Robustness with Limited Specificity of Telomere Binding Proteins Underlies Telomere Evolution in Nematoda

Hobum Song[1], Seon Hong Kim[2], Daisy S. Lim[3], Hee-Jung Choi[2, *], and Junho Lee[1, 2, 3, *]

1 Institute of Molecular Biology and Genetics, Seoul National University, Seoul 08826, Korea

2 Department of Biological Sciences, Seoul National University, Seoul 08826, Korea

3 Research Institute of Basic Sciences, Seoul National University, Seoul 08826, Korea

*Correspondences to:

Junho Lee

Department of Biological Sciences

Seoul National University, Seoul 08826, Korea

E-mail: elegans@snu.ac.kr

Phone: +821029442663

Keywords: nematodes, telomere-binding proteins, telomere repeat motif, evolution, binding specificity

Abstract

Telomeres, nucleoprotein complexes protecting the ends of linear chromosomes, are essential for ensuring the stability of eukaryotic genomes. As telomeres generally consist of repetitive DNA associated with specifically bound proteins, the evolution of telomeric repeat motifs hardly occurs. However, a recent study identified nematodes with telomeric repeats other than the canonical TTAGGC motif. Here, we investigated how telomere repeats were able to evolve despite the challenge posed by the specificity of telomere-binding proteins to the telomeric DNA in Nematoda, focusing on CEH-37. Initially, we assessed the conservation of telomere-binding proteins across diverse nematode species and obtained the well-conserved telomere-binding protein CEH-37 from each species with different telomeric repeats. The binding affinities of the proteins, measured by electrophoresis mobility shift assays, revealed that CEH-37 exhibits limited specificity, which cannot distinguish telomeric repeats found in nematodes except for TTAGGG motif. Furthermore, through the analysis of TBP conservation and the structure of CEH-37 in nematodes, we inferred that CEH-37 functioned as a double-stranded telomeric DNA binding protein before TEBP proteins emerged. Through this study, we suggest that telomeric sequences in Nematoda were able to evolve within the range of that does not exceed the sequence-specificity of telomere-binding proteins.

Introduction

Genome stability in eukaryotes is inherently challenged by the continuous shortening of chromosomal ends and their misrecognition as double-strand breaks (Muller, 1938; McClintock, 1938; Watson, 1972; Olovnikov, 1973). To address these problems, eukaryotes have nucleoprotein structures at the ends of their linear chromosomes called telomeres. Telomere-binding proteins (TBPs) specifically bind to DNA sequences generally composed of G-rich repeat motif, forming complexes that cap the chromosomal ends to prevent the activation of the DNA damage response and regulate the length of telomeres, ensuring functional telomeres. Consequently, changes in telomeric repeat motifs (TRMs) can lead to genomic instability by affecting the binding affinities of TBPs, making the evolution of telomeric sequences challenging (Steinberg-Neifach & Lue, 2015; Prušáková et al., 2021).

Despite the challenging nature of telomeric sequence evolution, it has been observed in several lineages, including yeast, plants, and insects (reviewed in Červenák et al., 2021; Procházková et al., 2016; Kuznetsova et al., 2019). Various hypotheses have been proposed to explain how this process is possible, and some of these hypotheses have been experimentally validated in certain organisms (Steinberg-Neifach & Lue, 2015; Saint-Leandre & Levine, 2020). In yeast, the adoption of novel TBPs to cope with rapid changes in telomeric sequences has been observed, while the coevolution of TRM and TBP has been the principle of telomere evolution in plants (Sepsiova et al., 2016; Červenák et al., 2021; Shakirov et al., 2009). In Animalia, however, no studies have been conducted to explain the relationship between TBPs and the evolution of telomeric sequences.

Recently, novel TRMs have been discovered within the Nematoda phylum, which includes one of the representative model organism *Caenorhabditis elegans* (Lim et al., 2023). While it has been well established that the nematode TRM is TTAGGC, *C. uteleia* TRM TTAGGT and some species of Panagrolaimidae family having TTAGAC TRM have been revealed by analyzing whole genome sequence (WGS) data. Since many TBPs of *C. elegans* have been identified, we may understand how TBPs had adapted to novel telomeric sequences by comparing highly conserved TBPs from nematodes with noncanonical TRMs to those with traditional TRMs.

In this study, we aimed to understand how telomeres evolved in Nematoda despite the specificity of TBPs by using the functional domain of CEH-37, a highly conserved TBP in nematodes. We obtained CEH-37 homeodomains from species with different TRMs to confirm its binding affinity to recently discovered nematode telomeric DNAs. The CEH-37 homeodomain flexibly binds to each nematode telomeric DNA while distinguishing the common TTAGGG tandem sequence. Additionally, our study provides insights into the role of CEH-37 before the emergence of TEBP proteins in clade V nematodes, which facilitated the emergence of non-canonical TRMs in Nematoda due to its limited specificity.

Experimental procedures

Strains

Nematodes were grown in nematode growth medium seeded with *Escherichia coli* strain OP50 at 20℃ as standard methods (Brenner, 1974). *C. elegans* Bristol N2, *C. uteleia* JU2585, and Panagrolaimidae isolates (LJ2284, LJ2285, LJ2400, and LJ2406) were used for this study.

Conservation analysis of telomere binding proteins

FASTA protein files of *C. elegans* (PRJNA13758) and *C. uteleia* (PRJEB12600) were obtained from WormBase ParaSite (Kevin et al., 2017; Kevin et al., 2016). Protein sequences of Panagrolaimidae isolates (LJ2284, LJ2285, LJ2400, and LJ2406) were kindly produced by Lim et al.

To analyze conservation of telomere binding proteins of *C. elegans*, we used DIAMOND (version 2.1.8; diamond blastp -d -q -o --threads 20 --very-sensitive and diamond blastp -d -q -o --threads 20 --ultra-sensitive) (Buchfink et al. 2021). The conservation was evaluated by using highest bit score protein sequence with conserved domains of original proteins identified using NCBI CD-search (CDD database, version 3.21) (Marchler-Bauer and Bryant 2004; Lu et al. 2020; Wang et al., 2023).

Construct generation

ceh-37 homeobox region of each strain was cloned in a modified pGEX-4T-1 vector containing a HRV 3C cleavage site after the GST tag. To generate cDNA of ceh-37 homeobox, RNAs isolated from nematodes with TRIzol were reverse transcripted with PrimeScript RT Master Mix (Takara) except for *Strongyloides ratti*. Since we were not able to culture *S. ratti*, ceh-37 homeobox cDNA of *S. ratti* was generated by ligating two annealed ~110bp oligos after phosphorylation. cDNAs were fused into the vector by Gibson assembly master mix (New England Biolabs).

Protein expression and purification

For protein expression, protein encoding plasmids were transformed into *E. coli* Rosetta (DE3). Cells were grown in Terrific Broth until OD600 reached ~0.6, and protein expression was induced for 18 hours at 20℃ by adding IPTG at final concentration of 0.2 mM. Cells were harvested by centrifugation and were resuspended in PBS. After adding PMSF and DNase I, cells were sonicated on ice. The supernatant, collected via centrifugation of the lysed sample, was then loaded onto equilibrated glutathione agarose resin. Resin was washed two times with wash buffer (20 mM HEPES

pH 7.5, 100 mM NaCl) and the bound protein were released by HRV 3C protease treatment. The eluates were further purified by cation exchange chromatography and size exclusion chromatography using HiTrap SP HP (Cytivia) and Superdex 200 Increase (Cytivia), respectively.

Electrophoretic mobility shift assay (EMSA)

Telomeric DNA substrates used as DNA probes were prepared by annealing oligos with forward strand having Cy5 fluorescent on its 5' end. 5 nM of DNA probes were incubated with appropriate concentration of proteins in binding buffer (20 mM Tris-HCl, 50 mM NaCl, 1 mM MgCl2, 5 mM DTT, 5% glycerol, 0.001% tween-20, and 50 ug/mL BSA) for 15 minutes in room temperature. Samples were loaded to 6% native gel that had been pre-run. Electrophoresis was performed for 90 minutes at 130 V in cold room with 0.5x TBE as a running buffer.

Structure prediction and alignment

To predict biomolecular structures, we utilized the AlphaFold Server powered by AlphaFold 3 model (Abramson et al., 2024). Alignments of CEH-37 homeodomains from species with distinct TRMs were performed using the Matchmaker tool in UCSF ChimeraX (version 1.5) (Goddard et al., 2018; Pettersen et al., 2021).

Structural alignments were conducted using the web-based pairwise structure alignment service provided by the RCSB PDB website, employing the TM-align method (Zhang & Skolnick, 2005). The structures of each protein were obtained from models stored in the AlphaFold Protein Structure Database (Jumper et al., 2021; Varadi et al., 2022). The regions of proteins were pre-determined for proteins whose DNA-binding domains are previously specified (Konig et al., 1996; Sigrist et al., 2005; UniProt Consortium, 2023; Marchler-Bauer and Bryant 2004; Lu et al. 2020; Wang et al., 2023; Padmanaban et al., 2024).

Results

CEH-37 is one of the well-conserved telomere binding proteins in Nematoda

Eukaryotes form protein complexes, such as shelterin, terminin, to cap their telomeric DNAs (de Lange, 2005; Raffa et al., 2011; Moser & Nakamura, 2009). In *C. elegans*, several TBPs have been identified, and TEBP-1, TEBP-2, POT-1, POT-2, and MRT-1 are thought to form a telomeric protein complex (Raices et al. 2008; Yamamoto et al., 2021; Im & Lee, 2003, 2005; Joeng et al., 2004; Meier

et al., 2009; Yu et al., 2024; Kim et al., 2003). However, there are many species other than *C. elegans* in Nematoda, some of which possess noncanonical TRMs (Lim et al., 2024). Since TBPs need to bind effectively to telomeric DNA for telomeres to function properly, we investigated which TBPs of *C. elegans* have been conserved across other nematode species despite differences in their TRMs.

Nematodes are classified into five clades based on small subunit rRNA phylogeny (Blaxter et al., 1998). By conducting BLAST analysis with representative species from each clade, it was observed that the TBPs of *C. elegans* are well conserved within clade V (Figure 1A; Supplementary table 1). However, most of the TBPs, including TEBP and POT proteins, are not conserved in other clades. Meanwhile, the double-stranded telomeric DNA binding protein (dsTBP) CEH-37 and single-stranded telomeric DNA binding protein (ssTBP) HRPA-1 are well conserved in Nematoda, regardless of clade.

Conserved homeodomain of CEH-37 robustly binds to telomeric DNAs with limited binding specificity

CEH-37 is a homeodomain protein that binds to DNA using its homeodomain, which is conserved regardless of the strain's TRM (Figure 1B). To investigate how this protein has remained conserved despite changes in telomeric sequences, we obtained the homeodomains of CEH-37 from nematodes with different TRMs and measured their binding affinities to various telomeric DNAs.

First, we obtained the CEH-37 homeodomain from the model organism *C. elegans* and its related species *C. uteleia*, which has the non-canonical TRM TTAGGT, and measured their binding affinities to DNAs consisting of telomeric repeats found in nematodes using electrophoresis mobility shift assay (EMSA) (Figure 2; Supplementary figure S1; Supplementary table 2, 3). Interestingly, CEH-37 homeodomains from both species bind to telomeric DNAs with TTAGGC, TTAGAC, and TTAGGT repeats, with no significant difference in their dissociation constants (K_d). However, both CEH-37 homeodomains do not effectively bind to DNA with TTAGGG repeats. These results indicate that CEH-37 does not distinguish TTAGGC, TTAGAC, and TTAGGT sequences, while it shows less preference on TTAGGG sequence.

We also measured the K_d values of telomeric DNAs and CEH-37 homeodomains from strains in clade IV, where nematodes have TRMs not found in clade V (Supplementary figure S2; Supplementary table 2, 3). We measured the K_d values of CEH-37 homeodomains from LJ2284 and LJ2285, which were expected to have the TTAGAC TRM based on a previous study, as well as from LJ2400, a related species with the canonical nematode TRM, to various telomeric DNAs (Figure 3A-C). Similar to the results observed in species from clade V, every CEH-37 homeodomains exhibit robust binding to DNA with limited sequence specificity, irrespective of the TRM of their origin species.

S. ratti have been speculated to have TTAGGG TRM. Nevertheless, according to the results above, CEH-37 does not bind to DNA consisting of TTAGGG repeats effectively (Figure 2, 3A-C). Therefore, we evaluated the binding affinities of CEH-37 homeodomain from *S. ratti* to nematode

telomeric DNAs (Figure 3D; Supplementary figure S2; Supplementary table 2, 3). Interestingly, the CEH-37 homeodomain from *S. ratti* also does not bind efficiently to TTAGGG-repeated DNA as other telomeric DNAs whose sequences were identified in nematodes. This result suggests that *S. ratti* either employs an end protection mechanism other than CEH-37 capping or does not possess telomeres composed of TTAGGG repeats.

CEH-37 is the only well-conserved double-stranded telomeric DNA binding protein that has the typical structure of double-stranded telomeric DNA binding proteins in nematodes.

Telomeric complexes contain dsTBP subunits, which generally have Myb-like domains as their DNA-binding domains (DBDs) (references for human, mouse, yeast, etc.). In *C. elegans*, TEBP proteins with Myb-containing domains, which use these domains for DNA binding, have also been identified. These proteins are conserved within the *Caenorhabditis* genus, although there may be variations in the number of these proteins (Yamamoto et al., 2021; Dietz et al., 2021; Padmanaban et al., 2024). Despite the expectation that these proteins form telomeric complexes, they are not conserved in nematodes outside of clade V (Figure 1A). To maintain functional telomeres, where protein complexes cap the telomeric DNA, other TBPs might bind to the telomeric DNA in nematodes from clade I to IV.

As mentioned above, CEH-37 is a well-conserved TBP. Although it utilizes a homeodomain instead of a Myb-related domain to bind double-stranded telomeric DNA, its DBD has been hypothesized to share structural homologs with TBPs in other species (Kim et al., 2003; Moon et al., 2014). To quantitatively analyze the structural alignments of *C. elegans* TBPs with Myb-related domains in TBPs from other species, we aligned the Myb domain of the representative human TBP TRF1 to *C. elegans* dsTBPs (Figure 4; Supplementary table 4). According to measured template modeling scores (TM-scores), the CEH-37 homeodomain shows structural homology, as well as TEBP-1 and TEBP-2, whose third Myb-containing domains (MCD) are discovered to act as DBDs, although it does not exhibit sequence homology.

Next, we aligned *C. elegans* CEH-37 homeodomain to the double-stranded DNA binding subunits of telomeric complexes in other species (Supplementary figure S3; Supplementary table 4). The similarities in protein folds between the homeodomain of CEH-37 and the DBDs of TBPs in other species are identified, as they had TM-score over 0.5.

Discussion

In this study, we found that the evolution of TRMs in Nematoda occurred within the range allowed by the binding specificity of the well-conserved TBP CEH-37. Furthermore, CEH-37 is speculated to

be the ancestral double-stranded telomeric DNA binding subunit of the telomeric complex. According to the TBP conservation pattern in Nematoda (Figure 1A), it can be interpreted that after the emergence of ancient TEBP and POT proteins in clade III, these proteins further evolved to form the TEBP-POT complex in clade V, while they regressed in clade IV.

We confirmed that CEH-37 homeodomain of every species, including *S. ratti*, does not effectively bind to TTAGGG-repeats compared to other telomeric DNAs discovered in Nematoda, raising questions about the nature of the telomere in *S. ratti*. Previous whole genome sequence (WGS) data suggested *S. ratti* possesses TTAGGG TRM, as this was the only TTAGGC-related motifs identified (Lim et al., 2024). However, the number of the TRM-containing concatemers was very low compared to TRM-containing reads from other nematodes WGS data analyzed in the previous study, ranking 67th out of 67 species. Based on this information, we hypothesized that (1) *S. ratti* possesses short telomeric DNAs consisting of TTAGGG repeats, possibly protected by TBPs other than CEH-37 or other mechanisms (e.g. adoption of rigid G-quadruplex (G4) structures), or that (2) *S. ratti* does not actually possess TTAGGG TRM, and the observed reads from the parasite may have originated from its mammalian host (Zahler et al., 1991; Smith et al., 2011; Bryan, 2020). If the first hypothesis is true, it would be interesting to investigate the role of G4 structure in telomeres by comparing telomeres of *S. ratti*, nematodes with TTAGGC TRMs, which are also known to form telomeric G4 structures, and nematodes with TTAGAC TRMs, which might be unable to form telomeric G4 structures (Marquevielle et al., 2022).

The evolution of telomeric sequences and their associated proteins has also been observed in yeasts. Yeasts exhibit a wide range of TRMs, from common TTAGGG repeats to long (>20 bp) and heterogenous TRMs. This extreme variability resulted in the adoption of flexible TBPs, such as Rap1 and Taz1, instead of the strictly sequence-specific Taz1/Teb1 (Sepsiova et al., 2016; Moser & Nakamura, 2009). Following this evolution, the original TBPs either disappeared, transformed into telomere-associated proteins composing the telomeric protein complex (i.e. Rap1), or dedicated to their original second role (i.e. Teb1 as a transcription factor). The last situation is reminiscent of the nematode TBP CEH-37, which is also known for its role as a transcription factor involved in sensory neuron specification and intestinal immunity (Lanjuin et al., 2003; Liu et al., 2023).

The main difference between CEH-37 and other dsTBPs is that CEH-37 uses a homeodomain instead of Myb-related domains to bind to DNA. However, its structure is similar with Myb-related domains of TBPs in other species, indicating the convergent evolution (Supplementary table 4). Meanwhile, the identification of dsTBPs in many species has been predominantly focused on proteins with Myb-related domains, even though the protein function is primarily determined by its structure rather than its sequence (Li et al., 2005; Yu et al., 2000; Fulcher & Riha, 2016). The presence of CEH-37 suggests the possibility of dsTBPs with domains other than Myb-related domains existing in other species as well.

This study is the first to investigate the mechanism of TBPs to handle the evolution of the TRM in animals. We expect our study will contribute not only to telomere research using nematodes as

multicellular organism models but also to understanding the evolution of DNA-protein interaction beyond telomeres.

References

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung CC, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Žemgulytė A, Arvaniti E, Beattie C, Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Žídek A, Bapst V, Kohli P, Jaderberg M, Hassabis D, Jumper JM. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024 Jun;630(8016):493-500. doi: 10.1038/s41586-024-07487-w. Epub 2024 May 8. PMID: 38718835; PMCID: PMC11168924.

Blaxter, M., De Ley, P., Garey, J. et al. A molecular evolutionary framework for the phylum Nematoda. Nature 392, 71–75 (1998). https://doi.org/10.1038/32160

Bryan TM. G-Quadruplexes at Telomeres: Friend or Foe? Molecules. 2020 Aug 13;25(16):3686. doi: 10.3390/molecules25163686. PMID: 32823549; PMCID: PMC7464828.

Brenner S. The genetics of Caenorhabditis elegans. Genetics. 1974 May;77(1):71-94. doi: 10.1093/genetics/77.1.71. PMID: 4366476; PMCID: PMC1213120.

Buchfink B, Reuter K, Drost HG, "Sensitive protein alignments at tree-of-life scale using DIAMOND", Nature Methods 18, 366–368 (2021). doi:10.1038/s41592-021-01101-x

Červenák F, Sepšiová R, Nosek J, Tomáška Ľ. Step-by-Step Evolution of Telomeres: Lessons from Yeasts. Genome Biol Evol. 2021 Feb 3;13(2):evaa268. doi: 10.1093/gbe/evaa268. PMID: 33537752; PMCID: PMC7857110.

de Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev. 2005 Sep 15;19(18):2100-10. doi: 10.1101/gad.1346005. PMID: 16166375.

Dietz S, Almeida MV, Nischwitz E, Schreier J, Viceconte N, Fradera-Sola A, Renz C, Ceron-Noriega A, Ulrich HD, Kappei D, Ketting RF, Butter F. The double-stranded DNA-binding proteins TEBP-1 and TEBP-2 form a telomeric complex with POT-1. Nat Commun. 2021 May 11;12(1):2668. doi: 10.1038/s41467-021-22861-2. PMID: 33976151; PMCID: PMC8113555.

Fulcher N, Riha K. Using Centromere Mediated Genome Elimination to Elucidate the Functional Redundancy of Candidate Telomere Binding Proteins in Arabidopsis thaliana. Front Genet. 2016 Jan 5;6:349. doi: 10.3389/fgene.2015.00349. PMID: 26779251; PMCID: PMC4700174.

Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. 2018 Jan;27(1):14-25. doi: 10.1002/pro.3235. Epub 2017 Sep 6. PMID: 28710774; PMCID: PMC5734306.

Im SH, Lee J. Identification of HMG-5 as a double-stranded telomeric DNA-binding protein in the nematode Caenorhabditis elegans. FEBS Lett. 2003 Nov 20;554(3):455-61. doi: 10.1016/s0014-5793(03)01191-8. PMID: 14623111.

Im SH, Lee J. PLP-1 binds nematode double-stranded telomeric DNA. Mol Cells. 2005 Oct 31;20(2):297-302. PMID: 16267406.

Joeng KS, Song EJ, Lee KJ, Lee J. Long lifespan in worms with long telomeric DNA. Nat Genet. 2004 Jun;36(6):607-11. doi: 10.1038/ng1356. Epub 2004 May 2. Erratum in: Nat Genet. 2005 Jan;37(1):106. PMID: 15122256.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583-589. doi: 10.1038/s41586-021-03819-2. Epub 2021 Jul 15. PMID: 34265844; PMCID: PMC8371605.

Kevin L. Howe, Bruce J. Bolt, Myriam Shafie, Paul Kersey, and Matthew Berriman. WormBase ParaSite − a comprehensive resource for helminth genomics, Molecular and Biochemical Parasitology 2017 215 2-10. 10.1016/j.molbiopara.2016.11.005

Kevin L. Howe, Bruce J. Bolt, Scott Cain, Juancarlos Chan, Wen J. Chen, Paul Davis, James Done, Thomas Down, Sibyl Gao, Christian Grove, Todd W. Harris, Ranjana Kishore, Raymond Lee, Jane Lomax, Yuling Li, Hans-Michael Muller, Cecilia Nakamura, Paulo Nuin, Michael Paulini, Daniela Raciti, Gary Schindelman, Eleanor Stanley, Mary Ann Tuli, Kimberly Van Auken, Daniel Wang, Xiaodong Wang, Gary Williams, Adam Wright, Karen Yook, Matthew Berriman, Paul Kersey, Tim Schedl, Lincoln Stein, and Paul W. Sternberg. WormBase 2016: expanding to enable helminth genomic research Nucleic Acids Research 2016 44(D1) D774-D780. 10.1093/nar/gkv1217

Kim SH, Hwang SB, Chung IK, Lee J. Sequence-specific binding to telomeric DNA by CEH-37, a homeodomain protein in the nematode Caenorhabditis elegans. J Biol Chem. 2003 Jul 25;278(30):28038-44. doi: 10.1074/jbc.M302192200. Epub 2003 Apr 23. PMID: 12711598.

Konig P, Giraldo R, Chapman L, Rhodes D. The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. Cell. 1996 Apr 5;85(1):125-36. doi: 10.1016/s0092-8674(00)81088-0. PMID: 8620531.

Kuznetsova V, Grozeva S, Gokhman V. Telomere structure in insects: A review. J Zool Syst Evol Res. 2020; 58: 127–158. https://doi.org/10.1111/jzs.12332

Lanjuin A, VanHoven MK, Bargmann CI, Thompson JK, Sengupta P. Otx/otd homeobox genes specify distinct sensory neuron identities in C. elegans. Dev Cell. 2003 Oct;5(4):621-33. doi: 10.1016/s1534-5807(03)00293-4. PMID: 14536063.

Li B, Espinal A, Cross GA. Trypanosome telomeres are protected by a homologue of mammalian TRF2. Mol Cell Biol. 2005 Jun;25(12):5011-21. doi: 10.1128/MCB.25.12.5011-5021.2005. PMID: 15923618; PMCID: PMC1140576.

Lim J, Kim W, Kim J, Lee J. Telomeric repeat evolution in the phylum Nematoda revealed by high-quality genome assemblies and subtelomere structures. Genome Res. 2023 Nov 2;33(11):1947–57. doi: 10.1101/gr.278124.123. Epub ahead of print. PMID: 37918961; PMCID: PMC10760449.

Liu F, Zhang H, Wang H, Zhu X, Li S, Jiang N, Yu C, Liu Y, Xiao Y. The homeodomain transcription factor CEH-37 regulates PMK-1/p38 MAPK pathway to protect against intestinal infection via the phosphatase VHP-1. Cell Mol Life Sci. 2023 Oct 5;80(11):312. doi: 10.1007/s00018-023-04970-x. PMID: 37796333; PMCID: PMC11072455.

Lu S et al. (2020), "The conserved domain database in 2020", Nucleic Acids Res.48(D)265-8.

Marchler-Bauer A et al. (2017), "CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.", Nucleic Acids Res.45(D)200-3.

Marquevielle J, De Rache A, Vialet B, Morvan E, Mergny JL, Amrane S. G-quadruplex structure of the C. elegans telomeric repeat: a two tetrads basket type conformation stabilized by a non-canonical C-T base-pair. Nucleic Acids Res. 2022 Jul 8;50(12):7134-7146. doi: 10.1093/nar/gkac523. PMID:

35736226; PMCID: PMC9262591.

McClintock, B.(1938). The fusion of broken ends of sister half-chromatids following chromatid breakage at meiotic anaphases. Missouri Agricultural Experiment Station Research Bulletin 290: 1–48.

Meier B, Barber LJ, Liu Y, Shtessel L, Boulton SJ, Gartner A, Ahmed S. The MRT-1 nuclease is required for DNA crosslink repair and telomerase activity in vivo in Caenorhabditis elegans. EMBO J. 2009 Nov 18;28(22):3549-63. doi: 10.1038/emboj.2009.278. Epub 2009 Sep 24. PMID: 19779462; PMCID: PMC2782091.

Muller HJ. The remaking of chromosomes. Collecting Net. 1938; 13:181-198

Moon S, Lee YW, Kim WT, Lee W. Solution structure of CEH-37 homeodomain of the nematode Caenorhabditis elegans. Biochem Biophys Res Commun. 2014 Jan 10;443(2):370-5. doi: 10.1016/j.bbrc.2013.11.133. Epub 2013 Dec 19. PMID: 24361878.

Moser BA, Nakamura TM. Protection and replication of telomeres in fission yeast. Biochem Cell Biol. 2009 Oct;87(5):747-58. doi: 10.1139/O09-037. PMID: 19898524; PMCID: PMC2854563.

Olovnikov AM. A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. J Theor Biol. 1973 Sep 14;41(1):181-90. doi: 10.1016/0022-5193(73)90198-7. PMID: 4754905.

Padmanaban S, Lambacher NJ, Tesmer VM, Zhang J, Shibuya H, Nandakumar J. Caenorhabditis elegans telomere-binding proteins TEBP-1 and TEBP-2 adapt the Myb module to dimerize and bind telomeric DNA. Proc Natl Acad Sci U S A. 2024 Apr 16;121(16):e2316651121. doi: 10.1073/pnas.2316651121. Epub 2024 Apr 8. PMID: 38588418; PMCID: PMC11032478.

Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 2021 Jan;30(1):70-82. doi: 10.1002/pro.3943. Epub 2020 Oct 22. PMID: 32881101; PMCID: PMC7737788.

Procházková Schrumpfová P, Schořová Š, Fajkus J. Telomere- and Telomerase-Associated Proteins and Their Functions in the Plant Cell. Front Plant Sci. 2016 Jun 28;7:851. doi: 10.3389/fpls.2016.00851. PMID: 27446102; PMCID: PMC4924339.

Prušáková D, Peska V, Pekár S, Bubeník M, Čížek L, Bezděk A, Čapková Frydrychová R. Telomeric DNA sequences in beetle taxa vary with species richness. Sci Rep. 2021 Jun 25;11(1):13319. doi: 10.1038/s41598-021-92705-y. PMID: 34172809; PMCID: PMC8233369.

Raffa GD, Ciapponi L, Cenci G, Gatti M. Terminin: a protein complex that mediates epigenetic maintenance of Drosophila telomeres. Nucleus. 2011 Sep-Oct;2(5):383-91. doi: 10.4161/nucl.2.5.17873. Epub 2011 Sep 1. PMID: 21989238.

Raices M, Verdun RE, Compton SA, Haggblom CI, Griffith JD, Dillin A, Karlseder J. C. elegans telomeres contain G-strand and C-strand overhangs that are bound by distinct proteins. Cell. 2008 Mar 7;132(5):745-57. doi: 10.1016/j.cell.2007.12.039. PMID: 18329362.

Saint-Leandre B, Levine MT. The Telomere Paradox: Stable Genome Preservation with Rapidly Evolving Proteins. Trends Genet. 2020 Apr;36(4):232-242. doi: 10.1016/j.tig.2020.01.007. Epub 2020 Feb 12. PMID: 32155445; PMCID: PMC7066039.

Sepsiova R, Necasova I, Willcox S, Prochazkova K, Gorilak P, Nosek J, Hofr C, Griffith JD, Tomaska L. Evolution of Telomeres in Schizosaccharomyces pombe and Its Possible Relationship to the Diversification of Telomere Binding Proteins. PLoS One. 2016 Apr 21;11(4):e0154225. doi:

10.1371/journal.pone.0154225. PMID: 27101289; PMCID: PMC4839565.

Sigrist CJ, De Castro E, Langendijk-Genevaux PS, Le Saux V, Bairoch A, Hulo N. ProRule: a new database containing functional and structural information on PROSITE profiles. Bioinformatics. 2005 Nov 1;21(21):4060-6. doi: 10.1093/bioinformatics/bti614. Epub 2005 Aug 9. PMID: 16091411.

Smith JS, Chen Q, Yatsunyk LA, Nicoludis JM, Garcia MS, Kranaster R, Balasubramanian S, Monchaud D, Teulade-Fichou MP, Abramowitz L, Schultz DC, Johnson FB. Rudimentary G-quadruplex-based telomere capping in Saccharomyces cerevisiae. Nat Struct Mol Biol. 2011 Apr;18(4):478-85. doi: 10.1038/nsmb.2033. Epub 2011 Mar 13. PMID: 21399640; PMCID: PMC3119813.

Steinberg-Neifach O, Lue NF. Telomere DNA recognition in Saccharomycotina yeast: potential lessons for the co-evolution of ssDNA and dsDNA-binding proteins and their target sites. Front Genet. 2015 May 1;6:162. doi: 10.3389/fgene.2015.00162. PMID: 25983743; PMCID: PMC4416457.

UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023 Jan 6;51(D1):D523-D531. doi: 10.1093/nar/gkac1052. PMID: 36408920; PMCID: PMC9825514.

Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Žídek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, Cowie A, Hobbs N, Kohli P, Kleywegt G, Birney E, Hassabis D, Velankar S. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022 Jan 7;50(D1):D439-D444. doi: 10.1093/nar/gkab1061. PMID: 34791371; PMCID: PMC8728224.

Wang J et al. (2023), "The conserved domain database in 2023", Nucleic Acids Res.51(D)384-8.

Watson JD. Origin of concatemeric T7 DNA. Nat New Biol. 1972 Oct 18;239(94):197-201. doi: 10.1038/newbio239197a0. PMID: 4507727.

Yamamoto I, Zhang K, Zhang J, Vorontsov E, Shibuya H. Telomeric double-strand DNA-binding proteins DTN-1 and DTN-2 ensure germline immortality in Caenorhabditis elegans. Elife. 2021 Jan 21;10:e64104. doi: 10.7554/eLife.64104. PMID: 33476260; PMCID: PMC7819708.

Yu EY, Kim SE, Kim JH, Ko JH, Cho MH, Chung IK. Sequence-specific DNA recognition by the Myb-like domain of plant telomeric protein RTBP1. J Biol Chem. 2000 Aug 4;275(31):24208-14. doi: 10.1074/jbc.M003250200. PMID: 10811811.

Yu X, Gray S, Ferreira HC. POT-3 preferentially binds the terminal DNA-repeat on the telomeric G-overhang. Nucleic Acids Res. 2023 Jan 25;51(2):610-618. doi: 10.1093/nar/gkac1203. Erratum in: Nucleic Acids Res. 2024 May 8;52(8):4759. doi: 10.1093/nar/gkae242. PMID: 36583365; PMCID: PMC9881156.

Zahler AM, Williamson JR, Cech TR, Prescott DM. Inhibition of telomerase by G-quartet DNA structures. Nature. 1991 Apr 25;350(6320):718-20. doi: 10.1038/350718a0. PMID: 2023635.

Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 2005 Apr 22;33(7):2302-9. doi: 10.1093/nar/gki524. PMID: 15849316; PMCID: PMC1084323.

Competing interest

The authors have declared no competing interest.

Acknowledgements

We are grateful to Dr. J. Lim providing the sequencing data of Panagrolaimidae family isolates (LJ2284, LJ2285, LJ2400), and Dr. M-A Félix for providing C. uteleia JU2585. We also thank the members of Dr. J. Lee (laboratory of genes and development) and Dr. H-J Choi (laboratory of structural biology) for helpful discussion during the study. This work was funded by the National Research Foundation of Korea [NRF-2020R1A2C3003352, NRF-2020R1C1C101220611].

Figures

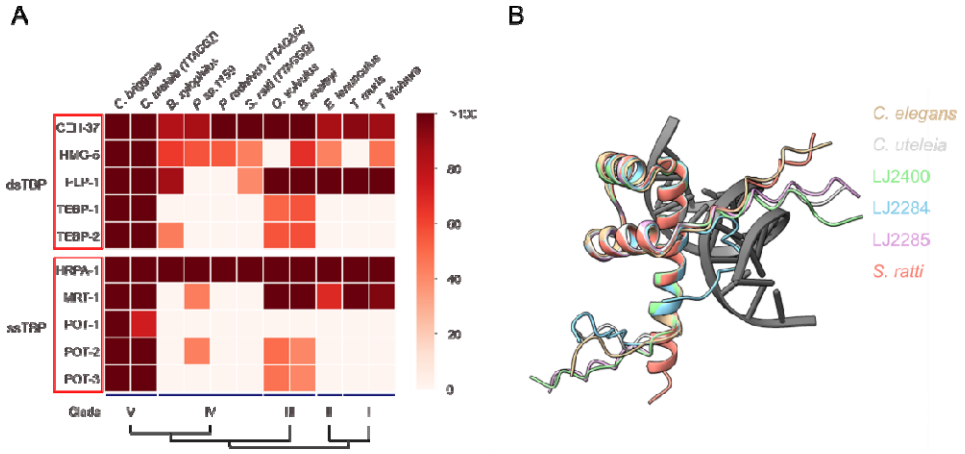


Figure 1. Conservation of telomere binding proteins in nematodes. (A) Heatmap illustrating the conservation of *C. elegans* TBPs in various nematodes according to the bit score. For nematodes with noncanonical TRMs, the TRMs are indicated in brackets after the name of the species. Nematodes with a canonical TTAGGC TRM are not explicitly labeled. (dsTBP: double-stranded telomeric DNA binding protein, ssTBP: single-stranded telomeric DNA binding protein.) (B) Model of the DNA-CEH-37 homeodomain complex of *C. elegans* with CEH-37 homeodomains from other nematodes aligned to it. The color of the species names corresponds to the color of the CEH-37 homeodomains.

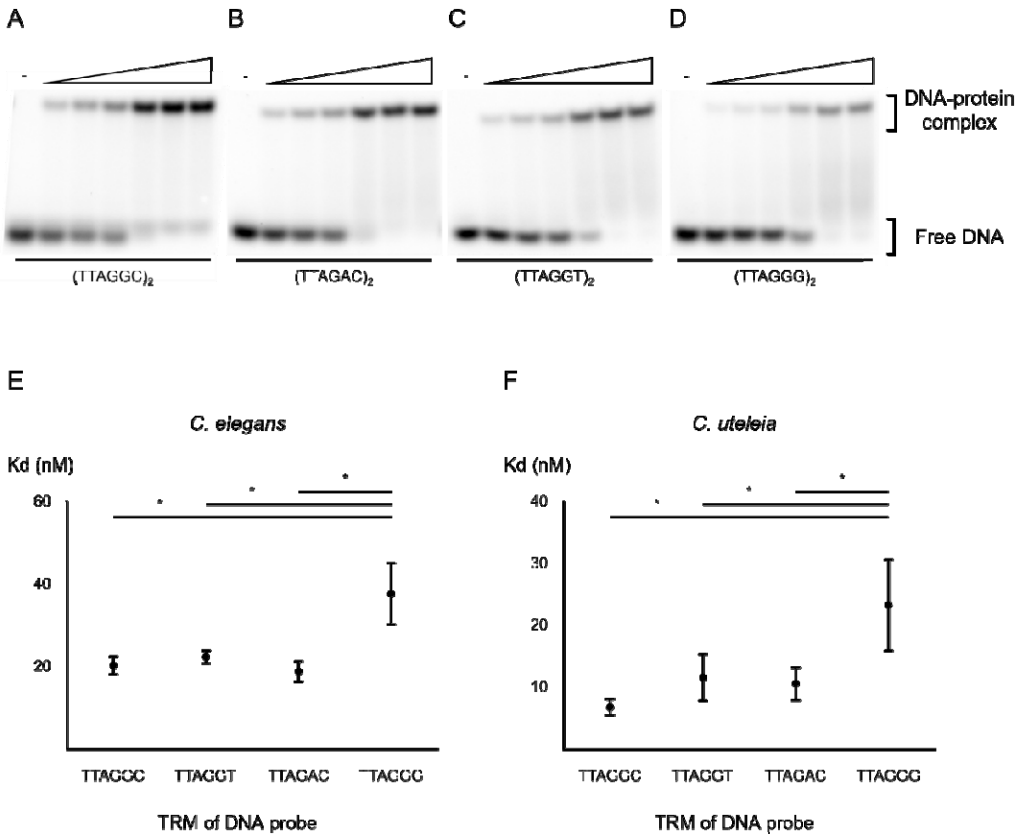


Figure 2. Binding affinity of CEH-37 homeodomains from the *Caenorhabditis* genus. (A-D) Representative images of EMSA demonstrating the binding of the *C. elegans* CEH-37 homeodomain to telomeric DNAs. 5 nM of DNA probes were incubated with increasing concentrations of proteins ranging from 5 nM to 160 nM. (E-F) Kd of *C. elegans* and *C. uteleia* CEH-37 homeodomains to each telomeric DNAs, respectively. DNA probes used for the Kd measurements consist of 2 repeats of the indicated sequences. Relationship without asterisks indicate non-significant differences (* P<0.05).

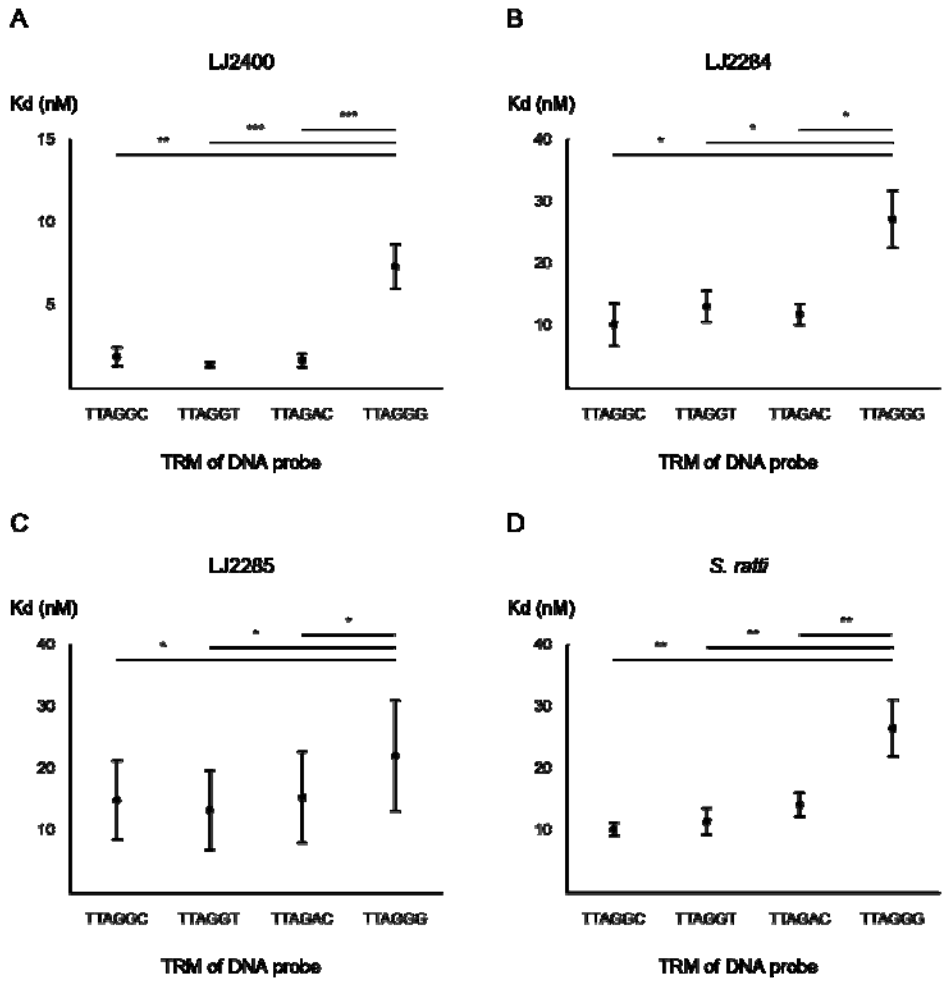


Figure 3. Binding affinity of CEH-37 homeodomains from clade IV nematodes. (A-D) Kd of CEH-37 homeodomains from LJ2400, LJ2284, LJ2285, and *S. ratti* binding to each telomeric DNA, respectively. DNA probes used for the Kd measurements consist of 2 repeats of the indicated sequences. Relationship without asterisks indicate non-significant differences (* 0.01<P<0.05, ** 0.005<P<0.01, *** P<0.005).

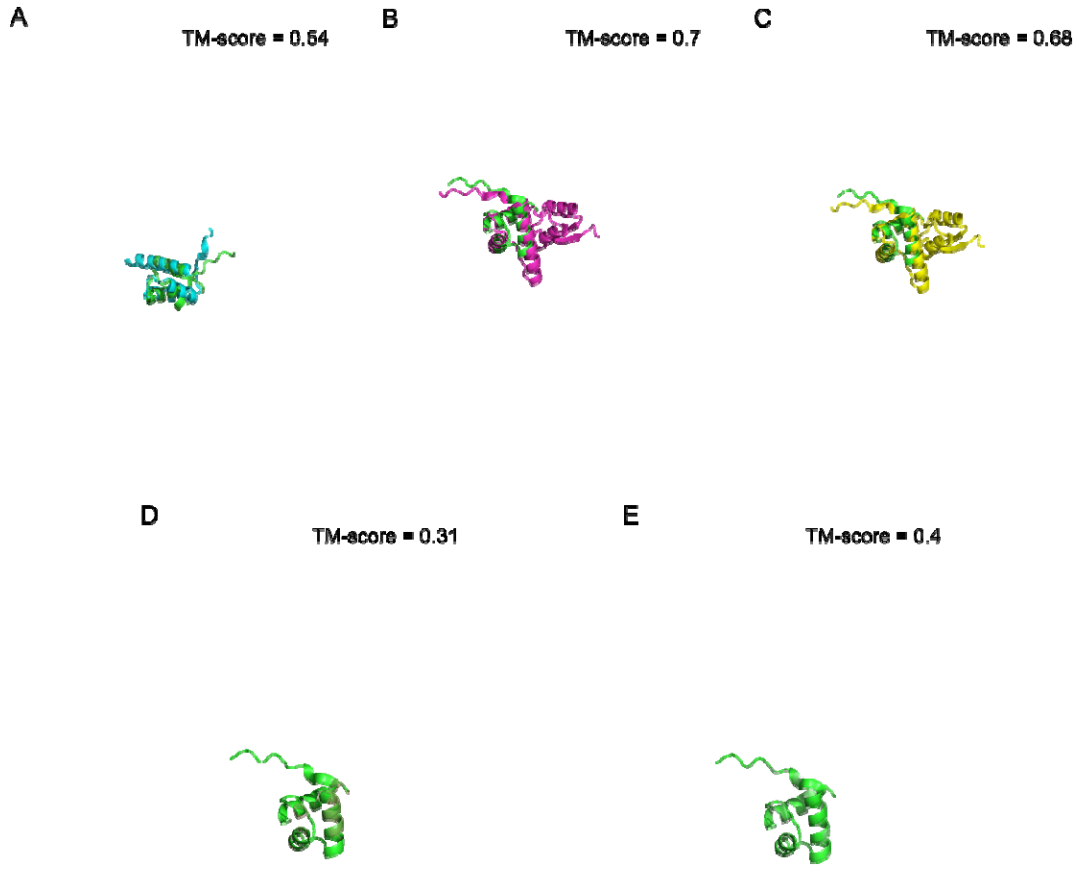
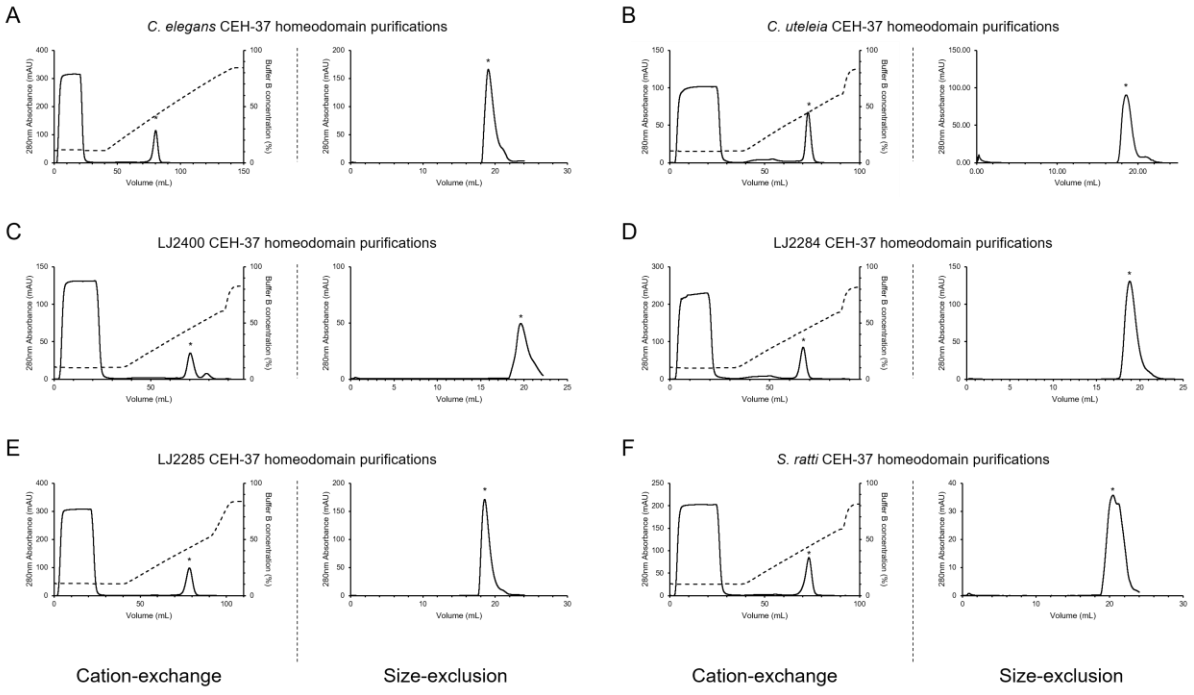
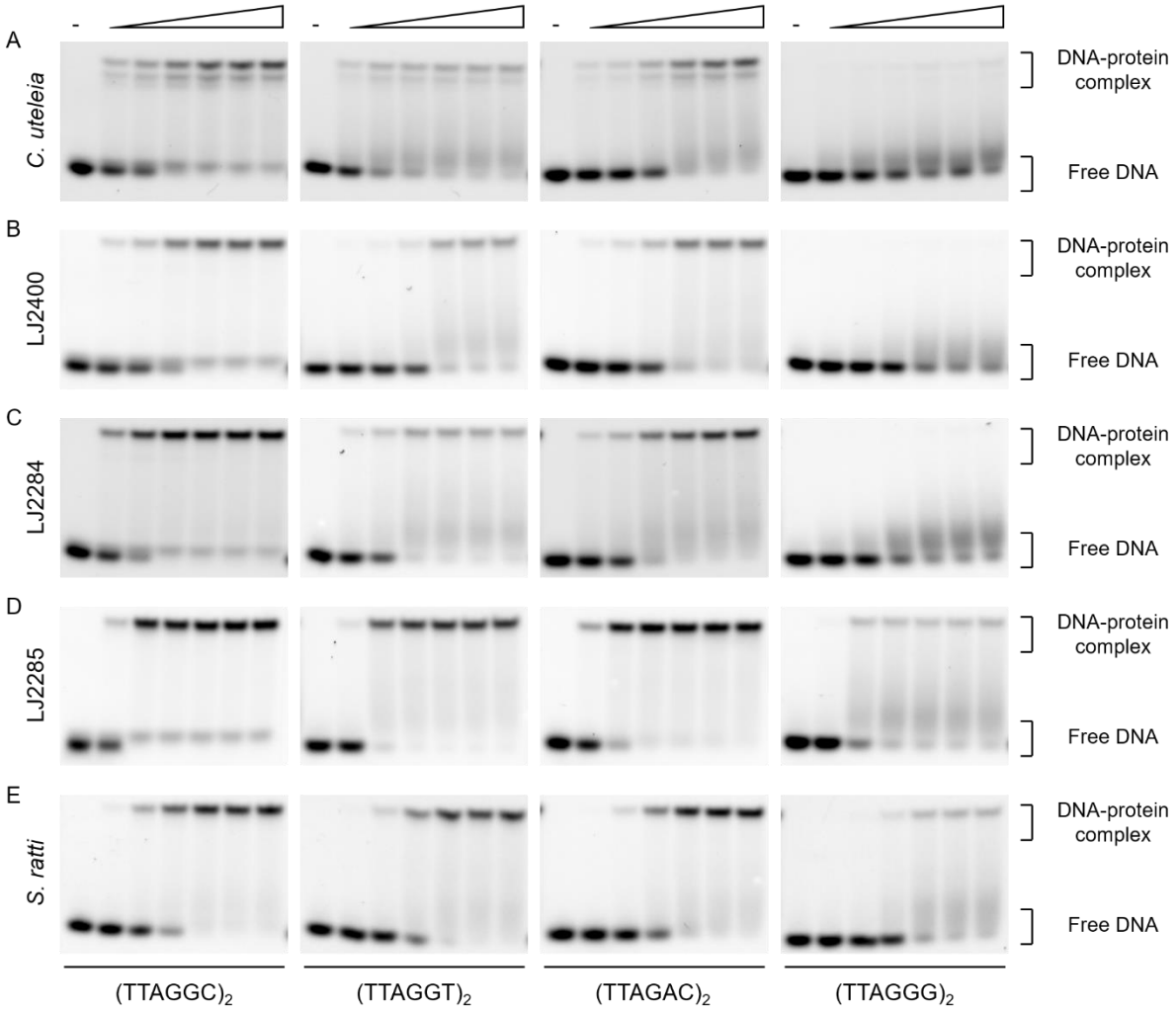


Figure 4. Models of a human TRF1 Myb domain aligned to *C. elegans* TBPs. The Myb domain of human TRF1 (green) is structurally aligned to (A) CEH-37 (cyan), (B) TEBP-1 (magenta), (C) TEBP-2 (yellow), (D) HMG-5 (salmon), and (E) PLP-1 (grey). The unaligned regions of each protein are faintly shown. TM-scores are 0.54, 0.7, 0.68, 0.31, and 0.4 respectively, where TM-score over 0.5 indicates a similar protein fold.

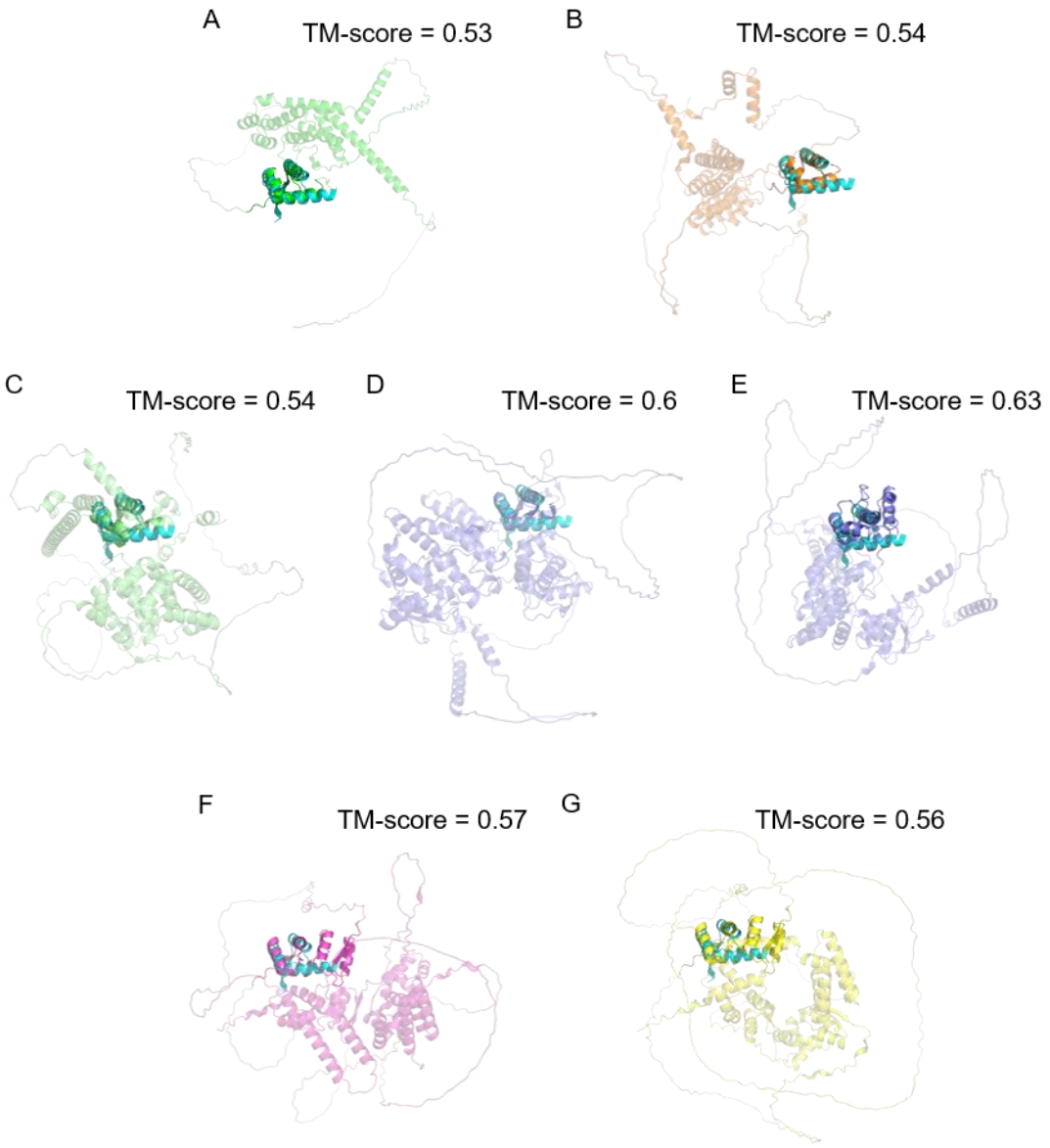
Supplementary figures



Supplementary figure 1. The purification of CEH-37 homeodomains. (A-F) After obtaining CEH-37 homeodomains using a GST-tag in a bacterial expression system, further purification was carried out sequentially using cation-exchange chromatography (left) and size-exclusion chromatography (right). Each graph illustrates the purification results of of CEH-37 homeodomains from (A) *C. elegans*, (B) *C. uteleia*, (C) LJ2400, (D) LJ2284, (E) LJ2285, and (F) *S. ratti,* respectively. The asterisks indicate the peaks of CEH-37 homeodomains.



Supplementary figure 2. Representative images of EMSA results of telomeric DNAs and CEH-37 homeodomains from various nematode species. (A-E) Representative EMSA images demonstrating the binding of the CEH-37 homeodomains from *C. uteleia*, LJ2400, LJ2284, LJ2285, and *S. ratti* to telomeric DNAs, respectively. DNA probes (5 nM) were incubated with increasing proteins concentrations ranging from 5 nM to 160 nM, except for LJ2400 CEH-37 homeodomain, whose concentration range was 0.625 nM to 20 nM.



Supplementary figure 3. Models of a CEH-37 homeodomain aligned to representative TBPs from various species. (A-E) The homeodomain of CEH-37 (cyan) is structurally aligned to (A-B) human TRF1 (green) and TRF2 (orange) Myb domains, (C) *S. pombe* Taz1 (lime) Myb domain, (D-E) *S. cerevisiae* RAP1 (blue) Myb domain and RAP1-DNA-bind domain (pfam09197), and (F-G) third Myb-containing domains of C. elegans TEBP-1 (magenta) and TEBP-2 (yellow) (Konig et al., 1996). The unaligned regions of each protein are faintly shown. TM-scores are 0.53, 0.54, 0.54, 0.6, 0.63, 0.57, and 0.56, respectively, where a TM-score over 0.5 indicates a similar protein fold.