcoRI* site shared by the four other clones (noted by a vertical dashed E site in Fig. 1b).

The 1.72 kb sequences upstream of the RPS sequence of the five clones were also highly homologous. In addition, they were highly homologous to the region upstream of the RPS sequence in Ca3 that included a portion of the C1 sequence (Lockhart *et al.*, 1995) and the C2 sequence (sequenced here). The average homology between the five clones and the homologous Ca3 sequence was 91.5%. The sequence of the C1 portion of this region has been published by Lockhart *et al.* (1995) and the C1 and C2 portions by Chindamporn *et al.* (1998) in a sequence analysis of the HOK fragment that overlaps a portion of the Ca3 probe. Physical maps comparing the Ca3 probe, the hypervariable fragments and the HOK fragment (Chindamporn *et al.*, 1998), which contains a partial RPS sequence similar to the one in C, are presented in Fig. 1.

Differences in the size of hypervariable fragments are due almost exclusively to insertions and deletions of full-size RPS units

Since the 5' and 3' ends of the five sequenced hypervariable genomic fragments were similar in size and

highly homologous, the differences in the total lengths of these fragments had to involve sequences within the RPS borders. To investigate the molecular basis of size variation, variable fragments that differed in size in two isolates of the same strain were cloned, sequenced and compared. The Ca3, C1 and RPS39 hybridization patterns of the two selected isolates from patient RP39 differed by a 13.7 kb band unique to the RP 39.1 pattern and a 9 kb band unique to the RP39.4 pattern (Fig. 2), suggesting that the two bands represented altered forms of the same genomic site. This interpretation was supported by the observation that the 5' and 3' terminal sequences of the two clones were identical. Clone RP39.4–9 kb was 8406 bp in length and contained one full-length RPS unit. Clone 39.1–13.7 kb was 12792 bp in length and contained three full-length tandem RPS units. The only difference, therefore, between RP39.4–9 kb and RP39.1–13.7 kb was the number of full-length RPS units in the RPS cluster.

The three tandem RPS units in RP39.1–13.7 kb defined by four *PstI* recognition sites (RPS 39-1, RPS 39-2 and RPS 39-3) were not identical (Fig. 4a). They differed by single base differences and by the number of alt sequences, small tandem repeat units of approximately 172 bp that have previously been demonstrated to differ in number between RPS units (Chibana *et al.*, 1994). While RPS 39-1 and RPS 39-2 contained the four alt sequences a, b, c and d, RPS 39-3 contained alt sequences a, c and d (Fig. 4a). When compared to published RPS sequences (Chibana *et al.*, 1994), RPS 39-1 and RPS 39-2 exhibited the highest homology with RPS 116 (99.5 and 99.1%, respectively), and RPS 39-3 exhibited the highest homology with RPS 620 (99.1%).

As noted previously, the larger the genomic fragment, the less stable it was when cloned into *E. coli* JM109. The instability was characterized by decreases in size that were compatible with the loss of one or more RPS units. To examine the validity of this interpretation, the 6.5 kb recombinant clone RP39.1–6.5 kb derived from RP39.1–13.7 kb after passage in *E. coli* was sequenced *in toto* and compared to the parental clone. The recombinant clone contained 6182 bp and was identical to RP39.1–13.7 kb between bp 1 and 2998, and between bp 9608 and 12792. The genesis of RP39.1–6.5 kb was explained by a recombinational event between the most 5' and the most 3' alt c sequences in the parental clone. The putative recombinatorial event is shown in Fig. 4 (b), where the predicted recombination sites are noted by arrowheads. This result suggested that alt clusters containing COM29 palindromic sequences represent hotspots for recombination inside RPS repeats.

RPS-containing loci dispersed throughout the *C. albicans* genome share similar upstream and downstream sequences

Chindamporn *et al.* (1998) presented evidence suggesting that RPS sequences were located between similar 5' and 3' sequences. In addition, we demonstrated that the RPS clusters in the hypervariable clones that were sequenced

Fig. 4. Sequence comparison of alt clusters. Comparison of alt clusters from three RPS tandem repeats, RPS39-1, RPS39-2 and RPS39-3, defined by four *Pst*I recognition sites from clone RP39.1-13.7 kb (a). Sequences in bold represent COM29 subrepeats. The start of each of the four alt elements a, b, c and d is represented by an arrow. The alt b sequence is not present in RPS 39-3 and the deletion is represented by a dashed line. Identical bases are represented by dots. In (b), the most 5' alt cluster of clone RP39.1-13.7 kb is compared to the alt cluster of clone RP39.1-6.5 kb and to the most 3' alt cluster of clone RP39.1-13.7 kb. Clone RP39.1-6.5 kb is a recombinant of clone RP39.1-13.7 kb generated in *E. coli* which has lost three RPS repeats. The arrowheads at base positions 490 and 491 indicate the predicted recombination site between the most 5' and 3' alt clusters of clone RP39.1-13.7 kb resulting in RP39.1-6.5 kb. Note that the recombination was unequal and resulted in the duplication of 54 bp in the newly created alt c sequence of clone RP39.1-6.5 kb.

were located between similar or identical 5' and 3' sequences. However, there are a number of lower-molecular-mass bands that contain RPS elements but that are not hypervariable within a strain. The following experiment was performed to test whether all RPS units throughout the genome are located between similar 3' and 5' sequences. A unique *Nsi*I restriction site was identified in the C2 region and a unique *Sal*I restriction site was identified in the non-RPS 3' region bordering RPS elements in the cloned and sequenced hypervariable fragments. *Nsi*I/*Sal*I-digested DNA preparations of 13 unrelated *C. albicans* isolates were probed with C2, the 5' end that borders RPS elements in hypervariable fragments, H3, the 3' end that borders RPS elements in hypervariable fragments, and the RPS element RPS39 (Figs 1 and 5). Strain relatedness was assessed by Ca3 fingerprinting (data not shown). The genetic diversity observed for this collection was comparable to that

observed for a previously studied collection of unrelated isolates (Pujol *et al.*, 1997). If all partial, single and tandem-repeat RPS elements dispersed throughout the genome are flanked by the same non-RPS 5' and 3' sequences, all bands identified by the RPS39 probe should also be identified by the C2 and H3 probes. This was the case for 12 out of the 13 tested isolates (Fig. 5). The exception was isolate RP41.1 (Fig. 5). In the Southern blot of this isolate, a 6.7 kb fragment hybridized to the RPS39 probe (Fig. 5b) and to the C2 probe (Fig. 5a) but not to the H3 probe (Fig. 5c). Conversely, a 6.1 kb fragment hybridized to the RPS39 probe (Fig. 5b) and to the H3 probe (Fig. 5c), but not to the C2 probe (Fig. 5a). This result is explained by the presence of a recognition site for either *Nsi*I or *Sal*I in an RPS unit of RP41.1 that separates a 12.8 kb fragment into a 6.1 and a 6.7 kb fragment. Together, these results demonstrate that in all cases, RPS sequences on different

Fig. 5. RPS-containing loci share similar upstream and downstream sequences. Southern blots of DNA from thirteen unrelated *C. albicans* isolates digested with *NsiI/SalI* were hybridized with the C2 region (a), which is upstream of the RPS sequences, RPS39 (b) and the H3 region (c), which is downstream of the RPS sequences. Boxes indicate the hybridization regions in relation to the *NsiI* restriction site in the 5' upstream region and a *SalI* restriction site in the 3' downstream region of the 8.4 kb cloned and sequenced hypervariable fragment RP39.4-9 kb. On the right-hand side of (b), estimates of the number of RPS units are provided. Molecular masses (kb) are shown to the left of (a). Note that except for strain RP41.1, every band present in the RPS pattern is also present in the C2 and H3 pattern.

Table 1. The molecular size (kb) and estimated number of RPS repeats (in parentheses) of bands containing one or more full-length repeats

The bottom line shows the minimum estimate of the total number of full-length RPS repeats per strain.

Strains ...	RP5.1	RP10.2	RP11.1	RP12.1	RP18.1	RP19.1	RP21.1	RP22.1	RP25.1	RP38.1	RP39.4	RP40.1	RP41.1
	29.6 (11)	16.2 (5)	31.2 (12)	27.3 (10)	27.3 (10)	30.4 (12)	30.4 (12)	31.2 (12)	28.8 (11)	27.3 (10)	24.6 (9)	29.6 (11)	28.8 (11)
	15.9 (5)		24.0 (9)	24.6 (9)	18.7 (6)	14.4 (4)		25.2 (9)	27.3 (10)	24.0 (9)	20.9 (7)	22.4 (8)	27.3 (10)
	6.4 (1)		22.4 (8)	18.7 (6)				22.0 (8)	24.0 (9)	7.5 (1)	17.5 (6)	17.2 (6)	24.0 (9)
			14.4 (4)					16.2 (5)		6.4 (1)	14.2 (4)	6.4 (1)	12.8 (4)
								11.9 (3)		6.1 (1)	6.8 (1)	6.1 (1)	7.4 (1)
											6.2 (1)		
Total	17	5	33	25	16	16	12	37	30	22	28	27	35

chromosomes are located between the same non-RPS 5' and 3' sequences.

Estimation of the number of RPS units per cluster and the minimum number of RPS-containing loci

Since RPS elements are flanked by the same non-RPS 5' and 3' sequences, bands of increasing molecular size in RPS-probed Southern blots of *NsiI/SalI*-digested DNA (Fig. 5b) should represent fragments containing increasing numbers of RPS tandem repeats. The set of bands between 4 and 5.1 kb (Fig. 5b), however, are too

small to contain a full-length RPS unit and still possess complete 5' and 3' flanking sequences. Using the known molecular sizes of the 3' and 5' flanking sequences and the size of RPS units so far sequenced, the minimum number of full-length RPS units per band was estimated for each of 13 *C. albicans* isolates (Table 1). On average, each isolate contained 3.5 ± 1.7 bands with one or more full-length RPS units. The range of bands per isolate with one or more full-length RPS units was 1–6. The average band with full-length sequences contained 6.7 RPS units. This results in a minimum estimate of 23.3 ± 9.6 complete units per *C. albicans* isolate, with a

Fig. 6. Microevolution of RPS-containing sequences *in vitro*. Four *C. albicans* isolates, RP39.2 (a), RP39.1 (b), RP10.3 (c) and RP5.1 (d) were grown in liquid medium for 3000 generations. Every 200 generations, cells were cloned and the DNA of one clone was processed for Southern-blot hybridization with the C1 fragment of Ca3. The number of generations is noted at the top of each lane. Molecular sizes (kb) are shown to the left of each gel.

range of 5–37 per isolate. These latter figures represent underestimates since full-length RPS-containing fragments can share the same molecular size, and this is suggested in a few cases by differences in the intensity of the same band in different isolates. Based upon densitometry measurements, Iwaguchi *et al.* (1992) estimated that strains FC18 and NUM812 contained 60 and 80 RPS repeats, respectively. Their estimates included partial as well as full-length RPS units and were made on the two strains with the highest number of RPS repeats.

In the molecular size range of 4.0–5.1 kb, each strain exhibited between 2 and 4 bands containing RPS sequences (Fig. 5b). The mean number of bands in this size range was 3.1 ± 0.6 . This again is an underestimate, since bands can contain fragments of different genomic origin. In addition, the C2 and H3 hybridization intensities of several of the bands in this range are greater than the high-molecular-mass bands containing tandem full-length RPS repeats. In contrast, the RPS hybridization intensities of the bands in this range are similar to those of the high-molecular-mass bands containing tandem full-length RPS repeats. These results suggest that while the high-molecular-mass bands represent one fragment or a low copy number per genome, the low-molecular-mass bands with partial RPS sequences between 4 and 5 kb represent higher copy numbers per genome.

Deletion and insertion of RPS units occur *in vitro*

By comparing isolates of a strain infecting an individual, we have demonstrated that reorganization of RPS-containing sequences occurs *in vivo*. To demonstrate that deletions and insertions of RPS-containing sequences occur *in vitro* and to obtain an estimate of frequency, four *C. albicans* isolates were grown for 3000 generations at 37 °C. A liquid growth culture of each strain was initiated from a single colony and, therefore, represented a clone. Every 200 generations, cells were plated and a single colony was used to initiate a new liquid growth culture, and was also grown and stored on agar slants for further analysis. Genomic DNA was digested with *EcoRI* and probed with the C1 fragment (Fig. 6). Each of the four test isolates exhibited microevolution within 3000 generations. In every case, the disappearance of a band at one molecular size was accompanied by the appearance of a band at another molecular size, suggesting a molecular size change at the same genomic locus. This conclusion is further supported by the relative stability of each band after a size change. The only example in which two bands changed at the same time was in isolate RP5.1, between 600 and 800 generations. In every case, the change in Ca3 pattern was observed in the C1 pattern (data not shown). In every case, increases or decreases in the size of a band over time occurred in multiples of approximately 2–2.5 kb, the approximate size range of the full-

length RPS units so far sequenced. These results demonstrate that spontaneous reorganization of sequences with tandemly repeated RPS units occurs at high frequency *in vitro*. The combined results of the four *C. albicans* isolates each monitored over 3000 generations provide a rough estimate of one reorganizational event per 1000 generations. These results also demonstrate that microevolutionary changes involving RPS-containing sequences can lead to divergence, convergence and reversion of patterns.

DISCUSSION

The complex fingerprinting probe Ca3 generates a Southern-blot hybridization pattern that contains invariable (monomorphic), moderately variable and hypervariable bands (Schmid *et al.*, 1990). The hypervariable bands have been successfully used to assess microevolution in infecting strains (Schroeppel *et al.*, 1994; Lockhart *et al.*, 1995, 1996). These bands represent genomic *EcoRI* fragments that hybridize to the C fragment of the 11 kb Ca3 probe. The C fragment, which is a 2.6 kb *EcoRI* digestion product of Ca3, can be cleaved with *SacI* to generate two fragments, C1 and C2. C1, which represents the 3' terminus of the Ca3 probe, was previously shown to contain a portion of the RPS repeat element (Lockhart *et al.*, 1995). RPS elements have been demonstrated to be dispersed throughout seven of the eight *C. albicans* chromosomes (Iwaguchi *et al.*, 1992) and to be present at least at two loci on chromosome seven (Chibana *et al.*, 1998). This suggests a minimum of 8 RPS loci or 16 RPS-containing regions in the diploid *C. albicans* genome. Since one of our long term objectives is to develop a similarity coefficient (S_{AB}) that incorporates the rate of change of each band in the Ca3 fingerprint pattern, and thus provide a more accurate measure of genetic distance between independent *C. albicans* isolates, we have begun to examine the molecular basis of band variability. Here, we describe experiments that were performed to elucidate the molecular mechanism responsible for the frequent changes in the hypervariable bands of the Ca3 fingerprint pattern used to monitor microevolution.

Microevolution involves the deletion/insertion of full-length RPS elements at specific genomic sites

The microevolutionary changes identified by Ca3 and the C1 fragment that commonly take place in infecting strains of *C. albicans* are restricted primarily to high-molecular-mass bands. We have demonstrated that the great majority of the genomic fragments that represent the high-molecular-mass hypervariable bands of the Ca3 pattern possess the same 3' and 5' non-RPS sequences bordering a single RPS unit or tandem sequence of RPS units. To test the molecular basis of size variability of these RPS-containing bands, we cloned and sequenced a fragment from two isolates of a single infecting strain that had undergone a size change, and demonstrated that the difference in size was due

Fig. 7. Model for unequal sister-chromatid exchange for a cluster of two RPS tandem repeats. RPSs are represented by arrows (RPS A and RPS B). The two sister chromatids are represented as black and grey. Unequal sister chromatid exchange involves the association and homologous recombination of out-of-phase tandem RPS repeats A and B located on sister chromatids of the same chromosome. The hypothetical centromere location is represented by a circle. In this model, deletions and insertions occur concomitantly resulting, after cell division, in one daughter cell receiving one chromatid with only one RPS unit, AB, and the other daughter cell receiving the other sister chromatid containing three RPSs, A, AB and B.

exclusively to a difference in the number of full-length tandem RPS units. By Southern analysis, we further demonstrated that over time, changes that occur *in vitro* in the size of RPS-containing fragments involve the deletion or addition of full-length RPS units. Additions and deletions occurred at equal frequencies, and the mean frequency of change was estimated to be one per 1000 cell divisions. This provides us with the first piece of information that can be used for weighting bands in a more complex computation of similarity between Ca3 fingerprint patterns.

Models for RPS reorganization

We have presented evidence that the majority of reorganizational events involving RPS units result in a single band change. If an unequal recombinational event occurred between homologous RPS-containing loci on chromatids of homologous chromosomes, the probabilities of generating daughter cells with a change in the molecular mass of two, one or no bands in the Southern blot hybridization pattern would be 25%, 50% and 25%, respectively. Since the daughter cells with no change cannot be discriminated, the proportions of band changes one would expect are 67% for single-band changes and 33% for double-band changes. We observed frequencies of 92% and 8%, respectively. Our results, therefore, suggest that reorganization during growth does not involve unequal recombination between chromatids of homologous chromosomes. It has been suggested that RPS elements may play a role in ectopic chromosomal reorganization (Chibana *et al.*,

Fig. 8. Model for slipped misalignment during DNA replication for a cluster of two RPS tandem repeats leading to deletion (a) or duplication (b) of an RPS unit. In both cases, the parental and daughter strands containing the RPS cluster are shown. RPSs are represented by thick black arrows. Vertical lines indicate the position of the COM29 elements. The position of the four alt subrepeats a, b, c and d of RPS are shown as a, b, c and d. This model requires pausing of DNA replication followed by unpairing of the nascent daughter strand from the parental strand and subsequent pairing in a different register. Deletion and duplication are not reciprocal events. In the deletion model (a), the synthesis of the first RPS copy is followed by a forward slippage of the daughter strand due to looping of the parental strand that results in deletion of one RPS copy. In the duplication model (b), both RPS copies are synthesized followed by backward slippage and looping of the daughter strand, and continued synthesis leading to duplication of one RPS copy. To the right of both panels is an enlargement of a putative secondary structure that could stabilize the looped-out single stranded DNA. The shaded regions on the loops correspond to the enlarged areas. In this schematic representation, the palindromic sequences of three COM29 sequences from the alt cluster of the first RPS copy are shown as they form intrastrand hybridization structures with three COM29 sequences from the alt cluster of the second RPS copy. Note that the second alt cluster is in a reverse orientation due to the looping out of the DNA single strand. All possible secondary structures that could form in a 2.3 kb single-stranded RPS DNA loop have not been represented. The remaining nucleotides in the loops are represented by dashed lines.

1994; Chu *et al.*, 1993). The electrophoretic karyotypes of each of the four strains (39.2, 39.1, 10.3 and 5.1) that were monitored *in vitro* for RPS reorganization events were compared at zero time and after 3000 generations (data not shown). The electrophoretic karyotypes were found in all cases to be stable, demonstrating that within the limited number of generations monitored, reciprocal ectopic recombination between RPS-containing sequences did not occur. This does not, however, rule out the possibility that they occur, but at frequencies too low to be observed in our experiments.

High-frequency recombinational events were observed only in RPS-containing regions with one or more full-length RPS units. Reorganizational events resulted in molecular mass changes consistent in each case with the addition or deletion of one or more full-length RPS units. In one case (strain RP5.1), a 7.5 kb band, estimated to contain one RPS unit, increased in size to 9.7 kb after 800 generations, then decreased back to 7.5 kb after 1400 generations. This appeared to represent first the addition of a complete RPS unit, then the deletion of a complete RPS unit. There was no case in which an RPS-

containing fragment decreased in size below that which accommodated one complete RPS unit. Therefore, duplication appears to require a minimum of one complete RPS unit, while deletion appears to require at least two RPS units.

Our results are consistent with two models of intra-chromosomal recombination: unequal sister-chromatid exchange and slipped misalignment at the replication fork. Unequal sister-chromatid exchange occurs during or soon after DNA replication (Kuzminov, 1996), and involves the association and homologous recombination of out-of-phase tandem repeat arrays. A model of unequal sister-chromatid exchange has been adapted to a cluster of two full-length RPS units in Fig. 7. In an unequal cross-over between RPS clusters, deletion and insertion occur concomitantly. After the unequal cross-over event and subsequent cell division, one daughter cell would receive the chromatid with deleted RPS and the other cell the chromatid with inserted RPS. Since the protocol we used to follow changes over 3000 generations in the four test strains involved cloning every 200 generations, the fingerprint obtained each 200

generations could include only one of the two genotypes resulting from sister chromatid exchange. Consistent with this model was the observation that the number of duplications was equal to the number of deletions in the combined 12000 generations of the four strains analysed.

Slipped misalignment occurs during replication and involves slippage of the daughter strand during the replication process (Levinson & Gutman, 1987; Kunkel, 1990). Models of slipped misalignment have been adapted to the replication of a cluster of two full-length RPS units for deletion in Fig. 8 (a) and duplication in Fig. 8 (b). In these models, a deletion and a duplication are not reciprocal outcomes. Slipped misalignment is frequently associated with additional symmetry elements flanking direct repeats that can promote pausing of DNA replication, promote strand slippage and stabilize the misaligned sequences (Glickman & Ripley, 1984; Trinh & Sinden, 1991, 1993; Ehrlich *et al.*, 1993). The requirement for stabilization may be particularly important for slippage at RPS clusters since the large size of RPS units (2.1–2.3 kb) would suggest a relatively unstable loop. The COM29 element at the end of each alt element in the RPS unit may play such a role (Fig. 8). Complete and partial COM29 elements are imperfect palindromes and, therefore, could promote pausing of replication by forming cruciform structures. Once the single-stranded DNA loop is formed, COM29 sequences of one RPS element could hybridize with COM29 elements of an adjacent RPS element on the same DNA strand aligned in the opposite direction as diagrammed in the loops in Fig. 8. The inverted palindromes in the model provide three intrastrand hybridizing areas of 12, 12 and 15 bp. In the case of a deletion, the coarrangement event would occur at the alt c sequence (Fig. 8a) leading to a free loop that is lost. Rearrangement at the alt c sequence was in fact observed in the deletion event that occurred in RP39.1–13.7 kb in *E. coli*, leading to the genesis of RP39.1–9 kb. In the case of the duplication model (Fig. 8b), the duplicate sequence would be localized between alt c sequences.

Since the outcomes in daughter cells are different in the two compatible models, they can be distinguished experimentally. In the case of an unequal sister-chromatid exchange, one daughter cell will contain a duplication and one a deletion. In the case of slipped misalignment, one daughter cell will contain either a duplication or deletion, and one will be unaltered. Experiments are now in progress to distinguish between the alternative models.

ACKNOWLEDGEMENTS

The authors are indebted to Dr S. Lockhart for valuable suggestions and to Dr B. Reed of the University of Michigan for the collections of recurrent vaginitis strains. This research was supported by Public Health grants AI39735 and DE10758 from the National Institutes of Health awarded to D.R.S.

REFERENCES

- Anderson, J., Srikantha, T., Morrow, B., Miyasaki, S. H., White, T. C., Agabian, N., Schmid, J. & Soll, D. R. (1993). Characterization and partial nucleotide sequence of the DNA fingerprinting probe Ca3 of *Candida albicans*. *J Clin Microbiol* **31**, 1472–1480.
- Chibana, H., Iwaguchi, S.-I., Homma, M., Chindamporn, A., Nakagawa, Y. & Tanaka, K. (1994). Diversity of tandemly repetitive sequences due to short periodic repetitions in the chromosomes of *Candida albicans*. *J Bacteriol* **176**, 3851–3858.
- Chibana, H., Magee, B. B., Grindle, S., Ran, Y., Scherer, S. & Magee, P. T. (1998). A physical map of chromosome 7 of *Candida albicans*. *Genetics* **149**, 1739–1752.
- Chindamporn, A., Nakagawa, Y., Homma, M., Chibana, H., Doi, M. & Tanaka, K. (1995). Analysis of the chromosomal localization of the repetitive sequences (RPSs) in *Candida albicans*. *Microbiology* **141**, 469–476.
- Chindamporn, A., Nakagawa, R., Mizuguchi, I., Chibana, H., Doi, M. & Tanaka, K. (1998). Repetitive sequences (RPSs) in the chromosomes of *Candida albicans* are sandwiched between two novel stretches, HOK and RB2, common to each chromosome. *Microbiology* **144**, 849–857.
- Chu, W.-S., Magee, B. B. & Magee, P. T. (1993). Construction of an *Sfi*I macrorestriction map of the *Candida albicans* genome. *J Bacteriol* **175**, 6637–6651.
- Church, G. M. & Gilbert, W. (1984). Genomic sequencing. *Proc Natl Acad Sci USA* **81**, 1991–1995.
- Ehrlich, S. D. (1989). Illegitimate recombination in bacteria. In *Mobile DNA*, pp. 799–832. Edited by D. E. Berg & M. M. Howe. Washington, DC: American Society for Microbiology.
- Ehrlich, S. D., Bierne, H., d'Alencon, E., Vilette, D., Petranovic, M., Noirot, P. & Michel, B. (1993). Mechanisms of illegitimate recombination. *Gene* **135**, 161–166.
- Glickman, B. W. & Ripley, L. S. (1984). Structural intermediates of deletion mutagenesis: a role for palindromic DNA. *Proc Natl Acad Sci USA* **81**, 512–516.
- Hellstein, J., Vawter-Hugart, H., Fotos, P., Schmid, J. & Soll, D. R. (1993). Genetic similarity and phenotypic diversity of commensal and pathogenic strains of *Candida albicans* isolated from the oral cavity. *J Clin Microbiol* **31**, 3190–3199.
- Iwaguchi, S.-I., Homma, M., Chibana, H. & Tanaka, K. (1992). Isolation and characterization of a repeated sequence (RPS1) of *Candida albicans*. *J Gen Microbiol* **138**, 1893–1900.
- Kleinegger, C., Lockhart, S. R., Vargas, K. & Soll, D. R. (1996). Frequency, intensity, species and strains of oral yeast vary as a function of host age. *J Clin Microbiol* **34**, 2246–2254.
- Kunkel, T. A. (1990). Misalignment-mediated DNA synthesis errors. *Biochemistry* **29**, 8003–8011.
- Kuzminov, A. (1996). *Recombinational Repair of DNA Damage*. Austin, TX: R. G. Landes.
- Levinson, G. & Gutman, G. A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* **4**, 203–221.
- Lockhart, S., Fritch, J. J., Meier, A. S., Schroppel, K., Srikantha, T., Galask, R. & Soll, D. R. (1995). Colonizing populations of *Candida albicans* are clonal in origin but undergo microevolution through C1 fragment reorganization as demonstrated by DNA fingerprinting and C1 sequencing. *J Clin Microbiol* **33**, 1501–1509.
- Lockhart, S., Reed, B. & Soll, D. R. (1996). Most frequent scenario for recurrent *Candida* vaginitis is strain maintenance with

'substrain shuffling': demonstration by sequential DNA fingerprinting with probes Ca3, C1, and CARE2. *J Clin Microbiol* **34**, 767–777.

Lockhart, S. R., Joly, S., Vargas, K., Swails-Wenger, J., Enger, L. & Soll, D. R. (1999). Defenses against oral *Candida* carriage break down in the elderly. *J Dent Res* **78**, 857–868.

Marco, F., Lockhart, S. R., Pfaller, M. A. & 12 other authors (1999). Elucidating the origins of nosocomial infections with *Candida albicans* by DNA fingerprinting with the complex probe Ca3. *J Clin Microbiol* **37**, 2817–2828.

Odds, F., Schmid, J. & Soll, D. R. (1990). Epidemiology of *Candida* infections in AIDS. In *Mycoses in AIDS Patients*, pp. 67–74. Edited by H. V. Bossche. New York: Plenum.

Pfaller, M. A., Lockhart, S. R., Pujol, C. A., Swails-Wenger, J. A., Messer, S. A., Edmund, M. B., Jones, R. N., Wenzel, R. P. & Soll, D. R. (1998). Hospital specificity, vaginal specificity and fluconazole-resistance of *Candida albicans* blood stream isolates. *J Clin Microbiol* **36**, 1518–1529.

Pujol, C., Joly, S., Lockhart, S., Noel, S., Tibayrenc, M. & Soll, D. R. (1997). Parity among the randomly amplified polymorphic DNA method, multilocus enzyme electrophoresis and Southern blot hybridization with the moderately repetitive probe Ca3 for fingerprinting of *Candida albicans*. *J Clin Microbiol* **35**, 2348–2358.

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

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Schmid, J., Voss, E. & Soll, D. R. (1990). Computer-assisted methods for assessing *Candida albicans* strain relatedness by Southern blot hybridization with repetitive sequence Ca3. *J Clin Microbiol* **28**, 1236–1243.

Schmid, J., Rotman, M., Reed, B., Pierson, C. L. & Soll, D. R. (1993).

Genetic similarity of *Candida albicans* strains from vaginitis patients and their partners. *J Clin Microbiol* **31**, 39–46.

Schmid, J., Tay, Y. P., Wan, L., Carr, M., Parr, D. & McKinney, W. (1995). Evidence for nosocomial transmission of *Candida albicans* obtained by Ca3 fingerprinting. *J Clin Microbiol* **33**, 1223–1230.

Schroeppel, K., Rotman, M., Galask, R., Mac, K. & Soll, D. R. (1994). The evolution and replacement of *Candida albicans* strains during recurrent vaginitis demonstrated by DNA fingerprinting. *J Clin Microbiol* **32**, 2646–2654.

Soll, D. R. (1993). DNA fingerprinting of *Candida albicans*. *J Mycol Med* **3**, 37–44.

Soll, D. R., Langtimm, C. J., McDowell, J., Hicks, J. & Galask, R. (1987). High frequency switching in *Candida* strains isolated from vaginitis patients. *J Clin Microbiol* **25**, 1611–1622.

Soll, D. R., Galask, R., Isley, S., Rao, T. V. G., Stone, D., Hicks, J., Schmid, J., Mac, K. & Hanna, C. (1989). 'Switching' of *Candida albicans* during successive episodes of recurrent vaginitis. *J Clin Microbiol* **27**, 681–690.

Soll, D. R., Galask, R., Schmid, J., Hanna, C., Mac, K. & Morrow, B. (1991). Genetic dissimilarity of commensal strains carried in different anatomical locations of the same healthy women. *J Clin Microbiol* **29**, 1702–1710.

Trinh, T. Q. & Sinden, R. R. (1991). Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*. *Nature* **352**, 544–547.

Trinh, T. Q. & Sinden, R. R. (1993). The influence of primary and secondary DNA structure in deletion and duplication between direct repeats in *Escherichia coli*. *Genetics* **134**, 409–422.

White, T. C., Pfaller, M. A., Rinaldi, M. C., Smith, J. & Redding, S. (1997). Stable azole drug resistance associated with a substrain of *Candida albicans* from an HIV-infected patient. *Oral Dis* **3**, S102–109.

Received 12 March 1999; revised 14 June 1999; accepted 17 June 1999.