Robust Binding Capability and Occasional Gene Loss of Telomere-Binding Proteins Underlying Telomere Evolution in Nematoda

Hobum Song [1], Seonhong Kim [2], Daisy Sunghee Lim [3], Hee-Jung Choi [2,*], Junho Lee [1,2,*]

[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

*Corresponding authors: E-mails: choihj@snu.ac.kr; elegans@snu.ac.kr.

Abstract

Telomeres, the nucleoprotein complexes that protect the ends of linear chromosomes, are essential for maintaining the stability of eukaryotic genomes. As telomeres generally consist of repetitive DNA associated with specifically bound proteins, telomeric repeat motifs are thought to be difficult to evolve. However, a recent study identified nematodes with telomeric repeats distinct from the canonical TTAGGC motif. Here, we investigated how telomere repeats could have evolved despite the challenge posed by the specificity of telomere-binding proteins (TBPs) to the telomeric DNA in Nematoda. We performed a phylogenetic analysis and electrophoresis mobility shift assays to assess the binding affinities of two TBPs, which displayed different conservation patterns. Our results revealed that the well-conserved protein CEH-37 exhibits limited specificity, unable to distinguish telomeric repeats found in nematodes except for the TTAGGG motif, while the less conserved POT proteins displayed rigid specificity. These findings suggest that the emergence of novel telomeric repeat motifs correlated with the characteristics and evolutionary outcomes of TBPs in Nematoda. Our study not only revealed the dynamics of telomere evolution but also enhanced the understanding of the evolutionary relationship between proteins and DNAs.

Key words: nematode, telomere-binding protein, telomere repeat motif, evolution, binding specificity.

Significance

Telomeres are chromosomal end structures composed of repetitive DNA where telomere-binding proteins (TBPs) specifically bind. This nucleoprotein complex is crucial for eukaryotes to maintain their genome integrity, making it difficult to evolve. However, a recent study identified nematodes with noncanonical telomeric repeat motifs (TRMs). Through an analysis of TBP conservation in Nematoda and assessments of their binding affinities, we observed that the binding capabilities of TBPs align with their conservation patterns. This correlation is also linked to the emergence of novel TRMs, explaining the evolutionary progress of TRMs in Nematoda. The evolutionary process discovered in this research is not limited to telomeres, providing further insights into evolutionary relationships between proteins and DNA.

Introduction

Genome stability in eukaryotes is inherently challenged by the continuous shortening of chromosomal ends and their misrecognition as double-strand breaks (Watson 1972; Olovnikov 1973; De Lange 2009). 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 motifs, 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 and Lue 2015).

Despite its challenging nature, telomeric sequence evolution has been observed in several lineages, including yeast, plants, and insects (Procházková Schrumpfová et al. 2016; Kuznetsova et al. 2020; Červenák et al. 2021). Various hypotheses have been proposed to explain how this process was possible, and some of these hypotheses have been experimentally validated in certain organisms (Steinberg-Neifach and Lue 2015; Saint-Leandre and Levine 2020). In yeast, the adoption of flexible 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 (Shakirov et al. 2009; Sepsiova et al. 2016; Červenák et al. 2021). In Animalia, however, no studies have been conducted to explain the relationship between TBPs and the evolution of telomeric sequences.

Nematodes are well-established model organisms with comprehensive databases and systems for research across diverse fields (Howe et al. 2016, 2017). Within the representative species Caenorhabditis elegans, research on telomeres has been conducted, which enabled the identification of C. elegans TBPs. While recent findings have confirmed that TEBP proteins, POT proteins, and MRT-1 interact to form telomeric protein complexes (Raices et al. 2008; Meier et al. 2009; Dietz et al. 2021; Yamamoto et al. 2021), the functions and molecular mechanisms of many C. elegans TBPs remain incompletely understood. For instance, while CEH-37, the first TBP identified in C. elegans, has demonstrated telomere-binding capacity and exhibits DNA-bending activity, functional studies using a deletion mutant have only suggested a potential role in maintaining chromosomal instability, based on a weak increase in the frequency of males (Kim et al. 2003). The presence of other TBPs with poorly defined functions, such as HMG-5, PLP-1, and HRPA-1, makes Nematoda an intriguing group for investigating their telomeres (Im and Lee 2003, 2005; Joeng et al. 2004; Yu et al. 2022).

Recently, novel TRMs have been discovered within the Nematoda phylum, which includes one of the representative model organisms Caenorhabditis elegans (Lim et al. 2023). While it has been well established that the nematode TRM is TTAGGC, some species, such as C. uteleia, Panagrellus redivivus and diverse isolates in Panagrolaimidae family (e.g. LJ2284, LJ2285), and Strongyloides ratti, have been revealed to possess noncanonical TRMs (TTAGGT, TTAGAC, and TTAGGG, respectively) by analysis of whole genome sequence (WGS) data (Fig. 1a). Given these findings, we aimed to understand how TRM evolved in Nematoda despite the challenges posed by sequence-specificities of TBPs. Based on our phylogenetic analysis, we focused on CEH-37, a well-conserved TBP, and less conserved POT proteins. The binding specificities of these proteins varied according to their conservation patterns—CEH-37 exhibited flexibility, while the POT proteins displayed relatively rigid specificity. These results indicate that the flexibility or the loss and gain of TBPs may coincide with the emergence of novel TRMs.

Results

Conservation of C. elegans TBPs in Nematoda

In eukaryotes, telomeric DNAs are capped by protein complexes (De Lange 2005; Moser and 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; Meier et al. 2009; Dietz et al. 2021; Yamamoto et al. 2021). However, there are some species in Nematoda found to possess noncanonical TRMs (Lim et al. 2023). 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.

As nematodes are classified into five clades based on their small rRNA phylogeny or orthology inference strategies (Blaxter et al. 1998; Smythe et al. 2019), we conducted BLAST analysis of TBPs from representative species of each clade, including the ones with noncanonical TRMs (Fig. 1a; supplementary table S1, Supplementary Material online). Since bit scores over 50-bits of BLASTP exhibit homology while bit scores below 40-bits are not statistically significant, we used a bit score threshold of 50-bits in BLASTP to indicate significant homology (Pearson 2013). Through the analysis, we observed that the TBPs of a clade V nematode species C. elegans are well conserved within clade V, while it was not the same in other clades. Notably, CEH-37 and HRPA-1 are well conserved across all clades of Nematoda. In contrast, the major subunits of the C. elegans telomeric protein complex, TEBP and POT
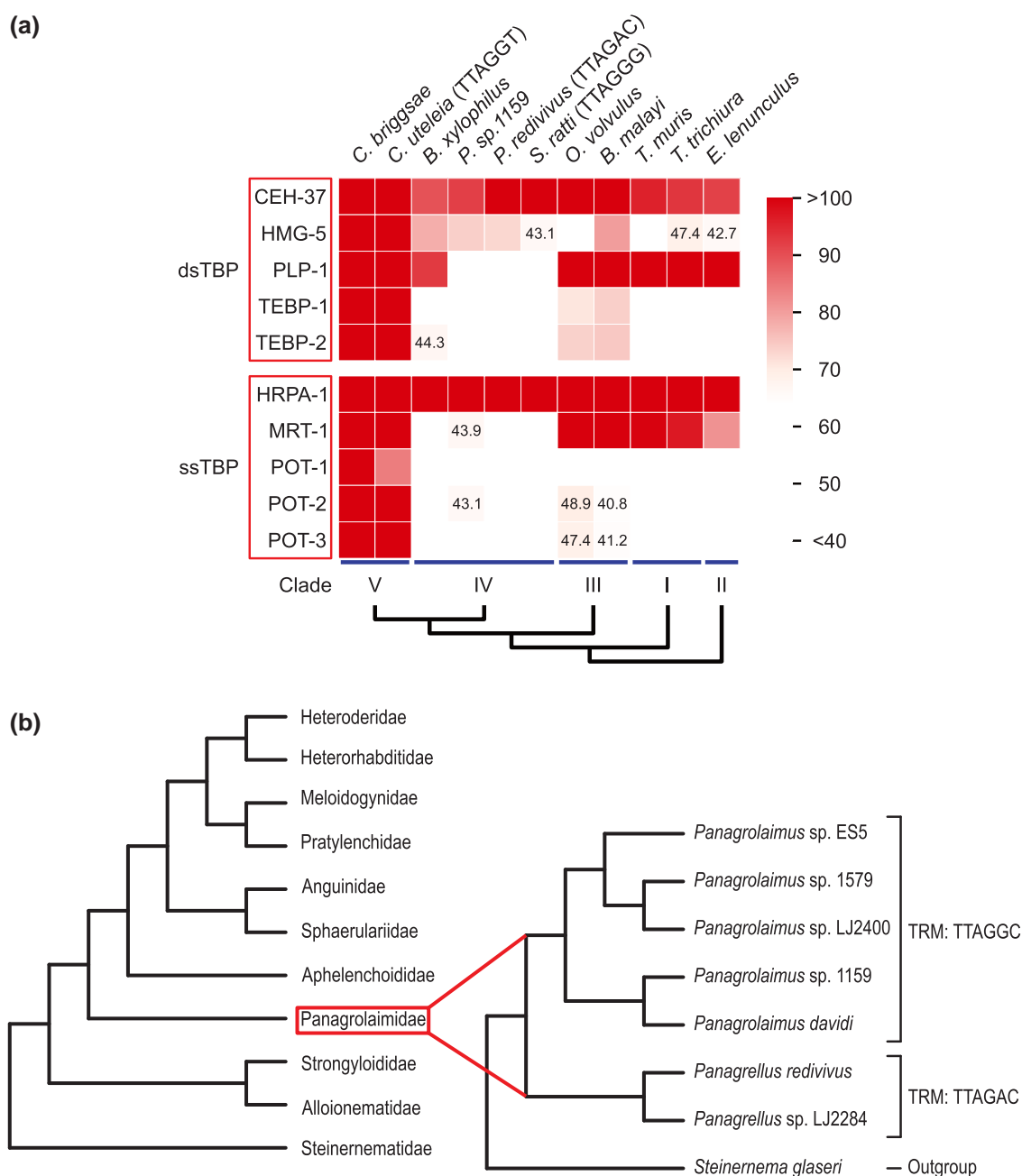
Fig. 1. Conservation of TBPs in nematodes and their phylogenetic relationships. a) Heatmap illustrating the conservation of *C. elegans* TBPs in various nematodes according to bit scores of BLASTP. For nematodes with noncanonical TRMs, the TRMs are indicated in brackets after the species names. Nematodes with a canonical TTAGGC TRM are not explicitly labeled. dsTBP, double-stranded telomeric DNA-binding protein; ssTBP, single-stranded telomeric DNA-binding protein. b) Phylogenetic tree of clade IV nematodes, shown at the family level. The phylogenetic tree of Panagrolaimidae, the only taxon that involves various species with a noncanonical TRM, is depicted in more detail, with *Steinernema glaseri* as an outgroup.

proteins, not seem to be conserved in clades I, II, and IV. This suggests that the species in clades other than III and V must possess a telomeric protein complex distinct from that of *C. elegans*. Our analysis shows that TBPs are varied in how conserved they are across Nematoda.

TTAGAC motif is the only noncanonical TRM that is identified in several species (Lim et al. 2023), and these

Panagrolaimidae species are involved in clade IV (Smythe et al. 2019; Lim et al. 2023) (Fig. 1b; supplementary table S2, Supplementary Material online). To examine the evolutionary sequence of the TRM and TBPs, we assessed the conservation of TEBP and POT proteins across all accessible clade IV nematodes using BLASTP via WormBase ParaSite (Howe et al. 2017). This analysis confirmed that TEBP and

POT proteins are not well conserved in clade IV nematodes, as no proteins from these species exhibited bit scores exceeding 50-bits (supplementary table S3, Supplementary Material online). These findings suggest that the evolution of TBPs may have preceded the alteration of TRMs.

Conserved Homeodomain of CEH-37 Robustly Binds to Telomeric DNAs With Limited Binding Specificity

CEH-37, the only well-conserved double-stranded telomeric DNA-binding protein (dsTBP) in Nematoda, was initially proposed as an evolutionarily conserved TBP due to the structural similarity of its homeodomain to dsTBPs in other species, which typically use Myb-related domains as DNA-binding domains (DBDs) (Kim et al. 2003; Moon et al. 2014). However, the discovery of TEBP proteins, which also possess Myb-related domains, raised doubts about this hypothesis. Notably, we observed that certain clades lack TEBP proteins and other subunits of the *C. elegans* telomeric protein complex. To investigate which TBPs might function like representative dsTBPs in other species, we examined structural homology between human TRF1 and *C. elegans* TBPs other than TEBP proteins, using TM-align (Zhang and Skolnick 2005). Given that a TM-score above 0.5 supports an evolutionary relationship, as demonstrated in Zhang and Skolnick (2005) and Xu and Zhang (2010), CEH-37 was the only dsTBP with a DBD structurally similar to the DBD of TRF1 (supplementary fig. S1 and table S4, Supplementary Material online). In addition, the CEH-37 homeodomain displayed structural similarity with other DBDs of representative dsTBPs such as *Schizosaccharomyces pombe* Taz1 and *Saccharomyces cerevisiae* RAP1 (supplementary table S4, Supplementary Material online). These unique characteristics of CEH-37 as a TBP raise questions about its role in the evolution of nematode telomeres.

To investigate how novel telomeric sequences emerged despite the conservation of CEH-37, we first compared sequences of the CEH-37 homeodomain in nematodes with different TRMs. The results indicated that CEH-37 appears to be well conserved at the sequence level, regardless of the TRM of the origin species (Fig. 2a). To further examine potential functional differences, we obtained the recombinant CEH-37 homeodomains from nematodes with different TRMs (supplementary fig. S2, Supplementary Material online) and measured their binding affinities to telomeric DNAs identified in Nematoda. First, we compared the binding affinities of CEH-37 homeodomains from *C. elegans* and *C. uteleia*, which have canonical TRM TTAGGC and noncanonical TRM TTAGGT, respectively, using electrophoresis mobility shift assay (EMSA) (Fig. 2b to d; supplementary tables S5 and S6, Supplementary Material online). Interestingly, CEH-37 homeodomains from both species bound to telomeric DNAs with TTAGGC, TTAGAC, and TTAGGT repeats, with no significant difference in their dissociation constants (K_d). However, both CEH-37 homeodomains did not effectively bind to DNA with TTAGGG repeats. These results indicate that CEH-37 does not distinguish TTAGGC, TTAGAC, and TTAGGT motifs, while it shows less preference for the TTAGGG motif.

We also measured the binding affinities of CEH-37 homeodomains from strains in clade IV, where species with other noncanonical TRMs were identified (Figs. 1b and 3; supplementary tables S5 and S6, Supplementary Material online). The Panagrolaimidae family in clade IV includes isolate LJ2284, which possesses noncanonical TRM TTAGAC, and a related species LJ2400, which has canonical TRM TTAGGC. Measurement of K_d values of CEH-37 homeodomains from LJ2284 and LJ2400 showed consistent results with the above experiments, indicating robust binding to telomeric DNA with limited sequence specificity, irrespective of the TRM of their origin species. These results suggest that TRMs in Nematoda evolved within the range of the well-conserved CEH-37 sequence specificity.

As our results show that CEH-37 does not effectively bind to telomeric DNA with TTAGGG repeats, we measured the binding affinities of CEH-37 homeodomains from *S. ratti*, the parasitic nematode in clade IV expected to possess noncanonical TRM TTAGGG (Lim et al. 2023). Interestingly, the CEH-37 homeodomain from *S. ratti* also did not bind efficiently to TTAGGG-repeated DNA compared with other nematode telomeric DNAs (Fig. 3a and d). Contrary to a previous study, this result suggests that *S. ratti* may not have TTAGGG as its TRM, or it may rely on a different protein instead of CEH-37.

Nematodes With Novel TRMs Need Alternative Mechanisms to Protect Their G-Rich Telomeric Overhangs Other Than Capping by POT Proteins

TEBP and POT proteins form a telomeric protein complex, which is essential for telomere protection in *C. elegans* (Dietz et al. 2021; Yamamoto et al. 2021). Interestingly, while these proteins are well conserved in *C. uteleia*, a clade V nematode species with distinct TTAGGT TRM, they are not conserved in clade IV nematodes, which include other noncanonical TRMs (Fig. 1a; supplementary tables S1 and S3, Supplementary Material online). To investigate how these proteins contributed to the evolution of TRMs, we obtained *C. elegans* POT proteins, POT-2 and POT-3, and assessed their binding affinities to nematode telomeric DNAs (Fig. 4a; supplementary fig. S3a to c and table S7, Supplementary Material online). The EMSA results indicate that both POT-2 and POT-3 bind efficiently to TTAGGC repeats. However, they did not bind to the DNA probe composed of TTAGGT repeats or TTAGAC repeats. These results indicate that POT proteins exhibit high sequence
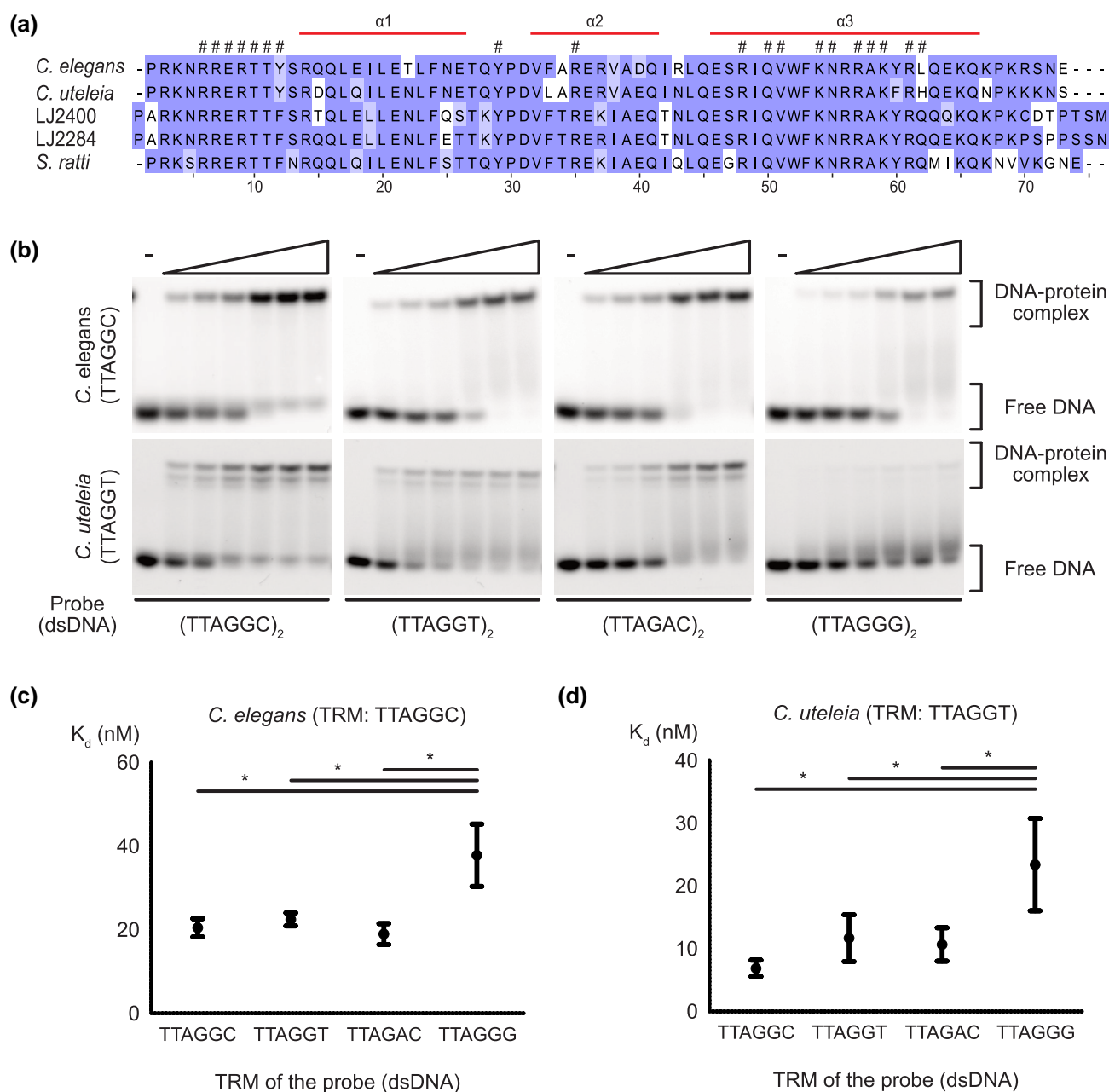
Fig. 2. Conservation of CEH-37 homeodomains from species with different telomeric repeat motifs and their binding affinities from *Caenorhabditis* species. a) The multiple sequence alignment result of CEH-37 homeodomains from nematodes used in this study. Each residue is shaded according to its degree of conservation. The α-helix regions of the domain are indicated above the multiple sequence alignment, and conserved residues involved in nucleic acid binding are indicated with hash-marks (#). b) Representative images of EMSA demonstrating the binding of the *C. elegans* and *C. uteleia* CEH-37 homeodomains to double-stranded telomeric DNAs. Five nanomolars of DNA probes indicated under each figure were incubated with increasing concentrations of proteins ranging from 5 to 160 nM. TRMs of each species are shown under their names. c and d) K_d of *C. elegans* and *C. uteleia* CEH-37 homeodomains to each telomeric DNA, respectively. DNA probes used for the K_d measurements consist of two repeats of the indicated sequences. Relationships without asterisks indicate nonsignificant differences (Newman–Keuls multiple comparison test, *$P < 0.05$).

specificity in binding telomeric DNA, failing to bind to the novel nematode telomeric repeats.

For the emergence of novel TRMs, POT proteins must either coevolve or be lost. As the results of TBP conservation indicate, POT proteins are not conserved in clade IV nematodes, suggesting that the TTAGAC TRM in clade IV arose from the loss of POT proteins (Fig. 1a; supplementary tables S1 and S3, Supplementary Material online). However, interpreting protein loss based solely on BLASTP can be ambiguous, especially for proteins with bit
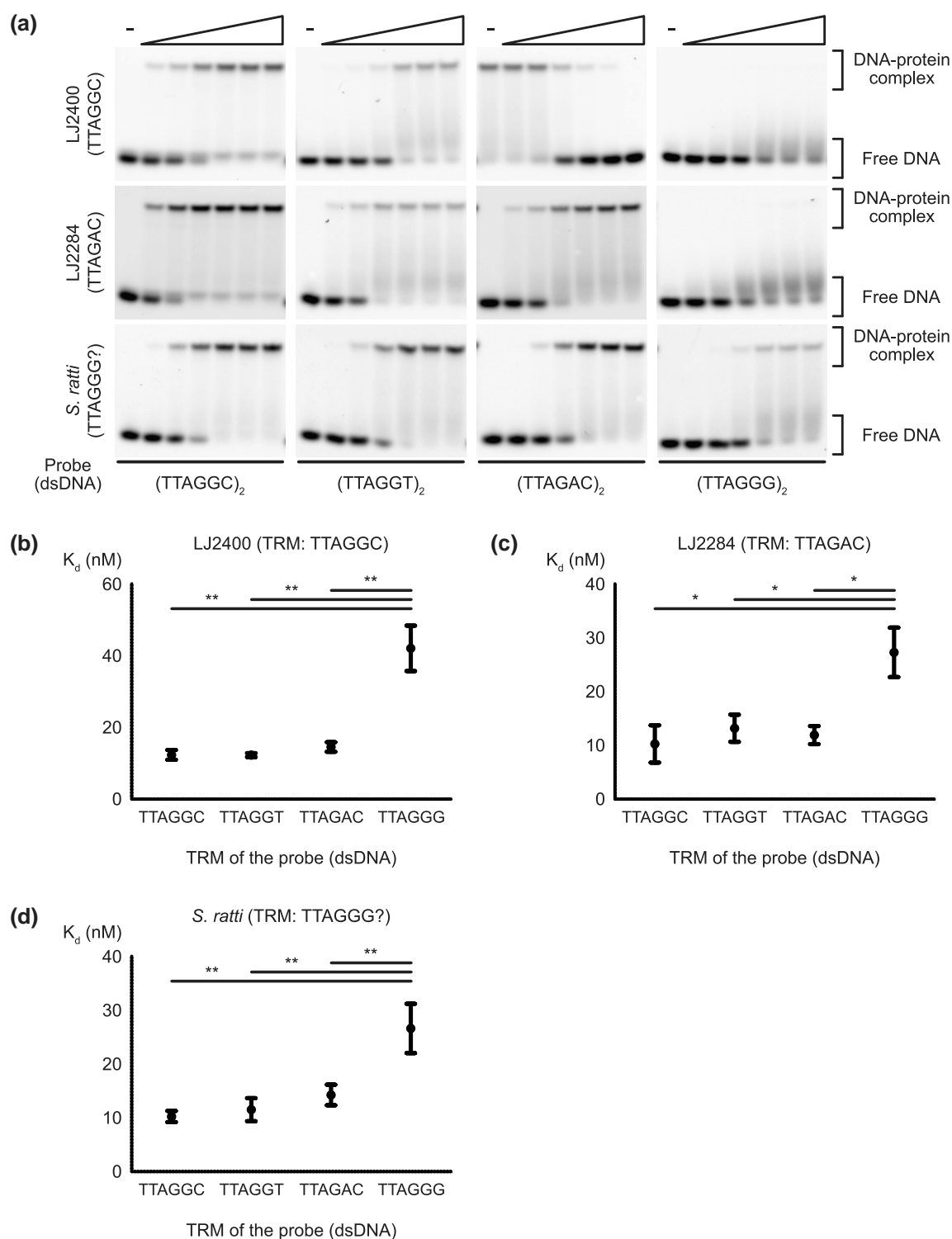
Fig. 3. Binding affinity of CEH-37 homeodomains from clade IV nematodes. a) Representative images of EMSA demonstrating the binding of the LJ2400, LJ2284, and *S. ratti* CEH-37 homeodomains to double-stranded telomeric DNAs. Five nanomolars of DNA probes indicated under each figure were incubated with increasing concentrations of proteins ranging from 5 to 160 nM for LJ2284 and *S. ratti* CEH-37 homeodomains, while the concentration range of the LJ2400 protein was 2.8 to 90 nM. TRMs of each species are shown under their names. b to d) K_d of CEH-37 homeodomains from LJ2400, LJ2284, and *S. ratti* binding to each telomeric DNA, respectively. DNA probes used for the K_d measurements consist of two repeats of the indicated sequences. Relationships without asterisks indicate nonsignificant differences (Newman–Keuls multiple comparison test, $*0.01 < P < 0.05$, $**0.005 < P < 0.01$, $***P < 0.005$).
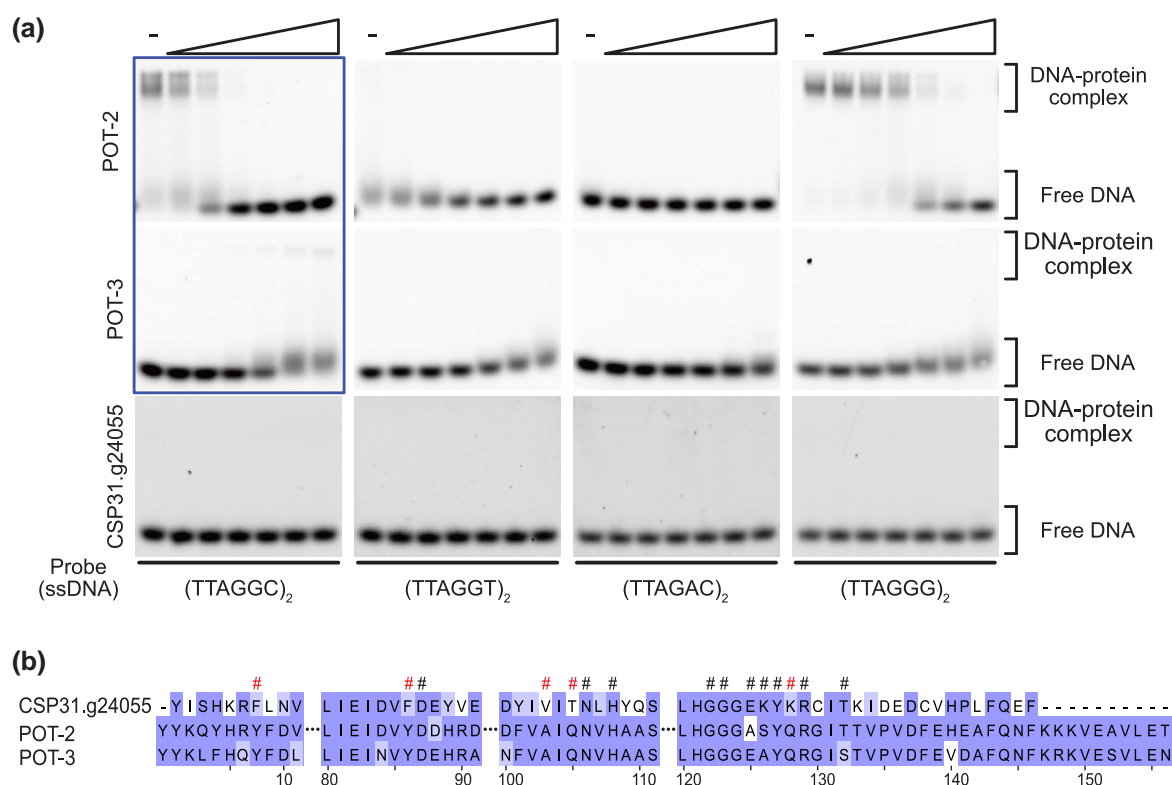
Fig. 4. Binding affinity of *C. elegans* and *C. uteleia* POT proteins and their sequence alignment. a) Representative images of EMSA demonstrating the binding of the *C. elegans* POT-2, POT-3, and *C. uteleia* CSP31.g24055 to single-stranded telomeric DNAs. Five nanomolars of DNA probes were incubated with increasing concentrations of proteins ranging from 20 to 640 nM except for the reaction by POT-2 and POT-3 to (TTAGGC)$_2$ probe, where the protein range was 5 to 160 nM (indicated with a box). b) The multiple sequence alignment result of OB-fold domains in POT proteins from *C. uteleia* and *C. elegans*. Each residue is shaded according to its degree of conservation. Conserved residues involved in nucleic acid binding are indicated with hash-marks (#), and the marks of residues not conserved between POT proteins from *C. uteleia* and *C. elegans* are colored in red.

scores between 40 and 50 (Pearson 2013). To address this, we further examined the presence of *C. elegans* TBP by searching for their orthologues in clade IV nematodes using available data from WormBase ParaSite (supplementary table S8, Supplementary Material online) (Howe et al. 2017). This analysis confirmed that POT and TEBP proteins are not conserved in any clade IV nematodes. These results suggest that the loss of *pot* genes preceded the emergence of the noncanonical TTAGAC TRM in clade IV. Similar to the earlier case of CEH-37, the loss of POT protein may have reduced selective pressure against the TTAGAC TRM, facilitating the emergence of the noncanonical TRM in Panagrolaimidae isolates.

While the emergence of the TTAGAC TRM can be attributed to the loss of major subunits of the telomeric protein complex, the emergence of the TTAGGT TRM in *C. uteleia* still raises the possibility of coevolution. Unlike clade IV nematodes, BLASTP analysis has shown that *C. uteleia* retains homologs of *C. elegans* POT-1 (CSP31.g1277) and POT-2 and POT-3 (CSP31.g24055). Since CSP31.g24055 (bit score = 137) showed a higher bit score with *C. elegans* POT-2 than CSP31.g1277 (bit score = 73.6) did with POT-1, we purified recombinant CSP31.g24055 to test whether it binds to its host telomeric DNA (supplementary fig. S3d and e, Supplementary Material online). Surprisingly, CSP31.g24055 failed to bind to any nematode telomeric DNAs (Fig. 4a), while it minimally bound regardless of the probe sequence in extreme conditions such as binding reaction at 37 °C, overnight (supplementary fig. S4, Supplementary Material online), indicating that it cannot function as a TBP. To investigate the differences between *C. uteleia* CSP31.g24055 and *C. elegans* POT proteins, we aligned the sequences of their DNA-binding domain, oligonucleotide/oligosaccharide-binding (OB)-fold. The alignment revealed that CSP31.g24055 exhibits alterations in DNA-binding sites, as well as a shorter C-terminal region (Fig. 4b). These differences might explain why CSP31.g24055 is unable to bind to telomeric repeats. Meanwhile, since the binding affinity of the *C. uteleia* POT-1 homolog CSP31.g1277 to telomeric DNAs has not yet been tested, it remains possible that *C. uteleia* possesses C-rich overhangs where CSP31.g1277 may bind with its well-conserved OB-fold domain (supplementary fig. S5,

Supplementary Material online), or alternative proteins or mechanisms are involved in protecting G-rich telomeric overhangs.

Discussion

Chromosome ends are protected by the intricate interaction between sequence-specific TBP and telomeric DNA. However, the relationship between the evolution of TRM and TBPs in animals has been elusive. In this study, we examined how TRMs could evolve in the challenge posed by the sequence specificity of TBPs in Nematoda. Phylogenetic analysis and measurements of binding affinities of CEH-37 and POT proteins suggested that the presence of TBPs with their sequence specificity exerts selective forces on TRM evolution.

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 heterogeneous TRMs. This extreme diversity accompanied the adoption of flexible TBPs, such as Rap1 and Taz1, instead of the strictly sequence-specific Teb1/Tay1 (Moser and Nakamura 2009; Sepsiova et al. 2016). As the new TBP emerged (i.e. Taz1) or the telomere-associated protein became a multifunctional TBP by gaining the ability to bind DNA directly (i.e. Rap1), the original TBPs either became dedicated to their secondary role as transcription factors or disappeared. The evolution of telomeres in yeasts, where TBPs have both telomeric functions and roles as transcription factors, is reminiscent of the nematode CEH-37, which is also known for its roles as a transcription factor involved in sensory neuron specification and intestinal immunity in C. elegans (Shore and Nasmyth 1987; Lanjuin et al. 2003; Liu et al. 2023).

Along with these findings, it is plausible that CEH-37 served as a main dsTBP in species lacking TEBP proteins and other C. elegans telomeric protein complex subunits. Prior to the discovery of TEBP proteins, CEH-37 was speculated to be a functionally conserved dsTBP based on its structural similarity to Myb-related domains of dsTBPs in other species, such as human TRF1 or S. pombe Taz1 (Kim et al. 2003; Moon et al. 2014). TBP conservation analysis in our study supports this view, as CEH-37 is the only TBP with this structural similarity in clades where TEBP proteins are not conserved (supplementary fig. S1 and table S4, Supplementary Material online). Therefore, we could cautiously hypothesize that CEH-37 is a result of convergent evolution that may have served as a main dsTBP in nematodes, which functions in species lacking TEBP proteins.

The emergence of the TTAGAC TRM in clade IV nematodes can be attributed not only to the flexibility of CEH-37 but also to the absence of specifically binding POT proteins, which hardly bind to TTAGAC repeats.

Unlike most eukaryotes, which possess OB-fold POT proteins to protect their telomeric overhangs, clade IV nematodes lack these POT proteins (Fig. 1a; supplementary tables S1, S3, and S8, Supplementary Material online). Clade IV nematodes likely evolved adaptations to compensate for the loss of POT proteins. Various strategies for protecting single-stranded telomeric DNA in the absence of POT proteins have been observed in other organisms. In budding yeast, telomeric overhangs are capped by the Cdc13-Stn1-Ten complex in place of POT proteins (Moser and Nakamura 2009). In plants, single-stranded telomeric DNA-binding proteins (ssTBPs) with or without OB-fold domains have been identified, while some species possess blunt-ended telomeres (Kazda et al. 2012; Luo et al. 2020). These precedents suggest that investigating how nematodes adapt to the loss of POT proteins could be a compelling area for further study, with the potential to discover new mechanisms.

Given that the nematode telomeric protein complex is known to comprise TEBP and POT proteins (Dietz et al. 2021; Yamamoto et al. 2021), it was surprising that a homolog of the 3′ telomeric overhang-binding POT proteins from C. uteleia, a unique clade V nematode with a distinct TTAGGT TRM, hardly binds to telomeric DNA (Fig. 4a). Since the 3′ telomeric overhang-binding POT proteins do not directly associate with the TEBP proteins in C. elegans, their functional significance might be reduced in the complex. Consequently, C. uteleia may employ other proteins to protect its 3′ telomeric overhangs or lack such overhangs. In contrast, C. elegans POT-1, which binds to the 5′ telomeric overhangs, directly interacts with TEBP proteins. Therefore, it is plausible that CSP31.g1277, the C. uteleia homolog of POT-1, binds to the novel telomeric DNA. This could have occurred either through the flexible binding specificity of C. elegans POT-1 or via the co-evolution of CSP31.g1277 within the emergence of the novel TTAGGT TRM. Evidence supporting the former possibility comes from the case of CEH-37 in this study, as well as the flexibility observed in yeast RAP1 and Taz1 (Moser and Nakamura 2009; Sepsiova et al. 2016). The latter one is supported by studies of plant evolution, where POT1 in green algae only binds to TTTAGGG repeats, while the POT1 of land plants possessing the TTAGGG TRM binds to both TTAGGG and TTTAGGG repeats (Shakirov et al. 2009). Identifying the mechanism by which the C. uteleia TRM arose would be an intriguing study, as would investigating the functions of CSP31.g24055 and the reason for its functional alteration.

We confirmed that CEH-37 homeodomains from every species, including S. ratti, do not effectively bind to TTAGGG repeats compared with other telomeric DNAs found in Nematoda, raising questions about the nature of the telomere in S. ratti. Previous WGS data speculated that S. ratti possesses a TTAGGG TRM, as this was the

only TTAGGC-related motif identified in the species (Lim et al. 2023). However, TRM-containing concatemers were rare compared to TRM-containing reads from WGS data of other nematodes analyzed in the study, ranking 67th out of 67 species. Based on this information, we hypothesize that (1) *S. ratti* possesses short telomeric DNAs consisting of TTAGGG repeats, possibly protected by TBPs other than CEH-37 or by other mechanisms (e.g. adoption of rigid G-quadruplex (G4) structures), or that (2) *S. ratti* does not possess TTAGGG TRM, and the observed reads from the parasite may have originated from its mammalian host (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 are unlikely to form telomeric G4 structures (Marquevielle et al. 2022).

By studying nematode telomere evolution using *C. elegans* TBPs as a reference, we revealed the background underlying telomere evolution at a molecular level. However, our understanding of species-specific telomere biology remains limited. Dietz et al. successfully identified TBPs in *C. elegans* using a DNA pull-down assay, and this approach could be applied to identify TBPs in other nematode species as well. Based on these results, comparing TBPs across nematode taxa with different TRMs and life histories will provide a more comprehensive understanding of nematode telomere evolution.

In summary, we demonstrated the dynamics of nematode telomere evolution and highlighted the role of TBPs in the evolution of telomere sequences across different nematode species. We expect our study to not only advance the field of telomere biology using nematodes as multicellular organism models but also enhance our understanding of the evolution of DNA–protein interactions beyond telomeres.

Methods

Strains

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

Conservation Analysis of TBPs

FASTA-format protein sequences of *C. elegans* (PRJNA13758), *C. briggsae* (PRJNA10731), *C. uteleia* (PRJEB12600), *Bursaphelenchus xylophilus* (PRJEA64437), *Panagrolaimus* sp. 1159 (PRJEB32708), *P. redivivus* (PRJNA186477), *S. ratti* (PRJEB125), *Onchocerca volvulus* (PRJEB513), *Brugia malayi*

(PRJNA10729), *Trichuris muris* (PRJEB126), *T. trichiura* (PRJEB535), and *Enoplolaimus lenunculus* (PRJNA953805) were obtained from WormBase ParaSite (Howe et al. 2016, 2017). Protein sequences of Panagrolaimidae isolates LJ2284 and LJ2400 (PRJNA845886), which were generated using BRAKER, were kindly provided by Lim et al. (2023).

To analyze the conservation of TBPs 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 domains of proteins were identified using NCBI CD-search (CDD database, version 3.21) (Marchler-Bauer et al. 2017; Lu et al. 2020; Wang et al. 2022). Multiple sequence alignments were performed by Clustal Omega (sequence type: Protein; output format: PHYLIP) and were visualized using Jalview 2.11.4 with the BLOSUM 62 color scheme (Waterhouse et al. 2009; Madeira et al. 2024). Additionally, BLAST+ provided by WormBase ParaSite on its web server or DIAMOND as mentioned above was used to analyze the conservation of TEBP-POT complex subunits in clade IV nematodes (Camacho et al. 2009). Query sequences were used as documented in supplementary table S3, Supplementary Material online, and BLAST was searched against the clade IV nematode protein database.

Phylogenetic Tree Construction

The cladogram of clade IV nematode families was generated based on the phylogeny tree presented in Smythe et al. (2019). To construct the tree of the Panagrolaimidae family, 18s rDNA sequences reported in Lim et al. (2023), along with the publicly available 18s rDNA sequence of *S. glaseri* (GenBank accession: AY284682.1) were aligned using Clustal Omega (sequence type: DNA; output format: PHYLIP) (Madeira et al. 2024). The resulting alignment was used as input for RaxML (version 8.2.12) via raxmlGUI 2.0 (version 2.0.13; analysis: ML+ rapid bootstrap; replications: 1,000; substitution model: GTRGAMMA) (Edler et al. 2021). The tree was visualized using Dendroscope (version 3.8.10) (Huson and Scornavacca 2012).

Construct Generation

ceh-37 homeobox regions of each strain were cloned in a modified pGEX-4T-1 vector containing an HRV 3C cleavage site after the GST tag, while *C. elegans pot-2* and *pot-3* were cloned in a modified pMJ806 vector containing a TEV cleavage site after the MBP tag. To generate cDNA of desired genes, RNAs isolated from nematodes with TRIzol were reverse transcribed with PrimeScript RT Master Mix (Takara) except for *S. ratti*. Since we were not able to culture *S. ratti*, *ceh-37* homeobox cDNA of *S. ratti* was generated by ligating two annealed ~110 bp oligos after

phosphorylation. cDNAs were fused into the vector by Gibson assembly master mix (New England Biolabs).

Protein Expression and Purification

The plasmid containing the desired gene was transformed into *E. coli* Rosetta (DE3) cells. Colonies were picked and inoculated into Terrific Broth, and the culture was grown at 37 °C until the OD_{600} reached approximately 0.6. Protein overexpression was induced by adding IPTG to a final concentration of 0.2 mM, followed by incubation at 20 °C for 18 h. The cells were harvested and resuspended in PBS, and cell lysis was performed using sonication, supplemented with PMSF and DNase I. The lysed cells were centrifuged at 14,000 × *g* for 15 min, and the supernatants were collected. The supernatants from cells expressing CEH-37 or POT proteins were incubated with glutathione agarose resin or amylose agarose resin, respectively, pre-equilibrated with PBS for 1 h with gentle rolling. The resin was washed with 10 column volumes of 20 mM HEPES pH 7.5, 100 mM NaCl buffer. While the proteins were bound to the resin, HRV3C or TEV protease was added for overnight digestion at 4 °C. The following day, the protein was released using the wash buffer, and subsequently loaded onto a HiTrap SP (CEH-37)/Q (POT proteins) HP 5 mL column (GE Healthcare). An appropriate salt gradient was applied using 20 mM HEPES pH 7.5 buffer, and peaks were analyzed by SDS-PAGE. The desired fractions were pooled and concentrated. Final purification was performed by loading the concentrated protein onto a Superdex 200 10/300 GL column (GE Healthcare) equilibrated with 20 mM HEPES pH 7.5, 150 mM NaCl. The peaks were analyzed by SDS-PAGE, and the desired fractions were collected and concentrated to the appropriate concentration. The concentrated protein solution was aliquoted, flash-frozen in liquid nitrogen, and stored at −80 °C.

During the purification of *C. uteleia* protein CSP31.g24055, the protein precipitated after protease treatment due to the minimal difference between the isoelectric point (7.93) of tag-free CSP31.g24055 and the pH of the buffer (7.5). To address this, we initially purified His-MBP-tagged CSP31.g24055 using a HiTrap Q HP 5 mL column followed by a Superdex 200 10/300 GL column. Subsequently, the buffer was exchanged to 20 mM HEPES pH 7, 150 mM NaCl using dialysis tubing, and His-tagged TEV protease was added for cleavage at 4 °C overnight. The tag and protease were removed by Ni-NTA resin (Qiagen), and the supernatant containing purified CSP31.g24055 was collected for further experiments.

Electrophoretic Mobility Shift Assay

One hundred micromolar stocks of DNA oligonucleotides in TE buffer (1 mM EDTA, 10 mM Tris-HCl pH 8.0) were used to prepare DNA probes. To generate 40 μM stocks of double-stranded DNA probes, forward strand oligos labeled with Cy5 fluorophore at the 5′ end were mixed with reverse strand oligos at a 1:1.2 ratio in the additional TE buffer. The solutions were heated to 95 °C for 5 min and then cooled to room temperature at a rate of 1 °C/min over 70 min. Single-stranded DNA probes were prepared by diluting the stock solutions of 5′ Cy5-labeled oligos in TE buffer. The prepared DNA probes were aliquoted and stored at −20 °C.

For binding reactions, 5 nM of DNA probes were incubated with appropriate concentrations of proteins in binding buffer (20 mM Tris-HCl, 50 mM NaCl, 1 mM $MgCl_2$, 5 mM DTT, 5% glycerol, 0.001% tween-20, and 50 ug/mL BSA) for 15 min at room temperature. The samples were loaded to 6% native gel that had been pre-run. Electrophoresis was performed for 90 min at 130 V in a cold room with 0.5× 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). Structural alignments were conducted using the web-based pairwise structure alignment service provided by the RCSB PDB website, employing the TM-align method (Zhang and 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 selected for the alignments were pre-determined DBDs specified in previous studies (König et al. 1996; Sigrist et al. 2005; Marchler-Bauer et al. 2017; Lu et al. 2020; The UniProt Consortium 2022; Wang et al. 2022).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Author Contributions

H.S., H.-J.C., and J.L. conceptualization; H.S. data curation; H.S. formal analysis; J.L. funding acquisition; H.S., S.K., and D.S.L. investigation; H.S., S.K., D.S.L., and H.-J.C. methodology; H.-J.C. and J.L. supervision; H.S. visualization;

H.S. writing—original draft; H.S., S.K., D.S.L., and J.L. writing—review and editing

Funding

Conflict of Interest

The authors declare that they have no conflicts of interest with the contents of this article.

Data Availability

Accession information of the datasets used in this work is provided in the Methods section.

Literature Cited

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024:630(8016):493–500. https://doi.org/10.1038/s41586-024-07487-w.

Blaxter ML, De Ley P, Garey JR, Liu LX, Scheldeman P, Vierstraete A, Vanfleteren JR, Mackey LY, Dorris M, Frisse LM, et al. A molecular evolutionary framework for the phylum Nematoda. Nature. 1998:392(6671):71–75. https://doi.org/10.1038/32160.

Brenner S. The genetics of Caenorhabditis elegans. Genetics. 1974:77(1):71–94. https://doi.org/10.1093/genetics/77.1.71.

Bryan TM. G-quadruplexes at telomeres: friend or foe? Molecules. 2020:25(16):3686. https://doi.org/10.3390/molecules25163686.

Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021:18(4): 366–368. https://doi.org/10.1038/s41592-021-01101-x.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinf. 2009:10(1):1–9. https://doi.org/10.1186/1471-2105-10-421.

Červenák F, Sepšiová R, Nosek J, Tomáška Ľ. Step-by-step evolution of telomeres: lessons from yeasts. Genome Biol Evol. 2021:13(2): evaa268. https://doi.org/10.1093/gbe/evaa268.

De Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev. 2005:19(18):2100–2110. https://doi.org/10.1101/gad.1346005.

De Lange T. How telomeres solve the end-protection problem. Science. 2009:326(5955):948–952. https://doi.org/10.1126/science.1170633.

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

Edler D, Klein J, Antonelli A, Silvestro D. raxmlGUI 2.0: a graphical interface and toolkit for phylogenetic analyses using RAxML. Methods Ecol Evol. 2021:12(2):373–377. https://doi.org/10.1111/2041-210X.13512.

Howe KL, Bolt BJ, Cain S, Chan J, Chen WJ, Davis P, Done J, Down T, Gao S, Grove C, et al. WormBase 2016: expanding to enable helminth genomic research. Nucleic Acids Res. 2016:44(D1): D774–D780. https://doi.org/10.1093/nar/gkv1217.

Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite—a comprehensive resource for helminth genomics. Mol Biochem Parasitol. 2017:215:2–10. https://doi.org/10.1016/j.molbiopara.2016.11.005.

Huson DH, Scornavacca C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol. 2012:61(6): 1061–1067. https://doi.org/10.1093/sysbio/sys062.

Im SH, Lee J. Identification of HMG-5 as a double-stranded telomeric DNA-binding protein in the nematode Caenorhabditis elegans. FEBS Lett. 2003:554(3):455–461. https://doi.org/10.1016/S0014-5793(03)01191-8.

Im SH, Lee J. PLP-1 binds nematode double-stranded telomeric DNA. Mol Cells. 2005:20(2):297–302. https://doi.org/10.1016/S1016-8478(23)13230-4.

Joeng KS, Song EJ, Lee K-J, Lee J. Long lifespan in worms with long telomeric DNA. Nat Genet. 2004:36(6):607–611. https://doi.org/10.1038/ng1356.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021:596(7873):583–589. https://doi.org/10.1038/s41586-021-03819-2.

Kazda A, Zellinger B, Rössler M, Derboven E, Kusenda B, Riha K. Chromosome end protection by blunt-ended telomeres. Genes Dev. 2012:26(15):1703–1713. https://doi.org/10.1101/gad.194944.112.

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. 2003:278(30):28038–28044. https://doi.org/10.1074/jbc.M302192200.

König P, Giraldo R, Chapman L, Rhodes D. The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. Cell. 1996:85(1):125–136. https://doi.org/10.1016/S0092-8674(00)81088-0.

Kuznetsova V, Grozeva S, Gokhman V. Telomere structure in insects: a review. J Zool Syst Evol Res. 2020:58(1):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. 2003:5(4):621–633. https://doi.org/10.1016/S1534-5807(03)00293-4.

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. 2023:33(11):1947–1957. https://doi.org/10.1101/gr.278124.123.

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. 2023:80(11):312. https://doi.org/10.1007/s00018-023-04970-x.

Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, GwadzM, Hurwitz DI, Marchler GH, Song JS, et al. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res. 2020:48(D1):D265–D268. https://doi.org/10.1093/nar/gkz991.

Luo M, Luo H, Li H, Liu D, Lu H. Single-stranded DNA-binding proteins in plant telomeres. Int J Biol Macromol. 2020:165:1463–1467. https://doi.org/10.1016/j.ijbiomac.2020.09.211.

Madeira F, Madhusoodanan N, Lee J, Eusebi A, Niewielska A, Tivey AR, Lopez R, Butcher S. The EMBL-EBI job dispatcher sequence analysis tools framework in 2024. Nucleic Acids Res. 2024:52(W1): W521–W525. https://doi.org/10.1093/nar/gkae241.

Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE:

functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 2017:45(D1):D200–D203. https://doi.org/10.1093/nar/gkw1129.

Marquevielle J, De Rache A, Vialet B, Morvan E, Mergny J-L, 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. 2022:50(12):7134–7146. https://doi.org/10.1093/nar/gkac523.

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. 2009:28(22):3549–3563. https://doi.org/10.1038/emboj.2009.278.

Moon S, Lee YW, Kim WT, Lee W. Solution structure of CEH-37 homeodomain of the nematode *Caenorhabditis elegans*. Biochem Biophys Res Commun. 2014:443(2):370–375. https://doi.org/10.1016/j.bbrc.2013.11.133.

Moser BA, Nakamura TM. Protection and replication of telomeres in fission yeast. Biochem Cell Biol. 2009:87(5):747–758. https://doi.org/10.1139/O09-037.

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. 1973:41(1):181–190. https://doi.org/10.1016/0022-5193(73)90198-7.

Pearson WR. An introduction to sequence similarity ("homology") searching. Curr Protoc Bioinformatics. 2013:42(1):3.1. 1–3.1. 8. https://doi.org/10.1002/0471250953.bi0301s42.

Procházková Schrumpfová P, Schořová Š, Fajkus J. Telomere- and telomerase-associated proteins and their functions in the plant cell. Front Plant Sci. 2016:7:851. https://doi.org/10.3389/fpls.2016.00851.

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. 2008:132(5):745–757. https://doi.org/10.1016/j.cell.2007.12.039.

Saint-Leandre B, Levine MT. The telomere paradox: stable genome preservation with rapidly evolving proteins. Trends Genet. 2020:36(4):232–242. https://doi.org/10.1016/j.tig.2020.01.007.

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. 2016:11(4):e0154225. https://doi.org/10.1371/journal.pone.0154225.

Shakirov EV, Song X, Joseph JA, Shippen DE. POT1 proteins in green algae and land plants: DNA-binding properties and evidence of co-evolution with telomeric DNA. Nucleic Acids Res. 2009:37(22):7455–7467. https://doi.org/10.1093/nar/gkp785.

Shore D, Nasmyth K. Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. Cell. 1987:51(5):721–732. https://doi.org/10.1016/0092-8674(87)90095-X.

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. 2005:21(21):4060–4066. https://doi.org/10.1093/bioinformatics/bti614.

Smith JS, Chen Q, Yatsunyk LA, Nicoludis JM, Garcia MS, Kranaster R, Balasubramanian S, Monchaud D, Teulade-Fichou M-P, Abramowitz L, et al. Rudimentary G-quadruplex-based telomere capping in *Saccharomyces cerevisiae*. Nat Struct Mol Biol. 2011:18(4):478–485. https://doi.org/10.1038/nsmb.2033.

Smythe AB, Holovachov O, Kocot KM. Improved phylogenomic sampling of free-living nematodes enhances resolution of higher-level nematode phylogeny. BMC Evol Biol. 2019:19(1):1–15. https://doi.org/10.1186/s12862-019-1444-x.

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. 2015:6:162. https://doi.org/10.3389/fgene.2015.00162.

The UniProt Consortium. UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Res. 2022:51:D523–D531. https://doi.org/10.1093/nar/gkac1052.

Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022:50(D1):D439–D444. https://doi.org/10.1093/nar/gkab1061.

Wang J, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, et al. The conserved domain database in 2023. Nucleic Acids Res. 2022:51(D1):D384–D388. https://doi.org/10.1093/nar/gkac1096.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009:25(9):1189–1191. https://doi.org/10.1093/bioinformatics/btp033.

Watson JD. Origin of concatemeric T7DNA. Nat New Biol. 1972:239(94):197–201. https://doi.org/10.1038/newbio239197a0.

Xu J, Zhang Y. How significant is a protein structure similarity with TM-score=0.5? Bioinformatics. 2010:26(7):889–895. https://doi.org/10.1093/bioinformatics/btq066.

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. 2021:10:e64104. https://doi.org/10.7554/eLife.64104.

Yu X, Gray S, Ferreira HC. POT-3 preferentially binds the terminal DNA-repeat on the telomeric G-overhang. Nucleic Acids Res. 2022:51(2):610–618. https://doi.org/10.1093/nar/gkac1203.

Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 2005:33(7):2302–2309. https://doi.org/10.1093/nar/gki524.