



# TAR30, a homolog of the canonical plant TTTAGGG telomeric repeat, is enriched in the proximal chromosome regions of peanut (*Arachis hypogaea* L.)

Dongying Gao · Eliza F. M. B. Nascimento · Soraya C. M. Leal-Bertioli · Brian Abernathy · Scott A. Jackson · Ana C. G. Araujo · David J. Bertioli

Received: 5 October 2021 / Revised: 7 January 2022 / Accepted: 11 January 2022 / Published online: 19 January 2022  
© Springer Nature B.V. 2022

**Abstract** Telomeres are the physical ends of eukaryotic linear chromosomes that play critical roles in cell division, chromosome maintenance, and genome stability. In many plants, telomeres are comprised of TTTAGGG tandem repeat that is widely found in plants. We refer to this repeat as canonical plant telomeric repeat (CPTR). Peanut (*Arachis hypogaea* L.) is a spontaneously formed allotetraploid and an important food and oil crop worldwide. In this study, we analyzed the peanut genome sequences and identified a new type of tandem repeat with 10-bp basic motif TTTT(C/T)TAGGG named *TAndem Repeat* (TAR) 30. TAR30 showed significant sequence identity to

TTTAGGG repeat in 112 plant genomes suggesting that TAR30 is a homolog of CPTR. It also is nearly identical to the telomeric tandem repeat in *Cestrum elegans*. Fluorescence in situ hybridization (FISH) analysis revealed interstitial locations of TAR30 in peanut chromosomes but we did not detect visible signals in the terminal ends of chromosomes as expected for telomeric repeats. Interestingly, different TAR30 hybridization patterns were found between the newly induced allotetraploid ValSten and its diploid wild progenitors. The canonical telomeric repeat TTTAGGG is also present in the peanut genomes and some of these repeats are closely adjacent to TAR30 from both cultivated peanut and its wild relatives. Overall, our work identifies a new homolog of CPTR and reveals the unique distributions of TAR30 in cultivated peanuts and wild species. Our results provide

Responsible Editor: Aurora Ruiz-Herrera

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10577-022-09684-7>.

D. Gao (✉)  
Small Grains and Potato Germplasm Research Unit,  
USDA ARS, Aberdeen, ID 83210, USA  
e-mail: dongying.gao@usda.gov

D. Gao · S. C. M. Leal-Bertioli · B. Abernathy ·  
S. A. Jackson · D. J. Bertioli (✉)  
Center for Applied Genetic Technologies, University  
of Georgia, Athens, GA 30602, USA  
e-mail: bertioli@uga.edu

E. F. M. B. Nascimento  
University of Brasilia, Campus Universitário Darcy  
Ribeiro, Brasília, DF 70910-900, Brazil

S. C. M. Leal-Bertioli  
Department of Plant Pathology, University of Georgia,  
Athens, GA 30602, USA

A. C. G. Araujo (✉)  
Embrapa Genetic Resources and Biotechnology, PqEB W5  
Norte Final, 70.770-917, Brasília, DF, Brazil  
e-mail: ana-claudia.guerra@embrapa.br

D. J. Bertioli  
Department of Crop and Soil Science, University  
of Georgia, Athens, GA 30602, USA

new insights into the evolution of tandem repeats during peanut polyploidization and domestication.

**Keywords** Tandem repeat · Chromosome · Peanut · Induced allotetraploid · FISH

### Abbreviations

CPTR	Canonical plant telomeric repeat
DAPI	4,6-Diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
DSBs	Double-strand breaks
FISH	Fluorescence in situ hybridization
LTR	Long terminal repeat
MYA	Million years ago
TERT	Telomerase reverse transcriptase
TEs	Transposable elements
TSD	Target site duplication

### Introduction

Repeated sequences, including transposable elements (TEs) or transposons, tandem repeats, and other repetitive elements, contribute large fractions of many eukaryotic genomes (Wessler 2006). Once considered “junk DNA,” repetitive sequences are now recognized to play essential roles in variation of functional genes (Kobayashi et al. 2004) and genome evolution (Gao et al. 2016). They also serve as important components to form and maintain functional centromeres and telomeres (Maxwell et al. 2006; reviewed in Grewal and Jia 2007; Gao et al. 2015) and provide raw materials for emergence of new genes and other genetic innovations (for review: Long et al. 2003).

Transposons are mobile genomic elements which have potential capability to move from one position to another place in host genome or even conduct horizontal transfer between distantly related organisms (El Baidouri et al. 2014; Gao et al. 2018). Except for a handful of TEs that show insertion preferences and were found in some specific regions such as telomeres (Maxwell et al. 2006), numerous transposons are dispersed throughout eukaryotic genomes. In plant genomes, TEs are very dynamic and homologous transposons have been difficult to identify, even in closely relative genomes (Piegu et al. 2006; Gao et al. 2009).

Tandem DNA repeats represent another type of repetitive sequences which are organized into

tandem array of multiple units ranging from a few to thousands or more. Centromeric tandem repeats (CTRs) and telomeric tandem repeats (TTRs) are two major types of tandem repeats in higher eukaryotic genomes. In primate and flowering plants, some centromeres can contain megabase arrays of tandem repeats; these extremely repetitive structures make the centromeric regions difficult to completely sequence, even with the latest long-read sequencers (Bzikadze and Pevzner 2020). CTRs usually evolve rapidly and lineage-specific centromeric repeats were found in many genomes (Lee et al. 2005). In contrast to CTRs, TTRs tend to be highly conserved. For example, the (TTAGGG)<sub>n</sub> repeat is present in the telomeric regions of human (*Homo sapiens*) and other vertebrates (Meyne et al. 1989; Podlevsky et al. 2008). In plants, the (TTTAGGG)<sub>n</sub> repeat, which was originally identified in *Arabidopsis thaliana* (Richards and Ausubel 1988), is present in a wide range of plants including dicots, monocots, and algae (Petracek et al. 1990; Fuchs et al. 1995; reviewed in Peska and Garcia 2020). We hereafter referred to as canonical plant telomeric repeat (CPTR). However, in some plants such as the order Asparagales and *Cestrum elegans*, the *Arabidopsis*-type telomeric repeat was replaced by TTAGGG, TTTTTTAGGG, or other types of tandem repeats (Peška et al., 2015; reviewed in Peska and Garcia 2020).

Despite the high conservation of plant telomeric tandem repeats in DNA sequence and cytological localization, the lengths of telomeric DNA tracts can be tremendously different among species and varieties as well as between different chromosomes within a nucleus (for review: Shippen and McKnight 1998). Additionally, the lengths of telomeres may vary dramatically between different developmental stages or tissues in some plants. For instance, in young embryos and undifferentiated calli of barley (*Hordeum vulgare*), the telomeres can reach 80 kb and 300 kb, respectively. However, the telomere lengths in mature embryos and old leaves were estimated to be 30 kb and 23 kb (Kilian et al. 1995). Stability of telomere length during plant growth and development was also found in white campion (*Melandrium album*) (Riha et al. 1998) and likely reflects various mechanisms of telomere length regulation in different plants. Maintenance of basic telomere lengths plays pivotal roles for cell differentiation as significant shortage of telomeres can cause higher frequency of chromosomal rearrangements and abnormal

cells (for review: Graham and Meeker 2017). Telomere lengths may also affect flowering time in some plants such as *Arabidopsis thaliana*, rice (*Oryza sativa*), and maize (*Zea mays*) (Choi et al. 2021). Additionally, mutations in the telomere lengths can result in progressive and severe developmental abnormalities in both germination and post germination growth of vegetative organs (Hong et al. 2007). In order to recover the losses of telomeres and to maintain the genomic stability, several mechanisms have evolved in eukaryotes, one of which is the telomerase-associated telomere elongation. During this process, the telomerase reverse transcriptase (TERT) catalyzes the synthesis of short tandem motifs and extends the telomeres by using the telomerase RNA subunit as the template (for review: Peska and Garcia 2020).

Peanut (*Arachis hypogaea* L. AABB genome type,  $2n=4x=40$ ) is an important oil and food crop in the world. The cultivated peanut is an allotetraploid which was derived from the interspecific hybridization of two diploid wild species, *A. duranensis* (AA genome type,  $2n=2x=20$ ) and *A. ipaensis* (BB genome type,  $2n=2x=20$ ), and followed by a spontaneous whole-genome duplication (Seijo et al. 2007). Like many other flowering plants, peanut telomeres also consist of multiple copies of TTTAGGG repeat (Du et al. 2016; Zhang et al. 2016). However, other types of repetitive sequences in peanuts and their evolution are still poorly understood. In this study, we analyzed the peanut genome sequences and identified a new tandem repeat named TAR30, [TTTT(C/T)TAGGG] $_n$ . We found that TAR30 shows significant sequence similarity to CPTR, (TTTAGGG) $_n$ , in over 100 plant genomes. Our fluorescence in situ hybridization (FISH) analysis revealed that TAR30 is enriched in the interstitial regions of peanut chromosomes. Different FISH hybridization patterns of TAR30 were observed between a newly induced allotetraploid and its two wild diploid progenitors. Our results indicate that TAR30 is a homolog of CPTR and provide insights into evolution of tandem repeats in peanuts.

## Materials and methods

### Plant materials

The seeds of cultivated peanut “Tifrunner” (Holbrook and Culbreath 2007) were provided by the

peanut breeding lab at USDA-ARS in Tifton, GA, USA. Another peanut variety “Runner IAC 886” was obtained from the Active Germplasm Bank of Embrapa Genetic Resources and Biotechnology (Embrapa-Cenargen, Brasília, Brazil). Seeds of two diploid wild species, *A. valida* (PI 468,154, BB genome type,  $2n=2x=20$ ) and *A. stenosperma* (PI666100, AA genome type,  $2n=2x=20$ ), were collected from the USDA seed bank. Additionally, a synthetic allotetraploid ValSten developed from the cross between the wild species *A. valida* and *A. stenosperma* (Gao et al. 2021) was also included. Seeds of ValSten used in this study were harvested from the earliest generation (S0) tetraploid plants.

### Sequence analysis

#### Identification of tandem repeat sequences

The genome sequence of cultivated peanut (Bertioli et al. 2019) was initially analyzed by the LTR-Finder software (Xu and Wang 2007) using the default parameters except that we set a 50 bp of minimum long terminal repeat (LTR) length and 100 bp of minimum distance between LTRs. The annotated sequences were then manually inspected based on their sequence structures including LTRs, target site duplication (TSD), and retrotransposase proteins. All sequences containing no retrotransposase protein were used for BLASTN searches against themselves to detect if they were tandemly repetitive. The selected tandem repeats were further analyzed by the Tandem Repeats Finder program (Benson 1999) to determine the consensus patterns of tandem repeats and the repeat unit sizes. The consensus sequence of the tandem repeat was generated with the WebLogo website (<https://weblogo.berkeley.edu/logo.cgi>).

#### Identification of transposons in tandem repeat sequences

The peanut repeat database including both transposons and tandem repeats were used to screen the assembled peanut genome (Bertioli et al. 2019) using RepeatMasker (<http://www.repeatmasker.org>) program with the default parameters but we used the “nolow” option to avoid masking the low-complexity DNA regions. The hits masked by the tandem repeat TAR30 were further checked to determine if TAR30

sequences were inserted by transposons according to the locations of both TAR30 and transposons in the peanut genome.

### PCR analysis

PCR analysis was performed by following our previous protocols (Gao et al. 2009). Briefly, the young leaves of Tifrunner plant were used to extract genomic DNA with the DNeasy Plant Mini Kit (QIAGEN, Venlo, Netherlands). Amplification was conducted in an MJ Research PTC-200 thermal cycler using 20 ng genomic DNA, 1.5 mM MgCl<sub>2</sub>, 1.0 unit Taq DNA polymerase, 0.2 mM dNTP, 0.2 mM primer, 1×PCR buffer, and ddH<sub>2</sub>O to a final volume of 25 µl. The temperature cycling conditions were 5 min at 95 °C, followed by 35 cycles of 95 °C for 50 s, 55 °C for 50 s, and 72 °C for 1 min, and a final extension at 72 °C for 5 min. The primers for amplifying TAR30 tandem repeat were 5'-ATTTGGAGTTTGGAGTTT AGG-3' (forward) and 5'-GGCGATATAAATAGG ACGAAT-3' (reverse). Amplicons were purified with QIAquick PCR purification kits (QIAGEN, Venlo, Netherlands) and it was used to obtain the probe for FISH.

### Fluorescence in situ hybridization (FISH)

#### Slide preparation

Around 5–10 mm of at least five root tips was collected from 4-week-old plants (5 plants for each genotype) and treated with 2 mM 8-hydroxyquinoline for 2 h at room temperature followed by an extra hour at 4 °C with fresh hydroxyquinoline solution (Fernández and Krapovickas, 1994). The samples were incubated in the fixative solution containing absolute ethanol:glacial acetic acid (3:1, v/v) for 12 h at 4 °C and the spreads of somatic chromosomes were prepared according to the previous protocol (Schwarzacher and Heslop-Harrison 2000). Meristems were digested in 10 mM citrate buffer containing 2% cellulase from *Trichoderma viridae* (Onozuka R-10; Serva, Heidelberg, Germany) and 20% pectinase from *Aspergillus niger* (Sigma-Aldrich, Darmstadt, Germany) for 2 h at 37 °C. Chromosomes of each root were spread in a drop of acetic acid of 45% (v/v) on a slide. The spread was obtained after a gentle pressure using the coverslip. The slides were selected using the phase contrast in the Axioskop

microscope (Zeiss, Oberkochen, Germany). Coverslips were removed; slides were air-dried for 24 h and kept at –20 °C until use.

#### Probe preparation, hybridization, and imaging

The purified TAR30 DNA sequence was labeled with digoxigenin-11-dUTP using the Nick translation kit (10,976,776,001, Roche Diagnostics GmbH, Mannheim, Germany). The 5S ribosomal DNA sequence (5S rDNA) was obtained from clones of *Lotus japonicus* (Pedrosa et al. 2002) and the correspondent probe was used as a positive technical control. The 5S rDNA sequence was isolated with the Illustra Plasmid Prep Midi Flow kit (28,904,269, GE Healthcare, Chicago, IL) and labeled with Cy3-dUTP (GEP53022, Roche Diagnostics GmbH, Mannheim, Germany) by nick translation.

FISH experiments were performed following the previous protocol (Schwarzacher and Heslop-Harrison 2000). Both probes of TAR30 and 5SrDNA were used simultaneously for hybridization (~50 ng of each probe/slide). Hybridizations were performed for 14 h at 37 °C, followed by washes with around 73% stringent washes, using saline citrate buffer (SSC) 2×. The chromosome sites that hybridized with the TAR30 probe labeled with digoxigenin-11-dUTP were immunocytochemically detected by the antibody anti-digoxigenin conjugated to fluorescein (11,207,741,910, Roche Diagnostics GmbH, Mannheim, Germany), while the 5S rDNA loci, in which probe was labeled with Cy3-dUTP, were detected by the direct observation in the Zeiss AxioPhot epifluorescence microscope (Zeiss, Oberkochen, Germany). All slides were counterstained with 4',6-diamidino-2-phenylindole (DAPI) before observation. Images were captured using Zeiss AxioCam MRc digital camera (Zeiss Light Microscopy, Göttingen, Germany) and the AxioVision Rel. 4.8 software and they were further processed as a whole image uniformly, using the Adobe Photoshop CS software.

## Results

### A new tandem repeat in cultivated peanut genome

In the process of annotating LTR retrotransposons in the genomes of wild rice species and other plants

(Gao et al. 2009, 2016), we recognized that not all sequences identified by the LTR\_Finder program (Xu and Wang 2007) were true LTR retrotransposons, and some of them were false positives including tandem repeats and others. We analyzed the cultivated peanut genome (Bertioli et al. 2019) with the LTR\_Finder program and identified a 6,030-bp sequence containing some typical features of LTR retrotransposons including the LTR with 5'TGT ... ACA3' and 5-bp TSD (Fig. 1a). To identify the putative retrotransposase protein and to define the superfamily of this "potential LTR retrotransposon," the 6,030-bp sequence was used as a query to conduct a BLASTX search against the NCBI non-redundant protein sequences (<https://blast.ncbi.nlm.nih.gov>). No significant hit was found even with a lower *E*-value cutoff (*E* value=0) suggesting that this sequence encodes no homologous protein to any protein in the NCBI protein database. No predicted protein was found for the sequence with the gene annotation program GENSCAN (Burge and Karlin, 1998). To further test if this sequence is related to any previously characterized LTR retroelements, the 6,030-bp sequence was used to search against the comprehensive LTR retrotransposon database, Gypsy Database (GyDB) (Llorens et al. 2011), and no hit was found. The BLAST searches against both NCBI and GyDB indicated that this is either a non-autonomous LTR retroelement or a false positive LTR retroelement.

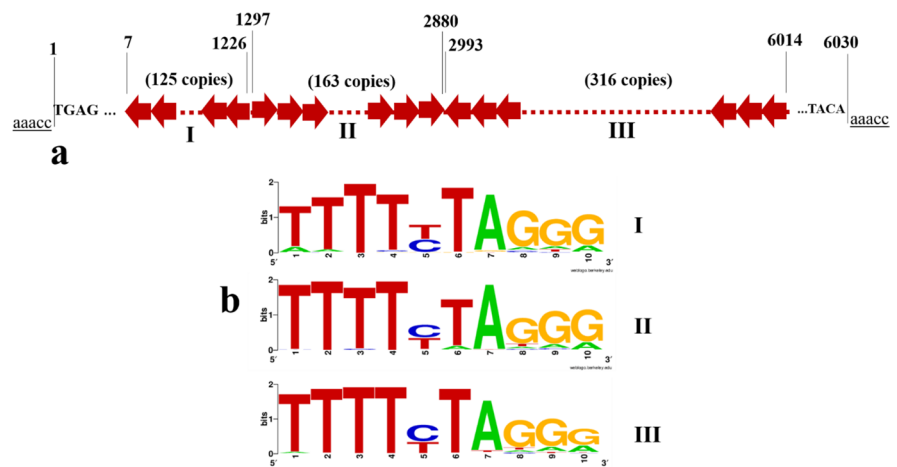
We further analyzed the 6,030-bp sequence with the Tandem Repeats Finder (Benson 1999) to determine if the sequence contained any tandem repeats. The results indicated that this sequence is organized

into three tracts and each of them contains multiple copies ranging from 125 to 316 of 10-bp monomers, TTTT(C/T)TAGGG (Fig. 1a, b). Therefore, this analysis indicated that this is a new tandem repeat sequence in peanut which was named as TAR30 (TAndem Repeat 30).

#### Identification of homologous sequences of TAR30 in plant genomes

To detect if there were homologous sequences of TAR30 in other plant genomes and to better understand its evolutionary origin, TAR30 DNA sequence was used to conduct BLASTN searches against the plant genomic sequences in GenBank. To avoid false positives, we used stringent criteria including over 200-bp size of homologous sequence alignment and a lower *E* value ( $\leq 1 \times e^{-10}$ ). For some genomes, if multiple significant hits were detected, the sequence with the best alignment score was designed as the TAR30 homolog. TAR30 homologous sequences were found in 160 plants including 139 flowering plants, three lycophytes, and 18 algae (Table 1, Supplementary Table 1). All the homologous sequences were analyzed with the Tandem Repeats Finder program (Benson 1999) to determine the basic motifs of tandem repeats. The homologous sequences from 70% (112/160  $\times$  100%) of the plant genomes contain the TTTAGGG repeat that has been detected in the telomeric regions in many plants including *Arabidopsis* (Richards and Ausubel 1988), barley (Röder et al. 1993) and rice (Mizuno et al. 2006). We refer to this sequence as (T)3A(G)3 where (T)3 and (G)3

**Fig. 1** Sequence structure of the 6014-bp sequence (a) and WebLogo of three tracts of TAR30 (b). The false positive element shows some structural features of LTR retrotransposons such as started with TGT and ended by ACA and flanked by 5-bp TSD. However, it contains multiple units of tandem repeat indicated by red arrows



**Table 1** Summary of different types of TAR30 homologs in 160 plant genomes

Homologous sequence	Angiosperms			Lycophyte	Algae	Total
	Eudicots	Monocots	<i>Amborella</i>			
(T) <sub>3</sub> A(G) <sub>3</sub>	73	28	1	3	7	112
(T) <sub>4-5</sub> A(G) <sub>3</sub>	11	0	0	0	6	17
Other TRs	22	1	0	0	4	27
Non-TR	3	0	0	0	1	4
Total	109	29	1	3	18	160

represent three thymines (TTT) and guanines (GGG), respectively. It should be noted that this number is undercounted as we used very strict criteria. For example, homologous sequences of TAR30 with the TTTAGGG motif was found in *Zizania latifolia*, *Cenchrus americanus*, *Ensete ventricosum*, *Brassica juncea*, and six other plants which were not included due to either short alignment size (less than 200 bp) or higher *E* value ( $> 1 \times e^{-10}$ ). Additionally, the homologs of TAR30 in 17 plant genomes contain the consensus pattern of TTTTAGGG or TTT TTAGGG (Table 1, Supplementary Table 1). Thus, our sequence comparisons suggest that TAR30 is the homologous sequence of CPTR and is likely related to the TTTAGGG telomeric repeat. We also identified significant hits of TAR30 in 31 plant genomes although these homologous sequences consisted of other types of tandem repeats or other sequences but their chromosomal distributions are not clear. The TTTAGGG repetitive sequences may be present in some of the 31 plant genomes but were not sequenced or assembled. Another possibility was that TTTAGG G repeat evolved rapidly or has been lost in these plants (Sýkorová et al., 2006).

TAR30 was also used to conduct BLASTN searches against the sequences from 245 plant mitochondria and 2216 plant chloroplasts in GenBank, but no homologous hit was detected suggesting the homologs of TAR30 are only present in plant nuclear genomes.

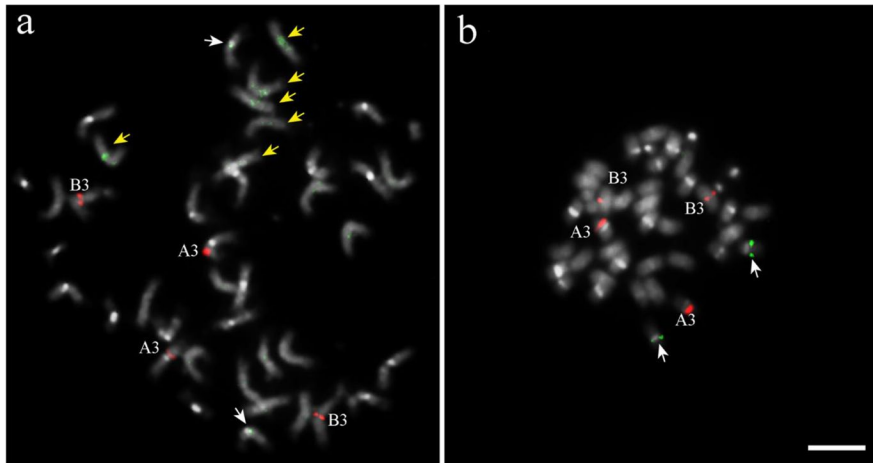
### Chromosomal locations of TAR30

TAR30 was used as the probe to conduct FISH analyses to determine its genomic distributions. 5S rDNA was included as a positive control and marker for chromosomes A3 and B3. In both peanut cultivars “RunnerIAC 886” (Fig. 2a) and “Tifrunner” (Fig. 2b), 40 chromosomes were observed, including

20 chromosomes from subgenome A since they all have DAPI<sup>+</sup> bands (bright white) on the centromere regions and another 20 chromosomes lacking DAPI<sup>+</sup> signals on centromeres thus corresponding to chromosomes of the subgenome B in both varieties. FISH signals of the TAR30 probe (green signals) were detected in the proximal region of one pair of chromosomes of the subgenome A (white arrows in Fig. 2) in both cultivars, and in similar region of three pairs of the subgenome B in “Runner IAC 886” (yellow arrows in Fig. 2a) but not in the subgenome B of “Tifrunner” (Fig. 2b). As expected, rDNA loci (red signals) were identified in the cyt-A3 and cyt-B3 chromosomes in both peanut cultivars. No overlap of TAR30 and 5S rDNA was detected in these chromosomes.

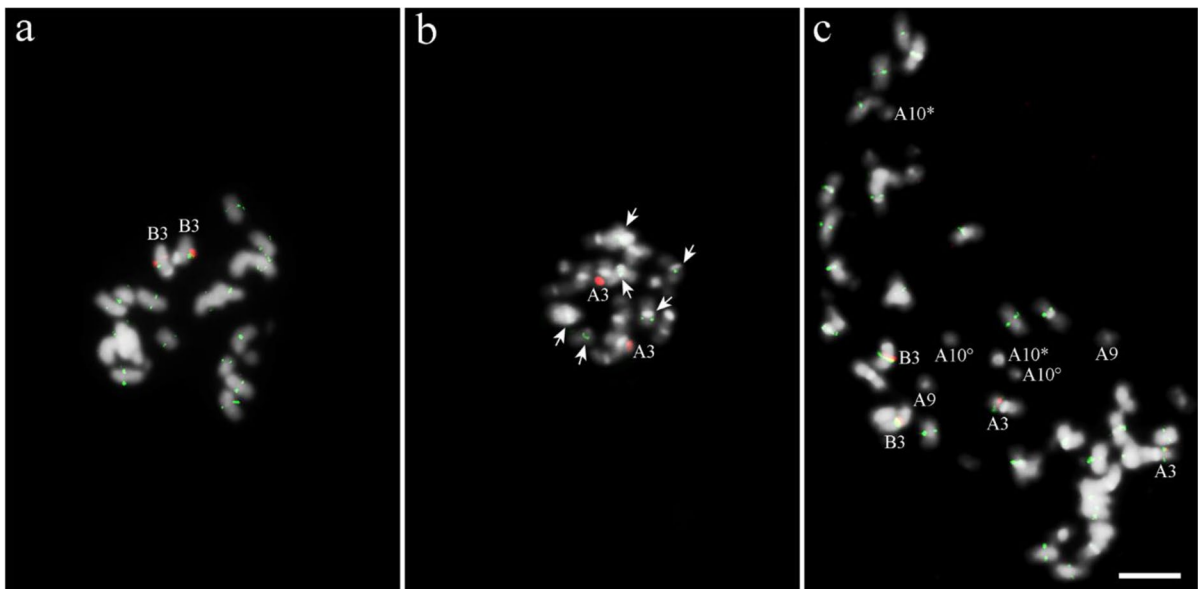
To compare the distribution patterns of TAR30 between the two diploid species that generated the synthetic allotetraploid ValSten and to better understand TAR30 distributions in chromosomes in wild diploids and in newly obtained allotetraploid genotype, further FISH analysis of TAR30 was conducted together with rDNA 5S loci detection. Eighteen chromosomes without DAPI<sup>+</sup> bands and two chromosomes with DAPI<sup>+</sup> bands (B3) were found in the mitotic cells of *A. valida* (BB-type genome) with TAR30 FISH signals observed in the proximal regions of all 20 chromosomes, but with the signals more evident in some of the chromosomes (Fig. 3a). An overlap of TAR30 and 5S rDNA hybridization signals was observed in the chromosome B3, indicating overlap or close proximity of this ribosomal DNA region with the TAR30 sequence. All twenty chromosomes of *A. stenosperma* (AA-type genome) had DAPI<sup>+</sup> bands and the TAR30 hybridization signals were found on the proximal regions of only three pairs of chromosomes (white arrows in Fig. 3b).

Forty mitotic chromosomes were observed in the new synthetic wild ValSten (Gao et al. 2021), equal



**Fig. 2** FISH analysis of two peanut cultivars, “Runner IAC 886” (a) and “Tifrunner” (b). DAPI<sup>+</sup> bands (bright white) on centromere regions of the chromosomes of subgenome A lacking in another 20 chromosomes of the subgenome B. Hybridization with TAR30 probe (green signals) on the proximal

region of one pair of chromosomes of the subgenome A (white arrows) and hybridization signals (green) in three pairs of chromosomes in “Runner IAC 886” (yellow arrows), together with the 5S rDNA probe (red signals) (a, b) in chromosomes cyt-A3 (A3) and cyt-B3 (B3). Scale bar: 5 μm



**Fig. 3** FISH analysis of TAR30 and 5S rDNA in *A. valida* (a), *A. stenosperma* (b), and the new synthetic allotetraploid ValSten (c). DAPI<sup>+</sup> bands (bright white) on the centromere regions of the 10 pairs of chromosomes of the AA wild type (b), lacking in 20 chromosomes of BB wild type (a), while present in half of the 40 chromosomes of ValSten (c). Hybridization with TAR30 probe (green signals) in all chromosomes of *A. valida* (a), in three pairs of chromosomes in *A. stenosperma* (b) (white arrows), and in 18 pairs of chromosomes in ValSten

(c). 5S rDNA loci (red signals) were detected in only one pair of chromosomes, cyt-B3 (B3) in *A. valida* (a) and cyt-A3 (A3) in *A. stenosperma* (b) and in both pairs of chromosomes (cyt-A3 and B3; A3 and B3) in ValSten (c). Hybridization signals of both probes (green and red) were detected in cyt-B3 of *A. valida* and cyt-B3 and A3 in ValSten (c). A10 with short arm and proximal segment of the long arm (A10\*) and the satellite region (A10°) and small pair A (A9). Scale bars: 5 μm

to the sum of number of chromosomes of its diploid parental species, *A. stenosperma* and *A. valida*. FISH signals were detected in 36 chromosomes of them, with varying signal strength among the chromosomes and lacking detection only in the chromosome pairs A9 (small pair) and A10 (SAT chromosomes) (Fig. 3c). The 36 signals more likely correspond to 18 pairs of chromosomes, which is more than the sum of the number of chromosomes with TAR30 signals in *A. stenosperma* (six chromosomes) and *A. valida* (twenty chromosomes). The colocalization of TAR30 and 5S rDNA hybridization signals on the chromosome B3 of ValSten was similar to that in the B3 of the *A. valida*. The TAR30 and 5S rDNA hybridization signals were also colocalized in the A3 of ValSten, but not in the A genome species, *A. stenosperma*, indicating differences of genome organization after tetraploidization.

The newly synthesized tetraploid wild ValSten showed FISH signals of TAR30 in more chromosomes than that in both cultivated peanut varieties. The hybridization signals in ValSten were evidently closer to the centromere region whereas the signals in the two peanut cultivars were dispersed along chromosome arms.

#### Coexistence of TAR30 with TTTAGGG telomeric repeat

Previous cytogenetic analyses revealed that the FISH signals of TTTAGGG tandem repeat were detected in several *Arachis* species including both cultivated peanut and its two wild progenitors (Du et al. 2016; Zhang et al. 2016); however, it is not clear about the abundance of the telomeric repeat in the peanut genomes. We used the telomeric repetitive sequence of *Arabidopsis* (Richards and Ausubel, 1988) to search against the genomes of wild and cultivated peanuts (Bertioli et al. 2016, 2019). Multiple significant hits ( $< 1 \times e^{-100}$ ) with the basic motif of TTTAGGG were identified, thus confirming the high conservation and presence of TTTAGGG repeat in peanuts.

Highly repetitive sequences make some regions including centromeres and telomeres difficult to sequence completely and assemble correctly (Bzikadze and Pevzner 2020). To provide more accurate estimation on their abundance, both TAR30 (PIVG01000013: 67,234,205–67,235,790) and TTT

AGGG repeats in *A. hypogaea* (PIVG01000012: 72,642–74,255) were used to screen the PacBio long-read sequencing data sets from both cultivated peanut and its diploid wild parents (Bertioli et al. 2016, 2019). Our analysis indicated that TTTAGGG telomeric repeat is more abundant than TAR30 in both wild and cultivated genomes (Supplementary Fig. 1). The coverage of TTTAGGG telomeric repeat in cultivated peanut was nearly equal to the sum of that detected in two diploid wild species. However, TAR30 in cultivated peanut genome is two times more than the sum of the repeat in two diploid wilds suggesting recent amplification of TAR30 in cultivated peanut after the polyploidization or domestication.

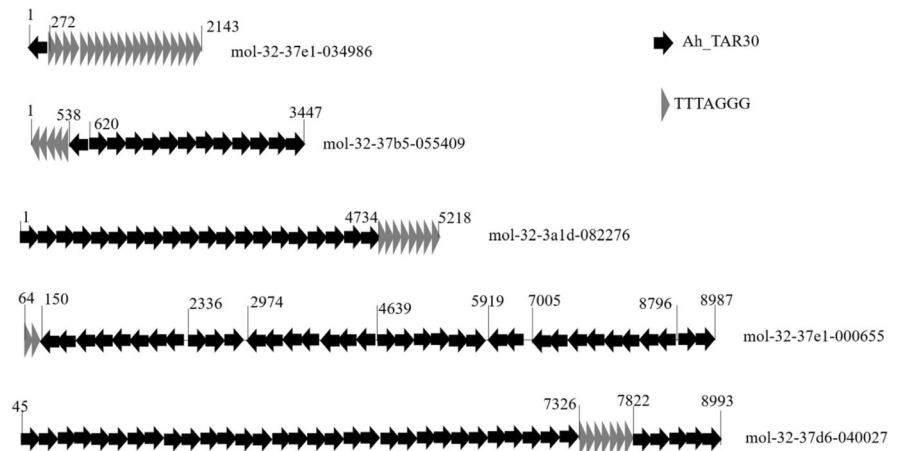
Most of the masked long reads contained either TAR30 or TTTAGGG repeat; however, some reads contained both tandem repeats. All these reads were extracted and manually inspected. Twenty-six reads from cultivated peanut harbored both TAR30 and TTTAGGG telomeric repeat which were organized into different tracts (Fig. 4). The coexistence of TAR30 and TTTAGGG repeat was also found in 21 reads from *A. ipaensis* and in 10 reads from *A. duranensis*. Our long-read surveys show that TAR30s and TTTAGGG repeats are mostly located in different genomic regions, although some of them colocalized.

#### Expression and genic contributions of TAR30

Tandem repeats play important roles not only in functional centromeres and telomeres but also in gene evolution as they may serve as coding sequences and generate structural and functional divergence of genes between relatives (Gemayel et al. 2012; Sulovari et al. 2019). We searched against the cDNA sequences in GenBank and identified 12 significant hits ( $< 1 \times e^{-10}$ ) including 11 cDNAs from developing embryo or young leaves in Tifrunner (Supplementary Table 2). These cDNAs were used to conduct BLASTX search against the NCBI protein sequences (<https://blast.ncbi.nlm.nih.gov>); no significant hit was found for 10 cDNAs suggesting that these cDNAs likely serve as non-encoding sequences such as untranslated regions (UTRs) or they encode new or unique proteins not yet annotated in GenBank. However, one TAR30-related cDNA sequence (LN672472), which was collected from the mixed samples treated by various biotic and abiotic stresses in a Chinese peanut cultivar Minhua6, showed sequence similarity to glutaminyl-peptide



**Fig. 4** Five PacBio long reads of cultivated peanut which contain both TAR30 and TTTAGGG telomeric repeats. The orientations of different tandem repeat tracts are indicated by the black arrows (TAR30) or gray triangles (TTTAGGG)



cyclotransferase ( $2 \times e^{-29}$ ). These results indicated that some TAR30 sequences were expressed and likely served as either non-coding RNA or parts of functional genes expressed during stress. To gain a global view on its impact on genes, we compared the locations of TAR30 sequences with the annotated genes in cultivated peanut (Bertioli et al. 2019). Thirteen genes were found to overlap with TAR30 sequences indicating the TAR30 sequences have been recruited as the parts of these genes (Table 2). It includes the gene encoding MYB family transcription factor (TF) which plays a key function in regulatory networks controlling development, metabolism, and responses to biotic and abiotic stresses (for review: Dubos et al. 2010).

#### Identification of transposons in TAR30 tracts

We screened the cultivated peanut genome with our repeat database including both transposons and tandem repeats and identified 92 tracts of TAR30 (here we define tracts as the tandem arrays of TAR30 which sizes are larger than 800 bp). We further manually inspected and compared the TAR30 tracts with the genomic distributions of transposons. Thirty-five TAR30 tracts were interrupted by transposons including both class I and class II transposons, but LTR retrotransposons were dominant in both abundance and coverage (Supplementary Table 3). A 48,271-bp tract on chromosome 1 for which a mutator transposon Ah\_mu4 inserted into TAR30 repeats and then

**Table 2** A list of TAR30 related annotated genes in *A. hypogaea* Tifrunner

Annotated genes	Chromosome	Locations of genes	Locations of repeats	Gene functions
Z9P2D3.1	Chr4	58,946,181–58,948,728	58,946,176–58,946,241	Uncharacterized protein
WMV57Q.1	Chr8	19,627,097–19,647,502	19,627,995–19,634,435	MYB family transcription factor
BUZ6CA.1	Chr9	120,474,641–120,493,717	120,480,199–120,493,741	Unknown protein
ZIF87W.1	Chr12	171,136–182,060	171,041–180,563	Uncharacterized protein
9RVJ5A.1	Chr12	254,881–291,127	268,170–290,571	Uncharacterized protein
6M2V35.1	Chr12	571,090–572,694	571,764–572,164	Uncharacterized protein
ZA9B2K.1	Chr12	814,777–816,525	815,406–815,793	Uncharacterized protein
AQSQ6B.1	Chr12	863,651–865,519	864,158–864,523	Unknown protein
H0WP65.1	Chr14	3,808,830–3,824,216	3,808,777–3,824,254	Glutamine-rich protein 2
63SMXT.1	Chr15	158,228,589–158,242,035	158,229,020–158,242,109	Glutamine-rich protein 2
GG1Z4T.1	Chr20	71,138–93,731	74,461–93,065	Unknown protein
V12SW3.1	Chr20	4,797,342–4,803,482	4,801,655–4,801,738	Unknown protein
6R9DCH.1	Chr20	141,413,078–141,426,967	141,425,592–141,425,748	Unknown protein

served as the bottom element for hosting four LTR retrotransposons (Fig. 5). Among the four LTR retrotransposons, Ah\_Feral is a complete retroelement with two intact LTRs and flanked by 5-bp TSD which allowed us to calculate the insertion time of this element based on the sequence divergence of its two LTRs (Gao et al. 2009). Our result indicates that this element inserted into the peanut genome about 0.35 million years ago (MYA). To further gain insights into the genomic dynamics mediated by TAR30 and transposons, 2-kb sequence for each end of the track (1–2000, 46,272–48,271) which includes both TAR30 and the mutator transposon sequences was used to search against GenBank; both 2-kb ends were detected in the subgenome A of *A. hypogaea* and *A. duranensis* (AA genome) but not in the subgenome B of *A. hypogaea* and *A. ipaensis* (BB genome); thus, all these transposons likely inserted into the TAR30 region after the divergence of the A and B genomes that occurred about 2.2 MYA (Bertioli et al. 2016).

## Discussion

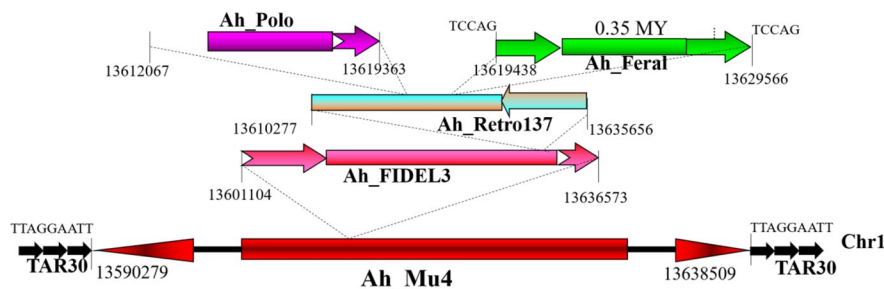
### Evolutionary origin of TAR30

In this study, we analyzed the cultivated peanut genome and identified a new tandem repeat TAR30 containing the 10-bp of basic motif of TTTT(C/T)TAGGG. Comparative analyses revealed that TAR30 is a homolog of the *Arabidopsis*-type telomeric repeat TTTAGGG in 112 plant genomes. Additionally, the homologous sequences of TAR30 in 17 other plants contained TG-rich tandem repeats such

as TTTTAGGG which represents another type of plant telomeric repeat in *Chlamydomonas* (Petracek et al. 1990; reviewed in Peska and Garcia 2020). It should be noted that strict criteria ( $>200$ -bp hit size and  $E$  value of  $\leq 1 \times 10^{-10}$ ) were used to define the homologous sequences of TAR30. All homologous sequences with TTTAGGG or other TG-rich motifs were the hits with the best alignment scores and lowest  $E$  values. From this, we conclude that TAR30 is a homolog of CPTR and it likely shared a common evolutionary origin with the canonical TTTAGGG repeat. No homolog of TAR30 was found in the sequenced eukaryotic mitochondria and plant chloroplasts suggesting that TAR30 was likely derived from variations of nuclear DNA sequence. Given that the basic motif of TAR30 is found in both A and B genomes of wild and cultivated peanuts but not in other sequence legumes such as soybean (*Glycine max*), chick pea (*Cicer arietinum*), and barrel medic (*Medicago truncatula*) (Supplementary Table 2), TAR30 likely emerged after the split of the *Arachis* genus and other legumes but before the divergence of two *Arachis* sub-genomes (A and B genome). As TAR30 is abundantly present in the *Arachis* genomes, it was possible that TAR30 loci served as fragile sites and play some role in *Arachis* genome evolution or instability via homologous exchanges, chromosomal rearrangements, and other mechanisms (Mondello et al. 2000; Dvořáčková et al. 2015).

### Genomic distributions of TAR30

In a wide range of plants, spanning flowering plants to algae, FISH signals of telomeric tandem repeats



**Fig. 5** Graphic display of nested transposons. An array of TAR30 is interrupted by a complete mutator transposon which is flanked by 9-bp TSD (TTAGGAATT) and serves as the host element for four LTR retrotransposons. The concave-pointed

rightwards arrows represent intact LTRs while the ribbon arrows mean truncated LTRs. The insertion time of the retrotransposon with two intact LTRs was calculated and is shown

including TTTAGGG and other variants were only detected at the ends of plant chromosomes, and not in other chromosome regions (Röder et al. 1993; Fuchs et al. 1995; Sýkorová et al., 2003; Mizuno et al. 2006; Peška et al., 2015; Yang et al. 2017). However, an increasing body of work shows that telomeric repeats are also located in other regions including the centromeres or pericentromeres in some *Solanum* species and other eukaryotes (Nergadze et al. 2004; Tek and Jiang 2004; He et al. 2013; Du et al. 2016; Zhang et al. 2016). The telomeric-like repeats found in non-terminal regions of chromosomes including centromeric, pericentromeric regions and between centromeres and telomeres are called interstitial telomeric repeats (ITRs) (Bolzán 2017). TAR30 exhibited different distributions from that of the plant telomeric repeat TTTAGGG. Here, the TAR30 FISH signals were detected in the proximal regions of the chromosomes, including centromeric and interstitial regions, and clear lack of hybridization signals in the terminal region of the chromosomes (Fig. 2, 3). Previous results indicated that the FISH signals of common plant TTTAGGG repeat were mostly detected at the ends of chromosomes in both wild and cultivated peanuts with few signals in pericentromeric regions and other locations (Du et al. 2016; Zhang et al., 2016). Thus, TAR30 showed distinct chromosomal distributions from the TTTAGGG telomeric repeat in cultivated peanut and the diploid wild *Arachis* species which is consistent with our in silico analysis as only a tiny fraction of long PacBio reads contained both TAR30 and TTTAGGG repeat (Fig. 4). Overall, the enrichments of TAR30 signals in proximal regions of peanut chromosomes suggest that TAR30 is an interstitial homolog of TTTAGGG telomeric repeat in peanuts although we cannot rule that there is a little amount of TAR30 in the terminal regions of peanut chromosomes that impaired the microscope detection.

#### Sequence convergence between TAR30 and the TTT TTTAGGG telomeric repeat

Telomeric tandem repeats tend to be highly conserved. The *Arabidopsis*-type TTTAGGG repeat has been found across the plant kingdom (Petracek et al. 1990; Fuchs et al. 1995; reviewed in Peska and Garcia 2020). Thus, this telomeric repeat probably existed in the common ancestor of flowering plants. Transitions/switches of telomeric repeats have also been

reported in many plants including Asparagaceae, *Zostera marina*, and Iridaceae in which the *Arabidopsis*-type repeat has been lost and replaced by the human-type repeat (Sýkorová et al. 2006; Peska et al. 2020). Thus, the TTAGGG repeat in these plants should be newly emerged after the divergence of monocots and dicots. The repeated appearances of the human-type telomeric repeat in distantly related lineages might be a consequence of independent events which can be considered evolutionary convergence or “homoplasmy” of telomeric sequences. The recovery of TTTAGGG repeat in these plants indicates the extreme importance of telomeres and implies that the evolution of telomeric repeats have undergone strong but similar selection pressures which often result in similar outcomes (Parker et al. 2013).

TAR30, TTTT(C/T)TAGGG, is nearly identical to the telomeric tandem repeat in *Cestrum elegans* with TTTTTTAGGG motif (Peška et al., 2015). It was possible that the original type of TAR30 was TTTTTTAGGG and then a single nucleotide substitution (T to C) within TTTTTTAGGG sequence occurred, generating TTTTCTAGGG unit which further expanded through strand slippage during DNA replication (Schlötterer and Tautz, 1992). Indeed, there are numerous TTTTTTAGGG motifs in the arrays of TAR30 (Supplementary Fig. 2). One scenario was that TAR30 and the TTTTTTAGGG repeat have independently emerged in two different lineages. However, TAR30 sequences are enriched in proximal regions of peanut chromosomes (Fig. 2, 3) that is distinct from the telomeric repeat in *C. elegans* (Peška et al., 2015).

#### Genomic changes during peanut polyploidization and domestication

Polyploidy has played a critical role in the divergence of higher plants and the innovations of new phenotypic traits (for review: Van de Peer et al. 2009). However, the genomic changes in the process of plant polyploidization and domestication are still not well understood. Induced allotetraploids provide a valuable resource for understanding the genomic differences in early generations after polyploidization. Previous studies have revealed genomic changes during polyploidization including sequence elimination (Guo and Han 2014), exchange of repeats between parental chromosomes

(Lim et al. 2007), meiotic instability (Chester et al. 2012), transposon activation (Ramachandran et al. 2020), and sequence replacements (Lim et al. 2007). We genotyped two wild diploid species, *A. valida* (BB) and *A. stenosperma* (AA), and the earliest generation (S0) of their synthetic tetraploid ValSten (AABB) with a SNP array, and identified AAAA or BBBB alleles at some loci in ValSten suggesting homologous recombination during inter-specific hybridization and polyploidization (Gao et al. 2021). Our FISH analysis detected different hybridization patterns of TAR30 between the two diploid parents and their synthetic allotetraploid and identified the emergence of TAR30 signals on A3 chromosome and a few other chromosomes in ValSten (Fig. 3). The appearance of TAR30 on A3 chromosome in ValSten may be caused by rapid burst of TAR30 or homologous exchanges between parental chromosomes (Lim et al. 2007). The FISH signals of TAR30 were stronger on some chromosomes in cultivated peanut than that in the synthetic wild tetraploid (Fig. 2, 3) and the coverage of TAR30 in cultivated peanut was higher than the sum of two wild progenitor diploids (*A. duranensis* and *A. ipaensis*) (Supplementary Fig. 1) implying possible amplification of TAR30 after peanut polyploidization.

In conclusion, TAR30 likely represents a homolog of the *Arabidopsis*-type telomeric repeat. Unlike the canonical telomeric repeat, TAR30 is enriched in the interstitial and centromeric regions of peanut chromosomes and lacks visible signals at the ends of chromosomes. Distinct FISH signals of TAR30 were detected between two diploid wild species and their newly synthetic allotetraploid as well as cultivated peanut implying rapid changes of repetitive sequences during peanut polyploidization and domestication. Overall, our efforts provide new insights into the chromosome evolution and the genomic dynamics of repetitive sequences in the early generation of synthetic polyploids.

**Acknowledgements** We thank Drs. Corley Holbrook and Gongshe Hu for their valuable comments. We also thank the three anonymous reviewers for their helpful comments.

**Funding** This research was funded by the grants from National Peanut Board in the USA and Georgia Peanut Commission.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Disclaimer** Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

## References

- El Baidouri M, Carpentier MC, Cooke R, Gao D, Lasserre E, Llauro C, Mirouze M, Picault N, Jackson SA, Panaud O (2014) Widespread and frequent horizontal transfers of transposable elements in plants.
- Benson G (1999) Tandem Repeats Finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27:573–580
- Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EK, Liu X, Gao D, Clevenger J, Dash S, Ren L, Moretzsohn MC, Shirasawa K, Huang W, Vidigal B, Abernathy B, Chu Y, Niederhuth CE, Umale P, Araújo AC, Kozik A, Kim KD, Burow MD, Varshney RK, Wang X, Zhang X, Barkley N, Guimarães PM, Isobe S, Guo B, Liao B, Stalker HT, Schmitz RJ, Scheffler BE, Leal-Bertioli SC, Xun X, Jackson SA, Michelmore R, Ozias-Akins P (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat Genet* 48:438–446
- Bertioli DJ, Jenkins J, Clevenger J, Dudchenko O, Gao D, Seijo G, Leal-Bertioli SCM, Ren L, Farmer AD, Pandey MK, Samoluk SS, Abernathy B, Agarwal G, Ballén-Taborda C, Cameron C, Campbell J, Chavarro C, Chitkineni A, Chu Y, Dash S, El Baidouri M, Guo B, Huang W, Kim KD, Korani W, Lanciano S, Lui CG, Mirouze M, Moretzsohn MC, Pham M, Shin JH, Shirasawa K, Sinharoy S, Sreedasyam A, Weeks NT, Zhang X, Zheng Z, Sun Z, Froenicke L, Aiden EL, Michelmore R, Varshney RK, Holbrook CC, Cannon EKS, Scheffler BE, Grimwood J, Ozias-Akins P, Cannon SB, Jackson SA, Schmutz J (2019) The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat Genet* 51:877–884
- Bolán AD (2017) Interstitial telomeric sequences in vertebrate chromosomes: origin, function, instability and evolution. *Mutat Res* 773:51–65
- Burge CB, Karlin S (1998) Finding the genes in genomic DNA. *Curr Opin Struct Biol* 8:346–354
- Bzikadze AV, Pevzner PA (2020) Automated assembly of centromeres from ultra-long error-prone reads. *Nat Biotechnol* 38:1309–1316
- Chester M, Gallagher JP, Symonds VV, Cruz da Silva AV, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE (2012) Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proc Natl Acad Sci U S A* 109:1176–1181
- Choi JY, Abdulkina LR, Yin J, Chastukhina IB, Lovell JT, Agabekian IA, Young PG, Razaque S, Shippen DE, Juenger TE, Shakirov EV, Purugganan MD (2021)

- Natural variation in plant telomere length is associated with flowering time. *Plant Cell* 33:1118–1134
- Du P, Li LN, Zhang ZX, Liu H, Qin L, Huang BY, Dong WZ, Tang FS, Qi ZJ, Zhang XY (2016) Chromosome painting of telomeric repeats reveals new evidence for genome evolution in peanut. *J Integr Agr* 15:2488–2496
- Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L (2010) MYB transcription factors in Arabidopsis. *Trends Plant Sci* 15:573–581
- Dvořáčková M, Fojtová M, Fajkus J (2015) Chromatin dynamics of plant telomeres and ribosomal genes. *Plant J* 83:18–37
- Fernández A, Krapovickas A (1994) Cromosomas y evolución en Arachis (Leguminosae). *Bonplandia* 8:187–220
- Fuchs J, Brandes A, Schubert I (1995) Telomere sequence localization and karyotype evolution in higher plants. *Plant Syst Evol* 196:227–241
- Fulnecková J, Sevcíková T, Fajkus J, Lukesová A, Lukes M, Vlcek C, Lang BF, Kim E, Eliás M, Sykorová E (2013) A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biol Evol* 5:468–483
- Gao D, Gill N, Kim HR, Walling JG, Zhang W, Fan C, Yu Y, Ma J, SanMiguel P, Jiang N, Cheng Z, Wing RA, Jiang J, Jackson SA (2009) A lineage-specific centromere retrotransposon in *Oryza brachyantha*. *Plant J* 60:820–831
- Gao D, Jiang N, Wing RA, Jiang J, Jackson SA (2015) Transposons play an important role in the evolution and diversification of centromeres among closely related species. *Front Plant Sci* 6:216
- Gao D, Li Y, Kim KD, Abernathy B, Jackson SA (2016) Landscape and evolutionary dynamics of terminal repeat retrotransposons in miniature in plant genomes. *Genome Biol* 17:7
- Gao D, Chu Y, Xia H, Xu C, Heyduk K, Abernathy B, Ozias-Akins P, Leebens-Mack JH, Jackson SA (2018) Horizontal Transfer of non-LTR retrotransposons from arthropods to flowering plants. *Mol Biol Evol* 35:354–364
- Gao D, Araujo A, Nascimento E, Chavarro MC, Xia H, Jackson S, Bertoli D, Leal-Bertoli S (2021) ValSten: a new wild species derived allotetraploid for increasing genetic diversity of the peanut crop. *Genet Resour Crop Evol* 68:1471–1485
- Gemayel R, Cho J, Boeynaems S, Verstrepen KJ (2012) Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. *Genes (basel)* 3:461–480
- Genome Res* 24:831–838
- Graham MK, Meeke A (2017) Telomeres and telomerase in prostate cancer development and therapy. *Nat Rev Urol* 14:607–619
- Grewal SI, Jia S (2007) Heterochromatin revisited. *Nat Rev Genet* 8:35–46
- Guo X, Han F (2014) Asymmetric epigenetic modification and elimination of rDNA sequences by polyploidization in wheat. *Plant Cell* 26:4311–4327
- He L, Liu J, Torres GA, Zhang H, Jiang J, Xie C (2013) Interstitial telomeric repeats are enriched in the centromeres of chromosomes in *Solanum* species. *Chromosome Res* 21:5–13
- Holbrook CC, Culbreath AK (2007) Registration of ‘Tifrunner’ peanut. *J Plant Registrations* 1:124
- Hong JP, Byun MY, Koo DH, An K, Bang JW, Chung IK, An G, Kim WT (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
- Hudson WH, Ortlund EA (2014) The structure, function and evolution of proteins that bind DNA and RNA. *Nat Rev Mol Cell Biol* 15:749–760
- Kilian A, Stiff C, Kleinhofs A (1995) Barley telomeres shorten during differentiation but grow in callus culture. *Proc Natl Acad Sci U S A* 92:9555–9559
- Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-induced mutations in grape skin color. *Science* 304:982
- Lee HR, Zhang W, Langdon T, Jin W, Yan H, Cheng Z, Jiang J (2005) Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species. *Proc Natl Acad Sci U S A* 102:11793–11798
- Lim KY, Kovarik A, Matyasek R, Chase MW, Clarkson JJ, Grandbastien MA, Leitch AR (2007) Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. *New Phytol* 175:756–763
- Llorens C, Futami R, Covelli L, Domínguez-Escribá L, Viu JM, Tamarit D, Aguilar-Rodríguez J, Vicente-Ripolles M, Fuster G, Bernet GP, Maumus F, Munoz-Pomer A, Sempere JM, Latorre A, Moya A (2011) The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res* 39(Database issue):D70–74
- Long M, Betrán E, Thornton K, Wang W (2003) The origin of new genes: glimpses from the young and old. *Nat Rev Genet* 4:865–875
- Maxwell PH, Belote JM, Levis RW (2006) Identification of multiple transcription initiation, polyadenylation, and splice sites in the *Drosophila melanogaster* TART family of telomeric retrotransposons. *Nucleic Acids Res* 34:5498–5507
- Meyne J, Ratliff RL, Moyzis RK (1989) Conservation of the human telomere sequence (TTAGGG)<sub>n</sub> among vertebrates. *Proc Natl Acad Sci U S A* 86:7049–7053
- Mizuno H, Wu J, Kanamori H, Fujisawa M, Namiki N, Saji S, Katagiri S, Katayose Y, Sasaki T, Matsumoto T (2006) Sequencing and characterization of telomere and sub-telomere regions on rice chromosomes 1S, 2S, 2L, 6L, 7S, 7L and 8S. *Plant J* 46:206–217
- Mondello C, Pirzio L, Azzalin CM, Giulotto E (2000) Instability of interstitial telomeric sequences in the human genome. *Genomics* 68:111–117
- Nergadze SG, Rocchi M, Azzalin CM, Mondello C, Giulotto E (2004) Insertion of telomeric repeats at intrachromosomal break sites during primate evolution. *Genome Res* 14:1704–1710
- Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ (2013) Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* 502:228–231
- Pedrosa A, Sandal N, Stougaard J, Schweizer D, Bachmair A (2002) Chromosomal map of the model legume *Lotus japonicus*. *Genet* 161:1661–1672

- Peška V, Fajkus P, Fojtová M, Dvořáčková M, Hapala J, Dvořáček V, Polanská P, Leitch AR, Sýkorová E, Fajkus J (2015) Characterisation of an unusual telomere motif (TTTTTTAGGG)<sub>n</sub> in the plant *Cestrum elegans* (Solanaceae), a species with a large genome. *Plant J* 82:644–654
- Peska V, Garcia S (2020) Origin, diversity, and evolution of telomere sequences in plants. *Front Plant Sci* 11:117
- Peska V, Mátl M, Mandáková T, Vitales D, Fajkus P, Fajkus J, Garcia S (2020) Human-like telomeres in *Zostera marina* reveal a mode of transition from the plant to the human telomeric sequences. *J Exp Bot* 71:5786–5793
- Petracek ME, Lefebvre PA, Silflow CD, Berman J (1990) *Chlamydomonas* telomere sequences are A+T-rich but contain three consecutive G-C base pairs. *Proc Natl Acad Sci U S A* 87:8222–8226
- Piegu B, Guyot R, Picault N, Roulin A, Sanyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, Panaud O (2006) Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res* 16:1262–1269
- Podlevsky JD, Bley CJ, Omana RV, Chen QX, JJ, (2008) The telomerase database. *Nucleic Acids Res* 36:D339–343
- Ramachandran D, McKain MR, Kellogg EA, Hawkins J (2020) Evolutionary dynamics of transposable elements following a shared polyploidization event in the tribe Andropogoneae. *G3 (Bethesda)* 10:4387–4398
- Richards EJ, Ausubel FM (1988) Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell* 53:127–136
- Riha K, Fajkus J, Siroky J, Vyskot B (1998) Developmental control of telomere lengths and telomerase activity in plants. *Plant Cell* 10:1691–1698
- Röder MS, Lapitan NL, Sorrells ME, Tanksley SD (1993) Genetic and physical mapping of barley telomeres. *Mol Gen Genet* 238:294–303
- Schlötterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Res* 20:211–215
- Schwarzacher T, Heslop-Harrison J (2000) Practical in situ hybridization. BIOS Scientific Publishers Ltd., Oxford
- Seijo GJ, Lavia GI, Fernandez A, Krapovickas A, Ducasse D, Bertioli DJ, Moscone DEA (2007) Genomic relationships between the cultivated peanut (*Arachis hypogaea*—Leguminosae) and its close relatives revealed by double GISH. *Am J Bot* 94:1963–1971
- Shippen DE, McKnight TD (1998) Telomeres, telomerases and plant development. *Trend Plant Sci* 3:126–130
- Sulovari A, Li R, Audano PA, Porubsky D, Vollger MR, Logsdon GA; Human Genome Structural Variation Consortium, Warren WC, Pollen AA, Chaisson MJP, Eichler EE (2019) Human specific tandem repeat expansion and differential gene expression during primate evolution. *Proc Natl Acad Sci U S A* 116:23243–23253
- Sýkorová E, Lim KY, Kunická Z, Chase MW, Bennett MD, Fajkus J, Leitch AR (2003) Telomere variability in the monocotyledonous plant order Asparagales. *Proc Biol Sci* 270:1893–1904
- Sýkorová E, Leitch AR, Fajkus J (2006) Asparagales telomerases which synthesize the human type of telomeres. *Plant Mol Biol* 60:633–646
- Tek AL, Jiang JM (2004) The centromeric regions of potato chromosomes contain megabase-sized tandem arrays of telomere-similar sequence. *Chromosoma* 113:77–83
- Van de Peer Y, Maere S, Meyer A (2009) The evolutionary significance of ancient genome duplications. *Nat Rev Genet* 10:725–732
- Wessler SR (2006) Transposable elements and the evolution of eukaryotic genomes. *Proc Natl Acad Sci U S A* 103:17600–17601
- Xu Z, Wang H (2007) LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* 35:W265–W268
- Yang QF, Liu L, Liu Y, Zhou ZG (2017) Telomeric localization of the *Arabidopsis*-type heptamer repeat, (TTTAGG)<sub>n</sub>, at the chromosome ends in *Saccharina japonica* (Phaeophyta). *J Phycol* 53:235–240
- Zhang L, Yang X, Tian L, Chen L, Yu W (2016) Identification of peanut (*Arachis hypogaea*) chromosomes using a fluorescence in situ hybridization system reveals multiple hybridization events during tetraploid peanut formation. *New Phytol* 211:1424–1439

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.