Duplex telomere-binding proteins in fungi with canonical telomere repeats: new lessons in the rapid evolution of telomere proteins

Authors: Neal F. Lue [1,2]

Affiliations: [1]Department of Microbiology & Immunology, W. R. Hearst Microbiology Research Center, [2]Sandra and Edward Meyer Cancer Center, Weill Medical College of Cornell University, 1300 York Avenue, New York, New York 10065, USA

Correspondence:
Neal F. Lue
Department of Microbiology and Immunology
W. R. Hearst Microbiology Research Center
Weill Medical College of Cornell University
1300 York Avenue
New York, NY 10065, USA
nflue@med.cornell.edu

Abstract

The telomere protein assemblies in different fungal lineages manifest quite profound structural and functional divergence, implying a high degree of flexibility and adaptability. Previous comparative analyses of fungal telomeres have focused on the role of telomere sequence alterations in promoting the evolution of corresponding proteins, particularly in budding and fission yeast. However, emerging evidence suggests that even in fungi with the canonical 6-bp telomere repeat unit, there are significant remodeling of the telomere assembly. Indeed, a new protein family can be recruited to serve dedicated telomere functions, and then experience subsequent loss in sub-branches of the clade. An especially interesting example is the Tay1 family of proteins, which emerged in fungi prior to the divergence of basidiomycetes from ascomycetes. This relatively recent protein family appears to have acquired its telomere DNA-binding activity through the modification of another Myb-containing protein. Members of the Tay1 family evidently underwent rather dramatic functional diversification, serving e.g., as transcription factors in fission yeast while acting to promote telomere maintenance in basidiomycetes and some hemi-ascomycetes. Remarkably, despite its distinct structural organization and evolutionary origin, a basidiomycete Tay1 appears to promote telomere replication using the same mechanism as mammalian TRF1, i.e., by recruiting and regulating Blm helicase activity. This apparent example of convergent evolution at the molecular level highlight the ability of telomere proteins to acquire new interaction targets. The remarkable evolutionary history of Tay1 illustrates the power of protein modularity and the facile acquisition of nucleic acid/protein-binding activity to promote telomere flexibility.

Introduction

Linear eukaryotic chromosome ends are stabilized by protein assemblies that organize the repetitive terminal DNA sequence (~5-20 base pairs per repeat unit) into protective structures that are resistant to aberrant degradation and recombination (1-3). The DNA component of this protective "cap", known as telomeres, usually consists of a duplex region of hundreds to thousands of nucleotides and a 3' overhang of tens to hundreds of nucleotides (also named the G-tail because of its G-rich nucleotide composition). Both the duplex region and the G-tail consist of the same short repeat unit, and both are bound by sequence-specific recognition proteins, which in turn recruit other proteins crucial for telomere protection. Collectively these proteins suppress the action of checkpoint and repair factors that can engender profound genomic instability.

Besides telomere protection, the other major function of telomere-bound proteins is to help preserve and replenish telomere DNAs. Despite their fundamental importance, telomere DNAs are subjected to progressive attrition owing to incomplete end replication (4, 5). Telomeres can also experience drastic truncation due to recombinational excision or replication fork collapse (4, 6). To compensate for such losses, eukaryotic cells employ telomerase and the primase-pol α complex to extend the G-tail and the complementary C-strand of telomeres, respectively (7-10). In addition, the cells are known to recruit a number of DNA helicases and repair proteins to overcome or alleviate problems arising from telomere replication fork stalling or collapse (11). Not surprisingly, these DNA maintenance pathways are under robust control by telomere-bound proteins in order to maintain telomere lengths within a size range that is optimal for telomere function.

Even though one might have imagined that the crucial importance of telomeres would make the nucleoprotein structures highly conserved in evolution, telomeres have in fact been subjected to rapid evolution, especially in selected clades. Nowhere is this malleability more evident than in the fungal phyla that include as their members some the most frequently employed model organisms. In both Saccharomycotina and Taphrinomycotina, which include *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, respectively, the telomere DNA repeat sequence are often irregular and variable, and they differ substantially from the canonical sequence 5'-TTAGGG-3'/5'-CCCTAA-3' (12, 13). While the underlying reasons for such telomere DNA sequence divergence remain obscure, it does highlight the adaptive capacity of fungal cells to stabilize the altered sequence at chromosome ends.

In contrast to the budding and fission yeasts, many other fungi in the Ascomycota phylum, including most of the filamentous fungi, have retained the canonical TTAGGG sequence. This telomere sequence is also widely conserved in the more basal branches of fungi such as Basidiomycota and Mucoromycota. To what extent the telomere assemblies in these "non-standard" fungi manifest structural and functional divergence is an open question. Indeed, while putative telomere-binding proteins can be readily identified in many such fungi, very few studies have experimentally interrogated the functions of these proteins. Despite this substantial knowledge gap, a few recent studies have begun to provide tantalizing hints of significant structural and functional divergence at basidiomycetes telomeres (14-16). In particular, it appears that even in the context of an invariant telomere repeat sequence, a new family of telomere DNA-binding protein can emerge and acquire telomere functions through the acquisition of new DNA sequence specificity and protein partners. It can also acquire non-telomeric functions, or be lost in some descendants. The potential of telomeres to evolve new regulatory mechanisms is thus not confined to scenarios that entail DNA sequence alterations.

In this focused review, I will first provide a very brief overview of Ascomycota telomere variability, highlighting the well characterized, co-evolving telomere DNA sequence and recognition proteins in this phylum. This will be followed by a more in-depth discussion of recent works on telomere regulations in *Ustilago maydis*, a member of the Basidiomycota phylum. A special emphasis will be on the *U. maydis* Tay1 protein, which appears to belong to a relatively new protein family in fungal evolution. Members of Tay1 are confined to Ascomycota and Basidiomycota, and while they all bind the 5'-TTAGGG-3'/5'-CCCTAA-3' repeat unit with high affinity and sequence specificity, these proteins evidently mediate distinct telomeric and non-telomeric functions in different fungi. Notably, the *U. maydis* Tay1 protein, despite being structurally different from mammalian TRF1 (a major double strand telomere binding protein), exhibits surprising mechanistic and functional similarities to this mammalian protein. The origin of Tay1 and the implications of Tay1 diversity for the malleability and adaptability of telomeres are discussed.

Organisms with canonical and variant telomere repeats: recognition of double-stranded telomeres by distinct Myb-containing proteins

The most prevalent telomere repeat unit and possibly the most ancient is 5'-TTAGGG-3'/5'-CCCTAA-3', which is found in various fungi, plant, metazoans, and protozoa. In organisms with this telomere repeat unit, the duplex region is typically recognized directly by a member of the TRF protein family, whereas the G-tail is bound by an OB-fold protein named POT1 (1, 17). In most mammalian cells, two structurally similar TRF homologs (TRF1 and TRF2) play partially overlapping and non-redundant functions in telomere protection and telomere maintenance (1). Deleting or depleting TRFs often triggers significant telomere length alterations as well as structural abnormalities. TRF homologs are bi-partite proteins that consist of an N-terminal TRFH domain responsible for dimerization and a C-terminal Myb motif responsible for DNA-binding (Fig. 1a). TRFs also employs multiple surface features within and outside the TRFH domain to interact with partners that regulate telomere functions (18, 19). For example, mammalian TRF1 is thought to utilize a basic patch (located in between its TRFH and Myb domain) to bind and recruit BLM helicase, which in turn promotes the complete replication of telomeres by unwinding G-rich replication barriers (20-22). Similarly, mammalian TRF2 has been shown to recruit another helicase (RTEL1) as well as replisome proteins (Claspin, DONSON, etc.) to facilitate telomere replication (17, 23-25). However, unlike TRF1, TRF2 is responsible for a key protective function of telomeres by virtue of its ability to suppress telomere-telomere fusions.

Notably, the TRF proteins possess a variant of the Myb motif that is highly specific for the canonical TTAGGG sequence (26). Hence, in budding and fission yeasts, where the telomere repeat units are made up of different sequences, the closest TRF homologs cannot recognize the telomere repeats with high affinity and do not function as the main duplex telomere-binding proteins. Instead, budding and fission yeasts utilize two other Myb-containing proteins, named Rap1 and Taz1, to coat and protect their respective telomeres (12, 13, 27-29). In contrast to Rap1 and Taz1, the most TRF-like genes in budding and fission yeasts (often named Tbf1) provide important functions not at telomeres, but at subtelomeres or elsewhere in the genome. *S. cerevisiae* Tbf1, for example, regulates subtelomere structure and function, while also acting as a transcription factor upstream of various snoRNA genes (30-32). In addition, *Sc*Tbf1 has been shown to regulate double strand break repair (33). These diverse and non-telomeric functions of budding yeast Tbf1 support the notion that this protein shares a common ancestry with TRFs, but has evolved and maintained functions away from the telomere terminal repeats due to the divergence of telomere sequences in this organism. Together these observations on Rap1, Taz1 and Tbf1 underscore one specific mechanism that promotes the remodeling of the telomere nucleoprotein complex.

Fungi with TTAGGG repeats: how prone are they to telomere nucleoprotein remodeling?

As illustrated in the preceding section, the flexibility of the fungal telomere complex is evident in sub-phyla that experienced substantial telomere sequence divergence. Many fungi in the more basal clades, however, have retained the canonical, 6-bp telomere repeat. An interesting question, then, is whether the telomere proteins in these clades might exhibit less flexibility and less remodeling. While many genomes in these clades have been sequenced and putative telomere-binding proteins can be readily identified (see e.g., (34)), there are as yet, very few studies that experimentally interrogate the functions of these proteins. The best studied organism in this regard is *Ustilago maydis,* a plant pathogen that forms corn galls. This yeast-like fungus was developed by Robin Holliday several decades ago as a model system for studying recombinational repair (35). It belongs to the Basidiomycota phylum, members of which also include the human pathogen *Cryptococcus neoformans*. As a model for telomere research, *U. maydis* offers a number of advantages beyond the standard budding and fission yeast model, including (i) its greater resemblance to the mammalian system with respect to the recombination and repair machinery (35) , and (ii) its retention of the same telomere repeat as the mammalian repeat (36, 37). Interestingly, with regard to shelterin-like telomere-binding proteins in the *U. maydis* genome, several initial surveys revealed a putative Pot1 ortholog, but nothing resembling the mammalian TRF proteins (34, 38). Instead, a protein bearing consecutive Myb motifs near its N-terminus (named *Um*Tay1 or *Um*Trf1) was postulated to be the most plausible candidate for binding the double-stranded region of telomeres (15, 38). Notably, in addition to having N-terminal Myb motifs, *Um*Tay1 differs from a standard TRF homolog in having a much larger size (~150 kD) and in lacking a TRFH dimerization domain (Fig. 1a).

These initial conjectures notwithstanding, a TRF/TBF-like gene (named *Um*Tbf1 or *Um*Trf2) was subsequently uncovered in *U. maydis*, suggesting that this fungus may harbor two structurally distinct duplex telomere binding proteins (13, 16). To simplify the discussion, I will henceforth refer to these two proteins as *Um*Tay1 and *Um*Trf2, respectively.

Experimental interrogation of *Um*Tay1 and *Um*Trf2 confirmed the roles of both proteins in telomere regulation and revealed an interesting division of labor that is different from other systems (16). *Um*Tay1, on the one hand, plays minimal roles in telomere protection — deletion of the gene triggers neither growth defects nor telomere structural abnormalities. Further analysis of *Umtay1Δ* revealed preferential loss of long telomeres and the suppression of telomere recombination in the context of *ku70* transcriptional repression, i.e., Ku70 depletion. These phenotypes are reminiscent of *Ustilago maydis* mutants with a knockout of *blm*, a conserved helicase with functions in both DNA repair and telomere regulation. Previous studies have shown that the *Um*Blm helicase positively stimulates both telomere replication and telomere recombination (15, 39). Indeed, *Um*Tay1 physically interacts with Blm and modulates Blm helicase activity in a substrate sequence-dependent manner (16). Thus, by regulating the Blm helicase activity, *Um*Tay1 appears to perform similar functions in these two pathways. In contrast to *Um*Tay1, *Um*Trf2 plays a pivotal role in telomere protection: deletion of *trf2* is lethal, and transcriptional repression of this gene triggers a constellation of telomere aberrations that are indicative of de-protection, including telomere length heterogeneity, accumulation of ssDNA and extra-chromosomal telomere repeats (16). Therefore, despite having similar affinity and binding specificity for double-stranded telomere repeats, *Um*Tay1 and *Um*Trf2 mediate largely non-overlapping functions in telomere regulation. Notably, this division of labor is somewhat different from that in mammals, where two structurally similar proteins (i.e., TRF1 and TRF2) execute broadly related functions. For example, while TRF1 is known to play a critical and preferential function in telomere replication by recruiting BLM helicase to unwind G-rich structural barriers (20, 22), TRF2 also makes notable contributions by interacting with the RTEL1 helicase as well as replisome-associated proteins (18, 24, 25, 40). In addition, even though TRF2 appears to be the main mediator of telomere protection by suppressing ATM activation and telomere fusions, TRF1 also contributes to chromosome end protection (1, 41). The structural and functional differences between the duplex telomere binding proteins in fungi and mammals indicate that the retention of the same telomere repeat does not preclude the evolution of new telomere regulatory factors or functional shuffling among proteins capable of binding telomere repeats.

The evolutionary origin of the Tay1 protein family

Given its structural resemblance to TRF/TBF, *Um*Trf2 most likely shares a common ancestry with prototypical TTAGGG-binding proteins, and has inherited its telomere functions from the same ancestor as TRF1/2. The origin of Tay1 is less clear. Bioinformatic analysis revealed closely-related family members with tandem Myb motifs in the Basidiomycota and Ascomycota phyla, but not in more basal lineages such as Mucoromycota and Glomeromycota (Fig. 1b). Thus, Tay1 may have originated through gene duplication and modification from a similar protein in the common ancestor of all of these fungal branches. In other words, the antecedent of Tay1 may be a structurally similar protein shared by Basidiomycota, Ascomycota, and the more basal phyla.

Based on this rationale, I performed psi-BLAST screening of Tay1-like proteins in Glomeromycota and Mucoromycota (the closest fungal phyla that occupy a more basal position than Basidiomycota and Ascomycota), and analyzed statistically significant hits with respect to (i) the presence of tandem Myb domains; and (ii) the conservation of putative TTAGGG repeat-binding residues (16, 26). Two protein families with the highest sequence similarity scores were uncovered in psi-BLAST and were named TayL1 (Tay1-like 1) and TayL2, respectively (Fig. 1b). Members of these two families are variably distributed in the two analyzed fungal clades. Glomeromycota, for example, harbors both TayL1 and TayL2, whereas Mucorales appears to harbor only TayL2. TayL1, a small protein that comprises just two copies of the Myb domain, is the prime candidate for being the closest relative of Tay1, given their notable sequence similarities (~35% identity, ~50% similarity, and ~5% gaps in the aligned Myb domain region) (Fig. 1b and 1c). TayL1 also appears to share almost all amino acids residues in TRF that are implicated in binding the TTAGGG repeats (Supp. Fig. 1), supporting its ability to recognize this sequence. Whether TayL1 actually executes a telomere function is an interesting question for future investigation. In contrast, TayL2, a much

larger protein (~550-700 amino acids) with extra-Myb regions, appears to be more distantly related to Tay1 (~17% identity, ~30% similarity, and ~15% gaps in the aligned Myb domain region), and its lack of several putative TTAGGG-binding residues renders its potential for a telomere function less plausible (Supp. Fig. 1). Indeed, a couple of TayL2-like proteins (UMAG 04101 and UMAG 10544), distinct from *Um*Tay1, can also be discerned in the genome of *Ustilago maydis* and those of other Basidiomycota/Ascomycota species, indicating that this represents a distinct protein family with an ancient evolutionary history. (Some of the TayL2 entries in the database are annotated as Bas1-like because of their similarities to the transcription factor Bas1 in *S. cerevisiae*. However, *S. cerevisia*e Bas1 has much weaker sequence similarity to *Um*Tay1 than the TayL2s identified in the psi-BLAST analysis.) It is also worth noting while psi-BLAST identified two Myb motifs in TayL2 that align well to the DNA-binding region of Tay1, the annotations of multiple TayL2s in the databases postulate the existence of a third Myb motif, which is not present in Tay1 and which further underscores the greater evolutionary distance between these two protein families (Fig. 1a).

Taken together, the distribution of Tay1-like proteins can be used to construct a parsimonious model for their evolutionary kinship, as follows (Fig. 1d). Prior to the divergence of Mucoromycota from Basidiomycota and Ascomycota, Tay1* (common ancestor of Tay1 and TayL1) emerged as a TTAGGG repeat binding protein, possibly through duplication and evolutionary tinkering of another protein with tandem Myb domains (e.g., TayL2). Subsequently, in some sub-lineages of Basidiomycota and Ascomycota, this protein acquired telomere function through its ability to localize to telomeres and interact with other telomere regulators such as Blm. However, in most Mucoromycota species, this gene duplication and neofunctionalization never transpired, and TAY1/TAYL1 either remained a protein with minimal telomere function, or was lost from the genome. It is worth noting that an underlying factor that enables this evolutionary scenario is the modular nature of proteins, which in this case allows the ancient Tay1 to acquire the necessary DNA-binding and telomere-regulatory activities in a step-wise fashion.

Discussion: implications for the evolution of Tay1 and protein-protein interactions at telomeres

The foregoing discussion argues for a relatively recent origin for Tay1 through alteration of another protein that did not possess telomere-binding activity or telomere function. As such, Tay1 emerged in an organism where the telomere functions are presumably well served by a TRF-like protein (e.g., the ancestor of UmTrf2). Any telomere function acquired by Tay1 in this context was likely to be non-essential or redundant – as evident from the phenotypes of the *Umtay1Δ* mutant. These conjectures have ramifications for the subsequent evolution of Tay1 in basidiomycetes and ascomycetes. In particular, given that there was no strong selection pressure for the telomere function(s) of ancient Tay1, this protein would have been relatively unconstrained in adopting other cellular functions. Indeed, the *S. pombe* Tay1 homolog, also known as *Sp*Teb1, has been shown to regulate transcription rather than telomeres (42). The ancient Tay1 could also be lost from the genome without great detriment –unless, of course, it had somehow acquired a more critical function. Indeed, no Tay1 homolog can be readily identified in most of the Saccharomycotina yeasts, including the *Kluyveromyces*, *Saccharomyces*, and *Candida* species. As noted before, these yeasts have experience drastic telomere sequence alterations that eliminated the canonical TTAGGG repeats from chromosome ends. Hence any Tay1 that was present in the ancestor of these yeasts would have been unable to remain telomere-bound in the descendants. Unless this Tay1 has somehow acquired critical non-telomeric functions, there would be little selection pressure for its retention. An interesting exception to this evolutionary scenario is *Yarrowia lipolytica*, an early branching Saccharomycotina yeast. Tay1 appears to be essential in this yeast, and deleting just one TAY1 allele in a diploid strain causes drastic telomere shortening, supporting an important function in telomere maintenance (14). The distinct fate of *Yl*Tay1 can be understood in light of the mild telomere sequence alteration that transpired in this yeast; while the telomere repeat in *Yarrowia lipolytica* deviated from the canonical repeat (TTAGTCAGGG rather than TTAGGG), the retention of the GGGTTA core element recognized by the Myb motif allowed *Yl*Tay1 to remain telomere-bound and to perform its telomere maintenance function (14). Thus, the presence or absence of Tay1, as well as its divergent functions in different fungi, can be largely rationalized by its evolutionary origin and its DNA recognition property.

Given that *Um*Tay1 has a fundamentally different structure and evolutionary origin from mammalian TRF1, it is perhaps surprising that these two proteins share the same molecular partner (Blm) and the same telomere

function (promoting replication). This apparent instance of convergent evolution at the molecular level suggests that telomere proteins may be quite adapt at evolving new protein-protein interactions. The greater propensity of telomere proteins to acquire interaction partners is also consistent with the growing list of proteins shown to bind shelterin subunits, especially TRF1 and TRF2. Both TRF1 and TRF2 contain within its TRFH domain a surface groove capable of binding a short peptide motif (TBM; Y/F/H-X-L-X-P) (19). An impressive list of DNA repair and replisome proteins have been shown to carry this motif, and to make functionally important interaction with TRF1 or TRF2 (18, 19, 40, 43). Notably, this motif is not reliably conserved in evolution in mammals (e.g., see for example (40)), suggesting that individual TRF-target interaction can be gained or lost quite recently. This again echoes the notion that telomere proteins are quite facile in evolving new interaction partners. Could there be some unique feature of telomeres that facilitate this? One possibility is the repetitive nature of the telomere sequence, which results in the clustering (and increased local concentration) of telomere proteins. In this setting, even a mutation that results in low affinity binding to a novel partner may be sufficient to support enough complex formation to result in a selectable phenotype. This is similar to a previous proposal that emphasizes the power of protein co-localization to drive evolutionary changes (44). Thus it would not be surprising if future studies of telomere regulation in different organisms were to uncover more examples of convergent evolution at the molecular level. A possible theme is that the key players in telomere regulation (e.g., Blm) will turn out to be well conserved in evolution but the molecular interactions through which their functions are executed may not be.

Acknowledgement

I thank Bill Holloman for introducing me to the excellent world of "Usti". Works on *Ustilago maydis* in my laboratory is supported by grants from NSF (MCB1817331) and NIH (GM107287).

References

1. T. de Lange: Shelterin-Mediated Telomere Protection. *Annu Rev Genet*, 52, 223-247 (2018) doi:10.1146/annurev-genet-032918-021921

2. D. Jain and J. P. Cooper: Telomeric strategies: means to an end. *Annu Rev Genet*, 44, 243-69 (2011) doi:10.1146/annurev-genet-102108-134841

3. R. J. O'Sullivan and J. Karlseder: Telomeres: protecting chromosomes against genome instability. *Nat Rev Mol Cell Biol*, 11(3), 171-81 (2010) doi:nrm2848 [pii]
10.1038/nrm2848

4. P. M. Lansdorp: Major cutbacks at chromosome ends. *Trends Biochem Sci*, 30(7), 388-95 (2005) doi:10.1016/j.tibs.2005.05.004

5. A. M. Olovnikov: Telomeres, telomerase, and aging: origin of the theory. *Experimental Gerontology*, 31, 443-448 (1996)

6. A. Lustig: Clues to catastrophic telomere loss in mammals from yeast telomere rapid deletion. *Nat Rev Genet*, 4(11), p916-23 (2003)

7. V. Pfeiffer and J. Lingner: Replication of telomeres and the regulation of telomerase. *Cold Spring Harb Perspect Biol*, 5(5), a010405 (2013) doi:10.1101/cshperspect.a010405

8. E. H. Blackburn and K. Collins: Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harb Perspect Biol*, 3(5) (2011) doi:cshperspect.a003558 [pii]

10.1101/cshperspect.a003558

9. C. Autexier and N. F. Lue: The Structure And Function of Telomerase Reverse Transcriptase. *Annu Rev Biochem*, 75, 493-517 (2006)

10. N. F. Lue, J. Chan, W. E. Wright and J. Hurwitz: The CDC13-STN1-TEN1 complex stimulates Pol alpha activity by promoting RNA priming and primase-to-polymerase switch. *Nat Commun*, 5, 5762 (2014) doi:10.1038/ncomms6762

11. P. Martinez and M. A. Blasco: Replicating through telomeres: a means to an end. *Trends Biochem Sci*, 40(9), 504-15 (2015) doi:10.1016/j.tibs.2015.06.003

12. O. Steinberg-Neifach and N. F. Lue: Telomere DNA recognition in Saccharomycotina yeast: potential lessons for the co-evolution of ssDNA and dsDNA-binding proteins and their target sites. *Front Genet*, 6, 162 (2015) doi:10.3389/fgene.2015.00162

13. R. Sepsiova, I. Necasova, S. Willcox, K. Prochazkova, P. Gorilak, J. Nosek, C. Hofr, J. D. Griffith and L. Tomaska: Evolution of Telomeres in Schizosaccharomyces pombe and Its Possible Relationship to the Diversification of Telomere Binding Proteins. *PLoS One*, 11(4), e0154225 (2016) doi:10.1371/journal.pone.0154225

14. J. Kramara, S. Willcox, S. Gunisova, S. Kinsky, J. Nosek, J. D. Griffith and L. Tomaska: Tay1 protein, a novel telomere binding factor from Yarrowia lipolytica. *J Biol Chem*, 285(49), 38078-92 (2010) doi:10.1074/jbc.M110.127605

15. E. Y. Yu, M. Hsu, W. K. Holloman and N. F. Lue: Contributions of recombination and repair proteins to telomere maintenance in telomerase-positive and negative Ustilago maydis. *Mol Microbiol*, 107(1), 81-93 (2018) doi:10.1111/mmi.13866

16. E. Y. Yu, S. S. Zahid, S. Ganduri, J. H. Sutherland, M. Hsu, W. K. Holloman and N. F. Lue: Structurally distinct telomere-binding proteins in Ustilago maydis execute non-overlapping functions in telomere replication, recombination, and protection. *Commun Biol*, 3(1), 777 (2020) doi:10.1038/s42003-020-01505-z

17. A. Cicconi and S. Chang: Shelterin and the replisome: at the intersection of telomere repair and replication. *Curr Opin Genet Dev*, 60, 77-84 (2020) doi:10.1016/j.gde.2020.02.016

18. H. Kim, O. H. Lee, H. Xin, L. Y. Chen, J. Qin, H. K. Chae, S. Y. Lin, A. Safari, D. Liu and Z. Songyang: TRF2 functions as a protein hub and regulates telomere maintenance by recognizing specific peptide motifs. *Nat Struct Mol Biol*, 16(4), 372-9 (2009) doi:10.1038/nsmb.1575

19. Y. Chen, Y. Yang, M. van Overbeek, J. R. Donigian, P. Baciu, T. de Lange and M. Lei: A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. *Science*, 319(5866), 1092-6 (2008) doi:10.1126/science.1151804

20. M. Zimmermann, T. Kibe, S. Kabir and T. de Lange: TRF1 negotiates TTAGGG repeat-associated replication problems by recruiting the BLM helicase and the TPP1/POT1 repressor of ATR signaling. *Genes Dev*, 28(22), 2477-91 (2014) doi:10.1101/gad.251611.114

21. K. Lillard-Wetherell, A. Machwe, G. T. Langland, K. A. Combs, G. K. Behbehani, S. A. Schonberg, J. German, J. J. Turchi, D. K. Orren and J. Groden: Association and regulation of the BLM helicase by the telomere proteins TRF1 and TRF2. *Hum Mol Genet*, 13(17), 1919-32 (2004) doi:10.1093/hmg/ddh193

22. A. Sfeir, S. T. Kosiyatrakul, D. Hockemeyer, S. L. MacRae, J. Karlseder, C. L. Schildkraut and T. de Lange: Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell*, 138(1), 90-103 (2009) doi:10.1016/j.cell.2009.06.021

23. G. Sarek, P. Kotsantis, P. Ruis, D. Van Ly, P. Margalef, V. Borel, X. F. Zheng, H. R. Flynn, A. P. Snijders, D. Chowdhury, A. J. Cesare and S. J. Boulton: CDK phosphorylation of TRF2 controls t-loop dynamics during the cell cycle. *Nature*, 575(7783), 523-527 (2019) doi:10.1038/s41586-019-1744-8

24. G. Sarek, J. B. Vannier, S. Panier, J. H. J. Petrini and S. J. Boulton: TRF2 recruits RTEL1 to telomeres in S phase to promote t-loop unwinding. *Mol Cell*, 57(4), 622-635 (2015) doi:10.1016/j.molcel.2014.12.024

25. R. Rai, P. Gu, C. Broton, C. Kumar-Sinha, Y. Chen and S. Chang: The Replisome Mediates A-NHEJ Repair of Telomeres Lacking POT1-TPP1 Independently of MRN Function. *Cell Rep*, 29(11), 3708-3725 e5 (2019) doi:10.1016/j.celrep.2019.11.012

26. R. Court, L. Chapman, L. Fairall and D. Rhodes: How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures. *EMBO Rep*, 6(1), 39-45 (2005) doi:10.1038/sj.embor.7400314

27. J. P. Cooper, E. R. Nimmo, R. C. Allshire and T. R. Cech: Regulation of telomere length and function by a Myb-domain protein in fission yeast. *Nature*, 385(6618), 744-7 (1997) doi:10.1038/385744a0

28. A. J. Lustig, S. Kurtz and D. Shore: Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length. *Science*, 250(4980), 549-53 (1990) doi:10.1126/science.2237406

29. A. R. Buchman, N. F. Lue and R. D. Kornberg: Connections between transcriptional activators, silencers, and telomeres as revealed by functional analysis of a yeast DNA-binding protein. *Mol Cell Biol*, 8(12), 5086-99 (1988) doi:10.1128/mcb.8.12.5086

30. M. Preti, C. Ribeyre, C. Pascali, M. C. Bosio, B. Cortelazzi, J. Rougemont, E. Guarnera, F. Naef, D. Shore and G. Dieci: The telomere-binding protein Tbf1 demarcates snoRNA gene promoters in Saccharomyces cerevisiae. *Mol Cell*, 38(4), 614-20 (2010) doi:10.1016/j.molcel.2010.04.016

31. C. E. Koering, G. Fourel, E. Binet-Brasselet, T. Laroche, F. Klein and E. Gilson: Identification of high affinity Tbf1p-binding sites within the budding yeast genome. *Nucleic Acids Res*, 28(13), 2519-26 (2000) doi:10.1093/nar/28.13.2519

32. G. Fourel, E. Revardel, C. E. Koering and E. Gilson: Cohabitation of insulators and silencing elements in yeast subtelomeric regions. *EMBO J*, 18(9), 2522-37 (1999) doi:10.1093/emboj/18.9.2522

33. D. Bonetti, S. Anbalagan, G. Lucchini, M. Clerici and M. P. Longhese: Tbf1 and Vid22 promote resection and non-homologous end joining of DNA double-strand break ends. *EMBO J*, 32(2), 275-89 (2013) doi:10.1038/emboj.2012.327

34. P. Sanchez-Alonso and P. Guzman: Predicted elements of telomere organization and function in Ustilago maydis. *Fungal Genet Biol*, 45 Suppl 1, S54-62 (2008) doi:S1087-1845(08)00069-8 [pii] 10.1016/j.fgb.2008.04.009

35. W. K. Holloman, J. Schirawski and R. Holliday: The homologous recombination system of Ustilago maydis. *Fungal Genet Biol*, 45 Suppl 1, S31-9 (2008) doi:10.1016/j.fgb.2008.04.006

36. P. A. Guzman and J. G. Sanchez: Characterization of telomeric regions from Ustilago maydis. *Microbiology*, 140 (Pt 3), 551-7 (1994)

37. P. Sanchez-Alonso and P. Guzman: Organization of chromosome ends in Ustilago maydis. RecQ-like helicase motifs at telomeric regions. *Genetics*, 148(3), 1043-54 (1998)

38. E. Y. Yu, M. Kojic, W. K. Holloman and N. F. Lue: Brh2 and Rad51 promote telomere maintenance in Ustilago maydis, a new model system of DNA repair proteins at telomeres. *DNA Repair (Amst)*, 12(7), 472-9 (2013) doi:S1568-7864(13)00121-3 [pii] 10.1016/j.dnarep.2013.04.027

39. E. Y. Yu, J. Perez-Martin, W. K. Holloman and N. F. Lue: Mre11 and Blm-Dependent Formation of ALT-Like Telomeres in Ku-Deficient Ustilago maydis. *PLoS Genet*, 11(10), e1005570 (2015) doi:10.1371/journal.pgen.1005570

40. A. Cicconi, R. Rai, X. Xiong, C. Broton, A. Al-Hiyasat, C. Hu, S. Dong, W. Sun, J. Garbarino, R. S. Bindra, C. Schildkraut, Y. Chen and S. Chang: Microcephalin 1/BRIT1-TRF2 interaction promotes telomere replication and repair, linking telomere dysfunction to primary microcephaly. *Nat Commun*, 11(1), 5861 (2020) doi:10.1038/s41467-020-19674-0

41. Y. W. Lee, R. Arora, H. Wischnewski and C. M. Azzalin: TRF1 participates in chromosome end protection by averting TRF2-dependent telomeric R loops. *Nat Struct Mol Biol*, 25(2), 147-153 (2018) doi:10.1038/s41594-017-0021-5

42. L. P. Valente, P. M. Dehe, M. Klutstein, S. Aligianni, S. Watt, J. Bahler and J. P. Cooper: Myb-domain protein Teb1 controls histone levels and centromere assembly in fission yeast. *EMBO J*, 32(3), 450-60 (2013) doi:10.1038/emboj.2012.339

43. B. Wan, J. Yin, K. Horvath, J. Sarkar, Y. Chen, J. Wu, K. Wan, J. Lu, P. Gu, E. Y. Yu, N. F. Lue, S. Chang, Y. Liu and M. Lei: SLX4 assembles a telomere maintenance toolkit by bridging multiple endonucleases with telomeres. *Cell Rep*, 4(5), 861-9 (2013) doi:10.1016/j.celrep.2013.08.017

44. J. Kuriyan and D. Eisenberg: The origin of protein interactions and allostery in colocalization. *Nature*, 450(7172), 983-90 (2007) doi:10.1038/nature06524

Figure Legend

Figure 1. *Duplex telomere binding proteins and their evolution in fungi*

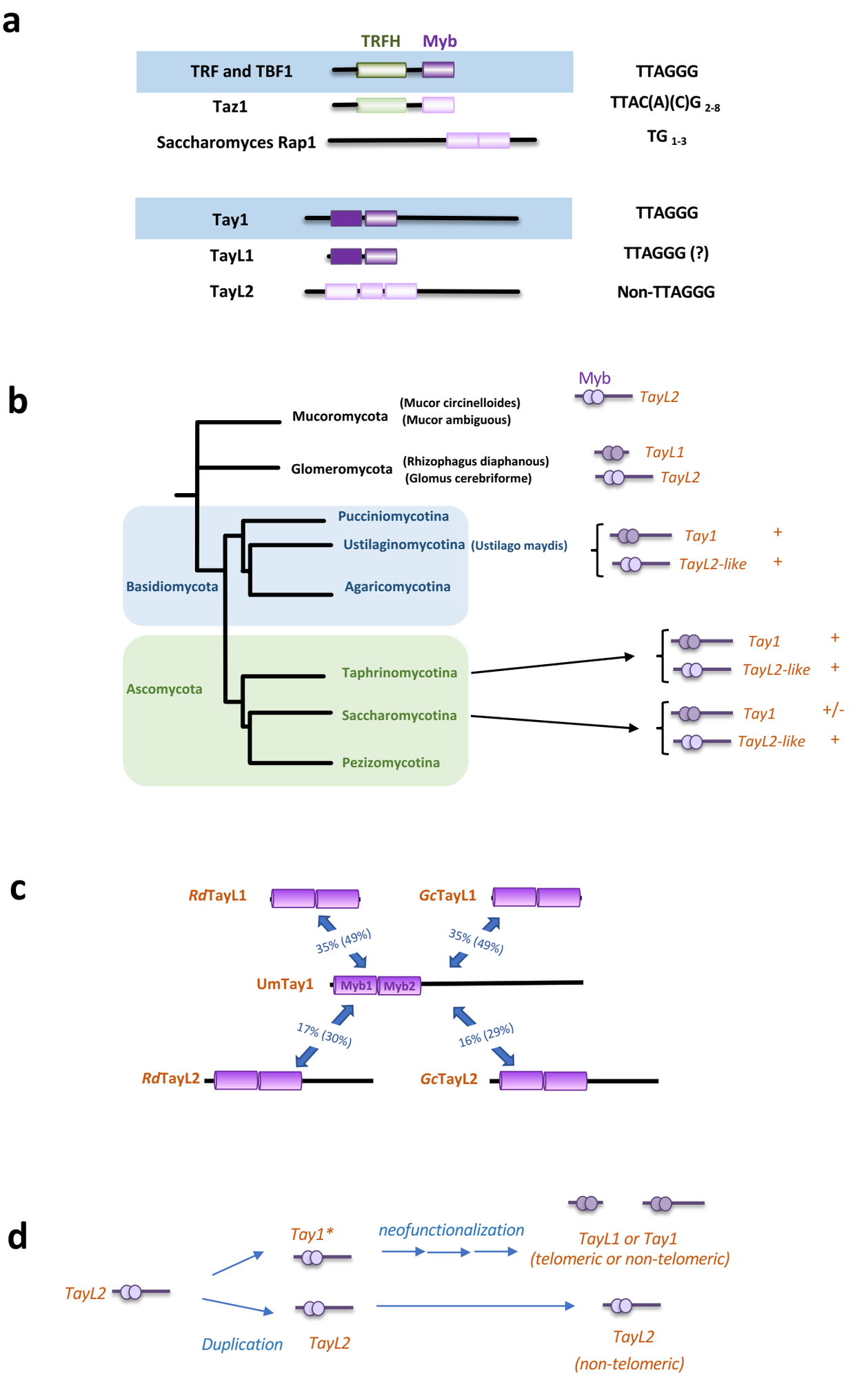
a. *The distinctive structures of duplex telomere binding proteins and their homologs* (top) The domain structures of the major duplex telomere-binding proteins in vertebrates (TRF1 and TRF2), fission yeast (Taz1), and budding yeast (Rap1) are displayed. Both TRF1 and TRF2 consist of an N-terminal dimerization and protein-protein interaction domain (TRFH) and a C-terminal DNA binding motif (Myb). Taz1 has a similar domain structure as TRF1/2, but with a variant TRFH and a variant Myb domain. Rap1 contains two variant Myb motifs near its C-terminus. The sequences recognized by these proteins are shown to the right. (bottom) The domain structures Tay1, TayL1, and TayL2 are illustrated. Tay1 is a fungus-specific telomere-binding protein with two consecutive Myb motifs neat its N-terminus. These two Myb motifs exhibit strong similarities to vertebrates TRF1 and TRF2, and specifically bind the TTAGGG repeats. TayL1 resembles Tay1 but does not contain the C-terminal, non-Myb region found in Tay1. TayL2 has a similar domain structure as Tay1, but carries more divergent Myb motifs that are not predicted to bind TTAGGG. In some annotations of TayL2 homologs, this protein is postulated to contain a third Myb motif in between the two motifs that align to the Tay1 protein. The sequences recognized or presumed to be recognized by these proteins are shown to the right.

b. *The distribution of Tay1, TayL1 and TayL2 in fungi* The distributions of Tay1, TayL1, and TayL2 in various fungal phyla are displayed. TayL2 is widely disseminated in Basidiomycota and Ascomycota, and is present in neighboring fungal branches including Glomeromycota and Mucoromycota. Tay1 is also widely disseminated in Basidiomycota and Ascomycota, but often absent in Saccharomycotina. TayL1 is apparently restricted to Glomeromycota.

c. *The degrees of sequence similarities between theTay1, TayL1 and TayL2 protein families* The extent of sequence identity and similarity (in parenthesis) between the Myb motifs of *Um*Tay1 and various TayL1 and TayL2 homologs are displayed. *Rd* and *Gc* designate *Rhizophagus diaphanous* and *Glomus cerebriforme*, respectively. Both belong to the Glomeromycota phylum.

d. *Model for the origin and subsequent evolution of Tay1* A common ancestor of Tay1 and TayL1 (Tay1*, with high affinity and sequence specificity for TTAGGG) is hypothesized to emerge through gene duplication and modification of a Tay1-like gene such as TayL2. Subsequently, Tay1* may evolve into either the miniaturized TayL1 in Glormeromycota, or Tay1 in Basidiomycota and Ascomycota, and acquire either telomeric or non-telomeric functions. See main text for more details.

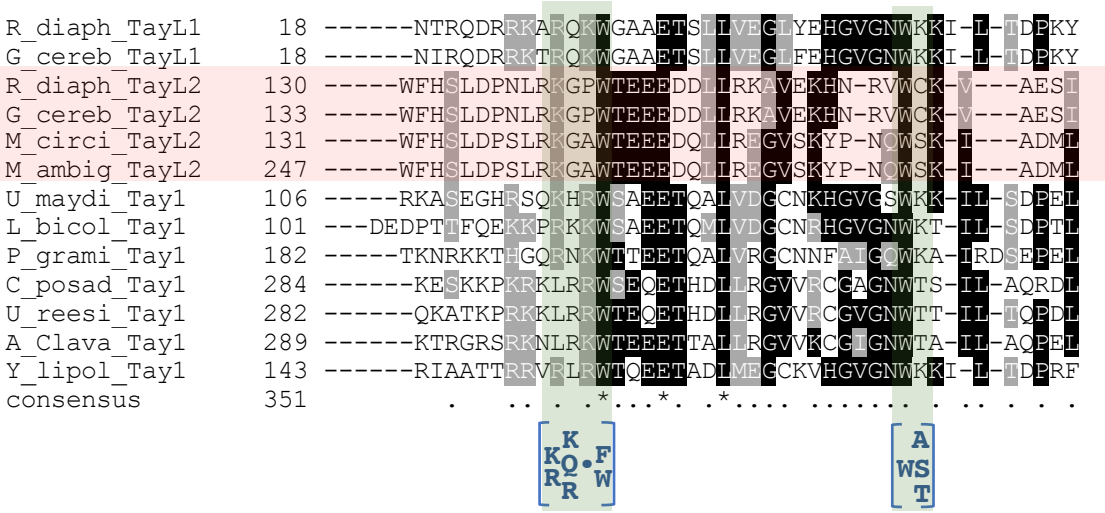
Fig. 1


a
TRFH
Myb
TRF and TBF1
TTAGGG
Taz1
TTAC(A)(C)G 2-8
Saccharomyces Rap1
TG 1-3
Tay1
TTAGGG
TayL1
TTAGGG (?)
TayL2
Non-TTAGGG
b
Myb
TayL2
Mucoromycota
(Mucor circinelloides)
(Mucor ambiguous)
Glomeromycota
(Rhizophagus diaphanous)
(Glomus cerebriforme)
TayL1
TayL2
Basidiomycota
Pucciniomycotina
Ustilaginomycotina
(Ustilago maydis)
Agaricomycotina
Tay1 +
TayL2-like +
Ascomycota
Taphrinomycotina
Saccharomycotina
Pezizomycotina
Tay1 +
TayL2-like +
Tay1 +/-
TayL2-like +
c
RdTayL1
GcTayL1
35% (49%)
35% (49%)
UmTay1
Myb1
Myb2
17% (30%)
16% (29%)
RdTayL2
GcTayL2
d
TayL2
Tay1*
neofunctionalization
TayL1 or Tay1
(telomeric or non-telomeric)
Duplication
TayL2
TayL2
(non-telomeric)

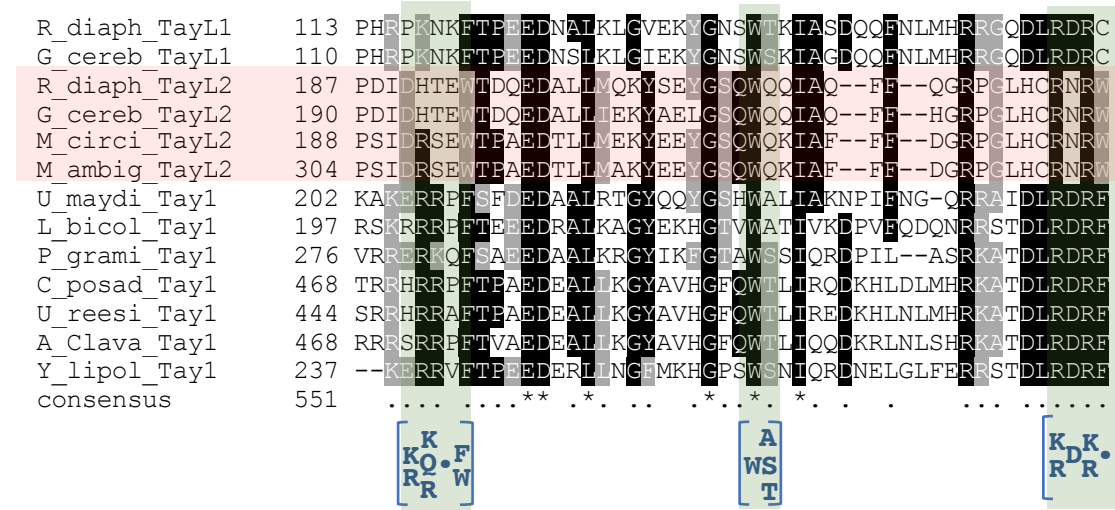
Supp. Fig. 1

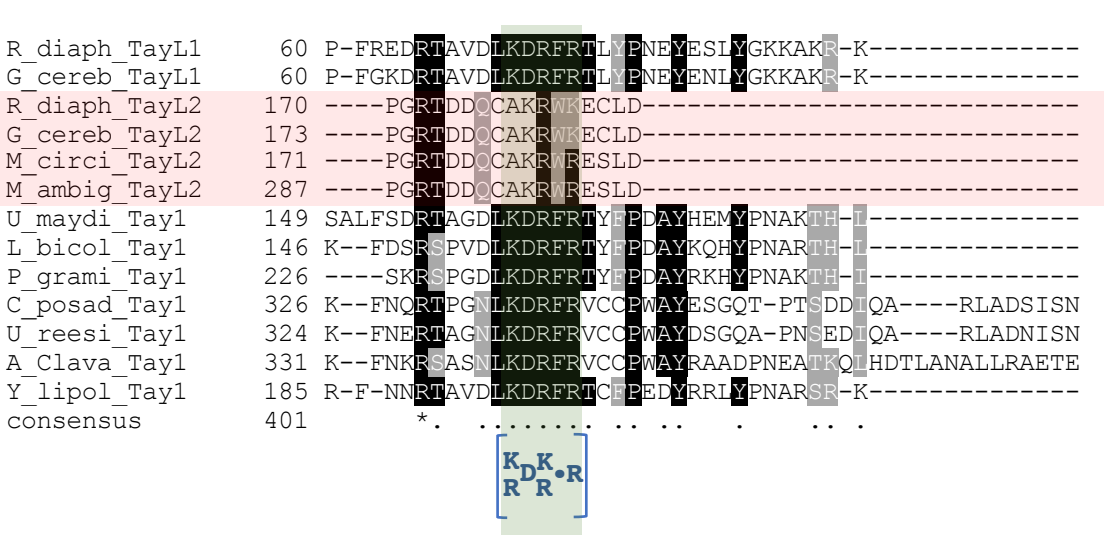
Lue

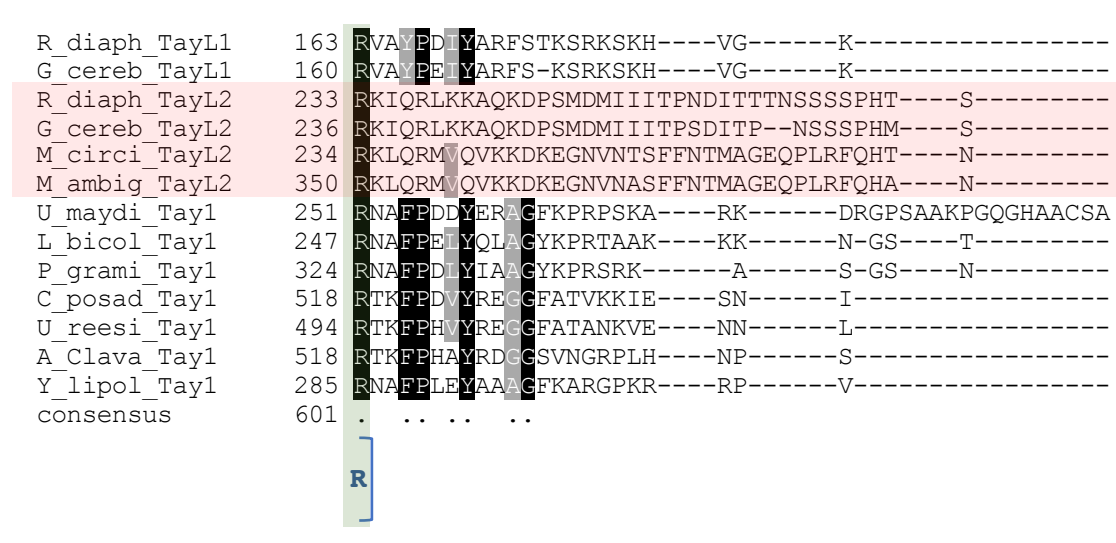
Myb1



Myb2







Supp. Fig. 1 Sequence alignments of Tay1, TayL1, and TayL2 family members

The Myb1 and Myb2 motifs for Tay1, TayL2 and TayL2 family members are aligned using T-coffee. The conserved motifs in TRF1 family members involved in binding the TTAGGG repeats (based on crystal structures in Court et al., 2005) are displayed below the alignments. The proteins included in the alignments are as follows: R_diaph_TayL1, from Rhizophagus diaphanous, RGB43698.1; G_cereb_TayL1, from *Glomus cerebriforme*, RIA97975.1; R_diaph_TayL2, from *Rhizophagus diaphanous*, RGB438099.1; G_cereb_TayL2, from *Glomus cerebriforme*, RIA96913.1; M_circi_TayL2, from *Mucor circinelloides*, EPB89450.1; M_ambig_TayL2, from *Mucor ambiguus*, GAN08276.1; U_maydi_Tay1; from *Ustilago maydis*, Um02326.1; L_bicol_Tay1, from *Laccaria bicolor*; XP_001886326; P_grami_Tay1, from *Puccinia graminis*, XP_003325585; C_posad_Tay1, from *Coccidioides posadasii*, EFW18452; U_reesi_Tay1, from *Uncinocarpus reesei*, XP_002540930; A_clava_Tay1, from *Aspergillus Clavatus*, XP_001270706; Y_lipol_Tay1, from *Yarrowia lipolytica*, XP_502676.