



## Telomeric repeat evolution in the phylum Nematoda revealed by high-quality genome assemblies and subtelomere structures

Jiseon Lim, Wonjoo Kim, Jun Kim, et al.

*Genome Res.* 2023 33: 1947-1957 originally published online November 2, 2023

Access the most recent version at doi:[10.1101/gr.278124.123](https://doi.org/10.1101/gr.278124.123)

---

**References** This article cites 89 articles, 17 of which can be accessed free at:  
<http://genome.cshlp.org/content/33/11/1947.full.html#ref-list-1>

**Open Access** Freely available online through the *Genome Research* Open Access option.

**Creative Commons License** This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

# Telomeric repeat evolution in the phylum Nematoda revealed by high-quality genome assemblies and subtelomere structures

Jiseon Lim,<sup>1,2</sup> Wonjoo Kim,<sup>1,2</sup> Jun Kim,<sup>1,3,4</sup> and Junho Lee<sup>1,2,3</sup>

<sup>1</sup>Department of Biological Sciences, Seoul National University, Gwanak-gu, Seoul 08826, South Korea; <sup>2</sup>Institute of Molecular Biology and Genetics, Seoul National University, Seoul 08826, South Korea; <sup>3</sup>Research Institute of Basic Sciences, Seoul National University, Seoul 08826, South Korea; <sup>4</sup>Department of Convergent Bioscience and Informatics, College of Bioscience and Biotechnology, Chungnam National University, Daejeon 34134, South Korea

Telomeres are composed of tandem arrays of telomeric-repeat motifs (TRMs) and telomere-binding proteins (TBPs), which are responsible for ensuring end-protection and end-replication of chromosomes. TRMs are highly conserved owing to the sequence specificity of TBPs, although significant alterations in TRM have been observed in several taxa, except Nematoda. We used public whole-genome sequencing data sets to analyze putative TRMs of 100 nematode species and determined that three distinct branches included specific novel TRMs, suggesting that evolutionary alterations in TRMs occurred in Nematoda. We focused on one of the three branches, the Panagrolaimidae family, and performed a de novo assembly of four high-quality draft genomes of the canonical (TTAGGC) and novel TRM (TTAGAC) isolates; the latter genomes revealed densely clustered arrays of the novel TRM. We then comprehensively analyzed the subtelomeric regions of the genomes to infer how the novel TRM evolved. We identified DNA damage–repair signatures in subtelomeric sequences that were representative of consequences of telomere maintenance mechanisms by alternative lengthening of telomeres. We propose a hypothetical scenario in which TTAGAC-containing units are clustered in subtelomeric regions and pre-existing TBPs capable of binding both canonical and novel TRMs aided the evolution of the novel TRM in the Panagrolaimidae family.

[Supplemental material is available for this article.]

Because linear chromosomes originated from a circular structure, a specific DNA–protein complex has evolved to protect the ends of the chromosomes, that is, the telomere. Telomeres are typically composed of tandem arrays of telomeric-repeat motifs (TRMs) and telomere-binding proteins (TBPs) that can bind to the tandem arrays (Zhong et al. 1992). TBP binding interferes with the accession of other proteins. Hence, proteins that respond to DNA damage signals cannot bind to the exposed ends of linear chromosomes, and the exposed ends are not recognized as DNA damage sites (Van Steensel et al. 1998; De Lange 2009; Lazzerrini-Denchi and Sfeir 2016). This enables the telomere to maintain chromosome integrity.

TRM sequences are typically well conserved because TBPs attach to arrays of TRMs in a sequence-specific manner. However, dozens of variant TRMs have evolved in many taxa, including 150 or more motifs in fungi (Červenák et al. 2021) and seven or more motifs in plants (Peska and Garcia 2020). Furthermore, except for Hymenoptera, many arthropod species have unique telomere structures. For example, most Insecta species have telomere structures composed of telomere-specific retrotransposons interposed between canonical TRMs, whereas Diptera species have telomeres that are exclusively composed of retrotransposons and lack TRMs (Abad et al. 2004; Garavís et al. 2013; Zhou et al. 2022). It is yet to be elucidated how TRMs have been altered while interacting with TBPs. These diverse TRMs and telomere structures developed

from the coevolution of a TRM and several TBPs (Shakirov et al. 2009; Steinberg-Neifach and Lue 2015; Sepsiova et al. 2016; Červenák et al. 2019); however, they may have also originated from TRM-specific changes regardless of TBP changes if TBPs were capable of binding to several types of TRMs (Rotková et al. 2004; Fajkus et al. 2005; Kramara et al. 2010; Visacka et al. 2012; Tomáška et al. 2018; Červenák et al. 2019, 2021).

We focused our analysis on subtelomeric regions and the DNA damage–repair signatures generated during telomere evolution, especially telomere and subtelomere reconstruction by alternative lengthening of telomeres (ALT) mechanisms (Heaphy et al. 2011; Sobinoff and Pickett 2017). ALT, which refers to alternative mechanisms that substitute telomerase function for telomere maintenance at the chromosome ends (Bryan et al. 1995), is crucial in telomerase mutant eukaryotic cells, such as those in yeast, *Caenorhabditis elegans*, the mouse, and human, and can thus be used to model and interpret the evolution of TRM changes (Lundblad and Blackburn 1993; Bryan et al. 1995; Seo et al. 2015; Kim et al. 2020, 2021a,b). In these ALT models, homology-directed repair (HDR) mechanisms can add new telomeric components. HDRs exploit the homology between shortened telomeric sequences and tandem arrays of TRMs in other genomic loci to replicate the tandem array and adjacent sequences to the shorter end (Lydeard et al. 2007; Roumelioti et al. 2016; Kramara et al. 2018; Kim et al. 2020). Furthermore, a unique sequence inserted between TRMs, known as the template for ALT, may be used to repair

**Corresponding authors:** [elegans@snu.ac.kr](mailto:elegans@snu.ac.kr), [junkim@cnu.ac.kr](mailto:junkim@cnu.ac.kr)

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.278124.123>. Freely available online through the *Genome Research* Open Access option.

© 2023 Lim et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

shorter telomeric repeats even in wild-type *C. elegans*, resulting in the formation of a new subtelomeric structure (Kim et al. 2019a). These findings suggest that ALT mechanisms play an essential role in the evolution of telomeres and that the evolutionary route can be reconstructed by studying ALT activity traces left in subtelomeric sequences.

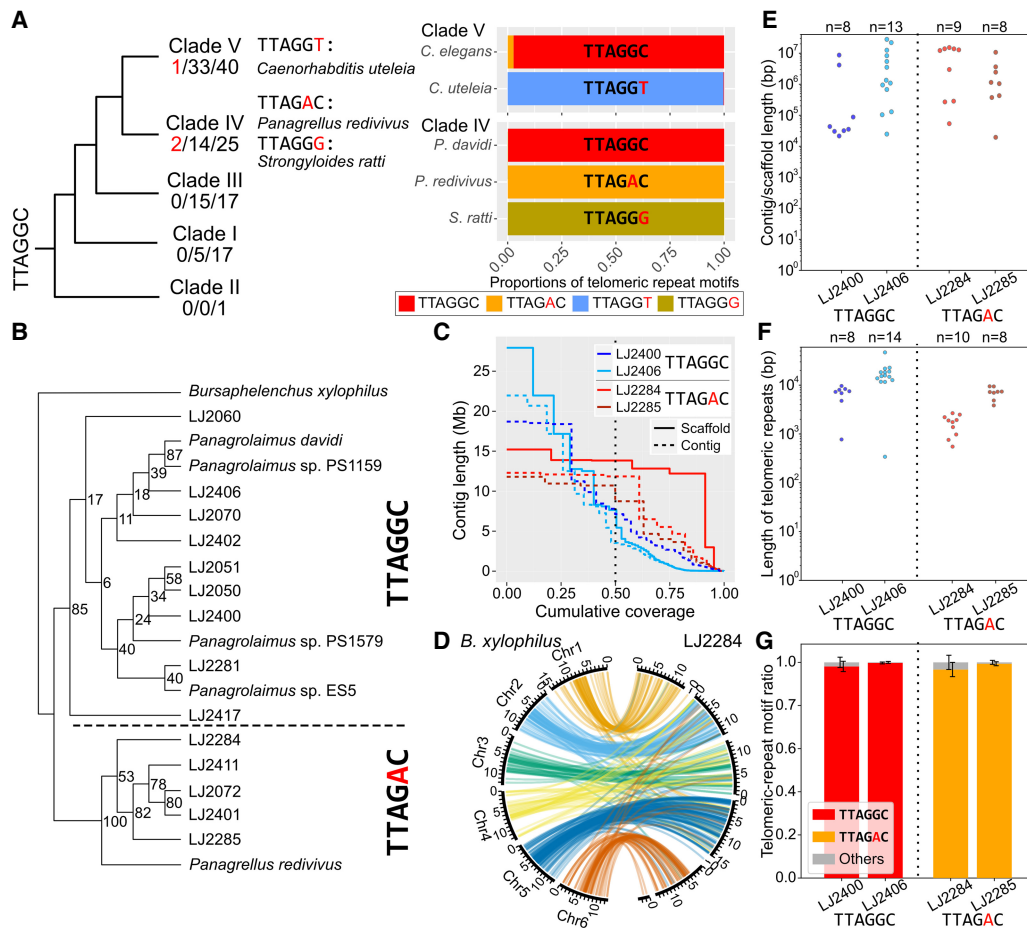
The present study explored TRM evolution in Nematoda using public whole-genome sequencing (WGS) data obtained from 100 nematode species and our high-quality genome assemblies. We evaluated the subtelomeric regions of TTAGGC-telomere isolates of Panagrolaimidae to infer possible evolutionary paths toward TTAGAC telomeres.

## Results

### Identification of three novel TRMs in Nematoda

We examined variations in TRMs using public short-read WGS data from Nematoda species that have matched public genome assemblies from all five major clades. A total of 100 species met the criteria, including at least 17 species from Clades I, III, IV, and V but only one from Clade II (Fig. 1A). The species and accession numbers used in this analysis are listed in Supplemental Tables S1 and S2.

The canonical TRM of Nematoda is known to be a 6-bp sequence (TTAGGC) that covers a >1-kb length in *C. elegans* telomeres (Yoshimura et al. 2019; Kim et al. 2019a). We examined the WGS



**Figure 1.** Putative novel TRMs in the phylum Nematoda and validation of the emergence of TTAGAC in the family Panagrolaimidae. (A) The cladogram was adapted and modified from Smythe et al. (2019) with permission from *BMC Ecology and Evolution* 2019, CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>). The three numbers separated by “/” under each clade indicate the number of species with novel TRMs/the number of species whose TRM was identified in our analysis/total number of species we analyzed, respectively. Changes from TTAGGC, the canonical Nematoda TRM, to the novel TRMs are highlighted in red. Below each novel TRM is a species name with the corresponding TRM. Each bar indicates the proportion of the canonical Nematoda TRM (TTAGGC) and novel TRMs in each species. The upper bar of each pair represents a control species that harbors the canonical TRM. The TRMs are represented as follows: red, TTAGGC; orange, TTAGAC; blue, TTAGGT; and khaki, TTAGGG. (B) Phylogenetic relationships of 19 Panagrolaimidae species/isolates based on their 18S rDNA sequences and their putative TRMs. *Bursaphelenchus xylophilus* was used as an outgroup species. The number on each node represents the bootstrap support value. (C) Contig/scaffold length distributions of the genome assemblies. The vertical dotted line indicates N50 contig/scaffold size. (D) Synteny plot of Lj2284 compared to *B. xylophilus*. Each colored line represents a BUSCO gene shared among the genome assemblies of our four isolates and *B. xylophilus*. Orange lines indicate that corresponding BUSCO genes are in Chromosome 1 in *B. xylophilus*. Sky blue, bluish green, yellow, blue, and vermillion represent Chromosomes 2, 3, 4, 5, and 6, respectively. We used Wong’s color palette designed for color-blind individuals (Wong 2011). (E–G) The vertical dotted line separates the TTAGGC-telomere isolates (left) and the TTAGAC-telomere isolates (right). (E) Length distributions of the contigs/scaffolds containing highly clustered telomeric repeats at the end. Each dot represents each contig/scaffold, and the total number of contigs/scaffolds of each isolate (n) is indicated at the top of the graph. (F) Estimated length distributions of clustered telomeric repeats at the end of the contigs/scaffolds. Each dot indicates a telomeric-repeat cluster for each contig/scaffold, and the total number of telomeric-repeat clusters for each isolate (n) is indicated at the top of the graph. (G) The proportion of TRM types in clustered telomeric repeats. Error bars represent the SD for all clustered telomeric repeats at the end of contigs/scaffolds in each isolate.

data from 100 species to identify novel TRMs with a concatemer of 5–7 bp, a high copy number, and a pattern similar to TTAGGC. We determined the TRM in 67 species using the following criterion: TRM repeats should be long concatemers for TBP binding, so TRM-like *k*-mer counts should be high enough (100 or more in 5 million short reads in our analysis) (Fig. 1A; Supplemental Table S1). However, 26 of the remaining 33 species showed low *k*-mer counts of TRM-like motifs, so we could not determine their TRMs (Supplemental Table S2). The remaining seven species had high *k*-mer counts, but we excluded them from the TRM-determined group because of the following reasons: We detected potential sample or host contamination in *Enoplus brevis*, two *Strongyloides*, and three *Trichinella* species, and *Diploscapter pachys* had only short concatemers (10 or fewer copies) in the middle of its reads (Supplemental Table S2; for more information, see Supplemental Note).

TTAGGC, the canonical Nematoda TRM, was identified in 64 of the 67 TRM-determined species (Fig. 1A; Supplemental Table S1). We determined that the other three nematode species had few TTAGGC copies but a high copy number of one of three putative TRMs: TTAGGT, TTAGAC, and TTAGGG in *Caenorhabditis uteleia*, *Panagrellus redivivus*, and *Strongyloides ratti*, respectively (Fig. 1A; Supplemental Fig. S1; Supplemental Tables S1, S3; for further validation of the TRMs, see Supplemental Note; Supplemental Table S4). The identified TRMs were new to Nematoda and had not been reported previously. We assumed that these three motifs evolved independently because they occurred in phylogenetically distant branches.

### The emergence of a new TRM, TTAGAC, in the family Panagrolaimidae

We focused on the TTAGAC sequence of the family Panagrolaimidae to confirm the noncanonical TRM sequence identified by short-read WGS data. This is because many different nematodes of this family were widely isolated from regions across South Korea, and some were easily cultured under laboratory conditions. We used 19 short-read WGS data sets of the family Panagrolaimidae to determine whether the unique motif TTAGAC emerged once or multiple times (Supplemental Table S5). WGS data sets of five species were available from publicly accessible databases, whereas the remainder were derived from sequencing data of the isolates newly collected from South Korea, which were probably distinct species as they had distinguishable ribosomal DNA (rDNA) sequences (Supplemental Table S6). *Bursaphelenchus xylophilus* was also used as the outgroup species.

We determined that *P. redivivus* and five out of 14 collected isolates had the novel TRM, TTAGAC, whereas all four *Panagrolaimus* species and the remaining nine isolates possessed the canonical TRM, TTAGGC. Moreover, the five collected TTAGAC-TRM isolates clustered with *P. redivivus*; however, the nine TTAGGC-TRM isolates were grouped with the *Panagrolaimus* species (Fig. 1B; Supplemental Table S5). Furthermore, the TTAGGC concatemer was not observed in any of the six TTAGAC-TRM species/isolates, nor was the TTAGAC concatemer found in any of the 13 TTAGGC-TRM species/isolates. These results suggest that TTAGAC sequences in Panagrolaimidae may have arisen from a single ancestor that had TTAGAC as its TRM.

### High-quality genome assemblies confirm that TTAGAC repeats are clustered at the ends of contigs or pseudochromosome molecules in TTAGAC-TRM isolates

We constructed de novo genome assemblies of Panagrolaimidae isolates using Pacific Biosciences (PacBio) high-fidelity (HiFi)

long-read sequencing technology and Arima Hi-C technology to assess whether the newly identified TRMs presented clusters at the ends of the chromosomes. Two TTAGAC-TRM isolates, LJ2284 and LJ2285, and two TTAGGC-TRM isolates, LJ2400 and LJ2406, were selected for the study. We produced 3.1 Gb (43×), 2.9 Gb (55×), 17.4 Gb (26×), and 36.1 Gb (93×) HiFi reads for LJ2284, LJ2285, LJ2400, and LJ2406, respectively. The HiFi reads were then assembled into high-quality draft genomes in which the length of the longest contig and contig N50 varied from 11.80 Mb to 21.96 Mb and 3.28 Mb to 11.84 Mb, respectively. We also generated Hi-C data for LJ2284 and LJ2406 and scaffolded their contigs into larger chunks (Supplemental Fig. S2; Supplemental Table S7). The longest scaffold lengths increased to 15.2 Mb and 27.9 Mb, and their scaffold N50 increased to 13.8 Mb and 7.6 Mb for LJ2284 and LJ2406, respectively (Fig. 1C; Supplemental Table S8). Additionally, 91.4% of the LJ2284 contigs was connected as five scaffolds (Supplemental Fig. S2).

All of our genome assemblies had ~75% BUSCO completeness, which was comparable to that of the nearly complete genome assembly of *B. xylophilus* (71%) (Supplemental Fig. S3; Dayi et al. 2020). In addition, we analyzed the synteny relationships between *B. xylophilus* and our genome assemblies using their common single-copy orthologs. Pseudo-chromosome-level scaffolds of LJ2284 and contigs of LJ2285 showed marked collinearity for all the chromosomes, except Chromosome 4 of *B. xylophilus* (Fig. 1D; Supplemental Fig. S4; Supplemental Note). Chromosomes 5 and 6 of *B. xylophilus* also showed collinearity with clusters of LJ2400 contigs and LJ2406 scaffolds, but the collinearity of other chromosomes was much weaker (Supplemental Fig. S4; Supplemental Note). These findings suggest that our Panagrolaimidae genome assemblies were highly contiguous.

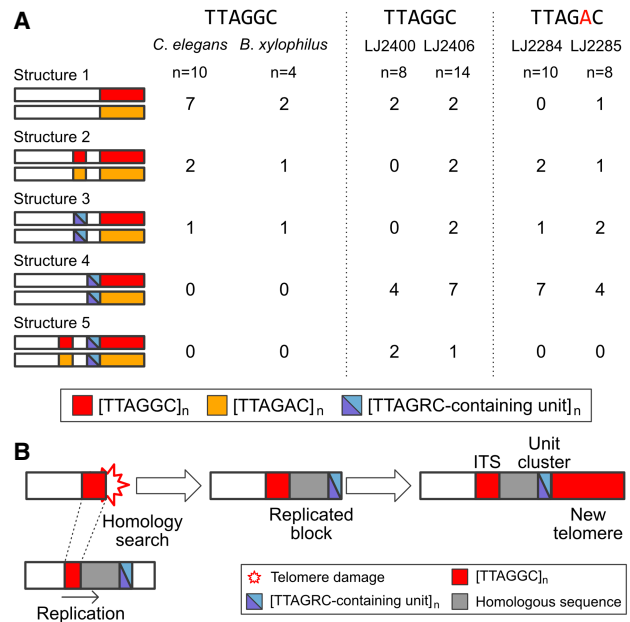
The constructed high-quality genome assemblies revealed that the telomeric repeats of both LJ2284 and LJ2285 changed from TTAGGC to TTAGAC. First, we confirmed whether the telomeric repeats had been adequately assembled. Our analysis revealed that each genome assembly had eight to 13 contigs/scaffolds ending with telomeric repeats (Fig. 1E). Among these contigs/scaffolds, one each of LJ2284 and LJ2406 was assembled at the telomere-to-telomere level (Supplemental Tables S7, S8). The telomeric repeats at the ends of the contigs/scaffolds in each genome assembly showed various ranges of length in the Panagrolaimidae isolates (Fig. 1F; Supplemental Table S9). Their lengths in LJ2284 ranged from 0.5 to 2.7 kb, but they were much longer in LJ2285 and LJ2400, ranging from 3.8 to 9.7 kb (except for one in LJ2400). LJ2406 showed telomeric-repeat clusters >10 kb (except for one cluster). We identified that 84.9%–99.4% of raw HiFi reads were >10 kb in all four isolates and that LJ2406, which has much longer telomeres than the other three isolates, had 328 reads consisting only of 10- to 26-kb telomeric repeats (Supplemental Fig. S5). Therefore, the TRM cluster lengths at each contig/scaffold end would be highly correlated with their actual lengths (for detailed description, see Supplemental Note). The lengths of the clustered TRMs were comparable to those of the two *C. elegans* assemblies constructed using long-read sequencing technologies (2.3–5.7 kb) (Kim et al. 2019a; Yoshimura et al. 2019). Furthermore, the telomeric repeats consisted mainly of TTAGAC in LJ2284 (96.67%) and LJ2285 (99.35%) and of TTAGGC in LJ2400 (98.11%) and LJ2406 (99.73%) (Fig. 1G; Supplemental Fig. S6; Supplemental Table S9). These findings supported our TRM study using short-read WGS data and revealed that the TTAGAC telomere had evolved in the family Panagrolaimidae.

## TTAGAC-containing units cluster near telomeric regions of TTAGGC-telomere isolates

Because most of the nematode species that we analyzed had a canonical Nematoda TRM sequence, TTAGGC, the TTAGAC telomere probably evolved from its ancestral TTAGGC telomere via a single-nucleotide mutation of the telomerase RNA component gene. We hypothesized that subtelomeric sequences can provide evidence regarding the evolutionary mechanism because these can carry historical records of telomere dynamics. For example, it has previously been documented that ALT mechanisms have been used to reconstruct subtelomeric and telomeric regions in *C. elegans* and that templates for ALT in subtelomeric regions can be used to repair telomere damage or maintain the integrity of the telomere (Seo et al. 2015; Kim et al. 2019a, 2020, 2021b; Lee et al. 2022b). Thus, we analyzed subtelomeric regions of our four Panagrolaimidae genome assemblies to trace ancestral telomere damage and repair events that could explain how the TTAGAC telomere evolved from the TTAGGC telomere.

Most telomeric repeats in telomeric regions (seven of 10 in LJ2284, four of eight in LJ2285, six of eight in LJ2400, and eight of 14 in LJ2406) were attached to unit clusters containing TTAGGC or TTAGAC (e.g., [TTAGGCCTAATTGC]<sub>n</sub> or [TTAGACTTATTCGC]<sub>n</sub>) (Fig. 2A; Supplemental Table S10). This finding was striking, as none of the subtelomeric regions of *C. elegans* and *B. xylophilus* showed TRM-containing unit clusters directly attached to the telomeric regions (Fig. 2A; Supplemental Table S11). Furthermore, both TTAGGC-containing unit clusters and TTAGAC-containing unit clusters were identified in all genomes of the TTAGGC- and TTAGAC-telomere isolates, with the exception of the LJ2285 genome (TTAGAC telomere), which contained only TTAGAC-containing unit clusters. For LJ2284, LJ2400, and LJ2406, these clusters consisted of repetitive units with length distributions ranging from 14 bp to ~2 kb, and these repetitive units were tandemly repeated on an average of 19 or more copies (Supplemental Tables S10, S12). The LJ2285 genome did not contain such same-length unit clusters; however, it contained sequence clusters consisting of five to nine copies of units containing TTAGAC and similar sequences (10–74 bp). These TRM-containing unit clusters were not found outside the subtelomeric region in the telomere-containing contigs/scaffolds (for a description of a contrasting case in *C. elegans*, see Supplemental Note), raising the possibility that these clusters are associated with telomere maintenance in Panagrolaimidae. Notably, we also identified a similar subtelomeric pattern in *C. uteleia*, the noncanonical TTAGGT-TRM species, in which TTAGGT- and/or TTAGGC-containing unit clusters were attached to or close to long TTAGGT repeat arrays (obtained from the NCBI Sequence Read Archive [SRA; <https://www.ncbi.nlm.nih.gov/sra>] under accession number ERR8978452) (Supplemental Fig. S7; Supplemental Note; The Darwin Tree of Life Project Consortium 2022).

In addition to the TRM-containing unit clusters, interstitial telomeric sequences (ITSs) were frequently observed in the subtelomeric regions of the constructed genome assemblies, which were consistent with the ITS enrichment present in *C. elegans* chromosome arms (The *C. elegans* Sequencing Consortium 1998). All subtelomeric structures were categorized based on the location of the unit clusters and/or the ITSs in subtelomeric regions (up to 200 kb from the end of the telomeric contig/scaffold) (Supplemental Tables S10, S11). We annotated a total of 12 subtypes (Supplemental Fig. S8), but for simplicity, we merged them into five major subtelomeric structures by considering only telomeric TTAGRC-containing unit



**Figure 2.** Subtelomere structures and a proposed model to explain clustered blocks of TRMs or TTAGRC-containing unit clusters in the subtelomeric regions. (A) Schematic representation of subtelomere structures in the contigs/scaffolds listed in Supplemental Table S10. We categorized the structure of subtelomeric regions (up to 200 kb from the end of the telomeric contig/scaffold) using the following criteria: (1) whether it has ITSs (shorter red blocks) and/or TTAGRC-containing unit clusters (half-sky blue/half-violet blocks) and (2) whether ITSs, unit clusters, and telomeric regions were directly attached or separated by other sequences (white, empty blocks). Each horizontal bar represents a subtelomere structure with only a telomeric TTAGRC-containing unit cluster (shown in structures 4 and 5) and an ITS or a TTAGRC-containing unit cluster closest to the telomere (shown in structures 2 or 3). Each value refers to the number of contigs/scaffolds with the corresponding structure type in each species/isolate. TTAGGC-telomere species and isolates had ITSs that were composed of TTAGGC, rather than TTAGAC, and TTAGAC-telomere species and isolates had only TTAGAC-type ITSs, too. (B) A proposed model for generating subtelomere structures responsible for structure 5. After telomere damage, HDR mechanisms can exploit homology between shortened telomeric repeats and ITSs at other loci to repair the damaged telomere. A HDR mechanism, break-induced replication (BIR), may replicate sequences near the ITS, creating a new homologous block of the original sequence block. If the original template block shows a TTAGRC-containing unit cluster, the cluster would also be replicated. Finally, a new telomere composed of TTAGGC repeats can be replenished by active telomerase.

clusters and ITSs or TTAGRC-containing unit clusters closest to the telomere (Fig. 2A, structures 1–5; for examples of structures 2–5, see Supplemental Fig. S9). Notably, the motif in all ITSs was the same as their TRM. The four genome assemblies showed similar patterns. For most of the subtelomeric structures in the four genome assemblies, the TRM-containing unit cluster was positioned directly adjacent to the telomeric region. Structure 1, which did not contain any unit cluster or ITS, accounted for fewer than a quarter of all the subtelomeric structures (the lengths of telomeres in structure 1 differ from those of ITSs in *C. elegans* and our four isolates) (see Supplemental Note). In contrast, structures 2–5 had at least one unit cluster or ITS separated from the telomeric region, which may have resulted from ancestral telomere damage and repair processes.

## TTAGAC-containing units may be used to maintain the integrity of TTAGGC telomeres

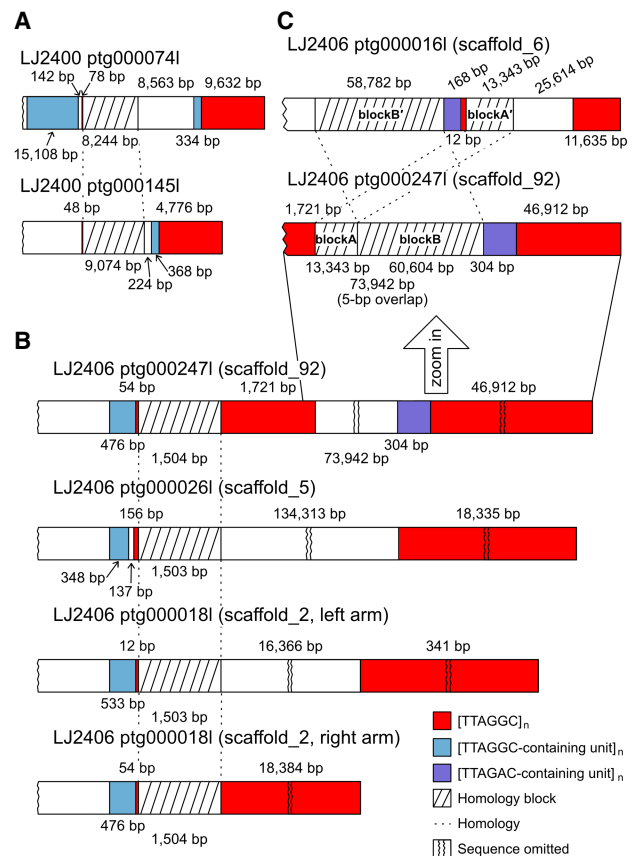
We postulated that the identified subtelomeric clusters of the TTAGRC-containing units and ITSs were evidence of ancestral

telomere damage and repair via ALT mechanisms and that, if true, this would help understand the role of TTAGRC-containing unit clusters in telomere maintenance. Among structures 2–5, structure 5 may be explained by break-induced replication (BIR), one of the most well-known ALT mechanisms for telomere repair (Malkova et al. 1996; Bosco and Haber 1998; Kim et al. 2019a, 2020). Telomeres that have been damaged or shortened should be repaired, and BIR may rely on homology between shortened telomeric repeats and ITSs at other loci (Fig. 2B, left side). The replication was then activated, and occasionally, sequences beside the ITS would also be replicated (Fig. 2B, middle). Furthermore, if the original homology block contained a TTAGRC-containing unit cluster, the cluster would also be duplicated simultaneously, resulting in a new subtelomere structure with a homology block interposed between the ITS and the TTAGRC-containing unit cluster. Finally, a new telomere composed of TTAGGC repeats could be replenished by active telomerase (Fig. 2B, right side). We then hypothesized that if structure 5 cases in our genomes had been generated via BIR, we should be able to identify another sequence block homologous to the interposed sequence between the ITS and the TTAGRC-containing unit cluster.

To test this hypothesis, we searched in whole-genome sequences for homology blocks of the sequences representing structure 5 (ptg000074l and ptg000145l of LJ2400 and ptg000247l of LJ2406) (Figs. 2, 3; Supplemental Fig. S8; Supplemental Table S10). ptg000074l and ptg000145l of LJ2400 shared 8-kb and 9-kb homology blocks adjacent to the ITS and had a TTAGGC-containing unit cluster ~9-kb and ~200-bp away from the shared block, respectively (Fig. 3A). These 8-kb and 9-kb homology blocks pointed to the possibility of their replication via BIR using the ITS as homology.

The ptg000247l of LJ2406 had a more evident BIR signature than the ptg000074l and ptg000145l of LJ2400, indicating that the subtelomeric TTAGRC-containing unit cluster was built via the ALT process. The subtelomeric region of ptg000247l of LJ2406 consisted of a TTAGGC-containing unit cluster, a 54-bp ITS, a 1.5-kb sequence block, a 1.7-kb ITS, a 74-kb sequence block, and a TTAGAC-containing unit cluster close to the telomeric region (Fig. 3B). By searching for this complex subtelomeric sequence in the whole genome using BLAST, we determined that the 1.5-kb and 74-kb blocks had homologous sequences in other subtelomeric regions. The 1.5-kb block between the 54-bp ITS and 1.7-kb ITS had almost identical homologous sequences in three different subtelomeric regions. These shared blocks were directly adjacent to short and variable-length ITSs near TTAGGC-containing unit clusters (Fig. 3B). The TTAGGC-containing cluster and ITS may have been used for homology to replicate the 1.5-kb homology blocks.

The 74-kb block between the 1.7-kb ITS and the TTAGAC-containing unit cluster possessed another trace of HDR and an additional replication of a TTAGRC-containing unit cluster after telomere damage (Fig. 3C). As the 1.7-kb length of the ITS is too long to be created by random mutation, it could represent evidence of ancestral telomere damage and repair. The 74-kb block contained at least two distinct homology blocks: the 13-kb blockA and the 61-kb blockB (Fig. 3C). The 13-kb blockA directly linked to the ITS in ptg000247l had an identical blockA' in ptg000016l that was also directly attached to another 180-bp sequence of an ITS and a TTAGC-containing unit cluster. The 61-kb blockB was homologous but not identical to the 59-kb blockB' located on the opposite side of the 180-bp block of the ITS and the TTAGAC-containing unit cluster in ptg000016l. Of note, the last 5 bp of the 13-kb blockA



**Figure 3.** Traces of telomere damage and repair processes through ALT mechanisms identified in subtelomeric regions of genome assemblies of TTAGGC-telomere isolates. (A–C) Schematic representations of subtelomeres in the genome assemblies of LJ2400 and LJ2406. Hatched boxes indicate homology blocks between two or more subtelomeres, and dashed lines indicate homologous relationships. The red boxes represent telomeric-repeat clusters. Violet and sky blue boxes represent TTAGAC-containing unit clusters and TTAGGC-containing unit clusters, respectively. Not to scale.

and the first 5 bp of the 61-kb blockB were identical (i.e., TAAAT). The 61-kb blockB was further divided into four blocks, each of which also showed microhomologies with the adjacent blocks (Supplemental Fig. S10; Supplemental Table S13). These data imply that telomere damage and repair may have occurred in the following order: First, the 13-kb blockA in ptg000247l was replicated by BIR using homology between the 1.7-kb ITS in ptg000247l and the 180-bp block of the ITS and the TTAGAC-containing unit cluster in ptg000016l. Next, the 61-kb blockB in ptg000247l was replicated by multiple rounds of microhomology-mediated BIR (MMBIR) and template switching events using microhomology sequences at the end of the blocks (Supplemental Fig. S10; Supplemental Note).

This 59-kb blockB' in ptg000016l was directly attached to the TTAGAC-containing unit cluster, which was duplicated in ptg000247l. Notably, the original TTAGAC-containing unit cluster in ptg000016l was shorter than the duplicated cluster in ptg000247l, implying that the duplicated cluster nearly doubled after MMBIR was completed. The duplicated TTAGAC-containing unit cluster in ptg000247l was directly linked to the TTAGGC telomere, indicating that the duplicated cluster was exposed at the

end. These lines of evidence support the notion that even in TTAGGC-telomere species, TTAGAC-containing units could be incorporated to constitute a telomere or at least serve as a DNA template for replenishing TRM repeats by telomerase. This use of TTAGAC-containing units in the TTAGGC-telomere isolate may have contributed to the evolution of the TTAGAC telomere from the TTAGGC telomere.

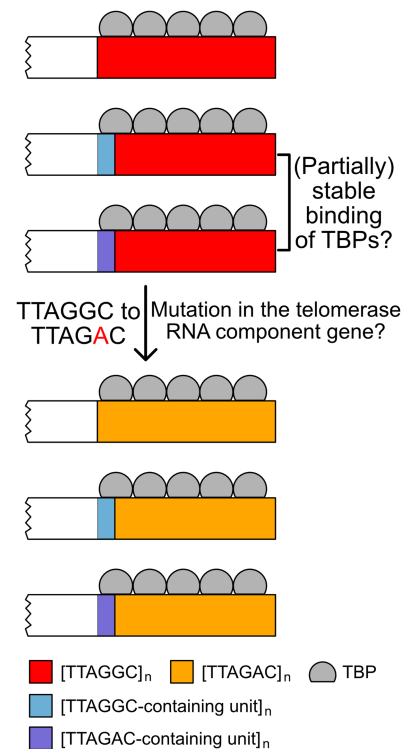
### Telomere-associated protein genes show similar conservation patterns among Panagrolaimidae species/isolates

We hypothesized that TTAGAC-containing unit clusters served as a component of partially stable telomere in TTAGGC-telomere species and that TBPs for TTAGGC telomere could have bound TTAGAC repeats and have been conserved in both TTAGGC-telomere and TTAGAC-telomere species. To support this hypothesis, we analyzed protein homologies in Panagrolaimidae with *B. xylophilus* as an outgroup species, in which proteins have been described to maintain or bind telomeres in *C. elegans* (Ahmed and Hodgkin 2000; Hofmann et al. 2002; Im and Lee 2003, 2005; Kim et al. 2003; Joeng et al. 2004; Meier et al. 2006, 2009; Boerckel et al. 2007; Raices et al. 2008; Ferreira et al. 2013; Shtessel et al. 2013; Dietz et al. 2021; Yamamoto et al. 2021). We identified 12 proteins that showed similar homology patterns between TTAGGC- and TTAGAC-telomere species/isolates (Supplemental Fig. S11). In contrast, four proteins (POT-2, TEBP-2 [also known as DTN-2], HPR-9, and MRT-1) displayed inconsistent conservation patterns. POT-2 was conserved in only two of the three TTAGGC-telomere species/isolates, and TEBP-2 remained in only one of the three TTAGAC-telomere species/isolates. HPR-9 was observed in only one TTAGAC-telomere isolate. MRT-1, required for telomerase activity in *C. elegans* (Meier et al. 2009), was not detected in *B. xylophilus* or in any of the three TTAGAC-telomere species/isolates but was present in all three TTAGGC-telomere species/isolates. However, MRT-1 functional domains were not conserved in the TTAGGC-telomere species/isolates. Nonetheless, homologs of the four TBPs were mostly conserved in all Panagrolaimidae species and isolates (Supplemental Table S14). Despite these variations, TRT-1 was conserved in all species and isolates, suggesting that they may have active telomerase and that changes in its RNA template may have been involved in TRM evolution (see Discussion) (Fig. 4).

## Discussion

TRMs and their protein partners are well conserved and mostly co-evolving because they must be connected with each other to stably maintain the DNA–protein complex in telomeres (Shakirov et al. 2009; Steinberg-Neifach and Lue 2015; Sepsiova et al. 2016; Červenák et al. 2019). Nevertheless, TRMs have evolved across several taxa (Fulnečková et al. 2013; Garavís et al. 2013; Peska and García 2020; Červenák et al. 2021). In this study, we identified three novel TRMs in Nematoda and confirmed that TTAGAC, one of the three novel TRMs, composes telomeres in a subset of Panagrolaimidae isolates. We also hypothesized that TTAGAC evolved through TTAGAC-containing unit clusters and the robustness of the TTAGGC-suited TBPs to bind tandem arrays of TTAGAC repeats in a canonical TTAGGC-telomere ancestor.

The consistent pattern of TRM changes observed in other multicellular organisms supported our hypothesis. First, sequence changes in plant TRMs also show key characteristics (Peska and García 2020). In particular, the basal TRM in plants, TTTAGGG (*Arabi-*



**Figure 4.** A model for TRM evolution in Panagrolaimidae. If TBPs with affinity for TTAGGC telomeres could bind to any of the TTAGGC telomere, TTAGGC-containing unit clusters, or TTAGAC-containing unit clusters, they were able to construct a fully or partially stable telomere. This binding robustness of TBPs may have facilitated the TRM conversion from TTAGGC to TTAGAC. Even after TRM changed via mutation of the telomerase RNA component gene, if it occurred, TBPs could still bind to altered telomeric repeats, which possibly contributed to the evolution of a novel TRM, TTAGAC, in the Panagrolaimidae family.

*dopsis*-type) (Richards and Ausubel 1988) evolved into TTTAGGG (vertebrate-type) (Weiss and Scherthan 2002; Sýkorová et al. 2003, 2006), TTTTAGGG (*Chlamydomonas*-type) (Fulnečková et al. 2012), TTTCAGG\_ (*Genlisea*-type) (Tran et al. 2015), TTTTTCAGG\_ (*Genlisea*-type) (Tran et al. 2015), TTTTTAGG\_ (*Klebsormidium*-type) (Fulnečková et al. 2013), and TTTTTTTAGGG (*Cestrum*-type) (Peška et al. 2015). Briefly, these variant TRMs showed alterations in the lengths of the T and G arms or changes in pyrimidine (T-to-C) nucleotides close to the middle A. Pyrimidine nucleotide conversion, rather than changes between pyrimidine and purine nucleotides and vice versa, has also been observed in insects in which TTAGG changed to TCAGG (Mravinac et al. 2011). Similarly, in the current study, TTAGGC changed to TTAGGT (pyrimidine conversion) and TTAGAC (purine conversion) in nematodes. Considering pyrimidines (T and C) have comparable base sizes and purines (G and A) are of similar sizes, we propose that conversions between pyrimidines or between purines do not sterically hinder TBP binding, allowing pre-existing TBPs to attach to the modified TRM. Furthermore, in yeasts and plants, TBPs are sufficiently robust to bind to various types of TRM despite several base pair changes (Rotková et al. 2004; Fajkus et al. 2005; Kramara et al. 2010; Visacka et al. 2012; Tomáška et al. 2018; Červenák et al. 2019, 2021). In yeast, a TBP binds to at least six different types of TRMs, implying binding robustness of TBPs during TRM evolution (Červenák et al. 2019). Here we suggest that this idea could be further supported by TTAGAC-containing unit clusters in TTAGGC-telomere species.

It is still unclear whether TTAGGC-binding TBP's also bind to TTAGAC in TTAGGC-telomere Panagrolaimidae species/isolates. Traces of ALT action identified in our genome assemblies revealed that TTAGAC-containing units were used to repair telomere damage in TTAGGC-telomere isolates. Two TTAGGC-telomere isolates and one TTAGAC-telomere isolate examined harbored both clusters of consecutive TTAGGC- and TTAGAC-containing units. Some clusters were attached directly or were adjacent to the telomeres. Specifically, the subtelomeric sequences of ptg0002471 in LJ2406, a TTAGGC-telomere isolate, showed evidence of direct usage of a TTAGAC-containing unit for the repair of a damaged telomere. In the vicinity of a damaged and shortened telomere, there were homologous sequence blocks and a TTAGAC-containing unit cluster, which were probably replicated by BIR and MMBIR, and the TTAGAC-containing unit cluster was directly adjacent to its telomeric TTAGGC cluster. Thus, at least over a brief period of time, the TTAGAC-containing unit cluster was exposed at the end of the chromosome before the TTAGGC repeats were replenished.

Furthermore, the TTAGAC-containing unit cluster was not only duplicated but also elongated at the end because it was nearly twice as long as its initial cluster template in a separate subtelomeric region. This cluster might have been elongated via replication slippage based on its repetitive nature or via ALT mechanisms that can replicate TRM-containing units at the end of the chromosome. We could not determine the specific mechanisms involved in the elongation of the TTAGAC-containing unit cluster. If the elongation was achieved via ALT mechanisms, this suggests that the TTAGAC-containing unit cluster shows its own replication capability in the telomeric region to maintain the telomere.

If the TTAGAC-containing unit cluster acted as a component of the telomere, the TTAGGC-suited TBP's would bind to the telomere; otherwise, the end would be recognized as a DNA damage site. This binding robustness of TBP's, if it exists as reported in fungi, may explain the subtelomere structure and TRM evolution. This robustness could be achieved via two potential ways. First, the TTAGGC-suited TBP's also showed sufficient binding affinity for TTAGAC, such that TTAGAC-unit clusters could be used for repairing telomere damage. This binding affinity stabilized the TTAGAC-containing unit–TBP complex as a telomere. Second, TBP's showed a weak binding affinity to TTAGAC and formed a partially stable telomere complex, and the affinity strengthened through frequent use of TTAGAC-containing unit clusters to repair telomere damage and ensure fitness. In both cases, TTAGGC-suited TBP's could also stabilize TTAGAC telomeres, which thus facilitated the conversion from TTAGGC telomeres to TTAGAC telomeres.

Our results suggest a plausible evolutionary model for the TRM conversion in the family Panagrolaimidae. First, ancestral TTAGGC-telomere species may have used various ALT mechanisms to counteract telomere damage when the active telomerase cannot access the damaged telomere (Fig. 4). Replication of TTAGGC-containing unit clusters could represent one such ALT mechanism and would have been used to repair damaged telomeres. Consequently, active telomerase may use unit clusters as a starting template to fully replenish the partially repaired telomere. Because TBP's can fully or partially stabilize telomere complexes consisting of TTAGAC-containing unit clusters, the evolution of the TRM conversion to TTAGAC had a greater fitness than other TTAGGC variants. We could not identify the telomerase RNA component gene in Panagrolaimidae. However, conservation of TRT-1 and highly accurate replication of TRMs at the

telomere probably support the presence of an active telomerase in these species/isolates (Fig. 1G; Supplemental Figs. S6, S11). TRM would have converted to TTAGAC via mutation in the telomerase RNA component gene, leading to TTAGAC telomeres, although it remains unclear whether this change involved a single mutation in the gene or coexistence of the original and mutated genes. This transition may have occurred long ago, because TTAGAC- and TTAGGC-telomere isolates had only TTAGAC- and TTAGGC-ITSS, respectively. Moreover, both telomeres and ITSS consisted of the same TRM, and sequences of their ITSS were not much degenerated, indicating that these ITSS resulted from recent, and probably frequent, telomere damage events in these isolates.

It is unclear whether TTAGGC-suited TBP's could bind to the novel telomeric sequence; however, if this were possible, the TTAGAC cluster–TBP complex would be partially stabilized, preventing the new telomere from being recognized as a DNA damage site. This hypothesis can be evaluated directly by inducing a mutation in the telomerase RNA component gene and then measuring the robustness of TBP in terms of sequence specificity. Unfortunately, we could not identify the telomerase RNA component gene for these isolates, as the gene has not been identified even in *C. elegans*. Further studies are required to support our hypothesis. Nonetheless, we identified TTAGGC- and/or TTAGGT-containing unit clusters in putative subtelomeric regions of *C. utelicia*, which suggests that TRM-containing unit clusters are associated with telomere maintenance and that their usage in telomere maintenance contributes to TRM evolution (see also Supplemental Note).

Our results describe the independent evolution of three novel TRMs in Nematoda. We found that an ALT-mediated mechanism may have been used to repair ancestral telomere damage in Panagrolaimidae nematodes. Further, we hypothesized that the use of TTAGAC-containing units and the robustness of TBP's facilitated the evolution of the novel TRM. As the robustness of the TBP's may depend on the same-sized bases between canonical and novel TRMs, the evolutionary path we propose for Panagrolaimidae can also be applied to plants, insects, or other nematodes in which mutations occur only between pyrimidine nucleotides or between purine nucleotides. Although our study cannot explain all TRM changes in Nematoda, it provides new insight into telomere evolution through ALT and the robustness of TBP's.

## Methods

### Worm sampling and culture

Nematodes were collected from rotten fruits in South Korea (for more information, see Supplemental Table S5). They were grown in plates containing nematode growth medium seeded with *Escherichia coli* strain OP50 and contaminated by their natural microbes.

### DNA/RNA extraction and sequencing

Mixed-stage worms were lysed, and genomic DNA was extracted using the Genra Puregene cell and tissue kit (Qiagen) to obtain WGS data for Panagrolaimidae isolates LJ2281, LJ2284, LJ2285, LJ2400, LJ2402, and LJ2406. MacroGen (<https://www.macrogen.com/en/main>) prepared DNA sequencing libraries using TruSeq Nano DNA and performed 151-bp paired-end DNA sequencing on the Illumina NovaSeq 6000 platform. One adult female worm was lysed, and its genomic DNA was amplified in the same 0.2-mL PCR tube using the REPLI-g single-cell WGA kit (Qiagen) to

obtain single-worm WGS data from LJ2050, LJ2051, LJ2060, LJ2070, LJ2072, LJ2401, LJ2411, and LJ2417. Paired-end DNA sequencing was performed using the Illumina NovaSeq 6000 platform by TheraGen Bio (<https://www.theragenbio.com/en/>).

Long-read DNA sequencing and short-read RNA sequencing of LJ2284, LJ2285, LJ2400, and LJ2406 were conducted as previously described (Kim et al. 2019a; Lee et al. 2022a). In summary, genomic DNA was extracted from mixed-stage worms using phenol/chloroform/isoamyl alcohol (25:24:1) and sequenced using the HiFi mode of the PacBio Sequel II platform by MacroGen. RNA was extracted from mixed-stage worms using the TRIzol method. RNA sequencing libraries were prepared using TruSeq nano DNA and sequenced by MacroGen using the Illumina NovaSeq 6000 platform with paired-end reads.

### Preparing publicly available WGS data

We included Nematoda species in our study, whose genome assembly data were available in the NCBI GenBank database (<https://www.ncbi.nlm.nih.gov/genbank/>) and short-read sequencing data were available in the NCBI Sequence Read Archive (SRA; <https://www.ncbi.nlm.nih.gov/sra/>; RRID:SCR\_004891). The following criteria were used to filter the short-read sequencing data: instrument, *Illumina* (if HiSeq data were available, we used HiSeq; otherwise, MiSeq, NextSeq, and Genome Analyzer II data were selected); source, *genomic*; layout, *PAIRED*; number of bases, >5 Gb; spot number, >5 million; and read length, >70 bp. Nematode clades were classified according to the phylogenetic tree (Smythe et al. 2019). Any genus that was not included in the phylogenetic tree was excluded. WGS data for *E. brevis*, which lacked genome assembly information in the GenBank database, and WGS data for *P. redivivus* (Srinivasan et al. 2013) that were supplied by Dr. P. W. Sternberg (now it can be accessible via NCBI SRA under accession number SRR25684730) were added to the filtered data sets. Detailed accession information is available in Supplemental Tables S1 and S2.

### Identifying TRMs using short-read WGS data

To normalize the public WGS data, we trimmed all reads to 60 bp and used only 5 million subsampled reads from the R1 files using Seqtk (version 1.3-r106; *seqtk sample -s 11 - 5000000*; <https://github.com/lh3/seqtk>). For each data set, 23-mers were counted using Kounta (version 0.2.3; *kounta -kmer 23 --out*; <https://github.com/tseemann/kounta>). Only tandemly repeated sequences with unit lengths of 5–7 bp and counts greater than 99 were used to identify putative TRMs that were most frequent and most similar to the canonical TTAGGC sequence. Methods used to further validate the novel putative TRMs are described in the Supplemental Methods. We averaged the counts of the 23-mers that contained the corresponding TRM concatemers to compare the number of TRMs across species. For the Panagrolaimidae isolates, we used 20 million subsampled reads and all 23-mers, as telomeric repeats were not evenly amplified during whole-genome amplification.

### Genome assembly

HiFi reads were de novo assembled using Hifiasm (version 0.13-r308; *hifiasm -l0*) (Cheng et al. 2021). Contigs that were potentially contaminated with bacterial sequences were removed as previously described (Kim et al. 2019a). Scaffolding and visualization based on the Hi-C data are described in the Supplemental Methods. The completeness of the genome assembly was evaluated using BUSCO (version 4.0.6; *busco -m genome -l nematoda\_odb10*) (Simão et al. 2015) with the nematoda data set in the OrthoDB release 10. Isolate-specific repetitive sequences were identified using BuildDatabase (version 2.0.1; default options) (Flynn

et al. 2020) and RepeatModeler (version 2.0.1; *RepeatModeler -database -LTRStruct*) (Flynn et al. 2020). Isolate-specific and known metazoan-repetitive sequences were masked using RepeatMasker (version 4.1.0; *RepeatMasker -lib -s* for species-specific repeats and *RepeatMasker -species metazoa -s* for metazoan repeats) (Chen 2004). We mapped the RNA-seq reads to the repeat-masked genome using HISAT2 (version 2.2.1; *hisat2-build* for genome indexing and *hisat2* with default options to map the RNA-seq reads to the corresponding genome) (Kim et al. 2019b). RNA-mapping data were used to annotate genes using BRAKER (version 2.1.5; *braker.pl --genome -bam --softmasking*) (Stanke et al. 2006, 2008; Li et al. 2009; Barnett et al. 2011; Lomsadze et al. 2014; Buchfink et al. 2015; Hoff et al. 2016, 2019; Brůna et al. 2021).

### Preparation of 18S rDNA sequences to generate the Panagrolaimidae phylogenetic tree

For *Panagrolaimus* sp. PS1159, *Panagrolaimus davidi*, *P. redivivus*, and *B. xylophilus*, we used publicly available 18S rDNA sequences (GenBank accessions U81579.1, AJ567385.1, AF083007.1, and KJ636306.1, respectively). For the HiFi-based de novo genome assemblies, we extracted 18S rDNA sequences by searching known nematode 18S rDNA PCR primers (nSSU\_F\_04: 5'-GCTTGCTCAAA GATTAAGCC-3' and nSSU\_R\_82: 5'-TGATCCTTCTGCAGGTTT ACCTAC-3') (Medlin et al. 1988; Blaxter et al. 1998) in genome assemblies using BLAST+ (version 2.7.1; *makeblastdb -input\_type fasta -dbtype nucl* and *blastn -task blastn-short -outfmt 6*) (Camacho et al. 2009).

For the other TTAGAC-TRM isolates used in this study, we mapped their short-read DNA sequencing data to the LJ2285 genome assembly using BWA-MEM (version 0.7.17; *bwa mem*, default option), and the 18S rDNA sequence variation of each isolate were identified using BCFtools (version 1.13; *bcftools mpileup -Ou -f | bcftools call -Ou -mv | bcftools norm -f -Oz -o*) (Li 2011a). The indexed output VCF files were obtained using Tabix (version 1.13; default option) (Li 2011b), and the LJ2285 18S rDNA sequence was replaced with their variants of each isolate using SAMtools (version 1.13) and BCFtools (*samtools faidx -r | bcftools consensus -o*) (Li et al. 2009). For the other TTAGGC-TRM isolates, we used the LJ2400 genome assembly as a reference and repeated the procedure described above. All 18S rDNA sequences of Panagrolaimidae obtained in this study are listed in Supplemental Table S6. We generated an alignment file using all 18S rDNA sequences as input for Clustal Omega (sequence type: DNA; output alignment format: PHYLIP; RRID:SCR\_001591) (Sievers and Higgins 2021; <https://www.ebi.ac.uk/Tools/msa/clustalo/>). We used this alignment as an input to RAxML (version 8.2.12) (Stamatakis 2014) using raxmlGUI 2.0 (version 2.0.10; options: GTRGAMMA and ML+ rapid bootstrap with 1000 replications) (Edler et al. 2021) to infer the phylogenetic relationships, which were visualized using Dendroscope (version 3.8.4) (Huson and Scornavacca 2012).

### Telomere and subtelomere structure in the genome assemblies

We selected candidate telomere-containing contigs using two tandemly repeated copies of TTAGRC that appeared three or more times in the 600-bp region at each end of the contig. We manually validated these telomeric regions using the following criteria: TRM cluster length  $\geq 500$  bp (the longest ITS length in *C. elegans*) (Yoshimura et al. 2019) or TRM cluster still located at the end after scaffolding (only for LJ2284 and LJ2406). We analyzed whether each telomere-containing contig/scaffold had clusters with six or more copies of TRMs or whether TTAGRC-containing units repeated tandemly in its subtelomeric region (up to 200 kb from the end of the contig/scaffold). For LJ2285, we considered sequences

consisting of five to nine copies of different units that contained TTAGRC and similar sequences as unit clusters because LJ2285 did not contain clusters composed of the same units. Subsequently, we investigated whether the unit cluster (containing six or more copies of units, the same TTAGRC, and identity  $\geq 85\%$  of units) existed outside the subtelomeric region of telomere-containing contigs/scaffolds using BLAST+ (version 2.12.0; *makeblastdb -input\_type fasta -dbtype nucl; blastn -task blastn-short -outfmt 6* for sequences  $< 30$  bp and *blastn -task megablast -outfmt 6* for sequences  $\geq 30$  bp) (Camacho et al. 2009). Unit sequences of LJ2284, LJ2400, and LJ2406 and unit cluster sequences in LJ2285 were denoted starting with TTAGRC (Supplemental Table S12). The BIR traces were analyzed by searching for homologous sequences between ITSs, TTAGRC-containing unit clusters, or telomeric repeats against their corresponding genome assembly using BLAST+ (version 2.7.1; *makeblastdb -input\_type fasta -dbtype nucl and blastn -task megablast -outfmt 6*) (Camacho et al. 2009).

### Conservation analysis of telomere-associated proteins

Protein FASTA sequences were downloaded from WormBase for *C. elegans* (release WS281) and WormBase ParaSite (Howe et al. 2017) for *Panagrolaimus* sp. PS1159, *P. davidi*, *P. redivivus*, and *B. xylophilus* (release WBPS15). Protein sequences for *Panagrolaimus* sp. PS1579 and *Panagrolaimus* sp. ES5 were not publicly available. For the four *Panagrolaimidae* isolates (LJ2284, LJ2285, LJ2400, and LJ2406), we used protein FASTA sequences obtained through gene annotation using BRAKER. For the non-*C. elegans* nematodes, we searched for protein sequences that were conserved in *C. elegans* using DIAMOND (version 2.0.11; *diamond blastp -d -q -o --threads 20 --very-sensitive and diamond blastp -d -q -o --threads 20 --ultra-sensitive*) (Buchfink et al. 2021) to characterize telomere-associated protein sequences that included TRT-1, POT-1, POT-2, POT-3, MRT-1, MRT-2, TEBP-1 (also known as DTN-1), TEBP-2 (also known as DTN-2), HPR-9, HPR-17, HUS-1, SUN-1, CEH-37, HMG-5, HSPA-1, and PLP-1. We filtered the nematode's highest bit score protein sequence for each *C. elegans* telomere-associated protein and identified the conserved domains using NCBI CD-search (CDD database, version 3.19) (Marchler-Bauer and Bryant 2004; Lu et al. 2020).

### Data access

Raw sequencing data and genome assemblies generated in this study have been submitted to the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject>) under accession number PRJNA845886 and to the Korean Nucleotide Archive (KoNA; <https://kobic.re.kr/kona>) under the BioProject accession number KAP220348 (sequencing data only).

### Competing interest statement

The authors declare no competing interests.

### Acknowledgments

This study was supported by the Samsung Science and Technology Foundation (SSTF-BA1501-52) and the National Research Foundation of Korea (2019R1A6A1A10073437 to J.K.). J. Lim was also supported by a scholarship for basic research, Seoul National University. We thank Dr. P.W. Sternberg for providing and publishing valuable sequencing data of *P. redivivus* and thank the Darwin Tree of Life Project Consortium for releasing the sequencing data of *C. uteleia*. We also thank Namhee Kim and Sungyeol Ahn for providing access to their farms to collect rotten fruits and Dr.

D.S. Lim for collaboration in collection of nematode species. We appreciate Dr. S. Sung for providing the suggestions on generating the Hi-C library and Dr. C. Kim for the help in Hi-C data processing.

**Author contributions:** J.L. did conceptualization, methodology, formal analysis, investigation, writing of the original draft, and reviewing and editing. W.K. did investigation. J.K. did conceptualization, methodology, investigation, writing of the original draft, and reviewing and editing. J.L. did conceptualization, reviewing and editing, funding acquisition, and supervision.

### References

- Abad JP, De Pablos B, Osoegawa K, De Jong PJ, Martín-Gallardo A, Villasante A. 2004. TAHRE, a novel telomeric retrotransposon from *Drosophila melanogaster*, reveals the origin of *Drosophila* telomeres. *Mol Biol Evol* **21**: 1620–1624. doi:10.1093/molbev/msh180
- Ahmed S, Hodgkin J. 2000. MRT-2 checkpoint protein is required for germline immortality and telomere replication in *C. elegans*. *Nature* **403**: 159–164. doi:10.1038/35003120
- Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27**: 1691–1692. doi:10.1093/bioinformatics/btr174
- Blaxter ML, De Ley P, Garey JR, Liu LX, Scheldeman P, Vierstraete A, Vanfleteren JR, Mackey LY, Dorris M, Frisse LM, et al. 1998. A molecular evolutionary framework for the phylum Nematoda. *Nature* **392**: 71–75. doi:10.1038/32160
- Boerckel J, Walker D, Ahmed S. 2007. The *Caenorhabditis elegans* Rad17 homolog HPR-17 is required for telomere replication. *Genetics* **176**: 703–709. doi:10.1534/genetics.106.070201
- Bosco G, Haber JE. 1998. Chromosome break-induced DNA replication leads to nonreciprocal translocations and telomere capture. *Genetics* **150**: 1037–1047. doi:10.1093/genetics/150.3.1037
- Brúna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform* **3**: lqaa108. doi:10.1093/nargab/lqaa108
- Bryan TM, Englezou A, Gupta J, Bacchetti S, Reddel R. 1995. Telomere elongation in immortal human cells without detectable telomerase activity. *EMBO J* **14**: 4240–4248. doi:10.1002/j.1460-2075.1995.tb00098.x
- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**: 59–60. doi:10.1038/nmeth.3176
- Buchfink B, Reuter K, Drost H-G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* **18**: 366–368. doi:10.1038/s41592-021-01101-x
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421. doi:10.1186/1471-2105-10-421
- The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**: 2012–2018. doi:10.1126/science.282.5396.2012
- Červenák F, Juríková K, Devillers H, Kaffe B, Khatib A, Bonnell E, Sopkovičová M, Wellinger RJ, Nosek J, Tzfatí Y, 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**: 13365. doi:10.1038/s41598-019-49628-6
- Červenák F, Sepšiová R, Nosek J, Tomáška L. 2021. Step-by-step evolution of telomeres: lessons from yeasts. *Genome Biol Evol* **13**: evaa268. doi:10.1093/gbe/evaa268
- Chen N. 2004. Using repeat masker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* **5**: 4.10.11–14.10.14. doi:10.1002/0471250953.bi0410s05
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**: 170–175. doi:10.1038/s41592-020-01056-5
- The Darwin Tree of Life Project Consortium. 2022. Sequence locally, think globally: the Darwin tree of life project. *Proc Natl Acad Sci* **119**: e2115642118. doi:10.1073/pnas.2115642118
- Dayi M, Sun S, Maeda Y, Tanaka R, Yoshida A, Tsai JJ, Kikuchi T. 2020. Nearly complete genome assembly of the pinewood nematode *Bursaphelenchus xylophilus* strain Ka4C1. *Microbiol Resour Announc* **9**: e01002–e01020. doi:10.1128/MRA.01002-20
- De Lange T. 2009. How telomeres solve the end-protection problem. *Science* **326**: 948–952. doi:10.1126/science.1170633
- Dietz S, Almeida MV, Nischwitz E, Schreier J, Viceconte N, Fradera-Sola A, Renz C, Ceron-Noriega A, Ulrich HD, Kappel D, et al. 2021. The

- double-stranded DNA-binding proteins TEBP-1 and TEBP-2 form a telomeric complex with POT-1. *Nat Commun* **12**: 2668. doi:10.1038/s41467-021-22861-2
- Eidler D, Klein J, Antonelli A, Silvestro D. 2021. raxmlGUI 2.0: a graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods Ecol Evol* **12**: 373–377. doi:10.1111/2041-210X.13512
- Fajkus J, Sýkorová E, Leitch AR. 2005. Telomeres in evolution and evolution of telomeres. *Chromosome Res* **13**: 469–479. doi:10.1007/s10577-005-0997-2
- Ferreira HC, Towbin BD, Jegou T, Gasser SM. 2013. The shelterin protein POT-1 anchors *Caenorhabditis elegans* telomeres through SUN-1 at the nuclear periphery. *J Cell Biol* **203**: 727–735. doi:10.1083/jcb.201307181
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci* **117**: 9451–9457. doi:10.1073/pnas.1921046117
- Fulnečková J, Hasíková T, Fajkus J, Lukešová A, Eliáš M, Sýkorová E. 2012. Dynamic evolution of telomeric sequences in the green algal order Chlamydomonadales. *Genome Biol Evol* **4**: 248–264. doi:10.1093/gbe/evs007
- Fulnečková J, Ševčíková T, Fajkus J, Lukešová A, Lukeš M, Vlček Č, Lang BF, Kim E, Eliáš M, Sýkorová E. 2013. A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biol Evol* **5**: 468–483. doi:10.1093/gbe/evt019
- Garavís M, González C, Villasante A. 2013. On the origin of the eukaryotic chromosome: the role of noncanonical DNA structures in telomere evolution. *Genome Biol Evol* **5**: 1142–1150. doi:10.1093/gbe/evt079
- Heaphy CM, Subhawong AP, Hong S-M, Goggins MG, Montgomery EA, Gabrielson E, Netto GJ, Epstein JI, Lotan TL, Westra WH, et al. 2011. Prevalence of the alternative lengthening of telomeres telomere maintenance mechanism in human cancer subtypes. *Am J Pathol* **179**: 1608–1615. doi:10.1016/j.ajpath.2011.06.018
- Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**: 767–769. doi:10.1093/bioinformatics/btv661
- Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. In *Gene prediction: methods and protocols* (ed. Kollmar M), pp. 65–95. Humana, New York.
- Hofmann ER, Milstein S, Boulton SJ, Ye M, Hofmann JJ, Stergiou L, Gartner A, Vidal M, Hengartner MO. 2002. *Caenorhabditis elegans* HUS-1 is a DNA damage checkpoint protein required for genome stability and EGL-1-mediated apoptosis. *Curr Biol* **12**: 1908–1918. doi:10.1016/S0960-9822(02)01262-9
- Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. 2017. Wormbase ParaSite: a comprehensive resource for helminth genomics. *Mol Biochem Parasitol* **215**: 2–10. doi:10.1016/j.molbiopara.2016.11.005
- Huson DH, Scornavacca C. 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol* **61**: 1061–1067. doi:10.1093/sysbio/sys062
- Im SH, Lee J. 2003. Identification of HMG-5 as a double-stranded telomeric DNA-binding protein in the nematode *Caenorhabditis elegans*. *FEBS Lett* **554**: 455–461. doi:10.1016/S0014-5793(03)01191-8
- Im SH, Lee J. 2005. PLP-1 binds nematode double-stranded telomeric DNA. *Mol Cell* **20**: 297–302.
- Joeng KS, Song EJ, Lee K-J, Lee J. 2004. Long lifespan in worms with long telomeric DNA. *Nat Genet* **36**: 607–611. doi:10.1038/ng1356
- Kim SH, Hwang SB, Chung IK, Lee J. 2003. Sequence-specific binding to telomeric DNA by CEH-37, a homeodomain protein in the nematode *Caenorhabditis elegans*. *J Biol Chem* **278**: 28038–28044. doi:10.1074/jbc.M302192200
- Kim C, Kim J, Kim S, Cook DE, Evans KS, Andersen EC, Lee J. 2019a. Long-read sequencing reveals intra-species tolerance of substantial structural variations and new subtelomere formation in *C. elegans*. *Genome Res* **29**: 1023–1035. doi:10.1101/gr.246082.118
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019b. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**: 907–915. doi:10.1038/s41587-019-0201-4
- Kim C, Sung S, Kim J, Lee J. 2020. Repair and reconstruction of telomeric and subtelomeric regions and genesis of new telomeres: implications for chromosome evolution. *Bioessays* **42**: 1900177. doi:10.1002/bies.201900177
- Kim C, Sung S, Kim J-S, Lee H, Jung Y, Shin S, Kim E, Seo JJ, Kim J, Kim D, et al. 2021a. Telomeres reforged with non-telomeric sequences in mouse embryonic stem cells. *Nat Commun* **12**: 1097. doi:10.1038/s41467-020-20314-w
- Kim E, Kim J, Kim C, Lee J. 2021b. Long-read sequencing and de novo genome assemblies reveal complex chromosome end structures caused by telomere dysfunction at the single nucleotide level. *Nucleic Acids Res* **49**: 3338–3353. doi:10.1093/nar/gkab141
- Kramara J, Willcox S, Gunisova S, Kinsky S, Nosek J, Griffith JD, Tomaska L. 2010. Tay1 protein, a novel telomere binding factor from *Yarrowia lipolytica*. *J Biol Chem* **285**: 38078–38092. doi:10.1074/jbc.M110.127605
- Kramara J, Osia B, Malkova A. 2018. Break-induced replication: the where, the why, and the how. *Trends Genet* **34**: 518–531. doi:10.1016/j.tig.2018.04.002
- Lazzerini-Denchi E, Sfeir A. 2016. Stop pulling my strings: What telomeres taught us about the DNA damage response. *Nat Rev Mol Cell Biol* **17**: 364–378. doi:10.1038/nrm.2016.43
- Lee BY, Kim J, Lee J. 2022a. Intraspecific *de novo* gene birth revealed by presence-absence variant genes in *Caenorhabditis elegans*. *NAR Genom Bioinform* **4**: lqac031. doi:10.1093/nargab/lqac031
- Lee BY, Kim J, Lee J. 2022b. Long-read sequencing infers a mechanism for copy number variation of template for alternative lengthening of telomeres in a wild *C. elegans* strain. *MicroPubl Biol* **2022**: 10.17912/micropub.biology.000563. doi:10.17912/micropub.biology.000563
- Li H. 2011a. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–2993. doi:10.1093/bioinformatics/btr509
- Li H. 2011b. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* **27**: 718–719. doi:10.1093/bioinformatics/btq671
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079. doi:10.1093/bioinformatics/btp352
- Lomsadze A, Burns PD, Borodovsky M. 2014. Integration of mapped RNA-seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res* **42**: e119. doi:10.1093/nar/gku557
- Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, et al. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* **48**: D265–D268. doi:10.1093/nar/gkz991
- Lundblad V, Blackburn EH. 1993. An alternative pathway for yeast telomere maintenance rescues est1– senescence. *Cell* **73**: 347–360. doi:10.1016/0092-8674(93)90234-H
- Lydeard JR, Jain S, Yamaguchi M, Haber JE. 2007. Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature* **448**: 820–823. doi:10.1038/nature06047
- Malkova A, Ivanov EL, Haber JE. 1996. Double-strand break repair in the absence of RAD51 in yeast: a possible role for break-induced DNA replication. *Proc Natl Acad Sci* **93**: 7131–7136. doi:10.1073/pnas.93.14.7131
- Marchler-Bauer A, Bryant SH. 2004. CD-search: protein domain annotations on the fly. *Nucleic Acids Res* **32**: W327–W331. doi:10.1093/nar/gkh454
- Medlin L, Elwood HJ, Stickel S, Sogin ML. 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* **71**: 491–499. doi:10.1016/0378-1119(88)90066-2
- Meier B, Clejan I, Liu Y, Lowden M, Gartner A, Hodgkin J, Ahmed S. 2006. trt-1 is the *Caenorhabditis elegans* catalytic subunit of telomerase. *PLoS Genet* **2**: e18. doi:10.1371/journal.pgen.0020018
- Meier B, Barber LJ, Liu Y, Shtessel L, Boulton SJ, Gartner A, Ahmed S. 2009. The MRT-1 nuclease is required for DNA crosslink repair and telomerase activity in vivo in *Caenorhabditis elegans*. *EMBO J* **28**: 3549–3563. doi:10.1038/emboj.2009.278
- Mravinac B, Meštrović N, Čavrak VV, Plohl M. 2011. TCAGG, an alternative telomeric sequence in insects. *Chromosoma* **120**: 367–376. doi:10.1007/s00412-011-0317-x
- Peska V, Garcia S. 2020. Origin, diversity, and evolution of telomere sequences in plants. *Front Plant Sci* **11**: 117. doi:10.3389/fpls.2020.00117
- 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 (TTTTTAGGG)<sub>n</sub> in the plant *Cestrum elegans* (Solanaceae), a species with a large genome. *Plant J* **82**: 644–654. doi:10.1111/tpj.12839
- Raices M, Verdun RE, Compton SA, Haggblom CI, Griffith JD, Dillin A, Karlseder J. 2008. *C. elegans* telomeres contain G-strand and C-strand overhangs that are bound by distinct proteins. *Cell* **132**: 745–757. doi:10.1016/j.cell.2007.12.039
- Richards EJ, Ausubel FM. 1988. Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell* **53**: 127–136. doi:10.1016/0092-8674(88)90494-1
- Rotková G, Skleničková M, Dvořáčková M, Sýkorová E, Leitch AR, Fajkus J. 2004. An evolutionary change in telomere sequence motif within the plant section Asparagales had significance for telomere nucleoprotein complexes. *Cytogenet Genome Res* **107**: 132–138. doi:10.1159/000079584
- Roumelioti FM, Sotiropoulos SK, Katsini V, Chiourea M, Halazonetis TD, Gagos S. 2016. Alternative lengthening of human telomeres is a conservative

- DNA replication process with features of break-induced replication. *EMBO Rep* **17**: 1731–1737. doi:10.15252/embr.201643169
- Seo B, Kim C, Hills M, Sung S, Kim H, Kim E, Lim DS, Oh H-S, Choi RMJ, Chun J, et al. 2015. Telomere maintenance through recruitment of internal genomic regions. *Nat Commun* **6**: 8189. doi:10.1038/ncomms9189
- Sepsiova R, Necasova I, Willcox S, Prochazkova K, Gorilak P, Nosek J, Hofr C, Griffith JD, Tomaska L. 2016. Evolution of telomeres in *Schizosaccharomyces pombe* and its possible relationship to the diversification of telomere binding proteins. *PLoS One* **11**: e0154225. doi:10.1371/journal.pone.0154225
- Shakirov EV, Song X, Joseph JA, Shippen DE. 2009. POT1 proteins in green algae and land plants: DNA-binding properties and evidence of co-evolution with telomeric DNA. *Nucleic Acids Res* **37**: 7455–7467. doi:10.1093/nar/gkp785
- Shtessel L, Lowden MR, Cheng C, Simon M, Wang K, Ahmed S. 2013. *Caenorhabditis elegans* POT-1 and POT-2 repress telomere maintenance pathways. *G3 Genes Genomes Genetics* **3**: 305–313. doi:10.1534/g3.112.004440
- Sievers F, Higgins DG. 2021. The clustal omega multiple alignment package. In *Multiple sequence alignment: methods and protocols* (ed. Katoh K), pp. 3–16. Humana, New York.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212. doi:10.1093/bioinformatics/btv351
- Smythe AB, Holovachov O, Kocot KM. 2019. Improved phylogenomic sampling of free-living nematodes enhances resolution of higher-level nematode phylogeny. *BMC Evol Biol* **19**: 121. doi:10.1186/s12862-019-1444-x
- Sobinoff AP, Pickett HA. 2017. Alternative lengthening of telomeres: DNA repair pathways converge. *Trends Genet* **33**: 921–932. doi:10.1016/j.tig.2017.09.003
- Srinivasan J, Dillman AR, Macchietto MG, Heikkinen L, Lakso M, Fracchia KM, Antoshechkin I, Mortazavi A, Wong G, Sternberg PW. 2013. The draft genome and transcriptome of *Panagrellus redivivus* are shaped by the harsh demands of a free-living lifestyle. *Genetics* **193**: 1279–1295. doi:10.1534/genetics.112.148809
- Stamatakis A. 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313. doi:10.1093/bioinformatics/btu033
- Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**: 62. doi:10.1186/1471-2105-7-62
- Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntetically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* **24**: 637–644. doi:10.1093/bioinformatics/btn013
- Steinberg-Neifach O, Lue NF. 2015. Telomere DNA recognition in *Saccharomycotina* yeast: potential lessons for the co-evolution of ssDNA and dsDNA-binding proteins and their target sites. *Front Genet* **6**: 162. doi:10.3389/fgene.2015.00162
- Sýkorová E, Lim K, Kunická Z, Chase M, Bennett M, Fajkus J, Leitch A. 2003. Telomere variability in the monocotyledonous plant order Asparagales. *Proc Biol Sci* **270**: 1893–1904. doi:10.1098/rspb.2003.2446
- Sýkorová E, Fajkus J, Mezníková M, Lim KY, Neplechová K, Blattner FR, Chase MW, Leitch AR. 2006. Minisatellite telomeres occur in the family Alliaceae but are lost in *Allium*. *Am J Bot* **93**: 814–823. doi:10.3732/ajb.93.6.814
- Tomáška L, Nosek J, Sepsiova R, Červenák F, Juríková K, Procházková K, Neboháčová M, Willcox S, Griffith JD. 2018. Commentary: single-stranded telomere-binding protein employs a dual rheostat for binding affinity and specificity that drives function. *Front Genet* **9**: 742. doi:10.3389/fgene.2018.00742
- Tran TD, Cao HX, Jovtchev G, Neumann P, Novák P, Fojtová M, Vu GT, Macas J, Fajkus J, Schubert I, et al. 2015. Centromere and telomere sequence alterations reflect the rapid genome evolution within the carnivorous plant genus *Genlisea*. *Plant J* **84**: 1087–1099. doi:10.1111/tpj.13058
- Van Steensel B, Smogorzewska A, De Lange T. 1998. TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**: 401–413. doi:10.1016/S0092-8674(00)80932-0
- Visacka K, Hofr C, Necasova I, Pavlouskova J, Sepsiova R, Wimmerova M, Simonicova L, Nosek J, Fajkus J, et al. 2012. Synergism of the two Myb domains of Tay1 protein results in high affinity binding to telomeres. *J Biol Chem* **287**: 32206–32215. doi:10.1074/jbc.M112.385591
- Weiss H, Scherthan H. 2002. Aloe spp.: plants with vertebrate-like telomeric sequences. *Chromosome Res* **10**: 155–164. doi:10.1023/A:1014905319557
- Wong B. 2011. Color blindness. *Nat Methods* **8**: 441. doi:10.1038/nmeth.1618
- Yamamoto I, Zhang K, Zhang J, Vorontsov E, Shibuya H. 2021. Telomeric double-strand DNA-binding proteins DTN-1 and DTN-2 ensure germline immortality in *Caenorhabditis elegans*. *eLife* **10**: e64104. doi:10.7554/eLife.64104
- Yoshimura J, Ichikawa K, Shoura MJ, Artiles KL, Gabdank I, Wahba L, Smith CL, Edgley ML, Rougvie AE, Fire AZ, et al. 2019. ReCompleting the *Caenorhabditis elegans* genome. *Genome Res* **29**: 1009–1022. doi:10.1101/gr.244830.118
- Zhong Z, Shiue L, Kaplan S, de Lange T. 1992. A mammalian factor that binds telomeric TTAGGG repeats in vitro. *Mol Cell Biol* **12**: 4834–4843. doi:10.1128/mcb.12.11.4834
- Zhou Y, Wang Y, Xiong X, Appel AG, Zhang C, Wang X. 2022. Profiles of telomeric repeats in Insecta reveal diverse forms of telomeric motifs in Hymenoptera. *Life Sci Alliance* **5**: e202101163. doi:10.26508/lsa.202101163

Received May 24, 2023; accepted in revised form October 16, 2023.