# Origin, Diversity, and Evolution of Telomere Sequences in Plants

*Vratislav Peska[1]\* and Sònia Garcia[2]\**

[1] *Department of Cell Biology and Radiobiology, The Czech Academy of Sciences, Institute of Biophysics, Brno, Czechia,*
[2] *Institut Botànic de Barcelona (IBB, CSIC-Ajuntament de Barcelona), Barcelona, Spain*

Telomeres are basic structures of eukaryote genomes. They distinguish natural chromosome ends from double-stranded breaks in DNA and protect chromosome ends from degradation or end-to-end fusion with other chromosomes. Telomere sequences are usually tandemly arranged minisatellites, typically following the formula $(T_xA_yG_z)_n$. Although they are well conserved across large groups of organisms, recent findings in plants imply that their diversity has been underestimated. Changes in telomeres are of enormous evolutionary importance as they can affect whole-genome stability. Even a small change in the telomere motif of each repeat unit represents an important interference in the system of sequence-specific telomere binding proteins. Here, we provide an overview of telomere sequences, considering the latest phylogenomic evolutionary framework of plants in the broad sense (Archaeplastida), in which new telomeric sequences have recently been found in diverse and economically important families such as Solanaceae and Amaryllidaceae. In the family Lentibulariaceae and in many groups of green algae, deviations from the typical plant telomeric sequence have also been detected recently. Ancestry and possible homoplasy in telomeric motifs, as well as extant gaps in knowledge are discussed. With the increasing availability of genomic approaches, it is likely that more telomeric diversity will be uncovered in the future. We also discuss basic methods used for telomere identification and we explain the implications of the recent discovery of plant telomerase RNA on further research about the role of telomerase in eukaryogenesis or on the molecular causes and consequences of telomere variability.

**Keywords:** *Allium*, *Cestrum*, circular chromosomes, *Genlisea*, green algae, linear chromosomes, telomerase, telomeres

## INTRODUCTION

Telomeres are nucleoprotein structures at the very ends of linear eukaryotic chromosomes. They solve two major end-problems at the same time. The first is about chromosome end protection. It is estimated that normal human cells must repair at least 50 endogenous double-stranded breaks (DSBs) per cell per cell-cycle (Vilenchik and Knudson, 2003). Telomeres distinguish the natural chromosomal ends from harmful DSBs and prevent their ectopic repair, e.g., by end-to-end fusions of chromosomes (vanSteensel and deLange, 1997). The second is the end-replication problem that
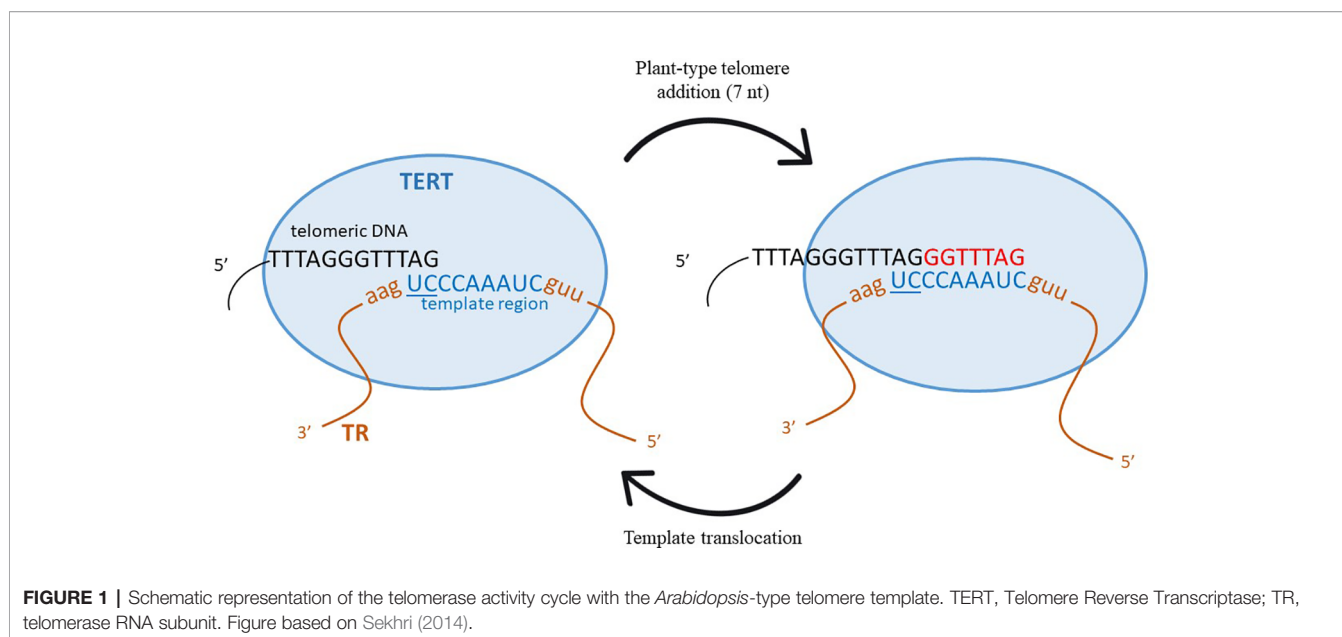
deals with the maintenance of proper telomere lengths. This was recognized independently by two researchers (Watson, 1972; Olovnikov, 1973). Since replicative DNA-dependent DNA polymerases cannot complete DNA synthesis at the very ends of chromosomes, compensation for replicative telomere sequence loss must come from an RNA-dependent DNA polymerase. This enzyme, called telomerase, together with the first telomere minisatellite sequence, was discovered in the ciliate *Tetrahymena* (Blackburn and Gall, 1978; Greider and Blackburn, 1985). However, this is only one aspect of telomere length maintenance. The epigenetic regulation of telomere length homeostasis, including interaction of long noncoding telomeric repeat containing RNA and exonuclease activity pathways, have also been extensively studied due to its therapeutic potential (Wellinger et al., 1996; Polotnianka et al., 1998; Pfeiffer and Lingner, 2012).

Telomerase, the enzyme in charge of adding telomere repeat sequences to the 3' end of telomeres, is a conserved complex enzyme with numerous components [its structure has been recently reviewed by (Wang et al., 2019), and specifically for plants, by (Majerska et al., 2017)]. In principle, only two main components are essential for telomerase enzymatic activity, a catalytically active protein component, called telomere reverse transcriptase (TERT), and a template component, formed by the telomerase RNA subunit (TR). While TERT is evolutionarily quite well conserved, TR is very variable, with lengths ranging from ca. 150 nt (*Tetrahymena*) to more than 2,000 nt (fungi from genus *Neurospora*). Only a short region in the whole TR molecule serves as a template for newly synthesized telomere DNA (Greider and Blackburn, 1985; Qi et al., 2013). This region in TR is usually formed by a complete telomere motif followed by a partial one, the latter serving as an annealing region for the existing telomere DNA. Although, in principle, only a single extra nucleotide is needed (as a partial motif), usually more than

one is found. For example, two extra nucleotides form the annealing motif in mice or five in human (Blasco et al., 1995; Feng et al., 1995). In plants, however, the size of the template region is variable, e.g., two in *Arabidopsis thaliana*, seven in *Arabis* sp. or six in *Nicotiana* (Fajkus et al., 2019). The other TR regions have structural, regulatory and protein interactive functions [reviewed in (Podlevsky and Chen, 2016)]. See also a schematic depiction of telomerase and its activity cycle in **Figure 1**.

## HOW VARIABLE ARE TELOMERE SEQUENCES?

Telomere sequences are usually short minisatellites tandemly arranged, typically following the formula $(T_xA_yG_z)_n$. The minisatellite arrangement originates from the way in which telomerase synthesizes the DNA, in short, and mostly identical motifs, one by one. Several hypotheses consider that such an arrangement is important because it promotes the recognition of telomere specific proteins by homo- and heterodimers [e.g., (Hofr et al., 2009; Visacka et al., 2012)] and for the potential to form G-quadruplexes that may stabilize chromosome ends or serve as substrates for telomere-specific proteins (Spiegel et al., 2020; Tran et al., 2013). Telomere sequences are well conserved through evolution, and large groups of organisms use the group-typical telomere motif to build their telomere DNA. A gradually increasing number of studies and large screenings have shown that all tested vertebrates and many basal metazoans use TTAGGG (Meyne et al., 1989; Traut et al., 2007) while Euarthropoda (arthropods), including Hexapoda (insects), have TTAGG (Frydrychova et al., 2004; Vitkova et al., 2005). Steadily, numerous exceptions are accumulating over time, e.g., $(A(G)_{1-8})$ in *Dictyostelium* (Emery and Weiner, 1981), TTAGGC in *Ascaris*



**FIGURE 1 |** Schematic representation of the telomerase activity cycle with the *Arabidopsis*-type telomere template. TERT, Telomere Reverse Transcriptase; TR, telomerase RNA subunit. Figure based on Sekhri (2014).

*lumbricoides* (Nematoda) (Muller et al., 1991), TCAGG in Coleoptera (beetles) (Mravinac et al., 2011), TAGGG/TAAGG/TAAGGG in *Giardia* (diplomonads) (Uzlikova et al., 2017), or TTNNNNAGGG in *Yarrowia* clade (yeasts) (Cervenak et al., 2019). Moreover, telomerase-independent systems, in which the minisatellite telomere sequence has been lost and substituted by complex repeats, are represented, for example, by *Diptera* and Chironomidae (reviewed in (Mason et al., 2016)). For a general review on eukaryotic telomere sequence see (Fajkus et al., 2005; Fulneckova et al., 2013).

Telomere composition in plants is even more diverse. Here we use the term "plants" in a broad sense, also known as Archaeplastida or kingdom Plantae *sensu lato*, and comprising Rhodophyta (red algae), Glaucophyta, the Chlorophyte algae grade and the Streptophyte algae grade (altogether known as green algae), and Embryophyta (land plants) (One Thousand Plant Transcriptomes Initiative, 2019). The typical telomere plant sequence is TTTAGGG, also called *Arabidopsis*-type (or simply, plant-type) since it was discovered in *Arabidopsis thaliana* (Richards and Ausubel, 1988) and now in many other species across almost all plant orders. Although TTTAGGG is still the most frequent, there is significant variability in telomere sequences in red and green algal lineages. As for red algae (Rhodophyta), telomere sequence information is mostly missing or fragmentary, although some telomere candidates have been discovered *in silico*, such as AATGGGGGG for *Cyanidioschyzon merolae* (Nozaki et al., 2007), TTATT(T)AGGG for *Galdieria sulphuraria* (Fulneckova et al., 2013); TTAGGG has been found in genomic reads of *Porphyra umbilicalis* (Fulneckova et al., 2013), but more evidence is needed to confirm their terminal position on chromosomes. Telomere diversity in green algae reflects both dynamic changes and its paraphyletic character. Although TTTAGGG prevails in Chlorophyta, such as in genera *Ostreococcus* (Derelle et al., 2006) and *Chlorella* (Higashiyama et al., 1995), many other divergent motifs have been detected there too, such as TTAGGG in genus *Dunaliella* and *Stephanosphaeria* (Fulneckova et al., 2012), and TTTTAGGG in *Chlamydomonas* (Petracek et al., 1990). In basal Streptophyta (Klebsormidiophyceae) progressive changes in motifs from TTTAGGG to TTTTAGGG and TTTTAGG have been described. The presence of TTAGGG in Rhodophyta and Glaucophyta leads to the hypothesis that this is the ancestral motif in plants (Archaeplastida) (Fulneckova et al., 2013).

Concerning land plants, one of the first screenings performed showed that the *Arabidopsis*-type sequence was the most common and was mostly conserved through their phylogeny (Cox et al., 1993; Fuchs et al., 1995), although some of these authors had already detected several exceptions in the family Amaryllidaceae (former Alliaceae), in which the *Arabidopsis*-type sequence was absent in several species. Later, the first telomere sequence unusual for land plants, the vertebrate-type TTAGGG, was characterized in *Aloe* and in some other Asparagales (Weiss and Scherthan, 2002; Puizina et al., 2003; Sykorova et al., 2003c). A hypothesis about repeated losses and recoveries of the TTTAGGG and TTAGGG telomere sequence

in Asparagales was formulated (Adams et al., 2001). With the postrefinement of order Asparagales in the APGIII (Angiosperm Phylogeny Group 2009) (Bremer et al., 2009), it was shown that only two major evolutionary switches in telomere sequence composition occurred (rather than several repeated losses and gains), in the following order: the first one in family Iridaceae, in which a shift from the plant-type TTTAGGG to the vertebrate-type TTAGGG happened, followed by families Xeronemataceae, Asphodelaceae and the core Asparagales (including Amarillidaceae s.l and Asparagaceae s.l.); and the second one within subfamily Allioideae (formerly treated as a separate family, Alliaceae) in which a completely new telomere sequence emerged, CTCGGTTATGGG (Fajkus et al., 2016). Outside Asparagales, new telomere sequences have also been detected in land plant groups as disparate as (i) Solanaceae, in which the telomere sequence of *Cestrum elegans* TTTTTTAGGG was described (Sykorova et al., 2003a; Sykorova et al., 2003b; Peska et al., 2008; Peska et al., 2015) and (ii) Lentibulariaceae, where genus *Genlisea* showed a remarkable diversity with some species characterized by the *Arabidopsis*-type telomere repeats while others exhibited intermingled sequence variants TTCAGG and TTTCAGG (Tran et al., 2015).

Despite all the telomere motif exceptions detected, the real diversity in telomeric sequences in land plants is probably greatly underestimated. A recent publication (Vitales et al., 2017), in which a screening of land plant telomere sequences was performed, found that telomere sequences were only known clearly for less than 10% of the species and 40% of the genera contained in the Plant rDNA database (www.plantrdnadatabase.com), a resource providing molecular cytogenetics information on land plants (Garcia et al., 2012). A summary of telomere sequence distribution in plants, following APG IV (The Angiosperm Phylogeny Group, 2016) (Byng et al., 2016), as well as the most recent plant phylogeny (One Thousand Plant Transcriptomes Initiative, 2019) is found in **Figure 2**.

## FROM SCREENINGS TO DISCOVERY: HOW TELOMERIC MOTIFS CAN BE IDENTIFIED?

The evidence that a given candidate sequence is a real telomeric one includes several steps that properly declare its localization at all chromosomal termini, and eventually the involvement of telomerase in its synthesis. Molecular cytogenetics (mostly by Fluorescence *in situ* Hybridization, FISH) has become important for visualizing the terminal localization of labeled probes of candidate sequences at all chromosomal termini. However, standalone FISH it is not enough to prove the very terminal position. For example, AcepSAT356 [a 356bp-long satellite from *Allium cepa*, (Peska et al., 2019)] was proposed in onion as the telomere candidate, based on results from FISH analysis (Pich and Schubert, 1998). Nevertheless, its apparent terminal location by *in situ* has never been convincingly linked to telomere function. Actually, the discovery of the *Allium* minisatellite
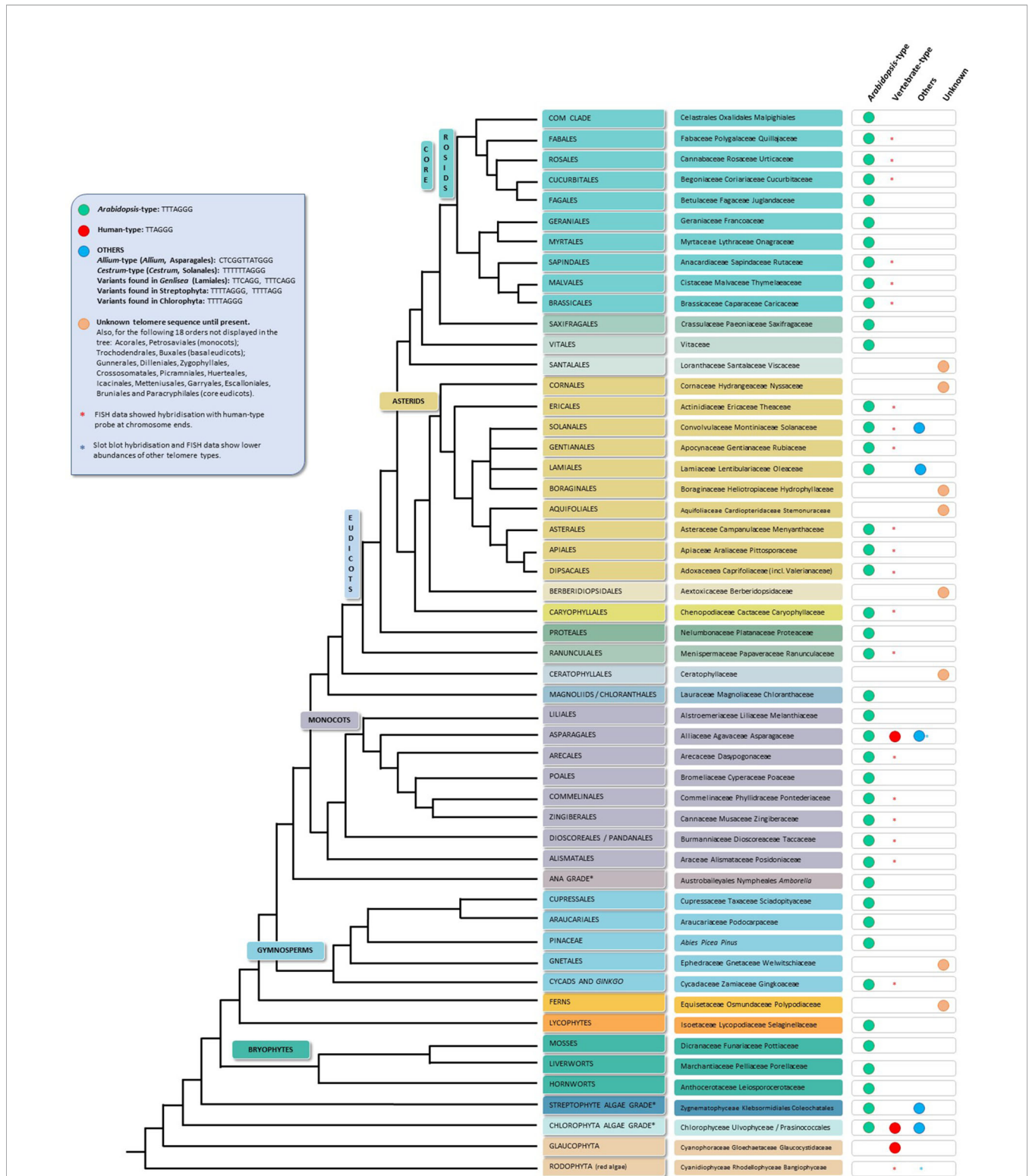
**FIGURE 2 |** Telomere motifs in Archaeplastida (plants in the broad sense), based on the APG IV (The Angiosperm Phylogeny Group 2016) and on the One Thousand Plant Transcriptomes Initiative (2019). Branch lengths do not express real time scales. For simplicity and to save space, certain polyphyletic "groups" (grades) marked with an asterisk in the tree have been represented by a single branch; for the same reason, several minor orders (listed in the blue square at the left upper side of the figure) are not depicted on the tree. The first tip label usually refers to plant orders and in a few cases, to divisions, grades and even families; the second label displays representative families and in a few cases, representative orders or genera.

telomere sequence CTCGGTTATGGG and telomerase would mean that AcepSAT356 is subterminal (Fajkus et al., 2019). Positive FISH telomeric signals can also mask tiny changes in telomere motifs such as single nucleotide polymorphisms, or false-negative results may result from short telomeres being beneath the detection limit of the technique.

There are two additional approaches that determine the terminal position at greater resolution than FISH; these are based on exonuclease BAL31 activity. The first is the classical Terminal Restriction Fragment (TRF) analysis, in which samples treated by BAL31 show progressive shortening of terminal fragments and a decrease in signal intensity with increasing time of exonuclease treatment. The subsequent analysis of fragment lengths is performed by Southern-blot hybridization (Fojtova et al., 2015). The second is comparative genome skimming (NGS data) of nondigested and BAL31-digested genomic DNA, in parallel. In the BAL31 treated dataset, there is a significant under-representation of telomere sequences, therefore the terminal sequences are identified by comparison with the untreated dataset, using bioinformatics tools RepeatExplorer or Tandem Repeats Finder [a pipeline called BAL31-NGS (Benson, 1999; Novak et al., 2010; Peska et al., 2017)].

The other important test of a given telomere sequence candidate in a species is the demonstration of telomerase activity. In this, a useful experimental approach, developed first for human cells, is the Telomere Repeat Amplification Protocol (TRAP) (Kim et al., 1994), followed by sequencing of the detected products (Peska et al., 2015; Fajkus et al., 2016), which is a little less sensitive to false-positive results than FISH. All these methods, including FISH (Fuchs et al., 1995; Shibata and Hizume, 2011) and others such as slot-blot hybridization (Sykorova et al., 2003c), and TRAP (Fulneckova et al., 2012; Fulneckova et al., 2016), can be used to screen for telomeres across wide groups of complex organisms, including plants. However, only a combination of suitably chosen methods can convincingly lead to a conclusion about the telomere function of a candidate sequence, since results base on a single approach might be misleading. A more complete overview of the strategies for *de novo* telomere candidate sequence identification, including the very first attempt in *Tetrahymena* (Greider and Blackburn, 1985) are summarised in a methodological article, with emphasis on the NGS approach used in plants with extremely large genomes (Peska et al., 2017).

## IS THERE HOMOPLASY IN TELOMERE SEQUENCES?

The ancestral telomere sequence is thought to be TTAGGG and is the most commonly found across the tree of life (Fulneckova et al., 2013). Yet, it seems clear that the frequency of homoplasy in telomere motif evolution is relatively high. For example, short, simple motifs like the plant-type TTTAGGG have appeared independently and repeatedly in cryptomonads, oomycete fungi, and alveolates; similarly, the vertebrate-type TTAGGG

has emerged secondarily in certain groups of plants (Asparagales, Rodophyta and Chlorophyta algae) (Sykorova et al., 2003c; Fulneckova et al., 2012; Fulneckova et al., 2013; Somanathan and Baysdorfer, 2018). The reason some telomere sequences have emerged more frequently than other, usually more complex sequences is probably related to selection pressures, which would favor accuracy for a particular sequence-specific DNA-protein interaction (Forstemann et al., 2003). If there was a change in each telomere motif, interference in the telomeric nucleoprotein structure would necessarily lead to genome instability. This is the reason telomere sequences are so evolutionary stable, comprising very few novel and successful sequences, a pattern consistent with the idea of repeated losses and the emergence of the typical telomere sequences, as proposed for Asparagales (Adams et al., 2001).

The finding of homoplasy across telomere sequences raises the question, what are the molecular causes and processes taking place during these shifts? A change in telomere sequence, despite seeming trivial in some cases (e.g., one extra T), may cause serious interference with genome integrity, because of a disturbed balance in the telomere DNA-protein interactions. It is also unclear whether a change in telomere sequence may have any evolutionary advantage; in this regard, (Tran et al., 2015) suggested that the appearance of a "methylatable" cytosine in a G-rich telomere strand would raise the possibility of regulation by epigenetic modification.

## WHAT ARE THE MOLECULAR REASONS FOR CHANGES IN THE TELOMERE MOTIFS?

To explain telomere sequence change, the first candidate is the template subunit of telomerase, telomerase RNA (TR). The previously identified TR from yeast and vertebrates belongs to a different group of transcripts, whose connecting feature was that they were transcribed by RNA polymerase II (Pol II)—in all but ciliates; this used to be the single exception from Pol II transcripts before publication of the land plant TR identification [reviewed in (Podlevsky and Chen, 2016)]. By using the relatively long telomere motif of *Allium* to look for its TR within the total RNA sequence data pool, Fajkus et al. (2019) showed that a previously characterized noncoding RNA involved in the stress reaction in *A. thaliana,* called AtR8, was indeed the telomerase RNA subunit (Wu et al., 2012; Fajkus et al., 2019). It was a transcript of RNA polymerase III (Pol III) containing the corresponding regulatory elements in its promoter structure. For a long time, researchers expected that plant TR would be so divergent that it would be impossible to identify it based on a homology search (Cifuentes-Rojas et al., 2011). However, a certain degree of similarity was successfully used to identify a common TR in several *Allium* species with comparative Blast. Surprisingly, sequence homology, the presence of the same regulatory elements, and a corresponding template region led to the identification of TRs in *Allium*, *Arabidopsis* and more than 70 other distantly related plants, including those with diverged

telomere motifs like *Genlisea*, *Cestrum*, and *Tulbaghia*. As far as we know, there is still no data on any algal TR, which would elucidate whether Pol III transcription of TR is a general feature for all plants or not. This work (Fajkus et al., 2019), based on CRISPR knock-out and other experiments, also showed that a previously identified telomerase RNA candidate in *A. thaliana* (Cifuentes-Rojas et al., 2011; Beilstein et al., 2012) was not a functional template subunit of telomerase, as was also demonstrated shortly after by (Dew-Budd et al., 2019). Assuming that the Pol II/Pol III dependency for TR transcription is a reliable evolutionary marker, future TR research in other main eukaryotic lineages will probably open new insights into the origin of eukaryotes. Telomerase genes and telomere sequences are unrecognized sources of information in this direction, and the finding of a Pol III dependent TR biogenesis pathway in ciliate and plant lineages may represent the first steps in this direction (Greider and Blackburn, 1989; Fajkus et al., 2019).

## HOW DID CHROMOSOMES BECOME LINEAR?

A vast majority of prokaryotes contain circular chromosomes while linear chromosomes are the rule in eukaryotes. Therefore there are two possible scenarios in which either (i) linearization was performed by a primitive telomerase, preceding other processes which led to current linear chromosomal features and functions or (ii) linearization of a pre-eukaryotic circular chromosome was initially telomerase independent, but just before current eukaryotes diverged, a primitive telomerase started to occupy chromosome ends and became essential for the newly formed linear chromosomes (Nosek et al., 2006). Villasante et al. (2007) proposed an evolutionary scenario in which the breakage of the ancestral prokaryotic circular chromosome activated a transposition mechanism at DNA ends, allowing the formation of telomeres by a recombination-dependent replication mechanism: consequences of this hypothesis led to the surprising conclusion that eukaryotic centromeres were derived from telomeres.

Interestingly, the opposite process to linearization, i.e., formation of circular chromosomes (also termed ring chromosomes) has emerged from time to time during the evolution of eukaryotes, although being highly unstable. For example, in the case of *Amaranthus tuberculatus*, ring chromosomes appeared as a stress-induced response, carrying resistance against a herbicide (glyphosate); these extra ring chromosomes did not show hybridization with telomere probes in the karyotype analysis (Koo et al., 2018). The almost universal telomerase system and the exceptionality of circular chromosomes in eukaryotes do not allow us to support one hypothesis over the other. However, the recombinational machinery used in the alternative lengthening of telomeres (ALT), a telomerase-independent pathway, associated with certain human cancers

(Zhang et al., 2019), is already present in prokaryotes. In addition, there is evidence of chromosome linearization occurring independently in distinct prokaryote lineages (Ferdows and Barbour, 1989; Nosek et al., 1995; Volff and Altenbuchner, 2000). Therefore, the hypothesis that the first linear eukaryotic chromosome (originating from a prokaryote ancestor) was telomerase-independent seems more likely. There are some examples that show that the telomerase-based system is not essential for telomere maintenance in all eukaryotes: retrotransposons in *Drosophila* telomeres, satellite repeats in *Chironomus*, another insect (Rubin, 1978; Biessmann and Mason, 2003), and ALT in telomerase-negative human cancers (Hu et al., 2016; Zhang et al., 2019). Yet, some of these systems may not be as different, and may perhaps share a common origin: in *Drosophila*, the telomere maintenance, based in retrotransposition, is not too distinct from the telomerase-based mechanism (Danilevskaya et al., 1998), leading to the hypothesis that the telomerase itself may be a former retrotransposon. But certainly, telomerase-negative plant species have not been discovered to date and all exceptions, in which the typical plant-type telomere was absent, were later shown to have different, but still telomerase-synthesized, motifs. Nevertheless, the ALT machinery is present in plants in parallel to the telomerase activity (Watson and Shippen, 2007; Ruckova et al., 2008). Interesting questions about the role of telomerase, telomeres and their maintenance in plant tumors arise from that. An attractive one is about the absence of metastasis in plants, despite the presence of ALT, perhaps related with plant tissue rigidity or different immune systems than in animals (Seyfried and Huysentruyt, 2013).

Although we are gaining increasing knowledge of telomere biology, we are still unable to explain the emergence of telomerase in eukaryotes. Current evidence supports the hypothesis that the emergence of eukaryotes together with their linear chromosomes, telomeres, and telomerase was related to the appearance of spliceosomal introns in archaeal hosts (Koonin, 2006; Fajkus et al., 2019). The similarity between TERT and other retroelements has been discussed for some time (Pardue et al., 1997). Remarkably, a relatively recent study showed that TERT, as a probable member of progeny group II introns, is sequentially close to *Penelope*-like element retrotransposons (Gladyshev and Arkhipova, 2007). But TERT is only one of the two essential telomerase components, and TR is, in its origin, even more enigmatic due to its low sequence conservation across all eukaryotes [see review (Podlevsky and Chen, 2016; Fajkus et al., 2019)].

## CONCLUSION

At the beginning of the plant genomics era, the telomere sequence was considered almost changeless. The general conservation of telomeres and the telomerase system suggested that all plants may have the TTTAGGG plant-type telomere. The identification of unusual telomere sequences in complex plant

genomes, in many cases with giant C-values (such as in *Cestrum* and *Allium* sp.), was worth the effort, since the exceptionally long *Allium* telomere motif was the clue in looking for a genuine TR in land plants. The newly described TR in plants and further telomere/telomerase research in basal clades of algae might reveal valuable information about early evolution, therefore plant telomere research can significantly contribute to hypotheses on the emergence of eukaryotes.

## AUTHOR CONTRIBUTIONS

VP and SG have contributed equally to the writing, editing, and preparation of this mini-review.

## REFERENCES

Adams, S. P., Hartman, T. P., Lim, K. Y., Chase, M. W., Bennett, M. D., Leitch, I. J., et al. (2001). Loss and recovery of Arabidopsis-type telomere repeat sequences 5'-(TTTAGGG)(n)-3' in the evolution of a major radiation of flowering plants. *Proc. Biol. Sci.* 268 (1476), 1541–1546. doi: 10.1098/rspb.2001.1726

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

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27 (2), 573–580. doi: 10.1093/nar/27.2.573

Biessmann, H., and Mason, J. M. (2003). Telomerase-independent mechanisms of telomere elongation. *Cell Mol. Life Sci.* 60 (11), 2325–2333. doi: 10.1007/s00018-003-3247-9

Blackburn, E. H., and Gall, J. G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *J. Mol. Biol.* 120 (1), 33–53. doi: 10.1016/0022-2836(78)90294-2

Blasco, M. A., Funk, W., Villeponteau, B., and Greider, C. W. (1995). Functional characterization and developmental regulation of mouse telomerase RNA. *Science* 269 (5228), 1267–1270. doi: 10.1126/science.7544492

Bremer, B., Bremer, K., Chase, M. W., Fay, M. F., Reveal, J. L., Soltis, D. E., et al. (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161 (2), 105–121. doi: 10.1111/j.1095-8339.2009.00996.x

Byng, J. W., Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Judd, W. S., Mabberley, D. J., et al. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 181 (1), 1–20. doi: 10.1111/boj.12385

Cervenak, F., Jurikova, K., Devillers, H., Kaffe, B., Khatib, A., Bonnell, E., et al. (2019). Identification of telomerase RNAs in species of the Yarrowia clade provides insights into the co-evolution of telomerase, telomeric repeats and telomere-binding proteins. *Sci. Rep.* 9 (1), 13365. doi: 10.1038/s41598-019-49628-6

Cifuentes-Rojas, C., Kannan, K., Tseng, L., and Shippen, D. E. (2011). Two RNA subunits and POT1a are components Arabidopsis telomerase. *Proc. Natl. Acad. Sci. U. S. A.* 108 (1), 73–78. doi: 10.1073/pnas.1013021107

Cox, A. V., Bennett, S. T., Parokonny, A. S., Kenton, A., Callimassia, M. A., and Bennett, M. D. (1993). Comparison of plant telomere locations using a Pcr-generated synthetic probe. *Ann. Bot.* 72 (3), 239–247. doi: 10.1006/anbo.1993.1104

Danilevskaya, O. N., Lowenhaupt, K., and Pardue, M. L. (1998). Conserved subfamilies of the *Drosophila* HeT-A telomere-specific retrotransposon. *Genetics* 148 (1), 233–242. https://www.genetics.org/content/148/1/233.short

Derelle, E., Ferraz, C., Rombauts, S., Rouze, P., Worden, A. Z., Robbens, S., et al. (2006). Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features. *Proc. Natl. Acad. Sci. U. States America* 103 (31), 11647–11652. doi: 10.1073/pnas.0604795103

## FUNDING

## ACKNOWLEDGMENTS

Dew-Budd, K., Cheung, J., Palos, K., Forsythe, E. S., and Beilstein, M. A. (2019). Evolutionary and biochemical analyses reveal conservation of the Brassicaceae telomerase ribonucleoprotein complex. *BioRxiv* 760785. doi: 10.1101/760785

Emery, H. S., and Weiner, A. M. (1981). An irregular satellite sequence is found at the termini of the linear extrachromosomal rDNA in Dictyostelium discoideum. *Cell* 26 (3 Pt 1), 411–419. doi: 10.1016/0092-8674(81)90210-5

Fajkus, J., Sýkorová, E., and Leitch, A. R. (2005). Telomeres in evolution and evolution of telomeres. *Chromosome Res.* 13 (5), 469–479. doi: 10.1007/s10577-005-0997-2

Fajkus, P., Peska, V., Sitova, Z., Fulneckova, J., Dvorackova, M., Gogela, R., et al. (2016). Allium telomeres unmasked: the unusual telomeric sequence (CTCGGTTATGGG)n is synthesized by telomerase. *Plant J.* 85 (3), 337–347. doi: 10.1111/tpj.13115

Fajkus, P., Peska, V., Zavodnik, M., Fojtova, M., Fulneckova, J., Dobias, S., et al. (2019). Telomerase RNAs in land plants. *Nucleic Acids Res.* 47 (18), 9842–9856. doi: 10.1093/nar/gkz695

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 (5228), 1236–1241. doi: 10.1126/science.7544491

Ferdows, M. S., and Barbour, A. G. (1989). Megabase-sized linear DNA in the bacterium Borrelia burgdorferi, the Lyme disease agent. *Proc. Natl. Acad. Sci. U. S. A.* 86 (15), 5969–5973. doi: 10.1073/pnas.86.15.5969

Fojtova, M., Sykorova, E., Najdekrova, L., Polanska, P., Zachova, D., Vagnerova, R., et al. (2015). Telomere dynamics in the lower plant Physcomitrella patens. *Plant Mol. Biol.* 87 (6), 591–601. doi: 10.1007/s11103-015-0299-9

Forstemann, K., Zaug, A. J., Cech, T. R., and Lingner, J. (2003). Yeast telomerase is specialized for C/A-rich RNA templates. *Nucleic Acids Res.* 31 (6), 1646–1655. doi: 10.1093/nar/gkg261

Frydrychova, R., Grossmann, P., Trubac, P., Vitkova, M., and Marec, F. (2004). Phylogenetic distribution of TTAGG telomeric repeats in insects. *Genome* 47 (1), 163–178. doi: 10.1139/g03-100

Fuchs, J., Brandes, A., and Schubert, I. (1995). Telomere sequence localization and karyotype evolution in higher-plants. *Plant Syst. Evol.* 196 (3-4), 227–241. doi: 10.1007/Bf00982962

Fulneckova, J., Hasikova, T., Fajkus, J., Lukesova, A., Elias, M., and Sykorova, E. (2012). Dynamic evolution of telomeric sequences in the green algal order Chlamydomonadales. *Genome Biol. Evol.* 4 (3), 248–264. doi: 10.1093/gbe/evs007

Fulneckova, J., Sevcikova, T., Fajkus, J., Lukesova, A., Lukes, M., Vlcek, C., et al. (2013). A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biol. Evol.* 5 (3), 468–483. doi: 10.1093/gbe/evt019

Fulneckova, J., Sevcikova, T., Lukesova, A., and Sykorova, E. (2016). Transitions between the Arabidopsis-type and the human-type telomere sequence in green algae (clade Caudivolvoxa, Chlamydomonadales). *Chromosoma* 125 (3), 437–451. doi: 10.1007/s00412-015-0557-2

Garcia, S., Garnatje, T., and Kovarik, A. (2012). Plant rDNA database: ribosomal DNA loci information goes online. *Chromosoma* 121 (4), 389–394. doi: 10.1007/s00412-012-0368-7

Gladyshev, E. A., and Arkhipova, I. R. (2007). Telomere-associated endonuclease-deficient Penelope-like retroelements in diverse eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 104 (22), 9352–9357. doi: 10.1073/pnas.0702741104

Greider, C. W., and Blackburn, E. H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* 43 (2 Pt 1), 405–413. doi: 10.1016/0092-8674(85)90170-9

Greider, C. W., and Blackburn, E. H. (1989). A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature* 337 (6205), 331–337. doi: 10.1038/337331a0

Higashiyama, T., Noutoshi, Y., Akiba, M., and Yamada, T. (1995). Telomere and LINE-like elements at the termini of the Chlorella chromosome I. *Nucleic Acids Symp. Ser.* (34), 71–72. https://europepmc.org/article/med/8841557

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 (1), 221–228. doi: 10.1042/BJ20082195

Hu, Y., Shi, G., Zhang, L. C., Li, F., Jiang, Y. L., Jiang, S., et al. (2016). Switch telomerase to ALT mechanism by inducing telomeric DNA damages and dysfunction of ATRX and DAXX. *Sci. Rep.* 6. doi: 10.1038/Srep32280

Kim, N. W., Piatyszek, M. A., Prowse, K. R., Harley, C. B., West, M. D., Ho, P. L., et al. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266 (5193), 2011–2015. doi: 10.1126/science.7605428

Koo, D. H., Molin, W. T., Saski, C. A., Jiang, J., Putta, K., Jugulam, M., et al. (2018). Extrachromosomal circular DNA-based amplification and transmission of herbicide resistance in crop weed Amaranthus palmeri. *Proc. Natl. Acad. Sci. U. S. A.* 115 (13), 3332–3337. doi: 10.1073/pnas.1719354115

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

Majerska, J., Schrumpfova, P. P., Dokladal, L., Schorova, S., Stejskal, K., Oboril, M., et al. (2017). Tandem affinity purification of AtTERT reveals putative interaction partners of plant telomerase in vivo. *Protoplasma* 254 (4), 1547–1562. doi: 10.1007/s00709-016-1042-3

Mason, J. M., Randall, T. A., and Capkova Frydrychova, R. (2016). Telomerase lost? *Chromosoma* 125 (1), 65–73. doi: 10.1007/s00412-015-0528-7

Meyne, J., Ratliff, R. L., and Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)n among vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* 86 (18), 7049–7053. doi: 10.1073/pnas.86.18.7049

Mravinac, B., Mestrovic, N., Cavrak, V. V., and Plohl, M. (2011). TCAGG, an alternative telomeric sequence in insects. *Chromosoma* 120 (4), 367–376. doi: 10.1007/s00412-011-0317-x

Muller, F., Wicky, C., Spicher, A., and Tobler, H. (1991). New telomere formation after developmentally regulated chromosomal breakage during the process of chromatin diminution in Ascaris lumbricoides. *Cell* 67 (4), 815–822. doi: 10.1016/0092-8674(91)90076-b

Nosek, J., Dinouel, N., Kovac, L., and Fukuhara, H. (1995). Linear mitochondrial DNAs from yeasts: telomeres with large tandem repetitions. *Mol. Gen. Genet.* 247 (1), 61–72. doi: 10.1007/bf00425822

Nosek, J., Kosa, P., and Tomaska, L. (2006). On the origin of telomeres: a glimpse at the pre-telomerase world. *Bioessays* 28 (2), 182–190. doi: 10.1002/bies.20355

Novak, P., Neumann, P., and Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinf.* 11, 378. doi: 10.1186/1471-2105-11-378

Nozaki, H., Takano, H., Misumi, O., Terasawa, K., Matsuzaki, M., Maruyama, S., et al. (2007). A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga Cyanidioschyzon merolae. *BMC Biol.* 5, 28. doi: 10.1186/1741-7007-5-28

Olovnikov, A. M. (1973). Theory of marginotomy - incomplete copying of template margin in enzymic-synthesis of polynucleotides and biological significance of phenomenon. *J. Theor. Biol.* 41 (1), 181–190. doi: 10.1016/0022-5193(73)90198-7

One Thousand Plant Transcriptomes Initiative (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574 (7780), 679–685. doi: 10.1038/s41586-019-1693-2

Pardue, M. L., Danilevskaya, O. N., Traverse, K. L., and Lowenhaupt, K. (1997). Evolutionary links between telomeres and transposable elements. *Genetica* 100 (1–3), 73–84. doi: 10.1007/978-94-011-4898-6_7

Peska, V., Sykorova, E., and Fajkus, J. (2008). Two faces of solanaceae telomeres: a comparison between Nicotiana and Cestrum telomeres and telomere-binding proteins. *Cytogenetic. Genome Res.* 122 (3-4), 380–387. doi: 10.1159/000167826

Peska, V., Fajkus, P., Fojtova, M., Dvorackova, M., Hapala, J., Dvoracek, V., et al. (2015). Characterisation of an unusual telomere motif (TTTTTTAGGG)(n) in the plant Cestrum elegans (Solanaceae), a species with a large genome. *Plant J.* 82 (4), 644–654. doi: 10.1111/tpj.12839

Peska, V., Sitova, Z., Fajkus, P., and Fajkus, J. (2017). BAL31-NGS approach for identification of telomeres de novo in large genomes. *Methods* 114, 16–27. doi: 10.1016/j.ymeth.2016.08.017

Peska, V., Mandakova, T., Ihradska, V., and Fajkus, J. (2019). Comparative dissection of three giant genomes: allium cepa, allium sativum, and allium ursinum. *Int. J. Mol. Sci.* 20 (3). doi: 10.3390/ijms20030733

Petracek, M. E., Lefebvre, P. A., Silflow, C. D., and Berman, J. (1990). Chlamydomonas telomere sequences are A+T-rich but contain three consecutive G-C base pairs. *Proc. Natl. Acad. Sci. U. S. A.* 87 (21), 8222–8226. doi: 10.1073/pnas.87.21.8222

Pfeiffer, V., and Lingner, J. (2012). TERRA promotes telomere shortening through exonuclease 1-mediated resection of chromosome ends. *PLoS Genet.* 8 (6). doi: 10.1371/journal.pgen.1002747

Pich, U., and Schubert, I. (1998). Terminal heterochromatin and alternative telomeric sequences in Allium cepa. *Chromosome Res.* 6 (4), 315–321. doi: 10.1023/a:1009227009121

Podlevsky, J. D., and Chen, J. J. (2016). Evolutionary perspectives of telomerase RNA structure and function. *RNA Biol.* 13 (8), 720–732. doi: 10.1080/15476286.2016.1205768

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 (14), 831–834. doi: 10.1016/S0960-9822(98)70325-2

Puizina, J., Weiss-Schneeweiss, H., Pedrosa-Harand, A., Kamenjarin, J., Trinajstic, I., Riha, K., and Schweizer, D. (2003). Karyotype analysis in Hyacinthella dalmatica (Hyacinthaceae) reveals vertebrate-type telomere repeats at the chromosome ends. *Genome* 46 (6), 1070–6. doi: 10.1139/g03-078

Qi, X. D., 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 (1), 450–462. doi: 10.1093/nar/gks980

Richards, E. J., and Ausubel, F. M. (1988). Isolation of a higher eukaryotic telomere from Arabidopsis thaliana. *Cell* 53 (1), 127–136. doi: 10.1016/0092-8674(88)90494-1

Rubin, G. M. (1978). Isolation of a telomeric DNA sequence from Drosophila melanogaster. *Cold Spring Harb. Symp. Quant. Biol.* 42 Pt 2, 1041–1046. doi: 10.1101/sqb.1978.042.01.104

Ruckova, E., Friml, J., Schrumpfova, P. P., and Fajkus, J. (2008). Role of alternative telomere lengthening unmasked in telomerase knock-out mutant plants. *Plant Mol. Biol.* 66 (6), 637–646. doi: 10.1007/s11103-008-9295-7

Sekhri, K. (2014). Telomeres and telomerase: understanding basic structure and potential new therapeutic strategies targeting it in the treatment of cancer. *Postgrad. Med. J.* 60 (3), 303. doi: 10.4103/0022-3859.138797

Seyfried, T. N., and Huysentruyt, L. C. (2013). On the origin of cancer metastasis. *Crit. Rev. Oncog.* 18 (1-2), 43–73. doi: 10.1615/critrevoncog.v18.i1-2.40

Shibata, F., and Hizume, M. (2011). Survey of arabidopsis- and human-type telomere repeats in plants using fluorescence in situ Hybridisation. *Cytologia* 76 (3), 353–360. doi: 10.1508/cytologia.76.353

Somanathan, I., and Baysdorfer, C. (2018). A bioinformatics approach to identify telomere sequences. *Biotechniques* 65 (1), 20–25. doi: 10.2144/btn-2018-0057

Spiegel, J., Adhikari, S., and Balasubramanian, S. (2020). The structure and function of DNA G-quadruplexes. *Trends Chem.* 2 (2), 123–136. doi: 10.1016/j.trechm.2019.07.002

Sykorova, E., Lim, K. Y., Fajkus, J., and Leitch, A. R. (2003a). The signature of the Cestrum genome suggests an evolutionary response to the loss of (TTTAGGG) n telomeres. *Chromosoma* 112 (4), 164–172. doi: 10.1007/s00412-003-0256-2

Sykorova, E., Lim, K. Y., Chase, M. W., Knapp, S., Leitch, I. J., Leitch, A. R., et al. (2003b). The absence of Arabidopsis-type telomeres in Cestrum and closely related genera Vestia and Sessea (Solanaceae): first evidence from eudicots. *Plant J.* 34 (3), 283–291. doi: 10.1046/j.1365-313x.2003.01731.x

Sykorova, E., Lim, K. Y., Kunicka, Z., Chase, M. W., Bennett, M. D., Fajkus, J., et al. (2003c). Telomere variability in the monocotyledonous plant order Asparagales. *Proc. Biol. Sci.* 270 (1527), 1893–1904. doi: 10.1098/rspb.2003.2446

Tran, P. L., De Cian, A., Gros, J., Moriyama, R., and Mergny, J. L. (2013). Tetramolecular quadruplex stability and assembly. *Top. Curr. Chem.* 330, 243–273. doi: 10.1007/128_2012_334

Tran, T. D., Cao, H. X., Jovtchev, G., Neumann, P., Novak, P., Fojtova, M., et al. (2015). Centromere and telomere sequence alterations reflect the rapid genome evolution within the carnivorous plant genus Genlisea. *Plant J.* 84 (6), 1087–1099. doi: 10.1111/tpj.13058

Traut, W., Szczepanowski, M., Vitkova, M., Opitz, C., Marec, F., and Zrzavy, J. (2007). The telomere repeat motif of basal Metazoa. *Chromosome Res.* 15 (3), 371–382. doi: 10.1007/s10577-007-1132-3

Uzlikova, M., Fulneckova, J., Weisz, F., Sykorova, E., Nohynkova, E., and Tumova, P. (2017). Characterization of telomeres and telomerase from the single-celled eukaryote Giardia intestinalis. *Mol. Biochem. Parasitol.* 211, 31–38. doi: 10.1016/j.molbiopara.2016.09.003

vanSteensel, B., and deLange, T. (1997). Control of telomere length by the human telomeric protein TRF1. *Nature* 385 (6618), 740–743. doi: 10.1038/385740a0

Vilenchik, M. M., and Knudson, A. G. (2003). Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. *Proc. Natl. Acad. Sci. U. S. A.* 100 (22), 12871–12876. doi: 10.1073/pnas.2135498100

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 (25), 10542–10547. doi: 10.1073/pnas.0703808104

Visacka, K., Hofr, C., Willcox, S., Necasova, I., Pavlouskova, J., Sepsiova, R., et al. (2012). Synergism of the two Myb domains of Tay1 protein results in high affinity binding to telomeres. *J. Biol. Chem.* 287 (38), 32206–32215. doi: 10.1074/jbc.M112.385591

Vitales, D., D'Ambrosio, U., Galvez, F., Kovarik, A., and Garcia, S. (2017). Third release of the plant rDNA database with updated content and information on telomere composition and sequenced plant genomes. *Plant Syst. Evol.* 303 (8), 1115–1121. doi: 10.1007/s00606-017-1440-9

Vitkova, M., Kral, J., Traut, W., Zrzavy, J., and Marec, F. (2005). The evolutionary origin of insect telomeric repeats, (TTAGG)n. *Chromosome Res.* 13 (2), 145–156. doi: 10.1007/s10577-005-7721-0

Volff, J. N., and Altenbuchner, J. (2000). A new beginning with new ends: linearisation of circular chromosomes during bacterial evolution. *FEMS Microbiol. Lett.* 186 (2), 143–150. doi: 10.1111/j.1574-6968.2000.tb09095.x

Wang, Y., Susac, L., and Feigon, J. (2019). Structural biology of telomerase. *Cold Spring Harb. Perspect. Biol.* 11, a032383. doi: 10.1101/cshperspect.a032383

Watson, J. M., and Shippen, D. E. (2007). Telomere rapid deletion regulates telomere length in Arabidopsis thaliana. *Mol. Cell. Biol.* 27 (5), 1706–1715. doi: 10.1128/Mcb.02059-06

Watson, J. D. (1972). Origin of Concatemeric T7 DNA. *Nature-New Biol.* 239 (94), 197–201. doi: 10.1038/newbio239197a0

Weiss, H., and Scherthan, H. (2002). Aloe spp.–plants with vertebrate-like telomeric sequences. *Chromosome Res.* 10 (2), 155–164. doi: 10.1023/a:1014905319557

Wellinger, R. J., Ethier, K., Labrecque, P., and Zakian, V. A. (1996). Evidence for a new step in telomere maintenance. *Cell* 85 (3), 423–433. doi: 10.1016/S0092-8674(00)81120-4

Wu, J., Okada, T., Fukushima, T., Tsudzuki, T., Sugiura, M., and Yukawa, Y. (2012). A novel hypoxic stress-responsive long non-coding RNA transcribed by RNA polymerase III in Arabidopsis. *RNA Biol.* 9 (3), 302–313. doi: 10.4161/rna.19101

Zhang, J. M., Yadav, T., Ouyang, J., Lan, L., and Zou, L. (2019). Alternative lengthening of telomeres through two distinct break-induced replication pathways. *Cell Rep.* 26 (4), 955–95+. doi: 10.1016/j.celrep.2018.12.102