Supplemental Material for

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

Jiseon Lim, Wonjoo Kim, Jun Kim, and Junho Lee

Junho Lee and Jun Kim

Email: elegans@snu.ac.kr and junkim@cnu.ac.kr

Table of Contents

Supplemental Note

Identification of sample contamination in *Enoplus brevis*

E. brevis was the only species in Clade II with short-read sequencing data that met the TRM analysis criteria (except for the lack of a published genome assembly). Nonetheless, we identified a sequence difference of 1.2% (22 single-nucleotide variants among the 1797-bp sequence), indicating that *E. brevis* data would be contaminated, by mapping its data to the public 18S rDNA sequence (GenBank: U88336.1) and calling variants. We did not find any uncontaminated sequencing data for this species, so we included the results of the analysis of this species in the TRM-undetermined species group (Supplemental Table S2).

Validation of the putative novel TRM in the vertebrate parasites *Strongyloides* and *Trichinella*

Strongyloides and *Trichinella* species, in which TTAGGG or TTACGG were discovered as their putative TRMs, are vertebrate parasites (Abadie 1963; Dick and Belosevic 1978; Nwaorgu and Connan 1980; Pozio et al. 1989; Yamaguchi 1991; Kapel 2000; Mitreva and Jasmer 2006). Thus, we tested whether these putative TRMs originated from their host contamination or from their nematode telomeres with the following procedure. (1) For each species, TRM analysis was performed using additional short-read sequencing data that were produced using different samples, and then we checked whether these independent datasets of the same species produced consistent TRM discoveries. (2) If TRM results were consistent, we checked whether any of the data was prepared by host contamination-free methods, such as nematode tissue separation before sequencing. For species that satisfied this criterion, we concluded that the

species probably utilized the novel TRM. (3) If their TRM results were consistent, but the tissue source information did not confirm whether the sequencing data were host contamination-free, we extracted the putative telomeric and subtelomeric reads that contained the novel TRM but were not mapped to the host genome.

In *S. ratti*, TTAGGG was dominant in one of two additional sequencing datasets, similar to our previous TRM analysis, and the dataset was obtained by a possible host contamination-free method, sperm extraction before sequencing (Supplemental Table S4) (Kulkarni et al. 2016). Therefore, the TRM change to TTAGGG in *S. ratti* is probably real, which could be confirmed in the future by analyzing a high-quality reference genome based on long-read sequencing data. In *S. papillosus* and *S. stercoralis*, additional analysis results were inconsistent with their putative TRMs; thus, we could not determine the TRM or any other complex repeats constituting the telomere in these species.

In *T. nativa*, we found that the putative TRM, TTAGGG, was also dominant in the two additional sequencing datasets (Supplemental Table S4). We could not find high-quality sequencing data for *T. britovi* and *Trichinella* sp. T9, but we analyzed their single data, too. All data for the three species were probably not prepared by host contamination-free methods, as they were sequenced using the entire worms freshly collected from host muscles (Korhonen et al. 2016; Feng et al. 2021).

In the three *Trichinella* species, we searched for reads that were not mapped to the host genome and whose pairs started with TRM repeats. Sequencing data for *T. britovi* and *Trichinella* sp. T9 did not have unmapped reads of this type, so we could not conclude whether TTAGGG is the TRM in these species. *T. nativa* had only 8–9 unmapped reads despite the high

sequencing depth of its genome. However, the hosts used to collect *T*. *nativa* samples have missing subtelomeric regions, so these reads may come from the unknown host subtelomeric sequences (Kim et al. 2020). We could not confirm the TRM of *T*. *nativa*.

TTAGGT clusters and TTAGGT-containing unit clusters were directly attached in *Caenorhabditis uteleia*

We analyzed TTAGGT repeats in short-read sequencing data and short-read-based genome assembly, independent of the previous *k*-mer-based method. Of the 20 million 125-bp reads, 2415 reads were composed of 20 copies of TTAGGT tandem repeats. This suggested that TTAGGT could be the true TRM in this species. No scaffolds ending with TTAGGT repeats were identified probably because of the poor genome assembly quality. However, we identified TTAGGT-containing unit clusters at four ends of contigs (two ends with six copies of unit1 and two ends with ≥10 copies of unit2; 16-bp unit1 = TTAGGT TTAGGT TTAC; 20-bp unit2 = TTAGGT TTAGGT TTAACTTC), which are similar to the TRM-containing unit clusters in Panagrolaimidae. Using public long-read sequencing data of *C*. *uteleia*, we validated whether TTAGGT was the TRM, whether TTAGGT-unit clusters were located close to the TRM cluster and whether TTAGGC-unit clusters existed in its genome (SRA accession: ERR8978452) (The Darwin Tree of Life Project Consortium 2022). First, 50 of 750,000 reads ended with an average of 179 copies of TTAGGT (the minimum copy number was 28), and none of the reads ended with six copies of TTAGGC repeats, indicating that TTAGGT, rather than TTAGGC, was the TRM of *C*. *uteleia*. Second, the TRM clusters were adjacent to an average of 297 copies of unit1 (30–665 copies), and among the 50 reads, 19 reads ended with ≥1-kb TTAGGT repeats attached to ≥1-kb

unit1 clusters. This implied that the unit1 cluster was located in the subtelomeric region. Finally, we identified at least one read composed of a unit cluster containing both TTAGGT and TTAGGC, an ITS, a TTAGGT-containing unit cluster, and a putative telomeric region, sequentially (Supplemental Fig. S7). This was similar to the results obtained in Panagrolaimidae. The above data suggest that TTAGGT-containing unit clusters in subtelomeric regions are associated with the evolution of TRM in *Caenorhabditis*. In future studies, we will be able to analyze the evolution of long-read sequencing-based genome assemblies of *C. uteleia* and its relative species. Currently, long-read sequencing-based genome assembly of *C. uteleia* and any sequencing data of its relative species are not available.

TAGGG is not the TRM of *Diploscapter pachys*

D. pachys is expected to have a unichromosome, in which all chromosomes fuse, possibly in a circular form, and no TRM has been identified in previous studies investigating its genome assembly (Fradin et al. 2017; Schwarz 2017; Eweis et al. 2022). We identified TAGGG as the putative TRM in *D. pachys* using short-read sequencing data; however, the TAGGG repeats were too short, up to ten copies. Moreover, most TAGGG repeats were located in the middle of reads rather than at any ends; therefore, we concluded that TAGGG might not act as the TRM in this species.

Synteny relationships between *Bursaphelenchus xylophilus* and Panagrolaimidae isolates

To better understand synteny relationships, we selected 375 BUSCO genes that were shared among all five genome assemblies of *B. xylophilus*, and our four isolates, namely LJ2284, LJ2285, LJ2400, and LJ2406. The contigs/scaffolds of our genome assemblies containing any of the 375 genes covered 96%, 90%, 79%, and 65% of LJ2284, LJ2285, LJ2400, and LJ2406 genome assemblies, respectively. Many of the contigs/scaffolds in the LJ2284 and LJ2285 genome assemblies exhibited highly conserved synteny relationships with Chromosomes 1, 2, 3, 5, and 6 in *B. xylophilus* (Fig. 1D and Supplemental Fig. S4). We suspected that Chromosome 4 of their ancestor was integrated or fused to other chromosomes, such as Chromosome 3. In contrast, most contigs/scaffolds in the LJ2400 and LJ2406 genome assemblies exhibited more mixed patterns with the chromosomes of *B. xylophilus* but still exhibited conserved patterns with Chromosomes 5 and 6 (Supplemental Fig. S4).

Validation of the extraordinarily long telomeres in LJ2406

Telomere lengths of the LJ2406 genome assembly ranged from 11.6 kb to 46.9 kb (except for the extremely short 341 bp telomere), which are much longer than those of our other three isolates, *C. elegans* and *B. xylophilus* (Supplemental Tables S9 and S11) (Kim et al. 2019; Yoshimura et al. 2019; Dayi et al. 2020). To confirm that the longer telomere lengths were comparable to their real lengths, we analyzed raw HiFi read lengths that consisted only of telomeric repeats. Among the raw HiFi reads of LJ2406, 98 percent of telomeric reads were longer than 10 kb, and one telomeric read was 25.6-kb long (telomeric reads are defined by the following criteria: composed of >95% of TTAGGC repeats; starting and ending with two copies of TTAGGC repeats) (Supplemental Fig. S5). These long telomeric reads can explain all

<23 kb telomeres in the genome assembly of LJ2406. Due to limitations in HiFi sequencing, which produces size-selected reads of 10–20 kb on average, we were unable to find any telomeric reads close to 47 kb. However, we can assume that the genome assembler arranged and assembled telomeric reads to the 47-kb telomere by considering their internal telomeric variants.

Length difference between telomere and ITS

All ITSs are shorter than or equal to 461 bp in the *C. elegans* VC2010 genome and 156 bp in our four genome assemblies, except for a 1.7-kb ITS in LJ2406. However, all telomeres with Structure 1 subtelomeres are longer than 2 kb, except for two: telomeres of ptg000189l in LJ2400 (771 bp) and BXYJ5_Chr3 in *B. xylophilus* (1.3 kb) (Supplemental Tables S10 and S11). This suggests that there is a difference between the length distributions of ITSs and telomeres and that >2-kb telomeres are genuine. The two <2 kb telomeres are shorter than the exceptionally long 1.7-kb ITS of LJ2406; however, they are longer than typical ITSs and located at the end of contig/scaffold. Therefore, all telomeres with Structure 1 were considered to exhibit the characteristics of telomeres rather than ITSs.

Subtelomeric TTAGRC-containing unit clusters in *Caenorhabditis elegans* and *Bursaphelenchus xylophilus*

We analyzed subtelomeric regions of *C. elegans* and *B. xylophilus* to identify TRM-containing unit clusters in these species. In *C. elegans*, a cluster of a 14-mer unit (eight copies of GGCCCTAAGCCTAA and its variants) was identified ~150-kb away from the left arm telomere

of Chromosome IV (chrIV_pilon:160109-160220) (Supplemental Table S11). In contrast to the Panagrolaimidae, the *C. elegans* unit was found outside of the subtelomeric region (chrII_pilon:2613443-2613540). We also searched for units containing TTAGGT, which contain the putative TRM of *C. uteleia* in the *C. elegans* genome but could not identify any clusters. In *B. xylophilus*, a cluster of a 2250-mer unit (12 copies) was identified ~1-kb away from the left arm telomere of Chromosome I (BXYJ5_Chr1:2404-27903) (Supplemental Tables S11 and S12). This unit was not identified outside the subtelomeric region.

MMBIR signatures that may have generated the new subtelomeric sequences in ptg000247l in LJ2406

We searched homologous sequences of the duplicated block directly attached to the ITS in LJ2406 ptg000247l and found that the duplicated block could be divided into five homologous sequence blocks (gray, pink, blue, yellow, and bluish green blocks in Supplemental Fig. S10A). These five sequence blocks were consecutively overlapped by the four microhomologous sequences between each pair of homologous blocks identified in ptg000016l and ptg000849l (1–6 bp, Supplemental Fig. S10B). This consecutive replication and the remaining microhomologous sequences between homologous blocks were indicative of the MMBIR reaction. Other mechanisms, such as illegitimate recombination, may not adequately explain the consecutive DNA repair signature.

Alternative explanation for the existence of TRM-containing unit clusters

The presence of both canonical and novel TRM-containing unit clusters in the Panagrolaimidae isolates could be just a consequence, rather than a cause supporting the TRM evolution from TTAGGC to TTAGAC. Since nematodes are typically not haploid species, a mutated novel allele of the telomerase RNA component gene should be utilized together with the canonical allele. It is also possible that the gene was duplicated and the duplicated gene was mutated responsible for TTAGAC. In any case, both canonical and mutated alleles responsible for TTAGGC and TTAGAC, respectively, would have been used together in the same species. It could have facilitated the usage of both canonical and novel TRM-containing unit clusters, as these unit clusters might have supported telomere maintenance. These hypotheses should be further addressed by future biochemical investigations.

Supplemental Methods

Validation of novel candidate TRMs identified by the *k*-mer-based approach

For *Caenorhabditis uteleia*, in which TTAGGT was identified as a putative TRM, we counted the short reads that consisted of only TTAGGT repeats and analyzed short-read–based scaffolds to identify any scaffold end that contained ≥6 copies of TTAGGT repeats or TTAGGT-containing unit clusters (see Supplemental Table S1 for SRA and genome assembly accessions). Moreover, we analyzed the long reads ending with ≥20 copies of TTAGGT repeats attached to the TTAGGT-containing unit cluster (SRA accession: ERR8978452) and verified whether any of the reads had ≥1 kb of TTAGGT repeats and ≥1 kb of TTAGGT-containing cluster. For each species of *Strongyloides* and *Trichinella* genus, TRM analysis was performed using additionally obtained short-read sequencing data (data size ≥1 Gb) from different BioProject accessions (see Supplemental Table S4 for SRA accession numbers). We

also investigated whether the sequenced sample in each dataset was prepared by host contamination-free methods, such as specific tissue isolation. To obtain nematode telomeric and subtelomeric reads that did not originate from the host, we collected read pairs containing ≥6 copies of TRM repeats from each dataset and mapped the reads to the host genome using BWA-MEM (version 0.7.17; *bwa mem*, default option). Then, we extracted read pairs, of which one read started with CCCTAA (i.e. the reverse complement form of a TTAGGG telomeric read) and the other was unmapped to the host genome (a non-host subtelomeric read), using SAMtools (Li et al. 2009) (version 1.13; *samtools fasta -f 4*). We used Toplevel unmasked genomic sequences of *Rattus norvegicus* and *Mus musculus* from Ensembl (release 108) as host genome assemblies. For *Diploscapter pachys*, we determined the length of the longest TAGGG repeat in its short-read sequencing data and the number of short reads containing the repeat sequence of that length. Thereafter, we investigated whether its genome assembly has a repeat sequence of that length.

Hi-C-based scaffolding of HiFi genome assemblies

We prepared ~30,000 worms from LJ2284 and LJ2406 and generated their Hi-C libraries using the Arima-HiC+ kit and the Arima Library Prep kit following the manufacturer's protocols. Each library was sequenced on the Illumina NovaSeq 6000 platform. Then, we indexed the contig-level genome assemblies using SAMtools (Li et al. 2009) (version 1.13; *samtools faidx*) and mapped Hi-C reads to the corresponding genome assembly using the Arima-HiC Mapping Pipeline v02 script (https://github.com/ArimaGenomics/mapping_pipeline). Second, we sorted output BAM files

using SAMtools (*samtools sort -n*) and utilized the alignment information to scaffold the contigs using YaHS (Zhou et al. 2023) (version 1.2a.2; *yahs -q 30 -l 300000 --no-contig-e -e GATC,GANTC*). We manually removed scaffolded bacterial sequences. We annotated their mitochondrial genomes by analyzing similarities between specific contigs/scaffolds and the mitochondrial genomes of *Caenorhabditis elegans* (Yoshimura et al. 2019) and *Bursaphelenchus xylophilus* (GenBank accession: AY508069.1) using BLAST+ (Camacho et al. 2009) (version 2.12.0; *blastn -outfmt 6*). We linearized the circularly repeated mitochondrial genomes of LJ2284 and LJ2406 by aligning these using BLAST+ (version 2.12.0; *blastn -outfmt 6*).

Visualizing Hi-C contact maps

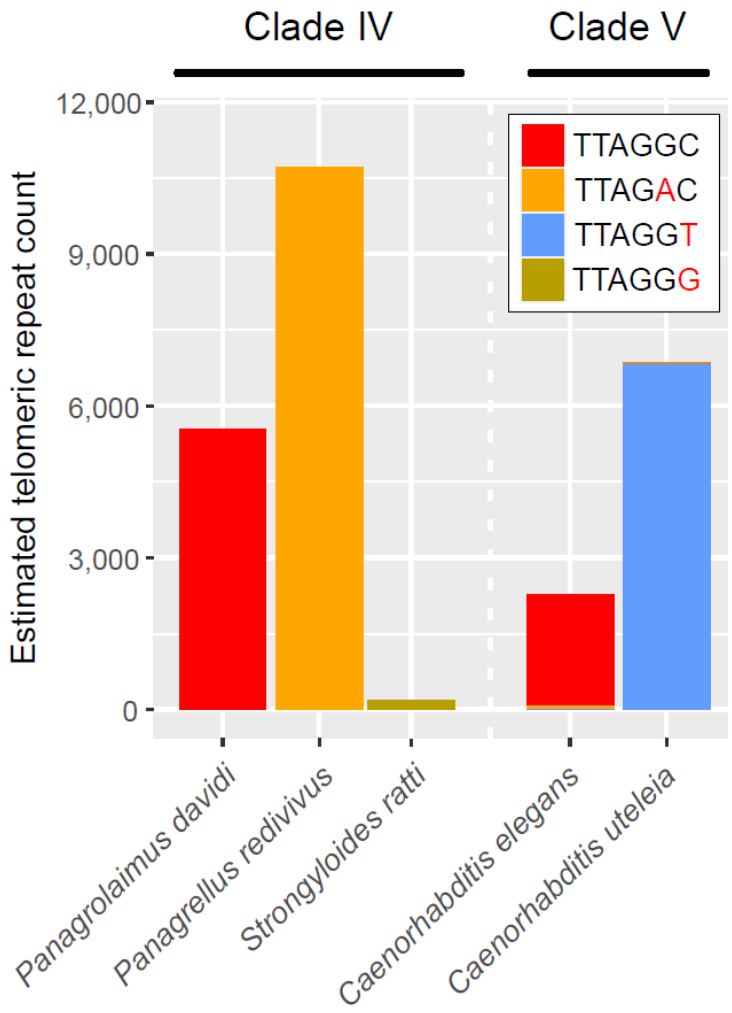
We visualized Hi-C contact maps using the Micro-C protocol to generate a pairs file and Hi-C contact map (https://github.com/dovetail-genomics/Micro-C) as follows: first, we indexed scaffold-level genome assemblies using SAMtools (*samtools faidx*) and mapped Hi-C reads to the corresponding genome using BWA (version 0.7.17; *bwa index* and *bwa mem -5SP -T0*). Next, we processed Hi-C read pairs using pairtools (https://github.com/mirnylab/pairtools) (version 0.3.0; *pairtools parse --min-mapq 40 --walks-policy 5unique --max-inter-align-gap 30 --chroms-path sacffold_size_file* to identify ligation junctions and generate a pairsam_file, *pairtools sort* for sorting pairsam_file, *pairtools dedup --mark-dups --output-stats stats_file* for removing PCR duplicates and *pairtools split --output-pairs pairs_file --output-sam* for generating pairs files). With these Hi-C read pairs files, we generated a contact map using

JuicerTools (Durand et al. 2016b) (version 1.22.01; *juicer_tools_1.22.01.jar pre*) and the output files were visualized using Juicebox (Durand et al. 2016a) (version 1.11.08).

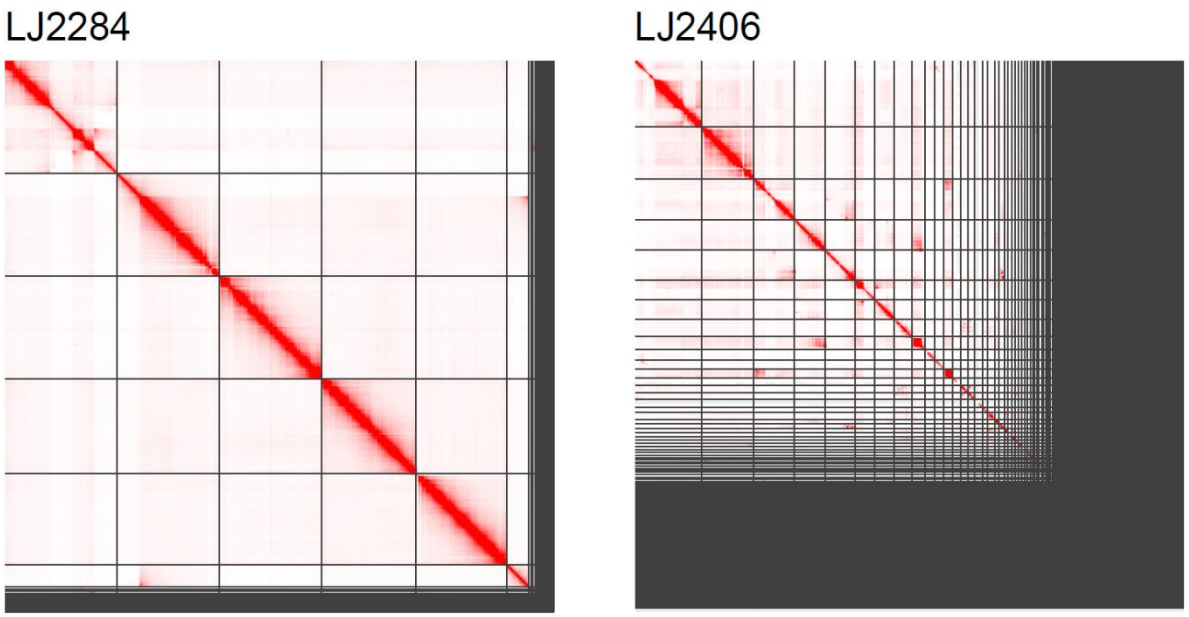
Visualizing synteny relationships

Complete and single-copy BUSCO genes were filtered to select common genes in the genome assemblies of *B. xylophilus* and Panagrolaimidae. The common, complete and single-copy BUSCO genes in the constructed genome assemblies were further categorized according to the chromosomes of *B. xylophilus* and localized by assigning specific colors to each chromosome. We visualized their synteny relationships using Circos (Krzywinski et al. 2009) (version v 0.69-8; *circos -conf*).

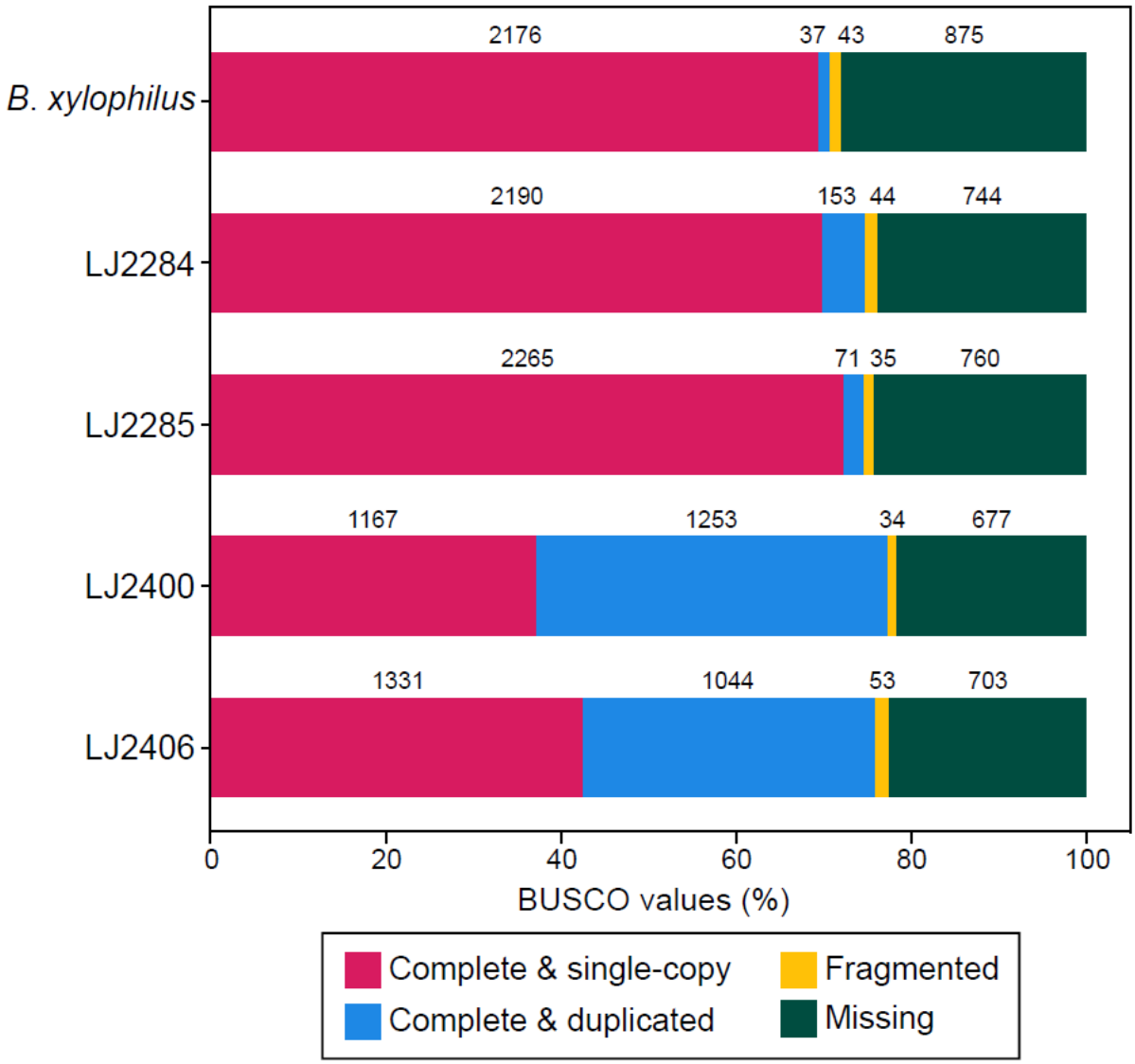
Supplemental Figures



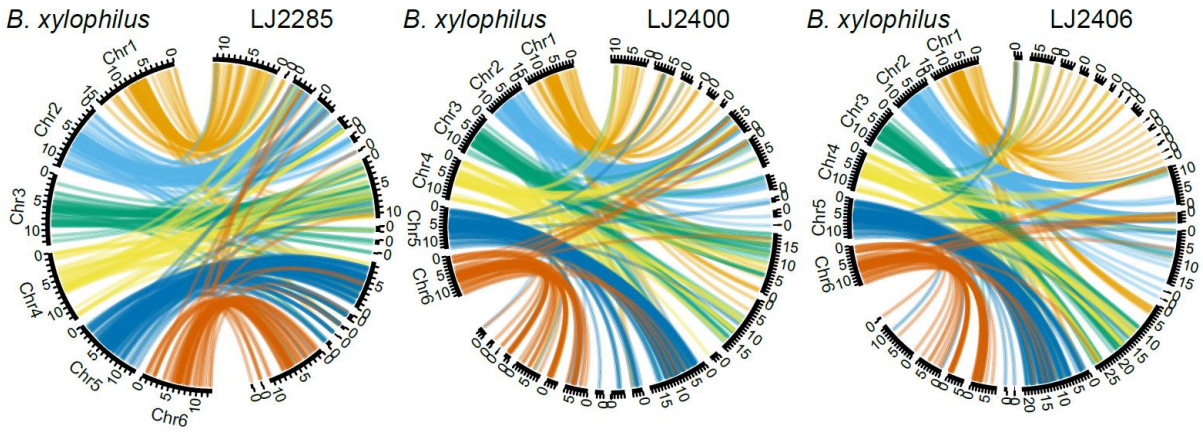
Supplemental Figure S1. Estimated telomeric-repeat counts. Each telomeric-repeat count was obtained by averaging the counts of 23-mers that contained each TRM among sub-sampled 5 million reads of short-read sequencing data using the *k*-mer-based method. Each bar indicates the count of the canonical Nematoda TRM (TTAGGC) and the novel TRMs for each species. The first species of each clade is a control species that harbors the canonical TRM. Each color represents a specific TRM as follows: red, TTAGGC; orange, TTAGAC; blue, TTAGGT; khaki, TTAGGG.



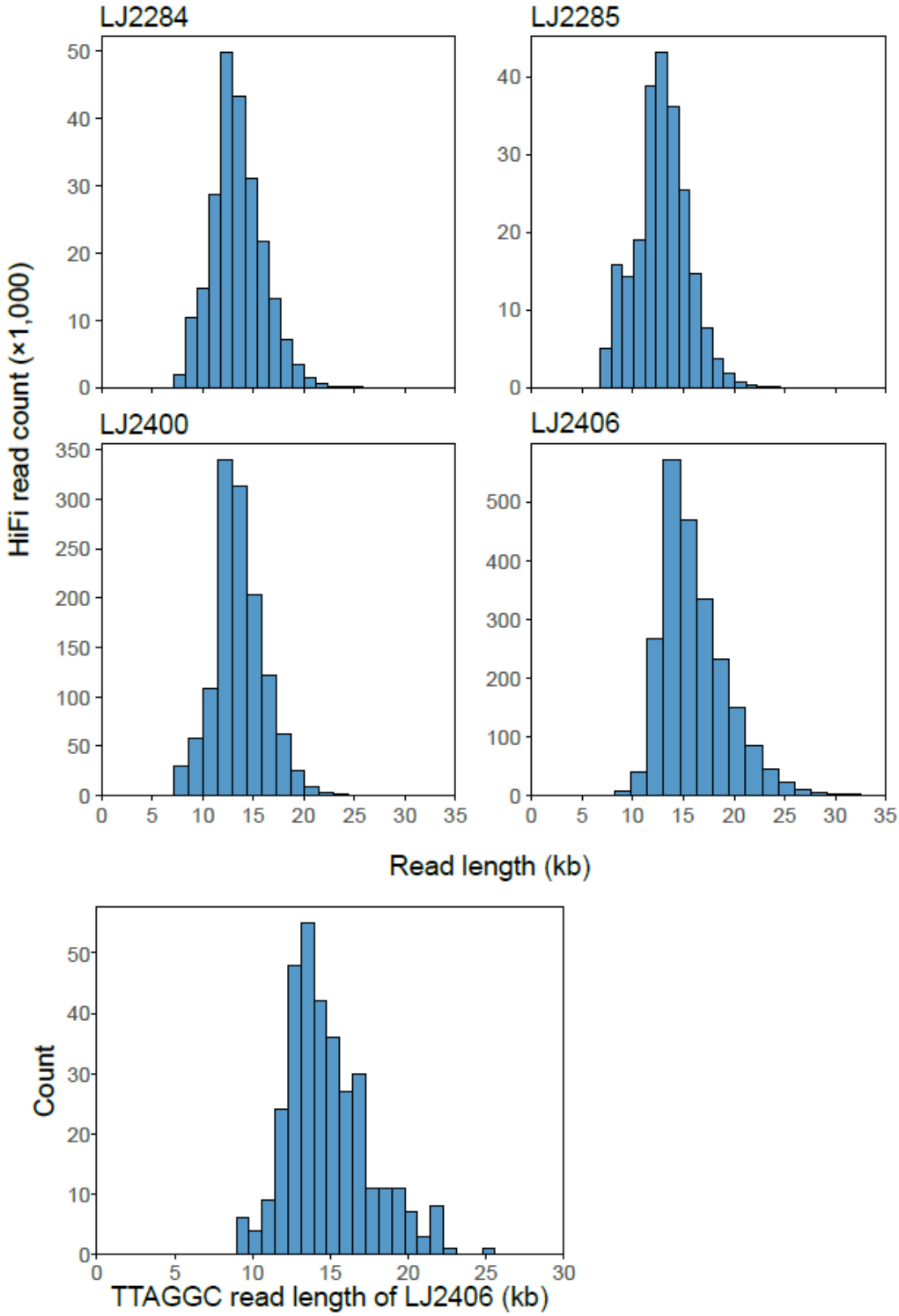
Supplemental Figure S2. Heatmaps for Hi-C contact maps of LJ2284 and LJ2406 genome assemblies. Scaffolds were ordered from top to bottom and from left to right according to lengths, and interactions between each position are shown in red.



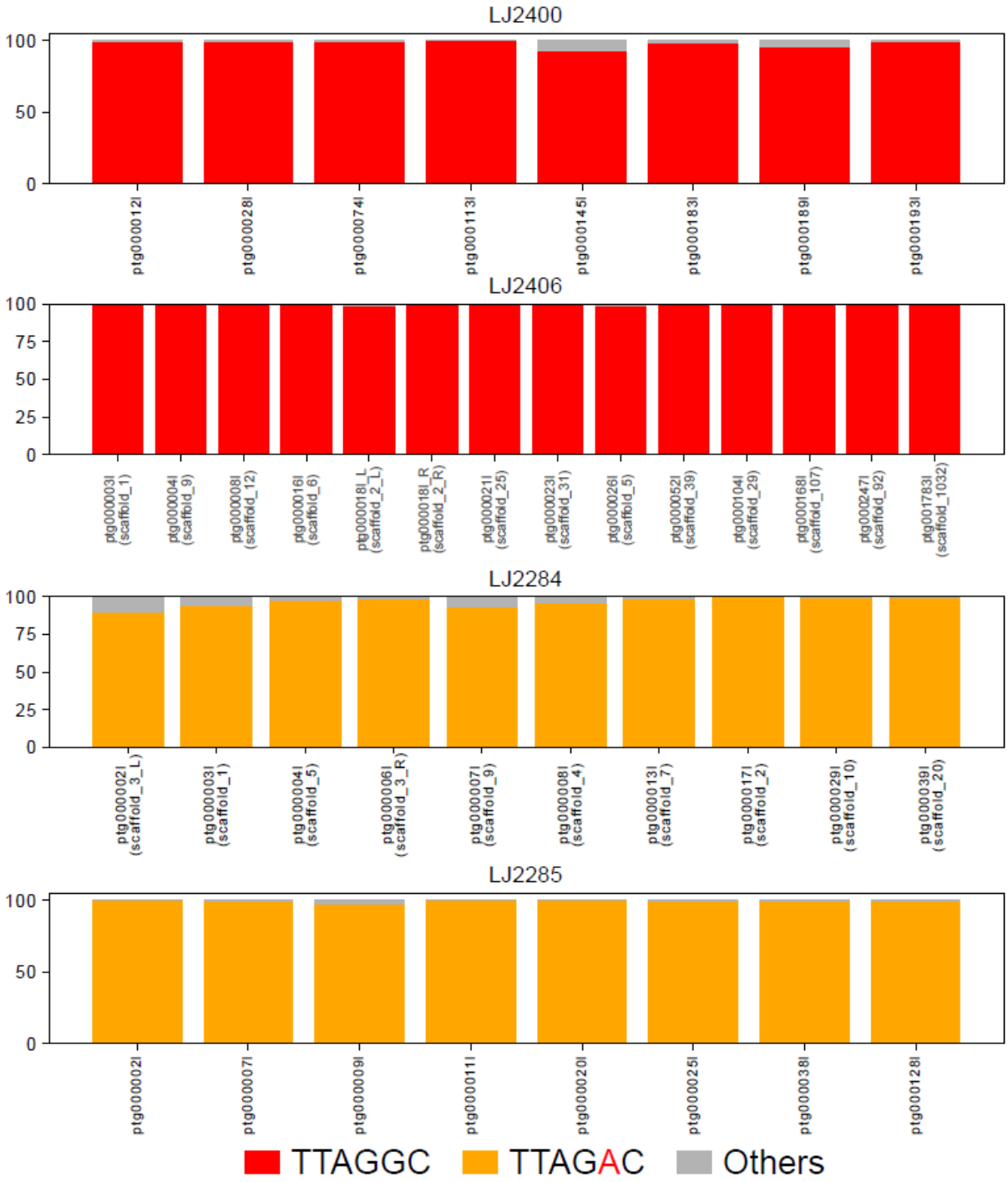
Supplemental Figure S3. BUSCO values of LJ2284, LJ2285, LJ2400, and LJ2406. *B. xylophilus* was used as a control.



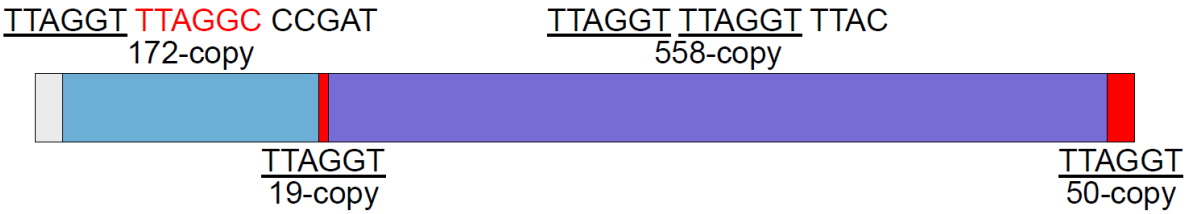
Supplemental Figure S4. Synteny plots of LJ2285, LJ2400, and LJ2406 compared to *B. xylophilus*. Each colored line represents a BUSCO gene shared among the genome assemblies of LJ2284, LJ2285, LJ2400, LJ2406, and *B. xylophilus*. Orange lines indicate that corresponding BUSCO genes are in Chromosome 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).



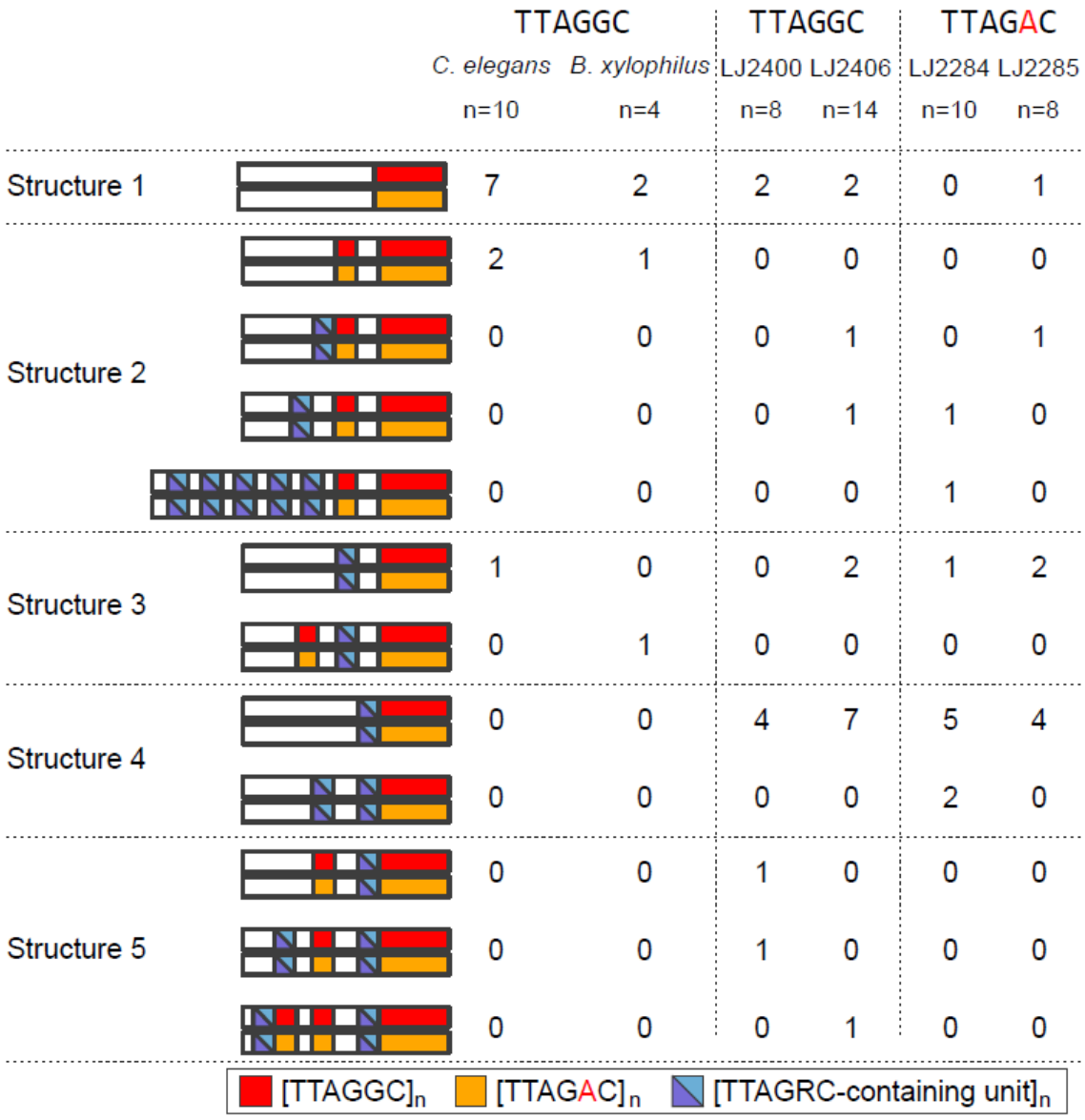
Supplemental Figure S5. HiFi read length distributions for four Panagrolaimidae isolates and telomeric read length distribution of LJ2406



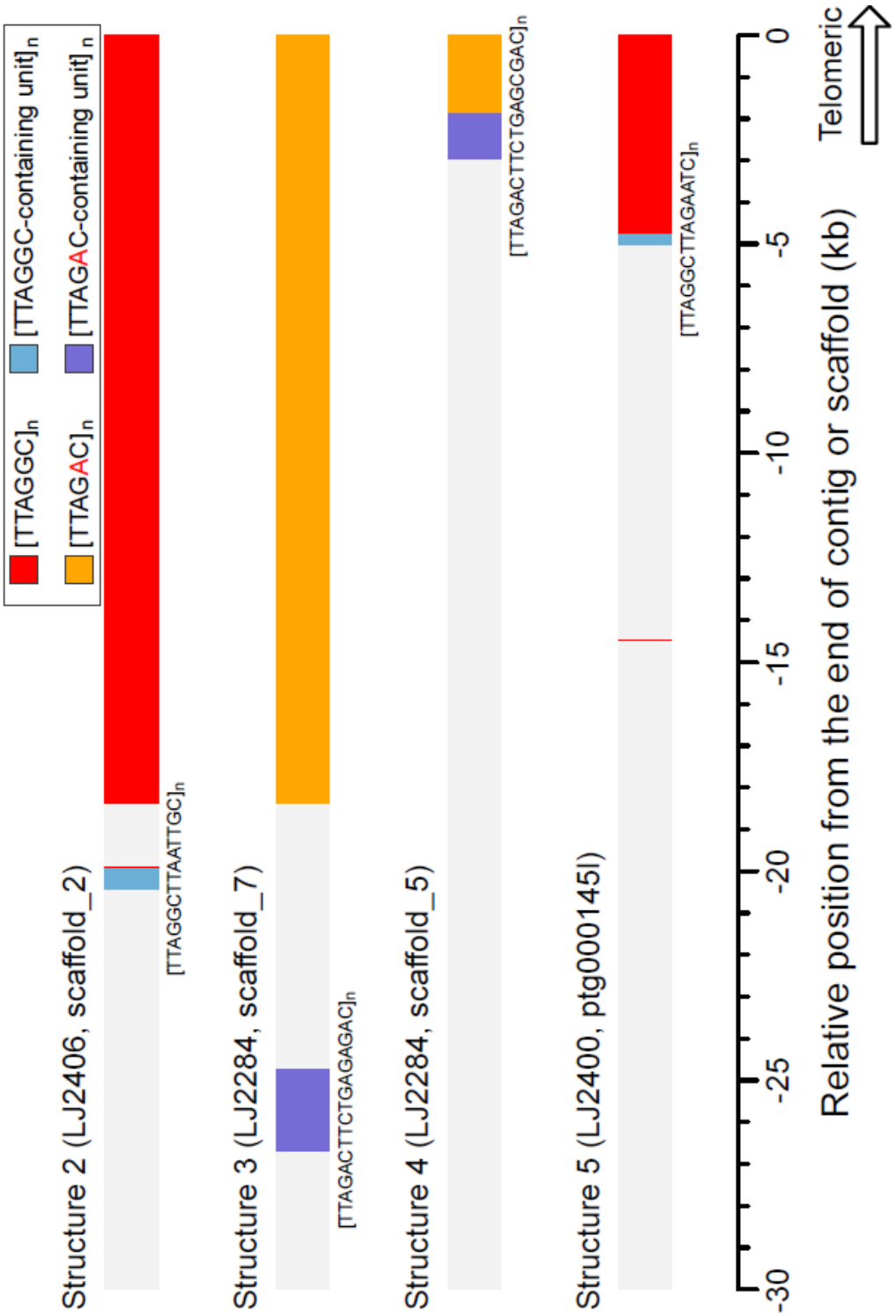
Supplemental Figure S6. The proportion of TRM types in clustered telomeric repeats at the end of each contig/scaffold



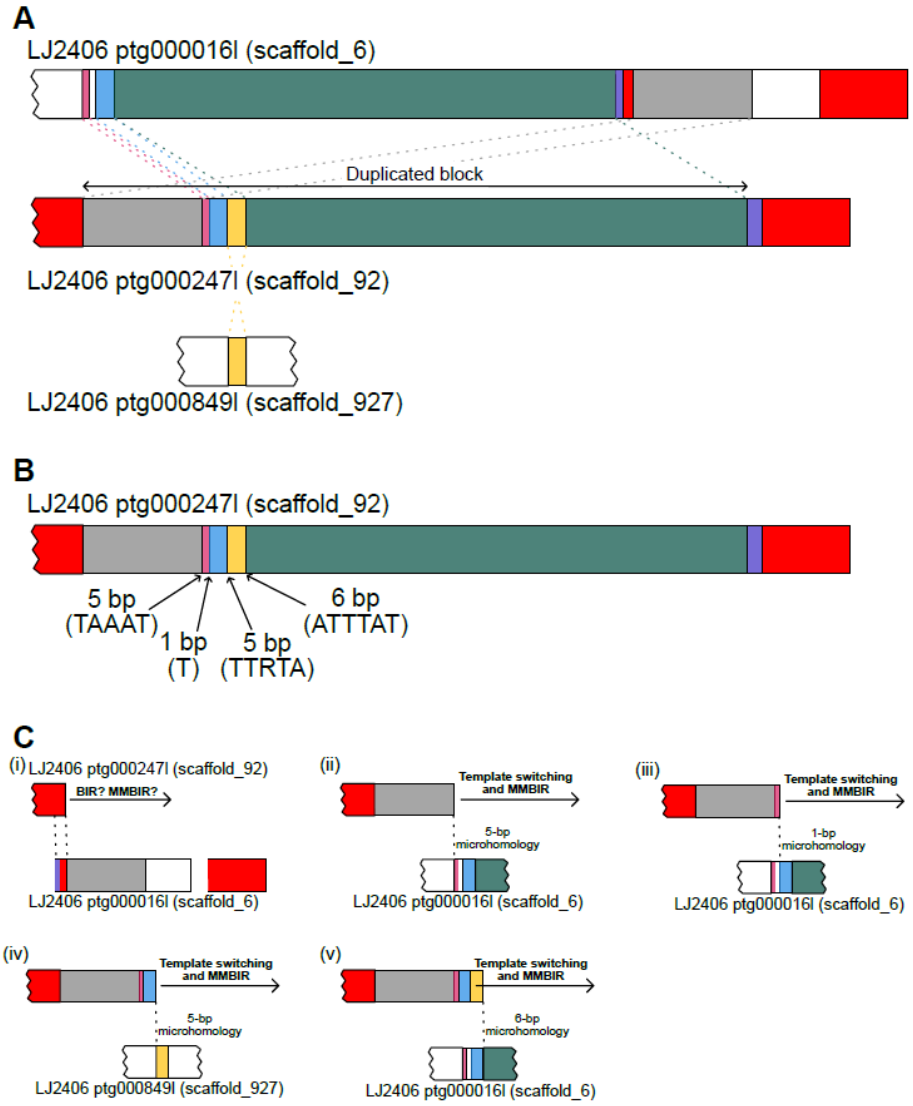
Supplemental Figure S7. A 13.3-kb read of *Caenorhabditis uteleia* consisting of two TTAGGT- and/or TTAGGC-containing unit clusters and two TTAGGT clusters



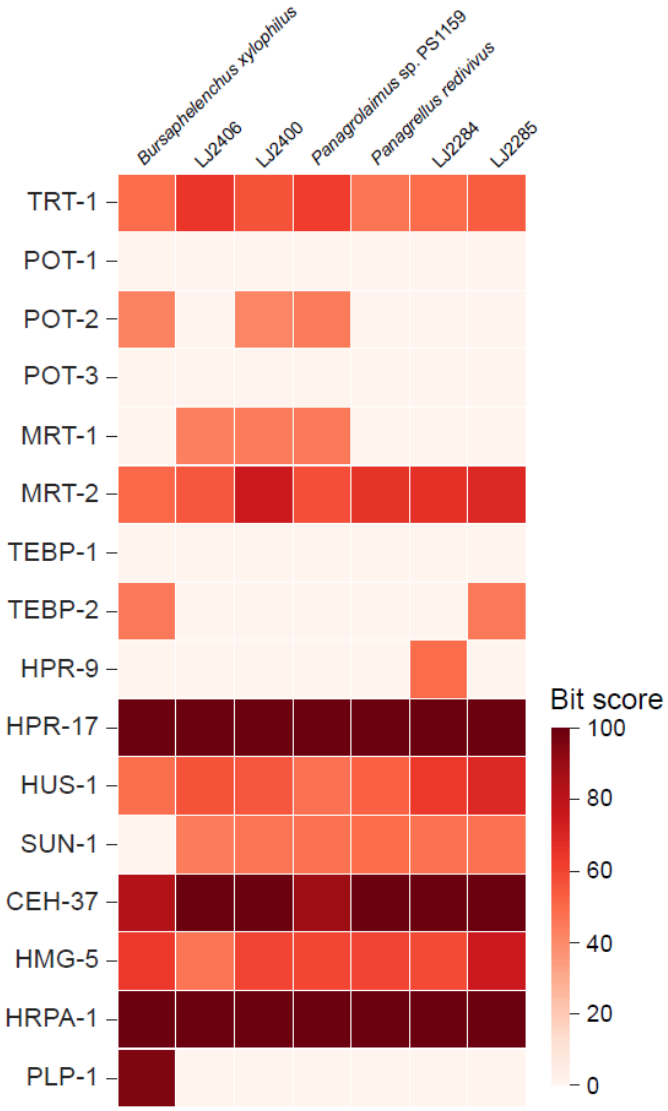
Supplemental Figure S8. Categorization of subtelomeric structures into 12 subdivided types. Each horizontal bar represents a subtelomeric region that is up to 200 kb from the end of the contig/scaffold. TTAGGC-telomere species and isolates had ITSs that were composed of TTAGGC, rather than TTAGAC, and TTAGAC-telomere species and isolates had only TTAGAC-type ITSs, too.



Supplemental Figure S9. Examples of subtelomere structure types shown only up to 30 kb from the end of each contig or scaffold



Supplemental Figure S10. Microhomology between homologous blocks. (A) Duplicated block and homologous sequence blocks found in ptg000016l and ptg000849l. Each color represents a homologous block pair, except red (TTAGGC cluster) and purple (TTAGAC-containing unit cluster) blocks. (B) Microhomology lengths and sequences that overlapped between their homologous sequence blocks in other contigs/scaffolds. (C) A model for consecutive events of template switching and MMBIR that could have generated the new subtelomeric region in ptg000247l in LJ2406.



Supplemental Figure S11. A heat map showing whether *C. elegans* telomere maintenance/binding proteins are conserved in Panagrolaimidae species/isolates. The color index represents the bit score of each *C. elegans* protein alignment with the corresponding Panagrolaimidae protein of each species/isolate. The maximum value of the index was set to 100; thus, values over 100 had the same color as values = 100. *Bursaphelenchus xylophilus* was used as an outgroup species.

Supplemental Tables Legends

Supplemental Table S1. Accession numbers for genome assembly and short-read sequencing data, genome sizes, and TRMs for species for which we identified the putative TRM. Five million sub-sampled reads were used. Orange box, genome assembly accessions belonging to a different BioProject from short-read sequencing data for the species; green box, novel TRM. ND: not determined.

Supplemental Table S2. Accessions numbers of genome assembly and short-read sequencing data, genome sizes, and TRM analysis results for species for which we could not determine the TRM. Five million sub-sampled reads were used. Orange box, genome assembly accessions belonging to a different BioProject from short-read sequencing data for the species. ND: not determined.

Supplemental Table S3. Average counts of 23-mers containing concatemers of each TRM among sub-sampled five million reads of short-read sequencing data.

Supplemental Table S4. Validation of putative TRMs in parasitic nematodes.

Supplemental Table S5. Summary data of species/isolates of Panagrolaimidae that contain species/isolate names, sampling information for substrate and location, TRMs and TBPs homologous to those of *C. elegans*. Twenty million sub-sampled reads were used. ND: not determined.

Supplemental Table S6. 18S ribosomal DNA sequences in Panagrolaimidae.

Supplemental Table S7. Summary of Hi-C scaffolding of HiFi-based genome assemblies of LJ2284 and LJ2406.

Supplemental Table S8. HiFi- and/or HiC-based *de novo* genome assembly statistics for four Panagrolaimidae isolates.

Supplemental Table S9. Raw data of TRM proportions for clustered telomeric repeats at the end of each contig/scaffold.

Supplemental Table S10. Distributions of telomere, ITS and subtelomeric TTAGRC-containing units in telomere-containing contigs/scaffolds.

Supplemental Table S11. Distributions of telomeric repeats, ITS and subtelomeric TTAGGC-containing units in telomeric repeat-containing chromosomes in *C. elegans* and *B. xylophilus*.

Supplemental Table S12. List of subtelomeric units/clusters containing TTAGRC.

Supplemental Table S13. Microhomology and homologous sequence blocks in ptg000247l in LJ2406, potentially generated by MMBIR.

Supplemental Table S14. Conservation patterns of homologs of TEBP-2, POT-2, MRT-1, or HPR-9. Each value represents the bit score of each TBP homolog alignment for the corresponding protein of each species/isolate.

Supplemental Material References

Abadie SH. 1963. The life cycle of Strongyloides ratti. *The Journal of parasitology*: 241-248.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC bioinformatics* **10**: 1-9.

The Darwin Tree of Life Project Consortium. 2022. Sequence locally, think globally: The Darwin tree of life project. *Proceedings of the National Academy of Sciences* **119**: e2115642118.

Dayi M, Sun S, Maeda Y, Tanaka R, Yoshida A, Tsai IJ, Kikuchi T. 2020. Nearly complete genome assembly of the pinewood nematode Bursaphelenchus xylophilus strain Ka4C1. *Microbiology resource announcements* **9**: e01002-01020.

Dick T, Belosevic M. 1978. Observations on a Trichinella spiralis isolate from a polar bear. *The Journal of Parasitology* **64**: 1143-1145.

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016a. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell systems* **3**: 99-101.

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. 2016b. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell systems* **3**: 95-98.

Eweis DS, Delattre M, Plastino J. 2022. Asymmetry is defined during meiosis in the oocyte of the parthenogenetic nematode Diploscapter pachys. *Developmental Biology* **483**: 13-21.

Feng Y, Liu X, Liu Y, Tang B, Bai X, Li C, Wang X, Deng Y, Gao F, Liu M. 2021. Comparative epigenomics reveals host diversity of the Trichinella epigenomes and their effects on differential parasitism. *Frontiers in Cell and Developmental Biology* **9**: 681839.

Fradin H, Kiontke K, Zegar C, Gutwein M, Lucas J, Kovtun M, Corcoran DL, Baugh LR, Fitch DH, Piano F. 2017. Genome architecture and evolution of a unichromosomal asexual nematode. *Current Biology* **27**: 2928-2939. e2926.

Kapel CM. 2000. Host diversity and biological characteristics of the Trichinella genotypes and their effect on transmission. *Veterinary parasitology* **93**: 263-278.

Kim C, Kim J, Kim S, Cook DE, Evans KS, Andersen EC, Lee J. 2019. Long-read sequencing reveals intra-species tolerance of substantial structural variations and new subtelomere formation in C. elegans. *Genome research* **29**: 1023-1035.

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.

Korhonen PK, Pozio E, La Rosa G, Chang BC, Koehler AV, Hoberg EP, Boag PR, Tan P, Jex AR, Hofmann A. 2016. Phylogenomic and biogeographic reconstruction of the Trichinella complex. *Nature communications* **7**: 10513.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome research* **19**: 1639-1645.

Kulkarni A, Holz A, Rödelsperger C, Harbecke D, Streit A. 2016. Differential chromatin amplification and chromosome complements in the germline of Strongyloididae (Nematoda). *Chromosoma* **125**: 125-136.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009.

The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078-2079.
Mitreva M, Jasmer DP. 2006. Biology and genome of Trichinella spiralis. *WormBook: the online review of C elegans biology*: 1-21.
Nwaorgu O, Connan R. 1980. The migration of Strongyloides papillosus in rabbits following infection by the oral and subcutaneous routes. *Journal of helminthology* **54**: 223-232.
Pozio E, La Rosa G, Rossi P, Fico R. 1989. Survival of Trichinella muscle larvae in frozen wolf tissue in Italy. *The Journal of Parasitology* **75**: 472-473.
Schwarz EM. 2017. Evolution: a Parthenogenetic nematode shows how animals become sexless. *Current Biology* **27**: R1064-R1066.
Wong B. 2011. Color blindness. *nature methods* **8**: 441.
Yamaguchi T. 1991. Present status of trichinellosis in Japan. *Southeast Asian J Trop Med Public Health* **22**: 295-301.
Yoshimura J, Ichikawa K, Shoura MJ, Artiles KL, Gabdank I, Wahba L, Smith CL, Edgley ML, Rougvie AE, Fire AZ. 2019. Recompleting the Caenorhabditis elegans genome. *Genome research* **29**: 1009-1022.
Zhou C, McCarthy SA, Durbin R. 2023. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**: btac808.