Yeast
Yeast 2002; **19**: 991–994.
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/yea.890

Yeast Sequencing Report

Genomic differences between *Candida glabrata* and *Saccharomyces cerevisiae* around the *MRPL28* and *GCN3* loci

David W. Walsh,[1] Kenneth H. Wolfe[2] and Geraldine Butler[1]*

[1] *Department of Biochemistry and Conway Institute of Biomolecular and Biomedical Science, University College Dublin, Belfield, Dublin 4, Ireland*
[2] *Department of Genetics, Smurfit Institute, University of Dublin, Trinity College, Dublin 2, Ireland*

**Correspondence to: Geraldine Butler, Department of Biochemistry, University College Dublin, Belfield, Dublin 4, Ireland. E-mail: gbutler@ucd.ie*

Received: 16 March 2002
Accepted: 10 May 2002

Abstract

We report the sequences of two genomic regions from the pathogenic yeast *Candida glabrata* and their comparison to *Saccharomyces cerevisiae*. A 3 kb region from *C. glabrata* was sequenced that contains homologues of the *S. cerevisiae* genes *TFB3, MRPL28* and *STP1*. The equivalent region in *S. cerevisiae* includes a fourth gene, *MFA1*, coding for mating factor a. The absence of *MFA1* is consistent with *C. glabrata*'s asexual life cycle, although we cannot exclude the possibility that a-factor gene(s) are located somewhere else in its genome. We also report the sequence of a 16 kb region from *C. glabrata* that contains a five-gene cluster similar to *S. cerevisiae* chromosome XI (including *GCN3*) followed by a four-gene cluster similar to chromosome XV (including *HIS3*). A small-scale rearrangement of gene order has occurred in the chromosome XI-like section. The sequences have been deposited in the GenBank database with Accession Nos AY083606 and AY083607.

Keywords: ***Candida glabrata*; *Saccharomyces cerevisiae*; gene order; mating pheromone**

Introduction

The yeast *Candida glabrata* has historically been considered as a commensal organism, and is part of the normal flora of healthy individuals. However, in recent years the incidence of infection caused by *C. glabrata* has greatly increased, particularly in immunocompromised patients. Although candidiasis is usually associated with *C. albicans*, recent reports have shown that *C. glabrata* is now the second or third most common cause, accounting for 12–20% of infections (Pfaller *et al*., 1999). In some US hospitals *C. glabrata* is now more frequently isolated from bloodstream infections than *C. albicans* (Berrouane *et al*., 1999). The increasing incidence of infection has been associated with widespread use of azole antifungal drugs (specifically fluconazole), as *C. glabrata* is inherently less susceptible than other *Candida* species. *C. glabrata*, like all *Candida* species, is an imperfect yeast lacking an apparent sexual cycle. However, while *C. albicans* and other related species are always diploid when isolated, *C. glabrata* is haploid (Whelan *et al*., 1984). *C. glabrata* is also much more closely related to *S. cerevisiae* and other members of the genus *Saccharomyces* family than it is to other *Candida* species (Cai *et al*., 1996). This suggests that *C. glabrata* may have lost the ability to mate relatively recently. To date, the available data from *C. glabrata* suggests that gene order and gene sequence are strongly conserved with *S. cerevisiae* (e.g. Nagahashi *et al*., 1998). Here we report two cases of disruption to conserved gene order, caused by probable gene loss in *C. glabrata (MFA1)*, and by a local rearrangement within a five-gene cluster near the *GCN3* locus.

Materials and methods

Plasmids pH1 and pH4, with overlapping inserts totalling 16.4 kb surrounding the *C. glabrata HIS3* locus (Kitada *et al*., 1995), were gifts from Dr K. Kitada. The region between *TFB3* and *STP1* was isolated on a 3.1 kb fragment from *C. glabrata* strain CBS138 by PCR. Degenerate oligonucleotide primers were designed using CODEHOP (Rose *et al*., 1998) from multiple alignments of proteins from several species. The primers used were 5′-ATTTGAAGATGCTTAAGTTGAAAAAGAR-GTNGAYRT-3′ (for *TFB3*) and 5′-AATAACCT-CTAATTCTAAATCTAGCATCACARTARTGR-CA-3′ (for *STP1*). The reaction was performed at an annealing temperature of 45 °C using a mixture of Taq and Pwo DNA polymerises (Expand, Roche Diagnostics). The resulting fragment was ligated into *Eco*RV-digested pBluescript to generate the plasmid pDW1. The DNA sequence of the pH1/pH4 and pDW1 inserts was determined commercially by Agowa (Berlin, Germany). ORFs were located using the NCBI ORF Finder (**www.ncbi.nlm.nih.gov**). Sequence alignments were performed using ClustalW (Thompson *et al*., 1996).

Results and discussion

The biochemical basis of the apparent mating defect in *C. glabrata* is not known, but if this species has been asexual for a significant evolutionary period, it is likely to have lost homologues of *S. cerevisiae* genes that function exclusively in mating. To investigate this, we searched for a *C. glabrata* locus homologous to *S. cerevisiae MFA1*. In *S. cerevisiae*, the mating pheromone **a**-factor is encoded by two duplicated genes, *MFA1* and *MFA2* (Brake *et al*., 1986). The pheromone genes have no known role outside of the mating process. We tried to isolate the *C. glabrata MFA1* locus by virtue of sequence conservation in neighbouring genes. Sequence data from multiple alignments with related proteins was used to design oligonucleotide primers from conserved parts of the genes *TFB3* and *STP1*, which flank *MFA1* and *MRPL28* on *S. cerevisiae* chromosome IV (Figure 1).

A 3.1 kb fragment of genomic DNA from *C. glabrata* was isolated by PCR as described in Materials and methods. Sequence analysis indicated that this region encodes two partial and one complete ORF (Figure 1, Table 1). One end of the fragment contains the 3′ end (234 residues) of a homologue of *TFB3* (component of TFIIH). This is followed by a long intergenic region of 1.2 kb with no large ORFs, and then a homologue of the mitochondrial ribosomal protein gene *MRPL28*

Table 1. Sequence identity between *C. glabrata* and *S. cerevisiae* open reading frames

	Identity %	
Open reading frame	**Protein**	**Nucleic acid**
CgTFB3*	68	65
CgMRPL28	52	53
CgYKR023W*	28	50
CgDBP7	63	67
CgRPC37	51	60
CgGCN3	81	73
CgYKR021W	30	29
CgHIS3	74	54
CgDED1	72	70
CgYOR205C	43	50
CgNOC2*	69	70

* Incomplete open reading frames.

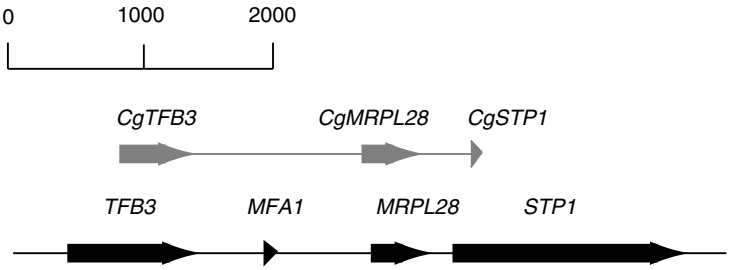

Figure 1. Comparison of the *TFB3–STP1* interval in *C. glabrata* and *S. cerevisiae*. The scale bar indicated the distance in base pairs. Only partial sequence is available for the *CgTFB3* and *CgSTP1* ORFs

(146 residues). The end of the fragment encodes a short partial ORF which is homologous to *STP1* (pre tRNA splicing). The similarity is clear when the sequence corresponding to the oligonucleotide used in the PCR reaction is included. The gene order in this region is identical with part of chromosome IV in *S. cerevisiae* (Figure 1), except that there is no equivalent of *MFA1* in *C. glabrata*.

The **a**-factor protein is small (36 residues) but the gene is well-conserved in *Saccharomyces castellii* and *Zygosaccharomyces rouxii* (71% and 65% identity, respectively; data from GenBank Accession Nos AZ927101 and AL394565; Cliften *et al*., 2001; de Montigny *et al*., 2000). As *Z. rouxii* is probably more distantly related to *S. cerevisiae* than is *C. glabrata* (Belloch *et al*., 2000), we should have been able to identify a *C. glabrata* homologue of *MFA1* if it were present in this part of the genome. The 1.2 kb spacer in *C. glabrata* contains several ORFs 30–40 codons in size, but none has significant sequence similarity to *MFA1* and none has strong codon bias like *MFA1*. Neither is a *MFA1* pseudogene present. We cannot, however, exclude the possibility that *C. glabrata* produces **a**-factor either from an *MFA2* locus, or from an *MFA1* gene that has transposed to somewhere else in the genome. Further analysis of the *C. glabrata* genome will be necessary to determine whether it has a cryptic sexual cycle, as has been proposed for *C. albicans* (Tzung *et al*., 2001). In *Z. rouxii*, the **a**-factor gene identified in Accession No. AL394565 is adjacent to a homologue of *YNL144C*, similar to *S. cerevisiae MFA2*. *Z. rouxii TFB3* and *MRPL28* genes are linked to each other (at the two ends of plasmid AR0AA004F02; de Montigny *et al*., 2000) but the region between them has not been sequenced so we do not know whether a *MFA1* homologue is present at the syntenic position in that species.

Our results show that apart from the loss of *MFA1* the order of genes in the *TFB3–STP1* region is co-linear in *C. glabrata* and *S. cerevisiae.* This is also true for almost all published examples from *C. glabrata* where the gene order is known. To test how widespread this conservation is, we analysed gene order in a larger (16 kb) region surrounding the *HIS3* gene in *C. glabrata*. The fragment contains nine partial or complete ORFs (Figure 2, Table 1). The first five are homologous to genes on *S. cerevisiae* chromosome XI. The fragment begins with a partial ORF encoding 51 amino acids from the C-terminal region of a protein with 28% identity to YKR023Wp (a protein of unknown function). This is followed by homologues of *DBP7* (a DEAD box RNA helicase involved in biogenesis of the 60S ribosomal subunit; 715 residues), *RPC37* (C37 subunit of RNA polymerase III; 241 residues) *GCN3* (α-subunit of translation initiation factor

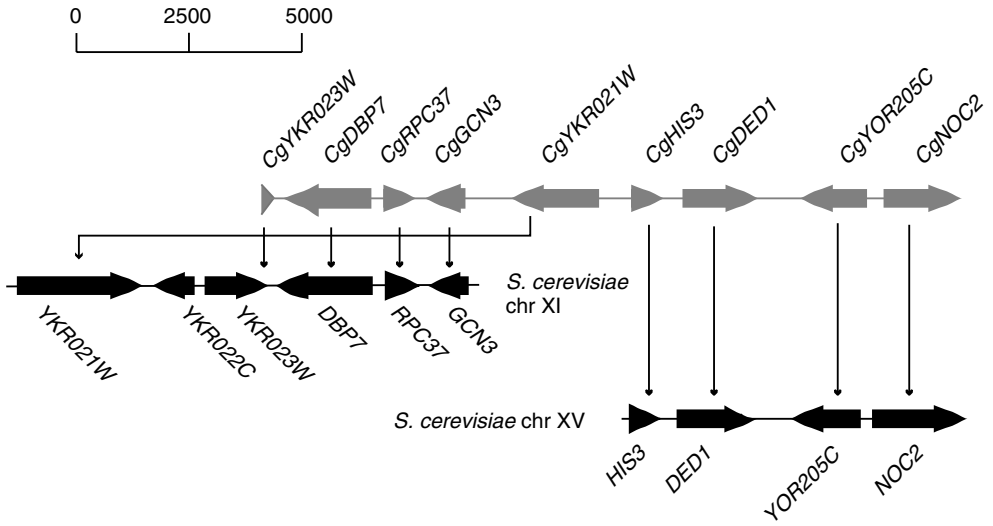

Figure 2. Comparison of a *C. glabrata* region containing *CgGCN3* and *CgHIS3* to parts of *S. cerevisiae* chromosomes XI and XV. The ORF *YOR203W* on *S. cerevisiae* chromosome XV, which overlaps both *HIS3* and *DED1*, is not shown because it is designated as a 'spurious ORF' by Wood *et al.* (2001) and as a 'questionable ORF' in the MIPS database

eIF2B; 305 residues) and *YKR021W* (unknown function; 694 residues). The first four genes are co-linear in *S. cerevisiae* and *C. glabrata* (Figure 2). *CgYKR021W*, however, is out of position and in inverted orientation. This was probably caused by either a short-range transposition of *CgYKR021W* or by inversion of a five-gene region (*YKR022W* to *GCN3*) in one of the species. The remaining genes are co-linear with part of chromosome XV of *S. cerevisiae*. These include the previously reported *CgHIS3* and *CgDED1* (Kitada, *et al*., 1995; Cormack and Falkow, 1999). These are followed by *CgYOR205C*, predicted to encode a protein of 526 amino acids with 43% identity to *S. cerevisiae YOR205C*, a gene of unknown function. The remainder of the fragment contains an incomplete ORF encoding 633 residues of CgNoc2p, with 69% identity to *S. cerevisiae* Noc2p, another protein involved in biogenesis of the 60S ribosome subunit.

Acknowledgements

We thank Dr K. Kitada for plasmids. This study was supported by the Health Research Board (to G.B.) and Science Foundation Ireland (to K.W.).

References

Belloch C, Querol A, Garcia MD, Barrio E. 2000. Phylogeny of the genus *Kluyveromyces* inferred from the mitochondrial cytochrome *c* oxidase II gene. *Int J Syst Evol Microbiol* **50**: 405–416.

Berrouane YF, Herwaldt LA, Pfaller MA. 1999. Trends in antifungal use and epidemiology of nosocomial yeast infections in a university hospital. *J Clin Microbiol* **37**: 531–537.

Brake A, Brenner C, Najarian R, Laybourn P, Merryweather J. 1986. Structure of genes encoding precursors of the yeast peptide mating pheromone **a**-Factor. In *Protein Transport and Secretion*, Gething MJ (ed.). Cold Spring Harbor Laboratory Press: New York; 103–108.

Cai J, Roberts IN, Collins MD. 1996. Phylogenetic relationships among members of the ascomycetous yeast genera *Brettanomyces*, *Debaryomyces, Dekkera* and *Kluyveromyces* deduced by small-subunit rRNA gene sequences. *Int J Syst Bacteriol* **46**: 542–549.

Cliften PF, Hillier LW, Fulton L *et al*. 2001. Surveying *Saccharomyces* genomes to identify functional elements by comparative DNA sequence analysis. *Genome Res* **11**: 1175–1186.

Cormack BP, Falkow S. 1999. Efficient homologous and illegitimate recombination in the opportunistic yeast pathogen *Candida glabrata*. *Genetics* **151**: 979–987.

de Montigny J, Straub M, Potier S *et al*. 2000. Genomic exploration of the hemiascomycetous yeasts: 8. *Zygosaccharomyces rouxii*. *FEBS Lett* **487**: 52–55.

Hanic-Joyce PJ, Joyce PBM. 1998. A high-copy-number *ADE2*-bearing plasmid for transformation of *Candida glabrata*. *Gene* **211**: 395–400.

Kitada K, Yamaguchi E, Arisawa M. 1995. Cloning of the *Candida glabrata TRP1* and *HIS3* genes, and construction of their disruptant strains by sequential integrative transformation. *Gene* **165**: 203–206.

Nagahashi S, Lussier M, Bussey H. 1998. Isolation of *Candida glabrata* homologues of the *Saccharomyces cerevisiae KRE9* and *KNH1* genes and their involvement in cell wall β-1,6-glucan synthesis. *J Bacteriol* **180**: 5020–5029.

Pfaller MA, Jones RN, Doern GV *et al*. 1999. International surveillance of blood stream infections due to *Candida* species in the European SENTRY Program: species distribution and antifungal susceptibility including the investigational triazole and echinocandin agents. SENTRY Participant Group (Europe). *Diagn Microb Infect Dis* **35**: 19–25.

Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S. 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res* **26**: 1628–1635.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **11**: 4673–4680.

Tzung KW, Williams RM, Scherer S *et al*. 2001. Genomic evidence for a complete sexual cycle in *Candida albicans*. *Proc Natl Acad Sci USA* **98**: 3249–3253.

Whelan WL, Simon S, Beneke ES, Rogers AL. 1984. Auxotrophic variants of *Torulopsis glabrata*. *FEMS Microbiol Lett* **24**: 1–4.

Wood V, Rutherford KM, Ivens A, Rajandream M-A, Barrell B. 2001. A re-annotation of the *Saccharomyces cerevisiae* genome. *Comp Funct Genom* **2**: 143–154.

