Efficient Homologous and Illegitimate Recombination in the Opportunistic Yeast Pathogen *Candida glabrata*

Brendan P. Cormack* and Stanley Falkow*,†

**Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305-5402 and †Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, Hamilton, Montana 59840*

Manuscript received September 21, 1998
Accepted for publication December 7, 1998

ABSTRACT

The opportunistic pathogen *Candida glabrata* causes significant disease in humans. To develop genetic tools to investigate the pathogenicity of this organism, we have constructed *ura3* and *his3* auxotrophic strains by deleting the relevant coding regions in a *C. glabrata* clinical isolate. Linearized plasmids carrying a *Saccharomyces cerevisiae URA3* gene efficiently transformed the *ura3* auxotroph to prototrophy. Homologous recombination events were observed when the linearized plasmid carried short terminal regions homologous with the chromosome. In contrast, in the absence of any chromosomal homology, the plasmid integrated by illegitimate recombination into random sites in the genome. Sequence analysis of the target sites revealed that for the majority of illegitimate transformants there was no microhomology with the integration site. Approximately 0.25% of the insertions resulted in amino acid auxotrophy, suggesting that insertion was random at a gross level. Sequence analysis suggested that illegitimate recombination is nonrandom at the single-gene level and that the integrating plasmid has a preference for inserting into noncoding regions of the genome. Analysis of the relative numbers of homologous and illegitimate recombination events suggests that *C. glabrata* possesses efficient systems for both homologous and nonhomologous recombination.

CANDIDA species are the causative agents for both mucosal and systemic human infections. Significant progress has been made in our understanding of Candida virulence. Factors thought to contribute to virulence include the ability to grow in different morphological forms and at elevated temperature, the ability to switch between different colony/cellular phenotypes, the ability to adhere to host tissue, and the production and secretion of a variety of hydrolytic enzymes (Odds 1994). The genetic analysis of these characteristics or of other virulence attributes has been complicated in *C. albicans* because it is diploid and asexual. Thus, in general, two copies of a particular gene of interest must be altered to observe the mutant phenotype. In contrast to *C. albicans*, *C. glabrata*, an important opportunistic pathogen responsible for 10–15% of symptomatic vaginal and systemic candidiasis (Redondo-Lopez *et al.* 1990; Pfaller 1996), is haploid and therefore more amenable to genetic analysis of yeast pathogenesis. A number of molecular genetic tools exist already in *C. glabrata*. The *URA3* gene has been cloned, and *ura3* auxotrophs isolated (Zhou *et al.* 1994). Disruption of the *HIS3* and *TRP1* coding regions with *URA3* was used to generate *his3* and *trp1* auxotrophic strains (Kitada *et al.* 1995). *C. glabrata* autonomously replicating sequences (ARS) and centromere (CEN) sequences have been identified and can function to direct replication and segregation of episomal plasmids (Mehra *et al.* 1992; Kitada *et al.* 1996).

Here we analyze the fate of transforming DNA in *C. glabrata* using a virulent clinical isolate (Fidel *et al.* 1996). We describe unexpectedly high levels of nonhomologous recombination in *C. glabrata* and characterize the insertion site specificity for these nonhomologous events. We exploit this nonhomologous recombination as a method of insertional mutagenesis.

Corresponding author: Brendan Cormack, Department of Molecular Biology and Genetics, Johns Hopkins Medical School, PCTB 522, 725 N. Wolfe St., Baltimore, MD 21205-2185.
E-mail: bcormack@jhmi.edu

MATERIALS AND METHODS

Strains: The parental *C. glabrata* isolate (strain B, a gift of Paul Fidel and Jack Sobel) was a clinical isolate from a case of vaginitis that did not respond to fluconazole or boric acid treatment; it is virulent in a murine model of vaginitis (Fidel *et al.* 1996). All derivative strains are summarized in Table 1. Media were prepared as described (Sherman *et al.* 1986). YP media contained glucose (2%); synthetic complete media was supplemented with casamino acids (0.6%), adenine (25 mg · liter^{-1}), and uracil (25 mg · liter^{-1}), or with all amino acids (100 mg · liter^{-1}) except those being tested for specific auxotrophies.

Transformation of *C. glabrata*: Transformations were carried out using a variation of the modified LiAc protocol (Gietz *et al.* 1992). Cells were grown in YPD to early log phase and collected by centrifugation. They were washed twice with water, and resuspended in 0.1 m LiAc, 1 mm EDTA, and 10 mm Tris, pH 7.5. Transforming DNA, was added to the cells with

TABLE 1

Strains used in this study

Strain	Parent	Genotype
BG2	B	Clinical isolate (Fidel *et al.* 1996)
BG14	BG2	*ura3Δ(−85 + 932)::Tn903NeoR*
BG87b	BG2	*his3Δ(1 + 631) ura3Δ(−85 + 932)::Tn903NeoR*
BG88b	BG2	*his3Δ(1 + 631) ura3Δ(−85 + 932)::Tn903NeoR*
BG98	BG87b	*his3Δ(1 + 631)*
BG99	BG88b	*his3Δ(1 + 631)*

100 μg of denatured salmon sperm DNA and 0.5 ml of 0.1 m LiAc, 40% PEG (3350), 1 mm EDTA, and 10 mm Tris, pH 7.5 (1 μg). The mix was incubated at 30° for 30 min and heat-shocked at 45° for 15 min.

***URA3* deletion construct:** The neomycin-resistance gene (*neo*) of Tn903 was amplified by PCR using primers Tn903-1 and Tn903-2 (sequences shown in Table 2). The Tn903-1 primer incorporates 55 nucleotides upstream of the ATG of the *CUP1* gene of *C. glabrata*. This promoter region is sufficient to transcribe the *neo* gene at levels sufficient to give resistance to 500 μg G418/ml; the parental strain BG2 is sensitive to G418 at 50 μg/ml. We cloned a 2.2-kb *Pst*I fragment containing the *C. glabrata URA3* locus into Puc19 to generate pBC34.1. pBC34.1 was digested with *Xho*I and *Hin*dIII and ligated to the *Sal*I-*Bam*HI PCR fragment containing the *neo* gene and a *Bam*HI-*Hin*dIII fragment containing the 3′ region of the *URA3* locus, which was generated by PCR amplification with the primers URA3-940 and URA3PST (Table 2). This PCR fragment was sequenced to verify that there were no mutations introduced by PCR amplification. The final construct (pBC-39.1) consists of the *URA3* gene carried on a *Pst*I fragment in which the coding region of the *URA3* gene, from the *Xho*I site at −85, with respect to the ATG, to the *Ssp*I site at 932 has been replaced with the *neo* gene. To generate the *ura3* auxotroph, BG14, strain BG2 was transformed with *Pst*I-digested pBC39.1. The *Pst*I fragment of pBC39.1 contained the 5′ and 3′ UTR of *URA3* flanking the *neo* gene of Tn903. G418-resistant and 5-fluoroorotic acid (5-FOA)-sensitive recombinants were isolated. Southern analysis showed that the *URA3* locus had been replaced by the deleted gene containing the neomycin resistance cassette (data not shown).

***HIS3* deletion construct:** We cloned the *HIS3* locus as a *Sph*I-*Bgl* II fragment (2.5 kb) into Puc19 to generate pBC102.1. The sequence of this 2.5-kb region is available under GenBank accession number AF107116. Primer HIS3PST1, homologous to the 3′ end of the *His3* coding region, was used to amplify

TABLE 2

Oligonucleotide primers for PCR

	Sequences
Tn903-1	GGCAAAGTCGACAAAAACAATAGGCATATATAAAGCAGAGAGCACAGCAGTAC ACACATTTGCAGTATGAGCCATATTCAACGGGAAA
Tn903-2	AAAGGATCCAAGCCGCCGTCCCGTCAAGTCAG
URA3PST	AAAGCTTACTGCAGAACAAATATTCTGTCTATC
URA3-940	AAGGATCCTAAAACTAATATAATGGC
HIS3PST1	AAAACTGCAGTAGAACACAGCCCACAGCTACC
HIS3PST2	AAAACTGCAGCCTTTGCTCGATGCTTCTCTT
TARGET 15-1	CGGGATCCTGAGCTTCCAGCAGTCTCCC
TARGET 15-2	CGGAATTCACGTTACCCGGATAGAAGTG
TARGET 16-1	CGGGATCCAAACTGCTGTGGAGGAATAG
TARGET 16-2	CGGAATTCTGTCTTCCTTTCATCAGTGACT
TARGET 17-1	CGGAATTCTAAACTATCTGGTTAGTAAATTG
TARGET 17-2	CGGGATCCGGGTAGACATGTTTATATAGT
ATn903 100	GGAATTCGGGCAATCAGGTGCGACAATC
ATn903 500	GGGATCCAAAATCACTCGCATCAACC
BTn903 200	GGAATTCGATGGTCAGACTAAACT
BTn903 400	GGGATCCTCGAATGCAACCGGCGCAGGA
CTn903 360	AAGGATCCTATTTTATCCTGAATCAGGATA
DTn903 280	AAGAATTCATGGTTACTCACCACTGCGA
DTn903 330	AAGGATCCTAATACCTGGAATGCTGTTTTC
ETn903 450	GGAATTCTCAGGCGCAATCACGAATGAAT
ETn903 650	GGGATCCATCAATACAACCTATTAATTTCC
FTn903 500	GCGAATTCATGACGAGCGTAATGGCT
FTn903 600	GCGGATCCGGTTATCAAGTGAGAAATC
GTn903 520	GCGAATTCTGAACAAGTCTGGAAAG
GTn903 575	GCGGATCCGAATCCGGTGAGAATGGCA

a *Pst*I-*Bgl*II fragment and primer HIS3PST2, homologous to the 5′ end of the *HIS3* coding region, was used to amplify a *Sph*I-*Pst*I fragment. The primer sequences are in Table 2. These fragments were combined with *Sph*I-*Bam*HI-digested YIplac211 (Gietz and Sugino 1988) in a ligation to generate pBC104. Yiplac211 carries the *Saccharomyces cerevisiae URA3* gene that functions in *C. glabrata*. pBC104 plasmid carries the 5′- and 3′-untranslated regions of *HIS3*, but is deleted for the entire coding region (nucleotide 1 to 631, with respect to the "A" of the first methionine codon). The flanking region was sequenced to verify that there were no mutations introduced by PCR. PBC104 was linearized by digestion with *Not*I, for which there is a site in the 3′ UTR of the *HIS3* gene, and used to transform BG14. Ura^+ prototrophs were colony purified and then plated on SC plates containing 5-fluoroorotic acid (5FOA; 1.1 g · $liter^{-1}$). FOA^R segregants were tested for growth in the absence of histidine. Approximately half of the segregants were His^- and half His^+, as expected. The structure of the deleted locus for two isolates, BG87b and BG88b, was verified by Southern blot analysis (data not shown). Transformation of BG87b and BG88b with the 2.2-kb fragment of the *C. glabrata URA3* gene restored the *URA3* locus and generated, respectively, the *his3* auxotrophs BG98 and BG99.

Sequence of illegitimate integration sites for YIplac211: The DNA flanking an insertion site was rescued by digestion of genomic DNA from the transformant with *Eco*RI or *Hin*dIII for which there are no recognition sites in the integrating plasmid. Ligation of this digestion mixture generates a plasmid carrying the genomic DNA flanking the insertion site. Plasmid was recovered by transformation into *Escherichia coli* strain DH12 (GIBCO/BRL, Rockville, MD). For those events in which two plasmids had integrated in tandem, the ligation mix generated a dimeric plasmid in which there were two copies of the integrating vector as well as the genomic sequences flanking the integration site. In these cases, the rescued plasmid was transformed into the Rec^+ strain MC1061. The dimeric plasmid resolved into the two component plasmids: a vector-sized plasmid and a vector carrying the chromosomal DNA flanking the original genomic integration site. The flanking genomic DNA in all rescued plasmids was sequenced using primers that hybridize to the vector. The flanking DNA was sequenced on one strand only, but each integrant was sequenced twice from independently prepared templates. The overall sequence quality was excellent, making it unlikely that the lack of ORFs was an artifact of frameshifting because of poor sequence quality. The DNA sequence was used in homology searches using the BLASTX and BLASTN algorithms. For all but two sequences, there were no significant homologies in GenBank.

We obtained the sequence of flanking DNA for 37 integrants. For 17 of these, we obtained the sequence of the undisrupted genomic target site. Of these 17, 14 were insertions in a single 1.4-kb locus. For these, the sequence of the flanking DNAs for each of the insertions was assembled to give the sequence of the entire locus, including the integration sites for each insertion. For inserts 15, 16, and 17 (Figure 4) the sequence of the flanking DNA was used to synthesize primers, which were subsequently used to amplify (via PCR) the integration site from genomic DNA of the parental wild-type strain. The primers used for amplification of the insertion sites are shown in Table 2. Thus, for each insertion, we obtained the sequence of both junctions for the integrant into genomic DNA and for the undisrupted integration site.

Construction of Tn903 directed integration vectors: Fragments of the *neo* gene of Tn903 were amplified by PCR and cloned as *Eco*RI-*Bam*HI fragments into a version of YIPLAC-211 (Gietz and Sugino 1988) from which the *Hin*dIII site had previously been deleted by incubation with dNTPs and the Klenow fragment of *E. coli* polymerase I and religation. The primers include *Eco*RI sites and *Bam*HI sites and were chosen so that the natural *Sma*I or *Hin*dIII sites of the *neo* gene would be in the exact middle of the amplified fragment. The primers (A-GTn903 xxx) are shown in Table 2. The numbers refer to the nucleotide position of the neomycin resistance gene with respect to the ATG. This nucleotide is at the 5′ end of the PCR oligo and, after PCR amplification, is at the end of the generated fragment. Primers with the same letter designation were used in one PCR amplification. The size of the PCR fragments for a given primer pair can be obtained by subtracting the two nucleotide number designations. Of the seven resulting plasmids, then, three contain inserts that flank the *Hin*dIII site, and four contain inserts flanking the *Sma*I site.

RESULTS

Construction of auxotrophic strains of *C. glabrata*: To generate a *ura3* auxotroph of *C. glabrata*, a wild-type clinical isolate of *C. glabrata* [strain BG2, isolate B (Fidel *et al.* 1996)] was transformed with a linear DNA fragment consisting of the *neo* gene of Tn903 flanked by 5′ and 3′ *URA3* sequences. In four isolates (strains BG12, BG13, BG14, BG15), the *URA3* locus had been replaced by the the neomycin resistance cassette (data not shown). To generate a *his3* auxotroph of *C. glabrata*, the *HIS3* coding region was precisely deleted in a standard two-step disruption (Struhl 1983; details of strain construction in materials and methods). The structures of the *HIS3* and *URA3* loci in these strains (Figure 1a) were verified by Southern analysis (data not shown).

Nonhomologous recombination in *C. glabrata*: Consistent with what has been previously reported (Zhou *et al.* 1994), we found that a *S. cerevisiae CEN URA3* vector (pYCplac33, Gietz and Sugino 1988) transforms a *C.*

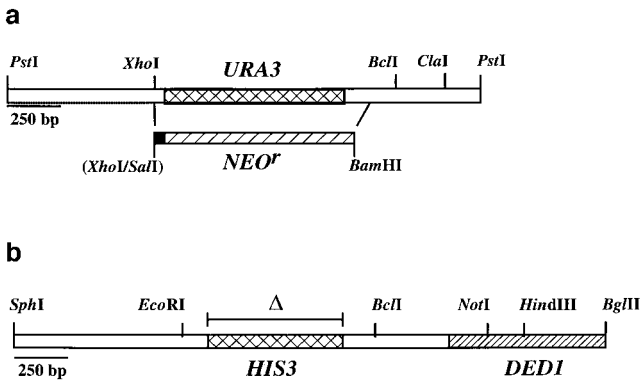


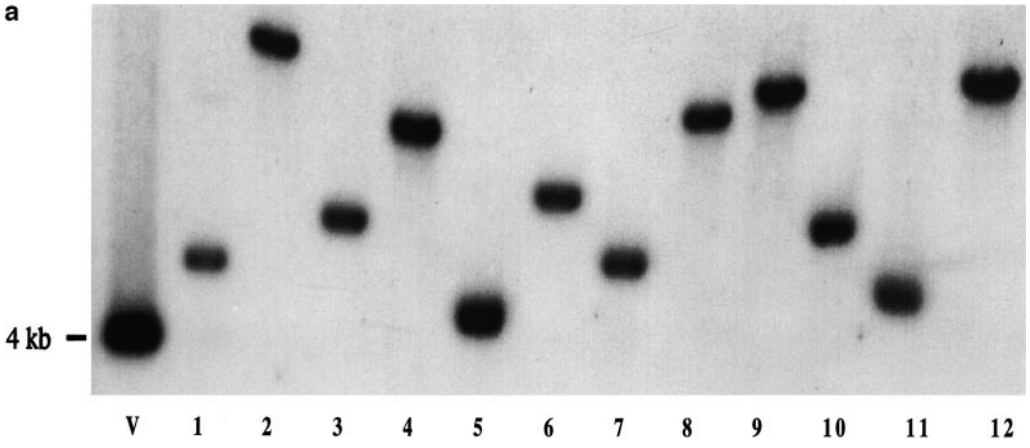
Figure 1.—Structure of disrupted loci. (a) Structure of disrupted *URA3* locus. The 2.2-kb *Pst*I fragment is shown. The *URA3* coding region is shown as a crosshatched box. The *neo* cassette used to disrupt the *URA3* gene is shown. The region of the *URA3* locus replaced by the *neo* cassette is indicated by connecting lines. The coding region for the *neo* gene is preceded by 54 bp of promoter from the *C. glabrata CUP1* gene (indicated by a filled box). (b) Structure of disrupted *HIS3* locus. The 2.8-kb *Sph*I-*Bgl*II fragment is shown with the *HIS3* and *DED1* ORFs indicated by hatched boxes. The region deleted in his3Δ strains is indicated.

glabrata ura3 mutant strain (BG14) at high efficiency ($2–5 \times 10^5/\mu g$) and replicates as a circular plasmid in these transformants (data not shown).

To examine efficiencies of illegitimate integative transformation, the *ura3* strain BG14 was transformed with the *S. cerevisiae URA3* integrating vector YIplac211, which has no homology with the *C. glabrata* genome. Before transformation, the vector was linearized with *Bam*HI or *Xba*I, enzymes that recognize polylinker sites in the plasmid and do not cleave within the *URA3* gene. Transformation by illegitimate recombination occurred with a frequency of $5 \times 10^2/\mu g$, compared to $1–5 \times 10^0/\mu g$ seen in *S. cerevisiae* (Schiestl *et al.* 1993). Genomic DNA from transformants was analyzed by Southern blot analysis probing with the *S. cerevisiae URA3* gene (Figure 2a). The sites of integration showed no obvious bias for a particular locus, and for most transformants only a single copy of the transforming vector had integrated. Of 59 total transformants analyzed, 45 (76%) were simple insertions, 9 (15%) were tandem insertions of two copies of the plasmid, and 4 (7%) were nontandem integrations of two copies of the plasmid. Ten of 4000 transformants (0.25%) were unable to grow on minimal media supplemented with only adenine and uracil. This frequency of amino acid auxotrophy is close to the predicted frequency (1–2%) for random integration if we assume 100 genes capable of mutating to auxotrophy. Several different auxotrophies were found (Figure 2b), again suggesting a fairly random distribution of the integrated plasmid in the genome.

To examine the target site specificity, we rescued the integrated plasmid from 37 integrants and determined the flanking DNA sequence. The 37 integrants sequenced included 2 colony morphology mutants, 6 auxotrophs, and 13 mutants with no apparent phenotype; in addition, we had identified 16 insertions that had lost the ability to adhere to eukaryotic cells (B. P. Cormack and S. Falkow, unpublished results). Only 2 of the insertions (both with no apparent phenotype) were in coding regions; the other 35 mutants apparently inserted in noncoding regions because only small ORFs of less than 250 nucleotides were present at the insertion site. None of these small ORFs had homology to genes in the *S. cerevisiae* database, on the basis of BLASTN or BLASTX homology searches (see materials and methods). The loci targeted in the mutants with adherence phenotypes were characterized in detail. Of the 16 mutants, 14 were insertions in the same 1.4-kb locus [the Epithelial Adhesin 1 (*EPA1*) gene] but were all independent insertion events because the precise insertion points were not the same. A physical map of the insertions is shown in Figure 3. The majority (13/14) of insertions form one clustered group ~1.2 kb upstream of the start site of translation. To assess whether there was any microhomology between the ends of the integrating plasmids and the genomic target site, we determined the DNA sequence of the undisrupted genomic target site in a total of 17 of the insertions. Of this total, 14 (1–14; Figure 4) were insertions at the *EPA1* locus; 3 others were insertions with no phenotype (15–17; Figure 4). For nine (53%) of the integration events, there was no homology between either end of the integrating plasmid and the site of integration. For eight (47%) of the integrations, there was microhomology between one end of the integrating plasmid and the target site.

Similar to *S. cerevisiae* (Schiestl *et al.* 1993), a small number of illegitimate events were not simple insertions. In one case (recombinant 10; Figure 4) the integration generated a target site deletion of 98 bp. In another (recombinant 12; Figure 4), there was an insertion of 125 bases of nonhomologous DNA at the locus



b

Auxotrophy	Number
Arg	1
His	1
Lys	3
Met/Cys	2
Trp	1
Val	1
Total/4000	10

Figure 2.—Fate of transformed YIplac211 in BG14. (a) Genomic DNAs from 12 transformants were prepared and digested with *Pst*I, which cleaves once in the integrating plasmid. Lane 1, YIplac211 digested with *Pst*I. Lanes 2–13, DNA from transformants digested with *Pst*I. The Southern blot was probed with [^{32}P]dCTP-labeled YIPlac 211. (b) Auxotrophs isolated among illegitimate integrants of YIPlac 211 in BG14. A total of 4000 transformants were tested by replica plating onto media lacking a given amino acid. Ten transformants were unable to grow without supplementation of the indicated amino acids.

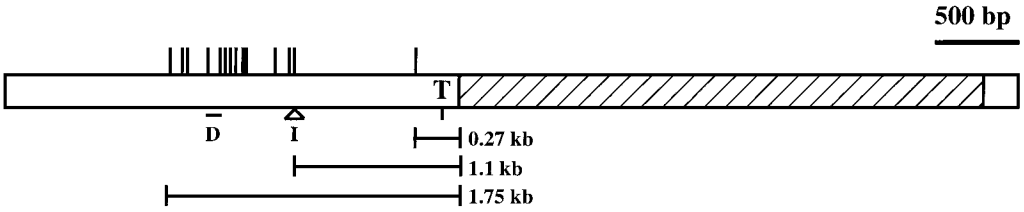


Figure 3.—Physical map of 14 insertions at the *EPA1* locus. Insertion sites are marked with vertical hatch marks. The open reading frame is shown as a hatched box. The insertions correspond to sequences 1–14 of Figure 4. The extent and position of the deleted fragment in transformant 10 is shown by a bar. The position and size of the insertion in transformant 12 is shown by the small triangle. A "T" marks a consensus TATA box.

immediately adjacent to the plasmid insertion. This inserted sequence had no significant homology to sequences in the database. In an additional four events (recombinants 2, 8, 10, and 14; Figure 4) a nonencoded A or T was added to the integrating end of the plasmid as part of the overall process of integration.

To show that the phenotype of a given strain was the result of the inserted plasmid, we chose one of the lysine auxotrophs, one of the mutants with an altered colony morphology, and all the mutants with a strong adherence phenotype and reintroduced the YIplac211 plasmid into the same locus in the parental strain BG14. To do this, we used the rescued plasmid (YIplac211 carrying the flanking genomic DNA for the insertion site) for each strain to transform the parental strain to Ura^+, which disrupts the gene by homologous recombination between the flanking DNA in the vector and the chromosomal DNA. Of such transformants tested, ~90% had the same phenotype as the original mutants, showing that the phenotype in these strains was indeed linked to the insertion of the plasmid. The remaining 10% are presumably due to the insertion of the plasmid by nonhomologous recombination, which is 10-fold less efficient than homologous recombination (see below).

Homologous and nonhomologous transformation: To characterize the sequence requirements for homologous and nonhomologous recombination in *C. glabrata*, a set of seven integrating plasmids was made carrying DNA from the Tn903 *neo* gene that we had used to disrupt the *URA3* locus to generate the strain BG14. Plasmids integrating by homologous recombination disrupted the *neo* gene rendering the strain sensitive to G418; on the other hand, plasmids integrating by nonhomologous recombination left the strain resistant to G418. The *neo* fragments ranged in size from 50–400 bp and were chosen so that in the middle of each fragment was either the naturally occurring *Sma*I site or the naturally occurring *Hin*dIII site. Each of these plasmids was linearized with *Hin*dIII or *Sma*I and used to transform the strain BG14. A simple comparison of the ratio of G418-resistant and G418-sensitive integrants for each plasmid gave the frequency of homologous *vs.* nonhomologous integration. The frequency of total, homologous, and nonhomologous integrants for each of the plasmids is shown in Figure 5. With <50 bases of homology, the frequency of homologous recombination is very low. With >200 bases of homology, total transformation efficiency increases dramatically and >90% of the integrants are homologous recombinants. We confirmed by PCR analysis (data not shown) that 20/20 neomycin-sensitive integrants were indeed integrants at the *ura3*::neoR locus, while 20/20 neomycin-resistant integrants were not integrants at the *ura3*::neoR locus.

DISCUSSION

C. glabrata is a significant cause of human disease in spite of the fact that it is phylogenetically closely related to the nonpathogenic *S. cerevisiae*. In this investigation, we used standard one- and two-step disruptions to generate *ura3* and *his3* auxotrophs in which the coding regions are deleted. These *ura3* and *his3* strains of *C. glabrata* should be useful in molecular-genetic analysis of multiple aspects of *C. glabrata* virulence, including generating prototrophic hybrids for use in parasexual analysis of *C. glabrata* after spheroplast fusion (Whelan and Kwon-Chung 1987).

C. glabrata can be transformed with circular plasmids carrying ARS elements from *S. cerevisiae* (Zhou *et al.* 1994) or from *C. glabrata* itself (Mehra *et al.* 1992); these plasmids replicate as unstable, high-copy extrachromosomal elements. Transformation of *C. glabrata* is extremely efficient: we found here that linearized plasmids not carrying an ARS element transformed *C. glabrata* at high efficiency, integrating by homologous recombination when the plasmid carried as little as 100 bp of DNA homologous with the genome. Plasmids carrying no *C. glabrata* genomic DNA integrated by nonhomologous recombination at a frequency ~10% of that seen for homologous integration. Thus, *C. glabrata* possesses efficient systems for homologous and nonhomologous recombination.

In some respects, illegitimate recombination in *C. glabrata* is similar to that in *S. cerevisiae*. Integration is relatively random in overall distribution in the genome, as suggested by Southern analysis, by the auxotroph frequencies among illegitimate recombinants, and by the sequence of target sites. No consensus site for integration was apparent with the possible exception of a slight preference for C (18/34) and T (18/34) as the first and second nucleotides 3′ to the integration site, respectively (Figure 4).

Figure 4.—Flanking sequences and target sites for *Bam*HI- and *Xba*I-digested YIplac211 transformed into BY14. Each target sequence is shown as the center double-stranded DNA. Above and below the target sequence is the sequence of the digested ends of YIplac211. The flanking sequences are aligned so that the single strand of the integrating end is immediately adjacent to the first base unambiguously derived from the target sequence. Underlined are nucleotides in the overhang of the integrating end that are homologous to the target sequence. The SED (single end directed), NED (neither end directed), or BED (both ends directed) classification is based how many ends of the integrating plasmid have microhomology with the target site. For the sake of consistency and comparison with the work in S. *cerevisiae* of Schiestl *et al.* (1994), we score as microhomologies those events involving at least 2 bp immediately adjacent to the recombination point. In bold are sequences immediately adjacent to the site of insertion that match the consensus for topoisomerase recognition [(G/C)(A/T)T]. In capitals are nucleotides present in the junction sequence that are not templated by target or plasmid, and which were presumably added during the recombination reaction. For recombinant 10, two target sites are shown because in this recombinant, the intervening 98 nucleotides between the two target sites were deleted. For recombinant 12, only one of the ends shown is plasmid derived, because there was a 125-bp fragment of unknown origin inserted in between the target and the second plasmid end.

Three lines of evidence suggest that illegitimate recombination in *C. glabrata* differs significantly from that in *S. cerevisiae.* First, our sequence of the target sites showed illegitimate recombination in *C. glabrata* has only a slight dependence on microhomology between plasmid ends and target site. For example, none of the illegitimate recombination events in *C. glabrata* had homology between both ends of the plasmid and the

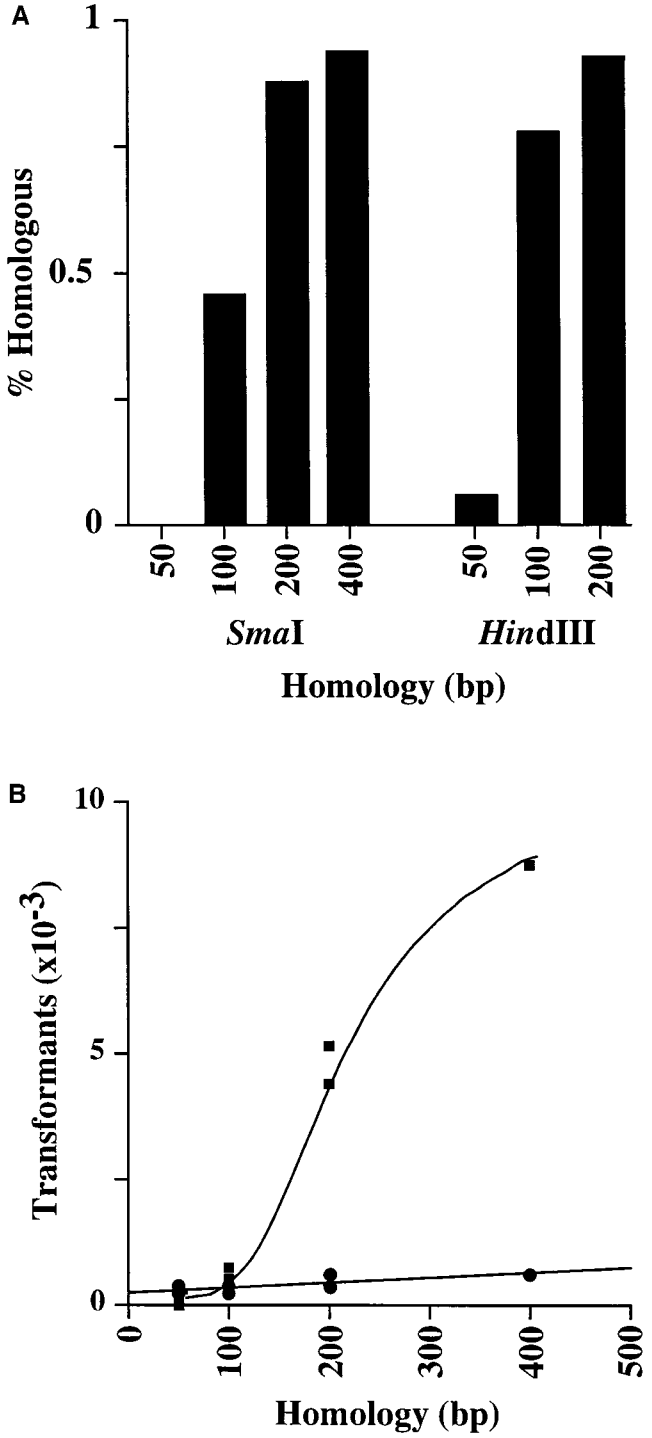


Figure 5.—Influence of amount of homology on rates of homologous and illegitimate recombination. (A) Number of homologous recombination events expressed as a percentage of total transformants plotted against the exact size (in base pairs) of the insert carried on the integrating plasmid. For details on insert size, see materials and methods. *Hin*dIII and *Sma*I refer to the enzyme used to linearize the plasmid in the precise middle of the *neo* gene fragment (see materials and methods). (B) Number of transformants (●, illegitimate recombinants; ■, homologous recombinants) per microgram of integrating plasmid plotted against size (in base pairs) of the insert carried on the integrating plasmid.

TABLE 3

Target sites for illegitimate recombination in *S. cerevisiae* and *C. glabrata*

	C. glabrata		*S. cerevisiae*	
	No. of events	Top1	No. of events	Top1
NED	9 (0.53)	1	6 (0.26)	5
SED	8 (0.47)	4	7 (0.30)	4
BED	0	—	10 (0.44)	2
Total	17	5	23	11

Classification of homology between the ends of the integrating plasmid and the target site in the genome. The events are classified according to the system of Schiestl *et al.* (1994). The SED (single end directed), NED (neither end directed), or BED (both ends directed) classification is based on how many ends of the integrating plasmid have homology with the target site. For the sake of consistency and comparison with the work of Schiestl *et al.* (1994) in *S. cerevisiae*, we count as homologous those events involving at least two base pairs immediately adjacent to the recombination point; a single nucleotide of homology (for example, insertion 1) is not counted as microhomology because this will occur on a purely random basis. The Top1 column summarizes the frequency of finding the Top1p consensus recognition site [(C/G)(A/T)T] immediately adjacent to the insertion point. The data for *S. cerevisiae*, shown for purposes of comparison taken from Figure 1A (Schiestl *et al.* 1994), Figure 5 (Schiestl and Petes 1991), Figure 1 (Schiestl *et al.* 1993), and Figure 4 (Schiestl *et al.* 1993).

target DNA, compared with 43% (10/23) in *S. cerevisiae*; 47% (8/17) of *C. glabrata* events had microhomology between one end and the target site compared to 74% (17/23) in *S. cerevisiae* (Schiestl and Petes 1991; Schiestl *et al.* 1993, 1994; Zhu and Schiestl 1996; Table 3). Indeed, in *C. glabrata*, the average number of homologous bases between the overhanging four bases of the integrating end and the target site was only 1.3, close to what is expected (1.0) on a purely random basis.

The second major difference with illegitimate recombination in *S. cerevisiae* is the role of topoisomerase I. In *S. cerevisiae*, toposiomerase I has been implicated in illegitimate recombination by several independent lines of evidence (Zhu and Schiestl 1996), including a higher-than-expected proportion of integration events immediately adjacent to consensus sites for topoisomerase I cleavage [consensus (C/G) (A/T) (T)]. By contrast, in *C. glabrata*, this bias for integration at Top1p sites was not observed. We do not know the consensus sequence for Top1p recognition in *C. glabrata*. However, (C/G) (A/T) (T) is the consensus site for the eukaryotic topoisomerase I in species ranging from *S. cerevisiae* (Zhu and Schiestl 1996) and *Tetrahymena pyriformis* (Andersen *et al.* 1985) to rats (Been *et al.* 1984). It is therefore likely that the Top1p site in *C. glabrata* falls within this consensus. For the 17 events analyzed, only 5 (15% of target sites) had adjacent (C/G) (A/T) (T) sites (Table 3). This is only marginally higher than the

number expected on a random basis (~2/17 target sites). In *S. cerevisiae* CTT is the primary site for Top1p action in mediating illegitimate recombination, because, in cells overexpressing Top1p, 12/16 consensus sequences adjacent to integration sites are CTT (Zhu and Schiestl 1996). In *C. glabrata*, by contrast, the 5 Top1p sites adjacent to insertion sites are distributed between all 4 possible sites (GTT, CTT, GAT, and CAT; Figure 4). While we cannot rule out that the presence of these 5 Top1p consensus sites reflects a modest role for *TOP1* in illegitimate recombination in *C. glabrata*, the representation of all 4 possible sites suggests that they are present fortuitously.

The third and most important difference with illegitimate recombination in *S. cerevisiae* is that illegitimate events in *C. glabrata* were targeted to noncoding regions. The ratio of coding and noncoding sequences in *C. glabrata* is probably similar to that found in *S. cerevisiae* because the genome size of *C. glabrata* is very similar to that of *S. cerevisiae* (Maleszka and Clark-Walker 1993). Moreover, for at least two loci, *HIS3–DED1* (Figure 1) and *CUP2-YGL164c*, the intergenic regions are approximately the same size and gene organization is conserved. Strikingly, for 37 integration events analyzed, all but 2 (which were in open reading frames homologous to *S. cerevisiae* genes YIL049 and YPL224C) were targeted to noncoding sequence. The 35 insertions in noncoding regions are likely in promoter regions. We have detailed information about three loci in which insertions afffected adherence to epithelial cells. The 14 insertions at the *EPA1* locus are between 0.3 and 1.75 kb upstream of the translational start (Figure 3). Strikingly, in screening for such insertion mutants, we did not isolate a single insertion in the coding region of the *EPA1* gene, even though such an insertion would result in a nonadherent phenotype (B. P. Cormack and S. Falkow, unpublished results). This emphasizes the bias of illegitimate recombination to noncoding regions. We also sequenced 2.5 kb of two other loci in which insertions had strong adherence phenotypes. In one case, the insertion was 250 bases upstream of a long (>1 kb) ORF. In the other case, the insertion was 0.5 kb upstream of a 500-bp ORF.

The bias of illegitimate recombination to noncoding regions in *C. glabrata* could result from several different mechanisms: (1) the recombination complex might be recruited to the site of assembly of interaction with specific host factors; (2) the targeted intergenic regions might be generally more accessible than coding regions of the genome. These same mechanisms have been proposed to explain targeting of members of the Ty retrotransposon family to regions upstream of RNA polymerase III genes or to regions of silenced chromatin (Chalker and Sandmeyer 1992; Ji *et al.* 1993; Zou and Voytas 1997; Kim *et al.* 1998). An analysis of illegitimate recombination in *C. glabrata* might reveal aspects of chromosome structure important in targeting the recombination machinery. Moreover, because insertions are exclusively or nearly exclusively targeted to intergenic regions, mutagenesis by illegitimate recombination may be a useful method of altering transcriptional regulation of endogenous genes.

Insertion by illegitimate recombination appears to be a useful method of general mutagenesis. The sequence of the integration sites and the variety of auxotrophic mutants that we recovered suggest a fairly random genomic distribution of insertions. Moreover, we were able to isolate insertions upstream of *EPA1*, a gene important in adherence to epithelial cells (B. P. Cormack and S. Falkow, unpublished results), on the basis of phenotype caused by the insertion. We demonstrated that, for strains showing a phenotype, the rescued plasmid/flanking DNA vector could be used to reintroduce the insertion at the original locus and regenerate a strain with the same phenotype. Because *C. glabrata* is asexual, putative mutants cannot be analyzed to see whether the insertion and the phenotype cosegregate in the spores resulting from a sexual cross. However, reintroduction of the plasmid into the same locus of the parental strain gives the same information and should be useful as a rapid second screen in *C. glabrata* mutant hunts.

We thank Dennis Thiele for the *C. glabrata URA3* gene and Paul Fidel and Jack Sobel for providing the original clinical *C. glabrata* isolate. We thank Jeff Corden, Carol Greider, Jef Boeke, and Lalita Ramakrishnan for comments on the manuscript. This work was supported by a grant to S.F. from SmithKline Beecham. B.P.C. was supported by a postdoctoral fellowship from the Helen Hay Whitney Foundation.

LITERATURE CITED

Andersen, A. H., E. Gocke, B. J. Bonven, O. F. Nielsen and O. Westergaard, 1985 Topoisomerase I has a strong binding preference for a conserved hexadecameric sequence in the promoter region of the rRNA gene from Tetrahymena pyriformis. Nucleic Acids Res. **13:** 1543–1557.

Been, M. D., R. R. Burgess and J. J. Champoux, 1984 Nucleotide sequence preference at rat liver and wheat germ type 1 DNA topoisomerase breakage sites in duplex SV40 DNA. Nucleic Acids Res. **12:** 3097–3114.

Chalker, D. L., and S. B. Sandmeyer, 1992 Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes Dev. **6:** 117–128.

Fidel, P. L., Jr., J. L. Cutright, L. Tait and J. D. Sobel, 1996 A murine model of Candida glabrata vaginitis. J. Infect. Dis. **173:** 425–431.

Gietz, R. D., and A. Sugino, 1988 New yeast-Escherichia coli shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. Gene **74:** 527–534.

Gietz, D., A. St. Jean, R. A. Woods and R. H. Schiestl, 1992 Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res. **20:** 1425.

Ji, H., D. P. Moore, M. A. Blomberg, L. T. Braiterman, D. F. Voytas *et al.*, 1993 Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. Cell **73:** 1007–1018.

Kim, J. M., S. Vanguri, J. D. Boeke, A. Gabriel and D. F. Voytas, 1998 Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. Genome Res. **8:** 464–478.

Kitada, K., E. Yamaguchi and M. Arisawa, 1995 Cloning of the

Candida glabrata TRP1 and HIS3 genes, and construction of their disruptant strains by sequential integrative transformation. Gene **165:** 203–206.

Kitada, K., E. Yamaguchi and M. Arisawa, 1996 Isolation of a Candida glabrata centromere and its use in construction of plasmid vectors. Gene **175:** 105–108.

Maleszka, R., and G. D. Clark-Walker, 1993 Yeasts have a four-fold variation in ribosomal DNA copy number. Yeast **9:** 53–58.

Mehra, R. K., J. L. Thorvaldsen, I. G. Macreadie and D. R. Winge, 1992 Cloning system for Candida glabrata using elements from the metallothionein-IIa-encoding gene that confer autonomous replication. Gene **113:** 119–124.

Odds, F. C., 1994 Pathogenesis of Candida infections. J. Am. Acad. Dermatol. **31:** S2–5.

Pfaller, M. A., 1996 Nosocomial candidiasis: emerging species, reservoirs, and modes of transmission. Clin. Infect. Dis. **22**(Suppl. 2): S89–94.

Redondo-Lopez, V., M. Lynch, C. Schmitt, R. Cook and J. D. Sobel, 1990 Torulopsis glabrata vaginitis: clinical aspects and susceptibility to antifungal agents. Obstet. Gynecol. **76:** 651–655.

Schiestl, R. H., and T. D. Petes, 1991 Integration of DNA fragments by illegitimate recombination in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA **88:** 7585–7589.

Schiestl, R. H., M. Dominska and T. D. Petes, 1993 Transformation of *Saccharomyces cerevisiae* with nonhomologous DNA: illegitimate integration of transforming DNA into yeast chromosomes and in vivo ligation of transforming DNA to mitochondrial DNA sequences. Mol. Cell. Biol. **13:** 2697–2705.

Schiestl, R. H., J. Zhu and T. D. Petes, 1994 Effect of mutations in genes affecting homologous recombination on restriction enzyme-mediated and illegitimate recombination in Saccharomyces cerevisiae. Mol. Cell. Biol. **14:** 4493–4500.

Sherman, F., G. R. Fink and J. B. Hicks, 1986 *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Struhl, K., 1983 The new yeast genetics. Nature **305:** 391–397.

Whelan, W. L., and K. J. Kwon-Chung, 1987 Parasexual genetics of *Torulopsis glabrata*. J. Bacteriol. **169:** 4991–4994.

Zhou, P., M. S. Szczypka, R. Young and D. J. Thiele, 1994 A system for gene cloning and manipulation in the yeast Candida glabrata. Gene **142:** 135–140.

Zhu, J., and R. H. Schiestl, 1996 Topoisomerase I involvement in illegitimate recombination in Saccharomyces cerevisiae. Mol. Cell. Biol. **16:** 1805–1812.

Zou, S., and D. F. Voytas, 1997 Silent chromatin determines target preference of the Saccharomyces retrotransposon Ty5. Proc. Natl. Acad. Sci. USA **94:** 7412–7416.

Communicating editor: M. Johnston