The evolution of metazoan shelterin

Logan R. Myler,[1] Charles G. Kinzig,[1] Nanda K. Sasi, George Zakusilo, Sarah W. Cai, and Titia de Lange

Laboratory for Cell Biology and Genetics, The Rockefeller University, New York, New York 10021, USA

The mammalian telomeric shelterin complex—comprised of TRF1, TRF2, Rap1, TIN2, TPP1, and POT1—blocks the DNA damage response at chromosome ends and interacts with telomerase and the CST complex to regulate telomere length. The evolutionary origins of shelterin are unclear, partly because unicellular organisms have distinct telomeric proteins. Here, we describe the evolution of metazoan shelterin, showing that TRF1 emerged in vertebrates upon duplication of a TRF2-like ancestor. TRF1 and TRF2 diverged rapidly during vertebrate evolution through the acquisition of new domains and interacting factors. Vertebrate shelterin is also distinguished by the presence of an HJRL domain in the split C-terminal OB fold of POT1, whereas invertebrate POT1s carry inserts of variable nature. Importantly, the data reveal that, apart from the primate and rodent POT1 orthologs, all metazoan POT1s are predicted to have a fourth OB fold at their N termini. Therefore, we propose that POT1 arose from a four-OB-fold ancestor, most likely an RPA70-like protein. This analysis provides insights into the biology of shelterin and its evolution from ancestral telomeric DNA-binding proteins.

[*Keywords*: evolution; telomere; telomerase; shelterin; TRF1; TRF2; Rap1; TIN2; TPP1; POT1; NBS1; Apollo; PNUTS; MCPH1]

Supplemental material is available for this article.

Received July 13, 2021; revised version accepted October 5, 2021.

Telomeres have been studied in a wide variety of eukaryotes, revealing that some aspects of telomere biology are highly conserved, whereas others have diverged considerably (de Lange 2009; Lue 2018). A highly conserved aspect of telomere biology is the telomerase reverse transcriptase (TERT), which is similar in all eukaryotes (Wu et al. 2017; Nguyen et al. 2019) and evolved from the reverse transcriptases of ancient retroelements (Nakamura et al. 1997; de Lange 2015). Likewise, telomerase template RNAs, which dictate the synthesis of TTAGGG-like repeats onto the 3′ ends of metazoan chromosomes, share several conserved structural elements in highly divergent eukaryotes (Wu et al. 2017; Nguyen et al. 2019; Ghanim et al. 2021; He et al. 2021). Finally, the Ctc1–Stn1–Ten1 (CST) complex responsible for polymerase α/primase-mediated fill-in at the 5′ ends of telomeres is also highly conserved and probably evolved from a single-stranded (ss) DNA-binding complex related to replication protein A (RPA) (Lue 2018; Lim et al. 2020).

In contrast, the proteins that associate with the duplex telomeric repeat array have diversified more, despite the presence of TTAGGG-like repeats (including variants such as TTGGGG, TTTTGGGG, and TTAGG) at the telomeres of most eukaryotes. Telomeric proteins have been identified in mammals, plants, fungi, ciliates, trypanosomes, and other species. Apart from two common DNA-binding domains (the oligonucleotide/oligosaccharide-binding [OB] fold and the Myb/SANT domain, referred to here as Myb), the telomere-specific TRFH (telomeric repeat factor homology) domain, and the BRCT (BRCA1 C-terminal) domain (de Lange 2018), the telomeric protein complexes appear to lack conserved features. Because the species studied are from disparate clades, it is difficult to discern how telomeric proteins evolved from common ancestors. Therefore, we set out to trace the evolution of shelterin, first identified in mammals (de Lange 2005), through the vertebrate and invertebrate metazoan lineages.

Shelterin binds to the TTAGGG repeats that denote the telomeres of most metazoans, where it ensures that telomere ends are not recognized and/or processed by pathways that act on double-strand breaks (DSBs) (de Lange 2018). Shelterin also regulates the extension of telomeric sequences by telomerase and facilitates the semiconservative replication of telomeric DNA (Hockemeyer and Collins 2015). Despite its myriad functions, shelterin is comprised of just six distinct proteins: TRF1, TRF2, Rap1, TIN2, TPP1, and POT1.

The architecture of shelterin and the function of its subunits have been studied extensively in human and mouse cells (O'Sullivan and Karlseder 2010; Lazzerini-Denchi and Sfeir 2016; de Lange 2018). In these organisms, shelterin contains subunits that bind to the double-stranded

[1]**These authors contributed equally to this work.**

Corresponding author: delange@rockefeller.edu

Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.348835.121. Freely available online through the *Genes & Development* Open Access option.

(ds) telomeric DNA and subunits that engage telomeric DNA in single-stranded (ss) form, a conserved feature of telomere ends. Two paralogous homodimeric DNA-binding proteins—TRF1 and TRF2—anchor shelterin onto the ds telomeric TTAGGG repeats. Both TRF1 and TRF2 bind TIN2, which promotes the stability of TRF1 and TRF2 at telomeres. TRF2 (but not TRF1) binds to Rap1. The ss telomeric repeats are bound by POT1, which forms a heterodimer with TPP1. TIN2 binds to TPP1, thereby connecting POT1/TPP1 to TRF1 and TRF2.

Repression of the DNA damage response at telomeres is primarily achieved through TRF2 and POT1. TRF2 prevents ATM kinase activation at chromosome ends and ensures that the nonhomologous end joining pathway does not ligate one telomere to another (van Steensel et al. 1998; Karlseder et al. 1999; Smogorzewska et al. 2002; Celli et al. 2006). TRF2 blocks these pathways through the formation of the t-loop structure, in which the 3′ telomeric overhang strand invades the duplex part of the telomere (Griffith et al. 1999; Doksani et al. 2013). Although TRF1 is a paralog of TRF2, it is not directly involved in the protection of telomeres. Rather, TRF1 facilitates the duplication of telomeres by preventing the stalling/arrest of the replication fork in the TTAGGG repeat array (Martínez et al. 2009; Sfeir et al. 2009). ATR signaling, the second DNA damage signaling pathway that endangers telomeres, is repressed by POT1, which excludes the ssDNA sensor RPA from telomeric DNA (Denchi and de Lange 2007; Guo et al. 2007; Gong and de Lange 2010; Kratz and de Lange 2018).

Shelterin has a dual role in the maintenance of telomeric DNA. Telomerase-mediated telomere elongation requires the recruitment of the enzyme by TPP1, which interacts directly with the TERT component (Nandakumar et al. 2012; Zhong et al. 2012; Hockemeyer and Collins 2015; Lim and Cech 2021). In addition, shelterin regulates telomere length homeostasis in a process that is thought to involve a negative feedback loop whereby telomerase is inhibited at hyperelongated telomeres (Hockemeyer and Collins 2015). The main players in this negative feedback loop are TRF1, TIN2, POT1, and the CST complex, but the mechanism of telomerase inhibition is unclear.

Human and mouse shelterin complexes are similar in many respects, allowing cross-species extrapolation of functional studies. However, some differences are notable. For instance, mice and rats have two POT1 proteins (POT1a and POT1b) rather than the single POT1 protein present in primate shelterin (Hockemeyer et al. 2006; Wu et al. 2006). Mouse and human shelterin also have distinct shelterin accessory factors, which make an important contribution to shelterin function (Diotti and Loayza 2011).

Here, we describe how metazoan shelterin evolved. We report that the shelterin complexes of invertebrates contain only one telomeric repeat binding factor (TRF) with features of TRF2, including sequences related to the iDDR (inhibitor of the DNA damage response) (Okamoto et al. 2013). Therefore, we infer that TRF1 was formed through a duplication of a TRF2-like gene near the emergence of the vertebrate lineage. Possibly, the replication function specific to vertebrate TRF1 is executed by the TRF2-like shelterin subunit of invertebrates. Once duplicated, TRF1 and TRF2 diverged rapidly during vertebrate evolution, resulting in their carrying distinct N termini and having different interacting partners. Another major change occurred in the C-terminal OB fold of POT1, which is split by a Holliday junction resolvase-like (HJRL) domain in vertebrates but has a different insert in other metazoans. Finally, we report that the three-OB-fold architecture of primate and rodent POT1 proteins is an exception. Most metazoans, including most mammals, have POT1 orthologs with an additional N-terminal domain that is predicted to be an OB fold, which we term OB-N. This suggests that POT1 first evolved as a four-OB-fold protein, potentially from an RPA70-like ancestor. These findings should lend impetus to unraveling the perplexing evolution of telomeric protein complexes in other eukaryotes, potentially revealing how early eukaryotes solved the problems inherent to linear chromosomes.

Results

TRF1 arose in vertebrates

The TRF1 and TRF2 subunits of shelterin (Supplemental Fig. S1A) are encoded by paralogous genes that are annotated in most vertebrate genomes, but the presence of invertebrate TRF1/2 orthologs, defined as proteins containing a TRFH domain paired with a C-terminal TRF1/2-like Myb domain, has not been examined. To catalog TRF1 and/or TRF2 paralogs in all branches of metazoans, we identified genes annotated as TRF1- or TRF2-like. Additional candidate genes were uncovered through targeted genome BLASTp searches with the TRFH and Myb domains of human TRF2 and TRF1. Only genes with a C-terminal Myb domain paired with a TRFH domain were considered TRF1/2 candidates (Supplemental Fig. S1A) because TRFH-mediated homodimerization is required for the telomere binding activity of TRF1 and TRF2 (Bianchi et al. 1997; van Steensel and de Lange 1997; van Steensel et al. 1998). Although our analysis was limited by the availability of high-quality genome sequences, we identified TRF genes in 16 invertebrates and compared these with the TRF orthologs of 11 vertebrates (Supplemental Figs. S1, S2). We excluded arthropods and nematodes because they have noncanonical telomeric repeat DNA, often lack telomerase, and have a distinct set of telomeric DNA-binding proteins (Fulcher et al. 2014). A sequence identities and similarities (SIAS) tool (http://imed.med.ucm.es/Tools/sias.html) indicated that the Myb domains of invertebrate TRF orthologs had, on average, 50% ± 10% amino acid identity to the Myb domains of vertebrate TRF1/2 orthologs. The TRFH domains of invertebrate TRFs appeared to be less conserved, showing 36% ± 10% identity compared with vertebrate TRF1/2.

The Myb domains of all candidate TRF1/2 proteins have the conserved K-D-R/K-W/Y-R motif in helix 3, which is critical for recognition of the telomeric

TTAGGG repeats (Supplemental Fig. S1B; Court et al. 2005). Because it has been reported that TRF1 evolved the ability to bind to telomeric DNA in the therian lineage (Kappei et al. 2017), we determined whether two TRF1 proteins from nontherian vertebrates (the lizard *Anolis carolinensis* and the platypus *Ornithorhynchus anatinus*) have the ability to localize to telomeres (Supplemental Fig. S3A,B). Both proteins localized to telomeres in mouse embryonic fibroblasts (MEFs), and their Myb domains allowed mouse TRF1 (missing its Myb domain) to localize telomeres (Supplemental Fig. S3B). Importantly, the telomeric localization of the lizard and platypus TRF1s was abolished by the equivalent of the R425V mutation, which inactivates DNA binding by the human Myb domain (Supplemental Fig. S3B; Fairall et al. 2001). Based on these data, we consider it likely that all vertebrate TRF1 proteins can bind telomeric DNA.

Each invertebrate species examined contained only a single candidate TRF gene. The absence of a second TRF gene in each invertebrate was further supported by BLASTp searches with the Myb domain of the TRF gene from that species. Bayesian phylogenetic analysis segregated the TRFH domains of all TRF genes according to phylum, as expected (Fig. 1A). However, the TRF1 and TRF2 orthologs of vertebrates clustered together independently of the phylogenetic tree, indicating that TRF1 and TRF2 diverged from a common ancestor within the vertebrate lineage (Fig. 1A, arrowhead). This implies that the ancestral gene of TRF1 and TRF2 was duplicated around the time that vertebrates diverged from the cephalochordates. Because we have been unable to find TRF orthologs in tunicates, the exact point in evolution when the duplication took place remains to be determined.

To determine whether the common ancestor of TRF1 and TRF2 was more like TRF1 or more like TRF2, we examined the invertebrate orthologs for features that distinguish the two vertebrate paralogs. TRF1 and TRF2 differ in their N terminus, which is acidic for TRF1 and basic for TRF2. However, as detailed below (see Fig. 2A,B), these domains evolved rapidly in vertebrates, making them unsuitable for this analysis. Instead, we focused on another discriminating feature: the TRF2-specific iDDR domain, which resides immediately upstream of the Myb domain (Okamoto et al. 2013). The iDDR of human TRF2 opposes RNF168-mediated histone ubiquitination to block the accumulation of the DNA damage factor 53BP1 at dysfunctional telomeres (Okamoto et al. 2013). Alignment of the iDDRs of vertebrate TRF2 proteins revealed a short motif rich in acidic residues surrounding a conserved tryptophan (Fig. 1B,C). This iDDR motif was also present in the cephalochordates, suggesting that their single TRF protein is similar to TRF2 (Fig. 1B).

The iDDR sequence motif was not sufficiently conserved to definitively determine whether it was present in the achordate TRF proteins (Fig. 1D). However, all vertebrate TRF2 orthologs were distinct from TRF1 in that they possessed clusters of glutamate and aspartate residues within 60 amino acids upstream of the Myb domain (Fig. 1D). Calculation of the local isoelectric point (pI) upstream of the Myb domains of the vertebrate TRFs revealed a markedly low local pI immediately upstream of the Myb domain in TRF2 but not in TRF1 (Fig. 1E). The iDDR motif and the low local pI were found in the cephalochordates' TRFs, further arguing that their TRF is TRF2-like (Fig. 1C,E).

Consistent with the invertebrate TRFs displaying TRF2-like features, the cnidarians had an iDDR-like region upstream of the Myb domain of their TRFs, featuring clusters of acidic residues flanking a tryptophan residue and a pronounced drop in the local pI (Fig. 1D,E). The echinoderms also had these features, although some lacked the tryptophan associated with acidic residues, and the low local pI segment was farther upstream of the Myb domain. Only the mollusks' TRFs did not have an obvious acidic region upstream of the Myb domain, suggesting that the iDDR-like feature may have been lost or repositioned in this clade.

Based on the presence of an iDDR-like region upstream of the Myb domain of the cephalochordates, cnidarians, and echinoderms, we propose that the single TRF gene in invertebrates resembles TRF2 more than TRF1 (Fig. 1F). Consistent with this view, these TRFs also feature a highly conserved proline near the end of the TRFH domain that is specific to vertebrate TRF2 paralogs (Supplemental Fig. S2A, arrow). Finally, vertebrate TRF1 and TRF2 proteins differ significantly in length, primarily due to TRF1 having a shorter hinge region between the TRFH and Myb domains (Supplemental Fig. S2B). In agreement with their similarity to TRF2, the invertebrate TRFs are significantly longer than the vertebrate TRF1 orthologs and have a longer hinge region (Supplemental Fig. S2B).

In conclusion, except for the vertebrates, all metazoans have a shelterin that contains a single double-stranded (ds) telomeric DNA-binding protein with features of TRF2. In vertebrates, shelterin gained a second ds telomeric DNA-binding protein, TRF1, that likely arose through the duplication of the gene encoding the ancestral metazoan TRF2-like protein. TRF1 presumably lost the iDDR-like region and the Rap1- and TIN2-binding sites from the hinge region, possibly resulting in its smaller size. We assume that, as in mammals, the TTAGGG repeats in invertebrate telomeres challenge replication forks, potentially requiring a shelterin-based solution to this problem. It is possible that the single TRF of invertebrate shelterin has this function and that, in vertebrates, TRF1 retained the ability to promote the duplication of the telomeric DNA while TRF2 lost this feature.

TRF1 and TRF2 evolved distinct N-terminal domains in mammals

The single TRF of cephalochordates possesses neither of the N-terminal domains typical of mammalian TRF1 and TRF2 (Fig. 2A), indicating that the N termini of TRF1 and TRF2 evolved during vertebrate evolution, as others have suggested (Poulet et al. 2012). The basic N terminus of mammalian TRF2 is rich in arginine and glycine residues and is referred to as the Basic or the GAR (glycine arginine-rich) domain (Supplemental Fig. S1A). The TRF2

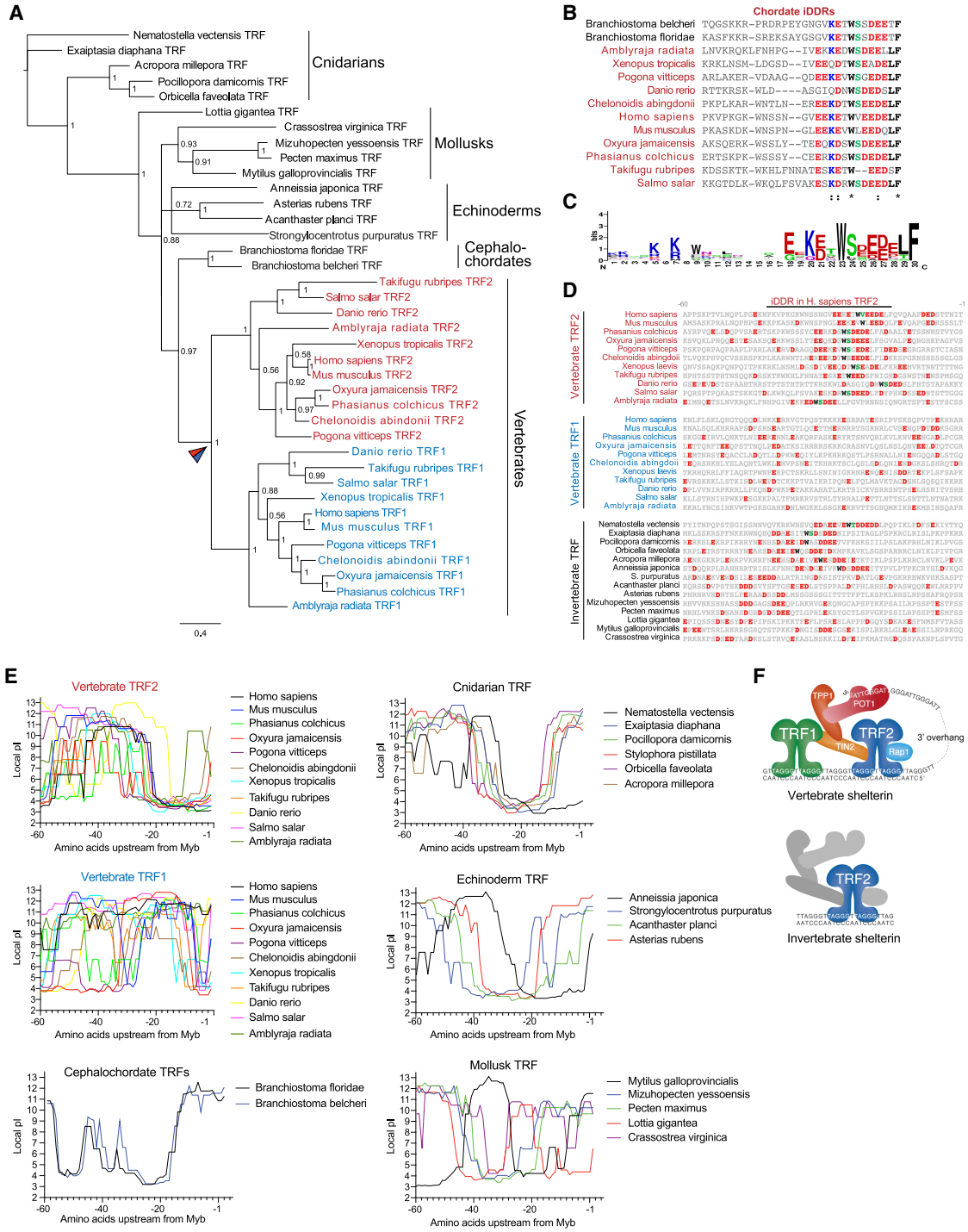

Figure 1. TRF1 arose in vertebrates from a duplication of a TRF2-like gene. (*A*) Phylogenetic tree of the TRFH domains of TRF orthologs from representative metazoans, including invertebrate TRF orthologs (black) and vertebrate TRF2 (red) and TRF1 (blue) orthologs. The arrowhead indicates the predicted gene duplication event. The scale bar indicates branch length in substitutions per sequence site. Node probabilities are indicated. (*B*) MUSCLE alignment of iDDR domains from the indicated TRF2 (red) or TRF (black) orthologs. Conserved residues from the iDDR domain are indicated in black (hydrophobic or aromatic residues), red (acidic residues), blue (basic residues), and green (V/S/T). (*) Identical residues, (:) conserved substitutions. (*C*) Sequence logo of conserved residues in the iDDR domains from *B*. "N" represents sequence position starting from the *left*. Bits indicate relative frequency of each amino acid at that position. (*D*) Comparison of the 60 amino acids immediately N-terminal of the Myb domains in the indicated TRF1/TRF2 orthologs. (Red) Acidic residues, (black) tryptophan (W), (green) V/S/T positions that approximate the motif in *C*. (*E*) Graphs showing the local pI of the sequence immediately upstream of the Myb domain of the indicated TRF1/2 orthologs. (*F*) Model of vertebrate and invertebrate shelterin highlighting the TRF2-like nature of the single TRF in invertebrate shelterin.

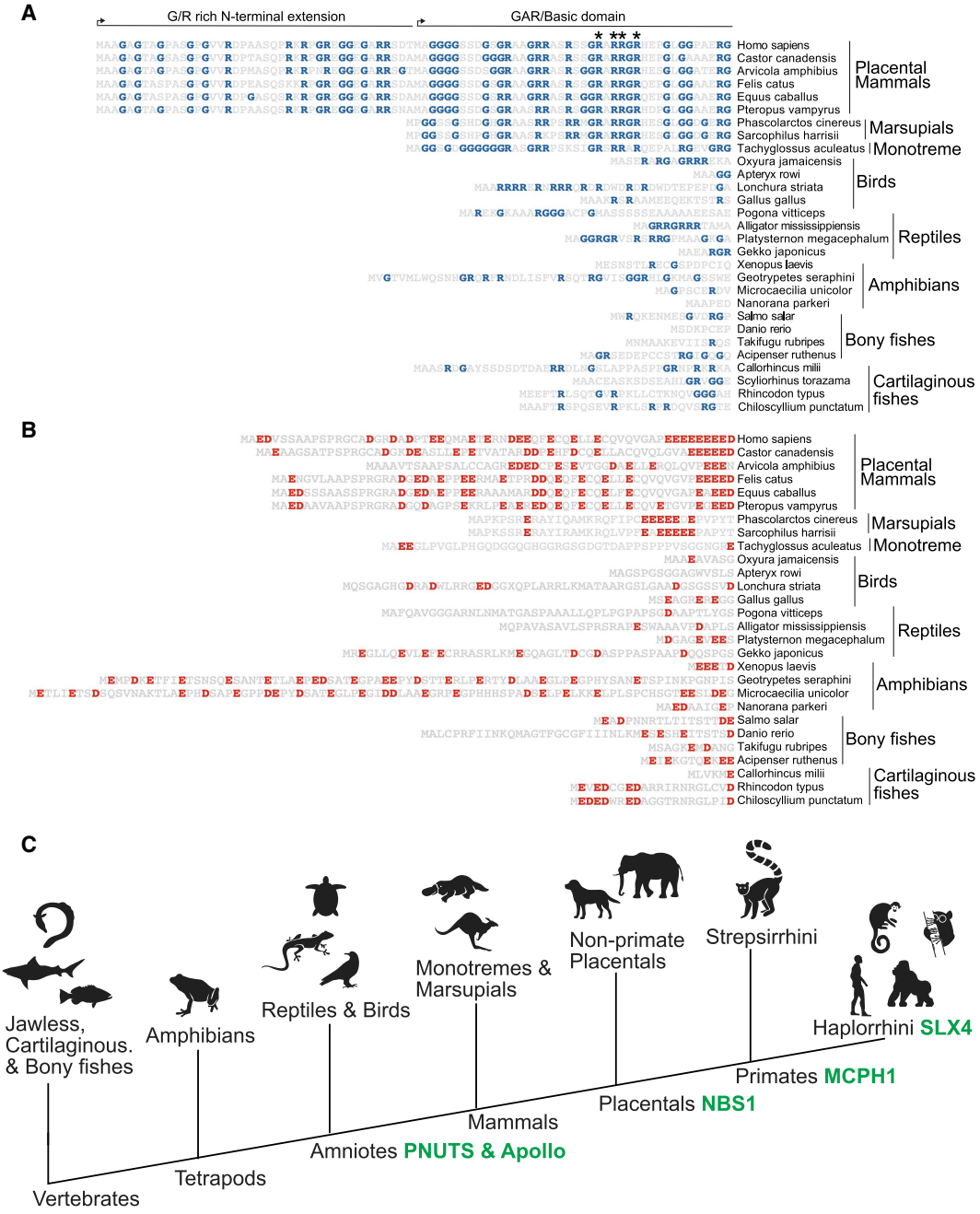

Figure 2. Emergence of the N-terminal domains in vertebrate TRFs and the Y/H-x-L-x-P motifs in human TRF2-binding partners. (*A*) Right-aligned Basic/GAR domains of vertebrate TRF2 orthologs, with glycines (G) and arginines (R) indicated in blue. (*) LANA-related motif residues. (*B*) Right-aligned acidic domains of vertebrate TRF1 orthologs with acidic (D/E) residues indicated in red. (*C*) Cladogram showing the acquisition of the Y/H-x-L-x-P motif in shelterin accessory proteins that interact with the TRFH domain of TRF2.

paralogs of placental mammals express two forms of TRF2 that differ in the length of the GAR/Basic domain (Fig. 2A; Timashev and de Lange 2020). Marsupials and monotremes lack the long form of TRF2 but have a GAR/Basic domain very similar to that of the placental mammals (Fig. 2A). Each of these GAR/Basic domain sequences features the LANA-related motif that is thought to allow the N terminus of TRF2 to bind to the acidic patch of the nucleosome (Konishi et al. 2016). However, most other vertebrate TRF2 paralogs lack a region that can be qualified as a GAR or Basic domain, or the N-terminal domain is very short (Fig. 2A).

The acidic domain of TRF1 is also a mammalian idiosyncrasy (Fig. 2B). With the exception of a few amphibians, nonmammalian vertebrates contain TRF1 orthologs with an N terminus that either is very short or lacks the

preponderance of acidic amino acids typical of mammalian TRF1 orthologs. This is consistent with the previously noted lack of an acidic N terminus in the TRF1 orthologs of *Gallus gallus*, *Xenopus laevis*, and *Danio rerio* (Bassham et al. 1998; De Rycker et al. 2003; Xie et al. 2011). Thus, the acidic and basic nature of the TRF1 and TRF2 N termini, respectively, likely evolved around the time that mammals diverged from other vertebrates.

Emergence of TRF2 TRFH-binding partners

The TRFH domain of human TRF2 has a binding site for several shelterin accessory factors, including Apollo/SNM1B, SLX4, and NBS1 (Demuth et al. 2004; Lenain et al. 2006; van Overbeek and de Lange 2006; Chen et al. 2008). These factors use a Y/H-x-L-x-P motif to bind to the same site in TRF2 (surrounding F162 in human TRF2), which is present only in vertebrates (Supplemental Fig. S2A). TRF1 has a similar binding site surrounding F142 that is required for its interaction with TIN2 (Supplemental Fig. S2A; Chen et al. 2008).

The TRF2-interacting partners evolved their Y/H-x-L-x-P motifs at different points in vertebrate evolution (Fig. 2C; Supplemental Fig. S4). In the PNUTS phosphatase (Kim et al. 2009) and Apollo/SNM1B orthologs, this motif is present in amniotes; in PNUTS, the acquisition of the TRF2-binding motif may have occurred earlier, since some amphibians and fish have a closely related Y-x-I-x-P sequence (Supplemental Fig. S4). The second addition to the TRF2 arsenal is NBS1, which evolved its TRF2-binding motif in placental mammals. The Y/H-x-L-x-P motif in MCPH1 evolved in the primates, with the TRF2-binding motif in SLX4 being the last to appear, arising in the haplorrhines (simians and tarsiers) (Fig. 2C). Since detailed information on TRF2-binding proteins largely derives from studies of mouse and human TRF2, it is possible that other TRF2 interactions evolved in non-mammalian vertebrates through acquisition of Y/H-x-L-x-P motifs in proteins that are distinct from the mammalian TRF2 interactors.

Evolution of the TRF2–Rap1 interface

Rap1 has long been known to be one of the most conserved shelterin subunits, and it is the only shelterin subunit found at the telomeres of budding yeast (Shore 1994; Kabir et al. 2010, 2014). The most conserved regions of Rap1 are the N-terminal BRCT domain (pfam16589) and the central Myb domain (pfam08914), which are found not only in all metazoans but also in many unicellular organisms (Fig. 3A). The Rap1 C-terminal (RCT) domain, which interacts with TRF2, is much less conserved, yet structural analysis has revealed that this region folds into a similar six-helix bundle in mammals and fungi (Fig. 3B,C; Chen et al. 2011). The interaction of the RCT domain with its binding partners is primarily mediated by helices 1 and 2 (Fig. 3B,C; Chen et al. 2011). Consistent with the conservation of the RCT in vertebrate Rap1 orthologs, the Rap1 binding site in the hinge region of TRF2 is conserved in mammals, birds, most reptiles, and some amphibians (Fig. 3C). The sequence conservation of the Rap1-binding site is accompanied by the structural conservation of the two helices involved in Rap1 binding (as predicted by RaptorX) (Fig. 3C). In contrast, most TRF2 orthologs in bony and cartilaginous fish appear to lack the Rap1-binding site, and it is not found in invertebrates (Fig. 3C; data not shown). Although there is some sequence similarity between the vertebrate Rap1 binding site and the sequences in the TRF2 hinge domain of *Branchiostoma belcheri* and *Acanthaster planci*, this is unlikely to represent a Rap1 binding site, since these regions are not predicted to fold into the two helices required for binding to the Rap1 RCT (data not shown).

These data do not imply, however, that the Rap1 orthologs from fish and invertebrates do not bind to their cognate TRF2 proteins. In this regard, Gaullier et al. (2016) proposed that a Y-x-L-x-P motif in the BRCT domain of Rap1 binds to the TRF2 F162 site. As this motif in Rap1 is absent from many vertebrate Rap1 orthologs (Supplemental Fig. S5), it is unlikely to represent an ancestral Rap1–TRF2 interface. An alternative Rap1–TRF binding mode, involving the Myb domain of Rap1, was noted in the unicellular parasite *Trypanosoma brucei* (Afrin et al. 2020). Whether this represents an ancestral state remains to be determined.

Only the TRFH domain of TIN2 is highly conserved

The TRFH domain of TIN2 is conserved in all metazoans (23% ± 9% identity) (Fig. 4A,B; http://imed.med.ucm.es/Tools/sias.html), consistent with the ancient origin of the TRFH domain (Hu et al. 2017; Xue et al. 2017). In contrast, the F-x-L-x-P TRF1-binding motif (cd11741), C-terminal to the TRFH domain of TIN2, is found only in mammals (Fig. 4C; data not shown). Nonetheless, in *Xenopus*, TRF1 has been shown to interact with TIN2 (Vizlin-Hodzic et al. 2009). The short region where dyskeratosis congenita (DC) (Savage 2018) mutations cluster is conserved in mammals, but the homology is strongly reduced in reptiles and amphibians (Fig. 4C). No significant homology with the DC cluster was detected in other vertebrates (Fig. 4C; data not shown).

Consistent with the conservation of the TRFH domain, the regions in TRF2 and TPP1 that interact with the TRFH domain of TIN2 are conserved in all vertebrates (33% ± 18% identity for the TIN2-binding motif [TBM] in TPP1 and 46% ± 24% identity for the TBM in TRF2) (Fig. 4D,E; http://imed.med.ucm.es/Tools/sias.html). However, the TBM in TRF2 could not be identified in 12 invertebrates examined by alignment to the relevant sequences from vertebrate TRF2s, indicating a low level of conservation (Fig. 4D; data not shown). The helical motif that allows TPP1 to bind to TIN2 was also absent from most invertebrates, except for the echinoderms *Acanthaster planci* and *Asterias rubens* (Fig. 4E; data not shown). In these two starfish, the C terminus of TPP1 is predicted to be α-helical and has sequence similarity to the TBM in vertebrate TPP1s (Fig. 4E).

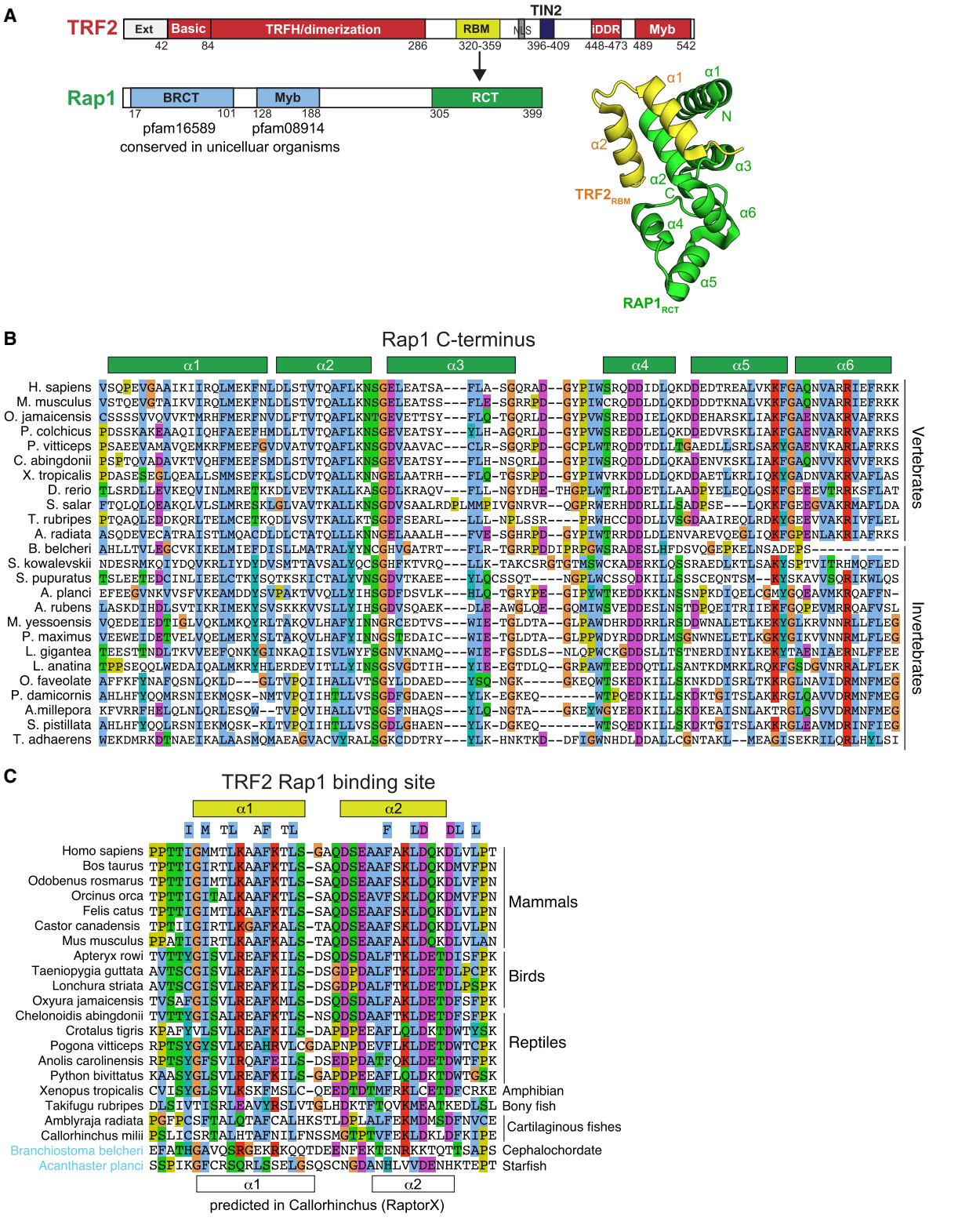

Figure 3. Evolution of the TRF2–Rap1 interface. (*A*) Domain architecture of human TRF2 and Rap1 with the indicated RBM–RCT interaction (arrow) and the structure of the TRF2 RBM (yellow) with the Rap1 RCT (green) (from PDB 3K6G). (*B*) Alignment of the RCT of Rap1. Helices are indicated by green boxes. (*C*) Alignment of the RBM of TRF2. Helices are indicated by yellow boxes. The helices in the presumed RBM of *Callorhinchus milii* are predicted by RaptorX. All alignments were performed with MUSCLE and are shown with ClustalX coloring.

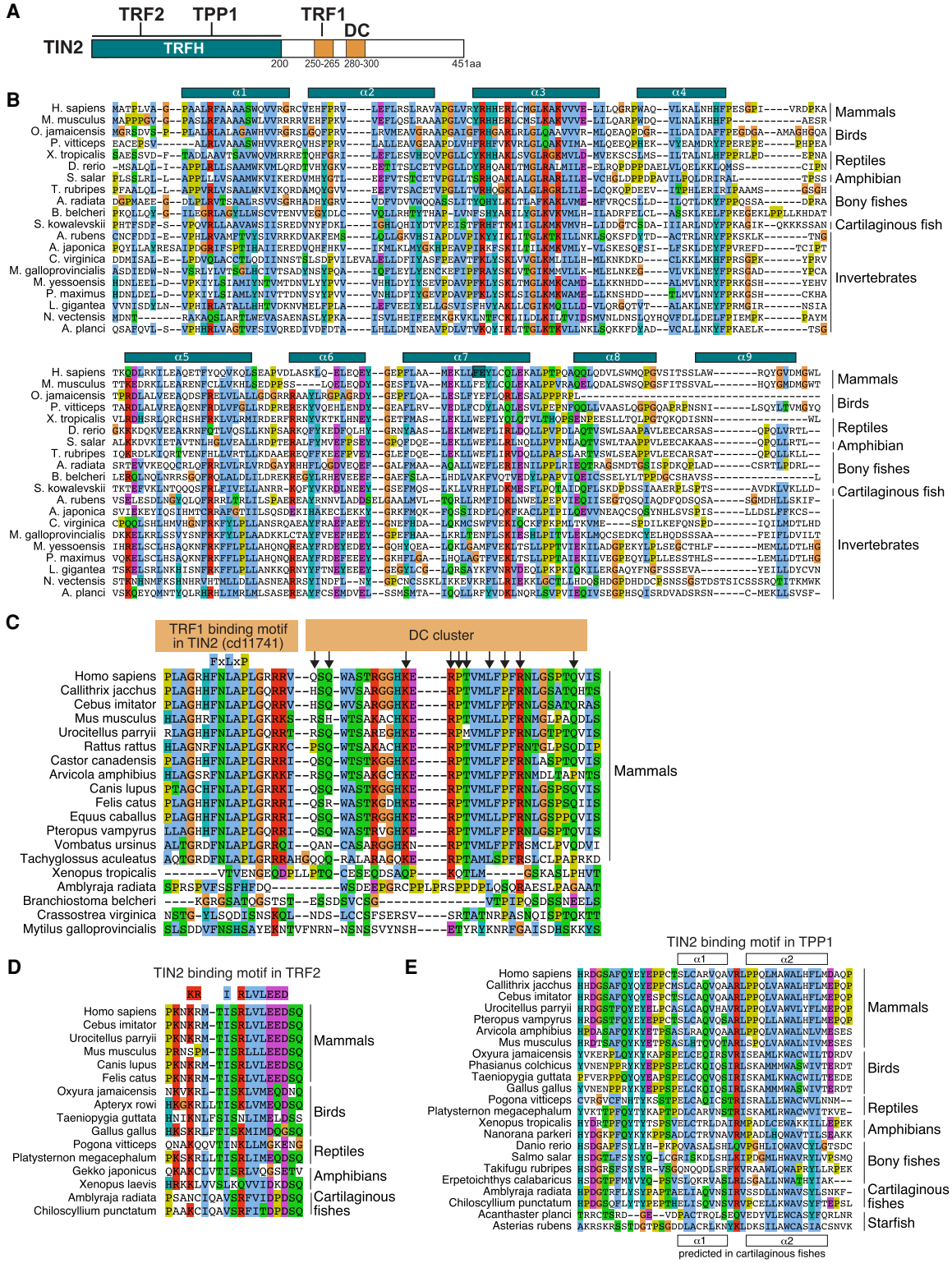

Figure 4. Conservation of the TIN2 TRFH domain and its interaction motifs in TRF2 and TPP1. (*A*) Domain architecture of human TIN2. Sites of interaction with other shelterin subunits are indicated. (*B*) Alignment of the TIN2 TRFH domains with conserved residues is indicated. Helices are indicated by teal boxes. (*C*) Alignments of the TBM and the DC patch. Arrows indicate DC mutations (Q269X, Q271X, K280E, K280fs, R282H, R282S, P283S, P283A, P283H, T284A, T284fs, L287P, P289S, R291G, and Q298fs). (*D*) Alignment of the TIN2-binding motif of TRF2. (*E*) Alignment of the TIN2-binding motif of TPP1. Helices from the human structure are noted with the white boxes *above* the sequences. Predicted helices (RaptorX) in TPP1 orthologs from cartilaginous fish are shown in white boxes *below* the sequences. Alignments in *B*–*E* were done with MUSCLE with ClustalX coloring.

Conserved motifs of TPP1 and its highly variable linker

The TIN2 interaction motif (Fig. 4E) in the C terminus of TPP1 is separated from two other functional domains: the N-terminal OB fold and the adjacent recruitment domain (RD), which interacts with POT1 (Fig. 5A). The OB fold and RD are separated from the TIN2-binding site by linkers of variable length, ranging from 50 to 700 amino acids (Fig. 5A,B), which is the least conserved part of TPP1 (Supplemental Fig. S6A). While the linker in the vertebrate TPP1 orthologs has conserved sequences (Supplemental Fig. S6B), this conservation does not extend to invertebrates (data not shown). In the vertebrates, the

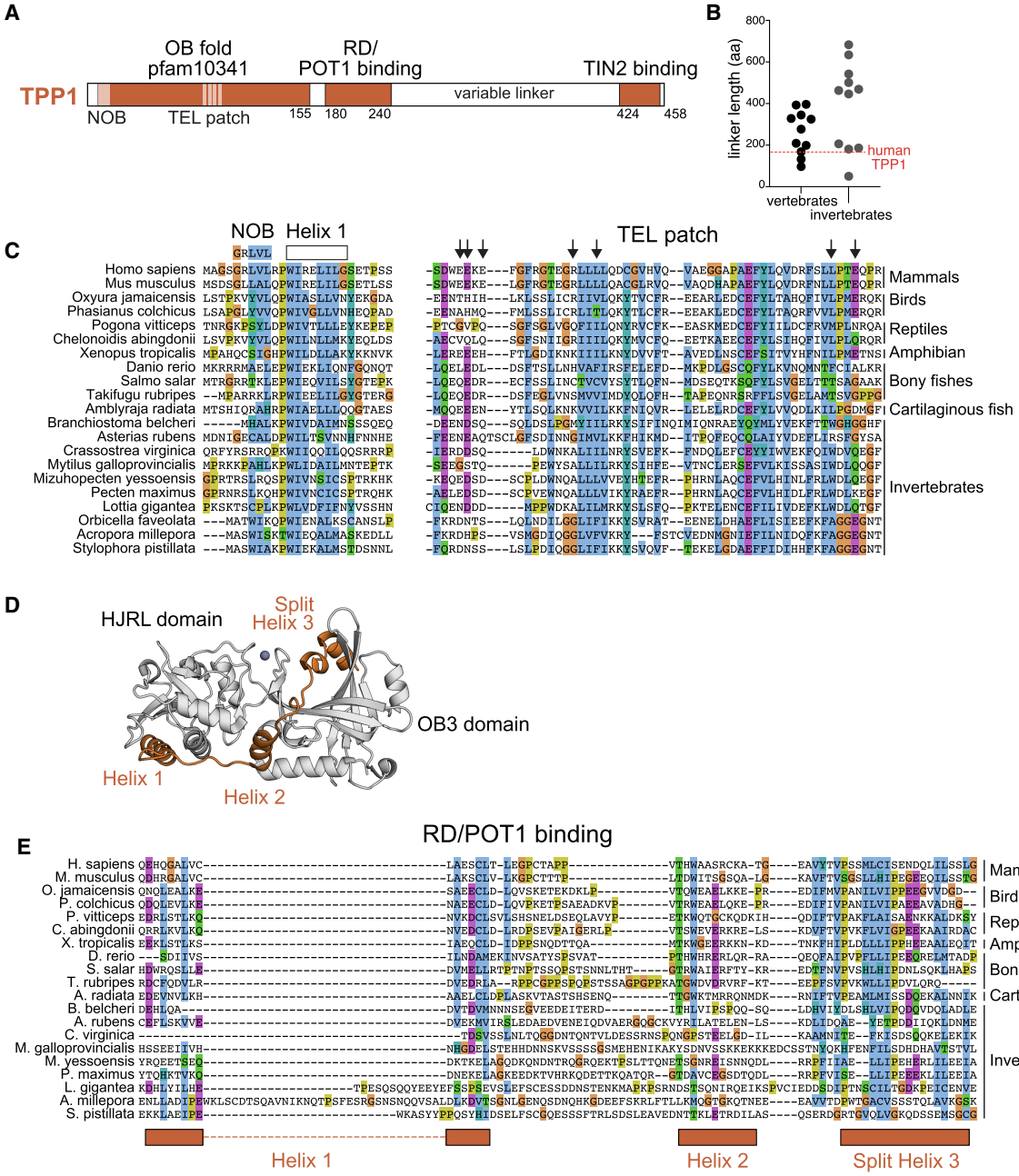

Figure 5. Conserved motifs and the variable linker of TPP1. (*A*) Domain architecture of the main isoform of human TPP1. (*B*) Plot showing the variability in the length of the TPP1 linker. (*C*) Alignment of two telomerase recruitment motifs. (*Left*) The N-terminal OB (NOB) motif is indicated *above*. Helix 1 of the human TPP1 OB fold is indicated with a white box. (*Right*) The telomerase-binding TEL patch is indicated with arrows pointing to residues (E81, E82, E84, R93, L96, L125, and E128) critical for the TPP1–telomerase interaction. (*D*) Structure (PDB 5UN7) of the OB3/HJRL domain of POT1 (gray) showing the three helices of TPP1 critical for POT1 binding (orange). (*E*) Alignment of the RD/POT1-binding domain of TPP1. Positions of the three helices in the human TPP1 sequence are indicated in orange *below*. Alignments in *C* and *E* were done with MUSCLE with ClustalX coloring.

linker is predicted to lack secondary structure, and it has a large number of serine and proline residues with low helical propensity, suggesting that it is intrinsically disordered (Supplemental Fig. S6C).

The OB fold of TPP1 is highly conserved in all metazoans as well as in a number of unicellular organisms (pfam10341). In mammals, the OB fold recruits telomerase to telomeres using the most N-terminal sequences (NOB) and the centrally located TEL patch, which bind to the TEN domain of TERT (Nandakumar et al. 2012; Zhong et al. 2012; Sexton et al. 2014; Grill et al. 2018, 2021). As previously noted, the TEL patch and the NOB are conserved in mammals (Fig. 5C; Grill et al. 2018, 2021). However, the residues known to contribute to the interaction of TPP1 with TERT in human cells are not conserved in other vertebrates (Fig. 5C). It remains to be addressed whether and how TPP1 brings telomerase to telomeres in nonmammalian animals.

Mammalian TPP1 binds POT1 through the interaction of its central RD domain with the HJRL domain of POT1, which interrupts the third OB fold in vertebrates (Figs. 5D,E, 6; Chen et al. 2017; Rice et al. 2017). The sequence of the RD domain of TPP1 is conserved in vertebrates but less so in invertebrates, consistent with the HJRL domain being absent from invertebrate POT1 orthologs (see Fig. 6). Nonetheless, the invertebrate RD domains are predicted to contain the second helix and the third (split) helix that provide the POT1-binding interface in vertebrates (Fig. 5E).

Most metazoan POT1 proteins have a predicted fourth OB fold

The ssDNA–binding domain of POT1 contains two highly conserved OB folds (OB1 and OB2; pfam 02765 and 16686, respectively), the first of which allowed the identification of both fission yeast and human POT1 orthologs based on sequence similarity to the ciliate telomere-binding protein TEBPα (Fig. 6A; Baumann and Cech 2001). Crystallography revealed the presence of a third OB fold in human POT1, which is split by an HJRL domain (Fig. 6A; Chen et al. 2017; Rice et al. 2017). The function of the HJRL has not been established.

Interestingly, our analysis indicates that the three-OB-fold architecture of POT1 in rodents and primates differs from what is seen in all other animals (Fig. 6B). Most vertebrate POT1 proteins have a conserved N-terminal extension that is also present in the POT1 proteins of invertebrates (Fig. 6B). SWISS-MODEL analysis predicted that in each case, this N-terminal extension folds into a canonical OB fold (Fig. 6C). We conclude that most metazoan POT1 proteins likely have a fourth OB fold at their N termini, which we call OB-N.

The loss of OB-N occurred during mammalian evolution ~100 million years ago when the Euarchontoglires (which include primates, rodents, and lagomorphs) diverged from Laurasiatherian mammals (which include carnivores, insectivores, and bats) (Supplemental Fig. S7A). Analysis of the exon–intron structure of the human POT1 gene indicates that the absence of OB-N is due to the loss of a small exon (e.g., exon 3 in the *Felis catus* POT1 gene), which results in a frameshift and premature termination of the OB-N reading frame (Supplemental Fig. S7B,C). Downstream from this stop codon, a second ATG allows expression of the three-OB-fold version of POT1. This second ATG is also present in the feline POT1 mRNA, raising the possibility that cats, and perhaps other mammals, express a three- as well as a four-OB-fold POT1 isoform. Among the Euarchontoglires, we identified one species (the arctic ground squirrel, *Urocitellus parryii*) that has restored the four-OB-fold architecture of POT1, presumably by mutations that revive the lost exon.

The C-terminal OB fold (OB3) of all metazoan POT1 orthologs is well conserved and often interrupted by a region with a much lower level of conservation (Fig. 6D,E). In vertebrate POT1 orthologs, OB3 contains a conserved HJRL domain, whereas the insert is highly variable in sequence and length in invertebrates (Fig. 6E,F; Supplemental Fig. S8A,B). For instance, the insert in several corals is predicted to be an OB fold, raising the odd possibility of an OB fold within an OB fold (Supplemental Fig. S8B). Structural analysis may therefore reveal a matryoshka doll version of POT1 in these animals. Despite the variability of the insert in the C-terminal OB fold, it always contains cysteine residues that coordinate Zn^{2+} in conjunction with cysteine residues in the N-terminal half of the third OB fold (Fig. 6E).

Based on these findings, we propose that POT1 arose as a four-OB-fold protein, containing OB-N, OB1, OB2, and the split OB3 (Fig. 6F). Most likely, the original eukaryotic POT1 evolved from a duplication of a precursor similar to RPA70, which has four OB folds and is part of an ssDNA-binding complex (Fig. 6F; Caldwell and Spies 2020). The four-OB-fold POT1 proteins have the same domain architecture as RPA70, and in both proteins, the second and the third OB fold have high-affinity ssDNA-binding activity. Furthermore, the fourth (C-terminal) OB folds of both proteins have a Zn^{2+}-binding motif and provide binding sites for interacting partners (TPP1 for POT1, and RPA32 and RPA14 for RPA70). Consistent with the idea that POT1 arose from a four-OB-fold RPA70-like protein, there are four predicted OB folds in the POT1 of *Capsaspora owczarzaki* (Fig. 6G), a protist that is thought to be one of the closest unicellular relatives of animals (Ferrer-Bonet and Ruiz-Trillo 2017). Furthermore, there are four OB folds in POT1 ortholog TEB1, which is a component of the *Tetrahymena* telomerase holoenzyme (He et al. 2021). If the original eukaryotic POT1 indeed had four OB folds, the N-terminal OB fold must have been lost in some fungi and ciliates, since there are only three OB folds in *Schizosaccharomyces pombe* pot1 and *Oxytricha nova* TEBPα (Horvath et al. 1998; Lloyd et al. 2016).

Duplication of POT1 in the rodent lineage

Rats and mice have two POT1 proteins at their telomeres, POT1a and POT1b, whose functions have diverged. POT1b recruits the CST complex to telomeres while POT1a serves as the primary repressor of ATR signaling (Hockemeyer et al. 2006; Denchi and de Lange 2007; Wu

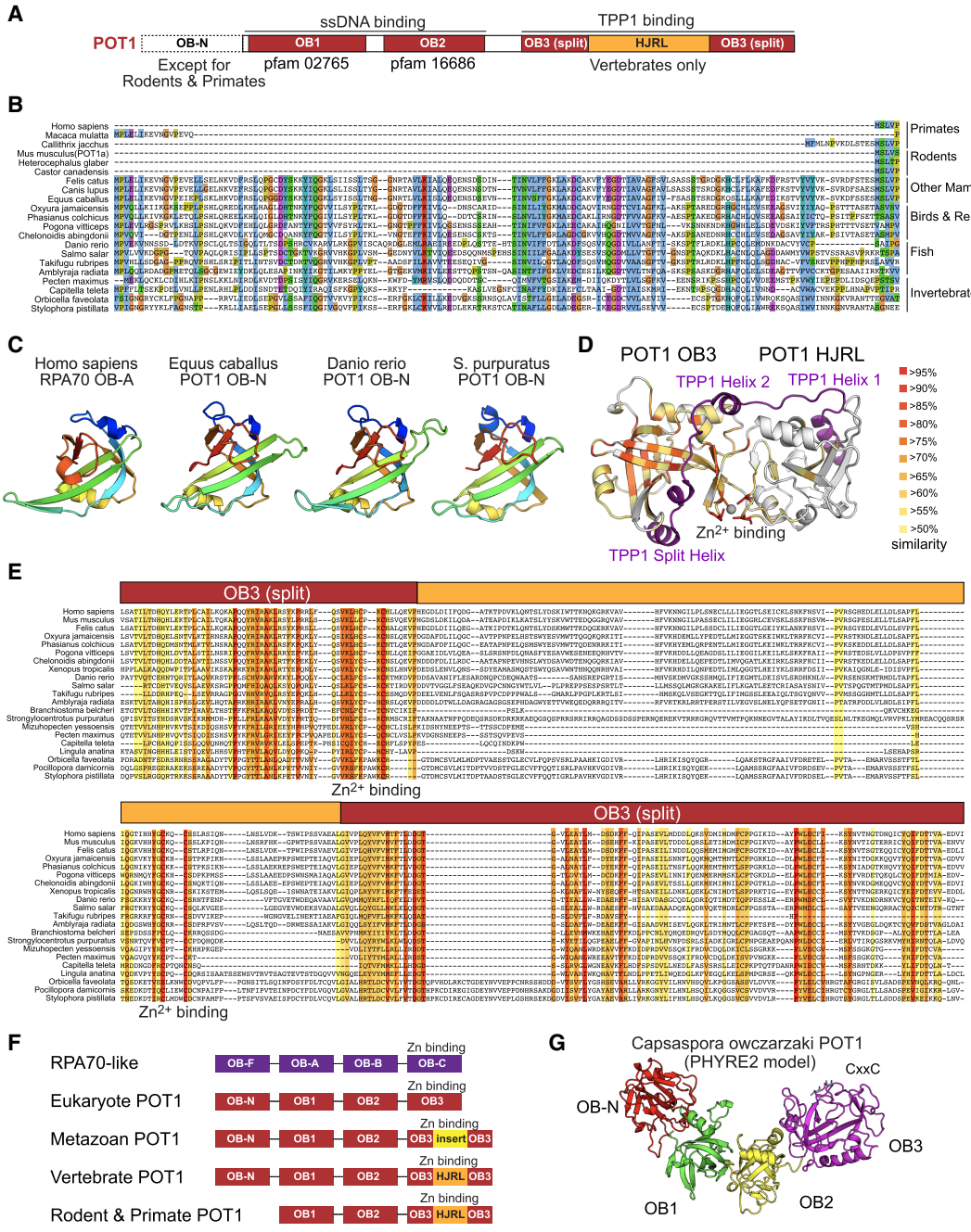

Figure 6. A predicted fourth OB fold in most metazoan POT1 orthologs. (*A*) Domain architecture of human POT1, with the N-terminal OB fold (OB-N), which is absent in rodents and primates, indicated by a dotted box. Pfam database numbers for OB1 and OB2 of POT1 are indicated. The HJRL domain, specific to vertebrates, is depicted in orange. (*B*) MUSCLE alignment with ClustalX coloring of the N-terminal OB fold region (OB-N) of POT1 from representative species. (*C*) Predicted OB fold structures of the N-terminal POT1 sequences of the three indicated species. Templates were selected by alignment by SWISS-MODEL, and the sequence of the input proteins was modeled against the known template structure. All three were modeled against OB-A of human RPA70 (PDB 1FGU; shown at the *left*) with a Qmean score greater than −4.00, indicating a high-quality estimation. Structures are colored in PyMOL using rainbow spectrum coloring to indicate the position along the sequence from the N (red) to the C terminus (violet). (*D*) Structure of the human POT1 OB3/HJRL in complex with the RD/POT1-binding domain (violet) of TPP1 (PDB 5UN7). Helices of TPP1 are indicated. POT1 is colored by sequence similarity according to the BLOSUM62 matrix as indicated in the legend. The four cysteine residues that coordinate a single Zn^{2+} ion (gray) are shown as sticks. (*E*) MUSCLE alignment with ClustalX coloring of the POT1 OB3/HJRL C termini of representative species colored as in *D*. The split OB3 (red) and HJRL insert (orange) are indicated *above*. Residues that coordinate Zn^{2+} are indicated. (*F*) Model for the evolution of POT1 from an RPA70-like protein with four OB folds. (*G*) Template-modeled structure of the full-length *Capsaspora owczarzaki* POT1 generated using intensive PHYRE2 modeling. Four distinct OB fold-like domains are indicated in red (OB-N), green (OB1), yellow (OB2), and violet (OB3). The C-x-x-C motif in OB3 is indicated.

et al. 2012). Our analysis of the POT1 genes in rodents shows that this duplication occurred during the divergence of Muridae (e.g., rats, mice, and gerbils) and Cricetidae (e.g., voles, lemmings, and hamsters) from Spalacidae (mole rats) and Dipodidae (jerboas) (Supplemental Fig. S9A,B). POT1b binds to CST with two sequence motifs that are distinct from POT1a (Supplemental Fig. S9C; Wu et al. 2012). In the rodents with a single POT1, this protein has the C-terminal motif that allows POT1b to bind CST. Therefore, it is likely that the CST binding ability was the ancestral state and that POT1a lost this feature.

Another difference between shelterin in mice and rats versus humans is that these rodents lack a tankyrase 1-binding site in the N terminus of TRF1 (Smith et al. 1998; Donigian and de Lange 2007). Remarkably, the presence of a tankyrase 1-binding site in rodent TRF1 orthologs is closely correlated with the presence of a single POT1 gene (Supplemental Fig. S9D). Thus, during the evolution of the Muridae and Cricetidae, the interaction between TRF1 and tankyrase 1 was lost, and POT1 was duplicated. Whether these two changes are functionally related remains to be determined.

Discussion

The data presented here illuminate how shelterin evolved in metazoans, revealing unanticipated modifications of the six shelterin subunits (summarized in Fig. 7). Although it remains to be determined whether the invertebrate shelterin subunits identified here actually bind to and function at telomeres, our findings indicate that in invertebrates, shelterin contains a single TRF. These invertebrate TRFs are more like TRF2 than TRF1, arguing that the original shelterin was built on a TRF2-like factor. In addition to the TRFH and Myb domains that define the TRF proteins, most metazoan TRFs contain an iDDR-like segment immediately upstream of the Myb domain. Given the conservation of this feature, it will be of interest to understand the full spectrum of functions associated with the iDDR. It will also be important to understand whether the role of TRF1 in promoting the replication of telomeric DNA is a feature it shares with the ancestral TRF or a novelty.

The duplication that gave rise to TRF1 and TRF2 occurred during the divergence of vertebrates. Following this duplication, several nonshelterin proteins (e.g., PNUTS, Apollo, and SLX4) evolved the Y/H-x-L-x-P motif, which interacts with the common protein interaction site in the TRFH domain of TRF2. PNUTS was the first and SLX4 was the last protein to become shelterin accessory factors by gaining the TRF2-binding site. The distinct N-terminal domains of TRF1 and TRF2, which are characterized by acidic and basic residues, respectively, did not appear until the divergence of mammals. Whereas the function of the acidic domain of TRF1 is not clear, the Basic/GAR domain of TRF2 is critical for the protection of the t-loop from inappropriate processing by branch migration and Holliday junction resolution (Wang et al. 2004; Saint-Léger et al. 2014; Rai et al. 2016; Schmutz et al. 2017). Possibly, this protective function is mediated by a TRF2-interacting protein in nonmammalian vertebrates.

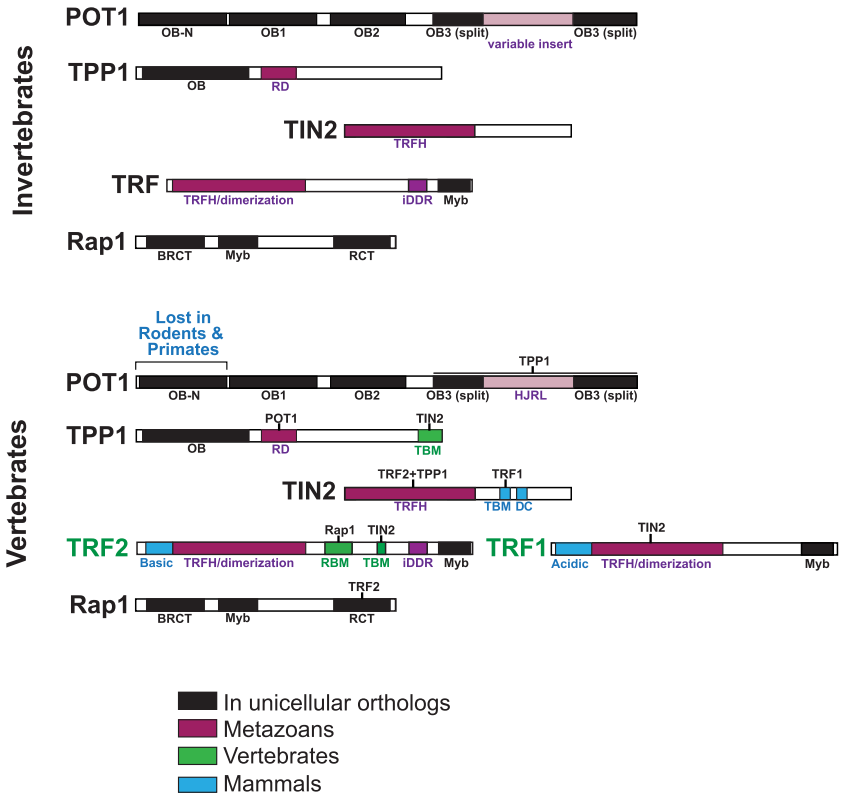

Figure 7. Summary of the evolution of metazoan shelterin. (*Top*) Schematic of the five-subunit invertebrate shelterin complex and its conserved domains. (*Bottom*) Schematic of the six-subunit vertebrate shelterin, highlighting the domains absent from the invertebrate shelterin subunits. The indicated interactions are observed in human and mouse shelterin and have not been verified in most other vertebrates. The colors of the domains are clarified in the color legend.

TIN2 is present in all metazoan shelterin complexes, and its TRFH domain, where TRF2 and TPP1 interact, is highly conserved. In contrast, the C-terminal half of TIN2, including the TRF1-binding site and the neighboring cluster of amino acids mutated in DC, is poorly conserved outside mammals. However, the interaction of *Xenopus* TIN2 and TRF1 suggests that TRF1 can bind to TIN2 in the absence of the TRF1-binding motif defined in mammals (Vizlin-Hodzic et al. 2009).

TPP1 has three functional domains that are well conserved in all metazoans: the N-terminal OB fold where telomerase binds, the RD where POT1 binds, and the TIN2-binding site at its C terminus. The region between the POT1- and TIN2-binding sites is highly divergent in sequence and shows remarkable variation in length. This part of TPP1 may represent a flexible linker allowing POT1 and telomerase to be positioned in different orientations vis-à-vis the core of shelterin.

Finally, the data reveal important aspects of metazoan POT1. Our data indicate that most metazoans have POT1 proteins with four rather than three OB folds, suggesting that POT1 evolved from a precursor that had an additional OB fold (OB-N) at its N terminus. We propose that POT1 evolved from a protein related to RPA70, the largest subunit of the RPA ssDNA-binding complex. RPA70 has four OB folds and, like POT1, interacts with its binding partners through a C-terminal OB fold with a Zn^{2+}-binding site. OB-N was lost in primates and rodents, and it will be of interest to determine the function of OB-N in other vertebrates. The most C-terminal OB fold (OB3), which is split by the insertion of an HJRL domain in vertebrates, is also split in invertebrates. While the nature of the inserted sequence varies, the Zn^{2+}-binding feature is conserved. This suggests that it is the split nature of OB3 and Zn^{2+} binding, rather than the added sequence, that is functionally significant.

These data show that the basic architecture of shelterin —including the presence of duplex and single-stranded DNA-binding factors and subunits that allow the DNA-binding proteins to form a complex—is highly conserved in metazoans. Nonetheless, shelterin underwent major innovations, especially during vertebrate evolution, where one subunit was duplicated and others lost or gained entire domains. The only shelterin subunit that changed little during metazoan evolution is Rap1. Although Rap1 is the most conserved shelterin subunit and has a counterpart at telomeres of unicellular organisms, it makes relatively little contribution to shelterin function in mammals.

It will be of interest to further trace the evolution of shelterin subunits in unicellular organisms to understand how this complex emerged from proteins that provided its basic building blocks: Myb domains, TRFH domains, and OB folds.

Materials and methods

Identification of genes encoding shelterin subunits

All accession numbers used in this study are listed in Supplemental Table S1. Metazoan shelterin subunits were identified using BLASTp and PSI-BLAST searches of the nonredundant protein sequence (nr) database. A core set of 11 vertebrate species (representing mammals, reptiles, birds, amphibians, and fish) that contained all six shelterin subunits was identified and used as a starting point for alignments. Invertebrate TRF orthologs were identified based on searches against the TRFH and Myb domains of TRF1 and TRF2. Other shelterin subunits were identified in these organisms by searching for sequences homologous to the BRCT, Myb, and/or RCT domains of human Rap1; the TRFH domain of human TIN2; the OB fold of human TPP1; the OB1–OB2 (ssDNA-binding) domains of human POT1; and the respective full-length proteins. An *E*-value threshold of 0.005 was used as a cutoff for alignment. Shelterin subunits from additional species in certain clades (e.g., primates and rodents) were identified in representative species chosen by intermediate evolutionary distance and high annotation quality. DNA sequences of POT1 exons 1–6 of *Homo sapiens* and *Felis catus* were identified from the EMBL-EBI Ensembl database and compared using SnapGene.

Alignments

Protein sequences obtained from BLAST were aligned using multiple sequence comparison by log expectation (MUSCLE) in either SnapGene (GSL Biotech) or Jalview (Waterhouse et al. 2009). To define individual domains, the full-length sequences were aligned, and then sections of the alignment were realigned to refine segments misaligned due to gaps. Alignments were colored using ClustalX unless otherwise indicated. Domains were identified by alignment to the corresponding *Homo sapiens* sequences. Some alignments were confirmed using secondary structure prediction using JPred4 (Drozdetskiy et al. 2015) or RaptorX (Källberg et al. 2014). For visualizing the variable linker region of TPP1, the NCBI Multiple Sequence Alignment (MSA) viewer was used by importing a MUSCLE alignment of the TPP1 proteins and set to color by "conservation" in Jalview. Helix propensity was visualized in Jalview using original values described by Chou and Fasman (1978). JPred secondary structure prediction of the TPP1 linker was performed in Jalview.

Phylogenetic tree inference and evolutionary time-scale analysis

Phylogenetic trees were inferred using MrBayes version 3.2 (Ronquist et al. 2012). Alignments were created using MUSCLE and formatted as a NEXUS file. MrBayes sampled across fixed amino acid rate matrices using one to three chains per analysis. After 10^5–10^6 generations, the models converged with a low standard deviation and were summarized into a tree format file. Trees were visualized in FigTree. For species-level analysis, TimeTree (Hedges et al. 2015; Kumar et al. 2017) was used.

iDDR logo and local pI analysis

An iDDR sequence logo was generated using WebLogo (Crooks et al. 2004). A custom script for analyzing the local pI of a protein sequence was written in R. A 20-amino-acid sliding window was used to calculate the pI at a given position using the EMBOSS pK scale for each amino acid.

Structure viewing and model building

Published structures were downloaded from the Protein Data Bank and visualized in PyMOL. The following PDB structures were used: 3K6G (TRF2-Rap1), 5UN7 (POT1-TPP1), and 1FGU [RPA70(OB-A)]. Protein modeling was run in PHYRE2 (Kelley

et al. 2015) or SWISS-MODEL (Bordoli et al. 2009). For the *Capsaspora owczarzaki* model in Figure 6F and for verification of other four-OB-fold POT1 species (data not shown), PHYRE2 modeling was run in "intensive" mode to model the full protein structure rather than simply the well-aligned part. Identification of HJRL insert sequences was performed by first aligning the sequences of OB3 and then modeling the insert sequence in PHYRE2.

Mapping conservation of the POT1 OB3-HJRL onto the structure

The POT1 OB3-HJRL was extracted from the POT1 OB3-HJRL/TPP1 RD(PBM) structure (PDB 5UN7), and an alignment matching the sequence was generated in Jalview. A custom Perl script calculated the sequence similarity scores at each position using the BLOSUM62 matrix. These similarity scores were mapped onto the structure using the indicated cutoff values for each color. The colors were further mapped onto the sequence alignment using a custom script to output color commands and then visualized on the alignment using AlScript (Barton 1993).

Cell culture and expression constructs

MEFs were cultured in Dulbecco's modified Eagle medium (DMEM; Corning) supplemented with 15% fetal bovine serum (FBS; Gibco), nonessential amino acids (Gibco), 2 mM L-glutamine (Gibco), 100 U/mL penicillin, and 100 µg/mL streptomycin (Gibco). Phoenix cells were cultured in DMEM supplemented with 10% bovine calf serum (BCS). The doxycycline-inducible transactivator rtTA was delivered into MEFs using pQCXIB (Clontech) retroviral vector. Nontherian TRF1 cDNAs (*Anolis carolinensis* ENSACAT00000014506.4 and *Ornithorhynchus anatinus* ENSOANT00000002746.3) were synthesized as gBlocks gene fragments (IDT) and cloned (alongside mouse TRF1) into pRetroX-TRE3G (Takara) using Gibson assembly adding an N-terminal Myc tag. The TRF1-expressing retroviral constructs were delivered into rtTA-MEFs by retroviral transduction. Protein expression was induced with 2 µg/mL doxycycline for 24 h.

IF-FISH and immunoblotting

IF-FISH with Myc (CST 9B11, #2276) was performed as described previously (Dimitrova et al. 2008). Digital images were acquired on a DeltaVision system (Applied Precision) equipped with a cooled charge-coupled device camera (CoolSnap QE, Photometrics), a PlanApo 60× 1.40 NA objective (Olympus America, Inc.), and SoftWoRx software. For immunoblotting, whole-cell lysates were prepared using RIPA buffer (150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris-HCl at pH 8) containing protease inhibitors (100 µM PMSF, 1 mM benzamidine, 2.5 µg/mL pepstatin A, 10 µg/mL leupeptin, 10 µg/mL aprotinin) and phosphatase inhibitors (1 mM each NaF, Na_3VO_4, $Na_2P_2O_7$). Proteins were fractionated by SDS/PAGE and blotted onto membranes. Membranes were blocked in 5% milk/0.1% Tween-20 in TBS (TBST), incubated in primary antibody for 1 h at room temperature, washed three times in TBST, and incubated for 1 h at room temperature with secondary antibodies. After three washes in TBST, membranes were developed using enhanced chemiluminescence (Amersham).

Competing interest statement

T.d.L. is a member of the Scientific Advisory Board of Calico LLC and a venture partner of Catalio.

Acknowledgments

We thank Dr. Li Zhao and members of the de Lange laboratory for helpful comments on this manuscript. This work is supported by grants to T.d.L. from the National Institutes of Health (NIH; R35 CA210036 and AG016642). L.R.M. is supported by the National Cancer Institute of the NIH (K00CA212452). C.G.K. is supported by an F30 Predoctoral Fellowship from the National Cancer Institute of the National Institutes of Health under award number F30CA257419. G.Z. and C.G.K. are supported by a Medical Scientist Training Program grant from the National Institute of General Medical Sciences of the NIH under award number T32GM007739 to the Weill Cornell/Rockefeller/Sloan Kettering Tri-Institutional MD-PhD Program. S.W.C. is supported by the David Rockefeller Graduate Program and a National Science Foundation Graduate Research Fellowship Program under grant number 1946429. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Author contributions: The study was conceived by T.d.L, L.R.M., and C.G.K. L.R.M. and C.G.K executed most of the in silico analysis. G.Z. contributed to the analysis of the iDDR. S.W.C. contributed to the analysis of OB3 of POT1. N.K.S. performed the in vivo analysis of platypus and lizard TRF1s. T.d.L. wrote the paper with input of all coauthors.

References

Afrin M, Kishmiri H, Sandhu R, Rabbani MAG, Li B. 2020. Trypanosoma brucei RAP1 has essential functional domains that are required for different protein interactions. *mSphere* **5:** e00027-20. doi:10.1128/mSphere.00027-20

Barton GJ. 1993. AlScript: a tool to format multiple sequence alignments. *Protein Eng* **6:** 37–40. doi:10.1093/protein/6.1.37

Bassham S, Beam A, Shampay J. 1998. Telomere variation in *Xenopus laevis*. *Mol Cell Biol* **18:** 269–275. doi:10.1128/MCB.18.1.269

Baumann P, Cech TR. 2001. Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* **292:** 1171–1175. doi:10.1126/science.1060036

Bianchi A, Smith S, Chong L, Elias P, de Lange T. 1997. TRF1 is a dimer and bends telomeric DNA. *EMBO J* **16:** 1785–1794. doi:10.1093/emboj/16.7.1785

Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. 2009. Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* **4:** 1–13. doi:10.1038/nprot.2008.197

Caldwell CC, Spies M. 2020. Dynamic elements of replication protein A at the crossroads of DNA replication, recombination, and repair. *Crit Rev Biochem Mol Biol* **55:** 482–507. doi:10.1080/10409238.2020.1813070

Celli GB, Denchi EL, de Lange T. 2006. Ku70 stimulates fusion of dysfunctional telomeres yet protects chromosome ends from homologous recombination. *Nat Cell Biol* **8:** 885–890. doi:10.1038/ncb1444

Chen Y, Yang Y, van Overbeek M, Donigian JR, Baciu P, de Lange T, Lei M. 2008. A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. *Science* **319:** 1092–1096. doi:10.1126/science.1151804

Chen Y, Rai R, Zhou ZR, Kanoh J, Ribeyre C, Yang Y, Zheng H, Damay P, Wang F, Tsujii H, et al. 2011. A conserved motif within RAP1 has diversified roles in telomere protection and

regulation in different organisms. *Nat Struct Mol Biol* **18:** 213–221. doi:10.1038/nsmb.1974
Chen C, Gu P, Wu J, Chen X, Niu S, Sun H, Wu L, Li N, Peng J, Shi S, et al. 2017. Structural insights into POT1-TPP1 interaction and POT1 C-terminal mutations in human cancer. *Nat Commun* **8:** 14929. doi:10.1038/ncomms14929
Chou PY, Fasman GD. 1978. Empirical predictions of protein conformation. *Annu Rev Biochem* **47:** 251–276. doi:10.1146/annurev.bi.47.070178.001343
Court R, Chapman L, Fairall L, Rhodes D. 2005. How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures. *EMBO Rep* **6:** 39–45. doi:10.1038/sj.embor.7400314
Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. Weblogo: a sequence logo generator. *Genome Res* **14:** 1188–1190. doi:10.1101/gr.849004
de Lange T. 2005. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev* **19:** 2100–2110. doi:10.1101/gad.1346005
de Lange T. 2009. How telomeres solve the end-protection problem. *Science* **326:** 948–952. doi:10.1126/science.1170633
de Lange T. 2015. A loopy view of telomere evolution. *Front Genet* **6:** 321. doi:10.3389/fgene.2015.00321
de Lange T. 2018. Shelterin-mediated telomere protection. *Annu Rev Genet* **52:** 223–247. doi:10.1146/annurev-genet-032918-021921
Demuth I, Digweed M, Concannon P. 2004. Human SNM1B is required for normal cellular response to both DNA interstrand crosslink-inducing agents and ionizing radiation. *Oncogene* **23:** 8611–8618. doi:10.1038/sj.onc.1207895
Denchi EL, de Lange T. 2007. Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Nature* **448:** 1068–1071. doi:10.1038/nature06065
De Rycker M, Venkatesan RN, Wei C, Price CM. 2003. Vertebrate tankyrase domain structure and sterile α motif (SAM)-mediated multimerization. *Biochem J* **372:** 87–96. doi:10.1042/bj20021450
Dimitrova N, Chen YC, Spector DL, de Lange T. 2008. 53BP1 promotes non-homologous end joining of telomeres by increasing chromatin mobility. *Nature* **456:** 524–528. doi:10.1038/nature07433
Diotti R, Loayza D. 2011. Shelterin complex and associated factors at human telomeres. *Nucleus* **2:** 119–135. doi:10.4161/nucl.2.2.15135
Doksani Y, Wu JY, de Lange T, Zhuang X. 2013. Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation. *Cell* **155:** 345–356. doi:10.1016/j.cell.2013.09.048
Donigian JR, de Lange T. 2007. The role of the poly(ADP-ribose) polymerase tankyrase1 in telomere length control by the TRF1 component of the shelterin complex. *J Biol Chem* **282:** 22662–22667. doi:10.1074/jbc.M702620200
Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* **43:** W389–W394. doi:10.1093/nar/gkv332
Fairall L, Chapman L, Moss H, de Lange T, Rhodes D. 2001. Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2. *Mol Cell* **8:** 351–361. doi:10.1016/S1097-2765(01)00321-5
Ferrer-Bonet M, Ruiz-Trillo I. 2017. *Capsaspora owczarzaki*. *Curr Biol* **27:** R829–R830. doi:10.1016/j.cub.2017.05.074
Fulcher N, Derboven E, Valuchova S, Riha K. 2014. If the cap fits, wear it: an overview of telomeric structures over evolution. *Cell Mol Life Sci* **71:** 847–865. doi:10.1007/s00018-013-1469-z
Gaullier G, Miron S, Pisano S, Buisson R, Le Bihan YV, Tellier-Lebègue C, Messaoud W, Roblin P, Guimarães BG, Thai R, et al. 2016. A higher-order entity formed by the flexible assembly of RAP1 with TRF2. *Nucleic Acids Res* **44:** 1962–1976. doi:10.1093/nar/gkv1531
Ghanim GE, Fountain AJ, van Roon AM, Rangan R, Das R, Collins K, Nguyen THD. 2021. Structure of human telomerase holoenzyme with bound telomeric DNA. *Nature* **593:** 449–453. doi:10.1038/s41586-021-03415-4
Gong Y, de Lange T. 2010. A Shld1-controlled POT1a provides support for repression of ATR signaling at telomeres through RPA exclusion. *Mol Cell* **40:** 377–387. doi:10.1016/j.molcel.2010.10.016
Griffith JD, Comeau L, Rosenfield S, Stansel RM, Bianchi A, Moss H, de Lange T. 1999. Mammalian telomeres end in a large duplex loop. *Cell* **97:** 503–514. doi:10.1016/S0092-8674(00)80760-6
Grill S, Tesmer VM, Nandakumar J. 2018. The N terminus of the OB domain of telomere protein TPP1 is critical for telomerase action. *Cell Rep* **22:** 1132–1140. doi:10.1016/j.celrep.2018.01.012
Grill S, Padmanaban S, Friedman A, Perkey E, Allen F, Tesmer VM, Chase J, Khoriaty R, Keegan CE, Maillard I, et al. 2021. TPP1 mutagenesis screens unravel shelterin interfaces and functions in hematopoiesis. *JCI Insight* **6:** 138059. doi:10.1172/jci.insight.138059
Guo X, Deng Y, Lin Y, Cosme-Blanco W, Chan S, He H, Yuan G, Brown EJ, Chang S. 2007. Dysfunctional telomeres activate an ATM-ATR-dependent DNA damage response to suppress tumorigenesis. *EMBO J* **26:** 4709–4719. doi:10.1038/sj.emboj.7601893
He Y, Wang Y, Liu B, Helmling C, Sušac L, Cheng R, Zhou ZH, Feigon J. 2021. Structures of telomerase at several steps of telomere repeat synthesis. *Nature* **593:** 454–459. doi:10.1038/s41586-021-03529-9
Hedges SB, Marin J, Suleski M, Paymer M, Kumar S. 2015. Tree of life reveals clock-like speciation and diversification. *Mol Biol Evol* **32:** 835–845. doi:10.1093/molbev/msv037
Hockemeyer D, Collins K. 2015. Control of telomerase action at human telomeres. *Nat Struct Mol Biol* **22:** 848–852. doi:10.1038/nsmb.3083
Hockemeyer D, Daniels JP, Takai H, de Lange T. 2006. Recent expansion of the telomeric complex in rodents: two distinct POT1 proteins protect mouse telomeres. *Cell* **126:** 63–77. doi:10.1016/j.cell.2006.04.044
Horvath MP, Schweiker VL, Bevilacqua JM, Ruggles JA, Schultz SC. 1998. Crystal structure of the Oxytricha nova telomere end binding protein complexed with single strand DNA. *Cell* **95:** 963–974. doi:10.1016/S0092-8674(00)81720-1
Hu C, Rai R, Huang C, Broton C, Long J, Xu Y, Xue J, Lei M, Chang S, Chen Y. 2017. Structural and functional analyses of the mammalian TIN2–TPP1–TRF2 telomeric complex. *Cell Res* **27:** 1485–1502. doi:10.1038/cr.2017.144
Kabir S, Sfeir A, de Lange T. 2010. Taking apart Rap1: an adaptor protein with telomeric and non-telomeric functions. *Cell Cycle* **9:** 4061–4067. doi:10.4161/cc.9.20.13579
Kabir S, Hockemeyer D, de Lange T. 2014. TALEN gene knockouts reveal no requirement for the conserved human shelterin protein Rap1 in telomere protection and length regulation. *Cell Rep* **9:** 1273–1280. doi:10.1016/j.celrep.2014.10.014
Källberg M, Margaryan G, Wang S, Ma J, Xu J. 2014. RaptorX server: a resource for template-based protein structure modeling. In: *Protein structure prediction* (ed. Kihara D), pp. 17–27. Humana Press, New York. doi:10.1007/978-1-4939-0366-5_2

Kappei D, Scheibe M, Paszkowski-Rogacz M, Bluhm A, Gossmann TI, Dietz S, Dejung M, Herlyn H, Buchholz F, Mann M, et al. 2017. Phylointeractomics reconstructs functional evolution of protein binding. *Nat Commun* **8:** 14334. doi:10.1038/ncomms14334

Karlseder J, Broccoli D, Dai Y, Hardy S, de Lange T. 1999. p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. *Science* **283:** 1321–1325. doi:10.1126/science.283.5406.1321

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10:** 845–858. doi:10.1038/nprot.2015.053

Kim H, Lee OH, Xin H, Chen LY, Qin J, Chae HK, Lin SY, Safari A, Liu D, Songyang Z. 2009. TRF2 functions as a protein hub and regulates telomere maintenance by recognizing specific peptide motifs. *Nat Struct Mol Biol* **16:** 372–379. doi:10.1038/nsmb.1575

Konishi A, Izumi T, Shimizu S. 2016. TRF2 protein interacts with core histones to stabilize chromosome ends. *J Biol Chem* **291:** 20798–20810. doi:10.1074/jbc.M116.719021

Kratz K, de Lange T. 2018. Protection of telomeres 1 proteins POT1a and POT1b can repress ATR signaling by RPA exclusion, but binding to CST limits ATR repression by POT1b. *J Biol Chem* **293:** 14384–14392. doi:10.1074/jbc.RA118.004598

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. Timetree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol* **34:** 1812–1819. doi:10.1093/molbev/msx116

Lazzerini-Denchi E, Sfeir A. 2016. Stop pulling my strings—what telomeres taught us about the DNA damage response. *Nat Rev Mol Cell Biol* **17:** 364–378. doi:10.1038/nrm.2016.43

Lenain C, Bauwens S, Amiard S, Brunori M, Giraud-Panis MJ, Gilson E. 2006. The Apollo 5′ exonuclease functions together with TRF2 to protect telomeres from DNA repair. *Curr Biol* **16:** 1303–1310. doi:10.1016/j.cub.2006.05.021

Lim CJ, Cech TR. 2021. Shaping human telomeres: from shelterin and CST complexes to telomeric chromatin organization. *Nat Rev Mol Cell Biol* **22:** 283–298. doi:10.1038/s41580-021-00328-y

Lim CJ, Barbour AT, Zaug AJ, Goodrich KJ, McKay AE, Wuttke DS, Cech TR. 2020. The structure of human CST reveals a decameric assembly bound to telomeric DNA. *Science* **368:** 1081–1085. doi:10.1126/science.aaz9649

Lloyd NR, Dickey TH, Hom RA, Wuttke DS. 2016. Tying up the ends: plasticity in the recognition of single-stranded DNA at telomeres. *Biochemistry* **55:** 5326–5340. doi:10.1021/acs.biochem.6b00496

Lue NF. 2018. Evolving linear chromosomes and telomeres: a C-strand-centric view. *Trends Biochem Sci* **43:** 314–326. doi:10.1016/j.tibs.2018.02.008

Martínez P, Thanasoula M, Muñoz P, Liao C, Tejera A, McNees C, Flores JM, Fernández-Capetillo O, Tarsounas M, Blasco MA. 2009. Increased telomere fragility and fusions resulting from TRF1 deficiency lead to degenerative pathologies and increased cancer in mice. *Genes Dev* **23:** 2060–2075. doi:10.1101/gad.543509

Nakamura TM, Morin GB, Chapman KB, Weinrich SL, Andrews WH, Lingner J, Harley CB, Cech TR. 1997. Telomerase catalytic subunit homologs from fission yeast and human. *Science* **277:** 955–959. doi:10.1126/science.277.5328.955

Nandakumar J, Bell CF, Weidenfeld I, Zaug AJ, Leinwand LA, Cech TR. 2012. The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity. *Nature* **492:** 285–289. doi:10.1038/nature11648

Nguyen THD, Collins K, Nogales E. 2019. Telomerase structures and regulation: shedding light on the chromosome end. *Curr Opin Struct Biol* **55:** 185–193. doi:10.1016/j.sbi.2019.04.009

Okamoto K, Bartocci C, Ouzounov I, Diedrich JK, Yates JR III, Denchi EL. 2013. A two-step mechanism for TRF2-mediated chromosome-end protection. *Nature* **494:** 502–505. doi:10.1038/nature11873

O'Sullivan RJ, Karlseder J. 2010. Telomeres: protecting chromosomes against genome instability. *Nat Rev Mol Cell Biol* **11:** 171–181. doi:10.1038/nrm2848

Poulet A, Pisano S, Faivre-Moskalenko C, Pei B, Tauran Y, Haftek-Terreau Z, Brunet F, Le Bihan YV, Ledu MH, Montel F, et al. 2012. The N-terminal domains of TRF1 and TRF2 regulate their ability to condense telomeric DNA. *Nucleic Acids Res* **40:** 2566–2576. doi:10.1093/nar/gkr1116

Rai R, Chen Y, Lei M, Chang S. 2016. TRF2-RAP1 is required to protect telomeres from engaging in homologous recombination-mediated deletions and fusions. *Nat Commun* **7:** 10881. doi:10.1038/ncomms10881

Rice C, Shastrula PK, Kossenkov AV, Hills R, Baird DM, Showe LC, Doukov T, Janicki S, Skordalakes E. 2017. Structural and functional analysis of the human POT1-TPP1 telomeric complex. *Nat Commun* **8:** 14928. doi:10.1038/ncomms14928

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. Mrbayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61:** 539–542. doi:10.1093/sysbio/sys029

Saint-Léger A, Koelblen M, Civitelli L, Bah A, Djerbi N, Giraud-Panis MJ, Londoño-Vallejo A, Ascenzioni F, Gilson E. 2014. The basic N-terminal domain of TRF2 limits recombination endonuclease action at human telomeres. *Cell Cycle* **13:** 2469–2474. doi:10.4161/cc.29422

Savage SA. 2018. Beginning at the ends: telomeres and human disease. *F1000Res* **7:** 524. doi:10.12688/f1000research.14068.1

Schmutz I, Timashev L, Xie W, Patel DJ, de Lange T. 2017. TRF2 binds branched DNA to safeguard telomere integrity. *Nat Struct Mol Biol* **24:** 734–742. doi:10.1038/nsmb.3451

Sexton AN, Regalado SG, Lai CS, Cost GJ, O'Neil CM, Urnov FD, Gregory PD, Jaenisch R, Collins K, Hockemeyer D. 2014. Genetic and molecular identification of three human TPP1 functions in telomerase action: recruitment, activation, and homeostasis set point regulation. *Genes Dev* **28:** 1885–1899. doi:10.1101/gad.246819.114

Sfeir A, Kosiyatrakul ST, Hockemeyer D, MacRae SL, Karlseder J, Schildkraut CL, de Lange T. 2009. Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell* **138:** 90–103. doi:10.1016/j.cell.2009.06.021

Shore D. 1994. RAP1: a protean regulator in yeast. *Trends Genet* **10:** 408–412. doi:10.1016/0168-9525(94)90058-2

Smith S, Giriat I, Schmitt A, de Lange T. 1998. Tankyrase, a poly (ADP-ribose) polymerase at human telomeres. *Science* **282:** 1484–1487. doi:10.1126/science.282.5393.1484

Smogorzewska A, Karlseder J, Holtgreve-Grez H, Jauch A, de Lange T. 2002. DNA ligase IV-dependent NHEJ of deprotected mammalian telomeres in G1 and G2. *Curr Biol* **12:** 1635–1644. doi:10.1016/S0960-9822(02)01179-X

Timashev LA, de Lange T. 2020. Characterization of t-loop formation by TRF2. *Nucleus* **11:** 164–177. doi:10.1080/19491034.2020.1783782

van Overbeek M, de Lange T. 2006. Apollo, an Artemis-related nuclease, interacts with TRF2 and protects human telomeres in S phase. *Curr Biol* **16:** 1295–1302. doi:10.1016/j.cub.2006.05.022

van Steensel B, de Lange T. 1997. Control of telomere length by the human telomeric protein TRF1. *Nature* **385:** 740–743. doi:10.1038/385740a0

van Steensel B, Smogorzewska A, de Lange T. 1998. TRF2 protects human telomeres from end-to-end fusions. *Cell* **92:** 401–413. doi:10.1016/S0092-8674(00)80932-0

Vizlin-Hodzic D, Ryme J, Simonsson S, Simonsson T. 2009. Developmental studies of *Xenopus* shelterin complexes: the message to reset telomere length is already present in the egg. *FASEB J* **23:** 2587–2594. doi:10.1096/fj.09-129619

Wang RC, Smogorzewska A, de Lange T. 2004. Homologous recombination generates T-loop-sized deletions at human telomeres. *Cell* **119:** 355–368. doi:10.1016/j.cell.2004.10.011

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

Wu L, Multani AS, He H, Cosme-Blanco W, Deng Y, Deng JM, Bachilo O, Pathak S, Tahara H, Bailey SM, et al. 2006. Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres. *Cell* **126:** 49–62. doi:10.1016/j.cell.2006.05.037

Wu P, Takai H, de Lange T. 2012. Telomeric 3′ overhangs derive from resection by Exo1 and Apollo and fill-in by POT1b-associated CST. *Cell* **150:** 39–52. doi:10.1016/j.cell.2012.05.026

Wu RA, Upton HE, Vogan JM, Collins K. 2017. Telomerase mechanism of telomere synthesis. *Annu Rev Biochem* **86:** 439–460. doi:10.1146/annurev-biochem-061516-045019

Xie Y, Yang D, He Q, Songyang Z. 2011. Zebrafish as a model system to study the physiological function of telomeric protein TPP1. *PLoS One* **6:** e16440. doi:10.1371/journal.pone.0016440

Xue J, Chen H, Wu J, Takeuchi M, Inoue H, Liu Y, Sun H, Chen Y, Kanoh J, Lei M. 2017. Structure of the fission yeast S. pombe telomeric Tpz1–Poz1–Rap1 complex. *Cell Res* **27:** 1503–1520. doi:10.1038/cr.2017.145

Zhong FL, Batista LF, Freund A, Pech MF, Venteicher AS, Artandi SE. 2012. TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends. *Cell* **150:** 481–494. doi:10.1016/j.cell.2012.07.012