# Step-by-Step Evolution of Telomeres: Lessons from Yeasts

Filip Červenák[1],*, Regina Sepšiová[1], Jozef Nosek[2], and Ľubomír Tomáška[1],*

[1]Department of Genetics, Comenius University in Bratislava, Faculty of Natural Sciences, Bratislava, Slovakia
[2]Department of Biochemistry, Comenius University in Bratislava, Faculty of Natural Sciences, Bratislava, Slovakia

*Corresponding authors: E-mails: lubomir.tomaska@uniba.sk; cervenak3@uniba.sk.

## Abstract

In virtually every eukaryotic species, the ends of nuclear chromosomes are protected by telomeres, nucleoprotein structures counteracting the end-replication problem and suppressing recombination and undue DNA repair. Although in most cases, the primary structure of telomeric DNA is conserved, there are several exceptions to this rule. One is represented by the telomeric repeats of ascomycetous yeasts, which encompass a great variety of sequences, whose evolutionary origin has been puzzling for several decades. At present, the key questions concerning the driving force behind their rapid evolution and the means of co-evolution of telomeric repeats and telomere-binding proteins remain largely unanswered. Previously published studies addressed mostly the general concepts of the evolutionary origin of telomeres, key properties of telomeric proteins as well as the molecular mechanisms of telomere maintenance; however, the evolutionary process itself has not been analyzed thoroughly. Here, we aimed to inspect the evolution of telomeres in ascomycetous yeasts from the subphyla Saccharomycotina and Taphrinomycotina, with special focus on the evolutionary origin of species-specific telomeric repeats. We analyzed the sequences of telomeric repeats from 204 yeast species classified into 20 families and as a result, we propose a step-by-step model, which integrates the diversity of telomeric repeats, telomerase RNAs, telomere-binding protein complexes and explains a propensity of certain species to generate the repeat heterogeneity within a single telomeric array.

**Key words:** telomeric repeat, telomere-binding protein, telomerase, evolution.

## Significance

The terminal regions of eukaryotic nuclear chromosomes usually consist of an array of short tandem repeats representing variants of the sequence 5′-TTAGGG-3′. However, this feature of nuclear telomeres is not preserved in ascomycetous yeasts exhibiting a great diversity of telomeric DNA motifs. The evolutionary origin of this variability is still puzzling, as well as the means of co-evolution of telomeric DNA and telomere-binding proteins. In this study, we explored the variability of telomeric repeats in ascomycetous yeasts and proposed a tangible model of their evolution, describing hypothetical steps of diversification of systems protecting the chromosomal ends in yeasts.

## Introduction

Telomeric DNA is a key constituent of the system responsible for chromosome end-protection (Levy et al. 1992; de Lange 2009). The overall structure of different types of telomeres might vary significantly between species (Huang et al. 1998; Monaghan et al. 2018), extrachromosomal genetic elements, such as plasmids and viruses (e.g., Hinnebusch and Barbour 1991; Qin and Cohen 1998) or even between different compartments of a single cell (Gilson and McFadden 1995; Nosek et al. 1998; Zauner et al. 2000). However, their principal function is always to provide a solution to both the end-replication and end-protection problems, thus contributing to genome stability (McKnight et al. 2002; Bhattacharyya and Lustig 2006). Telomeric DNA of nuclear chromosomes usually consists of a double-stranded (ds) region composed of an array of short tandem repeats, which ends with a single-

stranded (ss) 3′ overhang (McEachern et al. 2000). Both ss and ds regions of telomeric DNA are bound by telomeric DNA-binding proteins (TBPs), which facilitate the crucial processes of telomere maintenance—replication, addition of new telomeric repeats and protection against undue DNA recombination and repair (Griffith et al. 1999; Smogorzewska and de Lange 2004; de Lange 2005; de Lange 2009; Stewart et al. 2012; Tomáška et al. 2020).

New telomeric repeats are usually synthesized at the 3′ ss overhang by telomerase, followed by the synthesis of complementary strand by conventional DNA polymerase (Blackburn and Collins 2011; Wu et al. 2017; Lue 2018). In most eukaryotes with known telomeric sequence, including all vertebrates, telomeric repeat is represented by a hexanucleotide 5′-TTAGGG-3′ or its single nucleotide substitution/indel variants (5′-TTAGGG-3′-like repeats) (Meyne et al. 1989; Weiss and Scherthan 2002; Vítková et al. 2005). As shown previously, this type of sequence is recognized by various DNA-binding proteins, such as mammalian TRF1, TRF2 (dsTBPs), and Pot1 (ssTBP) (de Lange 2005; de Lange 2018). Furthermore, several unrelated TBPs, including yeast telomeric proteins Tay1 and Taz1 (dsTBPs from Yarrowia lipolytica and Schizosaccharomyces pombe, respectively) bind 5′-TTAGGG-3′-like sequences with high affinity, even though the natural telomeric repeats of both yeasts differ from this motif (Kramara et al. 2010; Višacká et al. 2012; Sepšiová et al. 2016). In addition, Teb1p, a homolog of Tay1p responsible for the regulation of gene expression in S. pombe, also exhibit a high affinity toward the 5′-TTAGGG-3′ sequence (Brigati et al. 1993; Valente et al. 2013; Sepšiová et al. 2016), making this motif a versatile basis for the assembly of several types of telomere-protecting complexes.

## The Evolutionary Paths to Telomere Diversity

In most eukaryotic lineages, including basal branches, telomeres are composed of 5′-TTAGGG-3′ repeats (fig. 1), suggesting the ancestral origin of this motif (Blackburn and Challoner 1984; Chiurillo et al. 2000; Fulnečková et al. 2013). However, for reasons that still remain unclear, in certain lineages, telomeres evolved rapidly, leading to the emergence of novel types of repeats (Lue 2010; Červenák et al. 2017).

In most flowering plants, telomeric sequences are very similar (5′-TTTAGGG-3′) (Richards and Ausubel 1988; Richards et al. 1992; Shippen and McKnight 1998; Peška and Garcia 2020) or even identical to the vertebrate motif (e.g., several species belonging to the order Asparagales or Zostera marina) (Sýkorová, Lim, Kunická, et al. 2003; Sýkorová et al. 2006; Fajkus et al. 2019; Peška et al. 2020). However, there are two specific lineages, where these motifs evolved into more complex sequences. The first is represented by the genus Allium, including A. cepa, A. ursinum, and A. fistulosum. In these species, the telomeric repeat (5′-CTCGGTTATGGG-3′) is about two times longer than the vertebrate motif and includes a G/C-rich sequence at the 5′ end (Fajkus et al. 2016, 2019). The second has been discovered in Cestrum elegans, a species belonging to the Cestrum clade, which also includes the genera Sessea and Vestia, of the family Solanaceae. In this species, the sequence of the telomeric repeat (5′-TTTTTTAGGG-3′) is longer than the common plant motif, harboring additional thymine nucleotides (Peška et al. 2015; Fajkus et al. 2019). Interestingly, in closely related species Sessea stipulata and Vestia foetida, telomeres do not contain typical plant telomeric repeats as well, suggesting the unusual repeat might be conserved in the entire Cestrum clade (Sýkorová, Lim, Chase, et al. 2003; Fajkus et al. 2019). Similarly, the variant repeats were also found at the chromosomal ends of green algae, such as Chlorella vulgaris (5′-TTTAGGG-3′) and Chlamydomonas reinhardtii (5′-TTTTAGGG-3′) (Petracek et al. 1990; Higashiyama et al. 1995).

Another demonstration of the evolutionary tinkering at the chromosomal ends can be found in invertebrate animals, such as worms and insects. Specifically, in contrast to most flatworms, whose telomeres are composed of 5′-TTAGGG-3′ repeats (Bombarová et al. 2009), some chromosomal ends of the tapeworm Hymenolepis microstoma terminate with extremely long (median unit length of 179 nt) repeats, which exhibit several traits typical for centromeric sequences (Olson et al. 2020). This observation further underlines a possible functional and evolutionary relationship between centromeres and telomeres (Villasante et al. 2007) and suggests that under specific conditions, one type of the repetitive DNA element can be replaced by the other (Olson et al. 2020). Moreover, a slight deviation from the 5′-TTAGGG-3′ motif is also typical for some of the phylogenetically distant worm species, such as Caenorhabditis elegans and Ascaris lumbricoides, which possess a single substitution variant (5′-TTAGGC-3′) of the canonical repeat (Müller et al. 1991; Wicky et al. 1996).

In most insects, a telomeric repeat similar to the canonical vertebrate motif (5′-TTAGG-3′ or 5′-TCAGG-3′) was reported (Sahara et al. 1999; Vítková et al. 2005; Kuznetsova et al. 2012). However, in the fruit fly Drosophila melanogaster and the midge Chironomus tentans, the insect-type telomeric repeats were replaced by either retrotransposons or complex satellite sequences, respectively (Levis et al. 1993; Nielsen and Edström 1993; Kordyukova et al. 2018). In D. melanogaster, the three highly specialized terminal retrotransposons (HeT-A, TART, TAHRE) act as the protectors of the upstream unique sequence by inserting their copies preferentially at the chromosomal termini (Kordyukova et al. 2018). This type of genomic symbiosis enables the mobile element to propagate itself without reducing the fitness of its host, which is beneficial for both partners. In C. tentans, the repeat unit is 350 nt long and is likely to originate from short (7–10 nt) sequences similar to canonical telomeric repeats (Saiga and Edström

**Fig. 1.**—Telomere diversity in eukaryotic nuclear chromosomes. The phylogenies of eukaryotes, opisthokonts, and holozoans were inferred from Burki et al. (2020), Li et al. (2020), and Torruella et al. (2012), respectively. Black branches represent the lineages where the human-type (i.e., 5′-TTAGGG-3′) telomeric repeats were identified. Red branches indicate the lineages in which the 5′-TTAGGG-3′ telomeric repeats have not been identified so far. The lineages with uncharacterized telomeric sequences are shown in gray. Examples of nonhuman telomeric repeats are indicated above the branches (i.e., [1]*Cryptosporidium parvum* [Liu et al. 1998], [2]*Eimeria tenella* [Shirley 1994], [3]*Theileria parva* [Sohanpal et al. 1995], [4]*Tetrahymena thermophila* [Blackburn and Gall 1978], [5]*Oxytricha trifallax* [Klobutcher et al. 1981], [6]*Guillardia theta* [Zauner et al. 2000], [7]*Arabidopsis thaliana* [Richards and Ausubel 1988], [8]*Cestrum elegans* [Peška et al. 2015], [9]*Allium cepa* [Fajkus et al. 2016], [10]*Dictyostelium discoideum* [Rodriguez-Centeno et al. 2019], [11]*Giardia lamblia* [Morrison et al. 2007], [12]*Giardia intestinalis* [Le Blancq et al. 1991], [13]*Aspergillus oryzae* [Kusumoto et al. 2003], [14]various species [see the main text, fig. 2 and supplementary table 1, Supplementary Material online], [15]*Schizosaccharomyces cryophilus* and [16]*Schizosaccharomyces japonicus* [Rhind et al. 2011], [17]*Schizosaccharomyces pombe* [Murray et al. 1986], [18]*Phanerochaete chrysosporium* [Ramirez et al. 2011], [19]*Cryptococcus neoformans* [Edman 1992], [20]*Postia placenta* [Ramirez et al. 2011], [21]*Wallemia sebi* [Fulnečková et al. 2013], [22]*Rhizopus oryzae* [Ma et al. 2009], [23]*Allomyces macrogynus* [Fulnečková et al. 2013], [24]*Encephalitozoon cuniculi* [Peyret 2001], [25]*Creolimax fragrantissima* [Denbo et al. 2019], [26]*Caenorhabditis elegans* [Wicky et al. 1996], [27]*Apis melifera* [Sahara et al. 1999], [28]*Tribolium castaneum* [Osanai et al. 2006], [29]*Hymenolepis microstoma* [Olson et al. 2020], [30]*Drosophila melanogaster* [Levis et al. 1993]).

1985; Nielsen and Edström 1993). Presumably, several inversions and insertions changed the overall structure of these repeats, transforming them into long telomeric blocks that are no longer maintained by telomerase (Nielsen and Edström 1993).

Finally, one of the most puzzling cases of runaway evolution in telomere biology is represented by the telomeric repeats of fungi. Although in all major fungal lineages, the 5′-TTAGGG-3′-like motifs are common (fig. 1), telomeric repeats of species belonging to the ascomycete subphyla Saccharomycotina (e.g., *S. cerevisiae*, *Candida albicans*, *Kluyveromyces lactis*) and Taphrinomycotina (e.g., *S. pombe*) vary in both primary sequence and length (supplementary table 1, Supplementary Material online) (Gunišová et al. 2009; Lue 2010; Červenák et al. 2017). Telomeric motifs of

these species might be over 20 nucleotides long (e.g., *C. albicans*, *K. lactis*), heterogenous (e.g., *S. cerevisiae*, *S. pombe*) and appear unrelated to canonical 5′-TTAGGG-3′ repeat (Gunišová et al. 2009; Rhind et al. 2011). Moreover, the variability in telomeric sequences is in concert with the variability of telomeric proteins, ranging from conserved components similar to their counterparts in vertebrates (*S. pombe*) to systems based on unrelated DNA-binding proteins (*S. cerevisiae*, *Y. lipolytica*) and their interacting partners (Smogorzewska and de Lange 2004; Longhese et al. 2012; Červenák et al. 2017). This unprecedented variability hinders our understanding of the evolution of telomeres in certain fungal lineages. Nevertheless, the accessible data describing telomeric repeats of many different species (Cohn et al. 1998; Gunišová et al. 2009) along with several large scale genome

sequencing projects focused on yeasts (e.g., The 1000 yeast genome project; http://www.y1000plus.org/, last accessed January 6, 2021) enabled us to inspect the evolutionary process in greater detail. Based on these data, we present here a model describing a series of steps, which might have modified the ancestral fungal telomeric motifs into the present-day palette of variable telomeric repeats found in ascomycetous yeasts.

## 5′-TTAGGG-3′-like Motifs Served as a Basis for the Diversification of Fungal Telomeric Repeats

In Basidiomycota, canonical 5′-TTAGGG-3′ telomeric repeat is typical for most species (Guzmán and Sánchez 1994; Ramírez et al. 2011; Fulnečková et al. 2013). Similarly to some plants and insects, both longer (5′-TTTAGGG-3′ in *Phanerochaete chrysosporium*) and shorter (5′-TTAGG-3′ in *Postia placenta*, *Heterobasidion annosum*) variants are present, providing evidence for a slight variability of telomeric repeats in the phylum (Guzmán and Sánchez 1994; Ramírez et al. 2011). In Ascomycota, most species belonging to the subphylum Pezizomycotina (e.g., *Neurospora crassa*, *Podospora anserina*, *Aspergillus fumigatus*) share the canonical telomeric motif as well (Podlevsky et al. 2007; Fulnečková et al. 2013), although in *Aspergillus oryzae*, a longer repeat (5′-TTAGGGTCAACA-3′) was reported (Kusumoto et al. 2003). In the subphylum Taphrinomycotina, several species of the genera *Taphrina*, *Pneumocystis*, and *Saitoella* also possess 5′-TTAGGG-3′ repeats (Underwood et al. 1996; Cisse et al. 2013; Tsai et al. 2014). On the other hand, in fission yeast species of the genus *Schizosaccharomyces*, 5′-TTACTTGGG-3′ repeats were identified in both *S. cryophilus* and *S. octosporus* and heterogenous repeats 5′-TTACA$_{0-1}$G$_{2-8}$-3′ are typical for *S. pombe* (Murray et al. 1986; Rhind et al. 2011).

In contrast to most fungal lineages, 5′-TTAGGG-3′ telomeric motifs were not identified in the available genomic sequences of the Saccharomycotina species. As mentioned above, telomeric repeats in this subphylum vary greatly in terms of sequence motif, heterogeneity and overall length of telomeric arrays. The shortest predicted telomeric motifs (in case of heterogenous repeats, a sequence of a single repeat is deduced from the template domain of telomerase RNA [TER]) are 5 nt long (5′-TCAGG-3′) and were identified in several species belonging to the genus *Hanseniaspora*, whereas one of the longest putative telomeric motifs is almost nine times as long (43 nt) and was found in *Pichia norvegensis* (supplementary table 1, Supplementary Material online). Between these two extremes lies a wide range of different motifs with limited similarity to either 5′-TTAGGG-3′ or its variants. As a result, it is difficult to get a reliable alignment of yeast telomeric sequences that would point out the evolutionary conserved sites. However, several pieces of indirect evidence suggest that similarly to other eukaryotic lineages, the common ancestor of the subphylum Saccharomycotina possessed a 5′-TTAGGG-3′-like telomeric repeats.

Firstly, the putative telomeric repeats of *Lipomyces oligophaga* and *L. suomiensis*, two species from the *Lipomycetaceae* family, representing a basal branch of the Saccharomycotina subphylum (Shen et al. 2016; Shen et al. 2018), resemble the 5′-TTAGGG-3′ repeat (fig. 2; Krassovski et al. 2019). Moreover, we searched the genomic assemblies of other species of this family for putative telomeric repeats and identified four additional motifs, located predominantly as tandem repeat arrays at the ends of several scaffolds, which share this resemblance (supplementary table 1, Supplementary Material online). Similarly, we identified a candidate for telomeric repeat of *Trigonopsis vinaria*, a species belonging to the family *Trigonopsidaceae*, which is another deep branching lineage of this subphylum. Although this repeat is slightly more divergent, including a single thymine nucleotide inserted into the original triplet of guanosines, the overall similarity to 5′-TTAGGG-3′-like telomeric motifs is preserved. Next, in 12 out of 13 species belonging to the *Yarrowia* clade, the telomeric motif is 5′-TTNNNNAGGG-3′ with variable central four nucleotides, showing the conserved portion of the sequence is identical to the canonical repeat (Červenák et al. 2019). Importantly, in some of the yeast species with longer telomeric repeats (e.g., *Debaryomyces hansenii*, *Clavispora lusitaniae*), the 5′-TTAGGG-3′-like sequence is partially preserved at the 5′ end of the repeat (fig. 2). These 5′ elements may represent the remnants of the ancestral 5′-TTAGGG-3′-like telomeric motif and were eventually lost in some lineages (in most species of the *Saccharomycetaceae* family, telomeric repeats lack these elements). In addition, the 5′-TTAGGG-3′ repeats located at the subtelomeres of *S. cerevisiae* (Louis and Haber 1992) possibly represent a molecular fossil that might have served as the telomeric sequence before the current *S. cerevisiae* telomeric motif emerged and the chromosomal ends expanded. To inspect this hypothesis in greater detail, we analyzed the genomic sequences of several species where telomeric arrays have been unambiguously assigned to the ends of the corresponding chromosomes (e.g., *S. pombe*, *Y. lipolytica*), but we did not detect substantial enrichment for 5′-TTAGGG-3′ repeats in the subterminal regions of chromosomal contigs. Therefore, it is likely that in these cases the ancestral telomeric repeats were lost due to a lack of positive selection. In *S. cerevisiae*, subtelomeric 5′-TTAGGG-3′ repeats, in concert with DNA-binding proteins, such as Tbf1p and Reb1p mediate the formation of heterochromatin (Fourel et al. 1999), which may in turn also play a role in the preservation of chromosomal termini.

The telomere-binding proteins are another crucial factor in the evolutionary process. In *Y. lipolytica*, a species belonging to one of the basal clades of Saccharomycotina, the ds part of telomeric DNA is bound by Tay1p, a DNA-binding protein containing two tandem Myb domains exhibiting similarity to those of human TRF1 and TRF2 (Kramara et al. 2010).

**Fig. 2.**—Variability of telomeric repeats and TBPs in the subphylum Saccharomycotina. Topology of the phylogenetic tree was inferred from Shen et al. (2016). Blue text represents the 5′-TTAGGG-3′ sequences, orange boxes mark the conserved Rap1p-binding sites (according to Wahlin and Cohn 2000) and gray boxes mark the degenerate Rap1p-binding sites (up to 2 substitutions/indels). Tay1p and/or Rap1p homologs were identified in genomic sequences using the TBlastN algorithm (blast.ncbi.nlm.nih.gov) and the sequence of ScRap1p or YlTay1p as query. Putative homologs (E-value $< 1.10^{-12}$) are indicated (+). The species with confirmed telomeric repeats are listed in bold, asterisks indicate the presence of heterogenous repeats.

| Species | Telomeric repeat | Tay1p/Teb1p | Rap1p |
|---|---|---|---|
| *Lipomyces oligophaga* | TTTGAGGG | + | |
| *Lipomyces suomiensis* | TTTAGTAGGG | + | |
| *Trigonopsis vinaria* | TTAGCGAGGTG | | |
| **Candida hispaniensis** | TTGACGAGAG | + | |
| **Yarrowia lipolytica** | TTAGTCAGGG | + | |
| *Wickerhamiella cacticola* | TTACCATTCGG | + | |
| *Wickerhamiella infanticola* | TTAGGAGAGGGGAATAAGACAC | + | + |
| *Pichia kudriavzevii* | TTACAATATGAACTAGGAGCGAGGTGT | | + |
| *Pichia norvegensis* | TTAACGTAAACTAGGAGCGCGTGTGGGTTGAGCAGACTATCAC | | + |
| *Cephaloascus albidus** | TCTGGG(TG)0-4 | | + |
| *Cephaloascus fragrans* | TGGTGTAAGGGTGAG | | + |
| **Meyerozyma guilliermondii** | TGGTGTGTGTAC | | + |
| **Debaryomyces hansenii** | TTGAGGTGTAGGGATG | | + |
| **Scheffersomyces stipitis** | TTTTCACGTCTTGCGGTATGGATC | | + |
| *Spathaspora passalidarium* | TTCTGGGGTACTCTCTTATGTTGCGGGTAGGATG | | + |
| **Candida albicans** | TTCTTGGTGTACGGATGTCTAAC | | + |
| **Candida parapsilosis** | TTGATTATACTGAGGTCCGGATG | | + |
| *Hyphopichia burtonii* | TTTTTGCCCTAACTAAATACTAGATGTACGGAGG | | + |
| **Clavispora lusitaniae** | TTCTTTAGGGAGGTACTGATG | | + |
| *Candida auris** | TTCTTAGGTGGTGTCTGGGT | | + |
| *Ascoidea asiatica* | TTGATGTTGGATGACAG | | + |
| *Saccharomycopsis fodiens* | TGTGGATG - TGTGGATG | | + |
| *Saccharomycopsis malanga* | TCAGTGGGGTAAGGGTG | | + |
| *Cyberlindnera saturnus* | TTGTGGTG - TTGTGGTG | | + |
| *Wickerhamomyces canadensis* | TTCTGATGGTG - TTCTGATGGTG | | + |
| *Hanseniaspora meyeri* | TCAGG - TCAGG - TCAGG | | + |
| *Hanseniaspora uvarum* | TCAGG - TCAGG - TCAGG | | + |
| **Kluyveromyces lactis** | TTGATTAGGTATGTGGTGTACGGAT | | + |
| **Saccharomyces cerevisiae *** | TGGTGTGTGGGTG | | + |

Importantly, Tay1p was reported to bind 5′-TTAGGG-3′ repeats with higher affinity than the natural telomeric sequence of *Y. lipolytica* (5′-TTAGTCAGGG-3′) (Višacká et al. 2012), suggesting that either 1) the optimal affinity of this protein toward telomeric sequences is not necessarily the highest possible ( Tomáška et al. 2019), or 2) in its evolutionary history, Tay1p was selected as the 5′-TTAGGG-3′-binding protein and as soon as the telomeric repeats of yeasts started to accumulate mutations, its affinity toward the new motifs decreased. The second scenario relies on the presumption that telomeric repeats evolved faster than the corresponding TBPs and is consistent with the fact that besides telomeres, TBPs usually bind various DNA sequences located inside the chromosome (such as promoter regions of specific genes) and interact with different protein partners (e.g., transcription activators). As a result, yeast TBPs had to adapt to new telomeric repeats while maintaining the already existing interactions. Understandably, a combination of both aforementioned scenarios is possible, assuming a wider range of affinities between telomeric repeats and the TBP is acceptable for the TBP (in this case Tay1p) to fulfill its telomeric functions. Nevertheless, in many yeast lineages with further modified telomeric repeats, *TAY1* gene was not identified in the genome (these species recruited Rap1p as TBP; fig. 2), suggesting the gradual decrease in affinity of the ancestral TBP toward ever-changing telomeric repeats led to its replacement and (in some cases) elimination.

Taken together, the currently available data concerning the evolutionary constrains of different yeast species, properties of specific TBPs and the sequences of several telomeric repeats indicate that the ancient telomeric repeat present in the common ancestor of Saccharomycotina species was a 5′-TTAGGG-3′-like sequence.

## Gradual Divergence of Yeast Telomeric Repeats: A Model of Evolution

Looking at the present variability of telomeric repeats in ascomycetous yeasts, it is obvious that these sequences evolved rapidly, and as a result, 1) all Saccharomycotina species analyzed so far lack the 5′-TTAGGG-3′ telomeric repeats, and 2) protein complexes of yeast telomeres were restructured several times in distinct phylogenetic branches. In our previous work (Červenák et al. 2017), we proposed a general evolutionary scenario, where the initial changes in the sequences of telomeric repeats triggered the transition from ancestral TBPs with strict preference for 5′-TTAGGG-3′ repeats to their more flexible counterparts. Here, we aimed to inspect the process in greater detail with special focus on the specific changes in the sequences of telomeric repeats from a large palette of species.

Therefore, we systematically analyzed the telomeric repeats in Saccharomycotina species, put them into the context of their evolutionary history and supported the resulting model with newly identified putative telomeric sequences.

As mentioned above, most of the *Yarrowia* clade species share the 5′-TTNNNNAGGG-3′ telomeric motif, suggesting that in the early stages of the evolutionary process the 5′-TTAGGG-3′ sequence became prone to the insertion of spacer (Červenák et al. 2019). This hypothesis is in line with the data from other basal lineages, such as the *Lipomycetaceae* family, where the putative telomeric repeats also include short insertions (supplementary table 1, Supplementary Material online). Next, we searched for the similarities in telomeric sequences (both putative and experimentally confirmed) from species belonging to the families *Debaryomycetaceae* (e.g., *D. hansenii*) and *Metschnikowiaceae* (e.g., *C. lusitaniae*, *Candida auris*) and found out that in some of them, an arrangement similar to that of the *Yarrowia* clade species is preserved (fig. 2). In *D. hansenii*, the spacer element is longer than in *Y. lipolytica* and three additional nucleotides are present at the 3′ end, but the overall structure is conserved. Moreover, the predicted telomeric repeats of several related species showed similar structural pattern with variation in the spacer length. In *C. lusitaniae* and *C. auris*, the 3′ ends of the repeats are expanded, but the *Yarrowia*-like sequence is still present at the 5′ end. The sequences of putative telomeric motifs, predicted for other species of the family *Metschnikowiaceae* are structurally similar, including the expanded 3′ end (supplementary table 1, Supplementary Material online). Importantly, the 3′ sequences of telomeric motifs from these species represent potential binding sites for Rap1p (fig. 2). This particular feature is crucial for our understanding of the evolutionary process, suggesting the ancestral telomeric repeats expanded at their 3′ ends, which led to the emergence of a new protein-binding site. In some species (e.g., *Hyphopichia burtonii*, *Spathaspora passalidarum*), it is likely that the expansion of the repeat was due to gradual elongation of spacer, which eventually provided a platform for the generation of Rap1p-binding site (fig. 2).

According to our hypothesis, up to certain point in the evolution, these complex telomeric repeats were still bound by the ancestral 5′-TTAGGG-3′-binding proteins (such as Tay1p), although their affinity gradually decreased as the mutations accumulated. After some time, the decrease in affinity compromised the telomeric functions of the dsTBP, which would lead to a drastic decrease in cell fitness. Presumably, this situation was resolved by the recruitment of Rap1p, a novel telomere-binding protein, which recognized the complex repeats and took over the position of the dsTBP in some yeast lineages (Višacká et al. 2012). This scenario is supported by the sequence composition of specific telomeric repeats and can be experimentally tested in several ways, for example, by confirming the role of repeats identified in silico,

comparing the kinetic properties of different TBPs toward various types of telomeric repeats and their variants, and/or reprogramming telomeres in a species to carry repeats typical for another species (similarly to the case of "humanized" *S. cerevisiae* telomeres) and testing the ability of native TBP to maintain such telomeres in vivo (Henning et al. 1998).

The evolution of Rap1p itself is an interesting story, that needs to fit into our model. In several species, such as human and *S. pombe*, it does not bind DNA directly but is associated with TBPs via protein–protein interactions and it is a part of the telomere-protecting complex called shelterin (de Lange 2005; Fujita et al. 2012). In these species, Rap1p possesses just one Myb domain, which is insufficient for stable DNA-binding and thus cannot associate with telomeric DNA directly (Li et al. 2000; Chikashige and Hiraoka 2001). However, in several species from the families *Saccharomycetaceae* (*S. cerevisiae* and related species), *Debaryomycetaceae* (e.g., *D. hansenii*, *Meyerozyma guilliermondii*, *S. passalidarum*, *Scheffersomyces stipitis*) and *Metschnikowiaceae* (e.g., *C. lusitaniae*), Rap1p includes two Myb domains, enabling it to effectively bind telomeric repeats. Similarly, some of the Rap1p homologs from basidiomycetes with 5′-TTAGGG-3′ telomeric repeats were also predicted to possess two Myb domains (Yu et al. 2013), although none of these proteins was reported to be involved in telomere maintenance so far. Instead, telomeres in *U. maydis* seem to be bound by two dsTBPs, homologous to *Yl*Tay1p (*Um*Trf1p) and *Sc*Tbf1p (*Um*Trf2p), respectively, and each having a distinct role in telomere maintenance (Yu et al. 2020). This situation can be explained by two possible scenarios: 1) the ancestral Rap1p possessed one Myb domain and acquired another in several independent lineages, or 2) the ancestral protein possessed two Myb domains and some of the later orthologs lost one of them. Either way, the ancestral *RAP1* gene was present in the common ancestor of both ascomycetes and basidiomycetes alongside the other TBP-encoding genes, which might have emerged via the neutral evolutionary ratchet (Višacká et al. 2012), and became associated with telomeric DNA via protein–protein interactions (as in human or *S. pombe*). Afterwards, the initial proximity of the protein to telomeric DNA in combination with two functional Myb domains made it possible for Rap1p to become a TBP in certain lineages. In others (e.g., *Lipomycetaceae*, *Yarrowia* clade), Rap1p did not acquire any important function and was lost. Moreover, one specific feature of Rap1p, which might have been useful during the Tay1p/Rap1p transition period is its flexibility in DNA-binding, allowing it to bind a wider spectrum of potential substrates than its ancient counterparts (Steinberg-Neifach and Lue 2015). Specifically, the binding of two Myb domains tolerates different spacing between recognition sequences, allowing for flexible binding to a spectrum of motifs. Thus, the DNA-binding activity of Rap1p may be modulated by binding to specific variants of consensus sequence and the protein may adopt several conformations, enabling a

recruitment of different cofactors (König et al. 1996; Steinberg-Neifach and Lue 2015).

Nevertheless, the fundamental reconstruction of the telomere-associated complexes in many yeast species had to be an extremely challenging process. Specific interactions between the DNA-binding protein and its interacting partners had to be rebuilt, many of the interacting partners themselves had to be replaced and the communication between the telomeric proteins and other components of the telomere maintenance systems, such as telomerase and DNA repair machinery needed to be maintained. Clearly, Rap1p was able to fulfill this complex task and from that point on, it served as a basis for the assembly of telomere-protecting complexes in the corresponding yeast species.

After the Tay1p/Rap1p transition, we can observe the gradual disappearance of the 5′-TTAGGG-3′ sequence from yeast telomeric sequences, which is probably due to the relaxation of selection pressure (fig. 2). As a result, in many species belonging to the families *Debaryomycetaceae*, *Metschnikowiaceae* and *Phaffomycetaceae* (e.g., *Wickerhamomyces canadensis*, *Cyberlindnera saturnus*), Rap1p-binding site is conserved, whereas the 5′ sequence is variable (Gunišová et al. 2009). In most species of the family *Saccharomycetaceae*, including *S. cerevisiae*, the former 5′ sequence of the telomeric repeat is completely lost (Cohn et al. 1998), suggesting the remnants of the 5′-TTAGGG-3′ repeat were not essential for the binding of Rap1p. Eventually, although the transition from ancestral 5′-TTAGGG-3′-binding telomeric proteins to flexible Rap1p was caused by the initial changes in telomeric sequences, it resulted in a situation, were the TBP was much more likely to tolerate new mutations. Such arrangement ultimately accelerated the evolutionary process and led to the present variability of telomeric motifs in ascomycetous yeast. It is also possible, that in some species (e.g., *S. cerevisiae*), the 5′-TTAGGG-3′-like repeats might have been bound by Tbf1p, which either served as the initial dsTBP instead of Tay1, or took part in a more complex Tay1p/Tbf1p/Rap1p transition period (Červenák et al. 2017). However, with an exemption of its potential role in maintenance of artificially humanized telomeres in *S. cerevisiae* (Alexander and Zakian 2003; Berthiau et al. 2006; Fukunaga et al. 2012; Ribaud et al. 2012), its function as the native TBP was so far not documented in any species.

## Role of TERs in the Evolution of Telomeric Repeats

When examining the sequences of telomeric repeats and the process of their evolution, some of the crucial pieces of evidence might lie in the TERs. These noncoding RNAs contain several important domains responsible for RNA–protein interactions, as well as the template domain, which specifies the sequence of a telomeric repeat (Romero and Blackburn 1991; Theimer and Feigon 2006). Most regions of TERs evolve

extremely rapidly and even in relatively closely related species, it is complicated to align the entire sequences (Gunišová et al. 2009; Červenák et al. 2019). On the other hand, the template domain is usually conserved (Chen et al. 2000; Červenák et al. 2019). As the sequence complementary to G-strand of every telomeric repeat, the template domain of TER is crucial in our understanding of telomeres, because 1) if a mutation should affect the sequence of every telomeric repeat in the cell, it needs to occur in the template domain of TER, and 2) the secondary structure of the template domain might affect the activity of the entire telomerase complex.

Telomeric arrays undergo frequent mutations resulting from an inefficient repair of damaged DNA, causing local problems to DNA-binding proteins, replication and transcription machineries, and other regulatory systems (Chen et al. 2000; Fumagalli et al. 2012). However, these are never systematic and do not occur repeatedly at the same positions. In contrast, if the template domain of TER is affected by a mutation, each newly synthesized telomeric repeat is affected, which inflicts a much broader impact on the cell fitness and ultimately the evolutionary process. Importantly, for every telomeric repeat, there are many variants of a template domain (different permutations of 1–2 repeat units) potentially acceptable for its synthesis by telomerase. This leaves a wide space for a variability among the primary structures of the templates. However, if the yeast telomeric repeats evolved step-by-step from an ancient 5′-TTAGGG-3′-like sequences, as we propose, then so should have done the template domains of their TERs. Consequently, when aligning the template domains of TERs from several unrelated yeast species, it is apparent that these sequences show the same pattern of changes as the telomeric repeats (fig. 3). This suggests that the evolutionarily modern templates evolved from the ancient sequences via gradual accumulation of mutations and the expansion of 5′ end of the template domain, followed by the Tay1p/Rap1p transition and the shortening of the 3′ end. Considering the mechanism of telomere-lengthening by telomerase (Wu et al. 2017), the reason why the 5′ end of the template accumulated mutations faster than the 3′ end might have been based on the role of 3′ end in the telomere-TER annealing.

The secondary structure of the template domain and possibly the entire TER molecule is another important factor in the evolutionary process. Specifically, the ability of mutated TERs to form a catalytically active complex with the other subunits of telomerase is critical for the synthesis of new telomeric repeats (Bachand and Autexier 2001; Lin et al. 2004). Therefore, if the other types of changes (e.g., severe rearrangements of the original template domain or the expansion of its 3′ end) prevented TERs from creating a secondary structure suitable for the telomerase assembly, such changes would have been detrimental to the telomere maintenance system even though the telomeric repeat itself was suitable for the binding of a particular TBP.
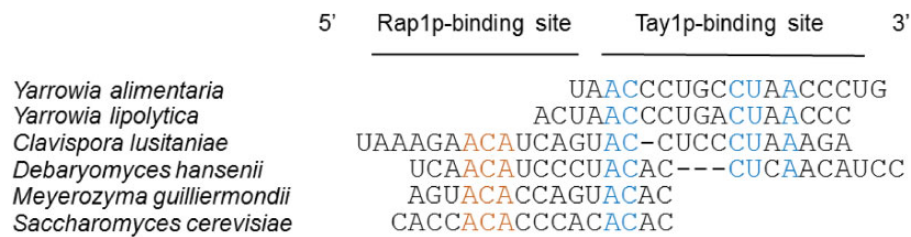
|  | 5' | Rap1p-binding site | Tay1p-binding site | 3' |

```
                         5'  Rap1p-binding site    Tay1p-binding site  3'
                             _____        _____

Yarrowia alimentaria                          UAACCCUGCCUAACCCUG
Yarrowia lipolytica                          ACUAACCCUGACUAACCC
Clavispora lusitaniae        UAAAGAACAUCAGUAC-CUCCCUAAAGA
Debaryomyces hansenii         UCAACAUCCCUACAC---CUCAACAUCC
Meyerozyma guilliermondii          AGUACACCAGUACAC
Saccharomyces cerevisiae          CACCACACCCACACAC
```

Fig. 3.—Evolution of template domains of yeast TERs. Putative binding sites for Tay1p and Rap1p are indicated, the bases conserved in all species are marked by blue (in Tay1p-binding site) or orange (in Rap1p-binding site) color.

## Heterogenous Telomeric Repeats in Yeasts

One particular feature of yeast telomeric repeats, which emerged alongside their rapid evolution, is the heterogeneity of repetitive units inside a single telomeric array (Cohn and Blackburn 1995; McCormick-Graham et al. 1997; Cohn et al. 1998). It is typical for *S. cerevisiae* and several mechanisms contributing to this type of variability have been described in the past, including stuttering, stalling, random misincorporation of bases or a premature dissociation of telomerase (McCormick-Graham et al. 1997; Förstemann and Lingner 2001). In vitro characterization of telomerases from *S. cerevisiae* and *Naumovozyma castellii* suggests the variations in *S. cerevisiae* telomeric repeats are due to premature dissociation and stuttering along the template (Cohn and Blackburn 1995). Interestingly, in some strains of *Candida tropicalis*, telomeric arrays are composed of two types of repeats, differing in a single nucleotide. It was proposed that this type of heterogeneity might have been introduced by two distinct TER alleles with different template domains (McEachern and Blackburn 1994).

In *S. pombe*, heterogeneity can be explained as a combination of effects where in a certain extension cycle only a part of RNA template is used, or, as in *S.cerevisiae*, RNA template aligns to telomeric DNA in different registers in different cycles of extension (Förstemann and Lingner 2001; Webb and Zakian 2008; Wellinger and Zakian 2012). The analyses of *S. pombe* TER also revealed that polymerization can in some cycles extend into the template boundary element and cause the heterogeneity at the 3' end of the repeat (Webb and Zakian 2008).

In other eukaryotes (e.g., the ciliate *Tetrahymena thermophila* from the superphylum Alveolata), the insertion of additional cytosine residue in TER template can lead to telomerase stuttering, resulting in the addition of multiple guanosines (Gilley et al. 1995). But how did this sequence heterogeneity arise and why was it tolerated (or possibly even favored) in the evolution? There are two possible situations, where heterogeneity in telomeric repeats might be tolerated: 1) the telomere-binding proteins are flexible binders, recognizing a broader palette of target sequences, or 2) a pool of Myb domain encoding sequences is present in the genome, enabling further adaptation to sequence variability at telomeres by

substituting the former TBP for a more flexible contender. It was also proposed that under certain conditions, heterogeneity-generating mechanisms might be positively selected (Tomáška and Nosek 2009; Červenák et al. 2017).

At *S. pombe* telomeres, the heterogenous repeats are bound by Taz1, a dsTBP, which possesses only a single Myb domain in its sequence and forms a homodimer complexes (Cooper et al. 1997; Deng et al. 2015). In comparison to human telomeric proteins TRF1 and TRF2, Taz1p exhibits more flexibility and binds various DNA substrates with comparable affinity (Sepšiová et al. 2016). Importantly, the origin of gene encoding Taz1p still remains enigmatic, making it a possible Myb domain-containing contender which replaced a Tay1-like protein at *S. pombe* telomeres. However, since the telomeric repeats of *S. pombe* and related species resemble the 5'-TTAGGG-3' motif, the reason for the Tay1p/Taz1p transition might have been different in comparison with Saccharomycotina species. According to our scenario, Taz1p was able to replace Tay1p even though the affinity of Tay1p toward telomeric DNA was still relatively high and no novel protein-binding site emerged (fig. 4). The exact reasons for this change of guards in Taphrinomycotina are still unclear, although the flexibility in DNA-binding, which is typical for Taz1p (similarly to Rap1p) might have played a key role. Moreover, it is possible that in comparison to Saccharomycotina, the evolutionary process is still in its early stage and the Tay1p homologs will eventually disappear in some Taphrinomycotina species as well. On the other hand, the role of *Sp*Teb1p in transcription regulation of several essential genes (e.g., those encoding histones) suggests it might simply adapt to a new function (Valente et al. 2013). Importantly, the binding flexibility of Taz1p allowed for the emergence of heterogeneity in *S. pombe* telomeric sequence later on, whereas the Tay1p/Rap1p transition served as a basis for the emergence of telomeric heterogeneity in *S. cerevisiae*. However, the transition to the novel types of TBPs was clearly not the only reason for the emergence of telomeric heterogeneity, since the other related species (to both *S. cerevisiae* and *S. pombe*), which possess similar TBPs, contain homogenous telomeric repeats. It is therefore possible, that the flexible TBPs only created the environment, where the additional mutations affecting the processivity and fidelity of yeast

FIG. 4.—Telomeric repeats and TBPs in the subphylum Taphrinomycotina. Topology of the phylogenetic tree was inferred from Milo et al. (2019). Blue text represents the 5'-TTAGGG-3' sequences. Teb1p and/or Taz1p homologs were identified in genomic sequences using the TBlastN algorithm (blast.ncbi.nlm.nih.gov) and the sequence of SpTeb1p or SpTaz1p as query. Putative homologs (E-value $< 1.10^{-5}$) are indicated (+).The species with confirmed telomeric repeats are listed in bold, asterisks indicate the presence of heterogenous repeats.

telomerases were tolerated. The heterogeneity itself might have been introduced later by different types of error-prone telomerase complexes, establishing a dynamic balance between the properties of specific TBP and corresponding telomerase. According to this hypothesis, the emergence of telomeric heterogeneity might be expected in more yeast species with flexible TBPs in the future.

## Conclusions

In recent years, the variability of yeast telomeric repeats, TBPs and TERs was addressed several times, bringing up new experimental data and new points of view. The identification of telomeric sequences in most Saccharomycotina clades and comparative analyses of telomeric complexes from different lineages pointed out general trends in genome evolution, which served as the basis for our model (Lue 2010; Červenák et al. 2017). Later, the structural and functional analyses of telomeric proteins revealed the flexibility in DNA-binding of specific TBPs (Steinberg-Neifach and Lue 2015; Sepšiová et al. 2016; Tomáška, et al. 2019), which was instrumental in our understanding of processes, such as Tay1p/Rap1p transition. In this article, we combined the available data with our own analysis of several yeast telomeric repeats and template domains of TERs, in order to create a tangible step-by-step model of evolution. According to this model, the 5' end of the template domains in yeast TERs expanded, leading to the emergence of complex telomeric repeats. Consequently, at the 3' ends of these repeats, a site appeared which eventually led to Tay1p/Rap1p transition and the shortening of novel repeats from their 5' end. As a result, large diversity of telomeric repeats emerged in ascomycetous yeasts and was accompanied by formation of several types of protein complexes, which were adapted for their effective maintenance.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

## Data Availability

The data underlying this article are available in the article and in the Supplementary Material online.

## Literature Cited

Alexander MK, Zakian VA. 2003. Rap1 telomere association is not required for mitotic stability of a $C_3TA_2$ telomere in yeast. EMBO J. 22(7):1688–1696.

Bachand F, Autexier C. 2001. Functional regions of human telomerase reverse transcriptase and human telomerase RNA required for telomerase activity and RNA-protein interactions. Mol Cell Biol. 21(5):1888–1897.

Berthiau A-S, et al. 2006. Subtelomeric proteins negatively regulate telomere elongation in budding yeast. EMBO J. 25(4):846–856.

Bhattacharyya MK, Lustig AJ. 2006. Telomere dynamics in genome stability. Trends Biochem Sci. 31(2):114–122.

Blackburn EH, Challoner PB. 1984. Identification of a telomeric DNA sequence in *Trypanosoma brucei*. Cell 36(2):447–457.

Blackburn EH, Collins K. 2011. Telomerase: an RNP enzyme synthesizes DNA. Cold Spring Harb Perspect Biol. 3(5):a003558–a003558.

Blackburn EH, Gall JG. 1978. A tandemly repeated sequence at the termini of extrachromosomal ribosomal RNA genes in *Tetrahymena*. J Mol Biol. 120(1):33–35.

Bombarová M, Vítková M, Spakulová M, Koubková B. 2009. Telomere analysis of platyhelminths and acanthocephalans by FISH and Southern blot hybridization. Genome 52(11):897–1706.

Brigati C, Kurtz S, Balderes D, Vidali G, Shore D. 1993. An essential yeast gene encoding a TTAGGG repeat-binding protein. Mol Cell Biol. 13(2):1306–1314.

Burki F, Roger AJ, Brown MW, Simpson A. 2020. The new tree of eukaryotes. Trends Ecol Evol. 35(1):43–55.

Červenák F, et al. 2017. Double-stranded telomeric DNA binding proteins: diversity matters. Cell Cycle 16(17):1568–1577.

Červenák F, et al. 2019. Identification of telomerase RNAs in species of the *Yarrowia* clade provides insight into the co-evolution of telomerase, telomeric repeats and telomere-binding proteins. Sci Rep. 9(1):13365.

Chen JL, Blasco MA, Greider CW. 2000. Secondary structure of vertebrate telomerase RNA. Cell 100(5):503–514.

Chikashige Y, Hiraoka Y. 2001. Telomere binding of the Rap1 protein is required for meiosis in fission yeast. Curr Biol. 11(20):1618–1623.

Chiurillo MA, et al. 2000. Cloning and characterization of *Leishmania donovani* telomeres. Exp Parasitol. 94(4):248–258.

Cisse OH, et al. 2013. Genome sequencing of the plant pathogen *Taphrina deformans*, the causal agent of peach leaf curl. mBio 4(3):e00055-13.

Cohn M, Blackburn EH. 1995. Telomerase in yeast. Science 269(5222):396–400.

Cohn M, McEachern MJ, Blackburn EH. 1998. Telomeric diversity within the genus *Saccharomyces*. Curr Genet. 33(2):83–91.

Cooper JP, Nimmo ER, Allshire RC, Cech TR. 1997. Regulation of telomere length by a Myb-domain protein in fission yeast. Nature 385(6618):744–747.

de Lange T. 2005. Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev. 19(18):2100–2110.

de Lange T. 2009. How telomeres solve the end-protection problem. Science 326(5955):948–952.

de Lange T. 2018. Shelterin-mediated telomere protection. Annu Rev Genet. 52(1):223–247.

Denbo S, et al. 2019. Revision of the *Capsaspora* genome using read mating information adjust the view on premetazoan genome. Dev Growth Differ. 61(1):34–42.

Deng W, et al. 2015. Fission yeast telomere-binding protein Taz1 is functional but not a structural counterpart of human TRF1 and TRF2. Cell Res. 25(7):881–884.

Edman JC. 1992. Isolation of telomerelike sequences from *Cryptococcus neoformans* and their use in high-efficiency transformation. Mol Cell Biol. 12(6):2777–2783.

Fajkus P, et al. 2016. *Allium* telomeres unmasked: the unusual telomeric sequence (CTCGGTTATGGG)n is synthesized by telomerase. Plant J. 85(3):337–347.

Fajkus P, et al. 2019. Telomerase RNAs in land plants. Nucleic Acids Res. 47(18):9842–9856.

Förstemann K, Lingner J. 2001. Molecular basis for telomere repeat divergence in budding yeast. Mol Cell Biol. 21(21):7277–7286.

Fourel G, Revardel E, Koering CE, Gilson E. 1999. Cohabitation of insulators and silencing elements in yeast subtelomeric regions. EMBO J. 18(9):2522–2537.

Fujita I, Tanaka M, Kanoh J. 2012. Identification of the functional domains of the telomere protein Rap1 in *Schizosaccharomyces pombe*. PLoS One 7(11):e49151.

Fukunaga K, Hirano Y, Sugimoto K. 2012. Subtelomere-binding protein Tbf1 and telomere-binding protein Rap1 collaborate to inhibit localization of the Mre11 complex to DNA ends in budding yeast. Mol Cell Biol. 23(2):347–359.

Fulnečková J, et al. 2013. A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryots. Genome Biol Evol. 5(3):468–483.

Fumagalli M, et al. 2012. Telomeric DNA damage is irreparable and causes persistent DNA-damage-response activation. Nat Cell Biol. 14(4):355–365.

Gilley D, Lee MS, Blackburn EH. 1995. Altering specific telomerase RNA template residues affects active site function. Genes Dev. 9(18):2214–2226.

Gilson P, McFadden GI. 1995. The chlorarachniophyte: a cell with two different nuclei and two different telomeres. Chromosoma 103(9):635–641.

Griffith JD, et al. 1999. Mammalian telomeres end in a large duplex loop. Cell 97(4):503–514.

Gunišová S, et al. 2009. Identification and comparative analysis of telomerase RNAs from *Candida* species reveal conservation of functional elements. RNA 15(4):546–559.

Guzmán PA, Sánchez JG. 1994. Characterization of telomeric regions from *Ustilago maydis*. Microbiology 140(Pt3):551–557.

Henning KA, Moskowitz N, Ashlock MA, Liu PP. 1998. Humanizing the yeast telomerase template. Proc Natl Acad Sci USA.. 95(10):5667–5671.

Higashiyama T, Noutoshi Y, Akiba M, Yamada T. 1995. Telomere and LINE-like elements at the termini of the *Chlorella* chromosome I. Nucleic Acids Symp Ser. 34:71–72.

Hinnebusch J, Barbour AG. 1991. Linear plasmids of *Borrelia burgdorferi* have a telomeric structure and sequence similar to those of a eukaryotic virus. J Bacteriol. 173(22):7233–7239.

Huang CH, Lin YS, Yang YL, Huang SW, Chen CW. 1998. The telomeres of *Streptomyces* contain conserved palindromic sequences with potential to form complex secondary structures. Mol Microbiol. 28(5):905–916.

Klobutcher LA, Swanton MT, Donini P, Prescott DM. 1981. All gene-sized DNA molecules in four species of hypotrichs have the sam terminal sequence and an unusual 3′ teminus. Proc Natl Acad Sci USA. 78(5):3015–3019.

König P, Giraldo R, Chapman L, Rhodes D. 1996. The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. Cell 85(1):125–136.

Kordyukova M, Olovnikov I, Kalmykova A. 2018. Transposon control mechanisms in telomere biology. Curr Opin Genet Dev. 49:56–62.

Kramara J, et al. 2010. Tay1 protein, a novel telomere binding factor from *Yarrowia lipolytica*. J Biol Chem. 285(49):38078–38092.

Krassovski T, et al. 2019. Multiple reinventions of mating-type switching during budding yeast evolution. Curr Biol. 29(15):2555–2562.

Kusumoto K-I, Suzuki S, Kashiwagi Y. 2003. Telomeric repeat sequence of *Aspergillus oryzae* consists of dodeca-nucleotides. Appl Microbiol Biotechnol. 61(3):247–251.

Kuznetsova VG, Grozeva SM, Anokhin BA. 2012. The first finding of (TTAGG)n telomeric repeat in chromosomes of true bugs (Heteroptera, Belostomatidae). Comp Cytogenet. 6(4):341–346.

Le Blancq SM, Kase RS, Van der Ploeg LH. 1991. Analysis of a *Giardia lamblia* rRNA encoding telomere with [TAGGG]n as the telomere repeat. Nucleic Acids Res. 19(20):5790–5790.

Levis RW, Ganesan R, Houtchens K, Tolar LA, Sheen F. 1993. Transposons in place of telomeric repeats at a *Drosophila* telomere. Cell 75(6):1083–1093.

Levy MZ, Allsopp RC, Futcher AB, Greider CW, Harley CB. 1992. Telomere end-replication problem and cell aging. J Mol Biol. 225(4):951–960.

Li B, Oestreich S, de Lange T. 2000. Identification of human Rap1: implications for telomere evolution. Cell 101(5):471–483.

Li Y, et al. 2020. A genome-scale phylogeny o Fungi; insights into early evolution, radiations, and the relationship between taxonomy and phylogeny. BioRxiv preprint. doi:10.1101/2020.08.23.262857.

Lin J, et al. 2004. A universal telomerase RNA core structure includes structured motifs required for binding the telomerase reverse transcriptase protein. Proc Natl Acad Sci USA. 101(41):14713–14718.

Liu C, Schroeder AA, Kapur V, Abrahamsen MS. 1998. Telomeric sequences of Cryptosporidium parvum. Mol Biochem Parasitol. 94(2):291–296.

Longhese MP, Anbalagan S, Martina M, Bonetti D. 2012. The role of shelterin in maintaining telomere integrity. Front Biosci. 17(1):1715–1728.

Louis EJ, Haber JE. 1992. The structure and evolution of subtelomeric Y' repeats in Saccharomyces cerevisiae. Genetics 131(3):559–574.

Lue NF. 2010. The plasticity of telomere maintenance mechanisms in yeast. Trends Biochem Sci. 35(1):8–17.

Lue NF. 2018. Evolving linear chromosomes and telomeres: a C-strand-centric view. Trends Biochem Sci. 43(5):314–326.

Ma LJ, et al. 2009. Genomic analysis of the basal lineage fungus Rhizopus oryzae reveals a whole-genome duplication. PLoS Genet. 5(7):e1000549.

McCormick-Graham M, Haynes WJ, Romero DP. 1997. Variable telomeric repeat synthesis in Paramecium tetraurelia is consistent with misincorporation by telomerase. EMBO J. 16(11):3233–3242.

McEachern MJ, Blackburn EH. 1994. A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts. Proc Natl Acad Sci USA. 91(8):3453–3457.

McEachern MJ, Krauskopf A, Blackburn EH. 2000. Telomeres and their control. Annu Rev Genet. 34(1):331–358.

McKnight TD, Říha K, Shippen DE. 2002. Telomeres, telomerase, and stability of the plant genome. Plant Mol Biol. 48(4):331–337.

Meyne J, Ratliff RL, Moyzis RK. 1989. Conservation of the human telomere sequence (TTAGGG)n among vertebrates. Proc Natl Acad Sci USA. 86(18):7049–7053.

Milo S, Harari-Misgav R, Hazkani-Covo E, Covo S. 2019. Limited DNA repair gene repertoire in ascomycete yeast revealed by comparative genomics. Genome Biol Evol. 11(12):3409–3423.

Monaghan P, Eisenberg DTA, Harrington L, Nussey D. 2018. Understanding diversity in telomere dynamics. Philos Trans R Soc B. 373(1741):20160435.

Morrison HG, et al. 2007. Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. Science 317(5846):1921–1926.

Müller F, Wicky C, Spicher A, Tobler H. 1991. New telomere formation after developmentally regulated chromosomal breakage during the process of chromatin diminution in Ascaris lumbricoides. Cell 67(4):815–822.

Murray AW, Schultes NP, Szostak JW. 1986. Chromosome length controls mitotic chromosome segregation in yeast. Cell 45(4):529–536.

Nielsen L, Edström JE. 1993. Complex telomere-associated repeat units in members of the genus Chironomus evolve from sequences similar to simple telomeric repeats. Mol Cell Biol. 13(3):1583–1589.

Nosek J, Tomáška Ľ, Fukuhara H, Suyama Y, Kováč L. 1998. Linear mitochondrial genomes: 30 years down the line. Trends Genet. 14(5):184–188.

Olson PD, et al. 2020. Complete representation of a tapeworm genome reveals chromosomes capped by centromeres, necessitating a dual role in segregation and protection. BMC Biol. 18(1):165.

Osanai M, Kojima KK, Futahashi R, Yaguchi S, Fujiwara H. 2006. Identification and characterization of the telomerase reverse transcriptase of Bombyx mori (silkworm) and Tribolium castaneum (flour beetle). Gene 376(2):281–289.

Petracek ME, Lefebvre PA, Silflow CD, Berman J. 1990. Chlamydomonas telomere sequences are A+T-rich but contain three consecutive G-C base pairs. Proc Natl Acad Sci USA. 87(21):8222–8226.

Peška V, et al. 2015. Characterisation of an unusual telomere motif (TTTTTTAGGG)n in the plant Cestrum elegans (Solanaceae), a species with large genome. Plant J. 82(4):644–654.

Peška V, et al. 2020. Human-like telomeres in Zostera marina reveal a mode of transition from the plant to the human telomeric sequences. J Exp Bot. 71(19):5786–5793.

Peška V, Garcia S. 2020. Origin, diversity, and evolution of telomere sequences in plants. Front Plant Sci. 11:117.

Peyret P. 2001. Sequence and analysis of chromosome I of the amitochondriate intracellular parasite Encephalitozoon cuniculi (Microspora). Genome Res. 11(2):198–207.

Podlevsky JD, Bley CJ, Omana RV, Qi X, Chen J-L. 2007. The telomerase database. Nucleic Acids Res. 36(Database):D339–343.

Qin Z, Cohen SN. 1998. Replication at the telomeres of the Streptomyces linear plasmid pSLA2. Mol Microbiol. 28(5):893–903.

Ramírez L, Pérez G, Castanera R, Santoyo F, Pisabarro AG. 2011. Basidiomycetes telomeres—a bioinformatic approach. In: Mahdavi MA, editor. Bioinformatics: trends and methodologies. Rijeka (Croatia): InTech. p. 393–424.

Rhind N, et al. 2011. Comparative functional genomics of the fission yeasts. Science 332(6032):930–936.

Ribaud V, Ribeyre C, Damay P, Shore D. 2012. DNA-end capping by the budding yeast transcription factor and subtelomeric binding protein Tbf1. EMBO J 31(1):138–149.

Richards EJ, Ausubel FM. 1988. Isolation of higher eukaryotic telomere from Arabidopsis thaliana. Cell 53(1):127–136.

Richards EJ, Chao S, Vongs A, Yang J. 1992. Characterization of Arabidopsis thaliana telomeres isolated in yeast. Nucleic Acids Res. 20(15):4039–4046.

Rodriguez-Centeno J, Manguán-García C, Perona R, Sastre L. 2019. Structure of Dictyostelium discoideum telomeres. Analysis of possible replication mechanisms. PLoS One 14(9):e0222909.

Romero DP, Blackburn EH. 1991. A conserved secondary structure for telomerase RNA. Cell 67(2):343–353.

Sahara K, Marec F, Traut W. 1999. TTAGG Telomeric repeats in chromosomes of some insects and other arthropods. Chromosome Res. 7(6):449–460.

Saiga H, Edström JE. 1985. Long tandem arrays of complex repeat units in Chironomus telomeres. EMBO J 4(3):799–804.

Sepšiová R, et al. 2016. Evolution of telomeres in Schizosaccharomces pombe and its possible relationship to the diversification of telomere binding proteins. PLoS One 11(4):e0154225.

Shen X-X, et al. 2016. Reconstructing the backbone of the Saccharomycotina yeast phylogeny using genome-scale data. G3 6(12):3927–3939.

Shen X-X, et al. 2018. Tempo and mode of genome evolution in the budding yeast subphylum. Cell 175(6):1533–1545.

Shippen DE, McKnight TD. 1998. Telomeres, telomerase and plant development. Trends Plant Sci. 3(4):126–130.

Shirley MW. 1994. The genome of Eimeria tenella: further studies on its molecular organisation. Parasitol Res. 80(5):366–373.

Smogorzewska A, de Lange T. 2004. Regulation of telomerase by telomeric proteins. Annu Rev Biochem. 73(1):177–208.

Sohanpal BK, Morzaria SP, Gobright EI, Bishop RP. 1995. Characterisation of the telomeres at opposite ens of a 3 Mb Theileria parva chromosome. Nucleic Acids Res. 23(11):1942–1947.

Steinberg-Neifach O, Lue NF. 2015. 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.

Stewart JA, Chaiken MF, Wang F, Price CM. 2012. Maintaining the end: roles of telomere proteins in end-protection, telomere replication and length regulation. Mutat Res. 730(1–2):12–19.

Sýkorová E, Leitch AR, Fajkus J. 2006. Asparagales telomerases which synthesize the human type of telomeres. Plant Mol Biol. 60(5):633–646.

Sýkorová E, Lim KY, Chase MW, et al. 2003. The absence pf *Arabidopsis*-type telomeres in *Cestrum* and closely related genera *Vestia* and *Sessea* (Solanaceae): first evidence from eudicots. Plant J. 34(3):283–291.

Sýkorová E, Lim KY, Kunická Z, et al. 2003. Telomere variability in the monocotyledonous plant order Asparagales. Proc R Soc Lond B. 270(1527):1893–1904.

Theimer CA, Feigon J. 2006. Structure and function of telomerase RNA. Curr Opin Struct Biol. 16(3):307–318.

Tomáška Ľ, et al. 2019. Commentary: single-stranded telomere-binding protein employs a dual rheostat for binding affinity and specificity that drives function. Front Genet. 9:742.

Tomáška Ľ, Cesare AJ, AlTurki TM, Griffith JD. 2020. Twenty years of t-loops: a case study for the importance of collaboration in molecular biology. DNA Repair 94:102901.

Tomáška Ľ, Nosek J. 2009. Telomere heterogeneity: taking advantage of stochastic events. FEBS Lett. 583(7):1067–1071.

Torruella G, et al. 2012. Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. Mol Biol Evol. 29(2):531–544.

Tsai IJ, et al. 2014. Comparative genomics of *Taphrina* fungi causing varying degrees of tumorous deformity in plants. Genome Biol Evol. 6(4):861–872.

Underwood AP, Louis EJ, Borts RH, Stringer JR, Wakefield AE. 1996. *Pneumocystis carinii* telomere repeats are composed of TTAGGG and the subtelomeric sequence contains a gene encoding the major surface glycoprotein. Mol Microbiol. 19(2):273–281.

Valente LP, et al. 2013. Myb-domain protein Teb1 controls histone levels and centromere assembly in fission yeast. EMBO J. 32(3):450–460.

Villasante A, Abad JP, Méndez-Lago M. 2007. Centromere were derived from telomeres during the evolution of the eukaryotic chromosome. Proc Natl Acad Sci USA. 104(25):10542–10547.

Višacká K, et al. 2012. Synergism of the two Myb domains of Tay1 protein results in high affinity binding to telomeres. J Biol Chem. 287(38):32206–32215.

Vítková M, Král J, Traut W, Zrzavý J, Marec F. 2005. The evolutionary origin of insect telomeric repeats (TTAGG)*n*. Chromosome Res. 13(2):145–156.

Wahlin J, Cohn M. 2000. *Saccharomyces cerevisiae* RAP1 binds to telomeric sequences with spatial flexibility. Nucleic Acids Res. 28(12):2292–2301.

Webb CJ, Zakian VA. 2008. Identification and characterization of the *Schizosaccharomyces pombe* TER1 telomerase RNA. Nat Struct Mol Biol. 15(1):34–42.

Weiss H, Scherthan H. 2002. *Aloe* spp.—plants with vertebrate-like telomeric sequences. Chromosome Res. 10(2):155–164.

Wellinger RJ, Zakian VA. 2012. Everything you ever wanted to know about *Saccharomyces cerevisiae* telomeres: beginning to end. Genetics 191(4):1073–1105.

Wicky C, et al. 1996. Telomeric repeats (TTAGGC)n are sufficient for chromosome capping function in *Caenorhabditis elegans*. Proc Natl Acad Sci USA. 93(17):8983–8988.

Wu RA, Upton HE, Vogan JM, Collins K. 2017. Telomerase mechanism of telomere synthesis. Annu Rev Biochem. 86(1):439–460.

Yu EY, et al. 2020. Structurally distinct duplex telomere repeat-binding proteins in *Ustilago maydis* execute specialized, non-overlapping functions in telomere recombination and telomere protection. Commun Biol. 3(1):777.

Yu EY, Kojic M, Holloman WK, Lue NF. 2013. Brh2 and Rad51 promote telomere maintenance in *Ustilago maydis*, a new model system of DNA repair proteins at telomeres. DNA Repair 12(7):472–479.

Zauner S, et al. 2000. Chloroplast protein and centrosomal genes, a tRNA intron, and odd telomeres in an unusually compact eukaryotic genome, the cryptomonad nucleomorph. Proc Natl Acad Sci USA. 97(1):200–205.

**Associate editor:** Wolfe Kenneth