Sequence Finishing and Gene Mapping for Candida albicans Chromosome 7 and Syntenic Analysis Against the Saccharomyces cerevisiae Genome

Hiroji Chibana,*1 Nao Oka,* Hironobu Nakayama,† Toshihiro Aoyama,† B. B. Magee,‡ P. T. Magee† and Yuzuru Mikami†

*Research Center for Pathogenic Fungi and Microbial Toxizes, Chiba University, Chiba 260-8673, Japan, †Suzuka National College of Technology, Suzuka 510-0294, Japan and ‡Department of Genetics, Cell Biology, and Development, University of Minnesota, Saint Paul, Minnesota 55108

Manuscript received August 10, 2004
Accepted for publication April 23, 2005

ABSTRACT

The size of the genome in the opportunist fungus Candida albicans is 15.6 Mb. Whole-genome shotgun sequencing was carried out at Stanford University where the sequences were assembled into 412 contigs. C. albicans is a diploid basically, and analysis of the sequence is complicated due to repeated sequences and to sequence polymorphism between homologous chromosomes. Chromosome 7 is 1 Mb in size and the best characterized of the 8 chromosomes in C. albicans. We assigned 16 of the contigs, ranging in length from 7309 to 267,590 bp, to chromosome 7 and determined sequences of 16 regions. These regions included four gaps, a misassembled sequence, and two major repeat sequences (MRS) of >16 kb. The length of the continuous sequence attained was 949,626 bp and provided complete coverage of chromosome 7 except for telomeric regions. Sequence analysis was carried out and predicted 404 genes, 11 of which included at least one intron. A 7-kb indel, which might be caused by a retrotransposon, was identified as the largest difference between the homologous chromosomes. Synteny analysis revealed that the degree of synteny between C. albicans and Saccharomyces cerevisiae is too weak to use for completion of the genomic sequence in C. albicans.

Candida albicans is an important pathogenic fungus, causing diseases ranging from superficial thrush and vaginal infections in overtly healthy humans to systemic infections of immunocompromised hosts, such as patients who have undergone organ transplants or those undergoing intensive chemotherapy. In this diploid organism, mating is regulated by a mating-type-like locus (Hull et al. 2000; Magee and Magee 2000; Lott et al. 2001), but a perfect meiotic or haploid phase has not been identified. It is supposed that reproduction is primarily by clonal propagation (Pujol et al. 1993; Graser et al. 1996; Lott et al. 1999; Xu et al. 1999), which has led to sequence polymorphisms between homologous chromosomes (Lott et al. 1999). Because meiotic analysis is not available in this fungus, the whole-genome sequence will facilitate efficient and expedient research advances. However, to attain it, we had to overcome several obstacles. Major repeat sequences (MRSs), the length of which are at least 16 kbp, are located on all chromosomes except chromosome 3 (Chibana et al. 1998; Chindamporn et al. 1998), and partial sequences of MRS are distributed in the genome, making it impossible to join contigs that flank them. Additionally, sequence polymorphisms were identified between homologous chromosomes (Chibana et al. 2000). For these reasons, it appeared that it would be difficult to complete the sequence of the entire C. albicans genome after whole-genome shotgun sequencing.

As part of characterizing the genome, a physical mapping project in C. albicans strain 1161 is underway at the University of Minnesota (Chibana et al. 1998). A macro-restriction map was constructed using SfiI, revealing that the chromosome number is 8 and that the size of the genome is 16 Mb (Chu et al. 1993). Chromosome 7 was shown to be composed of four SfiI fragments named 7C, 7A, 7F, and 7G (Chu et al. 1993). A more accurate physical map for chromosome 7 was constructed using fosmid contiguous clones and random breakage mapping (Chibana et al. 1998). In parallel, whole-genome shotgun sequencing and assembly for the C. albicans strain SC5314 genome was carried out at the Stanford Genome Technology Center. The most recent assembly of the sequences, assembly 19, is composed of 412 supercontigs, including 146 homologous pairs (http://www-sequence.stanford.edu/group/candida/index.html). For this study, one from each pair of the homologous contigs was arbitrarily discarded, and the remaining contigs were used to create a reference haploid genome consisting of 266 supercontigs (Jones et al. 2004). We used this haploid set of supercontigs as
our starting point and closed up the polymorphisms between the homologous pairs of the supercontigs.

Although the strain used for the mapping project is different from the strain used for the sequence project, major differences between these strains on the chromosome level have not been identified. In this work we demonstrate how the physical maps help with completion of the chromosomal sequence, e.g., the completion of the whole sequence of chromosome 7. In another aspect of this work, we examine whether there are obvious syntenic regions between the Saccharomyces cerevisiae and C. albicans genomes. The degree of synteny that we identified did not provide gene linkages useful for gap closing. This work is a pilot study for completion of the whole-genome sequence of C. albicans.

MATERIALS AND METHODS

DNA amplification for gap closing: PCR with each primer pair (shown in supplementary data at http://www.genetics.org/supplemental/) was carried out with Ready-To-Go PCR beads (Amersham Biosciences) using genomic DNA of C. albicans SC5314 as a template DNA. PCR was carried out using a hotstart of 3 min at 94°C followed by 35 cycles of 94°C for 10 sec, 50°C for 10 sec, and 68°C for 1 min, concluding with 68°C for 10 min. Long PCR was carried out with LA PCR kit ver.2.1 (Takara, Tokyo). Conditions used were a hotstart of 3 min at 94°C followed by 35 cycles of 98°C for 10 sec and 68°C for 20 min, concluding with a final extension of 72°C for 10 min. Genomic DNA from C. albicans strain SC5314 (Fonzi and Irwin 1993) was used for all sequence analysis in this work.

Determination of DNA sequence: DNA products amplified by long PCR were randomly fragmented into 1- to 1.5-kb fragments with HydroShear (Genomachines). The fragmented DNA was blunted and ligated with pNotI linker (Takara), using a DNA blunting kit (Takara). The ligated product was digested with NotI and ligated with pBlueScriptSK+. Determination of sequence was performed with primers pBF, GAGCGGATAA CAATTTTCACACAGGAAACAG; and pBRN, CCGTCAAGGT CGACGCGTACT. DNA was labeled with BigDye terminator cycle sequencing kits version 1.1 (ABI), and the sequence was read using ABI5100 (ABI) sequencer.

Test analyses for gene prediction: All the genome sequences and coding sequence data in S. cerevisiae currently exhibited by EMBL were used for evaluation of the gene-finding programs Glimmer 2.10 and Critica version 1.05. We selected the set of open reading frames (ORFs) that met the criteria of having the start codon ATG and the termination codon as TAA, TAG, or TGA, and that extracted a domain of 303 bases or greater (100 aa or greater) in any of the six reading frames.

Glimmer 2.10 (Delcher et al. 1999) has a function that permits training of the program, using data from a known set of ORFs to set parameters for that species and allowing better predictions for related species. For training purposes, we used the S. cerevisiae data set of ORFs of 303 or greater bases (100 aa or greater). For the Critica version 1.05 (http://geta.life.uiuc.edu/~gary/programs/CRITICA/critica105b/critica.html) analysis, sequence data of release 1 of RefSeq of microbial and fungi minus the sequence data of S. cerevisiae were used as reference sequences for BLASTN. When the termination position predicted with the gene-domain prediction tools and the termination position of the gene domain of S. cerevisiae as registered in EMBL corresponded, it was judged that the domain prediction was correct.

Syntenic analysis: BLASTX on NCBI with default values was used to compare the whole sequence of chromosome 7 against the S. cerevisiae genome. Sequences of open reading frames and also sequences of intergenic spaces were surveyed. If the score was ≥50, it was taken to indicate that the corresponding sequence was an orthologous gene. If at least two alleles were located within 20 kb on C. albicans chromosome 7 were also located within 20 kb in the genome of S. cerevisiae, then that region was defined as a synteny block.

RESULTS

Identification and mapping of supercontigs on chromosome 7: In previous work, 39 DNA probes were sequenced and mapped to chromosome 7, and the chromosomal location of 11 of these probes was determined accurately using random breakage mapping (Chibana et al. 1998). The sequences of these probes were used to perform BLASTN searches against the Stanford Candida genome website (http://www-sequence.stanford.edu:8080/bncontigs19super.html) and 15 contigs were assigned to chromosome 7 (Figure 1). Of the supercontigs, 19-10248 and 19-20248, 19-10187 and 19-20187, 19-10219 and 19-20219, 19-10110 and 19-20110, and 19-10253 and 19-20253 were identified as homologous pairs, and 19-10262, 19-2485, 19-2505, 19-2175, and 19-2506 were identified as haploid contigs. To fill the gaps between the supercontigs, a haploid set of the supercontigs (Jones et al. 2004), comprising 19-10248, 19-10187, 19-10219, 19-10110, and 19-10253, was derived from the diploid set of supercontigs. Recognition sites for SfiI were identified on 19-10262, 19-2485, 19-10187, and 19-10219. The supercontigs were mapped precisely on chromosome 7 using the SfiI map (Chiu et al. 1993) and an accurate physical map derived from fosmid contig and random breakage maps for chromosome 7 (Figure 1) (Chibana et al. 1998).

Conserved orientation of SfiI fragment 7F: Two MRSs exist on chromosome 7, one between 7A and 7F and another between 7F and 7G. Since the sequences of MRSs are highly conserved across the chromosomes, the process of assembling the contigs across the MRSs would be error prone. Indeed, there are inconsistencies in the sequences across the MRSs among the supercontigs. The orientation of fragment 7F was determined for strain 1161 (Chibana et al. 1998, 2000). However, since 7F is flanked by two inverted MRSs, it is possible that the entire 7F region could be inverted in SC5314 on one or both homologs due to homologous recombination between the MRSs. To resolve the inconsistencies and confirm the orientation of 7F on chromosome 7, long PCR amplification was performed with two pairs of PCR primers based on unique sequences flanking both the MRSs. The sequence of the products obtained from the PCR amplifications, which were ~16 kb, was determined. When the counterpart of the primer pairs was swapped, the PCR band did not appear. Thus, the results indicate that the orientation of 7F in strain
Figure 1.—Overview of gap closing. A physical map, previously published by Chibana et al. (1998), is presented with the shaded bar. Probes assigned to chromosome 7 are located on the map. Open boxes on the shaded bar describe sites for the probes that are mapped with random breakage mapping. SfiI sites are described with vertical open bars on the shaded bar. The location for ODP2 was corrected by the authors after publication. Two arrows on the shaded bar describe MRS. Although the MRSs are represented with a single repetitive sequence (RPS) (Iwaguchi et al. 1992) in this map, RPS is a tandem repeated sequence in some other MRS. The total length of chromosome 7 is 995 kbp in this map. Supercontigs assigned on chromosome 7 are described with a closed arrow headed by a solid bar. Probes with a known DNA sequence and mapped on the supercontigs are described with an open box. R2B9, SfiI, RBPI, ARS3, ARG4, and G2E10 were used as landmarks to assign the supercontigs on the physical map. To adjust the position of the landmarks, supercontig 19-10262 was cleaved with a spacer. A sequence deletion was identified on supercontig 19-20248 and indicated with a spacer. Gaps between the supercontigs and region including ambiguous sequences were amplified with PCR and are indicated by numbers 1–16 in parentheses. Ca35A5 represents a whole sequence of cosmid inserted DNA no. AL033396. The inserted DNA of the cosmid was isolated from strain 1161 and the sequence was determined using the methods described by Tait et al. (1997).
SC5314 is the same as in strain 1161 and in other strains for which the orientation was determined by Chibana et al. 1998, 2000).

PCR amplification and sequencing to fill the gaps between the supercontigs and to elucidate ambiguous sequences in the supercontigs: To obtain the sequence of the gaps between the supercontigs, appropriate PCR primers were designed near the ends of each supercontig. PCR was carried out using those primers, and the sequence of the PCR products was determined. Since supercontigs 19-10248, 19-2305, and 19-10187 overlapped each other, they were assumed to contain assembly errors. About 10 kb extending from the end of 19-10248 to the middle of 19-10187 was amplified by PCR and the sequence was determined. Another assembly error, caused by repetitive sequences located at three points, was found on 19-10253. The assembly errors were corrected, and the gaps between 19-10110 and 19-10253 and between 19-10253 and 19-2335 were closed. Although supercontig 19-2335 was not identified with BLASTN search, it was assigned between 19-10253 and 19-2506 because it shares homologous sequences with 19-10253.

Undetermined sequences, which were depicted in the Stanford assembly 19 using the letter n, ranged in length from a single nucleotide to 360 nucleotides and had a total length of 659 bp distributed across seven regions on the supercontigs. The unread regions were recovered from the GenBank database, once the sequences had been integrated with assembly 19 (Jones et al. 2004). These regions were amplified with flanking primers, and their sequences were determined in this work.

Gene prediction of chromosome 7 in C. albicans: Gene-finding tools Glimmer 2.10 (Salzberg et al. 1998) and Critica 1.05 work with high reliability for finding genes within sequences, provided that the sequences do not include introns. Thus, these tools have been used for sequence analyses on prokaryotic genomes (Aggarwal and Ramaswamy 2002; McHardy et al. 2004). Unlike some other fungal species, only a small fraction of genes in the genome of C. albicans carry introns (Braun et al. 2005). The Candida intron structure is generally similar to that of S. cerevisiae (Jones et al. 2004). For these reasons, Glimmer 2 and Critica were employed to begin gene finding. To apply these programs to gene finding in fungal genomes, their reliability first has to be evaluated. The genomic sequence of S. cerevisiae was surveyed using these tools. In this test evaluation, ORFs were classified into three groups according to the robustness of the results. Class 1 ORFs were identified with both Glimmer 2 and Critica and class 2 ORFs were identified by only one of the programs (Table 1). The third group contained ORFs that were predicted by neither program.

On chromosome 7, 516 open reading frames encoding >100 amino acids were identified. According to our method for ORF classification, 373 class 1 and 18 class 2 ORFs were identified. For these ORFs, the specificity is >95%. The remaining 125 ORFs were not suggested by either program. Of this remainder, 107 ORFs were not counted as coding sequences, since the complementary chain encoded an overlapping gene predicted with higher probability. It was not possible to confirm that the other 18 ORFs encoded polypeptides. BLASTX analysis showed that 2 ORFs <100 amino acids had a score >50 against S. cerevisiae or Schizosaccharomyces pombe. A total of 20 ORFs were classified as class 3 genes. At 16 sites a coding region was possibly divided by an intron or sequence error on the supercontigs. For these sites, sequence determination and intron identification were performed, and an intron was suggested in 9 coding sequences (CDSs) and two introns in each of 2 other CDSs. In the remaining 5 sites, sequencing errors were identified and corrected. A gene encoding a Leu-trNA was also identified. A total of 404 CDSs were predicted as genes (Table 2, Figure 2). The gene maps, to which BLAST and Pfam information are attached, are available as supplementary data at http://www.genetics.org/supplemental/.

Syntenv between C. albicans chromosome 7 and the genome of S. cerevisiae: In previous analyses of synteny between C. albicans chromosome 7 and the S. cerevisiae genome, few syntenic regions were identified because of the low resolution of sequence for the chromosome. The gene map was based on the fosmid contig map, which is composed of a tiling set of fosmid clones. The average length of the fosmid DNA insert was 40 kbp, and only 39 probes were mapped with their sequence information (Chibana et al. 1998). The continuous DNA sequence information covering chromosome 7 allowed us to perform syntenic analysis at a higher resolution. The gene arrangement of the 282 C. albicans ORFs identified by sequence similarity to the genome of S. cerevisiae was compared in the two fungi. A group of ORFs with close linkage in both fungi was called a synteny block. A total of 32 synteny blocks were identified between chromosome 7 in C. albicans and the genome in S. cerevisiae. The number of ORFs found in shared

Figure 2.—Gene maps for C. albicans chromosome 7. Class 1 and class 2 ORFs are shown as rectangles with thick and thin outlines, respectively. Class 3 ORFs are shown as rectangles with no outline. Each ORF was colored the same color as the best-hit species: red, blue, green, and yellow for S. cerevisiae, S. pombe, N. crassa, and Homo sapiens, respectively. Solid arrowheads pointing toward the coding sequence indicate the position of the intron. CDSs, which are reorganized in this work, are represented by asterisks. ORF number of assembly 19 is assigned in the column of C. albicans. Results of BlastX, Pfam, and more information are available as supplementary data at http://www.genetics.org/supplemental/
Finishing of *C. albicans* Chromosome 7
Figure 2.—Continued.
Figure 2.—Continued.
Figure 2.—Continued.
Figure 2.—Continued.
Figure 2.—Continued.
TABLE 1

<table>
<thead>
<tr>
<th></th>
<th>ORF &gt;100 aa</th>
<th>Class 1: Glimmer 2 or Critica but not Glimmer 2</th>
<th>Predicted by neither Glimmer 2 nor Critica</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of prediction</td>
<td>7467 (100)</td>
<td>4702 (77.1)</td>
<td>2006 (10.5)</td>
</tr>
<tr>
<td>Correct prediction</td>
<td>5831 (78.1)</td>
<td>4498 (95.7)</td>
<td>612 (30.5)</td>
</tr>
</tbody>
</table>

Numbers in parentheses are percentages.

synteny blocks ranged from 2 to 8 ORFs. The average of the number of ORFs per synteny block was 2.68 ORFs. A synteny block composed of eight ORFs is the largest area of synteny between chromosome 7 and the S. cerevisiae genome. However, the order and direction of the ORFs were jumbled, with at least three indels and four inversions being found in this region (Figure 3).

DISCUSSION

In previous work, probes G2E10 and R2B9 were assigned the edges of chromosome 7 (Chibana et al. 1998). G2E10 was assigned 35 kbp away from one end of chromosome 7. The sequence of G2E10 was identified 29 kbp away from the end of contig 19-10262. This indicates that the end of contig 19-10262 is 6 kbp from the end of chromosome 7. Using similar reasoning, we suggest that the end of contig 19-2506 reaches to within 20 kbp of the other end of chromosome 7. The total length of the continuous sequence composed of contigs and filled gaps comes to 950 kb. The remaining sequences, including telomeres and subtelomeres, amount to 26 kb. The telomere sequence in C. albicans is composed of 23 bp repeated sequences (Sadhu et al. 1991). Long PCR was carried out using PCR primers based on the sequence of the telomere and supercontigs to amplify the telomeres and subtelomeres, but the expected product was not detected (data not shown). The same problem is likely to happen in other chromosome gap closings. Other approaches, e.g., cloning into a cosmId vector, are needed to determine the sequence of these regions.

The supercontigs released by Jones et al. (2004) were assigned to chromosome 7, and the gap closing and sequence correction of chromosome 7 in C. albicans was carried out in this work. The length of the continuous sequence is 949,626 bp and covers the entire chromosome 7 except for telomeric and subtelomeric regions. The total length of the missing sequences was only 289 bp in the gaps and 659 bp in continuous sequence of the supercontigs. Thus, the coverage of the supercontigs in assembly 19 was 99.9% of the continuous sequences of chromosome 7.

The differences in length between the homologous pairs of the supercontigs of chromosome 7 are <1 kb, except between 19-10248 and 19-20248. A region causing the difference was identified and is depicted in Figure 4. The difference in length is 7249 bp, due to a section in 19-20248 that contains three defective ORFs flanked by long terminal repeats (LTRs) of the class described by Goodwin and Poulter (2000) as LTR η. Because there were no LTR sequences on the corresponding region in 19-20248, this region has been inserted on only one homolog. Three ORFs included a sequence homologous to pol- and gag-like elements and to STA1, respectively. STA1 is associated with initiation of the developmental programs of pseudohyphal formation and invasive growth response in S. cerevisiae (Vivier et al. 1997). Although the similarity is not high between the ORF on chromosome 7 and STA1 in S. cerevisiae, this retrotransposon-like element might contribute to cell morphology polymorphism and pathogenesis in C. albicans. We suggest that this region should be investigated further.

TABLE 2

<table>
<thead>
<tr>
<th></th>
<th>Class 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF &gt;100 aa</td>
<td>Class 1</td>
</tr>
<tr>
<td>No. of prediction</td>
<td>516</td>
</tr>
</tbody>
</table>

* BLASTX analysis showed that two ORFs of <100 amino acids had a score >50.
The greatest synteny block between chromosome 7 in \textit{C. albicans} and all chromosomes in \textit{S. cerevisiae} includes \textit{LEU2} and \textit{NFS1} (Figure 3). Interestingly, that region has received attention from other groups. It was published that close linkage of these genes was conserved in \textit{Ashbya gossypii}, \textit{C. albicans}, \textit{C. maltosa}, \textit{C. rugosa}, \textit{Pichia anomala}, \textit{S. cerevisiae}, \textit{S. servozii}, \textit{Yamadazyma ohmeri}, and \textit{Zygosaccharomyces rouxii} (Keogh et al. 1998; De la Rosa et al. 2001). We surveyed the linkage of both genes in other fungi. The linkage of the genes was not conserved in \textit{S. pombe} (Wood et al. 2002) (http://www.sanger.ac.uk/Projects/S_pombe/) or in \textit{Neurospora crassa} (Galagan et al. 2003) (http://www-genome.wi.mit.edu/annotation/fungi/neurospora/index.html). The linkage of these genes is thus likely conserved only in closely related Ascomycetes species. In this work, it was revealed that the conserved region extends beyond \textit{LEU2} and \textit{NFS1} and comprises the longest syntenic block in chromosome 7 in \textit{C. albicans} and all chromosomes in \textit{S. cerevisiae}.

**Figure 3.**—The longest syntenic region between the genome in \textit{S. cerevisiae} and chromosome 7 in \textit{C. albicans}. The region between \textit{DCC1} and \textit{PGS1} includes 11 ORFs encoding polypeptides with $>100$ aa. Of these, 8 ORFs were identified as sharing similarity in amino acid sequence with genes in \textit{S. cerevisiae} with a score $>50$ in BLASTX results. Although linkage of these genes is conserved in \textit{C. albicans} and \textit{S. cerevisiae}, at least four inversions and three indels were identified.

The greatest synteny block between chromosome 7 in \textit{C. albicans} and all chromosomes in \textit{S. cerevisiae} includes \textit{LEU2} and \textit{NFS1} (Figure 3). Interestingly, that region has received attention from other groups. It was published that close linkage of these genes was conserved in \textit{Ashbya gossypii}, \textit{C. albicans}, \textit{C. maltosa}, \textit{C. rugosa}, \textit{Pichia anomala}, \textit{S. cerevisiae}, \textit{S. servozii}, \textit{Yamadazyma ohmeri}, and \textit{Zygosaccharomyces rouxii} (Keogh et al. 1998; De la Rosa et al. 2001). We surveyed the linkage of both genes in other fungi. The linkage of the genes was not conserved in \textit{S. pombe} (Wood et al. 2002) (http://www.sanger.ac.uk/Projects/S_pombe/) or in \textit{Neurospora crassa} (Galagan et al. 2003) (http://www-genome.wi.mit.edu/annotation/fungi/neurospora/index.html). The linkage of these genes is thus likely conserved only in closely related Ascomycetes species. In this work, it was revealed that the conserved region extends beyond \textit{LEU2} and \textit{NFS1} and comprises the longest syntenic block in chro-

**Figure 4.**—Sequence indel between 19-10248 and 19-20248. The length of this indel is 7249 bp. LTR \( \eta \) (Goodwin and Poultier 2000) flanks this region. The best hits of three ORFs are \textit{YLT1}, hypothetical protein \textit{YLT1} (TrEMBL no. Q9COU0) suggested to encode a fungal retroelement polypolymer in \textit{Yarrowia lipolytica} (1e-44); \textit{YLS8}, putative protein (TrEMBL no. Q64LP8) suggested to encode Gag or capsid-like proteins from LTR retrotransposons \textit{Oryza sativa} (1e-25); and \textit{STA1}, putative protein (TrEMBL no. P08640) suggested to encode glucoamylase S1/S2 precursor (EC 3.2.1.3) (glucan 1,4- \( \alpha \)-glucosidase) (1,4- \( \alpha \)-d-glucan) in \textit{S. cerevisiae} (9e-06).

**Figure 3.**—The longest syntenic region between the genome in \textit{S. cerevisiae} and chromosome 7 in \textit{C. albicans}. The region between \textit{DCC1} and \textit{PGS1} includes 11 ORFs encoding polypeptides with $>100$ aa. Of these, 8 ORFs were identified as sharing similarity in amino acid sequence with genes in \textit{S. cerevisiae} with a score $>50$ in BLASTX results. Although linkage of these genes is conserved in \textit{C. albicans} and \textit{S. cerevisiae}, at least four inversions and three indels were identified.

The greatest synteny block between chromosome 7 in \textit{C. albicans} and all chromosomes in \textit{S. cerevisiae} includes \textit{LEU2} and \textit{NFS1} (Figure 3). Interestingly, that region has received attention from other groups. It was published that close linkage of these genes was conserved in \textit{Ashbya gossypii}, \textit{C. albicans}, \textit{C. maltosa}, \textit{C. rugosa}, \textit{Pichia anomala}, \textit{S. cerevisiae}, \textit{S. servozii}, \textit{Yamadazyma ohmeri}, and \textit{Zygosaccharomyces rouxii} (Keogh et al. 1998; De la Rosa et al. 2001). We surveyed the linkage of both genes in other fungi. The linkage of the genes was not conserved in \textit{S. pombe} (Wood et al. 2002) (http://www.sanger.ac.uk/Projects/S_pombe/) or in \textit{Neurospora crassa} (Galagan et al. 2003) (http://www-genome.wi.mit.edu/annotation/fungi/neurospora/index.html). The linkage of these genes is thus likely conserved only in closely related Ascomycetes species. In this work, it was revealed that the conserved region extends beyond \textit{LEU2} and \textit{NFS1} and comprises the longest syntenic block in chro-

---

**LITERATURE CITED**


Chindamporn, A., Y. Nakagawa, I. Mizuguchi, H. Chibana, M. Doi et al., 1998 Repetitive sequences (RPSs) in the chromosomes of \textit{Candida albicans} are sandwiched between two novel stretches, HOK and RB2, common to each chromosome. Microbiology 144 (Pt. 4): 849–857.


Graser, Y., M. Volovsek, J. Arrington, G. Schonian, W. Presser et al., 1996 Molecular markers reveal that population structure

Communicating editor: M. Sachs