IDENTIFICATION AND CHARACTERIZATION OF TELOMERASE COMPONENTS IN PLANTS

A Dissertation

by

JIARUI SONG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee, Jorge Cruz-Reyes
Committee Members, Dorothy E. Shippen
Junjie Zhang
Wayne K. Versaw
Head of Department, A. Joshua Wand

May 2021

Major Subject: Biochemistry

ABSTRACT

Telomerase is a specialized ribonucleoprotein (RNP) complex responsible for telomere maintenance. Telomerase consists of three essential components: the catalytic subunit telomerase reverse transcriptase TERT, the telomerase RNA subunit TR, and telomerase accessory proteins. Studies of telomerase aim to define the components of telomerase RNP complexes and characterize their relations in terms of biogenesis, subunit interactions, and structures, which all provide significant implications for cellular aging and cancer.

In the plant kingdom, TERT was defined 20 years ago in the model organism *Arabidopsis thaliana*. However, due to technological limitations, the precise RNA component and the complete list of accessory proteins are unknown. In this dissertation, I describe the identification and characterization of the *bona fide* telomerase RNA, termed AtTR, in *A. thaliana*, which lead to the discovery of 85 TR from three major clades of plants with a conserved secondary structure serving as an evolutionary bridge to connect divergent TRs in vertebrates and ciliates. In addition to TR, I characterized the previously identified telomerase-associated proteins, AtPOT1a and dyskerin. I found that AtPOT1a physically associates with TERT and telomeric DNA to promote telomerase recruitment on telomeres. I also detected a specific interaction between dyskerin and AtTR mediated by a unique three-way junction element supported by cryo-EM analysis. This study reinforced the conclusion that plant telomerase is an evolutionary bridge by presenting a chimeric RNP complex with ciliate-like AtTR and vertebrate-like telomerase-associated protein dyskerin and AtPOT1a. Furthermore, we re-evaluated a previously defined telomerase-associated RNA molecule AtTER2. We found that AtTER2 completely overlaps with the gene encoding tRNA Deaminase 3 (*TAD3)*, which indirectly participates in telomere maintenance. Finally, we utilized quantitative mass spectrometry to define the associated proteins of telomerase core in *A. thaliana*.

This study provided a framework to uncover telomerase-associated proteins and indications for telomerase RNP components in the plants.

In summary, the studies presented in this dissertation reveal that Arabidopsis telomerase serves as an evolutionary bridge to unite vertebrate and ciliate telomerase RNP complexes by maintaining a chimeric complex, making it an exciting model system for study of telomere and telomerase biology.

DEDICATION

This dissertation is dedicated

To my parents, for their endless love and support.

To Liqi, for the unconditional love throughout my life.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Dorothy Shippen for your diligent nature and your selfless guidance to me. You are my American mother who hosts the warm Shippen lab family. Your mentorship has allowed me to become the scientist I am today and encourage me to explore more. You trained me with critical thinking and taught me how to interact with other scientists. Through your leadership I obtained a deep understanding of the telomerase world. I am especially grateful to have you encourage me to take challenges and recover from low confidence.

I would like to thank my committee Dr. Junjie Zhang, Dr. Jorge Cruz-Reyes, and Dr. Wayne Versaw and collaborators Dr. Julian Chen, Dr. Ophelia Papoulas, and Dr. Edward Marcotte. Each of you promoted my growth as a scientist by driving my research with thought provoking questions and leading collaborations. I view you as mentors and friends.

I also would like to thank Dr. Claudia Castillo-González, Dr. Ana Victoria Suescun and Dr. Hengyi (Peter) Xu for being great mentors. I would not be the same person I am today without your support and instruction. Countless times you taught me invaluable lessons. I cannot thank Dr. Sreyashree Bose enough for sharing her expertise and answering thousands of questions in both professional and personal circumstances. A thank you to all members of the Shippen lab, you have made my life better and have also taught me to be a good scientist.

My graduate career cannot be smooth without the effort of the wonderful people surrounding me in the Biochemistry and Biophysics department. Many thanks to Tillie, Divina, and Sherry for helping me numerous times. I am grateful to be a member of BGA and to meet so many friends. You helped me so much in my personal life.

Additionally, I would like to thank my family and my friends for their support throughout my life. My parents provided an environment in which they heavily encourage educations. I would

like to thank the family I chose. I cannot be more grateful to have Liqi willing to spend the rest of life with me. To my friends Cai, Borja, Zhicheng, and Kailun, thank you for pushing me to be a better person through competition and instruction. Thanks for all the laughs and great times. You are all truly great people.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professors Jorge Cruz-Reyes, Dorothy E. Shippen and Junjie Zhang of the Department of Biochemistry and Biophysics and Professor Wayne K. Versaw of the Department of Biology.

Part of the bioinformatic data and the RNA structure model presented in Chapter II were provided by Professor Julian Chen's laboratory and were published in 2019. The cryo-EM structure in Chapter III was generated by Professor Junjie Zhang's laboratory. The analyses depicted in Chapter IV were conducted in collaboration with other Shippen lab members and were published in 2020. The quantitative mass spectrometry data in Appendix was generated in collaboration with Professor Edward Marcotte's laboratory. All other work conducted for the dissertation was completed by the student independently.

Funding sources

This work was supported by NIH Grant R01 GM065383 and NSF Grant MCB151787.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

The transition from circular chromosomes in prokaryotes to linear chromosomes in eukaryotes simultaneously led to multiple challenges posed by the new mode of genome architecture. First, the end of the chromosome cannot be fully copied by conventional DNA polymerases. These enzymes require an RNA primer for the initiation step and fail to completely replicate the 3' end of linear DNA. This leads to the progressive loss of sequences from DNA termini, a phenomenon described as the 'end-replication problem' (1, 2). Second, genetic information encoded at the ends of DNA strands is vulnerable to exonuclease attack because the ends of linear chromosomes mimic DNA double-strand breaks. Without proper protection, chromosome ends trigger the repair machinery to perform illegitimate repair reactions like end-to-end chromosome fusion (3). Thus, the 'end-protection problem' is a serious dilemma for linear genomes.

To address the 'end-replication' and 'end-protection' problems, specialized DNA sequences with designated architectures to cap chromosomal ends are essential. The telomere consists of stretches of simple repetitive DNA sequences that form the physical ends of linear chromosomes. For example, vertebrates maintain telomeric DNA with TTAGGG repeats (4), while in *Tetrahymena*, telomeres consist of approximately 50 copies of TTGGGG (5). Therefore, one strand of telomeres containing multiple G nucleotides is considered as a G-rich strand and the complementary strand is a C-rich strand. To distinguish telomeres from DNA double-strand breaks, telomeres maintain a single-stranded overhang of the G-rich strand at the 3′ end of the chromosome, which assembles into a lariat-like configuration known as the telomere-loop (t-loop). T-loops form by strand invasion of the 3′ overhang into the duplex region of the telomere tract (Figure 1A) (6).

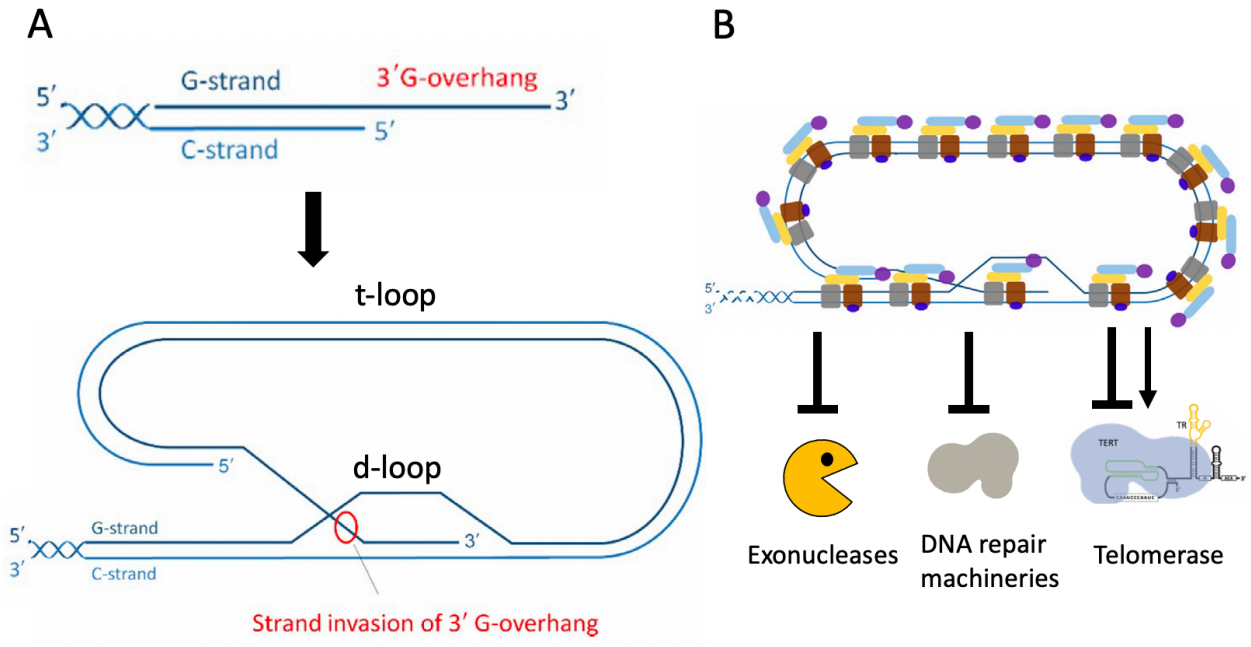


Figure 1. Telomere structure.

(A) Top: The telomere consists of a region simple G/C rich repeats of double-stranded and single-stranded DNA. The G-rich strand protrudes toward the end of the chromosome as a 3' overhang. Bottom: The telomere forms a t-loop configuration that arises by strand invasion of the telomeric 3' G-overhang into the upstream telomeric double-stranded DNA. The 3' overhang displaces one strand and anneals to other strand of DNA forming a d-loop. (B) Single-stranded and double-stranded telomere binding proteins associate with telomeric DNA to assemble into a complex that inhibits exonucleases and DNA repair activities, and also regulates access of the telomerase enzyme when telomeres are replicated.

Single-stranded and double-stranded telomere binding proteins shield the exposed chromosomal termini from exonucleases and DNA repair machineries (Figure 1B) (7, 8). Telomere and telomere-associated proteins form highly ordered and dynamic complexes responsible for telomere maintenance and length regulation (Figure 1B) (6).

In proliferative cells and unicellular organisms, telomeric DNA replication predominantly relies on telomerase as a specialized reverse transcriptase (9). Telomerase was initially identified in the Elizabeth Blackburn's laboratory from the ciliate *Tetrahymena thermophila* by the activity of telomeric repeat synthesis (10). Subsequent characterization revealed that telomerase is a ribonucleoprotein (RNP) complex containing two core components: the catalytic subunit telomerase reverse transcriptase (TERT) and the RNA subunit telomerase RNA (TR) (11, 12). TERT utilizes a template sequence embedded in TR to synthesize the telomeric repeats (Figure 2A) (13). The TR template can be dissected into two regions: an alignment region to hybridize with telomeric DNA and a templating region to dictate the telomeric repeats (Figure 2A). In addition to the template, TR harbors essential structural elements that serve as a scaffold to retain TERT and accessory proteins for telomerase's catalytic process.

A complete cycle of telomerase-mediated telomere elongation can be divided into three stages: binding, extension, and translocation (Figure 2B) (14). Thus, the specialized reverse transcriptase TERT has to secure critical TR interactions to maintain a stable RNP complex compatible for both the addition of multiple nucleotides by this single active site, and the translocation of the primer 3' terminus relative to the template sequence for multiple rounds of processive nucleotide addition (termed telomere repeat addition processivity) (15, 16).

Although TERT and TR are sufficient to reconstitute active telomerase *in vitro*, additional species-specific subunits perform significant functions in vivo, including telomerase activity and

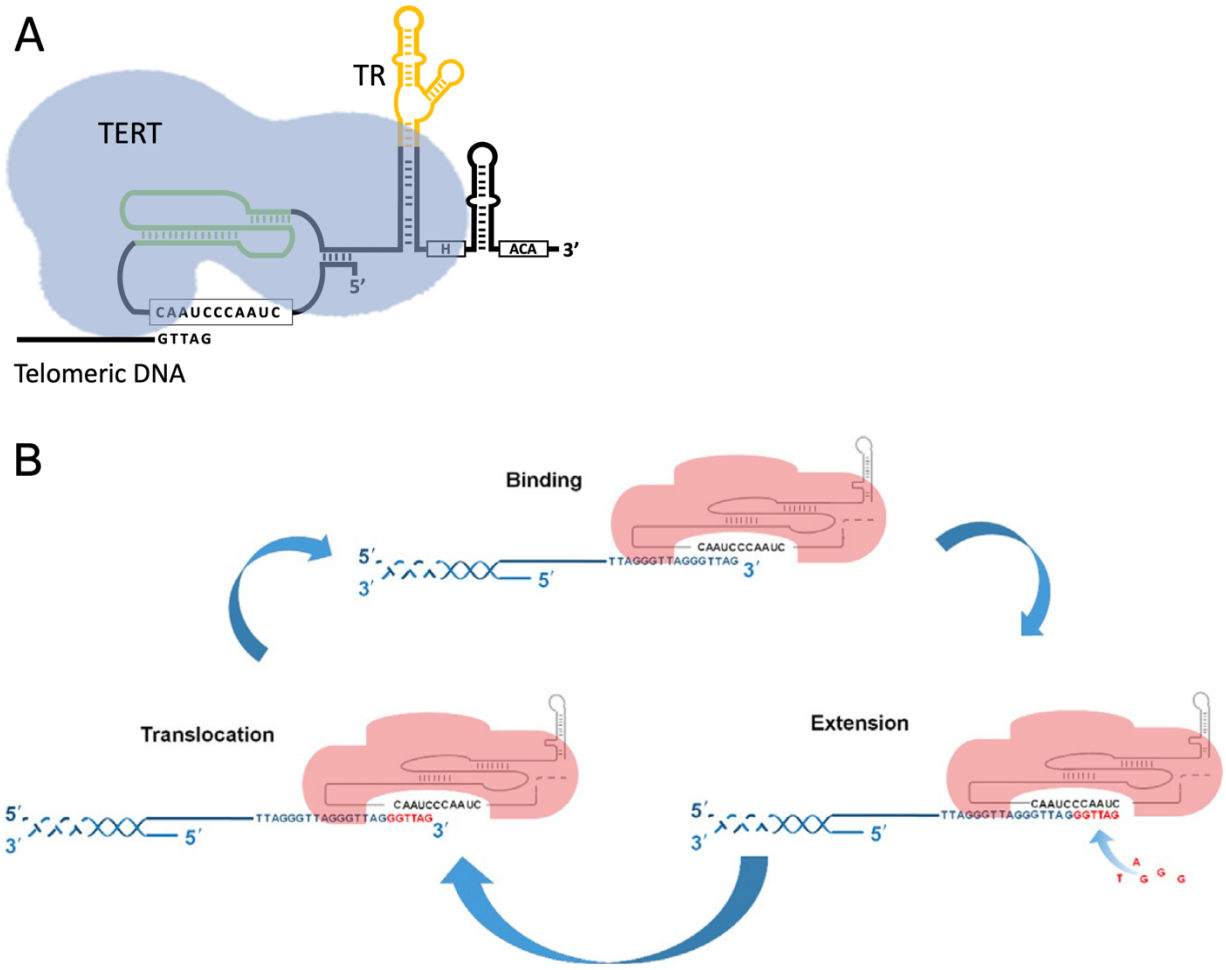


Figure 2. Telomerase is designated for telomere elongation.

(A) The core components of telomerase are the telomerase reverse transcriptase (TERT) and telomerase RNA (TR). TR contains a template for telomeric DNA synthesis. In this diagram the TR template sequence and structure are shown for vertebrate telomerase. (B) The catalytic cycle of telomerase can be divided into three steps. Initially, the telomeric 3' G-overhang binds to the TR template (Binding). Next, TERT utilizes TR template to polymerize one copy of the telomeric repeat by adding nucleotides onto the 3' end of the telomeric DNA (Extension). Then, the active site of TERT translocates to the 3' end of the newly formed telomeric repeat (Translocation). Another round of nucleotide addition is then initiated. (B) is adapted from Giardini et al. (3).

processivity regulation, telomerase recruitment, TR biogenesis, and RNP maturation (17). As discussed below, these accessory factors are highly divergent, and often lineage specific. In addition, telomerase activity, processivity, and fidelity vary among different organisms and in different developmental stages of the same organism (18). Importantly, deficiencies in telomerase components lead to profound genome instability (19). In human beings, several genetically inherited diseases result from the loss of telomerase, including aplastic anemia, dyskeratosis congenita, and idiopathic pulmonary fibrosis (20). Conversely, reactivation of telomerase is associated with the majority of human cancers (21). Therefore, precise control of telomerase is an inevitable task for most of the organisms.

Below I introduce several unique and shared features of telomerases from ciliates, yeasts, vertebrates and plants with an emphasis on the telomere complex structure, telomerase RNP structure and telomerase RNP biogenesis.

Telomere maintains a highly ordered architecture

The concept of the telomere was uncovered independently in *Drosophila* and maize from the work of Hermann Muller and Barbara McClintock, respectively (22, 23). These groundbreaking studies showed that the intact chromosomal ends were resistant to end-to-end fusion while broken chromosomes were not, which indicates that the natural ends harbor a protective structure to inhibit chromosomal rearrangements. Muller first used the term 'telomere' to define the physical ends of linear (22). We now realize that the telomere and its binding proteins assemble into a highly ordered, effective and dynamic machinery to secure the chromosomal ends from nucleolytic attack and illegitimate DNA repair (3). Telomere deprotection and insufficient maintenance culminate in chromosomal abnormalities, cell cycle arrest, and senescence (20).

Below I focus on the structure and dynamics of the telomere as a highly ordered nucleoprotein complex.

Telomeres consist of simple G-rich repeats with a 3' overhang

As mentioned above, in most eukaryotic species, the telomere is composed of stretches of tandemly repeated G-rich sequences (3). However, in some groups of fungi including *Saccharomyces cerevisiae*, the telomere sequence is not composed of perfect repetitive arrays. Instead, telomeres consist of TG_{1-3} repeats that are considered as degenerate sequences, resulting from stuttering of the telomerase enzyme during telomere replication (24). Interestingly, the telomeres of plant species can be subdivided into three types. The model organism *Arabidopsis thaliana* and many other land plant species maintain a seven nucleotide repeat of TTTAGGG (25). In contrast, in some monocotyledonous plants TTAGGG repeats are found when the family *Iridaceae* diverged (26). The third group of plant telomere sequences is found in the *Allium* genus and consists of a CTCGGTTATGGG repeat (27). Different telomere sequences raise interesting questions about the molecular basis of these evolutionary switches, and particularly about the evolution of the telomere template with telomerase in these organisms. Nonetheless, telomeres in the majority of eukaryotic species consist of G-rich sequences.

Zooming on the architecture of telomeres, it is clear that the chromosome terminus consists of both double-stranded and single-stranded regions of telomeric repeats (6). The length of the G-rich overhang (termed the G-overhang) is relatively short compared to the double-stranded region, and varies from 20 nucleotides in *Tetrahymena* to 400 nucleotides in vertebrates (28). The current model for G-overhang maintenance is a three-step process (3). First, telomerase, which is predominantly responsible for telomere synthesis in most of organisms, elongates the 3' terminus

of the G-rich strand. Then, DNA polymerase alpha fills in the C-rich strand to replicate the 5' end of telomeres. Finally, the mature G-overhang is created by exonuclease processing.

In reality, formation and maintenance of the 3' G-overhang requires a more complicated process (29), because telomeres replicated by leading-strand DNA replication result in a blunt-ended terminus, while telomeres replicated by lagging-strand DNA replication give rise to 3' overhang after removal of the final RNA primer (Figure 3). A dynamic model has been proposed in which two distinct nucleases, Exo1 and Apollo, coordinate the processing of the two different types of telomeric ends (29). The double-stranded telomere binding protein TRF2 associates with the blunt end of leading-end telomere and then specifically recruits Apollo for end resection, while single-stranded telomere binding protein POT1b located on the lagging-end telomere inhibits Apollo to facilitate Exo1-mediated resection (Figure 3). After resection, the CST complex composed of CTC1/Cdc13/STN1/TEN1, facilitates recruitment of DNA polymerase alpha for C-strand fill in (29). Interestingly, human CTC1 and STN1 are also known as α accessory factors AAF132 and AAF44 due to their ability to stimulate polymerase activity *in vitro* (30). In this way, both ends of the chromosome are symmetrical and contain G-overhangs.

The Ku 70/ Ku 80 heterodimer, rather than TRF2, specifically localizes at the blunt-ended terminus to assemble into a stable complex that protects the chromosome end (31). This observation is intriguing since Ku is best known for its critical role in the non-homologous end joining (NHEJ) pathway for DNA double-strand break repair (32). Thus, we would expect that Ku would promote end-to-end chromosome fusion if it were located at the extreme terminus of linear DNA. A detailed mechanism for how Ku governs telomere integrity in unknown.

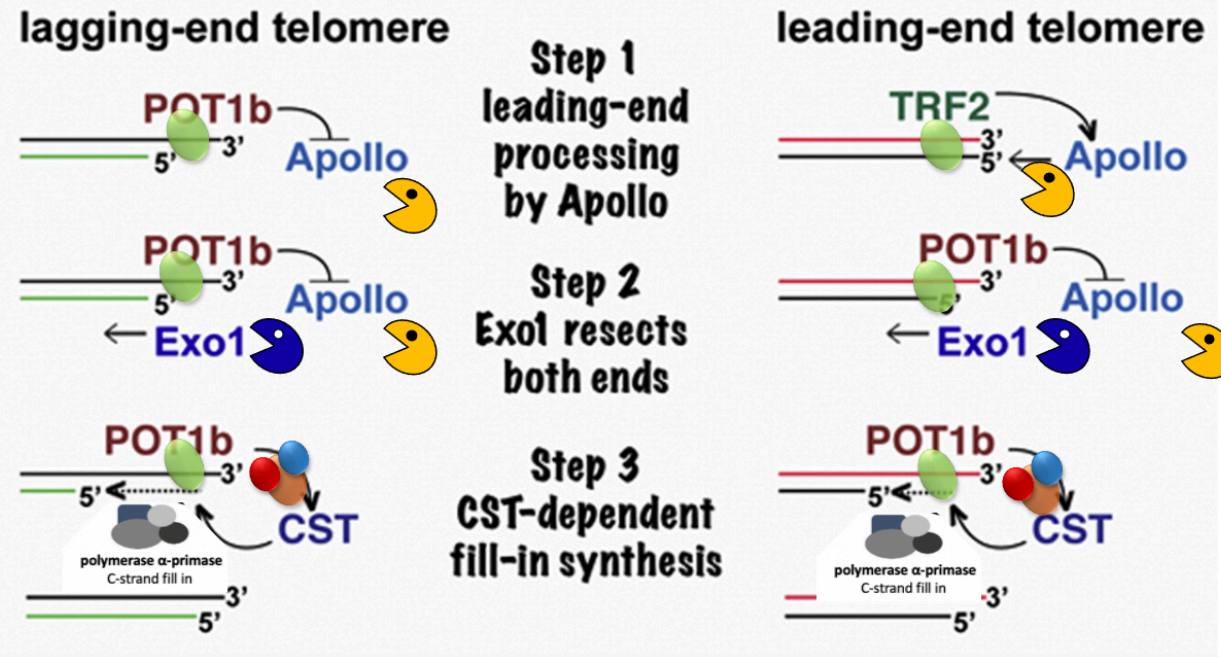


Figure 3. Steps in telomeric overhang generation in mice. Modified from Wu et al. (29).

Apollo interacts with the double-stranded telomere-associated protein TRF2 to initiate formation of the 3′ overhang at the chromosome terminus replicated by leading strand replication. In contrast, Apollo-mediated resection is blocked at the chromosome end replicated by lagging strand machinery by the single-stranded telomere-associated protein POT1b. Then, Exo1 extensively resects telomere ends in step 2 to transiently generate long 3′ overhangs. Finally, CST/AAF, a DNA polymerase accessory factor, binds POT1b and shortens the extended overhangs, likely through fill-in synthesis. Modified from Wu et al. (29).

Telomeres form a t-loop structure

Telomeres from the majority of eukaryotic species tend to assemble into high-order architectures maintained by telomere-associated proteins (33). The most well-characterized structure is a telomere-loop (t-loop). In vertebrates, the t-loop is formed by strand invasion of the 3' G-rich overhang (approximately 100-200 nucleotides in length) into the double-stranded region to form a lariat-like configuration (34). Within the t-loop, the local structure of the 3' G-rich overhang invading the duplex region is described as a displacement loop (d-loop) (see Figure 1A) (34). The structure of t-loop has been directly visualized using electron microscopy from photo-crosslinked human chromatin (33) and un-crosslinked chromatin from other vertebrates (35). In addition, the t-loop was described in Trypanosomes and fission yeast (33, 34, 36).

T-loops safeguard chromosomal ends from being recognized as DNA double-strand breaks. They also inhibit the accessibility of telomerase on telomeres by hiding the G-overhang. Therefore, dynamic regulation of t-loops during the cell cycle is essential to allow for transient disassembly during telomere replication (37). Recent studies indicate that dephosphorylation of a CDK-dependent phosphorylation site within the double-stranded telomere binding protein TRF2 is responsible for recruiting the RTEL1 helicase to transiently unwind t-loops in humans and facilitate telomere replication (37). This study not only provides solid evidence to support the existence of t-loops, but also indicates that telomeres and telomere-associated proteins assemble into a highly dynamic complex.

Telomere-associated proteins govern chromosomal ends and regulate telomere homeostasis

The telomere is a nucleoprotein complex composed of a variety of proteins that interact with double-stranded telomeric DNA, the single-stranded G-overhang, or with each other to form

a highly ordered and dynamic assembly (16). The protein-DNA and protein-protein interactions between telomeric DNA and its associated proteins are essential for chromosome protection, telomerase recruitment, and maintaining telomere homeostasis.

One of the most well-studied single-stranded end-binding complexes is the RPA-like (replication protein A-like) CST complex (Figure 4) (38). CST was originally discovered from the studies of *S. cerevisiae* telomeres with the identification of three core subunits: Cdc13, Stn1, and Ten1 (38). Stn1 and Ten1 are small, relatively conserved proteins that share structural similarities with and serve as functional homologs of RPA32 and RPA14, respectively (39). In contrast, Cdc13 is much larger and much less conserved. Structure studies revealed that OB-fold (oligonucleotide/oligosaccharide-binding fold) domains within Cdc13 have structural homology to another RPA protein, RPA70 (39). The OB-fold domains within Cdc13 anchor the entire CST complex on the G-overhang.

Functional studies of *S. cerevisiae* CST complex reveal a mutually exclusive dual-function model. In late S phase, Cdc13 contact with the G-overhang promotes the recruitment of telomerase via direct interaction with a telomerase-associated protein named Est1 (40, 41). The entire CST complex then facilitates replication termination and terminal processing of telomeric DNA (42). This dynamic model for the CST function highlights the significance of telomeric proteins in regulating telomere homeostasis.

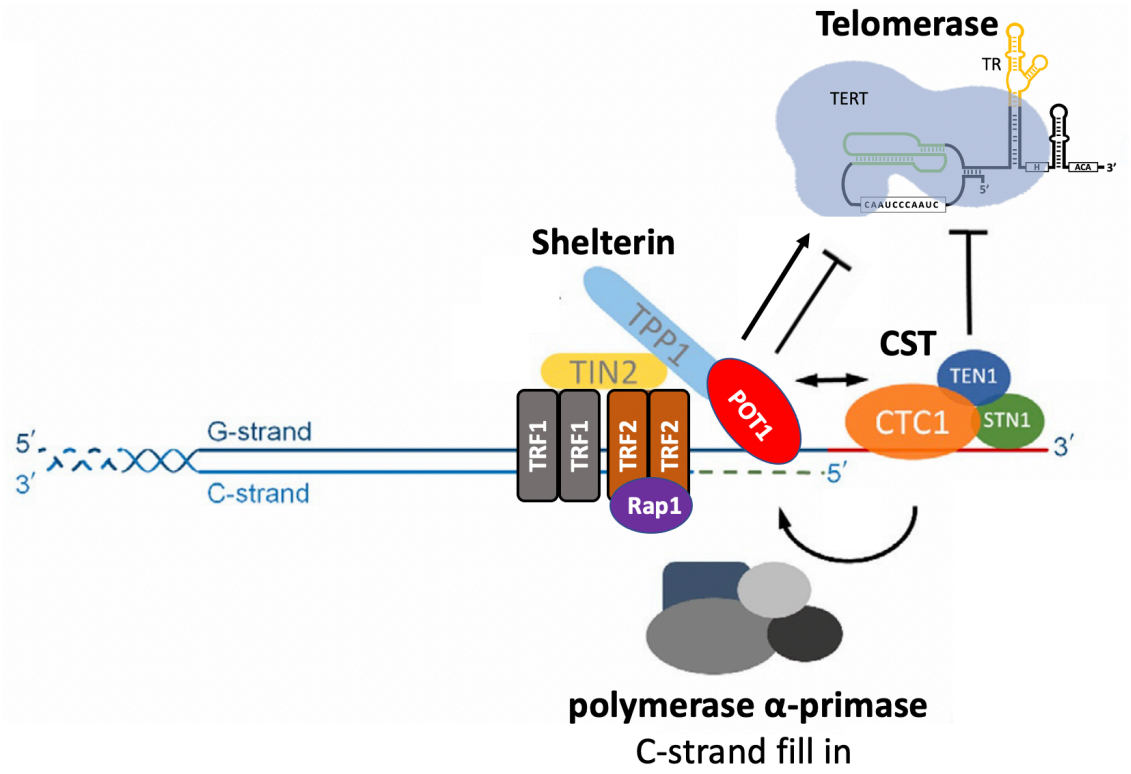


Figure 4. Mammalian telomeric chromatin.

Shelterin and CST complexes are the best characterized telomere-associated complexes. In vertebrate, CST is a trimer consisting of CTC1, STN1, and TEN1. CST is predominantly located on the telomere single-stranded end where it serves to regulate telomerase accessibility and to facilitate the C-rich strand fill-in process. Shelterin contains double-strand binding proteins TRF1 and TRF2, and the single-strand binding protein, POT1. Dynamic protein-protein interactions within shelterin differentially regulate telomere length.

Despite the evolution of telomeric DNA sequences, telomere-associated proteins maintain several common features. CST complexes have been reported in vertebrates and plants. In these species, CTC1 (conserved telomere maintenance component 1) replaces Cdc13 as the largest component of CST complexes (39). Similar to budding yeast, the vertebrate CST complex serves as a terminator for telomerase-mediated telomere replication in a size-dependent mechanism (43, 44). Telomere elongation by telomerase results in a long G-overhang, which in turn recruits sufficient CST to terminate telomere replication. Studies of the human CST complex revealed an additional function in C-rich strand fill-in synthesis of telomeres (Figure 4) (29). As described above, telomere proteins coordinate the activity of Exo1 and Apollo nucleases to control end resection for telomeres replicated by both leading and lagging strand synthesis. After Exo1 processing, the CST complex stimulates DNA polymerase α-primase and enhances its affinity for the DNA template to allow the final steps of telomeric DNA replication, C-rich strand fill-in synthesis (Figure 3 and Figure 4) (29).

The single-stranded and double-stranded regions of vertebrate telomeres are stably bound by a six-member protein complex termed shelterin. Shelterin consists of TRF1 (telomere repeat-binding factor 1), TRF2 (telomere repeat-binding factor 2), TIN2 (TRF-interacting protein 2), POT1 (Protection of Telomeres 1), TPP1 (TINT1-PIP1-PTOP1), and RAP1 (Repressor Activator Protein 1) (Figure 4) (45). Both TRF1 and TRF2 assemble into homodimers and associate with double-stranded telomere DNA via Myb-like homeodomains (46). Genetic studies showed that TRF1 and TRF2 function in telomere length regulation and telomere end protection, respectively (47, 48). To bridge other components of shelterin, TRF1 and TRF2 interact with TIN2 and RAP1 (Figure 4). RAP1 does not directly associate with telomeric DNA. It is assembled into shelterin by interactions with TIN2 and TRF2. RAP1 and TIN2 do not bind DNA, but they form a protein

bridge that unites TRF1 and TRF2, and tethers the single-stranded telomere-associated TPP1-POT1 complex to the TRF1-TRF2 complex (49, 50). (Figure 4). Functional studies indicate that TIN2 is also critical for stabilizing TRF2 on telomeres (50).

Similar to the single-stranded binding protein Cdc13 in the yeast CST complex, POT1 maintains RPA-like (replication protein A-like) OB-fold domains for G-overhang association (Figure 4) (51). Interestingly, unlike the RPA proteins, which are promiscuous in binding DNA substrates, POT1 displays high binding specificity for G-rich telomeric sequence (52). Together with TPP1, POT1 exhibits a mutually exclusive dual-function model similar to the yeast Cdc13. Generally, TPP1-POT1 binding to the telomeric terminus inhibits telomerase access to telomeres. In contrast, during telomere replication, TPP1-POT1 switches to a processivity factor for telomerase to facilitate telomerase recruitment and retention on telomeric DNA (53). A recent study proposed a G-overhang length-dependent model to explain the dynamic function of TPP1-POT1 (54). However, the detailed mechanism of how TPP1-POT1 differentially regulates telomere length as both positive and negative regulator is unclear.

From these data, it is clear that telomeres maintain a highly ordered, efficient and dynamic complex via DNA-protein and protein-protein interactions. Both double-stranded and single-stranded regions of the telomere are occupied by dedicated proteins with essential roles in telomere homeostasis. Dynamic regulation of the telomere complex provides a window into cell cycle control of chromosome end biochemistry.

Telomerase is designated for telomere maintenance

Eukaryotes employ telomeres to protect the ends of linear chromosomes from illegitimate DNA repair and exonuclease attack thereby solving the 'end-protection problem' (7, 8). However, a second problem faced by linear chromosomes, the 'end replication problem' occurs because telomeres are not fully replicated by conventional DNA polymerases. These enzymes require an RNA primer for the initiation step and fail to completely replicate the 3' end of linear DNA, which leads to the progressive loss of telomeric DNA (1, 2). To address the 'end replication problem', most of eukaryotes utilize telomerase (9). Telomerase is a ribonucleoprotein (RNP) complex consisting of two essential components: the catalytic subunit telomerase reverse transcriptase (TERT) and the RNA subunit, telomerase RNA (TR) (11, 12). TR is classified as a long non-coding RNA and it serves as the template for catalytic TERT in telomere synthesis. In addition, telomerase contains several species-specific associated proteins with critical functions in RNP biogenesis and regulation. This section will focus on the general features of telomerase regulation and the telomerase catalytic mechanism.

TERT is the limiting molecule of telomerase

In most unicellular organisms, the core telomerase components are expressed ubiquitously with little regulation. In contrast, in multicellular organisms, including humans, telomerase is active only in certain tissues or developmental stages, including embryogenesis, epithelial and lymphoid progenitors and the germline (55). Telomerase activity is undetectable or extremely low in somatic cells. The spatial and developmental profile of telomerase expression primarily depends on regulation of TERT protein. In humans, TR is much more abundantly and ubiquitously expressed than TERT with approximately 23,000 TR molecules per cell in telomerase positive

cells compared to only 5-10 TERT molecules (56). Therefore, TERT serves as the limiting molecule of telomerase. Understanding mechanisms of TERT regulation can therefore provide insights into telomerase dynamics in biological settings.

In general, TERT is controlled at three levels: transcriptional regulation, post-transcriptional regulation and protein localization (57, 58). I will use the studies in human telomerase to demonstrate these three perspectives. First, human TERT is transcriptionally repressed in somatic cells (59). The transcription factor C-myc specifically activates TERT transcription in certain cell types, including primary fibroblasts (60). Other transcription factors such as p53, Mad1 and pRB are proposed to interact with the TERT promoter to negatively regulate expression. However, a detailed regulatory pathway of TERT transcription is unclear. Second, human TERT is post-transcriptionally modified by phosphorylation and ubiquitination (59). Several kinases, including protein kinase B/Akt, phosphorylate TERT to positively regulate telomerase activity (61). In addition, a ubiquitin ligase, MKRN1 interacts with TERT *in vitro* and serves as a negative regulator of telomerase activity and telomere length *in vivo* (62). Finally, telomerase localizes at different subcellular compartments during the cell cycle (63). The current model is that in S phase TERT translocates into cajal bodies, where the TR is processed. TERT is also subjected to structure-based quality control in this compartment (63). Therefore, TERT localization in cajal bodies is essential for telomerase RNP assembly and telomerase activity. Taken together, these three modes of TERT regulation modulate telomerase activity through different stages of the cell cycle and in different types of cells.

The mechanism of telomere elongation by telomerase

Telomerase is the primary mechanism used for the replication of telomeres. After recruitment to the G-overhang, telomerase undergoes three steps to achieve a complete catalytic cycle: binding, extension and translocation (Figure 5) (3, 14). First, telomerase RNP complex anchors on the telomeric 3' end via DNA-RNA hybridization mediated by the template region of TR and a DNA-protein interaction maintained by TERT. This binding step allows telomerase to form a stable complex on the telomere and to position the 3' end into the active site of TERT (Figure 5, binding stage). Next, nucleotides are added onto telomeric DNA by the reverse transcriptase activity of TERT using the templating sequence within TR as a guide (Figure 5, extension stage). Because multiple nucleotides (for example, 6 nucleotides in human and 7 nucleotides in Arabidopsis) are added by telomerase in a single extension step, nucleotide addition processivity (NAP) is critical for telomerase to synthesize a complete telomeric repeat. In the third step, the active site of telomerase is translocated so that the active site associates once again with the 5' end of the TR template allowing the newly formed 3' end of the telomeric end to be properly positioned to initiate the next round of telomere elongation (Figure 5, translocation stage). The repeat addition processivity (RAP) of telomerase describes the ability of telomerase to elongate the same substrate for multiple translocation rounds without dissociation from the telomeric DNA. In general, the accessory proteins of telomerase RNP complex are responsible for recruiting telomerase and regulating RAP by promoting the interaction between the telomerase and telomeric end.

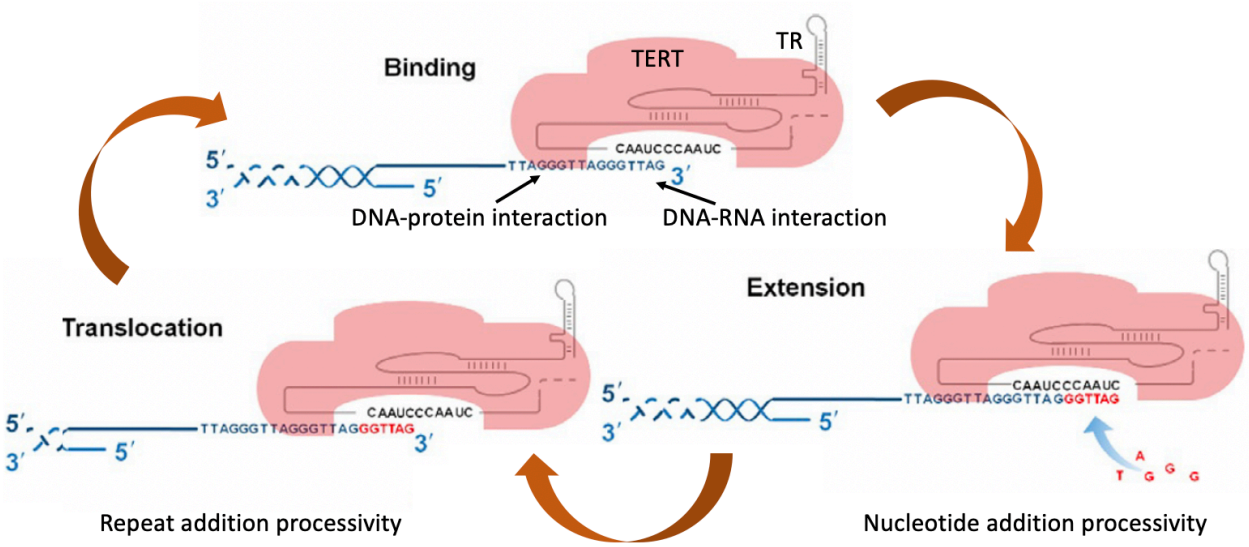


Figure 5. A complete cycle of telomerase-mediated telomere replication.

Initially, the telomeric 3' G-overhang binds to the TR template and TERT protein (Binding). Next, TERT utilizes TR to reverse transcribe one copy of the telomeric repeat by adding nucleotides onto the 3' end of the telomeric DNA (Extension). Then, the active site of TERT translocates to the 3' end of the newly formed telomeric repeat (Translocation). Another round of nucleotide addition is initiated.

TERT is a specialized reverse transcriptase

The previous section introduced the general mechanism of telomerase-mediated telomere replication. Here, I will focus on structural perspectives to demonstrate the detailed models of TERT interaction with TR and substrate DNA. As the catalytic and largest subunit of telomerase, TERT serves to coordinate other components and substrates to maintain a stable and active RNP complex.

A conserved RT domain

The amino acid sequence of TERT is similar to ordinary reverse transcriptases, especially for residues involved in nucleotide recognition and catalysis (64). In general, four functional domains are present in most of TERT proteins: the reverse transcriptase domain (RT), the telomerase 'essential' N-terminal domain (TEN), the telomerase RNA-binding domain (TRBD, and the C-terminal extension (CTE) (Figure 6A) (17, 65).

In addition, as with other reverse transcriptases, TERT maintains seven conserved reverse transcriptase-specific motifs in the RT domain. Aspartic residues responsible for catalysis of dNTP addition are located in motifs A and C (66). Following the analogy of telomerase being the "right hand" form, two sub-domains 'palm' and 'fingers' constructed of β-sheets and α-helices can be distinguished (Figure 6B). One unique feature of TERT is an insertion in the 'fingers' sub-domain (IFD) separating motifs A and B (Figure 6C) (67). From the crystal structure of the *Tribolium castaneum* TERT (*T. castaneum* TERT lacks a TEN domain), it is evident that the IFD motif consists of two anti-parallel α-helices that are responsible for the structural arrangement of the other two helices (Figure 6C). This arrangement might be employed by TERT to contact the RNA-DNA duplex (68). Study of yeast TERT revealed that mutations in the RT domain impact

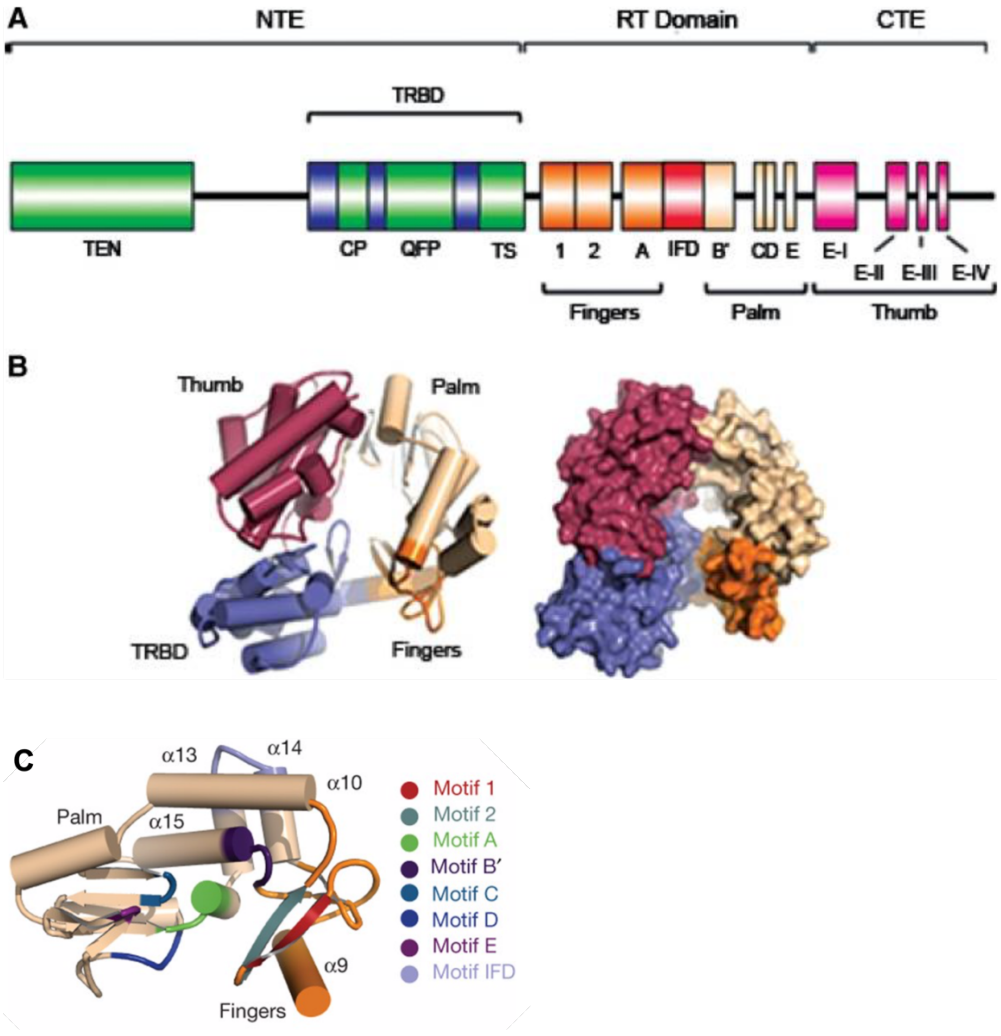


Figure 6. Structural organization of the TERT protein. Reprint from Gillis et al. (68).

(A) Architecture of human TERT. (B) Domain organization of the *T. castaneum* TERT. Taken from (65). (C) The reverse transcriptase domain of *T. castaneum* TERT. The palm subdomain contains motifs A (green), B' (dark purple), C (blue), D (navy blue), E (magenta) and IFD (light blue). Reprint from Gillis et al. (68).

telomerase function and, more interestingly, specific mutations in motif E of the RT domain enhance enzymatic processivity (69).

A novel TEN domain captures the DNA substrate

The telomerase 'essential' N-terminal (TEN) domain of TERT is required for telomerase function *in vitro* and in vivo. It interacts with both single-stranded telomeric DNA and TR (70-72). Studies using mutagenesis and chemical cross-linking demonstrated that the TEN domain contains a specific anchor motif that interacts with substrate DNA to retain telomerase complex on the telomeric end (71, 73). The capacity of TEN domain to capture DNA promotes telomerase repeat addition processivity (RAP) by maintaining association with the elongation product (70, 74). Interestingly, the X-ray structure of the recombinant TEN domain from *T. thermophila* TERT reveals no structural homology to other known proteins (75), indicating this domain is unique to TERT.

Specialized TRBD and CTE domain confining TR interaction specificity

TERT is distinct from ordinary reverse transcriptases because of its ability to use an internal RNA template (TR) for telomere synthesis. The TR binding domain (TRBD) of TERT is responsible for conferring the specificity of TR-TERT interaction (76). In general, telomerase RNA maintains two essential RNA domains: the template-pseudoknot (T-PK) and the stem terminus element (STE), also known as CR4/5 in humans (Figure 7A, bottom). Both T-PK and STE directly interact with different domains of TERT and are necessary for optimal telomerase activity. In humans, biochemical experiments and structural studies showed that the P6 stem within CR4/5 is physically associated with TRBD and this interactions is required for telomerase RNP

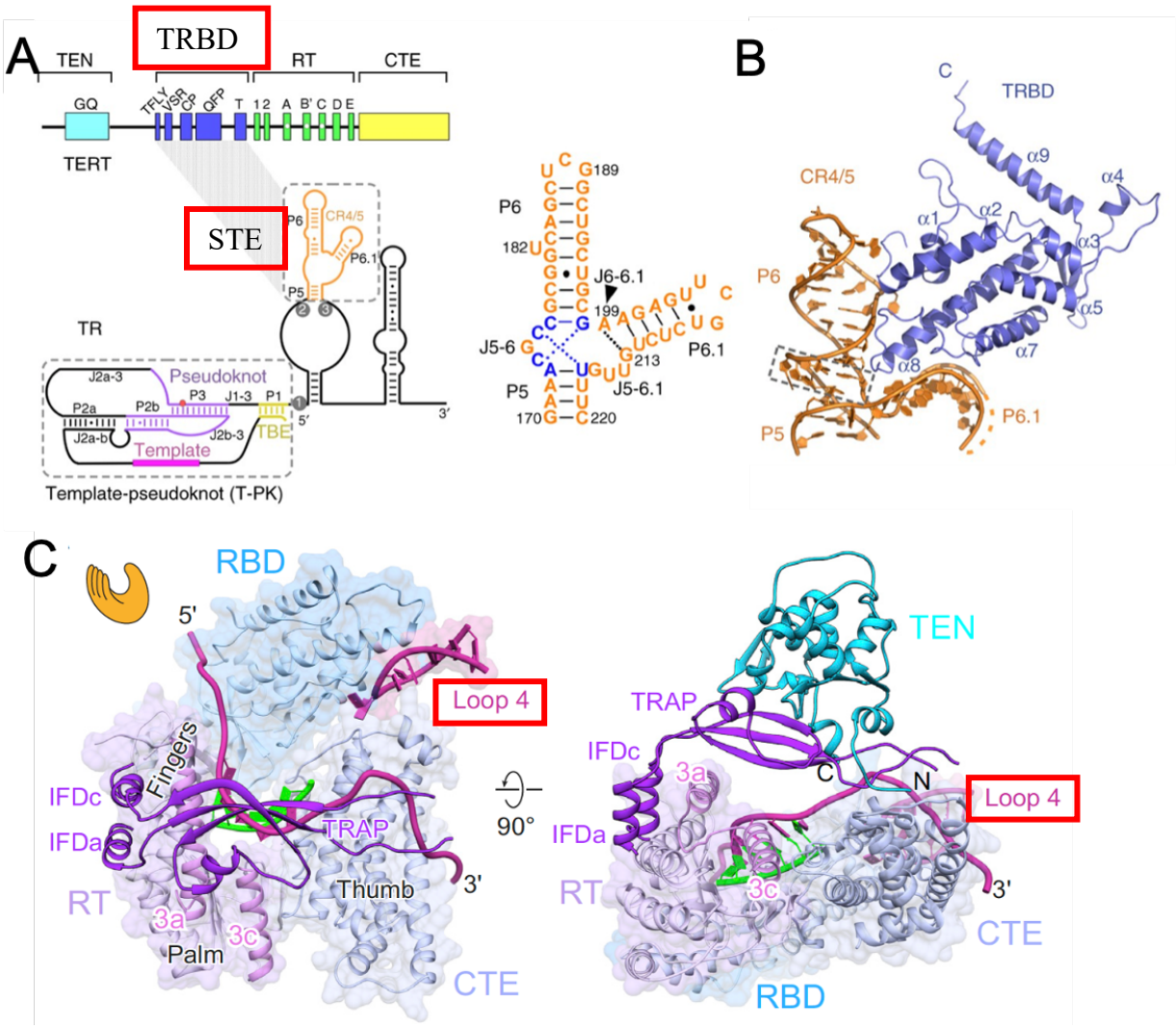


Figure 7. Domain organization of the TERT–TR interaction. Reprinted from Huang et al. (78) and Jiang et al. (79).

(A) Top: domain organization of TERT. Bottom: secondary structures of TR. Conserved domains and motifs are denoted. The TRBD-CR4/5 interaction is indicated with a connecting gray shadow. (B) View of the structure of the CR4/5–TRBD complex. TRBD and CR4/5 are colored in blue and orange, respectively. The base triples are denoted with a dashed box. Reprinted from Huang et al. (78). (C) ‘‘Hand’’ view of TERT with template-DNA duplex and TR. The RT (palm and fingers) and CTE (thumb) form the hand, with the TRBD between fingers and thumb. Helix IV (indicated as Loop 4 in the figure) between TRBD and CTE closes the TERT ring. Reprinted from Jiang et al. (79).

assembly (Figure 7A) (72, 76, 77). This conclusion is supported by the *Oryzias latipes* (medaka) CR4/5-TRBD crystal structure (Figure 7B) (78). Interestingly, the 4.8-Å cryo-EM structure of *T. thermophila* telomerase indicates a different TRBD-TR binding arrangement (Figure 7C) (79). In this structure, the loop 4 (also known as SL4), a functionally equivalent RNA motif of the human P6 stem, is not long enough to reach the TRBD binding pocket. Instead, the helix IV is bound at the interface between the TRBD and CTE domains (72, 79).

Adjacent to TRBD, TERT has a C-terminal extension (CTE) domain which also enhances nucleic acid association (65). From the crystal structure of *T. castaneum* TERT and cryo-EM structures of human and Tetrahymena telomerase RNP, the CTE closely contacts TRBD to create a ring-like protein structure for nucleic acid interaction (Figure 7C) (68, 72, 79, 80). This model is supported by experimental data based on mutagenesis (81, 82). Especially in yeast, mutagenesis of CTE residues demonstrate the importance of this domain for TERT stability and the efficiency of DNA substrate elongation (81). In humans, mutations in the CTE domain impact telomerase function, similar to some mutations in the TEN domain (83, 84).

Overall, all four domains of TERT have evolved specific roles, and make telomerase a unique reverse transcriptase. Collaboration among all four domains enable TERT to maintain a stable RNP complex and actively engage with the telomeric DNA.

TR is more than a telomeric template

In telomerase RNP, TR contains the telomeric template sequence and RNA elements necessary for binding proteins in the RNP complex. However, unlike TERT, TR is vastly divergent in sequence, length, structure and biogenesis pathway (85). For example, the length of TR can range from about 150 nucleotides in Tetrahymena to over 2000 nucleotides in yeasts (85).

Nonetheless, divergent TR molecules maintain highly conserved RNA structural domains that allow TR to function as an essential component of the RNP instead of merely a template for reverse transcription (Figure 8). The conserved TR elements serve as a scaffold for connecting TERT and accessory proteins responsible for RNP biogenesis, engagement with the chromosome terminus and regulation of telomerase enzyme activity (18). How TR coordinates with TERT and accessory proteins makes telomerase a valuable model to study the co-evolution of RNA and proteins. This section will focus on structural perspectives of TR.

An RNA template for telomeric DNA synthesis

Two conserved domains within TR are critical for telomerase catalysis (86). The first is the template-pseudoknot domain (T-PK) which bears a single-stranded template region closed by a long-range base-pairing of sequence near the 5' end of TR (Figure 8). The TR template typically corresponds to 1.5 to 2 copies of the telomeric repeat with two distinguishable segments: the alignment region (0.5-1 copy) and the templating region (1 copy) (Figure 9) (12, 86-88). During the catalytic cycle, the alignment region hybridizes to the 3' end of DNA substrate and then the templating region is utilized to determine the nucleotide sequence synthesized (Figure 9). For example, the human TR has a 11-nucleotide template including a 5-nucleotide alignment region (5'-CUAAC-3') and a 6-nucleotide templating region (5'-CUAACC-3') that corresponds to the human telomeric DNA sequence (5'-GGTTAG$_n$-3') (89). Although the telomere repeat sequences of yeast, ciliates and plants are distinct from the vertebrate repeat due to insertion or mutations, the function of the template domains is conserved (Figure 9). Data from rodents indicate that the length of the alignment region directly affects telomerase repeat addition processivity (RAP) by regulating the dissociation of the DNA substrate (90). A 2-nucleotide alignment region of rodent

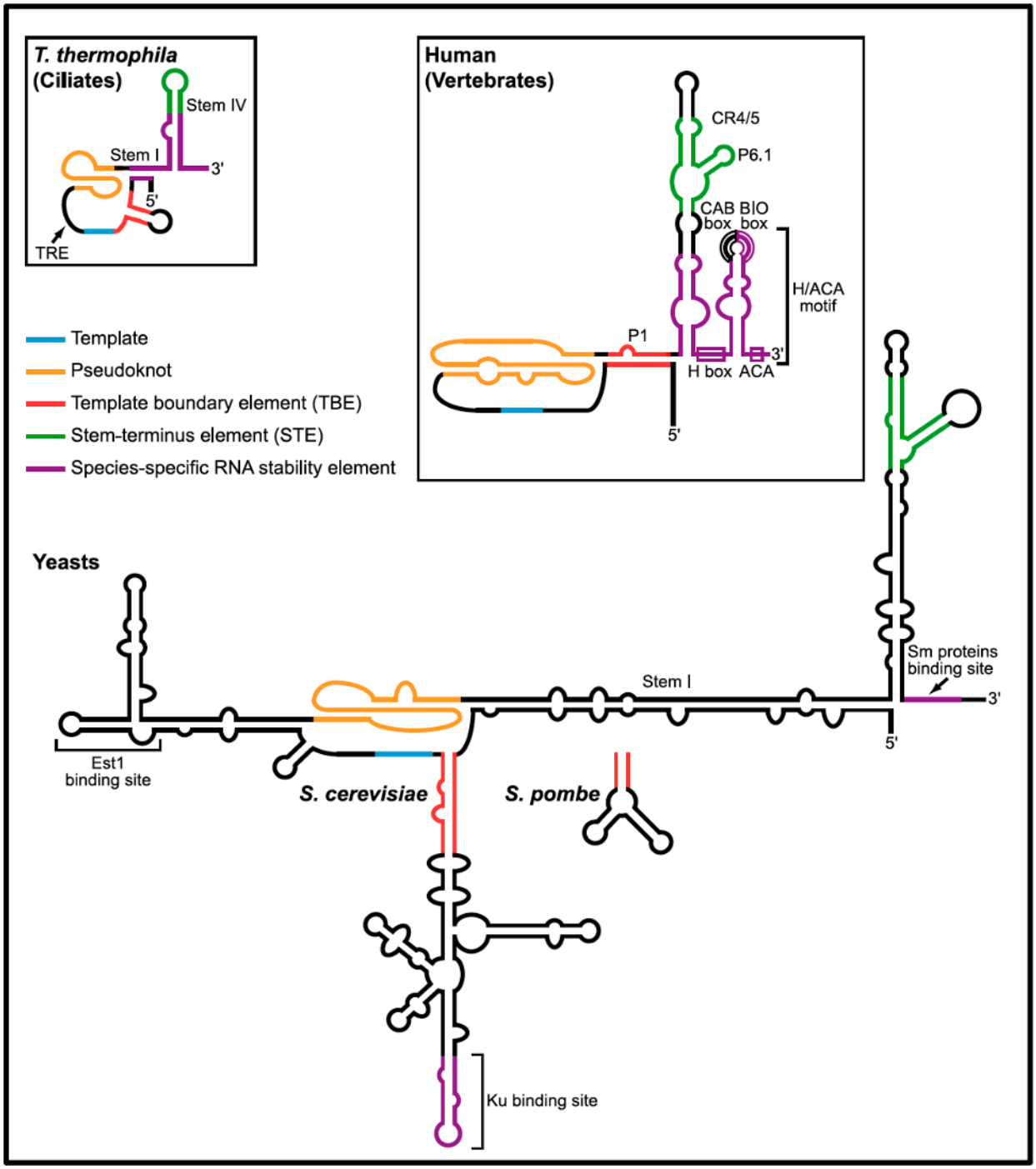


Figure 8. Diagram of TR secondary structures highlighting functional motifs. Reprinted from Egan and Collins (18).

The template, pseudoknot, TBE and STE are common to ciliate, yeast, and vertebrate TR. See text for details. Reprinted from Egan and Collins (18).

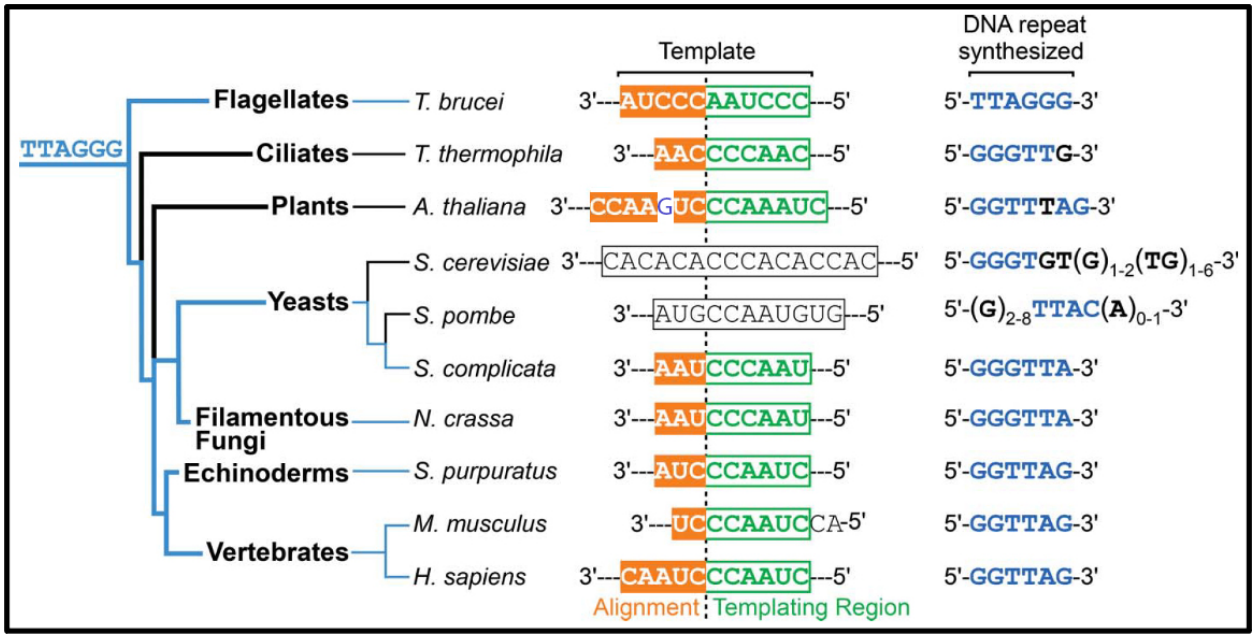


Figure 9. Diagram of telomeric DNA repeat and TR template sequences. Reprinted from Podlevsky and Chen (85).

TR leads to relatively low telomerase processivity, which can be significantly enhanced via extending the alignment length (90, 91). In addition, artificially decreasing the length of the human TR alignment region results in reduced telomerase repeat addition processivity (91). Other mutations within the template of human TR adjust the rate of repeat synthesis due to a pause signal pathway derived from the wild-type template sequence (92, 93).

A template boundary element governing template

The 5′ end of the TR template is defined by a template boundary element (TBE) that enforces polymerase fidelity by preventing incorporation of flanking non-telomeric sequence into telomeric DNA (see Figure 8) (91, 94, 95). This is critical because synthesizing non-telomeric sequences could abolish the binding of telomeric proteins on telomeres and severely compromise telomere functions. Two major structures of TBE are revealed from studies of the known TRs: a local template-adjacent helix and a long-range template-enclosing helix.

The local helix TBE is found in ciliates, flagellates, fungi and echinoderms (94-97). It is constructed of a stem-loop element to restrict the single-stranded template region. The Tetrahymena TR utilizes a short A-form helix TBE (helix II, also known as SL2 or loop2) with 5-nucleotide distance upstream of the 5' boundary of the template (see Figure 8, top left) (94, 96). A 2.7-Å crystal structure of Tetrahymena TRBD-TBE complex revealed that the CP2 motif within the TERT-TR binding domain (TRBD) physically interacts with TBE and mediates the protein-RNA interactions governing template-boundary definition (Figure 10A) (77, 98). This conclusion is consistent with the recent cryo-EM structure of the Tetrahymena telomerase (79, 80).

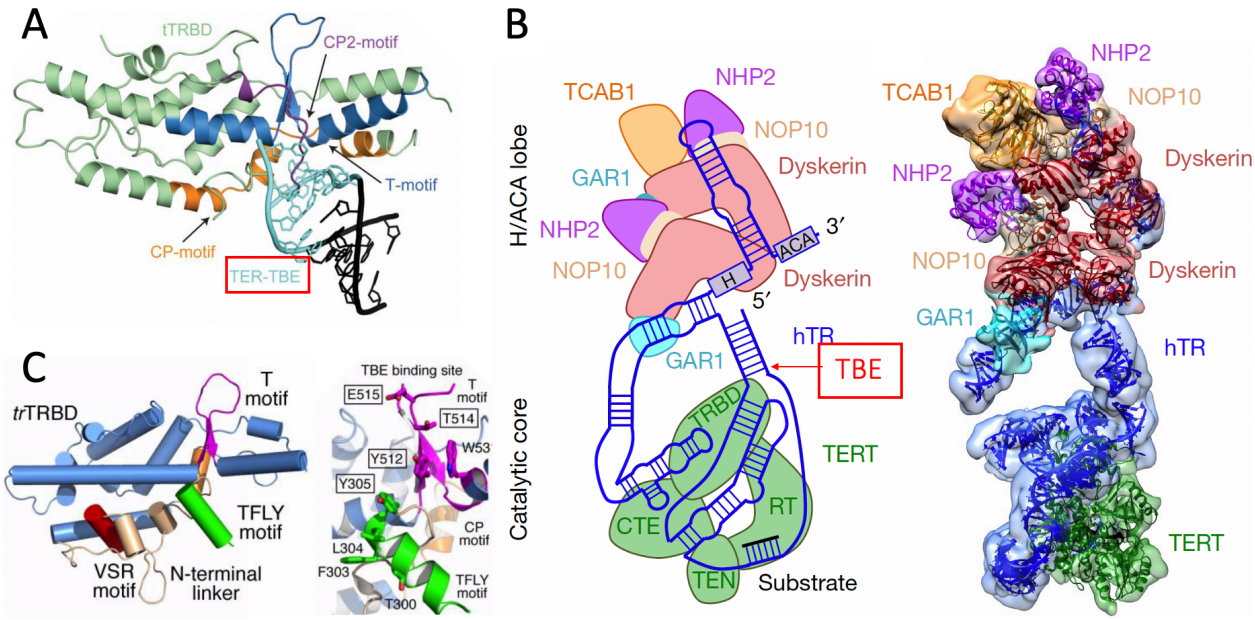


Figure 10. Structure studies of TBE associating TERT. Reprinted from Jansson et al. (98) and Harkisheimer et al. (100).

(A) Structure of the Tetrahymena TRBD-TBE complex. Tetrahymena TRBD is shown in light green with the CP2, CP, and T-motifs shown in purple, orange and blue, respectively. The TBE is shown in cyan and the remaining stem II RNA in black. Reprinted from Jansson et al. (98). (B) Cryo-EM structure of the substrate-bound human telomerase holoenzyme. TEB is indicated by the red arrow. Modified from Nguyen et al. (72). (C) Structure of *T. rubripes* TRBD and essential residues for TBE interaction. Left: diagram representation of the *T. rubripes* TRBD. The N-terminal linker (beige), the TFLY (green), T (magenta) and VSR (red) motifs are shown. Right: T (magenta) and TFLY (green) motifs involved in TBE binding. Conserved residues mutated in this study are shown in boxes. Reprinted from Harkisheimer et al. (100).

In addition to the template-adjacent helix TBE, the template-enclosing helix TBE is found in vertebrate TR (91, 99). A long-range base-paired interaction, termed P1 helix, seals off the T-PK core and serves as a double-stranded barrier for the template (see Figure 8, top right). A 10-Å cryo-EM structure of human telomerase indicates that human TBE serves as a bridge connecting the catalytic core to the H/ACA RNP with only limited interaction to TERT (Figure 10B) (72). However, a 2.4-Å crystal structure of the *Takifugu rubripes* (Japanese puffer) TRBD domain indicates that a TFLY motif within TRBD forms part of the T-CP pocket, and is implicated in TBE binding (Figure 10C) (100). Mutagenesis of conserved residues within the *T. rubripes* TRBD TFLY motif disrupts TBE binding and telomerase activity, supporting a role for this domain in TBE function (100). Low resolution cryo-EM structure of the human telomerase may have limited detection of detailed interactions. Interestingly, rodent TR lacks a structural TBE. Instead the 5' end of TR is located only 2 nucleotides upstream of the template and serves as a functional boundary (101).

In summary, a template-boundary element is essential for TR to facilitate faithful and accurate incorporation of telomeric repeats. During TR evolution, two distinct structures have emerged to fulfill TBE function with various binding mechanisms for TERT interaction.

A conserved pseudoknot element embedded in the template core domain

In addition to the template and TBE, the T-PK domain is another critical region of TR and contains a pseudoknot (PK) located downstream of the template (see Figure 8) (102). PK structures are divided into two main groups: in vertebrates and yeast, the PK structure is generally extensive and more stable (86, 87, 103), harboring large helices and long single-stranded loops. In contrast, ciliate PK is relatively primitive and less stable (see Figure 8) (104, 105). Nuclear magnetic

resonance (NMR) studies of TR fragments reveal a unique triple-helix structure in the PK that plays an essential but poorly understood role in promoting telomerase RNP assembly and telomerase activity (106-109).

To address the function of PK from a structural perspective, cryo-EM studies have been applied to provide insight into how PK interacts with TERT. The 4.8-Å cryo-EM structure of Tetrahymena telomerase provides a high-resolution examination of the T-PK domain in a substrate-loaded complex (79). Most of the TBE (also known as the helix II or loop2 in this TR) and the template strongly interact with TERT. Unexpectedly, the PK maintains only a few contacts with the basic surface of the TERT CTE domain on the back side of the TERT ring (Figure 11A). $A_{80}U_{81}$ residues are responsible for forming the triple-helix and serve as the main interaction of PK-CTE. A limited resolution (10 Å for the holoenzyme and 7.7 Å for the catalytic core) of human telomerase cryo-EM structure precludes us from further interpretation (72). However, NMR data indicate that the T-PK forms a rigid, arc-like structure (108). Consistent with the Tetrahymena telomerase structure, the PK approaches the surfaces of TRBD and CTE domains with few interactions (Figure 11B).

It can be concluded that although the PK is a conserved element of TR with essential functions in telomerase assembly and activity (105, 107, 109), it has only limited RNA-protein interactions with TERT. Therefore, PK might play more important role in the initial TERT recognition stage rather than in stabilizing the final complex.

The stem terminus element facilitates TERT binding and stimulates telomerase activity

The second essential domain of TR is a ‘stem terminus’ element (STE), which specifically interacts with TERT to stabilize the RNP. Notably, the STE is sufficient to reconstitute telomerase

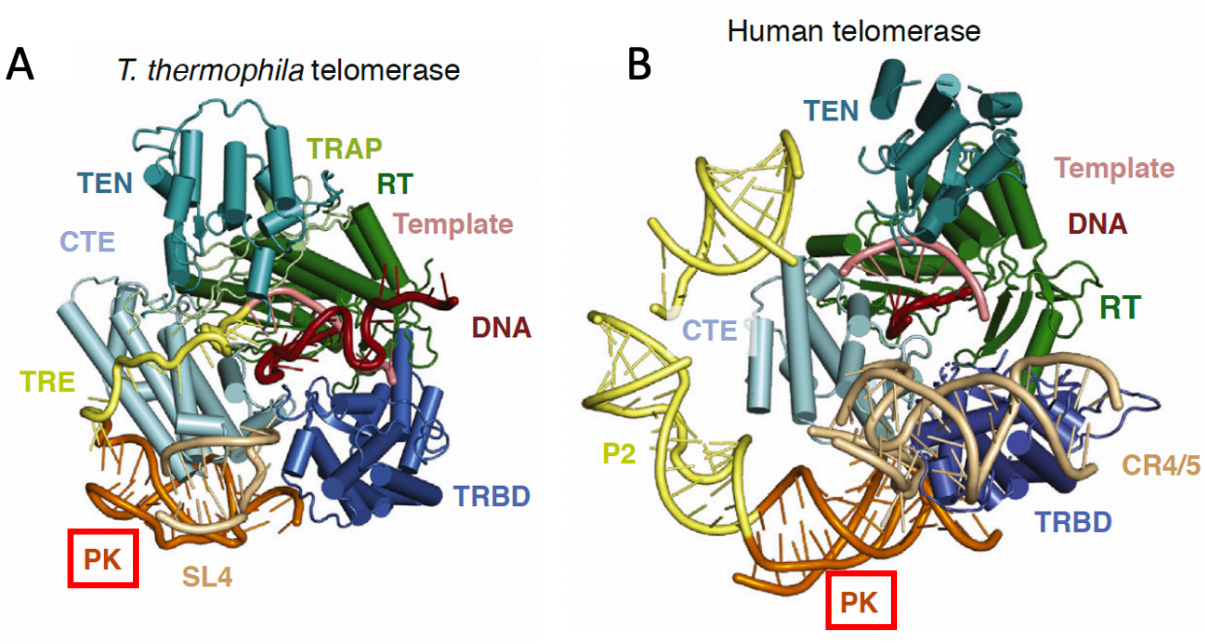


Figure 11. Cryo-EM structures of the catalytic core in Tetrahymena and human telomerase. Adapted from Nguyen et al. (110).

Views show the details of TERT ring interacting the pseudoknot (PK) structure of TR. The four TERT domains are distinguished. Orange color indicates the triple helix structure obtained for PK, which unexpectedly has limited interactions with the interface of TRBD-CTE domains of TERT protein in both organisms. Adapted from Nguyen et al. (110).

activity in trans together with the T-PK domain (86, 111-114). The STE domain is known as helix IV (also as SL4) in ciliates, conserved region 4/5 (CR4/5) in vertebrates, and three-way junction (TWJ) in budding yeasts (Figure 8). The vertebrate CR4/5 and budding yeast TWJ comprise an intersection of three RNA helixes (86, 111, 113-115). In marked contrast, the ciliate helix IV is basically a short stem-loop element (see Figure 8) (112). Interestingly, in Tetrahymena, proper positioning of the helix IV into the TERT binding pocket requires the La-related protein p65 to bend this RNA stem (116-118).

Precisely how STE stimulates telomerase activity is unknown. UV crosslink-based mapping and a 3-Å crystal structure of the Medaka CR4/5-TRBD complex demonstrate the binding mechanism of STE-TERT interaction in vertebrates (78). The CR4/5 of Medaka TR forms an L-shaped three-way-junction, maintained by helix P5, helix P6 and helix P6.1. The P6 and P6.1 form two arms clamping onto the TRBD (Figure 12A). The human telomerase cryo-EM structure revealed a similar conformation to the Medaka CR4/5-TRBD and further showed that the P6.1 helix clamps TRBD by inserting into the interface between the TRBD and CTE (Figure 12B) (72). In addition, human P6 is longer and composed of two consecutive stems: P6a and P6b. P6a is responsible for interactions with TRBD, while the distal stem P6b provides no RNA-protein or RNA-RNA interaction (Figure 12B) (72). Tetrahymena telomerase maintains a different STE-TERT interaction. Unlike its vertebrate equivalent P6, helix IV (also known as loop 4) physically associates with the interface between TRBD and CTE (Figure 12C), corresponding to the vertebrate P6.1 binding pocket. Therefore, Tetrahymena helix IV and vertebrate P6.1 serve an identical structural role in telomerase catalytic core architecture (72, 79).

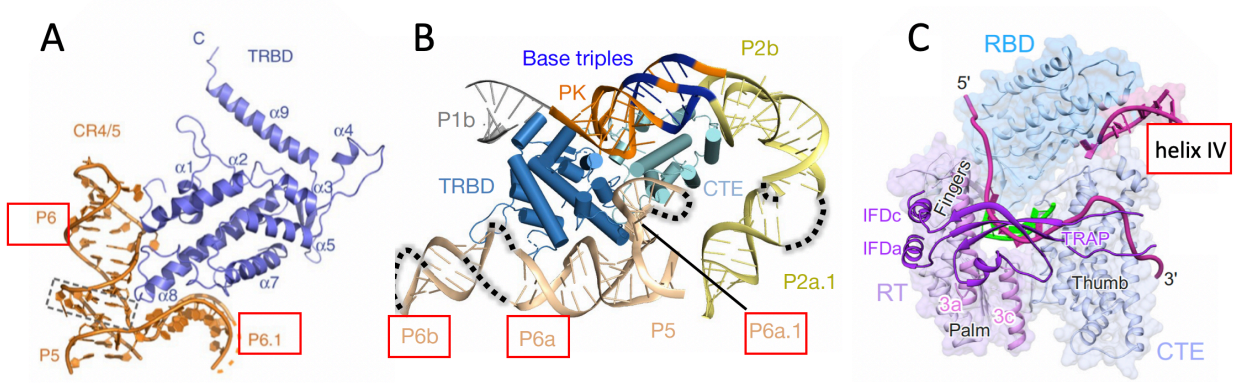


Figure 12. Structure of STE interacting TERT. Modified from Huang et al. (78), Nguyen et al. (72), and Jiang et al. (79).

(A) X-ray structure of the recombinant Medaka CR4/5-TRBD complex. Two RNA helixes, P6 and P6.1, clamp onto the TRBD domain of TERT. Modified from Huang et al. (78). (B) Cryo-EM structure of human telomerase focused on CR4/5-TRBD-CTE. The human TRBD and CTE domains are encircled by TR P2 (yellow) and PK (orange) and CR4/5 (wheat) domains. P6.1, P6a, and P6b are distinguished. Modified from Nguyen et al. (72). (C) Cryo-EM structure of Tetrahymena telomerase focused on helix IV-TRBD. Modified from Jiang et al. (79).

Overall, the requirement of TR to maintain a STE for the optimal telomerase activity is universally conserved among all major groups of eukaryotes. STE from both vertebrates and ciliates interacts with TERT to facilitate telomerase assembly and stimulate telomerase activity (86, 97, 112, 114). However, interestingly, the structural mechanism of STE-TERT interaction is dramatically distinct in vertebrates and ciliates. Obtaining more information from other species, especially yeasts and plants, will be very helpful in building a more robust general model to explore TR-TERT co-evolution.

TR biogenesis results in telomerase RNP assembly with distinct accessory proteins

In addition to the conserved T-PK and STE domains of TR which contribute a defined template and a stable TR-TERT assembly, supplemental RNA elements of TR are extremely divergent in different groups of the evolutionary lineage (85). These structural elements predominantly serve as an scaffold to recruit species-specific accessory proteins responsible for TR biogenesis, RNP assembly and telomerase enzymatic regulation (18). One of the most obvious examples of the divergence in telomerase RNP components is seen in the TR biogenesis elements located within the 3' portion of TR. The distinct maturation pathways these structures engaged largely account for the origin of the disparities in TR sequence, length, secondary structure, and most importantly, telomerase RNP composition (Figure 13). Therefore, interpreting the mechanisms of TR biogenesis is a powerful strategy to explore the evolutionary relationships of telomerase across different eukaryotic lineages, and to dissect the biological function of accessory proteins in these divergent telomerase RNPs. This section will focus on TR biogenesis pathways in ciliates, fungi, flagellates and vertebrates.

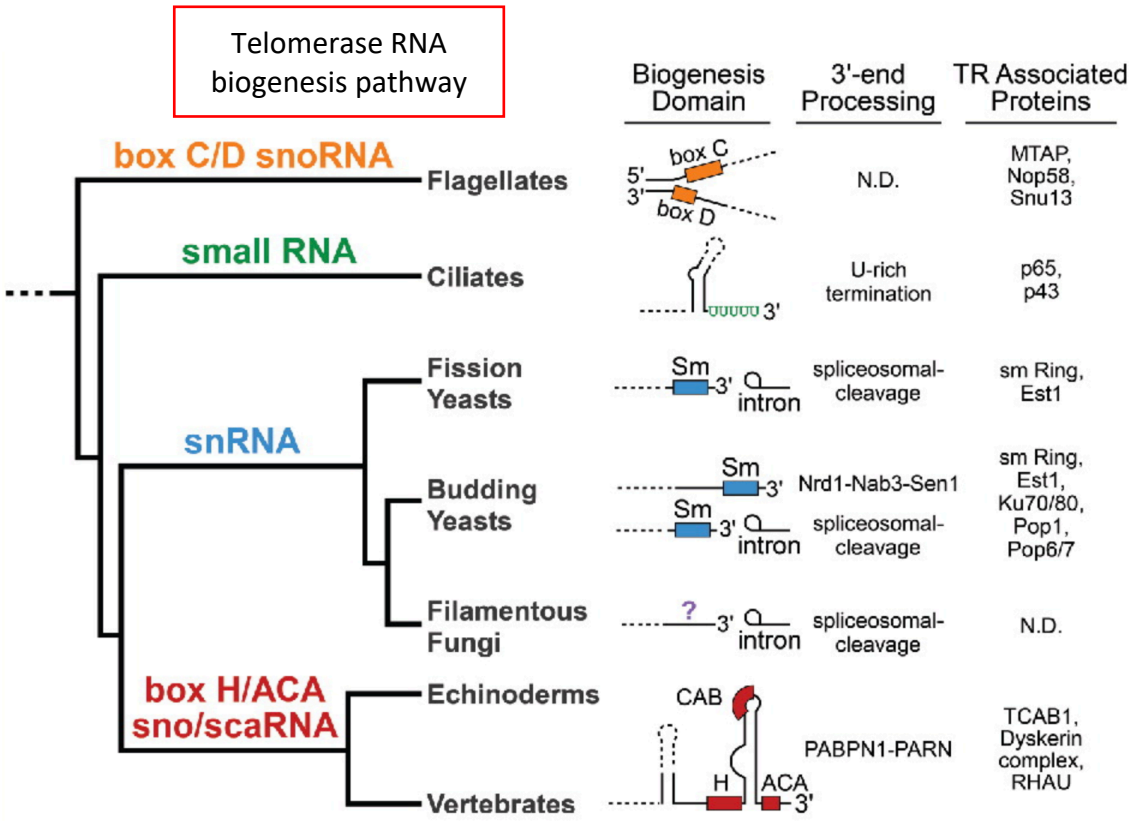


Figure 13. Four predominant pathways for TR maturation. Modified from Podlevsky and Chen (85).

Multiple RNA biogenesis pathways evolved for TR. These pathways include mechanisms illustrated the box C/D snoRNA, pol III transcribed small RNA, snRNA, and box H/ACA sno/scaRNA in flagellates, ciliates, yeasts and vertebrates, respectively. Modified from Podlevsky and Chen (85).

A small RNA pathway for ciliate TR

Ciliates employ RNA polymerase III (Pol III) for telomerase RNA transcription, leading to relatively small RNA molecule (140 to 210 nucleotides) with a 3'-poly(U) tail (12, 119). The nascent poly(U) tail of ciliate TR is sufficient to recruit La-related protein family 7 (LaRP-7) proteins: P65 in *T. thermophila* and P43 in *Euplotes aediculatus* (120-123).

LaRP-7 belongs to the large La protein family that is characterized by an extremely well-conserved La-motif (124). La protein was firstly identified in the so-called La autoantigen (or genuine La protein), an abundant RNA-binding factor detected in all eukaryotic species studied (125). The La-motif and following RNA recognition motifs (RRMs) are responsible for the genuine La protein to interact with Pol III transcripts including U6 snRNA, tRNA precursors, and pre-5S rRNA to facilitate their biogenesis (126). Recently, a group utilized phylogenetic analyses with structural motif alignments of La proteins from a large number of highly divergent eukaryotes and classified La proteins into five distinct families: genuine La proteins, LaRP-1, LaRP-4, LaRP-6, and LaRP-7 (124). LaRP-7 structurally similar to genuine La proteins with designated substrates specificities and functions (124). Evidence suggests the majority of LaRP-7 has involved in telomerase RNP complexes. In addition to P65 and P43 from ciliates, *Schizosaccharomyces pombe* telomerase retains LARP-7 protein Pof8 within the active telomerase core to facilitate RNP biogenesis (127-129). In human, Alazami syndrome patients who have LaRP7 deficiency showed a phenotype of impaired telomere maintenance, which indirectly indicates the function of LaRP7 in telomere biology (130).

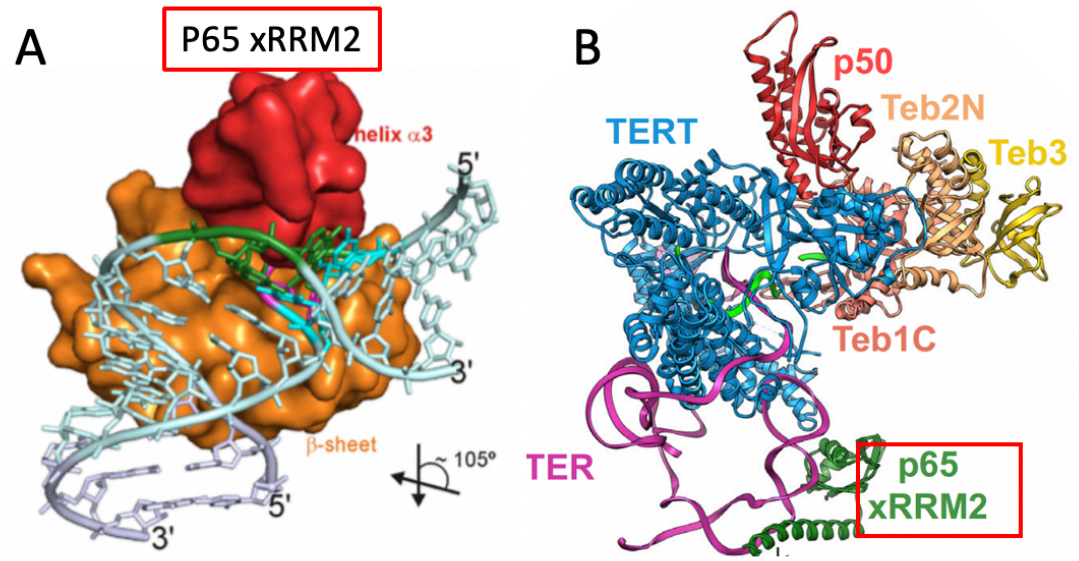


Figure 14. Structure of the P65 association with Tetrahymena TR. Modified from Singh et al. (118) and Jiang et al. (79).

(A) Crystal Structure of P65 C-terminal domain (xRRM) with TR helix IV. Stick rendering of the RNA on the surface of the protein, illustrating the 105° bend induced by protein binding. Modified from Singh et al. (118). (B) View of the molecular model of Tetrahymena TERT–TR–P65 RNP core from cryo-EM studies. The P65 xRRM2 indicated in the Figure is termed xRRM. Modified from Jiang et al. (79).

In ciliates, LARP7 proteins are constitutive components of the telomerase catalytic core essential for the TR maturation and telomerase activity (120-123). For example, association of P65 results in a significant bend (105°) in helix IV of Tetrahymena TR which enhance the interaction with TERT (116, 117). A 2.5-Å resolution crystal structure of the P65 xRRM domain (C-terminal RNA binding domain) provides insight into the molecular mechanism of Tetrahymena TR conformational changes (Figure 14A) (118). Although most of P65 is invisible in the cryo-EM structure of the Tetrahymena telomerase RNP due to its flexibility, the xRRM domain is clearly distinguished as part of the catalytic core (Figure 14B) (79, 80).

In conclusion, LaRP7 proteins promote telomerase RNP assembly through their specific interactions with TR, which leads to enhanced protein binding. Interestingly, unlike other La family proteins, LaRP7 proteins are retained in mature telomerase RNP complexes. Therefore, examining how LaRP7 proteins evolved substrate specificity and adapted to become essential and stable subunits of the telomerase complex can provide new information about the evolution of telomerase and the co-evolution of RNA and proteins within RNP complexes, more broadly.

An snRNA pathway for fungal TR

In contrast to ciliate TRs, fungi encode large TR molecules (900 to 2400 nucleotides) that are transcribed by RNA polymerase II (Pol II) (131). The maturation of fungal TR requires components of the canonical small nuclear RNA (snRNA) biogenesis pathway and results in RNP assembly with a Sm/Lsm heptameric ring located on the single-stranded uridine-rich region at the 3’ end of the RNA(132, 133). Sm/Lsm proteins protect the TR termini by promoting cap hypermethylation via a protein-protein interaction with the TMG synthase Tgs1. Interestingly, the 3’ processing mechanisms of TR vary between budding yeast and fission yeasts.

Saccharomyces cerevisiae utilizes a Nrd1/Nab3-based non-coding RNA termination pathway for TR processing which requires a terminator sequence upstream of the natural polyadenylation sites (134). In contrast in fission yeast, *Candida* budding yeast, and other filamentous fungal species, TR molecules processed by spliceosome-mediated intron splicing machinery (135). Furthermore, as mentioned above in *S. pombe* telomerase, the LaRP7 protein Pof8 plays an unexpected and critical role in TR biogenesis and RNP assembly (127, 128). Pof8 is a constitutive component of the active *S. pombe* telomerase enzyme. It is responsible for promoting the binding of the Lsm complex to TR, which in turn enhances the binding of TR to TERT. A recent study suggests that Pof8 plays a key role in telomerase RNA folding quality control to secure an efficient assembly of telomerase RNP complex (136). Notably, crystal structure and NMR analyses of the Pof8 xRRM domain demonstrated structural homology between Pof8 and P65 (129). Therefore, although fungal TR currently utilizes the snRNA pathway for biogenesis, the LaRP7 protein Pof8 might retain an ancestral function of interaction with TR to facilitate RNP assembly.

A C/D snoRNA pathway for flagellate TR

The biogenesis of flagellate TR is dramatically different from ciliate and fungal TRs as it employs the box C/D-mediated small nucleolar RNA (snoRNA) maturation pathway (137, 138). Components of the canonical C/D snoRNA biogenesis pathway including Nop58 and Snu13 are involved in processing flagellate TR. Although information is limited, it seems a *trans* splicing event is required for 5' end processing of the flagellate TR (139). This leads to TR association with a methyltransferase-associated protein (MTAP). Interestingly, MTAP is functionally related to the vertebrate TCAB1 protein (137).

A H/ACA snoRNA pathway for vertebrate TR

Vertebrates also use Pol II to transcribe a TR with sizes ranging from 312 to 559 nucleotides (87). Similar to the C/D snoRNA pathway used in the production of flagellate TR, vertebrate TR shares a biogenesis pathway with H/ACA snoRNAs (140). A highly conserved H/ACA domain was identified near the 3' terminus of vertebrate TR, which comprises a tandem array of stem-loops (P4-P4.1-P4.2 and P7-P8) interspersed by a box H motif (ANANNA) and an ACA sequence (see Figure 8) (141-143). Mutations in the H/ACA domain of human TR have been identified in patients with dyskeratosis congenita, aplastic anemia and idiopathic pulmonary fibrosis (144). Two sets of H/ACA RNP proteins consisting of dyskerin, NOP10, Gal1, and NHP2 are recruited on the H/ACA domain and assembled into the telomerase holoenzyme (Figure 15) (72, 145). Unlike all other H/ACA snoRNAs, association of dyskerin complex does not trigger 5' cleavage of the vertebrate TR to the 5' boundary of the H/ACA domain. In this way the essential T-PK domain is retained in TR. Also, dyskerin does not serve canonically as an enzyme for targeted RNA pseudouridylation. The current model is that H/ACA facilitates accurate TR 3' processing and serves to stabilize the telomerase RNP (143). The nascent TR is processed in two steps with the initial longer forms being trimmed by RRP6 (eleventh subunit of the exosome) and the resulting shorter forms then being processed by PARN (poly(A)-specific ribonuclease). Binding of the dyskerin complex resolves an inaccurately folded triple-helix structure of the TR 3' end to promote this tandem processing event (143).

The 3'-apical loop of the human TR H/ACA domain contains a Cajal body box (CAB box) that mediates the trafficking of telomerase to Cajal bodies for telomerase RNP modification and assembly (146). The telomerase Cajal body protein 1 (TCAB1) specifically recognizes the CAB box (Figure 15). Depletion of TCAB1 using RNA interference (RNAi) compromises trafficking

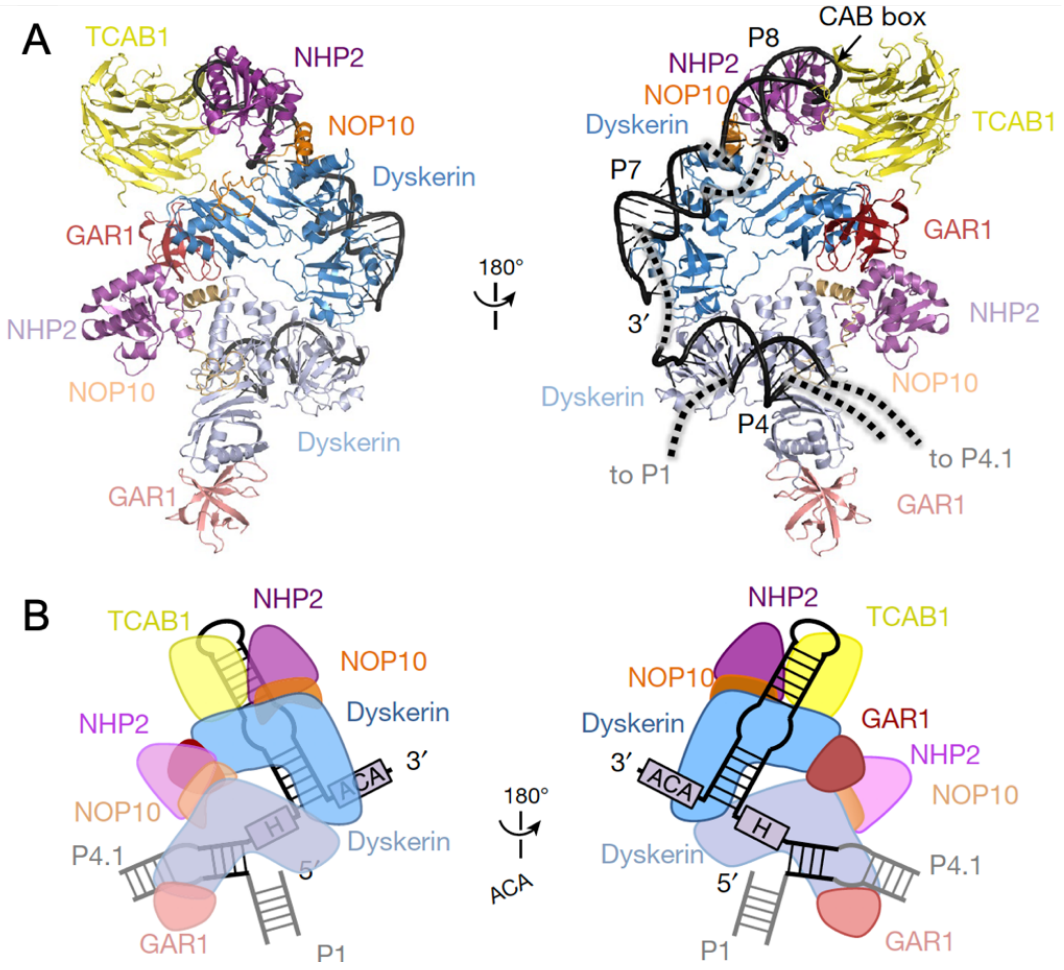


Figure 15. The cryo-EM structure of H/ACA RNP. Reprinted from Nguyen et al. (72).

(A) Front (left) and back (right) views of the H/ACA lobe within the human telomerase structure, with color-coded subunits: TCAB1, yellow; dyskerin, blue; GAR1, red; NOP10, orange; NHP2, magenta; hTR, black. To distinguish between the two sets of H/ACA proteins, the first set bound to the 5′ hairpin (P4 stem) is in a lighter shade. (B) Schematics of subunit arrangements within the H/ACA lobe. Reprinted from Nguyen et al. (72).

to Cajal bodies, mis-localizes telomerase to nucleoli and results in telomere-shortening (147). A recent study utilizing an advanced technique for RNA structure measurement (icSHAPE) revealed an alternative function of TCAB1 in TR CR4/5 folding. This alternative function ensures CR4/5-TERT interaction and subsequent RNP assembly (148). The new data for TCAB1 helps to explain the underlying mechanism for two germline mutations in CR4/5 that cause dyskeratosis congenita, G305A and C287G. Specifically, it is proposed that TCAB1depletion disrupts the association of TERT with TR and hence telomerase catalytic function.

Cryo-EM studies of human telomerase holoenzyme have significantly improved our understanding of H/ACA RNP structure (Figure 15) (72). Such studies have revealed a bilobal organization that encompasses the assembly of ten protein subunits connected by hTR as a scaffold. TERT, TR template and substrate DNA are located in one lobe of the holoenzyme termed the catalytic core, while a single TCAB1 protein and two sets of H/ACA proteins (dyskerin, NOP10, Gal1, and NHP2) occupy the other lobe as the H/ACA RNP (Figure 15). The discovery of this asymmetric bilobed structure for telomerase roundly rejected a prevailing model obtained from low-resolution negative-stain EM indicating that human telomerase was a dimer (72, 149). In addition, the human telomerase cryo-EM structure yielded the first structural insight into an intact eukaryotic H/ACA RNP. Interestingly, the two H/ACA protein heterotetramers bind to RNA elements with distinct binding mechanisms. The first set of H/ACA proteins closely contacts the 5′ stem-loop (P4) by exclusive dyskerin interaction, while the second set forms more extensive interactions with the 3′ hairpin (P7-P8) including weak dyskerin-P7, NOP10-P8 and NHP2-P8 interactions (Figure 15). A strong interaction between the two H/ACA protein heterotetramers is mediated by the two dyskerin proteins, arguing binding of one dyskerin protein at the base of the 5′ stem-loop (P4) is sufficient to anchor and assemble the entire H/ACA RNP. The structural

independence of H/ACA RNP (i.e. the fact that it does not directly contact TR or other proteins in the catalytic core) indicates that the H/ACA RNP has evolved to serve an important role in the telomerase core complex. Indeed, numerous functional studies confirm the H/ACA RNP domain of human telomerase is an essential component with multiple functions in TR biogenesis, telomerase RNP assembly and translocation (72, 150, 151).

Additional components of telomerase are critical for telomerase functions

Although proteins involved in TR biogenesis represent the primary group of accessory proteins for telomerase, other constitutive subunits have been identified with indispensable functions in RNP assembly, engagement with the chromosome terminus and regulation of telomerase activity (18). Unlike factors involved in TR biogenesis, these proteins are much more divergent making it hard to corelate them evolutionarily and functionally. Therefore, in the following section will provide a description of some of the other well-studied accessory proteins from yeasts, vertebrates and ciliates.

Fungi maintain large TR to harbor accessory proteins for distinct functions

As previously described, fungi employ RNA Pol II to generate very large TR molecules with extensive protein-binding motifs for accessory proteins. One of the best studied of these accessory proteins in budding yeast is Ever-shorter Telomere 1 (Est1p) (152), which was identified in a genetic screen for genes necessary for telomere maintenance (153). Est1p recognizes a yeast-specific TR element located downstream of the template sequence (154). Est1p is required for telomere maintenance in vivo and is sufficient to stimulate the telomerase catalytic activity (155-157). Besides TR, *S. cerevisiae* Est1p also interacts with the single-stranded telomeric DNA-

binding protein Cdc13 (from CST), and this interaction physically bridges the telomerase RNP to the telomere ends (158). The Est1p interaction with telomerase at telomeres is also proposed to activate the enzyme (159). Ever-shorter telomere 3 (Est3), which was identified in the same screen as Est1 does not physically bind to *S. cerevisiae* TR and instead directly interacts with the TERT protein (160). Limited information from yeast *Candida albicans* suggests Est3p stimulates in vitro reconstituted telomerase activity for some oligonucleotide primers (161). However, complete characterization of Est3p and its function is missing.

The development of new techniques for targeted RNP purification has facilitated the identification of novel subunits of telomerase. Using recombinant *S. cerevisiae* TR attached to a 10X MS2 tag, the processing of precursor 1 (Pop1) protein and the Pop6/7 heterodimer proteins were uncovered as stable interactive partners of TR (162). Embedded in *S. cerevisiae* TR is a structural element functionally interchangeable with the P3 domain of the RNase P and mitochondrial RNase P (MRP) RNA component. This RNA element, termed the P3-like subdomain, is responsible for the association of Pop1 and Pop6/7. Interestingly, providing recombinant Pop1 protein to in vitro telomerase reconstitution assays was sufficient to boost enzyme activity to detectable levels, something not previously accomplished using the full-length *S. cerevisiae* TR in this system (162). A recent study explored the biological function of Pop proteins associating telomerase (163). RNase P and MRP are conserved RNP complexes with essential roles in processing rRNA and tRNA. Unexpectedly, the *pop* mutants result in more abundant TR with a cellular mis-location in the cytoplasm and a subsequent deficiency of telomere maintenance. Limited information suggests that Pop proteins might mediate the stabilization of Est1p binding to TR. What is clear is that Pop proteins affect TR and canonical RNase P/MRP substrates in very different ways (163).

In addition to TR-based telomerase purification, an independent group generated a telomerase-overexpression strain of *S. cerevisiae* with a green fluorescent protein (GFP) tag linking Est1p and TERT (Est2p) proteins (164). Proteomic profiling of the purified telomerase revealed previously identified accessory proteins including Est3p and Sm proteins. More important, a novel mechanism of telomerase regulation was uncovered for the Cdc48-Npl4-Ufd1 complex (164). Cdc48-Npl4-Ufd1 complex is evolutionarily conserved and targets ubiquitinated proteins for degradation (165). Intriguingly, in Cdc48-deficient cells, Est1p is ubiquitinated and its cell cycle-regulated abundance is lost, which demonstrates a non-canonical function of Cdc48-Npl4-Ufd1 complex responsible for Est1p abundance, activity and cell cycle regulation (164). These studies indicate that telomerase is a highly complicated RNP subjected to different mechanisms of regulation.

The Ku heterodimer is well characterized for its functions in the non-homologous end-joining (NHEJ) pathway for double-stranded DNA break repair, but it also plays a critical role in telomere biology (166). In *S. cerevisiae*, the Ku heterodimer physically interacts with TR at the apical stem-loop of the template-adjacent TBE helix (167, 168), and this interaction is essential for telomerase association with telomeres during the G1 stage of the cell cycle (169). Although the function of Ku heterodimer is to recruit telomerase to telomeric DNA, Ku binding TR and DNA is mutually exclusive (170), suggesting that the engagement of Ku in the telomerase RNP with telomeric DNA is a dynamic process. However, a detailed mechanism of how Ku bridges telomerase and telomeric DNA is not clear.

To define the essential RNA elements necessary for telomerase function in yeast, a miniature yeast TR (Mini-T: 500 nucleotides in length) was generated by retaining the protein-binding elements but deleting much of the RNA separating them (171). These studies revealed that

the template-core domain (TERT-binding domain), Ku-binding TBE, Est1p-binding domain and Sm protein-recognition motif are all required and sufficient for TR to retain function in vivo. Although examination of the Mini-T RNA was ten years prior to the identification of Pop1 and Pop6/7 as critical telomerase subunits, the Pop-recognizing P3-like subdomain is included within the essential Est1p-binding domain (162). The generation of the functional Mini-T RNA strongly support the function of TR as a scaffold harboring essential accessory proteins.

Because of its large size, it is not surprising that yeast telomerase RNA maintains an extensive scaffold that retain more proteins than telomerase from other species. However, yeast telomerase is not strictly regulated compared to vertebrates, which raises an interesting evolutionary question of how telomerase-associated and regulatory proteins in higher eukaryotes to adapt and switch to more efficient mechanisms. Nevertheless, for all telomerase enzymes, the accessory proteins employ multiple functions in telomerase RNP assembly, telomerase activity regulation, and recruitment to telomeric terminus.

Structural studies uncovered accessory proteins of Tetrahymena telomerase

Utilizing cryo-EM as a powerful technique, the Feigon lab explored the structure of Tetrahymena telomerase holoenzyme with publication of a 9-Å resolution structure (80) and a 4.8-Å telomerase loaded with substrate DNA (79). The studies not only provided the structural insight into the telomerase catalytic core, but also uncovered previously unknown subunits and their interactions (Figure 16). Since the size of Tetrahymena TR (150 nucleotides) is not sufficient to retain additional proteins besides TERT and P65, cryo-EM identified accessory proteins physically anchored to TERT.

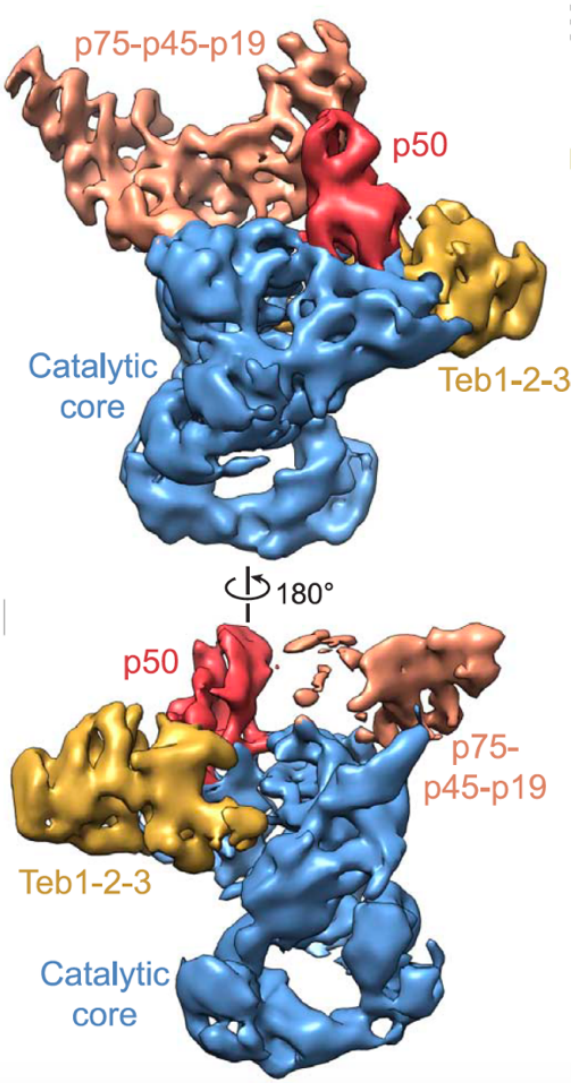


Figure 16. Cryo-EM reconstructions of Tetrahymena telomerase holoenzyme. Reprinted from Jiang et al. (80). Views of the 9.4 Å cryo-EM map, with the catalytic core, Teb1C-Teb2N-Teb3 (TEB), p75C-p45N-p19 (CST), and p50N colored in blue, gold, copper, and red, respectively. Reprinted from Jiang et al. (80).

The single-stranded telomere DNA-binding protein Teb1, paralogous to heterotrimeric replication protein A (RPA), was shown to form a TEB complex with two additional proteins (Teb2 and Teb3) (Figure 16). The cryo-EM map also revealed that the Teb1 C-terminal domain, TERT TEN domain and P50 contact each other in a triangular arrangement. On exciting result was that by fitting human TPP1 OB fold into the cryo-EM density of P50, P50 appears to be a structural paralog to TPP1. Therefore, the TEB-P50 might be the functional equivalent of human TPP1-POT1 resulting from substituting POT1 with RPA-paralogous proteins.

The second identified sub-complex in the Tetrahymena structure is a CST (CTC1-STN1-TEN1)-like complex consisting of P75-P45-P19 (Figure 16). The CST complex has been proposed to be a telomere-specific RPA (39). Human CST serves as an inhibitor of telomerase activity and determines the telomeric 3′ overhang structure (43). Recently, a 3-Å cryo-EM structure of human CST was released from the Cech lab which demonstrated a decameric assembly bound to telomeric DNA (172). As described above, the budding yeast CST subunit Cdc13 recruits the telomerase holoenzyme to telomeres by an interaction with Est1p (173). Analysis of the Tetrahymena P75-P45-P19 sub-complex indicated that P19 is a structural homolog of TEN1 (80), while P75 and P45 are homologous to CTC1 and STN1, respectively. Consistent with in vitro biochemical studies showing P75 is necessary and sufficient to bind P50 (174), cryo-EM density maps revealed close contact between P75 and P50. This interaction is required to anchor P75-P45-P19 on TERT (174, 175). Therefore, the CST-like P75-P45-P19 complex and the bridge protein P50 have critical functions in telomerase RNP complex. Overall, cryo-EM sturdies provide a new framework to identify and characterize telomerase RNP components. It also examines the interactions of components to demonstrate their biological significance.

Accessory proteins bridge the human telomerase catalytic core to telomere substrates

In human, the shelterin component TPP1 is required for telomerase recruitment to telomeres (see Figure 4). Assembling with Protection of Telomeres 1 (POT1) protein, the POT1-TPP1 heterodimer binds the 3' single-stranded DNA extension of human telomeres to promote telomerase repeat addition processivity (RAP) (176). POT1-TPP1 binding to the DNA substrate reduces primer dissociation rate and enhances the translocation efficiency (176). Mutagenesis analyses revealed that residues within the TEN domain of human TERT, especially the Gly100, are responsible for the physical connection between TPP1 and the telomerase catalytic core (177). However, POT1 is not required for telomerase recruitment to telomeres when assayed by the chromatin immunoprecipitation (ChIP) assay or the fluorescence in situ hybridization (178). In addition to the evidence of TPP1 interacting with the double-stranded telomere binding protein TRF1 and TRF2 (Figure 4), one model proposed that the recruitment role of TPP1 might not require POT1-mediated linkage to single-stranded telomeric DNA (178-180). Unfortunately, the cryo-EM structure of the affinity-purified human telomerase holoenzyme does not contain TPP1 or POT1 protein (72). In summary, POT1-TPP1 serves to stimulate telomerase repeat addition processivity. It communicates between telomerase and telomeric binding complex shelterin to facilitate the dynamic regulation of telomere homeostasis.

In addition to POT1-TPP1, the ATPases pontin and reptin were identified as telomerase-associated proteins through affinity purification of TERT from human cells (181). Pontin interacts directly with both TERT and the H/ACA RNP component dyskerin. The current functional model of pontin and reptin interacting telomerase is to regulate TERT protein in a cell-cycle dependent way, supported by the evidence that the amount of TERT bound to pontin and reptin peaks in S phase (181).

Consequences of Telomerase Dysfunction

Unlike unicellular eukaryotes that maintain a ubiquitously expressed telomerase, multicellular eukaryotes have harnessed regulatory machineries to control the abundance and activity of telomerase in different stages of the cell cycle, in different cell types and in development stages (3). In humans, telomerase is not active in most somatic cells, which restricts the proliferation potential of cells by limiting the telomere elongation. In contrast, telomerase activity is upregulated in most cancers (182). In addition, abnormal telomerase expression and telomerase mutations have been associated with many different stem-cell related diseases including dyskeratosis congenita, bone marrow failure and idiopathic pulmonary fibrosis (183). Mutations have been distinguished from all of components of telomerase that significantly contribute to the telomerase-mediated diseases (Figure 17). Collectively, telomere-related diseases have been termed telomeropathies (184).

Dyskeratosis congenita (DKC) is a severe and inherited multisystem disorder appearing at a ratio of 1 in 1,000,000 individuals (185). It is diagnosed by the triad of reticulated skin pigmentation, nail dystrophy and white patches in the mouth. Mutations in TERT and TR genes have been identified in most of DKC patients with the phenotype of suppressed telomerase activity (186). Telomerase deficiency in these patients leads to insufficient telomere maintenance that is detrimental to the renewal of regenerative tissues. In addition to the core components, a cluster of mutations concentrated in the N-terminal domain and the archaeosine-specific transglycosylase (PUA) domain of telomerase accessory protein dyskerin also contribute to DKC (187). As expected, dyskerin mutations lead to decreased TR abundance and reduced telomerase activity, corresponding to the function of dyskerin in TR biogenesis (144). Dyskerin interacts TR by

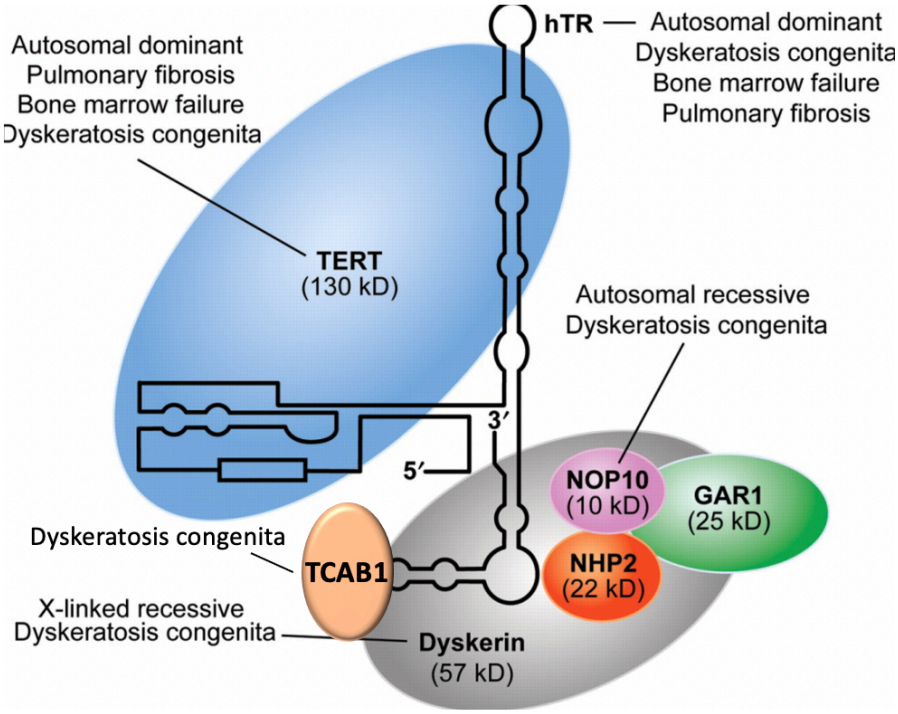


Figure 17. Mutations of telomerase components lead to severe diseases. Modified from Garcia et al. (183).
The human telomerase RNP and the diseases-caused mutations in each telomerase component are indicated. Modified from Garcia et al. (183).

assembling into a H/ACA RNP domain with three other proteins: Nop10, Gal1, and Nhp2 (see Figure 15) (72). Studies have linked mutations of both Nop10 and Nhp2 to autosomal recessive DKC patients (188, 189). Similar to other patients, Nop10 or Nhp2 mutation leads to telomerase deficiency and telomere shortening. A recent study of another telomerase accessory protein TCAB1 revealed the function of TCAB1 in TR folding (148), a finding that helps to explain the two germline mutations identified from DCK patients caused by TCAB1-dependent disruption of telomerase function.

Approximately 80% of DKC patients exhibit bone marrow failure, characterized by aplastic anemia or a reduction in blood cell lineages (190). Sequencing of TERT and TR genes from patients with various bone marrow failure syndromes have revealed multiple mutations either overlapping with DKC patients or specific for bone marrow failure (191). Most mutations identified lead to reduced telomerase activity by haploinsufficiency. Overall, telomerase dysfunction leads to severe diseases.

Arabidopsis serves as a model in telomere and telomerase biology

Arabidopsis thaliana is classified as a mustard weed belonging to the Brassicaceae family of angiosperms, and it has become the most-frequently used model organism for plant biology. First, *A. thaliana* has a short generation time and can generate a large number of progenies from one plant. This large yield of progeny significantly facilitates genetics experiments. Second, *A. thaliana* maintains a small genome (125 Mb) that has been thoroughly sequenced and well annotated (192, 193). Third, a variety of Arabidopsis mutant collections including T-DNA insertion lines and EMS-mutagenized lines are available. Researchers can easily access these resources for both forward genetics and reverse genetics studies. Finally, *A. thaliana* is compatible

with the efficient CRISPR-Cas9 system for genome editing, which creates numerous target-specific loss-of-function or missense mutations for downstream analyses. Therefore, Arabidopsis is an optimal model to explore telomere biology in the plant kingdom.

Telomeres in Arabidopsis consist of TTTAGGG repeats and are relatively short, ranging from 2 kilobases to 5 kilobases in the Columbia ecotype (194). While the overall organization of telomeres are similar to other eukaryotes, with double- and single-stranded regions, the termini display a unique non-symmetric structure: one telomeric end maintains a G-rich 3' overhang similar to other eukaryotes, while the other end of the same chromosome exhibits a blunt-ended terminus (195). Maintaining a blunt end should display more subtle telomere length fluctuations by reducing the dynamic nature of both single-stranded G-overhang and G-overhang associated proteins, which might be advantageous for the proliferation potential of cells lacking telomerase (196), as well for governing the genome stability of cells facing environmental stresses.

Although plant telomeres were first studied by Barbara McClintock 80 years ago (23), less is known about the telomerase enzyme in plants than in other major eukaryotic lineages. The section below will focus on the current state of knowledge of *A. thaliana* telomerase RNP complex in terms of TERT, TR and telomerase-associated proteins.

Arabidopsis has a conserved TERT protein

The structure of Arabidopsis TERT is highly similar to vertebrate TERT (197). Moreover, as in humans, Arabidopsis TERT is expressed only in reproductive organs and other highly proliferative cells (197). A null mutation in the *A. thaliana TERT* gene generated with a T-DNA insertion completely abolishes active telomerase causing telomeres to progressively shorten and yet mutant plants can survive for up to ten generations (19). However, with each passing generation,

plants accumulate more genome instability as seen by the increasing presence of end-to-end chromosome fusions during anaphase. The capacity of Arabidopsis to withstand the loss of telomerase for so many generations despite worsening genome instability might be an evolutionary advantage, and a reflection of the plasticity of plant development and genome structure.

Identification of telomerase-associated RNAs in A. thaliana

Although TERT is well characterized in Arabidopsis, the identity of the true RNA subunit of telomerase has been problematic. Previously two long non-coding RNA, AtTER1 and AtTER2, were identified as telomerase-associated RNA in *A. thaliana* (198). Specifically, AtTER1 was proposed as the functional TR that provided a template for telomere synthesis. However, new pieces of evidence indicate that AtTER1 is not a *bona fide* TR. First, the identification of AtTER1 was problematic at the beginning because it based on a biased assumption that the template sequence would allow perfect Watson-Crick base pairing with the telomere. In the original experiment, active telomerase was purified through tandem chromatography to enrich telomerase components including TR. Then, the RNA from purified fractions was amplified with primers complementary to only the perfectly aligned template sequence. However, a small group of vertebrates including hamster and mouse utilize a wobble G-T base pair in the alignment region of template (87). Therefore, this experimental design would miss a potential true TR that harbored a wobble G-T base pair and favor an artificial candidate. Second, the experiment used to identify AtTER1 lacked a strong control. Enrichment of active telomerase was conducted only in the wild-type background without a telomerase mutant control. Therefore, AtTER1 could simply be a contamination from non-specifically co-purified RNA molecules. Third, the data from genetics and biochemistry experiments to confirm AtTER1 function were subsequently unreproducible by

other members of the Shippen lab. Finally, using the CRISPR-Cas9 based gene editing technique, two independent groups mutated *AtTER1* gene and removed the entire template sequence from AtTER1 (199, 200). These mutations did not lead to inactive telomerase or telomere shortening, establishing that AtTER1 is not required for telomerase function and telomere maintenance *in vivo* (199, 200). Therefore, an urgent task and a major focus of my dissertation research was to design an unbiased approach to uncover the *bona fide* TR in Arabidopsis (Chapter II).

The second telomerase-associated RNA AtTER2 was identified by sequence similarity of AtTER1 (198). AtTER2 was described to negatively regulate telomerase activity in response to DNA damage (201). The very low abundance of AtTER2 and discovery that AtTER1 was not the true TR prompted a re-examination of this locus. I participate in the team with Shippen lab colleagues to carefully characterize AtTER2 and its overlapping gene in Chapter IV.

Plant telomerase consists more than TERT and TR

Although the information about plant telomerase is relatively limited, several proteins were identified as critical components of the *A. thaliana* telomerase RNP. Immunoprecipitation experiments indicate that AtPOT1a physically associates with the active telomerase RNP (202). In addition, genetic experiments demonstrated that unlike vertebrate POT1, AtPOT1a is required for telomere maintenance. Telomeres are progressively shortened at the same rate as in *tert* mutants (203). Finally, biochemical studies indicated that recombinant OB1 domain of AtPOT1a is not only sufficient to specifically bind single-stranded telomeric DNA in vitro, as has been shown for other POT1 orthologs (204), but also stimulates telomerase repeat addition processivity (RAP). Low levels of AtPOT1 were detected at telomeres in unsynchronized cells and in cells arrested in G2, while AtPOT1 binding telomeres was significantly enhanced during S-phase, when telomerase

is thought to act at telomeres. These evidences suggest that by bridging telomerase and telomeric DNA, AtPOT1a serves as a recruitment factor and RAP factor for telomerase in a particular stage of cell cycle. An important part of my dissertation research has been to investigate how POT1a engages the telomerase enzyme (Chapter III)

In addition to POT1a, analysis of dyskerin (AtNAP57) has provided evidence that it may be a constitutive component of the telomerase RNP in both eudicots *A. thaliana* and monocots *Allium cepa* (205, 206). A missense mutation of dyskerin dominantly reduced telomerase activity and impacted telomere maintenance (205). Unlike humans, plant TR does not harbor a canonical H/ACA motif. Hence, if dyskerin binds TR the mechanism of dyskerin-telomerase interaction is unclear (99). In Chapter III, I describe my analysis of dyskerin in the Arabidopsis telomerase RNP.

Finally, other proteins have been implicated as accessory factors for plant telomerase. These include the shelterin-like proteins TRB1-3, RuvBL1 and RuvBL 2a (plant homologs of pontin and reptin), which were uncovered using truncated and inactive TERT as a bait (207, 208). However, rigorous confirmation that these proteins are function components of the telomerase RNP in plants is lacking. An unbiased strategy of functional telomerase RNP purification, and a thorough characterization of plant telomerase holoenzyme is required for understanding accessory protein compositions and functions. In Appendix, I describe initial experiments to identify the full set of core components of Arabidopsis telomerase.

Overview

This dissertation describes the identification and characterization of essential components of the telomerase RNP from the plant kingdom. In Chapter II, I present the *bona fide* TR from the *A. thaliana*, named AtTR, identified through unbiased sequencing of RNAs that copurify with

Arabidopsis TERT (AtTERT) protein. The identification of this molecule enabled the discovery of 85 AtTR orthologs from three major clades of plants. Bioinformatic analyses and experimental verification allowed construction of a conserved and robust secondary structure model for plant TR revealing essential T-PK and STE domains as well as novel plant-specific elements. Zooming in the chimeric plant PK element indicates that the plant TR is an evolutionary bridge connecting the disparate structures of previously characterized TRs from ciliates and vertebrates.

In Chapter III, I describe a detailed examination of the associated proteins of *A. thaliana* telomerase RNP. AtPOT1a was previously proposed to be an accessory protein of active telomerase that stimulates its repeat addition processivity (202). My work revealed a physical interaction between AtPOT1a and AtTERT, which might facilitate telomerase interaction with the substrate DNA. In addition, I uncovered a direct interaction between the snoRNP biogenesis factor dyskerin with AtTR, and demonstrated that dyskerin promotes telomerase repeat addition processivity in vitro. Binding studies revealed that a plant-specific three-way junction embedded in AtTR is responsible for dyskerin interaction with this RNA. An 8.6-Å cryo-EM structure of a recombinant dyskerin-AtTR complex revealed an identical protein organization of Arabidopsis dyskerin to the human telomerase H/ACA RNP. Overall, my analysis of the association of dyskerin and POT1a in the Arabidopsis telomerase RNP provides important new insight into the evolutionary relationship between plant and vertebrate telomerase, in which plant telomerase RNP retains associated proteins identical to vertebrates by a plant-specific assembly.

In Chapter IV, I present a collaborative study in which members of the Shippen lab re-evaluate the telomerase-associated lncRNA, TER2 and demonstrate this RNA is neither a component of telomerases or a stable transcript. Re-annotation of the TER2 locus revealed that it is completely embedded within a gene encoding tRNA Adenosine Deaminase 3 (*TAD3*). Further

analyses revealed that the function of TAD3, but not TER2, is important for telomere maintenance. Our studies showed that the function of TAD3 in telomere maintenance is independent of telomerase and telomere architecture, and instead reflects a broader, likely indirect, role for TAD3 in modulating cellular metabolism.

In the Appendix, I describe the first set of results from an unbiased approach to define telomerase-associated proteins in Arabidopsis. Taking advantages of the sensitive quantitative mass spectrometry (qMS) technique, we identified genuine La family protein AtLa1 as a putative constitutive subunit of Arabidopsis telomerase holoenzyme. Interestingly, we also found that AtCPN60B, the beta subunit of the chloroplast chaperonin 60, associates with Arabidopsis telomerase. This result opens a framework for studying telomerase biogenesis and translocation in plants.

In total, this dissertation supports the conclusion that studies of plant telomerase have the potential to provide critical new insights into the evolution and function of divergent telomerase RNP complexes.

CHAPTER II

THE CONSERVED STRUCTURE OF PLANT TELOMERASE RNA PROVIDES THE MISSING LINK FOR AN EVOLUTIONARY PATHWAY FROM CILIATES TO HUMANS*

Summary

Telomerase is essential for maintaining telomere integrity. Although telomerase function is widely conserved, the integral telomerase RNA (TR) that provides a template for telomeric DNA synthesis has diverged dramatically. Nevertheless, TR molecules retain two highly conserved structural domains critical for catalysis: a template-proximal pseudoknot (PK) structure and a downstream stem-loop structure. Here we introduce the authentic TR from the plant *Arabidopsis thaliana* called AtTR identified through next-generation sequencing of RNAs co-purifying with Arabidopsis TERT. This RNA is distinct from the RNA previously described as the templating telomerase RNA, AtTER1. AtTR is a 268 nt Pol III transcript, necessary for telomere maintenance *in vivo* and sufficient with TERT to reconstitute telomerase activity *in vitro*. Bioinformatics analysis identified 85 AtTR orthologs from three major clades of plants: angiosperms, gymnosperms and lycophytes. Through phylogenetic comparison, a secondary structure model conserved among plant TRs was inferred and verified using *in vitro* and *in vivo* chemical probing. The conserved plant TR structure contains a template-PK core domain enclosed by a P1 stem and a 3' long stem P4/5/6, both of which resemble a corresponding structural element in ciliate and

---

* Reprinted with permission from 'The conserved structure of plant telomerase RNA provides the missing link for an evolutionary pathway from ciliates to humans' by Song, J., Logeswaran, D., Castillo-González, C., Li, Y., Bose, S., Aklilu, B., Ma, Z., Polkhovskiy, A., Chen, J., and Shippen, D.E., 2019. Proceedings of the National Academy of Sciences, 116 (49) 24542-24550. Copyright 2019 National Academy of Sciences.

vertebrate TRs. However, the plant TR contains additional stems and linkers within the template-PK core, allowing for expansion of PK structure from the simple PK in the smaller ciliate TR during evolution. Hence, the plant TR provides an evolutionary bridge that unites the disparate structures of previously characterized TRs from ciliates and vertebrates.

Introduction

Many non-coding RNAs (ncRNAs) function as integral components of ribonucleoprotein (RNP) complex enzymes that govern cellular processes such as translation, RNA splicing and telomere maintenance (209). The telomerase RNA (TR or TER) assembles with the telomerase reverse transcriptase (TERT) protein to form the catalytic core of an enzyme that maintains telomere function and genome integrity by continually adding telomeric DNA repeats onto chromosome ends (210). TR contains a template for the synthesis of G-rich telomere repeat arrays catalyzed by TERT. In addition, TR harbors highly conserved structural domains that serve as a scaffold for binding accessory proteins that facilitate RNP biogenesis, engagement with the chromosome terminus and regulation of telomerase enzyme activity (85).

The essential role of telomerase in telomere maintenance is universally conserved across Eukarya, except for a small group of insect species that evolved a retrotransposon-mediated mechanism (211). Nevertheless, key aspects of the telomerase RNP have diverged dramatically, including the sequence and length of TR, the protein composition of the holoenzyme and the mechanism of RNP maturation (18). For example, TR genes in ciliated protozoa encode relatively small RNAs (140-210 nt. in length) that are transcribed by RNA polymerase III (Pol III) (12, 119). The La-related protein P65 in Tetrahymena recognizes the 3’ poly-U tail of TR and bends the RNA to facilitate telomerase RNP assembly (80, 118). In contrast, fungi maintain much larger TR

molecules (900 to 2,400 nt.) that are transcribed by RNA polymerase II (Pol II) (85). The 3' end maturation of fungal TRs requires components of the canonical snRNA biogenesis pathway and results in RNP assembly with Sm and Lsm proteins (134, 135). Like fungi, vertebrates also utilize Pol II to transcribe a TR with a size ranging from 312 to 559 nt (87). However, vertebrate telomerase RNP processing and biogenesis proceeds via a small nucleolar RNA (snoRNA) maturation pathway (140). In vertebrates, a highly conserved structural motif in the 3' H/ACA domain of TR binds the protein components of the H/ACA snoRNP (Dyskerin, NOP10, NHP2, and GAR1) which then protect the 3'end of the mature TR from exonuclease degradation (141-143).

Within TR, two conserved domains are critical for telomerase catalysis (86). The first is the template-pseudoknot domain which bears a single-stranded template region typically corresponding to 1.5-2 copies of the telomeric repeat (85). The 5' boundary of the TR template is defined by a template boundary element (TBE) that promotes polymerase fidelity by preventing incorporation of non-telomeric nucleotides into telomeric DNA (91, 94, 95, 98). In addition to the template and TBE, the pseudoknot (PK) structure located downstream of the template is essential for TERT-TR interaction and enzyme activity (9, 102). The PK structures from vertebrates and yeast TRs are generally larger and more stable (86, 87), harboring longer helices than the PK structures of ciliate TR, which are relatively primitive and less stable (104, 105). NMR studies of TR reveal a unique triple-helix structure in the PK which plays an essential, but poorly understood, role in promoting telomerase activity (108). Another essential domain of TR, called helix IV in ciliates or CR4/5 in vertebrates, can reconstitute telomerase activity in trans together with the template-PK domain (111-114). TRs from other groups of eukaryotes including echinoderms and trypanosomes also possess a second structural domain called eCR4/5 that can bind independently

to TERT in trans and is functionally equivalent to the vertebrate CR4/5. The requirement of two conserved structural TR domains for telomerase activity is therefore universally conserved among all major groups of eukaryotes from Trypanosome to vertebrates (97).

We previously described the identification of two telomerase-associated RNAs from A. thaliana termed AtTER1 and AtTER2 (198, 201). AtTER1 was proposed to serve as the template for telomeric DNA synthesis by telomerase (198). However, recent data has refuted the role of AtTER1 in telomere maintenance (199, 212). Moreover, Fajkus and colleagues recently reported the identification of a novel telomerase RNA from A. thaliana termed AtTR that is required for telomere maintenance and is conserved across land plants (199). Here we present results of a next-generation sequencing analysis of TERT-associated RNAs, which independently led to the identification of AtTR as the bona fide RNA component for Arabidopsis telomerase. We show that AtTR is crucial for telomere maintenance in vivo and sufficient to reconstitute telomerase activity with A. thaliana TERT (AtTERT) protein in vitro. In addition, by employing phylogenetic sequence analysis of homologous TRs from the three distantly related plant lineages including angiosperms, gymnosperms and the early branching lycophytes, we determine a conserved structural model for plant TRs that was verified using chemical probing and mutagenesis. Our findings provide an evolutionary bridge to unite the disparate structures of the previously characterized TRs from ciliates and vertebrates as well as a new platform to explore the evolution of the telomerase RNP enzyme.

Results

AtTR is the predominant RNA associated with active telomerase in Arabidopsis

Prompted by collaborative work with the Beilstein lab, which indicated that AtTER1 was not the authentic TR component for *A. thaliana* telomerase (212), we developed an unbiased approach to identify ncRNAs associated with the AtTERT protein through RNA immunoprecipitation (RIP) analysis using anti-AtTERT antibody. RIP was performed under native conditions with mild salt and detergent concentrations to retain weak interactions. Next-generation sequencing of co-purified RNAs identified 177 RNA sequences that were significantly enriched in the wild-type (WT) but not *tert* null samples (Figure. 18*A*). The previously reported telomerase RNA template AtTER1 and the TERT-associated RNA AtTER2 were not found among these AtTERT-associated RNAs. To address the possibility that AtTER1 was masked by other more abundant RNAs, we used more stringent conditions to purify active telomerase by size exclusion chromatography prior to RIP (Figure. 18*B*). Telomerase activity was detected by quantitative telomere repeat amplification protocol (qTRAP) with the peak activity in a fraction corresponding to an apparent molecular mass of ~300 kDa (Figure. 18*C*). A scatter plot of RNAs purified and sequenced from fractions with peak telomerase activity revealed a single RNA that was enriched more than 100-fold above background (Figure. 18*D*). This is the same RNA independently reported by Fajkus et al and dubbed AtTR (199). Since AtTER1 overlaps with the 5' region of RAD52 locus (213) (supplementary Figure. S1*A*), we performed additional TERT RIP experiments to directly test if RAD52 mRNA was present in the IP. While RAD52 mRNA could be amplified from the IP, an RNA corresponding to the previously described AtTER1 could not (supplementary Figure. S1*B* and S1*C*). These results are inconsistent with AtTER1 being a

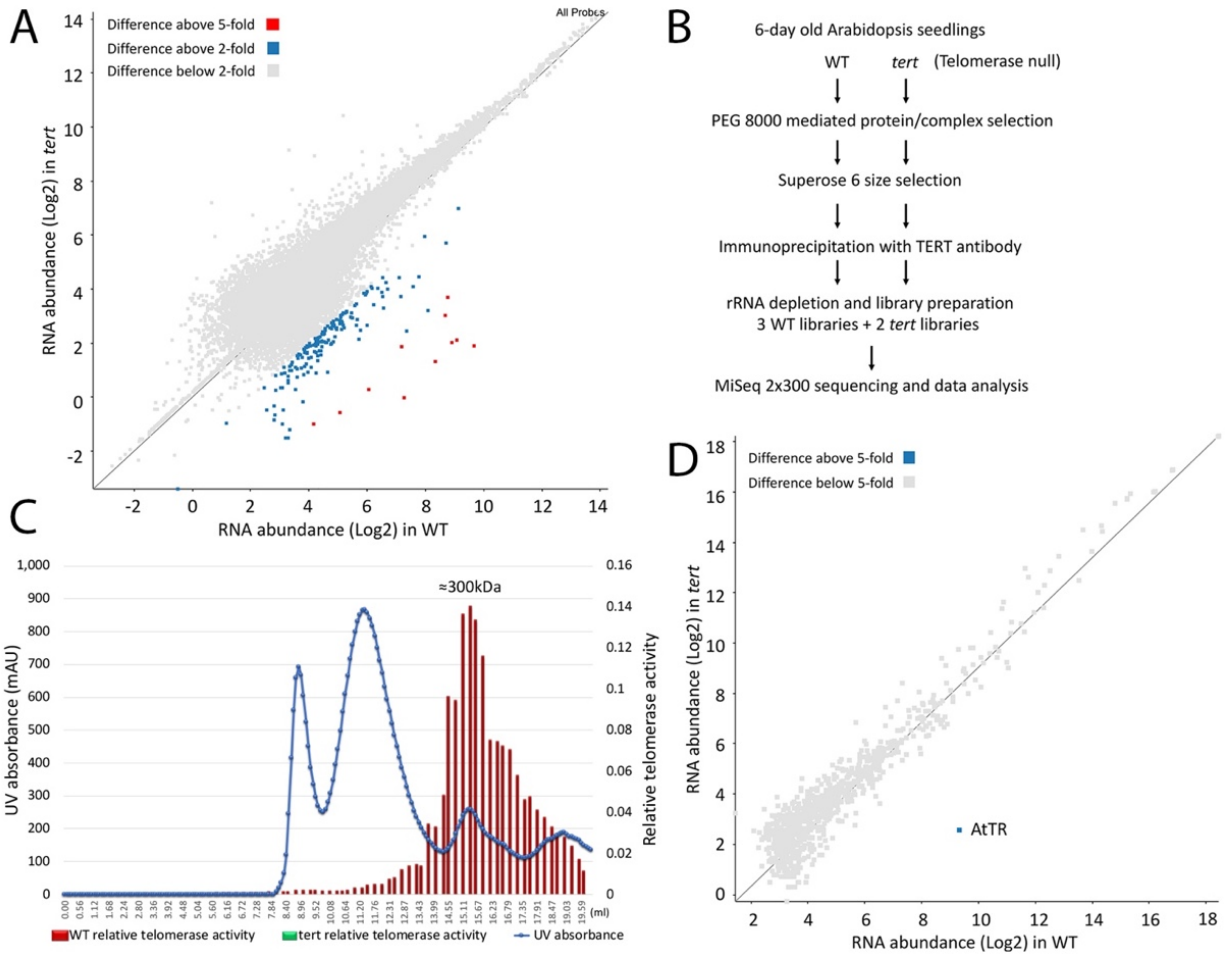


Figure 18. A single RNA species is enriched in active telomerase complexes.

(*A*) Scatter plot representing RNA targets enriched in a direct RIP seq experiment. WT and *tert* null mutant samples are compared to identify potential AtTERT-associated RNAs labelled as blue or red according to their relative enrichment in WT greater than 2-fold or 5-fold, respectively. (*B*) Schematic representation of the experimental design for identification of telomerase-associated RNAs. (*C*) Size exclusion chromatogram of *A. thaliana* protein lysate. Blue curve shows the elution profile and red bars the relative telomerase activity from each fraction. (*D*) Scatter plot of RNAs copurified with the active *A. thaliana* telomerase complex. AtTR is the only RNA molecule significantly enriched in WT samples compared to *tert* mutants.

functional telomerase RNA and instead support the recent findings of Fajkus et al (199) and Dew-Budd et al (212) that AtTER1 is not required for telomere maintenance.

AtTR is required for telomere repeat synthesis by A. thaliana telomerase

AtTR was originally described as a noncoding Pol III transcript involved in the stress response (214). AtTR bears a 9-nt sequence of 5'-CUAAACCCU-3' complementary to the *A. thaliana* 7-nt telomeric DNA sequence $(TTTAGGG)_n$ (25). Mapping of its 5' and 3' ends by rapid amplification of cDNA ends (RACE) revealed that AtTR is 268 nt in length (supplementary Figure. S2*A*). The size of endogenous AtTR was verified by northern blotting (supplementary Figure. S2*B*). Using direct terminator exonuclease treatment in combination with pyrophosphohydrolase, we found that AtTR bears a 5' triphosphate structure (supplementary Figure. S2*C*). AtTR is widely expressed, but most abundant in actively dividing cell culture. Notably, AtTR is also abundant in mature leaves where AtTERT is conspicuously absent and telomerase activity is negligible (supplementary Figure. S2*D*).

We used two genetic approaches to determine if AtTR is required for telomerase activity and telomere maintenance *in vivo*. First, we found that a homozygous T-DNA insertion allele of AtTR (Flag_410H04) completely abolished AtTR RNA production as well as telomerase activity detected by qTRAP, while plants bearing a heterozygous mutation had ~50% of the WT level of AtTR and 50% of the WT telomerase qTRAP activity (Figure. 19*A* and supplementary Figure. S3*A* and S3*B*). Terminal Restriction Fragment (TRF) analyses showed progressive shortening of the telomere tract in homozygous Flag_410H04 mutants over five generations (Figure. 19*B*), reminiscent of *tert* null mutants (19). Second, two independent CRISPR-mediated deletions that either remove a 49 nt sequence including the template or a 14 nt sequence downstream of the

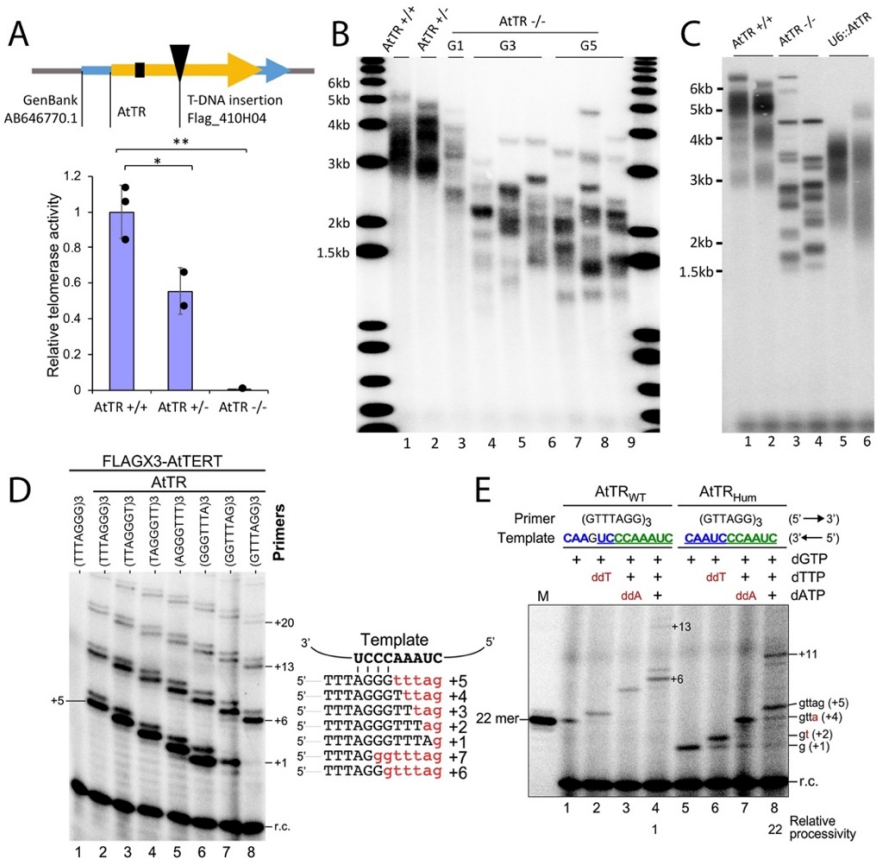


Figure 19. AtTR is the RNA template for Arabidopsis telomerase.

(*A*) Top, schematic representation of the *AtTR* gene showing the template domain (black box) and the location of a T-DNA insertion. Bottom, relative telomerase activity of WT, heterozygous and homozygous AtTR mutants determined by quantitative TRAP assay. (*B*) TRF analysis of telomere length in AtTR mutants across multiple generations. (*C*) TRF results for genetic complementation with AtTR driven by the U6 promoter. Third generation *AtTR*$^{-/-}$ mutants untransformed or transformed with U6::AtTR. **(***D***)** *In vitro* reconstitution of *A. thaliana* telomerase activity. Sequences of the putative template with the annealing position of seven circular permuted telomeric DNA primers are shown (right). The predicted primer-extended products are shown in red. *A. thaliana* telomerase is reconstituted *in vitro* from synthesized FLAGx3-AtTERT and 1.5 µM of T7 transcribed full-length AtTR (268nt). The affinity-purified telomerase was assayed for activity in the presence of ^{32}P-dGTP, dTTP, dATP and seven plant telomeric DNA primers with permuted sequences. A radiolabeled 18-mer recovery control (r.c.) was added before product purification and precipitation. Numbers to the right of the gel denote the number of nucleotides added to the primer. (*E*) Template-directed nucleotide addition by *A. thaliana* telomerase. Telomerase was reconstituted *in vitro* with AtTERT and either AtTR$_{WT}$ or AtTR$_{Hum}$. The reconstituted telomerase was assayed for activity in the presence of ^{32}P-dGTP and different combinations of dTTP, dATP, ddTTP or ddATP. A 21 nt plant telomeric DNA primer $(GTTTAGG)_3$ was used for AtTR, and an 18 nt human telomeric DNA primer $(GTTAGG)_3$ was used for the AtTR$_{hum}$. A radiolabeled 18-mer recovery control (r.c.) was added before product purification and precipitation. Numbers and sequences of nucleotides added to the primers are indicated.

template disrupted telomere maintenance (supplementary Figure. S4). We performed genetic complementation experiments on Flag_410H04 AtTR null mutants using an AtTR construct driven by the U6 promoter (U6::AtTR). Transformants with U6::AtTR expression had restored telomerase activity and increased telomere length (Figure. 19*C* and supplementary Figure. S3*C* and S3*D*). These findings confirm that AtTR is necessary for both telomerase enzyme activity and telomere maintenance in *A. thaliana*.

AtTR and AtTERT reconstitute active telomerase in vitro

We next asked whether AtTR can assemble with AtTERT *in vitro* to reconstitute active telomerase. As shown in Figure. 19*D*, recombinant FLAGx3-AtTERT protein synthesized in rabbit reticulocyte lysate was assembled with T7 RNA polymerase transcribed AtTR *in vitro* and the reconstituted telomerase was immuno-purified followed by a direct primer extension assay (Figure. 19*D*). Importantly, the primer extension activity is AtTR-dependent as no activity was detected in the absence of AtTR (Figure. 19*D*, lane 1). Seven *A. thaliana* telomeric DNA primers with permuted sequences of TTTAGGG bearing different 3' terminal sequences were examined using *in vitro* reconstituted telomerase enzyme. The reaction with $(GTTTAGG)_3$ generated a 7-nt ladder pattern of products with major bands at positions +6, +13 and +20 (Figure. 19*D*, lane 8), consistent with the 7-nt telomeric DNA repeats synthesized by *A. thaliana* telomerase. *A. thaliana* telomerase exhibited similar levels of activity with the different permuted telomeric DNA primers and generated the expected offset banding patterns (Figure. 19*D*, lanes 2-7), indicating correct primer-template alignment and specific usage of the template.

To further examine the templating function of AtTR, we generated an AtTR template mutant ($AtTR_{hum}$) with a template sequence similar to the human TR (hTR) template that allows

the synthesis of 6-nt TTAGGG repeats. The telomeric TTAGGG repeats are ubiquitously conserved in most lineages of eukaryotes (85). The 9-nt AtTR template sequence 5'-CUAAACCCU**G**AACC-3' for the synthesis of 7-nt repeats (TTTAGGG)n is flanked by a G residue at it 3' boundary and could potentially be expanded to a longer 14 nt template by mutating the G residue to A. To convert the native *A. thaliana* template sequence to a human-like template, we simply deleted one A residue in the polymerization template sequence and the non-conserved G residue in the alignment sequence, which resulted in a 12-nt 5'-CUAACCCUAACC-3' template for synthesizing TTAGGG repeats. As expected, the telomerase reconstituted from the $AtTR_{hum}$ template mutant generates the first major bands at position +5(+gttag) and the second major band at +11, indicating the addition of a 6-nt DNA repeat using the human-like template (Figure. 19*E*, lane 8). Moreover, the inclusion of dideoxy-ribonucleotides, either ddTTP or ddATP, terminated the primer extension reaction at the expected positions on the template of the $AtTR_{WT}$ and $AtTR_{hum}$ (Figure. 19*E*, lanes 2-3 and 6-7). In addition, under processive conditions with all three nucleotides, the $AtTR_{hum}$ template with a long 6-nt alignment region led to a significantly high processivity based on the ratio of +11/+5 products (Figure. 19*E*, lanes 4 and 8), consistent with a previous finding that longer templates correlate with high repeat addition processivity (90). Altogether, these data demonstrate that the template sequence 43-CUAAACCCU-51 within AtTR is a *bona fide* template for telomeric DNA repeat synthesis by *A. thaliana* TERT.

Plant TRs share a conserved secondary structure

To discern the structure of AtTR, we employed phylogenetic comparative analysis to infer a secondary structure model from the sequence alignment of plant TR homologs identified from three major clades of land plant species: angiosperms, gymnosperms and lycophytes (Figure. 20*A*).

Orthologs of AtTR were identified by searching genomic sequence data from National Center for Biotechnology using sequence homology search tools including BLAST, Fragrep2 (215) and Infernal (216). While the BLAST was able to find TR homologs from closely related species, Fragrep2 allowed identification of TR homologs from more distantly related species by utilizing position specific weight matrix (PWM) based searches with PWMs derived from multiple sequence alignments, as opposed to using the primary sequence as the search query. Collectively, we identified 85 AtTR orthologs, 70 from angiosperms, 11 from gymnosperms and 4 from lycophytes (supplementary Table S1). To infer secondary structure, multiple sequence alignment analysis was performed with 16 representative TR sequences (12 angiosperms, 3 gymnosperms and 1 lycophytes) selected from the 85 sequences to allow at least one representative from each individual order spanning three distinct clades (supplementary Figure. S5). All TR sequences including those from the basal groups, gymnosperms and lycophytes, can be reliably aligned with the TR sequence from angiosperms, revealing universally conserved structural elements of plant TRs. From the alignment of 16 divergent plant TR sequences, universal or group-specific nucleotide covariations were identified to infer base-paired structural elements (Figure. 20*B*-20*D*, supplementary Figure. S6). Comparison of TR secondary structures from three representative species, *A. thaliana* from angiosperms, *Picea glauca* (spruce) from gymnosperms and *S. kraussiana* from lycophytes, revealed three common structural features: a conserved template-PK core domain enclosed by stem P1c, a long stem that comprises consecutive short base-paired regions termed P4, P5 and P6, and a long-range base-paired stem P1a formed between the extreme distal 5' and 3' sequences (Figure. 20*B*-20*D*).

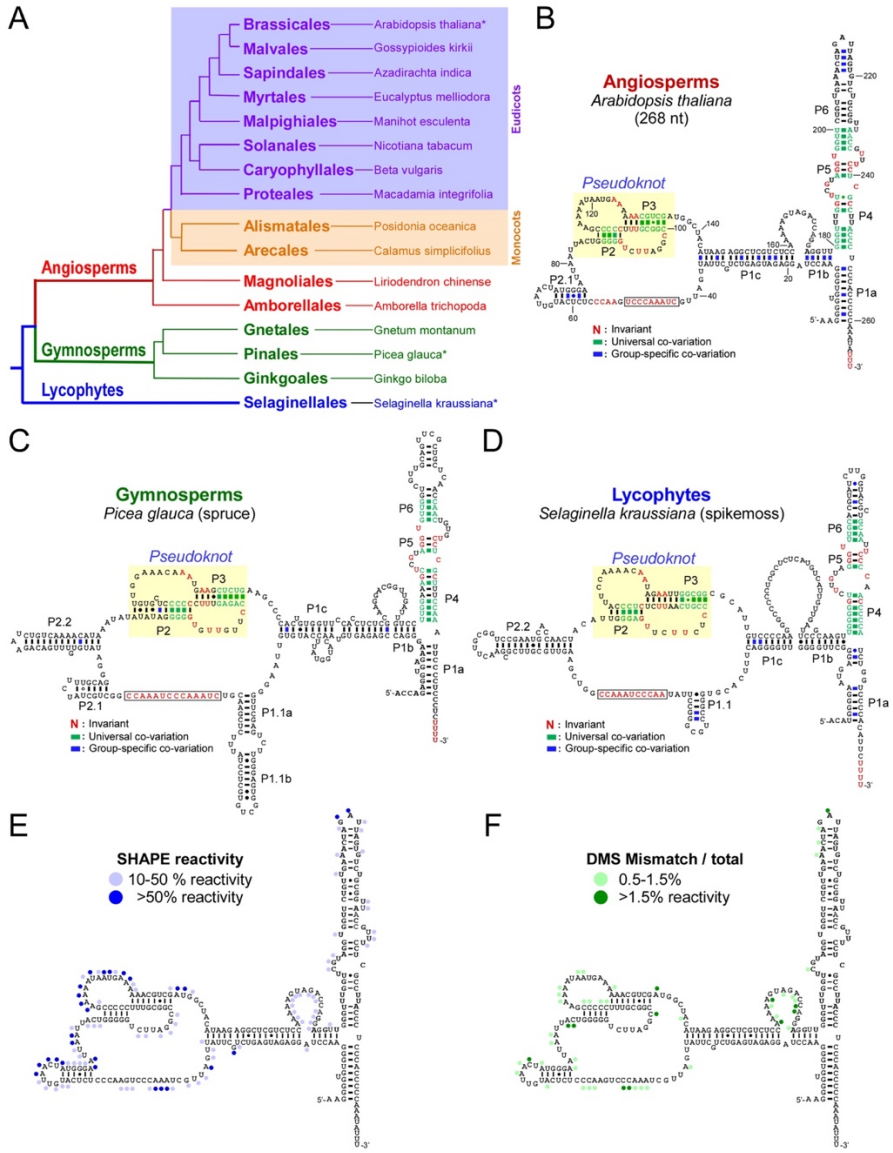


Figure 20. Plant TRs share a conserved secondary structure.

(A) Evolutionary relationship between major land plant clades. A single representative species of each order is included. An asterisk denotes the species with the secondary structure models shown in B, C, D. Representative TR secondary structures determined by phylogenetic sequence analysis are shown for (B) A. thaliana from angiosperms, (C) Picea glauca (spruce) from gymnosperms and (D) S. kraussiana (spike moss) from lycophytes. The characteristic TR pseudoknot (PK) is shaded in yellow. Universal co-variations (green line), group-specific co-variations (blue line) and plant invariant residues (red) are indicated and based on sequence alignment of 16 divergent plant species spanning 8 eudicots, 2 monocots, 2 early branching angiosperms, 3 gymnosperms and 1 lycophyte. The aligned sequences are shown in SI Appendix Figure. S5. (E) In vitro chemical probing of AtTR secondary structure by SHAPE. Chemical reactivities per nucleotide are plotted on the AtTR secondary structure. (F) In vivo chemical probing of AtTR structure by DMS-MaPseq. Average mutation frequencies per nucleotide are plotted on the AtTR secondary structure.

The plant template-PK (T-PK) core domain resembles those from ciliate, fungal and vertebrate TRs, consisting of a template, a universal PK structure formed by stems P2 and P3, and a core-enclosing stem P1c (Figure. 20*B*-20*D*). However, the plant T-PK core domain contains additional plant-specific stems, namely P1.1 (in *P. glauca* and *S. kraussiana*), P2.1 (in *A. thaliana* and *P. glauca*) and P2.2 (in *P. glauca* and *S. kraussiana*) (Figure. 20*B*-20*D*). The P1.1 stem can be found in the invertebrate echinoderm and fungal TRs, and could potentially function as a TBE (86, 88). The P2.1 and P2.2 stems are not present in all plant TRs, suggesting that they are more adaptable and maybe important for a function specific to some plant groups. One possible role for the variable P2.1 and P2.2 stems is to maintain the length of the linker between the template and the pseudoknot structure within the T-PK core domain.

In addition to the T-PK core domain, the plant TR contains a long helical structure with three consecutive short stems, P4, P5 and P6, located near the 3'end between P1a and P1b (Figure. 20*B*-20*D*). The location and structure of the plant P4/P5/P6 stem resembles the vertebrate CR4/5 domain, echinoderm eCR4/5 domain or ciliate helix IV, all of which are essential for telomerase activity (88, 111, 112). The three-way junction formed between P1a, P1b and P4/5/6 appears to be a conserved feature of plant TR (Figure. 20*B*-20*D*). This P1a-mediated three-way junction is unique to plant TR and is not found in other known TRs.

This conserved secondary structure model of AtTR is supported by chemical modification probing analysis. Selective 2'-hydroxyl acylation analysis by primer extension (SHAPE) was employed to examine the accessibility of each nucleotide in the *in vitro* folded RNA (217). N-methylisatoic anhydride (NMIA) modification of individual nucleotides was monitored and SHAPE activity plotted on the structural model to identify unpaired residues (Figure. 20*E* and

supplementary Figure. S7). Consistent with our AtTR structural model, most unpaired nucleotides showed significant SHAPE activity.

We also probed the AtTR structure *in vivo* by dimethyl sulfide (DMS) footprinting and mutational profiling (DMS-MaPseq). DMS methylates the base-pairing faces of single-stranded, unprotected adenosines and cytidines. Such modifications cause the stalling of conventional reverse transcriptases during cDNA synthesis, allowing for footprinting studies. These modifications can also result in mismatches in cDNA when TGIRT reverse transcriptase is used (218). DMS modifications were analyzed by primer extension (DMS footprinting), while DMS-induced mutational rates per position were calculated by coupling TGITR cDNA synthesis with high throughput sequencing. DMS footprinting identified 38 accessible nucleotides that mapped to predicted single-stranded residues (supplementary Figure. S8). Results of DMS MaPseq extended these findings and revealed a detailed map of nucleotide accessibility (Figure. 20*F* and supplementary Figure. S9). Accessible nucleotides were concentrated in the predicted single-stranded regions within the T-PK and P1b-P1c linker. Altogether, these *in vitro* and *in vivo* structural probing results provide strong support for our AtTR secondary structure model.

In addition to inferring the conserved secondary structure, the multiple sequence alignment of the 16 representative plant TRs spanning land plant evolution revealed five highly conserved regions (CR), CR1 to CR5, containing nucleotides that are invariant among these 16 distantly related species (supplementary Figure. S5). Such remarkable conservation of nucleotide identity usually predicts essential functions of these regions as evident in vertebrate TRs (87). CR1 corresponds to the template of AtTR. CR2 and CR3 form the universal P2 and P3 stems of the PK, while CR4 and CR5 form a P5 structural element that includes the short 3-bp P5 stem, an asymmetric internal loop and the upper part of stem P4 (Figure. 20*B*-20*D*). While lacking the P6.1

stem-loop, the universal P5 structural element of the plant TR resembles the CR4/5 domain conserved in vertebrate, fission yeast and filamentous fungal TRs (86, 111). This highly conserved P5 stem may serve as a protein binding site or play a crucial role in telomerase function.

The AtTR PK domain is essential for telomerase function and homologous to human TR

With a robust secondary structure model for AtTR, we sought to map the structural elements essential for telomerase activity. Full-length or truncated AtTR constructs were assembled with recombinant FLAGx3-AtTERT *in vitro* and the immuno-purified enzymes were analyzed for telomerase activity by direct primer extension. Analysis of three truncated AtTR fragments, 11-179, 25-153 and 42-136 (Figure. 21*A*), showed that AtTR-25-153 is the minimal PK fragment sufficient to reconstitute about 40% of wild-type activity without the P4/5/6 domain (Figure. 21*B*, lanes 2 and 3). The core-enclosing P1c stem appeared to be important for telomerase function as the AtTR-42-136 fragment with P1c removed was unable to reconstitute any significant activity (Figure. 21*B*, lane 4). Equivalent to the CR4/5 domain of human TR, the 3' P1a/4/5/6 domain of AtTR can also function *in trans* as a separate RNA molecule to stimulate the reconstituted activity from the basal 40% to 66% of wild-type level (Figure. 21*C*). A basal activity of telomerase reconstituted from the T-PK domain alone was previously reported with Trypanosome and Echinoderm TRs (88, 97), indicating an evolutionary transition of functional dependence for the two conserved TR domains.

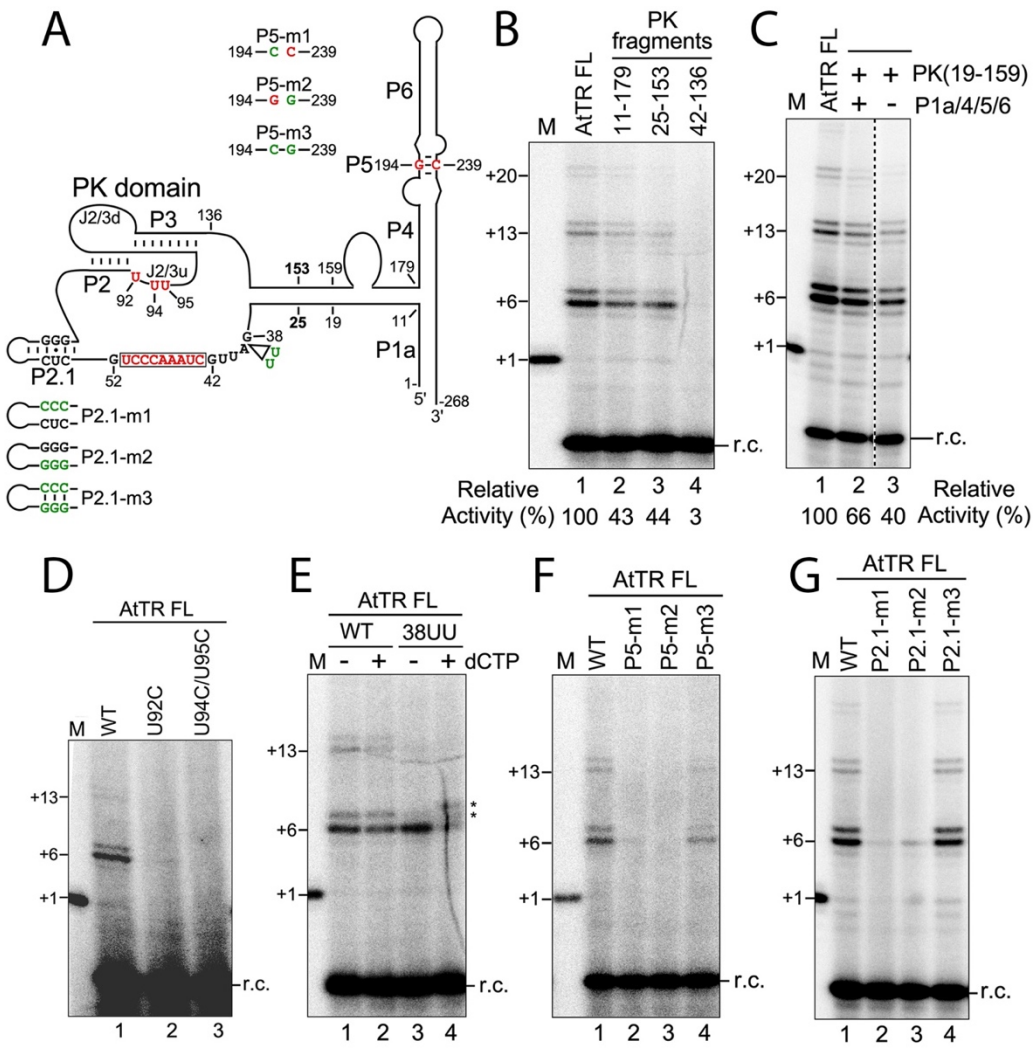


Figure 21. Functional characterization of critical structural elements in AtTR.

(*A*) A schematic of AtTR secondary structure. The 5' and 3' residues of truncated AtTR fragments are denoted on the AtTR structure. The positions and identities of specific point mutations introduced are indicated. (*B*) Identification of a minimal PK fragment and (*C*) functional analysis of stem P1a/4/5/6. Full-length AtTR (AtTR-FL) and various AtTR truncated fragments were assembled with AtTERT *in vitro* and analyzed for activity by primer extension assay. The number of nucleotides (+6, +13 or +20) added in each major band of product are indicated. The P1a/4/5/6 fragment was generated by deleting residues 25-153 from the AtTR-FL and replacing with a GAAA tetraloop. The relative activities of the reactions are indicated under the gel. A recovery control (r.c.) is shown. (*D*) The functional requirement of invariant U residues in PK domain. (*E*) The effect of P1c linker length on template boundary definition. (*F*) Compensatory mutagenesis analysis of stem P5. (*G*) Compensatory mutagenesis analysis of stem P2.1. AtTR-FL constructs bearing specific point mutations are assembled with AtTERT *in vitro* and analyzed for telomerase activity. For analyzing template boundary definition with AtTR-38UU, the reconstituted enzyme was analyzed in the absence (-) or presence (+) of dCTP in addition to dGTP, dATP and dTTP.

The PK structure of plant TRs highly resembles the PK structures in ciliate and vertebrate TRs with differences in size and complexity. In human TR PK structure, the invariant U residues in the J2/3 upstream region (J2/3u) are essential to telomerase activity (107). To determine if the invariant U residues in plant TR PK are functionally homologous to the human TR, we reconstituted telomerases with two AtTR mutants, U92C and UU94/95CC. The activity assays of the mutant enzyme showed no activity (Figure. 21*D*, lanes 2 and 3), indicating these U residues in the AtTR PK domain are absolutely required for telomerase activity. Therefore, the T-PK domains of AtTR and hTR are both structurally and functionally homologous.

Another critical function provided by the T-PK domain is defining the functional template boundary through specific structural elements, i.e. the P1 stem in vertebrate TR (91). The P1c stem in the T-PK domain of AtTR resembles the P1 stem in human TR, and presumably functions as the template boundary element. To test this idea, we generated an AtTR mutant 38UU with two U residues inserted between the P1c stem and the template to increase the linker length, a critical determinant of the template boundary. In the wild-type AtTR template, a G residue immediately flanks the 5' boundary and does not serve as a template even in the presence of dCTP substrate (Figure. 21*E*, lanes 1 and 2). However, in the presence of dCTP, the telomerase enzyme reconstituted with the AtTR mutant 38UU utilized the G residue as a template beyond the template boundary (Figure. 21*E*, lanes 3 and 4). Thus, *A. thaliana* and human telomerases share a homologous mechanism for template boundary definition.

While the overall secondary structure of AtTR is well supported by co-variation evidence and chemical probing data, we performed mutagenesis analysis to provide additional support for the highly conserved P5 stem and the plant-specific P2.1 stem (Figure. 21*A*). The 3-bp P5 stem is formed by two highly conserved regions, CR4 and CR5, with only limited co-variation support for

one of the 3 base-pairs. We thus generated AtTR full-length constructs, P5-m1 and -m2, with two single point mutations, G194C and C239G, introduced to disrupt the invariant G:C base-pairing in the P5 stem, or a compensatory mutant P5-m3 with both point mutations to restore the base-pairing (Figure. 21*A*). The activity assay showed that P5-m1 and -m2 single point mutations abolished telomerase activity (Figure. 21*F*, lanes 2 and 3), while the compensatory mutation P5-m3 restored activity (Figure. 21*F*, lane 4), consistent with the essential base-paired structure of stem P5. A similar mutagenesis approach was employed to confirm the base-paired structure and the functional importance of stem P2.1 (Figure. 21*G*). Altogether, these *in vitro* studies strongly support the robustness of the phylogenetic comparative analysis for inferring RNA secondary structure in plant TR.

Discussion

Telomerase emerged in early eukaryotes as a specialized reverse transcriptase with an integral RNA template to counteract the end-replication problem and maintain genomic integrity. While the catalytic TERT component of telomerase is conserved among eukaryotes, the TR component has diverged significantly during evolution. A missing piece in the evolutionary history of telomerase has been plant TR. Recent studies from the Fajkus (199) and Beilstein (212) labs indicated that the previously identified AtTER1 (198) was not the authentic TR in A. thaliana. The results from our independent study support this conclusion. We were unable to detect AtTER1 using two purification schemes, one designed to identify RNAs loosely associated with AtTERT, and a second more stringent approach to identify RNAs associated with partially purified, enzymatically active telomerase. The misidentification of AtTER1 in the previous study may have resulted from a primer extension strategy that employed biased primers corresponding to predicted

Arabidopsis TR template, and which inadvertently recovered a low-abundance RNA molecule derived from the RAD52 locus that co-purified with telomerase. Our next-generation sequencing approach also failed to recover AtTER2, a second telomerase-associated RNA proposed to negatively regulate enzyme activity in response to DNA damage (198, 201). Re-evaluation of the AtTER2 locus in relation to telomerase and telomeres is now underway.

Nevertheless, the single RNA enriched by 100-fold in enzymatically active telomerase fractions from our more stringent purification scheme was AtTR, the same RNA molecule uncovered independently by the Fajkus lab using an *in silico* strategy to find plant TRs (199). To investigate the function of AtTR, we employed a combination of Arabidopsis genetics and in vitro reconstitution experiments using a rigorous non-PCR assay of direct primer extension to test the authenticity of this putative telomerase RNA template. We determined that AtTR was not only required for telomere maintenance in vivo, but also possessed a functional template for telomeric DNA synthesis by AtTERT in vitro. Our observations agree with those of Fajkus et al. and confirm that AtTR is the bona fide telomerase RNA subunit for A. thaliana.

AtTR was first described in 2012 by Wu and collaborators as a root-specific, conserved Pol III-dependent ncRNA (214). The ATTR gene (Genbank AB646770.1) includes a U6-like Type III promoter and poly(T) terminator. The promoter has a consensus cis upstream sequence element (USE) and a TATA box-like element 25 bp upstream of the transcription start site (TSS). The discovery of plant TRs being Pol III RNA transcripts leads to an interesting question: was the first TR a Pol II or Pol III transcript? TR was originally identified in ciliates as a small Pol III RNA transcript with sizes ranging from 140 to 210 nt (Figure. 22). RNA polymerase III is generally employed for transcribing small RNA such as 5S rRNA and tRNA due to its sequence-dependent termination at a U-rich termination site. A large RNA would encounter a high frequency of U-rich

sequences and suffer premature termination with Pol III transcription, which is consistent with the small size of ciliate TR (119). Surprisingly, TRs identified later in vertebrates and fungi are larger Pol II transcripts with sizes of 312-559 nt and 920-2425 nt, respectively (86, 87). While it seems reasonable to assume that the Pol III TR transcript is more ancestral, TRs from early branching flagellates, including Trypanosomes, are large Pol II transcripts ranging between 781-993 nt (Figure. 22). Discerning the origin of TR will require discovery of TRs from the early branching lineages of eukaryotes, a daunting task considering the extremely divergent nature of TR.

The conserved secondary structures of plant TRs presented in this study were determined by employing phylogenetic comparative analysis, a gold standard for inferring RNA secondary structures (87, 219). Moreover, the secondary structure of AtTR was verified by in vitro and in vivo chemical probing approaches under native conditions as well as mutagenesis analysis using an in vitro reconstitution system. In the AtTR structure, the most crucial structural element is the PK, which is conserved in all known TRs except Trypanosome (Figure. 22). Trypanosome TR contains two structural domains, the template-core and eCR4/5, both of which are required for telomerase activity in vitro and can function in trans as two separate RNA fragments (97). However, the minimal template core domain of Trypanosome TR does not contain a PK, arguing that the critical TR PK was a later adaptation. Nevertheless, helix III of Trypanosome TR is potentially homologous to the PK forming helix III of Tetrahymena TR as both helices are located between the template and the core enclosing helix, i.e., helix I in Tetrahymena TR or P1 stem in other TRs.

The PK structure of Tetrahymena TR only requires formation of a 4 bp stem between the loop sequence of helix III and an upstream complementary sequence (Figure. 22). This 4 bp stem is structurally equivalent to the vertebrate P2 stem which is longer and contains two consecutive stems, P2a and P2b, and with an additional P2a.1 stem in the mammalian TR PK (Figure. 22).

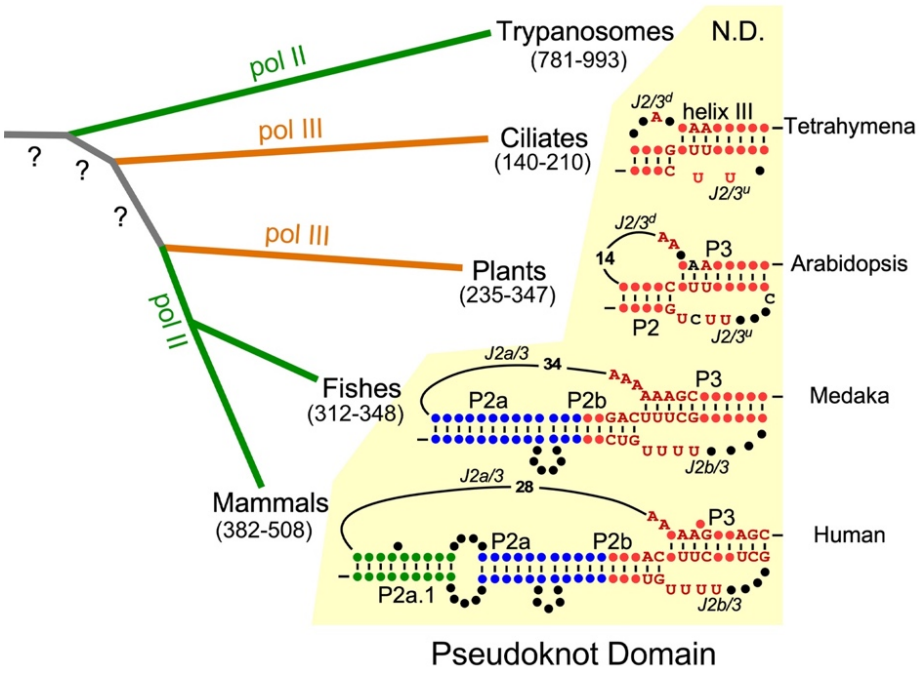


Figure 22. Evolution of TR pseudoknot structures.

A simplified phylogenetic tree of major eukaryotic lineages is shown in the left panel. Branch length in the tree does not reflect evolutionary distance. The lineages with TR transcribed by Pol II (green) and Pol III (orange) are depicted. The size range of TRs from each group is indicated. The PK structures of TRs from the major groups of eukaryotes including ciliates, plants, fishes and mammals are shown in the right panel. Trypanosome TR does not have a PK structure in the template core domain (97). The P2 and P3 stems conserved from ciliates to mammals are shown in red with highly conserved nucleotides explicitly denoted. The vertebrate-specific stem extension P2a is shown in blue while the mammal-specific stem extension P2a.1 is shown in green. The length of joining sequences, J2/3 upstream (J2/3u) or downstream (J2/3d) regions, between stems P2 and P3 are indicated.

How this primitive ciliate TR PK evolved to the more complex vertebrate TR PK has been unclear. The structure of plant TR PK now provides an explanation for the structural transition from ciliate to vertebrate PK. Similar to ciliate PK, plant PK contains a short unstable 4 bp P2 stem and a longer 8-9 bp P3 stem. DMS chemical probing of the *A. thaliana* TR PK reveals mild modification of the P2 stem, consistent with a more unstable helix (Figure. 20F). Notably, the ciliate and plant PK structures differ in the length of the joining sequences, J2/3 upstream (J2/3u) and J2/3 downstream (J2/3d) (Figure. 22). The length of J2/3u increases from 3 nt in Tetrahymena to 8 nt in plants, similar to the 8 nt J2b/3 in vertebrate TR PK (Figure. 22). The length of J2/3d sequence also increases from 4 nt in Tetrahymena to 14 nt in the *A. thaliana* PK. We propose that the longer J2/3d makes it possible to expand the short 4 bp P2 stem to a longer P2a/P2b stem in vertebrate PK during evolution. Notably, plant TR contains additional stems (P2.1 and P2.2) located between the template and the P2 stem (Figure 22). These additional stems may reflect selective pressure to maintain the spatial constraints for the enzyme active site as the P2 stem expands during evolution. Therefore, the plant TR PK provides an evolutionary bridge for the structural transition from ciliate TR to vertebrate TR.

Materials and methods

Plant material, growth conditions, and transformation

Arabidopsis thaliana accession Col-0, WS, *attr* (Flag_410H04), and *tert* (SALK_041265C) were used in this study. Cell line T87 was obtained from ABRC and was originally derived from *A. thaliana* accession Col-0. The cell culture was maintained as indicated by the ABRC, passed every seven days in NT-1 media and grown under continuous light at room temperature with constant shaking at 120 rpm. Seeds were sterilized in 50% bleach with 0.1% Triton X-100 and

then plated on half Murashige and Skoog (half MS) medium with 0.8% agar. Plants were grown at 22°C under long day light conditions. AtTR was placed under the control of the U6 promoter in the pHSN6A01 vector. Guide RNAs targeting the sequences surrounding the template of AtTR were cloned into the pDs-Sa-Cas9 vector for transformation into Col-0 plants using *Agrobacterium*-mediated transformation as described (220). For genetic complementation, third generation $AtTR^{-/-}$ were transformed with *A. tumefasciens* GV3101 containing pHSN6A01 U6::AtTR. Transformants were selected on hygromycin in T1 and analyzed for telomere phenotypes. In parallel, untransformed fourth generation $AtTR^{+/+}$ and $AtTR^{-/-}$ plants were analyzed.

Gel filtration of active telomerase

Five-day-old WT (Col) and *tert* seedlings were ground in liquid nitrogen and homogenized in extraction buffer (50 mM Tris-OAC pH7.5, 100 mM KGlu, 5 mM $MgCl_2$, 20 mM EGTA, 15 g/L Polyvinylpyrrolidone (PVP), 10% Glycerol, 20 ul/ml Plant protease inhibitor cocktail (Sigma), 1 ul/ml RNase OUT recombinant ribonuclease inhibitor (Thermo Fisher), 0.2 mM PMSF, and 1.5 mM DTT). The homogenates were centrifuged at 14,000 r.p.m for 15 min at 4°C and the supernatant was combined with 10% final concentration of PEG8000 to precipitate protein complexes for 45 min at 4°C. The precipitation was collected and resuspended with buffer TERT (50 mM Tris-OAC pH7.5, 150 mM KGlu, 5 mM $MgCl_2$, 5% Glycerol, 20 ul/ml Plant protease inhibitor cocktail (Sigma), 1 ul/ml RNase OUT (Thermo Fisher), 0.1 mM PMSF, and 1.5 mM DTT). After three rounds of centrifugation, supernatant was injected into an AKTA FPLC system, and the proteins were fractionated through a Superose6 Increase 10/300 GL column (GE Healthcare) driven by buffer TERT. Fractions were collected to measure telomerase activity by qTRAP.

RIP seq

Anti-AtTERT antibody was affinity purified with an EpiMAX affinity purification kit (abcam) following the manufacturer's protocol. It was preincubated with protein A magnetic beads (Dynabeads) before IP experiments. For the direct RIP seq, 1.2 g of WT (Col-0) and tert Arabidopsis flowers were ground in liquid nitrogen and homogenized in buffer RIP (100 mM Tris-OAC pH7.5, 100 mM KGlu, 1 mM MgCl2, 0.5% Triton X-100, 0.1% Tween 20, 20 ul/ml Plant protease inhibitor cocktail (Sigma), 1 ul/ml RNase OUT (Thermo Fisher) and 2.5 mM DTT). After clearing by centrifugation, protein complexes were immunoprecipitated using preincubated anti-AtTERT magnetic beads for 2.5 h at 4°C. After incubation, beads were washed with buffer RIP for seven times and resuspended with 1 ml TRIzol reagent (Invitrogen) to extract RNA. For RIP seq after gel filtration, fractions with peak telomerase activity were incubated in buffer TERT (50 mM Tris-OAC pH7.5, 150 mM KGlu, 5 mM MgCl2, 5% Glycerol, 20 ul/ml Plant protease inhibitor cocktail (Sigma), 1 ul/ml RNase OUT (Thermo Fisher), 0.1 mM PMSF, and 1.5 mM DTT) with preincubated anti-AtTERT magnetic beads for 3 h at 4°C. Beads were washed with buffer RIP seven times, and the remaining RNA was extracted following Direct-zol RNA kits (ZYMO research) including in-column DNase treatment. After rRNA depletion, construction of Illumina sequencing libraries was performed with the NEBNext Ultra Directional RNA Library Prep Kit, and libraries were sequenced on an 300x2 Illumina MiSeq platform by Texas A&M AgriLife Genomics and Bioinformatics Service.

In vivo *DMS modification*

DMS treatment was performed as described with a few modifications (221, 222). For DMS footprinting, Arabidopsis cell culture in growth medium was mixed with DMS (Sigma, Cat#:

D186309) to a final concentration of 0.75%. Incubation was applied with gentle shake in vacuum condition for 5 or 10 min. After adding β-mercaptoethanol to quench the reaction, materials were washed five times with miracloth wrap. The dry materials were immediately frozen in liquid nitrogen and stored at -80°C.

For DMS MaPseq, four-day-old WT (WS) seedlings were treated with 1% DMS or water (Mock samples) in DMS reaction buffer (40 mM HEPES pH7.5, 100 mM KCl and 0.5 mM MgCl2). 7.5 min of vacuum incubation was applied twice with a thorough mix in between. Materials after DMS incubation were washed and collected as described previously (221).

DMS footprinting

10 µg total RNA extracted from each DMS-treated sample was mixed with ^{32}P-radiolabelled gene-specific primers in annealing buffer (10 mM Tris-HCl pH7.5 and 20 mM KCl) of total 10 µl volume. The mixture was heated at 75°C for 3 min, annealed at 55°C for 15 min, and stabilized at 4°C for 2 min. After annealing, 10 µl reverse transcription (RT) reaction including 1x SuperScript IV buffer, 1 mM DTT, 1 mM dNTPs, 1 µl RNaseOUT (Thermo Fisher) and 1 µl SuperScript IV reverse transcriptase (Thermo Fisher) was added. The reaction proceeded for 1 h at 60°C. Reaction products were alkali-treated to hydrolyze the RNA, neutralized and precipitated before loading into a 7M Urea 8% (19:1) polyacrylamide gel. The gel image was collected with a Typhoon FLA 9500 (GE Healthcare) and bands were quantified using Quantity One (Bio-Red).

Target-specific DMS-MaPseq

Target-specific DMS-MaPseq was performed as described (218, 223) with modifications. Total RNA was extracted from DMS or Mock treated samples using RNA Clean & Concentrator-

5 (ZYMO research) with in-column DNase digestion. RNA quality was analyzed on agarose gels. 5 μg high quality RNA was combined with gene-specific primers (5 pmol each) in a total volume of 11 μl. The mixture was heated at 75°C for 3 min and annealed at 55°C for 15 min. TGIRT reaction buffer including 4 μl 5x First-Strand buffer (Thermo Fisher), 1 μl 0.1M DTT, 1 μl RNaseOUT (Thermo Fisher) and 1 μl TGIRT-III (Ingex, Cat#: TGIRT50) was added and the solution was incubated at room temperature for 30 min. Then, 2 μl 10 mM dNTP was added and the well-mixed reaction was processed at 60°C for 2.5 h. After RT, 1 μl cDNA solution was directly added into a 50 μl PCR reaction using Phusion High-Fidelity DNA Polymerase (NEB) to amply AtTR or ACT2 mRNA with an approximate product size of 260 bp. PCR products were gel purified and quantified by the Qubit dsDNA BR assay kit (Thermo Fisher). Without fragmentation, the cleaned PCR products were directly assembled into Illumina sequencing libraries using the NEBNext Ultra II DNA Library Prep Kit with 25 ng input. One Mock library and two DMS libraries were built for each genotype. Finally, the libraries were quantified using Agilent TapeStation before sequencing on an 150x2 Illumina NextSeq 500 platform at Texas A&M University.

SHAPE

SHAPE was performed as described with modifications (217). 2 pmol gel purified AtTR was folded in SHAPE buffer (100 mM HEPES pH8.0, 100 mM NaCl, and 7 mM $MgCl_2$) at 37°C for 30 min. NMIA of 6.5 mM final concentration was used for RNA modification. After resolving primer extension products on a 7M Urea 8% (19:1) polyacrylamide gel, the image was collected with a Typhoon FLA 9500 (GE Healthcare) and bands were quantified using Quantity One (Bio-Red).

Northern Blotting

Total RNA was extracted from Arabidopsis cell culture by Direct-zol RNA kits (ZYMO research) including in-column DNase treatment. 15 μg total RNA was fractionated on a 7M Urea 4% (19:1) polyacrylamide gel together with *in vitro* transcribed AtTR as a molecular weight marker. RNA was semi-dry transferred to a Hybond+ membrane (GE Healthcare) and hybridized for 16 h at 65°C with a combination of three ^{32}P-radiolabelled oligonucleotides complementary to AtTR. After the membrane was washed, the gel image was collected with a Typhoon FLA 9500 (GE Healthcare).

Terminal Restriction Fragment (TRF), TRAP, and quantitative TRAP (qTRAP)

TRF, TRAP, and qTRAP assays were performed as previously described (202, 205, 224) with one modification. For TRAP and qTRAP, partially purified telomerase was incubated with corresponding reactions at room temperature for 30 min instead of 37°C.

Bioinformatics analysis

AtTR orthologs were identified by standalone BLAST (version 2.2.31+) searches initially using AtTR as query from closely related species. The BLASTN search was performed with the -task dc-megablast parameter to allow for identification of more variable sequences. For more distantly related species, position weight matrix (PWM) search using fragrep 2 (215) was performed for candidate identification. The PWM was created using sequence alignment from AtTR orthologs identified via BLAST and the match scores were relaxed during PWM searches to allow for identification of more divergent sequences. Once a reliable secondary structure was established using the TRs identified via BLAST and fragrep2, secondary structure-based searches

were performed using Infernal (216) for identification of orthologs from more distantly related species.

Sequence alignment analysis

Multiple sequence alignment of land plant TRs was performed initially using the ClustalW algorithm of the Bioedit program. Manual refinements were made to preliminary alignments with highly conserved regions and invariant primary sequence motifs as anchor points. Sequences from closely related species of the Brassicaceae family were aligned first and the alignment was expanded by including sequences in order of phylogenetic relationships to the existing alignment.

In vitro *reconstitution of Arabidopsis telomerase*

3xFLAG tagged Arabidopsis TERT (AtTERT) was expressed in rabbit reticulocyte lysate (RRL) from the p3xFLAG-AtTERT plasmid using the TNT Quick Coupled transcription/translation kit (Promega) following manufacturer's instructions. The AtTR fragments were *in vitro* transcribed by T7 RNA polymerase, gel purified and assembled with TERT protein for 30 min at 30˚C at a final concentration of 1.5 µM (97).

Telomerase direct primer extension

12 µl of in vitro reconstituted telomerase enzyme was immuno-purified with 3 µl of anti-FLAG M2 magnetic beads (Sigma M8823) at room temperature for 1 hr. The telomerase enzyme on beads was assayed in a 10 µl reaction containing 1X telomerase reaction buffer (50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 0.5 mM MgCl2, 5 mM BME and 1mM spermidine), 1µM DNA primer, and specified dNTPs or ddNTPs and 0.18µM of 32P-dGTP (3,000 Ci/mmol, 10 mCi/ml; Perkin-

Elmer). Reactions were incubated at 30˚C for 60 min and terminated by phenol/chloroform extraction, followed by ethanol precipitation. The 22-mer size marker was prepared in a 10 µl reaction containing (GGGTTTA)3 oligo, 1x TdT reaction buffer, 5 units of terminal deoxynucleotidyl transferase (TdT, Affymetrix) and 0.1 µM of 32P-dGTP. The reaction was incubated at room temperature for 3 sec and terminated by addition of 10 µl 2x formamide loading buffer (10mM Tris-HCl, pH8.0, 80% (vol/vol) formamide, 2 mM EDTA, 0.08% bromophenol blue, and 0.08% Xylene cyanol). The DNA products were resolved on a 10% (wt/vol) polyacrylamide/8 M urea denaturing gel, dried, exposed to a phosphorstorage screen and imaged on a Typhoon gel scanner (GE Healthcare).

Supplementary data

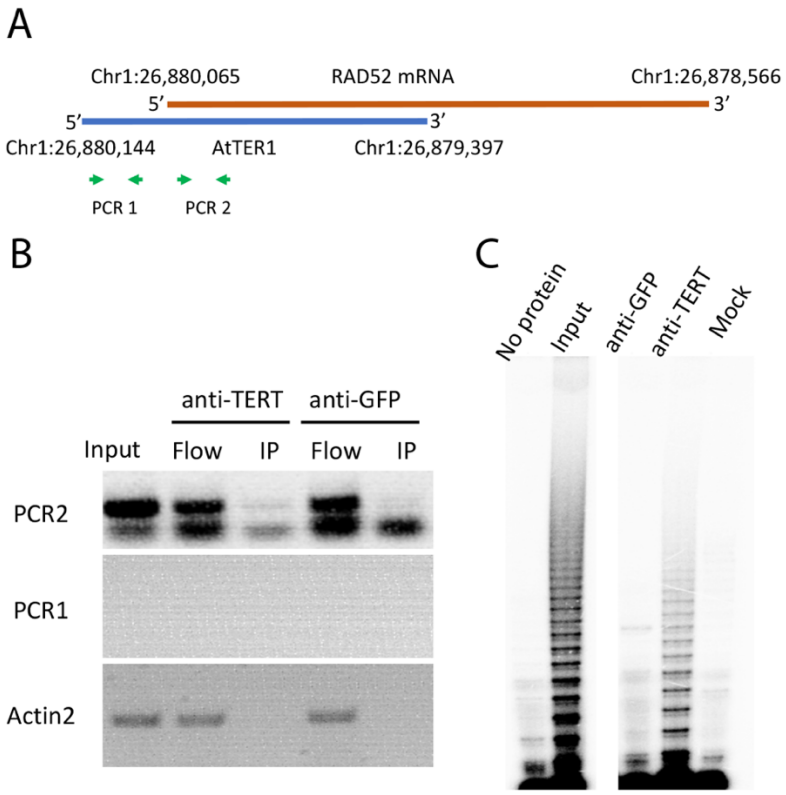


Figure S1. AtTER1 is not recovered in an AtTERT IP. (A) Schematic representation of the physical map of AtTER1 and RAD52 mRNA. (B) RT-PCR experiments were conducted using RNAs independently collected from AtTERT-IP and GFP-IP. Primers used in the experiments are indicated in panel A. (C) TRAP was performed with identical samples in B to verify that active telomerase was purified from AtTERT-IP but not the GFP-IP.

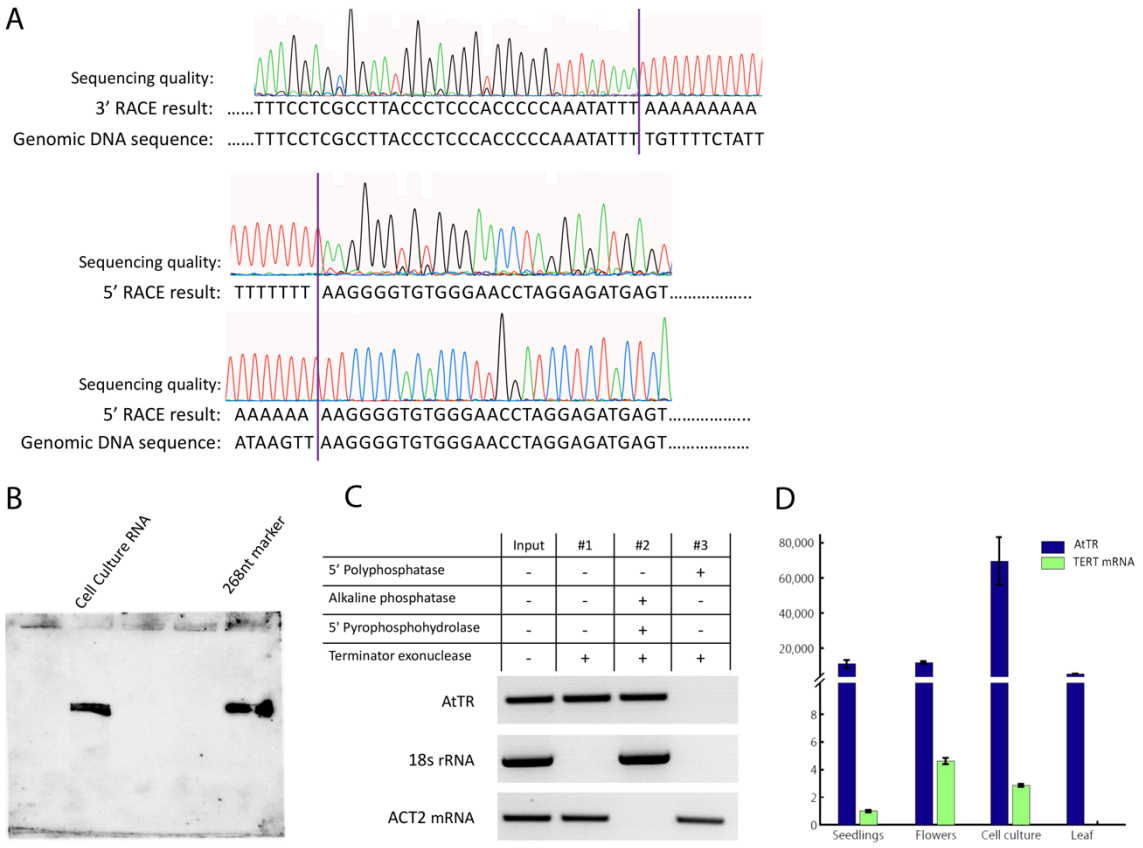


Figure S2. Characterization of AtTR. (A) 5’ and 3’ RACE of AtTR define it as a 268nt lncRNA derived from Chr 2 position 12619067 to 12619334. Both polyA and polyT tails were used in 5’ RACE to precisely map the transcriptional start site. (B) Northern blotting using total RNA from *A. thaliana* cell line T87 confirmed that AtTR is 268 nt in length. (C) Enzymatic probing of 5' end structure shows that AtTR has a 5’ triphosphate. 18s rRNA and ACT2 mRNA served as controls for 5' monophosphate or capped RNAs, respectively. (D) qPCR indicates that AtTR is expressed throughout the plant life cycle and it is enriched in rapidly dividing *A. thaliana* cell culture.

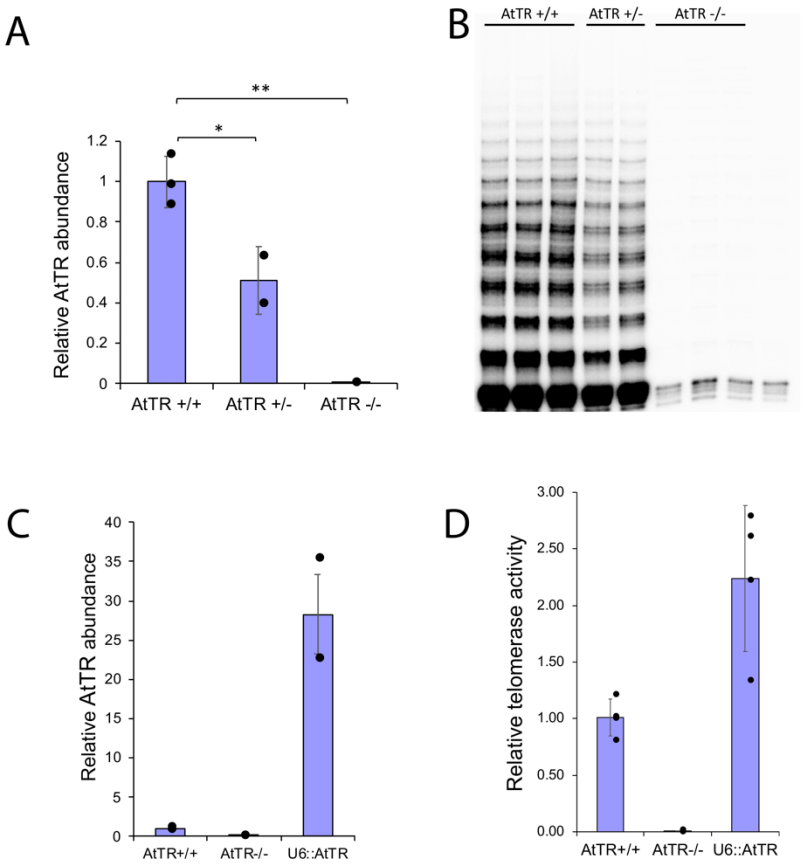


Figure S3. AtTR is the *bona fide* template of *A. thaliana* telomerase. (A) qPCR analysis of the Flag_410H04 T-DNA insertion line defines this an AtTR null mutant. (B) Telomerase activities of WT (AtTR +/+), heterozygous (AtTR +/-), and homozygous (AtTR -/-) Flag_410H04 segregants were determined by TRAP. (C) AtTR abundance was measured by RT-qPCR in untransformed AtTR +/+, AtTR -/- and U6::AtTR complementation lines. AtTR expressed from the U6 promoter in the AtTR -/- background results in a ~28-fold average overexpression of AtTR as compared to WT plants. (D) Telomerase activity was measured by qTRAP in AtTR +/+, AtTR -/- and U6::AtTR complementation lines. Overexpression of AtTR results in increased telomerase activity.

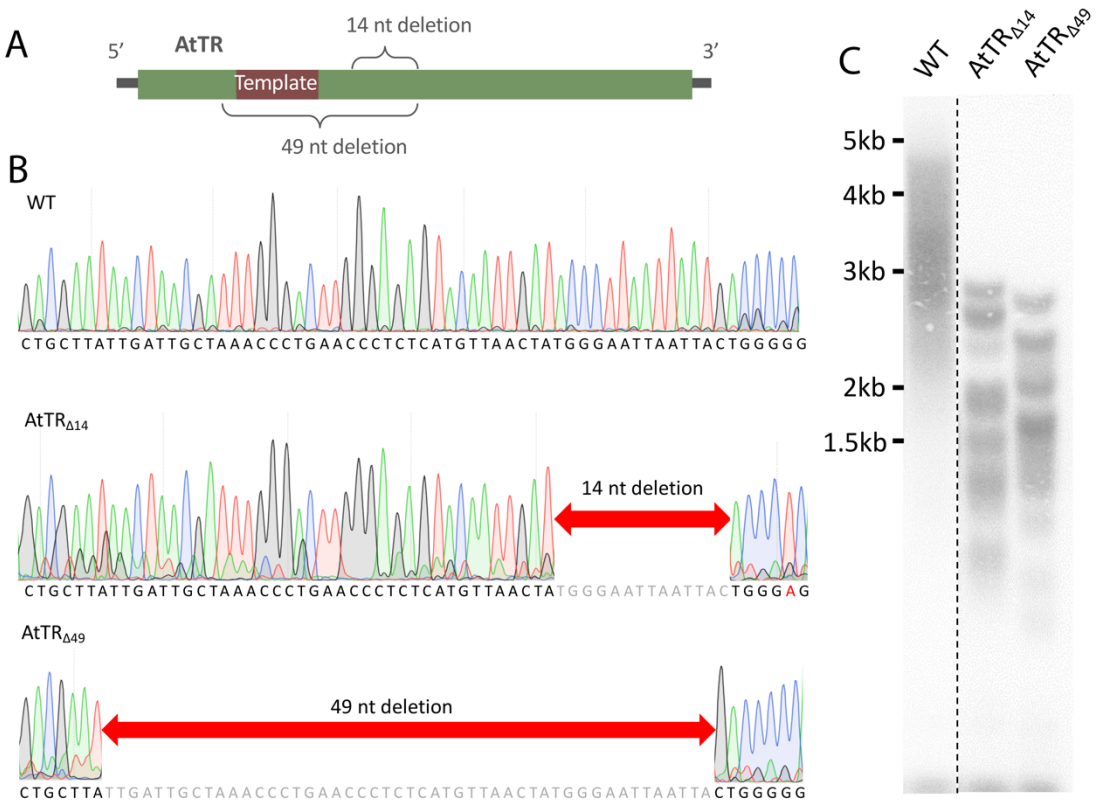


Figure S4. Two independent CRISPR alleles of AtTR abolish telomere maintenance. (A) Schematic representation of the two CRISPR alleles of AtTR (AtTR$_{\Delta14}$ and AtTR$_{\Delta49}$). (B) Sequence profiles of independent homozygous plants for each CRISPR AtTR allele. (C) TRF analysis shows plants homozygous for the CRISPR AtTR alleles have shorter, more homogeneous telomere tracts than the WT.

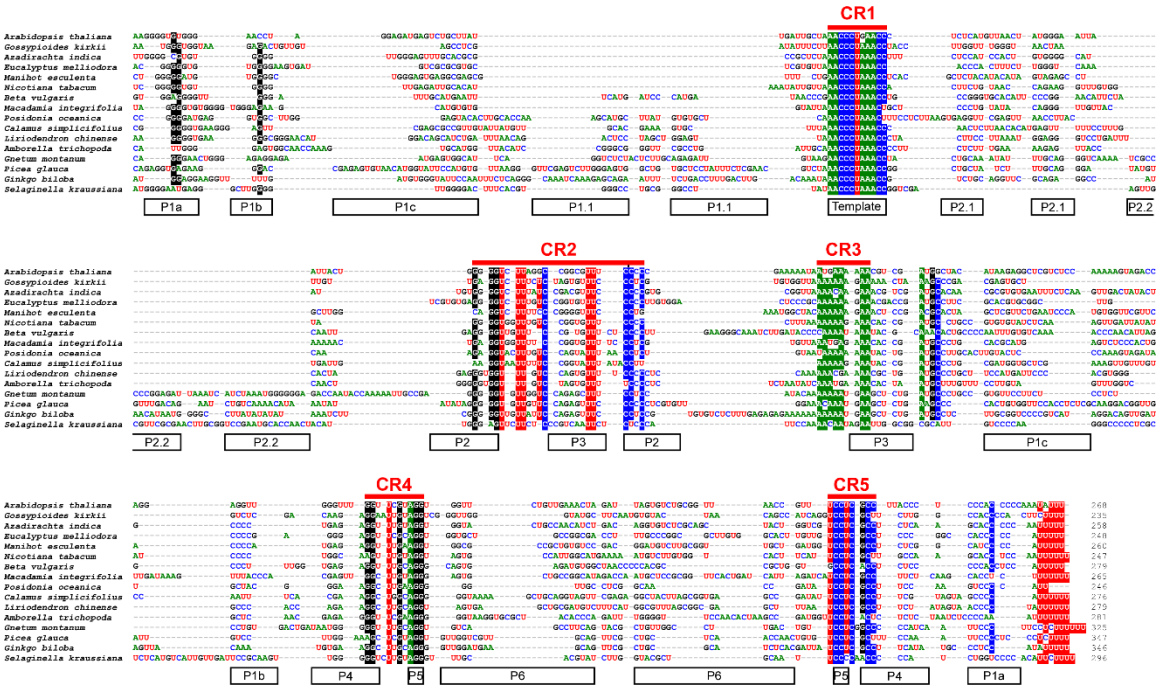
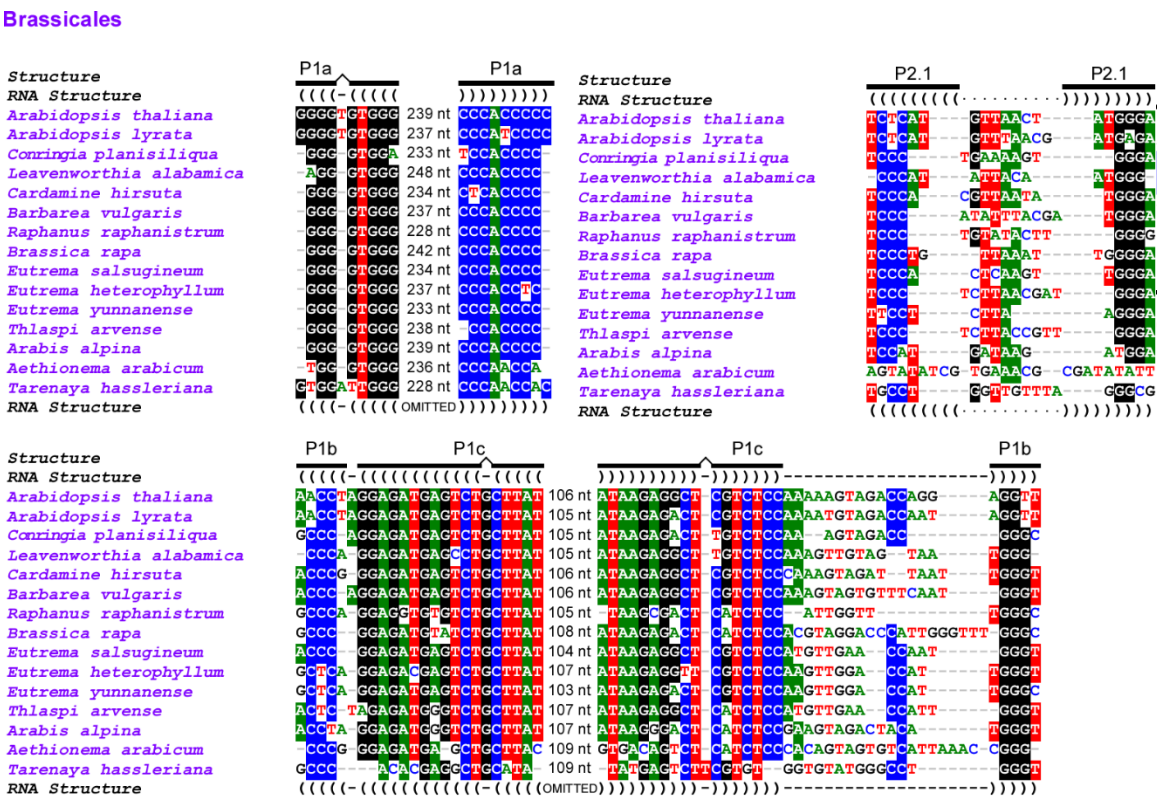
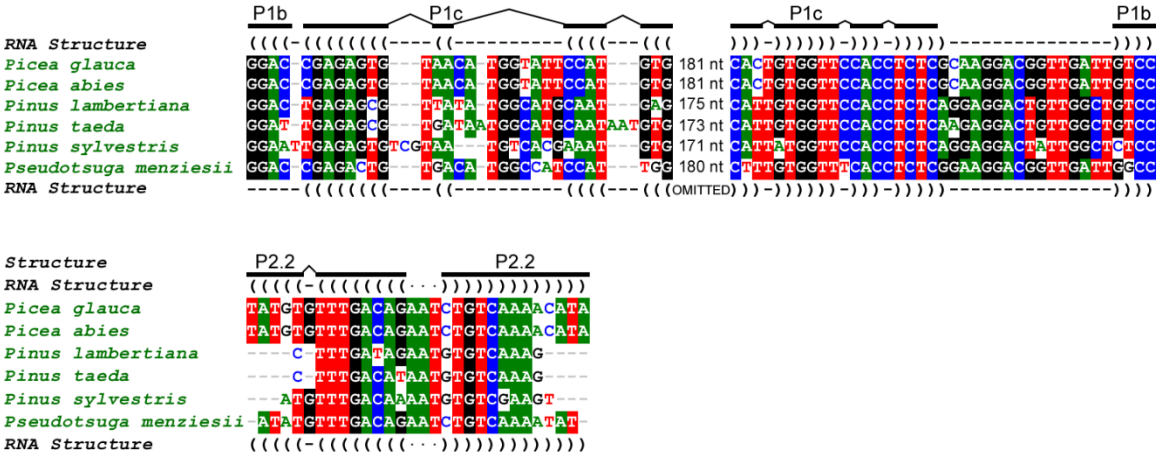


Figure S5. Multiple sequence alignment of plant TRs. Alignment of TR sequences from 8 eudicots, 2 monocots, 2 early branching angiosperms, 3 gymnosperms and 1 lycophyte species representative of land plants. Multiple sequence alignment was performed using the ClustalW algorithm in the BioEdit program. Highly conserved regions and motifs were aligned first followed by alignment of intervening sequences using conserved regions as anchors. The total number of nucleotides in each TR is indicated at the end of the respective sequence. Individual nucleotides are colored by identity (A; green, G; black, U; red, C; blue) and nucleotides that are conserved in ≥ 75% of given plant species are shaded (White text on colored background). Five conserved regions (CRs) are indicated with red lines above the alignment. The template and base-paired helices (P1-P6) in the secondary structures are denoted within white boxes below the alignment.

A


Brassicales
Structure
RNA Structure
Arabidopsis thaliana
Arabidopsis lyrata
Conringia planisiliqua
Leavenworthia alabamica
Cardamine hirsuta
Barbarea vulgaris
Raphanus raphanistrum
Brassica rapa
Eutrema salsugineum
Eutrema heterophyllum
Eutrema yunnanense
Thlaspi arvense
Arabis alpina
Aethionema arabicum
Tarenaya hassleriana
P1a
P1a
P2.1
P2.1
P1b
P1c
P1c
P1b
P6
P6
239 nt
237 nt
233 nt
248 nt
234 nt
237 nt
228 nt
242 nt
234 nt
237 nt
233 nt
238 nt
239 nt
236 nt
228 nt
106 nt
105 nt
105 nt
106 nt
106 nt
105 nt
108 nt
104 nt
107 nt
103 nt
107 nt
107 nt
109 nt
109 nt
OMITTED

B


Pinales
RNA Structure
Picea glauca
Picea abies
Pinus lambertiana
Pinus taeda
Pinus sylvestris
Pseudotsuga menziesii
P1b
P1c
P1c
P1b
P2.2
P2.2
181 nt
181 nt
175 nt
173 nt
171 nt
180 nt
OMITTED

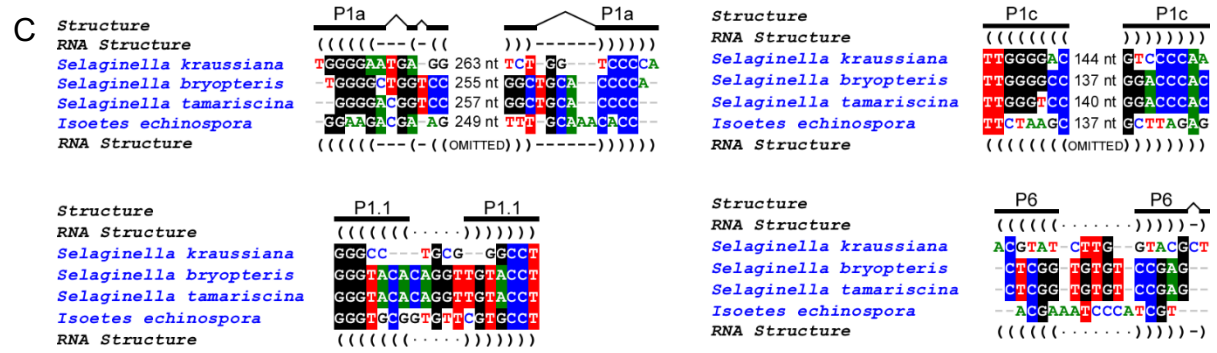


Figure S6. Sequence alignments of TR structural elements from respective clades to identify group-specific co-variations. Individual nucleotides are colored by identity (A; green, G; black, T; red, C; blue) with shaded residues shown as white text in colored background. Variable shading was applied to show clarity of co-variation. Individual TR elements are indicated above each alignment block with secondary structure representation shown using dot-bracket notations at the bottom. Intervening residues of structural elements that form long range base pairing are omitted and the number of nucleotides omitted are shown between the base paired regions. (A) Sequence alignments of TR structural elements from 15 species belonging to the Brassicales order including AtTR (Figure 3A). Shading of P1a (80%), P2.1 (75%), P1b/P1c (60%) and P6 (80%) are shown. (B) Sequence alignments of TR structural elements of 6 species from order Pinales including *P. glauca* TR (Figure 3C). Shading of P1b/P1c (50%) and P2.2 (65%) are shown. (C) Sequence alignments of TR structural elements of 4 species from division lycophyta including *S. kraussiana* TR (Figure 3D). Shading of 50% shown for all elements P1a, P1c, P1.1 and P6.

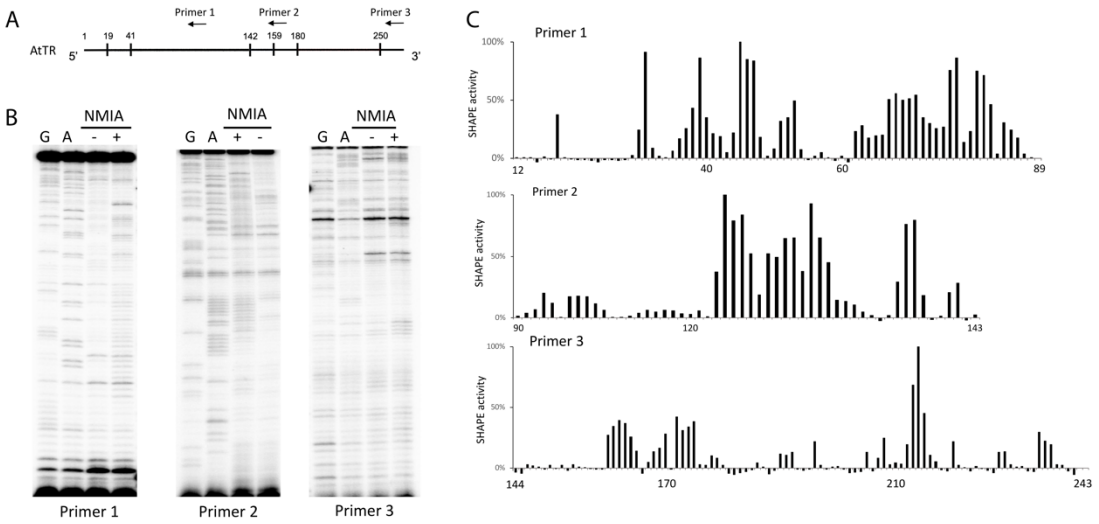


Figure S7. SHAPE data support the structural model of AtTR. (A) Diagram of AtTR and primers used in the SHAPE assay. (B) Primer extension results for *in vitro* transcribed AtTR in the presence and absence of NMIA. (C) Quantified SHAPE activities are plotted along the AtTR sequence.

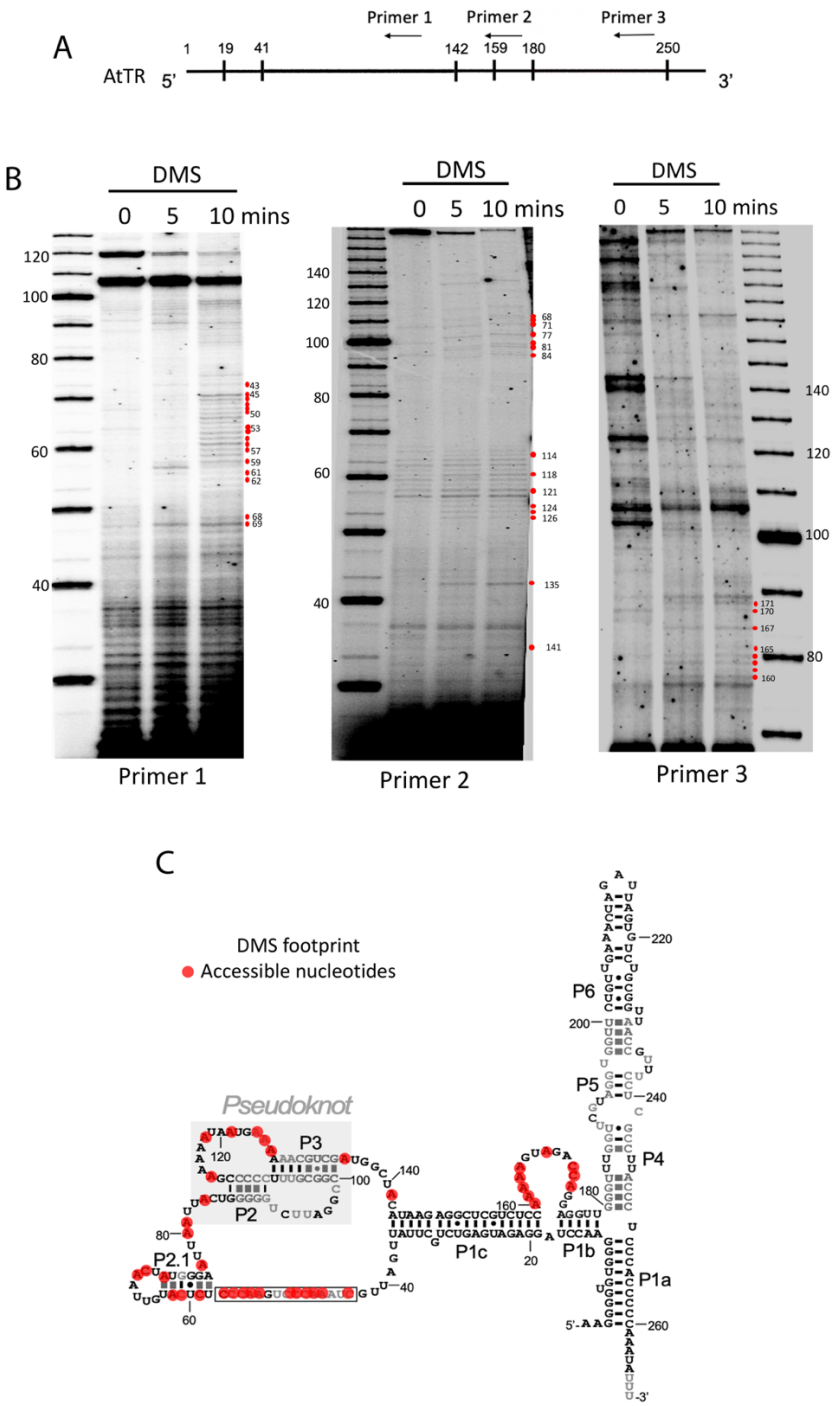


Figure S8. *In vivo* DMS footprinting uncovers accessible nucleotides in AtTR. (A) Schematic representation of AtTR and the primers used in the DMS footprinting assay. (B) Primer extension results using total RNA extracted from *A. thaliana* cell culture. A time course of DMS treatment is indicated. Red dots denote accessible nucleotides. Molecular weight markers (nts) are shown. (C) The DMS-accessible nucleotides (red) are mapped on the AtTR structure.

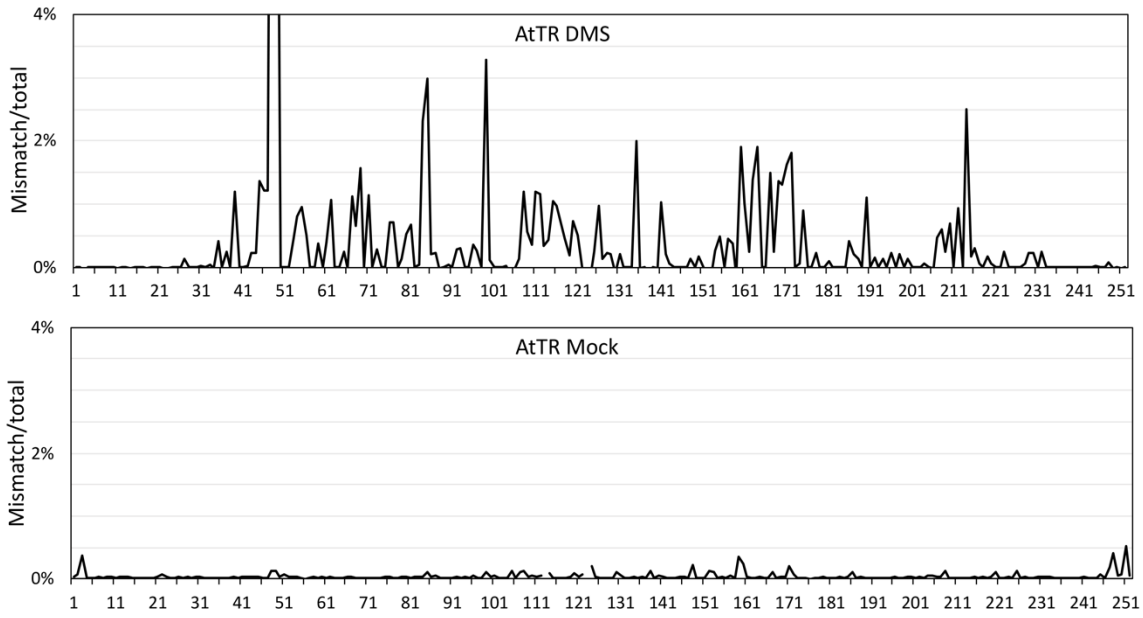


Figure S9. DMS MaPseq provides detailed information on accessible nucleotides in AtTR. Average mutation frequencies are plotted along AtTR sequences.

Table S1. Species with TR identified in this study.

Clade	Order	Species	Accession	Start Coordinates[a]	End Coordinates[b]	Source
Angiosperms	Apiales	*Daucus carota*	NC_030389.1	27,698,101	27,698,376	NCBI
	Asterales	*Chrysanthemum seticuspe*	BDUE01009703.1	11,137	11,419	NCBI
	Boraginales	*Echium plantagineum*	QFAX02000220.1	135,925	136,171	NCBI
	Brassicales	*Aethionema arabicum*	KE151693.1	19,486	19,752	NCBI
	Brassicales	*Arabidopsis halleri*	FJVB01000013.1	273,652	273,920	NCBI
	Brassicales	*Arabidopsis lyrata*	NW_003302193.1	6,235	6,501	NCBI
	Brassicales	*Arabis alpina*	LT669791.1	32,604,818	32,605,077	NCBI
	Brassicales	*Arabis montbretiana*	LNCH01009117.1	36,530	36,800	NCBI
	Brassicales	*Arabis nordmanniana*	LNCG01220153.1	3,675	3,942	NCBI
	Brassicales	*Barbarea vulgaris*	LXTM01001115.1	52,644	52,908	NCBI
	Brassicales	*Boechera stricta*	MLHT01000206.1	2,167,990	2,168,256	NCBI
	Brassicales	*Brassica cretica*	QGKV01138583.1	73	347	NCBI
	Brassicales	*Brassica juncea*	CM007199.1	42,541,815	42,542,082	NCBI
	Brassicales	*Brassica rapa*	NC_024798.1	13,226,512	13,226,780	NCBI
	Brassicales	*Capsella bursa-pastoris*	MPGU01000291.1	544,071	544,341	NCBI
	Brassicales	*Cardamine hirsuta*	Chr4	14,202,139	14,202,402	MPIPZ
	Brassicales	*Conringia planisiliqua*	FNXX01000004.1	7,410,188	7,410,452	NCBI
	Brassicales	*Crucihimalaya himalaica*	SMJT01000124.1	207,463	207,727	NCBI
	Brassicales	*Euclidium syriacum*	FPAK01000008.1	2,642,798	2,643,063	NCBI
	Brassicales	*Eutrema heterophyllum*	PKMM01021225.1	255,915	256,145	NCBI
	Brassicales	*Eutrema salsugineum*	NW_006256908.1	4,817,520	4,817,781	NCBI
	Brassicales	*Eutrema yunnanense*	PKML01061038.1	473	736	NCBI
	Brassicales	*Leavenworthia alabamica*	KE157026.1	94,536	94,815	NCBI
	Brassicales	*Raphanus raphanistrum*	JRQH01003943.1	2,932	3,186	NCBI
	Brassicales	*Raphanus sativus*	NW_017353142.1	35,244,840	35,245,094	NCBI
	Brassicales	*Schrenkiella parvula*	CM001190.1	12,955,768	12,956,035	NCBI
	Brassicales	*Sisymbrium irio*	KE156162.1	139,418	139,686	NCBI
	Brassicales	*Tarenaya hassleriana*	NW_010971389.1	564,830	565,090	NCBI
	Brassicales	*Thlaspi arvense*	AZNP01000142.1	120,620	120,884	NCBI
	Caryophyllales	*Beta vulgaris*	NC_025816.2	46,989,999	46,990,277	NCBI
	Cucurbitales	*Cucurbita argyrosperma*	SDJN01000158.1	222,899	223,147	NCBI
	Fabales	*Pisum sativum*	PUCA014342884.1	375	635	NCBI
	Fagales	*Casuarina equisetifolia*	RDRV01000354.1	115,161	115,414	NCBI
	Gentianales	*Coffea eugenoides*	NC_040043.1	6,019,019	6,019,286	NCBI
	Lamiales	*Olea europaea*	NW_019237129.1	278,546	278,801	NCBI
	Malpighiales	*Caryocar brasiliense*	STGP01026219.1	4,821	5,089	NCBI
	Malpighiales	*Manihot esculenta*	NC_035172.1	30,848,292	30,848,551	NCBI
	Malpighiales	*Populus simonii*	CM017472.2	14,153,485	14,153,758	NCBI
	Malpighiales	*Viola pubescens*	NBIL01136792.1	11,205	11,463	NCBI
	Malvales	*Aquilaria agallochum*	KK907007.1	4,840	5,116	NCBI
	Malvales	*Aquilaria sinensis*	SMDT01003036.1	616,167	616,432	NCBI
	Malvales	*Corchorus capsularis*	AWWV01006766.1	15,731	16,000	NCBI
	Malvales	*Corchorus olitorius*	AWUE01012270.1	7,870	8,137	NCBI
	Malvales	*Durio zibethinus*	NW_019167871.1	10,632,362	10,632,624	NCBI
	Malvales	*Gossypioides kirkii*	CM008983.1	32,191,550	32,191,784	NCBI
	Malvales	*Gossypium aroboreum*	NC_030666.1	87,580,015	87,580,260	NCBI

Angiosperms	Malvales	*Gossypium australe*	CM016621.1	70,003,453	70,003,727	NCBI
	Malvales	*Gossypium thurberi*	CM013381.1	23,878,808	23,879,056	NCBI
	Malvales	*Kokia drynaroides*	NTFQ01013625.1	69,702	69,937	NCBI
	Malvales	*Theobroma cacao*	NC_030859.1	13,601,850	13,602,113	NCBI
	Myrtales	*Eucalyptus camaldulensis*	BADO01007437.1	1,766	2,014	NCBI
	Myrtales	*Eucalyptus melliodora*	SISH01000046.1	4,217,044	4,217,293	NCBI
	Oxalidales	*Aristotelia chilensis*	VEXP01036680.1	842	1,093	NCBI
	Oxalidales	*Cephalotus follicularis*	BDDD01000524.1	142,301	142,563	NCBI
	Proteales	*Macadamia integrifolia*	UZVR01001767.1	83,061	83,327	NCBI
	Rosales	*Rosa chinensis*	NC_037093.1	60,849,275	60,849,535	NCBI
	Sapindales	*Atalantia buxifolia*	MKYR01004417.1	843,867	844,129	NCBI
	Sapindales	*Azadirachta indica*	AMWY02057456.1	1,105	1,362	NCBI
	Sapindales	*Cintrus hindsii*	QWBT01000927.1	5,076,792	5,077,050	NCBI
	Sapindales	*Citrus clementina*	NW_006261964.1	4,968,658	4,968,914	NCBI
	Sapindales	*Xanthoceras sorbifolium*	CM010616.1	13,811,758	13,812,007	NCBI
	Solanales	*Cuscuta australis*	NQVE01000092.1	783,927	784,177	NCBI
	Solanales	*Nicotiana rustica*	ML520654.1	26,960	27,208	NCBI
	Solanales	*Nicotiana tabacum*	NW_015926110.1	63,719	63,965	NCBI
	Solanales	*Solanum tuberosum*	NW_006239035.1	694,552	694,810	NCBI
	Zingiberales	*Musa balbisiana*	CM017189.1	27,011,643	27,011,918	NCBI
	Alismatales	*Posidonia oceanica*	GGFN01190223.1	5	250	NCBI*
	Arecales	*Calamus simplicifolius*	UESW01003909.1	1,490,526	1,490,801	NCBI
	Amborellales	*Amborella trichopoda*	NW_006494910.1	7,781,446	7,781,726	NCBI
	Magnoliales	*Liriodendron chinense*	PVNU02000262.1	764,788	765,066	NCBI
Gymnosperms	Gnetales	*Gnetum montanum*	scaffold866741	96,878	97,202	DRYAD
	Cupressales	*Sequoia sempervirens*	VDFB01200574.1	58,757	59,059	NCBI
	Ginkgoales	*Ginkgo biloba*	Chr9	251,572,510	251,572,855	GIGA
	Pinales	*Abies balsamea*	aalba5_s00030163	46,478	46,823	TG DB
	Pinales	*Larix sibirica*	NWUY0100044616.1	10,666	11,015	NCBI
	Pinales	*Picea abies*	CBVK0101923023.1	6,762	7,110	NCBI
	Pinales	*Picea glauca*	ALWZ04S1636083.1	4,036	4,382	NCBI
	Pinales	*Pinus lambertiana*	LMTP010003768.1	303,995	304,339	NCBI
	Pinales	*Pinus sylvestris*	contig_7214027	1,193	1,533	NCBI
	Pinales	*Pinus taeda*	APFE031443769.1	20,896	21,241	NCBI
	Pinales	*Pseudotsuga menziesii*	LPNX010568464.1	175,954	176,301	NCBI
Lycophytes	Isoetales	*Isoetes echinospora*	GGKY01093994.1	1,209	1,488	NCBI*
	Selaginellales	*Selaginella kraussiana*	LDJE01041645.1	1,146	1,441	NCBI
	Selaginellales	*Selaginella bryopteris*	GEMU01091170.1	1	305	NCBI*
	Selaginellales	*Selaginella tamariscina*	PUQB01000486.1	141,633	141,932	NCBI

[a]: 5' end predicted based on multiple sequence alignment with AtTR
[b]: 3' end of TR inferred based on the presence of a poly 'U' tract
NCBI: National center for Biotechnology Information – Genome Database, URL: www.ncbi.nlm.nih.gov/genome/
NCBI*: National center for Biotechnology Information – Transcriptome Shotgun Assembly Sequence Database, URL : www.ncbi.nlm.nih.gov/genbank/tsa/
MPIPG : Max Planck Institute for Plant Breeding Research – Genomic Resource, URL : chi.mpipz.mpg.de
DRYAD : Dryad digital repository, URL : datadryad.org
GIGA DB : GigaDB data repository, URL : gigadb.org
TGDB : TreeGenes database, URL : treegenesdb.org

Table S2. List of oligonucleotides used.

Primer name	Sequence	Purpose
AtTR-SHAPE-1	ATTTTTCGGGGGAAACGCCG	SHAPE primer extension
AtTR-SHAPE-2	CCTCCTGGTCTACTTTTTGGAGA	SHAPE primer extension
AtTR-SHAPE-3	AAATATTTGGGGGTGGGAGGG	SHAPE primer extension
AtTR-DMS footprinting-1	ATTTTTCGGGGGAAACGCCG	DMS footprinting
AtTR-DMS footprinting-2	CCTCCTGGTCTACTTTTTGGAGA	DMS footprinting
AtTR-DMS footprinting-3	GAGGGTAAGGCGAGGAAACG	DMS footprinting
AtTR-MaPseq-F	AAGGGGTGTGGGAACCTAGGAG	DMS MaPseq
AtTR-MaPseq-R	GAGGGTAAGGCGAGGAAACG	DMS MaPseq
ACT2-MaPseq-F	GCTGGAATCCACGAGACAACCTAT	DMS MaPseq
ACT2-MaPseq-R	GGAGATCCACATCTGCTGGAATG	DMS MaPseq
AtTR-Northern-1	TGGGAGGGTAAGGCGAGGAAACGGTTAACCGCAGACACTAATCTAGTTTC	Northern Blot
AtTR-Northern-2	TCAGGGTTTAGCAATCAATAAGCAGACTCATCTCCTAGGTTCCCACACCCCTT	Northern Blot
AtTR-Northern-3	AACCAAACCCAACCTCCTGGTCTACTTTTTGGA	Northern Blot
18S rRNA-RT-F	TAACTCGACGGATCGCATGG	18s rRNA RT PCR
18S rRNA-RT-R	CGACCCATCCCAAGGTTCAA	18s rRNA RT PCR
AtTR-F1	TGCTAAACCCTGAACCCTCTC	AtTR qPCR
AtTR-R1	AAATATTTGGGGGTGGGAGGGT	AtTR qPCR
AtTR-PS1-F	ATTGATAGTTAACATGAGAGGGTT	AtTR CRISPR construct cloning
AtTR-PS1-R	AAACAACCCTCTCATGTTAACTAT	AtTR CRISPR construct cloning
AtTR-PS2-F	ATTGTGCTTATTGATTGCTAAACC	AtTR CRISPR construct cloning
AtTR-PS2-R	AAACGGTTTAGCAATCAATAAGCA	AtTR CRISPR construct cloning
AtTR-PS3-F	ATTGAACTATGGGAATTAATTACT	AtTR CRISPR construct cloning
AtTR-PS3-R	AAACAGTAATTAATTCCCATAGTT	AtTR CRISPR construct cloning
T7-AtTR-F	GAAATTAATACGACTCACTATAAAGGGGTGTGGGAACCTAGGAG	In vitro transcription
Oligo dA anchor primer	GACCACGCGTATCGATGTCGACAAAAAAAAAAAAAAAA	5' RACE
AtTR-R2	GAGCCTCTTATGTAGCCATCG	5' RACE
AtTR-R3	CGGCCTAAGACCCCCAGTAA	5' RACE
AtTR-R4	GAGAGGGTTCAGGGTTTAGCA	5' RACE
AtTR-F2	TGCTAAACCCTGAACCCTCTC	3' RACE
AtTR-F3	GGTGTGGGAACCTAGGAGATG	3' RACE
AtTR-F4	TGGTTCGTAGGTGGTTCTGTT	3' RACE

CHAPTER III

THE PLANT TELOMERASE RETAINS VERTEBRATE-LIKE ASSOCIATED PROTEINS VIA A PLANT-SPECIFIC ASSEMBLY

Summary

Telomerase is a ribonucleoprotein (RNP) complex that compensates for the loss of telomeric DNA during chromosome replication using the catalytic telomerase reverse transcriptase (TERT) and its intrinsic templating telomerase RNA (TR). The recent discovery of the *bona fide* TR from the plant kingdom represents an important advance as the plant TR exhibits both conserved and unique secondary structure elements that provide novel insight into TR evolution. Nevertheless, the protein composition of plant telomerase remains mostly unclear. Here we use a multidisciplinary approach to examine two accessory proteins of *Arabidopsis thaliana* telomerase that promote enzyme activity *in vivo*, AtPOT1a and dyskerin. We report that AtPOT1a associates with telomerase via direct interaction with TERT. Loss of AtPOT1a does not impact AtTR stability or structure, implying that AtPOT1a is not critical for assembly of the core complex. However, we demonstrate the function of AtPOT1a in telomerase recruitment on telomeres, which is functionally equivalent to the human POT1-TPP1 heterodimer. In addition, we show that dyskerin directly binds AtTR via a plant-specific RNA element that is required to stimulate telomerase activity *in vitro* and *in vivo*. We present an 8.6-Å cryo-EM structure of the AtTR-dyskerin-GAL1 complex showing that dyskerin assembles via a novel interaction with a unique three-way junction element in AtTR, but forms a dimer with identical arrangement and architecture as dyskerin in the human telomerase RNP. Thus, the plant telomerase with its ciliate-like Pol III transcribed TR and vertebrate-like accessory protein, dyskerin, provides a unique perspective for telomerase RNP evolution.

Introduction

Telomerase is a ribonucleoprotein (RNP) complex that maintains telomere integrity, and is thus a key invention that helped enable the transition from prokaryotic circular chromosomes to eukaryotic linear chromosomes (6). The essential core of telomerase consists of a catalytic subunit telomerase reverse transcriptase (TERT) and a long noncoding RNA (TR) (9). TR serves as a template for synthesis of telomere repeats by TERT. TR also encodes a structural scaffold for binding accessory proteins that facilitate TR biogenesis, RNP assembly, engagement with the chromosome terminus, and regulation of telomerase enzymatic activity (85).

Studies of TR secondary structure revealed two essential domains for telomerase catalysis (86, 97, 114). The first is a template-pseudoknot domain bearing a single-stranded template, a double-stranded template boundary element (TBE) and a pseudoknot (PK) motif (104-106). The second critical domain is a 'stem-terminus' element (STE) near the 3' end termed helix IV in ciliate TR or CR4/5 in vertebrates (111, 112, 115). The template-PK and STE are sufficient to reconstitute telomerase activity in trans (97, 112, 114). In addition to these two essential domains, the remainder of the TR has diverse structural motifs that act as a scaffold to connect species-specific accessory proteins (18). For example, in Tetrahymena, the La-related protein P65 recognizes the 3' poly-U tail of RNA Polymerase III (Pol III) transcribed TR and bends the RNA to facilitate telomerase RNP assembly (116-118). In contrast, fungal TR is transcribed by RNA Polymerase II (Pol II) and requires components of the canonical snRNA biogenesis pathway for 3' end maturation, resulting in RNP assembly with Sm and Lsm proteins (134, 135). Interestingly, in fission yeast another La-related protein, Pof8, associates with active telomerase via a non-canonical interaction with this Pol II transcript (127-129). Vertebrate telomerase RNP biogenesis proceeds via yet another maturation pathway devised for small nucleolar RNAs (snoRNA) (87).

In vertebrate TR, the protein components of the H/ACA snoRNP (dyskerin, NOP10, NHP2, and GAR1) recognize a conserved H/ACA motif that protects the 3' end of the mature TR from exonuclease degradation (140, 143).

The emergence of cryo-EM technology has facilitated dramatic new insights into the telomerase RNP. A 9-Å resolution structure of Tetrahymena telomerase unexpectedly revealed novel accessory proteins, including a large RPA-like heterotrimer analogous to the telomere replication complex, CTC1/STN1/TEN1 (CST) (80). Subsequently, a 4.8-Å structure of Tetrahymena telomerase bound to telomeric DNA provided novel information on regulation of the catalytic cycle (79). Most recently, a cryo-EM structure of substrate-loaded human telomerase holoenzyme roundly rejected the prevailing model that telomerase was a dimer, and instead showed a bilobed structure consisting of a catalytic core (7.7-Å resolution) on one side, and on the other side a H/ACA RNP (8.2-Å resolution) containing two sets of dyskerin-NOP10-NHP2-GAR1 tetramer (72). These studies confirmed that TR is a scaffold for essential accessory proteins (110).

Analysis of telomerase in the plant kingdom has lagged relative to other organisms. However, new studies using an unbiased approach for active telomerase purification uncovered the *bona fide* TR from *Arabidopsis thaliana* and enabled identification of 85 TR homologs across the plant kingdom (99). Fajkus et al. independently uncovered the same TR using a bioinformatic approach (206). A robust structural model for plant TR was created and revealed a conserved PK domain with a ciliate-like unstable stem and a vertebrate-like extensive single-stranded loop. Thus, the chimeric plant PK serves as a kind of evolutionary bridge for the structural transition of the highly divergent TR. In conjunction with the highly informative PK architecture of TR, additional plant-specific RNA elements were distinguished, including a 5'-3' long-range interaction, P1a

stem, and a flexible linker connecting P1b to P1c (99). This observation suggests that plant TR may assemble with other proteins beside TERT using these unique RNA elements as a scaffold.

In support of this conclusion, size exclusion chromatography of *A. thaliana* telomerase suggests that the core RNP is approximately 300 kDa in size (99), indicating that besides 130 kDa AtTERT and the 85 kDa AtTR, other accessory proteins remain to be identified. POT1 (Protection of Telomeres) is one of the most conserved constituents of the telomere complex (225). All POT1 proteins harbor oligosaccharide/oligonucleotide-binding fold (OB-fold) domains that facilitate interaction with single-stranded telomeric DNA. In vertebrates, POT1 is a stable component of telomeres that associates with five additional proteins to form a highly dynamic, higher-order assembly termed shelterin (226). Shelterin controls telomerase access to the chromosome terminus and prevents the ends from engaging in illegitimate DNA repair (227). POT1 together with its binding partner TPP1 collaboratively engage telomerase through physical contacts with the TERT protein. POT1-TPP1 binding to the DNA substrate reduces primer dissociation rate and enhances the translocation efficiency of telomerase to promote telomerase repeat addition processivity (RAP) (53, 54, 176). There are two POT1 orthologs in *A. thaliana* (203). AtPOT1a associates with the enzymatically active telomerase RNP and is required for telomere maintenance in vivo (202, 203). Interestingly, unlike human POT1, AtPOT1a is not a stable component of telomeric chromatin and instead its interaction with chromosome ends peaks in S phase (202). This observation is consistent with the proposal that AtPOT1a is an accessory factor for telomerase rather than a shelterin component. Nevertheless, like the POT1-TPP1 complex, AtPOT1a can bind telomeric DNA and stimulate telomerase RAP *in vitro* (204). Mechanistic details for how AtPOT1a engages telomerase and promotes RAP in plants are unknown.

Another potential subunit of plant telomerase is dyskerin (*NAP57*). Previous studies showed that dyskerin is physically associated with the telomerase RNP in both eudicots *A. thaliana* and monocots *Allium cepa* (205, 206). Arabidopsis plants bearing a homozygous null mutation in *AtNAP57* are inviable. However, transgenic plants carrying a single wild-type *AtNAP57* allele with a second allele bearing a T66A mutation are viable yet exhibit decreased telomerase activity and insufficient telomere maintenance (205). This finding argues that dyskerin is critical for telomerase function in plants. Notably, in contrast to vertebrate TR molecules, plant TR does not harbor a canonical H/ACA motif to anchor dyskerin (99). Thus, the interaction between dyskerin and plant telomerase may be distinct from the vertebrate enzyme.

Here we examine the interactions of AtPOT1a and dyskerin with Arabidopsis telomerase RNP using a combined approach of biochemistry and structural biology. We show that AtPOT1a physically interacts with TERT in a TR-independent manner. Further, we report that neither TR biogenesis nor structure are impacted by the loss of AtPOT1a. In contrast, lacking AtPOT1a reduced telomerase recruitment on telomeres, which exhibits the functional similarity between AtPOT1a and human POT1-TPP1 dimer. We also observed a direct interaction between dyskerin and a plant-specific three-way junction within AtTR; dyskerin association with this element stimulates telomerase repeat addition processivity *in vitro*. Moreover, deletion of this region of AtTR significantly reduces TR stability and telomerase activity *in vivo*. We further report an 8.6-Å cryo-EM structure that provides new insight into dyskerin-AtTR interaction and demonstrates the difference and similarity of telomerase RNP assembly between plants and vertebrates. These findings argue that the plant telomerase represents a chimeric RNP, harboring a ciliate-like Pol III transcribed TR and a vertebrate-like accessory protein, dyskerin. Thus, the Arabidopsis telomerase serves as a new platform for exploring telomerase evolution.

Results

AtPOT1a engages telomerase via an AtTR independent mechanism

With the recent discovery of Arabidopsis TR (99), we set out to test whether AtPOT1a and AtTR assemble into the same RNP complex using RNA immunoprecipitation (RIP) (Figure. 23a, left). As expected, qPCR revealed that TERT and AtPOT1a IP significantly enriched AtTR, but not GAPDH mRNA, relative to the non-specific control GFP IP. Moreover, telomerase activity was enriched in both the AtPOT1a and TERT IP samples, verifying the association between AtPOT1a and active telomerase (Figure. 23a, right). To examine the function of AtPOT1a in AtTR biogenesis and stability, we measured AtTR abundance in a *pot1a* homozygous mutant (Figure. 23b). There was no obvious difference in AtTR abundance in *pot1a* versus wild-type Arabidopsis extracts. These findings argue that AtPOT1a is not required for AtTR biogenesis or its assembly with TERT.

While TR acts as a framework to connect most of the accessory factors that constitute the telomerase RNP, some auxiliary proteins directly interact with TERT including the POT1-TPP1 heterodimer in human and the CST (CTC1-STN1-TEN1)-like P75-P45-P19 complex in Tetrahymena (53, 80). Therefore, co-IP experiments were performed with recombinant FLAG-tagged TERT and T7-tagged AtPOT1a expressed in a rabbit reticulocyte system in the presence or absence of AtTR to test if these proteins interact with each other *in vitro*, and if this interaction is dependent on AtTR. Pull-down using FLAG antibody followed by detection of AtPOT1a indicated that AtPOT1a physically interacts with TERT (Figure. 23c). Notably, the same amount of AtPOT1a copurified in the presence of AtTR, in the presence of non-specific RNA, or in the absence of any RNA. Therefore, we conclude that AtPOT1a interacts with TERT and this interaction is not dependent on AtTR.

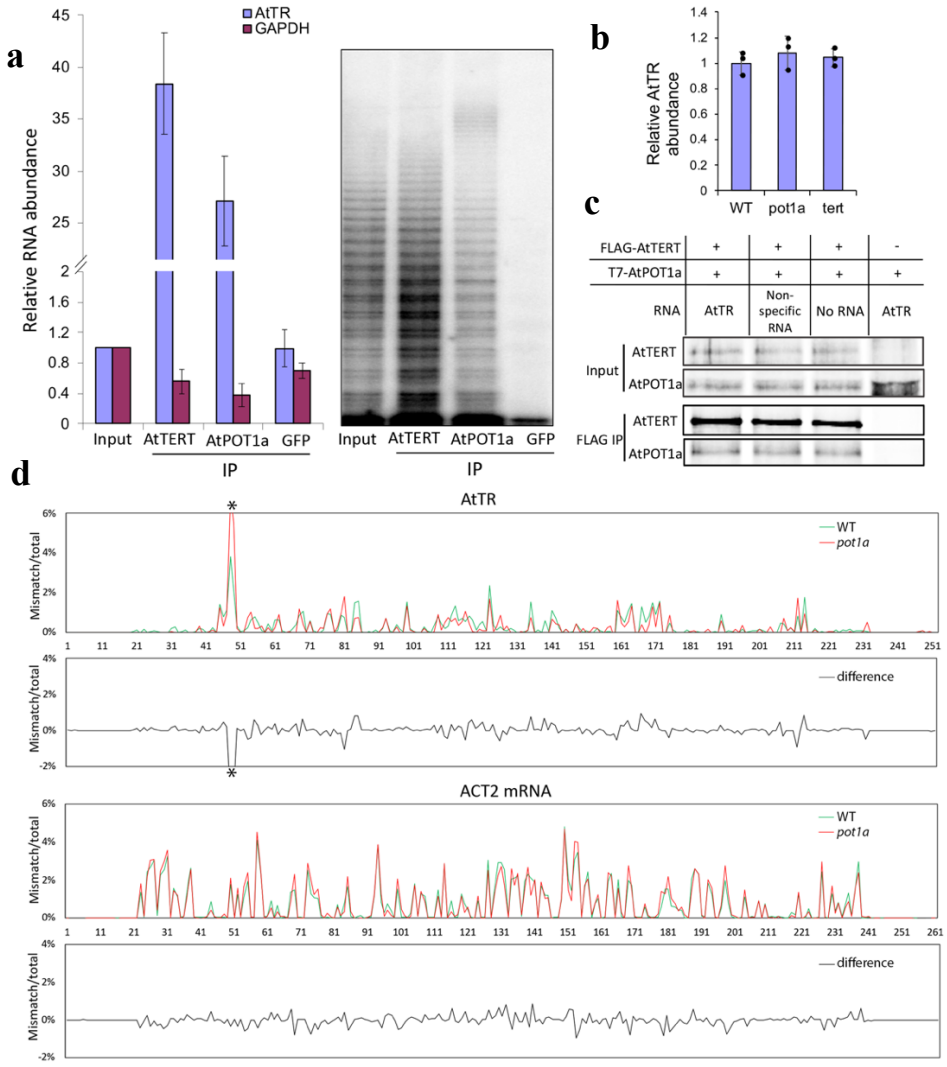


Figure 23. AtPOT1a does not engage Arabidopsis telomerase via interaction with AtTR.

a, qPCR (left) and TRAP (right) analyses of immunoprecipitated samples were conducted to measure the copurification of AtTR (left), GAPDH mRNA (left), and active telomerase (right). Anti-AtTERT and Anti-GFP IPs serve as positive and negative controls respectively. **b**, qPCR results show abundance of AtTR in WT, *tert*, and *pot1a* mutants. **c**, in vitro co-IP experiments were performed in a coupled transcription/translation system in the presence or absence of different RNA molecules. Co-purification of AtPOT1a was measured after Anti-FLAG IP. **d**, in vivo chemical probing of RNA secondary structure by targeting specific DMS-MaPseq from wild type (WT) Arabidopsis and *pot1a* mutants. Average mutation frequencies are plotted along AtTR and ACT2 mRNA sequences separately. Asterisk labels the template region of AtTR with increased flexibility in the *pot1a* mutant.

To investigate whether AtPOT1a impacts AtTR structure and nucleotides accessibility *in vivo*, we employed DMS-MaPseq assay in wild-type plants and plants lacking AtPOT1a (Figure. 23d and supplementary Figure. S10). The DMS-induced mismatch rate correlates with the respective flexibility and accessibility of A and C residues (218). We observed no significant difference in nucleotides accessibility between wild type and *pot1a* mutants for the majority of AtTR with the notable exception of nucleotides 48 and 49 which are embedded in the telomere template region (Figure. 23d, asterisk points on nucleotides 48 and 49). Given the ability of AtPOT1a to bind telomeric DNA and to stimulate telomerase repeat addition processivity (RAP) (204), we hypothesize that the increased TR template accessibility in *pot1a* mutants reflects inefficient recruitment of telomerase to telomeric DNA, leading to more telomerase particles unbound to substrate. Taken together, our data support the conclusion that AtPOT1a physically associates with telomerase in a AtTR-independent manner, and does not regulate AtTR abundance and structure. These results further suggest that the main function of AtPOT1a is to stimulate telomerase processivity and recruitment by enhancing interaction with telomeric DNA perhaps by engaging TERT.

Dyskerin tightly binds AtTR and promotes telomerase processivity in vitro

Previous biochemical and genetic experiments indicate that dyskerin associates with the plant active telomerase RNP (205, 206). To examine whether the dyskerin-telomerase association is mediated by AtTR, we designed two independent approaches. First, an *in vivo* co-IP experiment using anti-TERT antibody was conducted in the presence or absence of AtTR (Figure. 24a). AtTERT and PEPC, the positive and negative controls, respectively were detected by western blotting. Notably, dyskerin copurified with TERT in extracts from wild-type plants, but not from

mutants that lack AtTR. This result argues that AtTR is necessary to bridge the dyskerin-AtTERT interaction.

As an alternate strategy, we performed a pull-down experiment using two antisense locked nucleic acid (LNA) oligonucleotides to capture proteins that directly bind AtTR (Figure. 24b and c). The oligonucleotides were designed to target accessible regions within AtTR (99) (supplementary Figure. S11a). qPCR analysis of the pull-down samples showed a dramatic enrichment of AtTR, but not ACT2 mRNA or GAPDH mRNA (Figure. 24c and supplementary Figure. 11b). This observation confirms the stringent specificity of antisense LNA oligonucleotides. Western blotting indicated that both dyskerin and AtTERT were enriched in the AtTR pull-down (Figure. 24b). These experiments support the conclusion that dyskerin associates with Arabidopsis telomerase via a direct interaction with AtTR.

We next examined the dyskerin-AtTR interaction *in vitro* using recombinant dyskerin protein. When expressed in *Escherichia coli*, full-length dyskerin formed aggregates unless maintained in a low concentration (supplementary Figure. S12a). To increase soluble protein yield, we removed the C-terminal disordered region (residues 440-565), which is predicted as a nuclear location signal (NLS) (supplementary Figure. 12c). Dyskerin$_{\Delta C}$ vastly enhanced the protein solubility (Extended Data Figure. 3b). Electrophoretic mobility shift assay (EMSA) was used to monitor the dyskerin$_{\Delta C}$-AtTR interaction (Figure. 24d). A reliable, single-band higher molecular weight signal was observed after titration of 12.5 nM dyskerin$_{\Delta C}$ protein with radiolabeled AtTR, indicative of a stable dyskerin-AtTR complex. A 100-fold excess of specific non-radiolabeled competitor, but not non-specific yeast tRNA, resolved the high molecular weight signal into free AtTR molecules (Figure. 24d, lane 9-10). Quantification of EMSA with plots fit to the Hill

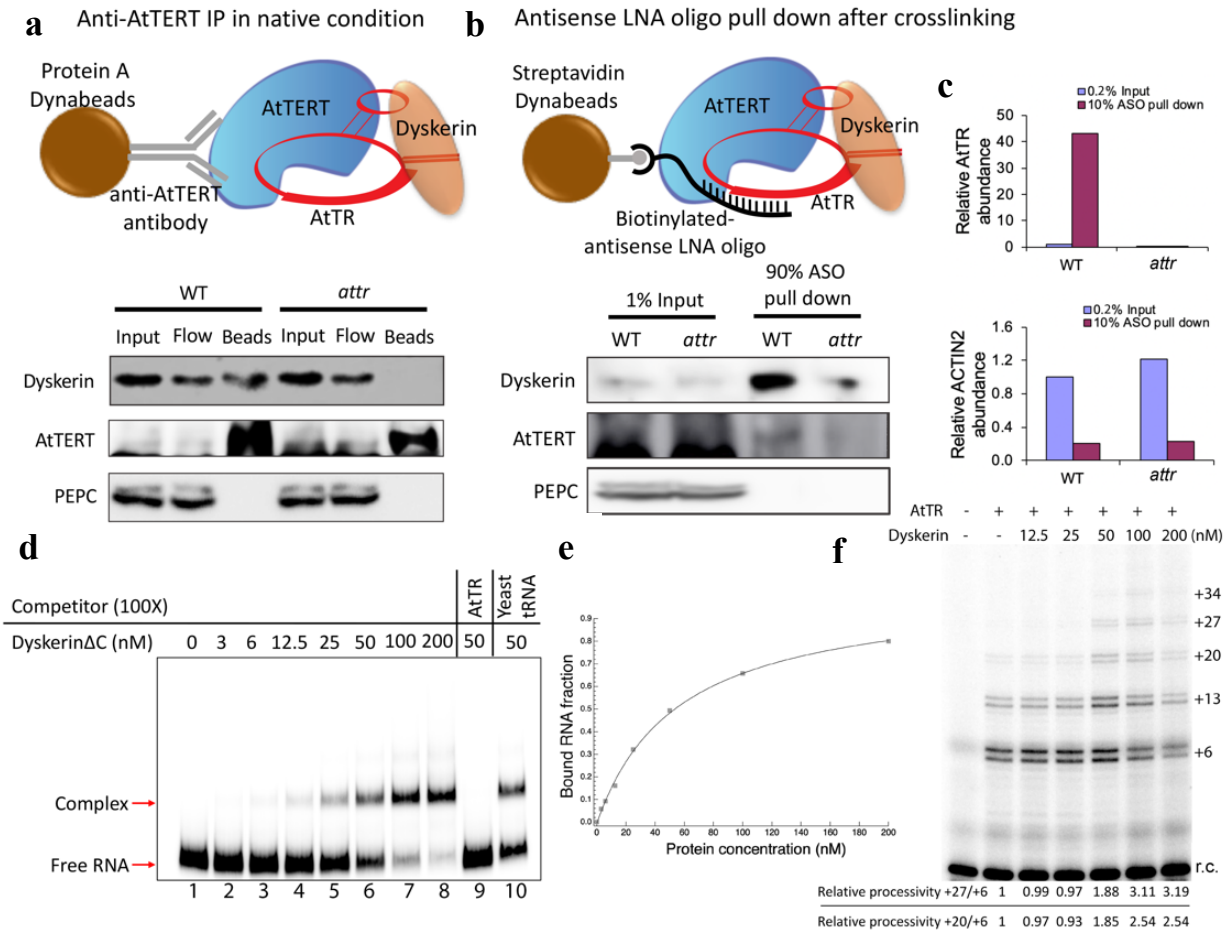


Figure 24. Dyskerin physically associates with AtTR and stimulates telomerase processivity *in vitro*.

a, Top: schematic representation of the in vivo co-IP experiment. Bottom: western blot analysis of dyskerin, AtTERT and PEPC in corresponding fractions. **b**, Top: schematic representation of the antisense LNA oligo pull down experiment. Bottom: western blot analysis of dyskerin, AtTERT and PEPC in corresponding fractions. **c**, qPCR analysis of AtTR and ACT2 mRNA abundance in the antisense LNA oligo pull down. **d**, EMSA to examine the dyskerinΔC-AtTR interaction in vitro. 100X specific (AtTR) or non-specific (yeast tRNA) competitors was provided to test binding specificity. **e**, quantification of EMSA results plotted and fitted to the Hill Equation with the Hill coefficient as 1. The bound RNA fraction was calculated using the ratio of RNP signal intensity to the total signal intensity. **f**, in vitro telomerase reconstitution assay with a titration of dyskerin$_{FL}$ protein. Relative telomerase processivity was calculated and indicated at the bottom of the panel. A radiolabeled 18-mer recovery control (r.c.) was added before product purification and precipitation. This experiment was independently reproduced twice with identical results.

equation revealed a binding affinity (K_d) of 55.97 (±4.58) nM (Figure. 24e). These results indicate a specific, high-affinity interaction between Dyskerin$_{\Delta C}$ and AtTR *in vitro*.

To investigate the function of dyskerin association with AtTR, we conducted *in vitro* telomerase reconstitution using full-length dyskerin (Dyskerin$_{FL}$) (Figure. 24f). Despite aggregation issues, we were reluctant to use Dyskerin$_{\Delta C}$ because of the potential loss of function. Therefore, different concentrations of Dyskerin$_{FL}$ were added to a reconstitution reaction containing rabbit reticulocyte lysate (RRL)-expressed 3XFLAG-AtTERT and *in vitro* folded AtTR. Using the direct telomerase extension assay (228), we observed a substantial increase in signal intensity for the longer extension products (+20, +27, and +34) relative to shorter products (+6), indicating that telomerase repeat addition processivity (RAP) is stimulated by the addition of dyskerin (Figure. 24f). While the overall activity of the reaction containing 200 nM dyskerin$_{FL}$ was diminished, we suspect as a consequence of protein aggregation, the extent of RAP was proportional to the amount of dyskerin (Figure. 24f, lane 7). Altogether, these findings indicate that dyskerin physically interacts with AtTR within *A. thaliana* telomerase and that this interaction promotes telomerase RAP.

A plant-specific long-range interaction P1a stem is required for dyskerin interaction with AtTR.

Using competitive EMSA with recombinant Dyskerin$_{\Delta C}$ protein, we screened 15 AtTR truncation constructs to map the RNA element responsible for dyskerin association (Figure 25). The truncations were designed by deleting different RNA motifs identified from our secondary structure model (Figure. 25a). We observed that removing the P1a stem abolished Dyskerin$_{\Delta C}$ binding (construct 1), while truncation of the non-conserved distal stem-loop of STE (construct 2), the single-stranded linker connecting P1b to P1c (construct 3), the template region and stem P2.1

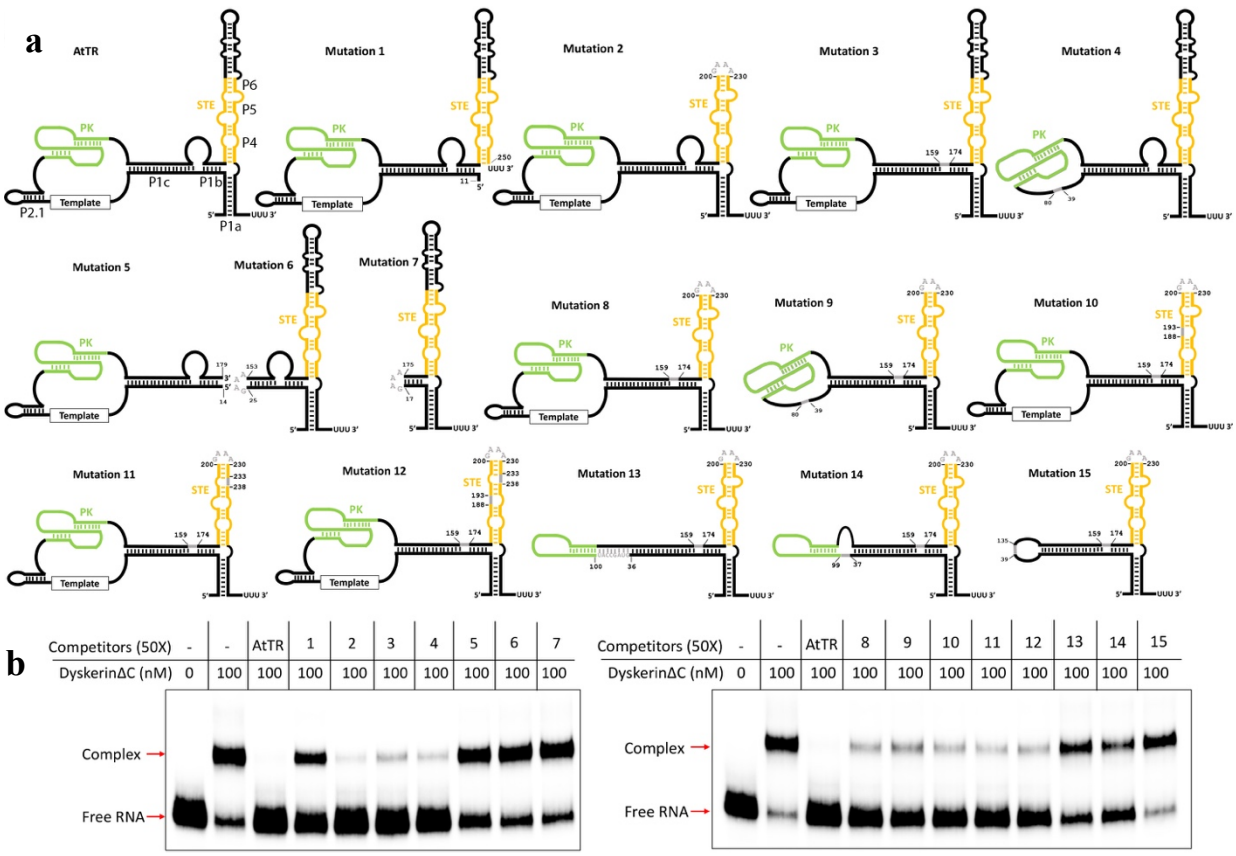


Figure 25. Mapping the dyskerin binding site on AtTR.

a, Truncation constructs of AtTR. **b**, Competitive EMSA performed with 50X excess non-radiolabeled AtTR or individual truncations. Mutation constructs 2, 3, 4, 8, 9, 10, 11, and 12 show strong competition; constructs 13 and 14 show medium competition; constructs 5, 6, 7, and 15 show weak competition or no competition.

(construct 4), and two conserved single-stranded loops within P4 and P5 (constructs 10 and 11) did not impede the interaction (Figure. 25b).

P1a is a plant-specific RNA element formed by a long-range interaction of the 5' and 3' ends of AtTR. Connecting P1a to P1b and P4-P5-P6 generates a unique three-way junction (TWJ) structure. In construct 5, we removed P1a together with P4-P5-P6 to completely disturb the TWJ. Consistent with construct 1, construct 5 did not compete with full-length AtTR for Dyskerin$_{\Delta C}$ binding. We also observed that reactions with constructs 13, 14, and 15 displayed intermediate competition, consistent with the importance of maintaining a complete PK structure for dyskerin association. However, an important caveat of this interpretation is that we introduced multiple modifications in these constructs, which could impact the overall structure of AtTR. Nonetheless, our results indicate that a plant-specific TWJ including the unique P1a stem is essential for dyskerin interaction with AtTR.

The P1a stem is essential for dyskerin-mediated stimulation of telomerase processivity in vitro and AtTR stability in vivo

In vertebrates and yeast, dyskerin associates with NOP10, GAL1, Nhp2 and a target RNA subunit to assemble into a H/ACA RNP complex (229, 230). Since our initial experiments employed only the dyskerin subunit, the absence of interaction partners could explain the low solubility of recombinant dyskerin protein. To obtain a more stable complex, we co-expressed three protein subunits of the plant H/ACA complex including Dyskerin$_{\Delta C}$, NOP10, and a truncation of GAL1 (residues 53-145; GAL1$_{\Delta}$) in *E. coli*. Consistent with the reconstitution of yeast H/ACA complex (230), co-expression of NOP10 and GAL1$_{\Delta}$ with dyskerin$_{\Delta C}$ significantly enhanced complex stability and solubility.

To verify the role of dyskerin in stimulating telomerase activity, *in vitro* telomerase reconstitution was performed with the addition of recombinant Dyskerin$_{\Delta C}$-NOP10-GAL1$_{\Delta}$ heterotrimer (Figure. 26a and b). As expected, longer extension products (+20, +27, and +34) were more abundant indicating stimulation of RAP by Dyskerin$_{\Delta C}$-NOP10-GAL1$_{\Delta}$ (Figure. 26b, lane 2-4). In contrast, substituting full-length AtTR with truncation construct 1, which lacks P1a and abolishes dyskerin interaction, did not result in increased RAP (Figure. 26b, lane 5-7). This result supports the importance of the plant-specific element P1a in the dyskerin-mediated stimulation of telomerase RAP *in vitro*.

To test the function of the P1a stem *in vivo*, we conducted a genetic complementation assay. For these experiments, we transformed *attr* null mutants with constructs containing U6 promoter-driven AtTR with either wild-type sequence (U6::AtTR) or one of three P1a variants: a non-complementary strand on the 5' side of P1a (U6::m16), a non-complementary strand on the 3' side of P1a (U6::m17) or a compensatory mutation with identical structure to wild-type AtTR (U6::m18) (Figure. 26c). Although expression of all four constructs was driven by the same U6 promoter, perturbation of P1a dramatically impacted RNA stability (Figure 26d). U6::m16 and U6::m17 accumulated much less AtTR than wild-type plants, while U6::m18 restored to the wild-type level (Figure 26d). We performed the TRAP assay to measure telomerase activity in the corresponding transgenic lines. Consistent with their AtTR levels, U6::m16 and U6::17, but not U6::m18, displayed a significant reduction in telomerase activity compared to wild type (Figure 26e). Altogether these findings indicate that preserving the plant-specific P1a is not only critical for dyskerin-AtTR interaction *in vitro*, but disruption of P1a severely compromises AtTR stability and telomerase activity *in vivo*.

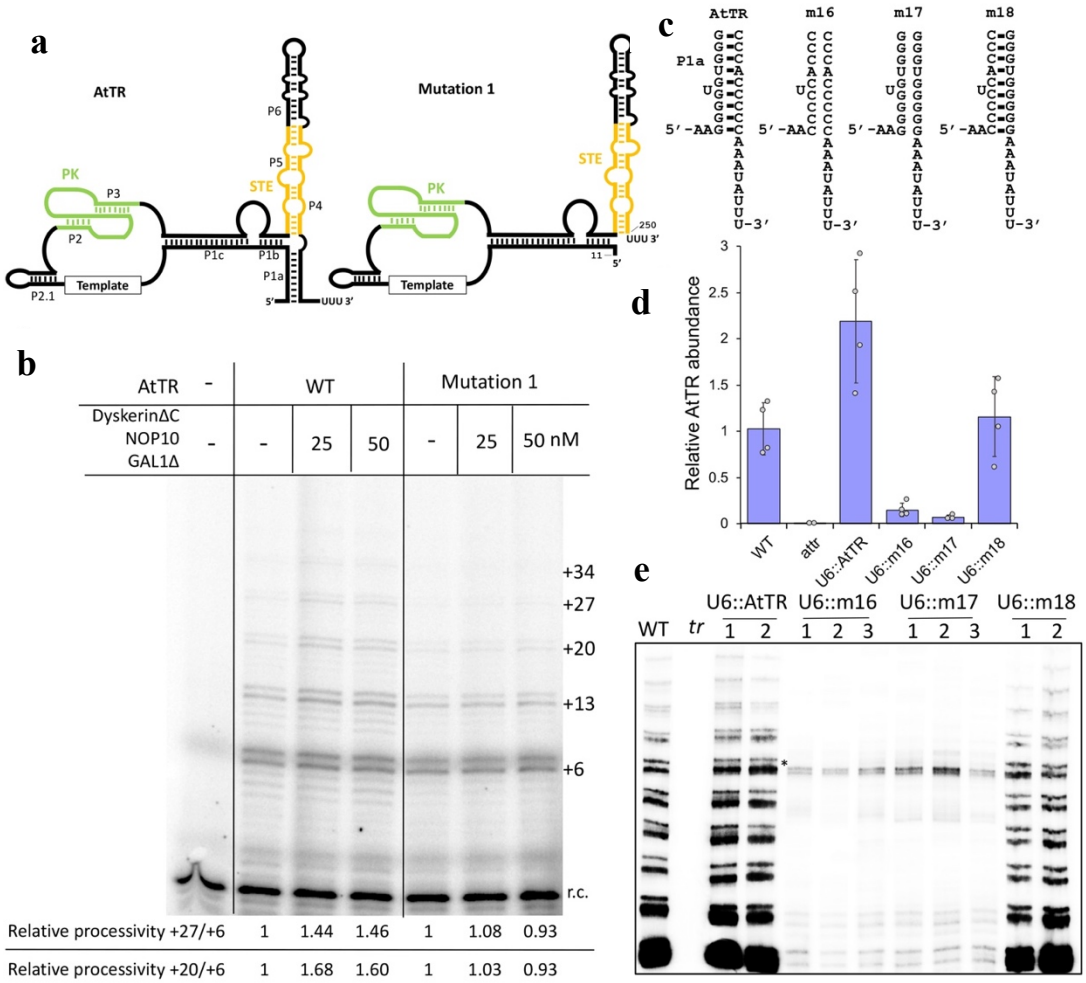


Figure 26. P1a is required for dyskerin stimulation of telomerase processivity *in vitro* and AtTR stability *in vivo*. **a**, schematic representation of the of the WT AtTR and mutation construct 1 secondary structures. **b**, *in vitro* telomerase reconstitution assay with the titration of DyskerinΔC-NOP10-GAL1Δ heterotrimer. WT AtTR or mutation 1 was used independently as TR in this assay. Relative telomerase processivity was calculated and indicated at the bottom of the panel. A radiolabeled 18-mer recovery control (r.c.) was added before product purification and precipitation. **c**, schematic representation of the AtTR and AtTR variants used for genetic complementation. **d**, qPCR analysis of AtTR abundance in transgenic plants expressing WT AtTR or AtTR variants. **e**, TRAP assay to measure telomerase activity in transgenic plants.

Cryo-EM structure of dyskerin-GAL1-AtTR complex

The cryo-EM structure of human telomerase holoenzyme was highly informative and revealed two copies of H/ACA RNP components (dyskerin-NOP10-GAL1-NHP2) located on the H/ACA motif within hTR (72). This structure indicated that dyskerin dimerization and interactions between dyskerin and hTR were essential for assembling the entire holoenzyme complex. To examine the binding mechanism of dyskerin interaction with *A. thaliana* TR, we co-expressed Dyskerin$_{\Delta C}$-NOP10-GAL1$_{\Delta}$ and assembled them with *in vitro*-folded, full-length AtTR (Extended Data Figure. 4a). After assembly, the RNP was further purified by size-exclusion chromatography to remove unassembled subunits (supplementary Figure. S13b and c) and cryo-EM imaging was performed.

From 58,138 particle images, the three-dimensional asymmetry refinement yielded a density map of the AtTR-dyskerin-NOP10-GAL1 RNP complex at 8.6-Å resolution. At this resolution, we could distinguish protein α-helices, β-barrels and double-stranded RNA, and fit this information into homology models of the corresponding proteins and RNA densities. Similar to the human telomerase cryo-EM structure (72), our data indicated an RNP complex is formed by two sets of the H/ACA RNP components containing one molecule of dyskerin and GAR1 in each complex (Figure. 27). The density of NOP10 is unclear, perhaps because this molecule is flexible. From the overall cryo-EM density, we distinguished two clear RNA helical stems that associate with two respective dyskerin protein molecules and an additional RNA density inserted into the interface of dyskerin proteins. Notably, the *A. thaliana* dyskerin protein dimerizes in an architecture identical to the cryo-EM structure of human telomerase holoenzyme with RNA stems located at similar positions (Figure. 27), which indicates the conserved structural basis of dyskerin dimerizing and interacting with TR between Arabidopsis and human telomerase RNP.

Based on our cryo-EM map, we propose a model for AtTR binding with dyskerin. Because the gel filtration chromatography of our recombinant complex enriched at a size of 170 kDa, we predict that the two sets of dyskerin complexes anchor on a single AtTR molecule by associating with two separate RNA helical stems. The cryo-EM density map supports the conclusion that the P1b-P1c stem and P4-P5-P6 stem derived from the TWJ of AtTR serve as the corresponding 5' and 3' hairpin stems responsible for dyskerin binding (Figure. 27), which are the structural equivalences of P4 stem and P7-P8 stem, respectively, in the cryo-EM structure of human telomerase. Because we did not include TERT protein in our recombinant RNP complex, the T-PK domain of AtTR was unbound. Therefore, it is not surprising that we were unable to unambiguously identify density of T-PK in our map due to high flexibility. However, what is striking is that the cryo-EM density map suggests that the 5'-3' long-range interaction P1a stem interacts with both dyskerin proteins by inserting to the interface of the dyskerin dimer (Figure. 27). In this model, the position of the AtTR 3' end is different from the canonical model proposed from the archaea and the human H/ACA RNP (72, 187). Further experiments will be required to confirm the model of the AtTR binding. Nevertheless, our cryo-EM structure of recombinant AtTR-dyskerin-GAL1 indicates that a dimerized dyskerin complex interacts with the plant-specific TWJ of AtTR.

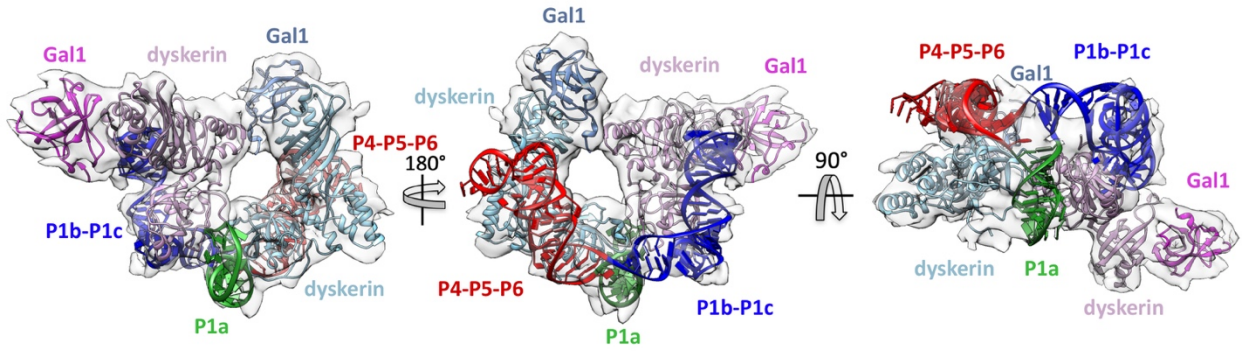


Figure 27. Cryo-EM structure of recombinant AtTR-dyskerin-Gal1 complex.

Cryo-EM density was modeled using the crystal structure of the Shq1-Cbf5-Nop10-Gar1 complex obtained from *Saccharomyces cerevisiae*. RNA fragments were inserted into unoccupied density. The two dyskerin proteins are in cyan and pink. The homology built GAR1 proteins are in blue and magenta. The AtTR is colored by its fragments (P1a: green, P1b-P1c: blue, and P4-P5-P6: red).

Discussion

Telomerase is a specialized RNP complex predominantly designated for maintaining telomere length homeostasis. While the telomerase catalytic component TERT is conserved among eukaryotes, the TR subunit and accessory proteins have diverged significantly during evolution (18, 85). A recent breakthrough uncovered by our team and independently by the Fajkus laboratory revealed the *bona fide* TR from plants (99, 206). This discovery fills an important gap in the evolutionary history of TR. With AtTR, we were able to identify additional 85 TR molecules across three major clades in the plant kingdom and to define a conserved secondary structure model for plant TR. We distinguished plant-specific RNA elements in our structure model, most notably a TWJ motif. The TWJ is retained in all known plant TR molecules with evidence of multiple base-pair covariations that provide strong support for its biological function. We hypothesize that the function of the TWJ is to engage accessory proteins necessary for telomerase RNP activity similar to the unique secondary structure elements found in TR counterparts in other eukaryotes (231). The goals of the current study were to determine how two previously identified telomerase-associated proteins, AtPOT1a and dyskerin, engage AtTR, and to test whether either of them associate with the TWJ.

AtPOT1a is one of two *A. thaliana* orthologs of human POT1 protein. hPOT1 stimulates telomerase RAP together with TPP1 by retaining hTERT on the same single stranded telomeric DNA substrate (54, 176, 203). hPOT1-TPP1 heterodimer physically interacts the TEN domain of hTERT, which is also conserved in *A. thaliana* TERT (177). Previous studies of AtPOT1a and its role in promoting telomerase activity in Arabidopsis were constrained by our limited knowledge of the core components of the enzyme. However, the discovery of AtTR and the development of a robust *in vitro* reconstitution assay for Arabidopsis telomerase made it possible to dissect how

AtPOT1a promotes telomerase enzyme activity. Our results indicate that AtPOT1 associates with TERT. Specifically, we found that the loss of AtPOT1 did not impact AtTR abundance or secondary structure. However, DMS-MaPseq data indicated that the template region of AtTR is more accessible in the absence of AtPOT1a, an observation that is consistent with a role for AtPOT1a in promoting telomerase recruitment and association with telomeric DNA (202). We note that AtPOT1a is sufficient to specifically bind single-stranded telomeric DNA and to promote telomerase RAP in Arabidopsis *in vitro* (204). Thus, AtPOT1a may be the functional equivalent of the vertebrate POT1 serving to provide a physical connection between TERT and telomeric DNA to stimulate telomerase recruitment and RAP. Interestingly, the POT1 binding partner in vertebrates, TPP1, is not conserved in *A. thaliana*. Accordingly, it is unknown whether AtPOT1a bears the full function of POT1-TPP1 or interacts with a functional alternate of TPP1.

Dyskerin is an RNP maturation factor conserved from yeast to mammals (232). Human dyskerin participates in hTR biogenesis via a conserved H/ACA motif and is retained in the active telomerase RNP complex (140) (143). The function of dyskerin facilitating H/ACA containing RNA biogenesis and 3' end maturation has thus far been described only for RNA Pol II transcripts. Notably, AtTR like the other plant TR molecules is expressed by RNA Pol III (206), and does not contain a canonical H/ACA motif. Despite this, genetic data in Arabidopsis (205) and biochemical data in both Arabidopsis and in *Allium Cepa* (205) (206) argue that dyskerin contributes to telomerase function in plants. These observations led us to investigate whether AtTR mediates the dyskerin association with telomerase. Using two different affinity purification strategies, we determined that AtTR is required for dyskerin association. We also found that recombinant dyskerin specifically binds full-length AtTR with high binding affinity and promotes telomerase

RAP *in vitro*. These results argue that a plant-specific, non-canonical mechanism is employed for dyskerin interaction with AtTR.

How Arabidopsis dyskerin came to bind a Pol III RNA that lacks an H/ACA motif is unknown. However, it is notable that the polymerase responsible for TR synthesis appears to have switched more than once during TR evolution (85). Since Arabidopsis contains canonical H/ACA snoRNA species (233), it is possible that dyskerin developed an additional binding mode as it co-evolved with AtTR. Perhaps AtTR was originally a Pol II transcript that acquired an internal stretch of U-rich sequence downstream of the dyskerin binding site. This change together with modification to the promoter could be sufficient to enable Pol III-mediated transcription and termination, yet allow dyskerin binding to be retained.

Mapping the dyskerin binding site within AtTR demonstrated that the plant-specific TWJ consisting of P1a with two other stems, P1b-P1c stem and P4-P5-P6 stem is the binding site. While mammalian and yeast TR maintain a TWJ, this structure is constructed by local branches completely embedded within the 'stem-terminus' element (STE, also known as CR4/5 in mammals and yeasts) (Figure 28). In contrast, the plant TR TWJ utilizes one stem from the STE, one stem from template core-enclosing P1b-P1c, and one stem from plant-specific long-range interaction, P1a (Figure 28). While the mammalian TWJ interacts with TERT to promote telomerase RNP assembly (234), we hypothesize the plant TWJ serves a different function because of its structural divergence. Further, we found that removing P1a abolished the dyskerin-mediated stimulation of telomerase RAP *in vitro* while disrupting the P1a structure reduced AtTR stability and telomerase activity *in vivo*. Our data argue that P1a is an essential RNA element of AtTR, which assembles into a plant-specific TWJ motif, and associates with dyskerin to promote telomerase RAP *in vitro* and AtTR stability *in vivo*.

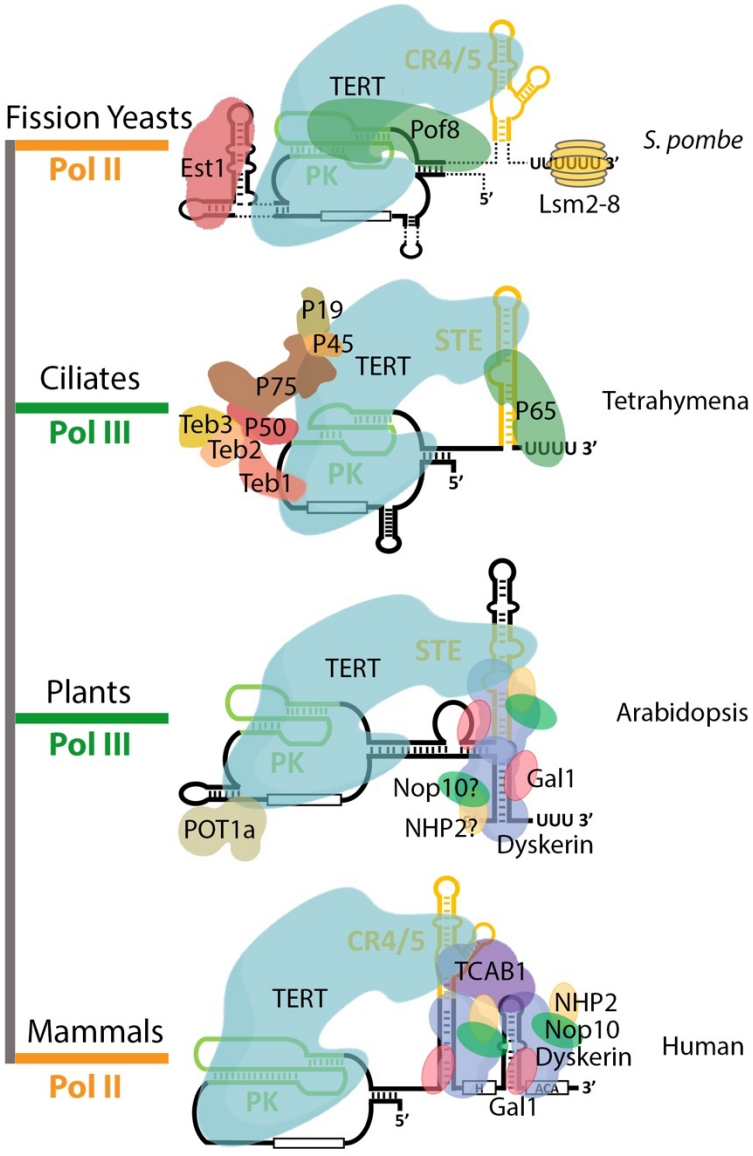


Figure 28. Plant telomerase RNP harbors unique and conserved components that bridge ciliate and mammalian enzymes.

Left: the diagram indicating TR transcribed by Pol II (orange) or Pol III (green). Right: schematic models of telomerase complexes in *S. pombe*, Tetrahymena, Arabidopsis, and human. TR secondary structures are labeled in green (PK) and yellow (CR4/5 or STE) to indicate essential RNA domains for TERT recognition. Accessory proteins presented have been determined by direct interaction with active telomerase.

Finally, we used cryo-EM to investigate the AtTR-dyskerin interaction in detail. We present an 8.6-Å cryo-EM structure of recombinant AtTR-dyskerin-NOP10-GAL1 complex (Figure 27). Our data reveal two copies of dyskerin and two GAL1 molecules retained on a single AtTR. Strikingly, the plant complex displays an identical dyskerin and GAL1 architecture as the human telomerase RNP (72). We found that dyskerin proteins dimerize and bind two separate RNA helixes, which anchors the entire complex. This consistency of the dimerized dyskerin architecture in human and Arabidopsis telomerase is not unexpected given the highly conserved structure of dyskerin (232). However, what is surprising is the arrangement of AtTR after modelling. We found that two arms of the TWJ, P1b-P1c and P4-P5-P6, serve as the 5' and 3' helices structurally equivalent to the P4 and P7-P8 stems within human H/ACA motif. However, the plant-specific P1a inserts to the interface between the dyskerin proteins (Figure 27). This model not only provides a mechanism for how dyskerin engages the AtTR TWJ, but it also provides additional support for the importance of P1a in dyskerin binding. While we cannot conclude that the AtTR TWJ is the structural or functional equivalent of the human TR H/ACA motif, our cryo-EM structure provides a non-canonical mechanism for dyskerin interaction with RNA, and suggests that dyskerin engages other RNA targets yet to be defined.

It is not unusual to discover accessory factors associated with telomerase that do not behave in canonical ways. For example, the La-related protein family 7 (LaRP7) protein Pof8 has been described as a constitutive subunit of the fission yeast *Schizosaccharomyces pombe* telomerase by direct interaction with the Pol II transcribed yeast TR (Figure 28) (127, 128). Pof8 is responsible for facilitating the binding of the Lsm complex to TR, which in turn promotes the entire telomerase RNP assembly (128). However, LaRP7 and other La family proteins are best known for recognizing and binding of Pol III transcripts (124), including the well-characterized LaRP7

protein P65 in Tetrahymena (Figure 28) (120-123). Although crystal structures demonstrated the structural homology between Pof8 and P65 (129, 136), a non-canonical model was recently proposed that Pof8 associates with the unexpected pseudoknot (PK) structure of yeast TR and plays a key role in telomerase RNA folding quality control (136). Furthermore, Alazami syndrome patients with LaRP7 deficiency showed a phenotype of insufficient telomere maintenance (130), which indirectly suggest the function of human LaRP7 in telomere and telomerase biology. Overall, more information, especially from a complete identification of components of telomerase RNP complexes, is needed to generate a general model and evolutionary explanation for the divergence of telomerase-associated proteins.

We present a model for telomerase holoenzyme in Arabidopsis (Figure 28). Compared to other characterized telomerase RNP complexes, Arabidopsis telomerase consists of Pol III transcribed TR similar to Tetrahymena. However, its association with a POT1 ortholog and the dyskerin complex is more akin to human telomerase. Our discovery that plant telomerase retains vertebrate-like associated proteins via a plant-specific assembly reinforces the conclusion that plant telomerase provides a unique evolutionary bridge uniting the highly divergent telomerase RNP complexes described to date (99).

Materials and methods

Plant material, growth conditions, and transformation

Arabidopsis thaliana accession Col-0, WS, *attr* (Flag_410H04), *attert* (SALK_041265C), and *atpot1a* (ref) were used in this study. Seeds were sterilized in 50% bleach with 0.1% Triton X-100 and then plated on half Murashige and Skoog (half MS) medium with 0.8% agar. Plants were grown at 22°C under long day light conditions. For genetic complementation, pHSN6A01-

AtTR, pHSN6A01-m16, pHSN6A01-m17, and pHSN6A01-m18 were transformed into second generation AtTR-/- (Flag_410H04) plants using Agrobacterium-mediated (*A. tumefasciens* GV3101) transformation as described. Transformants were selected on hygromycin in T1 and analyzed for telomerase phenotypes. In parallel, untransformed AtTR+/+ and AtTR-/- plants were analyzed.

Plasmids

The DNA sequences that encode full-length dyskerin (AtNAP57) (residues 1–565) or dyskerinΔC (residues 1–439) were cloned into a kanamycin-resistant pET28a plasmid to achieve a N-terminal 6xHis tag using the NEBuilder HiFi DNA assembly cloning kit. For co-expression of dyskerinΔC-NOP10-GAL1Δ, coding sequences of dyskerinΔC and full length NOP10 (residues 1–64) were cloned into an ampicillin-resistant pETDuet-1 plasmid to occupy independent expression cassettes using the NEBuilder HiFi DNA assembly cloning kit. DyskerinΔC, but not NOP10, was fused with a N-terminal 6xHis tag. GAL1Δ (AT3G03920) (residues 53–145) was cloned separately into a kanamycin-resistant pET28a plasmid without any tag fused to its N or C-terminal. For genetic complementation, AtTR was cloned into a binary vector pHSN6A01 under the control of the U6 promoter using the NEBuilder HiFi DNA assembly cloning kit. Mutagenesis was applied to pHSN6A01-AtTR to produce 18 AtTR variants using the Q5 site-directed mutagenesis kit (NEB). For in vitro Co-IP and telomerase reconstitution assays, full length AtPOT1a coding sequence was cloned into a pET28a plasmid to allow attachment with an N-terminal T7 tag. The pCITE-3xFLAG-AtTERT plasmid was a generous gift from Dr. Julian J.-L. Chen at Arizona State University.

Protein expression and purification

All proteins were expressed in *E. coli* Rosetta (DE3) strains that provide additional tRNAs. For purification of individual full-length dyskerin or dyskerinΔC protein, cells were resuspended in 50 mM Tris-Cl pH 7.3, 400 mM NaCl, 5% glycerol, and 2 mM DTT and lysed by a microfluidizer. Proteins were purified through HisTrap-HP column (GE Healthcare) and Superdex 200 increase 10/300 gel filtration chromatography (GE Healthcare).

For reassembly and purification of AtTR-dyskerinΔC-NOP10-GAL1Δ RNP, the dyskerinΔC-NOP10-GAL1Δ heterotrimer was co-expressed from a pETDuet-1 plasmid expressing dyskerinΔC and Nop10 and a pET28a plasmid expressing GAL1Δ. The complex was purified through a HisTrap-HP column (GE Healthcare) and Superdex 200 increase 10/300 gel filtration chromatography (GE Healthcare) in buffer (50 mM Tris-Cl pH 7.3, 350 mM KCl, and 2 mM DTT). AtTR and corresponding AtTR variants were prepared by an AmpliScribe T7-Flash transcription kit using PCR products as templates and were PAGE purified. The dyskerinΔC-NOP10-GAL1Δ complex was incubated with in vitro folded AtTR at 30°C for 45 min to enable RNA-protein interaction. The reassembled RNP was further purified through Superdex 200 increase 10/300 gel filtration chromatography (GE Healthcare) to remove unassembled subunits and aggregates. An additional 5 mM DTT was provided before storage at -80°C.

RIP

Anti-AtTERT antibody was affinity-purified as previously described (99). Briefly, antibody was preincubated with protein A magnetic beads (Dynabeads) at 4°C for 2h before IP. 1.5 g of WT (Col-0) Arabidopsis seedlings were ground in liquid nitrogen and homogenized in RIP buffer (100 mM Tris-OAC pH 7.5, 150 mM KGlu, 1 mM $MgCl_2$, 0.5% Triton X-100, 0.1%

Tween 20, 20 μL/mL Plant protease inhibitor mixture [Sigma-Aldrich], 1 μL/mL RNaseOUT [Thermo Fisher Scientific], and 2.5 mM DTT). After clearing by centrifugation, protein complexes were immunoprecipitated using preincubated anti-AtTERT magnetic beads at 4°C for 3 h. After incubation, beads were washed 7 times with RIP buffer and then resuspended with 1 mL of TRIzol reagent (Invitrogen) to extract RNA. For the TRAP measurement of immunoprecipitated samples, prior to resuspending in TRIzol, 5% beads were transferred into TRAP extension reaction (50 mM Tris-Cl pH 8.0, 50 mM KCl, 2 mM DTT, 3 mM $MgCl_2$, 1 mM spermidine, 1 μM TRAP-F primer, and 0.5 mM each ddNTPs) and incubated at 30°C for 45 min. Extended products were ethanol precipitated and used for direct TRAP assay as previously described (99).

In vitro *Co-IP*

Co-expression of T7-AtPOT1a and 3xFLAG-AtTERT was conducted using a TNT Quick Coupled transcription/translation kit (Promega) following the manufacturer's instructions with 35S-methionine labelling. In vitro folded AtTR, non-specific RNA (yeast tRNA) or mock (water) was provided independently into the respective reactions. In vitro reconstituted complexes were immunopurified with anti-FLAG M2 magnetic beads (Sigma-Aldrich) at room temperature for 1.5 h to enrich AtTERT and copurified proteins. The products were resolved on a 10% SDS-PAGE gel, dried, exposed and detected by a Typhoon FLA 9500 phosphorimager (GE Healthcare).

Target-specific DMS-MaPseq

The DMS-MaPseq assay was adapted from (222) and performed as previously described (99). Briefly, four-day-old WT (Col-0) or *atpot1a* seedlings were treated with 1% DMS or water (mock samples) in DMS reaction buffer (40 mM HEPES pH 7.5, 100 mM KCl and 0.5 mM $MgCl_2$)

with two times of 7.5 min vacuum. Total RNA was extracted from respective samples using RNA Clean & Concentrator-5 (Zymo Research) including in-column DNase digestion. 5 μg of high-quality RNA combined with gene-specific primers (AtTR or ACT2 mRNA) (5 pmol each) was heated and annealed for primer binding. TGIRT (Ingex) reaction was assembled following manufacturer's instructions. After RT, 1 μL of cDNA solution was directly added into a 50-μL PCR using Phusion High-Fidelity DNA Polymerase (New England BioLabs) to amply AtTR or ACT2 mRNA. PCR products were gel-purified, quantified, and directly assembled into Illumina sequencing libraries using the NEBNext Ultra II DNA Library Prep Kit (NEB) with 25 ng input. One mock library and two DMS libraries were built for each sample. Sequencing was performed on an 150x2 Illumina NextSeq 500 platform at Texas A&M University.

In vivo *Co-IP*

1.5 g of WT (WS) or *attr* seedlings were ground in liquid nitrogen and homogenized in Co-IP buffer (100 mM Tris-OAC pH 7.5, 150 mM KGlu, 1 mM $MgCl_2$, 0.5% Triton X-100, 0.1% Tween 20, 20 μL/mL Plant protease inhibitor mixture [Sigma-Aldrich], 1 μL/mL RNaseOUT [Thermo Fisher Scientific] and 2.5 mM DTT). After clearing by centrifugation, protein complexes were immunoprecipitated using preincubated anti-AtTERT magnetic beads at 4°C for 3 h. After incubation, beads were washed seven times with Co-IP buffer and then resuspended in SDS-loading buffer. Copurified proteins were resolved on an 8% SDS-PAGE gel and analyzed by western blot.

Antisense LNA oligo pull down

An antisense LNA oligo pull down assay was adapted from (235) with modifications. Two antisense LNA oligos were designed to target the accessible regions within AtTR and labelled with Biotin at the 3' end. 1.5g of formaldehyde-crosslinked four-day-old WT (WS) or *attr* seedlings were homogenized in lysis buffer (50 mM Tris-OAC pH 7.5, 10 mM EDTA, 1% SDS, 20 μL/mL Plant protease inhibitor mixture [Sigma-Aldrich], 1 μL/mL RNaseOUT [Thermo Fisher Scientific], and 6 mM DTT). The pre-warmed 2x hybridization buffer (50 mM Tris-OAC pH 7.5, 750 mM NaCl, 15% formamide, 1 mM EDTA, 20 μL/mL Plant protease inhibitor mixture [Sigma-Aldrich], 1 μL/mL RNaseOUT [Thermo Fisher Scientific], and 6 mM DTT) and 100 pmol antisense LNA oligos were provided and incubated at 42°C for 2 h to achieve oligo annealing. 100 μL Dynabeads MyOne Streptavidin C1 beads (Thermo Fisher Scientific) was added and incubated for additional 45 min at 42°C. Beads were washed five times with wash buffer (2x SSC, 0.5% SDS, 1 mM PMSF, 5 mM DTT) before resuspending 90% of beads in SDS-loading buffer for protein analysis by western blot. The remaining 10% of beads was subjected to protease K digestion (10 mM Tris-Cl pH 7, 100 mM NaCl, 1 mM EDTA, 0.5% SDS and 5% 20 mg/ml protease K) at 50°C for 45 min prior to the RNA extraction by TRIzol reagent. The extracted RNA was analyzed by RT-qPCR for RNA abundance.

EMSA

EMSA was adapted from (204) with modifications. WT AtTR and AtTR variants were prepared by an AmpliScribe T7-Flash transcription kit and purified by 6% denaturing PAGE. Initially purified RNAs were 5' end labeled with (γ-^{32}P) ATP (Perkin Elmer) using T4 Polynucleotide kinase (NEB) and PAGE purified again to remove free ATP. Radiolabeled RNAs

or non-radiolabeled competitors were folded by heating at 98°C for 2 min and slow cooling to room temperature. Full-length dyskerin or dyskerinΔC protein was diluted and mixed with RNA in EMSA reaction (25 mM Tris-Cl pH 7.5, 50 mM KCl, 2 mM $MgCl_2$, 10% glycerol, 1 mM DTT, 50 μg/ml BSA, 50 μg/ml yeast tRNA and 1 μL/ 20 μL RNaseOUT [Thermo Fisher Scientific]) at 30°C for 30 min. EMSA products were separated on 5% native PAGE in glycine buffer pH 9.0 at 4°C. Gels were dried under vacuum at 80°C for 1 h prior to exposure to a phosphor imager screen. Data were collected on Typhoon FLA 9500 phosphorimager (GE Healthcare) and quantified using Quantity One software (Bio-Rad).

In vitro *telomerase reconstitution*

In vitro telomerase reconstitution was performed as previously described (99) with modifications. Briefly, 3xFLAG-AtTERT was expressed in rabbit reticulocyte lysate (RRL) (TNT Quick Coupled transcription/translation kit; Promega) following the manufacturer's instructions. WT AtTR or AtTR variants were *in vitro* transcribed, folded, and assembled with AtTERT in RRL for 20 min at 30˚C at a final concentration of 1.5 μM (236). Serial dilutons of full-length dyskerin or dyskerinΔC-NOP10-GAL1Δ heterotrimer were added to the TERT-AtTR RRL mixture and incubated for an additional 20 min. Reconstituted telomerase was immunopurified with anti-FLAG M2 magnetic beads (Sigma-Aldrich) at room temperature for 1 h. On-beads telomerase was measured for telomerase activity via direct primer extension assay as described (236). Briefly, beads were resuspended in a reaction containing telomerase reaction buffer (50 mM Tris-Cl pH 8.0, 50 mM NaCl, 0.5 mM $MgCl_2$, 2 mM DTT, and 1 mM spermidine), 1 μM extension primer, ddNTPs and 0.18 μM of ^{32}P-dGTP (PerkinElmer). Reactions were incubated at 30°C for 90 min and terminated by phenol/chloroform extraction, followed by ethanol precipitation. A radiolabeled

recovery control was mixed with each reaction before phenol/chloroform extraction. DNA products were resolved by 10% denaturing PAGE gel, dried, exposed and imaged on a Typhoon FLA 9500 phosphorimager (GE Healthcare). Signal intensities were quantified using Quantity One software (Bio-Rad).

TRAP

The TRAP assay was conducted as described (204) with modifications. Unopened Arabidopsis flower bundles were collected for partial purification of telomerase. Purified telomerase was mixed in Go Taq Master Mix (Promega) containing a TRAP forward primer and incubated for 30 min at room temperature instead of 37 °C. After extension, 0.4 μM reverse primer was added to the reaction followed by 20 PCR cycles. Products were resolved by 6% denaturing PAGE, dried, exposed, and imaged on Typhoon FLA 9500 phosphorimager (GE Healthcare).

Cryo-EM specimen preparation and data collection

Cryo-EM specimens were prepared by applying 3 μL of the freshly reconstituted complex to a glow-discharged C-Flat 2/1 400-mesh Holey Carbon Grid and vitrified using a Vitrobot Mark III (FEI Company) at 22°C with 100% relative humidity. The images of the complex of H/ACA lobe and AtTR were recorded under a Titan Krios microscope (FEI Company) operated at 300 kV (UTHSC). Data were collected using EPU on a K2 Summit direct detection camera (Gatan) in the super-resolution mode with a sub-pixel size of 0.535 Å. Beam shift was enabled to encompass 5 exposures per hole. The beam intensity was adjusted to a dose rate of 7 e- per pixel per second on the camera. A 35-frame movie stack was recorded for each exposure with 0.2 s per frame for a total exposure time of 7 s. An in-column energy filter was used with a slit width of 20 eV.

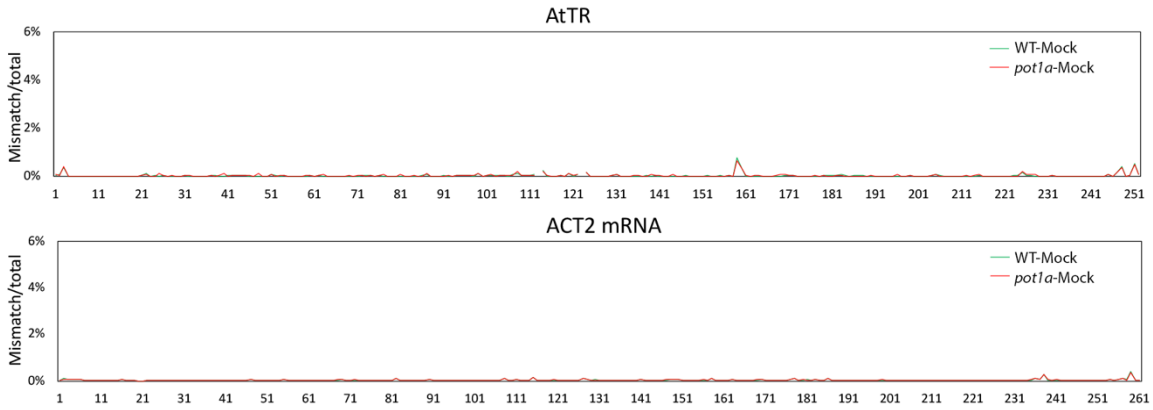
Cryo-EM data processing

A total of 4843 collected super-resolution movie stacks were aligned and summed by MotionCor2. The images were binned by 2 to yield a pixel size of 1.07 Å. The defocus value of each summed micrograph was determined using Gctf. Particles were then automatically picked and cleaned by 2D classifications in Relion. After the 3D classification, 58,138 particles were selected for the final 3D refinement. The refinement yielded the EM map at 8.6-Å resolution.

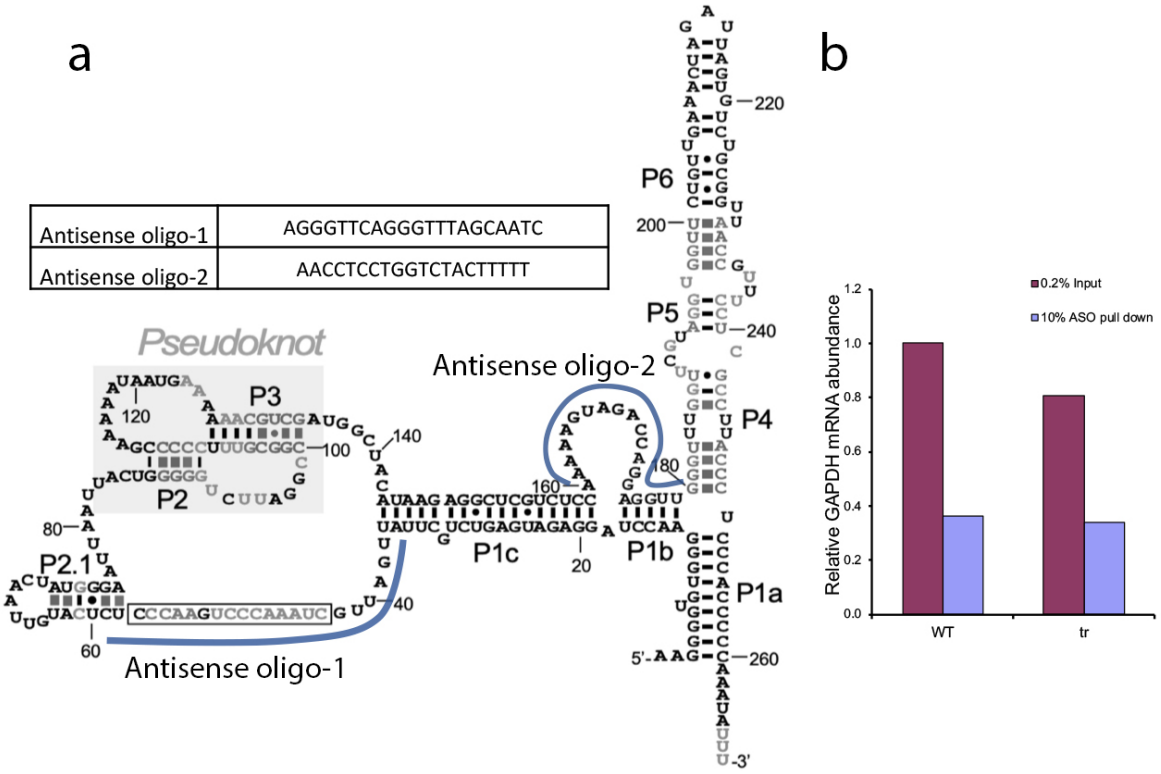
Model building

Protein models were homology-built using SWISS-MODEL with the crystal structure of the Shq1-Cbf5-Nop10-Gar1 complex from *Saccharomyces cerevisiae* as a template (PDB ID: 3UAI). The RNA model was first built by Rosetta RNA Denovo and refined into its densities by molecular dynamics flexible fitting (MDFF).

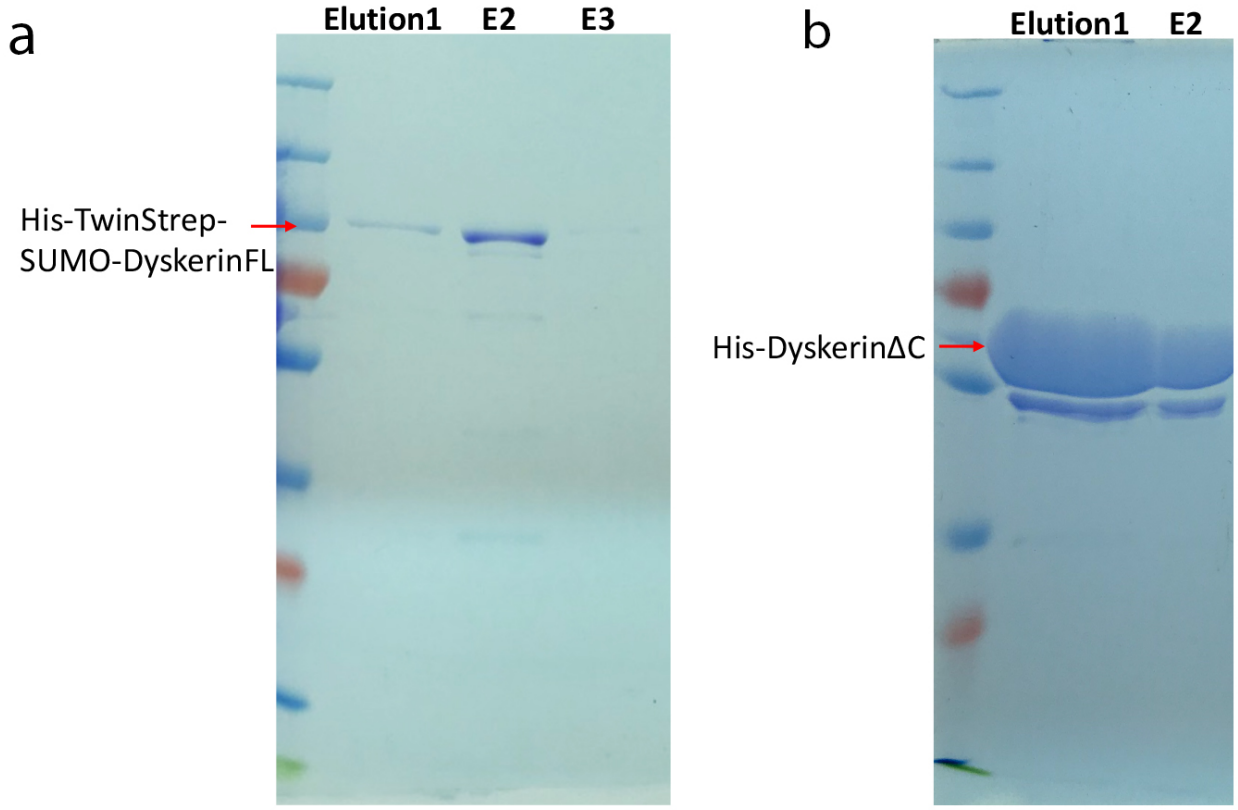
Supplementary data



Supplementary figure S10. DMS-MaPseq results of mock samples from WT and *pot1a* mutant. Average mutation frequencies are plotted along AtTR and ACT2 mRNA sequences.



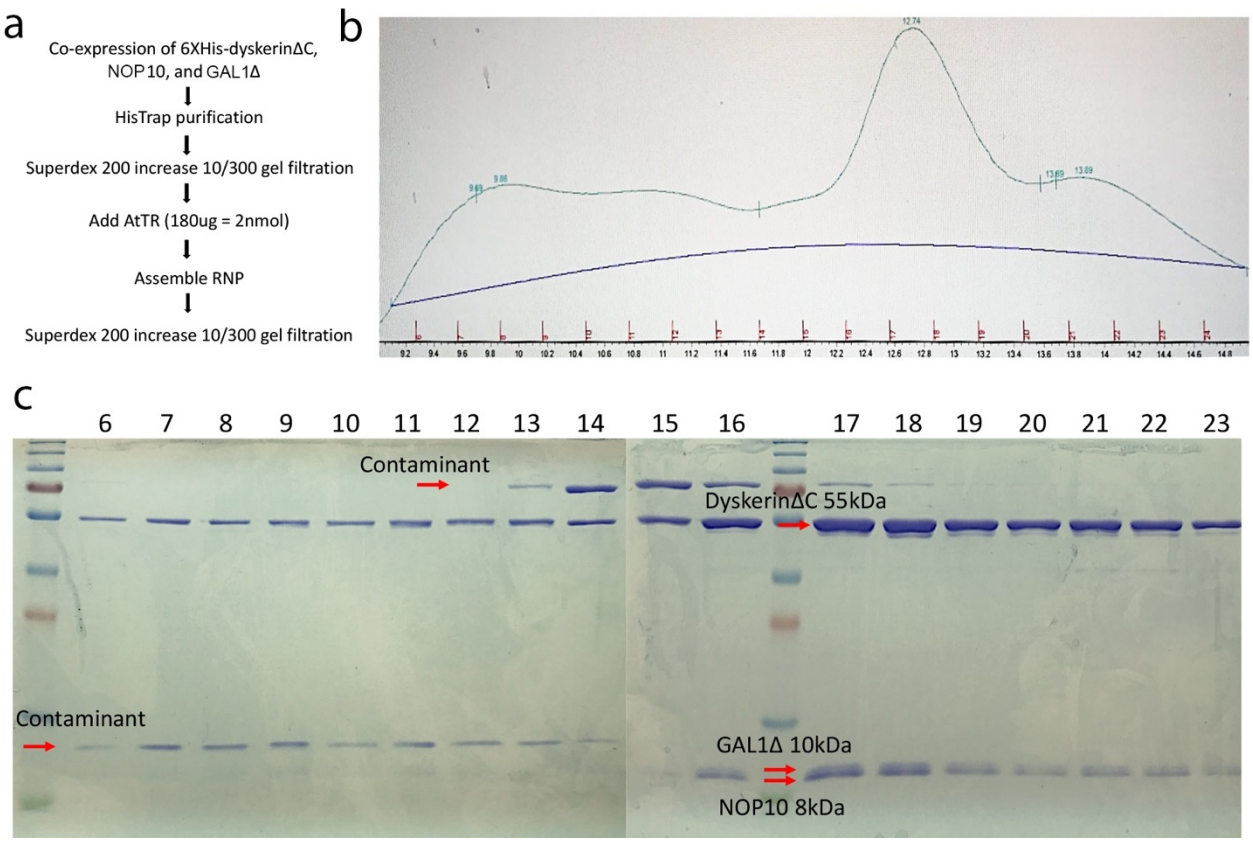
Supplementary figure S11. Antisense LNA oligo pull down of AtTR. a, schematic representation and sequences of the antisense LNA oligos used to specifically enrich AtTR. **b**, GAPDH mRNA was measured by qPCR as a negative control to support the specificity of antisense LNA oligos.



c

MAEVDISHSKKKKQDKTENDAADTGDYMIKPQSFTPAIDTSQWPILLKNYDRLNVRTGHYTPISAGHS
PLKRPLQEYIRYGVINLDKPANPSSHEVVAWIKRILRVEKTGHSGTLDPKVTGNLIVCIDRATRLVKSQQG
AGKEYVCVARLHSAVPDVAKVARALESLTGAVFQRPPLISAVKRQLRIRTIYESKLLEYDADRHLVVFWV
SCEAGTYIRTMCVHLGLLLGVGGHMQELRRVRSGILGENNNMVTMHDVMDAQFVYDNSRDESYLR
RVIMPLEMILTSYKRLVVKDSAVNAICYGAKLMIPGLLRFENDIDVGTEVVLMTTKGEAIAVGIAEMTTS
VMATCDHGVVAKIKRVMDRDTYPRKWGLGPRASMKKKLIADGKLDKHGKPNEKTPVEWSRNVVLP
TGGDAIIAGAAAAPEEIKADAENGEAGEARKRKHDDSSDSPAPVTTKKSKTKEVEGEEAEEKVKSSKKK
KKKDKEEEKEEEAGSEKKEKKKKKDKKEEVIEEVASPKSEKKKKKKSKDTEAAVDAEDESAAEKSEKKKK
KKDKKKKNKDSEDDEE

Supplementary figure S12. SDS-PAGE analysis of recombinant dyskerin full length (**a**) and dyskerinΔC (**b**). **c**, dyskerin full length sequence revealed a disordered C-terminus containing a putative nuclear location signal (red).



Supplementary figure S13. Purification of reconstituted dyskerinΔC-NOP10-GAL1Δ-AtTR. a, schematic representation of experimental design for dyskerinΔC-NOP10-GAL1Δ-AtTR complex purification. **b**, Superdex 200 increase 10/300 gel filtration chromatography results of dyskerinΔC-NOP10-GAL1Δ-AtTR complex. c, SDS-PAGE analysis of collected fractions from gel filtration chromatography. Fraction #18 was used for cryo-EM imaging.

CHAPTER IV

tRNA ADENOSINE DEAMINASE 3 IS REQUIRED FOR TELOMERE MAINTENANCE IN *ARABIDOPSIS THALIANA**

Summary

Telomere length maintenance is influenced by a complex web of chromatin and metabolism-related factors. We previously reported that a lncRNA termed AtTER2 regulates telomerase activity in *Arabidopsis thaliana* in response to DNA damage. AtTER2 was initially shown to partially overlap with the 5' UTR of the *tRNA ADENOSINE DEAMINASE 3* (TAD3) gene. However, updated genome annotation showed that *AtTER2* was completely embedded in *TAD3*, raising the possibility that phenotypes ascribed to *AtTER2* could be derived from *TAD3*. Here we show through strand-specific RNA-Seq, strand-specific qRT-PCR and bioinformatic analyses that *AtTER2* does not encode a stable lncRNA. Further examination of the original *tad3* (*ter2-1/tad3-1*) mutant revealed expression of an antisense transcript driven by a cryptic promoter in the T-DNA. Hence, a new hypomorphic allele of TAD3 *(tad3-2*) was examined. *tad3-2* mutants showed hypersensitivity to DNA damage, but no deregulation of telomerase, suggesting that the telomerase phenotype of *tad3-1* mutants reflects an off-target effect. Unexpectedly, however, *tad3-2* plants displayed progressive loss of telomeric DNA over successive generations that was not accompanied by alteration of terminal architecture or end protection. The phenotype was exacerbated in plants lacking the telomerase processivity factor POT1a, indicating that TAD3

---

* Reprinted with permission from 'tRNA ADENOSINE DEAMINASE 3 is required for telomere maintenance in *Arabidopsis thaliana*' by Bose, S., Suescún, A., Song, J., Castillo-González, C., Aklilu, B., Branham, E., Lynch, R., and Shippen, D.E., 2020. Plant Cell Reports volume 39, pages 1669–1685(2020). Copyright 2020 Springer Nature.

promotes telomere maintenance in a noncanonical, telomerase-independent pathway. The transcriptome of *tad3-2* mutants revealed significant dysregulation of genes involved in auxin signaling and glucosinolate biosynthesis, pathways that intersect the stress response, cell cycle regulation and DNA metabolism. These findings indicate that the *TAD3* locus indirectly contributes to telomere length homeostasis by altering the metabolic profile in *Arabidopsis*.

Introduction

Telomeres safeguard the genome by preventing chromosome ends from eliciting a DNA damage response and ensuring that terminal DNA sequences can be faithfully maintained (6). Due to the nature of eukaryotic DNA replication, telomeres culminate in a single-stranded extension termed the G-overhang (237), which acts as a substrate for the addition of telomeric repeats by telomerase. Plant telomeres are unusual in that one-half of their chromosome ends terminate in a G-overhang, and the other half in a blunt end bound by the Ku complex (195). Loss of Ku triggers extensive telomerase-dependent telomere elongation, presumably because blunt ends are converted to telomerase-accessible G-overhangs (31, 195). This unusual telomere architecture may further enhance genome stability, which seems advantageous given the sessile lifestyle of plants (196).

Telomere length homeostasis is modulated by a host of factors. At the telomere, components of the shelterin complex, particularly the TTP1/POT1 heterodimer, enhance telomerase activity and processivity on human telomeric DNA (53). In *Arabidopsis* POT1a associates with the telomerase ribonucleoprotein complex (RNP) and stimulates its repeat addition processivity (202, 204). Plants deficient in POT1a undergo telomeric DNA attrition at a rate similar to the amorphic telomerase (*AtTERT*) mutant (202). The progressive loss of telomeric

DNA in telomerase mutants ultimately causes a critical length threshold to be breached, activating a DNA damage response that leads to telomere fusion and genome-wide instability. *Arabidopsis* telomeres normally span 2–5 kb in length; telomere tracts shorter than 1kb have an increased probability of being recruited into end-to-end chromosome fusions (238). Thus, an optimal telomere length setpoint must be established to maintain genome integrity (239, 240).

In addition to canonical telomere-associated factors, genetic screens performed in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* demonstrate that telomere length is also influenced by a wide variety of "non-telomeric" genes that function in various aspects of DNA metabolism, chromatin modification, vesicular trafficking, RNA metabolism, ribosome metabolism and translation (241, 242). Perturbation of cell cycle progression can also alter telomere length. For example, mutation of *RAD1*, a component of the intra-S DNA damage checkpoint, leads to telomere shortening in *S. pombe* (243). Similarly, Rad1 functions as a positive regulator of telomere length in mammals, working in concert with Hus1 and Rad9 in the 911 complex (244).

Telomere dysfunction induces Programmed Cell Death (PCD) in plant meristems to eliminate genetically unstable cells (245, 246). PCD activation is essential for cell differentiation and proper development and is also involved in pathogen and environmental stress responses (247). PCD activation involves various kinds of molecular signals including plant hormones, calcium and reactive oxygen species (ROS) (248). Different hormonal pathways are interconnected to fine-tune PCD via transcriptional regulation. The auxin hormone regulates plant growth, and under normal conditions concentrates at the quiescent center of the root stem cell niche. Under abiotic stresses, many of which induce the accumulation of reactive oxygen species (ROS), auxin levels decline causing PCD in root tissues (249, 250). Auxin signaling also controls cell cycle progression by

mediating activation of CdC2. CdC2/Cdk2 kinase activity is necessary for expression of telomerase activity at early S phase (251-253). Thus, telomerase is a downstream target of auxin signaling pathway.

Telomerase is comprised of two core components, the catalytic subunit TERT and a long non-coding RNA (lncRNA) TER/TR (254) that serves as a template for telomere repeat addition (18). In *A. thaliana*, two lncRNAs were initially identified as telomerase subunits (198). AtTER1 was uncovered through partial purification of telomerase, and proposed to be the canonical telomerase RNA subunit. AtTER2, expressed from a locus partially overlapping the *tRNA Adenosine Deaminase 3* (*TAD3*) gene, was uncovered by BLAST based on its high sequence similarity to AtTER1 (198, 201). Subsequent studies indicated that AtTER2 was stabilized and functioned to down-regulate telomerase activity in response to DNA double-strand breaks (201, 224). We recently employed an unbiased RIP-seq approach to identify lncRNAs associated with active telomerase under native conditions and failed to recover AtTER1 (99). Instead, a single lncRNA, AtTR, was significantly enriched. Further analysis by our lab and others revealed that AtTR was the *bona fide* telomerase RNA subunit in *A. thaliana* (99, 200, 206).

A new annotation of the *A. thaliana* genome, Araport11, extended the 5' UTR of *TAD3* to now fully embed *AtTER2*. This updated annotation prompted us to re-examine the *TER2/TAD3* locus to assess whether the phenotypes originally ascribed to *TER2* might instead result from mutation of *TAD3*. tRNA Adenosine Deaminase 3 (TAD3) catalyzes the deamination of adenosine at position 34 of the tRNA anticodon loop into Inosine to facilitate wobble base pairing (255). Yeast and plant TAD3 amorphic mutants are inviable (256, 257). Similarly, loss of TAD3 in fission yeast compromises cell survival by affecting cell cycle progression (258). Decreased expression of human *TAD3* impacts RNA editing for several tRNA species and is associated with intellectual

disability (259). Notably, TAD3 was uncovered in a genetic screen in *S. cerevisiae* as one of the essential genes that impacts telomere length maintenance (242).

Here we show through strand-specific RNA-Seq, strand-specific qRT-PCR, and bioinformatic analyses that *AtTER2* does not encode a stable lncRNA, and the telomere-related functions from this locus derive from the *TAD3* gene. Through analysis of additional *TAD3* mutant alleles, we report that hypomorphic *tad3* mutants are hypersensitive to DNA damage, but *TAD3* is not required to regulate telomerase activity in response to DNA damage. However, *TAD3* is required for telomere length maintenance. This unanticipated function is independent of telomerase, and appears to reflect a broader role for *TAD3* in modulating cellular metabolism.

Results

Reexamination of AtTER2 locus

The initial characterization of *AtTER2* was based on annotation of the Arabidopsis genome published by The Arabidopsis Information Resource, TAIR10 (Release date, November 2010) (260). *AtTER2* is located in the Crick strand on Chromosome 5, partially overlapping the 5' UTR of *TAD3*, encoded in the Watson strand (Figure 29A) (201). The non-overlapping region of *AtTER2* was used to design AtTER2-specific primers to trace the molecule by RT-PCR. Given that the current genome annotation for *A. thaliana*, Araport11 (Release date, June 2016), extended the 5' UTR of *TAD3* to fully embed *AtTER2* (Figure 29A), we designed a strand-specific RT-PCR approach to exclusively detect the AtTER2 transcript. We were unable to detect AtTER2 in flowers, leaves, and seedlings from wild type plants grown under normal conditions (supplementary Figure S14A). Cq values > 31 were obtained for AtTER2 amplification compared to Cq ≅ 19 for the internal control ACT2 (AT3G18780) (supplementary Figure S14A).

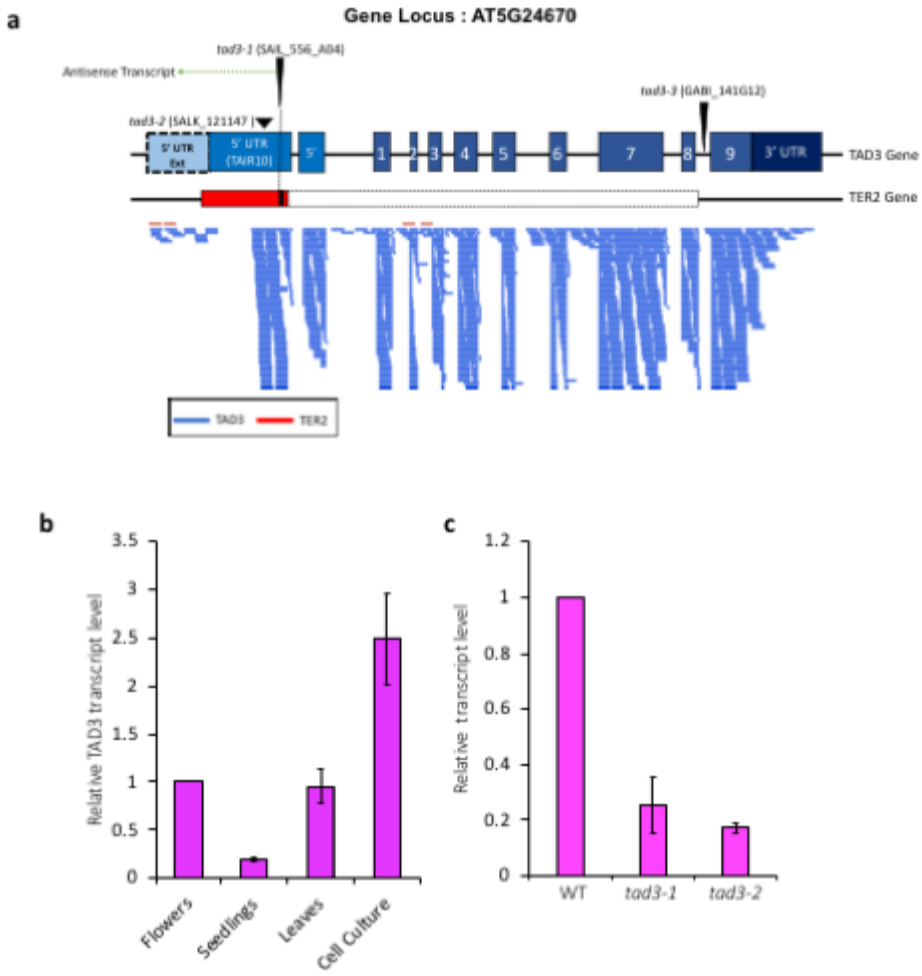


Figure 29. Reannotation of the TER2 locus based on TAIR10_v90.

(a) Schematic representation of the *TER2* and *TAD3* loci in *Arabidopsis thaliana* based on the Araport11 version of genome annotation. The *TAD3* gene (AT5G24670) is represented in blue and *TER2* in red. The previous genome annotation (TAIR10 + Araport11 5' Ext) placed TER2 within the 5' UTR of TAD3. The putative promoter would span the *TAD3* gene. The positions of the *tad3-1*, *tad3-2* and *tad3-3* T-DNA insertions are indicated by the black triangles. The short horizontal blue (TAD3) and red (TER2) lines below top panel denote stranded RNA-Seq reads from six-day-old wild type (WT) Col-0 seedlings. A cryptic antisense transcript emanating from the *tad3-1* insertion is indicated by the dotted green line. (b) qRT-PCR data for TAD3 mRNA in flowers, seedlings, leaves and cell culture. The mean of two biological replicates are shown as fold change with respect to WT flowers. The WT plants are homozygous for the *TAD3* allele. Error bars indicate standard deviation. (c) qRT-PCR data for TAD3 mRNA in flowers from WT, *tad3-1* and *tad3-2 plants*. The mean of two biological replicates are shown as fold change with respect to WT samples. Error bars indicate standard deviation.

Previous studies indicated that AtTER2 was stabilized and accumulated in response to DNA damage (201). To thoroughly explore AtTER2 expression, we performed total RNA sequencing on *tad3-2* mutants (see below) and wild type seedlings with and without zeocin treatment. Stranded RNAseq libraries were prepared from total RNA after depletion of ribosomal RNAs. Sequencing of untreated *tad3-2* and wild type seedlings produced a total of 51,194.244 (91.17%) and 59,996.775 (91.47%) reads, respectively, uniquely mapped to the reference genome. While sequencing of zeocin treated seedlings produced a total of 46,127.084 (91.59%) and 53,093.124 (89.75%) uniquely mapped to the reference genome in the *tad3-2* and wild type, respectively. *TAD3* expression in *tad3-2* mutants was ~33% of wild type (see below). However, we found no change in *TAD3* expression in wild type plants upon zeocin treatment. Moreover, no reads aligned to *AtTER2* in either the *tad3-2* or wild type datasets from mock (Figure 29A) or zeocin treated seedlings. Together, these data indicate that AtTER2 is not a stable lncRNA, and the previously detected PCR products likely reflect artifactual amplification of the Crick strand of the *TAD3* 5' UTR. Therefore, any functions previously ascribed to this locus derive from *TAD3*.

Identification of TAD3 mutant alleles

TAD3 is widely expressed, with peaks during bolting, formation of mature flowers and silique development (supplementary Figure S14B). *In silico* metanalysis of publicly available transcriptomic data using Genevestigator (261) indicated *TAD3* is most highly expressed in leaves, flowers and root apical meristem (supplementary Figure S14C). We verified this finding experimentally using qRT-PCR, and also found high *TAD3* expression in cell culture (Figure 29B).

Our previous analyses of the *TAD3* locus utilized the T-DNA insertion line SAIL_556_A04 (*ter2-1*) (201), now designated *tad3-1*, which resides in the 5' UTR

of *TAD3* (Figure 29A). Although no transcript spanning the T-DNA insertion in *tad3-1* could be detected (Figure 29A), qRT-PCR with strand-specific primers targeting a region 770 nt downstream of the T-DNA revealed the presence of an RNA transcript (supplementary Figure S15A and 2B), suggesting the activation of a cryptic promoter within the T-DNA (262). As this transcript could have indirect effects, we considered the *tad3-1* allele suboptimal for further studies, and characterized two additional T-DNA lines. One allele termed *tad3-3* carries a T-DNA in the intron between exons 8 and 9, but embryonic lethality was previously reported in homozygous mutants (256). The third T-DNA line (SALK_121147) termed *tad3-2* contains a T-DNA 902 nt downstream from the start of the *TAD3* 5' UTR (Figure 29A). In contrast to *tad3-1*, *tad3-2* does not produce an antisense transcript (supplementary Figure S15B). qRT-PCR analysis of floral RNA indicated that TAD3 mRNA is reduced by ~75% (p-value=0.06) in *tad3-1* and by 83% (p-value=0.01) in *tad3-2* mutants respectively (Figure 29C). Because of the higher knockdown and the absence of a potentially confounding antisense transcript, downstream analyses were performed using the *tad3-2* allele.

Plants deficient in TAD3 exhibit hypersensitivity to DNA damage and elevated programmed cell death

It was previously reported that *tad3-1* mutants exhibit an increased incidence of programmed cell death (PCD) in the Root Apical Meristem (RAM) after zeocin treatment (201). We re-examined this response in *tad3-2* mutants by imaging the RAM of seedlings stained with Propidium Iodide (PI) four- and six-hours post-treatment with 20 μM zeocin. At four hours, 70% of the *tad3-2* seedlings displayed PCD, compared to 0% of wild type seedlings (Figure 30A and B). Thus, *tad3-2* mutants are hypersensitive to DNA damage, consistent with the previous results

obtained in *tad3-1* mutants (201). The prior study indicated that *tad3-1* mutants have an intrinsically elevated accumulation of DDR-related transcripts, including *BRCA1, PARP1* and *PARP2* (201). However, transcriptomic analysis of *tad3-2* and wild type seedlings grown under normal conditions revealed only a slight increase in BRCA1 expression (1.8-fold (FDR<0.05)) and PARP2 expression (1.96-fold (FDR<0.05)) in *tad3-2* mutants. Data from the RNA-Seq experiment was confirmed through RT-qPCR analysis. While BRCA1 showed a consistent increase in gene expression in *tad3-2* mutants, expression levels of PARP1 and PARP2 remained unchanged upon loss of TAD3 (supplementary Figure S16A).

To test if increased PCD in *tad3-2* seedlings correlates with accumulation of endogenous DNA damage, we performed a modified version of the single cell comet assay using protoplasts extracted from 7 days old seedlings. We measured Percentage DNA in the comet Tail (PDT) and Tail Length (TL) to calculate Tail Moment (TM) <Olive, 2006 #2001>. Statistical analysis of any of these three parameters gauges the level of DNA damage (263) and can be confirmed by the other two parameters. For convenience, we represented DNA damage as a function of PDT (%PDT). As a positive control, assays were performed on cells from plants lacking ATR, a master regulator of the DNA damage response machinery (264). As expected, PDT was significantly higher in *atr* mutants compared to wild type (Figure 30C and D). However, the level of PDT observed in *tad3-2* mutant was similar to wild type. We conclude that loss of TAD3 does not lead to accumulation of damaged DNA, and existing DNA damage is not an underlying cause of the PCD in *tad3-2* mutants. To test whether DNA damage sensing and repair capabilities were affected in *tad3-2* mutants, we performed gene ontology analysis on the zeocin-induced differentially expressed genes (DEGs) in wild type and *tad3-2* seedling. We did not find any find conspicuous difference in the GO term enrichment between the genotypes (supplementary Figure S16C).

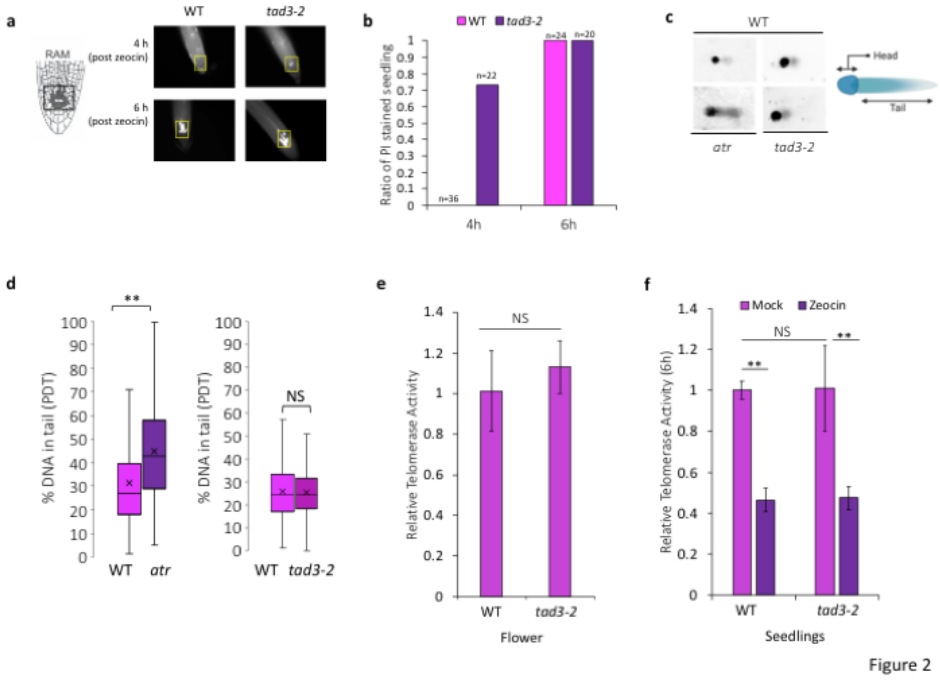


Figure 2

Figure 30. The TAD3 locus does not modulate DNA damage related pathways.

(a) Schematic representation of an *A. thaliana* root tip with the Root Apical Meristem (RAM) highlighted in gray (left). On the right, images of roots from *A. thaliana* seedlings stained with Propidium Iodide solution (PI) following treatment with 20 µM zeocin. Photos are shown of representative 4-days old WT and *tad3-2* seedlings treated with zeocin for 4 hours and 6 hours followed by PI staining. Yellow box highlights the RAM and Programmed Cell Death. (b) Percentage number of seedlings with PI staining in the RAM of WT and *tad3-2* seedlings treated with zeocin for 4 or 6 hours. Numerical values indicate total number of roots imaged for each condition. 0 out of the 36 WT seedlings showed RAM PCD at 4 hours post-zeocin treatment. (c) Representative images of data obtained from comet assays performed on protoplasts extracted from seedlings. The length and intensity of the comet tail indicates the level of DNA damage. (d) Values for percentage DNA in tail (%PDT) from the comet assay plotted using a box and whisker plot. Top and bottom edges of the box represent the first and the third quartiles, respectively. The length of the whisker spans the minimum to maximum values. The straight line inside the box represents the median and 'X' stands for sample mean. Normally distributed data have an overlapping mean and median. More than 1000 comets were scored for each genotype. **p-value <0.001 and NS = not significant based on students t- test. (e) Data obtained for quantitative Telomere Repeat Amplification Processivity (qTRAP) assays performed with flower bundles from WT and *tad3-2* mutants. The mean of three biological replicates are shown as fold change with respect to WT samples. Error bars indicate standard deviation. NS = not significant based on u-test. (f) qTRAP results for 7-day old seedlings untreated (mock) or treated with 20 µM zeocin for 6 hours. The mean of three biological replicates is shown as fold change with respect to WT samples at 0 h mock treated. Error bars indicate standard deviation. ** = p-value <0.01 based on u-test.

We further analyzed our RNA-Seq data to compare the transcriptional response to zeocin at the gene level, 77 of the top 100 zeocin-induced DEGs in wild type and *tad3-2* are commonly deregulated; notably, these genes are deregulated in the same magnitude (Supplementary Table 2). This provides ample and detailed evidence of the unaltered transcriptional response to zeocin treated *tad3-2* mutants, as compared to wild type.

Finally, since AtTER2 was reported to negatively regulate telomerase activity in response to DNA double-strand breaks (224), we re-assessed this conclusion using the *tad3-2* allele. We found no difference in telomerase activity levels of *tad3-2* flowers or seedlings relative to wild type (Figure 30E and F). Two hours of zeocin treatment induced a robust DNA damage response as evidenced by ~100-fold increase in *BRCA1* expression (supplementary Figure S16B). Telomerase activity was decreased by ~50% in both wild type and *tad3-2* mutants, after six hours of zeocin treatment (Figure 30F), arguing that the telomerase response to zeocin is not dependent on *TAD3.* Altogether, these findings indicate that *tad3* mutants are hypersensitive to DNA damage, but TAD3 does not regulate the response to DNA damage.

A telomere maintenance defect in tad3-2 mutants is independent of telomerase

As part of our characterization of the *TAD3* locus, we monitored bulk telomere length over three consecutive generations in *tad3-2* mutants. Terminal Restriction Fragment (TRF) analyses revealed a subtle but progressive loss of high molecular weight telomere tracts in the *tad3-2* mutants relative to wild type siblings (Figure 31A). Genetic complementation was used to test if the telomere maintenance defect was due to the loss of TAD3. *tad3-2* mutants were transformed with a full-length *TAD3* gene under the control of its native promoter (P_{TAD3}::TAD3) (Figure 31B).

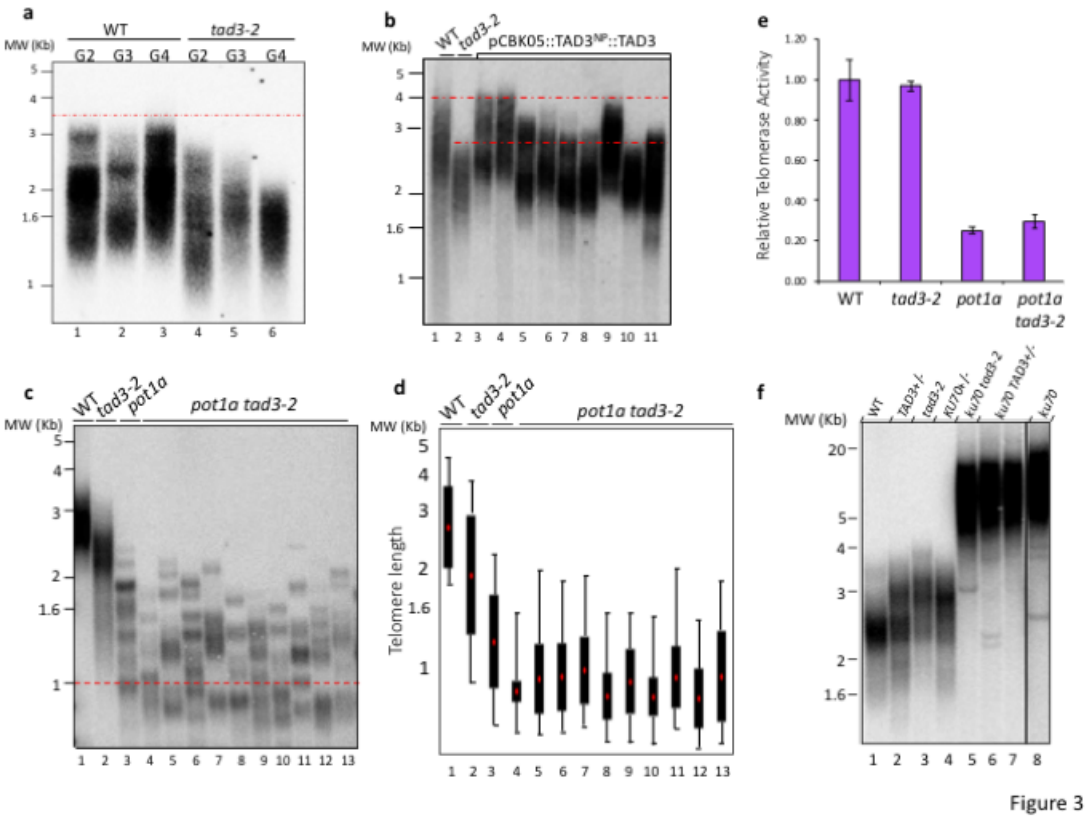


Figure 31. TAD3 maintains telomeres via a telomerase-independent pathway.

(a) Results of Terminal Restriction Fragment (TRF) analysis to measure bulk telomere length in *tad3-2* mutants from second (G2), third (G3) and fourth (G4) generations of homozygosity along with the segregating WT siblings followed in parallel through successive generations. Red dotted line indicates the maximum telomere length for the WT samples from this cross. (b) Results of TRF analysis performed for genetic complementation of *tad3-2* mutants. 4-week-old *tad3-2* mutants were transformed with pCBK05::NP^{TAD3}::TAD3. Results are shown for WT (lane 1), DNA from the same generation of untransformed *tad3-2* mutants as the control (lane 2), and complementation lines (lanes 3-11). Red dashed line indicates maximum telomere lengths for WT and *tad3-2* samples. Lanes 1 and 2 contains DNA derived from a pool of ~100 seedlings and lane 3-11 contains DNA from individual transformants. (c) Results from TRF analysis of WT, G2 *tad3-2*, G2 *pot1a* and G2 *pot1a tad3-2* mutants. DNA samples are derived from individual plants of each genotype. Red line indicates the critical telomere length threshold of 1 Kb. (d) Quantification of the TRF gel from panel C determined by TeloTool. Data are represented as box and whisker plot. Red dot within the box represents the mean value**.** (e) Results from qTRAP assays performed with flowers from WT, G2 *tad3-2*, G2 *pot1a* and G2 *pot1a tad3-2* samples. The mean of three biological replicates are shown as fold change with respect to WT samples (f) Results of TRF analysis for WT, *tad3-2*, *tad3-2 ku70* and *ku70* mutants. DNA was analyzed from individual segregating siblings belonging to G1 of the *tad3-2* X *ku70* cross. The gel has been sliced to highlight the lane for *ku70*. A vertical line separates the *ku70* lane from the rest of the gel.

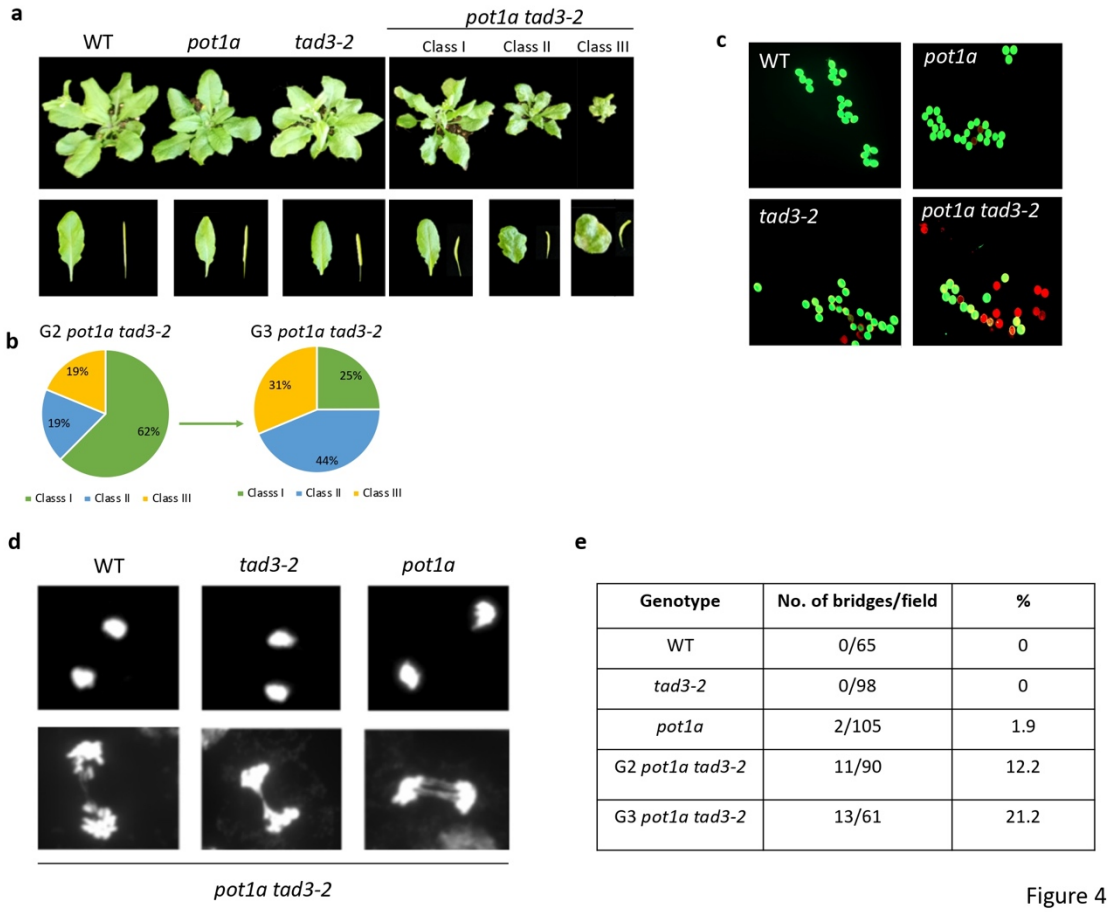
Within a single generation, three of the nine independent transformants showed complete recovery of telomere length to wild type, and in five others some telomere tracts were longer than in *tad3-2* mutants (Figure 31B), supporting a role for TAD3 in telomere length maintenance.

When telomerase activity is limiting, shorter telomeres are preferentially elongated (265-267). To investigate if depletion of long telomeres in *tad3-2* mutants reflects a defect in telomerase, we generated double mutant plants. Our initial goal was to obtain plants lacking *TAD3* and *TERT*. Both genes are situated on chromosome 5, approximately 2.9 Mb apart with *TAD3* proximal to the centromere (260). Linkage calculations indicated that Mendelian segregation of the two loci was possible, and predicted 6.25% of the offspring of *TAD3-2*$^{+/-}$ *TERT*$^{+/-}$ would be *tad3-2*$^{-/-}$ *tert*$^{-/-}$. Nevertheless, we failed to recover any homozygous double mutants among ~200 offspring analyzed, suggesting that *TERT* and *TAD3* may cooperate for some essential non-telomeric function. As an alternative strategy, we made crosses to generate plants doubly deficient in *TAD3* and *POT1a*. First-generation (G1) *pot1a tad3-2* plants were readily obtained and were self-pollinated to produce second-generation (G2) *pot1a tad3-2* mutants. In parallel, we propagated wild type, *pot1a* and *tad3-2* single mutants. Each line was grown for several consecutive generations (G2-G4).

We assessed how the combined loss of POT1a and TAD3 impacted telomere length using TRF (Figure 31C). As expected, telomeres in G2 *pot1a* mutants were shorter than wild type and displayed a discrete banding pattern indicative of a telomerase deficiency (Figure 31C). Strikingly, telomeres in G2 *pot1a tad3-2* were even shorter than the *pot1a* single mutants (Figure 31C). A banding pattern was visible for longer telomeres, but telomere tracts shorter than 1kb were more heterogeneous (Figure 31C). Quantification of telomere length using TeloTool (268) showed wild type spanned 2.0-5.0 kb with a mean telomere length (MTL) of 3kb (Figure 31D). *tad3-2*

telomeres were similar though slightly shorter (range=1.2-4.0 kb; MTL= 2.1 kb). In contrast, telomeres in G2 *pot1a tad3-2* plants were significantly shorter (range=0.5-2.1 kb; MTL=1 kb) than telomeres in G2 *pot1a* mutants (range=0.8-2.8 kb; MTL=1.7 kb) (Figure 31D). We conclude that combined loss of TAD3 and POT1a accelerates telomere shortening relative to the loss of POT1a alone.

Progressive telomere shortening ultimately causes profound developmental defects as a consequence of genome instability (19). Consistent with the hypothesis that TAD3 acts synergistically with telomerase, there was accelerated shortening of telomeres in *pot1a tad3-2* mutants (Figure 31C and supplementary S17A), which correlated with an early onset of stem cell-related defects (Figure 32A and supplementary S17B). *tad3-2* mutants displayed no visible developmental defects for three generations (Figure 32A and supplementary S17B). Conversely, gross morphological abnormalities were evident in *pot1a tad3-2* mutants beginning in G2 and worsened over the generations (Figure 32A and supplementary S17B). Importantly, *pot1a* single mutants were indistinguishable from wild type in G2 (Figure 32A and supplementary S17B). We categorized *pot1a tad3-2* plants into three groups: class I mutants were similar to wild type; class II plants had stunted growth with leaf abnormalities, constricted rosettes, and occasional hook-shaped siliques; and class III mutants were more severely impacted than class II (Figure 32A). The number of class II and class III mutants increased with each generation (Figure 32B). Pollen viability of G2 *pot1a tad3-2* was decreased relative to WT or either single mutants (Figure 32C), and later generation *pot1a tad3* plants were sterile, failing to produce any siliques (Figure 32A).



Genotype	No. of bridges/field	%
WT	0/65	0
tad3-2	0/98	0
pot1a	2/105	1.9
G2 *pot1a tad3-2*	11/90	12.2
G3 *pot1a tad3-2*	13/61	21.2

Figure 4

Figure 32. Exacerbated reproductive and developmental defects and genome instability in *pot1a tad3-2* mutants. (a) Photos of rosettes, individual leaves and siliques from three-week-old WT, G2 *tad3-2*, G2 *pot1a* and G2 *pot1a tad3-2* plants. Siliques and leaves were collected from the same position for all samples. For G2 *pot1a tad3-2* mutants representative images from three phenotypic classes (I, II, and III) are shown. (b) Pie chart illustrating the relative fraction of plants belonging to each phenotypic class of G2 *pot1a tad3-2* double mutants. (c) Viability of pollen grains produced by WT, G2 *tad3-2*,G2 *pot1a* and G2 *pot1a tad3-2* assessed with the FDA staining protocol in combination with PI staining. Live pollen metabolizes the FDA into green colored fluorescein. PI stains dead pollen. (d) Mitotic spreads of anaphase were made from flower pistils of four-week-old WT, G2 *tad3-2*, G2 *pot1a* and G2 *pot1a tad3-2* plants using previously published protocol. Chromatin was stained with DAPI and observed with 100X magnification on a fluorescent microscope. (e) Quantification of anaphase bridges obtained from analyzing mitotic fields in pistils from genotypes as indicated.

The worsening of developmental phenotypes correlated with an increased incidence of telomere tracts below the critical 1kb length threshold (238) (supplementary Figure S17A). Analysis of mitotically dividing cells revealed 12% of the anaphases in G2 *pot1a tad3-2* harbored bridged chromosomes, consistent with telomere-to-telomere fusion, compared to 1.9% in *pot1a* and 0% in *tad3-2* and wild type siblings (Figure 32D and E). The percentage of anaphase bridges increased to 21% in G3 *pot1a tad3-2* (Figure 32E). Telomere fusion PCR experiments confirmed that the chromatin bridges reflected end-to-end chromosome joining through telomeres (supplementary Figure S17C and D).

The data presented thus far suggest that TAD3 and telomerase act in parallel pathways to maintain telomere length. However, an alternative possibility is that TAD3 acts in a pathway overlapping with telomerase. Although repeat addition processivity of telomerase is severely compromised in *pot1a* mutants, enzyme activity is not entirely abrogated (202). Thus, with both TAD3 and POT1a simultaneously inactivated, telomerase activity could be entirely abolished. To test this, qTRAP was performed with *pot1a tad3-2* mutants. There was no difference in telomerase activity in *pot1a tad3-2* mutants compared to *pot1a* (Figure 31E), indicating that TAD3 is not required for maximal telomerase stimulation.

Finally, we asked if TAD3 was required for telomerase recruitment and enzymology at chromosome ends *in vivo* by assessing how the loss of TAD3 impacted telomere elongation in plants lacking Ku70. If telomere elongation in *ku70* mutants requires TAD3, then plants doubly deficient in both Ku70 and TAD3 should not have ultra-long telomeres. To test this hypothesis, we crossed *ku70* and *tad3-2* single mutants and segregated double mutants from *Ku70+/- TAD3-2+/-* parents. TRF analysis performed with the G1 siblings, revealed no difference in telomere length in G1 *ku70 tad3-2* plants compared to G1 *ku70* (Figure 31F). Thus, TAD3 does not appear

to play a critical role in promoting telomerase engagement and extension at chromosome ends. Altogether, our results support the conclusion that TAD3 acts independently of telomerase for telomere length maintenance.

Telomere terminal architecture is unperturbed in tad3-2 mutants

Another explanation for the telomere shortening phenotype is that telomere architecture is compromised in *tad3-2* mutants, leaving chromosome ends vulnerable to inappropriate nucleolytic processing. Telomere integrity cannot be grossly altered since *tad3-2* mutants do not suffer end-to-end fusions, but to test for subtle perturbation, we measured the status of the G-overhang using in-gel hybridization (269). The G-overhang signal was increased in *ku70* mutants by 2.5-fold (Figure 33A), consistent with the conversion of blunt-end telomeres into G-overhangs (195). In contrast, we found no difference in the G-overhang signal in G2 *pot1a tad3-2*, G2 *tad3-2,* and G2 *pot1a* mutants compared to wild type (Figure 33A). Next, we examined the integrity of blunt end telomeres using the dUTP-PENT assay (195). As expected, approximately 55% of the signal was retained in wild type samples after UDG treatment, confirming half the telomeres are blunt ended, while in *ku70* mutants, the signal was reduced by ~89%, consistent with conversion of most blunt ends into G overhangs (Figure 33B). *tad3-2* mutants exhibited a wild type level signal (~50%) after UDG treatment. We verified blunt end telomeres in *tad3-2* mutants using a hairpin ligation assay (31, 195). Blunt-ended telomeres migrate as a higher molecular weight smear and then are lost upon BamHI digestion. A high molecular weight smear sensitive to BamHI was observed in both wild type and *tad3-2* samples, but not in *ku70* (Figure 33C). We conclude that TAD3 does not play an essential role in maintaining the proper architecture of chromosome termini.

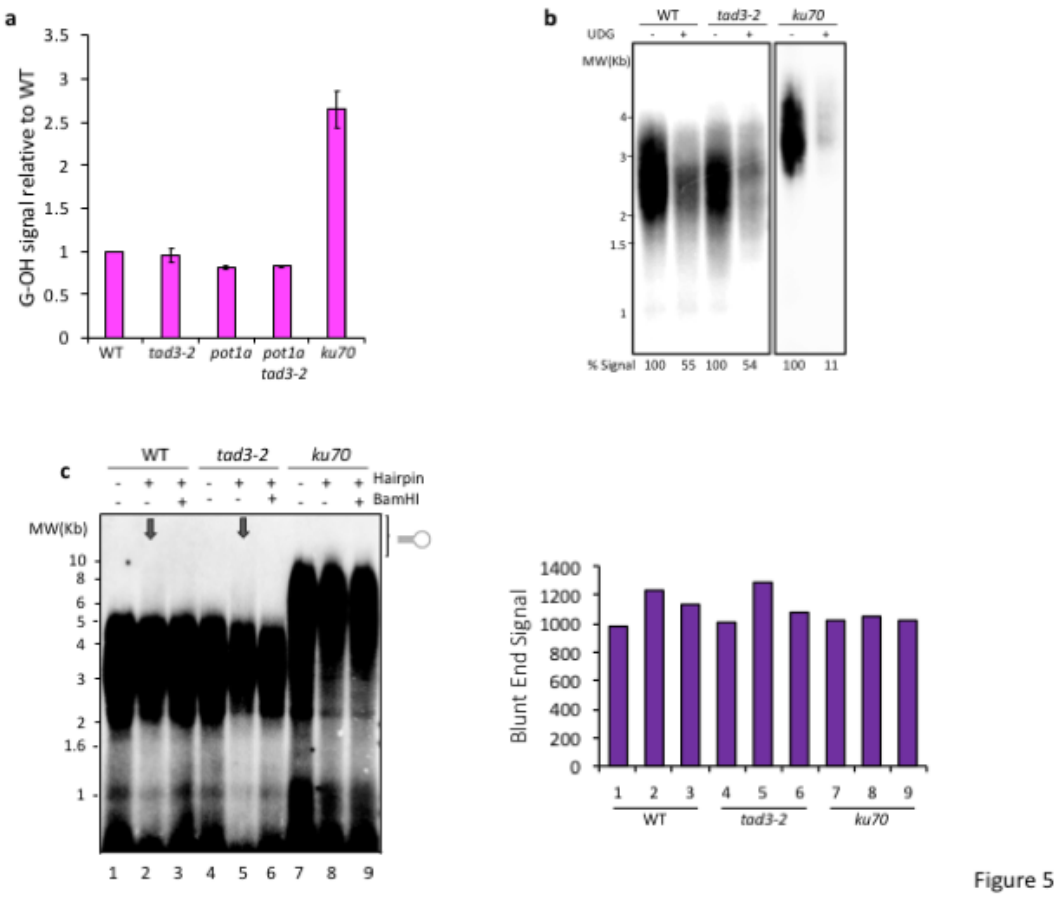


Figure 33. Loss of TAD3 does not affect the G-overhang or blunt-end architecture of telomeres.

(a) Quantification of the G-overhang assay (G-OH) performed using the in-gel hybridization technique with a radioactive probe complementary to the telomeric G-rich strand. The mean of two biological replicates are shown as fold change with respect to WT samples. Error bars indicate standard deviation. DNA from WT and *ku70* mutants serve as the negative and positive controls, respectively. (b) Results of UDG–PENT assays performed to assess telomere end architecture. The % signal was calculated using QuantityOne Software. DNA from a *ku70* mutant served as the positive control. (c) Results for a hairpin ligation assay to confirm the presence of blunt ended telomeres. For each genotype, the first lane (lanes 1, 4, 7) shows untreated telomeric DNA; the second lane (lanes 2, 5, 8) shows telomeric DNA ligated to a hairpin; the third lane (lanes 3, 6, 9) shows DNA cleaved with BamHI enzyme. The downward arrows highlight evidence of hairpin ligation. The higher molecular weight products in the *ku70* samples are expected since telomeres are elongated in the absence of Ku. The smeared area above the main hybridization signal (right side bracket) was quantified using QuantityOne and results are represented using the bar graph on the right side of the gel. Data correspond to quantitation for the reactions in each lane as described.

Loss of TAD3 impacts many cellular pathways

Given the essential role of tRNA deaminases in translation (255), TAD3 is expected to impinge on many cellular pathways. To gain insight into the global impact of TAD3 mutation, we further analyzed RNA-seq data from *tad3-2* and wild type seedlings to identify differentially expressed genes (DEGs). We used Limma-Voom on the web-based program Galaxy (270), with FDR<0.05. DEG with more than two-fold change in *tad3-2* compared with wild type was fed into G: Profiler to determine the functional enrichment of gene ontology (GO) terms. A total of 980 RNAs were differentially accumulated in *tad3-2* mutants; 598 were upregulated and 382 were downregulated. Notably, no telomere-related gene was identified as a DEG.

GO terms are categorized by Molecular Function (MF), Biological Pathway (BP) and Cellular Compartment (CC). We observed significant enrichment of GO terms in the BP category, with a large number of downregulated genes associated with auxin signaling, auxin transport, and cellular response to auxin. Other downregulated genes were associated with the cellular response to chemicals and growth, both of which are also related to auxin-related processes (Figure 34A). In contrast, upregulated genes showed significant enrichment of GO terms related to secondary metabolic processes, secondary metabolite synthesis, and particularly with the glucosinolate biosynthetic pathway (Figure 34A).

Since the transcriptional responses to zeocin in *tad3-2* mutants are almost indistinguishable from those of wild type plants, we re-examined our RNA-seq dataset in an effort to find more direct targets of TAD3 by looking at DEGs between zeocin-treated *tad3-2* and wild type seedlings. This stringent analysis resulted in 166 differentially accumulated RNAs in *tad3-2* mutants, of which 105 were upregulated and 61 were downregulated (supplementary Figure S18D). GO analysis of the new gene pool was consistent with the previous analysis: the downregulation of

a

Source	Term Name	$-\log_{10}(P_{adj})$
GO:BP	response to chemical	1.404×10^{-30}
GO:BP	response to auxin	2.867×10^{-30}
GO:BP	cellular response to chemical stimulus	2.567×10^{-29}
GO:BP	response to organic substance	3.046×10^{-25}
GO:BP	response to endogenous stimulus	2.898×10^{-24}
GO:BP	response to hormone	1.341×10^{-23}
GO:BP	cellular response to auxin stimulus	9.773×10^{-23}
GO:BP	cellular response to organic substance	3.440×10^{-22}
GO:BP	response to stimulus	4.223×10^{-22}
GO:BP	cellular response to endogenous stimulus	4.223×10^{-22}
GO:BP	cellular response to hormone stimulus	5.668×10^{-20}
GO:BP	hormone-mediated signaling pathway	2.060×10^{-8}
GO:BP	cellular response to stimulus	8.262×10^{-7}
GO:BP	growth	1.924×10^{-6}
GO:BP	signaling	6.851×10^{-6}
GO:BP	signal transduction	8.365×10^{-6}
GO:BP	cell communication	2.044×10^{-5}
GO:BP	secondary metabolic process	1.658×10^{-12}
GO:BP	secondary metabolite biosynthetic process	3.949×10^{-10}
GO:BP	S-glycoside biosynthetic process	8.074×10^{-10}
GO:BP	glycosinolate biosynthetic process	8.074×10^{-10}
GO:BP	glycosinolate metabolic process	6.038×10^{-10}
GO:BP	S-glycoside metabolic process	6.038×10^{-10}
GO:BP	glycosyl compound biosynthetic process	2.034×10^{-7}
GO:BP	sulfur compound metabolic process	1.241×10^{-6}
GO:BP	glycosyl compound metabolic process	1.498×10^{-6}
GO:BP	lipid metabolic process	7.361×10^{-6}
GO:BP	carbohydrate metabolic process	8.503×10^{-6}
GO:BP	sulfur compound biosynthetic process	9.470×10^{-6}
GO:BP	ion transport	1.110×10^{-5}
GO:BP	cell wall biogenesis	2.455×10^{-5}
GO:BP	cell wall organization or biogenesis	1.110×10^{-3}
GO:BP	transition metal ion transport	1.712×10^{-3}
GO:BP	metal ion homeostasis	3.239×10^{-3}

b

Cell Cycle Genes	TAIR ID	Fold Enrichment in *tad3-2* (FDR<0.05)
CDC6B	AT1G07270	2.81
ATTSO2	AT3G27060	2.1
CDC6	AT2G29680	2.02
CDKB1;1	AT3G54180	1.75
CDC45	AT3G25100	1.68
HAC1	AT1G79000	1.66
CDT1A	AT2G31270	1.62
MCM2	AT1G44900	1.61
MCM3	AT5G46280	1.6
MCM7	AT4G02060	1.54
MCM5	AT2G07690	1.42
MCM4	AT2G16440	1.41

Figure 6

Figure 34. Transcriptomic analysis reveals changes in auxin signaling, plant secondary metabolism and cell cycle-related genes due to loss of TAD3.

(a) Gene ontology analysis performed with differentially regulated genes in 6-day-old *tad3-2* seedlings compared to 6-day-old WT seedlings. Table contains GO term source, term name and with the numerical p-value expressed as a function of intensity of the green color. p-values greater than 10^-16 are highly significant. Red and the green arrows represent genes downregulated or upregulated in the *tad3-2* mutants, respectively. (b) Expression data for some critical cell cycle and DNA replication related genes in *tad3-2* mutants derived from transcriptome data.

auxin homeostasis and signal transduction pathway and upregulation of glucosinolate biosynthesis as the most affected processes in the hypomorphic *tad3-2* mutant (supplementary Figure S18A).

Finally, given the importance of TAD3 in cell cycle progression in fission yeast (258), we specifically looked for changes in expression of 150 critical cell cycle regulators and DNA replication factors in *tad3-2* mutants. We found that the MCM gene cluster (MCM2, MCM3, MCM4, MCM5, MCM7) exhibited a 1.5 - 1.8-fold increase in *tad3-2*, while CDC6 and CDC6B expression increased by almost 2.1-fold (Figure 34B). Both CDC6 and MCM gene clusters initiate S-phase by licensing origins for DNA replication (271, 272). Loss of TAD3 also led to elevated expression for some cell cycle regulators including CDKB11 (1.6 fold), HAC1 (1.66), CDC45 (1.69) and CDT1 (1.62) (Figure 34B). Thus, TAD3 modulates expression of numerous cell cycle related genes.

We also investigated the expression pattern of TAD3 gene expression across the cell cycle in Arabidopsis using synchronized T87 *A. thaliana* cell culture (273). Cells were treated with Aphidicolin to arrest them in G1/early S-phase. FACS analysis was done at various points after releasing the block to monitor cell cycle progression (supplementary Figure S18B) and transcript levels were measured using qRT-PCR. Although, TERT and TAD3 mRNA levels peaked during early S phase, based on statistical analysis, changes in expression levels for these genes across the cell cycle were not significant (supplementary Figure S18C). In contrast, POT1a mRNA level peaked during early S/G2 phase transition and statistical analysis revealed that POT1a expression changed significantly across the cell cycle (supplementary Figure S18D). Overall, based on the FACS analysis, TAD3 seems to be constitutively expressed across the cell cycle, possibly owing to its fundamental role in tRNA editing.

Discussion

Telomere length maintenance is essential for the stability of linear genomes. Over the past two decades, multiple genetic screens, interactome assays, and QTL mapping experiments illustrate the influence of "non-canonical" pathways in telomere length regulation. Remarkably, genome-wide studies in *S. cerevisiae* revealed that >5% of nonessential (241) and >11% of essential (242) genes are necessary for telomere maintenance. Recently, translation-related factors have emerged as critical determinants of telomere length homeostasis. One of essential gene affecting telomere length in budding yeast is YLR317W, a transcript produced from the *TAD3* locus (242). Here we demonstrate the importance of *TAD3* in telomere length maintenance in *A. thaliana*. We further show that this function is mediated by a noncanonical, telomerase-independent mechanism, highlighting the importance of cross-functional pathways in telomere biology.

Previously we described a telomerase regulatory function for the long non-coding RNA *AtTER2* encoded on the opposite strand and partially overlapping with the 5' UTR of *TAD3* (201). In considering updated *A. thaliana* genome annotation (Araport 11) showing that *AtTER2* is fully embedded into the 5' UTR of *TAD3* and the demonstration that TER1 was not the true telomerase RNA subunit (200, 206) led us to revisit the *TER2* locus using strand-specific qRT-PCR and transcriptomic analyses. We report that *TAD3* does not give rise to a stable lncRNA, and hence telomere-related functions derive from the *TAD3* gene itself.

Because a null mutation in *TAD3* leads to embryonic lethality (279), we obtained a new hypomorphic *tad3* mutant (*tad3-2)* to further explore its function in telomere biology. We discovered that in *tad3-2* mutants, the longest telomere tracts shortened progressively over successive generations, while shorter telomeres remained unchanged. A similar profile is observed

in cells haploinsufficient for key telomerase components (265-267). However, *ex vivo* qTRAP assays indicated wild type levels of telomerase activity in *tad3-2* mutants. In addition, analysis of *ku70 tad3-2* mutants revealed that telomerase can fully access and extend telomeres in plants deficient in TAD3. Strikingly, defective telomere maintenance in *tad3-2* mutants is strongly exacerbated in plants also lacking the telomerase processivity factor POT1a, with double mutants exhibiting an early onset of developmental defects and genome instability arising from telomere dysfunction. Thus, TAD3 facilitates telomere length homeostasis via a telomerase-independent pathway.

How could TAD3 promote telomere maintenance? TAD3 encodes a tRNA-editing deaminase that converts adenosine to inosine at the wobble 34 position of the tRNA anticodon loop (255). This modification expands pairing to A, U, C at the 3rd position of a codon (280, 281). I34 is critical for reading and translating C-ended codons (282) for Ala, Ser, Pro, and Thr (283). Consequently, compromising TAD3 is expected to impact many cellular pathways (284). Analysis of human transcriptome and proteome data confirm the importance of adenosine deaminases (ADATs) in translating transcripts rich in these same four codons (285). Because such translation-related data are unavailable for *A. thaliana*, we performed a transcriptome analysis on *tad3-2* mutants to examine how decreased expression of *AtTAD3* impacts plant metabolism.

Over 6000 genes are differentially regulated upon loss of TAD3, but intriguingly none are associated with known telomere pathways. Instead the genes are concentrated in two major areas with significant downregulation of the auxin signal transduction pathways and significant upregulation of the glucosinolate biosynthetic pathway. Notably, both metabolic processes intersect stress response, cell cycle regulation and DNA metabolism. Reduced auxin signaling may account for the elevated PCD in the RAM of *tad3-2* mutants in response to zeocin. Our RNA-seq

data and comet assays showed that *tad3-2* mutants mount a normal DDR and do not accumulate more DNA damage than wild type under normal conditions. Auxin inhibits PCD during plant development and in response to stress (286). Under normal conditions, auxin concentrations in root stem cell niche peak in the quiescent center and follow a local gradient at the root tip. However, in response to environmental stress, auxin levels decline, leading to PCD induction in roots (250). Thus, lower levels of auxin in *tad3-2* mutants may sensitize plants to PCD in response to stress. Alternatively, down regulation of auxin signaling may render chromatin more vulnerable to zeocin treatment. Auxin has recently been shown to increase chromatin compaction, and its inhibition results in increased DNA damage upon zeocin treatment (287).

Our transcriptomic data analyses also revealed that *tad3-2* mutants significantly upregulate genes in the glucosinolate biosynthetic pathway. Glucosinolates are secondary metabolites in cruciferous plants that serve as antimicrobials and defend against herbivory. Interestingly, glucosinolate accumulation regulates cell cycle progression in *Arabidopsis* and reduces the rate of DNA replication in wild type plants, causing cells to accumulate in S phase (288, 289). Despite the wide array of mutant phenotypes expected for TAD3 mutation, the predominant feature of *tad3* mutation in fission yeast is a cell cycle defect (258). While the changes were not as dramatic as in other metabolic pathways, we observed a surge in expression of genes that regulate cell cycle and promote DNA replication. Telomere replication and processing require a dynamic switch from a protective state to an open conformation and back again (290), and thus cell cycle perturbation can alter telomere length and terminal architecture (37, 291-293). Although we saw no obvious change in the status of G-overhangs or blunt end telomeres in *tad3* mutants, our experiments were performed on asynchronously growing seedlings. It is possible that a subtle shift in cell cycle

progression in *tad3* mutants decreases telomerase access to telomeres or increases access for nucleolytic processing enzymes, either of which would lead to telomere shortening.

We conclude that the *TAD3* locus indirectly contributes to telomere length homeostasis in *Arabidopsis* by altering the metabolic profile. Understanding precisely how cross-functional pathways influence telomere biology may shed new light on how telomeres serve as both sentinels and elicitors of physiological stress.

Material and Methods

Plant materials, genotyping and genetic complementation

Seeds for *tad3-1* (SAIL_556_A04), *tad3-2* (SALK_121147) and WT Col-0 accessions along with T87 cell culture for the Col-0 accession were obtained from the ABRC stock center. Seeds were sterilized using 70% ethanol, 10% bleach and 0.1% Triton X-100 followed by vernalization for 2 days at 4°C. Seeds were plated on half Murashige and Skoog (RPI M10500) and 1% agar (Caisson A038) supplemented with 1% sucrose. Plants were grown in soil in controlled growth chambers maintained at 22°C under long day light conditions. Photographs to assess plant growth and development were captured using a digital camera.

Genotyping (primer sequences in S1 table) was performed with leaf DNA and emerald enzyme master mix (Clontech RR310A). *pot1a tad3-2* double mutants and *ku70 tad3-2* double mutants were generated by crossing plants heterozygous for *pot1a* and *tad3-2 or ku70* and *tad3-2* followed by segregating progeny for multiple generations. For genetic complementation, 3-week-old G2 *tad3-2* plants were transformed with Agrobacterium (GV3101) cells harboring the plasmid pCBK05::NP^{TAD3}::TAD3 using the floral dip method (Zhang et al. 2006). Resistance to BASTA and Carbenicillin was used to select for true transformants in the next generation (G3).

RNA-Seq, transcriptome data visualization and analysis, and qRT-PCR

RNA extracted from 6-day-old seedlings was used to make RNA libraries in triplicate using the Illumina TruSeq® Stranded Total RNA Library Prep Plant (Catalog no. 2002061). After trimming the raw sequences using the Trimomatic program (Galaxy Europe), datasets were concatenated for each biological replicate and aligned to the *A. thaliana* reference genome sequence (TAIR10_v90) using RNA_STAR. The Bed file generated by RNA_STAR was visualized in SeqMonk to determine the density of raw reads aligning to various locations in the genome. To obtain the dataset for the Differentially Expressed Genes (DEG), the bed file was processed using the featurecounts program followed by the limma-voom software. For Gene Ontology analysis, the list was fed into G:Profiler (Reimand et al. 2007). For qRT-PCR, the Zymo Research kit (R2051) was used for RNA extraction. Strand-specific qRT-PCR was performed using cDNA synthesized from 1 µg total RNA using Super Scriptase IV (Thermo Fisher:18090050) and strand-specific primers (primer sequences in Supplemental Table I) followed by qPCR using PowerUp SyBr Green (Thermo Fisher: A25741). For non-stranded cDNA synthesis, a cDNA synthesis kit (Quanta:95047) was used with the same protocol for qPCR.

Zeocin treatment, PI staining and pollen viability assays

4- or 5-day-old seedlings grown on 0.5X MS media with 1% sucrose and 1% agar were transferred to six well plates containing MS media (Mock) or MS media plus 20 µM of Zeocin (Thermo Fisher Scientific - R25001). Plates were wrapped in aluminum foil and left on a shaker (100 RPM) for 2, 4 and 6 h. After treatment, seedlings were transferred to six well plates filled with PI stain solution (10 mg/ml; Sigma P4170) dissolved in H_20. After 30 sec, seedlings were washed in ddH_2O, transferred to slides in a droplet of H_20, sealed with a cover slip and imaged at

10X using a dsRED filter and brightfield of a Zeiss fluorescence microscope. Pollen viability was assessed as described (Li 2011). For accuracy and highest yield, the assay was performed with flowers collected between 6 AM and 8 AM. Slides containing pollen grains were imaged using a GFP filter (blue light, wavelength = 495 nm) on a Zeiss fluorescence microscope.

Comet Assay

The comet assay was performed with protoplasts using a comet assay kit from Trevigen (4250-050-K) following the manufacturer's directions with minor modifications. Protoplasts were extracted (He et al. 2007) from 6- or 7-day-old WT, *atr (*At5g40820) and *tad3-2* seedlings. A concentration of 2 x 10^5 cells/ml was used for the assay. Slides were run in an electrophoretic set up at 18 V for 10 minutes in complete darkness. After drying the agarose, slides were stained with PI stain (100 μg/ml), sealed with a cover slip and imaged using a Zeiss fluorescence microscope at 5X magnification with a dsRED filter. The parameters (Percentage DNA in tail, Tail Length, Tail Moment) were calculated using Open Comet Software (Gyori et al. 2014).

Telomere and telomerase analysis

TRF assays were performed with 3- to 4-week-old plants as described (Kobayashi et al. 2019). To obtain high quality DNA, phenol chloroform extraction was performed twice while extracting the DNA from plant tissues. Telomere length was quantified using TeloTool (Göhring et al. 2014). Telomere fusion PCR was performed using 2 μg of DNA as described [10]. Fusions were monitored between the right arm of chromosome 1 (1R) and left arm of chromosome 2 (2L), and between 1R and the left arm of chromosome 3 (3L) using primer indicated in Supplemental Table I. G-overhangs were assessed using in-gel hybridization as described previously (Riha et al.

2000) with slight modifications. Plants no older than 3 weeks were used for the assay to obtain high-quality DNA. To assess blunt end telomeres, the hairpin-ligation assay and the UDG PENT assays were performed using 150 μg of high quality DNA quantified using a Qubit Analyzer as described (Kazda et al. 2012). Quantitative TRAP was conducted as described (Song et al. 2019) with two minor modifications. Buffer W+ (1M Tris-Acetate pH 7.5, 1M $MgCl_2$, 2M KGlu, 0.5M EGTA, 30% PVP, Gylcerol, 1μM DTT, 0.6 nM VRC, 1μM PMSF) was used to suspend ground tissues (flowers or seedlings). The protein pellet was resuspended in buffer W+ supplemented with RNaseOUT (Thermo – 10777019). Debris were removed before measuring the protein concentration using Bradford reagent. Primer extension was performed with primer sequences in Supplemental Table I for 45 min at 25°C followed by qPCR using Dynamo SyBr mix (Thermo: F410L).

Anaphase bridges

Mitotic spreads from flower pistils were prepared and analyzed as described (Heslop-Harrison 1998; Surovtseva et al. 2009a). The spreads were stained with commercial DAPI solution (IHC-Tek 1W-1404), and imaged at 100X using a DAPI filter in Nikon Ti fluorescence microscope.

Cell culture synchronization and flow cytometry

T87 cell culture was maintained in NT-1 media on a rotary shaker (120 RPM) under continuous light for 24 h and every 7 days cells were subcultured into fresh NT-1 media (1:2 v/v). For cell synchronization, 5 mL of early stationary phase T87 cell suspension (7 days after previous subculture) was subcultured into 75 ml fresh NT-1 medium in a 250 mL Erlenmeyer flask. The flask was incubated at 24°C, 120 rpm under constant light for 7 days. 12 mL of the cell suspension

was transferred into 60 ml fresh NT-1 medium to achieve a dilution of 1:5. 10 mL cell suspension was cleaned by filtration through sterilized miracloth. Excess liquid was removed with a paper towel, and an aliquot of unsynchronized cells was frozen in liquid nitrogen. To block cells in G1/early S-phase 173 µl aphidicolin stock solution of 5mg/ml (Sigma Aldrich, Catalog no. A0781) was added to 72 mL of diluted cell suspension to obtain a final concentration of 12 µg/mL. The culture was incubated at 24°C, 120 rpm under constant light for 23 h. To release the block, cells were filtered through miracloth, washed vigorously with 500 ml NT-1 medium and resuspended in 60 ml NT-1 medium. Aliquots were taken at various times for DNA content analysis. The first aliquot was labeled "T0". The remaining cell culture continued to incubate at 24°C, 120 RPM under constant light and samples were collected each hour. For FACS analysis, frozen cells were transferred to a clean petri dish and 1 ml of cold homogenization buffer (25 mM PIPES (pH 7), 10 mM NaCl, 5 mM EDTA (pH 8), 250 mM Sucrose, 0.15 mM Spermine, 0.5 mM Spermidine, 20 mM -mercaptoethanol, 1% NP-40, 1 mM PMSF) was added. Cells were chopped with a razor blade to release nuclei followed by addition of 1 ml homogenization buffer. Cells were resuspended using a p1000 pipet and transferred into a new tube for 2 min. A 40 µm cell strainer (Merck or BD Falcon) was placed into a 50 ml falcon tube and the tube placed on ice. Resuspended cells were strained and collected into the cold falcon tube. Nuclei were collected by centrifugation at 7000 RPM for 20 min at 4°C then resuspended in homogenization buffer. Samples were treated with RNaseA at a final concentration of 15 µg/mL followed by incubation at RT for 10 min. Nuclei were stained with 60 µg/ml of propidium iodine (PI) and samples were run on a Becton-Dickinson FACSCalibur at 488 nm at the Flow Cytometry Core Facility, VMBS, Texas A&M University. DNA content was analyzed using CellQuest (Becton-Dickinson) and ModFit LT (Verity) programs.

Supplementary data

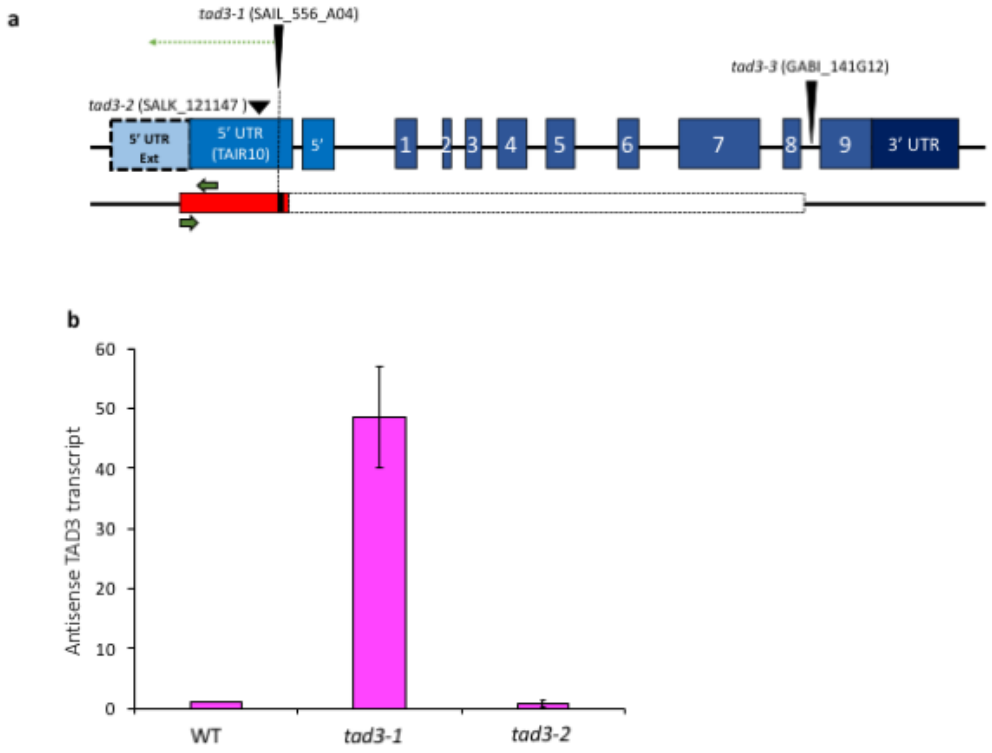
a

Gene	Cq Value
Flowers	
TER2	31.47
ACT2	19.92
Leaves	
TER2	32.40
ACT2	20.28
Seedlings	
TER2	32.39
ACT2	17.86

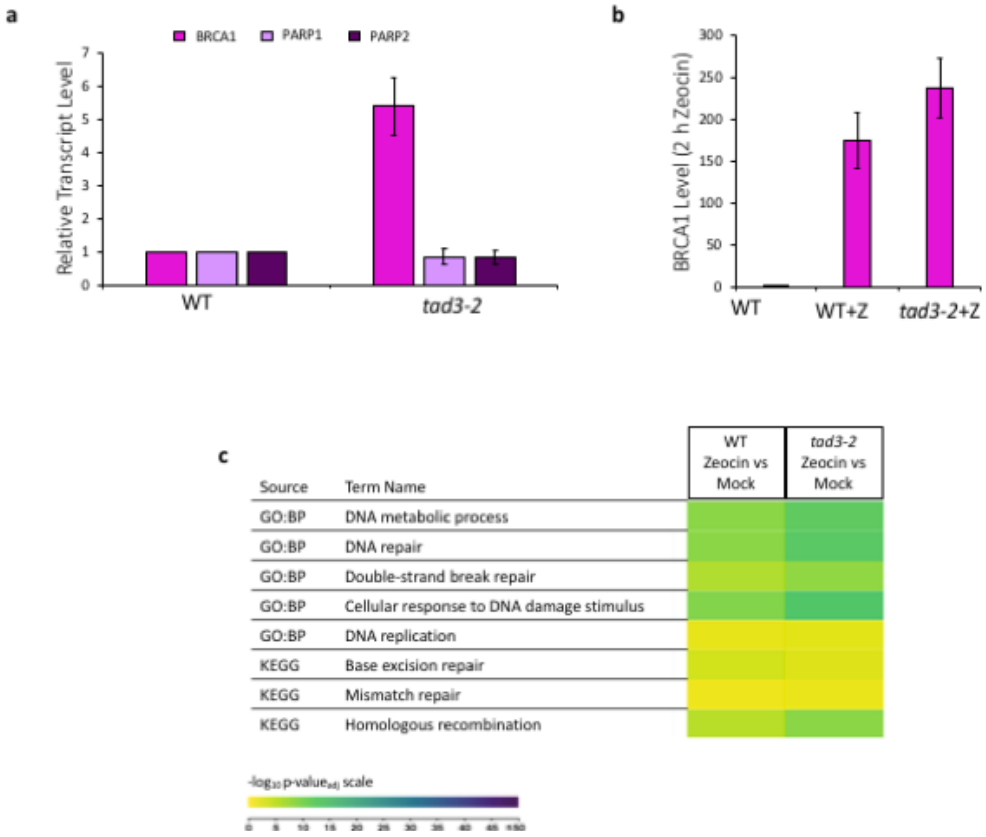
b

c

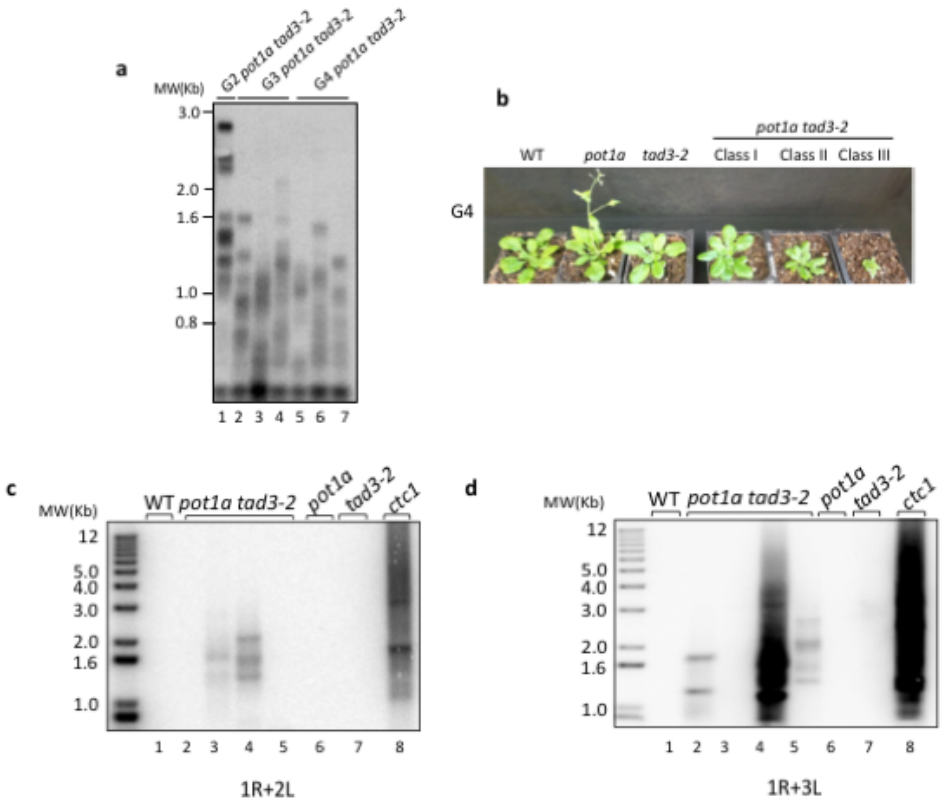
Supplementary figure S14. TAD3 mRNA expression is regulated during plant development. (a) Results from strand-specific qPCR. Cq values for *TER2* and the *ACT2* gene amplified using WT flowers, leaves and seedlings are shown. (b) Genevestigator-based analysis of TAD3 mRNA expression during different stages of plant growth and development. (c) Genevestigator analysis of organ-specific expression of TAD3 mRNA.



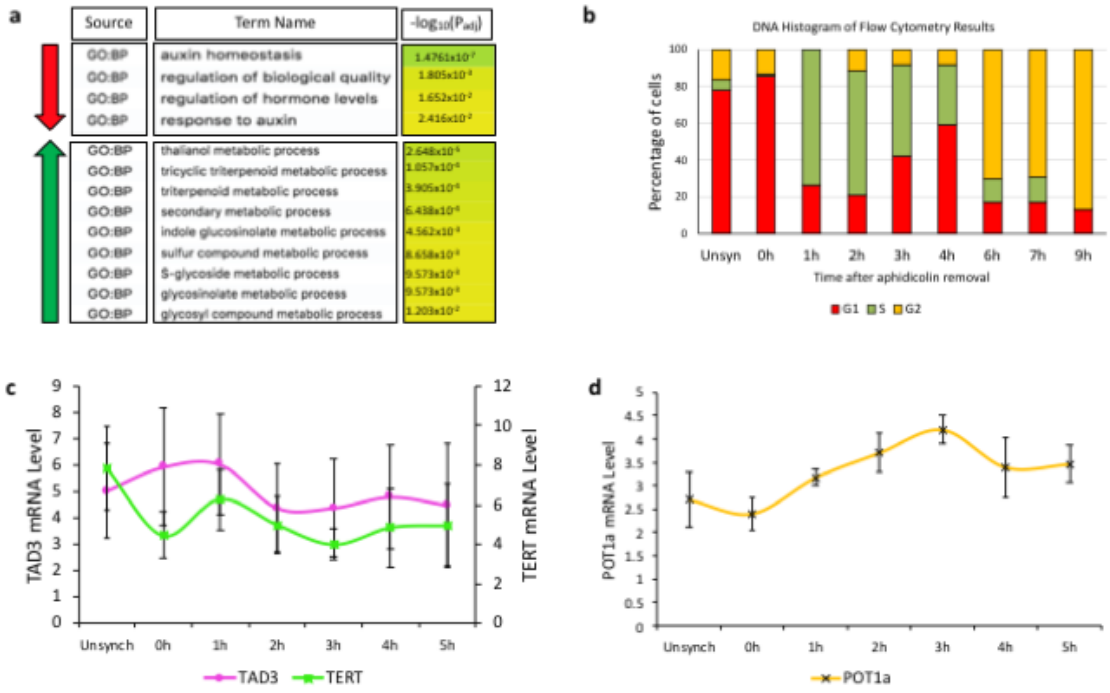
Supplementary figure S15. Identification of a cryptic transcript produced from the *TAD3* locus in *tad3-1* mutants. (a) Schematic representation of the *TAD3* locus (see legend for Figure 1). Green arrows denote forward and reverse primers used to detect expression. (b) qRT-PCR results obtained with these primers with WT, *tad3-1* and *tad3-2* samples. PCR product size = 108 nts. The mean of two biological replicates are shown as fold change with respect to WT samples**.** Error bars indicate standard deviation.



Supplementary figure S16. Results of zeocin treatment of WT and *tad3-2* seedlings. (a) Results for qRT-PCR experiments performed to detect *BRCA1*, *PARP1* and *PARP2* gene expression in *tad3-2* and WT seedlings under normal conditions. The mean of three biological replicates is shown as fold change with respect to WT samples. Error bars indicate standard deviation. (b) Results for qRT-PCR experiments performed to detect *BRCA1* gene expression samples treated with 20 µM zeocin for 2 hours. The mean of three biological replicates is shown as fold change with respect to untreated WT samples. Error bars indicate standard deviation. (c) Gene ontology analysis performed using G profiler with the genes upregulated in 6-day-old WT and *tad3-2* seedlings treated with 20 µM for 2 hours.



Supplementary figure S17. Combined loss of TAD3 and POT1a accelerates the onset of telomere dysfunction. (a) Results of TRF analysis for consecutive generations of individual *pot1a tad3-2* mutants from G2 (lane 1), G3 (lanes 2-4) and G4 (lanes 5-7). (b) Images of rosettes from three-week-old WT, G4 *tad3-2*, G4 *pot1a* and G4 *pot1a tad3-2* plants. Examples of plants from the different classes of G4 *pot1a tad3-2* mutants are shown. (c) and (d) Results from Telomere Fusion PCR assays with WT, *tad3-2*, *pot1a* and *pot1a tad3-2* samples. DNA from a *ctc1* null mutant served as the positive control. The subtelomeric primers used for PCR amplification are indicated below each blot.



Supplementary figure S18. Gene ontology analysis of WT and *tad3-2* transcriptomics data and expression of TAD3 and telomerase components across the cell cycle.(a) Gene ontology analysis performed with differentially regulated genes in zeocin treated six days old *tad3-2* seedlings compared to zeocin treated 6-day-old WT seedlings. (b) FACS data obtained from Aphidicolin-synchronized T87 cell culture. Graph shows a time course of the fraction of cells in each phase of the cell cycle post release from the drug. (c) RT-qPCR analysis of *TAD3* and *TERT* performed on the RNA isolated from synchronized T87 cell culture at different timepoints. ΔCt values were normalized to the reference gene AT4G26410. Each data point represents the mean value ± SD (n=4 independent assays). $p>0.05$ (ANOVA). (d) RT-qPCR analysis of *POT1a* performed on RNA isolated from synchronized T87 cell culture at different timepoints. ΔCt values were normalized to the reference gene AT4G26410. Each data point represents the mean value ± SD (n=4 independent assays). *$p<0.05$ (ANOVA).

Table S1

Primer Name	Primer Sequence
Genotyping Primers	
Lb1.3	ATTTTGCCGATTTCGGAAC
LB1 SAIL1	GCCTTTTCAGAAATGGATAAATAGCCTTGCTTCC
tad3-1_LP_SAIL	GACGACAACTAAACCCTACGCTTACA
tad3-1_RP_SAIL	CGATGTTGTTTTTCTGCTTAGGACACA
tad3-2_LP	ATGGTCAGGTGACAATGAAGG
tad3-2_RP	ACCTTAGCCACTAACAACCCC
ku70_LP	TTACTTTGTTGTTTCGGGTGC
ku70_RP	CTCTTGGCAAGTACACGCTTC
pot1a_LP	ATGGCGAAGAAGAGAGAGAGTCCCA
pot1a_RP	CTATCTTGATCTCTCTCAAGAAGGA
pot1a_T-DNA	CATTTTATAATAACGCTGCGGACATC
ctc1_LP	GGTGCTTGAGAAGATGTTTGC
ctc1_RP	ACTTTTAAATTCGCAGGGTGG
qPCR Primers	
TAD3_F	TAGATGGCATCGGTGGGTTCAG
TAD3_R	TCACTGCAGCATTCACAAGTGG
TER2_F	CGCTTACATAAAAACGCGACC
TER2_R	CACGTCTCTCTCTTACGTCGTC
TER2_3'_F	CAAACTTGTGTCCTAAGCAG
TER2_3'_R	GCAAACTGATTCCTGAAGTTGG
TERT_F	ACCGTTGCTTCGTTGTACTTCACG
TERT_R	CGACCCGCTTGAGAAGAAACTCC
POT1A_F	TTCCGTCACCTCCTGGATTAACAG
POT1A_R	TCTGTACTTGCATGTAACCTTGGG
Fusion PCR Primers	
1R	CTATTGC CAGAACCTTGATATTCAT
1L	ACAAGGATAGAAATAGAGCATCGTC
2R	CAACATGGCCCATTTAAGATTGAACGGG
3L	CATAATTCTCACAG CAGCACCGTAGA
4R	TGGGTGATTGTCATGCTACATGGTA
Others	
qTRAP_F Primer	CACTATCGACTACGCGATCAG
qTRAP_R Primer	CCCTAAACCCTAAACCCTAAA
UDG-PENT Primer	CCCTAAACCC TAAA
HairPin Assay Oligo	P-GGATCCGACTTTTGTCGGATCC
4X G-probe	(TTTAGGG)X4
4X C-probe	(CCCTAAA)X4

Table S2

Zeocin WT DEG	Zeocin *tad3-2* DEG	Common Genes (77 out of 100)
AT3G27060	AT3G27060	YES
AT5G48720	AT5G48720	YES
AT5G40840	AT4G21070	YES
AT2G21790	AT5G40840	YES
AT4G21070	AT2G21790	YES
AT5G24280	AT2G18193	YES
AT2G18193	AT5G24280	YES
AT1G17360	AT1G17360	YES
AT5G48020	AT5G48020	YES
AT2G31320	AT4G02390	YES
AT1G08260	AT5G23910	YES
AT5G23910	AT1G08260	YES
AT4G02390	AT2G31320	YES
AT4G02110	AT4G02110	YES
AT1G31280	AT2G30362	YES
AT2G30362	AT2G30360	YES
AT2G30360	AT1G31280	YES
AT1G49980	AT5G05730	YES
AT3G10500	AT1G49980	YES
AT2G45460	AT3G10500	YES
AT5G66130	AT5G49480	YES
AT5G49480	AT2G45460	YES
AT5G07610	AT5G66130	YES
AT4G35740	AT4G35740	YES
AT3G07800	AT3G58270	YES
AT3G58270	AT3G07800	YES
AT5G49110	AT5G07610	YES
AT4G01450	AT1G18570	YES
AT4G13370	AT3G45730	YES
AT1G51130	AT3G52115	YES
AT1G20750	AT3G13080	YES
AT3G52115	AT4G01450	YES
AT5G66140	AT2G41630	YES
AT2G47680	AT4G13370	YES
AT3G13080	AT2G47680	YES
AT4G37490	AT4G37490	YES
AT5G51580	AT1G51130	YES

AT1G56510	AT2G18690	YES
AT5G22010	AT5G49110	YES
AT5G51585	AT1G56510	YES
AT5G07165	AT5G55490	YES
AT5G65360	AT4G19130	YES
AT1G77120	AT1G09815	NO
AT5G55490	AT3G16530	YES
AT3G01680	AT2G26560	YES
AT4G19130	AT5G51580	YES
AT5G20850	AT4G39940	YES
AT5G27030	AT5G20850	YES
AT3G21850	AT5G51585	YES
AT1G09815	AT5G22010	YES
AT2G41630	AT5G66140	YES
AT5G51730	AT2G36780	YES
AT5G19890	AT5G60250	NO
AT1G17460	AT1G02920	YES
AT4G22960	AT1G59660	YES
AT4G25330	AT1G20750	YES
AT1G59660	AT5G27030	YES
AT3G45730	AT3G16525	YES
AT3G42860	AT4G25330	YES
AT4G34510	AT5G65360	YES
AT2G36780	AT5G11460	YES
AT3G25250	AT5G41750	YES
AT2G37240	AT3G01680	NO
AT5G11460	AT1G17460	YES
AT5G41750	AT3G54640	YES
AT4G24610	AT4G22960	NO
AT1G49050	AT3G42860	NO
AT2G30250	AT4G37030	YES
AT5G64060	AT5G51730	NO
AT5G18270	AT2G30250	YES
AT5G60250	AT1G21130	YES
AT5G43630	AT3G21850	NO
AT5G15540	AT4G27280	YES
AT3G27630	AT3G25250	YES
AT4G11740	AT5G07165	YES
AT5G03780	AT5G15540	YES

AT3G26720	AT2G23830	NO
AT3G09020	AT2G23150	YES
AT1G07500	AT2G41105	NO
AT3G20490	AT3G09020	YES
AT4G37030	AT1G57990	YES
AT1G78150	AT4G11740	NO
AT3G21860	AT1G26380	NO
AT5G58070	AT5G58070	YES
AT1G52315	AT3G53280	NO
AT1G18570	AT3G57550	YES
AT3G53280	AT5G03780	YES
AT1G08270	AT5G57220	YES
AT2G18600	AT1G17290	NO
AT5G47950	AT1G08270	NO
AT3G47540	AT5G18270	NO
AT4G33420	AT3G27630	NO
AT4G29170	AT4G12720	NO
AT1G62570	AT4G23140	NO
AT5G56780	AT2G41100	NO
AT2G23150	AT3G20490	YES
AT1G60500	AT3G13235	NO
AT1G12240	AT4G34510	NO
AT5G52750	AT5G25250	NO
AT1G24150	AT3G15356	NO

GeneID	**logFC WT**	**logFC *tad3-2***	**Absolute difference**
AT3G27060	6.33596894	5.54584746	0.790121478
AT5G48720	5.08095859	5.224730517	0.143771928
AT5G40840	4.3658556	3.928001004	0.437854601
AT2G21790	1.9802578	1.973281523	0.006976278
AT4G21070	6.02354949	5.568674609	0.45487488
AT5G24280	3.52034777	3.443262572	0.077085198
AT2G18193	7.38149903	6.489068328	0.892430701
AT1G17360	2.89987645	2.694787199	0.205089246
AT5G48020	2.23715482	2.427748515	0.190593696
AT2G31320	3.32849885	3.33938652	0.010887674
AT1G08260	2.8202274	2.82334538	0.003117982
AT5G23910	3.33036395	3.198746202	0.131617748
AT4G02390	5.93270397	5.395332281	0.537371689

AT4G02110	4.49367068	4.009913727	0.483756953
AT1G31280	2.74123515	2.873114407	0.131879261
AT2G30362	2.40441714	3.091197264	0.686780126
AT2G30360	2.40601684	3.091553483	0.685536644
AT1G49980	2.15217358	2.315270168	0.163096591
AT3G10500	1.52784254	1.556518282	0.028675746
AT2G45460	2.94637682	2.678912996	0.267463826
AT5G66130	3.90019117	3.34883069	0.551360475
AT5G49480	2.58863507	2.92338817	0.334753103
AT5G07610	3.94294741	3.998143956	0.05519655
AT4G35740	1.9971104	2.256395853	0.259285449
AT3G07800	5.16748476	5.432171096	0.264686336
AT3G58270	3.90205834	3.900138314	0.001920021
AT5G49110	3.49728931	3.161364354	0.335924954
AT4G01450	1.93869019	1.670933452	0.267756742
AT4G13370	2.92121252	2.680503554	0.240708966
AT1G51130	2.18925275	2.157547014	0.031705738
AT1G20750	7.0822289	6.552148748	0.530080151
AT3G52115	4.98208898	4.343039776	0.6390492
AT5G66140	1.37176657	1.313770847	0.057995719
AT2G47680	2.55092291	2.599251016	0.048328105
AT3G13080	1.04732894	1.221302792	0.173973854
AT4G37490	4.18816989	3.783795798	0.404374096
AT5G51580	3.84620323	4.037953465	0.191750231
AT1G56510	1.4769957	1.549340867	0.072345167
AT5G22010	1.41207653	1.418890082	0.006813555
AT5G51585	3.84497895	4.03697566	0.191996712
AT5G07165	9.36219987	8.007989538	1.354210337
AT5G65360	1.77136774	1.629030165	0.14233757
AT5G65360	1.77136774	1.629030165	0.14233757
AT5G55490	4.1493767	4.320930316	0.17155362
AT3G01680	1.35194707	1.297228408	0.054718659
AT4G19130	3.96219636	3.802396569	0.159799789
AT5G20850	5.4516323	5.391811721	0.059820574
AT5G27030	1.48427226	1.641351574	0.15707931
AT3G21850	5.4470332	5.54373238	0.096699181
AT1G09815	1.48647529	1.839419918	0.352944629
AT2G41630	1.0434402	1.34856621	0.305126009
AT5G51730	1.99431997	1.918514447	0.075805527
AT1G17460	2.4975162	2.532098907	0.034582711

AT4G22960	6.73949959	7.177493687	0.437994095
AT4G25330	6.75938742	7.546242788	0.786855372
AT1G59660	2.77247195	3.432443503	0.659971553
AT3G45730	1.90157468	3.03869228	1.1371176
AT3G42860	3.32223161	3.037423465	0.284808143
AT4G34510	6.55618871	5.664327444	0.891861264
AT2G36780	2.33328821	3.19228883	0.859000624
AT3G25250	4.21262146	4.238492757	0.025871296
AT5G11460	2.4975162	3.301707	0.804190804
AT5G41750	1.195776	1.506071813	0.310295814
AT2G30250	1.19279025	1.437269935	0.244479683
AT5G18270	1.90457996	2.012553302	0.107973347
AT5G60250	2.23145401	3.10991683	0.878462818
AT5G15540	0.985351	1.164683733	0.179332737
AT3G27630	7.24626581	7.097035835	0.149229978
AT4G11740	1.24524305	1.263880751	0.018637699
AT5G03780	4.53414402	4.200952919	0.333191098
AT3G09020	2.09802126	2.281350399	0.183329134
AT3G20490	1.83086588	1.998978725	0.168112843
AT4G37030	3.1093264	3.560827516	0.451501112
AT5G58070	0.83785281	1.057429219	0.21957641
AT1G18570	0.67616887	1.318192843	0.642023975
AT3G53280	1.05056386	1.327788781	0.277224925
AT2G23150	1.12331053	1.591323705	0.468013176

CHAPTER V

SUMMARY AND FUTURE DIRECTIONS

Dissertation Overview

Studies of telomerase intensely focus on characterization of telomerase RNP complexes in terms of biogenesis, subunit interactions, structures, and regulations, which all provide critical implications for cellular aging and cancer. Numerous publications relevant to ciliate and vertebrate telomerase define telomerase holoenzymes with significantly distinct subunit identities. Ciliates use RNA Pol III for TR transcription leading to production of a short TR with a limited scaffold for accessory protein associations (102). Most accessory proteins, including an RPA-like CST complex, directly interact with TERT (80). In contract, vertebrate TR is RNA Pol II transcribed. It is a relatively large molecule harboring an independent H/ACA RNP that facilitates TR biogenesis, stability, and complex assembly (143). Such dramatic divergence of TR and accessory proteins raises many questions about telomerase evolution, including which complex is ancestral and how species-specific accessory proteins are adopted.

Knowledge relevant to plant telomerase complex is limited. Telomerase is a RNP complex consisting of catalytic subunit TERT, RNA subunit TR, and accessory proteins (18). Arabidopsis TERT has been characterized to serve a conserved function in telomere maintenance as shown by its counterparts in vertebrates and ciliates (197). In contrast, TR is more divergent and species-specific. Previously, two long non-coding RNA, AtTER1 and AtTER2, have been identified as telomerase associated RNA in *A. thaliana* (198). AtTER1 was proposed to serve as the functional template for telomerase, while AtTER2 negatively regulated telomerase activity in response to DNA damage (198, 201). However, recent publications from two independent groups indicated that AtTER1 is not required for telomerase activity and telomere maintenance *in vivo* (199, 200),

which prompted an urgent need to define the *bona fide* TR in the plant kingdom. In chapter II, we found that AtTER1 is not a stable transcript. The co-purification of AtTER1 with active telomerase might result from the contamination of mRNA RAD52 overlapping with AtTER1 (Chapter II, supplementary figure 1). Chapter II also demonstrates the identification of the *bona fide* TR in Arabidopsis (AtTR) with a model of RNA secondary structure. An independent group took bioinformatic approaches and consistently uncovered AtTR as the functional telomerase RNA in plants (199).

In addition to TERT and TR, the telomerase associated proteins and their functions in plants are mostly unknown. AtPOT1a and dyskerin have been previously described to associate with active telomerase and to be responsible for optimal telomerase activity and telomere maintenance (202, 205). However, lacking the knowledge of TR significantly restrained any detailed characterization of their interactions with telomerase RNP complex. Therefore, Chapter II uncovered the *bona fide* AtTR with a reliable secondary structure (Chapter II, Figure 3), which significantly promote the studies of telomerase associated proteins in Arabidopsis. We found conserved RNA motifs responsible for TERT binding and, interestingly, a plant-specific three-way junction vulnerable to accessory protein interactions. This information leads us to examine the interactions between AtPOT1a, dyskerin, AtTR and TERT in chapter III. We revealed that AtPOT1a physically interacts with TERT and bridges TERT to substrate telomeric DNA to enhance telomerase repeat addition processivity (RAP). More importantly, we found that dyskerin specifically interacts with the three-way junction of AtTR to promote telomerase RAP *in vitro* and to secure AtTR stability *in vivo*. Solving an 8.6 Å cryo-EM structure provides a detailed model for this non-canonical interaction (Chapter III, Figure 4). This examination of AtPOT1a and dyskerin

interacting telomerase provides a clearer interpretation of plant telomerase RNP complex for the function of telomere maintenance.

In parallel, we aimed to define the complete list of telomerase core-associated proteins in Arabidopsis. Appendix reported an unbiased strategy to combine telomerase purification with the sensitive quantitative mass spectrometry (qMS) technique. We found a genuine La family protein AtLa1 significantly enriched from positive samples compared to negative controls (Appendix, Figure 2A). In addition, a chloroplast chaperone AtCPN60B was distinguished from qMS analysis. Both AtLa1 and AtCPN60B are consistent between two replicates (Appendix, Figure 2B). This study provides a framework to uncover telomerase-associated proteins and hints for telomerase RNP components.

As described above, AtTER1 and AtTER2 were identified simultaneously as telomerase-associated lncRNA with high sequence identity (198). Because AtTER1 was reported to not be required for telomerase activity and telomere maintenance *in vivo* (199, 200), it is reasonable to reevaluate AtTER2 for its function in telomerase activity regulation (201). In chapter IV, we revealed that AtTER2 is not a stable transcript and is completely embedded within the gene encoding tRNA Deaminase 3 (*TAD3)*. Unexpectedly, TAD3 dysfunction led to progressive loss of telomeric DNA over successive generations. Further examination suggested that TAD3 indirectly contributes to telomere length homeostasis by manipulating the metabolic profile without altering telomere terminal architecture and telomerase enzymatic activity.

In this dissertation, I extend the study of telomerase RNP into the plant kingdom. I demonstrate that plant telomerase RNP complex resides as a critical intermediate in telomerase evolution by maintaining a ciliate-like Pol III transcribed TR with vertebrate-like accessory

proteins. This dissertation unites the disparate lineage of telomerase and provides new insights into how RNA and protein subunits co-evolved in RNP complexes.

Plant TR is more than a DNA template

Chapter II reports identification of the Arabidopsis TR and how a secondary structure model of TR from the plant kingdom was derived. Using an unbiased approach of active telomerase purification followed by RNA sequencing, AtTR was identified as the first *bona fide* plant telomerase RNA. This important breakthrough enabled the subsequent identification of 85 TR from three major clades of the plant kingdom. Bioinformatic analyses and chemical probing revealed a conserved secondary structure for plant TR consisting of two essential domains: template-PK and STE. Analysis of the plant PK domain revealed a unique evolutionary connection that linked ciliate and vertebrate TR.

Chapter III focused on the biological significance of a plant-specific RNA motif. These studies revealed that a long-range interaction giving rise to the P1a stem of AtTR enabled the H/ACA RNP component dyskerin to bind AtTR, to stimulate telomerase repeat addition processivity *in vitro*, and to stabilize AtTR *in vivo*. Strikingly, the rest of the plant TR harbors both conserved and, more excitingly, plant-specific structural elements with uncharacterized functions. Examining those plant-specific structural elements with characterizations of their function will improve the understanding of plant telomerase which serves as a unique model in TR evolution. It will also provide insights into the TR-TERT coevolution and help to interpret the maximum flexibility of TR structural variants that TERT is compatible with.

The A. thaliana *TR template contains an unusual wobble base-pair*

Unexpectedly, my analysis of the template alignment region within AtTR revealed an A to G substitution at position 52 (Chapter II, Figure 3). The same change was identified in a group of vertebrates including mouse and hamster (87). This observation raises an interesting question of whether maintaining an unusual wobble base-pair is beneficial for telomerase activity and repeat addition processivity (RAP).

The G-T wobble base-pair that forms when telomeric DNA binds to this RNA residue consists of two hydrogen bonds similar to an A-T base-pair. Since there is no evidence that a ribonucleotide G-deoxyribonucleotide T base-pair is more stable than a ribonucleotide A-deoxyribonucleotide T base-pair, the driving force for retention of this nucleotide substitution in AtTR is unknown. If this G-T wobble base-pair is included, I hypothesize that *A. thaliana* contains a 7-nt alignment region and 7-nt templating region. Compared to the Tetrahymena (3-nt alignment and 6-nt templating) and human TR (5-nt alignment and 6-nt templating) molecules, Arabidopsis would therefore require more energy to dissociate the total 14 base-pairs in the process of telomerase translocation. Maintaining a less stable G-T wobble base-pair might therefore provide advantages for DNA substrate dissociation. This hypothesis is supported by the TR of plant species *Selaginella kraussiana*, which maintains a 4-nt alignment sequence without the G-T wobble base-pair. Therefore, the G-T wobble base-pair might only provide advantages when embedded in an extended alignment sequence. *In vitro* telomerase reconstitution assay would be an effective system to test this hypothesis. To this end, one could generate AtTR variants with different alignment lengths and numbers of wobble nucleotides and then examine the activity and repeat addition processivity of telomerase. I expect to observe reduced activity when the G-T wobble base-pair is substituted with a canonical A-T base-pair. In addition, a single molecule-based microscopy technique was recently applied to human telomerase to directly measure the speed of

telomerase engaging substrate DNA (294). This type of approach could be very useful to study the functional implications of the G-T wobble base-pair in AtTR.

A. thaliana *TR contains a STE with a novel structure*

The second essential domain of TR is a 'stem terminus' element (STE), which specifically interacts with TERT to stabilize the RNP and reconstitute telomerase activity in trans together with the T-PK domain (97). Interestingly, the structural arrangements of STE are distinct between species. Vertebrates have a three-way junction mediated STE, while ciliates maintain a compact stem loop structure (85). In addition, structural studies showed that STE from vertebrates and ciliates interact with different domains of TERT (110), which raises the question of the evolution of STE structure and STE-TERT interactions. Characterizing plant STE might provide an answer to unit them.

In plant TR, a conserved P4-P5-P6 stem has been identified as the functional equivalence of STE (Chapter II, Figure 4). However, the architecture of plant STE differs from other known STE domains as it forms two additional single-stranded loops rather than a three-way junction (Chapter II, Figure 3, P4-P5-P6). Most of the nucleotide residues within loops are conserved across plant TRs, which suggests a functional importance for this structure. It is unclear which region of Arabidopsis STE is responsible for TERT binding. However, functional dissection of the STE could be conducted in telomerase reconstitution assays with truncated STE in trans together with the T-PK domain. Solving a high-resolution cryo-EM structure for Arabidopsis telomerase holoenzyme should be the ultimate goal to reveal structural arrangements and interactions of the plant telomerase RNP.

A single-stranded sequence within the PK domain of AtTR is required for optimal telomerase activity and telomere maintenance

Pseudoknot (PK) is one of the most important RNA elements of TR responsible for RNP assembly. However, cryo-EM structural studies of human and Tetrahymena telomerase holoenzymes revealed limited interactions between PK and TERT (72, 79). This observation raises a question of which region of PK is more important for its function. Chapter II showed that the plant PK shares features of both vertebrates and ciliates (Chapter II, Figure 5). Therefore, *A. thaliana* serves as an ideal model for PK characterization.

Genetic complementation assays demonstrated that removing the AtTR PK domain (U6::dPK) completely abolished telomerase activity, while substituting unconserved nucleotides within the single-stranded loop connecting P2 to P3 (nt. position: 114-118) (U6::PK_{sub}) resulted in a partially functional TR (J. Song, B. Aklilu and D. Shippen, unpublished data). This result indicates that the unconserved P2-P3 loop is important for PK function *in vivo*. We then tested if PK_{sub} mutation changed AtTR structure. MaPseq experiments measuring residue accessibility revealed that the AtTR PK_{sub} maintains an overall wild-type (WT) structure. (J. Song, Z. Ma and D. Shippen, unpublished data). However, although PK_{sub} maintains a WT-like stem-loop in the PK region, we found a significant reduction of triple helix resulting in more dissociated stem-loop. This result suggests that substituting nucleotides as position 114-118 leads to an instable triple helix interaction. Finally, we asked whether PK_{sub} mutation impacts TR stability. RT-qPCR measurements of transgenic lines indicated that the partial complementation of U6::PK_{sub} was not caused by changes in RNA abundance or stability (J. Song and D.Shippen, unpublished data). This result was unexpected because substituting unconserved nucleotides that are not predicted to change TR structure should not perturb TR function unless the sequence is biologically important.

Further examination of AtTR PK_{sub} will be necessary to explain this observation, and may increase understand RNA folding and TERT-PK interaction in vivo.

Defining a minimal TR core and an ancestral TR in the plant kingdom

In budding yeast, TR is transcribed by RNA Pol II to generate a large telomerase scaffold, longer than 1000 nucleotides. The secondary structures of yeast TR are hard to predict because of this large size and flexibility (131). Nevertheless, functional examination and structural predictions revealed a central core region that contains the template-PK domain and three long helical arms emanating from the core. TR binding proteins including Ku heterodimer and Sm/Lsm proteins were shown to recognize the distal stem-loop elements of three helical arms (131). This finding raised the question of whether it is essential to maintain a complete TR scaffold. Truncation experiments produced a streamlined 500 nt version of yeast TR, called mini-T RNA, which contained only the central core and the protein-binding motifs (171). Mini-T is able to maintain telomeres *in vivo* and reconstitute telomerase activity *in vitro*, indicating that the most of yeast TR is unnecessary for telomerase function. The evolutionary reason for yeast to retain such a large TR is unknown, but it is possible that the segments of the RNA skeleton are involved in alternate non-telomerase functions.

One observation that leads us to examine core elements within Arabidopsis TR is our unexpected discovery of a plant-specific single-stranded RNA element connecting P1b and P1c. this structure is conserved in all the plant TR we identified, although this element varies in length and sequence (Chapter II, Figure 3). Interestingly, preliminary data from our collaborator's lab indicates that the P1b-P1c linker is not required for telomerase activity in reconstituted telomerase in vitro (J. Chen, personal communication). In addition, we found that this region is not required

for dyskerin association with AtTR (Chapter III, Figure 4). Clearly, evolutionary pressure has been applied to the P1b-P1c linker so that it can be maintained in all three of the major clades of plants. In the future, *in vivo* genetic complementation assays could help to decipher the biological relevance of this element, and to explore the possibility that it functions in one of the alternative 'moonlighting' functions of telomerase (295).

One important goal in defining the TR core in Arabidopsis telomerase is to generate a model for the ancestral plant TR. In Chapter II, several RNA elements including P1.1, P2.1 and P2.2 were identified as conserved in certain plant species (Chapter II, Figure 3). If a secondary structure model can be defined for the ancestral plant TR, it will be interesting to test whether the RNA elements we now designate as plant-specific evolved independently or were originally present in some common ancestor but then lost when plant and animal lineages diverged. A recent study by the Fajkus lab extends our knowledge of plant TR into monocot species (206). Further analysis of TR in divergent plant species will assist in defining nucleotide conservation and base-pair coevolution, which serves as the gold standard for RNA structural modeling and ancestor prediction.

Exploring the interaction of dyskerin with TR and other RNA targets in plants

Dyskerin is an RNP maturation factor conserved from yeast to mammals (232). Mammalian dyskerin is best known for its role in processing RNA polymerase II derived snoRNAs, but it is also required for telomerase biogenesis and directly contacts TR via a conserved H/ACA box motif (140). In Chapter III, we provide data showing that although AtTR does not harbor an H/ACA motif, dyskerin physically associates with AtTR through its plant-specific three-way junction. In this study, we examined the function of dyskerin interaction with AtTR by disturbing

a key component of the dyskerin binding site in AtTR, the P1a stem. We found that the absence of P1a abolished the dyskerin-mediated stimulation of telomerase repeat addition processivity *in vitro* and significantly impacted AtTR stability *in vivo* (Chapter III, Figure 4). A caveat of our experiment is that we cannot rule out the possibility that P1a is involved in additional functions besides dyskerin recognition. One way to address this is to study loss-of-function dyskerin mutants that fail to bind AtTR. Such a mutant would provide a valuable system to test the function of dyskerin in TR biogenesis, TR translocation, and RNP assembly. Finally, our data indicate that Arabidopsis dyskerin can bind RNA structures that do not conform strictly to the H/ACA motif found in mammalian TR. Therefore, it would be interesting to investigate how the dyskerin-RNA interface is evolving by looking for non-TR RNA targets of dyskerin in *A. thaliana* through RIP-seq or CLIP-seq. Such studies could provide new insight into the role of dyskerin in the telomerase RNP and the evolution of RNA recognition by this essential protein complex.

Examining the function of Arabidopsis dyskerin in AtTR biogenesis and RNP assembly

Dyskerin is an essential gene in plants as it is mammals (296), making it challenging to study the function of this interesting telomerase-associated protein. The Shippen lab previously reported construction of a bi-allelic dyskerin loss-of-function mutant to study the function of dyskerin in telomere maintenance (205). Unfortunately, the seeds for this mutant are no longer viable, and therefore a new dyskerin mutant must be created.

Using CRISPR-Cas9 technology it should be possible to generate a viable loss-of-function dyskerin mutant (297). Once that mutant is available, we can test if dyskerin is important for AtTR stability using northern blotting or qPCR. The mechanism of 3’ maturation of AtTR could be gauged by 3’ RACE or RACE-seq. Notably, there are four polyT terminator sequences downstream of the mature 3’ end of TR (214). Thus, it is possible that dyskerin association at the

proximal polyT sequence within the nascent transcript ensures that extended transcripts are degraded. We can also assess the contribution of dyskerin in AtTR biogenesis by measuring the fraction of AtTR molecules recovered from anti-TERT IP in dyskerin mutants. If TR recovery is reduced relative to IP in wild-type plants, the results would indicate dyskerin promotes RNP assembly. In parallel, one could attempt to identify dyskerin mutations that exclusively disrupt TR-dyskerin binding without disturbing core telomerase activity *in vitro*. These mutations could be used for *in vivo* rescue experiments to determine how the loss of dyskerin interaction with telomerase impacts TR abundance, telomerase processivity and telomere maintenance.

Examining how dyskerin stimulates telomerase repeat addition processivity

Telomerase is strictly regulated for its activity and repeat addition processivity in response to different stations of cycle cell and developmental stages (18). Therefore, examining the mechanism of how telomerase is regulated will facilitate the interpretation of cellular dynamics. It will also provide insights into therapeutic targets design for dysregulation of telomerase. In Chapter III, I provide evidence that dyskerin stimulates the repeat addition processivity of telomerase *in vitro*. While the mechanism of stimulation is unknown, I can propose two hypotheses.

First, dyskerin could act as a TR chaperone, facilitating conformational changes to favor TERT association as has been shown for Tetrahymena P65 (116, 118). Alternatively, dyskerin may promote enzyme activity after the TERT-TR complex is formed. To distinguish between these two hypotheses, *in vitro* "order-of-addition" reconstitution experiments can be used. In the 'dyskerin first' experiment, AtTR would be pre-incubated with dyskerin before TERT is added. *In vitro* folded AtTR would be incubated with *E. coli* expressed dyskerin prior to TERT addition

(expressed in RRL). RNP complexes could then be enriched by anti-TERT IP followed by direct telomerase activity assay. For the 'TERT first' experiment, AtTR should be pre-incubated with TERT-programmed RRL lysate. Then dyskerin would be added at the same concentration as AtTR followed by anti-TERT IP and direct telomerase activity assay. If telomerase activity increases when dyskerin is added to preassembled TERT-TR complexes, the results would imply that dyskerin plays a direct role in stimulating enzyme activity. This outcome would be fascinating as our preliminary data indicate that dyskerin does not directly contact TERT.

Alternatively, if telomerase activity is increased when dyskerin is pre-incubated with AtTR, dyskerin may assist in TR folding to favor TERT binding. The Tetrahymena P65 protein bends a 3' long terminal stem that is analogous to P4-P5-P6 in AtTR (118). Thus, Arabidopsis dyskerin, through its interaction with the three-way junction, might alter the conformation of this element to favor TERT binding. We could test this hypothesis using fluorescence resonance energy transfer (FRET) with a strategy analogous to the one used for Tetrahymena (116). Specifically, we could label residue 211 with Cy3 and residue 140 with Cy5. If dyskerin association causes a conformational change in AtTR, FRET signal distribution is expected to shift. Tetrahymena TR and P65 could serve as a positive control to verify the technique. If we observe a positive result, we could then generate truncations in dyskerin to determine if RNA binding and chaperone functions can be distinguished. Evidence for an RNA chaperone function for dyskerin would set the stage for scientists to assembly and function of larger, more complex telomerase RNP particles like human telomerase.

Examining how the dyskerin-RNA interface is evolving in plants

Unlike TR, dyskerin is a highly conserved molecule (296), making the non-canonical dyskerin-AtTR interaction in Arabidopsis all the more surprising. In chapter II, we showed the 3' three-way junction element that is responsible for dyskerin association is maintained among all identified plant TR (Chapter II, Figure 3), suggesting that dyskerin recognizing this special RNA element might be much more conserved in plants. This observation raises a question of whether dyskerin interacts with other substrates mediated by a similar three-way junction mechanism. The dyskerin-RNA interface could evolve in plants to include broader substrates.

To interstage this hypothesis, I propose to develop an unbiased approach to identify a complete list of RNA targets bound by dyskerin in Arabidopsis. I propose that RIP-seq or CLIP-seq be performed with Arabidopsis expressing Twin-Strep tagged dyskerin. Co-purified RNAs could be sequenced, quantified, and analyzed to identify significantly enriched RNAs between test and control groups. Structural modeling for RNAs could then be conducted using RNAstructure tools (6.0.1) with the goal of defining conserved nucleotide sequence and/or structural motifs within the RNAs. As proof of principle, we would expect to find Arabidopsis H/ACA snoRNAs previously identified via bioinformatics (233). Enriched RNAs could be studied to determine if they are transcribed predominantly by RNA Pol II or Pol III using promoter characterization. Any novel RNAs uncovered could then be characterized genetically as described above. Obtaining a complete list of dyskerin-associated RNA with secondary structure model will significantly improve the interpretation of how dyskerin selecting substrates. It might also lead to discovery of unknown functions for dyskerin.

Defining the complete complement of protein subunits in the Arabidopsis telomerase RNP

In Chapter III, I presented analysis of two telomerase-associated proteins AtPOT1a and dyskerin in Arabidopsis. These results and previous work from the Shippen lab indicate that AtPOT1a physically associates with AtTERT and telomeric DNA to facilitate the telomerase repeat addition processivity by retaining telomerase on the substrate (202, 204). In addition to AtPOT1a, I presented evidence that dyskerin directly interacts AtTR via a plant-specific assembly. This study provided support that dyskerin association with telomerase stimulates telomerase processivity *in vitro* and promotes TR stability *in vivo*. Despite these results, the full subunit composition of Arabidopsis telomerase RNP remains unknown. Revealing the complete components of plant telomerase will verify the functions of AtPOT1a and dyskerin. It will also significantly improve our understanding of telomerase RNP evolution and provide new insight into the dramatic divergence of telomerase RNP composition among different eukaryotic lineages.

TERT-mediated strategies to identify telomerase-associated proteins

To define the complete subunit composition of Arabidopsis telomerase RNP, we initiated an unbiased approach to purify the tagged-AtTERT protein and identify co-purified proteins as shown in Appendix. Because telomerase is not an abundant enzyme, quantitative MS analysis can be a powerful tool to identify low abundance proteins by looking at the relative enrichment of proteins in tagged sample versus an appropriate control, in our case a *tert* mutant line. As an initial strategy, we created a binary 'super-telomerase' vector (pHSN 35S-TSgeneAtTERT-U6-AtTR) that carries a 35S promoter-driven TwinStrep-tagged-AtTERT and a U6 promoter-driven AtTR, (Appendix, Figure 1A). A direct purification was conducted by using TwinStrep-affinity resin to capture tagged-AtTERT and associated proteins. A similar approach was used for yeast telomerase,

which revealed both tightly associated RNase P/MRP components (Pop1, Pop6, Pop7) and more loosely associated proteins, including the Cdc48-Npl4-Ufd1 complex (162, 164).

The Appendix reported the identification of two unexpected telomerase-associated proteins, AtLa1 and AtCPN60B, from the quantitative MS analysis (Appendix, Figure 2A). Both AtLa1 and AtCPN60B are significantly enriched from purification of the 'super-telomerase' plants compared to untransformed control (*tert* mutants). Statistical analysis of two biological replicates indicated the enrichments of AtLa1 and AtCPN60B are reproducible (Appendix, Figure 2B). We are currently working on digesting this observation with detailed interaction mechanisms and their biological functions.

In parallel with experiments to identify core components of the Arabidopsis telomerase RNP, we devised a sensitive purification strategy to capture a more comprehensive complement of loosely associated, potential regulatory subunits for Arabidopsis telomerase (Appendix). The initial plan is to use extracts prepared from so called 'native telomerase' transgenic plants for qMS. The 'native telomerase' plants were prepared similarly to the 'super-telomerase' with the difference of using a native TERT promoter (3000 bp upstream of transcription start site) to drive expression of TwinStrep-tagged-AtTERT and reconstitute telomerase *in vivo* (Appendix, Figure 1B). If we cannot identify proteins other than core subunits, we can attempt to preserve transient interactions and intermediate stages using cross-linking. We expect that "native telomerase" plants expressing a wild-type level of TERT and AtTR will reconstitute biologically relevant interactions for the telomerase RNP complex.

TR-mediated strategies to identify telomerase-associated proteins

An alternate approach to define the complete subunit composition of Arabidopsis

telomerase RNP is using tagged-AtTR as a bait. I propose to perform TR-based purification using a U6 promoter-driven AtTR construct fused with at least five copies of the MS2 bacteriophage coat protein binding sequence (pHSN-U6-AtTR-5xMS2). The position for installing 5xMS2 tag should be informed by AtTR secondary structure and nucleotide accessibility. Either the 5' or 3' end is a logical starting position for tagging. After a tag is engineered, it will be important to ensure the tag does not interfere with AtTR function and this can be tested via genetic complementation. MS2 affinity purification will enrich for AtTR binding proteins, including those that might not be present in TERT complexes. UV crosslinking can be performed prior to purification to preserve transient interactors including TR modifiers.

An alternate approach requires anti-sense oligos to specifically capture AtTR and AtTR-associated proteins. This approach is compatible with crosslinking. It was successfully applied to human telomerase RNA to identify hTR-binding proteins (235). In Chapter III, two anti-sense oligos were designed to target accessible regions of AtTR and revealed dyskerin as a co-purified protein by western blot (Chapter III, Figure 1). However, the amounts of proteins recovered from the anti-sense oligo purification were extremely limited. Therefore, optimizing an efficient purification protocol with large amount of input and a sensitive downstream qMS analysis will be required to generate reliable and reproducible results.

APEX-mediated strategies to identify telomerase-associated proteins

New studies with human telomerase have employed a proximity-dependent labeling technique to identify novel telomerase-associated proteins (298). Targeting the promiscuous biotin ligase APEX2 to MS2-tagged TR or TERT protein can enable biotinylation of physically close

proteins and subsequent proteomic identification of both endogenous subunits and interaction partners (299). This method represents a technological advance to detect transiently associated subunits with quicker temporal resolution because of the rapid (1-min) biotinylation catalyzed by APEX. Targeting of human TR by APEX2 identified a novel hTR interactor ALKBH5, which might regulate hTR function through its m^6A demethylase activity (298). Thus, the APEX-mediated labeling technique could serve as an alternate strategy to define telomerase accessory proteins in Arabidopsis. This approach might also uncover additional proteins transiently involving TR biogenesis and processing. Comparing the results from TERT-mediated "super-telomerase" purification and APEX-mediated purification could provide a means to dissect telomerase-associated proteins functionally and temporally.

Identification of other putative telomerase associated proteins

Dyskerin typically forms a heterotetramer complex consisting of dyskerin, Gar1, Nop10 and Nhp2. In Chapter III, I describe reconstitution of a dyskerin-Gal1-AtTR RNP to a stable state for cryo-EM analysis. Although Nop10 was co-expressed in the same construct as dyskerin, it was not detected in the final structure *in vitro*. This observation raises the question of whether Nop10 or Nhp2 stably associates with the Arabidopsis telomerase holoenzyme and whether these proteins are required for dyskerin function. If the Arabidopsis enzyme is similar to human telomerase, Nop10 and Nhp2 should be associated with the Arabidopsis telomerase (141). The Cryo-EM structure of the human telomerase-H/ACA RNP complex indicates that a single dyskerin associates with the 5' stem of hTR H/ACA domain (Chapter I, Figure 15), and this association is sufficient to recruit two full sets of dyskerin-Gar1-Nop10-Nhp2 heterotetramer (72). Dyskerin-TR and dyskerin-dyskerin interactions are the predominant forces that anchor the complex.

In Chapter III, I showed that Arabidopsis dyskerin not only associates with AtTR with high affinity, but also it dimerizes in a manner that is identical to dyskerin in the human telomerase complex. Therefore, I predict that Gar1, Nop10 and Nhp2 should also be present in the Arabidopsis telomerase complex through physical interactions with dyskerin. I propose genetic and biochemical approaches to test this hypothesis. Gar1, Nop10, and Nhp2 are conserved in Arabidopsis. However, AtGar1 and AtNhp2 are duplicated with no annotation to distinguish the proteins participating in telomerase H/ACA RNP. Nonetheless, it should be possible to generate mutants and antibodies to target the corresponding gene. IP experiments could then be used to test if Nop10 and Nhp2 co-purify with active Arabidopsis telomerase. In addition, knock-out or knock-down mutants of Nop10 or/and Nhp2 created by CRISPR-Cas9 could be valuable tools to test for a telomerase dysfunction phenotype.

Toward a high-resolution cryo-EM structure of Arabidopsis telomerase holoenzyme

In Chapter III, I present an 8.6-Å cryo-EM structure of the recombinant AtTR-dyskerin-Gal1 complex (Chapter III, Figure 4). Our complex exhibits an identical architecture as the dimerized dyskerin proteins within the human telomerase RNP, indicative of structural homology between the Arabidopsis and human dyskerin-TR complex. However, because our Arabidopsis complex lacks TERT, true structural homology for the entire core complex is only a hypothesis until a complete 3D model for the Arabidopsis telomerase RNP core can be obtained. More importantly, in the absence of TERT, much of the AtTR molecule will be highly flexible, precluding any solid structural interpretation. If we can ultimately obtain a cryo-EM structure from the entire purified active telomerase holoenzyme, it will provide both a framework for

understanding the molecular basis for telomerase holoenzyme assembly in plants and invaluable insight into telomerase RNP evolution.

Strategies to generate active telomerase core for cryo-EM

The preliminary cryo-EM analysis reported in Chapter III was part of a collaboration with Dr. Junjie Zhang. This collaboration should be extended in the future with telomerase enzyme purified from "super-telomerase" transgenic plants. Samples could be enriched using a *"direct"* affinity purification strategy identical to the purification which we conducted for qMS in the Appendix. Images could be obtained from either the Krios G3 microscope at the UTHealth science center at Houston or with the new Krios microscope recently obtained by our department. I estimate that 100 μL of 0.1 mg/mL telomerase particles will be required for imaging. Pilot cryo-EM images should be collected to ensure that particle integrity and concentration are sufficient for analysis. The collected images can then be processed similar to the approach used for analysis of the dyskerin-AtTR complex with cryoSPARC to generate a structural density and modeling based on published structures from homologous proteins.

One pitfall is that we may be unable to obtain sufficient material from *in vivo* reconstituted complexes for cryo-EM. If this is the case, an alternative approach would be to reconstitute recombinant telomerase particles *in vitro*. Because TERT is notoriously difficult to express in large quantities, we could begin with truncated TERT molecules to identify essential motifs required for RNA interaction. The RNA binding domain from Medaka TERT has been successfully purified from *E. coli* (78), and we could consider a similar strategy to obtain isolated domains of AtTERT for cryo-EM.

TAD3 is an indirect participator in telomere maintenance

In Chapter IV, we re-evaluated the identity of lncRNA AtTER2 that was previously reported as a telomerase activity regulator in response to DNA damage in Arabidopsis. Detailed analyses including updated genome annotation, strand-specific RNA-Seq, and strand-specific qRT-PCR indicated that *AtTER2* is not a stable transcript and was completely embedded in the gene encoding tRNA Deaminase 3 (*TAD3)*. Unexpectedly, TAD3 dysfunction led to progressive loss of telomeric DNA over successive generations. Further examination suggested that TAD3 indirectly contributes to telomere length homeostasis by manipulating the metabolic profile without altering telomere terminal architecture and telomerase enzymatic activity.

It is not unusual that a gene of interest indirectly participates in telomere maintenance. A genome-wide screen in yeast uncovered over 270 genes essential for telomere length maintenance (242). These genes span a variety of cellular processes, including protein degradation, pre-mRNA splicing and DNA replication, which suggests that telomere length is a sensitive indicator of cellular stresses. This interpretation could also apply to the phenotype we observed from the *tad3* mutants in Arabidopsis. Transcriptome analysis revealed that loss of TAD3 impacts many cellular pathways, especially auxin homeostasis, glucosinolate biosynthesis, and cell cycle progression. These pathways all have the potential to indirectly lead to telomere shortening by shared or independent mechanisms.

To obtain a more complete understanding of the factors that influence telomere maintenance in Arabidopsis, I propose to utilize genome-wide screens such as those performed in yeast (242). Scoring Arabidopsis T-DNA insertion collections by telomere length using TRF analysis would be an applicable approach. This strategy has recently been employed by Shippen lab in collaboration with Dr. Eugene Shakirov at Marshall University and Dr. Tom Juengar to find

genes that modulate telomere length set point. This study employed quantitative trait locus (QTL) mapping and transgenic manipulations that identified three QTL with the most distinguished on chromosome 5 explaining 42.2% of the telomere length variation (278). Using a genetic approach, a ribosomal RNA methyltransferase *NOP2a* was identified from this chromosome 5 QTL. Loss-of-function *nop2a* mutants exhibited a stable and shorter telomere length than the wild type. Similar to TAD3, NOP2a is involved in broad cellular pathways, including rRNA processing and ribosome biogenesis, which could impact telomere maintenance indirectly. Overall, a large network of genes participates in telomere homeostasis. Obtaining a complete list of genes that modulate telomere length will provide new insights into how the telomere is controlled by both intrinsic and extrinsic stresses on the cell.

Conclusions

In conclusion, this dissertation has provided important new information about the composition, interactions, and structure of the telomerase RNP complex from the plant model *A. thaliana*. It revealed similarities and intriguing differences in the telomerase RNP between plants, ciliates, and vertebrates. Future challenges will include defining the complete composition of the Arabidopsis telomerase RNP, elucidating the function of all the components, and the ultimate goal of solving a reliable cryo-EM structure of telomerase holoenzyme. The proposed studies will provide insight into the evolution of the telomerase complex, especially in the co-evolutions of RNA and its binding proteins.

REFERENCES

1. J. D. Watson, Origin of concatemeric T7 DNA. *Nat New Biol* **239**, 197-201 (1972).

2. A. M. Olovnikov, A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *J Theor Biol* **41**, 181-190 (1973).

3. M. A. Giardini, M. Segatto, M. S. da Silva, V. S. Nunes, M. I. Cano, Telomere and telomerase biology. *Prog Mol Biol Transl Sci* **125**, 1-40 (2014).

4. J. Meyne, R. L. Ratliff, R. K. Moyzis, Conservation of the human telomere sequence (TTAGGG)n among vertebrates. *Proc Natl Acad Sci U S A* **86**, 7049-7053 (1989).

5. E. H. Blackburn, J. G. Gall, A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *J Mol Biol* **120**, 33-53 (1978).

6. J. W. Shay, W. E. Wright, Telomeres and telomerase: three decades of progress. *Nat Rev Genet* **20**, 299-309 (2019).

7. J. Lingner, J. P. Cooper, T. R. Cech, Telomerase and DNA end replication: no longer a lagging strand problem? *Science* **269**, 1533-1534 (1995).

8. T. de Lange, A loopy view of telomere evolution. *Front Genet* **6**, 321 (2015).

9. E. H. Blackburn, K. Collins, Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harb Perspect Biol* **3** (2011).

10. C. W. Greider, E. H. Blackburn, Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* **43**, 405-413 (1985).

11. C. W. Greider, E. H. Blackburn, The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* **51**, 887-898 (1987).

12. C. W. Greider, E. H. Blackburn, A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature* **337**, 331-337 (1989).

13. D. Shippen-Lentz, E. H. Blackburn, Functional evidence for an RNA template in telomerase. *Science* **247**, 546-552 (1990).

14. C. Autexier, N. F. Lue, The structure and function of telomerase reverse transcriptase. *Annu Rev Biochem* **75**, 493-517 (2006).

15. T. M. Bryan, K. J. Goodrich, T. R. Cech, Telomerase RNA bound by protein motifs specific to telomerase reverse transcriptase. *Mol Cell* **6**, 493-499 (2000).

16. T. R. Cech, Beginning to understand the end of the chromosome. *Cell* **116**, 273-279 (2004).

17. C. I. Nugent, V. Lundblad, The telomerase reverse transcriptase: components and regulation. *Genes Dev* **12**, 1073-1085 (1998).

18. E. D. Egan, K. Collins, Biogenesis of telomerase ribonucleoproteins. *RNA* **18**, 1747-1759 (2012).

19. K. Riha, T. D. McKnight, L. R. Griffing, D. E. Shippen, Living with genome instability: plant responses to telomere dysfunction. *Science* **291**, 1797-1800 (2001).

20. C. M. Kong, X. W. Lee, X. Wang, Telomere shortening in human diseases. *FEBS J* **280**, 3180-3193 (2013).

21. M. A. Jafri, S. A. Ansari, M. H. Alqahtani, J. W. Shay, Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. *Genome Med* **8**, 69 (2016).

22. H. J. Muller, The remaking of chromosomes. *Collecting net* **13**, 181-198 (1938).

23. B. McClintock, The Stability of Broken Ends of Chromosomes in Zea Mays. *Genetics* **26**, 234-282 (1941).

24. K. Forstemann, J. Lingner, Molecular basis for telomere repeat divergence in budding yeast. *Mol Cell Biol* **21**, 7277-7286 (2001).

25. E. J. Richards, F. M. Ausubel, Isolation of a higher eukaryotic telomere from Arabidopsis thaliana. *Cell* **53**, 127-136 (1988).

26. E. Sykorova *et al.*, Telomere variability in the monocotyledonous plant order Asparagales. *Proc Biol Sci* **270**, 1893-1904 (2003).

27. P. Fajkus *et al.*, Allium telomeres unmasked: the unusual telomeric sequence (CTCGGTTATGGG)n is synthesized by telomerase. *Plant J* **85**, 337-347 (2016).

28. W. Chai, Q. Du, J. W. Shay, W. E. Wright, Human telomeres have different overhang sizes at leading versus lagging strands. *Mol Cell* **21**, 427-435 (2006).

29. P. Wu, H. Takai, T. de Lange, Telomeric 3' overhangs derive from resection by Exo1 and Apollo and fill-in by POT1b-associated CST. *Cell* **150**, 39-52 (2012).

30. D. E. Casteel *et al.*, A DNA polymerase-{alpha}{middle dot}primase cofactor with homology to replication protein A-32 regulates DNA replication in mammalian cells. *J Biol Chem* **284**, 5807-5818 (2009).

31. S. Valuchova *et al.*, Protection of Arabidopsis Blunt-Ended Telomeres Is Mediated by a Physical Association with the Ku Heterodimer. *Plant Cell* **29**, 1533-1545 (2017).

32. V. L. Fell, C. Schild-Poulter, The Ku heterodimer: function in DNA repair and beyond. *Mutat Res Rev Mutat Res* **763**, 15-29 (2015).

33. J. D. Griffith *et al.*, Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503-514 (1999).

34. T. de Lange, T-loops and the origin of telomeres. *Nat Rev Mol Cell Biol* **5**, 323-329 (2004).

35. R. M. Stansel, T. de Lange, J. D. Griffith, T-loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang. *EMBO J* **20**, 5532-5540 (2001).

36. L. Tomaska, S. Willcox, J. Slezakova, J. Nosek, J. D. Griffith, Taz1 binding to a fission yeast model telomere: formation of telomeric loops and higher order structures. *J Biol Chem* **279**, 50764-50772 (2004).

37. G. Sarek *et al.*, CDK phosphorylation of TRF2 controls t-loop dynamics during the cell cycle. *Nature* **575**, 523-527 (2019).

38. H. Gao, R. B. Cervantes, E. K. Mandell, J. H. Otero, V. Lundblad, RPA-like proteins mediate yeast telomere function. *Nat Struct Mol Biol* **14**, 208-214 (2007).

39. K. A. Lewis, D. S. Wuttke, Telomerase and telomere-associated proteins: structural insights into mechanism and evolution. *Structure* **20**, 28-39 (2012).

40. V. Faure, S. Coulon, J. Hardy, V. Geli, Cdc13 and telomerase bind through different mechanisms at the lagging- and leading-strand telomeres. *Mol Cell* **38**, 842-852 (2010).

41. S. K. Evans, V. Lundblad, Positive and negative regulation of telomerase access to the telomere. *J Cell Sci* **113 Pt 19**, 3357-3364 (2000).

42. A. Smogorzewska, T. de Lange, Regulation of telomerase by telomeric proteins. *Annu Rev Biochem* **73**, 177-208 (2004).

43. L. Y. Chen, S. Redon, J. Lingner, The human CST complex is a terminator of telomerase activity. *Nature* **488**, 540-544 (2012).

44. Y. Miyake *et al.*, RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway. *Mol Cell* **36**, 193-206 (2009).

45. T. de Lange, Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev* **19**, 2100-2110 (2005).

46. J. P. Cooper, E. R. Nimmo, R. C. Allshire, T. R. Cech, Regulation of telomere length and function by a Myb-domain protein in fission yeast. *Nature* **385**, 744-747 (1997).

47. B. van Steensel, A. Smogorzewska, T. de Lange, TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401-413 (1998).

48. B. van Steensel, T. de Lange, Control of telomere length by the human telomeric protein TRF1. *Nature* **385**, 740-743 (1997).

49. D. Liu, M. S. O'Connor, J. Qin, Z. Songyang, Telosome, a mammalian telomere-associated complex formed by multiple telomeric proteins. *J Biol Chem* **279**, 51338-51342 (2004).

50. J. Z. Ye *et al.*, TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. *J Biol Chem* **279**, 47264-47271 (2004).

51. M. Lei, E. R. Podell, T. R. Cech, Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nat Struct Mol Biol* **11**, 1223-1229 (2004).

52. D. Loayza, H. Parsons, J. Donigian, K. Hoke, T. de Lange, DNA binding features of human POT1: a nonamer 5'-TAGGGTTAG-3' minimal binding site, sequence specificity, and internal binding to multimeric sites. *J Biol Chem* **279**, 13241-13248 (2004).

53. F. Wang *et al.*, The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature* **445**, 506-510 (2007).

54. M. Xu, J. Kiselar, T. L. Whited, W. Hernandez-Sanchez, D. J. Taylor, POT1-TPP1 differentially regulates telomerase via POT1 His266 and as a function of single-stranded telomere DNA length. *Proc Natl Acad Sci U S A* **116**, 23527-23533 (2019).

55. W. E. Wright, M. A. Piatyszek, W. E. Rainey, W. Byrd, J. W. Shay, Telomerase activity in human germline and embryonic tissues and cells. *Dev Genet* **18**, 173-179 (1996).

56. X. Yi, J. W. Shay, W. E. Wright, Quantitation of telomerase components and hTERT mRNA splicing patterns in immortal human cells. *Nucleic Acids Res* **29**, 4818-4825 (2001).

57. X. Yi, V. M. Tesmer, I. Savre-Train, J. W. Shay, W. E. Wright, Both transcriptional and posttranscriptional mechanisms regulate human telomerase template RNA levels. *Mol Cell Biol* **19**, 3989-3997 (1999).

58. M. Gladych, A. Wojtyla, B. Rubis, Human telomerase expression regulation. *Biochem Cell Biol* **89**, 359-376 (2011).

59. Y. S. Cong, W. E. Wright, J. W. Shay, Human telomerase and its regulation. *Microbiol Mol Biol Rev* **66**, 407-425, table of contents (2002).

60. J. Wang, L. Y. Xie, S. Allan, D. Beach, G. J. Hannon, Myc activates telomerase. *Genes Dev* **12**, 1769-1774 (1998).

61. S. S. Kang, T. Kwon, D. Y. Kwon, S. I. Do, Akt protein kinase enhances human telomerase activity through phosphorylation of telomerase reverse transcriptase subunit. *J Biol Chem* **274**, 13085-13090 (1999).

62. J. H. Kim *et al.*, Ubiquitin ligase MKRN1 modulates telomere length homeostasis through a proteolysis of hTERT. *Genes Dev* **19**, 776-781 (2005).

63. R. L. Tomlinson, T. D. Ziegler, T. Supakorndej, R. M. Terns, M. P. Terns, Cell cycle-regulated trafficking of human telomerase to telomeres. *Mol Biol Cell* **17**, 955-965 (2006).

64. M. I. Zvereva, D. M. Shcherbakova, O. A. Dontsova, Telomerase: structure, functions, and activity regulation. *Biochemistry (Mosc)* **75**, 1563-1583 (2010).

65. H. D. Wyatt, S. C. West, T. L. Beattie, InTERTpreting telomerase structure and function. *Nucleic Acids Res* **38**, 5609-5622 (2010).

66. S. L. Weinrich *et al.*, Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTRT. *Nat Genet* **17**, 498-502 (1997).

67. T. W. Chu, Y. D'Souza, C. Autexier, The Insertion in Fingers Domain in Human Telomerase Can Mediate Enzyme Processivity and Telomerase Recruitment to Telomeres in a TPP1-Dependent Manner. *Mol Cell Biol* **36**, 210-222 (2016).

68. A. J. Gillis, A. P. Schuller, E. Skordalakes, Structure of the Tribolium castaneum telomerase catalytic subunit TERT. *Nature* **455**, 633-637 (2008).

69. Y. Peng, I. S. Mian, N. F. Lue, Analysis of telomerase processivity: mechanistic similarity to HIV-1 reverse transcriptase and role in telomere maintenance. *Mol Cell* **7**, 1201-1211 (2001).

70. T. J. Moriarty, R. J. Ward, M. A. Taboski, C. Autexier, An anchor site-type defect in human telomerase that disrupts telomere length maintenance and cellular immortalization. *Mol Biol Cell* **16**, 3152-3161 (2005).

71. N. F. Lue, A physical and functional constituent of telomerase anchor site. *J Biol Chem* **280**, 26586-26591 (2005).

72. T. H. D. Nguyen *et al.*, Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* **557**, 190-195 (2018).

73. E. Romi *et al.*, High-resolution physical and functional mapping of the template adjacent DNA binding site in catalytically active telomerase. *Proc Natl Acad Sci U S A* **104**, 8791-8796 (2007).

74. N. F. Lue, Y. Peng, Negative regulation of yeast telomerase activity through an interaction with an upstream region of the DNA primer. *Nucleic Acids Res* **26**, 1487-1494 (1998).

75. S. A. Jacobs, E. R. Podell, T. R. Cech, Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. *Nat Struct Mol Biol* **13**, 218-225 (2006).

76. C. M. O'Connor, C. K. Lai, K. Collins, Two purified domains of telomerase reverse transcriptase reconstitute sequence-specific interactions with RNA. *J Biol Chem* **280**, 17533-17539 (2005).

77. S. Rouda, E. Skordalakes, Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding. *Structure* **15**, 1403-1412 (2007).

78. J. Huang *et al.*, Structural basis for protein-RNA recognition in telomerase. *Nat Struct Mol Biol* **21**, 507-512 (2014).

79. J. Jiang *et al.*, Structure of Telomerase with Telomeric DNA. *Cell* **173**, 1179-1190 e1113 (2018).

80. J. Jiang *et al.*, Structure of Tetrahymena telomerase reveals previously unknown subunits, functions, and interactions. *Science* **350**, aab4070 (2015).

81. S. Hossain, S. Singh, N. F. Lue, Functional analysis of the C-terminal extension of telomerase reverse transcriptase. A putative "thumb" domain. *J Biol Chem* **277**, 36174-36180 (2002).

82. S. Huard, T. J. Moriarty, C. Autexier, The C terminus of the human telomerase reverse transcriptase is a determinant of enzyme processivity. *Nucleic Acids Res* **31**, 4059-4070 (2003).

83. S. S. Banik *et al.*, C-terminal regions of the human telomerase catalytic subunit essential for in vivo enzyme activity. *Mol Cell Biol* **22**, 6234-6246 (2002).

84. B. N. Armbruster, S. S. Banik, C. Guo, A. C. Smith, C. M. Counter, N-terminal domains of the human telomerase catalytic subunit required for enzyme activity in vivo. *Mol Cell Biol* **21**, 7775-7786 (2001).

85. J. D. Podlevsky, J. J. Chen, Evolutionary perspectives of telomerase RNA structure and function. *RNA Biol* **13**, 720-732 (2016).

86. X. Qi *et al.*, The common ancestral core of vertebrate and fungal telomerase RNAs. *Nucleic Acids Res* **41**, 450-462 (2013).

87. J. L. Chen, M. A. Blasco, C. W. Greider, Secondary structure of vertebrate telomerase RNA. *Cell* **100**, 503-514 (2000).

88. J. D. Podlevsky, Y. Li, J. J. Chen, Structure and function of echinoderm telomerase RNA. *RNA* **22**, 204-215 (2016).

89. J. Feng *et al.*, The RNA component of human telomerase. *Science* **269**, 1236-1241 (1995).

90. J. L. Chen, C. W. Greider, Determinants in mammalian telomerase RNA that mediate enzyme processivity and cross-species incompatibility. *EMBO J* **22**, 304-314 (2003).

91. J. L. Chen, C. W. Greider, Template boundary definition in mammalian telomerase. *Genes Dev* **17**, 2747-2752 (2003).

92. W. C. Drosopoulos, R. Direnzo, V. R. Prasad, Human telomerase RNA template sequence is a determinant of telomere repeat extension rate. *J Biol Chem* **280**, 32801-32810 (2005).

93. A. F. Brown *et al.*, A self-regulating template in human telomerase. *Proc Natl Acad Sci U S A* **111**, 11311-11316 (2014).

94. C. Autexier, C. W. Greider, Boundary elements of the Tetrahymena telomerase RNA template and alignment domains. *Genes Dev* **9**, 2227-2239 (1995).

95. Y. Tzfati, T. B. Fulton, J. Roy, E. H. Blackburn, Template boundary in a yeast telomerase specified by RNA structure. *Science* **288**, 863-867 (2000).

96. C. K. Lai, M. C. Miller, K. Collins, Template boundary definition in Tetrahymena telomerase. *Genes Dev* **16**, 415-420 (2002).

97. J. D. Podlevsky, Y. Li, J. J. Chen, The functional requirement of two structural domains within telomerase RNA emerged early in eukaryotes. *Nucleic Acids Res* **44**, 9891-9901 (2016).

98. L. I. Jansson *et al.*, Structural basis of template-boundary definition in Tetrahymena telomerase. *Nat Struct Mol Biol* **22**, 883-888 (2015).

99. J. Song *et al.*, The conserved structure of plant telomerase RNA provides the missing link for an evolutionary pathway from ciliates to humans. *Proc Natl Acad Sci U S A* **116**, 24542-24550 (2019).

100. M. Harkisheimer, M. Mason, E. Shuvaeva, E. Skordalakes, A motif in the vertebrate telomerase N-terminal linker of TERT contributes to RNA binding and telomerase activity and processivity. *Structure* **21**, 1870-1878 (2013).

101. C. S. Hinkley *et al.*, The mouse telomerase RNA 5"-end lies just upstream of the telomerase template sequence. *Nucleic Acids Res* **26**, 532-536 (1998).

102. J. D. Podlevsky, J. J. Chen, It all comes together at the ends: telomerase structure, function, and biogenesis. *Mutat Res* **730**, 3-11 (2012).

103. Y. Tzfati, Z. Knight, J. Roy, E. H. Blackburn, A novel pseudoknot element is essential for the action of a yeast telomerase. *Genes Dev* **17**, 1779-1788 (2003).

104. C. Autexier, C. W. Greider, Mutational analysis of the Tetrahymena telomerase RNA: identification of residues affecting telomerase activity in vitro. *Nucleic Acids Res* **26**, 787-795 (1998).

105. D. Gilley, E. H. Blackburn, The telomerase RNA pseudoknot is critical for the stable assembly of a catalytically active ribonucleoprotein. *Proc Natl Acad Sci U S A* **96**, 6621-6625 (1999).

106. D. D. Cash *et al.*, Pyrimidine motif triple helix in the Kluyveromyces lactis telomerase RNA pseudoknot is essential for function in vivo. *Proc Natl Acad Sci U S A* **110**, 10970-10975 (2013).

107. J. L. Chen, C. W. Greider, Functional analysis of the pseudoknot structure in human telomerase RNA. *Proc Natl Acad Sci U S A* **102**, 8080-8085; discussion 8077-8089 (2005).

108. C. A. Theimer, C. A. Blois, J. Feigon, Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol Cell* **17**, 671-682 (2005).

109. D. D. Cash, J. Feigon, Structure and folding of the Tetrahymena telomerase RNA pseudoknot. *Nucleic Acids Res* **45**, 482-495 (2017).

110. T. H. D. Nguyen, K. Collins, E. Nogales, Telomerase structures and regulation: shedding light on the chromosome end. *Curr Opin Struct Biol* **55**, 185-193 (2019).

111. J. L. Chen, K. K. Opperman, C. W. Greider, A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. *Nucleic Acids Res* **30**, 592-597 (2002).

112. D. X. Mason, E. Goneska, C. W. Greider, Stem-loop IV of tetrahymena telomerase RNA stimulates processivity in trans. *Mol Cell Biol* **23**, 5606-5613 (2003).

113. M. Xie *et al.*, Structure and function of the smallest vertebrate telomerase RNA from teleost fish. *J Biol Chem* **283**, 2049-2059 (2008).

114. J. R. Mitchell, K. Collins, Human telomerase activation requires two independent interactions between telomerase RNA and telomerase reverse transcriptase. *Mol Cell* **6**, 361-371 (2000).

115. Y. Brown *et al.*, A critical three-way junction is conserved in budding yeast and vertebrate telomerase RNAs. *Nucleic Acids Res* **35**, 6280-6289 (2007).

116. A. J. Berman, A. R. Gooding, T. R. Cech, Tetrahymena telomerase protein p65 induces conformational changes throughout telomerase RNA (TER) and rescues telomerase reverse transcriptase and TER assembly mutants. *Mol Cell Biol* **30**, 4965-4976 (2010).

117. B. M. Akiyama, J. Loper, K. Najarro, M. D. Stone, The C-terminal domain of Tetrahymena thermophila telomerase holoenzyme protein p65 induces multiple structural changes in telomerase RNA. *RNA* **18**, 653-660 (2012).

118. M. Singh *et al.*, Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65. *Mol Cell* **47**, 16-26 (2012).

119. J. Lingner, L. L. Hendrick, T. R. Cech, Telomerase RNAs of different ciliates have a common secondary structure and a permuted template. *Genes Dev* **8**, 1984-1998 (1994).

120. K. L. Witkin, K. Collins, Holoenzyme proteins required for the physiological assembly and activity of telomerase. *Genes Dev* **18**, 1107-1118 (2004).

121. B. Min, K. Collins, An RPA-related sequence-specific DNA-binding subunit of telomerase holoenzyme is required for elongation processivity and telomere maintenance. *Mol Cell* **36**, 609-619 (2009).

122. S. Aigner *et al.*, Euplotes telomerase contains an La motif protein produced by apparent translational frameshifting. *EMBO J* **19**, 6230-6239 (2000).

123. S. Aigner, J. Postberg, H. J. Lipps, T. R. Cech, The Euplotes La motif protein p43 has properties of a telomerase-specific subunit. *Biochemistry* **42**, 5736-5747 (2003).

124. C. Bousquet-Antonelli, J. M. Deragon, A comprehensive analysis of the La-motif protein superfamily. *RNA* **15**, 750-764 (2009).

125. S. L. Wolin, T. Cedervall, The La protein. *Annu Rev Biochem* **71**, 375-403 (2002).

126. Y. Huang, M. A. Bayfield, R. V. Intine, R. J. Maraia, Separate RNA-binding surfaces on the multifunctional La protein mediate distinguishable activities in tRNA maturation. *Nat Struct Mol Biol* **13**, 611-618 (2006).

127. A. K. Mennie, B. A. Moser, T. M. Nakamura, LARP7-like protein Pof8 regulates telomerase assembly and poly(A)+TERRA expression in fission yeast. *Nat Commun* **9**, 586 (2018).

128. D. J. Paez-Moscoso *et al.*, Pof8 is a La-related protein and a constitutive component of telomerase in fission yeast. *Nat Commun* **9**, 587 (2018).

129. R. Basu, C. D. Eichhorn, R. Cheng, R. D. Peterson, J. Feigon, Structure of S. pombe telomerase protein Pof8 C-terminal domain is an xRRM conserved among LARP7 proteins. *RNA Biol* 10.1080/15476286.2020.1836891, 1-12 (2020).

130. B. Holohan *et al.*, Impaired telomere maintenance in Alazami syndrome patients with LARP7 deficiency. *BMC Genomics* **17**, 749 (2016).

131. D. C. Zappulla, T. R. Cech, Yeast telomerase RNA: a flexible scaffold for protein subunits. *Proc Natl Acad Sci U S A* **101**, 10024-10029 (2004).

132. W. Tang, R. Kannan, M. Blanchette, P. Baumann, Telomerase RNA biogenesis involves sequential binding by Sm and Lsm complexes. *Nature* **484**, 260-264 (2012).

133. A. G. Seto, A. J. Zaug, S. G. Sobel, S. L. Wolin, T. R. Cech, Saccharomyces cerevisiae telomerase is an Sm small nuclear ribonucleoprotein particle. *Nature* **401**, 177-180 (1999).

134. J. F. Noel, S. Larose, S. Abou Elela, R. J. Wellinger, Budding yeast telomerase RNA transcription termination is dictated by the Nrd1/Nab3 non-coding RNA termination pathway. *Nucleic Acids Res* **40**, 5625-5636 (2012).

135. J. A. Box, J. T. Bunch, W. Tang, P. Baumann, Spliceosomal cleavage generates the 3' end of telomerase RNA. *Nature* **456**, 910-914 (2008).

136. X. Hu *et al.*, Quality-Control Mechanism for Telomerase RNA Folding in the Cell. *Cell Rep* **33**, 108568 (2020).

137. S. K. Gupta *et al.*, The Trypanosoma brucei telomerase RNA (TER) homologue binds core proteins of the C/D snoRNA family. *FEBS Lett* **587**, 1399-1404 (2013).

138. E. J. Vasconcelos *et al.*, The putative Leishmania telomerase RNA (LeishTER) undergoes trans-splicing and contains a conserved template sequence. *PLoS One* **9**, e112061 (2014).

139. R. Sandhu *et al.*, A trans-spliced telomerase RNA dictates telomere synthesis in Trypanosoma brucei. *Cell Res* **23**, 537-551 (2013).

140. J. R. Mitchell, J. Cheng, K. Collins, A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Mol Cell Biol* **19**, 567-576 (1999).

141. C. Wang, U. T. Meier, Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J* **23**, 1857-1867 (2004).

142. E. D. Egan, K. Collins, Specificity and stoichiometry of subunit interactions in the human telomerase holoenzyme assembled in vivo. *Mol Cell Biol* **30**, 2775-2786 (2010).

143. C. K. Tseng, H. F. Wang, M. R. Schroeder, P. Baumann, The H/ACA complex disrupts triplex in hTR precursor to permit processing by RRP6 and PARN. *Nat Commun* **9**, 5430 (2018).

144. J. R. Mitchell, E. Wood, K. Collins, A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* **402**, 551-555 (1999).

145. J. M. Vogan *et al.*, Minimized human telomerase maintains telomeres and resolves endogenous roles of H/ACA proteins, TCAB1, and Cajal bodies. *Elife* **5** (2016).

146. A. S. Venteicher *et al.*, A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science* **323**, 644-648 (2009).

147. F. Zhong *et al.*, Disruption of telomerase trafficking by TCAB1 mutation causes dyskeratosis congenita. *Genes Dev* **25**, 11-16 (2011).

148. L. Chen *et al.*, An Activity Switch in Human Telomerase Based on RNA Conformation and Shaped by TCAB1. *Cell* **174**, 218-230 e213 (2018).

149. A. Sauerwald *et al.*, Structure of active dimeric human telomerase. *Nat Struct Mol Biol* **20**, 454-460 (2013).

150. B. E. Jady, E. Bertrand, T. Kiss, Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. *J Cell Biol* **164**, 647-652 (2004).

151. D. E. MacNeil, P. Lambert-Lanteigne, C. Autexier, N-terminal residues of human dyskerin are required for interactions with telomerase RNA that prevent RNA degradation. *Nucleic Acids Res* **47**, 5368-5380 (2019).

152. J. Lingner, T. R. Cech, T. R. Hughes, V. Lundblad, Three Ever Shorter Telomere (EST) genes are dispensable for in vitro yeast telomerase activity. *Proc Natl Acad Sci U S A* **94**, 11190-11195 (1997).

153. V. Lundblad, J. W. Szostak, A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* **57**, 633-643 (1989).

154. A. G. Seto, A. J. Livengood, Y. Tzfati, E. H. Blackburn, T. R. Cech, A bulged stem tethers Est1p to telomerase RNA in budding yeast. *Genes Dev* **16**, 2800-2812 (2002).

155. S. K. Evans, V. Lundblad, The Est1 subunit of Saccharomyces cerevisiae telomerase makes multiple contributions to telomere length maintenance. *Genetics* **162**, 1101-1115 (2002).

156. H. T. Beernink, K. Miller, A. Deshpande, P. Bucher, J. P. Cooper, Telomere maintenance in fission yeast requires an Est1 ortholog. *Curr Biol* **13**, 575-580 (2003).

157. D. C. DeZwaan, B. C. Freeman, Is there a telomere-bound 'EST' telomerase holoenzyme? *Cell Cycle* **9**, 1913-1917 (2010).

158. T. M. Tucey, V. Lundblad, A yeast telomerase complex containing the Est1 recruitment protein is assembled early in the cell cycle. *Biochemistry* **52**, 1131-1133 (2013).

159. J. J. Lin, V. A. Zakian, An in vitro assay for Saccharomyces telomerase requires EST1. *Cell* **81**, 1127-1135 (1995).

160. T. R. Hughes, S. K. Evans, R. G. Weilbaecher, V. Lundblad, The Est3 protein is a subunit of yeast telomerase. *Curr Biol* **10**, 809-812 (2000).

161. J. M. Talley, D. C. DeZwaan, L. D. Maness, B. C. Freeman, K. L. Friedman, Stimulation of yeast telomerase activity by the ever shorter telomere 3 (Est3) subunit is dependent on direct interaction with the catalytic protein Est2. *J Biol Chem* **286**, 26431-26439 (2011).

162. B. Lemieux *et al.*, Active Yeast Telomerase Shares Subunits with Ribonucleoproteins RNase P and RNase MRP. *Cell* **165**, 1171-1181 (2016).

163. P. D. Garcia *et al.*, Stability and nuclear localization of yeast telomerase depend on protein components of RNase P/MRP. *Nat Commun* **11**, 2173 (2020).

164. K. W. Lin *et al.*, Proteomics of yeast telomerase identified Cdc48-Npl4-Ufd1 and Ufd4 as regulators of Est1 and telomere length. *Nat Commun* **6**, 8290 (2015).

165. M. Rape *et al.*, Mobilization of processed, membrane-tethered SPT23 transcription factor by CDC48(UFD1/NPL4), a ubiquitin-selective chaperone. *Cell* **107**, 667-677 (2001).

166. T. S. Fisher, V. A. Zakian, Ku: a multifunctional protein involved in telomere maintenance. *DNA Repair (Amst)* **4**, 1215-1226 (2005).

167. A. E. Stellwagen, Z. W. Haimberger, J. R. Veatch, D. E. Gottschling, Ku interacts with telomerase RNA to promote telomere addition at native and broken chromosome ends. *Genes Dev* **17**, 2384-2395 (2003).

168. M. M. Kabaha, B. Zhitomirsky, I. Schwartz, Y. Tzfati, The 5' arm of Kluyveromyces lactis telomerase RNA is critical for telomerase function. *Mol Cell Biol* **28**, 1875-1882 (2008).

169. T. S. Fisher, A. K. Taggart, V. A. Zakian, Cell cycle-dependent regulation of yeast telomerase by Ku. *Nat Struct Mol Biol* **11**, 1198-1205 (2004).

170. J. S. Pfingsten *et al.*, Mutually exclusive binding of telomerase RNA and DNA by Ku alters telomerase recruitment model. *Cell* **148**, 922-932 (2012).

171. D. C. Zappulla, K. Goodrich, T. R. Cech, A miniature yeast telomerase RNA functions in vivo and reconstitutes activity in vitro. *Nat Struct Mol Biol* **12**, 1072-1077 (2005).

172. C. J. Lim *et al.*, The structure of human CST reveals a decameric assembly bound to telomeric DNA. *Science* **368**, 1081-1085 (2020).

173. Y. Wu, V. A. Zakian, The telomeric Cdc13 protein interacts directly with the telomerase subunit Est1 to bring it to telomeric DNA ends in vitro. *Proc Natl Acad Sci U S A* **108**, 20362-20369 (2011).

174. K. Hong *et al.*, Tetrahymena telomerase holoenzyme assembly, activation, and inhibition by domains of the p50 central hub. *Mol Cell Biol* **33**, 3962-3971 (2013).

175. J. Jiang *et al.*, The architecture of Tetrahymena telomerase holoenzyme. *Nature* **496**, 187-192 (2013).

176. C. M. Latrick, T. R. Cech, POT1-TPP1 enhances telomerase processivity by slowing primer dissociation and aiding translocation. *EMBO J* **29**, 924-933 (2010).

177. A. J. Zaug, E. R. Podell, J. Nandakumar, T. R. Cech, Functional interaction between telomere protein TPP1 and telomerase. *Genes Dev* **24**, 613-622 (2010).

178. E. Abreu *et al.*, TIN2-tethered TPP1 recruits human telomerase to telomeres in vivo. *Mol Cell Biol* **30**, 2971-2982 (2010).

179. T. de Lange, How shelterin solves the telomere end-protection problem. *Cold Spring Harb Symp Quant Biol* **75**, 167-177 (2010).

180. A. M. Tejera *et al.*, TPP1 is required for TERT recruitment, telomere elongation during nuclear reprogramming, and normal skin development in mice. *Dev Cell* **18**, 775-789 (2010).

181. A. S. Venteicher, Z. Meng, P. J. Mason, T. D. Veenstra, S. E. Artandi, Identification of ATPases pontin and reptin as telomerase components essential for holoenzyme assembly. *Cell* **132**, 945-957 (2008).

182. B. Bernardes de Jesus, M. A. Blasco, Telomerase at the intersection of cancer and aging. *Trends Genet* **29**, 513-520 (2013).

183. C. K. Garcia, W. E. Wright, J. W. Shay, Human diseases of telomerase dysfunction: insights into tissue aging. *Nucleic Acids Res* **35**, 7406-7416 (2007).

184. B. Holohan, W. E. Wright, J. W. Shay, Cell biology of disease: Telomeropathies: an emerging spectrum disorder. *J Cell Biol* **205**, 289-299 (2014).

185. R. A. Drachtman, B. P. Alter, Dyskeratosis congenita. *Dermatol Clin* **13**, 33-39 (1995).

186. K. Terada *et al.*, TERT and TERC mutations detected in cryptic dyskeratosis congenita suppress telomerase activity. *Int J Lab Hematol* **42**, 316-321 (2020).

187. R. Rashid *et al.*, Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol Cell* **21**, 249-260 (2006).

188. A. J. Walne *et al.*, Genetic heterogeneity in autosomal recessive dyskeratosis congenita with one subtype due to mutations in the telomerase-associated protein NOP10. *Hum Mol Genet* **16**, 1619-1629 (2007).

189. T. Vulliamy *et al.*, Mutations in the telomerase component NHP2 cause the premature ageing syndrome dyskeratosis congenita. *Proc Natl Acad Sci U S A* **105**, 8073-8078 (2008).

190. I. Dokal, T. Vulliamy, Dyskeratosis congenita: its link to telomerase and aplastic anaemia. *Blood Rev* **17**, 217-225 (2003).

191. Z. T. Xin *et al.*, Functional characterization of natural telomerase mutations found in patients with hematologic disorders. *Blood* **109**, 524-532 (2007).

192. M. E. Schranz, M. A. Lysak, T. Mitchell-Olds, The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends Plant Sci* **11**, 535-542 (2006).

193. M. E. Schranz, B. H. Song, A. J. Windsor, T. Mitchell-Olds, Comparative genomics in the Brassicaceae: a family-wide perspective. *Curr Opin Plant Biol* **10**, 168-175 (2007).

194. E. V. Shakirov, D. E. Shippen, Length regulation and dynamics of individual telomere tracts in wild-type Arabidopsis. *Plant Cell* **16**, 1959-1967 (2004).

195. A. Kazda *et al.*, Chromosome end protection by blunt-ended telomeres. *Genes Dev* **26**, 1703-1713 (2012).

196. A. D. Nelson, D. E. Shippen, Blunt-ended telomeres: an alternative ending to the replication and end protection stories. *Genes Dev* **26**, 1648-1652 (2012).

197. M. S. Fitzgerald *et al.*, Disruption of the telomerase catalytic subunit gene from Arabidopsis inactivates telomerase and leads to a slow loss of telomeric DNA. *Proc Natl Acad Sci U S A* **96**, 14813-14818 (1999).

198. C. Cifuentes-Rojas, K. Kannan, L. Tseng, D. E. Shippen, Two RNA subunits and POT1a are components of Arabidopsis telomerase. *Proc Natl Acad Sci U S A* **108**, 73-78 (2011).

199. P. Fajkus *et al.*, Telomerase RNAs in land plants. *Nucleic Acids Res* 10.1093/nar/gkz695 (2019).

200. K. Dew-Budd, J. Cheung, K. Palos, E. S. Forsythe, M. A. Beilstein, Evolutionary and biochemical analyses reveal conservation of the Brassicaceae telomerase ribonucleoprotein complex. *PLoS One* **15**, e0222687 (2020).

201. C. Cifuentes-Rojas *et al.*, An alternative telomerase RNA in Arabidopsis modulates enzyme activity in response to DNA damage. *Genes Dev* **26**, 2512-2523 (2012).

202. Y. V. Surovtseva *et al.*, Arabidopsis POT1 associates with the telomerase RNP and is required for telomere maintenance. *EMBO J* **26**, 3653-3661 (2007).

203. E. V. Shakirov, Y. V. Surovtseva, N. Osbun, D. E. Shippen, The Arabidopsis Pot1 and Pot2 proteins function in telomere length homeostasis and chromosome end protection. *Mol Cell Biol* **25**, 7725-7733 (2005).

204. A. Arora, M. A. Beilstein, D. E. Shippen, Evolution of Arabidopsis protection of telomeres 1 alters nucleic acid recognition and telomerase regulation. *Nucleic Acids Res* **44**, 9821-9830 (2016).

205. K. Kannan, A. D. Nelson, D. E. Shippen, Dyskerin is a component of the Arabidopsis telomerase RNP required for telomere maintenance. *Mol Cell Biol* **28**, 2332-2341 (2008).

206. P. Fajkus *et al.*, Telomerase RNAs in land plants. *Nucleic Acids Res* **47**, 9842-9856 (2019).

207. P. P. Schrumpfova *et al.*, Telomere repeat binding proteins are functional components of Arabidopsis telomeres and interact with telomerase. *Plant J* **77**, 770-781 (2014).

208. S. Schorova, J. Fajkus, L. Zaveska Drabkova, D. Honys, P. P. Schrumpfova, The plant Pontin and Reptin homologues, RuvBL1 and RuvBL2a, colocalize with TERT and TRB proteins in vivo, and participate in telomerase biogenesis. *Plant J* **98**, 195-212 (2019).

209. J. E. Wilusz, H. Sunwoo, D. L. Spector, Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* **23**, 1494-1504 (2009).

210. J. W. Shay, W. E. Wright, Telomeres and telomerase: three decades of progress. *Nature Reviews Genetics* **20**, 299-309 (2019).

211. E. Casacuberta, Drosophila: Retrotransposons Making up Telomeres. *Viruses* **9** (2017).

212. K. Dew-Budd, J. Cheung, K. Palos, E. S. Forsythe, M. A. Beilstein, Evolutionary and biochemical analyses reveal conservation of the Brassicaceae telomerase ribonucleoprotein complex. *bioRxiv* 10.1101/760785, 760785 (2019).

213. A. Samach, C. Melamed-Bessudo, N. Avivi-Ragolski, S. Pietrokovski, A. A. Levy, Identification of plant RAD52 homologs and characterization of the Arabidopsis thaliana RAD52-like genes. *Plant Cell* **23**, 4266-4279 (2011).

214. J. Wu *et al.*, A novel hypoxic stress-responsive long non-coding RNA transcribed by RNA polymerase III in Arabidopsis. *RNA Biol* **9**, 302-313 (2012).

215. A. Mosig, K. Sameith, P. Stadler, Fragrep: an efficient search tool for fragmented patterns in genomic sequences. *Genomics Proteomics Bioinformatics* **4**, 56-60 (2006).

216. E. P. Nawrocki, S. R. Eddy, Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933-2935 (2013).

217. K. A. Wilkinson, E. J. Merino, K. M. Weeks, Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **1**, 1610-1616 (2006).

218. M. Zubradt *et al.*, DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat Methods* **14**, 75-82 (2017).

219. N. R. Pace, D. K. Smith, G. J. Olsen, B. D. James, Phylogenetic comparative analysis and the secondary structure of ribonuclease P RNA--a review. *Gene* **82**, 65-75 (1989).

220. X. Zhang, R. Henriques, S. S. Lin, Q. W. Niu, N. H. Chua, Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. *Nat Protoc* **1**, 641-646 (2006).

221. Z. Wang *et al.*, SWI2/SNF2 ATPase CHR2 remodels pri-miRNAs via Serrate to impede miRNA production. *Nature* **557**, 516-521 (2018).

222. Z. Wang, M. Wang, T. Wang, Y. Zhang, X. Zhang, Genome-wide probing RNA structure with the modified DMS-MaPseq in Arabidopsis. *Methods* **155**, 30-40 (2019).

223. Y. Ding *et al.*, In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696-700 (2014).

224. H. Xu, A. D. Nelson, D. E. Shippen, A transposable element within the Non-canonical telomerase RNA of Arabidopsis thaliana modulates telomerase in response to DNA damage [corrected]. *PLoS Genet* **11**, e1005281 (2015).

225. P. Baumann, C. Price, Pot1 and telomere maintenance. *FEBS Lett* **584**, 3779-3784 (2010).

226. J. N. Bandaria, P. Qin, V. Berk, S. Chu, A. Yildiz, Shelterin Protects Chromosome Ends by Compacting Telomeric Chromatin. *Cell* **164**, 735-746 (2016).

227. I. Schmutz, T. de Lange, Shelterin. *Curr Biol* **26**, R397-399 (2016).

228. S. B. Cohen, R. R. Reddel, A sensitive direct human telomerase activity assay. *Nat Methods* **5**, 355-360 (2008).

229. T. Hamma, A. R. Ferre-D'Amare, The box H/ACA ribonucleoprotein complex: interplay of RNA and protein structures in post-transcriptional RNA modification. *J Biol Chem* **285**, 805-809 (2010).

230. S. Li *et al.*, Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase. *Genes Dev* **25**, 2409-2421 (2011).

231. D. C. Zappulla, T. R. Cech, RNA as a flexible scaffold for proteins: yeast telomerase and beyond. *Cold Spring Harb Symp Quant Biol* **71**, 217-224 (2006).

232. A. Angrisani, R. Vicidomini, M. Turano, M. Furia, Human dyskerin: beyond telomeres. *Biol Chem* **395**, 593-610 (2014).

233. J. W. Brown *et al.*, Plant snoRNA database. *Nucleic Acids Res* **31**, 432-435 (2003).

234. C. J. Bley *et al.*, RNA-protein binding interface in the telomerase ribonucleoprotein. *Proc Natl Acad Sci U S A* **108**, 20333-20338 (2011).

235. C. Chu, J. Quinn, H. Y. Chang, Chromatin isolation by RNA purification (ChIRP). *J Vis Exp* 10.3791/3912 (2012).

236. Y. Chen, J. D. Podlevsky, D. Logeswaran, J. J. Chen, A single nucleotide incorporation step limits human telomerase repeat addition activity. *EMBO J* **37** (2018).

237. R. Sandhu, B. Li, Telomerase activity is required for the telomere G-overhang structure in Trypanosoma brucei. *Sci Rep* **7**, 15983 (2017).

238. M. Heacock, E. Spangler, K. Riha, J. Puizina, D. E. Shippen, Molecular analysis of telomere fusions in Arabidopsis: multiple pathways for chromosome end-joining. *EMBO J* **23**, 2304-2313 (2004).

239. J. M. Watson, D. E. Shippen, Telomere rapid deletion regulates telomere length in Arabidopsis thaliana. *Mol Cell Biol* **27**, 1706-1715 (2007).

240. Y. J. Chiang *et al.*, Telomere length is inherited with resetting of the telomere set-point. *Proc Natl Acad Sci U S A* **107**, 10148-10153 (2010).

241. S. H. Askree *et al.*, A genome-wide screen for Saccharomyces cerevisiae deletion mutants that affect telomere length. *Proc Natl Acad Sci U S A* **101**, 8658-8663 (2004).

242. L. Ungar *et al.*, A genome-wide screen for essential yeast genes that affect telomere length maintenance. *Nucleic Acids Res* **37**, 3840-3849 (2009).

243. T. M. Nakamura, B. A. Moser, P. Russell, Telomere binding of checkpoint sensor and DNA repair proteins contributes to maintenance of functional fission yeast telomeres. *Genetics* **161**, 1437-1452 (2002).

244. S. Francia, R. S. Weiss, F. d'Adda di Fagagna, Need telomere maintenance? Call 911. *Cell Div* **2**, 3 (2007).

245. K. A. Boltz, K. Leehy, X. Song, A. D. Nelson, D. E. Shippen, ATR cooperates with CTC1 and STN1 to maintain telomeres and genome integrity in Arabidopsis. *Mol Biol Cell* **23**, 1558-1568 (2012).

246. S. Amiard, O. Da Ines, M. E. Gallego, C. I. White, Responses to telomere erosion in plants. *PLoS One* **9**, e86220 (2014).

247. V. Locato, L. De Gara, Programmed Cell Death in Plants: An Overview. *Methods Mol Biol* **1743**, 1-8 (2018).

248. M. Huysmans, A. S. Lema, N. S. Coll, M. K. Nowack, Dying two deaths - programmed cell death regulation in development and disease. *Curr Opin Plant Biol* **35**, 37-44 (2017).

249. A. Krishnamurthy, B. Rathinasabapathi, Oxidative stress tolerance in plants: novel interplay between auxin and reactive oxygen species signaling. *Plant Signal Behav* **8**, doi: 10 4161/psb 25761 (2013).

250. J. H. Hong *et al.*, A Sacrifice-for-Survival Mechanism Protects Root Stem Cell Niche from Chilling Stress. *Cell* **170**, 102-113 e114 (2017).

251. K. Tamura, H. Liu, H. Takahashi, Auxin induction of cell cycle regulated activity of tobacco telomerase. *J Biol Chem* **274**, 20997-21002 (1999).

252. S. W. Yang, E. Jin, I. K. Chung, W. T. Kim, Cell cycle-dependent regulation of telomerase activity by auxin, abscisic acid and protein phosphorylation in tobacco BY-2 suspension culture cells. *Plant J* **29**, 617-626 (2002).

253. S. Ren, J. S. Johnston, D. E. Shippen, T. D. McKnight, TELOMERASE ACTIVATOR1 induces telomerase activity and potentiates responses to auxin in Arabidopsis. *Plant Cell* **16**, 2910-2922 (2004).

254. C. Musgrove, L. I. Jansson, M. D. Stone, New perspectives on telomerase RNA structure and function. *Wiley Interdiscip Rev RNA* **9** (2018).

255. A. G. Torres *et al.*, A-to-I editing on tRNAs: biochemical, biological and evolutionary implications. *FEBS Lett* **588**, 4279-4286 (2014).

256. W. Zhou, D. Karcher, R. Bock, Identification of enzymes for adenosine-to-inosine editing and discovery of cytidine-to-uridine editing in nucleus-encoded transfer RNAs of Arabidopsis. *Plant Physiol* **166**, 1985-1997 (2014).

257. A. P. Gerber, W. Keller, An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* **286**, 1146-1149 (1999).

258. S. Tsutsumi *et al.*, Wobble inosine tRNA modification is essential to cell cycle progression in G(1)/S and G(2)/M transitions in fission yeast. *J Biol Chem* **282**, 33459-33465 (2007).

259. A. G. Torres, E. Batlle, L. Ribas de Pouplana, Role of tRNA modifications in human diseases. *Trends Mol Med* **20**, 306-314 (2014).

260. T. Z. Berardini *et al.*, The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis* **53**, 474-485 (2015).

261. T. Hruz *et al.*, Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics* **2008**, 420747 (2008).

262. T. Mengiste, J. Paszkowski, Prospects for the precise engineering of plant genomes by homologous recombination. *Biol Chem* **380**, 749-758 (1999).

263. S. Beedanagari, S. V. Vulimiri, S. Bhatia, B. Mahadevan, "Chapter 43 - Genotoxicity biomarkers: Molecular basis of genetic variability and susceptibility" in Biomarkers in Toxicology, R. C. Gupta, Ed. (Academic Press, Boston, 2014), https://doi.org/10.1016/B978-0-12-404630-6.00043-9, pp. 729-742.

264. Z. Wang, R. Schwacke, R. Kunze, DNA Damage-Induced Transcription of Transposable Elements and Long Non-coding RNAs in Arabidopsis Is Rare and ATM-Dependent. *Mol Plant* **9**, 1142-1155 (2016).

265. M. Armanios *et al.*, Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita. *Proc Natl Acad Sci U S A* **102**, 15960-15964 (2005).

266. F. Goldman *et al.*, The effect of TERC haploinsufficiency on the inheritance of telomere length. *Proc Natl Acad Sci U S A* **102**, 17119-17124 (2005).

267. L. Harrington, Haploinsufficiency and telomere length homeostasis. *Mutat Res* **730**, 37-42 (2012).

268. J. Gohring, N. Fulcher, J. Jacak, K. Riha, TeloTool: a new tool for telomere length measurement from terminal restriction fragment analysis with improved probe intensity correction. *Nucleic Acids Res* **42**, e21 (2014).

269. K. Riha, T. D. McKnight, J. Fajkus, B. Vyskot, D. E. Shippen, Analysis of the G-overhang structures on plant telomeres: evidence for two distinct telomere architectures. *Plant J* **23**, 633-641 (2000).

270. E. Afgan *et al.*, The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* **46**, W537-W544 (2018).

271. L. R. Borlado, J. Mendez, CDC6: from DNA replication to cell cycle checkpoints and oncogenesis. *Carcinogenesis* **29**, 237-243 (2008).

272. M. Das, S. Singh, S. Pradhan, G. Narayan, MCM Paradox: Abundance of Eukaryotic Replicative Helicases and Genomic Integrity. *Mol Biol Int* **2014**, 574850 (2014).

273. M. Menges, J. A. Murray, Synchronous Arabidopsis suspension cultures for analysis of cell-cycle gene activity. *Plant J* **30**, 203-212 (2002).

274. N. S. Heiss *et al.*, X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat Genet* **19**, 32-38 (1998).

275. S. Maas, A. P. Gerber, A. Rich, Identification and characterization of a human tRNA-specific adenosine deaminase related to the ADAR family of pre-mRNA editing enzymes. *Proc Natl Acad Sci U S A* **96**, 8895-8900 (1999).

276. D. Fu, K. Collins, Purification of human telomerase complexes identifies factors involved in telomerase biogenesis and telomere length regulation. *Mol Cell* **28**, 773-785 (2007).

277. A. Gupta *et al.*, Telomere length homeostasis responds to changes in intracellular dNTP pools. *Genetics* **193**, 1095-1105 (2013).

278. L. R. Abdulkina *et al.*, Components of the ribosome biogenesis pathway underlie establishment of telomere length set point in Arabidopsis. *Nat Commun* **10**, 5479 (2019).

279. A. Agorio *et al.*, An Arabidopsis Natural Epiallele Maintained by a Feed-Forward Silencing Loop between Histone and DNA. *PLoS Genet* **13**, e1006551 (2017).

280. J. Ofengand, R. Liou, Correct codon--anticodon base pairing at the 5'-anticodon position blocks covalent cross-linking between transfer ribonucleic acid and 16S RNA at the ribosomal P site. *Biochemistry* **20**, 552-559 (1981).

281. H. Grosjean, V. de Crecy-Lagard, C. Marck, Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Lett* **584**, 252-264 (2010).

282. V. I. Lim, Analysis of action of the wobble adenine on codon reading within the ribosome. *J Mol Biol* **252**, 277-282 (1995).

283. A. Rafels-Ybern, A. G. Torres, X. Grau-Bove, I. Ruiz-Trillo, L. Ribas de Pouplana, Codon adaptation to tRNAs with Inosine modification at position 34 is widespread among Eukaryotes and present in two Bacterial phyla. *RNA Biol* **15**, 500-507 (2018).

284. P. Schimmel, The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nat Rev Mol Cell Biol* **19**, 45-58 (2018).

285. A. Rafels-Ybern, C. S. Attolini, L. Ribas de Pouplana, Distribution of ADAT-Dependent Codons in the Human Transcriptome. *Int J Mol Sci* **16**, 17303-17314 (2015).

286. F. Awwad, G. Bertrand, M. Grandbois, N. Beaudoin, Auxin protects Arabidopsis thaliana cell suspension cultures from programmed cell death induced by the cellulose biosynthesis inhibitors thaxtomin A and isoxaben. *BMC Plant Biol* **19**, 512 (2019).

287. J. Hasegawa *et al.*, Auxin decreases chromatin accessibility through the TIR1/AFBs auxin signaling pathway in proliferative cells. *Sci Rep* **8**, 7773 (2018).

288. S. E. Asberg, A. M. Bones, A. Overby, Allyl isothiocyanate affects the cell cycle of Arabidopsis thaliana. *Front Plant Sci* **6**, 364 (2015).

289. W. R. Chezem, N. K. Clay, Regulation of plant secondary metabolism and associated specialized cell development by MYBs and bHLHs. *Phytochemistry* **131**, 26-43 (2016).

290. E. Gobbini, C. Trovesi, C. Cassani, M. P. Longhese, Telomere uncapping at the crossroad between cell cycle arrest and carcinogenesis. *Mol Cell Oncol* **1**, e29901 (2014).

291. R. E. Verdun, L. Crabbe, C. Haggblom, J. Karlseder, Functional human telomeres are recognized as DNA damage in G2 of the cell cycle. *Mol Cell* **20**, 551-561 (2005).

292. M. D. Vodenicharov, R. J. Wellinger, The cell division cycle puts up with unprotected telomeres: cell cycle regulated telomere uncapping as a means to achieve telomere homeostasis. *Cell Cycle* **6**, 1161-1167 (2007).

293. J. A. Londono-Vallejo, R. J. Wellinger, Telomeres and telomerase dance to the rhythm of the cell cycle. *Trends Biochem Sci* **37**, 391-399 (2012).

294. H. Laprade *et al.*, Single-Molecule Imaging of Telomerase RNA Reveals a Recruitment-Retention Model for Telomere Elongation. *Mol Cell* **79**, 115-126 e116 (2020).

295. A. Romaniuk *et al.*, The non-canonical functions of telomerase: to turn off or not to turn off. *Mol Biol Rep* **46**, 1401-1411 (2019).

296. J. Maceluch, M. Kmieciak, Z. Szweykowska-Kulinska, A. Jarmolowski, Cloning and characterization of Arabidopsis thaliana AtNAP57--a homologue of yeast pseudouridine synthase Cbf5p. *Acta Biochim Pol* **48**, 699-709 (2001).

297. X. Liu, S. Wu, J. Xu, C. Sui, J. Wei, Application of CRISPR/Cas9 in plant biology. *Acta Pharm Sin B* **7**, 292-302 (2017).

298. S. Han *et al.*, RNA-protein interaction mapping via MS2- or Cas13-based APEX targeting. *Proc Natl Acad Sci U S A* **117**, 22068-22079 (2020).

299. H. Ummethum, S. Hamperl, Proximity Labeling Techniques to Study Chromatin. *Front Genet* **11**, 450 (2020).

300. C. Lee *et al.*, Functional partitioning of a liquid-like organelle during assembly of axonemal dyneins. *Elife* **9** (2020).

301. S. Fleurdepine, J. M. Deragon, M. Devic, J. Guilleminot, C. Bousquet-Antonelli, A bona fide La protein is required for embryogenesis in Arabidopsis thaliana. *Nucleic Acids Res* **35**, 3306-3321 (2007).

302. W. H. Shin, K. C. Chung, Human telomerase reverse transcriptase positively regulates mitophagy by inhibiting the processing and cytoplasmic release of mitochondrial PINK1. *Cell Death Dis* **11**, 425 (2020).

APPENDIX

IDENTIFICATION OF ATLA1 AND ATCPN60B AS TELOMERASE-ASSOCIATED SUBUNITS IN ARABIDOPSIS

Summary

Telomerase is a ribonucleoprotein (RNP) complex that prevents the loss of telomeric DNA during chromosome replication using a catalytic telomerase reverse transcriptase (TERT) and a templating telomerase RNA (TR). The recent discovery of the *bona fide* TR from the plant kingdom greatly extends knowledge of plant telomerase. However, the protein composition of plant telomerase RNP is largely unknown. Here, we present the results of quantitative mass spectrometry (qMS) to define accessory components of the *Arabidopsis thaliana* telomerase core. Using epitope-tagged TERT, we analyzed affinity-purified in vivo reconstituted telomerase holoenzyme and discovered that AtLa1 and AtCPN60B are telomerase-associated proteins. This finding not only provides new insight into the composition of plant telomerase, but also provides a new avenue for exploring the evolution of telomerase accessory proteins and coordination of subunits for RNP complexes.

Introduction

Telomerase functions to maintain telomere integrity. The telomerase RNA (TR) assembles with the telomerase reverse transcriptase (TERT) protein to form the catalytic core. TR contains a template for the synthesis of telomere repeat arrays catalyzed by TERT (210). In addition, TR harbors conserved structural domains that serve as a scaffold to retain species-specific accessory proteins that facilitate RNP biogenesis, engagement with the chromosome terminus, and regulation

of telomerase enzyme activity (85). Telomerase accessory proteins are divergent and species-specific. For example, ciliate telomerase retains La-related family protein 7 (LaRP-7) to facilitate RNP assembly (117), while vertebrate telomerase associates with H/ACA RNP components dyskerin-Nop10-Gal1-Nhp2 to promote TR biogenesis and cellular translocation (143). Although some studies have begun to investigate the composition of telomerase-associated proteins in plants (207, 208), the identity of the full complement of core telomerase subunits for the plant enzyme is still unknown. Here, we applied the quantitative mass spectrometry (qMS) technique for purified telomerase in *Arabidopsis thaliana*. Two proteins, AtLa1 and AtCPN60B, were identified as telomerase associated proteins with dramatically distinct function. Our study provides a new framework to uncover associated proteins of RNP complexes. Revealing AtLa1 and AtCPN60B as plant-specific telomerase associated proteins facilitates the interpretation of telomerase evolution and provides hints of the alternate function of telomerase.

Results and Discussion

To identity of the full complement of core telomerase subunits, we applied an unbiased approach to reconstitute and purify the active *A. thaliana* telomerase holoenzyme. Arabidopsis telomerase is not an abundant enzyme. Therefore, we reconstituted telomerase *in vivo* by overexpressing the two essential components: TERT and TR. We created a binary “super-telomerase" vector (pHSN6A01-35S-TSgeneAtTERT-U6-AtTR) that carries a 35S promoter-driven tagged-AtTERT and a U6 promoter-driven AtTR. Tagged-AtTERT was cloned using the genomic AtTERT sequence (5380bp) including all introns with a TwinStrep tag attached to the N-terminus. Because TERT serves as the limiting molecule in telomerase, other accessory

components are typically vastly more abundant and expected to be sufficient for “super-telomerase" *in vivo* reconstitution.

We transformed *tert* null mutants with this construct to create transgenic plants that over-express telomerase. We measured telomerase activity in 12 independent T1 transgenic plants using quantitative telomere repeat amplification protocol (qTRAP). Activity ranged from 2-fold (line 35S 12-5) to 40-fold (line 35S 10-3) higher than wild type (Figure 35A). In addition, for lines expressing the highest telomerase activity, we measured the telomere length of T3 generation plants by terminal restriction fragment analysis (TRF). These experiments confirmed that telomere length was increased relative to untransformed plants (Pierce Young and Dorothy Shippen, unpublished data). These results indicate that expression of AtTERT and AtTR from our “super-telomerase" vector is sufficient for assembly of active telomerase and rescue of short telomeres in *tert* mutants. Several independent T2 lines were selected to obtain a single-copy homozygous “super-telomerase” insertion and were propagated to the T3 generation.

In parallel, we created a “native-telomerase” vector (pHSN6A01-pTERT-TSgeneAtTERT-U6-AtTR) by substituting the 35S promoter for the native AtTERT promoter (3000 bp upstream of TERT transcriptional start). As expected, telomerase activity was restored to wild-type levels in most of these plants (Figure 35B). Similar to “super-telomerase” plants, we propagated the single-copy homozygous “native-telomerase” plants to the T3 generation for future analyses.

For purification of *in vivo* reconstituted telomerase RNP complexes, we modified the protocol for Twin-Strep tag affinity purification. 5 g of material from “super-telomerase” plants or the untransformed negative control (*tert* mutants) was collected, homogenized and purified using Strep-Tactin TX resin. This experiment was independently repeated to obtain a duplicate biological replicate. In collaboration with Drs. Ophelia Papoulas and Edward Marcotte at the

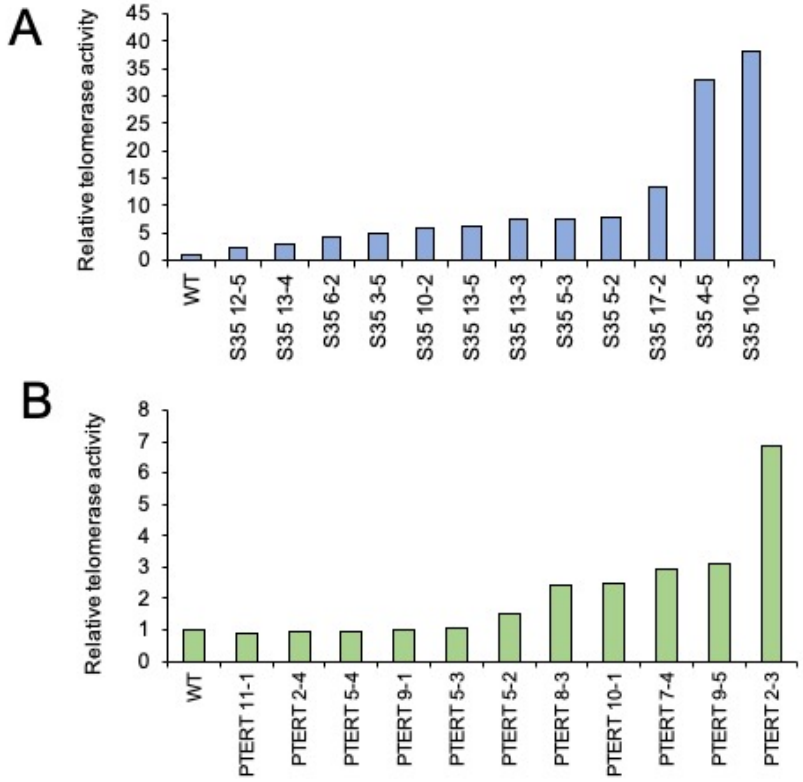


Figure 35. Telomerase activity from plants express "super-telomerase" or "native-telomerase".

(A) qTRAP assay results for 12 independent plant lines carrying the 'super-telomerase' insertion. (B) qTRAP results for 11 independent plant lines carrying the 'native-telomerase' insertion.

University of Texas at Austin, we performed the quantitative mass spectrometry (qMS) analysis with high sensitivity. Affinity purified proteins were trypsin digest to release peptides for the MS/MS detection. Data were collected using a data-dependent top speed HCD acquisition method with full precursor ion scans (MS1) collected at 120,000 m/z resolution. Raw MS/MS spectra were processed by Proteome Discoverer (v2.3) with the UniProt *A. thaliana* reference proteome (UP000006548). To identify proteins statistically significantly associated with each bait, we calculated both a log2 fold-change and a Z-score for each protein based on the observed PSMs in the bait 'super-telomerase' versus control '*tert*' pulldown, similar strategy as shown in (300).

Peptides were quantified to calculate the relative enrichment from "super-telomerase" plants compared to the *tert* null mutant. The results are presented in a scatter plot to highlight the enriched proteins (Figure 36A). In this plot, each axis is on a log-scale with the mean number of peptides spectrally matched across the two corresponding biological replicates. Although AtTERT was massively enriched relative to control plants, it was only the 50th most abundant protein in the "super-telomerase" purification and was far less abundant than Acetyl-CoA carboxylase 1 (ACC1), for example. This observation suggests that the majority of purified proteins were non-specific, because telomerase is not abundant and difficult to enrich during affinity purification.

In addition to TERT, two unexpected proteins were significantly enriched: the genuine La family protein AtLa1 and the protein chaperone AtCPN60B. AtLa1 exhibited a more than 10-fold enrichment compared to untransformed control (*tert* mutants) that barely had any contamination of AtLa1 (Figure 36A, red frame). AtCPN60B had less enrichment than AtLa1 and showed a non-specific contamination from the negative control (Figure 36A, orange frame). To statistically verify this observation, we calculated the relative enrichment of each protein from the individual biological replicates as a Z-score (Figure 36B). In this plot, each axis represents an independent

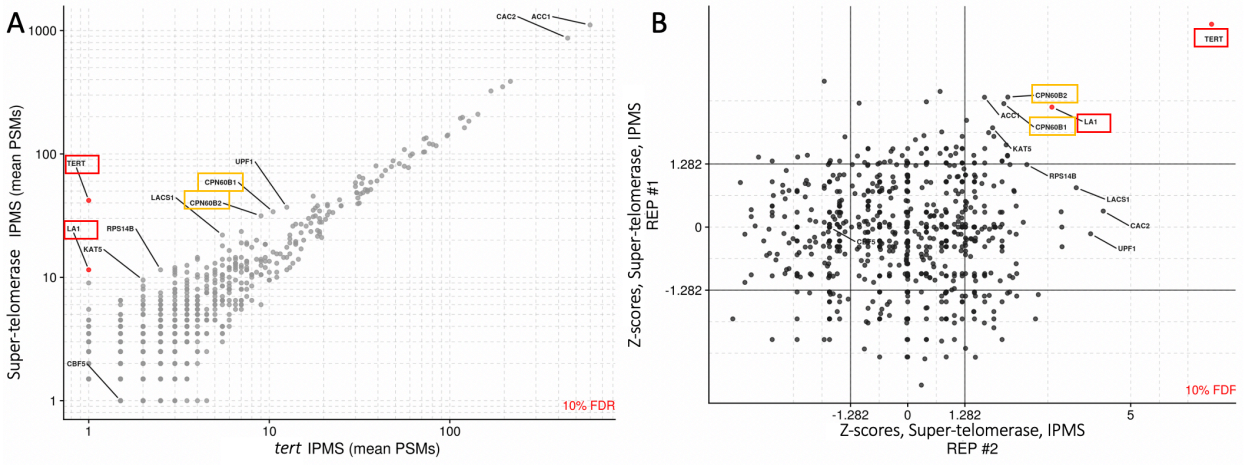


Figure 36. Quantitative MS analysis of proteins enriched from the 'super-telomerase' purification.

(A) a scatter plot to present all qualified peptides with respective protein IDs. AtTERT and AtLa1 are indicated in red to highlight their enrichments compared to the untransformed control (*tert*). AtCPN60B is labelled with orange to indicate a slight enrichment and show a non-specific contamination in the *tert.* (B) Z-scores of independent purifications are present to statistically examine the reproducibility of candidates.

replicate. We found that TERT, AtLa1, and AtCPN60B exhibit significant Z-scores in both replicates, while other proteins such as LACS1, CAC2 and UPF1 are not consistent between replicates. These analyses provide statistical support that AtLa1 and AtCPN60B are significantly co-purified, and hence are tightly associated with telomerase holoenzyme *in vivo*.

La family proteins share a conserved La motif and are found in several telomerase RNP complexes (116, 123, 127). In Tetrahymena, the LaRP-7 protein P65 bends TR helix IV to facilitate the TR assembly with TERT (116). Cryo-EM studies of the Tetrahymena telomerase RNP also identified a stable interaction between P65 and TR that helps to secure P65 as part of telomerase catalytic core (80). Recently, the LaRP-7 protein Pof8 was shown to associate with the fission yeast *Schizosaccharomyces pombe* telomerase (127). Pof8 recruits the Sm complex to the 3' end of yeast TR to promote TR biogenesis and TR protection (128).

No LaRP-7 proteins have been identified in Arabidopsis (124). However, two genuine La family proteins, AtLa1 and AtLa2, have been distinguished through bioinformatic analysis (301). AtLa1, but not AtLa2, restores *Saccharomyces cerevisiae* genuine La nuclear functions in the biogenesis of non-coding RNAs and is able to bind plant RNAs that terminate in 3'-UUU-OH. Consistent with a function in RNP biogenesis, AtLa1 is predominantly localized to the nucleoplasm and can also be detected in the nucleolar cavity. Inactivation of AtLa1 in Arabidopsis is lethal and leads to embryos arrested at an early globular stage of development (301). The biological function of AtLa2 is unclear. However, its inability to rescue the AtLa1 null mutant suggests that AtLa2 is not redundant with AtLa1 and thus may have a discrete function. Our analysis indicates that AtLa1 is tightly associated with the *A. thaliana* telomerase RNP. The notion that Arabidopsis telomerase contains a genuine La protein instead of a LaRP-7 protein sets this enzyme apart from other known telomerase RNP complexes. Nevertheless, this finding is not

entirely unexpected since AtTR is an RNA Polymerase III transcript with a 3' UUU-tail, the typical substrate for genuine La proteins. The detailed mechanism of AtLa1 interaction with AtTR is unknown. However, examining the model of P65-TR interaction in Tetrahymena might provide a useful framework for understanding the interaction and evolution of the AtLa1 interaction with AtTR.

AtCPN60B encodes the beta subunit of the chloroplast chaperonin 60, a homologue of bacterial GroEL. Mutations in this gene result in lesions on plant leaves, expression of systemic acquired resistance (SAR), and accelerated cell death in response to heat shock stress. Unpublished qMS data of Arabidopsis plants revealed an upregulation of AtCPN60B protein abundance in response to environmental stresses (Edward Marcotte' lab, unpublished data). Interestingly, the RNA component of Arabidopsis telomerase (AtTR) was originally described as a hypoxic stress-responsive RNA mostly accumulating in the cytoplasm (214). Whether AtCPN60B and AtTR co-localize and regulate under the same pathway is unknown. In addition, we cannot currently exclude the possibility that AtCPN60B associates reconstituted telomerase RNP in response to the stress of over-expressing telomerase components in the 'super-telomerase' background. The qMS analysis from 'native-telomerase' plants endogenously expressing AtTERT and AtTR will answer this question.

AtCPN60B association with chloroplasts is also interesting, as it indicates TERT might have an alternate function within this compartment that does not require telomerase. In human, TERT has been described to associate with a mitochondrial serine/threonine-protein kinase PINK1 by negatively regulating the cleavage of PINK1 and enhancing its mitochondrial localization (302). Arabidopsis TERT might also regulate AtCPN60B for its function in chloroplasts. In contrast, a different model of AtCPN60B associating telomerase suggests that AtCPN60B might help the

chloroplast translocation of AtTERT. Further analyses including genetics and biochemistry studies are necessary to verify this observation.

In conclusion, using the unbiased approach of telomerase purification and quantitative MS, we identified AtLa1 and AtCPN60B as telomerase-associated proteins in Arabidopsis. This discovery significantly extends knowledge of plant telomerase composition and raises interesting possible avenues for examining alternative functions for TERT and possibility AtTR. Moreover, if we can verify that the *A. thaliana* core RNP contains a genuine La protein, unlike LaRP-7 proteins that have been specifically designated for telomerase function in other organisms, we will have many new opportunities for exploring TR biogenesis, RNP evolution and the co-evolution of TR associated partners.

Materials and methods

Plant material, growth conditions, and transformation

Arabidopsis thaliana accession Col-0 and *attert* (SALK_041265C) were used in this study. Seeds were sterilized in 50% bleach with 0.1% Triton X-100 and then plated on half Murashige and Skoog (half MS) medium with 0.8% agar. Plants were grown at 22°C under long day light conditions. For genetic complementation, pHSN6A01-35S-TSgeneAtTERT-U6-AtTR and pHSN6A01-pTERT-TSgeneAtTERT-U6-AtTR were transformed into second generation AtTERT-/- plants using Agrobacterium-mediated (*A. tumefasciens* GV3101) transformation. Transformants were selected on hygromycin in T1 and analyzed for telomerase phenotypes. In parallel, untransformed AtTERT +/+ and AtTERT -/- plants were analyzed.

"Super-telomerase" purification

5 g of Arabidopsis flowers from "super-telomerase" plants or *tert* plants were ground in liquid nitrogen and homogenized in 20 mL lysis buffer (100 mM Tris-OAC pH 7.5, 150 mM KGlu, 5 mM $MgCl_2$, 0.1% Triton X-100, 20 mM EGTA, 15g/L PVP, 10% glycerol, 20 μL/mL Plant protease inhibitor mixture [Sigma-Aldrich], 1 μL/mL RNaseOUT [Thermo Fisher Scientific], and 2 mM DTT). After centrifugation to remove the precipitant, the supernatant was loaded onto a Strep-Tactin XT 4 Flow gravity column (0.2 mL column volume) (iba: 2-5031-005) at 4°C with a speed at 1-1.5 CV/min. Then, the column was washed with 2.5 mL wash buffer 1 (100 mM Tris-OAC pH 7.5, 150 mM KGlu, 5 mM $MgCl_2$, 0.1% Triton X-100, 10% glycerol, 20 μL/mL Plant protease inhibitor mixture [Sigma-Aldrich], 1 μL/mL RNaseOUT [Thermo Fisher Scientific], and 2 mM DTT), followed by 2 mL wash buffer 2 (100 mM Tris-OAC pH 7.5, 150 mM KGlu, 5 mM $MgCl_2$, 10% glycerol, 20 μL/mL Plant protease inhibitor mixture [Sigma-Aldrich], 1 μL/mL RNaseOUT [Thermo Fisher Scientific], and 2 mM DTT), and finished with 0.6 mL wash buffer 3 (100 mM Tris-OAC pH 7.5, 150 mM KGlu, 5 mM $MgCl_2$, 10% glycerol, and 2 mM DTT). Finally, the column was eluted with 8 CV of 20 mM NaOH. Each elution fraction was neutralized with 6 μL of 1M Bis-Tris-HCl pH 5.8 to pH to 7.0-7.5, combined, and concentrated to 50-100 μL by 5 kDa concentrator.

Terminal Restriction Fragment (TRF) and quantitative TRAP (qTRAP)

TRF was performed as previously described (202). qTRAP assays was modified from the previous publication (205). Briefly, unopened flower bundles were collected as initial materials. The partially purified telomerase was incubated with corresponding reactions at room temperature

for 45 min instead of 37°C. This room temperature incubation mimicked the growing condition of *A. thaliana*.

Quantitative mass spectrometry (qMS) analysis

Peptides were separated using reverse phase chromatography on a Dionex Ultimate 3000 RSLCnano UHPLC system (Thermo Scientific) with a C18 trap to Acclaim C18 PepMap RSLC column (Dionex; Thermo Scientific) configuration and eluted using a 3% to 40% gradient over 60 min. with direct injection into a Thermo Orbitrap Fusion Lumos Tribrid mass spectrometer using nano-electrospray. Data were collected using a data-dependent top speed HCD acquisition method with full precursor ion scans (MS1) collected at 120,000 m/z resolution. Monoisotopic precursor selection and charge-state screening were enabled using Advanced Peak Determination (APD), with ions of charge 2-6 selected for high energy-induced dissociation (HCD) with stepped collision energy of 30% +/- 3%. Dynamic exclusion was active for ions selected once with an exclusion period of 20 s. All MS2 scans were centroid and collected in rapid mode. Raw MS/MS spectra were processed using Proteome Discoverer (v2.3) using the UniProt *Arabidopsis thaliana* reference proteome, UP000006548.