THE EVOLVING TELOMERES

EDITED BY : Arthur J. Lustig and Kurt Runge
PUBLISHED IN : Frontiers in Genetics

ADAPT
RESIST
EXTINCT
CHANGE
EVOLVE

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: **researchtopics@frontiersin.org**

THE EVOLVING TELOMERES

Topic Editors:
Arthur J. Lustig, Tulane University, USA
Kurt Runge, Cleveland Clinic Foundation, USA
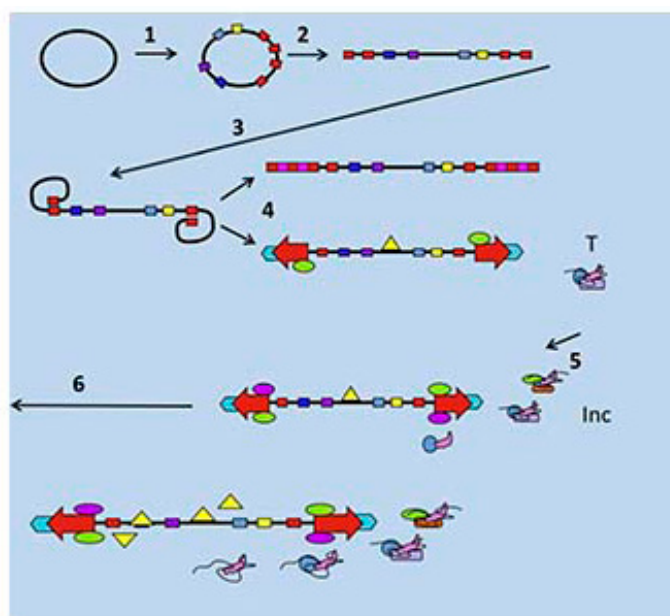
Hypotheses for the evolving telomere. The circular genomes that predominate in biology gradually acquire repeats (Step 1 in figure) in the form autocatalytic elements such as group II introns. Linearization (2) and stabilization by homology and protein binding (3) effectively caps the end in a structure similar to metazoan t-loops (see de Lange). The evolution of retrotransposons from group II introns provides a way to cap the end by repeated transposition, as in Drosophila (4 top, see Savant and Deininger) or leads to genesis of a telomerase reverse-transcriptase that can add repeats to chromosome ends that are bound by specific proteins (4 bottom). Selection in response to genomic stresses leads to duplication and exaptation of the telomerase long non-coding RNA, different DNA binding proteins, and alterations in telomere sequence (see Lue and Jiang) to provide specialized functions at the telomere and elsewhere in the genome, while eliminating others (5, see Shippen and Nelson, Riha and Fulcher, Lustig). This expansion of factors occurs in part by recruitment of chromosomal proteins (yellow triangle) to telomeres for specific telomere functions and, perhaps, as reservoir of factors to act at internal sites during genomic stress (6, see Mattarocci et al.).
Figure by Arthur J. Lustig and Kurt Runge.

Cover image: [iqoncept] © 123RF.COM

What controls the different rates of evolution to give rise to conserved and divergent proteins and RNAs? How many trials until evolution can adapt to physiological changes? Every organism has arisen through multiple molecular changes, and the mechanisms that are employed (mutagenesis, recombination, transposition) have been an issue left to the elegant discipline of evolutionary biology. But behind the theory are realities that we have yet to ascertain: How does an evolving cell accommodate its requirements for both conserving its essential functions, while also providing a selective advantage? In this volume, we focus on the evolution of the eukaryotic telomere, the ribo-nuclear protein complex at the end of a linear chromosome. The telomere is an example of a single chromosomal element that must function to maintain genomic stability. The telomeres of all species must provide a means to avoid the attrition from semi-conservative DNA replication and a means of telomere elongation (the telomere replication problem). For example, telomerase is the most well-studied mechanism to circumvent telomere attrition by adding the short repeats that constitutes most telomeres. The telomere must also guard against the multiple activities that can act on an unprotected double strand break requiring a window (or checkpoint) to compensate for telomere sequence loss as well as protection against non-specific processes (the telomere protection problem). This volume describes a range of methodologies including mechanistic studies, phylogenetic comparisons and data-based theoretical approaches to study telomere evolution over a broad spectrum of organisms that includes plants, animals and fungi. In telomeres that are elongated by telomerases, different components have widely different rates of evolution. Telomerases evolved from roots in archaebacteria including splicing factors and LTR-transposition. At the conserved level, the telomere is a rebel among double strand breaks (DSBs) and has altered the function of the highly conserved proteins of the ATM pathway into an elegant means of protecting the chromosome end and maintaining telomere size homeostasis through a competition of positive and negative factors. This homeostasis, coupled with highly conserved capping proteins, is sufficient for protection. However, far more proteins are present at the telomere to provide additional species-specific functions. Do these proteins provide insight into how the cell allows for rapid change without self-destruction?

Table of Contents

Editorial: The Evolving Telomeres

Kurt W. Runge[1] and Arthur J. Lustig[2]*

[1] *Department of Immunology, The Lerner Institute, Cleveland Clinic Foundation, Cleveland, OH, USA, [2] Department of Biochemistry and Molecular Biology, The Tulane Medical School and Tulane Cancer Center, New Orleans, LA, USA*

The Editorial on the Research Topic

The Evolving Telomeres

The study of the evolution of the end of chromosomes, or telomeres, has moved from the abstract to molecular observations and mechanistic possibilities. Although successful end-replication and end-protection are the primary driving forces acting at all telomeres (de Lange, 2009), the studies presented in this issue reveal apparent similarities, surprising differences, and new functions for telomere binding proteins (TeloBPs). These advances in molecular genetics of both common and more diverse organisms should lead to specific hypotheses for the roles of these proteins both at telomeres and throughout the genome and toward a broader view of how evolution solves different problems that occur in biology. The next step will be the experimental testing of evolutionary hypotheses.

As a reflection of the molecular advances, we framed the series "The Evolving Telomeres". We have covered information from multiple systems that use a variety of mechanisms. These include studies in Neal Lue's lab regarding the analysis of work in yeasts belonging to Saccharomycotina involving the co-evolution of single-stranded and double-stranded sequence TeloBPs as a function of telomeric sequence (Steinberg-Neifach and Lue). They find that proteins accommodate the differing sequence through duplication and divergence of functional proteins, combinatorial site recognition, and greater protein flexibility. David Shore's laboratory reviewed the apparent differences and similarities in the Rif1 protein (Mattarocci et al.) in yeasts and humans. Rif1 was first defined in budding yeast as a negative regulator of telomere size that counteracted the activation effects of Tel1 (ATM) binding to short telomeres (Hector et al., 2007; Sabourin et al., 2007). The multi-functional Rif1, on the other hand, is delivered to the terminus in greater amounts in longer telomeres that have a greater abundance of the major yeast TeloBP, Rap1, thereby displacing Tel1 (Chang et al., 2007; Hirano et al., 2009; Martina et al., 2012). These activities form a feedback mechanism that protects the telomere against non-productive repair such as the formation of end-to-end fusions. This dynamic homeostasis acts in a cap-like function, termed the anti-checkpoint (Ribeyre and Shore, 2012). Feedback mechanisms seem to be ubiquitous among telomeres.

One major issue is the source of the many discontinuities in the evolution in plant, fungal, and mammalian telomeres. Two studies probed some of the unique characteristics of plants. Dorothy Shippen's laboratory (Nelson and Shippen) studied the participation of long nuclear RNAs in plant telomere regulation. Among these is the telomerase RNA and an entire group of related RNAs, many of which act on telomerase, even as a negative regulator. These RNAs are absent from metazoans, illustrating how the metaphyta have likely adapted the system of RNA-based regulation to telomeres. This finding may reflect the high predominance of RNA-based defense mechanisms in plants, especially against transposons present in most of the genome (Shabalina and Koonin, 2008). Karel Riha's laboratory contributed an experimental study of another example of differing solutions to end-protection (Fulcher and Riha). One issue in Arabidopsis and many other plants has been

the lack of TRF-like (TRFL) factors that are so common in vertebrate cells. The major telomere binding proteins in vertebrates is TRF1, and often, TRF2. These proteins form the backbone of the shelterin complex, involved in both end-replication and protection (Karlseder et al., 2003; Wu and de Lange, 2008). The strangest observation is that TRFL are present and located at telomeres, but serve no obvious function. To rule out the possibility of functional redundancy, the authors' produced genetic knockouts of the possible functional TRF-like proteins with no effect on telomeres or growth. This result is in sharp contrast to the effects of TRF1 and TRF2 loss in vertebrates. Their data all but eliminate the chance for the presence that a homolog to the vertebrate telomere repeat factor (TRF1) that is important at Arabidopsis telomeres (Shakirov et al., 2008). Rather, a simple algal-related protein performs many of the TRF1 functions in Arabidopsis (Mozgova et al., 2008), leading to speculation on the odd rapid evolution of TeloBPs. Plants appear to have adapted telomeres to physiological requirements since the divergence of the original common ancestor that gave rise to metazoans.

Some components of telomeres are conserved such as the Mre11/Rad50/NBS complex and the Cdt1/Stn1Ten1 complex that assist in end protection. However, many others rapidly change with differing physiological and selective forces that maintain genome stability and cell survival. Art Lustig presented a hypothesis that evolution could cause rapid changes as a consequence of formation and divergence of paralogs (Lustig). The hypothesis argues that rapid evolution is driven by the requirement for genomic stability and, in some cases, by telomere stress response that increases the rate of paralogy and divergence. In fact, this result helps to explain the TeloBP divergence among fungal, invertebrates, vertebrate and plant species that have been investigated.

Evolution has provided multiple solutions to the end-replication problem of linear chromosomes besides telomerase and even telomeres. Some bacteriophages replicate the end by circularization or recombination (Lopes et al., 2010). Both adenovirus and the bacterium that causes Lyme disease, *Borrelia burgdorferi*, have chromosome ends capped by covalently bound proteins (Chaconas, 2005), and *Drosophila* and other dipterans have transposons at their chromosome termini (Villasante et al., 2008). The role of non-LTR retro-transposition in the evolution of telomerase has been controversial.

Indeed, in analyzing the origin of telomerase, (de Lange) proposes a theoretical scheme for type II introns, coupled with the formation of primitive t-loops, to evolve into telomerase,

independent of non-LTR retro-transpositions (Lambowitz and Belfort, 2015). Nevertheless, the review by Servant and Deininger focuses on the use in extant organisms of non-LTR retro-transposition in telomerase-positive cells, providing an example of a mechanism that persists and even co-exists with telomerase through evolution. The bottom line of these studies is the diversity of telomeric processes. This variety could be put into a broader context by a more extensive study of diverse organisms.

A major future goal, at least for microbes, is to test hypotheses regarding telomere evolution. These experiments use techniques for growth of cells at a constant density. One of these instruments used for these experiments is the turbidostat (Gresham and Dunham, 2014; Matteau et al., 2015; Takahashi et al., 2015) that can differentiate between the altered molecular changes that arise during the evolution of cells. Another exciting aspect of this work is that these experiments represent real-time (albeit manipulated) evolution. The artificial evolutionary approach is having signs of success in yeast and microbes under different conditions, such as oxidative stress (Raso et al., 2012) and these successes will undoubtedly continue.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Chaconas, G. (2005). Hairpin telomeres and genome plasticity in Borrelia: all mixed up in the end. *Mol. Microbiol.* 58, 625–635. doi: 10.1111/j.1365-2958.2005.04872.x

Chang, M., Arneric, M., and Lingner, J. (2007). Telomerase repeat addition processivity is increased at critically short telomeres in a Tel1-dependent manner in *Saccharomyces cerevisiae*. *Genes Dev.* 21, 2485–2494. doi: 10.1101/gad.1588807

de Lange, T. (2009). How telomeres solve the end-protection problem. *Science* 326, 948–952. doi: 10.1126/science.1170633

Gresham, D., and Dunham, M. J. (2014). The enduring utility of continuous culturing in experimental evolution. *Genomics* 104, 399–405. doi: 10.1016/j.ygeno.2014.09.015

Hector, R. E., Shtofman, R. L., Ray, A., Chen, B.-R., Nyun, T., Berkner, K. L. et al. (2007). Tel1p preferentially associates with short telomeres to stimulate their elongation. *Mol. Cell* 27, 851–858. doi: 10.1016/j.molcel.2007.08.007

Hirano, Y., Fukunaga, K., and Sugimoto, K. (2009). Rif1 and rif2 inhibit localization of tel1 to DNA ends. *Mol. Cell* 33, 312–322. doi: 10.1016/j.molcel.2008.12.027

Karlseder, J., Kachatrian, L., Takai, H., Mercer, K., Hingorani, S., Jacks, T., et al. (2003). Targeted deletion reveals an essential function for the telomere length regulator Trf1. *Mol. Cell Biol.* 23, 6533–6541. doi: 10.1128/MCB.23.18.6533-6541.2003

Lambowitz, A. M., and Belfort, M. (2015). Mobile bacterial Group II introns at the crux of eukaryotic evolution. *Microbiol. Spectr.* 3:MDNA3-0050-2014. doi: 10.1128/microbiolspec.mdna3-0050-2014

Lopes, A., Amarir-Bouhram, J., Faure, G., Petit, M. A., and Guerois, R. (2010). Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. *Nucleic Acids Res.* 38, 3952–3962. doi: 10.1093/nar/g kq096

Martina, M., Clerici, M., Baldo, V., Bonetti, D., Lucchini, G., and Longhese, M. P. (2012). A balance between Tel1 and Rif2 activities regulates nucleolytic processing and elongation at telomeres. *Mol. Cell Biol.* 32, 1604–1617. doi: 10.1128/MCB.06547-11

Matteau, D., Baby, V., Pelletier, S., and Rodrigue, S. (2015). A small-volume, lowcost, and versatile continuous culture device. *PLoS ONE* 10:e0133384. doi: 10.1371/journal.pone.0133384

Mozgova, I., Schrumpfova, P. P., Hofr, C., and Fajkus, J. (2008). Functional characterization of domains in AtTRB1, a putative telomere-binding protein in Arabidopsis thaliana. *Phytochemistry* 69, 1814–1819. doi: 10.1016/j.phytochem.2008.04.001

Raso, S., Van Genugten, B., Vermuë, M., and Wijffels, R. H. (2012). Effect of oxygen concentration on the growth of Nannochloropsis sp. at low light intensity. *J. Appl. Phycol.* 24, 863–871. doi: 10.1007/s10811-011-9706-z

Ribeyre, C., and Shore, D. (2012). Anticheckpoint pathways at telomeres in yeast. *Nat. Struct. Mol. Biol.* 19, 307–313. doi: 10.1038/nsmb.2225

Sabourin, M., Tuzon, C. T., and Zakian, V. A. (2007). Telomerase and Tel1p preferentially associate with short telomeres in *S. cerevisiae. Mol Cell.* 27, 550–561. doi: 10.1016/j.molcel.2007.07.016

Shabalina, S. A., and Koonin, E. V. (2008). Origins and evolution of eukaryotic RNA interference. *Trends Ecol. Evol.* 23, 578–587. doi: 10.1016/j.tree.2008.06.005

Shakirov, E. V., Salzberg, S. L., Alam, M., and Shippen, D. E. (2008). Analysis of carica papaya telomeres and telomere-associated proteins: insights into the evolution of telomere maintenance in brassicales. *Trop. Plant Biol.* 1, 202–215. doi: 10.1007/s12042-008-9018-x

Takahashi, C. N., Miller, A. W., Ekness, F., Dunham, M. J., and Klavins, E. (2015). A low cost, customizable turbidostat for use in synthetic circuit characterization. *ACS Synth. Biol.* 4, 32–38. doi: 10.1021/sb500165g

Villasante, A., de Pablos, B., Méndez-Lago, M., and Abad, J. P. (2008). Telomere maintenance in Drosophila: rapid transposon evolution at chromosome ends. *Cell Cycle* 7, 2134–2138. doi: 10.4161/cc.7.14.6275

Wu, P., and de Lange, T. (2008). No overt nucleosome eviction at deprotected telomeres. *Mol. Cell Biol.* 28, 5724–5735. doi: 10.1128/MCB.01764-07

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Telomere DNA recognition in Saccharomycotina yeast: potential lessons for the co-evolution of ssDNA and dsDNA-binding proteins and their target sites

Olga Steinberg-Neifach[1,2] and Neal F. Lue[1]*

[1] Department of Microbiology and Immunology, W. R. Hearst Microbiology Research Center, Weill Medical College, Cornell University, New York, NY, USA, [2] Hostos Community College, City University of New York, Bronx, NY, USA

In principle, alterations in the telomere repeat sequence would be expected to disrupt the protective nucleoprotein complexes that confer stability to chromosome ends, and hence relatively rare events in evolution. Indeed, numerous organisms in diverse phyla share a canonical 6 bp telomere repeat unit (5′-TTAGGG-3′/5′-CCCTAA-3′), suggesting common descent from an ancestor that carries this particular repeat. All the more remarkable, then, are the extraordinarily divergent telomere sequences that populate the Saccharomycotina subphylum of budding yeast. These sequences are distinguished from the canonical telomere repeat in being long, occasionally degenerate, and frequently non-G/C-rich. Despite the divergent telomere repeat sequences, studies to date indicate that the same families of single-strand and double-strand telomere binding proteins (i.e., the Cdc13 and Rap1 families) are responsible for telomere protection in Saccharomycotina yeast. The recognition mechanisms of the protein family members therefore offer an informative paradigm for understanding the co-evolution of DNA-binding proteins and the cognate target sequences. Existing data suggest three potential, inter-related solutions to the DNA recognition problem: (i) duplication of the recognition protein and functional modification; (ii) combinatorial recognition of target site; and (iii) flexibility of the recognition surfaces of the DNA-binding proteins to adopt alternative conformations. Evidence in support of these solutions and the relevance of these solutions to other DNA-protein regulatory systems are discussed.

Keywords: telomere, telomere-binding proteins, Saccharomycotina, co-evolution of DNA and binding proteins, gene duplication, dimerization, Rap1, Cdc13

Overview

Linear eukaryotic chromosome termini are stabilized by telomeres, which are specialized nucleoprotein complexes that suppress the recognition of the ends as double strand breaks (DSBs; de Lange, 2009; O'Sullivan and Karlseder, 2010; Jain and Cooper, 2011). This stabilization is mediated by a collection of telomeric proteins that associate directly or indirectly with the repetitive telomeric DNAs and that suppress the action of checkpoint and repair proteins. The DNA component of telomeres typically comprises a duplex region of hundreds to thousands of nucleotides and a 3′ overhang

of tens to hundreds of nucleotides (referred to as the G-tail because of its G-rich nucleotide composition). Both the duplex region and 3′ overhang are comprised 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. Because proteins recruited to the duplex telomere repeats and G-tails are both required for telomere stability, the duplex/G-tail DNA structural arrangement at chromosome ends is evidently essential for telomere function. Besides telomere protection, the other major function of telomere-bound proteins is to maintain telomere DNAs. Despite their fundamental importance, telomere DNAs can experience progressive loss owing to incomplete end replication (Olovnikov, 1996), as well as drastic truncation owing to recombinational excision or replication fork collapse (Lustig, 2003; Lansdorp, 2005). 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 (Autexier and Lue, 2006; Blackburn and Collins, 2011; Nandakumar and Cech, 2013; Pfeiffer and Lingner, 2013; Lue et al., 2014). Not surprisingly, these telomere extension activities are subjected to elaborate control by telomere-bound proteins in order to maintain telomere lengths within a size range that is appropriate for telomere function.

A particularly prevalent telomere repeat unit, found in various fungi, plant, metazoans, and protozoa, is 5′-TTAGGG-3′/5′-CCCTAA-3′. In organisms with this telomere repeat unit, the duplex region is typically recognized directly by a member of the telomere repeat binding factor (TRF) protein family, whereas the 3′ overhang bound directly by that of the protection of telomeres 1 (POT1) protein family (**Figure 1**). In most mammalian cells, for example, two TRF homologs (TRF1 and TRF2) and a POT1 homolog constitute the three direct DNA-binding components

FIGURE 1 | The distinctive telomere protective complexes in mammals and in Saccharomycotina yeast. (A) The mammalian telomeres are bound by a six-protein complex collectively named shelterin. Within the telomere nucleoprotein complex, duplex telomeres are bound directly by TRF1 and TRF2, and G-tails are bound directly by POT1. The mammalian CST (CTC1-STN1-TEN1) complex plays minimal roles in telomere protection, but is crucial for regulating telomere DNA synthesis. **(B)** The telomere complexes of Saccharomycotina yeast display considerable differences from those in other phyla; the duplex telomeres and G-tails in Saccharomycotina yeast are bound by Rap1 and Cdc13, respectively. Like CTC1, the fungal Cdc13 is part of the CST complex that also contains Stn1 and Ten1. However, unlike mammalian CST, the fungal CST complex is crucial both for telomere protection and for regulating telomere DNA synthesis.

of the six-protein "shelterin" complex that collectively protects the duplex telomeres and G-tails (**Figure 1**; de Lange, 2009). In fission yeast, on the other hand, a single TRF homolog (Taz1) and a POT1 homolog (Pot1) account for direct DNA-binding by a somewhat different version of the shelterin complex (Jain and Cooper, 2011). Both the TRF and POT1 family members have been subjected to extensive structural and functional investigations, and the molecular bases of their DNA recognition mechanisms are understood at the level of atomic resolution structures (Fairall et al., 2001; Lei et al., 2003, 2004; Court et al., 2005). TRF proteins form homodimers through their N-terminal TRF homology (TRFH) domain, and use the resulting tandem C-terminal Myb DNA-binding domains (DBDs) to make contacts with two adjacent repeat units. POT1 uses a pair of OB (oligonucleotide/oligosaccharide binding) folds to interact with ∼10 nt of the G-rich, 3′ end containing strand of telomeres [i.e., the (TTAGGG)n strand]. Sequence recognition by both proteins is highly specific such that most nucleotide substitutions in the target DNA cause a substantial loss in binding affinity. This sequence specificity is to be expected: given the capacity of the telomere proteins to "stabilize" DNA ends and prevent recombination and end-joining, promiscuous binding of these proteins to DNA DSBs would presumably be detrimental to the cells.

Implicit in the foregoing discussion are the substantial constraints imposed on the telomere nucleoprotein system during evolution. The greater constraints of the telomere system are evident when one compares its parts to those of a more circumscribed system consisting of, e.g., a transcription factor and its target site. In the latter case, a point mutation in the DNA target site could be readily accommodated by perhaps a few changes in the transcription factor DNA-binding surface. However, a comparable point mutation in the canonical telomere repeat unit is likely to cause greater disruption of cellular function and require greater compensatory adjustments. Loss of TRF or POT1 binding to the mutated repeat will probably cause extensive changes in the chromatin structure of telomeres. Conversely, restoration of normal telomere structure in this setting may require multiple changes in the binding surfaces of both TRF and POT1. Viewed in this light, it is perhaps not surprising that numerous present-day organisms in diverse phyla have retained the canonical, presumably ancient telomere repeat sequence and TRF and POT1 homologs. Examples of such organisms include fungi (e.g., basidiomycotina and pezizomycotina), metazoans (e.g., vertebrates), plants (e.g., *Aloe sp.*, *Hyacinthella dalmatica*, and *Othocallis siberica*), and even protists (e.g., trypanosome and *Leishmania*), where the TTAGGG repeat is relatively uncommon (Podlevsky et al., 2008).

The Unusual Telomere Repeats of Saccharomycotina Fungi

One group of organisms with telomere systems that deviate from the canonical system is found in the Saccharomycotina subphylum of budding yeast (**Figure 1**). They include a widely used model organism, several pathogenic fungi, and others (*Saccharomyces*, *Kluyveromyces*, and *Candida*). The telomere repeats in these organisms are extraordinarily divergent and differ from the canonical repeat in being long, occasionally
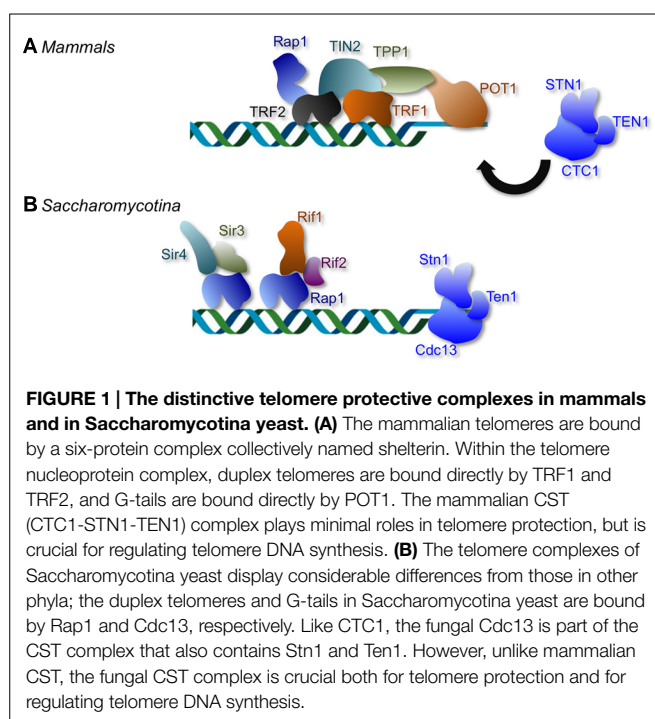
degenerate, and frequently non-G/C-rich. Notably, the telomeres of Saccharomycotina yeast are not bound directly by TRF and POT1 family members, but rather by two other distinct protein families named Rap1 and Cdc13, suggesting that the acquisition of atypical telomere DNA sequences was accompanied by a substantial remodeling of the telomere nucleoprotein structure (**Figure 1**). Remarkably, homologs or structural equivalents of Rap1 and Cdc13 also exist in organisms with the canonical telomere repeat sequence, but these homologs or equivalents clearly mediate distinct functions in these organisms. Mammalian RAP1, while a component of the shelterin complex, exhibits low affinity for telomere repeats and is localized to telomeres primarily through an interaction with TRF2 (Li et al., 2000; Arat and Griffith, 2012). The mammalian equivalent of Cdc13, named CTC1, is like Cdc13, a component of the CST (CTC1-STN1-TEN1) complex that also contains Stn1 and Ten1 (Miyake et al., 2009; Surovtseva et al., 2009). However, unlike Cdc13, CTC1 has little function in telomere protection, and appears to be primarily involved in regulating telomere DNA synthesis (Price et al., 2010; Stewart et al., 2012). The existence of mammalian CTC1 and RAP1 strongly suggests that fungal Cdc13 and Rap1 were not acquired *de novo*, but were co-opted to perform a new telomere function (i.e., direct telomere DNA-binding) as a pre-existing telomere component. Evolutionary models that account for the transition from the canonical telomere architecture to that found in Saccharomycotina yeast have been presented before, and will not be re-iterated in this review (Lue, 2010). Instead, we focus our discussion on a major evolutionary conundrum presented by the telomeres of this group of fungi, i.e., the DNA recognition challenge posed by rapidly evolving telomere sequence.

Interestingly, even though Rap1 exhibits little sequence similarity to TRF and has a distinct domain organization, it also utilizes Myb-like homeodomains for telomere DNA-binding. Likewise, Cdc13 can hardly be aligned to POT1 at the sequence level, yet both protein families employ the same OB fold scaffold for single-strand DNA (ssDNA) recognition. Unlike TRF and POT1, however, fungal Rap1s and Cdc13s are tasked with recognizing a very diverse collection of telomere target sequences. According to the estimates of evolutionary models, the Saccharomycotina yeasts share a common ancestor as recently as 300 million years ago, and yet collectively possess more than 20 distinct telomere repeats (Pesole et al., 1995; Hedges, 2002). *A priori*, this degree of evolutionary divergence can only be considered highly unusual. In terms of coding sequences, the *Candida* and *Saccharomyces* genomes are approximately as divergent as those of fish and humans, which possess the same canonical telomere sequence (Dujon et al., 2004). How then, do the major double-strand (ds) and ss telomere binding proteins (i.e., Rap1 and Cdc13) acquire the correct sequence-specificity for the rapidly changing telomere sequence? Even though we are far from having a complete answer, recent studies suggest a number of solutions to this challenge. In the following sections, we discuss in detail the structure, function and evolution of Rap1 and Cdc13, with a special emphasis on their evolutionary plasticity and their versatile DNA binding mechanisms that enables them to adapt to the multiplicity of target sequences. (In discussing the target sequence of Rap1, we will refer to the G-strand sequence such that the same strand is used in

describing both the Rap1 and Cdc13 targets. This is in contrast to the majority of previous articles that characterize Rap1 binding sites.)

Rap1

Rap1 (Repressor activator protein 1, also originally known as GRF1 or TUF1), a conserved telomere protection factor, exhibit remarkable functional versatility (Shore, 1994). Notably, it was first discovered in *Saccharomyces cerevisiae* as a transcriptional regulator of numerous metabolic genes (Huet et al., 1985). Subsequent studies implicate Rap1 as a key component of the mating type silencer as well as the major ds telomere DNA binding protein (Shore et al., 1987; Buchman et al., 1988). That a single factor mediates such diverse functions at distinct chromosomal locations certainly raises interesting mechanistic and evolutionary issues that remain incompletely resolved. The multi-functional nature of Rap1 is evidently conserved in evolution; mammalian Rap1 has also been reported to regulate transcription and protect telomeres (Li et al., 2000; Martinez et al., 2010; Sfeir et al., 2010). However, a recent study suggests that the telomere protection function of human Rap1 may be quite minor and perhaps nonexistent (Kabir et al., 2014). At telomeres, Rap1 displays striking malleability by interacting with different molecular targets in different organisms. In budding yeast, Rap1 binds ds telomere DNAs directly with high affinity and sequence specificity, whereas in fission yeast and mammals (and probably most other organisms), Rap1 is recruited to telomeres through interaction with other telomere proteins such as TRF2 and Taz1 (Li et al., 2000; Kanoh and Ishikawa, 2001). In keeping with its multi-functional nature, *S. cerevisiae* Rap1 possesses a complex domain organization (**Figure 2A**). Near its N-terminus is a BRCA1 C-terminus (BRCT) domain, a presumed protein interaction domain whose targets may include Gcr1, another transcription factor (Lopez et al., 1998). Located centrally is the DBD, which uses a pair of Myb motifs to interact with DNA (Giraldo and Rhodes, 1994; Wahlin and Cohn, 2000; **Figures 2A,B**). At the C-terminal end of Rap1 is a purely alpha helical structure Rap1 C-terminus (RCT) that has been shown to mediate interactions with other proteins required for proper telomere structure and function (e.g., Sir3, Sir4, Rif1, and Rif2; Feeser and Wolberger, 2008). Finally, a region between the DBD and RCT has been ascribed a transcriptional activation function (Shore, 1994). With a few exceptions (e.g., *C. albicans* Rap1 lacks RCT) this domain organization is conserved in other Saccharomycotina homologs. However, fission yeast and mammalian Rap1s display structural and functional differences, owing perhaps to their different means of telomere localization; these Rap1s carry a single Myb motif that binds DNA with low affinity, and an RCT that tethers Rap1 to a high-affinity DNA-binding protein (i.e., Taz1 in *S. pombe* and TRF2 in mammals; Li et al., 2000; Kanoh and Ishikawa, 2001; Arat and Griffith, 2012; **Figures 1** and **2A**).

The DNA-binding activity of the Rap1 DBD was first characterized for the *S. cerevisiae* protein, and the binding of *Sc*Rap1 to numerous DNA targets (~200–300 promoters, two silencers, and several telomeric variants) have been investigated individually and at genome-wide scale (Idrissi and Pina,
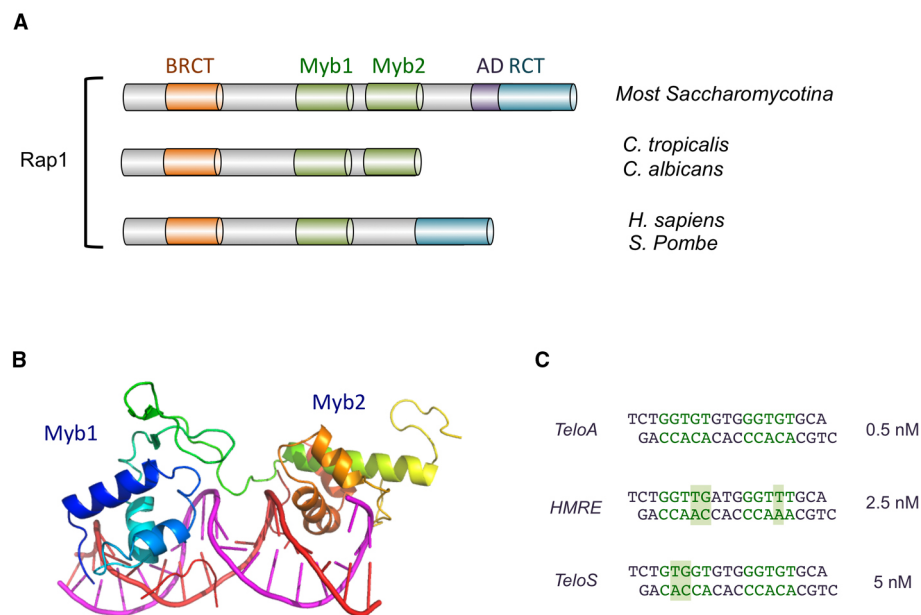
FIGURE 2 | The domain organization of Rap1 and the structure of Rap1_DBD-DNA complex. (A) The domain structures of Rap1 from various Saccharomycotina and other species are illustrated. The BRCT, Myb, AD (activation domain), and RCT (Rap1 C-terminal) domains are displayed in different colors. **(B)** The crystal structure of the Myb1 and Myb2 domains of *S. cerevisiae* Rap1 (shown in color spectrum from blue to orange) bound to its target DNA (shown in magenta and red; PDB ID: 1IGN). **(C)** The sequences of

the three duplex oligonucleotides bound by *Sc*Rap1_DBD in a crystallographic study are displayed. The half sites in each oligo are shown in green, and nucleotides that deviate from a canonical half site (5′-GGTGT-3′/5′-ACACC-3′) are shown with a shaded background. The affinities of Rap1_DBD for each sequence are shown on the right. Other variant targets (e.g., the site upstream of ribosomal protein genes: AAATGTATGGGTGT) have been reported to have comparable affinities (Idrissi et al., 1998; Idrissi and Pina, 1999).

1999; Lieb et al., 2001; Pina et al., 2003; Yarragudi et al., 2007; Rhee and Pugh, 2011). While several consensus sequences for Rap1 have been reported, a frequently noted version is $K_{13'} R_{12'} T_{11'} G_{10'} T_{9'} R_{8'} Y_{7'} G_{6'} G_{5'} G_{4'} T_{3'} G_{2'} T_{1'}$ (Lieb et al., 2001). This somewhat degenerate consensus consists of two half sites, $K_{13'} R_{12'} T_{11'} G_{10'} T_{9'}$ and $G_{5'} G_{4'} T_{3'} G_{2'} T_{1'}$, bound respectively by the second and first Myb motif in Rap1. Subsets of Rap1 targets (e.g., at ribosomal protein gene promoters) exhibit distinctive features with regard to their sequences and dispositions, suggesting that the activities of Rap1 at different chromosomal locations may be modulated by its binding to specific variants of the consensus, i.e., Rap1 may adopt different conformations, and hence recruit different co-factors depending on the specific target sequence to which it is bound (Pina et al., 2003).

As implied from the foregoing discussion, *Sc*Rap1 displays considerable flexibility in recognizing diverse target site sequences. This flexibility stems in part from the ability of the Myb motifs to tolerate many variations in the target sequence (especially the half site comprised of residues 13′–9′) without suffering a loss in binding affinity (Vignais et al., 1990; Idrissi and Pina, 1999). This is evident from the loose consensus reported for Rap1, and especially the more degenerate sequence reported for the first half site. The molecular basis for the flexibility of Rap1 has been investigated through crystallographic analysis of three complexes formed between *Sc*Rap1_DBD and different DNA target sites (**Figure 2C**; Konig et al., 1996; Taylor et al., 2000). Overall, the results indicate that recognition of base pairs that vary between the target sites is accomplished through the utilization

of alternative side-chain conformations and alternative contacts to the nucleotides. In other words, rather than altering its overall configuration, Rap1 modifies its fine surface structure to suit the demand of a particular target sequence. This inherent versatility is not unique to Rap1 (see, e.g., Schwabe et al., 1995), but appears to be highly developed in this protein, and may have allowed it to handle the challenge presented by the rapidly evolving telomere sequence in Saccharomycotina yeast (see below).

Another (probably minor) source of flexibility may be the number of nucleotides that separate the two half sites. In the vast majority of well-characterized target sites, this number is three such that the center-to-center distance between the two half sites is 8 bp (Pina et al., 2003). However, in a footprinting analysis utilizing a variant telomere sequence derived from *S. castellii*, *Sc*Rap1 produced a split footprint indicative of a center-to-center distance of 14 nt, suggesting that an atypical separation between the half sites can be tolerated in rare cases, possibly through looping out of the intervening DNA (Wahlin and Cohn, 2000).

Because all Saccharomycotina Rap1 homologs possess duplicated Myb motifs, it seems likely they all use such motif pairs for direct DNA-binding. This proposition is consistent with studies of two Rap1 family members, namely those in *S. castellii* and *C. albicans*. Specifically, the pairs of Myb motifs in each protein alone have been shown to be just as active in DNA-binding as the respective full-length protein (Wahlin and Cohn, 2002; Yu et al., 2010; Rhodin Edso et al., 2011). While not as well characterized as *Sc*Rap1, the DNA-binding mechanisms of *Scas*Rap1 and *Ca*Rap1 also appear to be quite similar to that for *Sc*Rap1 with respect to

Myb 1

Myb 2

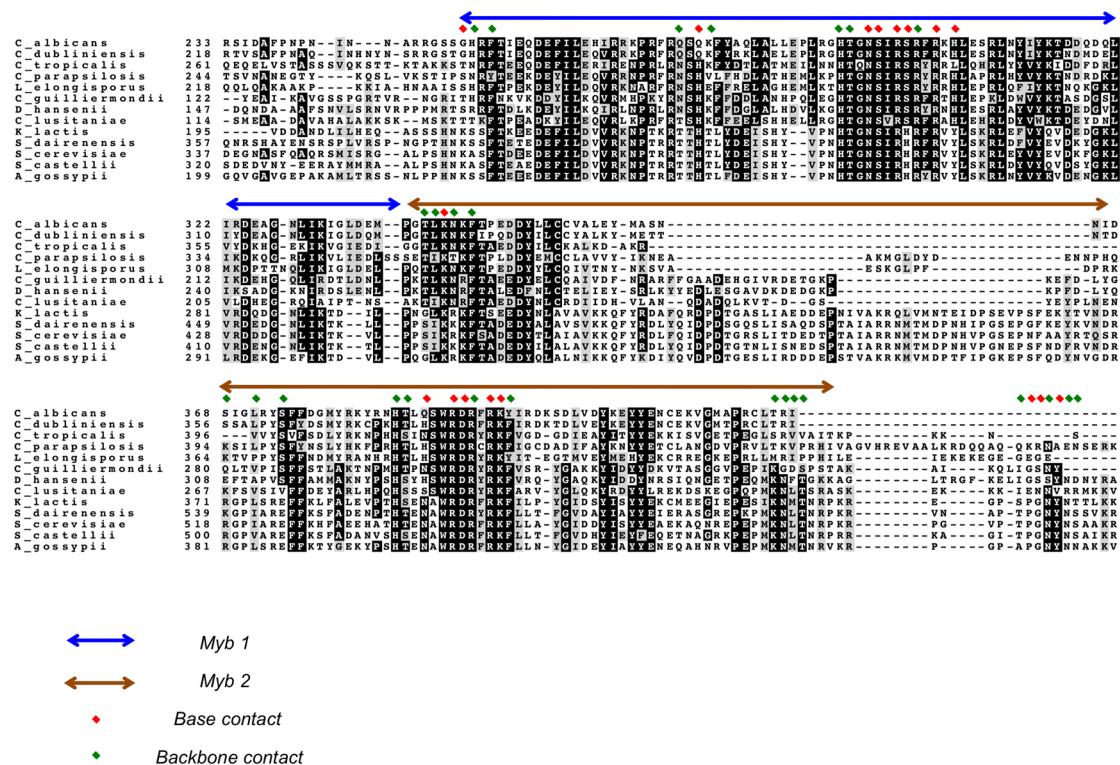◆ Base contact

◆ Backbone contact

FIGURE 3 | An alignment of Rap1$_{DBDs}$ showing the variable extent of sequence conservation of residues implicated in DNA contacts. An alignment of the DBDs of selected Saccharomycotina Rap1 homologs is displayed. The first and second Myb motifs are indicated by blue and brown arrows, respectively. Residues implicated in making base and backbone contacts are designated by red and green diamonds, respectively. Dark and gray shading of amino acids are used to highlight strict conservation and conservative substitution, respectively.

target site arrangement and sequence. For *Scas*Rap1, the minimal high affinity target is a 12-bp duplex (GGGTGTCTGGGT), within which just three positions (G1, C7, T12) appear to have non-stringent sequence requirement (Rhodin Edso et al., 2011). For *Ca*Rap1, the high affinity target consists of two 5-bp elements (GGTGT and GGATG) separated by two base pairs of random nucleotides (Yu et al., 2010). These observations are quite consistent with the notion of consecutive Myb motifs each recognizing 4–5 bp of G-rich elements. The exact identity of the first half site (GGTGT), which is the target of the second Myb motif according to the *Sc*Rap1$_{DBD}$-DNA crystal structure, suggests that the mechanisms of this second Myb motif in telomere DNA-binding may be quite well conserved in evolution. On the other hand, the half-site separations for *Scas*Rap1 and *Ca*Rap1 appear to be smaller than, and the consensus sequences for their second half sites quite different from that of *Sc*Rap1, consistent with significant adaptation of these Rap1 orthologs to their cognate telomere sequences. Notably, residues in the first Myb motif of *Sc*Rap1 implicated in direct base contact appear to exhibit greater sequence variation among all the Saccharomycotina homologs than comparable residues in the second Myb motif (**Figure 3**). This difference could reflect adaption of the first Myb to the more divergent target sites (i.e., the second half site). A notable difference between *Candida* and *Saccharomyces* Rap1s is that the former has a far less significant role in transcriptional regulation and does not appear to bind to the promoters of many metabolism-related genes

(Lavoie et al., 2010; Yu et al., 2010). Hence, it is unclear if *Ca*Rap1 possesses the same degree of target site recognition versatility as that possessed by *Sc*Rap1. Nevertheless, the versatility exhibited by *Sc*Rap1 indicates that members of this protein family has a variety of means to bind alternative sequences, and hence is well positioned to handle the challenge posed by the rapidly evolving telomere sequence in Saccharomycotina yeast.

Cdc13

Cdc13 (cell division cycle 13), the major G-tail binding protein in Saccharomycotina yeast, is like Rap1, a multifunctional protein with a complex domain organization (for reviews, see Giraud-Panis et al., 2010; Lue, 2010). As the name implies, it was initially characterized as a gene in *S. cerevisiae* that when mutated, causes cell cycle defects (Garvik et al., 1995). Subsequent studies uncovered not only the G-tail binding activity of *Sc*Cdc13, but also multiple functions for this protein at telomeres, including protecting telomeres against C-strand degradation, as well as regulation of both telomerase and Pol α in their telomere DNA synthesis activities (Nugent et al., 1996; Qi and Zakian, 2000; Pennock et al., 2001). For a subset of these functions, *Sc*Cdc13 works as part of a complex (CST) that also contains Stn1 and Ten1 (Giraud-Panis et al., 2010).

Structurally, *Sc*Cdc13 is quite large (924 aa) and complex, and is comprised of four OB fold domains that bind distinct
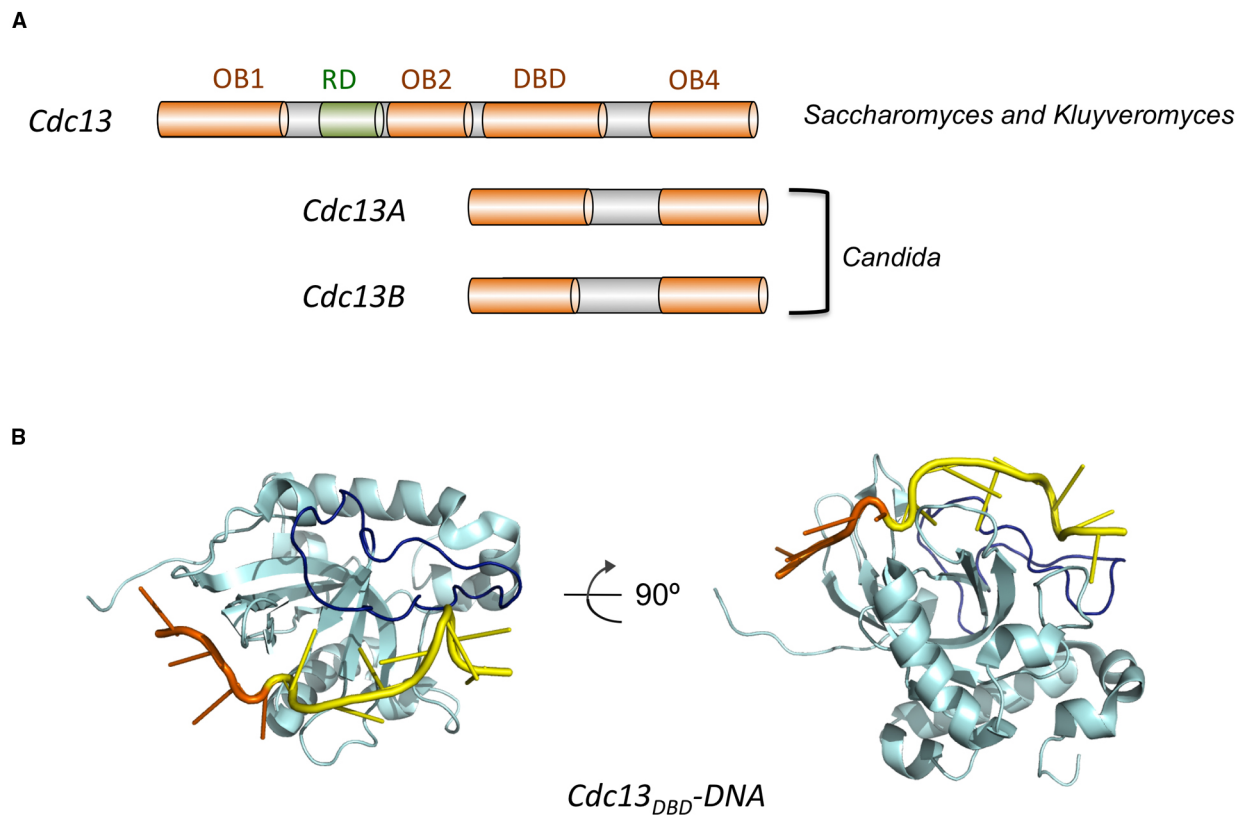
FIGURE 4 | The domain organization of Cdc13 and the structure of Cdc13$_{DBD}$-DNA complex. (A) The domain structures of Cdc13 homologs from various Saccharomycotina species are illustrated. The four OB folds (OB1, OB2, DBD, and OB4) are displayed in orange, and the RD (recruitment domain) is shown in green. **(B)** The structure of the DBD of *S. cerevisiae* Cdc13 bound to its target DNA (PDB ID: 1S40). Cdc13 is shown in cyan with the L23 loop highlighted in dark blue. The 5′ four nucleotides of the DNA target is shown in orange and the 3′ seven nucleotides shown in yellow.

molecular targets to mediate telomere protection and maintenance (**Figure 4A**). *Sc*Cdc13$_{OB1}$ forms dimers to create a binding groove for Pol1 (the catalytic subunit of DNA polymerase α), and may possess a low affinity G-strand-binding activity as well as binding sites for other proteins (Hsu et al., 2004; Mitchell et al., 2010; Sun et al., 2011). *Sc*Cdc13$_{OB2}$ also forms dimers and modulates interaction between Cdc13 and Stn1 (Mason et al., 2012). The third OB fold (*Sc*Cdc13$_{DBD}$) constitutes the high affinity G-strand-binding domain, and the final OB fold (*Sc*Cdc13$_{OB4}$) mediates interaction with Stn1 (Hughes et al., 2000; Sun et al., 2011; Yu et al., 2012). In addition to these OB fold domains, Cdc13 also carries a telomerase recruitment domain (RD) that binds to the telomerase regulatory subunit Est1 and that is required for telomere localization of telomerase (Pennock et al., 2001; Wu and Zakian, 2011).

Interestingly, analysis of other Cdc13s in Saccharomycotina yeast revealed a high degree of structural malleability and evolutionary plasticity. While all *Saccharomyces* and *Kluyveromyces spp.* carry just one Cdc13 homolog that resembles structurally *Sc*Cdc13, most *Candida spp.* carry two Cdc13 homologs (named Cdc13A and Cdc13B), each containing just two OB fold domains that align well to *Sc*Cdc13$_{DBD}$ and *Sc*Cdc13$_{OB4}$ (**Figure 4A**). The accumulated structural and functional evidence suggests that *Sc*Cdc13 (and other large Cdc13s) may arise through a fusion of

Cdc13A and Cdc13B in the common ancestor of Saccharomycotina yeast (Lue and Chan, 2013).

The G-tail binding activity of Cdc13 was first characterized (not surprisingly) for the *S. cerevisiae* protein. Proteolytic and deletion analyses defined a stable domain (*Sc*Cdc13$_{DBD}$, amino acid 557 to 692) that exhibits high affinity (sub nanomolar) for a variety of target sites that correspond to different variants of the irregular *Sc* G-strand repeats (G$_{1-3}$T; Hughes et al., 2000; Anderson et al., 2003). Even though full length *Sc*Cdc13 is naturally dimeric, the DBD domain behaves as a monomer in solution and binds DNA as a monomer. The minimal size for high affinity ligands is reported to be ~11 nt, and the affinities of *Sc*Cdc13$_{DBD}$ for these ligands are typically similar to or better than those of full length *Sc*Cdc13. Nuclear magnetic resonance (NMR) investigations of *Sc*Cdc13$_{DBD}$ revealed an OB fold structure, which is quite common for ss nucleic acid-binding proteins (Mitton-Fry et al., 2002, 2004; **Figure 4B**). A structural motif shared by numerous proteins, the OB fold is comprised of five beta strands (S1 through S5) that adopt the shape of a miniaturized barrel (Theobald et al., 2003; Bochkarev and Bochkareva, 2004). For most ssDNA-binding OB folds, residues in L12 (the loop connecting S1 and S2), L45 and the central beta strand (S3) are typically responsible for contacting a short (4–6 nt) ligand. A standard polarity prevails in the vast majority of OB-ssDNA complexes

such that L45 and L12 interact with the 5′ and 3′ portion of the target site, respectively. A distinctive feature of ScCdc13$_{DBD}$ is the presence of an extended and structurally well-defined L23 that makes contacts to nucleotides 3′ to the typical target site, thus expanding the ssDNA ligand to 11 nt (Mitton-Fry et al., 2004; Eldridge and Wuttke, 2008; **Figure 4B**). A combination of structural, biophysical and biochemical investigations have provided rich insights on the recognition mechanism of ScCdc13$_{DBD}$ for an 11-nt high affinity ligand (GTGTGGGTGTG; $K_d = 3$ pM; Anderson et al., 2003; Mitton-Fry et al., 2004; Eldridge et al., 2006; Eldridge and Wuttke, 2008). Like many other ssDNA and RNA-binding proteins, the hydrophobic and aromatic residues in ScCdc13$_{DBD}$ evidently make greater contribution to affinity than charged residues (Anderson et al., 2003). While amino acids that contribute significantly to binding can be identified throughout the DNA-protein interface, the most critical ones all interact primarily with the 5′-most four nucleotides (GTGT; Anderson et al., 2003; Mitton-Fry et al., 2004). The region surrounding the 5′ nucleotides appear to undergo conformational re-structuring upon DNA-binding, arguing for an induced fit mechanism that may enhance the specificity of interaction (Eldridge and Wuttke, 2008). In contrast, the 3′ nucleotides are bound chiefly by the extended L23 with less sequence specificity, which may allow ScCdc13$_{DBD}$ to interact optimally with the heterogeneous *S. cerevisiae* telomere repeats (Eldridge and Wuttke, 2008).

In addition to ScCdc13, several other family members in the *Saccharomyces* and *Candida* lineages have been investigated with respect to their DNA-binding properties, revealing interesting mechanistic variations in the recognition of G-tails. $Scas$Cdc13 is comparable in size to ScCdc13, but possesses a functional DBD domain that is more extended on the N-terminal side by ∼70 aa (Rhodin Edso et al., 2008). The structural basis for this additional requirement is not understood. Although the affinity of $Scas$Cdc13 for the cognate G-tail has not been determined quantitatively, the DBD domain appears to possess an affinity similar to that of the full length protein (Rhodin Edso et al., 2008). The 8 bp minimal target site (GTGTCTGG) is somewhat smaller than the 11 nt target site for ScCdc13, and the most critical nucleotide residues (positions 3, 4, 7, and 8) do not cluster near the 5′ end, suggesting substantial differences in the mechanism of binding (even though the GT-rich nature of the target site is conserved).

As described earlier, instead of carrying a large, 4-OB Cdc13, each *Candida* spp. possesses two Cdc13 homologs (Cdc13A and Cdc13B), both of which contain just 2 OB folds. Despite their small size, the *Candida* Cdc13s are clearly orthologs of the large 4-OB Cdc13s. Like the 4-OB Cdc13s, the *Candida* homologs are enriched at telomeres, and are required for normal telomere structure and function (Lue and Chan, 2013). Sequence alignments suggest that the small Cdc13s are structurally similar to the C-terminal half of the large Cdc13s, i.e., they consist of just the DBD and OB4 domains. In addition to the size difference, the small Cdc13s also exhibit distinct dimerization properties; whereas the large Cdc13s utilize OB1 for stable dimerization, the small Cdc13s appear to use primarily OB4 for this purpose (Yu et al., 2012; Lue and Chan, 2013). Moreover, in the two species for which both Cdc13A and Cdc13B dimerization have been subjected to

detailed analysis, the two paralogs appear to form preferentially heterodimers rather than homodimers (Lue and Chan, 2013; Steinberg-Neifach et al., 2015). Perhaps most interestingly, unlike ScCdc13, which uses a DBD monomer to mediate high affinity binding to G-tails, the *Candida* Cdc13s evidently require protein dimerization to achieve high affinity binding (Yu et al., 2012; Lue and Chan, 2013; Steinberg-Neifach et al., 2015).

The first *Candida* Cdc13 complex to be subjected to detailed DNA-binding analysis is the *C. tropicalis* Cdc13AA homodimer (Yu et al., 2012). (This analysis was performed prior to the discovery of CtCdc13B, and the activities of the CtCdc13AB and BB dimer, if any, remain uncharacterized.) Investigation of CtCdc13AA revealed two unexpected features. First, unlike both ScCdc13 and $Scas$Cdc13, the DBD domain of CtCdc13A alone is incapable of high affinity binding to the cognate G-tail. Instead, the formation of a stable DNA-protein complex requires dimerization of full length CtCdc13A mediated by the OB4 domain (Yu et al., 2012). Second, in keeping with the dimerization requirement, the high affinity DNA ligand consists of two copies of a 6-nt element (GGATGT) found within the *C. tropicalis* G-strand repeat unit. In the native *Ct* G-tail, the 6-nt elements are separated from one another by 17-nt, resulting in a minimal high affinity ligand (29-nt) that is far longer than those for ScCdc13 and $Scas$Cdc13. Additional characterization revealed substantial spatial flexibility between the two 6-nt elements in the high affinity complex: the distance can be as short as 10 nt (Yu et al., 2012). Thus, the individual DBDs of CtCdc13A evidently possess low affinity for a short ligand within the telomere repeat unit, requiring a pair of protein-ligand interactions conferred by the full length protein dimer to achieve high affinity Binding to G-tails.

As noted before, emerging data suggest that the *Candida* Cdc13s may exist preferentially as heterodimers, thus begging the question as to the recognition mechanism of this dimeric complex. This was first assessed in *Candida albicans* (Lue and Chan, 2013). Analysis of the *C. albicans* homodimers and heterodimers revealed substantial G-tail binding activities for both the AA and AB complex, but not the BB complex (Lue and Chan, 2013). However, the ligand requirements for CaCdc13AA and AB were not examine in detail due to the propensity of these complexes to form large aggregates. The second Cdc13 heterodimer to be analyzed was from *C. parapsilosis* (Steinberg-Neifach et al., 2015). Similar to *C. albicans*, the Cdc13 paralogs in *C. parapsilosis* can form homo-oligomeric complexes as well as heterodimers. Surprisingly, only the CpCdc13AB heterodimer exhibits robust G-tail binding activity. In contrast to CtCdc13AA, the formation of high affinity CpCdc13AB-DNA complex requires just one copy of the 6-nt consensus element. Additional studies revealed a minimal target site of ∼17 nt comprised of the 6-nt element and 11 nt on the immediate 5′ side of the element. Detailed investigation of the sequence specificity coupled with site-specific crosslinking assays uncovered an unprecedented "combinatorial" mechanism of G-tail recognition. In this mode of recognition, the DBDs of CpCdc13A and CpCdc13B make contacts to the 3′ and 5′ region of the repeat unit, respectively. Recognitions of both regions of the repeat are highly sequence-specific, thus enabling CpCdc13AB to bind its cognate target with much greater species-specificity

than the *Ct*Cdc13AA complex. In addition, the OB4 domains of *Cp*Cdc13A and *Cp*Cdc13B contribute to high affinity binding by forming a stable heterodimer to promote the dimerization of the DBDs. These results indicate that in some *Candida spp.*, the challenge of binding variant G-tails is met through the duplication of Cdc13, the hetero-dimerization of the paralogs, and the adaption of the DBDs to new target sequences. Studies of other additional *Candida* Cdc13s should provide insights on the general applicability of this proposal.

## Shared and Distinctive Features of ds and ss Telomere DNA Recognition in Saccharomycotina Yeast

As noted before, a unique attribute of the telomere system from the evolutionary standpoint is the need to maintain adequate recognition of the telomere DNA in both its double-stranded and single-stranded forms upon changes in the sequence of the DNA. The remarkable divergence of telomere repeat sequences in Saccharomycotina yeast indicates that the Rap1 and Cdc13 protein families are sufficiently versatile and malleable to meet the challenge. While the mechanisms used by each family for DNA recognition are clearly distinct, some general themes can nevertheless be discerned. Below I list common and distinctive strategies utilized by these protein families to enable recognition of diverse sequence targets by family members.

First, the utilization of a pair of DBDs, either as parts of the same polypeptide or a dimeric complex, is probably advantageous (**Figure 5**). A domain with a short DNA target site may be capable of forming only a low affinity complex; incorporating two low affinity interactions in a single complex can substantially increase

FIGURE 5 | The telomere repeat units of Saccharomycotina yeast and the putative or experimentally determined Rap1 and Cdc13 target sites. The phylogeny of *Candida spp.* and that of *Saccharomyces* and *Kluyveromyces spp.* are displayed separately along with the telomere repeat unit in each species. The putative Rap1 half-sites are displayed in green and the nucleotides that have been experimentally shown to contact Cdc13 or be required for Cdc13 binding are underlined in dark red. Note that *C. lusitaniae* has an unusual telomere repeat unit that carry just one obvious candidate half site for Rap1.

the overall affinity. The two-domain arrangement can also offer added flexibility to the system: variations in the spacing between the "half sites" are readily accommodated by two DBDs that can be flexibly positioned to each other. As illustrations, one can point to Rap1s in Saccharomycotina yeast, which have two Myb motifs and bind DNA with high affinity. In contrast, human and *S. pombe* Rap1s have just a single Myb motif and exhibit little or no DNA-binding activity. In addition, the apparent variations in the spacing between the Rap1 half sites in different organisms [e.g., 8 bp in *S. cerevisiae* and 7 bp in *C. albicans* (center-to-center distance)] are consistent with adaptions involving altered dispositions between the two Myb motifs (**Figure 5**). With regard to the Cdc13 family members, the utilization of two sets of protein–DNA contacts for high affinity binding is not universal. While *Candida* Cdc13 dimers probably all require two sets of DBD–DNA interactions to bind stably to G-tails, *S. cerevisiae* Cdc13 (despite forming dimers) binds G-tail with exceptionally high affinity using just one DBD–DNA interaction. This impressive feat of *Sc*Cdc13 is accomplished by expanding the typical OB-DNA interface through the acquisition of an extended and structurally well-defined L23. That is, rather than adding a second set of protein–DNA interaction, *Sc*Cdc13 was able to drastically expand the first set to enhance binding affinity. *S. castellii* Cdc13, another 4-OB fold Cdc13, also appears to need just one DBD–DNA interaction for high affinity binding. Whether this property applies to other large Cdc13s (e.g., *K. lactis* Cdc13) is an interesting question for future investigation.

A special case of achieving high affinity binding through two sets of protein–DNA interactions, employed by members of Cdc13 family only (specifically *Cp*Cdc13A and *Cp*Cdc13B), involves gene duplication and hetero-dimerization. Compared to homodimerization, this strategy has the advantage of allowing Cdc13 dimers to recognize a more complex target sequence made up of two distinct half sites. This advantage makes hetero-dimerization an especially adaptive strategy for the recognition of *Candida* telomere repeat units, which are long and complex.

The second common mechanistic feature that may enable ready adaption of Rap1 and Cdc13 to new telomere sequences is the ability of the DNA-binding surfaces of these proteins to undergo local conformational changes to accommodate different target sequence. This was implied by the huge number of Rap1 target sites in the *S. cerevisiae* genome and the very loose consensus sequence obtained for this protein. High resolution structural analyses of Rap1 bound to three target sequences provided amply illustration of this local flexibility at the molecular level (Taylor et al., 2000). In the case of Cdc13, there is no direct evidence yet for this local conformational flexibility. However, analysis of another ss telomere binding protein (TEBP from *Oxytricha nova*) revealed considerable tolerance of its binding surface to different sequences (Theobald and Schultz, 2003). Moreover, the intrinsically greater flexibility of ssDNA may further contribute to the ability of Cdc13 to accommodate sequence changes. An illustration of this, uncovered by investigation of *on*TEBP, is termed nucleotide shuffling, which involves the extrusion of a nucleotide away from the protein surface, and thus an alteration in the registry of the DNA (Theobald and Schultz, 2003). This phenomenon
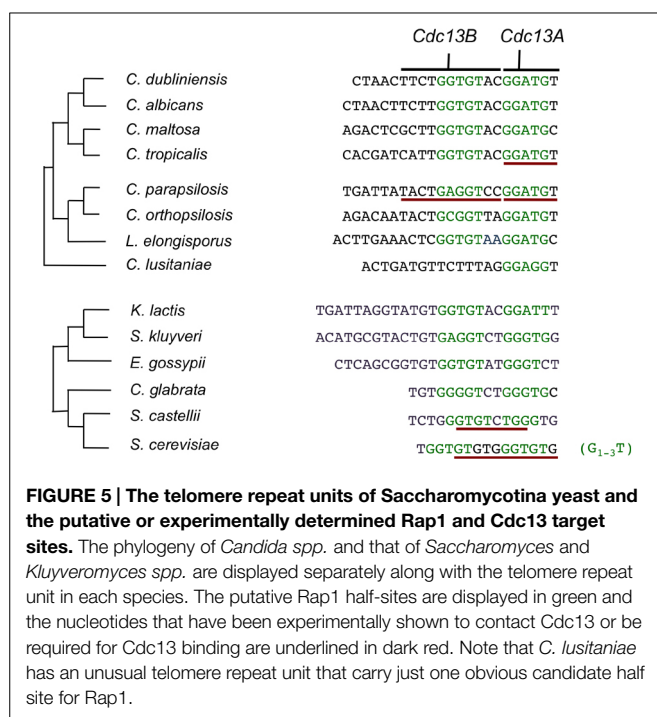
can conceivably allow insertional mutations in telomere DNA to be easily accommodated by Cdc13. Thus, both sequence-specific ss and dsDNA-binding proteins can exhibit limited versatility in binding multiple target sequences. Nevertheless, as noted earlier, promiscuous binding of telomere proteins to non-telomeric sites would probably be highly detrimental to cell physiology. Thus, limited versatility of Rap1 and Cdc13 in sequence-specific recognition probably is reflective of a finely calibrated evolutionary compromise.

Acknowledgment

References

Anderson, E. M., Halsey, W. A., and Wuttke, D. S. (2003). Site-directed mutagenesis reveals the thermodynamic requirements for single-stranded DNA recognition by the telomere-binding protein Cdc13. *Biochemistry* 42, 3751–3758. doi: 10.1021/bi027047c

Arat, N. O., and Griffith, J. D. (2012). Human Rap1 interacts directly with telomeric DNA and regulates TRF2 localization at the telomere. *J. Biol. Chem.* 287, 41583–41594. doi: 10.1074/jbc.M112.415984

Autexier, C., and Lue, N. F. (2006). The structure and function of telomerase reverse transcriptase. *Annu. Rev. Biochem.* 75, 493–517. doi: 10.1146/annurev.biochem.75.103004.142412

Blackburn, E. H., and Collins, K. (2011). Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harb. Perspect. Biol.* 3, a003558. doi: 10.1101/cshperspect.a003558

Bochkarev, A., and Bochkareva, E. (2004). From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold. *Curr. Opin. Struct. Biol.* 14, 36–42. doi: 10.1016/j.sbi.2004.01.001

Buchman, A., Lue, N., and Kornberg, R. (1988). Connections between transcriptional activators, silencers, and telomeres as revealed by functional analysis of a yeast DNA-binding protein. *Mol. Cell. Biol.* 8, 5086–5099.

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

de Lange, T. (2009). How telomeres solve the end-protection problem. *Science* 326, 948–952. doi: 10.1126/science.1170633

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., et al. (2004). Genome evolution in yeasts. *Nature* 430, 35–44. doi: 10.1038/nature02579

Eldridge, A. M., Halsey, W. A., and Wuttke, D. S. (2006). Identification of the determinants for the specific recognition of single-strand telomeric DNA by Cdc13. *Biochemistry* 45, 871–879. doi: 10.1021/bi0512703

Eldridge, A. M., and Wuttke, D. S. (2008). Probing the mechanism of recognition of ssDNA by the Cdc13-DBD. *Nucleic Acids Res.* 36, 1624–1633. doi: 10.1093/nar/gkn017

Fairall, L., Chapman, L., Moss, H., de Lange, T., and Rhodes, D. (2001). Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2. *Mol. Cell* 8, 351–361. doi: 10.1016/S1097-2765(01)00321-5

Feeser, E. A., and Wolberger, C. (2008). Structural and functional studies of the Rap1 C-terminus reveal novel separation-of-function mutants. *J. Mol. Biol.* 380, 520–531. doi: 10.1016/j.jmb.2008.04.078

Garvik, B., Carson, M., and Hartwell, L. (1995). Single-stranded DNA arising at telomeres in cdc13 mutants may constitute a specific signal for the RAD9 checkpoint. *Mol. Cell. Biol.* 15, 6128–6138.

Giraldo, R., and Rhodes, D. (1994). The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J.* 13, 2411–2420.

Giraud-Panis, M. J., Teixeira, M. T., Geli, V., and Gilson, E. (2010). CST meets shelterin to keep telomeres in check. *Mol. Cell* 39, 665–676. doi: 10.1016/j.molcel.2010.08.024

Hedges, S. B. (2002). The origin and evolution of model organisms. *Nat. Rev. Genet.* 3, 838–849. doi: 10.1038/nrg929

Hsu, C. L., Chen, Y. S., Tsai, S. Y., Tu, P. J., Wang, M. J., and Lin, J. J. (2004). Interaction of *Saccharomyces* Cdc13p with Pol1p, Imp4p, Sir4p and Zds2p is involved in telomere replication, telomere maintenance and cell growth control. *Nucleic Acids Res.* 32, 511–521. doi: 10.1093/nar/gkh203

Huet, J., Cottrelle, P., Cool, M., Vignais, M. L., Thiele, D., Marck, C., et al. (1985). A general upstream binding factor for genes of the yeast translational apparatus. *EMBO J.* 4, 3539–3547.

Hughes, T. R., Weilbaecher, R. G., Walterscheid, M., and Lundblad, V. (2000). Identification of the single-strand telomeric DNA binding domain of the *Saccharomyces cerevisiae* Cdc13 protein. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6457–6462. doi: 10.1073/pnas.97.12.6457

Idrissi, F. Z., Fernandez-Larrea, J. B., and Pina, B. (1998). Structural and functional heterogeneity of Rap1p complexes with telomeric and UASrpg-like DNA sequences. *J. Mol. Biol.* 284, 925–935. doi: 10.1006/jmbi.1998.2215

Idrissi, F. Z., and Pina, B. (1999). Functional divergence between the half-sites of the DNA-binding sequence for the yeast transcriptional regulator Rap1p. *Biochem. J.* 341(Pt 3), 477–482. doi: 10.1042/0264-6021:3410477

Jain, D., and Cooper, J. P. (2011). Telomeric strategies: means to an end. *Annu. Rev. Genet.* 44, 243–269. doi: 10.1146/annurev-genet-102108-134841

Kabir, S., Hockemeyer, D., and de Lange, T. (2014). TALEN gene knockouts reveal no requirement for the conserved human shelterin protein Rap1 in telomere protection and length regulation. *Cell Rep.* 9, 1273–1280. doi: 10.1016/j.celrep.2014.10.014

Kanoh, J., and Ishikawa, F. (2001). spRap1 and spRif1, recruited to telomeres by Taz1, are essential for telomere function in fission yeast. *Curr. Biol.* 11, 1624–1630. doi: 10.1016/S0960-9822(01)00503-6

Konig, P., Giraldo, R., Chapman, L., and Rhodes, D. (1996). The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. *Cell* 85, 125–136. doi: 10.1016/S0092-8674(00)81088-0

Lansdorp, P. M. (2005). Major cutbacks at chromosome ends. *Trends Biochem. Sci.* 30, 388–395. doi: 10.1016/j.tibs.2005.05.004

Lavoie, H., Hogues, H., Mallick, J., Sellam, A., Nantel, A., and Whiteway, M. (2010). Evolutionary tinkering with conserved components of a transcriptional regulatory network. *PLoS Biol.* 8:e1000329. doi: 10.1371/journal.pbio.1000329

Lei, M., Podell, E. R., Baumann, P., and Cech, T. R. (2003). DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA. *Nature* 426, 198–203. doi: 10.1038/nature02092

Lei, M., Podell, E. R., and Cech, T. R. (2004). Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nat. Struct. Mol. Biol.* 11, 1223–1229. doi: 10.1038/nsmb867

Li, B., Oestreich, S., and de Lange, T. (2000). Identification of human Rap1: implications for telomere evolution. *Cell* 101, 471–483. doi: 10.1016/S0092-8674(00)80858-2

Lieb, J. D., Liu, X., Botstein, D., and Brown, P. O. (2001). Promoter-specific binding of Rap1 revealed by genome-wide maps of protein–DNA association. *Nat. Genet.* 28, 327–334. doi: 10.1038/ng569

Lopez, M. C., Smerage, J. B., and Baker, H. V. (1998). Multiple domains of repressor activator protein 1 contribute to facilitated binding of glycolysis regulatory protein 1. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14112–14117. doi: 10.1073/pnas.95.24.14112

Lue, N. F. (2010). Plasticity of telomere maintenance mechanisms in yeast. *Trends Biochem. Sci.* 35, 8–17. doi: 10.1016/j.tibs.2009.08.006

Lue, N. F., and Chan, J. (2013). Duplication and functional specialization of the telomere-capping protein Cdc13 in *Candida* species. *J. Biol. Chem.* 288, 29115–29123. doi: 10.1074/jbc.M113.506519

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

Lustig, A. (2003). Clues to catastrophic telomere loss in mammals from yeast telomere rapid deletion. *Nat. Rev. Genet.* 4, 916–923. doi: 10.1038/nrg1207

Martinez, P., Thanasoula, M., Carlos, A. R., Gomez-Lopez, G., Tejera, A. M., Schoeftner, S., et al. (2010). Mammalian Rap1 controls telomere function and gene expression through binding to telomeric and extratelomeric sites. *Nat. Cell Biol.* 12, 768–780. doi: 10.1038/ncb2081

Mason, M., Wanat, J. J., Harper, S., Schultz, D. C., Speicher, D. W., Johnson, F. B., et al. (2012). Cdc13 OB2 dimerization required for productive Stn1 binding and efficient telomere maintenance. *Structure* 21, 109–120. doi: 10.1016/j.str.2012.10.012

Mitchell, M. T., Smith, J. S., Mason, M., Harper, S., Speicher, D. W., Johnson, F. B., et al. (2010). Cdc13 N-terminal dimerization, DNA binding, and telomere length regulation. *Mol. Cell. Biol.* 30, 5325–5334. doi: 10.1128/MCB.00515-10

Mitton-Fry, R., Anderson, E., Hughes, T., Lundblad, V., and Wuttke, D. (2002). Conserved structure for single-stranded telomeric DNA recognition. *Science* 296, 145–147. doi: 10.1126/science.1068799

Mitton-Fry, R. M., Anderson, E. M., Theobald, D. L., Glustrom, L. W., and Wuttke, D. S. (2004). Structural basis for telomeric single-stranded DNA recognition by yeast Cdc13. *J. Mol. Biol.* 338, 241–255. doi: 10.1016/j.jmb.2004.01.063

Miyake, Y., Nakamura, M., Nabetani, A., Shimamura, S., Tamura, M., Yonehara, S., et al. (2009). RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway. *Mol. Cell* 36, 193–206. doi: 10.1016/j.molcel.2009.08.009

Nandakumar, J., and Cech, T. R. (2013). Finding the end: recruitment of telomerase to telomeres. *Nat. Rev. Mol. Cell Biol.* 14, 69–82. doi: 10.1038/nrm3505

Nugent, C. I., Hughes, T. R., Lue, N. F., and Lundblad, V. (1996). Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance. *Science* 274, 249–252. doi: 10.1126/science.274.5285.249

Olovnikov, A. M. (1996). Telomeres, telomerase, and aging: origin of the theory. *Exp. Gerontol.* 31, 443–448. doi: 10.1016/0531-5565(96)00005-8

O'Sullivan, R. J., and Karlseder, J. (2010). Telomeres: protecting chromosomes against genome instability. *Nat. Rev. Mol. Cell Biol.* 11, 171–181. doi: 10.1038/nrm2848

Pennock, E., Buckley, K., and Lundblad, V. (2001). Cdc13 delivers separate complexes to the telomere for end protection and replication. *Cell* 104, 387–396. doi: 10.1016/S0092-8674(01)00226-4

Pesole, G., Lotti, M., Alberghina, L., and Saccone, C. (1995). Evolutionary origin of nonuniversal CUGSer codon in some *Candida* species as inferred from a molecular phylogeny. *Genetics* 141, 903–907.

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

Pina, B., Fernandez-Larrea, J., Garcia-Reyero, N., and Idrissi, F. Z. (2003). The different (sur)faces of Rap1p. *Mol. Genet. Genomics* 268, 791–798. doi: 10.1007/s00438-002-0801-3

Podlevsky, J. D., Bley, C. J., Omana, R. V., Qi, X., and Chen, J. J. (2008). The telomerase database. *Nucleic Acids Res.* 36, D339–D343. doi: 10.1093/nar/gkm700

Price, C. M., Boltz, K. A., Chaiken, M. F., Stewart, J. A., Beilstein, M. A., and Shippen, D. E. (2010). Evolution of CST function in telomere maintenance. *Cell Cycle* 9, 3157–3165. doi: 10.4161/cc.9.16.12547

Qi, H., and Zakian, V. A. (2000). The *Saccharomyces* telomere-binding protein Cdc13p interacts with both the catalytic subunit of DNA polymerase α and the telomerase-associated Est1 protein. *Genes Dev.* 14, 1777–1788.

Rhee, H. S., and Pugh, B. F. (2011). Comprehensive genome-wide protein–DNA interactions detected at single-nucleotide resolution. *Cell* 147, 1408–1419. doi: 10.1016/j.cell.2011.11.013

Rhodin Edso, J., Gustafsson, C., and Cohn, M. (2011). Single- and double-stranded DNA binding proteins act in concert to conserve a telomeric DNA core sequence. *Genome Integr.* 2, 2. doi: 10.1186/2041-9414-2-2

Rhodin Edso, J., Tati, R., and Cohn, M. (2008). Highly sequence-specific binding is retained within the DNA-binding domain of the *Saccharomyces castellii* Cdc13 telomere-binding protein. *FEMS Yeast Res.* 8, 1289–1302. doi: 10.1111/j.1567-1364.2008.00431.x

Schwabe, J. W., Chapman, L., and Rhodes, D. (1995). The oestrogen receptor recognizes an imperfectly palindromic response element through an alternative side-chain conformation. *Structure* 3, 201–213. doi: 10.1016/S0969-2126(01)00150-2

Sfeir, A., Kabir, S., Van Overbeek, M., Celli, G. B., and de Lange, T. (2010). Loss of Rap1 induces telomere recombination in the absence of NHEJ or a DNA damage signal. *Science* 327, 1657–1661. doi: 10.1126/science.1185100

Shore, D. (1994). RAP1: a protean regulator in yeast. *Trends Genet.* 10, 408–412. doi: 10.1016/0168-9525(94)90058-2

Shore, D., Stillman, D. J., Brand, A. H., and Nasmyth, K. A. (1987). Identification of silencer binding proteins from yeast: possible roles in SIR control and DNA replication. *EMBO J.* 6, 461–467.

Steinberg-Neifach, O., Wellington, K., Vazquez, L., and Lue, N. F. (2015). Combinatorial recognition of a complex telomere repeat sequence by the *Candida parapsilosis* Cdc13AB heterodimer. *Nucleic Acids Res.* 43, 2164-2176. doi: 10.1093/nar/gkv092

Stewart, J. A., Wang, F., Chaiken, M. F., Kasbek, C., Chastain, P. D. II, Wright, W. E., et al. (2012). Human CST promotes telomere duplex replication and general replication restart after fork stalling. *EMBO J.* 31, 3537–3549. doi: 10.1038/emboj.2012.215

Sun, J., Yang, Y., Wan, K., Mao, N., Yu, T. Y., Lin, Y. C., et al. (2011). Structural bases of dimerization of yeast telomere protein Cdc13 and its interaction with the catalytic subunit of DNA polymerase alpha. *Cell Res.* 21, 258–274. doi: 10.1038/cr.2010.138

Surovtseva, Y. V., Churikov, D., Boltz, K. A., Song, X., Lamb, J. C., Warrington, R., et al. (2009). Conserved telomere maintenance component 1 interacts with STN1 and maintains chromosome ends in higher eukaryotes. *Mol. Cell* 36, 207–218. doi: 10.1016/j.molcel.2009.09.017

Taylor, H. O., O'Reilly, M., Leslie, A. G., and Rhodes, D. (2000). How the multifunctional yeast Rap1p discriminates between DNA target sites: a crystallographic analysis. *J. Mol. Biol.* 303, 693–707. doi: 10.1006/jmbi.2000.4161

Theobald, D. L., Mitton-Fry, R. M., and Wuttke, D. S. (2003). Nucleic acid recognition by OB-fold proteins. *Annu. Rev. Biophys. Biomol. Struct.* 32, 115–133. doi: 10.1146/annurev.biophys.32.110601.142506

Theobald, D. L., and Schultz, S. C. (2003). Nucleotide shuffling and ssDNA recognition in *Oxytricha nova* telomere end-binding protein complexes. *EMBO J.* 22, 4314–4324. doi: 10.1093/emboj/cdg415

Vignais, M. L., Huet, J., Buhler, J. M., and Sentenac, A. (1990). Contacts between the factor TUF and RPG sequences. *J. Biol. Chem.* 265, 14669–14674.

Wahlin, J., and Cohn, M. (2000). *Saccharomyces cerevisiae* RAP1 binds to telomeric sequences with spatial flexibility. *Nucleic Acids Res.* 28, 2292–2301. doi: 10.1093/nar/28.12.2292

Wahlin, J., and Cohn, M. (2002). Analysis of the RAP1 protein binding to homogeneous telomeric repeats in *Saccharomyces castellii*. *Yeast* 19, 241–256. doi: 10.1002/yea.816

Wu, Y., and Zakian, V. A. (2011). The telomeric Cdc13 protein interacts directly with the telomerase subunit Est1 to bring it to telomeric DNA ends *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20362–20369. doi: 10.1073/pnas.1100281108

Yarragudi, A., Parfrey, L. W., and Morse, R. H. (2007). Genome-wide analysis of transcriptional dependence and probable target sites for Abf1 and Rap1 in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 35, 193–202. doi: 10.1093/nar/gkl1059

Yu, E. Y., Sun, J., Lei, M., and Lue, N. F. (2012). Analyses of *Candida* Cdc13 orthologues revealed a novel OB fold dimer arrangement, dimerization-assisted DNA binding, and substantial structural differences between Cdc13 and RPA70. *Mol. Cell. Biol.* 32, 186–198. doi: 10.1128/MCB.05875-11

Yu, E. Y., Yen, W. F., Steinberg-Neifach, O., and Lue, N. F. (2010). Rap1 in *Candida albicans*: an unusual structural organization and a critical function in suppressing telomere recombination. *Mol. Cell. Biol.* 30, 1254–1268. doi: 10.1128/MCB.00986-09

Rif1: A Conserved Regulator of DNA Replication and Repair Hijacked by Telomeres in Yeasts

Stefano Mattarocci[†], Lukas Hafner[†], Aleksandra Lezaja[†], Maksym Shyian and David Shore*

Department of Molecular Biology and Institute for Genetics and Genomics in Geneva, University of Geneva, Geneva, Switzerland

Rap1-interacting factor 1 (Rif1) was originally identified in the budding yeast *Saccharomyces cerevisiae* as a telomere-binding protein that negatively regulates telomerase-mediated telomere elongation. Although this function is conserved in the distantly related fission yeast *Schizosaccharomyces pombe*, recent studies, both in yeasts and in metazoans, reveal that Rif1 also functions more globally, both in the temporal control of DNA replication and in DNA repair. Rif1 proteins are large and characterized by N-terminal HEAT repeats, predicted to form an elongated alpha-helical structure. In addition, all Rif1 homologs contain two short motifs, abbreviated RVxF/SILK, that are implicated in recruitment of the PP1 (yeast Glc7) phosphatase. In yeasts the RVxF/SILK domains have been shown to play a role in control of DNA replication initiation, at least in part through targeted de-phosphorylation of proteins in the pre-Replication Complex. In human cells Rif1 is recruited to DNA double-strand breaks through an interaction with 53BP1 where it counteracts DNA resection, thus promoting repair by non-homologous end-joining. This function requires the N-terminal HEAT repeat-containing domain. Interestingly, this domain is also implicated in DNA end protection at un-capped telomeres in yeast. We conclude by discussing the deployment of Rif1 at telomeres in yeasts from both an evolutionary perspective and in light of its recently discovered global functions.

Keywords: Rif1, telomere, DNA replication timing, DNA repair, DNA recombination, telomere capping, Rap1

INTRODUCTION

Telomeres, the ends of linear eukaryotic chromosomes, pose two fundamental problems for the cell. First, the polarity of DNA synthesis, and its initiation by an RNA primer that must be subsequently replaced by DNA, means that conventional replication mechanisms cannot duplicate the termini of linear molecules (the so-called "end replication problem"; Watson, 1972; Olovnikov, 1973; Lingner et al., 1995). Second, chromosome ends physically resemble accidental DNA double-stranded breaks (DSBs), but must be treated differently by the cell to avoid DNA damage checkpoint activation and the genome instability caused by chromosome end fusions or translocations (the "end protection problem").

Organisms with linear chromosomes have thus had to evolve special mechanisms, carried out by a relatively conserved set of proteins, to replicate chromosome ends and to hide them from highly sensitive DNA damage checkpoint and repair systems (de Lange, 2009). In nearly all

eukaryotes the end replication problem is solved by the specialized reverse transcriptase enzyme called telomerase, which adds short G-rich repeated sequences [TG_{1-3} and $T_2AC(A)(C)G_{2-8}$, in *Saccharomyces cerevisiae* (Sc) and *Schizosaccharomyces pombe* (Sp), respectively; T_2AG_3 in metazoans] to chromosome 3′ ends, using an intrinsic RNA template. The regulated action of telomerase prevents the continual erosion of chromosome ends with succeeding cell divisions, and allows for the maintenance of a constant average length of telomere repeats at chromosome ends. A conserved complex of six proteins referred to as shelterin protects (or "caps") chromosome ends in metazoans (de Lange, 2005) thus solving the end protection problem (**Figure 1A**). Although the targets of shelterin throughout evolution appear to be highly conserved (e.g., ATM/ATR checkpoint pathways and the telomerase enzyme), the actual shelterin components themselves are less well conserved in yeasts, particularly in budding yeasts, where only one shelterin component, Rap1, is present (**Figure 1A**).

This Perspective article will focus on the Rif1 (Rap1-interacting factor 1) protein, a telomere-binding protein originally found in the budding yeast *Saccharomyces cerevisiae* (Hardy et al., 1992) and later in the distantly related fission yeast *Schizosaccharomyces pombe* (Kanoh and Ishikawa, 2001). More recently Rif1 has come to be recognized as a highly conserved protein in metazoans (Sreesankar et al., 2012). Surprisingly, though, there is no clear evidence that Rif1 is a telomere binding protein in any multicellular organism. Instead, recent discoveries in mammalian and yeast systems have pointed to two unanticipated and conserved functions of Rif1 that have dramatically altered our view of this protein. These studies reveal that Rif1 acts genome-wide to regulate DNA repair pathway choice and the temporal pattern of DNA replication. In the following sections, the telomeric functions of Rif1 and its

more widespread functions will be described with reference to conserved structural domains and motifs in Rif1 (**Figure 1B**). Finally, we will highlight and discuss unresolved questions related to the evolution of Rif1 as a telomeric protein in yeasts.

TELOMERIC FUNCTIONS OF Rif1 IN YEASTS

ScRif1 was first shown to negatively regulate telomere elongation, based on the observation that telomere repeat tracts in *rif1∆* cells are on average about twice the length of those in wild type cells (Hardy et al., 1992). A second Rap1-interacting factor, Rif2, has a smaller effect on telomere length and works in a parallel pathway (Wotton and Shore, 1997). The way in which Rif1 and Rif2 assemble on telomeric DNA has recently been elucidated in molecular detail by x-ray crystallography (Shi et al., 2013). Remarkably, both Rif1 and Rif2 employ a short alpha-helical peptide motif, referred as the Rap1-binding module (RBM; for Rif1$_{RBM}$ see **Figure 1B**) to bind to a conserved groove in the C-terminal domain of Rap1 (Rap1$_{RCT}$). Rif1 also contacts Rap1 at a different site on the RCT, though with lower affinity, through a tetramer-forming C-terminal domain (Rif1$_{CTD}$; see **Figure 1B**). Rif2 also contains a second Rap1-interacting domain that makes contact with a third region on the Rap1 C-terminus. This network of Rap1–Rif1–Rif2 interactions thus generates a "molecular Velcro" that promotes the cooperative binding of Rif1/Rif2 to the arrays of DNA-bound Rap1 found uniquely at telomeres (Shi et al., 2013). However, Rap1 binding alone is not sufficient for telomere length regulation by Rif1, since mutations in the conserved RVxF/SILK (involved in PP1 phosphatase binding; see **Figure 1B**) and the HEAT repeat domains cause telomere
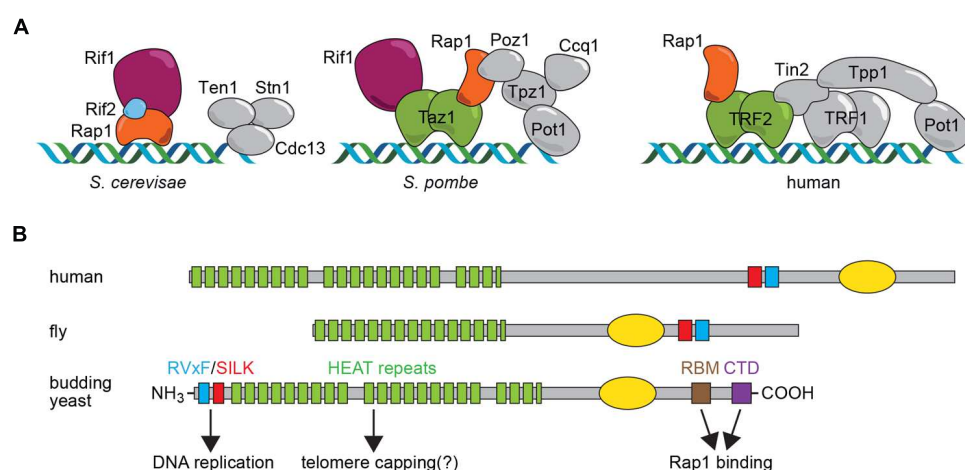
FIGURE 1 | (A) Shelterin complexes assembled on telomere-repeat sequences in budding yeast (*Saccharomyces cerevisiae*), fission yeast (*Schizosaccharomyces pombe*) and human cells. Proteins discussed here are highlighted in color. It should be noted that *Schizosaccharomyces pombe* and human also contain a CST complex involved in DNA replication at telomeres and, at least in humans, genome-wide. **(B)** Schematic representation of Rif1 motif structure in human, fly and budding yeast, with functional properties for the *Saccharomyces cerevisiae* protein indicated below. The yellow oval represents a region of homology to the alpha-CTD of bacterial polymerases that in hRif1 has been shown to have DNA-binding activity (Xu et al., 2010).

elongation (our unpublished results). Remarkably, the Rap1-interacting C-terminus of Rif1 is not required for some degree of telomere length regulation (Shi et al., 2013), suggesting that Rif1 may be able to localize to telomeres through a second mechanism, perhaps involving the large, conserved HEAT domain that occupies a significant portion of the Rif1 N-terminus (**Figure 1B**, see below). The targets of Rif1 and Rif2 in telomerase inhibition still remain to be clarified (Bianchi and Shore, 2008; Gao et al., 2010).

Although not essential for capping, recent studies show clearly that ScRif1 plays a role in protecting telomere ends. This was first revealed by its genetic interaction with Cdc13, a telomere-specific single-strand DNA-binding protein that forms part of the RPA-like Cdc13-Stn1-Ten1 (CST) complex essential for capping telomeres in the G2/M phase of the cell cycle (Anbalagan et al., 2011; Xue et al., 2011; see **Figure 1A**). When CST function is compromised, Rif1 becomes essential for telomere protection and survival. Even in cells where CST is perfectly functional, Rif1 is required for checkpoint inhibition at short telomeres (Ribeyre and Shore, 2012), where it works in parallel with Rif2 in the so-called telomeric anti-checkpoint (Michelson et al., 2005). Remarkably, these protective functions of Rif1 also do not require the C-terminal domains necessary for targeting to telomeric DNA through Rap1 interactions (Xue et al., 2011; our unpublished data). These observations point to a possible role of the N-terminal HEAT repeats in localizing Rif1 to its sites of action in chromatin.

In *Saccharomyces cerevisiae*, and indeed in many organisms where it has been examined, chromatin immediately internal to the telomere repeat tracts is transcriptionally silenced, or heterochromatic (Gottschling et al., 1990). This phenomenon, referred to as telomere position effect (TPE), is carried out by a set of SIR (Silent Information Regulator) proteins. SIR proteins are recruited to telomeres though interactions with both Rap1 and the Yku70/80 proteins, and spread along telomere-adjacent chromatin aided by the histone deacetylase activity of the highly conserved Sir2 protein (reviewed in Rusche et al., 2003). Interestingly, Rif1 counteracts the repressive function of SIR proteins at telomeres, at least in part by competing with Sir3, which also contains a RBM, for binding to the Rap1 C-terminus (Kyrion et al., 1993; Buck and Shore, 1995; Wotton and Shore, 1997; Shi et al., 2013). However, in *Candida glabrata,* the only other budding yeast where Rif1's telomeric silencing function has been examined, TPE is abolished by *rif1Δ*, despite the fact that this mutation has a similar telomere elongation phenotype to that observed in *Saccharomyces cerevisiae* (Castano et al., 2005).

The only other yeast in which Rif1 function has been directly examined, the fission yeast *Schizosaccharomyces pombe*, presents a very different picture. To begin with, SpRif1 is recruited to telomeres through an interaction with Taz1 (also a Myb domain DNA-binding protein, but more similar to human TRF1/TRF2), and not with SpRap1 (**Figure 1A**). Whereas SpRif1 also plays a role in limiting telomere elongation, though via a Rap1-independent pathway, there is no evidence that it prevents telomeres from activating DNA damage response (DDR) pathways (Kanoh and Ishikawa,

2001; Miller et al., 2005). Interestingly, SpRif1 and SpRap1 have opposite effects in *taz1Δ* cells, which are inviable at low temperatures due to chromosome entanglement. Deletion of Sp*Rif1⁺* restores normal growth in *taz1Δ* cells, suggesting that SpRif1 might block telomere recombination (Miller et al., 2005). With respect to TPE, SpRif1 appears to play a positive role at subtelomeric regions (Greenwood and Cooper, 2012).

Rif1 IS A REGULATOR OF DNA REPAIR

Building upon the early observations that human Rif1 (hRif1) localizes to damaged telomeres (Silverman et al., 2004; Xu and Blackburn, 2004) and also contributes to survival under DNA replication stress (Buonomo et al., 2009), a flurry of more recent reports have provided new molecular insights into the role of both human and mouse Rif1 in the DDR (Chapman et al., 2013; Di Virgilio et al., 2013; Escribano-Diaz et al., 2013; Zimmermann et al., 2013). Together, these studies showed that Rif1 is recruited to DNA double-strand breaks (DSBs) through an N-terminal phosphorylated domain of 53BP1, with which it cooperates to block DSB resection (**Figure 2A**). This action of Rif1 promotes break repair by non-homologous end-joining (NHEJ) in the G1 phase of the cell cycle and is opposed by the action of BRCA1 in S phase, which permits a switch to a homologous recombination (HR) mode of DNA repair. Given that HR is less error-prone than NHEJ, this conversion allows cells to profit from the availability of an intact sister chromatid during S phase.

Contrary to initial reports (Xue et al., 2011), it now appears that budding yeast Rif1 also localizes to DSBs (Martina et al., 2014; our unpublished results), strongly implying a role for Rif1 in some aspect of the DDR. Although yeast cells deleted for *RIF1* do not display any obvious increase in sensitivity to agents that damage DNA, the *rif1Δ* mutation displays "synthetic" phenotypes in combination with some mutations affecting replication or repair pathways, such as the MRX (Mre11-Rad50-Xrs2) complex, which is involved in both HR and NHEJ-mediated repair (Costanzo et al., 2010; Guenole et al., 2013; Martina et al., 2014). However, the precise role of Rif1 in the DDR in yeast cells is still not clear. Martina et al. (2014) have recently presented evidence that Rif1 promotes resection in yeast, thus, in principle, favoring HR over NHEJ.

Rif1 CONTROLS THE TEMPORAL PATTERN OF DNA REPLICATION INITIATION THROUGH THE PP1 PHOSPHATASE

One striking phenotype to emerge recently in studies of *RIF1* deletions in budding and fission yeasts, as well as knock-down experiments in mouse and human cells, is a global effect on the temporal pattern of chromosomal DNA replication. In all eukaryotes studied to date, replication in most cell types initiates at characteristic sites (origins) whose "firing" can occur
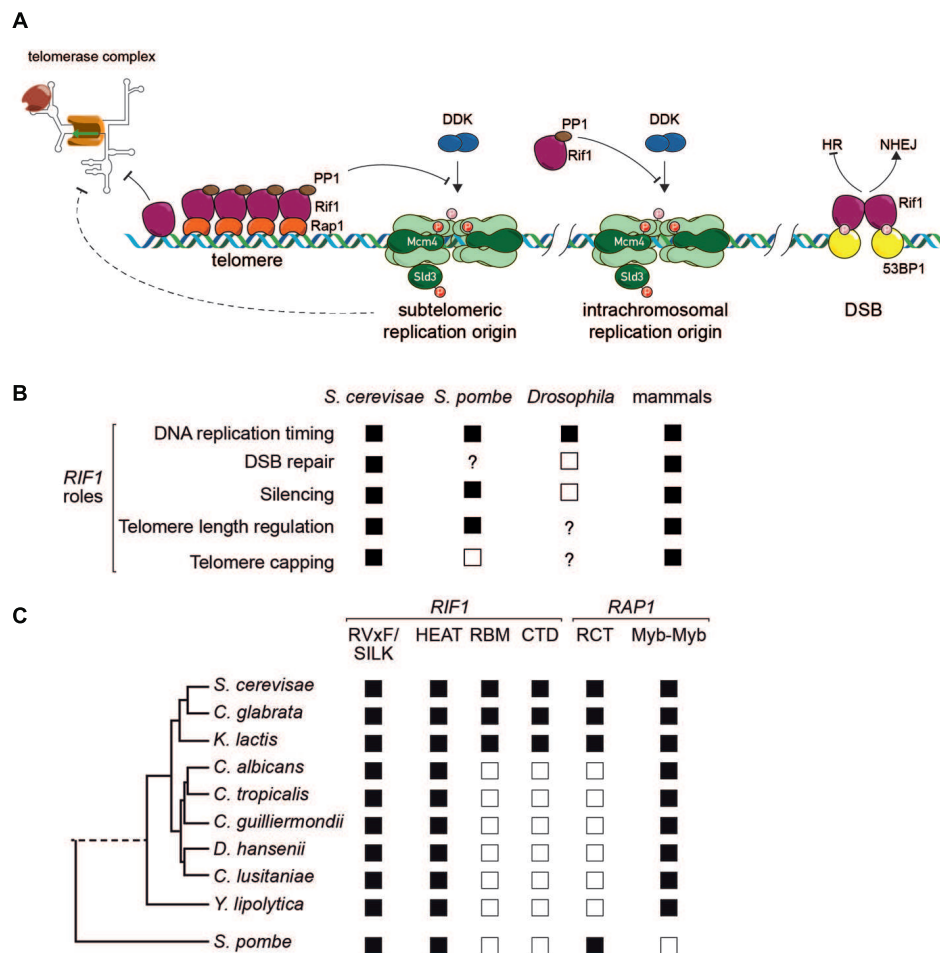
FIGURE 2 | (A) Schematic representation of Rif1 function at budding yeast telomeres (left), replication origins in yeasts (center) and double-strand breaks in mammalian cells (right). See text for details. **(B)** Conservation of Rif1 functions across species. Filled squares indicate that the function is present, according to at least one report; open squares indicate evidence for absence of function in at least one report; "?" indicates that the function has not been tested for in that organism. Note that for mammals, a Rif1 role in silencing, telomere length regulation and telomere capping has been examined only in mouse ES cells (Dan et al., 2014). **(C)** Conservation of Rif1 and Rap1 domains in yeasts. Filled square indicates presence and open square absence of indicated domain or motif.

either early during S phase, or at middle or late periods. This temporal pattern of replication initiation is highly controlled, but the underlying mechanisms are still poorly understood. The finding, that *rif1Δ* cells in both budding (Lian et al., 2011; Peace et al., 2014) and fission (Hayano et al., 2012) yeasts display major alterations in replication timing, was thus of considerable importance. Similar results were reported in studies of mouse and human cells in culture that were depleted for Rif1 (Cornacchia et al., 2012; Yamazaki et al., 2012). In *Schizosaccharomyces pombe* and mammalian cells the effects of Rif1 on replication timing were widespread, whereas in budding yeast initial studies suggested that they might be more restricted to telomere-proximal regions, where most late-firing origins are found.

Several lines of evidence provided clues to the mechanism by which Rif1 influences replication timing. The first of these, mentioned above, was the finding by Sreesankar et al. (2012) of the conserved SILK/RVxF motifs in Rif1, suggesting that

the protein might serve as a PP1 phosphatase co-factor or recruitment scaffold. A second key finding made in both fission and budding yeast, was that deletion of *RIF1* permits the growth of mutants with reduced Cdc7 (SpHsk1) protein kinase activity (Hayano et al., 2012; Dave et al., 2014; Hiraga et al., 2014; Mattarocci et al., 2014). Cdc7/Hsk1 kinase is the catalytic subunit of the Dbf4-dependent kinase (DDK) required for activation of the pre-Replication Complex (pre-RC). This genetic interaction suggests that Rif1 acts as a negative regulator of a process promoted by the DDK (**Figure 2A**). As predicted by this model, phosphorylation of two DDK targets in the pre-RC, Mcm4, part of the replicative helicase, and Sld3, a conserved adaptor protein involved in assembly of an active DNA polymerase on the pre-RC, is increased in point mutants affecting the Rif1 SILK/RVxF motifs (Dave et al., 2014; Hiraga et al., 2014; Mattarocci et al., 2014). Interestingly, suppression of *CDC7* mutation in budding yeast also requires the Rif1 HEAT motif region (Hiraga et al., 2014).

Given the presence of SILK/RVxF motifs in all Rif1 homologs, from yeast to human, it is tempting to speculate that the Rif1–PP1 interaction is ubiquitous. Indeed, this conclusion is supported by biochemical findings in human cells (Trinkle-Mulcahy et al., 2006). A strong prediction from the studies in both fission and budding yeast, but yet to be tested, is that SILK/RVxF mutations in mammalian Rif1 homologs will be defective in the PP1 interaction and display aberrant patterns of DNA replication.

One important mechanistic question that is still not fully understood is how Rif1 action is targeted so as to affect some but not all origins. In budding yeast this is partly resolved, since as pointed out above Rif1 localizes to telomeres through a network of interactions with Rap1, and firing of subtelomeric origins is strongly inhibited by Rif1 (see **Figure 2A**). Nevertheless, normally dormant chromosome-internal origins are activated in *rif1Δ* cells and there is so far no indication of how (or even if) Rif1 is targeted to these sites. In *Schizosaccharomyces pombe,* one very recent study provides evidence that Rif1 is recruited through an interaction with G-quadruplex DNA structures (Kanoh et al., 2015). An even more recent study in mouse embryonic stem cells (ESCs) indicates that Rif1 acts at the level of nuclear architecture to constrain late-replicating chromosomal domains to interact with each other exclusively during the period in G1 when replication timing is established (Foti et al., 2015).

A COMMON THREAD IN Rif1 FUNCTION THROUGHOUT EVOLUTION?

Recent studies thus now point to control of DNA replication initiation and DNA repair as highly conserved functions of eukaryotic Rif1 homologs (**Figures 2A,B**). The likely conservation of the Rif1–PP1 interaction throughout evolution, as well as the replication initiation targets identified in budding yeast (Mcm4 and Sld3), suggests that this Rif1 function may be the most conserved in mechanistic detail. The conservation of Rif1's function in the DDR is presently less clear. Here the role of mammalian Rif1 is better defined, with its recruitment to sites of damage requiring an interaction with 53BP1. In budding yeast the 53BP1 homolog, Rad9, counteracts the function of Rif1 (Martina et al., 2014), perhaps explaining why Rif1 in yeast appears to promote, rather than block 5′ end resection, at least in G1 cells. We find it interesting that data from both yeast and human cells are beginning to point to a role for the highly conserved HEAT repeat domain of Rif1 in localizing Rif1 to sites of damage (Xue et al., 2011; Escribano-Diaz et al., 2013). Although a C-terminal conserved domain with DNA-binding properties has been implicated in efficient hRif1 recruitment at stalled replication forks (Xu et al., 2010), the function of this domain in the DDR is still controversial (Escribano-Diaz et al., 2013). Furthermore, the possible role of the Rif1–PP1 interaction in the DDR has yet to be explored. Finally, the more general question of a possible relationship between the replication timing and DNA damage/repair functions of Rif1 has yet to be addressed. In this regard it is worth noting

that replication provides sister chromatids that can facilitate homologous repair

APPROPRIATION OF Rif1 AT YEAST TELOMERES: HOW AND WHY?

As pointed out above, and illustrated in **Figure 1A**, Rif1 appears to be localized to native (capped) telomeres only in yeasts. Yet again, though, the evolutionary scenario leading to this situation is uncertain, due to the different mechanisms for Rif1 telomere recruitment employed by fission and budding yeasts. In the budding yeast *Saccharomyces cerevisiae,* Rif1 localizes to telomeres through a network of interactions with ScRap1, as detailed above. However, the SpRif1 does not require SpRap1 for telomere binding, but instead localizes to telomeres through an interaction with Taz1, the duplex DNA telomere-binding protein in this organism. Thus, the most conserved partner of Rif1 in yeast shelterin complexes, Rap1, is not universally used for its recruitment. This curious fact may be explained by the observation that Rap1 probably emerged as a direct duplex DNA telomere-binding protein only in the *Saccharomycotina* yeasts where its Myb-like DNA-binding domain underwent duplication (**Figure 2C**). The budding yeasts still retain a Taz1/TRF2-like protein, called Tbf1, which itself retains telomere-capping functions (Ribaud et al., 2012). One plausible evolutionary scenario is that Taz1/Tbf1 recruited Rif1 to telomeres in the last common ancestor of fission and budding yeasts, with Rap1 acquiring this function as it replaced Tbf1 as the telomere-binding protein in budding yeasts. However, this scenario leaves open the question of how Rif1 is recruited by Rap1 in the large number of *Saccharomycotina* clades (including the well studied human pathogen *Candida albicans*) where Rap1 has no recognizable RCT domain (**Figure 2C**). Significantly, the Rif1 homologs in these organisms lack recognizable RBM and CTD domains (**Figure 2C**), implying that, if Rap1 does indeed recruit Rif1 to telomeres in these organisms (yet to be demonstrated experimentally), it does so through a different set of interactions.

It is interesting to consider what selective advantage telomeric Rif1 localization might afford to yeasts. One possibility is that modulation of replication timing at sub-telomeric regions by Rif1 provides a mechanism to regulate telomerase action as a function of telomere length, at least in part because early replication, which occurs at short telomeres, permits increased elongation in a given cell cycle (Bianchi and Shore, 2007) (**Figure 2A**). This may be particularly advantageous in yeasts where telomere repeat tracts are more than an order of magnitude shorter in length than in mammals and often have an irregular repeat sequence, both of which may limit t-loop formation. In addition, Rif1's still poorly understood end-capping function (Xue et al., 2011; Ribeyre and Shore, 2012; Martina et al., 2014) might also contribute to telomerase regulation (**Figure 2A**). It is also worth noting that the late-replicating sub-telomeric regions in *Saccharomyces cerevisiae* are at least partly heterochromatic and serve as a niche for gene families that play an important role in environmental adaptation (Brown et al., 2010). Their late replication causes a

higher rate of mutagenesis (Lang and Murray, 2011), which has been speculated to confer a selective advantage in fluctuating environmental conditions.

As a closing word of caution, we note that the unique presence of Rif1 at native telomeres in yeasts might be more apparent than real. It is possible that Rif1 is present at capped telomeres in metazoans, but in low amounts that have so far escaped detection, perhaps because it acts transiently during telomere replication and/or reassembly of the telomere cap, or in cell types that have not been carefully studied. In this regard it is worth noting that Rif1 is highly expressed in mouse ESCs and a recent report suggests that it is telomere-localized in these cells, where it plays a role in sub-telomeric heterochromatin formation (Dan et al., 2014). Interestingly, it appears that Rif1 represses a gene, *Zscan4*, a gene whose product promotes HR at telomere repeats. It seems clear that we are only beginning to understand the various functions of Rif1, much less their underlying mechanisms and evolutionary origins. The recent interest that Rif1 has attracted in both the DNA replication and DNA repair fields suggests that the coming years will bring new and important discoveries about this remarkably multifunctional protein.

AUTHOR CONTRIBUTIONS

SM, LH, and AL made equal contributions to this work. All authors listed have made substantial and direct intellectual contributions to this work and have approved it for publication.

FUNDING

REFERENCES

Anbalagan, S., Bonetti, D., Lucchini, G., and Longhese, M. P. (2011). Rif1 supports the function of the CST complex in yeast telomere capping. *PLoS Genet.* 7:e1002024. doi: 10.1371/journal.pgen.1002024

Bianchi, A., and Shore, D. (2007). Early replication of short telomeres in budding yeast. *Cell* 128, 1051–1062. doi: 10.1016/j.cell.2007.01.041

Bianchi, A., and Shore, D. (2008). How telomerase reaches its end: mechanism of telomerase regulation by the telomeric complex. *Mol. Cell* 31, 153–165. doi: 10.1016/j.molcel.2008.06.013

Brown, C. A., Murray, A. W., and Verstrepen, K. J. (2010). Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr. Biol.* 20, 895–903. doi: 10.1016/j.cub.2010.04.027

Buck, S. W., and Shore, D. (1995). Action of a RAP1 carboxy-terminal silencing domain reveals an underlying competition between HMR and telomeres in yeast. *Genes Dev.* 9, 370–384. doi: 10.1101/gad.9.3.370

Buonomo, S. B., Wu, Y., Ferguson, D., and de Lange, T. (2009). Mammalian Rif1 contributes to replication stress survival and homology-directed repair. *J. Cell Biol.* 187, 385–398. doi: 10.1083/jcb.200902039

Castano, I., Pan, S. J., Zupancic, M., Hennequin, C., Dujon, B., and Cormack, B. P. (2005). Telomere length control and transcriptional regulation of subtelomeric adhesins in *Candida glabrata*. *Mol. Microbiol.* 55, 1246–1258. doi: 10.1111/j.1365-2958.2004.04465.x

Chapman, J. R., Barral, P., Vannier, J. B., Borel, V., Steger, M., Tomas-Loba, A., et al. (2013). RIF1 is essential for 53BP1-dependent nonhomologous end joining and suppression of DNA double-strand break resection. *Mol. Cell* 49, 858–871. doi: 10.1016/j.molcel.2013.01.002

Cornacchia, D., Dileep, V., Quivy, J. P., Foti, R., Tili, F., Santarella-Mellwig, R., et al. (2012). Mouse Rif1 is a key regulator of the replication-timing programme in mammalian cells. *EMBO J.* 31, 3678–3690. doi: 10.1038/emboj.2012.214

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., et al. (2010). The genetic landscape of a cell. *Science* 327, 425–431. doi: 10.1126/science.1180823

Dan, J., Liu, Y., Liu, N., Chiourea, M., Okuka, M., Wu, T., et al. (2014). Rif1 maintains telomere length homeostasis of ESCs by mediating heterochromatin silencing. *Dev. Cell* 29, 7–19. doi: 10.1016/j.devcel.2014.03.004

Dave, A., Cooley, C., Garg, M., and Bianchi, A. (2014). Protein phosphatase 1 recruitment by Rif1 regulates DNA replication origin firing by counteracting DDK activity. *Cell Rep.* 7, 53–61. doi: 10.1016/j.celrep.2014.02.019

de Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* 19, 2100–2110. doi: 10.1101/gad.1346005

de Lange, T. (2009). How telomeres solve the end-protection problem. *Science* 326, 948–952. doi: 10.1126/science.1170633

Di Virgilio, M., Callen, E., Yamane, A., Zhang, W., Jankovic, M., Gitlin, A. D., et al. (2013). Rif1 prevents resection of DNA breaks and promotes immunoglobulin class switching. *Science* 339, 711–715. doi: 10.1126/science.1230624

Escribano-Diaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J. T., Tkáč, J., et al. (2013). A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Mol. Cell* 49, 872–883. doi: 10.1016/j.molcel.2013.01.001

Foti, R., Gnan, S., Cornacchia, D., Dileep, V., Bulut-Karslioglu, A., Diehl, S., et al. (2015). Nuclear architecture organized by Rif1 underpins the replication-timing program. *Mol. Cell* 61, 260–273. doi: 10.1016/j.molcel.2015.12.001

Gao, H., Toro, T. B., Paschini, M., Braunstein-Ballew, B., Cervantes, R. B., and Lundblad, V. (2010). Telomerase recruitment in *Saccharomyces cerevisiae* is not dependent on Tel1-mediated phosphorylation of Cdc13. *Genetics* 186, 1147–1159. doi: 10.1534/genetics.110.122044

Gottschling, D. E., Aparicio, O. M., Billington, B. L., and Zakian, V. A. (1990). Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* 63, 751–762. doi: 10.1016/0092-8674(90)90141-Z

Greenwood, J., and Cooper, J. P. (2012). Non-coding telomeric and subtelomeric transcripts are differentially regulated by telomeric and heterochromatin assembly factors in fission yeast. *Nucleic Acids Res.* 40, 2956–2963. doi: 10.1093/nar/gkr1155

Guenole, A., Srivas, R., Vreeken, K., Wang, Z. Z., Wang, S., Krogan, N. J., et al. (2013). Dissection of DNA damage responses using multiconditional genetic interaction maps. *Mol. Cell* 49, 346–358. doi: 10.1016/j.molcel.2012.11.023

Hardy, C. F. J., Sussel, L., and Shore, D. (1992). A RAP1-interacting protein involved in silencing and telomere length regulation. *Genes Dev.* 6, 801–814. doi: 10.1101/gad.6.5.801

Hayano, M., Kanoh, Y., Matsumoto, S., Renard-Guillet, C., Shirahige, K., and Masai, H. (2012). Rif1 is a global regulator of timing of replication origin firing in fission yeast. *Genes Dev.* 26, 137–150. doi: 10.1101/gad.178491.111

Hiraga, S., Alvino, G. M., Chang, F., Lian, H. Y., Sridhar, A., Kubota, T., et al. (2014). Rif1 controls DNA replication by directing Protein Phosphatase 1 to reverse Cdc7-mediated phosphorylation of the MCM complex. *Genes Dev.* 28, 372–383. doi: 10.1101/gad.231258.113

Kanoh, J., and Ishikawa, F. (2001). spRap1 and spRif1, recruited to telomeres by Taz1, are essential for telomere function in fission yeast. *Curr. Biol.* 11, 1624–1630. doi: 10.1016/S0960-9822(01)00503-6

Kanoh, Y., Matsumoto, S., Fukatsu, R., Kakusho, N., Kono, N., Renard-Guillet, C., et al. (2015). Rif1 binds to G quadruplexes and suppresses replication over long distances. *Nat. Struct. Mol. Biol.* 22, 889–897. doi: 10.1038/nsmb.3102

Kyrion, G., Liu, K., Liu, C., and Lustig, A. J. (1993). RAP1 and telomere structure regulate telomere position effects in *Saccharomyces cerevisiae*. *Genes Dev.* 7, 1146–1159. doi: 10.1101/gad.7.7a.1146

Lang, G. I., and Murray, A. W. (2011). Mutation rates across budding yeast chromosome VI are correlated with replication timing. *Genome Biol. Evol.* 3, 799–811. doi: 10.1093/gbe/evr054

Lian, H. Y., Robertson, E. D., Hiraga, S., Alvino, G. M., Collingwood, D., McCune, H. J., et al. (2011). The effect of Ku on telomere replication time is mediated by telomere length but is independent of histone tail acetylation. *Mol. Biol. Cell* 22, 1753–1765. doi: 10.1091/mbc.E10-06-0549

Lingner, J., Cooper, J. P., and Cech, T. R. (1995). Telomerase and DNA: no longer a lagging strand problem? *Science* 269, 1533–1534. doi: 10.1126/science.7545310

Martina, M., Bonetti, D., Villa, M., Lucchini, G., and Longhese, M. P. (2014). *Saccharomyces cerevisiae* Rif1 cooperates with MRX-Sae2 in promoting DNA-end resection. *EMBO Rep.* 15, 695–704. doi: 10.1002/embr.201338338

Mattarocci, S., Shyian, M., Lemmens, L., Damay, P., Altintas, D. M., Shi, T., et al. (2014). Rif1 controls DNA replication timing in yeast through the PP1 phosphatase Glc7. *Cell Rep.* 7, 62–69. doi: 10.1016/j.celrep.2014.03.010

Michelson, R. J., Rosenstein, S., and Weinert, T. (2005). A telomeric repeat sequence adjacent to a DNA double-stranded break produces an anticheckpoint. *Genes Dev.* 19, 2546–2559. doi: 10.1101/gad.1293805

Miller, K. M., Ferreira, M. G., and Cooper, J. P. (2005). Taz1, Rap1 and Rif1 act both interdependently and independently to maintain telomeres. *EMBO J.* 24, 3128–3135. doi: 10.1038/sj.emboj.7600779

Olovnikov, A. M. (1973). A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *J. Theor. Biol.* 41, 181–190. doi: 10.1016/0022-5193(73)90198-7

Peace, J. M., Ter-Zakarian, A., and Aparicio, O. M. (2014). Rif1 regulates initiation timing of late replication origins throughout the *S. cerevisiae* genome. *PLoS ONE* 9:e98501. doi: 10.1371/journal.pone.0098501

Ribaud, V., Ribeyre, C., Damay, P., and Shore, D. (2012). DNA-end capping by the budding yeast transcription factor and subtelomeric binding protein Tbf1. *EMBO J.* 31, 138–149. doi: 10.1038/emboj.2011.349

Ribeyre, C., and Shore, D. (2012). Anticheckpoint pathways at telomeres in yeast. *Nat. Struct. Mol. Biol.* 19, 307–313. doi: 10.1038/nsmb.2225

Rusche, L. N., Kirchmaier, A. L., and Rine, J. (2003). The establishment, inheritance, and function of silenced chromatin in *Saccharomyces cerevisiae*. *Annu. Rev. Biochem.* 72, 481–516. doi: 10.1146/annurev.biochem.72.121801.161547

Shi, T., Bunker, R. D., Mattarocci, S., Ribeyre, C., Faty, M., Gut, H., et al. (2013). Rif1 and Rif2 shape telomere function and architecture through multivalent Rap1 interactions. *Cell* 153, 1340–1353. doi: 10.1016/j.cell.2013.05.007

Silverman, J., Takai, H., Buonomo, S. B., Eisenhaber, F., and de Lange, T. (2004). Human Rif1, ortholog of a yeast telomeric protein, is regulated by ATM and

53BP1 and functions in the S-phase checkpoint. *Genes Dev.* 18, 2108–2119. doi: 10.1101/gad.1216004

Sreesankar, E., Senthilkumar, R., Bharathi, V., Mishra, R. K., and Mishra, K. (2012). Functional diversification of yeast telomere associated protein, Rif1, in higher eukaryotes. *BMC Genomics* 13:255. doi: 10.1186/1471-2164-13-255

Trinkle-Mulcahy, L., Andersen, J., Lam, Y. W., Moorhead, G., Mann, M., and Lamond, A. I. (2006). Repo-Man recruits PP1 gamma to chromatin and is essential for cell viability. *J. Cell Biol.* 172, 679–692. doi: 10.1083/jcb.200508154

Watson, J. D. (1972). Origin of concatemeric T7 DNA. *Nat. New Biol.* 239, 197–201. doi: 10.1038/newbio239197a0

Wotton, D., and Shore, D. (1997). A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in *Saccharomyces cerevisiae*. *Genes Dev.* 11, 748–760. doi: 10.1101/gad.11.6.748

Xu, D., Muniandy, P., Leo, E., Yin, J., Thangavel, S., Shen, X., et al. (2010). Rif1 provides a new DNA-binding interface for the Bloom syndrome complex to maintain normal replication. *EMBO J.* 29, 3140–3155. doi: 10.1038/emboj.2010.186

Xu, L., and Blackburn, E. H. (2004). Human Rif1 protein binds aberrant telomeres and aligns along anaphase midzone microtubules. *J. Cell Biol.* 167, 819–830. doi: 10.1083/jcb.200408181

Xue, Y., Rushton, M. D., and Maringele, L. (2011). A novel checkpoint and RPA inhibitory pathway regulated by Rif1. *PLoS Genet.* 7:e1002417. doi: 10.1371/journal.pgen.1002417

Yamazaki, S., Ishii, A., Kanoh, Y., Oda, M., Nishito, Y., and Masai, H. (2012). Rif1 regulates the replication timing domains on the human genome. *EMBO J.* 31, 3667–3677. doi: 10.1038/emboj.2012.180

Zimmermann, M., Lottersberger, F., Buonomo, S. B., Sfeir, A., and de Lange, T. (2013). 53BP1 regulates DSB repair using Rif1 to control 5' end resection. *Science* 339, 700–704. doi: 10.1126/science.1231573

Evolution of TERT-interacting lncRNAs: expanding the regulatory landscape of telomerase

Andrew D. L. Nelson[1]* and Dorothy E. Shippen[2]*

[1] School of Plant Sciences, University of Arizona, Tucson, AZ, USA, [2] Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX, USA

Long non-coding RNAs (lncRNAs) evolve rapidly and are functionally diverse. The emergence of new lncRNAs is driven by genome disturbance events, including whole genome duplication, and transposition. One of the few lncRNAs with a conserved role throughout eukaryotes is the telomerase RNA, TER. TER works in concert with the telomerase reverse transcriptase (TERT) to maintain telomeres. Here we discuss recent findings from *Arabidopsis thaliana* and its relatives illustrating the remarkable evolutionary flexibility within TER and the potential for non-canonical TERT-lncRNA interactions. We highlight the two TERs in *A. thaliana*. One is a conventional telomerase template. The other lncRNA negatively regulates telomerase activity in response to DNA damage, a function mediated by co-option of a transposable element. In addition, we discuss evidence for multiple independent TER loci throughout the plant family Brassicaceae, and how these loci not only reflect rapid convergent evolution, but also the flexibility of having a lncRNA at the core of telomerase. Lastly, we discuss the propensity for TERT to bind a suite of non-templating lncRNAs, and how such RNAs may facilitate telomerase regulation and off-telomere functions.

Keywords: telomerase, TER, lncRNA, evolution, *Arabidopsis*

Introduction

A major breakthrough in biology was the discovery that much of eukaryotic genomes are transcribed, yet only a small fraction of the transcripts derive from protein-coding genes. Most transcripts are long non-coding RNAs (lncRNAs). Generated from what were originally believed to be "dark" regions of the genome, lncRNAs number in the thousands to tens of thousands. Although only a few lncRNAs have been assigned a biological function, these molecules play essential roles in epigenetic regulation, stem cell biology and signal transduction and are emerging as important targets in human disease (Lee et al., 1996; Guttman et al., 2011; Wapinski and Chang, 2011; Scheuermann and Boyer, 2013). The molecular mechanisms of lncRNAs are varied, but appear to fall into four major categories: (1) molecular signals, (2) molecular decoys, (3) guides, and (4) scaffolds (Wang and Chang, 2011).

One of the best-studied lncRNAs is TER, the telomerase RNA. TER can be defined as a scaffolding lncRNA as it assembles into a ribonucleoprotein complex with several proteins including the reverse transcriptase TERT. TERT reiteratively copies a templating sequence embedded in TER to establish and maintain telomere repeats on chromosome ends. In stem and germline cells telomerase must continually replenish telomeric DNA to avoid cellular senescence, but in cells with limited proliferation programs the enzyme is repressed to avert tumorigenesis (Bernardes de Jesus and Blasco, 2013; Günes and Rudolph, 2013). Telomerase must also be precluded from acting at

double-strand breaks (DSBs) to promote faithful DNA repair. Consequently, telomerase is subjected to multiple levels of regulation that target both TERT and TER (Cifuentes-Rojas and Shippen, 2012; Egan and Collins, 2012).

TER is highly variable in nucleotide sequence and size, ranging from ~150 nucleotides in some ciliates to more than 1.2 kb in budding yeast (Egan and Collins, 2012). Despite its sequence variability, TER harbors conserved secondary and tertiary structures that are critical for TERT interaction and telomerase catalysis. These elements include a single-stranded region bearing the telomere template and a template boundary element that demarcates the 5′ end of the template. TERT binding is mediated by a pseudoknot adjacent to the telomere template (Zhang et al., 2011; Egan and Collins, 2012) and a stem terminus element (STE; Blackburn and Collins, 2011). Notably, the TER-TERT interaction does not require an intact telomere template, leaving open the opportunity for alternative lncRNAs to assemble into an RNP complex with TERT.

Although TERT and TER are sufficient to reconstitute telomerase enzyme activity *in vitro*, the essential domains of TER can be whittled down to a "Mini T" consisting of only ~150 nts (Chen and Greider, 2003; Zappulla et al., 2005; Cifuentes-Rojas et al., 2011). Because most of the structural similarity within eukaryotic TERs lies within these 150 nts, conforming to TERT's catalytic needs is a primary driver of TER conservation. TER assembles with suite of telomerase accessory proteins besides TERT that promote RNP maturation, modulate enzyme activity and facilitate telomerase recruitment to chromosome ends. More divergent than TERT, the accessory proteins typically are not shared between the major eukaryotic lineages (Collins, 2006). The ability of TER to accommodate a dynamic array of protein binding partners and yet retain its templating capacity demonstrates the advantage of having a lncRNA at the heart of the telomerase enzyme.

The Impact of Genome Dynamics on lncRNA Evolution

TER, like other lncRNAs, does not harbor an open reading frame and thus can readily absorb nucleotide changes without a cost to fitness (Ponting et al., 2009; Kutter et al., 2012). Indeed, lncRNAs evolve rapidly and their evolution is influenced by factors besides the accumulation of nucleotide changes. Referred to here as genome disturbance events, whole genome duplication (WGD), genome rearrangement, and transposition all contribute to the volatility of lncRNA repertoires in eukaryotes (Freeling et al., 2012; Kapusta et al., 2013). Studies in vertebrates suggest that as genomes evolve, the lncRNA population slowly changes due to accumulation of nucleotide changes and local rearrangements (**Figure 1A**). In contrast, a genome disturbance event can trigger a dramatic spike in the emergence of novel lncRNAs and decay of more ancient ones. Following WGD duplicated chromosomes undergo a process called fractionation, whereby genes and whole genomic regions accumulate mutations and decay at a rapid rate (Freeling et al., 2012). While this process often leads to gene loss, pseudogenization or promoter acquisition can give rise to novel lncRNAs (Ponting et al., 2009). Genome disturbances are
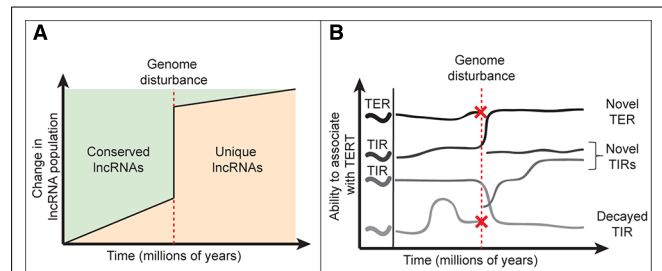
FIGURE 1 | Impact of genome dynamics on lncRNA evolution and TIR populations. (A) Model for lncRNA evolution. Normally, lncRNAs evolve gradually due to accumulation of nucleotide changes and localized genome rearrangement events. However, genome disturbance events (red dashed line) accelerate lncRNA evolution leading to decay or loss of conserved lncRNAs and birth of new lncRNAs. **(B)** Impact of genome disturbance on TERT interacting RNA populations. Within the pool of lncRNAs that bind TERT, TER likely remains stable (as seen in vertebrates). Non-canonical TERT-interacting RNAs (TIRs) are likely to be more dynamic, moving into and out of the pool over time (decaying TIRs). The canonical TER remains stable until a genome disturbance event occurs (red dashed line), where the possibility of TER loss is high. If the ancient TER locus is lost (red X), another lncRNA, presumably a TIR, will replace it as the templating telomerase RNA. A genome disturbance event can also lead to novel lncRNA emergence **(A)**, whereby some of these RNAs may become TIRs.

associated with rapid changes in lncRNA populations. Vertebrate genomes have remained relatively stable, and sequence orthologs for 20% of human lncRNAs are found in mice, including TER (Chen et al., 2000; Ponting, 2008; Necsulea et al., 2014). In contrast, less than 1% of *Arabidopsis thaliana* lncRNAs are evident in grape and poplar, two species with similar divergence times as that of human and mouse (Liu et al., 2012). The dramatic difference in identifiable lncRNA orthologs highlights the WGD and genome rearrangements that separate these plant species, and are consistent with the dynamic nature of plant genomes in general (Koenig and Weigel, 2015).

Transposable elements (TE) represent another means by which lncRNAs originate and diversify in vertebrates (Kapusta et al., 2013; Hoen and Bureau, 2015). Transposition can activate transcription adjacent loci, resulting in the birth of novel lncRNAs. TEs can also become incorporated into exons of lncRNAs in a process termed exaptation (Hoen and Bureau, 2015). TEs account for more than 30% of total lncRNA sequence. Moreover, roughly 70% of vertebrate lncRNAs contain at least some trace of repetitive elements. Unlike typical TEs that are silenced by cellular machinery, exapted elements may impart novel functions as well as contribute to integral facets of lncRNA maturation, such as transcription initiation, splicing, and polyadenylation (Keren et al., 2010; Kapusta and Feschotte, 2014). Additionally, exapted TEs are a common source of lineage-specific differential gene regulation (Lowe and Haussler, 2012). Johnson and Guigó (2014) argue that TEs have the potential to act as pre-formed functional RNA domains, endowing binding sites for novel interaction partners. For instance, TEs within XIST stimulate interactions with PRC2 and splicing factor ASF2 (Wutz et al., 2002; Jeon and Lee, 2011). As discussed below, TE exaptation into TER has dramatically influenced telomerase regulation in *A. thaliana*.

Given the volatile environment in which lncRNAs evolve, it is not surprising that TERs from different eukaryotic lineages bear little similarity to one another in both sequence and synteny (Chen et al., 2000; Cifuentes-Rojas et al., 2011; Qi et al., 2013). TERs from the major lineages likely represent convergent evolution, where unique and unrelated TERT-interacting RNA (TIR) molecules were adapted for use by the much more conserved TERT protein (**Figure 1B**). Despite their unique origins and disparate sequences, TERs from across much of eukarya have adapted similar core structural motifs and all require the templating domain in order to perform a very basic and conserved function: chromosome end maintenance (Chen et al., 2000; Qi et al., 2013).

Brassicaceae as a System for Comparative lncRNA and Telomere Analyses

Recent data from the plant kingdom is providing unanticipated new insights into TER evolution. Beginning with Barbara McClintock's pioneering work on maize telomeres in the 1930s (McKnight et al., 2002), plants have served as important models for chromosome biology. Their remarkable tolerance to genome instability and frequent WGD makes plants an important counterpoint to mammalian systems for analysis of genome dynamics and evolution. Brassicaceae is the most tractable of plant families and consequently the most valuable resource for comparative genomics. A large and diverse cadre of ∼3600 species, Brassicaceae grows throughout the world's temperate zones and is believed to have arisen ∼65 mya (Koenig and Weigel, 2015). Brassicaceae is home to many agriculturally important plants species, but the most well-known member is *A. thaliana*. Due to its powerful genetics, *A. thaliana* has become the reference species for all plant biology (Jones et al., 2008), and has served as a model for telomere analysis for over 15 years (Watson and Riha, 2010).

The *A. thaliana* genome is compact (130 mb), yet is characterized by three rounds of WGD. The most recent occurred at the base of the family (Koenig and Weigel, 2015). The speciation event that gave rise to *A. thaliana* was followed by genome rearrangement and a reduction in chromosome number. Several other lineages within Brassicaceae have undergone WGD, and chromosome painting reveals a litany of large-scale chromosomal rearrangements (Mandáková and Lysak, 2008; Kagale et al., 2014). Thus, Brassicaceae and *A. thaliana* in particular serve as excellent systems for understanding how telomeres and telomerase components evolve in an ever-changing genomic environment.

Despite the dynamic nature of plant genomes, telomeric DNA has remained remarkably resistant to change. The telomere repeat sequence (TTTAGGG)$_n$ is highly conserved throughout the plant kingdom, with a few interesting exceptions such as the order Asparagales (Sýkorová et al., 2003). Analysis of telomere length for twelve Brassicaceae species reveals some length variation, ranging from 850 bp to ∼9 kb (Nelson et al., 2014). However, this same degree of variation is observed among different ecotypes of *A. thaliana*, suggesting that factors modulating telomere length are conserved (Shakirov and Shippen, 2004). This conclusion is supported by the high degree of conservation associated with many telomere components [e.g., Cdc13/Stn1/Ten1 (CST) and TRF-like proteins; Karamysheva et al., 2004; Song et al., 2008; Surovtseva et al., 2009; Leehy et al., 2013; Nelson et al., 2014].

Duplication of TER: Adding to Nature's Toolbox of Telomerase Regulatory Mechanisms

The identification of telomerase protein components in *A. thaliana* has been driven largely by the conservation of subunits such as TERT, dyskerin and POT1 (Fitzgerald et al., 1999; Shakirov et al., 2005; Surovtseva et al., 2007; Kannan et al., 2008). TER, however, remained elusive until only a few years ago when telomerase-associated RNAs were identified by brute-force enzyme purification. These experiments unexpectedly uncovered more than one TER (Cifuentes-Rojas et al., 2011). TER1 (748 nt) and TER2 (784 nt) each contain 1.5 copies of the plant telomeric repeat sequence embedded in a 220 nt segment of ∼90% identity. In TER2 the conserved region is interrupted by a 529 nt unique sequence, subsequently shown to be a small transposon (see below). The transposon and the 3′ terminus are removed from TER2 to generate a smaller isoform termed TER2s (Cifuentes-Rojas et al., 2012). All three TER isoforms (TER1, TER2, and TER2s) assemble with TERT to reconstitute telomerase activity *in vitro*, indicating that the core elements required for catalysis are located in the conserved regions.

Whereas the discovery of multiple TERs in *A. thaliana* was unusual, there is precedent for alternative telomerase subunits. Moreover, subpopulations of unassembled TERT and TER can be found in human cells (Xi and Cech, 2014), making the exchange and/or incorporation of non-canonical telomerase subunits feasible (**Figure 2**). The ciliated protozoan *Euplotes crassus* encodes three TERT proteins, which presumably assemble with a single TER, and act in different developmental stages to facilitate telomere maintenance during vegetative growth or *de novo* telomere formation during sexual development (Karamysheva et al., 2003). There are also variant TERT isoforms in humans, produced by alternative splicing (Ulaner et al., 2000, 1998; Saebøe-Larssen et al., 2006). A major splice variant (*β*-deletion) that is abundantly expressed in cancer and stem cells lacks the conserved reverse transcriptase domains, and yet retains TER binding. This variant behaves as a dominant negative inhibitor of telomerase (**Figure 2**). It can also protect against apoptosis in cancer cells, likely through a telomerase-independent mechanism (Listerman et al., 2013). A growing list of non-telomeric functions have been ascribed to TERT (Ale-Agha et al., 2014). The influence of lncRNA binding partners on such activities is unclear.

Variant TER isoforms have also been reported. Some appear to be processing intermediates (Chapon et al., 1997; Box et al., 2008). Others including the non-canonical TERs in pig and cow were proposed to be pseudogenes based on the presence of a mutation in the templating domain and deletions in other conserved domains (Chen et al., 2000). However, like hTERT splice variants, these alternative TERs have the potential to serve as dominant negative regulators or to play non-canonical roles in telomere biology (**Figure 2**).
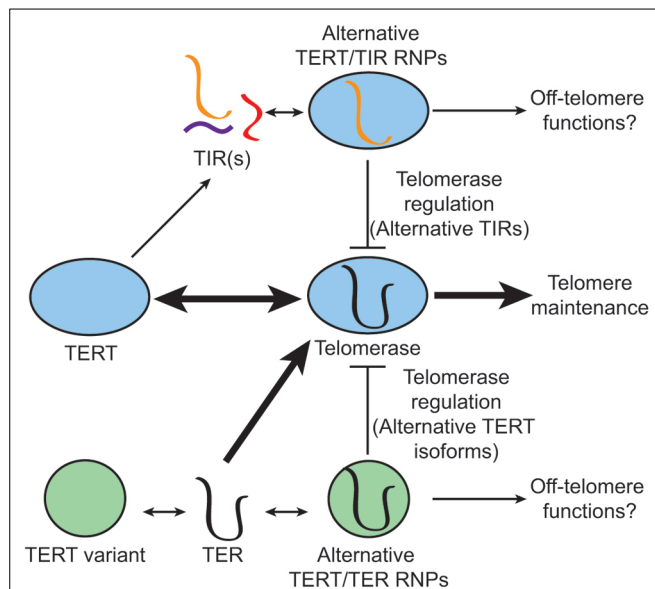
FIGURE 2 | Non-canonical telomerase subunits: alternative modes of enzyme regulation. The conventional telomerase enzyme contains the core subunits TERT and TER (middle), which cooperate in telomere maintenance. TERT can also assemble with non-canonical TERT-interacting lncRNAs (TIRs) (top). Such RNAs may hijack the function of TERT, and in the case of *A. thaliana* TER2, inhibit telomerase activity by sequestering TERT in a non-productive RNP complex. Alternative TERT isoforms (bottom) have also been described. In humans, a major TERT splice variant has lost its catalytic activity, but retains TER binding. Like TER2, this non-canonical TERT is proposed to inhibit telomerase by sequestering TER. Non-canonical telomerase RNP complexes may also have alternative functions off the telomere.

A particularly interesting example of alternative TERs is found in *A. thaliana*, where TER gene duplication provided a fertile breeding ground for the appearance of a novel mode of telomerase regulation. TER1 is the canonical telomere template required for telomere maintenance in *A. thaliana* (Cifuentes-Rojas et al., 2011). TER2, by contrast, negatively regulates the TER1 RNP (Cifuentes-Rojas et al., 2012). Telomerase activity is elevated in *ter2* mutants, while TER2 over-expression reduces the TER1 templating function leading to telomere shortening. Conversely, mutation of the templating domain of TER2 does not cause incorporation of mutant telomere repeats on chromosome ends, indicating that TER2, despite its capacity to direct telomere repeat addition *in vitro*, does not productively engage chromosome ends *in vivo*. Notably, TER2 serves as a lncRNA scaffold for a different set of accessory proteins than TER1, which may contribute to its distinct function *in vivo* (Cifuentes-Rojas et al., 2011, 2012). Furthermore, TERT has a higher affinity for TER2 than for TER1. Thus, TER2 has the ability to serve as a molecular decoy or sponge that sequesters the telomerase catalytic subunit in a non-functional complex.

Telomerase Regulation by Exaptation of a TE in TER

TER2 exhibits another of the lncRNA molecular paradigms: biological signal. Under standard growth conditions TER2 is

a low abundance RNA, more poorly expressed than TER1 or TER2s (Cifuentes-Rojas et al., 2012). However, in response to DSBs, TER2 is rapidly induced and becomes the predominant TER isoform. Telomerase activity is repressed as TER2 levels rise. Remarkably, TER2 induction is not mediated by increased transcription, but rather by increased RNA stability (Xu et al., 2015). Thus, TER2 serves as a rapid regulatory switch linking the DNA damage response directly to telomerase enzyme activity.

Clues for how TER2 might function as a DNA damage sensor came from inspection of another unique feature of this molecule: its 529 nt intervening sequence (removed during the formation of TER2s). The intervening sequence contains no obvious branch point site, and the 5′ and 3′ splice sites do not match mRNA splicing consensus sequences. Instead the boundaries of this element consist of short inverted repeats flanked by two 5 nt direct repeats. Further analysis of similar sequences throughout Brassicaceae indicated that the intervening sequence within TER2 is in fact a small TE, a solo long terminal repeat from a gypsy class of retrotransposons (Xu et al., 2015).

A TE is associated with the majority of TER2 loci in *A. thaliana* ecotypes but not all, providing an opportunity to assess if and how this element modulates telomerase behavior. The unique behavior of TER2 appears to be largely, if not entirely dependent on its TE (Xu et al., 2015). Without the TE, TER2 is a highly stable lncRNA that binds TERT with a lower affinity than TER1. Moreover, in *A. thaliana* ecotypes lacking the TER2 TE, telomerase regulation by DSBs is lost. Thus, exaptation of a TE into the TER2 locus profoundly influenced the regulation and behavior of this lncRNA by endowing it with a DNA damage sensor and the capacity to sequester TERT in a non-productive complex. This mode of telomerase regulation is expected to promote genome stability and may be especially beneficial during meiosis when genome-wide DSBs abound.

Evolution of TER as a TERT-associated lncRNA

Phylogenetic analysis, and particularly gene synteny, has revealed numerous lncRNAs orthologs, including TER, in several eukaryotic lineages (Chen et al., 2000; Qi et al., 2013). Beilstein et al. (2012) employed this strategy to identify an *A. thaliana* TER-like locus from 14 species sampling the breadth of Brassicaceae. However, three unanticipated findings were uncovered. First, *AtTER1* and *AtTER2* loci represent an *A. thaliana*-specific duplication event. In *A. lyrata*, the closest relative of *A. thaliana*, only a single *TER-like* locus was detected. Further analysis showed that the *A. thaliana TER1/TER2* duplication occurred as part of a large-scale genome rearrangement coinciding with *A. thaliana* speciation (Beilstein et al., 2012). Second, contrary to findings from yeast and mammals, there is no clear phylogenetic signature of conservation at the *TER-like* loci in Brassicaceae to infer critical structural and functional elements. The evolutionary pressures placed on each of these loci must be distinct. Third, and most surprisingly, the telomere templating domains of *TER-like* loci in multiple Brassicaceae species including *A. lyrata* carry point mutations that would

preclude synthesis of TTTAGGG repeats. Hence, an alternative locus must encode the canonical TER in many Brassicaceae species.

The Brassicaceae TERs and TIRs provide a fascinating window into both the molecular mechanisms and evolution of lncRNAs. Indeed TER2's emergence by TE exaptation may be only one example of how lncRNAs evolved to regulate TERT. We postulate that transformation of TER2 into a TERT decoy reflects TERT promiscuity for RNA. The ancient origin of TERT from a viral reverse transcriptase supports the notion that TERT evolved RNA specificity over time (Curcio and Belfort, 2007). Even now, sequencing of TIRs in human cells revealed >30 unique RNA species (Maida et al., 2009). In the event that a species' canonical TER locus is lost, a replacement is likely adapted from

the pool of TIRs (**Figure 1B**). Throughout the co-option of a novel TER, TERT would still have the capacity to assemble with a suite of non-templating TIRs and by extension their alternative accessory proteins. TIRs therefore have the potential to modulate conventional and non-conventional TERT-related activities (**Figures 1B** and **2**). Consequently, this intriguing class of lncRNAs provides new insights into regulating telomerase and potentially other cellular functions in cancer and age-associated diseases.

Acknowledgments

References

Ale-Agha, N., Dyballa-Rukes, N., Jakob, S., Altschmied, J., and Haendeler, J. (2014). Cellular functions of the dual-targeted catalytic subunit of telomerase, telomerase reverse transcriptase–potential role in senescence and aging. *Exp. Gerontol.* 56, 189–193. doi: 10.1016/j.exger.2014.02.011

Beilstein, M. A., Brinegar, A. E., and Shippen, D. E. (2012). Evolution of the *Arabidopsis* telomerase RNA. *Front. Genet.* 3:188. doi: 10.3389/fgene.2012.00188

Bernardes de Jesus, B., and Blasco, M. A. (2013). Telomerase at the intersection of cancer and aging. *Trends Genet.* 29, 513–520. doi: 10.1016/j.tig.2013.06.007

Blackburn, E. H., and Collins, K. (2011). Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harb. Perspect. Biol.* 3, a003558. doi: 10.1101/cshperspect.a003558

Box, J. A., Bunch, J. T., Tang, W., and Baumann, P. (2008). Spliceosomal cleavage generates the 3′ end of telomerase RNA. *Nature* 456, 910–914. doi: 10.1038/nature07584

Chapon, C., Cech, T. R., and Zaug, A. J. (1997). Polyadenylation of telomerase RNA in budding yeast. *RNA* 3, 1337–1351.

Chen, J.-L., Blasco, M. A., and Greider, C. W. (2000). Secondary structure of vertebrate telomerase RNA. *Cell* 100, 503–514. doi: 10.1016/S0092-8674(00)80687-X

Chen, J.-L., and Greider, C. W. (2003). Template boundary definition in mammalian telomerase. *Genes Dev.* 17, 2747–2752. doi: 10.1101/gad.1140303

Cifuentes-Rojas, C., Kannan, K., Tseng, L., and Shippen, D. E. (2011). Two RNA subunits and POT1a are components of *Arabidopsis* telomerase. *Proc. Natl. Acad. Sci. U.S.A.* 108, 73–78. doi: 10.1073/pnas.1013021107

Cifuentes-Rojas, C., Nelson, A. D. L., Boltz, K. A., Kannan, K., She, X., and Shippen, D. E. (2012). An alternative telomerase RNA in *Arabidopsis* modulates enzyme activity in response to DNA damage. *Genes Dev.* 26, 2512–2523. doi: 10.1101/gad.202960.112

Cifuentes-Rojas, C., and Shippen, D. E. (2012). Telomerase regulation. *Mutat. Res.* 730, 20–27. doi: 10.1016/j.mrfmmm.2011.10.003

Collins, K. (2006). The biogenesis and regulation of telomerase holoenzymes. *Nat. Rev. Mol. Cell Biol.* 7, 484–494. doi: 10.1038/nrm1961

Curcio, M. J., and Belfort, M. (2007). The beginning of the end: links between ancient retroelements and modern telomerases. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9107–9108. doi: 10.1073/pnas.0703224104

Egan, E. D., and Collins, K. (2012). Biogenesis of telomerase ribonucleoproteins. *RNA* 18, 1747–1759. doi: 10.1261/rna.034629.112

Fitzgerald, M. S., Riha, K., Gao, F., Ren, S., McKnight, T. D., and Shippen, D. E. (1999). Disruption of the telomerase catalytic subunit gene from *Arabidopsis* inactivates telomerase and leads to a slow loss of telomeric DNA. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14813–14818. doi: 10.1073/pnas.96.26.14813

Freeling, M., Woodhouse, M. R., Subramaniam, S., Turco, G., Lisch, D., and Schnable, J. C. (2012). Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Curr. Opin. Plant Biol.* 15, 131–139. doi: 10.1016/j.pbi.2012.01.015

Günes, C., and Rudolph, K. L. (2013). The role of telomeres in stem cells and cancer. *Cell* 152, 390–393. doi: 10.1016/j.cell.2013.01.010

Guttman, M., Donaghey, J., Carey, B. W., Garber, M., Grenier, J. K., Munson, G., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477, 295–300. doi: 10.1038/nature10398

Hoen, D. R., and Bureau, T. E. (2015). Discovery of novel genes derived from transposable elements using integrative genomic analysis. *Mol. Biol. Evol.* 32, 1487–1506. doi: 10.1093/molbev/msv042

Jeon, Y., and Lee, J. T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* 146, 119–133. doi: 10.1016/j.cell.2011.06.026

Johnson, R., and Guigó, R. (2014). The RIDL hypothesis: transposable elements as functional domains of long non-coding RNAs. *RNA* 20, 959–976. doi: 10.1261/rna.044560.114

Jones, A. M., Chory, J., Dangl, J. L., Estelle, M., Jacobsen, S. E., Meyerowitz, E. M., et al. (2008). The impact of *Arabidopsis* on human health: diversifying our portfolio. *Cell* 133, 939–943. doi: 10.1016/j.cell.2008.05.040

Kagale, S., Robinson, S. J., Nixon, J., Xiao, R., Huebert, T., Condie, J., et al. (2014). Polyploid evolution of the Brassicaceae during the cenozoic era. *Plant Cell* 26, 2777–2791. doi: 10.1105/tpc.114.126391

Kannan, K., Nelson, A. D. L., and Shippen, D. E. (2008). Dyskerin is a component of the *Arabidopsis* telomerase RNP required for telomere maintenance. *Mol. Cell. Biol.* 28, 2332–2341. doi: 10.1128/MCB.01490-07

Kapusta, A., and Feschotte, C. (2014). Volatile evolution of long non-coding RNA repertoires: mechanisms and biological implications. *Trends Genet.* 30, 439–452. doi: 10.1016/j.tig.2014.08.004

Kapusta, A., Kronenberg, Z., Lynch, V. J., Zhuo, X., Ramsay, L., Bourque, G., et al. (2013). Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long non-coding RNAs. *PLoS Genet.* 9:e1003470. doi: 10.1371/journal.pgen.1003470

Karamysheva, Z. N., Surovtseva, Y. V., Vespa, L., Shakirov, E. V., and Shippen, D. E. (2004). A C-terminal Myb extension domain defines a novel family of double-strand telomeric DNA-binding proteins in *Arabidopsis*. *J. Biol. Chem.* 279, 47799–47807. doi: 10.1074/jbc.M407938200

Karamysheva, Z., Wang, L., Shrode, T., Bednenko, J., Hurley, L. A., and Shippen, D. E. (2003). Developmentally programmed gene elimination in *Euplotes crassus* facilitates a switch in the telomerase catalytic subunit. *Cell* 113, 565–576. doi: 10.1016/S0092-8674(03)00363-5

Keren, H., Lev-Maor, G., and Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. *Nat. Rev. Genet.* 11, 345–355. doi: 10.1038/nrg2776

Koenig, D., and Weigel, D. (2015). Beyond the thale: comparative genomics and genetics of *Arabidopsis* relatives. *Nat. Rev. Genet.* 16, 285–298. doi: 10.1038/nrg3883

Kutter, C., Watt, S., Stefflova, K., Wilson, M. D., Goncalves, A., Ponting, C. P., et al. (2012). Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.* 8:e1002841. doi: 10.1371/journal.pgen.1002841

Lee, J. T., Strauss, W. M., Dausman, J. A., and Jaenisch, R. (1996). A 450 kb transgene displays properties of the mammalian X-inactivation center. *Cell* 86, 83–94. doi: 10.1016/S0092-8674(00)80079-3

Leehy, K. A., Lee, J. R., Song, X., Renfrew, K. B., and Shippen, D. E. (2013). MERISTEM DISORGANIZATION1 encodes TEN1, an essential telomere

protein that modulates telomerase processivity in *Arabidopsis*. *Plant Cell* 25, 1343–1354. doi: 10.1105/tpc.112.107425

Listerman, I., Sun, J., Gazzaniga, F. S., Lukas, J. L., and Blackburn, E. H. (2013). The major reverse transcriptase-incompetent splice variant of the human telomerase protein inhibits telomerase activity but protects from apoptosis. *Cancer Res.* 73, 2817–2828. doi: 10.1158/0008-5472.CAN-12-3082

Liu, J., Jung, C., Xu, J., Wang, H., Deng, S., Bernad, L., et al. (2012). Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* 24, 4333–4345. doi: 10.1105/tpc.112.102855

Lowe, C. B., and Haussler, D. (2012). 29 mammalian genomes reveal novel exaptations of mobile elements for likely regulatory functions in the human genome. *PLoS ONE* 7:e43128. doi: 10.1371/journal.pone.0043128

Maida, Y., Yasukawa, M., Furuuchi, M., Lassmann, T., Possemato, R., Okamoto, N., et al. (2009). An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA. *Nature* 461, 230–235. doi: 10.1038/nature08283

Mandáková, T., and Lysak, M. A. (2008). Chromosomal phylogeny and karyotype evolution in x=7 crucifer species (Brassicaceae). *Plant Cell* 20, 2559–2570. doi: 10.1105/tpc.108.062166

McKnight, T. D., Riha, K., and Shippen, D. E. (2002). Telomeres, telomerase, and stability of the plant genome. *Plant Mol. Biol.* 48, 331–337. doi: 10.1023/A:1014091032750

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., et al. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505, 635–640. doi: 10.1038/nature12943

Nelson, A. D. L., Forsythe, E. S., Gan, X., Tsiantis, M., and Beilstein, M. A. (2014). Extending the model of *Arabidopsis* telomere length and composition across Brassicaceae. *Chromosom. Res.* 22, 153–166. doi: 10.1007/s10577-014-9423-y

Ponting, C. P. (2008). The functional repertoires of metazoan genomes. *Nat. Rev. Genet.* 9, 689–698. doi: 10.1038/nrg2413

Ponting, C. P., Oliver, P. L., and Reik, W. (2009). Evolution and functions of long noncoding RNAs. *Cell* 136, 629–641. doi: 10.1016/j.cell.2009.02.006

Qi, X., Li, Y., Honda, S., Hoffmann, S., Marz, M., Mosig, A., et al. (2013). The common ancestral core of vertebrate and fungal telomerase RNAs. *Nucleic Acids Res.* 41, 450–462. doi: 10.1093/nar/gks980

Saebøe-Larssen, S., Fossberg, E., and Gaudernack, G. (2006). Characterization of novel alternative splicing sites in human telomerase reverse transcriptase (hTERT): analysis of expression and mutual correlation in mRNA isoforms from normal and tumour tissues. *BMC Mol. Biol.* 7:26. doi: 10.1186/1471-2199-7-26

Scheuermann, J. C., and Boyer, L. A. (2013). Getting to the heart of the matter: long non-coding RNAs in cardiac development and disease. *EMBO J.* 32, 1805–1816. doi: 10.1038/emboj.2013.134

Shakirov, E. V., and Shippen, D. E. (2004). Length regulation and dynamics of individual telomere tracts in wild-type *Arabidopsis*. *Plant Cell* 16, 1959–1967. doi: 10.1105/tpc.104.023093

Shakirov, E. V., Surovtseva, Y. V., Osbun, N., and Shippen, D. E. (2005). The *Arabidopsis* Pot1 and Pot2 proteins function in telomere length homeostasis and chromosome end protection. *Mol. Cell. Biol.* 25, 7725–7733. doi: 10.1128/MCB.25.17.7725-7733.2005

Song, X., Leehy, K., Warrington, R. T., Lamb, J. C., Surovtseva, Y. V., and Shippen, D. E. (2008). STN1 protects chromosome ends in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 19815–19820. doi: 10.1073/pnas.0807867105

Surovtseva, Y. V., Churikov, D., Boltz, K. A., Song, X., Lamb, J. C., Warrington, R., et al. (2009). Conserved telomere maintenance component 1 interacts with STN1 and maintains chromosome ends in higher eukaryotes. *Mol. Cell* 36, 207–218. doi: 10.1016/j.molcel.2009.09.017

Surovtseva, Y. V., Shakirov, E. V., Vespa, L., Osbun, N., Song, X., and Shippen, D. E. (2007). *Arabidopsis* POT1 associates with the telomerase RNP and is required for telomere maintenance. *EMBO J.* 26, 3653–3661. doi: 10.1038/sj.emboj.7601792

Sýkorová, E., Lim, K. Y., Kunická, Z., Chase, M. W., Bennett, M. D., Fajkus, J., et al. (2003). Telomere variability in the monocotyledonous plant order Asparagales. *Proc. Biol. Sci.* 270, 1893–1904. doi: 10.1098/rspb.2003.2446

Ulaner, G. A., Hu, J.-F., Vu, T. H., Oruganti, H., and Giudice, L. C. (2000). Regulation of telomerase by alternate splicing of human telomerase reverse transcriptase (hTERT) in normal and neoplastic ovary, endometrium and myometrium. *Int. J. Cancer* 85, 330–335. doi: 10.1002/(SICI)1097-0215(20000201)85:3<330::AID-IJC6>3.0.CO;2-U

Ulaner, G. A., Hu, J., Vu, T. H., Giudice, L. C., and Hoffman, a R. (1998). Telomerase activity in human development is regulated by human telomerase reverse transcriptase (hTERT) transcription and by alternate splicing of hTERT transcripts telomerase activity in human development is regulated by human telomerase reverse transc. *Cancer Res.* 58, 4168–4172.

Wang, K. C., and Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Mol. Cell* 43, 904–914. doi: 10.1016/j.molcel.2011.08.018

Wapinski, O., and Chang, H. Y. (2011). Long noncoding RNAs and human disease. *Trends Cell Biol.* 21, 354–361. doi: 10.1016/j.tcb.2011.04.001

Watson, J. M., and Riha, K. (2010). Comparative biology of telomeres: where plants stand. *FEBS Lett.* 584, 3752–3759. doi: 10.1016/j.febslet.2010.06.017

Wutz, A., Rasmussen, T. P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.* 30, 167–174. doi: 10.1038/ng820

Xi, L., and Cech, T. R. (2014). Inventory of telomerase components in human cells reveals multiple subpopulations of hTR and hTERT. *Nucleic Acids Res.* 42, 8565–8577. doi: 10.1093/nar/gku560

Xu, H., Nelson, A. D. L., and Shippen, D. E. (2015). A transposable element within the non-canonical telomerase RNA of *Arabidopsis thaliana* modulates telomerase activity in response to DNA damage. *PLoS Genet.* 11:e1005281. doi: 10.1371/journal.pgen.1005281

Zappulla, D. C., Goodrich, K., and Cech, T. R. (2005). A miniature yeast telomerase RNA functions *in vivo* and reconstitutes activity *in vitro*. *Nat. Struct. Mol. Biol.* 12, 1072–1077. doi: 10.1038/nsmb1019

Zhang, Q., Kim, N.-K., and Feigon, J. (2011). Architecture of human telomerase RNA. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20325–20332. doi: 10.1073/pnas.1100279108

Using Centromere Mediated Genome Elimination to Elucidate the Functional Redundancy of Candidate Telomere Binding Proteins in *Arabidopsis thaliana*

Nick Fulcher[1] and Karel Riha[2]*

[1] *Gregor Mendel Institute, Austrian Academy of Sciences, Vienna, Austria,* [2] *Central European Institute of Technology, Masaryk University, Brno, Czech Republic*

Proteins that bind to telomeric DNA form the key structural and functional constituents of telomeres. While telomere binding proteins have been described in the majority of organisms, their identity in plants remains unknown. Several protein families containing a telomere binding motif known as the telobox have been previously described in *Arabidopsis thaliana*. Nonetheless, functional evidence for their involvement at telomeres has not been obtained, likely due to functional redundancy. Here we performed genetic analysis on the TRF-like family consisting of six proteins (TRB1, TRP1, TRFL1, TRFL2, TRFL4, and TRF9) which have previously shown to bind telomeric DNA *in vitro*. We used haploid genetics to create multiple knock-out plants deficient for all six proteins of this gene family. These plants did not exhibit changes in telomere length, or phenotypes associated with telomere dysfunction. This data demonstrates that this telobox protein family is not involved in telomere maintenance in *Arabidopsis*. Phylogenetic analysis in major plant lineages revealed early diversification of telobox proteins families indicating that telomere function may be associated with other telobox proteins.

Keywords: telomeres, centromere, haploid, telobox, protein family

INTRODUCTION

Telomeres represent the nucleoprotein complexes that cap natural chromosome ends and function in the suppression of DNA damage signaling and control of cellular senescence. The classical telomere structure comprises tandem arrays of TTAGG-like sequences which contain G-rich 3′ overhangs at their termini. TRF1 and TRF2 represent the core duplex binding proteins of the mammalian telomere capping complex known as shelterin (de Lange, 2005); TRF1 is thought to be a regulator of telomere length (van Steensel and de Lange, 1997) and TRF2 has been shown to play a central role in protecting chromosome ends from end to end fusions and recombination (van Steensel et al., 1998; Wang et al., 2004). In contrast to the situation in a number of eukaryotic organisms which have extensively characterized chromosome-end capping protein complexes, the plant telomere binding components remain elusive (Watson and Riha, 2010). A hallmark of telomere binding proteins includes the presence of a single Myb domain containing the telobox, a motif that provides specificity to the telomeric sequence (Bilaud et al., 1996). Telobox containing proteins (TRF-like, TRFL) are present in genomes of all major groups of eukaryotes and they have

been considered the prime suspects for *bona fide* telomere binding proteins in plants. Indeed, functional analysis of TRFL proteins in rice and tobacco has indicated their involvement in telomere length homeostasis (Yang et al., 2004; Hong et al., 2007).

TRFL proteins have been extensively studied in *Arabidopsis*. The *Arabidopsis thaliana* genome encodes at least 15 proteins containing a single Myb domain with the telobox that are divided into three families (Zellinger and Riha, 2007). The Smh/TRB family consists of three proteins harboring a histone H1-like motif involved in multimerization, and the Myb domain at the N-terminus (Marian et al., 2003; Kuchar and Fajkus, 2004; Mozgova et al., 2008). The second family includes six proteins (TRFL3, 5, 6, 7, 8, 10; TRFL Group II) that are unable to bind telomeric DNA *in vitro*, and are also unable to form homo- and heterodimers, despite possessing the C-terminal Myb-telobox domain (Karamysheva et al., 2004). The third family also consists of six proteins with the C-terminal Myb domain (TBP1, TRP1, TRFL1, TRFL2, TRFL4 and TRFL9; TRFL Group I), but these proteins homo- and heterodimerize and can efficiently bind to telomeric DNA *in vitro* (Karamysheva et al., 2004). A key feature of this family is a ∼30 amino acid extension of the Myb-telobox domain that is likely responsible for specific binding to plant telomeric DNA. Structural studies of related tobacco and rice TRFL proteins determined that their binding to telomeric DNA occurs in a similar fashion as for human TRF1 (Ko et al., 2008, 2009). Thus, members of the TRFL Group I family have long been considered to act as putative telomere binding proteins in *Arabidopsis*. Nevertheless, plants containing single knockouts within members of this gene family have not shown drastic telomeric phenotypes (Karamysheva et al., 2004). The lack of severe telomere related phenotypes similar to mammalian TRF2 knock-outs suggested a functional redundancy among these proteins in *Arabidopsis*.

Reverse genetics based approaches have been used over many studies in *Arabidopsis* to target functional redundancy amongst gene families. Construction of lines with multiple T-DNA insertions in desired genes can, however, be time consuming requiring extensive genotyping of large populations of recombinant plants. Methods to improve the production of such mutant lines would be greatly beneficial to elucidate functional redundancy within gene families. Centromere mediated genome elimination has proven to be a powerful tool in *Arabidopsis* genetics allowing generation of haploid plants, rapid production of recombinant inbred lines, and reverse breeding approaches (Ravi and Chan, 2010; Seymour et al., 2012; Wijnker et al., 2012; Ravi et al., 2014). Crossing fertile male plants to the female *cenh3*/GFP-*tailswap* haploid inducer allows for the segregation of haploid plants containing genomes from the male parent. This technology also has the potential to easily generate multiple homozygous mutant combinations when crossing plants segregating for numerous T-DNA insertions to the haploid inducer (Ravi et al., 2014). In this case, haploid plants with interesting combinations can be analyzed directly for phenotypic defects, or diploids can also be recovered in the next generation due to spontaneous diploidization. This process would greatly reduce the genotyping workload that is normally

associated with the generation of quadruple or sextuple mutants by selfing alone.

In this study we tackled the functional redundancy thought to occur in *Arabidopsis* TRFL Group I family by using production of haploid plants via centromere mediated genome elimination. We have demonstrated that this method substantially facilitates generation of multiple quadruple, quintuple and sextuple mutants. Surprisingly, results show that multiple mutants do not display drastic telomeric length defects as shown for the mutants in other genes known to act at telomeres. This demonstrates that, at least in *Arabidopsis*, the TRFL protein family harboring the Myb extension does not contribute to telomere protection and/or maintenance. Furthermore, this study shows another use for centromere mediated genome elimination in the production of lines containing multiple mutations.

MATERIALS AND METHODS

Plant Lines

All T-DNA insertions used are shown in Supplementary Table S1 and Supplementary Figure S1. The *tbp1-1* mutant was obtained from the Institut National de la Recherche Agronomique Versailles (INRAV) collection and other alleles were from the European *Arabidopsis* stock centre (NASC). Plants were grown at 22°C in 16 h light/8 h dark cycles.

Centromere Mediated Genome Elimination

The *cenh3*/GFP-*tailswap* haploid inducer line was described previously by (Ravi and Chan, 2010). Homozygous *cenh3* mutant plants were confirmed by PCR genotyping using derived Cleaved Amplified Polymorphic Sequence (dCAPS) oligos (5′-GGTGCGATTTCTCCAGCA GTAAAAATC-3′ and 5′-CTGAG AAGATGAAGCACCGGCGATAT-3′). Resulting PCR products were digested with EcoRV, cleaved wild type (WT) alleles produced 191 and 24 bp fragments.

Haploid inducer *cenh3*/GFP-*tailswap* lines are mostly male sterile, but can be crossed as female. Heterozygous quadruple or sextuple mutants were crossed to *cenh3*/GFP-*tailswap* lines to produce haploid offspring that were homozygous for a combination of insertions derived from the male parent. Only plants that displayed the haploid phenotype as described by Ravi and Chan (2010) were selected for further analysis. These haploids were then subject to PCR genotyping using oligos shown in Supplementary Table S1. Diploid seeds can then be recovered from haploid plants due to spontaneous diploidization which allowed analysis of subsequent generations.

DNA Extraction and Telomere Analysis

One to two leaves were homogenized in 500 μl Extraction buffer (0.2 M Tris pH9, 0.3 M LiCl, 25 mM EDTA, and 1% SDS) tubes were centrifuged for 10 min at 4000 rpm (rcf 1756 g) and 350 μl was transferred to 350 μl isopropanol. Tubes were inverted to mix and centrifuged for 20 min at 4000 rpm. Supernatant was poured

away and the pellet was washed with 70% Ethanol. The remaining pellet was air dried and resuspended in 100 μl dH$_2$O. Telomere length was determined by terminal restriction fragment analysis, and statistical analysis of telomeric smears was performed using the TeloTool software (Gohring et al., 2014; Fulcher et al., 2015). Integrity of blunt ended telomeres was determined as previously described (Kazda et al., 2012).

Phylogenetic Analysis

Sequences of telobox containing proteins were obtained from indicated plant genomes by using http://www.phytozome.net, protein BLAST searches with the *A. thaliana* TRFL6 protein sequence as a query. Proteins were aligned by the ClustalW method and phylogenetic trees were constructed by Neighbor Joining method using CLC Main Workbench software (Qiagen).
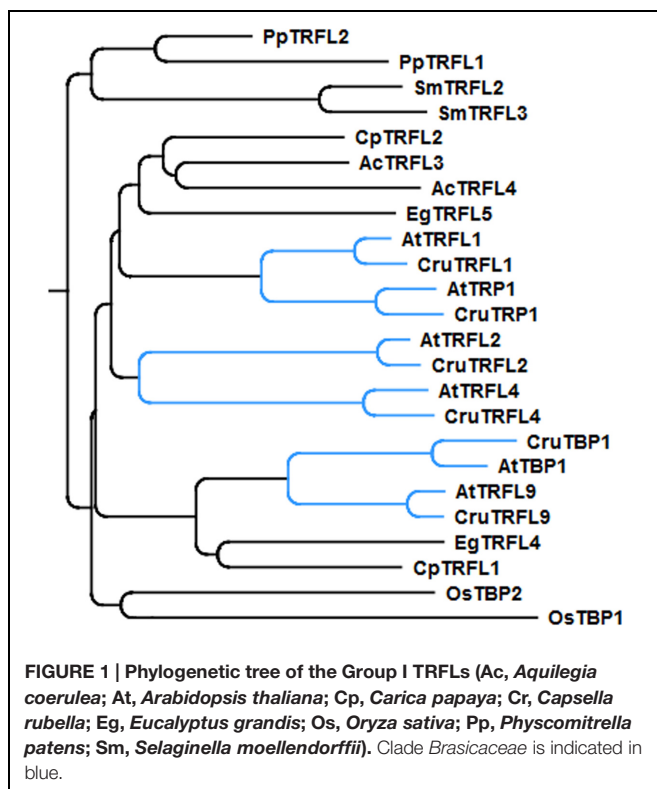
RESULTS

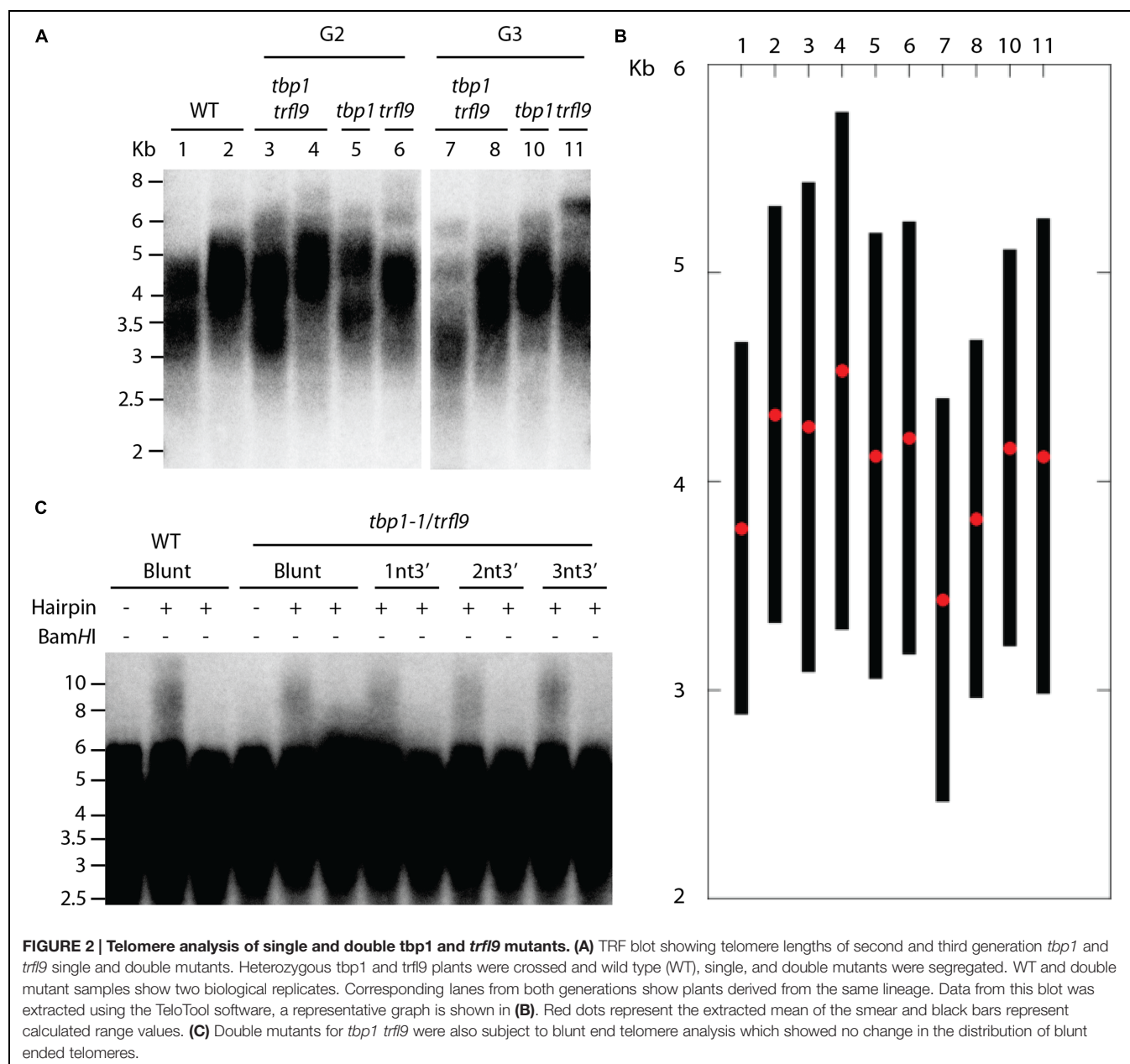## Knockouts of TBP1 and TRFL9 Showed No Changes in Telomere Length and Blunt End Distribution

Phylogenetic analysis indicated that *A. thaliana* Group I TRFL proteins result from relatively recent duplication events in *Brassicaceae* (**Figure 1**). Therefore, some paralogs may still retain overlapping functions. To begin elucidating the role of TRFL proteins at telomeres, we first examined the published allele of *tbp1-1* which has been reported to show telomere

FIGURE 1 | Phylogenetic tree of the Group I TRFLs (Ac, *Aquilegia coerulea*; **At,** *Arabidopsis thaliana*; **Cp,** *Carica papaya*; **Cr,** *Capsella rubella*; **Eg,** *Eucalyptus grandis*; **Os,** *Oryza sativa*; **Pp,** *Physcomitrella patens*; **Sm,** *Selaginella moellendorffii*). Clade *Brassicaceae* is indicated in blue.

elongation (Hwang and Cho, 2007). Within the TRFL family, TBP1 contains a closely related family member, TRFL9, which displays a high level of sequence conservation (**Figure 1**). We reasoned that double knockouts could exacerbate *tbp1-1* telomere phenotypes. Heterozygous plants containing the published *tbp1-1* allele (FLAG_072C05) were crossed to plants heterozygous for the *trfl9* (GK-036D11) mutation. Double heterozygous F1 plants were then selfed and First generation WT, double, and single mutants were segregated. DNA from five pooled plants was extracted from second and third generation of double mutants of the same lineage and subject to TRF analysis (**Figure 2**). To extract data from TRF blots, we used the recently published software TeloTool to measure telomere length and create graphs to better illustrate mean and range of telomeric smears (Gohring et al., 2014). No difference in telomere length was observed in second and third generation *tbp1-1* mutants compared to WT plants segregated from the same cross (**Figures 2A,B**). Double *tbp1 trfl9* mutants also did not appear to shown any great change in telomere length over three generations. Previous studies have shown that telomere lengthening occurs gradually in *tbp1-1* mutants over four generations (Hwang and Cho, 2007). Mutants for telomerase were also shown to show a loss of telomeric DNA of approximately 500bp per generation along with displaying a discrete banding pattern (Riha et al., 2001). However, it would be expected that knocking out core telomere associated proteins would lead to an immediate and severe effect. This has been shown in many studies where severe telomere defects were observed in *Ku70*, *stn1*, *ctc1*, and DNA polymerase α mutants, these are observed within one generation (Riha et al., 2002; Song et al., 2008; Surovtseva et al., 2009; Derboven et al., 2014).

We further examined telomere-end structure as depletion of telomere binding proteins may impair chromosome end protection and integrity of blunt-ended telomeres that are present in plants (Kazda et al., 2012). The current model for chromosome end protection in *Arabidopsis* suggests that telomeres at the leading end are protected from nucleotytic processing by the Ku heterodimer immediately after DNA replication. Because of this, lagging end telomeres in plants are thought to generate classical T-loop structures, whereas leading end telomeres remain blunt-ended and protected by Ku. A hairpin ligation assay was previously developed by Kazda et al. (2012) to detect the presence of blunt ends at *Arabidopsis* telomeres. Briefly, hairpin sequences containing a *Bam*HI site are ligated to blunt-ended telomeres and DNA is digested with *Alu*I to liberate telomeres. Hairpin structures are then subject to alkaline electrophoresis which produces a shift in the higher molecular weight TRF signal. Digestion with *Bam*HI shows that these events are the result of ligation of the hairpin to natural telomeric ends.

Because of the essential role of telomere binding proteins in telomere protection, we reasoned that their inactivation would lead to resection of blunt ended telomeres. However, no observable difference was seen in the presence of blunt ends in *tbp1 trfl9* double mutants using blunt end and short-overhang containing hairpins (**Figure 2C**). These data argue that absence of TBP1 and TRFL9 does not have any discernible effect on telomere structure.

FIGURE 2 | Telomere analysis of single and double tbp1 and *trfl9* mutants. (A) TRF blot showing telomere lengths of second and third generation *tbp1* and *trfl9* single and double mutants. Heterozygous tbp1 and trfl9 plants were crossed and wild type (WT), single, and double mutants were segregated. WT and double mutant samples show two biological replicates. Corresponding lanes from both generations show plants derived from the same lineage. Data from this blot was extracted using the TeloTool software, a representative graph is shown in **(B)**. Red dots represent the extracted mean of the smear and black bars represent calculated range values. **(C)** Double mutants for *tbp1 trfl9* were also subject to blunt end telomere analysis which showed no change in the distribution of blunt ended telomeres.

Multiple Combinations of Quadruple, Quintuple, and Sextuple Mutants Showed No Large Effect on Telomere Length

Because of the sequence similarities between the TRFL proteins, it is possible that other TRFL homologs compensate the functions of TBP1 and TRFL9 in their absence. Therefore, we decided to construct *Arabidopsis* plants with multiple mutant combinations of the genes in the group I TRFL family. Because generation of sextuple mutants would require extensive screening of a large number of plants in segregating populations, we decided to take advantage of centromere induced genome elimination to produce haploid F2 plants (Ravi and Chan, 2010). Frequency of any

quadruple mutant combination among such haploids is 1/16 as opposed to 1/256 in a diploid F2 population.

Centromere induced genome elimination involves generation of haploids by crossing diploid plants as male to the *cenh3*/GFP-*tailswap* haploid inducer. Single T-DNA insertion mutants were selected for each of the six candidate proteins. In addition to *trfl9* and *tbp1* alleles which were already mentioned, *trp1* (SALK_125033), *trfl1* (SALK_052864), *trfl2* (SAIL_73_G01), and *trfl4* (SAIL_73_F07) mutants were also obtained. In order to combine all alleles into the same plant, we first created three combinations of double heterozygous mutants ($trp1^{++/-}$ $trfl1^{+/-}$, $trfl2^{+/-}$ $trfl4^{+/-}$, and $tbp1^{+/-}$ $trfl9^{+/-}$). Next, we generated two combinations of quadruple mutants, and finally quintuple and sextuple mutants as illustrated in the

FIGURE 3 | Diagram illustrating crossing scheme for generating multiple TRFL mutants by using centromere mediated genome elimination. See text for details.
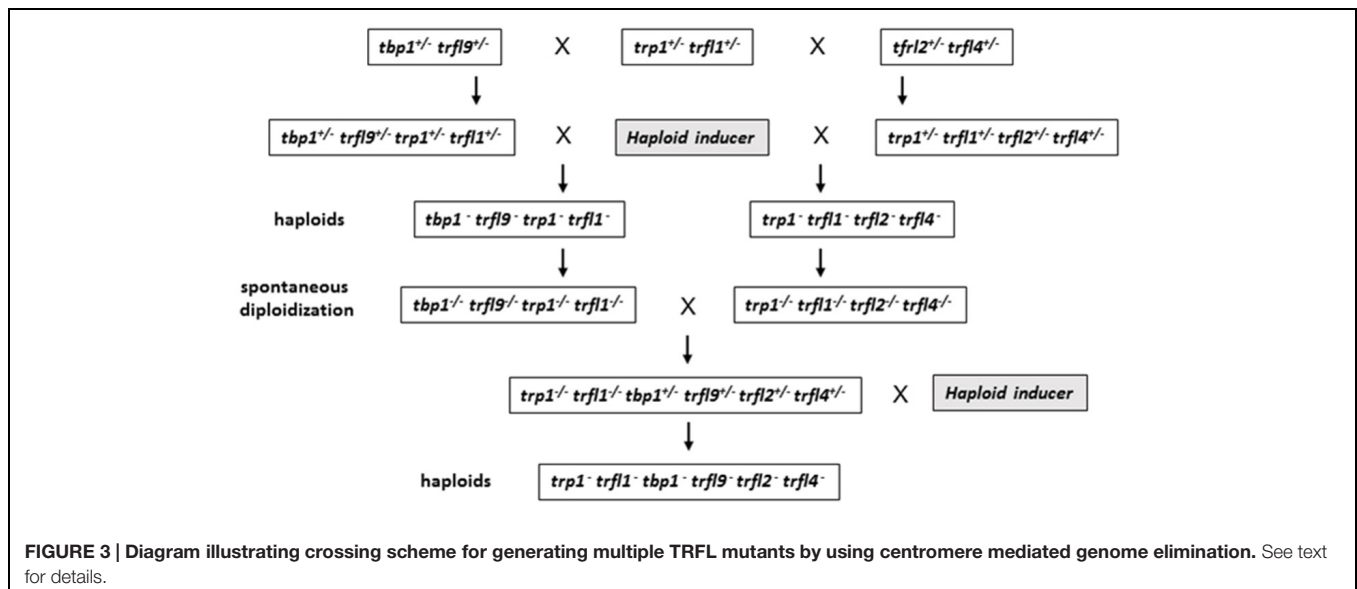
crossing scheme (**Figure 3**). Heterozygous quadruple mutants were then crossed as male to *cenh3*/GFP-*tailswap* plants; homozygous quadruple haploids were obtained along with the WT combination. Diploid seeds were obtained from mutant and WT haploids by spontaneous diploidization (**Figure 3**). First, two single doubled haploid plants were tested by TRF analysis for WT, $tbp1^{-/-} trfl9^{-/-} trp1^{-/-} trfl1^{-/-}$ and $trp1^{-/-} trfl1^{-/-} trfl2^{-/-} trfl4^{-/-}$ combinations (Second generation without functional protein, **Figure 4A**). Seeds were collected from these plants and pooled DNA from 5 plants was used for TRF analysis in the following generation (Third generation, **Figure 4B**). Terminal restriction fragment analysis of resulting $tbp1^{-/-} trfl9^{-/-} trp1^{-/-} trfl1^{-/-}$ and $trp1^{-/-} trfl1^{-/-} trfl2^{-/-} trfl4^{-/-}$ lines showed no effect on telomere length regulation (**Figure 4**).

Next, we created lines with disruptions in the entire gene family. For this, both quadruple homozygous mutant lines were crossed generating F1 plants that were homozygous for *trp1 trfl1* mutations, but segregating for the other four alleles (**Figure 3**). The haploid induction process was repeated by crossing these plants to the *cenh3*/GFP-*tailswap* plants and segregating quintuple and sextuple haploid plants. Individual quintuple and sextuple haploid plants were fully viable and exhibited neither retarded growth in comparison to haploid plants that segregated as WT, nor defects typical for plants with dysfunctional telomeres (Riha et al., 2001; Surovtseva et al., 2009; Derboven et al., 2014). TRF analysis did not reveal drastic changes in telomere length in these mutants (**Figure 5**), although observed variation seen among individual samples suggests that sextuple mutants could display a higher level of telomere length variation compared to WT. The telomere lengths observed here, however, all lie within the natural telomere length limits seen in Col-0 lines and natural variation amongst diverse *Arabidopsis* accessions was also shown to vary between approximately 1 and 9 kb (Fulcher et al., 2015). Normal growth and lack of a clear telomere length deviation in

sextuple mutants demonstrates that the Group II TRFL protein family does not play a major role in telomere maintenance in *A. thaliana*.

Phylogeny of Telobox Containing Proteins in the Plant Kingdom

Our genetic analysis excluded the possibility that the Group I TRFL protein family harbors functional counterparts of human TRF1/2. Thus, the candidate protein(s) may be encoded by one of the other two telobox families. It is expected that that the *bona fide* telomere binding protein will be highly conserved in plants. To look at evolution of telobox protein families, we performed systematic phylogenetic analysis of all telobox containing proteins in sequenced genomes representing different phylogenetic groups within plant kingdom. In this analysis we included *A. thaliana* and *Oryza sativa* as representatives of dicot and monocot angiosperm plants, respectively, *Selaginella moellendorffii* representing the oldest branch in the clade of vascular plants, moss *Physcomitrella patens* and two unicellular green algae, *Coccomyxa subellipsoidea* and *Ostreococcus lucimarinus*. Phylogeny based on whole protein alignments revealed presence of the all three telobox protein families already in the moss *P. patens* and separation of TRFL and Smh/TRB is apparent already in unicellular algae (**Figure 6**). This demonstrates ancient origin of the three telobox protein families and their diversification early in evolution of the plant lineage. Hence, telomere function can be associated with either of the remaining two telobox families.

DISCUSSION

Homologs of TRF1 and TRF2, the double stranded telomere binding proteins central to the shelterin complex, have not been clearly characterized so far in *Arabidopsis*. These proteins form the core part of shelterin and are essential for telomere
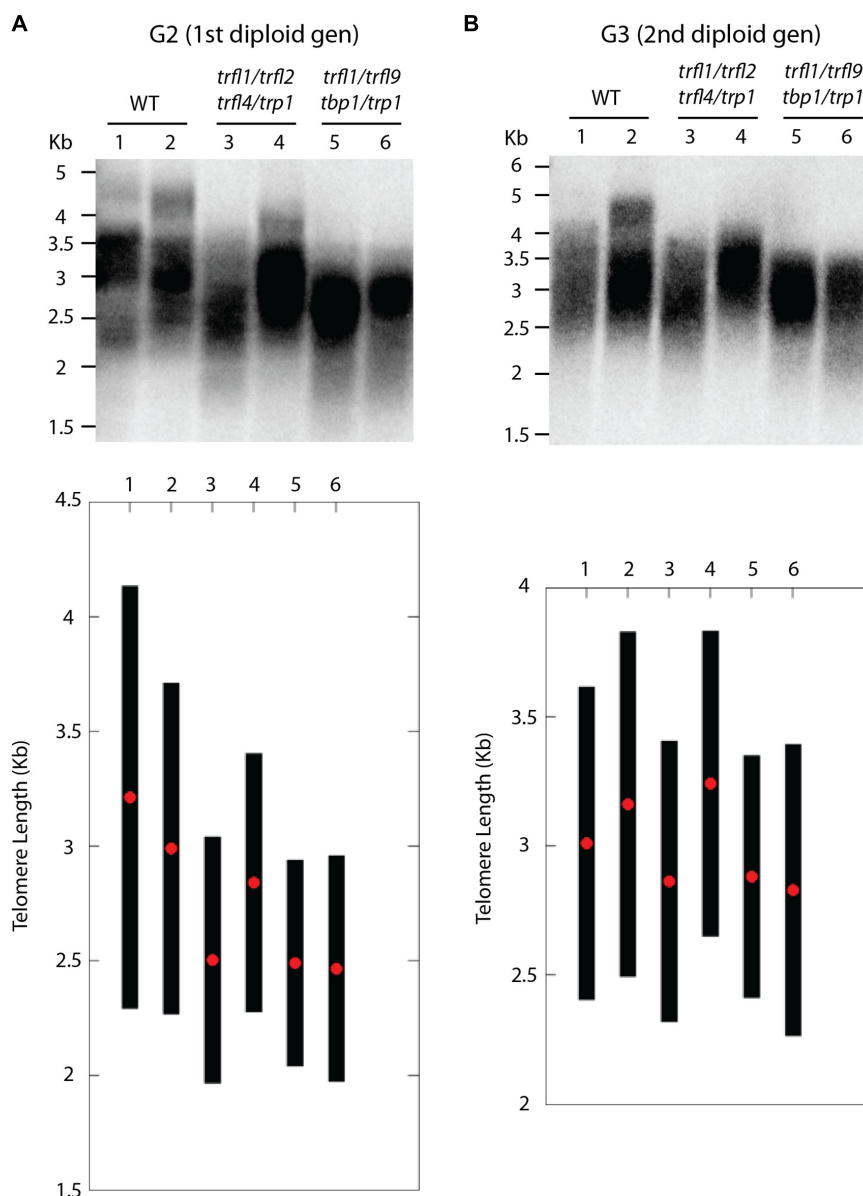
FIGURE 4 | Telomere length analysis of quadruple *trfl1 trfl2 trfl4 trp1* and *trfl1 trfl9 tbp1 trp1* mutants. TRF blots show telomere lengths from second generation **(A)** and third generation **(B)** plants along with WT controls. Blots from the Second generation show TRF profiles of two individual doubled haploid plants. DNA was pooled from 5 plants in the next generation and subject to TRF analysis in the third generation. Corresponding lanes of both blots represent plants of the same lineage. Graphs plotted from TeloTool analysis are shown below respective blots. Red dots represent the extracted mean and black bars show calculated range values.

maintenance and function. Cells expressing dominant negative alleles and conditional knockouts of TRF2 exhibit telomere fusions and telomere length defects demonstrating an essential role of TRF2 in telomere protection (van Steensel et al., 1998; Celli and de Lange, 2005). Functional studies of TRF1 indicate a role of the protein in telomere replication and length regulation (van Steensel and de Lange, 1997; Sfeir et al., 2009) TRFL proteins described in *Arabidopsis* highlighted a group of potential candidates containing C-terminal telobox motif and plant specific extension domain (Karamysheva et al., 2004). These

proteins also bind to telomeric DNA *in vitro* and the telobox domain is important for this interaction. In addition, studies have shown that disruption of similar proteins in rice, tobacco and tomato leads to telomeric and developmental phenotypes. Transformation of tobacco BY2 cells with 35S:*LeTBP1* from tomato was reported to result in telomere shortening from 15–55 kbps to 15–35 kbps (Moriguchi et al., 2006). In a later study, knockdowns of LeTBP1 in tomato showed defects in fruit development and genomic instability, no changes in telomere length were observed in these plants (Moriguchi et al.,

FIGURE 5 | Telomere length analysis of quintuple and sextuple mutants. TRF blot **(upper panel)** shows telomeric smears from WT, double *trfl1 trp1*, quintuple (noted gene represents WT locus) and complete sextuple knockouts. DNA was extracted from single segregated haploid plants. TeloTool analysis is shown **(lower panel)** where red dots show extrapolated mean and black bars shown range values.

FIGURE 6 | Phylogenetic tree of plant telobox protiens. (At, *Arabidopsis thaliana*; Cc, *Coccomyxa subellipsoidea*; Ol, *Ostreococcus lucimarinus*; Os, *Oryza sativa*; Pp, *Physcomitrella patens*; Sm, *Selaginella moellendorffii*).

2011). It could be, however, that in these studies, the TRF assay is not sensitive enough to detect small changes that occur in the already long telomeres of tobacco and tomato. Characterization of RICE TELOMERE BINDING PROTEIN1 (RTBP1) showed telomere elongation in first generation RTBP1 knockouts along with anaphase bridges, growth retardation, and floral defects in later generations (Hong et al., 2007). A similar result was reported in *Arabidopsis* showing knockouts of AtTBP1 undergoing telomere elongation over four generations (Hwang and Cho, 2007). However, the presence of *tbp1-1* in the Ws background complicates telomere length analysis as this accession has previously shown to display a bimodal telomere length distribution in WT plants (Shakirov and Shippen, 2004). Because of these previously reported phenotypes of these candidate telomere binding proteins in *Arabidopsis* and other plant species, *in vitro* telomeric duplex binding activity, and the high level of sequence conservation, it was expected that the Group I TRFL family comprises the canonical duplex telomere binding proteins.

However, in this study we show that knockouts of all six members of the family in *Arabidopsis* do not exhibit any obvious changes in telomere length or functionality. Thus, it
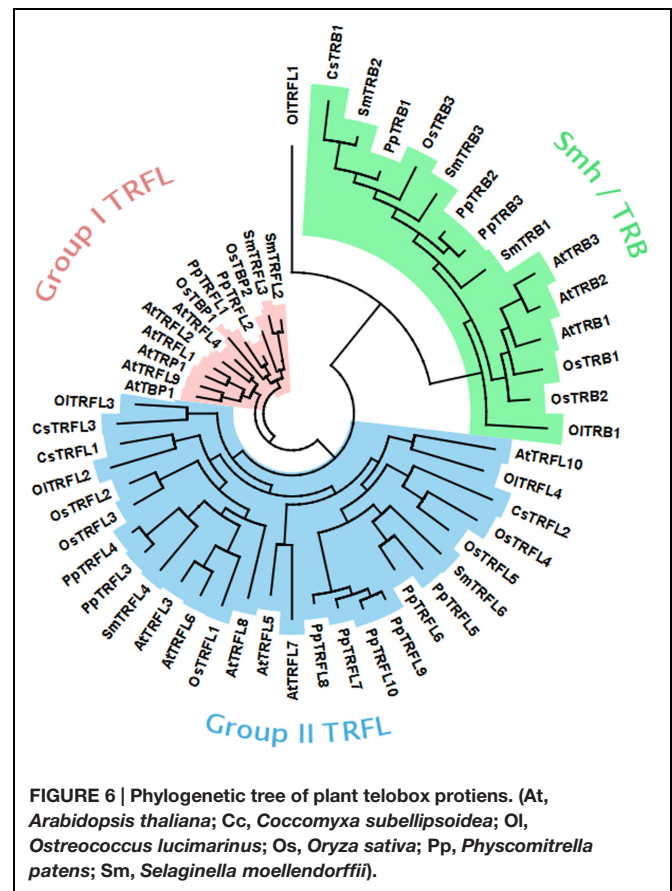
can be concluded that, at least in *Arabidopsis*, Group I TRFL family does not play a major role in telomere biology. The previously reported *in vitro* telomere binding of this group suggests there is association with telomeric DNA, although an effect on function has not been observed. Although studies in tobacco, rice, and tomato reported telomere phenotypes associated with knock-outs or overexpressing Group I TRFL proteins (Yang et al., 2004; Moriguchi et al., 2006; Hong et al., 2007), these effects are relatively mild and may reflect only an auxiliary function of these proteins at telomeres. Instead, these proteins may act as transcription factors as promoters of a number of genes are known to contain a short stretch of telomeric sequences (Tremousaygue et al., 1999). Hence, other proteins likely form the core structure of telomeric chromatin in plants.

The question remains as to what proteins comprise the telomere capping complex in *Arabidopsis*. The Smh/TRB proteins may be the next prime suspects. Phylogenetic analysis shows that these proteins are present in all plant taxonomic units including unicellular green algae suggesting that they may be associated with a fundamental biological function. Three Smh/TRB genes with an N-terminal telobox domain have been found in *Arabidopsis* and have shown to exhibit *in vitro* binding to telomeric DNA (Schrumpfova et al., 2004; Mozgova et al., 2008; Hofr et al., 2009). Recently, *Arabidopsis* TRB1 was found to bind to telomeric sequences *in vivo* through

immunolocalization studies in tobacco cells (Schrumpfova et al., 2014). One caveat with this approach is that telomeres in tobacco reach far greater lengths than with *Arabidopsis* (∼5 and 150 kb respectively). Association with telomeric DNA may, therefore, not be necessarily for telomere specific functions and can similarly colocalize with non-telomeric sequences. Chromatin Immunoprecipitation (ChIP) studies performed within the same paper, however, confirm binding to telomeric sequences in *Arabidopsis*. With this evident telomere binding capacity and interaction with Pot1b and the N-Terminus of TERT, SMH proteins also show promise as telomere binding components of *Arabidopsis* telomeres (Kuchar and Fajkus, 2004; Schrumpfova et al., 2014). Telomere length defects are also described for *trb1* mutants although the effect is relatively small after five generations of selfing (Schrumpfova et al., 2014). This could mean redundancy amongst the SMH family of proteins. Additionally, it is possible that members of the tested group 1 TRFL proteins are redundant with SMH/TRB proteins. Functional analysis of other members of this family should clarify the role of these proteins in telomere maintenance.

AUTHOR CONTRIBUTIONS

NF designed and performed the experiments and wrote the paper. KR designed the experiments, performed phylogenetic analysis and wrote the paper.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fgene. 2015.00349

REFERENCES

Bilaud, T., Koering, C. E., Binet-Brasselet, E., Ancelin, K., Pollice, A., Gasser, S. M., et al. (1996). The telobox, a Myb-related telomeric DNA binding motif found in proteins from yeast, plants and human. *Nucleic Acids Res.* 24, 1294–1303. doi: 10.1093/nar/24.7.1294

Celli, G. B., and de Lange, T. (2005). DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. *Nat. Cell Biol.* 7, 712–718. doi: 10.1038/ncb1275

de Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* 19, 2100–2110. doi: 10.1101/gad.1346005

Derboven, E., Ekker, H., Kusenda, B., Bulankova, P., and Riha, K. (2014). Role of STN1 and DNA polymerase alpha in telomere stability and genome-wide replication in *Arabidopsis*. *PLoS Genet.* 10:e1004682. doi: 10.1371/journal.pgen.1004682

Fulcher, N., Teubenbacher, A., Kerdaffrec, E., Farlow, A., Nordborg, M., and Riha, K. (2015). Genetic architecture of natural variation of telomere length in *Arabidopsis thaliana*. *Genetics* 199, 625–635. doi: 10.1534/genetics.114.172163

Gohring, J., Fulcher, N., Jacak, J., and Riha, K. (2014). TeloTool: a new tool for telomere length measurement from terminal restriction fragment analysis with improved probe intensity correction. *Nucleic Acids Res.* 42, e21. doi: 10.1093/nar/gkt1315

Hofr, C., Sultesova, P., Zimmermann, M., Mozgova, I., Prochazkova Schrumpfova, P., Wimmerova, M., et al. (2009). Single-Myb-histone proteins from *Arabidopsis thaliana*: a quantitative study of telomere-binding specificity and kinetics. *Biochem. J.* 419, 221–228. doi: 10.1042/BJ20082195

Hong, J. P., Byun, M. Y., Koo, D. H., An, K., Bang, J. W., Chung, I. K., et al. (2007). Suppression of RICE TELOMERE BINDING PROTEIN 1 results in severe and gradual developmental defects accompanied by genome instability in rice. *Plant Cell* 19, 1770–1781. doi: 10.1105/tpc.107.051953

Hwang, M. G., and Cho, M. H. (2007). *Arabidopsis thaliana* telomeric DNA-binding protein 1 is required for telomere length homeostasis and its Myb-extension domain stabilizes plant telomeric DNA binding. *Nucleic Acids Res.* 35, 1333–1342. doi: 10.1093/nar/gkm043

Karamysheva, Z. N., Surovtseva, Y. V., Vespa, L., Shakirov, E. V., and Shippen, D. E. (2004). A C-terminal Myb extension domain defines a novel family of double-strand telomeric DNA-binding proteins in *Arabidopsis*. *J. Biol. Chem.* 279, 47799–47807. doi: 10.1074/jbc.M407938200

Kazda, A., Zellinger, B., Rossler, M., Derboven, E., Kusenda, B., and Riha, K. (2012). Chromosome end protection by blunt-ended telomeres. *Genes Dev.* 26, 1703–1713. doi: 10.1101/gad.194944.112

Ko, S., Jun, S. H., Bae, H., Byun, J. S., Han, W., Park, H., et al. (2008). Structure of the DNA-binding domain of NgTRF1 reveals unique features of plant telomere-binding proteins. *Nucleic Acids Res.* 36, 2739–2755. doi: 10.1093/nar/gkn030

Ko, S., Yu, E. Y., Shin, J., Yoo, H. H., Tanaka, T., Kim, W. T., et al. (2009). Solution structure of the DNA binding domain of rice telomere binding protein RTBP1. *Biochemistry* 48, 827–838. doi: 10.1021/bi801270g

Kuchar, M., and Fajkus, J. (2004). Interactions of putative telomere-binding proteins in *Arabidopsis thaliana*: identification of functional TRF2 homolog in plants. *FEBS Lett.* 578, 311–315. doi: 10.1016/j.febslet.2004.11.021

Marian, C. O., Bordoli, S. J., Goltz, M., Santarella, R. A., Jackson, L. P., Danilevskaya, O., et al. (2003). The maize Single myb histone 1 gene, Smh1, belongs to a novel gene family and encodes a protein that binds telomere DNA repeats in vitro. *Plant Physiol.* 133, 1336–1350. doi: 10.1104/pp.103.026856

Moriguchi, R., Kanahama, K., and Kanayama, Y. (2006). Characterization and expression analysis of the tomato telomere-binding protein LeTBP1. *Plant Sci.* 171, 166–174. doi: 10.1016/j.plantsci.2006.03.010

Moriguchi, R., Ohata, K., Kanahama, K., Takahashi, H., Nishiyama, M., and Kanayama, Y. (2011). Suppression of telomere-binding protein gene expression represses seed and fruit development in tomato. *J. Plant Physiol.* 168, 1927–1933. doi: 10.1016/j.jplph.2011.05.009

Mozgova, I., Schrumpfova, P. P., Hofr, C., and Fajkus, J. (2008). Functional characterization of domains in AtTRB1, a putative telomere-binding protein in *Arabidopsis thaliana*. *Phytochemistry* 69, 1814–1819. doi: 10.1016/j.phytochem.2008.04.001

Ravi, M., and Chan, S. W. (2010). Haploid plants produced by centromere-mediated genome elimination. *Nature* 464, 615–618. doi: 10.1038/nature08842

Ravi, M., Marimuthu, M. P., Tan, E. H., Maheshwari, S., Henry, I. M., Marin-Rodriguez, B., et al. (2014). A haploid genetics toolbox for *Arabidopsis thaliana*. *Nat. Commun.* 5, 5334. doi: 10.1038/ncomms6334

Riha, K., Mcknight, T. D., Griffing, L. R., and Shippen, D. E. (2001). Living with genome instability: plant responses to telomere dysfunction. *Science* 291, 1797–1800. doi: 10.1126/science.1057110

Riha, K., Watson, J. M., Parkey, J., and Shippen, D. E. (2002). Telomere length deregulation and enhanced sensitivity to genotoxic stress in *Arabidopsis* mutants deficient in Ku70. *EMBO J.* 21, 2819–2826. doi: 10.1093/emboj/21.11.2819

Schrumpfova, P., Kuchar, M., Mikova, G., Skrisovska, L., Kubicarova, T., and Fajkus, J. (2004). Characterization of two *Arabidopsis thaliana* myb-like proteins showing affinity to telomeric DNA sequence. *Genome* 47, 316–324. doi: 10.1139/g03-136

Schrumpfova, P. P., Vychodilova, I., Dvorackova, M., Majerska, J., Dokladal, L., Schorova, S., et al. (2014). Telomere repeat binding proteins are functional

components of *Arabidopsis* telomeres and interact with telomerase. *Plant J.* 77, 770–781. doi: 10.1111/tpj.12428

Seymour, D. K., Filiault, D. L., Henry, I. M., Monson-Miller, J., Ravi, M., Pang, A., et al. (2012). Rapid creation of *Arabidopsis* doubled haploid lines for quantitative trait locus mapping. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4227–4232. doi: 10.1073/pnas.1117277109

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

Shakirov, E. V., and Shippen, D. E. (2004). Length regulation and dynamics of individual telomere tracts in wild-type *Arabidopsis*. *Plant Cell* 16, 1959–1967. doi: 10.1105/tpc.104.023093

Song, X., Leehy, K., Warrington, R. T., Lamb, J. C., Surovtseva, Y. V., and Shippen, D. E. (2008). STN1 protects chromosome ends in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 19815–19820. doi: 10.1073/pnas.0807867105

Surovtseva, Y. V., Churikov, D., Boltz, K. A., Song, X., Lamb, J. C., Warrington, R., et al. (2009). Conserved telomere maintenance component 1 interacts with STN1 and maintains chromosome ends in higher eukaryotes. *Mol. Cell* 36, 207–218. doi: 10.1016/j.molcel.2009.09.017

Tremousaygue, D., Manevski, A., Bardet, C., Lescure, N., and Lescure, B. (1999). Plant interstitial telomere motifs participate in the control of gene expression in root meristems. *Plant J.* 20, 553–561. doi: 10.1046/j.1365-313X.1999.00627.x

van Steensel, B., and de Lange, T. (1997). Control of telomere length by the human telomeric protein TRF1. *Nature* 385, 740–743. doi: 10.1038/385740a0

van Steensel, B., Smogorzewska, A., and De Lange, T. (1998). TRF2 protects human telomeres from end-to-end fusions. *Cell* 92, 401–413. doi: 10.1016/S0092-8674(00)80932-0

Wang, R. C., Smogorzewska, A., and De Lange, T. (2004). Homologous recombination generates T-loop-sized deletions at human telomeres. *Cell* 119, 355–368. doi: 10.1016/j.cell.2004.10.011

Watson, J. M., and Riha, K. (2010). Comparative biology of telomeres: where plants stand. *FEBS Lett.* 584, 3752–3759. doi: 10.1016/j.febslet.2010.06.017

Wijnker, E., Van Dun, K., De Snoo, C. B., Lelivelt, C. L., Keurentjes, J. J., Naharudin, N. S., et al. (2012). Reverse breeding in *Arabidopsis thaliana* generates homozygous parental lines from a heterozygous plant. *Nat. Genet.* 44, 467–470. doi: 10.1038/ng.2203

Yang, S. W., Kim, S. K., and Kim, W. T. (2004). Perturbation of NgTRF1 expression induces apoptosis-like cell death in tobacco BY-2 cells and implicates NgTRF1 in the control of telomere length and stability. *Plant Cell* 16, 3370–3385. doi: 10.1105/tpc.104.026278

Zellinger, B., and Riha, K. (2007). Composition of plant telomeres. *Biochim. Biophys. Acta* 1769, 399–409. doi: 10.1016/j.bbaexp.2007.02.001

frontiers
in Genetics

Hypothesis: Paralog Formation from Progenitor Proteins and Paralog Mutagenesis Spur the Rapid Evolution of Telomere Binding Proteins

Arthur J. Lustig*

Department of Biochemistry and Molecular Biology, Tulane University, New Orleans, LA, USA

Through elegant studies in fungal cells and complex organisms, we propose a unifying paradigm for the rapid evolution of telomere binding proteins (TBPs) that associate with either (or both) telomeric DNA and telomeric proteins. TBPs protect and regulate telomere structure and function. Four critical factors are involved. First, TBPs that commonly bind to telomeric DNA include the c-Myb binding proteins, OB-fold single-stranded binding proteins, and G-G base paired Hoogsteen structure (G4) binding proteins. Each contributes independently or, in some cases, cooperatively, to provide a minimum level of telomere function. As a result of these minimal requirements and the great abundance of homologs of these motifs in the proteome, DNA telomere-binding activity may be generated more easily than expected. Second, telomere dysfunction gives rise to genome instability, through the elevation of recombination rates, genome ploidy, and the frequency of gene mutations. The formation of paralogs that diverge from their progenitor proteins ultimately can form a high frequency of altered TBPs with altered functions. Third, TBPs that assemble into complexes (e.g., mammalian shelterin) derive benefits from the novel emergent functions. Fourth, a limiting factor in the evolution of TBP complexes is the formation of mutually compatible interaction surfaces amongst the TBPs. These factors may have different degrees of importance in the evolution of different phyla, illustrated by the apparently simpler telomeres in complex plants. Selective pressures that can utilize the mechanisms of paralog formation and mutagenesis to drive TBP evolution along routes dependent on the requisite physiologic changes.

Keywords: telomeres, evolution, non-LTR reverse transcription, telomerase, models, stress response

INTRODUCTION

Telomeres, the DNA-RNP structures present at the termini of all eukaryotic chromosomes, are essential for genome stability and function. The telomere serves two functions that are fundamental for viability. The first is to provide a solution to the end-replication problem. This problem refers to the inability of the lagging strand DNA of semi-conservative replication to maintain its terminal RNA primer at the $5'$ end of any replicating linear molecule (Levy et al., 1992). The leading

stand, in contrast, creates a blunt ended telomere at the other terminus. The lagging strand will form one 3′ overhang terminus (Kazda et al., 2012; Bonetti et al., 2013; Ghodke and Muniyappa, 2013). Continuing rounds of semi-conservative replication will result in the loss of DNA primers, leading to attrition and chromosome loss. The processing of the blunt-ended telomere is variable in different organisms (Kazda et al., 2012; Bonetti et al., 2013). Regardless, the loss of a terminal DNA primer predicts the inevitable attrition of terminal sequence, and, ultimately, cellular inviability.

The solution to this problem is based on the terminal 3′ overhang that serves as a substrate for recombination or telomerase. Telomerase is the RNP-reverse transcriptase that adds G + T-rich simple sequence onto the 3′ terminus using the RNA as a template. The core enzyme rate and processivity are regulated by a multiplicity of holoenzyme components and telomere binding proteins (TBPs; Tucey and Lundblad, 2014; Vogan and Collins, 2015). Telomerase can catalyze addition in a processive or a distributive mechanism. The repeats added are most often identical, but, in some organisms (e.g., fungal systems) can add inexact repeats. The irregular repeat is thought to be formed by misalignment of DNA on the RNA template (Petrov et al., 1998; Forstemann and Lingner, 2001). As an example of holoenzyme regulation, the budding yeast Cdc13 protein associates with and recruits the auxiliary protein, Est1. Est1, in turn, recruits the telomere reverse transcriptase (TERT, Est2 in yeast) and the complex with the RNA subunit (TR) finally recruits the Est3 subunit (Tucey and Lundblad, 2014).

The second function of a telomere is to overcome the end-protection problem (de Lange, 2009). That is, the telomere must not be accessible to non-specific enzymes, including nucleases, ligases, and recombinases that may lose, destabilize, or rearrange the telomere, respectively. In this sense, the telomere is a cap against activities that lead to genomic dysfunction, while allowing the access of positive and negative regulators of telomere addition.

One protective function is the feedback regulation of telomere size that is present in all organisms, although the mechanisms may vary (Evans and Lundblad, 2000). In yeast, a competition between negative and positive regulators of telomerase form a steady state using the ATM pathway (Lustig and Petes, 1986; Bianchi and Shore, 2007; Sabourin et al., 2007; Hirano et al., 2009; Martina et al., 2012; Ribeyre and Shore, 2012). ATM (Tel1) normally arrests cells, in response to double strand break, in the G2 phase of the cell cycle, until repair of DSBs is complete (Usui et al., 2001). However, the telomeric DSB is protected from both repair and genomic instability in part by this equilibrium creating an "anti-checkpoint," a part of the telomeric cap function (Carneiro et al., 2010).

In duplex DNA, the telomeric protein Rap1 forms the basic telomeric chromatin in yeast (Wright et al., 1992). Some of the major TBPs (e.g., in yeast Rap1 and the yKu70/80 heterodimer) protect the telomere from non-homologous end joining and inhibit end fusion (Frank-Vaillant and Marcand, 2002; Pardo and Marcand, 2005). Another cap structure, the Cst1/Stn1/Ten1 (CST) complex, also serves as a physical cap. Telomerase also appears to have the ability to block the end of the telomere

(Blackburn et al., 2000). Finally, in ciliates, Hoogsteen base-paired G4 structures, such as the G-quartet, associate with TBPs and telomerase to both act as a cap and as a regulator of telomerase addition in vitro and in vivo (Fang and Cech, 1993a; Oganesian et al., 2006). Taken together, the activities of homeostatic factors, telomerase, capping proteins, and G4 DNA TBPs control telomere size in context of the cell cycle.

The ATR pathway, however, is another part of the telomeric DNA checkpoint control. If telomerase does not add a compensatory amount of G + T repeats, cells will begin to senesce (Abdallah et al., 2009). If the telomere shortens beyond a threshold size, the cells will undergo a G2 arrest and a further loss of telomere sequences mediated by both recombinational and replicative DNA damage, leading to inviability. Ultimately, survivors use either a break-induced recombination or a rapid telomere elongation process to form elongated telomeres (Lustig, 2003; Pickett et al., 2011; Pickett and Reddel, 2012). The mechanistic details may differ along the evolutionary spectrum of organisms, but the basic paradigm remains unchanged. In this theoretical perspective, we will focus on the TBPs that associate with telomerase generated telomeres.

THE DIVERSITY OF TELOMERE BINDING PROTEINS

Evolutionary biologists and telomere researchers have long tried to explain the wide diversity of many proteins involved in telomere function and structure (Linger and Price, 2009). Models for the evolution of different modes of telomere maintenance are beginning to show promise. The major modes of telomere addition are telomerase and non-LTR reverse transcriptases. Telomerase may have formed from non-LTR reverse transcriptases with a specificity high in G + T content (Garavis et al., 2013). In contrast, reverse transcriptase possibly continued to be used when target site sequence bias is absent. These may well be the primary ancestral mechanisms of telomere formation, although the ancestral origin is, by definition, a matter of speculation. Evolution may at times repeat previously used mechanisms. For example, Drosophila arose long after primordial telomeres, yet uses telomeric non-LTR retrotransposons that are (telomere specific, Villasante et al., 2008). The mechanism used in Drosophila may lend insights in an evolutionary context, with some caution that Drosophila may use a variation on a theme.

Most non-LTR retrotransposons appear to have formed degenerate heterochromatin that was subsequently maintained by recombinational mechanisms (Villasante et al., 2007). Recombinational activity is used in extant organisms as an alternative telomere pathway in the absence of telomerase (Louis and Haber, 1990; Preiszner et al., 1994; Mizuno et al., 2008; Li et al., 2009; Torres et al., 2011). Investigators have observed rolling circle replication, unequal sister chromatid exchange, and mechanisms of simple sequence elongation (Tomaska et al., 2004a, 2009; Torres et al., 2011). We cannot exclude these uncharacterized mechanisms in ancestral telomere formation.

The mechanisms of telomere elongation are presented to provide context. Our focus, however, will be on the exploration of the curious rapid evolution of the TBPs in the telomerase-based systems. These data are not consistent with either a simple movement toward complexity or simplicity during evolution (Gould, 1996; de Lange, 2015). The complexity of the plant genome and its sophistication in development do not explain the simplicity of its telomere with little difference between complex plants and algae. We feel that rapid TBP evolution can be explained by a set of basic principles that governs diversity.

A Model for the Conservation and Diversity of TBF

Orthologs and Parologs

The major molecular biological means of describing closely related protein sequences is homology. However, the evolutionary significance of homology can be misinterpreted without a comparison among organisms of differing complexity. The significance of partial homology is difficult to interpret when applied to evolution. A protein having partial homology throughout all kingdoms and phyla tells us little about the directionality of inheritance during evolution. Homology and partial homology are anathema to many evolutionary biologists, providing information only about sequence identity, rather than evolutionary patterns.

The initial insights into evolutionary patterns were remarkable, having arisen independently of any knowledge of DNA. These theoretical and mathematical principles were based on abstract evolutionary concepts. The strongest hypotheses have weathered time to the genomic era. The field is finally in a position to test specific questions regarding the blueprints for telomere evolution at a molecular level.

Some specific terms that were last seen by most of us in a textbook require review. Two types of evolutionary relationships, orthologs and paralogs, are central to the outline of much of evolutionary change. The inheritance patterns and relative homology of proteins argue for a vertical process (as in an evolutionary tree) in evolution. In this way, a single ancestral progenitor can be envisioned by the orthologs among different organisms (Koonin, 2005).

Paralogs, on the other hand, are protein products of DNA or genomic duplication that lead to horizontal evolution; particularly two duplicate proteins, one of which evolves from the progenitor in a unique direction under strong selective conditions (**Figure 1**). Sometimes, both paralogs evolve into new products. Ultimately, sequence and evolutionary analysis are required to provide more evidence for the existence of a paralog. This paralog can subsequently become an ortholog of a long line of species. Examples of telomeric paralogs are shown in **Table 1**. We propose that telomere dysfunction creates a variety of stress responses and selection pressures that use elevated paralog formation and mutagenesis that lead to an exceedingly high rate of TBP evolution.
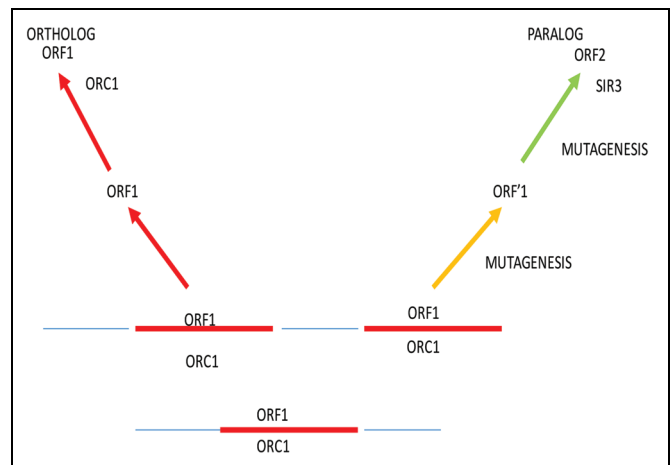
FIGURE 1 | **Paralog formation and mutagenesis of a single ORF1.** Under conditions of stress response and high selectivity, recombination and mutagenesis increase the frequency of paralog formation during evolution. In this process, recombination results in duplication of the ORF1 coding sequence. The first copy, when separated by recombination, remains stable as ORF1. Under selection, the paralog undergoes an elevated level of mutagenesis caused by stress in response to dysfunctional telomeres. Unknown multiple rounds of mutagenesis take place in evolutionary time to ultimately give rise to a unique functional protein, ORF2. In the ORC1 example, paralog formation gives rise to Sir3, a protein involved in silencing of genes and the structure of telomeres in several yeast strains.

The Conserved Elements of TBP

There is great diversity among proteins that bind to telomeric DNA and that associate with other telomeric proteins or G4 structures. However, there is a subclass of proteins and DNA structures that are present in most organisms and serve a conserved function. Since these are important in any analysis, we will first discuss the highly conserved telomere capping proteins and DNA structures.

The Conserved MR (X/N) Complex

The primary roles of MRX in the signaling and processing of DSBs are a major part of the highly conserved ATM checkpoint pathway (Foster et al., 2006; Dimitrova and de Lange, 2009; Amiard et al., 2011). However, the telomeres of extant organisms use ATM-MRX/N (The yeast Xrs2 is replaced with NBS in all other organisms). The genetic characterization of telomere homeostasis in *Saccharomyces cerevisiae* led to the discovery of ATM-mediated anti-checkpoints. Similar schemes are likely to be present in most organisms, including *Drosophila* (Ciapponi et al., 2004; Gao et al., 2009).

In yeast, the ATM ortholog, Tel1 (Lustig and Petes, 1986; Greenwell et al., 1995), coupled with MRX, associate exclusively to short telomeres (Chang et al., 2007; Sabourin et al., 2007). These associations lead to telomerase activation. The counteracting inhibitory activities, Rif1 and Rif2, are recruited to longer telomeres. Rif1 acts to displace the Tel1 molecule, while Rif2 inhibits Tel1 binding to telomeric DNA (Martina et al., 2012). This feedback cycle continues whenever telomeres fall into a range that is sensed by an unknown mechanism to be too short or too long, creating a telomere size homeostasis.

TABLE 1 | Examples of likely Tbp paralogs.

	Paralog1	Paralog 2	Paralog 3	Paralog 4	
S. cerevisiae	ORC1	SIR3			Liaw and Lustig, 2006
	EST1	EBS1			Zhou et al., 2000; Luke et al., 2007
	SIR2	HST1	HST2	HST3	SGD
	STN1*	RPA2			Sun et al., 2009
	TEN1*	RPA3*			Sun et al., 2009
Arabidopsis	TRFL1	TRFL2	TRFL4	TRF4	Fulcher and Riha, 2015
		TRB1	TRP1	TRF9	
	PARP1	PARP2	PARP3		Boltz et al., 2014
	POT1a	POT1b			Cifuentes-Rojas et al., 2010
Rodentia	POT1a	POT1b			Hockemeyer et al., 2006
	PARP1	PARP2			Cook et al., 2002
	TRF1*	TRF2*			
Humans	EST1a	EST1b	EST1c		
	TRF1*	TRF2*			Broccoli et al., 1997
	STN1*	RPA2*			Bryan et al., 2013
	TEN1*	RPA3*			Bryan et al., 2013

Paralog was formed in an ancestor and subsequently maintained. Others appear to have formed in a particular species. Rodentia refers to mice and rat species. This is not a complete list and database homologies indicate that there are likely to be more TBF paralogs.

Such an equilibrium between mechanisms of telomere attrition and deletion and mechanisms of telomere elongation is present in both normal and oncogenic cells (Lustig, 2003; Pickett et al., 2011; Pickett and Reddel, 2012). Details of this model are far more complex (Sreesankar et al., 2012). For example, Rif1 and Tel1 operate by altering the timing of replication (Peace et al., 2014; Sridhar et al., 2014) and, very likely, TBP binding is regulated temporally within the context of the cell cycle.

The NHEJ Protein Ku Complex Obstructs the Formation of Telomere Fusions

The third conserved feature of telomeres is terminal capping. One of these complexes is the Ku70/Ku80 heterodimer (Polotnianka et al., 1998; Baumann and Cech, 2000). Ku, paradoxically, plays a vital role in non-homologous recombination of blunt-ended DNA damage. However, Ku can also act as an inhibitor of ligation at telomeres. Indeed, Ku70/Ku80 acts to prevent the deleterious ligation of two telomeres. Inhibiting the formation of dicentric chromosomes (Polotnianka et al., 1998; Williams and Lustig, 2003). Dicentric chromosomes undergo a series of breakage-fusion-breakage cycles, as observed in maize (McClintock, 1942). While higher plants tolerate this damage during mitosis, very few other organisms are resistant to this process. Dicentric chromosomes in most organisms fail in meiosis.

CST, the Telomeric RPA Complex?

The terminal CST capping complex mimics the structure of Replication Factor A (RPA). However, their activities are functionally distinct (Wellinger, 2009). CST, as RPA, acts at multiple genomic sites (Miyake et al., 2009). However, rather than acting as a telomeric cap, RPA stabilizes single-stranded DNA at the telomere and elsewhere (Price et al., 2010; Chen et al., 2012; Wang et al., 2012). Both RPA and CST form complex trimeric structures but only contain small patches of sequence homology. However, crystal structure analyses have shown that the RPA2 and STN1 subunits of RPA and CST, respectively, have very similar structures, as do RPA3 and TEN1 (Sun et al., 2009). The maintenance of protein structure is also responsible for interaction in the absence of extensive homology. Given the prevalence of both CST and RPA in all eukaryotes, ancestral RPA subunits may have formed paralogs that subsequently diverged in primary sequence, while maintaining the structure of the RPA and CST subunits. In reality, this is probably often the case, but is usually reflected in the primary sequence. Hence, these "structural" paralogs can be missed in the absence of extensive sequence homology.

Telomeric Repeat-Containing RNA (TERRA) and T-Loops: Conserved Nucleic Acids

Several nucleic acids play important structural roles at many telomeres. First, t-loop structures, the result of intrachromatid invasion of the telomeric terminus into more proximal sequences, remain stable and may hide the single strand from telomere addition. It may also act as either a structural block or part of the telomere replication process (de Lange, 2002; Luke-Glaser et al., 2012). Second, in most organisms, unique telomeric repeat-containing RNA (TERRA) transcripts are initiated within a subtelomeric element and proceeds in a 5′ to 3′ direction toward the terminus. Very little is known about the function of TERRA at the telomere. (Maicher et al., 2014). However, in exciting new research, G4 DNA acts synergistically with TERRA to form complex structures, some of which could extend or shorten the telomere (Xu, 2012). TERRA also appears to regulate the very short and elongated telomeres of the alternative pathway of telomere addition (Arora and Azzalin, 2015). TERRA may protect the telomere and regulate telomerase addition, as well as participate in non-telomeric functions.

The Conservation of G4 DNA *In Vivo*

G4 DNA consists of non-canonical Hoogsteen base paired structures present in the high G + T content of the telomere. The formation of these structures has been postulated to be a conserved element in the evolution of telomeres. The evidence for the presence of G4 DNA is its ability DNA to bind unique ligands and clear histones from promoter regions.

G4 DNA can form at both regular and irregular repeated telomere sequences (such as yeast) *in vitro*. There is strong evidence for the function of G4 DNA at the telomere *in vivo*. In general, G4 DNA has a protective function, albeit redundant with other overlapping functions. G4 DNA also has a high binding affinity for Mre11. For example, in the absence of the normal capping mechanisms, G4 DNA can block exonuclease function (Smith et al., 2011). Both findings are consistent with the view that G4 DNA served as an initial cap early in evolution (Garavis et al., 2013). In some contexts, G4 structures alone can have a deleterious effect. For example, in yeast, the coating of the single-strand overhang with RFA prevents the interference of G4 structures with lagging strand semi-conservative DNA synthesis (Audry et al., 2015). Cdc13 has also been implicated as a G4 TBP, given the simultaneous loss of a G4 DNA cap function only in *cdc13-1* cells (Smith et al., 2011).

Both positive and negative G4 functions at the telomere have been substantiated in the context of a vast number of other studies. Studies in the ciliate Oxytricha provide the best evidence for a positive function of G4 DNA *in vivo*. Under a complex set of interactions between the major two telomere proteins, TEBF alpha and TEBP beta, TEBP beta coupled with G4 DNA structures can facilitate telomere elongation (Oganesian et al., 2006). Indeed, the G4 structure may serve as a primer for telomerase. These studies recapitulate earlier *in vitro* findings (Fang and Cech, 1993b). Similarly, G4 DNA in humans acts as a positive regulator of telomere elongation (Moye et al., 2015).

As noted, the presence of G4 DNA is not restricted to the telomere, but has activity in other regions. These regions include chromatin enriched for rDNA and promoters of genes encoding both transcriptional regulators and telomeric proteins (Paeschke et al., 2005). Indeed, Sgs1 helicase is required for transcriptional activation, suggesting that unwinding of the G4 DNA is needed for activation (Hershman et al., 2008). Supporting this view, multiple experiments in yeast and humans have shown that both Sgs1 and Pif1 helicases bind to and unwind the G4 DNA conformation (Han et al., 2000; Budhathoki et al., 2015; Duan et al., 2015). G4 DNA binding proteins (G4BP) are also likely to be regulators of telomeres through their action at promoters. Hence, the telomere may be influenced either directly through G4BP binding or indirectly through the regulation of the transcription of a TBP. Telomeric imperfect repeats can also form G4 structures that are thermodynamically distinct (Lustig, 1992). What is not known is what type of Hoogsteen base paired structures forms *in vivo*.

The Minimal Modular TBP

Previous investigators have postulated the least number of modules for a common functional TBP (Linger and Price, 2009). These modules consist of at least a c-myb (dsDNA) and/or an OB

(dsDNA) binding motif. In plants, a c-myb/histone H1 binding domain is a frequent telomere-binding element (Hwang and Cho, 2007). Hence, the combination of the DNA binding domains and G4 structures should be considered as an in *cis* telomere motif that has an essential role at the telomere. Many proteins that play widely different cellular roles can associate with one or more modules (**Figure 2**).

This modular structure helps to explain the finding that primary ciliate TBP (TEBF beta), the yeast Cst1 (Cdc13), and the human PPT1 TBP bind both to single-stranded DNA via OB-folds. Analogously, TEBF alpha shares homology with POT1 and binding to single-stranded DNA (Xin et al., 2007). G4 structures recruit MRX in yeast, thereby providing a source for homeostasis and a telomeric cap (Ghosal and Muniyappa, 2005). Whether this is a common phenomenon is not yet known.

The Diverse and Variable TBP: The Role of Stress Response and High Selection Pressure in Diversity

Stress response at the level of the cell cycle may initiate selection over an evolutionary time scale. In the context of the cell cycle, cells carrying a non-functional TBP may lead to dysfunctional telomeres that respond through a cellular stress mechanism. Results from the Lundblad lab suggest that after telomere loss, but before significant telomere loss, pathways with differential dependencies on telomeric regulators produce differing pathways of senescence (Ballew and Lundblad, 2013). Moreover, microarray studies reveal a major reprogramming of global gene expression after the loss of telomerase (Nautiyal et al., 2002), We have also generated evidence that argues for two pathways that retard the rate of senescence *in vivo*: the DSB and replicative repair pathways. The attempts to repair continue even under senescent conditions. These pathways may also be required in wild type cells. These data argue for multiple senescence-specific telomere loss pathways (Gao et al., 2014).

The physiological states that have conferred known cellular stress responses include replication stress response, heat shock stress response, and the oxidative stress response pathway. The oxidative response, for example, induces pathways that prevent the damage created by free radicals to a multiplicity of substrates. One of the response factors is the Ogg1 DNA glycosylase that catalyzes the repair of base excision damage induced by oxidation (Lu and Liu, 2010). Interestingly, *ogg1* mutants confer elongated telomeres, raising a possibility of a link between oxidative stress and telomeres (Akerfelt et al., 2010; Lushchak, 2011). In bacteria, the SOS response to massive DNA damage includes the activation of *recA* that coats single-stranded DNA and allows DNA repair (Witkin, 1991). The *recA* response clearly shows that stress response are common in all phyla (Jin et al., 2015).

We propose a stress response for telomere dysfunction that acts over an evolutionary level time frame. The telomere dysfunction would lead to a more continuous period of enhanced recombination and mutagenesis. In this context, cellular stress would be maintained through multiple generations.
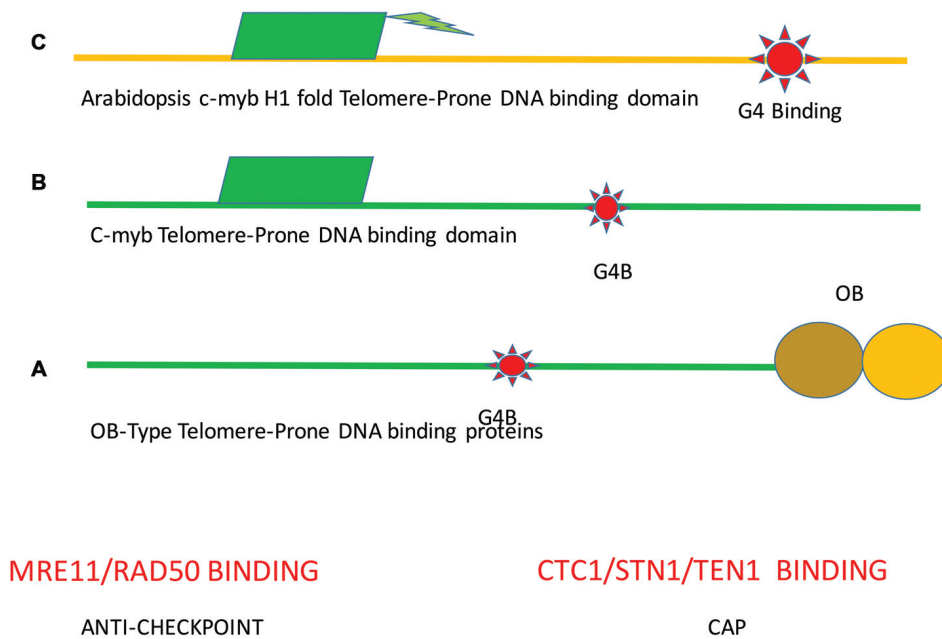
FIGURE 2 | Minimal function *cis*-acting elements at the telomere. A minimal telomere consists of different combinations of modular domains. **(A)** MR, referring to the MRX and MRN (duplex binding) complexes and the CTC complex that binds to single-stranded are required telomeric DNA binding proteins that are common to the DNA of most telomeres. Both proteins will result in far greater stability. **(B)** In the second module, the DNA duplex binding is mediated by c-myb binding sequences. TBP will also bind to G4DNA structures. Of course combinations of these modules are possible and in both cases MR binding to duplex DNA and the Cst1/Stn1/Ten1 (CST) binding to single stranded DNA will increase stability. Green indicates c-myb. This model is similar to the one proposed by Linger and Price (2009). Each modular protein can mutate or recombine with other modules to give rise to the possibility of modular based mutagenesis. **(C)** In plants, a different module is found. A single essential structure is formed that contains the c-myb domain with histone H1 folds to bind duplex DNA. Also, TBP associates with G4DNA. Other proteins that are present in plants and bind to ends include the MR and CTC complexes as in animal and fungal species. Ultimately the MR structure provides equilibrium of telomere sizes that serve to protect the end from double-strand break processing enzymes, resulting in the anti-checkpoint.

Several investigators have provided evidence for an elevation in recombination and mutagenesis in response to telomere replicative DNA stress (Shor et al., 2013; Meena et al., 2015). There is also evidence for TERRA-mediated replicative stress (Lopez de Silanes et al., 2014). Specifically, TERRA might participate in DNA-RNA G4 structures at telomeres and, at Watson-Crick based paired R-loops, forming G-loops (Duquette et al., 2004) The possibility of a G4 R-loop that could impede replication has also been a topic of speculation (Xu and Komiyama, 2012).

The induction of recombination under telomere stress could give rise to additional duplication events. One member of this pair would encode the progenitor protein of a telomere-independent nuclear chromatin protein (such as Orc1) that is maintained under selection. The second copy would be free to diverge into a TBP from Orc1. Alternatively, duplicated DNA encoding two diverged TBPs may alter their telomeric roles. We also propose an elevated rate of mutagenesis allowing rapid sequence divergence. In some situations, only a few essential residues may be necessary to form a distinct protein function. Following multiple generations under stress, partially stable proteins can attain incremental changes in protein function.

What might be a signal for a stress response that initiates the rapid evolution of TBP? For a signal to be effective, cells must be acutely sensitive to multiple indicators of telomere function. These indicators must measure parameters including (a) the state of the leading and lagging strands of semi-conservative replication, (b) the activity of telomerase, (c) the non-nucleosomal telomeric chromatin structure, (d) telomere size changes, (e) the nucleosomal subtelomeric heterochromatic state, (f) telomeric G2 cohesion, and (g) non-disjunction. We believe that the unique integration of telomeres into many cellular processes that contribute to and are influenced by telomere function may increase the rate of TBP evolution. The degree of telomeric damage cannot be so severe that the defect induces a cell checkpoint pathway within a single cell cycle. Rather, subtler defects may induce a response that leads to the formation of paralogs and novel factors that can resolve the stress over evolution.

Different modules may also respond differentially to stress response or selective pressure. An intramolecular recombination event with a homolog may lead to exon shuffling among the TBP. An additional class of paralogs may have domains that are differentially influenced by mutations (see Sir3 discussion below).

In addition to paralog formation and high levels of mutagenesis, rapid alterations in proteins can result in simple substitutions of other known proteins as well as protein loss. The data that support the former viewpoint has arisen from close examination and experimentation of the primeval yeast whole-genome duplication (WGD; Hufton and Panopoulou, 2009).

According to one Bayesian analysis of paralogs, WGD tends to be involved in generating paralogs of a similar function (Guan et al., 2007). However, a recent study has revealed that paralogs formed after yeast WGD undergo a wide range of divergence (Soria et al., 2014).

EVIDENCE FOR ELEVATED PARALOG AND MUTATIONS IN THE RAPID EVOLUTION OF TBP-A YEAST CASE STUDY

Gene Duplication and Divergence of One Paralog

Orc1 Paralog Formation with Sir3 in Budding Yeast

The yeast WGD serves as an outstanding model system for the study of the processes that lead to paralogs of differing function (Soria et al., 2014). An example (**Figure 3**) that has been examined in multiple fungal species (Capaldi and Berger, 2004) is the Origin of Replication Subunit 1 (Orc1). One of the paralogs of Orc1 in *S. cerevisiae* [and very closely related species (e.g., *S. byanus*)] is the Silencer Information Regulator 3 (Sir3; Liaw and Lustig, 2006).

Sir3 is a unique nuclear chromatin protein that functions in mating type and telomere silencing protein. In its mating-type silencing role, Sir3 maintain two of the three cassettes of mating-type information in a silent state, leaving only one of the cassettes in an expressed state (Lustig, 1998). Sir3 is essential for maintaining, but not establishing, the silencing of *HML* alpha and *HMR* a, present close to the left and right telomeres of chromosome III, respectively. Studies are conducted in the absence of mating type switching by using strains that lack the homothallic switching: gene, *HO*. In *ho* cells, incapable of mating type switching, only one mating type allele is expressed in haploid cells in the presence of the Sir3-dependent silent cassettes. The mating of *ho* haploids of different mating types produces diploids, permitting meiotic analyses. Meiosis is, of course, a significant selective force in evolution.

Sir3 is also essential for the silencing of ectopic telomere-adjacent genes associated with heterochromatic regions, a process termed telomere position effect (TPE; Gottschling et al., 1990). It is unlikely, however, that TPE plays a large role in cells lacking the ectopic silencing marker. Rather TPE is a quantitative read-out of the magnitude of heterochromatin formation in subtelomeric regions. In that regard, Sir3-dependent fold-back structures form at the subtelomeric/telomeric junction during maintenance of heterochromatin, a conclusion based on genetic and biochemical studies (Hecht et al., 1996) The fold-back structures resulted in homodimerization and heterodimerization of Sir3 and Sir4 in the telomeric regions and between telomeric and subtelomeric regions. At these sites, the heterochromatic proteins Sir3 and Sir4 interact with the C-terminal domain of the telomeric Rap1, and with N-termini of histones H3 and H4 (Kitada et al., 2012). Sir3 may also be important for the deletion of potential t-loops that may serve a sizing and protective functions (Bucholc et al., 2001)

Both the paralog Sir3 and the Sir4 protein associate with heterochromatic condensed chromatin and are necessary for maintenance, but not the establishment of silencing and heterochromatin. At higher concentrations, Sir3 has the unique property of spreading heterochromatin over an increasing distance from the telomere, a classic feature of eukaryotic heterochromatin (Buchberger et al., 2008)

The yeast Orc1 protein is a 914 amino acid (aa) protein with strong overall homology to other fungal Orc1 species. Orc1 contains the bromodomain adjacent homology (BAH) domain, an AAA ATP activity, and a Cdc6 winged helix domain (Wang et al., 1999; Capaldi and Berger, 2004). Orc1 has many of the features that are required to associate with the chromatin present during the initiation of DNA replication (Jiang et al., 2007; Prasanth et al., 2010; Thomae et al., 2011; **Figure 3**). Sir3 has 50% amino acids identity or similarity with these domains of Orc1. The most diverged portion of Sir3 primary sequence from Orc1 sequence is the 145aa C-terminal domain (CTD) present in Sir3. We have defined the CTD by the terminal sequences and the silencing activity displayed when the CTD is tethered to a specific chromosome (tethered silencing) and does not refer intrinsically to any structure (Liaw and Lustig, 2006; **Figure 3**).

The CTD has been investigated by (a) a tethered silencing assay of the domain containing Sir3 in trans, (b) CTD crystallization, (c) CTD mutational analysis, and (d) a study of the CTD in context of the full length protein (Liaw and Lustig, 2006; Oppikofer et al., 2013). Two major conclusions can be drawn from these studies. First, the CTD contains a dimerization domain composed of a winged helix structure. Second, the CTD has a mutation of unknown function that is likely to be redundant within the full length Sir3. This structure is likely to be required for the assembly of histones and Sir gene products (Liaw and Lustig, 2006). In addition, Both Orc1 and Cdc6 maintain residual function in tethered silencing assays, suggesting a significant, but insufficient, role of the Orc1 and Cdc6 winged helix in silencing (Liaw and Lustig, 2006). Cdc6 can also physically associate with Orc1, but not with Sir3 (**Figure 4**).

A close relative of *S. cerevisiae*, *S. byanus* can substitute for ScSir3 in a mating assay, despite its minimal CTD homology to ScSir3. We would predict that that domain of CTD also forms a winged helix domain, although this is uninvestigated. Such a rapid change in residues, however, may be due to a neutral effect of indels (mobile integrants) after the high levels of mutagenesis during the evolution of Sir3 (**Figure 5**).

Orc1/Sir3 Paralog Formation in Other Fungi

Our current studies show that a different form of Sir3 present in the Orc1 progenitor results in a pathogenic relative of *S. cerevisiae*, *Candida glabrata*. While ScSir3 behaves as a silencing protein (Liaw and Lustig, 2006), Cg Sir3 functions in a more elaborate silencing of many of the eicosapentaenoic acid (EPA) adherens. The adherens are under both positive and negative control for pathogenicity (Rosas-Hernandez et al., 2008; Halliwell et al., 2012).

Interestingly, the adheren silencer is very close to the telomere, implicating functional involvement (Liaw and Lustig,

2006). Pathogenicity is also dependent upon other telomeric proteins, including Ku and Rif1. Each telomere of *C. glabrata* behaves differently in the context of silencing. The cgSir3 CTD is divergent from ScOrc1 or ScSir3 (**Figures 6** and **7**). We analyzed the Sir3 phylogenetic tree using Phylome DB (www.phylomedb.org) (Huerta-Cepas et al., 2008; Huerta-Cepas et al., 2014; **Figure 7**). Curiously, *C. glabrata* and the closely related pathogen *Nakaseomyces delphensis* have very similar CTD domains, but both are highly diverged from *S. cerevisiae* Sir3 CTD to the level of insignificance (Lustig, unpublished data). We therefore have operationally termed this region the CTD2 region. The altered CTD2 function undoubtedly responds to a different set of selective pressures, the expression of EPA adherens that are necessary for pathogenicity (Ielasi et al., 2012). The *C. glabrata* obligate haploid also has three mating type cassettes, reminiscent of ScSir3, but not involved in mating type identity. Nonetheless, one of these near the telomere is also is under the control of Sir3 at a transcriptional level, remnants of a system that may be in the process of evolving into a new function (Yanez-Carrillo et al., 2014). Additional selection pressures, yet to be deduced, may be present to influence CgSir3. The functional residues of CTD2 have not been studied (**Figure 7**). Study of this region also suggests that CgSir3 in *S. cerevisiae* and *S. glabrata* have an ancient common ancestor.

The CTD, in this case, would not be expected to be highly sensitive to mutagenesis, since the function of active sites can be perturbed by only a few single mutations. However, CTD2 may be similar to CTD1 in providing a mutational buffer against functional change. Both types of Sir3 diverged from Orc1 after whole genome duplication. Alternatively, although remote, Orc 1 may act independently but at high levels in paralog formation. In either case, the two forms of Sir3 may have diverged rapidly to produce the extant unique proteins (Fabre et al., 2005). Heterochromatin proteins in other organisms (Sugiyama et al., 2005), such as HP1 of *S. pombe*, share homology and function between centromere and telomere heterochromatin but have no evolutionary relationship to Sir3.
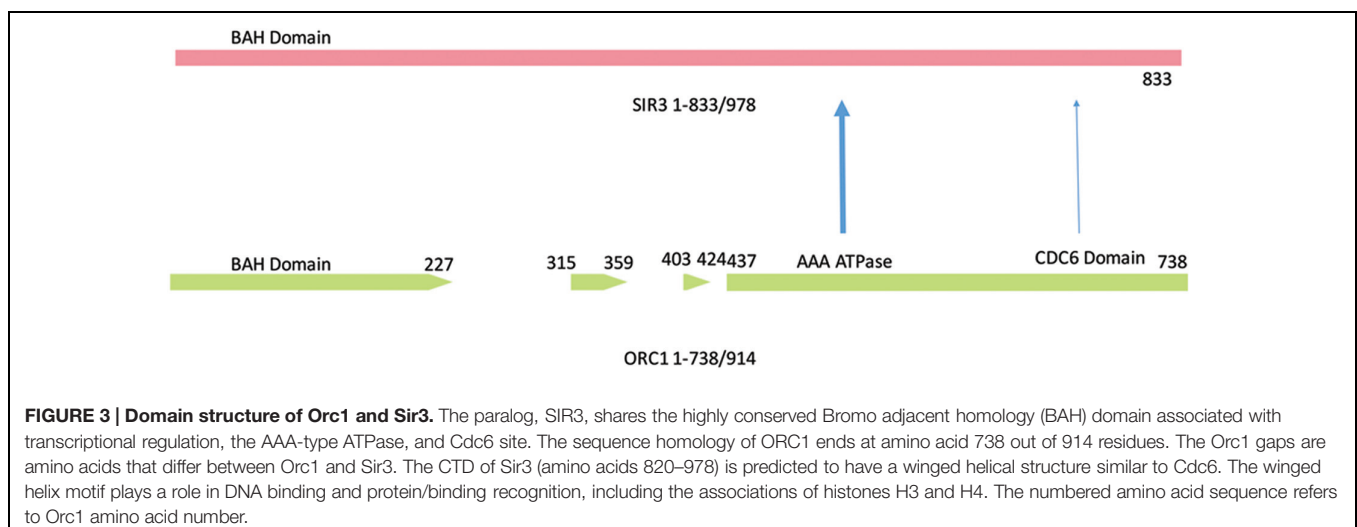
Thus, the Orc1/Sir3 system appears to be capable of two functional changes via the Sir3 CTD domain. Although a micro-evolutionary case, the paralogs are well suited examples of proteins with differing function. We propose that the elevation of paralog formation and mutagenesis at an evolutionary scale can promote rapid deviations in the related strains. Indeed, the divergence in CTD1 and CTD2 supports such an enhanced level of mutagenesis. Finally, we propose that this rate of adaptation is likely due to a yeast stress response that elevates recombination and mutagenesis.

The Separation of Two Telomeric Functions by Gene Duplication: Est1/Ebs1

Sir3 is not the only example of a paralog that can lead to altered activity after WGD. Est1, a part of the telomerase holoenzyme, has a paralog, Ebs1 (Zhou et al., 2000; Luke et al., 2007). Ebs1 is a component of the non-sense-mediated mRNA decay pathways. Indeed, non-sense-mediated mRNA decay reduces telomere size (Lew et al., 1998). Ebs1 shares only 27% homology with Est1 throughout the protein, so that the conserved domain involved in size control remains ambiguous. Ebs1 is also present in a single Est1/Ebs1 protein in the more distant pre-WGD *Kluyveromyces lactis* (Hsu et al., 2012). This fusion protein is likely to be closer to the common ancestral precursor protein. The precursor must have produced paralogs during or after WGD, diverging into separate ScEbs2 and ScEst1.

What Happened to RAP1? The Argument in Favor of Hypomorphs!

Most Rap1 molecules share the Rap1 C-terminal (RCT) domain (Chen et al., 2011). Rap1, in yeast, serves as the major functional yeast TBP that also is a DNA binding protein

FIGURE 3 | Domain structure of Orc1 and Sir3. The paralog, SIR3, shares the highly conserved Bromo adjacent homology (BAH) domain associated with transcriptional regulation, the AAA-type ATPase, and Cdc6 site. The sequence homology of ORC1 ends at amino acid 738 out of 914 residues. The Orc1 gaps are amino acids that differ between Orc1 and Sir3. The CTD of Sir3 (amino acids 820–978) is predicted to have a winged helical structure similar to Cdc6. The winged helix motif plays a role in DNA binding and protein/binding recognition, including the associations of histones H3 and H4. The numbered amino acid sequence refers to Orc1 amino acid number.
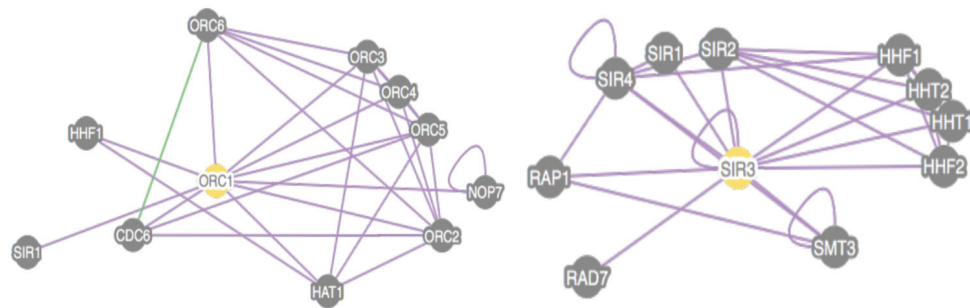
FIGURE 4 | Proteomic view of Association of Cdc6 with both Orc1 and Sir3 Protein-based associations are present between Orc1 and other yeast nuclear factors. We conducted an SGD search for physical interactions between Orc1 or Sir3 and other cellular proteins using at least four experiments. Orc1 is capable of associating with Cdc6 while Sir3 is not. In no experiment was Cdc6/Sir3 binding observed. One genetic interaction between Orc3 and Orc6 is also shown in this figure.

FIGURE 5 | Sequence Homology of CTD (CTT) in differing Saccharomyces species. The lines (from top to bottom) display Sir3 from *Saccharomyces cerevisiae* sequence, *S. byanus, S. milkatae* a, *S. paradoxis*, an independent *S. byanus, S. castilli*, and *S. kudriavzevii*. Yellow refers to identity, pink to high similarity, and green to statistical similarity.

and a transcriptional activator of glycolytic and ribosomal protein genes (Shore, 1994; Park et al., 2002). A great deal of evidence has amassed for the function of mammalian RAP1 through multiple assays (Li and de Lange, 2003; O'Connor et al., 2004; Bae and Baumann, 2007; Bombarde et al., 2010; Chen et al., 2011; Arat and Griffith, 2012) and is the most conserved protein at the telomere (Yang et al., 2011; Martinez et al., 2013; Yeung et al., 2013). However, recent data revealed

the unexpected result that loss of RAP1 in both mice and humans had no functional impacts at telomeres, but only in transcription (Martinez et al., 2010; Kabir et al., 2014). This could be the result of a requirement for the role in promoter activation in a limited number of transcripts (Bae and Baumann, 2007; Bombarde et al., 2010; Arat and Griffith, 2012) or the presence of a redundant telomere Rap1-like protein. Rap1may be present then at human telomeres as an artifact of the
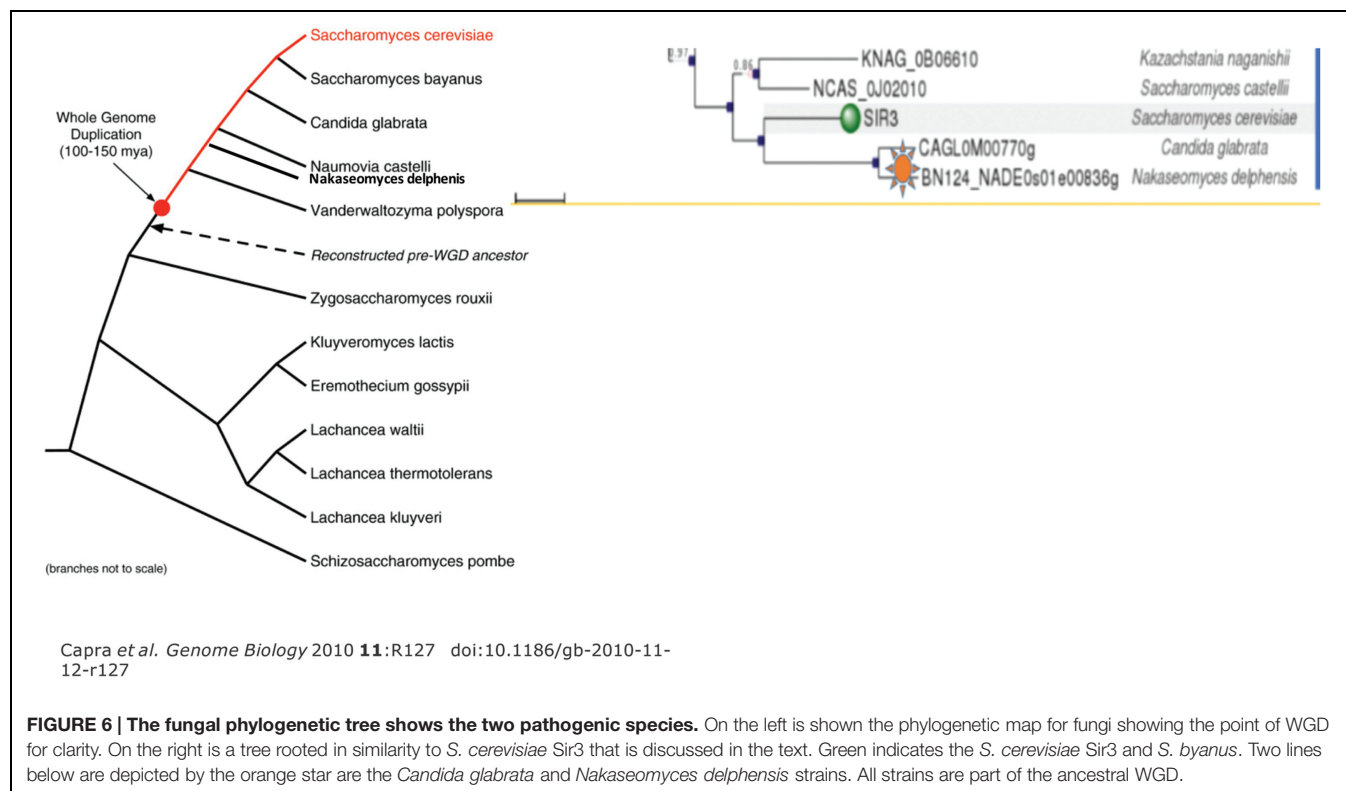
FIGURE 6 | The fungal phylogenetic tree shows the two pathogenic species. On the left is shown the phylogenetic map for fungi showing the point of WGD for clarity. On the right is a tree rooted in similarity to *S. cerevisiae* Sir3 that is discussed in the text. Green indicates the *S. cerevisiae* Sir3 and *S. byanus*. Two lines below are depicted by the orange star are the *Candida glabrata* and *Nakaseomyces delphensis* strains. All strains are part of the ancestral WGD.

FIGURE 7 | High sequence similarity is present between *C. glabrata* (Cg) and *N. delphensis* (Nr). The two different CTD sequences are shown in two close pathogenic species related to *Candida*. Red nucleotides shown identity and orange residues show similarities.

conserved heterodimer, TRF2/Rap1, at some promoters (Kabir et al., 2014).

How could RAP1 make such an evolutionary leap? Is this really due to a lack of function at telomeres? There are two other possible considerations. First, the RCT domain that is similar to the fission yeas *S. pombe* associates with the TRF2-like protein, Taz1, where deletion mutants have shown a high level of telomere involvement (Park et al., 2002). It seems unlikely that the lack of nucleosomes in *S. pombe* telomeric chromatin and its presence

in human telomeres governs this loss of Rap1 activity, Rap1 binding occurs via Taz1 and can function transcriptionally on nucleosomal DNA in mice or human cells (Wright et al., 1992; Park et al., 2002; Tomaska et al., 2004b; Galati et al., 2012).

We propose a number of solutions to this odd situation. The first, functional redundancy, is unattractive in its simplest form, since its presence would mask the phenotypes of rap1 mutants. Rather, we make a second proposal, albeit speculative, based on the inability to explain conservation

FIGURE 8 | Selected stress responses and surface interactions. *Selective Stress* **(A–D)** shows the proposed effect of stress responses over evolutionary time. **(A)** The initial modular telomere with a TBP (red). **(B)** Duplication of DNA (orange) encoding the protein. **(C)** Separation and mutagenesis (purple) of the two DNAs and the altered protein (orange and purple box). **(D)** The final telomere, bound by the mutated protein. Protein Legend: c-myb containing protein: green, G4 DNA binding protein (star), and OB-fold protein (yellow). *Protein interfaces*. In example **(E)**, protein (blue) associated with the c-myb protein with an unfavorable surface interaction shown by the x. **(F)** Protein interfaces that interact favorably with a second protein (red) to form a stable structure as indicated by the +. A simplified minimal modular telomere is shown just for reference.

in the absence of selection. Similarly, the transcriptional function in human cells do not appear extensive enough to induce such a strong selection. We therefore suggest differences intrinsic to hypomorphic and null alleles. In the presence of a horrendous telomeric damage event, viable cells could produce a "defect response system," not unlike many of the responses to serious cellular defects. A previous observation noted that a loss of RAP1 led to an increase in recombination (Sfeir et al., 2010), consistent with this idea. As noted, in the yeast *S. cerevisiae*, there is some evidence for rapid effects on recombination and mutagenesis in the face of telomere disaster (Shor et al., 2013; Meena et al., 2015). Recombinational induction has also been observed rapidly in yeast without the expected DNA damage response pathway (Lustig, unpublished data), consistent with effect found in human cells. We would like to propose that there is a telomere response system that is distinct from the DNA damage response pathway that can sense (through an unknown signal) an alteration in essential chromatin structure. A null allele might simply place too much stress on the cell, promoting the induction of specific proteins, one of which may have some functions of RAP1. Possibly, more information would be gained by the use of hypomorphic mutations that retains partial Rap1 function that may not be susceptible to this putative response. Under these hypothetical, conditions, the telomere damage may be below the sensitivity of detection, circumventing the effect of the response system. Under non-null conditions, the true effects of Rap1may be better determined, one way or another. This issue may be raised for a number of observations that seem to be signaling effects, rather than the original transient effect of the mutation

COMPLEX TELOMERES: SPECULATION ON THE FLEXIBLE DYNAMICS OF SHELTERIN

We normally think of shelterin an ordered set of proteins that are invariant in humans (de Lange, 2005). Shelterin is an outstanding model system to discuss the numerous ways of attaining a broader level of control. The conservation of shelterin function is likely to be a consequence of the interaction between the functional subunits (de Lange, 2005) that contain common motifs such as c-myb, OB, and G4 modules. Also, it is likely to involve the formation of only a subset of protein/protein junctions that are sterically and thermodynamically permissible. In addition, a subgroup of chromatin-associated proteins, TRF1, TRF2, and POT1 has probably evolved through a paralog-related process. So the overall constraints of variable TBPs include geometry, protein/protein interfaces, and the presence of proteins having truly unique functions. This set of constraints will vary through evolution in species having a multi-subunit shelterin-like structures. The nature and frequency of the multi-subunit protein interfaces would select for only steric and thermodynamic limitations, based on protein folding structures that fit the geometric and functional needs of the telomere.

When honing in on vertebrates (or mammals), it is clear that TRF1 is the ancestral protein to vertebrate TRF1 and TRF2 paralogs (Horvath, 2008). Similarly, TRF2, a paralog of TRF1, has become substantially specialized. TRF2 plays multiple roles in telomere maintenance and dynamics that are due to the unique chromatin structure (Broccoli et al., 1997). However, the

TRF1-nucleated class may have been derived by a TRF1 ortholog precursor to the major telomere proteins present throughout vertebrates (Horvath, 2008). Therefore, previous studies may not solve the telomere function in all complex vertebrates (except in mammals), but demonstrate one of many possible solutions that exist in extant organisms.

Paralog functions do play a role in some shelterin complex telomeres such as in the formation of Pot1a and Pot1b in rodents (Hockemeyer et al., 2006), but also in other organisms that have simpler telomeres, such as *Arabidopsis*, green algae, and the ciliate *Tetrahymena*. Pot1 forms homologs Pot1a and Pot1b in several species that are distant evolutionary, such as *Tetrahymena* (Jacob et al., 2007; Shakirov et al., 2009). The maintenance of the POT1 class of proteins is critical for shelterin function. POT1 plays a predominant role in the accessibility to and modulation of telomerase. Tankyrase, the protein that is responsible for the loading of TRF1 in vertebrates, also plays a role in plants. Importantly, this is a class of proteins with similar structure, but differing function, another possible outcome of paralog formation that both play a role at the telomere (Cook et al., 2002). In plants, tankyrases do not act as a TRF1 loading factors. That is not surprising given the evidence that TRF proteins are not functional in *Arabidopsis* (Boltz et al., 2014; Fulcher and Riha, 2015). A resolution of whether the tankyrases in plants are true paralogs and the nature of their specific function at telomeres will require future investigation. Telomerase holoenzyme also undergoes species-dependent paralog formation, particularly in Est1 and Pot1 (e.g., Est1a, b, c Pot1a, Pot1b). Est1a complements senescence in yeast and performs the telomerase function. The function of Est1b and Est1c are unclear (Sealey et al., 2011). Paralogs of Est1 are exclusively observed in humans. As expected, the conserved TBP components discussed in section "The Conserved Elements of TBP" are also present at human telomeres in addition to shelterin. This model coupling paralog formation and interface compatibility in the presence of a minimal number of conserved proteins is a proposal that tries to explain the rapid evolution of TBPs. Other ideas involving the cooperativity of processes are in no way mutually exclusive from our considerations.

Hence, the plethora of proteins present in a given cell type is likely to overcome a major thermodynamic barrier to the formation of shelterin. The formation of shelterin-like complexes may be the consequence of a trial and error process that may require sub-complexes. The shelterin complexes that are present in more complex organisms are under, as yet, uncharacterized selection pressures.

A MODEL FOR THE RAPID EVOLUTION OF TELOMERE BINDING PROTEINS

We propose five central principles that serve as the foundation for the rapid evolution of telomere-binding proteins. First, paralog formation seems to be a primary driving force in rapid evolution rather than ortholog formation. Second, telomere-binding proteins consist of a limited number of conserved motifs such as c-myb, OB, and G4 domains, which can initiate a minimal level of protection. Third, stress response at the evolutionary level may occur as the result of telomere dysfunction that increases the rate of recombination and mutagenesis. Fourth, the major limiting function in complex shelterins is the number of protein/protein interfaces needed to form a multi-subunit complex-as least at the structural level. Specific required functions may be under additional selection pressure. Fifth, some complexes provide novel functions (e.g., Pot2 access to telomerase) and the transducing of signals over a large portion of the telomere that may have effects that are greater than the sum of individual protein species. These five principles serve as the basis of any attempt to create a coherent evolutionary model.

We believe that the vastly different organismal requirements may alter selection patterns. For example, the abundance of telomeres, the cell cycle control of replication, the coordination of telomere and semi-conservative replication may have profound effects on the nature of telomere change (Horvath, 2008).

We propose, therefore, that the phenomenon of "rapid evolution" is the consequence of the high level of paralogs, producing distinct functional proteins through the induction of telomere stress response. While telomere evolution is clearly not the only case in which paralogs may evolve to form other functions, alterations in TBPs must be driven by the need for rapid response to physiological change (**Figure 8**).

A large number of experimental studies serve as the basis of these models. A complete solution to the patterns observed will require a greater knowledge of telomere protein/protein interactions and telomere protein domain structure. This level of understanding requires a collaborative effort to characterize more organisms for genetic analysis.

AUTHOR CONTRIBUTIONS

AJL is responsible for the content contained in the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Abdallah, P., Luciano, P., Runge, K. W., Lisby, M., Geli, V., Gilson, E., et al. (2009). A two-step model for senescence triggered by a single critically short telomere. *Nat. Cell Biol.* 11, 988–993. doi: 10.1038/ncb1911

Akerfelt, M., Morimoto, R. I., and Sistonen, L. (2010). Heat shock factors: integrators of cell stress, development and lifespan. *Nat. Rev. Mol. Cell Biol.* 11, 545–555. doi: 10.1038/nrm2938

Amiard, S., Depeiges, A., Allain, E., White, C. I., and Gallego, M. E. (2011). *Arabidopsis* ATM and ATR kinases prevent propagation of genome damage caused by telomere dysfunction. *Plant Cell* 23, 4254–4265. doi: 10.1105/tpc.111.092387

Arat, N. O., and Griffith, J. D. (2012). Human Rap1 interacts directly with telomeric DNA and regulates TRF2 localization at the telomere. *J. Biol. Chem.* 287, 41583–41594. doi: 10.1074/jbc.M112.415984

Arora, R., and Azzalin, C. M. (2015). Telomere elongation chooses TERRA ALTernatives. *RNA Biol.* 12, 938–941. doi: 10.1080/15476286.2015.1065374

Audry, J., Maestroni, L., Delagoutte, E., Gauthier, T., Nakamura, T. M., Gachet, Y., et al. (2015). RPA prevents G-rich structure formation at lagging-strand telomeres to allow maintenance of chromosome ends. *EMBO J.* 34, 1942–1958. doi: 10.15252/embj.201490773

Bae, N. S., and Baumann, P. (2007). A RAP1/TRF2 complex inhibits nonhomologous end-joining at human telomeric DNA ends. *Mol. Cell* 26, 323–334. doi: 10.1016/j.molcel.2007.03.023

Ballew, B. J., and Lundblad, V. (2013). Multiple genetic pathways regulate replicative senescence in telomerase-deficient yeast. *Aging Cell* 12, 719–727. doi: 10.1111/acel.12099

Baumann, P., and Cech, T. R. (2000). Protection of telomeres by the Ku protein in fission yeast. *Mol. Biol. Cell* 11, 3265–3275. doi: 10.1091/mbc.11.10.3265

Bianchi, A., and Shore, D. (2007). Increased association of telomerase with short telomeres in yeast. *Genes Dev.* 21, 1726–1730. doi: 10.1101/gad.438907

Blackburn, E. H., Chan, S., Chang, J., Fulton, T. B., Krauskopf, A., McEachern, M., et al. (2000). Molecular manifestations and molecular determinants of telomere capping. *Cold Spring Harb. Symp. Quant. Biol.* 65, 253–263. doi: 10.1101/sqb.2000.65.253

Boltz, K. A., Jasti, M., Townley, J. M., and Shippen, D. E. (2014). Analysis of poly(ADP-Ribose) polymerases in *Arabidopsis* telomere biology. *PLoS ONE* 9:e88872. doi: 10.1371/journal.pone.0088872

Bombarde, O., Boby, C., Gomez, D., Frit, P., Giraud-Panis, M. J., Gilson, E., et al. (2010). TRF2/RAP1 and DNA-PK mediate a double protection against joining at telomeric ends. *EMBO J.* 29, 1573–1584. doi: 10.1038/emboj.2010.49

Bonetti, D., Martina, M., Falcettoni, M., and Longhese, M. P. (2013). Telomere-end processing: mechanisms and regulation. *Chromosoma* [Epub ahead of print].

Broccoli, D., Smogorzewska, A., Chong, L., and de Lange, T. (1997). Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nat. Genet.* 17, 231–235. doi: 10.1038/ng1097-231

Bryan, C., Rice, C., Harkisheimer, M., Schultz, D. C., and Skordalakes, E. (2013). Structure of the human telomeric Stn1-Ten1 capping complex. *PLoS ONE* 8:e66756. doi: 10.1371/journal.pone.0066756

Buchberger, J. R., Onishi, M., Li, G., Seebacher, J., Rudner, A. D., Gygi, S. P., et al. (2008). Sir3-nucleosome interactions in spreading of silent chromatin in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* 28, 6903–6918. doi: 10.1128/MCB.01210-08

Bucholc, M., Park, Y., and Lustig, A. J. (2001). Intrachromatid excision of telomeric DNA as a mechanism for telomere size control in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* 21, 6559–6573. doi: 10.1128/MCB.21.19.6559-6573.2001

Budhathoki, J. B., Stafford, E. J., Yodh, J. G., and Balci, H. (2015). ATP-dependent G-quadruplex unfolding by Bloom helicase exhibits low processivity. *Nucleic Acids Res.* 43, 5961–5970. doi: 10.1093/nar/gkv531

Capaldi, S. A., and Berger, J. M. (2004). Biochemical characterization of Cdc6/Orc1 binding to the replication origin of the euryarchaeon *Methanothermobacter* thermoautotrophicus. *Nucleic Acids Res.* 32, 4821–4832. doi: 10.1093/nar/gkh819

Carneiro, T., Khair, L., Reis, C. C., Borges, V., Moser, B. A., Nakamura, T. M., et al. (2010). Telomeres avoid end detection by severing the checkpoint signal transduction pathway. *Nature* 467, 228–232. doi: 10.1038/nature09353

Chang, M., Arneric, M., and Lingner, J. (2007). Telomerase repeat addition processivity is increased at critically short telomeres in a Tel1-dependent manner in *Saccharomyces cerevisiae*. *Genes Dev.* 21, 2485–2494. doi: 10.1101/gad.1588807

Chen, L. Y., Redon, S., and Lingner, J. (2012). The human CST complex is a terminator of telomerase activity. *Nature* 488, 540–544. doi: 10.1038/nature11269

Chen, Y., Rai, R., Zhou, Z. R., Kanoh, J., Ribeyre, C., Yang, Y., et al. (2011). A conserved motif within RAP1 has diversified roles in telomere protection and regulation in different organisms. *Nat. Struct. Mol. Biol.* 18, 213–221. doi: 10.1038/nsmb.1974

Ciapponi, L., Cenci, G., Ducau, J., Flores, C., Johnson-Schlitz, D., Gorski, M. M., et al. (2004). The *Drosophila* Mre11/Rad50 complex is required to prevent both telomeric fusion and chromosome breakage. *Curr. Biol.* 14, 1360–1366. doi: 10.1016/j.cub.2004.07.019

Cifuentes-Rojas, C., Kannan, K., Tseng, L., and Shippen, D. E. (2010). Two RNA subunits and POT1a are components of *Arabidopsis* telomerase. *Proc. Natl. Acad. Sci. U.S.A* 108, 73–78. doi: 10.1073/pnas.1013021107

Cook, B. D., Dynek, J. N., Chang, W., Shostak, G., and Smith, S. (2002). Role for the related poly(ADP-Ribose) polymerases tankyrase 1 and 2 at human telomeres. *Mol. Cell Biol.* 22, 332–342. doi: 10.1128/MCB.22.1.332-342.2002

de Lange, T. (2002). Protection of mammalian telomeres. *Oncogene* 21, 532–540. doi: 10.1038/sj.onc.1205080

de Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* 19, 2100–2110. doi: 10.1101/gad.1346005

de Lange, T. (2009). How telomeres solve the end-protection problem. *Science* 326, 948–952. doi: 10.1126/science.1170633

de Lange, T. (2015). A loopy view of telomere evolution. *Front. Genet.* 6:321. doi: 10.3389/fgene.2015.00321

Dimitrova, N., and de Lange, T. (2009). Cell cycle-dependent role of MRN at dysfunctional telomeres: ATM signaling-dependent induction of nonhomologous end joining (NHEJ) in G1 and resection-mediated inhibition of NHEJ in G2. *Mol. Cell Biol.* 29, 5552–5563. doi: 10.1128/MCB.00 476-09

Duan, X. L., Liu, N. N., Yang, Y. T., Li, H. H., Li, M., Dou, S. X., et al. (2015). G-quadruplexes significantly stimulate Pif1 helicase-catalyzed duplex DNA unwinding. *J. Biol. Chem.* 290, 7722–7735. doi: 10.1074/jbc.M114.628008

Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F., and Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.* 18, 1618–1629. doi: 10.1101/gad.1200804

Evans, S. K., and Lundblad, V. (2000). Positive and negative regulation of telomerase access to the telomere. *J. Cell Sci.* 113(Pt 19), 3357–3364.

Fabre, E., Muller, H., Therizols, P., Lafontaine, I., Dujon, B., and Fairhead, C. (2005). Comparative genomics in hemiascomycete yeasts: evolution of sex, silencing, and subtelomeres. *Mol. Biol. Evol.* 22, 856–873. doi: 10.1093/molbev/msi070

Fang, G., and Cech, T. R. (1993a). The beta subunit of Oxytricha telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell* 74, 875–885. doi: 10.1016/0092-8674(93)90467-5

Fang, G., and Cech, T. R. (1993b). Characterization of a G-quartet formation reaction promoted by the beta-subunit of the Oxytricha telomere-binding protein. *Biochemistry* 32, 11646–11657. doi: 10.1021/bi00094a022

Forstemann, K., and Lingner, J. (2001). Molecular basis for telomere repeat divergence in budding yeast. *Mol. Cell Biol.* 21, 7277–7286. doi: 10.1128/MCB.21.21.7277-7286.2001

Foster, S. S., Zubko, M. K., Guillard, S., and Lydall, D. (2006). MRX protects telomeric DNA at uncapped telomeres of budding yeast cdc13-1 mutants. *DNA Repair (Amst.)* 5, 840–851. doi: 10.1016/j.dnarep.2006.04.005

Frank-Vaillant, M., and Marcand, S. (2002). Transient stability of DNA ends allows nonhomologous end joining to precede homologous recombination. *Mol. Cell* 10, 1189–1199. doi: 10.1016/S1097-2765(02)00705-0

Fulcher, N., and Riha, K. (2015). Using centromere mediated genome elimination to elucidate the functional redundancy of candidate telomere binding proteins in *Arabidopsis thaliana*. *Front. Genet.* 6:349. doi: 10.3389/fgene.2015.00349

Galati, A., Magdinier, F., Colasanti, V., Bauwens, S., Pinte, S., Ricordy, R., et al. (2012). TRF2 controls telomeric nucleosome organization in a cell cycle phase-dependent manner. *PLoS ONE* 7:e34386. doi: 10.1371/journal.pone.0034386

Gao, G., Bi, X., Chen, J., Srikanta, D., and Rong, Y. S. (2009). Mre11-Rad50-Nbs complex is required to cap telomeres during *Drosophila* embryogenesis.

Proc. Natl. Acad. Sci. U.S.A. 106, 10728–10733. doi: 10.1073/pnas.0902707106

Gao, H., Moss, D. L., Parke, C., Tatum, D., and Lustig, A. J. (2014). The Ctf18RFC clamp loader is essential for telomere stability in telomerase-negative and mre11 mutant alleles. *PLoS ONE* 9:e88633. doi: 10.1371/journal.pone.0088633

Garavis, M., Gonzalez, C., and Villasante, A. (2013). On the origin of the eukaryotic chromosome: the role of noncanonical DNA structures in telomere evolution. *Genome Biol. Evol.* 5, 1142–1150. doi: 10.1093/gbe/evt079

Ghodke, I., and Muniyappa, K. (2013). Processing of DNA double-stranded breaks and intermediates of recombination and repair by *Saccharomyces cerevisiae* Mre11 and its stimulation by Rad50, Xrs2, and Sae2 proteins. *J. Biol. Chem.* 288, 11273–11286. doi: 10.1074/jbc.M112.439315

Ghosal, G., and Muniyappa, K. (2005). *Saccharomyces cerevisiae* Mre11 is a high-affinity G4 DNA-binding protein and a G-rich DNA-specific endonuclease: implications for replication of telomeric DNA. *Nucleic Acids Res.* 33, 4692–4703. doi: 10.1093/nar/gki777

Gottschling, D. E., Aparicio, O. M., Billington, B. L., and Zakian, V. A. (1990). Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* 63, 751–762. doi: 10.1016/0092-8674(90)90141-Z

Gould, S. J. (1996). *Full House : The Spread of Excellence From Plato To Darwin*, 1st Edn. New York, NY: Harmony Books.

Greenwell, P. W., Kronmal, S. L., Porter, S. E., Gassenhuber, J., Obermaier, B., and Petes, T. D. (1995). TEL1, a gene involved in controlling telomere length in *S. cerevisiae*, is homologous to the human ataxia telangiectasia gene. *Cell* 82, 823–829. doi: 10.1016/0092-8674(95)90479-4

Guan, Y., Dunham, M. J., and Troyanskaya, O. G. (2007). Functional analysis of gene duplications in *Saccharomyces cerevisiae*. *Genetics* 175, 933–943. doi: 10.1534/genetics.106.064329

Halliwell, S. C., Smith, M. C., Muston, P., Holland, S. L., and Avery, S. V. (2012). Heterogeneous expression of the virulence-related adhesin Epa1 between individual cells and strains of the pathogen *Candida glabrata*. *Eukaryot. Cell* 11, 141–150. doi: 10.1128/EC.05232-11

Han, H., Bennett, R. J., and Hurley, L. H. (2000). Inhibition of unwinding of G-quadruplex structures by Sgs1 helicase in the presence of N,N'-bis[2-(1-piperidino)ethyl]-3,4,9,10-perylenetetracarboxylic diimide, a G-quadruplex-interactive ligand. *Biochemistry* 39, 9311–9316. doi: 10.1021/bi000482r

Hecht, A., Strahl-Bolsinger, S., and Grunstein, M. (1996). Spreading of transcriptional repressor SIR3 from telomeric heterochromatin. *Nature* 383, 92–96. doi: 10.1038/383092a0

Hershman, S. G., Chen, Q., Lee, J. Y., Kozak, M. L., Yue, P., Wang, L. S., et al. (2008). Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 36, 144–156. doi: 10.1093/nar/gkm986

Hirano, Y., Fukunaga, K., and Sugimoto, K. (2009). Rif1 and rif2 inhibit localization of tel1 to DNA ends. *Mol. Cell* 33, 312–322. doi: 10.1016/j.molcel.2008.12.027

Hockemeyer, D., Daniels, J. P., Takai, H., and de Lange, T. (2006). Recent expansion of the telomeric complex in rodents: two distinct POT1 proteins protect mouse telomeres. *Cell* 126, 63–77. doi: 10.1016/j.cell.2006.04.044

Horvath, M. P. (2008). "Evolution of telomere binding proteins," in *Origin and Evolution of Telomeres*, eds L. Tomaska and J. Nosek (Austin, TX: Landes Bioscience), 83–99.

Hsu, M., Yu, E. Y., Sprusansky, O., McEachern, M. J., and Lue, N. F. (2012). Functional analysis of the single Est1/Ebs1 homologue in kluyveromyces lactis reveals roles in both telomere maintenance and rapamycin resistance. *Eukaryot. Cell* 11, 932–942. doi: 10.1128/EC.05319-11

Huerta-Cepas, J., Bueno, A., Dopazo, J., and Gabaldon, T. (2008). PhylomeDB: a database for genome-wide collections of gene phylogenies. *Nucleic Acids Res.* 36, D491–D496. doi: 10.1093/nar/gkm899

Huerta-Cepas, J., Capella-Gutierrez, S., Pryszcz, L. P., Marcet-Houben, M., and Gabaldon, T. (2014). PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. *Nucleic Acids Res.* 42, D897–D902. doi: 10.1093/nar/gkt1177

Hufton, A. L., and Panopoulou, G. (2009). Polyploidy and genome restructuring: a variety of outcomes. *Curr. Opin. Genet. Dev.* 19, 600–606. doi: 10.1016/j.gde.2009.10.005

Hwang, M. G., and Cho, M. H. (2007). *Arabidopsis thaliana* telomeric DNA-binding protein 1 is required for telomere length homeostasis and its

Myb-extension domain stabilizes plant telomeric DNA binding. *Nucleic Acids Res.* 35, 1333–1342. doi: 10.1093/nar/gkm043

Ielasi, F. S., Decanniere, K., and Willaert, R. G. (2012). The epithelial adhesin 1 (Epa1p) from the human-pathogenic yeast *Candida glabrata*: structural and functional study of the carbohydrate-binding domain. *Acta Crystallogr. D Biol. Crystallogr.* 68, 210–217. doi: 10.1107/S0907444911054898

Jacob, N. K., Lescasse, R., Linger, B. R., and Price, C. M. (2007). Tetrahymena POT1a regulates telomere length and prevents activation of a cell cycle checkpoint. *Mol. Cell Biol.* 27, 1592–1601. doi: 10.1128/MCB.01975-06

Jiang, P. X., Wang, J., Feng, Y., and He, Z. G. (2007). Divergent functions of multiple eukaryote-like Orc1/Cdc6 proteins on modulating the loading of the MCM helicase onto the origins of the hyperthermophilic archaeon *Sulfolobus solfataricus* P2. *Biochem. Biophys. Res. Commun.* 361, 651–658. doi: 10.1016/j.bbrc.2007.07.073

Jin, L., Gaus, C., and Escher, B. I. (2015). Adaptive stress response pathways induced by environmental mixtures of bioaccumulative chemicals in dugongs. *Environ. Sci. Technol.* 49, 6963–6973. doi: 10.1021/acs.est.5b00947

Kabir, S., Hockemeyer, D., and de Lange, T. (2014). TALEN gene knockouts reveal no requirement for the conserved human shelterin protein Rap1 in telomere protection and length regulation. *Cell Rep.* 9, 1273–1280. doi: 10.1016/j.celrep.2014.10.014

Kazda, A., Zellinger, B., Rossler, M., Derboven, E., Kusenda, B., and Riha, K. (2012). Chromosome end protection by blunt-ended telomeres. *Genes Dev.* 26, 1703–1713. doi: 10.1101/gad.194944.112

Kitada, T., Kuryan, B. G., Tran, N. N., Song, C., Xue, Y., Carey, M., et al. (2012). Mechanism for epigenetic variegation of gene expression at yeast telomeric heterochromatin. *Genes Dev.* 26, 2443–2455. doi: 10.1101/gad.201095.112

Koonin, E. V. (2005). Paralogs and mutational robustness linked through transcriptional reprogramming. *Bioessays* 27, 865–868. doi: 10.1002/bies.20296

Levy, M. Z., Allsopp, R. C., Futcher, A. B., Greider, C. W., and Harley, C. B. (1992). Telomere end-replication problem and cell aging. *J. Mol. Biol.* 225, 951–960. doi: 10.1016/0022-2836(92)90096-3

Lew, J. E., Enomoto, S., and Berman, J. (1998). Telomere length regulation and telomeric chromatin require the nonsense-mediated mRNA decay pathway. *Mol. Cell Biol.* 18, 6121–6130. doi: 10.1128/MCB.18.10.6121

Li, B., and de Lange, T. (2003). Rap1 affects the length and heterogeneity of human telomeres. *Mol. Biol. Cell* 14, 5060–5068. doi: 10.1091/mbc.E03-06-0403

Li, J., Yang, F., Zhu, J., He, S., and Li, L. (2009). Characterization of a tandemly repeated subtelomeric sequence with inverted telomere repeats in maize. *Genome* 52, 286–293. doi: 10.1139/G09-005

Liaw, H., and Lustig, A. J. (2006). Sir3 C-terminal domain involvement in the initiation and spreading of heterochromatin. *Mol. Cell Biol.* 26, 7616–7631. doi: 10.1128/MCB.01082-06

Linger, B. R., and Price, C. M. (2009). Conservation of telomere protein complexes: shuffling through evolution. *Crit. Rev. Biochem. Mol. Biol.* 44, 434–446. doi: 10.3109/10409230903307329

Lopez de Silanes, I., Grana, O., De Bonis, M. L., Dominguez, O., Pisano, D. G., and Blasco, M. A. (2014). Identification of TERRA locus unveils a telomere protection role through association to nearly all chromosomes. *Nat. Commun.* 5:723. doi: 10.1038/ncomms5723

Louis, E. J., and Haber, J. E. (1990). Mitotic recombination among subtelomeric Y' repeats in *Saccharomyces cerevisiae*. *Genetics* 124, 547–559.

Lu, J., and Liu, Y. (2010). Deletion of Ogg1 DNA glycosylase results in telomere base damage and length alteration in yeast. *EMBO J.* 29, 398–409. doi: 10.1038/emboj.2009.355

Luke, B., Azzalin, C. M., Hug, N., Deplazes, A., Peter, M., and Lingner, J. (2007). *Saccharomyces cerevisiae* Ebs1p is a putative ortholog of human Smg7 and promotes nonsense-mediated mRNA decay. *Nucleic Acids Res.* 35, 7688–7697. doi: 10.1093/nar/gkm912

Luke-Glaser, S., Poschke, H., and Luke, B. (2012). Getting in (and out of) the loop: regulating higher order telomere structures. *Front. Oncol.* 2:180. doi: 10.3389/fonc.2012.00180

Lushchak, V. I. (2011). Adaptive response to oxidative stress: bacteria, fungi, plants and animals. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* 153, 175–190. doi: 10.1016/j.cbpc.2010.10.004

Lustig, A. J. (1992). Hoogsteen G-G base pairing is dispensable for telomere healing in yeast. *Nucleic Acids Res.* 20, 3021–3028. doi: 10.1093/nar/20.12.3021

Lustig, A. J. (1998). Mechanisms of silencing in *Saccharomyces cerevisiae*. *Curr. Opin. Genet. Dev.* 8, 233–239. doi: 10.1016/S0959-437X(98) 80043-9

Lustig, A. J. (2003). Clues to catastrophic telomere loss in mammals from yeast telomere rapid deletion. *Nat. Rev. Genet.* 4, 916–923. doi: 10.1038/nrg1207

Lustig, A. J., and Petes, T. D. (1986). Identification of yeast mutants with altered telomere structure. *Proc. Natl. Acad. Sci. U.S.A.* 83, 1398–1402. doi: 10.1073/pnas.83.5.1398

Maicher, A., Lockhart, A., and Luke, B. (2014). Breaking new ground: digging into TERRA function. *Biochim. Biophys. Acta* 1839, 387–394. doi: 10.1016/j.bbagrm.2014.03.012

Martina, M., Clerici, M., Baldo, V., Bonetti, D., Lucchini, G., and Longhese, M. P. (2012). A balance between Tel1 and Rif2 activities regulates nucleolytic processing and elongation at telomeres. *Mol. Cell Biol.* 32, 1604–1617. doi: 10.1128/MCB.06547-11

Martinez, P., Gomez-Lopez, G., Garcia, F., Mercken, E., Mitchell, S., Flores, J. M., et al. (2013). RAP1 protects from obesity through its extratelomeric role regulating gene expression. *Cell Rep.* 3, 2059–2074. doi: 10.1016/j.celrep.2013.05.030

Martinez, P., Thanasoula, M., Carlos, A. R., Gomez-Lopez, G., Tejera, A. M., Schoeftner, S., et al. (2010). Mammalian Rap1 controls telomere function and gene expression through binding to telomeric and extratelomeric sites. *Nat. Cell Biol.* 12, 768–780. doi: 10.1038/ncb2081

McClintock, B. (1942). The fusion of broken ends of chromosomes following nuclear fusion. *Proc. Natl. Acad. Sci. U.S.A.* 28, 458–463. doi: 10.1073/pnas.28.11.458

Meena, J. K., Cerutti, A., Beichler, C., Morita, Y., Bruhn, C., Kumar, M., et al. (2015). Telomerase abrogates aneuploidy-induced telomere replication stress, senescence and cell depletion. *EMBO J.* 34, 1371–1384. doi: 10.15252/embj.201490070

Miyake, Y., Nakamura, M., Nabetani, A., Shimamura, S., Tamura, M., Yonehara, S., et al. (2009). RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway. *Mol. Cell* 36, 193–206. doi: 10.1016/j.molcel.2009.08.009

Mizuno, H., Wu, J., Katayose, Y., Kanamori, H., Sasaki, T., and Matsumoto, T. (2008). Chromosome-specific distribution of nucleotide substitutions in telomeric repeats of rice (*Oryza sativa* L.). *Mol. Biol. Evol.* 25, 62–68. doi: 10.1093/molbev/msm227

Moye, A. L., Porter, K. C., Cohen, S. B., Phan, T., Zyner, K. G., Sasaki, N., et al. (2015). Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.* 6:7643. doi: 10.1038/ncomms8643

Nautiyal, S., DeRisi, J. L., and Blackburn, E. H. (2002). The genome-wide expression response to telomerase deletion in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9316–9321. doi: 10.1073/pnas.142162499

O'Connor, M. S., Safari, A., Liu, D., Qin, J., and Songyang, Z. (2004). The human Rap1 protein complex and modulation of telomere length. *J. Biol. Chem.* 279, 28585–28591. doi: 10.1074/jbc.M312913200

Oganesian, L., Moon, I. K., Bryan, T. M., and Jarstfer, M. B. (2006). Extension of G-quadruplex DNA by ciliate telomerase. *EMBO J.* 25, 1148–1159. doi: 10.1038/sj.emboj.7601006

Oppikofer, M., Kueng, S., Keusch, J. J., Hassler, M., Ladurner, A. G., Gut, H., et al. (2013). Dimerization of Sir3 via its C-terminal winged helix domain is essential for yeast heterochromatin formation. *EMBO J.* 32, 437–449. doi: 10.1038/emboj.2012.343

Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., and Lipps, H. J. (2005). Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.* 12, 847–854. doi: 10.1038/nsmb982

Pardo, B., and Marcand, S. (2005). Rap1 prevents telomere fusions by nonhomologous end joining. *EMBO J.* 24, 3117–3127. doi: 10.1038/sj.emboj.7600778

Park, M. J., Jang, Y. K., Choi, E. S., Kim, H. S., and Park, S. D. (2002). Fission yeast Rap1 homolog is a telomere-specific silencing factor and interacts with Taz1p. *Mol. Cells* 13, 327–333.

Peace, J. M., Ter-Zakarian, A., and Aparicio, O. M. (2014). Rif1 regulates initiation timing of late replication origins throughout the *S. cerevisiae* genome. *PLoS ONE* 9:e98501. doi: 10.1371/journal.pone.0098501

Petrov, A. V., Dokudovskaya, S. S., Sokolov, K. A., Lavrik, O. I., Favre, A., Dontsova, O. A., et al. (1998). Telomerase from *Saccharomyces cerevisiae*

contains several protein subunits and may have different activities depending on the protein content. *FEBS Lett.* 436, 35–40. doi: 10.1016/S0014-5793(98)01 091-6

Pickett, H. A., Henson, J. D., Au, A. Y., Neumann, A. A., and Reddel, R. R. (2011). Normal mammalian cells negatively regulate telomere length by telomere trimming. *Hum. Mol. Genet.* 20, 4684–4692. doi: 10.1093/hmg/ddr402

Pickett, H. A., and Reddel, R. R. (2012). The role of telomere trimming in normal telomere length dynamics. *Cell Cycle* 11, 1309–1315. doi: 10.4161/cc.19632

Polotnianka, R. M., Li, J., and Lustig, A. J. (1998). The yeast Ku heterodimer is essential for protection of the telomere against nucleolytic and recombinational activities. *Curr. Biol.* 8, 831–834. doi: 10.1016/S0960-9822(98)70325-2

Prasanth, S. G., Shen, Z., Prasanth, K. V., and Stillman, B. (2010). Human origin recognition complex is essential for HP1 binding to chromatin and heterochromatin organization. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15093–15098. doi: 10.1073/pnas.1009945107

Preiszner, J., Takacs, I., Bilgin, M., Gyorgyey, J., Dudits, D., and Feher, A. (1994). Organization of a Solanum brevidens repetitive sequence related to the TGRI subtelomeric repeats of Lycopersicon esculentum. *Theor. Appl. Genet.* 89, 1–8. doi: 10.1007/BF00226974

Price, C. M., Boltz, K. A., Chaiken, M. F., Stewart, J. A., Beilstein, M. A., and Shippen, D. E. (2010). Evolution of CST function in telomere maintenance. *Cell Cycle* 9, 3157–3165. doi: 10.4161/cc.9.16.12547

Ribeyre, C., and Shore, D. (2012). Anticheckpoint pathways at telomeres in yeast. *Nat. Struct. Mol. Biol.* 19, 307–313. doi: 10.1038/nsmb.2225

Rosas-Hernandez, L. L., Juarez-Reyes, A., Arroyo-Helguera, O. E., De Las Penas, A., Pan, S. J., Cormack, B. P., et al. (2008). yKu70/yKu80 and Rif1 regulate silencing differentially at telomeres in *Candida glabrata*. *Eukaryot. Cell* 7, 2168–2178. doi: 10.1128/EC.00228-08

Sabourin, M., Tuzon, C. T., and Zakian, V. A. (2007). Telomerase and Tel1p preferentially associate with short telomeres in *S. cerevisiae*. *Mol. Cell* 27, 550–561. doi: 10.1016/j.molcel.2007.07.016

Sealey, D. C., Kostic, A. D., LeBel, C., Pryde, F., and Harrington, L. (2011). The TPR-containing domain within Est1 homologs exhibits species-specific roles in telomerase interaction and telomere length homeostasis. *BMC Mol. Biol.* 12:45. doi: 10.1186/1471-2199-12-45

Sfeir, A., Kabir, S., van Overbeek, M., Celli, G. B., and de Lange, T. (2010). Loss of Rap1 induces telomere recombination in the absence of NHEJ or a DNA damage signal. *Science* 327, 1657–1661. doi: 10.1126/science.1185100

Shakirov, E. V., Song, X., Joseph, J. A., and Shippen, D. E. (2009). POT1 proteins in green algae and land plants: DNA-binding properties and evidence of co-evolution with telomeric DNA. *Nucleic Acids Res.* 37, 7455–7467. doi: 10.1093/nar/gkp785

Shor, E., Fox, C. A., and Broach, J. R. (2013). The yeast environmental stress response regulates mutagenesis induced by proteotoxic stress. *PLoS Genet.* 9:e1003680. doi: 10.1371/journal.pgen.1003680

Shore, D. (1994). RAP1: a protean regulator in yeast. *Trends Genet.* 10, 408–412. doi: 10.1016/0168-9525(94)90058-2

Smith, J. S., Chen, Q., Yatsunyk, L. A., Nicoludis, J. M., Garcia, M. S., Kranaster, R., et al. (2011). Rudimentary G-quadruplex-based telomere capping in *Saccharomyces cerevisiae*. *Nat. Struct. Mol. Biol.* 18, 478–485. doi: 10.1038/nsmb.2033

Soria, P. S., McGary, K. L., and Rokas, A. (2014). Functional divergence for every paralog. *Mol. Biol. Evol.* 31, 984–992. doi: 10.1093/molbev/msu050

Sreesankar, E., Senthilkumar, R., Bharathi, V., Mishra, R. K., and Mishra, K. (2012). Functional diversification of yeast telomere associated protein, Rif1, in higher eukaryotes. *BMC Genomics* 13:255. doi: 10.1186/1471-2164-13-255

Sridhar, A., Kediora, S., and Donaldson, A. D. (2014). At short telomeres Tel1 directs early replication and phosphorylates Rif1. *PLoS Genet.* 10:e1004691. doi: 10.1371/journal.pgen.1004691

Sugiyama, T., Cam, H., Verdel, A., Moazed, D., and Grewal, S. I. (2005). RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proc. Natl. Acad. Sci. U.S.A.* 102, 152–157. doi: 10.1073/pnas.0407641102

Sun, J., Yu, E. Y., Yang, Y., Confer, L. A., Sun, S. H., Wan, K., et al. (2009). Stn1-Ten1 is an Rpa2-Rpa3-like complex at telomeres. *Genes Dev.* 23, 2900–2914. doi: 10.1101/gad.1851909

Thomae, A. W., Baltin, J., Pich, D., Deutsch, M. J., Ravasz, M., Zeller, K., et al. (2011). Different roles of the human Orc6 protein in the replication initiation process. *Cell Mol. Life. Sci.* 68, 3741–3756. doi: 10.1007/s00018-011-0675-9

Tomaska, L., McEachern, M. J., and Nosek, J. (2004a). Alternatives to telomerase: keeping linear chromosomes via telomeric circles. *FEBS Lett.* 567, 142–146. doi: 10.1016/j.febslet.2004.04.058

Tomaska, L., Willcox, S., Slezakova, J., Nosek, J., and Griffith, J. D. (2004b). Taz1 binding to a fission yeast model telomere: formation of telomeric loops and higher order structures. *J. Biol. Chem.* 279, 50764–50772. doi: 10.1074/jbc.M409790200

Tomaska, L., Nosek, J., Kramara, J., and Griffith, J. D. (2009). Telomeric circles: universal players in telomere maintenance? *Nat. Struct. Mol. Biol.* 16, 1010–1015. doi: 10.1038/nsmb.1660

Torres, G. A., Gong, Z., Iovene, M., Hirsch, C. D., Buell, C. R., Bryan, G. J., et al. (2011). Organization and evolution of subtelomeric satellite repeats in the potato genome. *G3* (*Bethesda*) 1, 85–92. doi: 10.1534/g3.111.000125

Tucey, T. M., and Lundblad, V. (2014). Regulated assembly and disassembly of the yeast telomerase quaternary complex. *Genes Dev.* 28, 2077–2089. doi: 10.1101/gad.246256.114

Usui, T., Ogawa, H., and Petrini, J. H. (2001). A DNA damage response pathway controlled by Tel1 and the Mre11 complex. *Mol. Cell* 7, 1255–1266. doi: 10.1016/S1097-2765(01)00270-2

Villasante, A., Abad, J. P., and Mendez-Lago, M. (2007). Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10542–10547. doi: 10.1073/pnas.0703808104

Villasante, A., de Pablos, B., Mendez-Lago, M., and Abad, J. P. (2008). Telomere maintenance in *Drosophila*: rapid transposon evolution at chromosome ends. *Cell Cycle* 7, 2134–2138. doi: 10.4161/cc.7.14.6275

Vogan, J. M., and Collins, K. (2015). Dynamics of human telomerase holoenzyme assembly and subunit exchange across the cell cycle. *J. Biol. Chem.* 290, 21320–21335. doi: 10.1074/jbc.M115.659359

Wang, B., Feng, L., Hu, Y., Huang, S. H., Reynolds, C. P., Wu, L., et al. (1999). The essential role of *Saccharomyces cerevisiae* CDC6 nucleotide-binding site in cell growth, DNA synthesis, and Orc1 association. *J. Biol. Chem.* 274, 8291–8298. doi: 10.1074/jbc.274.12.8291

Wang, F., Stewart, J. A., Kasbek, C., Zhao, Y., Wright, W. E., and Price, C. M. (2012). Human CST has independent functions during telomere duplex replication and C-strand fill-in. *Cell Rep.* 2, 1096–1103. doi: 10.1016/j.celrep.2012.10.007

Wellinger, R. J. (2009). The CST complex and telomere maintenance: the exception becomes the rule. *Mol. Cell* 36, 168–169. doi: 10.1016/j.molcel.2009.10.001

Williams, B., and Lustig, A. J. (2003). The paradoxical relationship between NHEJ and telomeric fusion. *Mol. Cell* 11, 1125–1126.

Witkin, E. M. (1991). RecA protein in the SOS response: milestones and mysteries. *Biochimie* 73, 133–141. doi: 10.1016/0300-9084(91)90196-8

Wright, J. H., Gottschling, D. E., and Zakian, V. A. (1992). *Saccharomyces* telomeres assume a non-nucleosomal chromatin structure. *Genes Dev.* 6, 197–210. doi: 10.1101/gad.6.2.197

Xin, H., Liu, D., Wan, M., Safari, A., Kim, H., Sun, W., et al. (2007). TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase. *Nature* 445, 559–562. doi: 10.1038/nature05469

Xu, Y. (2012). Human telomere RNA: a potential target for ligand recognition. *Curr. Pharm. Des.* 18, 2096–2101. doi: 10.2174/1381612127999 58378

Xu, Y., and Komiyama, M. (2012). Structure, function and targeting of human telomere RNA. *Methods* 57, 100–105. doi: 10.1016/j.ymeth.2012. 02.015

Yanez-Carrillo, P., Robledo-Marquez, K. A., Ramirez-Zavaleta, C. Y., De Las Penas, A., and Castano, I. (2014). The mating type-like loci of *Candida glabrata*. *Rev. Iberoam. Micol.* 31, 30–34. doi: 10.1016/j.riam.2013.09.016

Yang, D., He, Q., Kim, H., Ma, W., and Songyang, Z. (2011). TIN2 protein dyskeratosis congenita missense mutants are defective in association with telomerase. *J. Biol. Chem.* 286, 23022–23030. doi: 10.1074/jbc.M111. 225870

Yeung, F., Ramirez, C. M., Mateos-Gomez, P. A., Pinzaru, A., Ceccarini, G., Kabir, S., et al. (2013). Nontelomeric role for Rap1 in regulating metabolism and protecting against obesity. *Cell Rep.* 3, 1847–1856. doi: 10.1016/j.celrep.2013.05.032

Zhou, J., Hidaka, K., and Futcher, B. (2000). The Est1 subunit of yeast telomerase binds the Tlc1 telomerase RNA. *Mol. Cell Biol.* 20, 1947–1955. doi: 10.1128/MCB.20.6.1947-1955.2000

Insertion of Retrotransposons at Chromosome Ends: Adaptive Response to Chromosome Maintenance

Geraldine Servant and Prescott L. Deininger*

Tulane Cancer Center, Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA

The telomerase complex is a specialized reverse transcriptase (RT) that inserts tandem DNA arrays at the linear chromosome ends and contributes to the protection of the genetic information in eukaryotic genomes. Telomerases are phylogenetically related to retrotransposons, encoding also the RT activity required for the amplification of their sequences throughout the genome. Intriguingly the telomerase gene is lost from the *Drosophila* genome and tandem retrotransposons replace telomeric sequences at the chromosome extremities. This observation suggests the versatility of RT activity in counteracting the chromosome shortening associated with genome replication and that retrotransposons can provide this activity in case of a dysfunctional telomerase. In this review paper, we describe the major classes of retroelements present in eukaryotic genomes in order to point out the differences and similarities with the telomerase complex. In a second part, we discuss the insertion of retroelements at the ends of chromosomes as an adaptive response for dysfunctional telomeres.

INTRODUCTION

In eukaryotic genomes, reverse transcriptase (RT) activity that leads to the synthesis of complementary DNA (cDNA) using an RNA template, is provided by two types of genetic elements, the telomerase gene and retroelements, also called retrotransposons. The telomerase reverse-transcribes a specific RNA template on to linear DNA ends to prevent the chromosome shortening caused by the replication mechanism (Blackburn, 1992). This is the first step of the formation of the complex nucleoprotein structures, the telomeres that cap and protect the chromosome ends (Muller, 1938; McClintock, 1941; Blackburn, 1992). Retrotransposons are mobile genetic elements that amplify their sequences throughout genomes, using an RNA intermediate and based on a "copy and paste" mechanism, termed retrotransposition (Boeke et al., 1985). Because these two genetic elements contain the same enzymatic activity and show some sequence similarity, it has been proposed that the telomerase complex has evolved from an ancestor retroelement and specialized to add nucleotides to the linear chromosome ends (**Figure 1**; Eickbush, 1997; Nakamura and Cech, 1998). The phylogenetic linkage between telomerases and retroelements has been reinforced by the identification of a group of retrotransposons, the Penelope-like elements, encoding a RT closely related to the telomerase enzyme (Arkhipova et al., 2003).

FIGURE 1 | Proposed model for the evolution of telomere elongation in eukaryotic genomes. Both telomerases and retrotransposons derived from an ancestor retroelement. Mutations of the telomerase complex or protein-associated telomeres, inactivating the telomerase function, cause a shortening of telomeres. Critically short telomeres induce cell cycle arrest that can lead to cell death. Some cells survive to dysfunctional telomerase because of the formation of alternative telomere structures, generated by either homologous recombination mechanism or an adaptive response involving the activation of retrotransposition and *de novo* inserts at the chromosome ends.

Retroelements have extensively colonized almost all eukaryotic organisms. For instance, 3% of the genome of the yeast *Saccharomyces cerevisiae* is made of retrotransposons (Kim et al., 1998). Retrotransposons also represent around 42, 37, and 3.6% of the genome of human, mouse and *Drosophila melanogaster,* respectively (Adams et al., 2000; Lander et al., 2001; Waterston et al., 2002). Because of their mobility and their high copy number, retrotransposons can generate gene disruption at the insertion site or cause genomic rearrangement by non-allelic homologous recombination. Therefore, they play an important role in the genome plasticity and they have a great impact on the architecture and evolution of eukaryotic genomes. In order for the elements to coexist with the cells, different strategies have been established to limit the damage caused by retrotransposition, including silencing of the elements (Hata and Sakaki, 1997; Bourc'His and Bestor, 2004) and destabilization of the new copies during the reverse transcription process by DNA repair proteins (Lee et al., 1998; Bryk et al., 2001; Gasior et al., 2008). A very efficient strategy to control the copy number in the genome is to direct the insertion in fairly safe regions, poor in genes, for example in heterochromatin or at telomeres (Okazaki et al., 1995; Zou et al., 1996; Takahashi et al., 1997).

Noteworthy in *Drosophila*, retrotransposons guarantee the protection of the chromosome ends because the telomerase

is absent, probably lost during evolution (Biessmann et al., 1990). This observation suggests that RT activity is necessary to assure the function of protection of the linear chromosome ends and that retroelements could provide this activity in case of a dysfunctional telomerase. In fact, either activation of retrotransposition or integration of retroelements at telomeres has been reported in cells that survive a mutation in the telomere function (Scholes et al., 2003; Morrish et al., 2007). It has been proposed that this process is an adaptive mechanism to maintain the chromosome ends (**Figure 1**). In this review paper, we discuss the insertion of retrotransposons at telomeres.

RETROTRANSPOSONS AND THE TELOMERASE COMPLEX

There are two major classes of retroelements: the long terminal repeat (LTR) retrotransposons, also called retrovirus-like elements, and the non-LTR retrotransposons. They are distinguishable based on structural features and the mechanism of retrotransposition.

LTR-Retrotransposons

Long terminal repeat elements share similarities of structure and mechanism of replication with retroviruses. However, LTR-retrotransposons do not have a functional *env* gene, coding for a protein involved in cellular membrane recognition and cell invasion. Therefore LTR-retrotransposons are trapped in cells and are not able to escape or infect other cells. The best described elements are ZAM and Idefix of *Drosophila*, Ty retrotransposon in yeast *S. cerevisiae*, and IAP in mouse (for review Morgan et al., 1999; Prudhomme et al., 2005; Curcio et al., 2015; Mager and Stoye, 2015; Sandmeyer et al., 2015).

Structure

Long terminal repeat-retrotransposons are flanked by LTRs, containing regulatory elements. These LTRs flank one or two open reading frames (ORFs), generally encoding GAG and POL proteins (**Figure 2A**). *GAG* and *POL* can be fused, as in the Ty5 element of *S. cerevisiae* (Zou et al., 1996; Neuveglise et al., 2002). Other LTR-retrotransposons contain two ORFs, either separated by a stop codon as in Tca2 of *Candida albicans* (Matthews et al., 1997; Neuveglise et al., 2002) or a frameshift as in Ty1 and Ty3 elements of *S. cerevisiae* (Clare et al., 1988; Neuveglise et al., 2002). As a consequence, both proteins are produced at different levels. GAG protein, the more abundant, is a structural protein that forms the virus-like particle (VLP). POL protein contains the protease (PR), RT associated with RNase H (RT/RH), and integrase (IN) activities. The organization of the domains in the POL protein is used for further classification of the LTR-elements in the two subfamilies, copia-Ty1 (PR-IN-RT/RH) and gypsy-Ty3 (PR-RT/RH-IN). LTRs possess the signals of initiation and termination of RNA polymerase (RNA pol) II transcription.
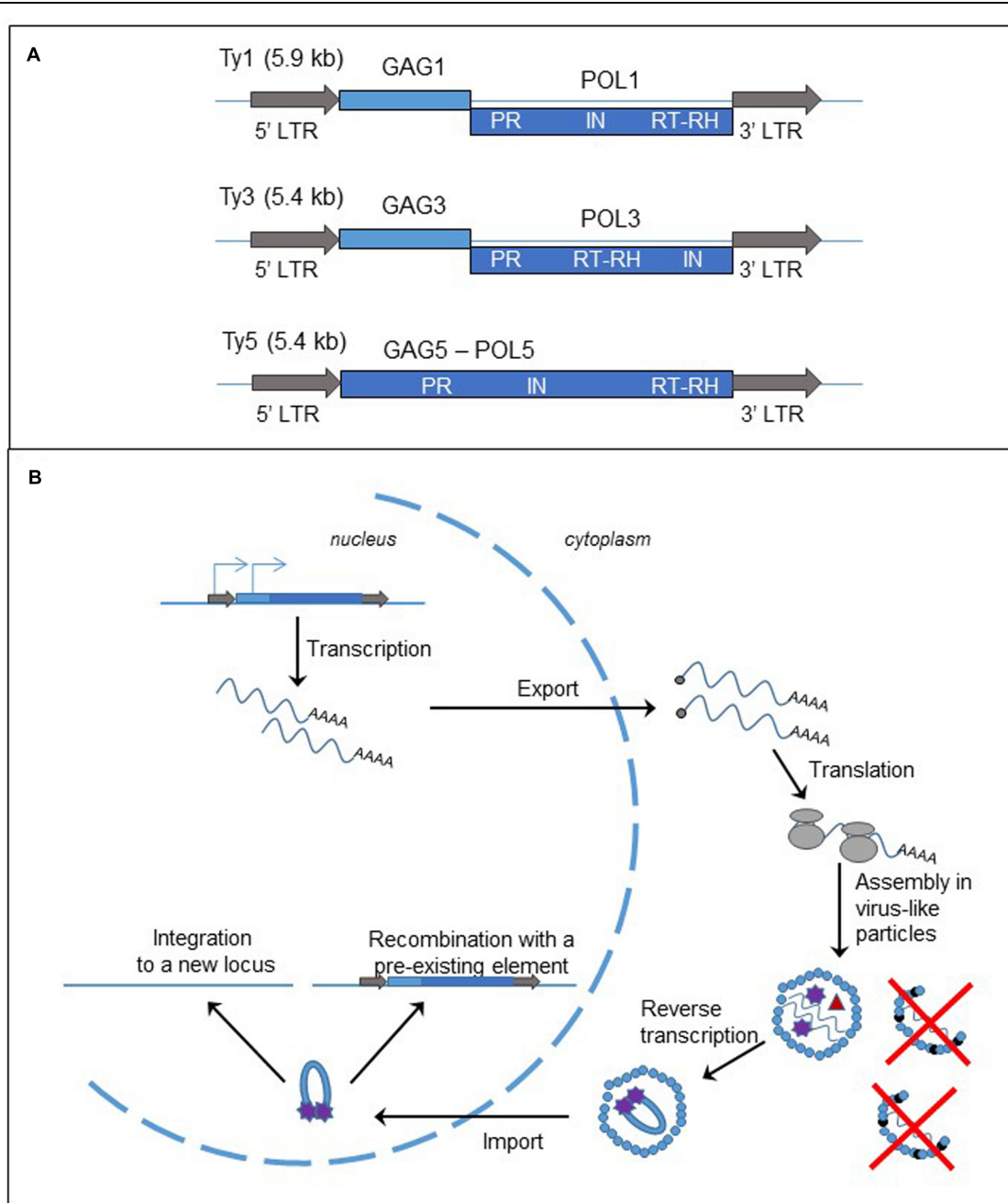
FIGURE 2 | Long terminal repeat retrotransposons, structure and replication cycle. (A) *Genomic organization of the yeast* Saccharomyces cerevisiae *retrotransposons, Ty1, Ty3, and Ty5. The gray arrows represent the LTRs; the light and dark blue boxes are the ORFs, GAG and POL, fused (Ty5) or separated by a frameshift (Ty1 and Ty3). LTR: long terminal repeat; PR, protease; IN, integrase; RT, reverse transcriptase; RH, RNase H.* **(B)** *Cycle of retrotransposition of LTR retrotransposons. The straight blue lines are the DNA strands. The light and dark blue boxes represent the two ORFs, GAG and POL, of the LTR-retrotransposon. The gray arrows are the LTRs flanking the two ORFs. The blue arrows on the left LTR and GAG represent the two initiation sites of the transcription of the element. The wavy blue lines represent mRNA of the element and the black dots at the left end is the cap. The gray circles are the ribosomes. The small blue circles represent GAGp and are organized in the VLP. The small black circles represent p22, the peptide responsible for Ty1 copy number control phenotype (destabilization of the VLPs). Inside the VLP the red triangle represent the reverse transcriptase, and the purple stars are the integrase.*

Retrotransposition Cycle

As described in **Figure 2B** and in several reviews (Curcio et al., 2015; Sandmeyer et al., 2015), the replication of LTR-retrotransposon starts with the transcription of a bicistronic RNA in the nucleus. The RNA is capped, polyadenylated and exported into the cytoplasm. Translation produces either a GAG protein or GAG-POL polyprotein. The polyprotein is processed by the protease encoded in the PR domain and the proteins are

associated with two RNA molecules to form the VLP. A tRNA is also encapsulated in the VLP and serves as a primer for the synthesis of the cDNA. The reverse transcription occurs in the cytoplasm inside the VLP. Then the complex cDNA – integrase is imported into the nucleus. There are two mechanisms for the insertion of the new copy of retrotransposon to the genome. First the cDNA can be integrated to a new locus by the integrase activity. Second it can recombine with a pre-existing element through the homologous recombination process.

Ty1 retrotransposition and expression are controlled by Ty1 copy number (Jiang, 2002; Garfinkel et al., 2003) through an original mechanism that has been deciphered recently. The RNA interference pathway limits many retrotransposons, but the budding yeast does not have the machinery. Ty1 copy number is instead limited by a peptide, p22, expressed from a shorter and alternative Ty1 transcript and corresponding to the C-terminal domain of the GAG protein (Nishida et al., 2015; Saha et al., 2015; Tucker et al., 2015). The peptide interacts with the GAG protein, inhibiting its function, and destabilizes the VLP, leading to the decrease in the retrotransposition frequency and the alteration of stability or maturation of Ty1 proteins (**Figure 2B**).

Endogenous Retroviruses

The endogenous retroviruses (ERV) are also classified as LTR-retrotransposons. As the name suggests, they are remnants of ancient retroviruses that have infected the germinal cells of an ancestor organism and lost the ability to escape the cells. ERVs make up 8% of human genome but they are not currently active (Lander et al., 2001). Too many mutations have accumulated in their sequences, rendering the elements unable to retrotranspose. Some human ERVs can still express proteins and have a significant role in the cellular metabolism, such as the syncytin, a protein specifically expressed in placenta from a degenerated ERV and has an important role in the formation of the syncytiotrophoblast, a tissue that allow exchanges between the mother and the embryo (Heidmann et al., 2009; Lavialle et al., 2013).

Non-LTR Retrotransposons

Non-LTR retrotransposons predominate in mammalian cells. In the human genome, the elements L1 and Alu are the most abundant and active mobile DNA species and constitute 17 and 11% of genome, respectively (Lander et al., 2001; de Koning et al., 2011). L1 is a long interspersed element (LINE) and encodes the activities required for its own retrotransposition. Alu element is a non-autonomous element, also called short interspersed element (SINE), and its replication relies on L1 protein expression.

Non-LTR retrotransposons represent a very broad group of retroelements, showing different features such as target-site specificity, enzymatic activities required for retrotransposition, or ORF number (Eickbush and Malik, 2002). In the present paper, we primarily focus on two model elements, the human L1 and Alu elements, in order to point out the differences with LTR-retrotransposons (**Figure 3**) and similarities and differences relative to the telomerase complex (for review Richardson et al., 2015).

Structure

The human genome contains about 500,000 copies of L1 elements (Lander et al., 2001). Out of them, only 6,000 are full-length, 6-kb long, and the others are generally 5′ truncated. L1 element consists of a 5′ untranslated region (UTR), two ORFs (ORF1 and ORF2), and a 3′UTR (**Figure 3A**). Inserts are flanked by target site duplications generated from the target site due to the mechanism of retrotransposition. ORF2 encodes the endonuclease (EN) and RT activities required to insert a new copy of the element to the genome (Mathias et al., 1991; Feng et al., 1996). In contrast, the function of ORF1 protein (ORF1p) is mostly unknown. However, ORF1p contains a nucleic acid binding domain, a chaperone activity, and a nucleolar localization signal (for review Martin, 2010). Both L1-encoded proteins are required for the mobility of autonomous elements (Moran et al., 1996). The L1 5′ UTR includes a RNA pol II promoter that assures the transcription of the element (Swergold, 1990; Severynse et al., 1992) and an antisense promoter (Speek, 2001). Recently, a third ORF, ORF0, has been discovered in the 5′ UTR of primate-specific L1 elements, expressed from an antisense promoter similar to the one previously described (Denli et al., 2015). The function of the protein still needs to be characterized but it seems that ORF0p modestly stimulates L1 retrotransposition. The L1 3′ UTR has a polyadenylation signal that is probably weak because some new L1 inserts include sequences from downstream of the original L1 elements (Moran et al., 1999). The process seems to be very frequent in cancer cells (Tubio et al., 2014). The L1 insert sequence ends with a poly (A) tail, a structure important for an efficient retrotransposition cycle (Moran et al., 1996; Doucet et al., 2015).

Alu elements, a 300 bp long, primate specific SINE, are related to 7SL RNA, the signal recognition particle (SRP) RNA (Quentin, 1992). They contain an internal promoter that allows them to be transcribed by the RNA pol III machinery. Alu inserts are flanked by TSDs and end with a poly (A) tail (**Figure 3A**). The presence of these structures, also important markers of L1 retrotransposition, supports the hypothesis that Alu elements share the same machinery as the L1 retrotransposon. However, enough differences in timing and factors influencing Alu retrotransposition, differentially from L1, indicate that their pathways diverge in many ways (Deininger and Batzer, 2002; Dewannieux et al., 2003; Wagstaff et al., 2013).

Retrotransposition Cycle

Based on the difference in the structure of the two groups of retroelements, it is not surprising that the elements do not share the same mechanism of retrotransposition. The main difference resides in the cellular location of the reverse transcription, occurring inside the VLP in the cytoplasm for LTR-retrotransposons and at the insertion site in the nucleus for LINEs and SINEs.

Briefly and as described in **Figure 3B** (for review Richardson et al., 2015), L1 mRNA, produced from the L1 promoter found within the 5′ UTR, is capped, polyadenylated and exported to the cytoplasm. L1 mRNA is translated into ORF1p and ORF2p as a bicistronic RNA. The proteins assemble with mRNA to
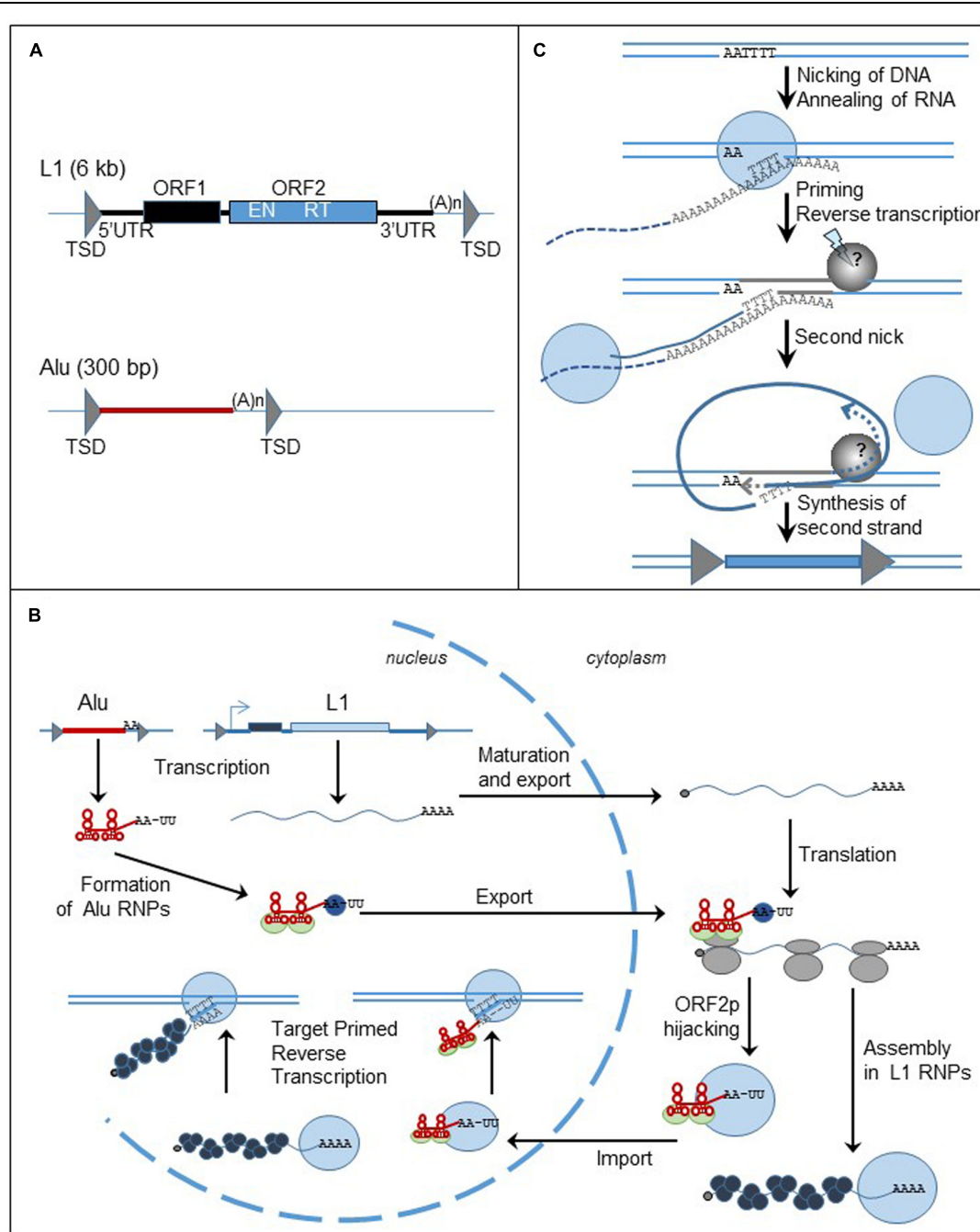
FIGURE 3 | Non-LTR retrotransposons, structure and replication cycle. (A) Genomic organization of L1 and Alu elements. Triangles represent TSDs; black and blue boxed are the ORFs. UTR: untranslated region; TSD: target site duplication; ORF: open reading frame; (A)n: poly (A) tail; EN: endonuclease; RT: reverse transcriptase. **(B)** Cycle of retrotransposition of L1 and Alu elements. The straight lines are the DNA strands. Black and blue boxes represent ORF1 and ORF2 of L1 retrotransposon. The red box represent Alu element. The gray triangles flanking the boxes are the TSDs. The wavy blue lines represent L1 mRNA and the black dots at the left extremity is the cap. Alu RNA is represented by the red line. Attached to the red line, the light green circles are the SRP9/14 protein complex, the blue circles are PABP. The gray circles are the ribosomes. The blue circles represent ORF2p and the black circles represent ORF1p. **(C)** Mechanism of insertion of L1 element in the genome, the TPRT process. The lines are the DNA strands; the dashed lines are the RNA template. Blue circles represent ORF2p; the gray circle is the unknown protein responsible for the formation of the second nick. Gray triangles represent the TDS. The blue box represent the new insert.

form ribonucleoprotein (RNP) particles. It is not clear if the whole RNP is imported to the nucleus, but at least ORF2p and mRNA must enter into the nucleus. The reverse transcription of

the mRNA occurs in the nucleus at the target site of insertion through a mechanism called target-primed reverse transcription (TPRT) (**Figure 3C**). The ORF2-EN domain recognizes and

cleaves an AT-rich region. The T-rich DNA 3′ overhang anneals to the poly (A) tail of L1 mRNA and serves as a primer for the reverse transcription. The next steps of the mechanism are less characterized but a second nick is generated in order to finalize the insertion of the new copy of the element. The reverse transcription process can be interrupted before the synthesis of the full-length cDNA, generating a 5′ end-truncated element. Microhomologies with the genome are often found at the 5′ end of the truncated inserts suggesting that DNA repair machinery can disrupt the TPRT process (Zingler et al., 2005; Babushok et al., 2006).

The sequence analogy between Alu and 7SL RNA supports the hypothesis that Alu RNA can associate with the ribosomes. Similar to 7SL RNA, Alu RNA binds to the protein heterodimer SRP9/14, part of the SRP complex that binds to ribosomes and recognizes the signal peptide of secreted proteins during their translation (Hsu et al., 1995; Chang et al., 1996; Ahl et al., 2015). Therefore it has been proposed that the SRP9/14 complex could bring Alu RNA near the ribosomes and allow it to hijack L1 proteins during their synthesis (Dewannieux et al., 2003). Additionally, the length of the poly (A) stretch in Alu RNA is another important factor for the ability of Alu element to retrotranspose and it has been proposed that the poly (A) binding protein (PABP) may bind the poly (A) stretch and facilitate Alu RNA to associate with the translation machinery and then with L1 retrotransposition machinery (Roy-Engel et al., 2002; Dewannieux and Heidmann, 2005; Comeaux et al., 2009; Wagstaff et al., 2013). It seems that only ORF2p is really required for Alu mobility (Dewannieux et al., 2003), however, the presence of ORF1p seems to improve the efficiency of Alu retrotransposition (Wallace et al., 2008). Therefore L1 and Alu mobility are regulated differently.

The Telomerase Complex, a Stringent Retrotransposon

The mechanism of telomere elongation is very similar to the non-autonomous, non-LTR retrotransposition process. In fact, the telomerase complex is organized in a complex RNP containing notably the telomerase (a RT enzyme), and a specific RNA template (**Figure 4A**; Greider and Blackburn, 1989; Feng et al., 1995; Harrington et al., 1997; Kilian et al., 1997; Lingner et al., 1997; Meyerson et al., 1997). The two components are located at two different loci in the genome and their expression is not linked. This system correlates with the RNP of a retrotransposon, constituted by a SINE RNA, such as human Alu RNA, associated with the LINE retrotransposition machinery. However, the two RNA templates are different. First the telomerase RNA template, including hTR in the human genome, is transcribed by the RNA pol II machinery and processed (Feng et al., 1995; Zaug et al., 1996; Mitchell et al., 1999). Second, the telomeric RNA template seems to be highly specialized, consisting in several domains necessary for both the assembly of the telomerase complex and notably catalytic activation of the telomerase: the telomerase binding domain, the template sequence for reverse transcription of telomere repeats, the telomerase-associated protein binding domains (for review Egan and Collins, 2012).

The telomerase protein, hTERT in human contains the RT activity. In contrast to RT encoded by retroelements, telomerase RT exists in one copy in the genome (Meyerson et al., 1997; Bryce et al., 2000). In addition, the enzyme does not bind and reverse transcribe its own mRNA with *cis* preference as the L1-ORF2p (Mitchell and Collins, 2000). In fact, the telomerase becomes active only after binding the telomerase RNA template and it has been identified that specific structures of the human RNA template are required for the catalytic activation of the enzyme (Mitchell and Collins, 2000). The telomerase complex assembles in the nucleus in Cajal bodies (Etheridge et al., 2002; Yang et al., 2002; Zhu et al., 2004; Venteicher et al., 2009). The two major components of the telomerase complex are associated with several proteins with multiple roles (for review Blackburn and Collins, 2011). The function of these proteins is really wide and diversified, and consists in the formation of the RNP, the regulation of telomerase activity, the regulation of the complex access to telomeres, and also the RNA stability, maturation and location.

The similarity between non-LTR retrotransposons and the telomerase complex is not only limited to the RNP structure because the reverse transcription of telomerase RNA template at chromosome ends utilizes a mechanism comparable to the TPRT process (Boeke, 1997; Eickbush, 1997), the insertion mechanism of non-LTR retrotransposon cDNA to the genome (Greider and Blackburn, 1989; Yu et al., 1990). However, in the case of the telomerase, the enzyme does not nick the DNA to prime the reverse transcription, but instead uses the 3′ OH end of the linear DNA to prime the reverse transcription. The RNA template is not entirely reverse transcribed at telomeres, only a small part of it, which also has some similarity to SINE TPRT. The elongation of telomeres is cell cycle dependent, and occurs during S-phase, when telomeres are uncapped and DNA is accessible (Jády et al., 2006; Tomlinson et al., 2006).

The role of the telomerase complex is essential for the maintenance of the genetic material because it allows for the synthesis of the chromosome extremities that the DNA polymerase is unable to amplify. Without this activity, replication would lead to chromosome shortening that could cause genome instability, senescence or apoptosis (Hayflick, 1979; Lundblad and Szostak, 1989; Harley et al., 1990; Levy et al., 1992). In humans, dysfunctional telomerase leads to diseases, such as dyskeratosis congenita, aplastic anemia, and pulmonary fibrosis (reviewed in Armanios and Blackburn, 2012). Alternatively, the length of the chromosome extremities are maintained through a mechanism of homologous recombination (for review Conomos et al., 2013). During the process, the 3′ OH end of the chromosome invades another chromosome end, and amplifies the repeats. Telomeres are thus dynamic structures and their sequence composition should be specific to prevent illegitimate recombination generating chromosomal rearrangements.
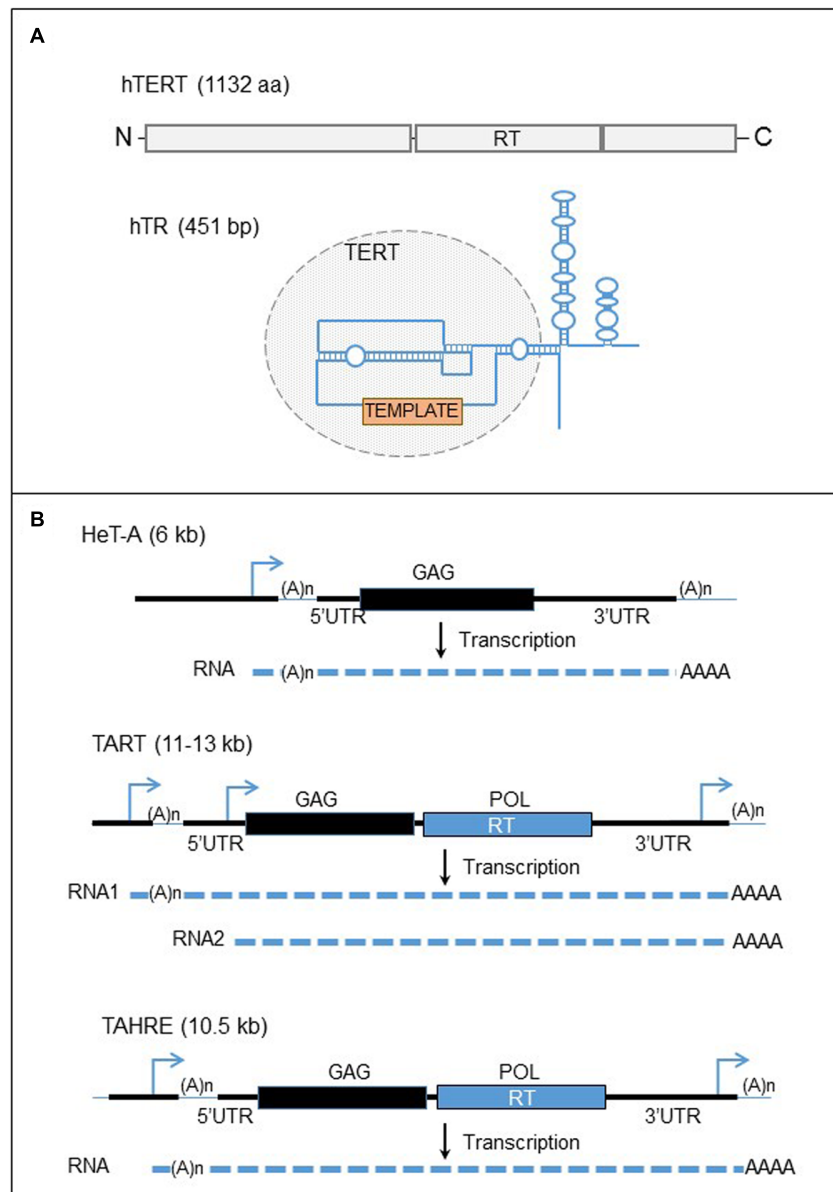
FIGURE 4 | Human telomerase complex and telomere-specific retrotransposons of *Drosophila*. (A) The major components of the human telomerase complex. Top panel: organization of the human telomerase enzyme (hTERT). The gray boxes represent the three domains of the protein, the N-terminal, the reverse transcriptase (RT), and the C-terminal domains from left to right. Bottom panel: structure of the telomerase RNA template. The blue line represent the telomerase RNA. The circle domain represent the domain recognized by the hTERT. The orange box represents the template motif. **(B)** Telomere-specific non-LTR retrotransposons of *Drosophila*. Black lines are the DNA strands. The blue arrows represent the promoters of the elements. Black and blue boxes represent the two ORFs, GAG and POL. The dashed lines are RNAs. UTR, untranslated region; RT, reverse transcriptase.

RETROTRANSPOSITION AT THE END OF THE CHROMOSOMES: SPECIFICITY OF INTEGRATION OR RESCUE OF DYSFUNCTIONAL TELOMERASE

Telomere-Specific Retrotransposons

As a specialized retroelement, the telomerase complex targets specifically the chromosome extremities to reverse transcribe the RNA template. Interestingly, the telomerase complex is recruited to chromosome ends through specific interactions between telomerase enzyme and the shelterin complex, the telomere-associated proteins that cap the DNA ends (for review Nandakumar and Cech, 2013). In the fission yeast *Schizosaccharomyces pombe*, the phosphorylation of telomere capping proteins by the DNA damage sensor kinases, ATM and ATR, is required for the interaction with the telomerase complex and the recruitment at telomeres (Moser et al., 2011; Yamazaki

et al., 2012). Such a regulation has not been yet characterized in mammalian cells but is suspected because ATM and ATR are also involved in telomere maintenance and notably telomere length regulation (for review Longhese, 2008; Diotti and Loayza, 2011). Intriguingly, important insights into the telomerase recruitment to chromosome ends were made by studying the mechanism of telomere healing, also called *de novo* telomere formation. Telomere healing is a very deleterious and rare process in the majority of eukaryote organisms that consists of adding telomere repeats at persisting DNA double strand breaks (DSBs) and leads to the loss of genetic information (for review Ribeyre and Shore, 2013). In budding yeast *S. cerevisiae*, telomere capping proteins and the telomerase complex are recruited to DSBs, in a comparable level as to telomeric ends, but the ATR ortholog, Mec1, limits their accumulation at DNA breaks and the *de novo* telomere formation (Zhang and Durocher, 2010; Ribaud et al., 2012). Therefore telomere healing can serve as a model to study the regulation of telomerase recruitment and activation in order to further determine the mechanism of protection of the linear DNA ends.

Retroelements have been identified and characterized in all sequenced eukaryotic genomes whereas they are a threat for the stability of the genomes. In human, their mobility, activated in germline cells, leads to diseases (for review Belancio et al., 2008; Hancks and Kazazian, 2012). The activity level of L1 elements is also very high but variable in a wide range of tumors (Iskow et al., 2010; Lee et al., 2012; Solyom et al., 2012; Tubio et al., 2014; Ewing et al., 2015). A very efficient way to prevent mobile DNA from generating gene mutations is to direct insertions in poor-gene regions. Subtelomeric and telomeric regions seem to represent a common "safe haven" for this purpose, although the multicopy rRNA cluster, and centromeric regions are used with some elements in some genomes. In this section, we examine the recruitment of telomere-specific retrotransposons, revealing similarities in the targeting mechanism of the telomerase complex, although the proteins involved may be different.

Target Specificity: Telomeres, Safe Harbor

The analysis of retrotransposons in genomes demonstrates that their distribution is not random and their location results of both integration specificity and selection pressure for the inserts that are less detrimental to the genome. The genome of *S. cerevisiae* is very condensed and retrotransposons are preferentially located in gene-poor regions of the chromosomes, either upstream of RNA pol III genes (Ty1, Ty2, Ty3) or at telomeres (Ty5) (Kim et al., 1998). In yeast, the integration bias is the consequence of a targeting strategy implying interaction between the integrase and cellular factors, rather than the recognition of a specific DNA sequence by the enzyme.

In the genome of *S. cerevisiae*, there are few insertions of Ty5 retrotransposons and only one copy is full-length but not active because the coding regions contain several mutations (Voytas and Boeke, 1992). The inserts are located in the heterochromatin near telomere regions of chromosomes. Using an active Ty5 element from the related yeast strain *Saccharomyces parodoxus*,

the Voytas laboratory has identified the mechanism of targeting specificity (Zou et al., 1996; Xie et al., 2001; Zhu et al., 2003). Ninety percent of *de novo* Ty5 elements are located in the silent chromatin at telomeres or silent mating loci and the integration is targeted through an interaction between the targeting domain of Ty5 integrase and the silent information regulator 4, Sir4p, a protein of the heterochromatin. Mutations in the targeting domain result in the loss of specificity of integration. Noteworthy, the integrase domain that interacts with Sir4p shares similarities with another protein interacting with Sir4p, Esc1p (Brady et al., 2008). Esc1p, a protein associated with the nuclear periphery, is also involved in chromatin silencing at telomeres (Andrulis et al., 2002). Additionally, the targeting domain is phosphorylated, and this post-translational modification mediates the interaction with Sir4p (Dai et al., 2007). The absence of phosphorylation results in a random integration of Ty5 elements in the genome and creates mutations. Intriguingly, the phosphorylation of integrase is regulated by stress conditions such as deprivation in nutrients (amino acids, nitrogen), suggesting that Ty5 retrotransposition is controlled for adaptive response to changes in environmental conditions.

Even if several copies of Ty1 retrotransposon of *S. cerevisiae* are recovered in subtelomeres, Ty1 is not a telomere-specific element. In fact, this location is a secondary target site selection and the targeting mechanism is not characterized. Ninety percent of Ty1 retrotransposons are preferentially targeted upstream of RNA pol III transcribed genes (Kim et al., 1998). The mechanism of this integration specificity has been recently identified and involves the interaction between Ty1 integrase and the cellular factor, AC40p, a subunit of RNA pol III complex (Bridier-Nahmias et al., 2015). When this interaction is lost, *de novo* Ty1 copies insert preferentially at chromosome ends. It has also been shown that the chromatin structure and chromatin remodeling complex are important components of the mechanism of the Ty1 integration upstream of RNA pol III transcribed genes (Bachman et al., 2005; Gelbart et al., 2005; Baller et al., 2012). Ty1 retrotransposons insert within 750 bases upstream of tRNA genes with a periodicity that depends on the nucleosome position in the region and more generally, Ty1 *de novo* inserts show a preference for nucleosome-rich sites, flanking RNA pol III transcribed genes (Baller et al., 2012). Therefore we can suppose that chromatin proteins can play a role in the insertion of Ty1 in heterochromatin at subtelomeric regions of chromosomes but the mechanism remains unknown and needs to be determined.

Retrotransposon to Compensate for the Absence of Telomerase in the Genome or a Low Expression Level of the Telomerase

The telomere-specific non-LTR retrotransposons of *Drosophila* represent an interesting case of domestication of transposable elements. The fly chromosome ends are not composed of canonical telomere repeats. The DNA component of the fly telomeres consists instead of three non-LTR retrotransposons arranged in tandem arrays, TAHRE, TART, and HeT-A (for review, Biessmann and Mason, 2003; Pardue et al., 2005; Pardue and Debarshe, 2011; Fujiwara, 2015). Additionally, the genome

of this organism does not encode a telomerase. The gene seems to have been lost in an ancestor of Diptera (Garavís et al., 2013). While some dipteran insects have maintained telomeric tandem repeats by homologous recombination, *Drosophila* genome has replaced the telomerase activity with the retrotransposition of the three telomere-specific retroelements. Therefore, RT activity from retrotransposons seems to be an adaptive cellular mechanism to recover a deficiency in the telomerase activity. Other *Drosophila* mobile elements are not found in the telomere arrays and the telomere-specific elements do not insert anywhere else in the genome, except for the broken ends of chromosomes (Biessmann et al., 1990; George et al., 2006).

The Pardue laboratory has described these elements and the telomere maintenance in *Drosophila* (**Figure 4B**). The sequence of the most abundant element, HeT-A, contains one ORF corresponding to a structure protein, ORF1, based on the domains present on the protein (Traverse and Pardue, 1988; Biessmann et al., 1990). Therefore HeT-A does not encode a RT activity and depends on another element for the retrotransposition. HeT-A is related to the latest discovered TAHRE element, encoding two ORFs (Abad et al., 2004). This element is less characterized because it is very rare at *Drosophila* telomeres. TART, the second most abundant element, has 2 ORFs and provide the retrotransposition machinery to the non-autonomous HeT-A (Sheen and Levis, 1994). Noteworthy, HeT-A ORF1p has a nuclear localization signal and the protein, fused to the green fluorescent protein (GFP), seems to form particles at chromosome ends in microscopy, whereas TART ORF1p does not have a specific cellular location (Rashkova et al., 2003). However, when the two proteins are overexpressed in *Drosophila* cells, both proteins co-localize at the end of chromosomes, suggesting that HeT-A ORF1p interacts with TART ORF1p and determines the intra-nuclear localization of TART proteins at the chromosome ends. The three non-LTR retroelements are assumed to insert specifically at the 3′ OH of the DNA end at the chromosome extremities. Therefore, an EN activity is dispensable for a retrotransposition event to occur. The promoter of HeT-A elements is in the 3′UTR whereas several promoters are located at both ends of the TART element (Danilevskaya et al., 1997, 1999). Therefore the transcription of an element can start from the 3′ end of the last element inserted at the end of chromosome, an apparent adaptation to retroelements appearing in tandem arrays.

Because *Drosophila* does not have canonical telomere repeats and telomerase complex, it is not surprising that proteins capping chromosome ends, constituting the terminin complex, are original and do not have sequence homology with proteins in human and yeasts (review Raffa et al., 2011, 2013). However, the function of terminin proteins such as HOAP and HipHop, is conserved: they are recruited to chromosome ends, accumulate, and prevent the action of DNA repair pathways on the chromosome extremities (Rashkova et al., 2002; Gao et al., 2010). The regulation of the recruitment of these proteins to telomeres is also conserved and involves the DNA sensor kinases ATM and ATR, which also regulate the formation and maintenance of telomeres in the other organisms (Bi et al., 2005;

Gao et al., 2010). The mechanism of recruitment of terminin proteins to chromosome ends is unknown and the interaction with the proteins of telomere-specific retrotransposons has never been characterized. Interestingly, the understanding of telomere maintenance in *Drosophila* has also benefited from studies of DSB repair by telomere healing. Actually chromosomes lacking telomere-specific retrotransposons are remarkably stable for several generations, even in natural fly populations (Biessmann et al., 1992; Ahmad and Golic, 1998; Kern and Begun, 2008). Additionally, while the process of *de novo* telomere addition involves the RT activity of the telomerase complex in most organisms, surprisingly the establishment of *Drosophila* caps at DNA ends does not require the retrotransposition of telomere-specific elements for the assembly and maintenance of a functional terminin complex (Gao et al., 2010; Beaucher et al., 2012). Therefore, even if the loss of telomerase complex in evolution changed the proteins involved in chromosome end cap, the function and mechanism of maintenance are conserved.

The silkworm, *Bombyx mori*, appears to be a hybrid of canonical telomeres with retrotransposon-based telomeres (for review Fujiwara, 2015). In this case, the telomere repeats are interrupted with two families of non-LTR retrotransposons, SART and TRAS (Okazaki et al., 1995; Takahashi et al., 1997). The telomerase activity in this organism is barely detectable and to maintain the length of the chromosome extremities, these autonomous retroelements target specifically the telomere repeats (Sasaki and Fujiwara, 2000). Intriguingly only full-length elements are identified at telomeres (Fujiwara et al., 2005). Some copies have been reported in other part of the chromosomes, mostly truncated and not at the target site (Monti et al., 2013). They may be the result of recombination events between elements at telomeres and sequences in the genome. SART and TRAS elements have a very similar structure to human L1 retrotransposons. They encode two ORFs, ORF1 and ORF2 (Okazaki et al., 1995; Takahashi et al., 1997). ORF2p has EN and RT activities. The EN domain recognizes the telomeric repeats, TTAGG, and cleaves specifically between T and A. TRAS ORF1p has a nuclear localization domain and is able to interact with ORF2p (Matsumoto et al., 2004). However, the specific role of ORF1p is not well understood. Both proteins are required for the mobility of SART and TRAS. Unlike L1, the 3′ UTR of the silkworm telomere-specific elements is also required for retrotransposition (Takahashi and Fujiwara, 2002). The 3′ UTR has specific motifs that are proposed to interact with the RT domain of ORF2p and to anneal to the target site (Osanai et al., 2004). Although these non-LTR retrotransposons are actively transcribed, promoter motifs have not been identified (Takahashi and Fujiwara, 1999). The activities of the telomerase complex and the telomere-specific retrotransposon may be in conflict if they occur at the same time. However, while the telomerase complex is regulated by the cell cycle, such a regulation has not been reported for SART and TRAS retrotransposons in *Bombyx mori*. Additionally little is known about the mechanism of the recruitment of these elements to the telomeric repeats and it is possible that cellular factors may direct the recognition of the target sequence by ORF2p.

Redirection of the Insertion in Case of Deficient in Telomere Maintenance: Impact on Genome Stability

It is intriguing to note that some telomere-specific retrotransposons seem to rescue partial or complete deficiencies of the telomerase activity. This observation suggests that retrotransposition may serve as a response to dysfunctional telomerases or to the absence of telomerase in cells.

The Curcio laboratory has studied the regulation of Ty1 retrotransposition in yeast strains defective in telomerase. In yeast, the telomere RT Est2p uses RNA template Tlc1 to polymerize telomere arrays at the chromosome extremities (for review Lundblad, 2002; Kupiec, 2014). In yeast strains deficient for the telomerase activity, the *est2∆* mutants, telomere length decreases with cell divisions until the telomere length becomes very short and causes the arrest of cell division (Lundblad and Szostak, 1989). Usually cells stop dividing after 50 to 100 generations. Rare cells survive and present alternative telomere structures (Lundblad and Blackburn, 1993). Type I survivors contain tandem arrays of subtelomeric repeat Y' and type II survivors have long and heterogeneous tracts of telomeric repeats. Scholes et al. (2003) has reported that Ty1 retrotransposition is induced in the *est2∆* mutant, before cell senescence and the appearance of survivors. The activation of Ty1 retrotransposition frequency occurs in parallel with telomere erosion and is characterized by an increase in Ty1 cDNA in cells. However, in survivors, the Ty1 retrotransposition rate decreases. Therefore Ty1 retrotransposition is induced as a response to telomere dysfunction and raise the question whether this activation plays a role in the formation of alternative telomeres. In another publication, the Curcio laboratory showed that chimeric Y'-Ty1 elements are identified in type I survivors (Maxwell et al., 2004). Ty1 retrotransposon contributes to the retrotransposition of the Y' repeats at subtelomeres in telomerase-deficient cells. Retrotransposition seems to be, in this case, one mechanism allowing for the extension of telomeres in telomerase-negative survivors. Intriguing the authors also showed that Y' RNA is enriched in Ty1 VLP fraction and that this enrichment is not regulated by telomere erosion because Y' RNA is present in the VLPs of telomerase-positive and negative cells. These data suggest that the integration events of Y' cDNA only occur in telomerase-deficient cells and raise the question of which cellular factors are involved in this control.

In contrast, L1 retrotransposition has not been reported to be activated in cells deficient in telomerase activity. However, there are EN-independent L1 events that have been reported to be inserted at the chromosome extremities. EN-independent events have been first characterized in the Moran laboratory, looking at the effect of the deficiency in the non-homologous end joining (NHEJ) DSB repair in mammalian cells (Morrish et al., 2002). They identified that normal L1 retrotransposition is not noticeably induced in this mutant, but they observed unusual events, that lack common marks of L1 retrotransposition such as TSDs, or common EN target site at the insertions. Additionally the *de novo* L1 copies are $3'$ end truncated, suggesting that these insertions have occurred at DNA lesions. In DNA PKcs-deficient cells, 30% of L1 EN-independent retrotransposition events have occurred at telomeres (Morrish et al., 2007). These events are not observed in another cell line deficient for XRCC4, an essential component of the NHEJ pathway (reviewed in Williams et al., 2014). DNA PKcs is very well identified as an essential kinase of the NHEJ pathway (for review Lees-Miller and Meek, 2003; Weterings and Van Gent, 2004). More recently, DNA PKcs has been reported as a component of the telomere maintenance. In fact, cells mutated in the kinase have uncapped dysfunctional telomeres, but unaffected in their length (Goytisolo et al., 2001; Williams et al., 2009). Morrish et al. (2007) showed that the new L1 inserts at the telomeres in DNA PKcs mutant can exhibit a poly (A) tail but the retrotransposition did not occur at common EN target sites. These observations imply that uncapped dysfunctional telomeres, but not shortened telomeres, are substrates for opportunistic L1 RT in mammalian cells. These data suggest that the L1 retrotransposition machinery is recruited to unprotected and persistent DNA ends and this phenomenon resembles the process described as *de novo* telomere formation at DSBs by the telomerase complex. Intriguingly, L1 retrotransposition at the chromosome ends, in this study, does not supply the absence of telomerase activity, revealing a more general response of the retrotransposons to the dysfunction of telomere maintenance.

CONCLUSION

Telomerases have likely evolved from an ancestor retroelement during genome evolution (**Figure 1**). They are essentially stringent non-autonomous retrotransposons, specialized to insert telomeric repeats at the linear chromosome ends. The description of telomerases and modern retrotransposons reveals the specificities of each group of genetic elements. Notably, the originality of the telomerase RT function is based on the exclusivity of the RNA template and this is a very unique mechanism of regulation. In fact, although retrotransposon enzymes preferentially bind and reverse transcribe their own encoding RNAs, they are able to recognize other RNAs. Therefore, they are responsible for the insertion of processed pseudogenes throughout the genome, and also they supply the machinery to amplify non-autonomous retroelements (Derr et al., 1991; Esnault et al., 2000; Wei et al., 2001). In contrast, telomerase complexes cannot reverse transcribe other sequences in the genome because the presence of the specific RNA template in the active site of the enzyme is necessary for the catalytic activation. Therefore the telomerase complexes are very unique genetic elements in eukaryotic genomes and mutations disrupting the telomerase function cause the shortening of telomeres and the arrest of the cell cycle. Telomerase-negative survivors need to develop alternative pathways to compensate for the shortening of the chromosome ends. We discussed in the present paper the possibility that retrotransposition might provide an adaptive mechanism for the formation of alternative telomere structures and compensate for the shortening of the chromosomes (**Figure 1**).

Two examples especially seem to validate this hypothesis: the *Drosophila* and silkworm telomere-specific non-LTR retrotransposons. These retrotransposons are specialized and are not inserted anywhere else in the genomes. Furthermore, the chromosome extremities of *Drosophila* and silkworm are also protected from the integration of other retrotransposons that are not telomere-specific. Because telomerase complexes are phylogenetically closer to non-LTR retrotransposons, notably based on the similarity of the insertion process, it is easy to imagine that non-LTR retrotransposons can counteract the shortening of the chromosomes in cells deficient for the telomerase function. However, in response to disrupted telomerase gene, the budding yeast *S. cerevisiae*, containing only LTR-retrotransposons, activates Ty1 RT, contributing to the formation of alternative telomere structures in survivor cells. Therefore, retrotransposition seems to be an evolutionary mechanism to compensate the telomerase deficiency. Intriguingly the comparison of the different mechanisms of chromosome end protection also reveals similarities in the recruitment of the telomerase complex and retrotransposons to the target sites, providing new perspectives for the investigation of telomere formation and maintenance.

AUTHOR CONTRIBUTIONS

GS and PD wrote the paper.

ACKNOWLEDGMENTS

REFERENCES

Abad, J. P., De Pablos, B., Osoegawa, K., De Jong, P. J., Martín-Gallardo, A., and Villasante, A. (2004). TAHRE, a novel telomeric retrotransposon from *Drosophila melanogaster*, reveals the origin of *Drosophila* telomeres. *Mol. Biol. Evol.* 21, 1620–1624. doi: 10.1093/molbev/msh180

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195. doi: 10.1126/science.287.5461.2185

Ahl, V., Keller, H., Schmidt, S., and Weichenrieder, O. (2015). Retrotransposition and crystal structure of an alu RNP in the ribosome-stalling conformation. *Mol. Cell* 60, 715–727. doi: 10.1016/j.molcel.2015.10.003

Ahmad, K., and Golic, K. G. (1998). The transmission of fragmented chromosomes in *Drosophila melanogaster*. *Genetics* 148, 775–792.

Andrulis, E. D., Zappulla, D. C., Ansari, A., Perrod, S., Laiosa, C. V., Gartenberg, M. R., et al. (2002). Esc1, a nuclear periphery protein required for Sir4-based plasmid anchoring and partitioning. *Mol. Cell. Biol.* 22, 8292–8301. doi: 10.1128/MCB.22.23.8292-8301.2002

Arkhipova, I. R., Pyatkov, K. I., Meselson, M., and Evgen'Ev, M. B. (2003). Retroelements containing introns in diverse invertebrate taxa. *Nat. Genet.* 33, 123–124. doi: 10.1038/ng1074

Armanios, M., and Blackburn, E. H. (2012). The telomere syndromes. *Nat. Rev. Genet.* 13, 693–704. doi: 10.1038/nrg3246

Babushok, D. V., Ostertag, E. M., Courtney, C. E., Choi, J. M., and Kazazian, H. H. (2006). L1 integration in a transgenic mouse model. *Genome Res.* 16, 240–250. doi: 10.1101/gr.4571606

Bachman, N., Gelbart, M. E., Tsukiyama, T., and Boeke, J. D. (2005). TFIIIB subunit Bdp1p is required for periodic integration of the Ty1 retrotransposon and targeting of Isw2p to *S. cerevisiae* tDNAs. *Genes Dev.* 19, 955–964. doi: 10.1101/gad.1299105

Baller, J. A., Gao, J., Stamenova, R., Curcio, M. J., and Voytas, D. F. (2012). A nucleosomal surface defines an integration hotspot for the *Saccharomyces cerevisiae* Ty1 retrotransposon. *Genome Res.* 22, 704–713. doi: 10.1101/gr.129585.111

Beaucher, M., Zheng, X. F., Amariei, F., and Rong, Y. S. (2012). Multiple pathways suppress telomere addition to DNA breaks in the *Drosophila* germline. *Genetics* 191, 407–417. doi: 10.1534/genetics.112.138818

Belancio, V. P., Hedges, D. J., and Deininger, P. (2008). Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res.* 18, 343–358. doi: 10.1101/gr.5558208

Bi, X., Srikanta, D., Fanti, L., Pimpinelli, S., Badugu, R., Kellum, R., et al. (2005). *Drosophila* ATM and ATR checkpoint kinases control partially redundant pathways for telomere maintenance. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15167–15172. doi: 10.1073/pnas.0504981102

Biessmann, H., Champion, L. E., O'Hair, M., Ikenaga, K., Kasravi, B., and Mason, J. M. (1992). Frequent transpositions of *Drosophila melanogaster* HeT-A transposable elements to receding chromosome ends. *EMBO J.* 11, 4459–4469.

Biessmann, H., and Mason, J. M. (2003). Telomerase-independent mechanisms of telomere elongation. *Cell Mol. Life Sci* 60, 2325–2333. doi: 10.1007/s00018-003-3247-9

Biessmann, H., Mason, J. M., Ferry, K., D'Hulst, M., Valgeirsdottir, K., Traverse, K. L., et al. (1990). Addition of telomere-associated HeT DNA sequences "heals" broken chromosome ends in *Drosophila*. *Cell* 61, 663–673. doi: 10.1016/0092-8674(90)90478-W

Blackburn, E. H. (1992). Telomerases. *Annu. Rev. Biochem.* 61, 113–129. doi: 10.1146/annurev.bi.61.070192.000553

Blackburn, E. H., and Collins, K. (2011). Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harb. Perspect. Biol.* 3:pii:a003558. doi: 10.1101/cshperspect.a003558

Boeke, J. D. (1997). LINEs and Alus–the polyA connection. *Nat. Genet.* 16, 6–7. doi: 10.1038/ng0597-6

Boeke, J. D., Garfinkel, D. J., Styles, C. A., and Fink, G. R. (1985). Ty elements transpose through an RNA intermediate. *Cell* 40, 491–500. doi: 10.1016/0092-8674(85)90197-7

Bourc'His, D., and Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 431, 96–99. doi: 10.1038/nature02886

Brady, T. L., Fuerst, P. G., Dick, R. A., Schmidt, C., and Voytas, D. F. (2008). Retrotransposon target site selection by imitation of a cellular protein. *Mol. Cell. Biol.* 28, 1230–1239. doi: 10.1128/MCB.01502-07

Bridier-Nahmias, A., Tchalikian-Cosson, A., Baller, J. A., Menouni, R., Fayol, H., Flores, A., et al. (2015). Retrotransposons. An RNA polymerase III subunit determines sites of retrotransposon integration. *Science* 348, 585–588. doi: 10.1126/science.1259114

Bryce, L. A., Morrison, N., Hoare, S. F., Muir, S., and Keith, W. N. (2000). Mapping of the gene for the human telomerase reverse transcriptase, hTERT, to chromosome 5p15.*33* by fluorescence in situ hybridization. *Neoplasia* 2, 197–201. doi: 10.1038/sj.neo.7900092

Bryk, M., Banerjee, M., Conte, D., and Curcio, M. J. (2001). The Sgs1 helicase of *Saccharomyces cerevisiae* inhibits retrotransposition of Ty1 multimeric arrays. *Mol. Cell. Biol.* 21, 5374–5388. doi: 10.1128/MCB.21.16.5374-5388.2001

Chang, D. Y., Hsu, K., and Maraia, R. J. (1996). Monomeric scAlu and nascent dimeric Alu RNAs induced by adenovirus are assembled into SRP9/14-containing RNPs in HeLa cells. *Nucleic Acids Res.* 24, 4165–4170. doi: 10.1093/nar/24.21.4165

Clare, J. J., Belcourt, M., and Farabaugh, P. J. (1988). Efficient translational frameshifting occurs within a conserved sequence of the overlap between the two genes of a yeast Ty1 transposon. *Proc. Natl. Acad. Sci. U.S.A.* 85, 6816–6820. doi: 10.1073/pnas.85.18.6816

Comeaux, M. S., Roy-Engel, A. M., Hedges, D. J., and Deininger, P. L. (2009). Diverse cis factors controlling Alu retrotransposition: what

causes Alu elements to die? *Genome Res* 19, 545–555. doi: 10.1101/gr.089789.108

Conomos, D., Pickett, H. A., and Reddel, R. R. (2013). Alternative lengthening of telomeres: remodeling the telomere architecture. *Front. Oncol.* 3:27. doi: 10.3389/fonc.2013.00027

Curcio, M. J., Lutz, S., and Lesage, P. (2015). The Ty1 LTR-Retrotransposon of budding yeast, *Saccharomyces cerevisiae*. *Microbiol. Spectr* 3, MDNA3–MDNA0053.

Dai, J., Xie, W., Brady, T. L., Gao, J., and Voytas, D. F. (2007). Phosphorylation regulates integration of the yeast Ty5 retrotransposon into heterochromatin. *Mol. Cell.* 27, 289–299. doi: 10.1016/j.molcel.2007.06.010

Danilevskaya, O. N., Arkhipova, I. R., Traverse, K. L., and Pardue, M. L. (1997). Promoting in tandem: the promoter for telomere transposon HeT-A and implications for the evolution of retroviral LTRs. *Cell* 88, 647–655. doi: 10.1016/S0092-8674(00)81907-8

Danilevskaya, O. N., Traverse, K. L., Hogan, N. C., Debaryshe, P. G., and Pardue, M. L. (1999). The two *Drosophila* telomeric transposable elements have very different patterns of transcription. *Mol. Cell. Biol.* 19, 873–881. doi: 10.1128/MCB.19.1.873

Deininger, P. L., and Batzer, M. A. (2002). Mammalian retroelements. *Genome Res.* 12, 1455–1465. doi: 10.1101/gr.282402

de Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A., and Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7:e1002384. doi: 10.1371/journal.pgen.1002384

Denli, A. M., Narvaiza, I., Kerman, B. E., Pena, M., Benner, C., Marchetto, M. C., et al. (2015). Primate-Specific ORF0 contributes to retrotransposon-mediated diversity. *Cell* 163, 583–593. doi: 10.1016/j.cell.2015.09.025

Derr, L. K., Strathern, J. N., and Garfinkel, D. J. (1991). RNA-mediated recombination in *S. cerevisiae*. *Cell* 67, 355–364. doi: 10.1016/0092-8674(91)90187-4

Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* 35, 41–48. doi: 10.1038/ng1223

Dewannieux, M., and Heidmann, T. (2005). Role of poly(A) tail length in Alu retrotransposition. *Genomics* 86, 378–381. doi: 10.1016/j.ygeno.2005.05.009

Diotti, R., and Loayza, D. (2011). Shelterin complex and associated factors at human telomeres. *Nucleus* 2, 119–135. doi: 10.4161/nucl.2.2.15135

Doucet, A. J., Wilusz, J. E., Miyoshi, T., Liu, Y., and Moran, J. V. (2015). A 3' Poly(A) tract is required for line-1 retrotransposition. *Mol Cell.* 60, 728–741.

Egan, E. D., and Collins, K. (2012). Biogenesis of telomerase ribonucleoproteins. *RNA* 18, 1747–1759. doi: 10.1261/rna.034629.112

Eickbush, T. H. (1997). Telomerase and retrotransposons: which came first? *Science* 277, 911–912. doi: 10.1126/science.277.5328.911

Eickbush, T. H., and Malik, H. S. (2002). "Origins and evolution of retrotransposons," in *Mobile DNA II*, eds N. L. Craig, C. R. M. Gellert, and A. M. Lambowitz (Washington, DC: ASM Press), 1111–1144.

Esnault, C., Maestre, J., and Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* 24, 363–367. doi: 10.1038/74184

Etheridge, K. T., Banik, S. S., Armbruster, B. N., Zhu, Y., Terns, R. M., Terns, M. P., et al. (2002). The nucleolar localization domain of the catalytic subunit of human telomerase. *J. Biol. Chem.* 277, 24764–24770. doi: 10.1074/jbc.M201227200

Ewing, A. D., Gacita, A., Wood, L. D., Ma, F., Xing, D., Kim, M. S., et al. (2015). Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. *Genome Res.* 25, 1536–1545. doi: 10.1101/gr.196238.115

Feng, J., Funk, W. D., Wang, S. S., Weinrich, S. L., Avilion, A. A., Chiu, C. P., et al. (1995). The RNA component of human telomerase. *Science* 269, 1236–1241. doi: 10.1126/science.7544491

Feng, Q., Moran, J. V., Kazazian, H. H. Jr., and Boeke, J. D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87, 905–916. doi: 10.1016/S0092-8674(00)81997-2

Fujiwara, H. (2015). Site-specific non-LTR retrotransposons. *Microbiol. Spectr* 3, MDNA3–MDNA1. doi: 10.1128/microbiolspec.MDNA3-0001-2014

Fujiwara, H., Osanai, M., Matsumoto, T., and Kojima, K. K. (2005). Telomere-specific non-LTR retrotransposons and telomere maintenance in the silkworm, *Bombyx mori. Chromosome Res.* 13, 455–467. doi: 10.1007/s10577-005-0990-9

Gao, G., Walser, J. C., Beaucher, M. L., Morciano, P., Wesolowska, N., Chen, J., et al. (2010). HipHop interacts with HOAP and HP1 to protect *Drosophila*

telomeres in a sequence-independent manner. *EMBO J.* 29, 819–829. doi: 10.1038/emboj.2009.394

Garavís, M., González, C., and Villasante, A. (2013). On the origin of the eukaryotic chromosome: the role of noncanonical DNA structures in telomere evolution. *Genome Biol. Evol.* 5, 1142–1150. doi: 10.1093/gbe/evt079

Garfinkel, D. J., Nyswaner, K., Wang, J., and Cho, J. Y. (2003). Post-transcriptional cosuppression of Ty1 retrotransposition. *Genetics* 165, 83–99.

Gasior, S. L., Roy-Engel, A. M., and Deininger, P. L. (2008). ERCC1/XPF limits L1 retrotransposition. *DNA Repair (Amst.)* 7, 983–989. doi: 10.1016/j.dnarep.2008.02.006

Gelbart, M. E., Bachman, N., Delrow, J., Boeke, J. D., and Tsukiyama, T. (2005). Genome-wide identification of Isw2 chromatin-remodeling targets by localization of a catalytically inactive mutant. *Genes Dev.* 19, 942–954. doi: 10.1101/gad.1298905

George, J. A., Debaryshe, P. G., Traverse, K. L., Celniker, S. E., and Pardue, M. L. (2006). Genomic organization of the *Drosophila* telomere retrotransposable elements. *Genome Res.* 16, 1231–1240. doi: 10.1101/gr.5348806

Goytisolo, F. A., Samper, E., Edmonson, S., Taccioli, G. E., and Blasco, M. A. (2001). The absence of the dna-dependent protein kinase catalytic subunit in mice results in anaphase bridges and in increased telomeric fusions with normal telomere length and G-strand overhang. *Mol. Cell. Biol.* 21, 3642–3651. doi: 10.1128/MCB.21.11.3642-3651.2001

Greider, C. W., and Blackburn, E. H. (1989). A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. *Nature* 337, 331–337. doi: 10.1038/337331a0

Hancks, D. C., and Kazazian, H. H. (2012). Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* 22, 191–203. doi: 10.1016/j.gde.2012.02.006

Harley, C. B., Futcher, A. B., and Greider, C. W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature* 345, 458–460. doi: 10.1038/345458a0

Harrington, L., Zhou, W., Mcphail, T., Oulton, R., Yeung, D. S., Mar, V., et al. (1997). Human telomerase contains evolutionarily conserved catalytic and structural subunits. *Genes Dev.* 11, 3109–3115. doi: 10.1101/gad.11.23.3109

Hata, K., and Sakaki, Y. (1997). Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene* 189, 227–234. doi: 10.1016/S0378-1119(96)00856-6

Hayflick, L. (1979). Cell biology of aging. *Fed. Proc.* 38, 1847–1850.

Heidmann, O., Vernochet, C., Dupressoir, A., and Heidmann, T. (2009). Identification of an endogenous retroviral envelope gene with fusogenic activity and placenta-specific expression in the rabbit: a new "syncytin" in a third order of mammals. *Retrovirology* 6:107. doi: 10.1186/1742-4690-6-107

Hsu, K., Chang, D. Y., and Maraia, R. J. (1995). Human signal recognition particle (SRP) Alu-associated protein also binds Alu interspersed repeat sequence RNAs. Characterization of human SRP9. *J. Biol. Chem.* 270, 10179–10186. doi: 10.1074/jbc.270.17.10179

Iskow, R. C., Mccabe, M. T., Mills, R. E., Torene, S., Pittard, W. S., Neuwald, A. F., et al. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* 141, 1253–1261. doi: 10.1016/j.cell.2010.05.020

Jády, B. E., Richard, P., Bertrand, E., and Kiss, T. (2006). Cell cycle-dependent recruitment of telomerase RNA and Cajal bodies to human telomeres. *Mol. Biol. Cell* 17, 944–954. doi: 10.1091/mbc.E05-09-0904

Jiang, Y. W. (2002). Transcriptional cosuppression of yeast Ty1 retrotransposons. *Genes Dev.* 16, 467–478. doi: 10.1101/gad.923502

Kern, A. D., and Begun, D. J. (2008). Recurrent deletion and gene presence/absence polymorphism: telomere dynamics dominate evolution at the tip of 3L in *Drosophila melanogaster* and *D. simulans*. *Genetics* 179, 1021–1027. doi: 10.1534/genetics.107.078345

Kilian, A., Bowtell, D. D., Abud, H. E., Hime, G. R., Venter, D. J., Keese, P. K., et al. (1997). Isolation of a candidate human telomerase catalytic subunit gene, which reveals complex splicing patterns in different cell types. *Hum. Mol. Genet.* 6, 2011–2019. doi: 10.1093/hmg/6.12.2011

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

Kupiec, M. (2014). Biology of telomeres: lessons from budding yeast. *FEMS Microbiol. Rev.* 38, 144–171. doi: 10.1111/1574-6976.12054

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062

Lavialle, C., Cornelis, G., Dupressoir, A., Esnault, C., Heidmann, O., Vernochet, C., et al. (2013). Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368, 20120507. doi: 10.1098/rstb.2012.0507

Lee, B. S., Lichtenstein, C. P., Faiola, B., Rinckel, L. A., Wysock, W., Curcio, M. J., et al. (1998). Posttranslational inhibition of Ty1 retrotransposition by nucleotide excision repair/transcription factor TFIIH subunits Ssl2p and Rad3p. *Genetics* 148, 1743–1761.

Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L. J., et al. (2012). Landscape of somatic retrotransposition in human cancers. *Science* 337, 967–971. doi: 10.1126/science.1222077

Lees-Miller, S. P., and Meek, K. (2003). Repair of DNA double strand breaks by non-homologous end joining. *Biochimie* 85, 1161–1173. doi: 10.1016/j.biochi.2003.10.011

Levy, M. Z., Allsopp, R. C., Futcher, A. B., Greider, C. W., and Harley, C. B. (1992). Telomere end-replication problem and cell aging. *J. Mol. Biol.* 225, 951–960. doi: 10.1016/0022-2836(92)90096-3

Lingner, J., Hughes, T. R., Shevchenko, A., Mann, M., Lundblad, V., and Cech, T. R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276, 561–567. doi: 10.1126/science.276.5312.561

Longhese, M. P. (2008). DNA damage response at functional and dysfunctional telomeres. *Genes Dev.* 22, 125–140. doi: 10.1101/gad.1626908

Lundblad, V. (2002). Telomere maintenance without telomerase. *Oncogene* 21, 522–531. doi: 10.1038/sj.onc.1205079

Lundblad, V., and Blackburn, E. H. (1993). An alternative pathway for yeast telomere maintenance rescues est1- senescence. *Cell* 73, 347–360. doi: 10.1016/0092-8674(93)90234-H

Lundblad, V., and Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* 57, 633–643. doi: 10.1016/0092-8674(89)90132-3

Mager, D. L., and Stoye, J. P. (2015). Mammalian endogenous retroviruses. *Microbiol. Spectr.* 3, MDNA3-0009-2014. doi: 10.1128/microbiolspec.MDNA3-0009-2014

Martin, S. L. (2010). Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. *RNA Biol.* 7, 706–711. doi: 10.4161/rna.7.6.13766

Mathias, S. L., Scott, A. F., Kazazian, H. H. Jr., Boeke, J. D., and Gabriel, A. (1991). Reverse transcriptase encoded by a human transposable element. *Science* 254, 1808–1810. doi: 10.1126/science.1722352

Matsumoto, T., Takahashi, H., and Fujiwara, H. (2004). Targeted nuclear import of open reading frame 1 protein is required for in vivo retrotransposition of a telomere-specific non terminal repeat retrotransposon, SART1. *Mol. Cell. Biol.* 24, 105–122. doi: 10.1128/MCB.24.1.105-122.2004

Matthews, G. D., Goodwin, T. J., Butler, M. I., Berryman, T. A., and Poulter, R. T. (1997). pCal, a highly unusual Ty1/copia retrotransposon from the pathogenic yeast Candida albicans. *J. Bacteriol.* 179, 7118–7128.

Maxwell, P. H., Coombes, C., Kenny, A. E., Lawler, J. F., Boeke, J. D., and Curcio, M. J. (2004). Ty1 mobilizes subtelomeric Y' elements in telomerase-negative *Saccharomyces cerevisiae* survivors. *Mol. Cell. Biol.* 24, 9887–9898. doi: 10.1128/MCB.24.22.9887-9898.2004

McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* 26, 234–282.

Meyerson, M., Counter, C. M., Eaton, E. N., Ellisen, L. W., Steiner, P., Caddle, S. D., et al. (1997). hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization. *Cell* 90, 785–795. doi: 10.1016/S0092-8674(00)80538-3

Mitchell, J. R., Cheng, J., and Collins, K. (1999). A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3′ end. *Mol. Cell. Biol.* 19, 567–576. doi: 10.1128/MCB.19.1.567

Mitchell, J. R., and Collins, K. (2000). Human telomerase activation requires two independent interactions between telomerase RNA and telomerase reverse transcriptase. *Mol. Cell* 6, 361–371. doi: 10.1016/S1097-2765(00)00036-8

Monti, V., Serafini, C., Manicardi, G. C., and Mandrioli, M. (2013). Characterization of non-LTR retrotransposable TRAS elements in the aphids Acyrthosiphon pisum and Myzus persicae (Aphididae, Hemiptera). *J. Hered.* 104, 547–553. doi: 10.1093/jhered/est017

Moran, J. V., Deberardinis, R. J., and Kazazian, H. H. Jr. (1999). Exon shuffling by L1 retrotransposition. *Science* 283, 1530–1534. doi: 10.1126/science.283.5407.1530

Moran, J. V., Holmes, S. E., Naas, T. P., Deberardinis, R. J., Boeke, J. D., and Kazazian, H. H. Jr. (1996). High frequency retrotransposition in cultured mammalian cells. *Cell* 87, 917–927. doi: 10.1016/S0092-8674(00)81998-4

Morgan, H. D., Sutherland, H. G., Martin, D. I., and Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* 23, 314–318. doi: 10.1038/15490

Morrish, T. A., Garcia-Perez, J. L., Stamato, T. D., Taccioli, G. E., Sekiguchi, J., and Moran, J. V. (2007). Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature* 446, 208–212. doi: 10.1038/nature05560

Morrish, T. A., Gilbert, N., Myers, J. S., Vincent, B. J., Stamato, T. D., Taccioli, G. E., et al. (2002). DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* 31, 159–165. doi: 10.1038/ng898

Moser, B. A., Chang, Y. T., Kosti, J., and Nakamura, T. M. (2011). Tel1ATM and Rad3ATR kinases promote Ccq1-Est1 interaction to maintain telomeres in fission yeast. *Nat. Struct. Mol. Biol.* 18, 1408–1413. doi: 10.1038/nsmb.2187

Muller, H. (1938). The remaking of chromosomes. *Collect. Net.* 13, 181–195,198.

Nakamura, T. M., and Cech, T. R. (1998). Reversing time: origin of telomerase. *Cell* 92, 587–590. doi: 10.1016/S0092-8674(00)81123-X

Nandakumar, J., and Cech, T. R. (2013). Finding the end: recruitment of telomerase to telomeres. *Nat. Rev. Mol. Cell Biol.* 14, 69–82. doi: 10.1038/nrm3505

Neuveglise, C., Feldmann, H., Bon, E., Gaillardin, C., and Casaregola, S. (2002). Genomic evolution of the long terminal repeat retrotransposons in hemiascomycetous yeasts. *Genome Res.* 12, 930–943. doi: 10.1101/gr.219202

Nishida, Y., Pachulska-Wieczorek, K., Błaszczyk, L., Saha, A., Gumna, J., Garfinkel, D. J., et al. (2015). Ty1 retrovirus-like element Gag contains overlapping restriction factor and nucleic acid chaperone functions. *Nucleic Acids Res.* 43, 7414–7431. doi: 10.1093/nar/gkv695

Okazaki, S., Ishikawa, H., and Fujiwara, H. (1995). Structural analysis of TRAS1, a novel family of telomeric repeat-associated retrotransposons in the silkworm, *Bombyx mori*. *Mol. Cell. Biol.* 15, 4545–4552. doi: 10.1128/MCB.15.8.4545

Osanai, M., Takahashi, H., Kojima, K. K., Hamada, M., and Fujiwara, H. (2004). Essential motifs in the 3′ untranslated region required for retrotransposition and the precise start of reverse transcription in non-long-terminal-repeat retrotransposon SART1. *Mol. Cell. Biol.* 24, 7902–7913. doi: 10.1128/MCB.24.18.7902-7913.2004

Pardue, M. L., and Debaryshe, P. (2011). Adapting to life at the end of the line: how *Drosophila* telomeric retrotransposons cope with their job. *Mob. Genet. Elements* 1, 128–134. doi: 10.4161/mge.1.2.16914

Pardue, M. L., Rashkova, S., Casacuberta, E., Debaryshe, P. G., George, J. A., and Traverse, K. L. (2005). Two retrotransposons maintain telomeres in *Drosophila*. *Chromosome Res.* 13, 443–453. doi: 10.1007/s10577-005-0993-6

Prudhomme, S., Bonnaud, B., and Mallet, F. (2005). Endogenous retroviruses and animal reproduction. *Cytogenet. Genome. Res.* 110, 353–364. doi: 10.1159/000084967

Quentin, Y. (1992). Origin of the alu family: a family of Alu-like monomers gave birth to the left and the right arms of the alu elements. *Nucleic Acids Res.* 20, 3397–3401. doi: 10.1093/nar/20.13.3397

Raffa, G. D., Cenci, G., Ciapponi, L., and Gatti, M. (2013). Organization and evolution of *Drosophila* termini: similarities and differences between *Drosophila* and human telomeres. *Front. Oncol.* 3:112. doi: 10.3389/fonc.2013.00112

Raffa, G. D., Ciapponi, L., Cenci, G., and Gatti, M. (2011). Terminin: a protein complex that mediates epigenetic maintenance of *Drosophila* telomeres. *Nucleus* 2, 383–391. doi: 10.4161/nucl.2.5.17873

Rashkova, S., Athanasiadis, A., and Pardue, M. L. (2003). Intracellular targeting of Gag proteins of the *Drosophila* telomeric retrotransposons. *J. Virol.* 77, 6376–6384. doi: 10.1128/JVI.77.11.6376-6384.2003

Rashkova, S., Karam, S. E., Kellum, R., and Pardue, M. L. (2002). Gag proteins of the two *Drosophila* telomeric retrotransposons are targeted to chromosome ends. *J. Cell Biol.* 159, 397–402. doi: 10.1083/jcb.200205039

Ribaud, V., Ribeyre, C., Damay, P., and Shore, D. (2012). DNA-end capping by the budding yeast transcription factor and subtelomeric binding protein Tbf1. *EMBO J.* 31, 138–149. doi: 10.1038/emboj.2011.349

Ribeyre, C., and Shore, D. (2013). Regulation of telomere addition at DNA double-strand breaks. *Chromosoma* 122, 159–173. doi: 10.1007/s00412-013-0404-2

Richardson, S. R., Doucet, A. J., Kopera, H. C., Moldovan, J. B., Garcia-Perez, J. L., and Moran, J. V. (2015). The influence of LINE-1 and SINE retrotransposons on mammalian genomes. *Microbiol. Spectr.* 3, MDNA3–0061-2014. doi: 10.1128/microbiolspec.MDNA3-0061-2014

Roy-Engel, A. M., Salem, A. H., Oyeniran, O. O., Deininger, L., Hedges, D. J., Kilroy, G. E., et al. (2002). Active alu element "A-tails": size does matter. *Genome Res.* 12, 1333–1344. doi: 10.1101/gr.384802

Saha, A., Mitchell, J. A., Nishida, Y., Hildreth, J. E., Ariberre, J. A., Gilbert, W. V., et al. (2015). A trans-dominant form of Gag restricts Ty1 retrotransposition and mediates copy number control. *J. Virol.* 89, 3922–3938. doi: 10.1128/JVI.03060-14

Sandmeyer, S., Patterson, K., and Bilanchone, V. (2015). Ty3, a position-specific retrotransposon in budding yeast. *Microbiol. Spectr.* 3, MDNA3–0057-2014. doi: 10.1128/microbiolspec.MDNA3-0057-2014

Sasaki, T., and Fujiwara, H. (2000). Detection and distribution patterns of telomerase activity in insects. *Eur. J. Biochem.* 267, 3025–3031. doi: 10.1046/j.1432-1033.2000.01323.x

Scholes, D. T., Kenny, A. E., Gamache, E. R., Mou, Z., and Curcio, M. J. (2003). Activation of a LTR-retrotransposon by telomere erosion. *Proc. Natl. Acad. Sci. U.S.A.* 100, 15736–15741. doi: 10.1073/pnas.2136609100

Severynse, D. M., Hutchison, C. A., and Edgell, M. H. (1992). Identification of transcriptional regulatory activity within the 5' A-type monomer sequence of the mouse LINE-1 retroposon. *Mamm. Genome* 2, 41–50. doi: 10.1007/BF00570439

Sheen, F. M., and Levis, R. W. (1994). Transposition of the LINE-like retrotransposon TART to *Drosophila* chromosome termini. *Proc. Natl. Acad. Sci. U.S.A.* 91, 12510–12514. doi: 10.1073/pnas.91.26.12510

Solyom, S., Ewing, A. D., Rahrmann, E. P., Doucet, T., Nelson, H. H., Burns, M. B., et al. (2012). Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res.* 22, 2328–2338. doi: 10.1101/gr.145235.112

Speek, M. (2001). Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell. Biol.* 21, 1973–1985. doi: 10.1128/MCB.21.6.1973-1985.2001

Swergold, G. D. (1990). Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.* 10, 6718–6729. doi: 10.1128/MCB.10.12.6718

Takahashi, H., and Fujiwara, H. (1999). Transcription analysis of the telomeric repeat-specific retrotransposons TRAS1 and SART1 of the silkworm *Bombyx mori*. *Nucleic Acids Res.* 27, 2015–2021. doi: 10.1093/nar/27.9.2015

Takahashi, H., and Fujiwara, H. (2002). Transplantation of target site specificity by swapping the endonuclease domains of two LINEs. *EMBO J.* 21, 408–417. doi: 10.1093/emboj/21.3.408

Takahashi, H., Okazaki, S., and Fujiwara, H. (1997). A new family of site-specific retrotransposons, SART1, is inserted into telomeric repeats of the silkworm, *Bombyx mori*. *Nucleic Acids Res.* 25, 1578–1584. doi: 10.1093/nar/25.8.1578

Tomlinson, R. L., Ziegler, T. D., Supakorndej, T., Terns, R. M., and Terns, M. P. (2006). Cell cycle-regulated trafficking of human telomerase to telomeres. *Mol. Biol. Cell* 17, 955–965. doi: 10.1091/mbc.E05-09-0903

Traverse, K. L., and Pardue, M. L. (1988). A spontaneously opened ring chromosome of *Drosophila melanogaster* has acquired He-T DNA sequences at both new telomeres. *Proc. Natl. Acad. Sci. U.S.A.* 85, 8116–8120. doi: 10.1073/pnas.85.21.8116

Tubio, J. M., Li, Y., Ju, Y. S., Martincorena, I., Cooke, S. L., Tojo, M., et al. (2014). Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* 345, 1251343. doi: 10.1126/science.1251343

Tucker, J. M., Larango, M. E., Wachsmuth, L. P., Kannan, N., and Garfinkel, D. J. (2015). The Ty1 retrotransposon restriction factor p22 targets gag. *PLoS Genet.* 11:e1005571. doi: 10.1371/journal.pgen.1005571

Venteicher, A. S., Abreu, E. B., Meng, Z., Mccann, K. E., Terns, R. M., Veenstra, T. D., et al. (2009). A human telomerase holoenzyme protein required for cajal body localization and telomere synthesis. *Science* 323, 644–648. doi: 10.1126/science.1165357

Voytas, D. F., and Boeke, J. D. (1992). Yeast retrotransposon revealed. *Nature* 358:717. doi: 10.1038/358717a0

Wagstaff, B. J., Kroutter, E. N., Derbes, R. S., Belancio, V. P., and Roy-Engel, A. M. (2013). Molecular reconstruction of extinct LINE-1 elements and their interaction with nonautonomous elements. *Mol. Biol. Evol.* 30, 88–99. doi: 10.1093/molbev/mss202

Wallace, N., Wagstaff, B. J., Deininger, P. L., and Roy-Engel, A. M. (2008). LINE-1 ORF1 protein enhances alu SINE retrotransposition. *Gene* 419, 1–6. doi: 10.1016/j.gene.2008.04.007

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562. doi: 10.1038/nature01262

Wei, W., Gilbert, N., Ooi, S. L., Lawler, J. F., Ostertag, E. M., Kazazian, H. H., et al. (2001). Human L1 retrotransposition: cis preference versus trans complementation. *Mol. Cell. Biol.* 21, 1429–1439. doi: 10.1128/MCB.21.4.1429-1439.2001

Weterings, E., and Van Gent, D. C. (2004). The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA Repair (Amst.)* 3, 1425–1435. doi: 10.1016/j.dnarep.2004.06.003

Williams, E. S., Klingler, R., Ponnaiya, B., Hardt, T., Schrock, E., Lees-Miller, S. P., et al. (2009). Telomere dysfunction and DNA-PKcs deficiency: characterization and consequence. *Cancer Res.* 69, 2100–2107. doi: 10.1158/0008-5472.CAN-08-2854

Williams, G. J., Hammel, M., Radhakrishnan, S. K., Ramsden, D., Lees-Miller, S. P., and Tainer, J. A. (2014). Structural insights into NHEJ: building up an integrated picture of the dynamic DSB repair super complex, one component and interaction at a time. *DNA Repair (Amst.)* 17, 110–120. doi: 10.1016/j.dnarep.2014.02.009

Xie, W., Gai, X., Zhu, Y., Zappulla, D. C., Sternglanz, R., and Voytas, D. F. (2001). Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. *Mol. Cell. Biol.* 21, 6606–6614. doi: 10.1128/MCB.21.19.6606-6614.2001

Yamazaki, H., Tarumoto, Y., and Ishikawa, F. (2012). Tel1(ATM) and Rad3(ATR) phosphorylate the telomere protein Ccq1 to recruit telomerase and elongate telomeres in fission yeast. *Genes Dev.* 26, 241–246. doi: 10.1101/gad.177873.111

Yang, Y., Chen, Y., Zhang, C., Huang, H., and Weissman, S. M. (2002). Nucleolar localization of hTERT protein is associated with telomerase function. *Exp. Cell Res.* 277, 201–209. doi: 10.1006/excr.2002.5541

Yu, G. L., Bradley, J. D., Attardi, L. D., and Blackburn, E. H. (1990). In vivo alteration of telomere sequences and senescence caused by mutated *Tetrahymena* telomerase RNAs. *Nature* 344, 126–132. doi: 10.1038/344126a0

Zaug, A. J., Linger, J., and Cech, T. R. (1996). Method for determining RNA 3' ends and application to human telomerase RNA. *Nucleic Acids Res.* 24, 532–533. doi: 10.1093/nar/24.3.532

Zhang, W., and Durocher, D. (2010). De novo telomere formation is suppressed by the Mec1-dependent inhibition of Cdc13 accumulation at DNA breaks. *Genes Dev.* 24, 502–515. doi: 10.1101/gad.1869110

Zhu, X. D., Niedernhofer, L., Kuster, B., Mann, M., Hoeijmakers, J. H., and De Lange, T. (2003). ERCC1/XPF removes the 3' overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes. *Mol. Cell.* 12, 1489–1498. doi: 10.1016/S1097-2765(03)00478-7

Zhu, Y., Tomlinson, R. L., Lukowiak, A. A., Terns, R. M., and Terns, M. P. (2004). Telomerase RNA accumulates in Cajal bodies in human cancer cells. *Mol. Biol. Cell* 15, 81–90. doi: 10.1091/mbc.E03-07-0525

Zingler, N., Willhoeft, U., Brose, H. P., Schoder, V., Jahns, T., Hanschmann, K. M., et al. (2005). Analysis of 5' junctions of human LINE-1 and Alu retrotransposons suggests an alternative model for 5'-end attachment requiring microhomology-mediated end-joining. *Genome Res.* 15, 780–789. doi: 10.1101/gr.3421505

Zou, S., Ke, N., Kim, J. M., and Voytas, D. F. (1996). The Saccharomyces retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. *Genes Dev.* 10, 634–645. doi: 10.1101/gad.10.5.634

A loopy view of telomere evolution

*Titia de Lange**

Laboratory for Cell Biology and Genetics, The Rockefeller University, New York, NY, USA

About a decade ago, I proposed that t-loops, the lariat structures adopted by many eukaryotic telomeres, could explain how the transition from circular to linear chromosomes was successfully negotiated by early eukaryotes. Here I reconsider this loopy hypothesis in the context of the idea that eukaryotes evolved through a period of genome invasion by Group II introns.

WHY LINEAR CHROMOSOMES?

Before the linear chromosomes of eukaryotes emerged ∼1 Gy ago, circular chromosomes had been successfully used for 2 Gy and they continue to predominate in the most common organisms on this planet (eubacteria and archaea). What were the disadvantages of circular chromosomes that could have ensured the supremacy of an incipient eukaryote with linear chromosomes?

It has been suggested that the answer lies in the first division of meiosis (Ishikawa and Naito, 1999). Meiosis may have first evolved as a mechanism to correct polyploidy arising from genome segregation mistakes. Furthermore, the counterpart of meiosis, the syngamic fusion of haploid cells to form diploids may have been advantageous to survive famine as well as providing greater resistance to the highly mutagenic environment that existed 1 Gy ago. It was argued that switching between diploid and haploid states poses a problem for organisms with circular chromosomes. In the reductional division of meiosis, the homologous chromosomes are held together by their chiasmata where recombination has generated a crossover between homologous chromatids. A single meiotic crossover (or any uneven number of crossovers) generates a dicentric circle, which will lead to non-disjunction of the homologs. As this problem is circumvented with linear chromosomes, linearization may have provided a selective advantage.

This argument ignores systems like the bacterial XerD/C resolution machinery, which efficiently cuts dimeric circular chromosomes at specific dif sites (Barre et al., 2001). A similar system could have been used to resolve dimeric circles in the meiosis of early eukaryotes. Below I propose that linear chromosomes arose as the consequence of the invasion of a circular genome with repeat sequences. Once formed, linear chromosomes may have had advantages under certain circumstances but their raison d'etre, I argue, is found in the way linear DNAs with repetitive sequences at their termini escape re-circularization through ligation.

T-LOOPS AS A PRIMORDIAL TELOMERE SYSTEM

The modern eukaryotic telomere is a complex system of two critical components (**Figure 1A**). The maintenance of telomeres requires the telomerase reverse transcriptase with its associated RNA template, which dictates the sequence repeats at the chromosome ends. This system ensures the presence of telomeric repeats despite their constant erosion with conventional replication [the "end-replication problem" (Watson, 1972; Olovnikov, 1973)]. Furthermore, the action of telomerase endows every chromosome end with binding sites for sequence specific telomere proteins. It is the presence of such telomeric proteins that protect the telomeres from being detected as sites of DNA damage, thus solving the "end-protection problem" (de Lange, 2009). Without telomeric proteins, the telomeric repeats do nothing to repress the DNA damage response and chromosome ends become

substrates for DNA repair. *Vice versa*, without the telomerase-derived telomeric repeats, the telomeric proteins are incapable of preventing genomic mayhem. It seems unlikely that both components of the telomere system, telomerase with its RNA template on the one hand and the sequence specific binding proteins that recognize the DNA version of this template sequence on the other, arose simultaneously. Of course, intermediate steps can be envisaged. For instance, the earliest telomerase RNA may have dictated sequences that happened to interact with a pre-existing DNA binding protein capable of some protection. But a much simpler scenario is suggested by the t-loop structure, the lariats found at present-day telomeres that play a critical role in telomere protection.

T-loops are double-stranded looped structures that are formed through the strand-invasion of the telomere terminus into the telomeric repeats (**Figure 1B**). Telomeres generally have a $3'$ overhang that facilitates the formation of t-loops and modern telomeres contain specific proteins that are critical for the formation/stabilization of this structure. In mammalian cells, t-loops block DNA repair by non-homologous end joining (NHEJ; see **Figure 1B**), which would generate end-to-end fused chromosomes. T-loops also represent a powerful mechanism for hiding the chromosome end from the ATM kinase-dependent DNA damage response, which would result in cell cycle arrest. Similarly, in the incipient eukaryote, a modified version of the t-loop structure could have protected the ends from resident nuclease and ligases (**Figure 1C**). Furthermore, if the structure at the base of the t-loop lacked single-stranded DNA, the chromosome ends would not have activated the bacterial SOS response, which detects DNA damage when ssDNA is formed (Baharoglu and Mazel, 2014). Although strand-invasion would require a single-stranded overhang and thus create a single stranded D (displacement) loop, a t-loop lacking ssDNA can be generated if the D loop is converted into double-stranded DNA by fill-in DNA synthesis (see below, **Figure 1C**).

As discussed in detail previously (de Lange, 2004), t-loops not only solve the end-protection problem, they also provide a mechanism for extending the terminal sequences without the aid of telomerase (**Figure 1C**). The structure at the base of the t-loop is identical to the structure of a replication fork. *De novo* recruitment of replication enzymes could ensure that the end is extended, solving the end-replication problem. These steps would not require evolution of new factors because the machinery that mediates replication restart events in bacteria is able to execute them.

The solution to the end-replication problem afforded by the t-loop structure is related to the telomere maintenance systems observed under certain circumstances in present-day eukaryotes. An example is the alternative lengthening of telomeres (ALT) pathway for telomere maintenance, which is a pathway active in a subset of human cancers that maintains telomeres by homologous recombination (HR). Although the exact mechanism of telomere elongation by ALT is not known, one of the proposed mechanisms involves extension of telomeres in the t-loop configuration (see de Lange, 2004).
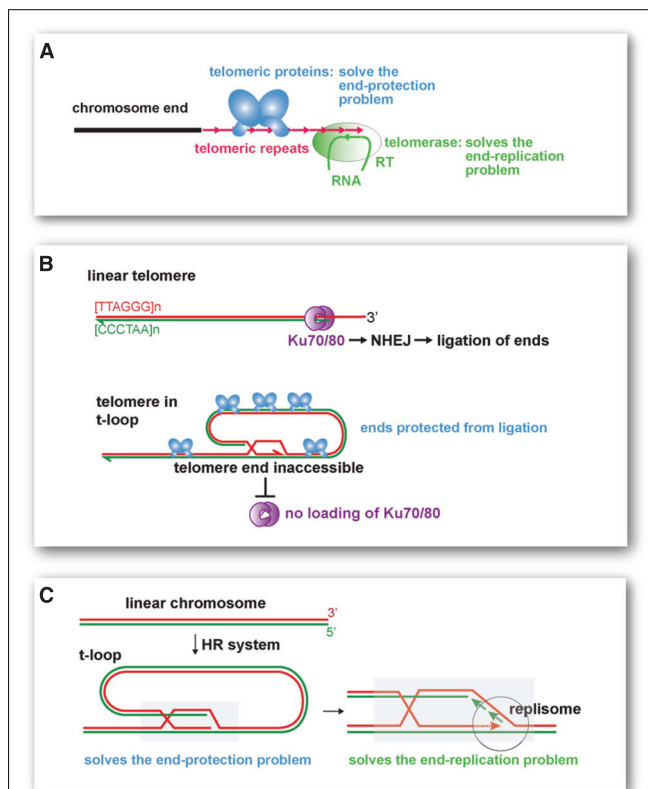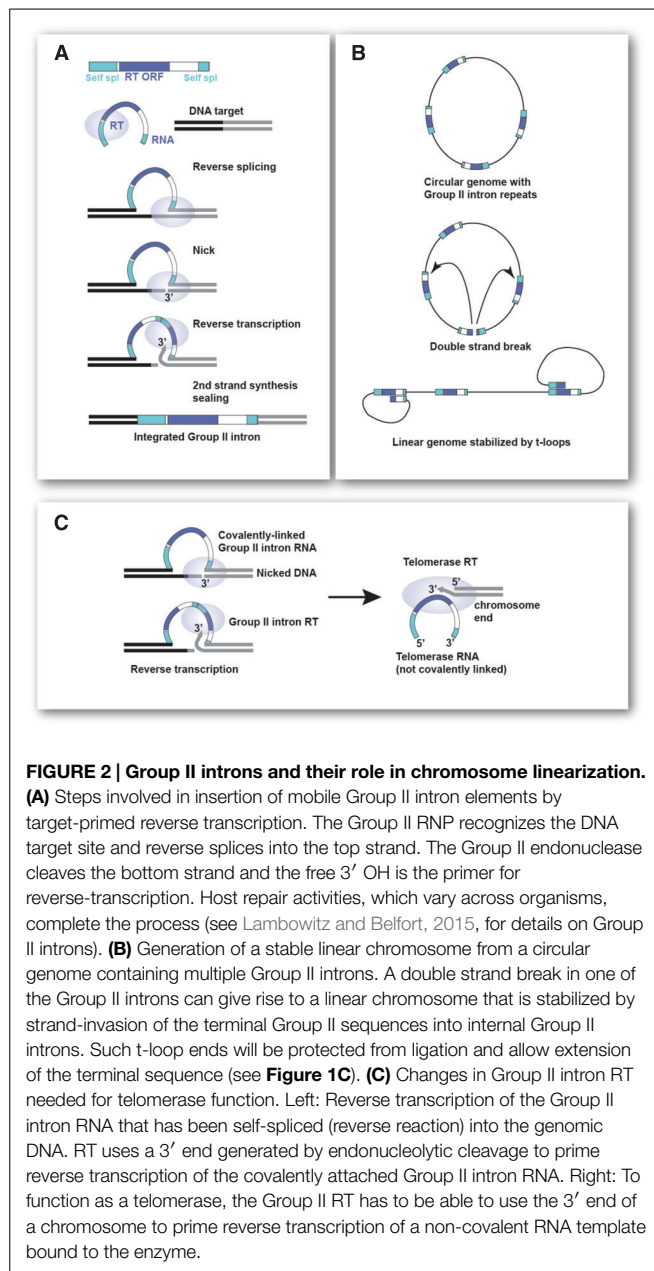

FIGURE 1 | Modern telomeres and their proposed t-loop precursor. (A) Current telomeres require a telomerase that synthesizes the telomeric repeats and counteracts the end-replication problem. They also require telomere specific proteins that recognize the telomerase products at chromosome ends and protect the ends from the DNA damage response (solving the end-protection problem). **(B)** Mammalian telomeres form t-loops, which sequester the telomere end and prevent ligation by NHEJ. Telomeric proteins (blue, e.g., TRF2) are needed to form the t-loop structure. Telomeric proteins also protect telomeres from other DNA repair pathways and prevent the activation of the DNA damage signaling pathways (not shown). **(C)** The t-loop based primordial telomere. The proposed precursor to modern telomeres is a t-loop structure as depicted. The critical aspect of the t-loop is the strand-invasion (mediated by homologous recombination factors) of the telomere end into a repeated homologous sequence (gray box). The invaded repeat could either be close to the end or chromosome-internal. Any repetitive sequence of sufficient length to allow homologous recombination can fulfill this function. Although the strand-invasion would require a $3'$ overhang, recruitment of a replisome and DNA synthesis would generate a structure lacking single-stranded DNA (shown on the left). The strand-invasion of the end blocks NHEJ and ssDNA recognition systems (e.g., SOS response), thus solving the end-protection problem. When the terminal sequences is extended by DNA replication, the end-replication problem is solved (right).

GROUP II INTRONS AND THE INEVITABILITY OF LINEAR CHROMOSOMES

As outlined above, the incipient eukaryotes could have had stable linear chromosomes without the need for telomerase or telomere specific proteins. The only necessity would have been terminal sequences that are homologous to more internal sequences so that the critical strand-invasion event can take place (**Figure 1C**). There is no need for an array of repeats at the

FIGURE 2 | Group II introns and their role in chromosome linearization.
(A) Steps involved in insertion of mobile Group II intron elements by target-primed reverse transcription. The Group II RNP recognizes the DNA target site and reverse splices into the top strand. The Group II endonuclease cleaves the bottom strand and the free 3′ OH is the primer for reverse-transcription. Host repair activities, which vary across organisms, complete the process (see Lambowitz and Belfort, 2015, for details on Group II introns). **(B)** Generation of a stable linear chromosome from a circular genome containing multiple Group II introns. A double strand break in one of the Group II introns can give rise to a linear chromosome that is stabilized by strand-invasion of the terminal Group II sequences into internal Group II introns. Such t-loop ends will be protected from ligation and allow extension of the terminal sequence (see **Figure 1C**). **(C)** Changes in Group II intron RT needed for telomerase function. Left: Reverse transcription of the Group II intron RNA that has been self-spliced (reverse reaction) into the genomic DNA. RT uses a 3′ end generated by endonucleolytic cleavage to prime reverse transcription of the covalently attached Group II intron RNA. Right: To function as a telomerase, the Group II RT has to be able to use the 3′ end of a chromosome to prime reverse transcription of a non-covalent RNA template bound to the enzyme.

ends. The ends could invade an internal copy of the repeat with exactly the same outcome of protection and replicative extension of the termini. But where did these repeats come from?

Although any repeat element of sufficient length and present at the required copy number would in principle provide circular genomes with the same high chance of becoming linear, I propose that mobile Group II self-splicing elements (Group II introns) are a good candidate for the repeats that led to chromosome linearization. Group II introns are the proposed ancestors of introns and non-LTR retrotransposons. These elements use reverse splicing and reverse transcription to efficiently integrate into specific DNA target sites (see **Figure 2A**). They can also spread through the genome by a similar, but less efficient reaction at ectopic sites.

Cavalier-Smith (1991) and Koonin (2006) have argued that the phagocytosis of an α-proteobacterial cell by archaeal (or actinobacterial) eukaryotic precursor could have been accompanied by massive invasion of Group II introns (Martin and Koonin, 2006). The Group II introns residing in the genome of the ingested future mitochondrion are proposed to have colonized the host genome resulting in a large number of repetitive elements (see **Figure 2A**, for schematic of Mobile Group II introns). The insertion of Group II introns into coding regions could have provided the selective pressure for the invention of the nucleus as a compartment where introns can be removed from pre-mRNAs before they are used by ribosomes (Koonin, 2006; Martin and Koonin, 2006; but see Cavalier-Smith, 2010, for a dissenting opinion). The removal of the Group II introns may have initially involved protein assisted self-splicing with protein-dependent splicing evolving later. The invasion of Group II introns may have also led to nonsense-mediated decay as a way to remove intron-bearing transcripts from ribosomes but generation of a nuclear envelope combined with a system that links mRNA transport to the completion of splicing is a more definitive solution (Koonin, 2006; Martin and Koonin, 2006).

I propose that accumulation of Group II introns in the genome could also have generated the condition under which linear chromosomes became inevitable. Consider a future eukaryote with a circular genome full of Group II introns (**Figure 2B**). If a double-strand break occurred in one of these repeats, the bacterial DNA repair machinery would have acted in one of two ways. Either the ends would be ligated back together by some form of NHEJ or the ends would have been processed by the HR machinery of the host. The strand-invasion by HR would have had to take place in other copies of Group II introns, since they would be the only homologous target. The initiation of recombination would have generated a terminal loop, a t-loop, of variable size and sequence composition (**Figure 2B**). A linear chromosome containing such t-loops at each end would be impervious to re-ligation and terminal sequence attrition would be counteracted by extension of the ends using the mechanism shown in **Figure 1**. Thus, once formed, such a linear chromosome would be stable. The chance of this scenario playing out is greater as the number of repeats increases. Once a linear with t-looped ends is formed, the path back to a circular genome is difficult because the ends are protected. During DNA replication of the t-loop, the ends would be free to undergo ligation thus reforming a circular chromosome. But this would only happen if the replication forks synchronously dislodged the t-loops. Even if re-circularization happened, there would be a good chance of another double strand break (DSB) occurring in a Group II intron leading again to a linear state. Thus, the earliest eukaryotic chromosomes may have existed predominantly as linears that occasionally were converted to a circular state.

FROM DISPERSED GROUP II INTRONS TO TELOMERE SPECIFIC REPEATS

After a period of semi-stable linear chromosomes, a more permanent linear state would have required the gradual evolution

toward the system used by modern telomeres. Two major steps are needed for this to happen. First, the telomerase system would have to evolve. The telomerase reverse transcriptase is likely derived from the reverse transcriptase (RT) of Group II introns (Nakamura and Cech, 1998; Dlakić and Mushegian, 2011). In order to become a true telomerase, the Group II RT would have had to gain the ability to use the 3′ end of a chromosome as a primer for reverse transcription of its associated RNA. Furthermore, it would have needed to use its RNA as a template even though it is not covalently linked to the target site (see **Figure 2C**). Once these modifications were made to one of the Group II intron RTs, the Group II intron sequences that became the future telomerase RNA could have evolved to cooperate with this enzyme. Thus, one Group II intron would encode the telomerase RT and another would encode the telomerase RNA. Both can now evolve into new genes that execute the terminal extension efficiently and repeatedly without the encoding genes being burdened by the requirements for self-splicing and other Group II intron functions. The RNA component can now change to 1. associate only with the telomerase RT; 2. specify a short sequence as a template for terminal sequence addition rather than the whole RNA; and 3. enable synthesis of an array of the same short repeats at every chromosome end.

The resulting system would have created linear chromosomes with arrays of short repeats that are telomere specific and no longer have homology to Group II introns. At this stage, the t-loops will only form within the telomeric repeat array since this is the only homologous sequence available in the genome.

Once all chromosome ends have the same sequence, the incipient eukaryote could evolve proteins that recognize this sequence. These early telomeric proteins are likely to be selected for their ability to mediate the t-loop structure since this was the critical aspect of telomere protection. They may also have had the ability to bind to the telomerase RT, thereby ensuring the maintenance of the telomeric repeats. These features are still present in modern telomeres. For instance, the telomeric

repeat binding factor 2 (TRF2) component of the mammalian telomeric complex (shelterin) enables t-loop formation whereas other factors in shelterin recruit telomerase (Nandakumar et al., 2012; Zhong et al., 2012; Doksani et al., 2013; Sexton et al., 2014).

WHY SUCH ELABORATE TELOMERES?

The scenario sketched above raised the question why telomeres became so elaborate. Why not stick with the simple t-loop mode? Why have telomerase and a host of telomeric proteins? The same question could be asked about intron splicing which evolved from simple self-splicing based on RNA catalysis to elaborate spliceosomal complexes with a myriad of RNA and protein components. In part, the answer must be that most processes in eukaryotes generally evolve toward complexity, presumably because complexity provides more regulatory opportunities and perhaps also because there is no selective pressure to enforce simplicity.

With regard to telomeres, there is an additional consideration. Telomeres need to adapt to the DNA repair pathways and DNA damage signaling pathways that evolve in their host cells. These pathways have become increasingly complex and more varied. In response, telomeres have attained additional bells and whistles to help protect chromosome ends from these pathways (de Lange, 2009). In contrast, the end-replication problem has remained the same. As a result, the way telomeres deal with the end-protection problem and the protein complexes used for this task are highly variable while telomerase has been conserved.

REFERENCES

Baharoglu, Z., and Mazel, D. (2014). SOS, the formidable strategy of bacteria against aggressions. *FEMS Microbiol. Rev.* 38, 1126–1145. doi: 10.1111/1574-6976.12077

Barre, F. X., Søballe, B., Michel, B., Aroyo, M., Robertson, M., and Sherratt, D. (2001). Circles: the replication-recombination-chromosome segregation connection. *Proc. Natl. Acad. Sci. U.S.A.* 98, 8189–8195. doi: 10.1073/pnas.111008998

Cavalier-Smith, T. (1991). Intron phylogeny: a new hypothesis. *Trends Genet.* 7, 145–148. doi: 10.1016/0168-9525(91)90102-V

Cavalier-Smith, T. (2010). Origin of the cell nucleus, mitosis and sex: roles of intracellular coevolution. *Biol. Direct.* 5, 7. doi: 10.1186/1745-6150-5-7

de Lange, T. (2004). T-loops and the origin of telomeres. *Nat. Rev. Mol. Cell Biol.* 5, 323–329. doi: 10.1038/nrm1359

de Lange, T. (2009). How telomeres solve the end-protection problem. *Science* 326, 948–952. doi: 10.1126/science.1170633

Dlakić, M., and Mushegian, A. (2011). Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. *RNA* 17, 799–808. doi: 10.1261/rna.2396011

Doksani, Y., Wu, J. Y., de Lange, T., and Zhuang, X. (2013). Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation. *Cell* 155, 345–356. doi: 10.1016/j.cell.2013.09.048

Ishikawa, F., and Naito, T. (1999). Why do we have linear chromosomes? A matter of Adam and Eve. *Mutat. Res.* 434, 99–107. doi: 10.1016/S0921-8777(99)00017-8

Koonin, E. V. (2006). The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate. *Biol. Direct.* 1, 22. doi: 10.1186/1745-6150-1-22

Lambowitz, A. M., and Belfort, M. (2015). Mobile bacterial group II introns at the crux of eukaryotic evolution. *Microbiol. Spectr.* doi: 10.1128/microbiolspec.mdna3-0050-2014 [Epub ahead of print].

Martin, W., and Koonin, E. V. (2006). Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 440, 41–45. doi: 10.1038/nature04531

Nakamura, T. M., and Cech, T. R. (1998). Reversing time: origin of telomerase. *Cell* 92, 587–590. doi: 10.1016/S0092-8674(00)81123-X

Nandakumar, J., Bell, C. F., Weidenfeld, I., Zaug, A. J., Leinwand, L. A., and Cech, T. R. (2012). The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity. *Nature* 492, 285–289. doi: 10.1038/nature11648

Olovnikov, A. M. (1973). A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *J. Theor. Biol.* 41, 181–190. doi: 10.1016/0022-5193(73)90198-7

Sexton, A. N., Regalado, S. G., Lai, C. S., Cost, G. J., O'Neil, C. M., Urnov, F. D., et al. (2014). Genetic and molecular identification of three human TPP1 functions in telomerase action: recruitment, activation, and homeostasis set point regulation. *Genes Dev.* 28, 1885–1899. doi: 10.1101/gad.246819.114

Watson, J. D. (1972). Origin of concatemeric T7 DNA. *Nat. New Biol.* 239, 197–201. doi: 10.1038/newbio239197a0

Zhong, F. L., Batista, L. F., Freund, A., Pech, M. F., Venteicher, A. S., and Artandi, S. E. (2012). TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends. *Cell* 150, 481–494. doi: 10.1016/j.cell.2012.07.012

Advantages of publishing in Frontiers

OPEN ACCESS

Articles are free to read, for greatest visibility

COLLABORATIVE PEER-REVIEW

Designed to be rigorous – yet also collaborative, fair and constructive

FAST PUBLICATION

85

Average 85 days from submission to publication (across all journals)

COPYRIGHT TO AUTHORS

No limit to article distribution and re-use

TRANSPARENT

Editors and reviewers acknowledged by name on published articles

SUPPORT

By our Swiss-based editorial team

IMPACT METRICS

Advanced metrics track your article's impact

GLOBAL SPREAD

5'100'000+ monthly article views and downloads

LOOP RESEARCH NETWORK

Our network increases readership for your article

Frontiers
EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on