RESULTS & DISCUSSION
III. Identification of C. albicans TAF12 Isoforms and Functional complementation in S. cerevisiae

III.1. Computational Identification of TAF12 Isoforms in C. albicans

TBP associated factors, TAFs, are highly conserved over evolution, and there is experimental evidence that conserved regions of hTAF12 can replace yTAF12 deletion (Moqtaderi et al., 1996), and likewise several other TAFs can function in heterologous systems. With the availability of genome sequences of several fungi, we were interested to explore if the various S. cerevisiae TAF protein sequences are conserved in its entirety in the fungal genomes or are selectively conserved only in their functional domains. To identify the various TAF orthologous sequences, we first chose the C. albicans genome sequence (Jones et al., 2004). We carried out a BLASTP search of the NCBI non-redundant database using each of the S. cerevisiae TAF proteins as query. The search identified a single orthologous sequence for twelve out of the thirteen S. cerevisiae TAF proteins in C. albicans. Unexpectedly, the S. cerevisiae TAF12 protein identified two significant hits in C. albicans, orf19.470 (E-value 2.7e-41) and orf19.6820 (E-value 3.5e-33). Both these hits, when used in a reciprocal BLASTP search of the S. cerevisiae genome, identified the S. cerevisiae TAF12 protein YDR145W as the top scoring hit.

The TAF12 protein contains a H2B-like histone-fold (HF) domain that is highly conserved across all eukaryotes (Lee and Young, 1998). A search for conserved Pfam domains using the SMART tool showed the presence of a HF domain at the C-terminus of both orf19.470 (638-710 aa) and orf19.6820 (407-479 aa). A schematic diagram of TAF12 sequences indicates a similar location of the histone H2B-like HF domain in yTAF12, the C. albicans TAF12-like proteins and hTAF12 (Fig III.1). Pair wise comparison of the two C. albicans proteins with yTAF12 showed that the orf19.6820 sequence, which is of nearly the same size (515 aa) as the
yTAF12 protein (539 aa), had 43% similarity/25% identity to yTAF12 across the whole ORF, and 76% similarity/57% identity within the HF domain (data not shown). On the other hand, the orf19.470 sequence was much larger (750 aa), and had 47% similarity/30% identity to yTAF12 at the whole protein level and 91% similarity/68% identity within the HF domain (data not shown).

We also identified similarity between orf19.470 and yTAF12 sequences in their amino-terminal Gln-rich regions. The orf19.6820 sequence did not have a Gln-rich segment. The Gln-rich segment in yTAF12 extended from amino acid residues 116 to 309, and contained two blocks of tandem Gln repeats, 9 and 23 residues long (Fig III.1.). The Gln residues in yTAF12 protein comprised 29.9% of the 193-amino acid Gln-rich region. The Gln-rich region in orf19.470 however, was found to be greatly expanded, spanning nearly two-thirds of the protein (amino acid residues 6-478). It contained three blocks of tandem Gln repeats, two of them with 10 Gln each and a third with 13 residues. Overall, the Gln residues comprised 19.3% of the total protein and 30.7% of the residues in the Gln-rich region. However, the significance of the expanded glutamine stretch in orf19.470 is not clear, since the yTAF12 HF

![Fig. III.1. Schematic diagram showing the various domains in TAF12-like proteins.](image-url)
domain alone is sufficient for genetic complementation of a ytafl2A mutant (Moqtaderi et al., 1996).

III.2. Phylogenetic analysis of TAF12 Isoforms from Candida clade

We then extended our analysis to look for the occurrence of TAF12-like sequences in 22 publicly available genome sequences of the fungal subphylum Saccharomycotina, which included Saccharomyces, Kluyveromyces, Debaryomyces and that of Candida clades. For this search, we used BLASTP using the orf19.470 or orf19.6820 protein sequences as query and obtained 29 hits with E-values ≤1e-10, and indeed TAF12-like sequence was identified from all the genomes examined. Interestingly, we found that five CTG group genomes, C. dubliniensis, C. tropicalis, C. parapsilosis, Lodderomyces elongisporus and Pichia stipitis, in addition to C. albicans gave two significant hits for TAF12. However, only one TAF12 hit was identified in the three remaining CTG group genomes C. lusitaniae, C. guilliermondii and D. hansenii (Dujon). Additional BLASTP search of 40 other publicly available fungal genome sequences yielded only a single hit in each organism (data not shown).

Next we carried out detailed examination of the sequence conservation of the various hits. For this we imported the complete protein sequence of each of the 29 hits into the Jalview software. We also included TAF12 sequences from human, Drosophila, mouse, rice, Arabidopsis and the histone H2B sequences encoded by HTB1 and HTB2 from both S. cerevisiae and C. albicans for comparison of the histone-fold domains. We carried out multiple sequence analysis (MSA) for all the protein sequences using the MUSCLE software. When gaps were present at the same position in ~50% or more of the sequences, they were removed from all sequences and realigned as before. This operation was performed till no gaps were present in the MSA. Next the MSA was used to construct phylogenetic tree using neighbor joining
method in the MEGA 4 software with the option of pair wise deletion and 1000 bootstrap replicates. The phylogenetic tree revealed three major clusters; first, named yTAF12, comprising the TAF12 sequences from \textit{Saccharomyces} and \textit{Kluyveromyces} clades. The second cluster separated into two subclusters, CaTAF12a and CaTAF12b, occupied by TAF12 sequences from CTG group organisms. The rest of the TAF12 sequences from \textit{S. pombe}, \textit{Y. lipolytica}, \textit{A. thaliana}, rice, fly, mouse and human formed the third cluster (Fig III.2.). The histone H2B sequences formed an outgroup, consistent with their sequence divergence from the TAF12-like sequences. Interestingly the single TAF12-like sequences from \textit{C. lusitaniae}, \textit{D. hansenii} and \textit{C. guilliermondii} clustered with the CaTAF12a. Because orf19.470 and orf19.6820 showed primary sequence similarity to HFD of yTAF12 and clustered as two separate branches in the phylogenetic tree, we named the two proteins as CaTAF12 isoform \textit{a} and isoform \textit{b} respectively.

Next we carried out MSA of the TAF12 HF domains from all of the above sequences using MUSCLE software. For this, we trimmed the complete protein MSA to remove all residues, barring 26 residues at the amino-terminal and 35 residues carboxy-terminal to the aligned HFD (Fig. III.3.). The MSA showed that all the TAF12 HFD sequences belonging to Saccharomycotina are highly conserved, but the \textit{S. pombe}, and the higher eukaryotic TAF12 sequences contained a five residue deletion. We used the histone fold domain MSA to construct a phylogenetic tree using neighbor joining method as described above. The phylogenetic tree essentially gave the same branching pattern of the HFD sequences as obtained in the complete protein phylogeny (Fig III.4.).

We also carried out secondary structure prediction of the HF domains from the MSA in Jalview (ver 2.5.1.) (Clamp et al., 2004), and processed them through
Fig. III.2. Phylogenetic tree of the TAF12 whole ORFs sequences and histone H2B sequences.

An MSA of TAF12 whole ORF sequences from Saccharomyces species and from higher eukaryotes, and S. cerevisiae and C. albicans histone H2B sequences (total taxa = 38) were used to construct a phylogenetic tree using MEGA 4 software. The Neighbor-Joining method was used and the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. All positions containing alignment gaps and missing data were eliminated in pairwise sequence comparisons. The name of the clusters, indicated on the right, refers to prominent sequences in each cluster. Higher Euk. refers to the TAF12 sequences from higher eukaryotes.
Fig. III.3. MSA of histone fold domain and flanking sequences of TAF12 proteins from Saccharomycotina genomes.

Histone fold domains of \textit{S. cerevisiae} and \textit{C. albicans} H2B proteins were used for comparison and the alignment was prepared using MUSCLE software. Secondary structure prediction was performed for SACE\_TAF12 using Jnetpred in JALVIEW. The red boxes indicate the predicted \(\alpha\)-helices based on SACE\_TAF12 sequence. The histone fold domain consists of a central long helix \(\alpha_2\) flanked by short helices \(\alpha_1\) and \(\alpha_3\) connected by loops L1 and L2.
Fig. III.4. Phylogenetic tree of the TAF12 histone-fold domain sequences and histone H2B.

An MSA of TAF12 HFD plus flanking sequences from Saccharomycotina species and from higher eukaryotes, and the HFD of S. cerevisiae and C. albicans histone H2B sequences (total taxa = 38) were used to construct a phylogenetic tree using MEGA4 software. The Neighbor-Joining method was used and the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. All positions containing alignment gaps and missing data were eliminated in pairwise sequence comparisons. The name of the clusters, indicated on the right, refers to prominent sequences in each cluster. Higher Euk. refers to the TAF12 sequences from higher eukaryotes.
JNetPred (Cole et al., 2008). The JNetPred algorithm uses primary sequence as well as the related sequences of known secondary structure, if available, to predict secondary structures (Cole et al., 2008). Both the orf19.470 and orf19.6820 were predicted to form a histone fold domain-like secondary structure (data not shown) characterized by a central long helix (α2) flanked by two short helices (α1 and α3) on either side, similar to the yTAF12 (Fig. III.4.). We also used the PSIPRED protein structure prediction server for comparative 3D modeling using pGenTHREADER (Lobley et al., 2009). This fold recognition program uses the input amino acid sequence to generate profiles of sequence-structure alignments and predicts a potential 3D structure. Both the orf19.470 and orf19.6820 HF domains gave the histone-fold domain structure (PDB id: lh3o) as the only significant hit. Furthermore, we obtained pair-wise distances for S. cerevisiae TAF12 and C. albicans TAF12a and TAF12b sequences from MEGA 4 software. We inferred from this analysis that TAF12a was evolutionarily closer to yTAF12 when compared against HFD (distance = 0.541) and whole protein (distance = 1.245). In contrast, the TAF12b HFD (distance = 0.708) and the whole protein (distance = 1.432) were more diverged from yTAF12. Moreover, TAF12a and TAF12b were also found to be significantly diverged from each other. The evolutionary origin and divergence of the two TAF12 isoforms in Candida is not clear. But it is possible that they may have independent origins through a horizontal gene transfer event. Alternately, both may have evolved from a common ancestor through a duplication event, and further undergone sequence divergence. In any case, the sequence divergence likely indicates a functional specialization of the two TAF12 isoforms.
III.3. Genetic Complementation analysis in *S. cerevisiae*

To examine if both TAF12 isoforms can function in *S. cerevisiae*, we decided to carry out a genetic complementation test. The TAF12 protein was shown previously to be essential for the viability of *S. cerevisiae* (Moqtaderi et al., 1996). Thus, rescue of the *ytaf12Δ* lethal phenotype by expression of TAF12a and/or TAF12b proteins would indicate functional complementation. It was also shown that the histone-fold domain of yTAF12 is sufficient for complementation of the *ytaf12Δ* strain (Moqtaderi et al., 1996). We therefore aimed to look for complementation by expressing the TAF12a and TAF12b whole ORFs under the *P_{GAL1}*, a galactose regulatable promoter in *S. cerevisiae*.

For these analyses, we used a haploid *S. cerevisiae* strain, YSS19, carrying a chromosomal TAF12 deletion covered by the TAF12 gene on a *URA3*-based plasmid (Sawhney, S. and Natarajan, K., unpublished). We planned to transform the different *Leu2*-based complementation constructs into YSS19 and then shuffle out the *URA3*-based plasmid containing the wild type yTAF12 gene using the drug 5-Fluoroorotic acid (5-FOA). The orotidine-5'-phosphate decarboxylase encoded by the *URA3* gene converts 5-FOA into a toxic product, 5-Fluorouracil. Therefore, Ura⁺ cells are 5-FOA sensitive and only the cells that lose the *URA3*-plasmid would be able to grow (5-FOA resistant). However, because yTAF12 is an essential gene, the strains obtained after the shuffle having lost the wild type yTAF12 gene on the *URA3*-based plasmid would only be able to grow if the expression of the *C. albicans* TAF12a and/or TAF12b genes from the *LEU2*-based plasmids complemented the function of yTAF12.

III.4. Construction of *P_{GAL1}.TAF12a* and *P_{GAL1}.TAF12b* plasmids.

We PCR amplified the *C. albicans* TAF12a whole ORF (+1 to +2250bp) from the SC5314 genomic DNA using ONC26 and ONC27, designed to exclude the stop
codon and include BamHI and XhoI sites at the 5’ and 3’ ends respectively. The TAF12b full length ORF was similarly amplified (+1 to +1545bp) with ONC28 and ONC29, designed to exclude the stop codon and include SalI sites at both the 5’ and the 3’ ends. The two amplicons were first cloned into pRS423 cut with EcoRV by a blunt-end ligation. Next, the TAF12a ORF was popped out with BamHI and XhoI restriction enzymes and cloned downstream of the \(P_{GAL1}\) promoter in the regulatable expression vector pESC-Leu (Stratagene) cut with BamHI and SalI to obtain the plasmid pGAL1-470. The TAF12b ORF was popped out with SalI and cloned downstream of the \(P_{GAL1}\) promoter in pESC-Leu (Stratagene) cut with SalI to obtain the plasmid pGAL1-6820. These constructs were verified by sequencing and would express the TAF12a and TAF12b ORFs under the \(P_{GAL1}\) promoter in fusion with a single MYC epitope to enable visualization by western blot.

III.5. Construction of plasmids bearing TAF12a and TAF12b histone fold domain

We replaced the yTAF12 ORF in the plasmid pKN56 (Natarajan et al., 1998), containing the coding region of yTAF12 cloned along with 1454bp upstream and 774bp downstream regions (the genomic region corresponding to the sequence between the native EcoRI and ClaI sites) in a Leu2-based single copy plasmid pRS416, with the HF domain regions of TAF12a and TAF12b. The TAF12a HF domain region was PCR amplified using a gene-specific forward primer, ONC42 that amplified the ORF starting at +1741bp while the TAF12b HF domain region was amplified using ONC44 as the forward primer that annealed to +1043 bp using plasmids pGAL1-470 and pGAL1-6820 respectively as template. We used a common reverse primer ONC43 designed to include a PacI restriction site at the 3’ end that annealed in a vector region common to both the template plasmids, and allowed the amplification of the MYC epitope in frame with the HF domain regions.
The PCR amplified TAF12a and TAF12b HF domains were cut with PacI and cloned into Tth111I cut, blunted and PacI cut pKN56 to obtain plasmids pRS416-470HF and pRS416-6820HF. Finally, the TAF12a and TAF12b HF coding sequences along with the yTAF12 upstream and downstream sequences were subcloned as BamHI-BglII fragments into BamHI cut YEplac181, a high copy Leu2-based plasmid to obtain plasmids YEplac181-470HF and YEplac181-6820HF. The resulting plasmids would express the HF domains in the context of the native yTAF12 promoter and downstream regions. We also used pKN58 (Natarajan et al., 1998) containing an N-terminal truncated version of yTAF12 that expressed an 148 aa protein containing the HF domain as control for the complementation assay.

III.6. Histone fold domain of TAF12a but not that of TAF12b complements yeast taf12Δ lethality

We transformed the plasmids pKN56, pKN58, YEplac181-470HF, YEplac181-6820HF and YEplac181 (empty vector) into YSS19 as described in Section II.6. and obtained the strains ISC25, ISC26, ISC27, ISC28 and ISC24 respectively. Several transformants from each of the strains were then replica printed onto SC-Leu-Ura and SC-Leu+5-FOA plates, incubated at 25°C and growth was monitored over several days. All the strains were able to grow equally well on the control SC-Leu-Ura media (Fig III.5.A.). The negative control strain ISC24 bearing empty vector was unable to grow on the 5-FOA plate. Strains ISC25 and ISC26, containing plasmids pKN56 and pKN58 respectively acted as positive controls and grew well on the 5-FOA plate, because expression of either the yTAF12 whole ORF or the HF domain complements the chromosomal taf12 deletion. Strain ISC27, expressing the TAF12a HFD was able to grow on the 5-FOA plate, but ISC28 expressing the TAF12b HFD could not grow indicating that the HF domain of C. albicans TAF12a but not that of TAF12b could complement the ytaf12 deletion.
Fig. III.5. Complementation of *S. cerevisiae* taf12Δ strain by TAF12a and TAF12b.

(A) Growth phenotype of TAF12a and TAF12b HFD expressed under the native yTAF12 promoter. Indicated strains were tested by replica printing for growth upon shuffle of the native yTAF12 on Sc-Leu-Ura and Sc-Leu+FOA plates respectively. The plates were incubated at 25°C and imaged using the Alphalnnotech Gel Documentation system. The yTAF12 HFD and TAF12a HFD complement the ytaf12 deletion but the TAF12b HFD does not. (B) Growth phenotype of TAF12a and TAF12b whole ORFs expressed under the *Pgal* promoter under activating (Gal/Raf) and basal expression (Raf) conditions in the presence (Sc-Leu-Ura) or absence (Sc-Leu+FOA) of native yTAF12. The indicated strains were replica printed onto the abovementioned plates, incubated at 25°C and imaged using the Alphalnnotech Gel Documentation system. Neither TAF12a nor TAF12b whole proteins complement the ytaf12 deletion. However, overexpression of TAF12a was dominant negative over yTAF12 and resulted in slow growth.
III.7. The TAF12a and TAF12b full length ORFs do not complement the yeast taf12Δ phenotype

The $P_{GAL1}$ promoter is a carbon source regulatable promoter, being tightly repressed by glucose, highly inducible by galactose and expressing at basal levels in the presence of raffinose in the growth media (Johnston, 1987). This enabled us to look at the effects of basal versus overexpressed levels of the heterologous proteins within the same strain. We transformed the plasmids pESC-Leu, pESC-Leu-yTAF12, pGAL1-470 and pGAL1-6820 into S. cerevisiae strain YSS19 and obtained strains ISC20, ISC21, ISC22 and ISC23 respectively. Several transformants from each of the strains were then replica printed onto SC-Leu-Ura and SC-L+5-FOA media containing 2% raffinose alone (Raf) or 2% galactose + 2% raffinose (Gal/Raf) as carbon source. The plates were incubated at 25°C and growth was monitored over several days. The strain ISC20, with the empty vector grew well in Gal/Raf or Raf containing plates but was unable to grow on the 5-FOA containing plates and served as a negative control (Fig III.5.B). Strain ISC25 was able to grow equally well on all the four kinds of plates, since the yTAF12 ORF was expressed from its native promoter independent of the carbon source. The strain ISC21 expressing the yTAF12 gene under the $P_{GAL1}$ promoter was also able to grow on all the four plates, although the growth was slow on the 5-FOA plate with only raffinose as a carbon source. It is possible that the basal level of yTAF12 expression in this condition was not sufficient to support robust growth. Nevertheless, the strain was growing well on Gal/Raf plates indicating that the $P_{GAL1}$ promoter was working as expected. Neither the TAF12a nor the TAF12b proteins, expressed in strains ISC22 and ISC23 respectively, complemented the ytaf12Δ deletion, as indicated by no growth on the 5-FOA plates even after five days (Fig III.5.B). However, the overexpression of TAF12a in the absence of 5-FOA (SC-Leu-Ura +Gal/Raf) in a strain expressing native yTAF12
caused slow growth, suggesting that the TAF12a might be interfering with the function of the yTAF12 protein (Fig III.5.B).

III.8. Summary

Our sequence analysis identified two TAF12 isoforms in C. albicans, named TAF12a and TAF12b. Phylogenetic analysis showed that CaTAF12a is present in six of the CTG group organisms and is evolutionarily closer to yTAF12. The TAF12b protein was found to be restricted to only a subset of the fungal species. Heterologous expression analysis in S. cerevisiae indicated that the histone-fold domain of TAF12a but not of TAF12b was able to rescue the ytaf12Δ lethal phenotype. These results suggest that CaTAF12a might be the functional ortholog of TAF12. However, under galactose induction, neither of the full length proteins complemented the ytaf12Δ deletion. It is possible that although the HF domains are very highly conserved, the function of the whole proteins in C. albicans have diverged. Thus it would be interesting to carry out genetic analysis in C. albicans to determine TAF12a and TAF12b functions in vivo.