GENE EVOLUTION AND FUNCTION IN *ARABIDOPSI*S TELOMERE BIOLOGY SYSTEM

A Dissertation

by

CALLIE REBEKAH KOBAYASHI

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,	Dorothy E. Shippen
Committee Members,	Mary Bryk
	Jean-Philippe Pellois
	Libo Shan
Head of Department,	Dorothy E. Shippen

December 2018

Major Subject: Biochemistry

ABSTRACT

Telomeres protect chromosome ends from being recognized as double-strand breaks, and respond to incomplete end-replication through telomerase-mediated extension. POT1 (Protection of Telomere 1) is a highly conserved protein required for capping chromosome ends and for regulating telomere extension by telomerase. *Arabidopsis thaliana* encodes three POT1 paralogues: POT1a, POT1b, and POT1c. POT1a functions to maintain telomere length homeostasis by promoting telomerase processivity, while POT1b functions in the DNA damage response. POT1c is derived from a recent duplication of the POT1a locus, but its function is unknown.

In this dissertation, I examined the function and evolution of POT1c using genetic and biochemical approach. Unlike *pot1a* mutants which show defects in telomere maintenance, plants lacking POT1c exhibit no obvious telomere-related or developmental phenotypes. Furthermore, the POT1c gene is not expressed under standard growth conditions. Transposable elements (TE) are embedded in the POT1c promoter region; yet, active silencing is not observed. Although POT1a and the dS17 gene, which was created in the same duplication event that gave rise to POT1c, are highly conserved among *A. thaliana* accessions, POT1c is not. Comparison of POT1a and POT1c loci from species closely related to *A. thaliana* and *A. lyrata* indicates that POT1c initially had a functional promoter that was subsequently disrupted by TE insertion. Together, these studies provide new insights into the fate of newly duplicated genes, and the importance of proper regulation of telomere proteins.

In addition to the study of the POT1c locus, I have analyzed a newly identified gene (NOP2A) that is implicated in the control of telomere length set point. NOP2A is a conserved rRNA methyltransferase protein that positively correlates with cell proliferation. Telomere length is variable across eukaryotes, but each species establishes a specific set point that allows full protection for chromosome ends. My research shows that mutation in NOP2A locus leads to shorter, but stable telomere length in the Col-0 accession of *A. thaliana*. These findings provide strong evidence that additional genes that regulate telomeres remain to be discovered.

DEDICATION

This dissertation is dedicated to my mother Sugie Kobayashi for her unconditional love and support, and my aunt and uncle Susan and Richard Frazier for being my second parents.

ACKNOWLEDGEMENTS

The past seven years have been a long journey. I cannot be grateful enough to all the people who have been involved in this process. From mentors who taught me how to be an independent and confident scientist, to friends and families who were there by my side during my ups and downs.

My first and foremost acknowledgement goes to my PI, Dr. Dorothy Shippen. The opportunity she has given me to join her lab and be mentored by her is something irreplaceable. She has always known how to motivate and push me to be the best scientist I can be. She has taught me confidence and perseverance, to never give up on something you believe in. While she was my academic mentor, she was also my life mentor: to be a strong female role model and to take charge of my life. She has always believed in me and has never given up on me, and for that I cannot thank her enough. As she is my role model, I will take her teaching and kindness to pass it on as I will move on to next chapter of my life.

I am also grateful for the guidance I have received from the faculty and staff in our department. Especially my committees: Dr. Mary Bryk, Dr. Jean-Philippe Pellois, and Dr. Libo Shan, who have challenged and critiqued me continuously to allow me to become the scientist I am today. In addition, Dr. Geoff Kapler, Dr. Xiuren Zhang, and Dr. Micheal Polymenis for their insights during our joint lab meetings. I also want to thank Rafael Almanzer for his consistent care and support he has given to all the graduate student in this department.

Thanks also goes to mentors who have inspired me in many ways. Dr. Claudia Castillo-Gonzalez was my first mentor/ friend in our department. She was still a graduate student when I rotated in Dr. Xiuren Zhang's lab, but she taught me consistency in science and kindness in life. She has also questioned and challenged me on my ideas to help become a creative thinker. Dr. Katie Leehy was my mentor during my rotation in Dr. Shippen's lab, and remained as my mentor after joining the Shippen lab. She never held anything back, and she taught me critical thinking and experimental planning. When she performed an experiment, she made sure it was executed perfectly with all the proper controls. Another mentor from Shippen lab is Dr. Andrew Nelson. It would be accurate to say that he was one of the reason I joined the Shippen lab. He taught me how to be excited with every single experiment you performed. He has also taught me the love for science even on days science is just not on your side. I will treasure this experience, and I hope I can become as great of a mentor to someone someday.

I would also like to thank other members of Shippen lab for their intellectual input and the friendship we have built: Kara, Jong-Ro, Kyle, Peter, Xintao, Eugene, Vicky, Behilu, Sreya, and Jerry. Thanks to my undergrads Haran, Juliana, Simon, and Eli for their help with my projects, and to allow me to be their mentor. Special thanks to Xiaoyuan and Borja. Xiaoyuan has become one of my very good friends. Something that started as wine tasting has become joy of eating food, and discussion of our project. Borja has given me smiles even on my bad days, and for that I cannot thank him enough.

Thank you to my friends, who I knew before graduate school and who I have bonded during graduate school. Without all of you, I would not have been able to go through this stressful time. Especially my best friend Charli and Lauren, who have always

been my side from the beginning to the end, and has understood me and stood by my side during my difficult times. You guys are the reason I have not given up.

Finally, I do not know how grateful I can be to have such an amazing family who has loved me unconditionally. My mother, who has worked hard so that I can have my education and encouraged me to follow my dreams. My aunt and uncle, Susan and Richard Frazier, who has been my second parents in the United States. They have made me feel welcome and comfortable in the United States, allowing me to never feel alone.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Professors Dorothy Shippen, Mary Bryk, and Jean-Philippe Pellois of the Department of Biochemistry and Biophysics and Professor Libo Shan of the Department of Plant Pathology and Microbiology.

The developmental analysis data for Chapter III was in part provided by Dr. Claudia Castillo-Gonzales and Borja Barbero.

The data for Chapter IV was in part provided by collaborators Dr. Eugene Shakirov, Dr. Thomas Juenger, Dr. John Lovell, Dr. Liliia Nigmatullina, Dr. Inna Chastukhina, Dr. Behilu Akilu, and Inna Agabekyan.

All other work for the dissertation was completed independently by the student. This work was made possible in part by N.I.H. under R01-GM065383 and CRFD Global to D. E. S., and N.I.H. under R01-GM127402-01 and R03-52891-02 to E. S.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Chromosome ends and their problems

In 1961, Leonard Hayflick proposed that there are limits to the number of times normal human fetal diploid cells can divide. He demonstrated that cells can divide 40-60 times before they arrest[1]. The idea of a limited cell life span was coined the "Hayflick Limit". In support of the concept that normal cells are not immortal, it was later shown that cells that stop dividing enter a physiological state termed senescence[1,2]. We now know that one of the reasons cells cannot divide infinitely is because a bit of genetic material from the end of chromosomes is lost after each cell division[3–5]. A mechanistic link between the termination of cell division and critically short chromosome ends has been proposed to trigger cellular senescence, and to be a cause of aging in humans [4,6,7]. However, this theory does not explain: (1) the difference between normal and immortal cells, and (2) why single celled organisms such as bacteria replicate infinitely, while normal human cells do not. To address these issues, research has focused on solving two major mysteries related to chromosome ends: the end replication problem and the end protection problem[8–12].

Unlike prokaryotic chromosomes, eukaryotic chromosomes are linear. As discussed below, the conversion of genomes from circular to linear demands a new mechanism for fully replicating the end of a chromosome and preventing the end from being recognized as a DNA double-strand break (DSB).

The end protection problem

The concept of the end protection problem surfaced in the 1940s when Barbara McClintock demonstrated that broken chromosomes in maize are unstable [13]. Subsequent studies showed that chromosomes lacking their natural ends are subjected to degradation by nucleolytic attack, and end-to-end fusion by being recognized as DSBs by DNA repair mechanisms[14–17]. DSB signaling activates the master kinases ATM and ATR, which induce cell cycle arrest. If not repaired, unprotected chromosome ends cause permanent cell cycle arrest and compromised genome integrity. Around the same time as McClintock's pioneering work in maize, Herman Muller demonstrated that chromosome ends in fruit flies were naturally stabilized. He termed the protective structure on chromosome ends, the telomere.

The end-replication problem

When the mechanism of DNA replication was revealed in the early 1970s, the question of the end replication problem arose. DNA replicates in a semi-conservative manner with DNA polymerase moving in the 5' to 3' direction, after the reaction is initiated by extension from an RNA primer[18]. Removal of the RNA primer creates a gap in the newly synthesized DNA between Okazaki fragments, which is then filled-in by a DNA polymerase (Figure 1-1). However, the extreme 5' end of lagging strand lacks an upstream Okazaki fragment to provide a 3' end for extension, which leaves the end of the nascent strand with a gap that cannot be filled (Figure 1-1). Thus, newly synthesized DNA strands become shorter. Each time the cell replicates there will be a gradual shortening of chromosomal DNA[3].

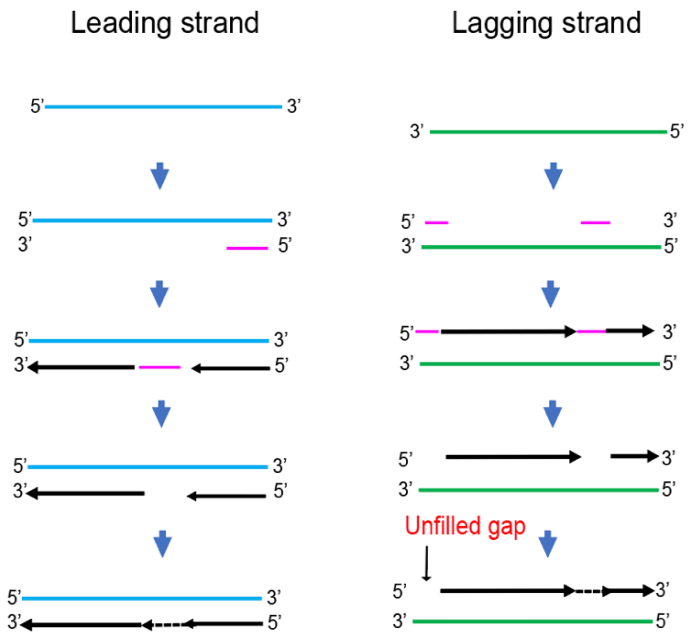


Figure 1-1. DNA replication. Model for semi-conservative DNA replication. Replication of the leading strand produces blunt ends, which are then processed by a 5' to 3' exonuclease to create a 3' overhang. Replication of the lagging strand naturally produces a 3' overhang once the final RNA primer is removed from the 5' end of the nascent strand.

In almost all eukaryotes, the end-replication problem is solved by the telomerase reverse transcriptase. Typically, the sequence at the end of the chromosome comprises simple tandem G-rich repeats. At the very ends of the chromosomes these G-rich repeats are synthesized by telomerase. The continual addition of telomere repeats by telomerase prevents the shortening of chromosome ends after DNA replication of the bulk chromosomes, thereby compensating for the end-replication problem[11,12]. In addition, telomere-associated protein complexes bind to the chromosome terminus, serving as a physical cap on the terminus and recruit telomerase. Thus, telomere-associated protein complexes help to solve both the end-protection problem and the end-replication problem.

Telomeres and disease

Telomeres and telomerase are important in maintaining genome integrity, and regulation of telomere length as well as telomerase enzyme activity is critical. Their misregulation leads to stem cell diseases and cancer. In the 1990's telomeres were linked to cellular senescence, as telomere length strongly correlated with the internal clock for the Hayflick limit. Thus, telomere length was shown to correlate with the number of times cells can divide. In somatic cells, where telomerase activity is low or absent, telomeres become shortened due to the end replication problem, causing cells to undergo replicative senescence or apoptosis[7,19]. This is not the case in germ cells where telomerase is highly active[20–22]. A large body of data indicates that a major reason why germ cells, stem cells, and cancer cells are highly proliferative is because they have a mechanism to solve the end-replication problem.

Conversely, when telomerase is inappropriately activated in non-proliferating cells, tumorigenesis can occur. Indeed, malignant cells, unlike their normal counterparts, maintain telomere length, primarily through the activation of telomerase[20,23–25]. Telomerase activity has been detected in more than 70% of *in vitro* immortalized human cell lines as well as ~85% of human tumors[26]. The remaining 15% of cancers maintain their telomere length through a DNA recombination mechanism termed alternative lengthening of telomeres (ALT)[26–33]. Mutations in genes that lead to short telomeres have been shown to cause a large number of proliferation-related diseases, including chronic lymphocytic leukemia, familial melanoma, and sporadic melanoma[34–38]. This is likely due to the genome instability that arises from telomere deprotection. Stabilization of telomeres is a

crucial process in gaining an immortal phenotype [3,39,40]; therefore, limiting the replicative potential of cells by inhibiting telomere maintenance aids in tumor suppression [40–42].

While misregulation of telomerase and/or telomere length is a rate-limiting step for cancer, critically short telomeres can also lead to age-related diseases and premature aging syndromes[22,24,25,43,44]. In mouse studies, loss of telomerase causes decreased tissue regeneration, viability and fertility, immune system failure, enhanced myocyte apoptosis and reactive hypertrophy leading to heart failure[45–49]. Age-related pathologies have also been associated with telomere dysregulation in humans. These include heart disease, ulcerative colitis, liver cirrhosis, and atherosclerosis[50–53]. Aplastic anemia and Coat's plus, a pleiotropic multisystem disorder, are examples of premature aging diseases caused by mutations in telomere-related genes[54–58]. Dyskeratosis congenita (DC), a progressive bone marrow failure disease, is another example of a premature aging disease caused by mutations in telomerase[59–64]. DC patients suffer from other organ failures, including pulmonary fibrosis, liver cirrhosis, and esophageal stricture[65–67]. Interestingly, recent studies in mice suggest a link between telomere gene mutations, and an alteration in metabolic pathways and obesity [68]. Collectively, the diseases associated with telomere failure have been termed "telomereopathies".

Telomere features and structure

The sequence and the number of telomere repeats found at chromosome ends vary among different species. Canonical G-rich telomere sequences are well conserved in eukaryotic evolution[69]. Specifically, TTAGGG is found in vertebrates, TTTAGGG in

plants, $(TG_{1-3})_n$ in *Saccharomyces cerevisiae*, and $(TTAC(A)(C)G_{1-8})_n$ in *Schizosaccharomyces pombe*[70–74].

While canonical telomeres and their extension seems to be the most common mechanism for solving end-replication and end-protection problems, there are exceptions[73,75–81]. These include terminal repeat (TR) sequences in linear viral DNA, covalently closed hairpin ends (hairpin-based termini) in linear bacterial DNA, and telomere-specific transposable elements in dipteran insects[77,82,91,83–90].

Telomere features and higher-order structure

Telomeres are structured in a way that allows chromosome ends to be protected, and also allows access for elongation when required. These structural features include a single stranded 3' extension termed the G-overhang and a t-loop[92] (Figure 1-2A).

The G-overhang acts as a substrate for telomerase to elongate telomeres. It also recruits proteins that are involved in protecting chromosome ends[93,94]. While the G-overhang is naturally produced in lagging-strand replication, leading strand replication yields blunt ends (Figure 1-1). The blunt end is converted into a G-overhang by exonucleases, such as Apollo, which removes nucleotides 5' to 3' to create a 3' extension, and thereby symmetrical chromosome ends[95–98]. Interestingly, Apollo exonuclease appears to be absent in *Arabidopsis thaliana*. In *A. thaliana* half of the telomeres are blunt-ended or possess extremely short 1-2 nt overhangs, while the other half contain a G-overhang of 20-30 nucleotides(Figure 1-2A)[99,100].

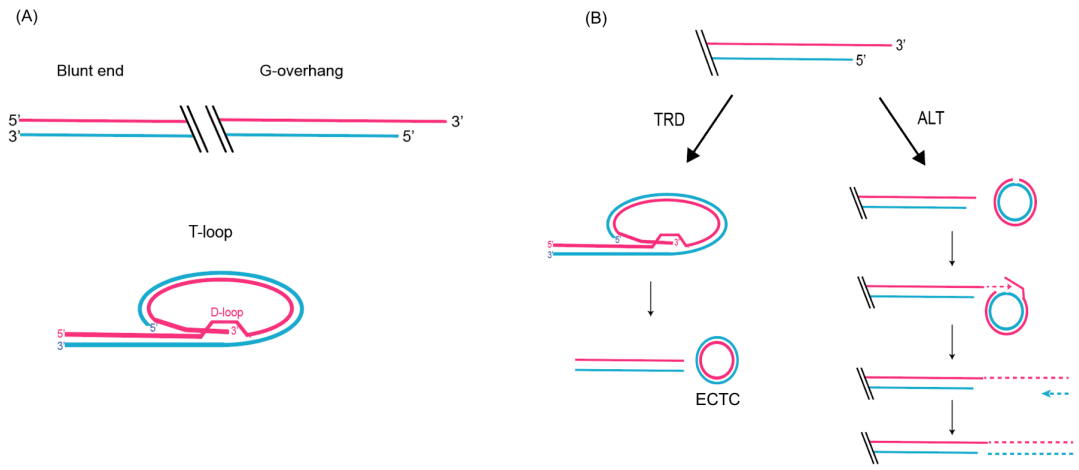


Figure 1-2. Telomere features and higher-order structures. (A) G-overhang (right): the 3' terminus contains a G-rich single-strand extension created by lagging strand replication. Blunt end (left): the chromosome terminus generated by leading strand replication. T-loop: the G-overhang can invade the DNA duplex to create a D loop structure. (B) TRD: telomere rapid deletion due to t-circle resolution by homologous recombination machinery. The telomere is shortened and an extra-chromosomal telomeric circle (ECTC) is generated. ALT: alternative lengthening of telomeres through rolling-circle replication using an ECTC as a template.

Single-stranded G-overhangs are able to fold back on themselves and invade into the double-stranded telomeric region to form a higher order structure called the t-loop. The t-loop sequesters and protects the chromosome end from being recognized as DNA damage response (DDR)[101,102]. A t-loop can also function to regulate telomeres that are too long through deletion of telomeric DNA sequences by TRD (Figure 1-2)[103].

One of the unique features of Arabidopsis is that unlike most organisms, which carry G-overhang on both ends of their chromosomes, *A. thaliana* chromosome ends are asymmetrical. Approximately 50% of the chromosome ends have G-overhangs, but the remaining telomeres are blunt-ended (Figure 1-2A)[99]. In Arabidopsis, the G-overhang is

protected by the telomere binding protein complex CST, while the blunt end is protected by Ku and is not processed into a G-overhang.

Telomere length set point

G-rich telomere repeats extend from the G-overhang into the duplex region of the chromosome terminus. The total number of G-rich repeats varies across evolution, but each species has a fixed average length of telomeric DNA that is maintained through successive generations[104,105]. Telomere lengths vary, ranging from 2-9 kb in *Arabidopsis thaliana* [72,106], 40-160kb in *Nicotiana tabacum*[74], 250-300bp in yeast[73], 50-150kb in mouse [107], and 5-15kb in humans[108]. Many genes have been implicated in determining the species-specific telomere length set point to establish telomere length homeostasis[109,110]. A balance between the replicative erosion and elongation by telomerase is essential to establish the set point[111].

While end replication and end protection problems play a major role in telomere length homeostasis, there are additional mechanisms that can alter telomere length including homologous recombination (Figure 1-2B). When telomeres become too long, they can undergo a process termed telomere rapid deletion (TRD) through t-loop resolution, causing dramatically shortened telomeres[112]. A t-loop resembles a Holliday junction intermediate and can be resolved by homologous recombination (HR). Control of TRD is important to prevent over-shortening of telomeres, and this is mediated by a component of the telomere capping complex[113–115]. When t-loop resolution occurs it creates extra-chromosomal telomere circles (ECTCs) as a by-product. ECTCs have been found in mammals, yeast, and plants[103,116–119]. On the flip side of TRD, Alternative

lengthening of telomere (ALT) can cause telomere elongation by using the ECTCs as a template for rolling circle extension [120]. The ALT pathway of telomere maintenance is independent of telomerase and has been seen among various eukaryotes[121,122].

Telomerase, its accessory proteins, and its regulation

Telomerase

Telomerase is a ribonucleoprotein (RNP) reverse transcriptase that is responsible for extending G-overhangs, thus solving the end-replication problem [123]. The minimal catalytic core of telomerase consists of a protein reverse transcriptase (TERT/TRT/Est2) and a template containing long noncoding RNA (lncRNA) (TER/TERC/TLC1) [123–132]. LncRNAs are defined as transcripts longer than 200 nt that do not code for protein[133–135]. Loss of either the catalytic protein or the template lncRNA component results in an "Ever Shorter Telomere" (EST) phenotype, where the telomeres become shorter and shorter each generation[130,136].

TERT contains four highly conserved domains: the telomerase essential N-terminal (TEN) domain, the telomerase RNA binding domain (TRBD), the reverse transcriptase (RT), and the C-terminal extension (CTE)[137]. In contrast to TERT, the TER lncRNA is highly variable in size and sequence. Nevertheless, all TER molecules have conserved structural core elements (Figure 1-3)[138–140]. First is the telomere complementary sequence that acts as a template to direct nucleotide addition onto the G-overhang[125,130,141]. The second structure is the pseudoknot domain.

While the pseudoknot in ciliate TERs is dispensable, a pseudoknot is critical for proper enzymatic activity of vertebrate and fungal telomerases[17–20]. Finally, the third

structure, the three-way-junction element, serves as the binding hub for accessory proteins. TERT and TER alone are sufficient for reconstitution of telomerase activity *in vitro*; however, the holoenzyme requires additional components *in vivo*. These accessory factors are important for proper RNP biogenesis, assembly, trafficking, and regulation of telomerase[137]. TER acts as a scaffold for many of these accessory factors.

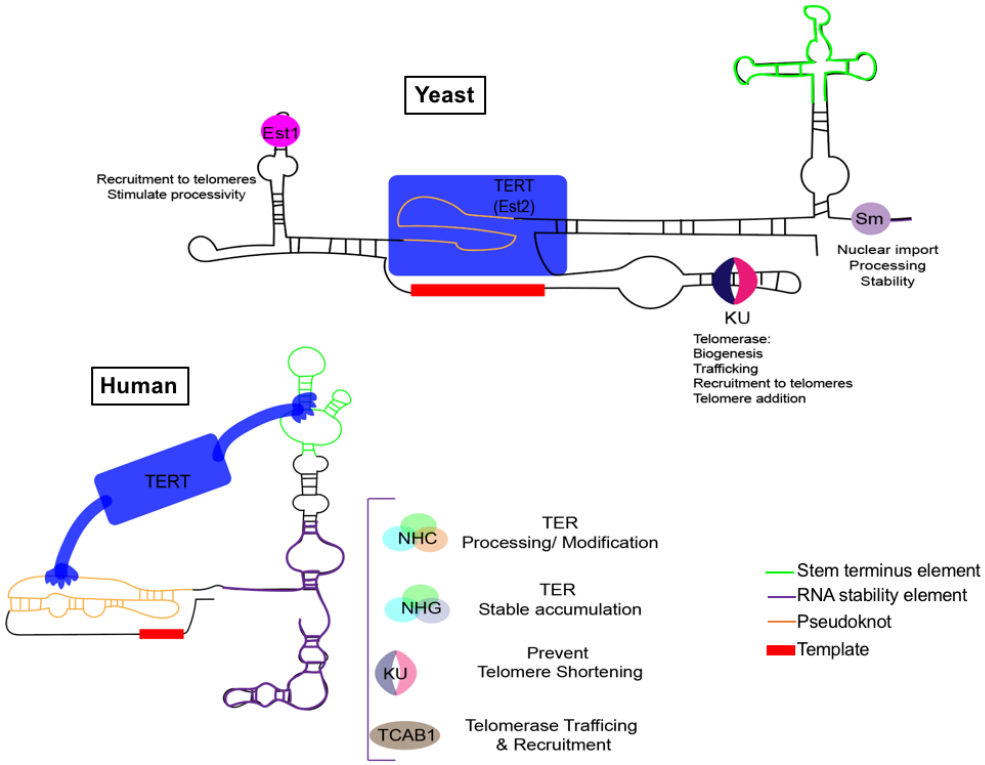


Figure 1-3.Telomerase RNA and Telomerase RNA binding proteins. The telomerase RNA and RNA binding proteins of *S. cerevisiae* and humans. Critical RNA elements are indicated in color: stem terminus element (green), Pseudoknot (orange), and template (red). Proteins responsible for RNA stability and biogenesis are bound at the RNA stability element (Blue). Yeast: Sm protein is responsible for nuclear import, processing, and stability of the telomerase RNA. Human: different proteins can associate with RNA stability element (purple). These proteins are responsible for RNA processing (NHC), stable accumulation (NHG), prevention of telomere shortening (KU), and telomerase trafficking and recruitment (TCAB1).

Telomerase accessory proteins in RNA biogenesis and trafficking

The well characterized yeast and human systems of telomerase biogenesis will be discussed in this section[139,141,146,147]. In yeast, TER (TLC1) is processed in nucleoli, assembled into an enzymatically active telomerase in the cytoplasm, trafficked into nucleoplasm, and then onto telomeres by the help of Ku[141,148–150]. Ku is a heterodimer that binds double-strand DNA ends and required for non-homologous end joining[151]. Ku interaction with the telomerase RNA is important for stable accumulation of telomerase, and recruitment of telomerase to the telomeres[147,152,153]. TLC1 is bound at its 3' end by the Sm protein, which functions in nuclear import, processing, and stability of RNA[131,154,155]. Depletion of Sm in yeast leads to a drastic decrease in telomerase RNP accumulation, but seems to be dispensable for human telomerase function (Figure 1-3)[154].

In humans, TER (hTR) is modified in the Cajal body, a nuclear sub-organelle largely consisting of proteins and RNA[141,156–161]. The H/ACA motif of hTR binds a series of accessory factors that promote telomerase RNA assembly. The NHG complex (NOP10, NHP2, and GAR1) processes hTR, while the NHC complex (NOP10, NHP2, and Cbf5/dyskerin) is required for stable accumulation of hTR[162–167]. Another H/ACA motif binding protein is TCAB1. While TCAB1 is not important for telomerase assembly and activity, it is a major factor for telomerase trafficking between nucleoli and Cajal bodies and recruitment to telomeres (Figure 1-3)[157,168]. Dyskerin is a component of active telomerase in humans[167]; however, TLC1 in yeast does not associate with dyskerin, and depletion of yeast dyskerin does not have an effect on the yeast TER level[164]. *Arabidopsis thaliana,* like humans, has a dyskerin homolog. Mutation of *A. thaliana* dyskerin leads to

deregulated short telomeres, suggesting that involvement of dyskerin in telomere maintenance is conserved in multicellular organisms[59,169].

Telomerase accessory proteins in regulation and activity

An enzyme possesses processivity when it has the ability to catalyze a reaction without releasing its substrate. The telomerase reverse transcriptase displays two types of processivity, nucleotide addition processivity and repeat addition processivity[170,171]. Nucleotide addition processivity requires the simultaneous translocation of the enzyme active site from the RNA-DNA duplex as each nucleotide is being added. Repeat addition processivity requires translocation of enzyme from the RNA-DNA duplex as each round of full repeat synthesis is completed. While nucleotide addition processivity is commonly associated with DNA polymerases, processive repeat addition is unique to telomerase, and requires telomerase-specific elements.

Proteins involved in the positive regulation of telomerase include Est1 and Est3 (budding yeast), p50 (*Tetrahymena thermophila*), and POT1a (*A. thaliana*)[141]. Cells lacking these proteins show an EST phenotype similar to cells lacking TERT or TER[5,130,136,172]. The Est1 and Est3 proteins from *S. cerevisiae* are dispensable *in vitro,* but are required *in vivo*[173,174]. Est1 binds TLC1[5,175,176], and was initially thought to be a recruitment factor for telomerase. It has since been shown to stimulate telomerase by activating telomerase at telomeres[177]. A protein with functional similarities to Est1 is *A. thaliana* POT1a. AtPOT1a is associated with the telomerase RNP, and is responsible for stimulating telomerase activity through enhancing repeat addition processivity[132,172,178]. Unlike Est1 and AtPOT1a, Est3 binds TERT instead of TLC1[179]. Est3 acts downstream of

TERT and Est1 for telomerase activation, and loss of Est3 leads to a decrease in nucleotide addition[180,181]. *T. thermophila* p50 is another telomerase accessory protein that stimulates processive repeat addition[182]. While Est1 and AtPOT1a seem to be functional homologs, Est3 and p50 are structural and functional homologs of human telomerase stimulating factor TPP1[182–185]. Unlike Est3 and p50, TPP1 is not a component of telomerase, but rather is a stable component of telomeres. TPP1 will be discussed further below.

Another interesting protein associated with telomerase is Ku. Ku has functions in diverse processes, most notably in DNA repair, immune system gene rearrangement, apoptosis, and telomere biology[153,186,187]. Although Ku mutants exhibit telomere defects in various organisms including *S. cerevisiae*, *S. pombe*, *A. thaliana*, mice, and trypanosomes, the precise role of Ku is slightly different among different organisms[153]. For example, in *S. cerevisiae* Ku is involved in telomerase biogenesis, trafficking, recruitment to telomeres, and telomere addition[150,152,153,188–191]. The absence of Ku leads to TLC1 accumulation in the cytoplasm and short telomeres[150,152]. In contrast, human Ku appears not to be involved in telomere length regulation, while the Ku protein in mice and *A. thaliana* is implicated in negative regulation of telomerase[192–196].

DNA double-strand breaks and de novo telomere formation by telomerase

When a DNA DSB occurs, the broken chromosome is typically repaired through the non-homologous end joining (NHEJ) or homologous recombination (HR) pathway[197]. Failure to properly repair a DSB is linked to increased cell death, cell-cycle arrest, telomere defects, and meiotic defects[198,199]. Therefore, a defect in DSB repair can compromise cell survival and the maintenance of genome integrity. Telomerase can

compete with DNA repair machinery at the site of DSBs, leading to the addition of telomere repeats and the establishment of a new telomere. This process, termed *de novo* telomere formation (DNTF) or chromosome healing, has been reported in many organisms including *T. thermophila*, yeast, mammals, and plants[200–205].

Although DNTF can stabilize centromere-containing DNA, it will lead to the loss of non-centromere-containing fragments. Typically, only very small terminal deletions are viable, and studies have revealed that DNTF is associated with genetic disorders and mental retardation in humans[206,207]. Thus, control of telomerase action at internal chromosome sites is vital for genome stability and organismal viability.

To prevent unwanted telomere formation, organisms have acquired mechanisms to prevent DNTF. In budding yeast, Mec1 (an ATR homolog and a check point protein kinase) and Pif1 (DNA helicase) are involved in DNTF inhibition[208–211]. Mec1 phosphorylates the telomere binding protein Cdc13, thus inhibiting Cdc13 accumulation at DSBs. Mec1 is responsible for Pif1 activation[210,212], which results in the unwinding of RNA-DNA hybrids *in vitro*, and is postulated to dislodge telomerase from its DNA substrate *in vivo*[213]. In addition, MLH1 (DNA mismatch repair) has been shown to inhibit DNTF in human cells, although the mechanism is still uncertain[214].

A. thaliana appears to have evolved a unique mechanism to regulate DNTF by directly inhibiting telomerase activity in response to DSBs using a lncRNA[215,216]. LncRNAs are best known for their role in regulating gene expression, through mechanisms that involve their use as signals, decoys, guides, and scaffolds[217,218]. *A. thaliana* is unusual as its telomerase is associated with a non-templating lncRNA termed TER2, that functions to negatively regulate telomerase activity (see below). Although the exact mechanism of

inhibition is still under investigation, loss of TER2 leads to an increase in telomerase activity and the failure to down regulate enzyme activity in response to DSBs[215,216]. Thus, TER2 is proposed to prevent unwanted telomerase activity at the DSB by globally down regulating enzyme activity.

Telomere-associated proteins

While telomerase is responsible for the physical extension of telomeric DNA, telomere binding protein complexes have the dual function of end protection and telomere length control (Figure 1-4)[219–221]. The first telomere end binding proteins (TEBP) were characterized in ciliate protozoa, and were shown to serve as a protective barrier for the G-overhang[222,223]. In budding yeast and in *A. thaliana*, the G-overhang is associated with CST (CTC1/ Cdc13, STN1, TEN1)[219,220]. In contrast, the G-overhang of mammals and fission yeast are capped by a single-stranded DNA binding protein POT1 (Pot1), which is a member of a different telomere protein complex termed shelterin[219,220]. Proteins that bind the G-overhang serve as a bridge to the duplex region of telomeres through protein contacts. Double-stranded telomere binding proteins include components of the shelterin complex (TRF1 and TRF2) and in budding yeast Rif/Rap1 protein[105,221,224]. Mutations in any shelterin or CST component leads to loss of telomere length maintenance and genome instability[221,224–228].

CST: end protection, telomere length regulation, and telomere replication

CST is a heterotrimeric protein complex composed of CTC1/Cdc13, STN1, and TEN1 (Figure 1-4)[105,220,226,227,229]. In multicellular eukaryotes, such as *A. thaliana*, Cdc13

is replaced by CTC1[230]. While shelterin contains both double and single strand DNA binding proteins, CST interacts with single-stranded telomeric DNA, and is structurally similar to RPA[231].

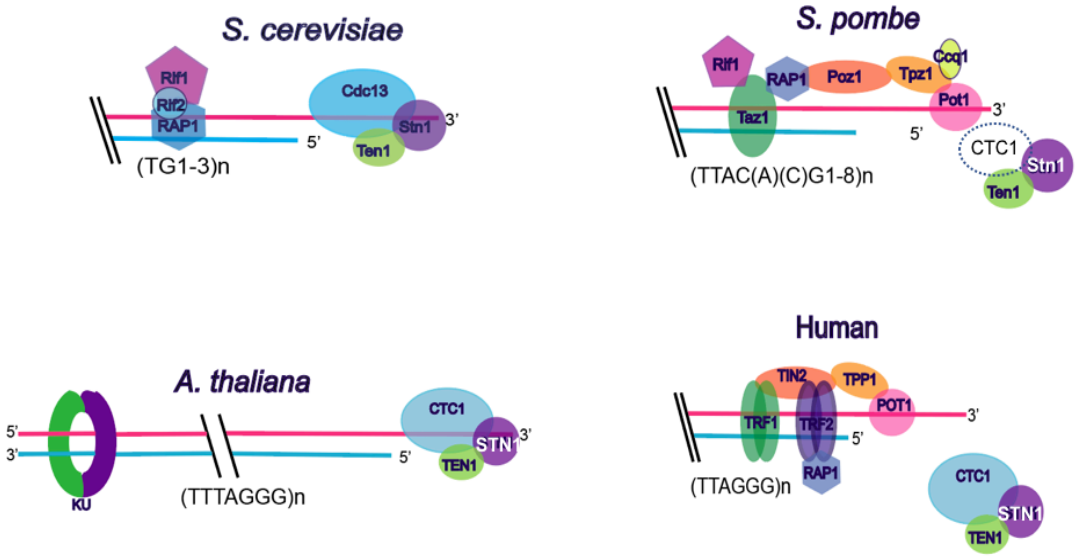


Figure 1-4. Telomere binding proteins. The telomere binding protein complexes of *S. pombe* (shelterin), *S. cerevisiae* (CST), human (shelterin), and *A. thaliana* (CST) are shown. The telomere sequence of each organism is indicated. Telomere proteins associate with both the duplex region of the telomere as well as the single strand G-overhang. Human CST and *S. pombe* Stn1/Ten1 are not stable components of telomeres, but rather associate with chromosome ends during S phase where they play a role in C-strand fill-in after telomerase has extended the telomere. Blunt-ends of *A. thaliana* telomeres are capped by KU70/80 heterodimer instead of the CST complex.

Loss of any one CST component in *S. cerevisiae* or *A. thaliana* leads to an increase in G-overhang signal, and telomere shortening. In plants, chromosome end-to-end fusion also occurs (Figure 1-5A)[230,232–239], indicative of chromosome deprotection and DDR

activation. CST also functions in telomere length regulation (Figure 1-5A)[105]. Yeast Stn1-Ten1 acts as a negative regulator of telomere length, while Cdc13 functions as a positive regulator[235,237,240]. However, the co-assembly of trimeric complex CST negatively regulates telomerase during late S/G2[240]. During telomere elongation by telomerase, the interaction of Cdc13 with the telomerase accessory protein Est1 is essential [234,241]. This interaction is enhanced by Cdk1 dependent phosphorylation of Cdc13, which favors Cdc13-Est1 interaction over Cdc13-Stn1 interaction[242]. The CST complex from *A. thaliana* appears to inhibit telomerase access to the telomeres, specifically through TEN1[178]. It is hypothesized that AtTEN1 is removed from CST and is replaced with POT1a to form a CSP (CTC1, STN1, POT1a) complex (Figure 1-5B)[178]. CSP would then allow telomerase access to telomeres. Thus, dynamic interactions among CST components are needed to properly regulate telomerase.

Initially, shelterin and CST were thought to be mutually exclusive telomere protein complexes, with shelterin confined to vertebrates and fission yeast, and CST to plants and budding yeast. However, homologs of the CST complex were recently identified in vertebrates, suggesting that CST may serve distinct functions in different organisms[230,243]. Unlike the CST complex from budding yeast, vertebrate CST seems to be involved exclusively in telomeric DNA replication, and not in chromosome end protection (Figure 1-5B)[244,245]. A role for vertebrate CST in DNA replication was not unexpected, since human STN1 and CTC1 were originally identified as part of the DNA polymerase-α stimulatory factor (AAF)[246]. Subsequent studies have shown that vertebrate CST components are not critical for telomere length regulation, but are important for G-overhang maintenance. Vertebrate CST is responsible for terminating telomerase action on

extended G-overhang and then promoting the fill-in of the C-strand by DNA polα/ primase[98,247,248]. While CST components are well conserved, shelterin seems to have undergone an evolutionary transformation, because aside from POT1, the other components of shelterin are not readily found in plant genomes[236,249–251].

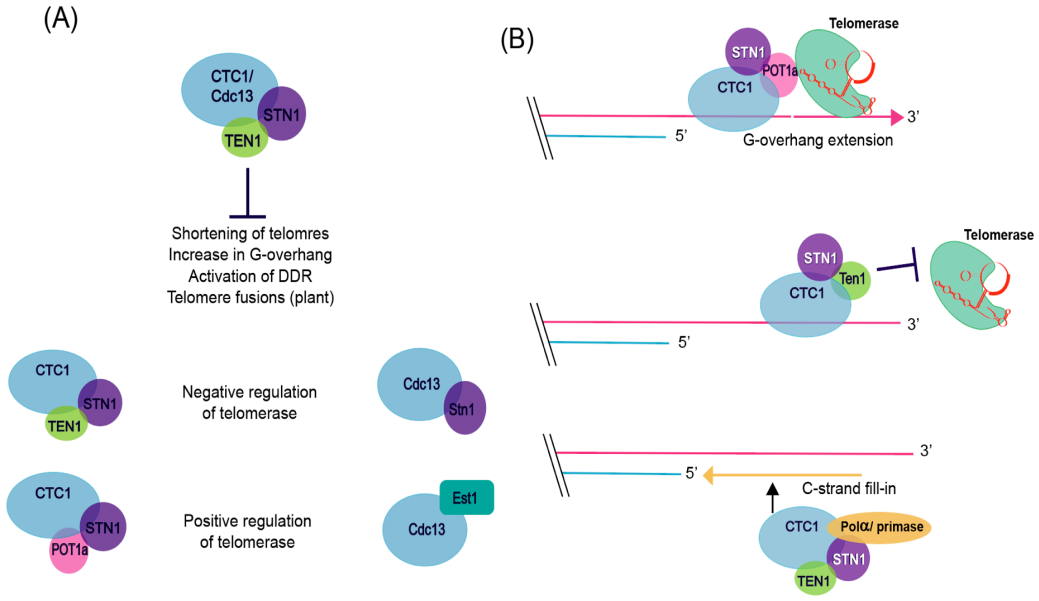


Figure 1-5. Functions of CST. (A) The CST (CTC1/STN1/TEN1) complex is important in chromosome end protection for both *S. cerevisiae* and *A. thaliana*. The dynamic assembly of distinct CST components leads to both positive and negative regulation of telomere length. (B) CSP (CTC1/STN1/POT1a) or Cdc13/Est1 functions to positively regulate telomerase for G-overhang extension (top). CST functions to negatively regulate telomerase (middle). Release of telomerase from telomeres, and CST association with Polα/primase allows C-strand fill-in (bottom).

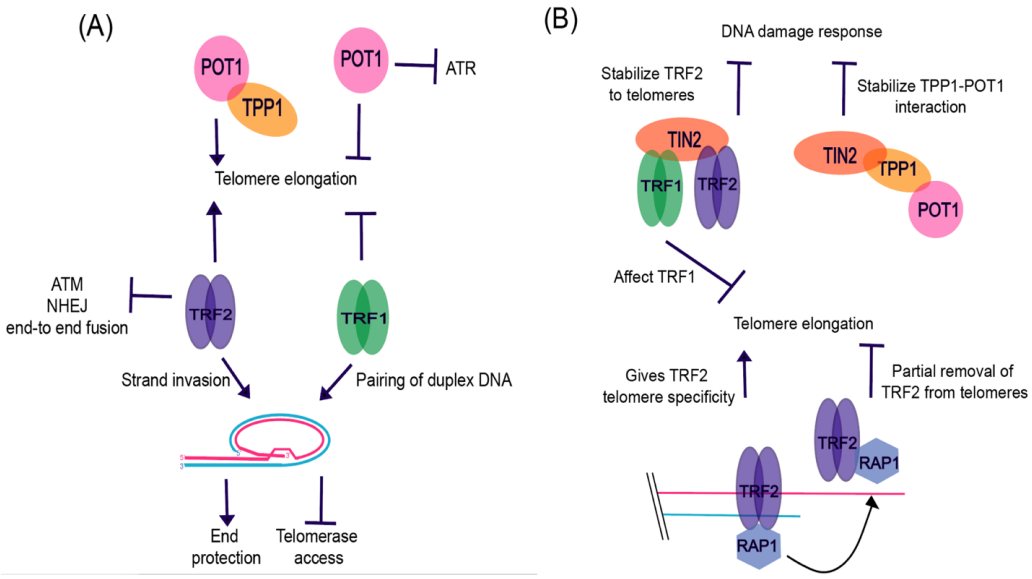


Figure 1-6. Shelterin functions. (A) Shelterin components POT1 and TRF2 are critical for inhibition of a DNA damage response at vertebrate telomeres. TRF1 and TRF2 both contribute to t-loop formation, which in turn protects chromosome ends and regulates telomerase access. While TRF1 and POT1 negatively regulate telomere length, TRF2 and POT1-TPP1 positively regulate telomere length. (B) TIN2 binds TRF1, TRF2, and TPP1. TIN2 plays an indirect role in telomere length regulation through TRF1 (TRF1 inhibition of telomere elongation as shown in A), and DNA damage response through TRF2 and TPP1-POT1 stabilization. RAP1 interaction with TRF2 gives TRF2 telomere specificity. RAP1 modulates TRF2 association with telomeres, thus affecting telomere length regulation.

Shelterin: end protection and telomere length regulation

Shelterin is composed of six proteins (TRF1/2, RAP1, TIN2, TPP1, and POT1) (Figure 1-4). Double-strand telomeric DNA binding proteins TRF1 and TRF2, and the single-strand binding protein POT1 along with its binding partner TPP1, are bridged through TIN2 and RAP1[252–254]. Each component makes a distinct contribution as it connects to the shelterin complex network (Figure 1-6).

TRF1 and TRF2 binds double strand telomeric DNA to regulate telomere length and protect the chromosome ends. TRF1 negatively regulates telomere length[255,256]. TRF2, by contrast, acts to repress an ATM-dependent DNA damage response (DDR), non-homologous end joining (NHEJ), and chromosome end-to-end fusion[17,257–259]. TRF1 and TRF2 accomplish these tasks by facilitating t-loop formation to promote both chromosome end protection and the regulation of telomerase access to telomeres[260–262]. During t-loop formation, TRF1 induces pairing of duplex telomeric DNA to allow bending and looping of the chromosome ends[262], while TRF2 is important for invasion of the single-strand G-overhangs into duplex telomeric DNA (Figure 1-6A) [260]. Rap1 functions to negatively regulate telomere length through TRF2 association[263–265]. TIN2 negatively regulates telomere length through TRF1/TRF2 interaction, as well as prevents DDR by TPP1-POT1 interaction (Figure 1-6B)[266–268]. Therefore, mutation in TRF1 or TRF2 leads to chromosome end deprotection and mis-regulation of telomere length[255,256,17,257–259].

POT1 is another shelterin protein that contributes both to telomere length regulation and chromosome end protection. POT1 is required to prevent activation of an ATR-dependent DDR[259], and is also proposed to compete with RPA binding at the telomeres[259,269]. Prevention of ATR-dependent DDR and RPA binding at the telomeres is important for chromosome end-protection, since DNA double strand repair is not favored at the chromosome ends. In addition to these protective functions, POT1 acts as a negative regulator of telomere length by blocking telomerase interactions with telomeric DNA[270,271]. However, when POT1 is bound to TPP1, the POT1-TPP1 heterodimer acts as a positive regulator of telomere length by physically recruiting telomerase to telomeres, and enhancing telomerase activity and processivity[272–275]. Analysis of POT1 protein in *A.*

thaliana is a major focus of this dissertation (Chapter 2), and will be discussed further below.

OB-fold proteins and evolution of POT1

OB-fold containing telomere proteins

One of the most common structural motifs seen in telomere proteins is the oligosaccharide/oligonucleotide binding domain (OB-fold). The three-dimensional structure of the OB-fold is highly conserved, but the amino acid sequence is not[276]. Therefore, proteins with an OB-fold are frequently identified based on structure prediction rather than through sequence similarity[277]. OB-fold proteins have diverse functions, and can interact with other proteins, DNA, and/or RNA[277–280]. OB-folds have been shown to play an important role in many different pathways, including DNA replication, DNA damage response and repair, transcription, translation, and telomere length homeostasis and end protection[281–283,284–286,287–289].

Telomere proteins that bind the G-overhang do so through an OB-fold. TEBPα/β from the ciliate *Oxytricha nova* was the first OB-fold containing telomere protein identified. It is a heterodimeric single-strand telomeric DNA binding protein, responsible for sequestering the 3' end of the chromosome to facilitate end protection and for presenting the 3' end during S phase for telomerase access[222,223]. TEBP functions as both a telomerase stimulator and inhibitor based on its dimerization partner: the TEBPα homodimer is an inhibitor and TEBPα/β heterodimer is a stimulator[290]. The ciliate TEBP proteins are thought to be ancestral to vertebrate and plant POT1. The TEBPα/β heterodimer is orthologous to *S. pombe* (Pot1-Tpz1) and mammalian (POT1-TPP1) (Figure

1-4) [272,291]. Notably, POT1 proteins typically carry two OB-folds at their N-terminus, which are structurally similar to the OB-folds of TEBPα from *O. nova*, and which directly bind the G-overhang[230,254,292–295]. In contrast, while each of the three CST subunits harbor OB-folds, this heterotrimer is more closely related to the RPA protein complex than to POT1 [226,296].

POT1 binds single-stranded telomeric DNA with sequence specificity conveyed by the first OB-fold[297–302].

The OB fold of POT1

POT1 is highly conserved across eukaryotes (Figure 1-7)[302–304]. DNA binding by human and *S. pombe* POT1 is mediated through a nucleic acid binding pocket primarily within the first OB-fold that excludes RNA[254,298,305]. Interestingly, while POT1 is well conserved across the plant kingdom (Figure 1-7), its interaction with nucleic acids appears to have diverged. *In vitro* telomeric DNA binding has been shown for the POT1 proteins from green algae and several monocots, but not with POT1 proteins from the Brassicaceae family (Figure 1-7)[300,302,306]. Although the two full-length POT1 proteins from *A. thaliana* (POT1a and POT1b) do not bind telomeric DNA *in vitro*[132], more recent studies with the first-OB fold of *A. thaliana* POT1a show that this domain in isolation has strong binding affinity for telomeric DNA revealing the conserved DNA binding component of POT1 protein[307].

Single-copy to multi-copy POT1

The POT1 protein family has undergone an expansion in some organisms. While humans possess only a single copy POT1 gene, other organisms possess more than one POT1 gene, including mice (POT1a and POT1b), *T. thermophila* (Pot1a and Pot2), *C. elegans* (POT-1, POT-2, POT-3, POT-4), and *A. thaliana* (POT1a, POT1b, and POT1c) (Figure 1-7 and Figure 1-8)[303]. The two mouse Pot1 proteins diverged to sub-functionalize, where each polypeptide contains a subset of the functions of single copy hPOT1[308]. The first OB-fold of mouse Pot1a is a negative regulator of telomere length and is important in preventing DNA damage response and DNA damage checkpoint activation[309,310]. In contrast, the C-terminal end of mouse Pot1b is important for the control of 5'-end resection[308,310,311]. *C. elegans* has four POT1 paralogues, and each contains only a single OB-fold (Figure 1-8).

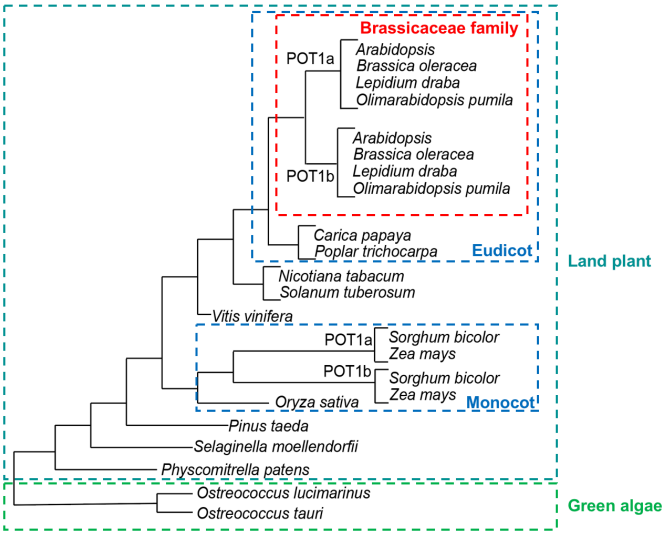


Figure 1-7. POT1 phylogenetic tree. Phylogenetic tree of POT1 from selected organisms in the Plantae kingdom. The early land plant *P. patens*, and the monocots *S. bicolor* and *Z. mays* POT1 protein bind telomeric DNA. Brassicaceae family (red box) including Arabidopsis POT1 proteins do not.

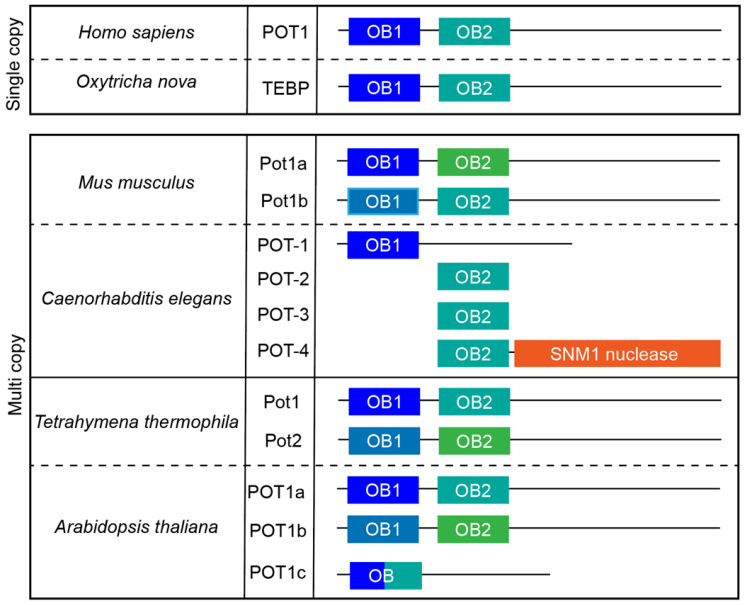


Figure 1-8.POT1 OB-fold. Human, mouse, *T. thermophila*, and *A. thaliana* POT1 consist of two OB-folds while *C. elegans* and *A.thaliana* POT1c consist of single OB-fold.

C. elegans POT-1 and POT-2 retain canonical POT1 function, where both proteins associate with single-strand telomeric DNA[292]. Both POT-1 and POT-2 are negative regulator of telomerase, but the two proteins function in a non-redundant manner to repress telomerase activity[312]. The third POT1-like protein from *C. elegans,* POT-3, harbors single OB fold that resembles OB2 rather than OB1; but the function of this fourth POT1 protein is still unknown[313].The fourth POT1-like protein POT-4, has a OB-fold and an SNM1 family nuclease domain[314]. The activity of telomerase *in vivo* is dependent on the OB-fold of MRT-1; however, the detailed mechanism is still unclear.

Another organism with two POT1 paralogues is *T. thermophila;* its POT1 genes are designated Pot1 and Pot2. *T. thermophila* Pot1 is an essential gene, whose absence causes growth arrest, telomere length de-regulation, DNA damage response, and telomere end de-

protection[315]. *T. thermophila* Pot2 plays a role in sexual reproduction, as well as recruitment of telomerase and/or endonuclease to micronuclear chromosome breakage sites for DNA cleavage[316]. Thus, although the extra copies of POT1 in mice and *C. elegans* retain a function in telomere biology, *T. thermophila* Pot2 has undergone neo-functionalization, acquiring a new function outside the conventional telomere pathway. Another example of an organism where POT1 paralogues may display neo-functionalization is *A. thaliana*.

In plant kingdom, there has been two independent duplication events that has led to two copies of POT1. First event was Panicoideae-specific in grasses 30 mya (monocot), and the second event at the base of Brassicaceae (eudicot)[317–319]. The first event occurred after the divergence of last common ancestor with rice, wheat, and barley, giving rise to two POT1 in *Zea mays* and *Sorghum bicolor* (Figure 1-7). The second event gave rise to two POT1 genes in Brassicaceae, which includes *A. thaliana* (Figure 1-7).

Three POT1 genes have been identified in *A. thaliana*, which were generated by two separate duplication events. The first duplication event occurred during the whole genome duplication of the Brassicaceae family approximately 34 million years ago (mya)[320,321]. The second duplication is specific to *A. thaliana,* placing the gene duplication approximately 10 mya.

Genetic data indicate that AtPOT1a is not required for telomere capping, but has maintained a function within telomere biology by regulating telomerase activity. POT1a does so by stimulating telomerase repeat addition processivity. The absence of POT1a protein leads to decreased telomerase activity and as a consequence, progressive telomere shortening similar to a *tert* mutant[172,178]. Furthermore, F65 is a highly conserved residue in

mammalian POT1 essential for telomeric DNA binding[309,322], and F65A mutation does not rescue the *pot1a* phenotype in *A. thaliana*[319]. Thus, the residue important for nucleic acid binding is an important component for POT1a function.

The phenylalanine at the amino acid position 65, located in the OB1, is important for telomeric DNA binding in AtPOT1a, and this residue is also found to be important for human POT1 to bind telomeric DNA[305,307]. In addition, POT1a OB1 is sufficient to stimulate telomerase processivity *in vitro*. In AtPOT1b, the critical amino acid for DNA binding in OB1 has been changed from phenylalanine to valine (V63). In marked contrast to POT1a OB1, POT1b OB1 cannot bind telomeric DNA or stimulate telomerase processivity i*n vitro*[307]. Another difference between POT1a and POT1b is the binding affinity for CTC1. While POT1a has a binding affinity toward CTC1, POT1b does not[319]. POT1a has evolved to have an enhanced association with CTC1 through three sites, E35, S212 , and E293, that went under positive selection while those residues in POT1b did not[319].

Finally, unlike POT1a, the absence of POT1b protein does not lead to an obvious telomere phenotype other than a slight increase in telomerase activity (B. Barbero and X. Xie, unpublished data). In addition, POT1b resides in cytoplasm instead of nucleus, where POT1a is seen. Detailed analysis of *pot1b* mutants suggests that the protein function in plant development. Specifically, POT1b is required for seed germination and pollen viability (B. Barbero and C. Castillo, unpublished data). These findings suggest that like the *T. thermophila* Pot2 protein, AtPOT1b has evolved a function in reproductive development. Thus, POT1a has maintained its ancestral POT1 function in telomere biology whereas POT1b has evolved to have a function outside telomere biology.

A third POT1 gene from *A. thaliana*, POT1c, has also been reported. Analysis of this gene is the focus of chapter 2 in this dissertation.

Telomere length homeostasis

Mechanism of telomere length regulation

While regulation of telomere length is modulated by trans-acting factors, such as telomerase and homologous recombination machinery; cis-acting telomere binding proteins also play a critical role in establishing and maintaining length homeostasis. As discussed above, each species has a fixed average telomere length that is maintained, which is referred to as the telomere length set point. The set point is established by a balance of mechanisms that promote elongation of short telomeres, or shortening of excessively long telomeres[225]. Examples of cis-acting factors that modulate telomere length set point include the telomere binding proteins TRF, Taz1, and Rap1 (Figure 1-4). A protein counting model has been proposed to control telomere length based on a negative feedback loop where telomere protein binding serves to negatively regulate extension. In this model, short telomeres are associated with fewer telomere binding proteins, while long telomeres bind more proteins, which ultimately inhibits telomere extension (Figure 1-9). In yeast, the telomere length "sensing mechanism" involves counting the number of Rap1 proteins bound to telomeres [323–326], while in humans it is the number of TRF1 molecules bound that correlates with telomere length [262,270]. As shelterin can stimulate t-loop formation, which inhibits telomerase access to the telomeres, t-loop prevents further telomere elongation[224]. The overall amount of shelterin components associated with telomeres also parallels the length of telomeres.

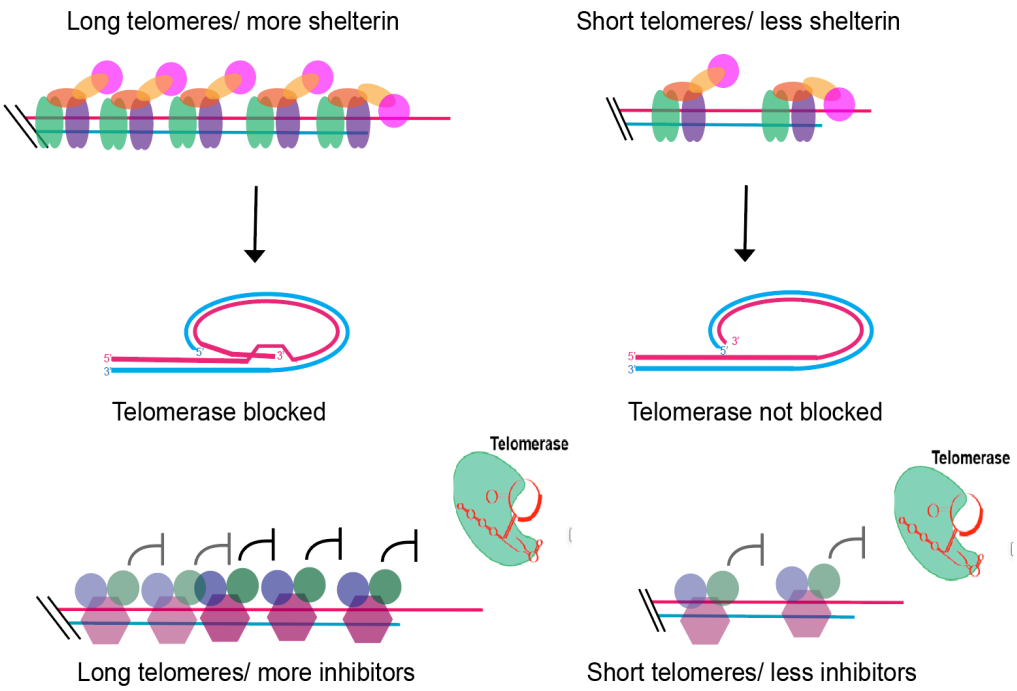


Figure 1-9.Telomere counting model. Longer telomeres allow more shelterin (top: mammals) and telomerase inhibitor (bottom: yeast) to associate with telomeres. Shorter telomeres have less shelterin (top: mammals) and telomerase inhibitor (bottom: yeast) associated with telomeres, thus allowing telomerase to access telomeres.

Despite these studies of cis-acting telomere binding proteins, relatively little is known about the fundamental mechanisms that control telomere length set point. Because telomere length is a quantitative trait, multiple genes are likely involved in determining telomere length set point[109]. In 2004, a genome wide screen of haploid yeast deletion sets identified more than 150 genes that affect telomere length, many of which had no obvious connection to telomere biology[110]. Since then, over yeast 400 genes have been identified that affect telomere length[109,327]. Characterization of newly identified genes that are involved in *A. thaliana* telomere length set point is a major objective of this dissertation (Chapter 3).

Qualitative traits and quantitative trait mapping

Quantitative traits (QT) and Quantitative trait loci (QTL)

Phenotypic traits can be either qualitative or quantitative (Figure 1-10). Qualitative traits follow simple Mendelian genetics and are easy to categorize[328]. They are often controlled by a single gene, thus the mode of inheritance is monogenic. Examples of qualitative traits are blood types and enzyme defects. On the other hand, quantitative traits (QT) show a range of variations in phenotype that are influenced by multiple genes. Thus, the phenotype observed is a result from the cumulative effect of multiple genes. Unlike qualitative traits, QT do not follow the Mendelian rules of single-gene inheritance, but rather phenotypes reflect a continuous gradient of phenotype seen in bell curve (Figure 1-10) and the mode of inheritance is polygenic[329]. Examples of qualitative traits are height and skin color, and telomere length set point is a quantitative trait.

While qualitative genetics looks at the change in the frequency of specific alleles, an alternate form of a gene of interest, quantitative genetics quantifies the change in the frequency distribution of the phenotype being studied[330–332]. Analysis of QT are done by searching a quantitative trait locus (QTL). A chromosomal region is termed a QTL when a segment of that DNA (the locus) is associated with a QT. For quantitative genetic traits, the relationship between genotype and phenotype are not simple; therefore, it is much difficult to study compared to qualitative genetic traits. Therefore, QTL analysis relies on detecting an association between phenotype and the genotype of marker such as single nucleotide polymorphisms (SNPs)[333].

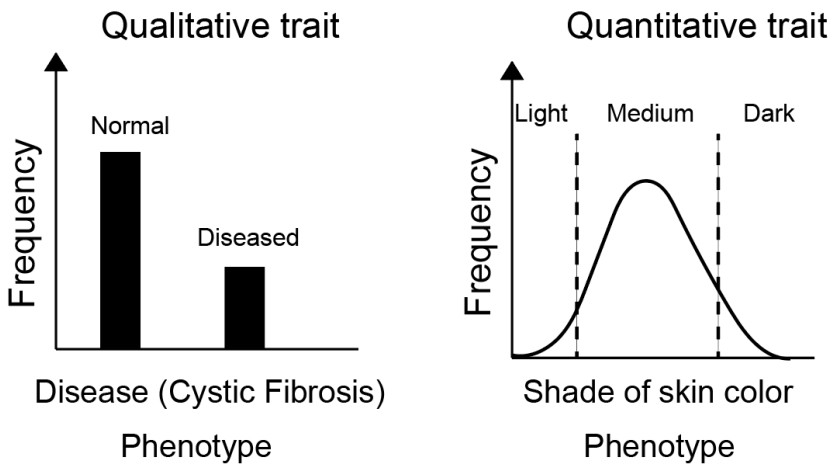


Figure 1-10. Quantitative vs qualitative trait. Qualitative trait is shown on the left, following the simple Mendelian inheritance of phenotype. Quantitative trait is shown on the right, showing the complex polygene inheritance as bell curve.

Mapping QTL

To detect a QTL, four components are required: 1) a genetically variable population for the trait of interest, 2) marker systems that allow genotyping of the population, 3) methods that can analyze a quantitative phenotype (trait), and 4) experimental and statistical methods for detecting and locating QTL[334].

In *A. thaliana* several variable populations are available for QTL mapping (Figure 1-11). These include naturally occurring inbred lines (accessions or strains), synthetic populations, Heterogeneous Stocks (HS), and Multiparent Advanced Generation Intercross (MAGIC) lines. Synthetic populations include Recombinant Inbred Lines (RIL), and Advanced Intercross Lines (AIL). RILs are generated by crossing two parent accessions to create recombinants, which are then inbred until they are isogenic. AILs are generated by randomly and sequentially intercrossing RIL populations. AILs are an extension of RILs

that provides higher mapping resolution due to more recombination events; thus, AILs are useful in refining a QTL[335]. HS is an extended approach from AIL, generated from crosses between multiple parental lines. HS are crossed for many generations to produce a highly recombinant heterozygous outbred population. Notably, HS from 8 parental mice strains and 8 parental drosophila have been successfully used for fine-mapping QTL[336–339]. MAGIC lines are similar to HS but RILs are generated from multiple parents[340]. MAGIC lines provide an intermediate niche between naturally occurring inbred lines and existing synthetic populations. MAGIC lines have been generated for several different plant species including *A. thaliana*, *Zea mays*, and *Oryza sativa*[341–343].

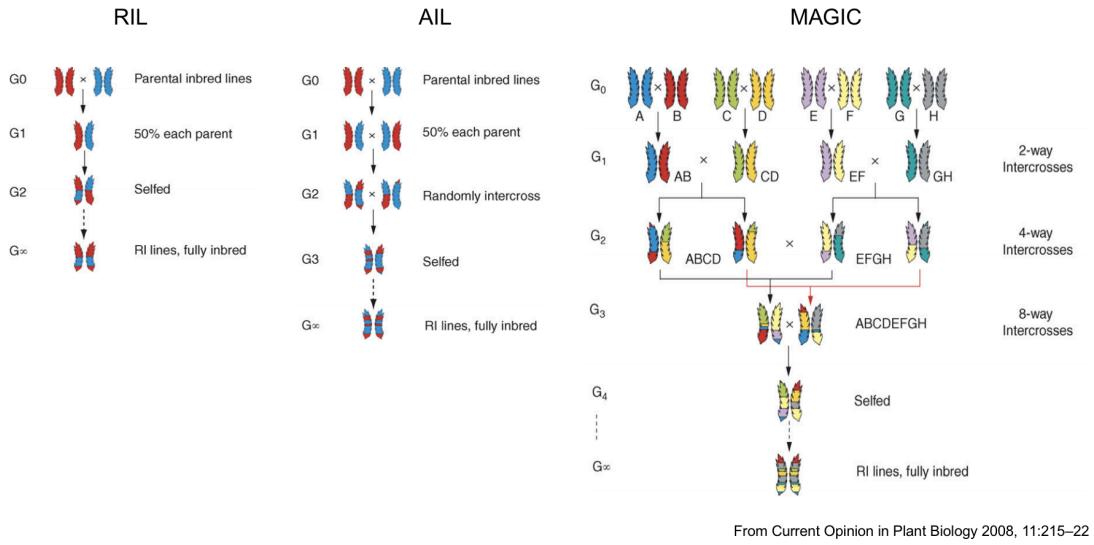


Figure 1-11. (reprinted form) Variable population for QTL mapping.[340] Recombinant Inbred Lines (RIL), Advanced Intercross Lines (AIL), and Multiparent Advanced Generation Intercross (MAGIC) lines.

QTLs are mapped through identification of molecular markers that correlate with the phenotype of interest. The main types of DNA markers commonly used for mapping populations are: 1) single nucleotide change/polymorphism (SNP); 2) insertion or deletions of various length; and 3) variations in the number of tandem repeats (VNTR)[344–346]. Methods to identify these differences in DNA sequences can be further categorized into three groups: bi-allelic dominant, bi-allelic co-dominant, and multi-allelic co-dominant[344,345,347].

Bi-allelic dominant markers are detected using Random amplified polymorphic DNA (RAPD) or Amplified fragment length polymorphism (AFLP). RAPD is a PCR method that uses random short primers to amplify genomic DNA to look for a semi-unique profile by analyzing the amplification pattern (different size amplicons), while in the AFLP method genomic DNA is fragmented with restriction enzymes and a subset of fragments is amplified[348,349]. Unlike RAPD that assesses product size, AFLP looks at the presence or absence of the product. Bi-allelic co-dominant markers can be detected using Restriction Fragment Length Polymorphisms (RFLP), a method that compares the size of fragmented DNA using restriction enzymes[350,351]. Multi-allelic co-dominant markers are detected by microsatellite analysis. Microsatellites (also called short tandem repeats (STRs) or simple sequence repeats (SSRs)) are repetitive DNA sequences[352,353]. The method determines the number of repeats in different samples.

Arabidopsis as a model organism

Arabidopsis is an established model system with unique features that allow it to be a reference organism for all of plant biology. First, the life span of Arabidopsis is short, *A*.

thaliana produces reproductive organs 4 weeks after germination, and requires a total of only 6-8 weeks to generate offspring. Mature plants produce more than 5000 seeds. It is possible to store genetic stocks as seeds, which minimizes the maintenance effort.

Second, Arabidopsis has a small compact genome that is fully sequenced. It contains 5 chromosomes with roughly 20,000 genes, and its genome is enriched with coding sequence (average of one gene in every 5kb)[354–359]. The Arabidopsis genome has been fully mapped, including mutant genes and molecular marker such as SNPs. Genetic analysis has also expanded to the study of epigenetic, gene silencing, centromere mapping, and reverse genetics. Moreover, at least half of these genes in the *A. thaliana* genome are conserved from bacteria to humans.

Third, the *A. thaliana* genome can be genetically manipulated in a simple and efficient manner for mutagenesis (including chemical and insertional), crosses, and DNA transformation.

A fourth advantage of *A. thaliana* is that a large number of molecular tools are also available. The Arabidopsis Information Resource (TAIR) is a website that contains locus information, genome annotation, results of micro array data, polymorphisms, and stock ordering[360]. The stock center stores seeds for thousands of mutants defective in all aspects of plant growth and development.

Arabidopsis Accessions, 1001 Genomes Project, and MAGIC Lines

There are over 855 Arabidopsis accessions. These natural inbred lines are a product of natural selection under various environmental conditions from around the world. The 1001 Genomes Project, which was launched in 2008, seeks to provide whole-

genome sequencing for individual accessions to explore sequence variation[361,362]. The accessions selected for analysis provide a wide range of natural variation in ecological and developmental traits. This large collection of intra-species genetic variants makes *A. thaliana* an invaluable source for the study of different traits. Different accessions have different phenotypes in morphology, fitness, flowering time, and metabolism. Thus, the 1001 *A. thaliana* genomes project can establish links between genotype and phenotype, facilitating a wide area of studies in evolutionary science, and plant breeding. Since 2010, a total of 1135 *A. thaliana* genomes have been analyzed[362].

For QTL mapping, MAGIC (Multiparent Advanced Generation Intercross) lines have been developed for *A. thaliana*. MAGIC lines are a mosaic of 19 founder accession genomes[341]. MAGIC lines have several advantages over the classical mapping populations. Naturally occurring inbred lines have poor mapping resolution, and synthetic populations have limited phenotypic diversity. While HS has been successful in fine-mapping QTL, individual genomes are unique and heterozygous. Consequently, high-density genotyping is required each time a population is phenotyped. Similar to HS, MAGIC lines provide high allelic and phenotypic diversity, and high mapping accuracy for detecting QTL. However, unlike HS, MAGIC lines eliminate the need to re-genotype the phenotyped population, saving time and money. There are over 700 MAGIC lines that are fully sequenced/genotyped, and the lines are genotyped at 1200 SNPs that are spaced approximately 100kb apart[341]. Thus, MAGIC lines have multiple advantages for fine-mapping QTL.

Arabidopsis telomere biology

Amongst all the advantages Arabidopsis has as model organism, its impact on telomere biology is most relevant to this dissertation. *A. thaliana* telomeres are relatively short, thus easy to analyze. Interestingly, telomere length set point varies among different *A. thaliana* accessions[106], making it possible to use QTL mapping to identify novel genes that regulate telomere length.

As in humans, Arabidopsis cellular proliferation capacity is correlated with telomerase activity, which is highest in actively dividing cells including flowers and young seedlings[363]. Overall, mechanisms of telomere and telomerase regulation are well conserved with vertebrates and yeast. However, some mutations in telomere-associated genes that are lethal in mammals and yeast can be tolerated in Arabidopsis. Thus, there are unique features in Arabidopsis telomere biology that allow study of telomere maintenance from an evolutionary standpoint.

Dissertation overview

There are two main research objectives described in this dissertation. The first is characterization of the *A. thaliana* POT1c locus. In Chapter II, genetic and evolutionary analyses of the AtPOT1c gene are presented. The second research focus of the dissertation is analysis of the role of NOP2A in telomere length set point. Chapter III covers both genetic and biochemical characterization of the NOP2A gene, while Chapter IV covers analysis of NOP2A locus in other *A. thaliana* accessions.

Chapter II describes the results of experiments aimed at investigating the function of POT1c and the fate of the POT1c locus after gene duplication. Using a genetic

approach, molecular and developmental phenotypes of *pot1c* mutants were examined. There were no obvious telomere-related defects or perturbations to plant reproduction or vegetative development in the mutants. In addition, analysis of POT1c expression and evolution of the POT1c locus was performed using a combination of *in silico* analysis, biochemistry and molecular genetics. The data indicate that POT1c is very poorly expressed, and likely to be a pseudogene. POT1c bears two transposable elements upstream of the start codon, and lacks a functional promoter. Unexpectedly, the POT1c locus shows no evidence of active gene silencing. POT1c is not well conserved among *A. thaliana* accessions, suggesting that it is not under selective pressure. Comparison of the POT1c locus with the corresponding region in close relatives of *A. thaliana* indicates that the transposons were inserted in the POT1c promoter very soon after the duplication event, and this transposable element insertion has led to the silencing of POT1c locus.

Chapter III describes the identification and characterization of new gene that plays a role in telomere length set point, NOP2A/OLI2. Genetic analysis was used to demonstrate that NOP2A/OLI2 is involved in telomere length set point. NOP2A/OLI2 mutant analysis shows decrease in telomerase activity, but *tert/nop2a* and *pot1a/nop2a* show phenotypes similar to *tert* and *pot1a* single mutants. This suggests that NOP2A/OLI2 stimulates telomerase activity in a different manner compared to POT1a. Finally, analysis of *A. thaliana* NOP2 paralogues do not show telomere shortening while mutants in OLI2 genetic pathway shows same telomere shortening as NOP2A/OLI2, indicating that telomere length set point phenotype is associated with OLI genetic pathway rather than NOP2.

In Chapter IV, genetic study is used to show that NOP2A/OLI2 is one of the major factor for determining telomere length set point in *A. thaliana*. Sf-2 and Col-0 accessions were crossed and segregated by their NOP2A locus. Three key observations are made from this experiment: 1) only the Sf-2/Sf-2 genotype has long telomeres, 2) not all the Sf-2/ Sf-2 genotypes have long telomeres, and 3) some Col-0/ Col-0 genotypes have telomeres longer than WT Col-0. The first observation indicates that NOP2A is necessary for long telomeres, the second observation proposes that there is another factor that functions with NOP2A, and the third observation suggests that there is another factor independent of NOP2A that plays a role in telomere length determination. Finally, the NOP2A locus is compared among 19 MAGIC parental accessions to investigate the difference in NOP2A locus, and how this difference may affect the telomeres of these accessions.

Altogether, this dissertation sheds new light on the fate of newly duplicated genes and the importance of proper regulation of genes involved in telomere biology. It also reveals novel genes that play a role in determining telomere length set point. These discoveries increase understanding of the complexity of telomere biology and its interconnectedness with fundamental processes that promote genome integrity.

CHAPTER II

FATE OF THE NEW DUPLICATE GENE POT1C

Summary

Protection of telomeres 1 (POT1) is a conserved telomere binding protein found in organisms as old as the earliest diverging land plants. POT1 proteins exhibit signature oligosaccharide/ oligonucleotide binding domains (OB-folds), typically two at their amino terminus. Although humans possess a single copy POT1 gene, POT1 is duplicated in other organisms, including rodents, ciliates and plants. Biochemical and genetic experiments reveal that the two mouse POT1 paralogs have sub-functionalized while *Tetrahymena thermophila* TtPot2 may have been subjected to neo-functionalization. The plant kingdom is characterized by at least two independent POT1 duplication events, one in the grasses and one in the Brassicaceae. *A. thaliana* carries three POT1-like loci, POT1a, POT1b and POT1c, that arose from two distinct duplication events. AtPOT1a has retain the ancestral function, while POT1b may have subjected to neo-functionalization. The current data for AtPOT1a and AtPOT1b genes argue that the *A. thaliana* POT1 gene family has been subjected to substantial, and ongoing functional divergence post duplication. This chapter explores the function and evolution of the third AtPOT1 paralog, POT1c. We provide evidence that the AtPOT1c locus is functionally silenced likely due to the acquisition of two transposable elements within the promoter and multiple deletions within the coding sequence. The data demonstrate a very different path POT1 duplicate in *A. thaliana* has taken: POT1c has non-functionalized through permanent silencing, while POT1b has neo-functionalized.

Introduction

Protection of telomeres 1 (POT1) is a conserved telomere binding protein found in organisms as old as the earliest diverging land plants[300,302]. Human POT1 is a component of telomere capping complex, shelterin, and directly contacts single-stranded telomeric DNA on the extreme chromosome terminus[254]. POT1 has dual functions in protecting telomeres from eliciting an ATR-dependent DNA damage response (DDR) and in regulating telomerase access to chromosome ends[254,259,271,364,365]. Emerging data suggest that POT1 also plays a role in regulating the efficiency and fidelity of DNA repair[366].

POT1 proteins exhibit signature oligosaccharide/ oligonucleotide binding domains (OB-folds), typically two at their amino terminus. The OB-folds convey sequence-specific recognition of single-stranded telomeric DNA. Although humans possess a single copy POT1 gene, POT1 is duplicated in other organisms, including rodents, ciliates and plants[308,315,367]. The two mouse POT1 genes exhibit 70-75% sequence similarity[308,311]. Biochemical and genetic experiments reveal that these two paralogs sub-functionalized, so that each expresses only a subset of the functions of single copy human POT1[308]. The first OB-fold of mPOT1a is a negative regulator of telomere length and is important in preventing DNA damage response and DNA damage checkpoint activation[309,310]. In contrast, the C-terminal domain of mPOT1b is required to control the resection of the 5'-end of the chromsosome[308,310,311]. *Tetrahymena thermophila* also encodes two POT1 genes, termed Pot1 and Pot2. TtPot1 is an essential gene, whose absence triggers growth arrest, telomere length de-regulation, DNA damage response, and telomere end de-protection[315]. In contrast, TtPot2 plays a role in sexual reproduction, as well as recruitment of telomerase and/or endonuclease to micronuclear chromosome breakage sites for DNA cleavage[316].

These findings suggest that TtPot2 may have been subjected to neo-functionalization.

The plant kingdom is characterized by at least two independent POT1 duplication events, one in the grasses and one in the Brassicaceae. A Panicoideae-specific POT1 duplication occurred in the grasses (~75 mya), resulting in two POT1 paralogs in *Zea mays* and *Sorghum bicolor*, which share approximately 75% amino acid sequence similarity[319,368]. A second duplication event occurred at the base of the Brassicaceae, a large family that includes *Arabidopsis thaliana*. Since POT1 genes are single copy in papaya and cotton, this POT1 duplication event was dated to approximately 100 mya[317–319].

A. thaliana carries three POT1-like loci, POT1a, POT1b and POT1c[319,369], that arose from two distinct duplication events. Strikingly, AtPOT1a and AtPOT1b display only 52% sequence similarity overall. Genetic analysis indicates that AtPOT1b cannot not complement AtPOT1a deficiency[319]. Moreover, the AtPOT1a gene is functionally distinct from the single copy POT1 gene in the moss *Physcomitrella patens*, which was previously shown to be critical for chromosome end protection in much the same way as hPOT1[2]. Finally, although the *Arabidopsis lyrata* POT1a (98% amino acid similarity to AtPOT1a) can fully rescue an AtPOT1a deficient mutant, the *Brassica oleracea* POT1a (74% amino acid similarity to AtPOT1a) can not[319].

AtPOT1a appears to retain only a subset of the ancestral functions ascribed to hPOT1. Plants null for POT1a show decreased telomerase activity and, as a consequence, an ever shorter telomere phenotype similar to plants lacking the telomerase catalytic subunit[172,363]. However, unlike hPOT1, which is a stable component of the shelterin telomere capping complex, AtPOT1a only transiently associates with telomeres and

instead is physically associated with the telomerase ribonucleoprotein (RNP) where it positively regulates enzyme activity by stimulating telomere repeat addition processivity[172,178].

This study explores the function and evolution of the third AtPOT1 paralog, POT1c. We show that AtPOT1c is derived from a very recent duplication event involving the POT1a locus and a neighboring gene encoding ribosomal protein S17 (dS17). The duplication was followed by deletion and inversion, leaving AtPOT1c with only a single OB-fold, rather than the two OB-folds associated with conventional POT1 proteins. We provide evidence that the AtPOT1c locus is functionally silenced likely due to the acquisition of two transposable elements within the promoter and multiple deletions within the coding sequence. Silencing of the POT1c locus has led to the accumulation of DNA mutations leading to non-conservation between accessions. Consistent with this conclusion, we found no discernable molecular or developmental defects in plants bearing a CRISPR mutation within the POT1c locus. This study has provided new insight for the consequences of a newly duplicated gene that is involved in telomere maintenance. The data demonstrate a very different path that a POT1 duplicate in AtPOT1 has taken: POT1c has non-functionalized through permanent silencing, while POT1b has neo-functionalized.

Materials and methods

POT1c transcript analysis

Total plant RNA was extracted from plant flower, leaf, and seedlings using the Zymo RNA mini-prep kit. RNA transcripts were amplified using cDNA synthesized with oligo dT. Nested PCR was performed to identify an RNA corresponding to the TAIR annotation

using different primer combinations (F1, F2, F3, F4, R1, R2, and R3) (Table 2-1). Quantitative POT1c transcript levels were analyzed by qPCR with GAPDH as a reference gene (Table 2-1).

POT1c promoter analysis and expression in vivo

A POT1c expression construct was created by cloning the ubiquitin promoter and genomic sequence of POT1c into pKGWFs7 destination vector using BamHI and SbfI, and SpeI (pKGWFs7:UBQp-POT1c) restriction enzymes, respectively. A GUS reporter construct for POT1c was created by cloning the POT1c promoter (845 bp upstream of the initial POT1c ATG) into pKGWFs7: GUS destination vector using EcoRI and XbaI (pKGWFs7:Pot1cp-GUS) restriction enzymes.

For transformation, small-scale *Agrobacterium tumefaciens* cultures were grown for 24 h at 30°C. 500 mL LB was inoculated with 0.5 mL of *A. tumefaciens* culture and was grown at 28°C until the OD_{595} reached 1.5. The culture was centrifuged for 10 min at 4,000 rpm, and was re-suspended in 15% sucrose 0.2% Silwet-77 solution. Plants with flowers were dipped in Agrobacterium solution for 1 min, and kept in the dark overnight. T1 seeds were selected on [20µg/ml] BASTA, and ½ MS media [25µg/ml] Kanamycin for pDs transformants.

For GUS detection, flowers, floral buds, seedlings and leaves were fixed using 90% acetone for 30 min. Samples were washed 3 times with citrate buffer for 10 min each. Samples were then treated with staining solution, vacuumed for 5 min and 15 min with each round releasing the vacuum slowly (2-3 min) to allow tissues to absorb the staining solution. Samples were then stained at 37C° for 48 h. Stained samples were washed with

50% and 70% ethanol for 3 h each, and then placed in 100% glycerol overnight at RT.

POT1c protein expression, purification, and gel shift analysis

A N-terminal SUMO tagged POT1c construct was created by cloning POT1c CDS sequence into pET28a vector containing N-terminal SUMO tag using BamHI and EcoRI (pET28a:His-SUMO-POT1c) restriction enzyme.

POT1c protein was expressed by inoculating six 1 litter LB media with 15 mL overnight *E. coli* culture containing pET28a:His-SUMO-POT1c. The inoculated culture was placed in 30 C° shaker until OD_{595} of 0.6 was reached. 500 mM IPTG was then added to allowed SUMO-POT1c expression at 16 C° for 18 hours. Culture was then spun down at 5000g for 15 minutes, and resuspended in lysis buffer (50mM Tris-Cl pH 7.5, 500mM NaCl, 10mM imidazole, 1mM DTT, and 1x sigma protease inhibitor cocktail). Samples were then sonicated for 10 seconds on and 10 seconds off for 15 minutes. Lysed sample was spun down at 15000g for 30 minutes and loaded on Ni-NTA column. The column was washed with 240mL wash buffer (20mM Tris-Cl pH 7.5, 150mM NaCl, 10mM imidazole, and 1mM DTT), and His-SUMO-POT1c was eluted with mL of 100 mM and 10 mL of 250 mM elution buffer (20mM Tris-Cl pH 7.5, 1500mM NaCl, 100 or 250 mM imidazole, and 1mM DTT). Elution samples were resolved by SDS-PAGE to select the fractions that contain His-SUMO-POT1c, and these samples were dialyzed over night at 4 C°.

Gel shit assay was performed by incubating the protein with 5' [^{32}P] labeled oligonucleotide probe $(TTTAGGG)_6$ in binding buffer (20mM Tris-Cl pH7.5, 200mM potassium glutamate, 10mM $MgCl_2$, 10% glycerol, 50µg/mL BSA, 1mM DTT, and 1µg/µL tRNA) for 30 min at 37 C°. Samples were resolved by non-denaturing 0.8% PAGE

at 120 volts, and the gel was exposed with a phosphor-screen and imaged. The image was analyzed using QuantityOne software.

In silico analysis

A secondary structure model for POT1c was derived using Jpred3 (http://www.compbio.dundee.ac.uk/jpred/). We assessed mRNA sequence, DNA methylation status, presence of small RNA, and transposable elements using data obtained from the Jacobsen epigenomics database (http://genomes.mcdb.ucla.edu/AthBSseq)[370–374]. Prediction of the POT1c promoter was performed using softberry, BDGP, and the plant promoter database 3.0. A total of 855 *Arabidopsis thaliana* accession sequences were obtained from the 1001 genomes project (http://signal.salk.edu/atg1001/3.0/gebrowser.php), and were analyzed using Geneious software.

CRISPR mutation

The protospacer sequence (TAGCAAAGCCGGGAGTAGGC) was used for generation of a CRISPR/Cas9 mediated *POT1c* mutant. The construct was cloned into pEN: St1gRNA entry vector, kindly provided by the Fauser lab (pEN: St1gRNA-Pot1c). pEN: St1gRNA-Pot1c was then cloned into pDs: St1Cas9 destination vector, generously provided by the Puchta lab (pDs: St1Cas9-St1gRNA-Pot1c). A melting curve of T1 transformants was obtained using the BioRAD CFX 96 system 3 step amplification melting curve protocol (Tm P1c CRISPR Fw and Tm P1c CRISPR Rv). DNA from plants that showed an altered melting curve was PCR amplified using gene-specific primers and

sequenced to confirm the mutation generated from CRISPR Cas9 system. Sequencing products were generated by amplification of the *POT1c* locus near the target site (Seq P1c CRISPR Fw and Seq P1c CRISPR Rv). The T2 generation of mutant plants was further screened to identify homozygous *POT1c* mutants. Homozygous mutants were then screened for the presence of the kanamycin gene using PCR (SS42 and SS102) to confirm a clean background, as well as the absence of the Cas9 nuclease system to prevent further alteration of the genome.

Telomerase activity assay (TRAP) and telomere length analysis (TRF)

Protein extract was prepared from 6 day-old seedlings. 1 mL Buffer W (50mM Tris Acetate pH7.5, 5 mM $MgCl_2$, 100 mM potassium glutamate, 20 mM EGTA, 10% glycerol, and 30% PVP) was added to ground tissue sample and incubated for 15 min at 4°C. Samples were centrifuged for 10 min at 15,000 rpm and the protein concentration was measured using the Bradford assay. 0.5µM of substrate telomeric DNA was extended by 150 ng of protein extract together with 12.5µL Dynamo SYBR master mix for 45 min at 37°C. Extended products were then measured by qPCR as output to gauge telomerase activity, and normalized to wild type (WT) activity for the fold-change difference.

Telomere length was determined using Terminal Restriction Fragment (TRF) analysis. Genomic DNA was digested overnight with MseI enzyme at 37°C and resolved on a 0.8% agarose gel overnight at 55V. DNA on the gel was denatured and transferred to a nitrocellulose or PVDF membrane. The membrane was hybridized with 5' [^{32}P] labeled oligonucleotide probe $(TTTAGGG)_4$ and exposed with phosphor-screen and imaged. The image was analyzed using QuantityOne software.

In vivo DNA methylation status

Chop PCR was performed to analyze the DNA methylation status of the POT1a and POT1c loci in different genetic backgrounds. Briefly, restriction enzymes SnaBI, HypCH4IV, and HhaI were used to digest genomic DNA. Digested DNA was PCR amplified using the primer sets ChopPCRP1cF1/ ChopPCRP1cR1, ChopPCRP1aF1/ ChopPCRP1aR1, ChopPCRP1aF2/ ChopPCRP1aR2, and ChopPCRP1aF3/ ChopPCRP1aR3 (Table 2-1).

Primer name	Primer sequence
POT1c (Long & TAIR) F1 **(F1)**	CTCCTCCACAAAACACACAAGG
POT1c (Long) F1 **(F2)**	CCTTGTGTGTTTTGTGGAGGAG
POT1c NESTED Fw **(F3)**	CCGAAGGAAAATGATCAAAATTCGTC
POT1c ATG Fw **(F4)**	ATGGCGAAGAAGAGAGACA
POT1c (TAIR) R1 **(R1)**	CCTCAAATCAAGAACCACGACG
POT1c(TAIR) R2 **(R2)**	TCACTTAGCCAGAATAGATGC
POT1c TGA Rv **(R3)**	CTAGTTTTTGCAACCTGTTAATCC
qPCR POT1c Fw	TCGCATTAAGGCACAAAGAGTCTTC
qPCR POT1c Rv	GTAAAGTATGAACAGAGAGACTGTTAGCCAG
qPCR POT1a Fw	CACCACCAAGTTTTGTTGACACTGG
qPCR POT1a Rv	GCGTCTCTGTGTCCTGAAACCCTAA
qPCR GAPDH Fw	TTGGTGACAACAGGTCAAGCA
qPCR GAPDH Rv	AAACTTGTCGCTCAATGCAATC
Chop PCR P1a F1	GGC AAA GTG CCT AAT CCC TGA AAA
Chop PCR P1a R1	CCATCTCCCGCTTTGCTACA
Chop PCR P1a F2	TGTAGCAAAGCGGGAGATGG
Chop PCR P1a R2	TGGCTACAACCCGCAAAAAC
Chop PCR P1a F3	GTTTTTGCGGGTTGT
Chop PCR P1a R3	GATGAGAATCGCCGTCACC
Chop PCR P1c F1	GTT CCT GAC ACT CAT CTA GGG AAC
Chop PCR P1c R1	GATCATCCTGAGCGATATGATGAACC
Tm P1c CRISPR Fw	GAC ATG TTT CCC GAA ATC AAA TGG TC
Tm P1c CRISPR Rv	TGTTCAAGGACGATACCAATGAGAC
Seq P1c CRISPR Fw	GTTGTAGCAAAGCCGGGAGA
Seq P1c CRISPR Rv	TGGTTGATGAGTTTAATGGCGT
SS42	TCCCAGGATTAGAATGATTAGG
SS102	CACCATGTTATCACATCAATCC

Table 2-1. PCR primer

Plant transfection

Small scale *Agrobacterium tumefaciens* cultures were grown for 24 h at 30°C. 500mL LB was inoculated with 0.5mL of *A. tumefaciens* culture and was grown at 28°C until the OD_{595} reached 1.5. The culture was spun down for 10 min at 4,000 rpm, and was

re-suspended in 15% sucrose 0.2% Silwet-77 solution. Plants with flowers were dipped in Agrobacterium solution for 1 min, and kept in the dark overnight. T1 seeds were selected on [20μg/ml] BASTA, and ½ MS media [25μg/ml] Kanamycin for pDs transformants.

Results

Identification of POT1c

In a BLAST search to identify POT1-like genes in *A. thaliana*, a truncated gene termed POT1c (At2g04395) located 255 kbp upstream of AtPOT1a (At2g05210) was found (Figure 2-1A)[369]. Although POT1a and POT1b paralogs exist throughout the Brassicaceae family, POT1c is unique to *Arabidopsis thaliana* (Figure 2-1B). The absence of POT1c in *Arabidopsis lyrata* dates the POT1c duplication event at ~10 MYA[375,376]. Sequence analysis suggests that a proximal duplication and inversion of At2g05210 (POT1a) and At2g05220 (encoding ribosomal protein S17) created At2g04395 (POT1c) and At2g04390 (dS17) (Figure 2-1A).

The POT1c gene is annotated in The Arabidopsis Information Resource (TAIR) as a 474 nt transcript encoding a protein with 158 amino acids (Figure 2-1C). To verify this information, qPCR was performed with cDNA from flowers, seedlings, and leaves using the mRNA sequence predicted from TAIR. PCR with a high number of cycles (80) and a high cDNA concentration (500ng), a PCR product of 717 bp was generated in reactions with RNA from flowers, seedlings, rosette leaves, and cauline leaves (Figure 2-1C).

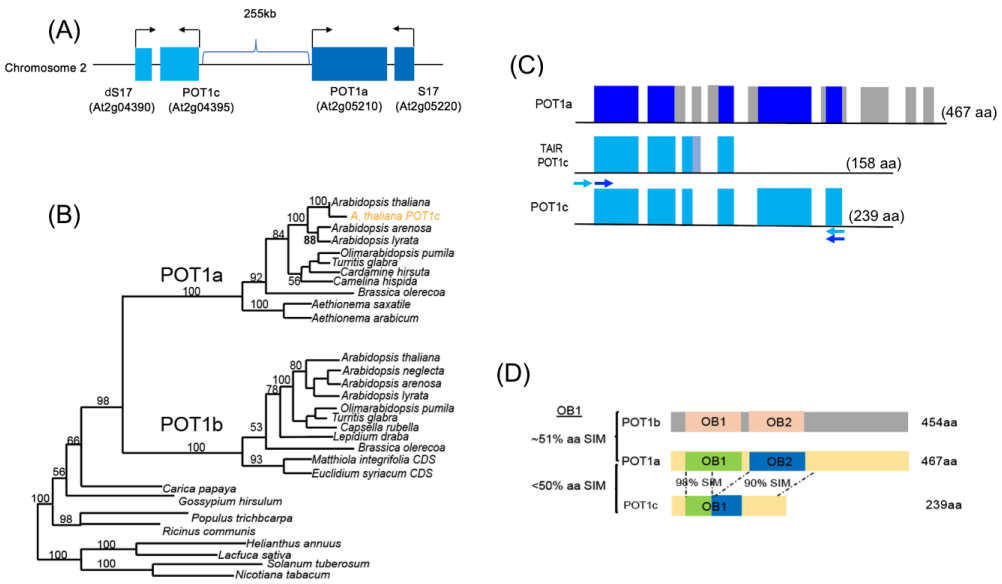


Figure 2-1. POT1c is a product of POT1a duplication. (A) The *A. thaliana* loci for POT1a (At2g05210) and S17 (At2g05220 (dark blue) and the duplicated locus for POT1c (At2g04395), and dS17 (At2g04390 (light blue) are shown. (B) Phylogenetic tree for the POT1 family in Brassicaceae and close relatives. POT1c (orange) is present only in *A. thaliana*. (C) Schematic of the POT1a, TAIR annotated POT1c, and POT1c loci. Boxes represent exons. Gray boxes in POT1a are missing in the POT1c transcript based on RT-PCR amplification. Arrows indicate two sets of primers used to amplify POT1c. (D) OB-fold domain outline of the *A. thaliana* POT1a, POT1b, and POT1c proteins. Percent amino acid similarity of the OB1 domains for POT1a-POT1b and POT1a-POT1c are indicated above (% AA SIM).

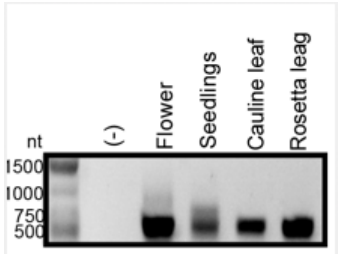


Figure 2-2. RT-PCR product of POT1c. A new PCR product derived from the POT1c locus is shown. The expected size product based on TAIR annotation is 474 nt.

Reactions with several different primer sets generated a POT1c product with a Cq value of 39, on par with background levels of expression (Figure 2-2). However, using standard This transcript encodes a 239 amino acid protein, longer than the POT1c polypeptide predicted from TAIR. The POT1c RNA we identified includes exons 5 and 6, while TAIR POT1c contains an in-frame stop codon after exon 4.

The secondary structure prediction for POT1c shows a single OB-fold domain that is lacking C-terminal α-helix (Figure 2-1D and Figure 2-3A). The N-terminal region of the POT1c OB-fold exhibits high amino acid sequence similarity to the N-terminus of POT1a OB1 (98%), while the C-terminal region of the POT1c OB-fold resembles the C-terminal region of POT1a OB2 (90% SIM) (Figure 2-1D). Overall, POT1c exhibits 47% ID/49% SIM to POT1a. While atypical, there are examples of OB-fold containing proteins that lack a C-terminal α-helix[279]. We generated a SUMO-tagged version of POT1c and were successful in expressing and purifying the recombinant protein from *E. coli*. These results indicate that POT1c can be expressed as a polypeptide that is not grossly unfolded (Figure 2-3B). Besides the presence of only a single OB-fold, a second major difference between POT1a and POT1c is the amino acid Phe at position 65 (Figure 2-3A). This amino acid is a highly conserved residue seen from Brassicaceae to plants with single copy POT1 that is important for telomeric DNA recognition[305,307]. In POT1c, Tyr is substituted for Phe at this position F65Y.

To ask if the POT1c protein can bind telomeric DNA *in vitro* similar to AtPOT1a OB1[307], gel shift experiments were performed. While 0.25μM AtPOT1a OB1 binds telomeric DNA, no DNA binding was observed with the same amount of SUMO-POT1c Figure 2-3C). When a higher concentration (up to 2μM) of SUMO-POT1c was tested, we

observed a weak association with telomeric DNA (Figure 2-3D). Although a faint band shift is observed at high protein concentration, the unbound signal remained unchanged. However, this weak binding is specific to telomeric DNA, as non-specific DNA does not compete with telomeric DNA (Figure 2-3E). To determine if the absence of robust telomeric DNA binding was due to the presence of Tyr instead of Phe in the nucleic acid binding pocket, we mutated POT1c to generate Y65F. We found no increase in telomeric DNA binding. Therefore, the low level of DNA binding by the POT1c protein is likely to reflect other differences in AtPOT1a OB1 and POT1c.

POT1c transcripts are undetectable

Because our PCR data indicated that POT1c was expressed at a very low level we analyzed the available *in silico* data for the POT1c locus[374]. In seedlings, there was no evidence of POT1c transcript accumulation. (Figure 2-4A), although a low level of POT1a transcript was detected, consistent with previous RT-PCR data[367]. POT1c has been reported to be expressed in the flowers from an F1 generation cross of the Civ x Ler accessions[377]. POT1c mRNA has also been detected in plants over-expressing IBL1, a protein involved in brassinosteroid signaling[378]. Nevertheless, these results suggest that POT1c is transcribed at very low to non-detectable level under standard growth conditions.

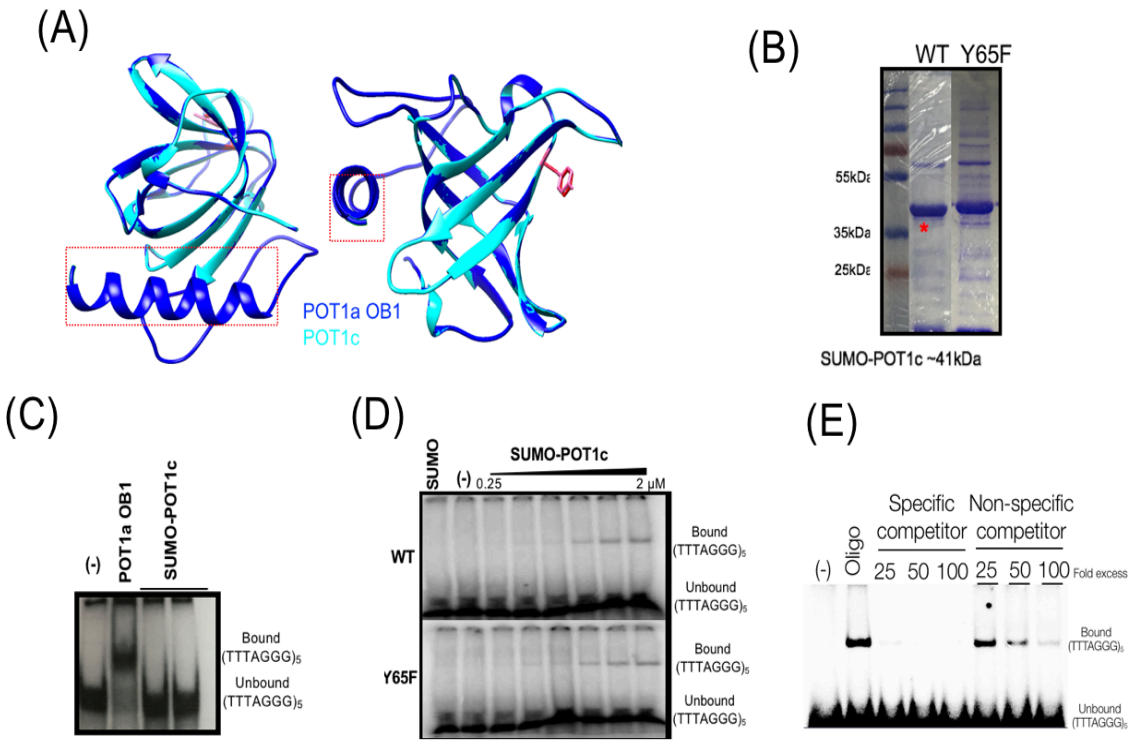


Figure 2-3. POT1c interaction with telomeric DNA *in vitro*. (A) 3D structure prediction overlay of POT1a OB1 (dark blue) and POT1c (light blue). Red box denotes the α-helix that is missing from POT1c. Conserved amino acid important for DNA recognition for POT1a (F65) is shown red and POT1c (Y65) shown in pink. (B) E. coli expressed and purified SUMO-POT1c. Y65F: POT1c with Y65F mutation. Expected size of SUMO-POT1c is 41 kDa. (C) Results from AtPOT1a OB1 and SUMO-POT1c electrophoretic mobility shift assay (EMSA). 0.25 μM protein was used to test for telomeric DNA binding. Substrate DNA is labeled with P^{32} at the 5'-end. (D) Results for WT and Y65F SUMO-POT1c EMSA. 0.25-2 μM protein was used to test for telomeric DNA binding. Substrate DNA is labeled with P^{32} at the 5'-end. (E) Results for WT SUMO-POT1c competition EMSA. 2 μM protein was used to test for DNA binding. Telomeric DNA is labeled with P^{32} at the 5'-end while non-specific DNA $(CCCTAAA)_5$ is not labeled.

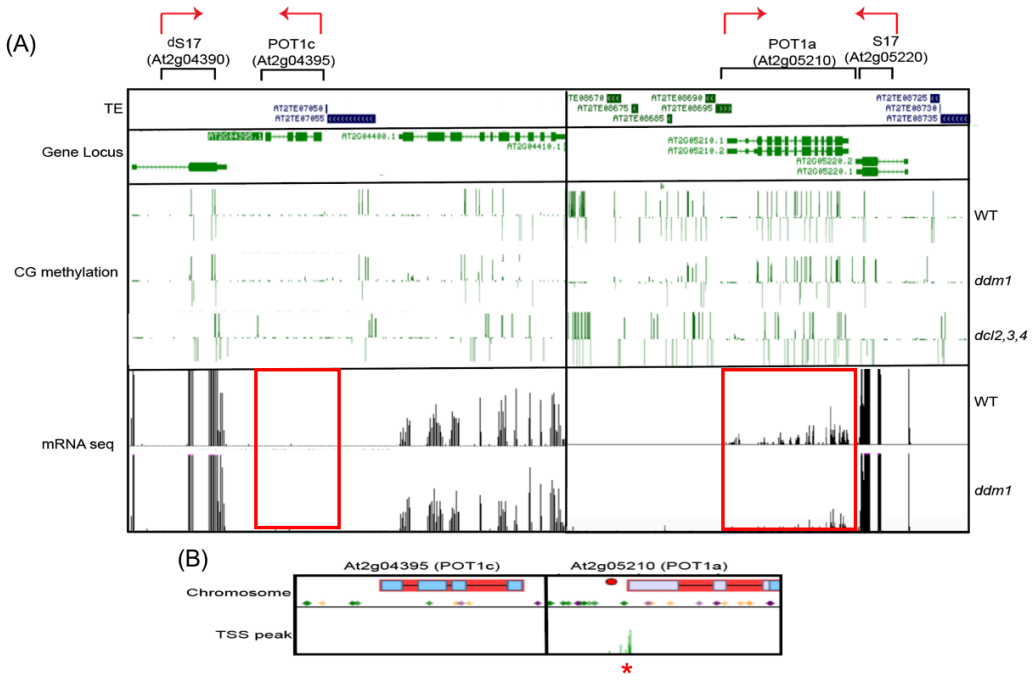


Figure 2-4. *In silico* analysis of the POT1c locus. (**A**) mCG methylation and RNA seq results for wild type (WT), *ddm1*, and *dcl2,3,4* mutant seedlings at the At2g04395 (POT1c), At2g04390 (dS17), At2gg05210 (POT1a), and At2g05220 (S17) loci from the Jacobsen database. Red arrows indicate the direction of transcription. (**B**) At2g04395 (POT1c) and At2gg05210 (POT1a) promoter regions. Red asterisk represents the detected Transcription Start Site (TSS).

POT1c mutants do not exhibit the molecular and developmental phenotypes associated with loss of POT1a or POT1b.

To ask if POT1c has a biological function, specifically a role in telomere maintenance or plant development, we generated a loss-of-function mutation in POT1c. A mutation in the At2g04395 locus was created with the *Streptococcus thermophilis* CRISPR/Cas9 system to target the first exon of the POT1c gene. The mutant allele, termed *pot1c-1,* had a deletion in the predicted start codon (Figure 2-5Figure 2-5A &B). If the *pot1c-1* allele were transcribed, the polypeptide derived from a downstream in-frame ATG

codon would lack the first two exons of POT1c, giving rise to a protein bearing only half of the residues needed for an OB-fold. Therefore, *pot1c-1* is very likely a null allele.

We tested if telomere maintenance was impacted in *pot1c-1* mutants using terminal restriction fragment (TRF) analysis to measure bulk telomere length and primer extension telomere repeat amplification (PETRA) to measure telomere length on individual chromosome arms. In contrast to the shortened telomeres in plants deficient in POT1a, *pot1c-1* telomeres were in the wild type size range and did not exhibit the sharp banding profile indicative of telomerase deficiency that is seen in plants lacking POT1a (Figure 2-6A &B)[172].

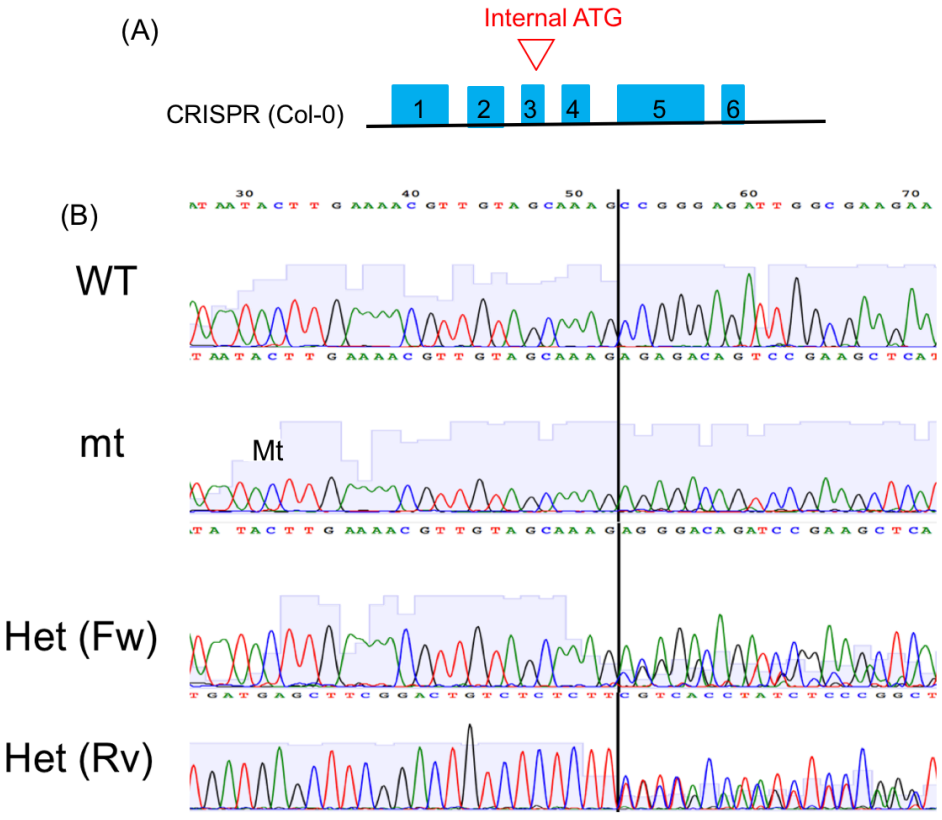


Figure 2-5. Generation of a POT1c mutant. (A) Location of deletion by CRISPR-Cas9 is indicated by the red triangle. (B) Chromatogram of wild type (WT), mutant (Mt), and heterozygous (Het) plants for POT1c locus. The vertical line indicates the start of deletion. Heterozygous chromatograms containing forward sequence (Het Fw) and reverse sequence (Het Rv) are shown.

To investigate whether telomerase activity was altered in plants deficient in POT1c, we performed the quantitative telomere repeat amplification protocol (TRAP). As expected, telomerase enzyme activity was significantly reduced in *pot1a* mutants, but was the same as wild type in the POT1c mutant (Figure 2-6C).

We next asked if *pot1c* depletion causes defects in plant growth and development similar to those associated with the loss of POT1b. Plants lacking POT1b exhibit decreased pollen viability. A second phenotype associated with *pot1b* mutants is short roots. Neither of these phenotypes is associated with the *pot1c* mutant (Figure 2-7A, B, and C). Indeed, we observed no remarkable difference in plant morphology or growth in *pot1c-1* mutants compared to wild type. One other notable phenotype in *pot1b* mutants is hypersensitivity to DNA damage.

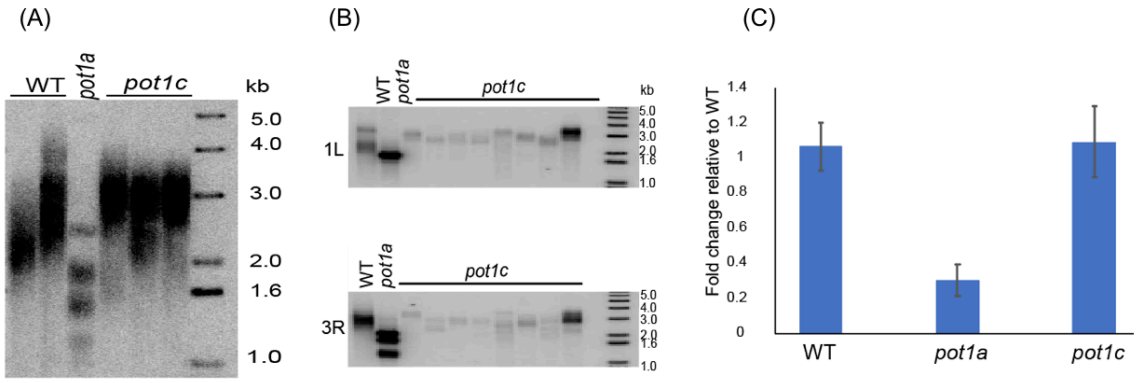


Figure 2-6. The *pot1c-1* mutant does not exhibit a defect in telomere maintenance or telomerase activity. (A, B) TRF and PETRA analysis were conducted to assess telomere length in WT plants and plants deficient in either POT1a or POT1c. (C) Telomerase activity levels in WT, *pot1a* and *pot1c* mutants were determined using quantitative TRAP. N=4.

We monitor DNA damage response by measuring programmed cell death in the root apical meristem after treatment with the radiomimetic drug zeocin. In contrast to *pot1b* mutants, the root apical meristem of *pot1c-1* mutants was indistinguishable from wild type following zeocin treatment (data not shown). Taken together, these data support the conclusion that POT1c does not act in the same manner as the other two *A. thaliana* POT1 paralogs, and they also argue that POT1c may be a non-functional gene.

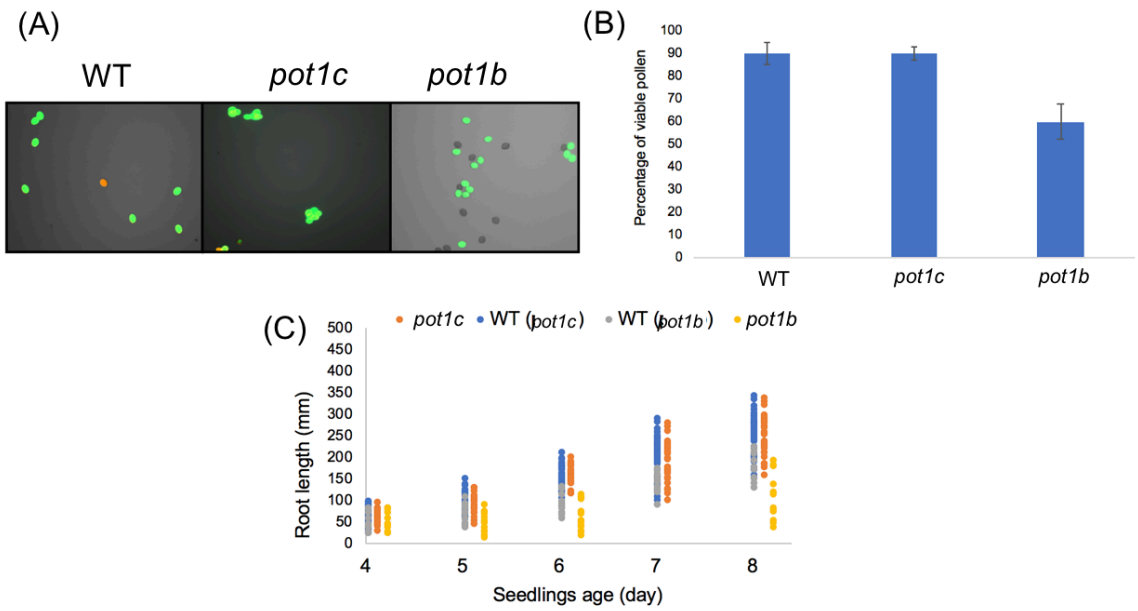


Figure 2-7. POT1c mutants exhibit no obvious developmental phenotypes. (A&B) Pollen viability assays conducted on WT and mutant plants. Fluorescein was used to identify the total number of live and dead pollen. (C) Root length measurements in WT and mutant seedlings.

The POT1c locus does not produce a proper mRNA

To test if expression of POT1c gives any phenotype, the entire 1,400 bp POT1c locus was expressed under ubiquitin promoter. Plants producing transcript from POT1c locus under ubiquitin promoter were identified by performing PCR for the POT1c transcript, and the transcripts produced from this construct were sequenced. Analysis of POT1c transcript sequence revealed several different splice product. Furthermore, all the different splice products are incomplete splicing where the location of the splice sites are not consistent among transcripts. This result indicates that POT1c locus has been altered so much from the original locus that current POT1c locus does not produce a proper transcript.

The POT1c locus is not conserved among different A. thaliana accessions

We leveraged the extensive *A. thaliana* sequence database to explore the evolutionary history of the POT1c locus to look for evidence of selective pressure. Our approach was to examine natural variation at this locus across 855 fully sequenced *A. thaliana* accessions (http://1001genomes.org/). We found that POT1a exhibits 86% nucleotide ID (87% amino acid ID), while for POT1c this value drops to only 12% nt identity (33% aa ID) (Table 2-2A). Analysis of the gene neighboring POT1a, S17, (At2g05220: S17) revealed 61% nt ID (74% aa ID), while the duplicated S17 (At2g04390: dS17) adjacent to POT1c exhibited 56% nt ID (60% aa ID) (Table 2-2A). Thus, both POT1a and S17 loci are more conserved across 855 *A. thaliana* accessions than the POT1c locus. The level of sequence divergence within the POT1c gene is particularly remarkable given that the dS17 locus, created at the same time as POT1c, is highly conserved, exhibiting 98% aa ID identity with its S17 parent (Table 2-2B).

A notable aspect of POT1c divergence is the prevalence of coding sequence mutations across *A. thaliana* accessions. 20% of the *A. thaliana* accessions contain mutations that will affect POT1c at the protein sequence level. In contrast, only 2% of the dS17 loci have mutations that will affect the dS17 protein sequence (Table 2-2C). In addition, the dS17 mutations that will affect dS17 protein sequence all appear to derive from an in-frame deletion located near the C-terminal end (Figure 2-8B). At the DNA sequence level, there are a few accessions that carry nucleotide substitutions at the dS17 locus, but these are silent mutations.

(A)

	At2g04390 (dS17)	At2g04395 (POT1c)	At2g05210 (POT1a)	At2g05220 (S17)
% Pairwise nt	96.9	96.9	99.8	99
% identical nt	55.8	11.9	86.1	60.9
% amino acid ID	60.7	33.5	87.6	74.3

(B)

	S17/ dS17	POT1a/POT1c
Overall % aa Identity	98	47

(C)

	dS17	POT1c
In-frame deletion	1.8	0
No ATG	0	9.1
Premature stop	0	2.8
Splice mutant	0	1.6
No stop codon	0	6.4
Total mutation	**1.8**	**19.9**

Table 2-2. POT1c is not conserved among 855 accessions. (A) % pairwise nucleotide, % identical nucleotide, and % amino acid identity between 855 accessions at indicated loci. (B) Overall amino acid % identity between S17 and dS17, and POT1a and POT1c. (C) The types of mutations in percentage are shown for each category. The total percentage of accessions with mutations is shown in red.

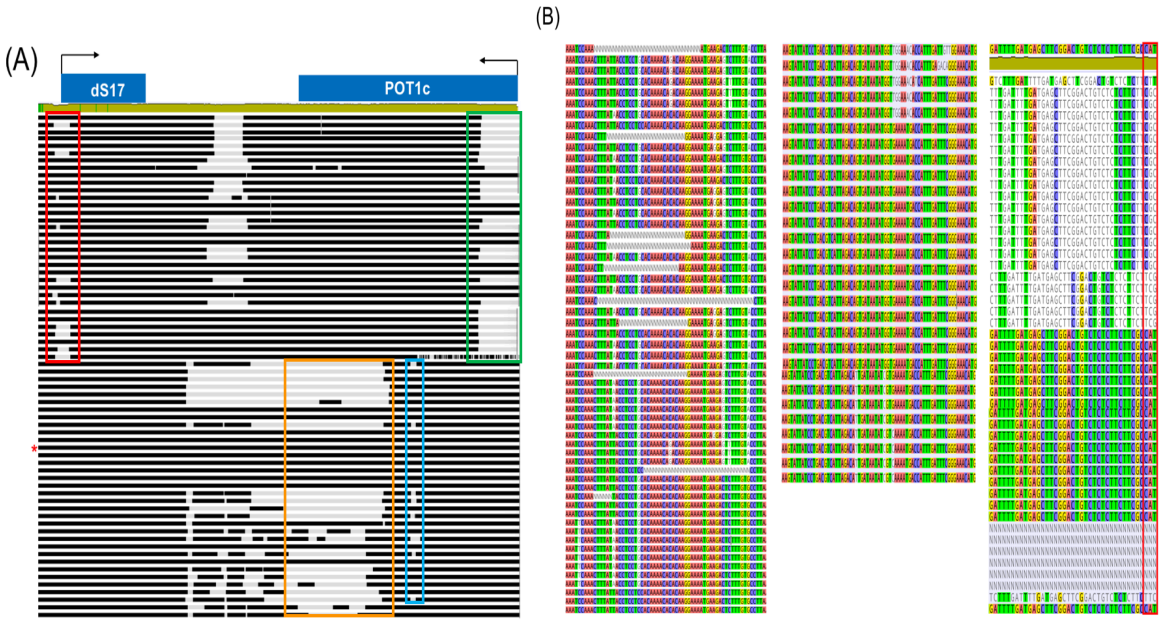


Figure 2-8. Comparison of POT1c and dS17 loci from different *A. thaliana* accessions. (A) The POT1c and dS17 coding regions are shown. Data for 67 representative accessions are presented to illustrate the different types of deletions within the POT1c and dS17 loci. Col-0 is shown with red asterisks. Boxes represent: dS17 deletion at ATG site (red box), POT1c deletion at ATG site (green box), POT1c deletion at exon 4 (orange box), and other internal deletions (blue box). (B) A blow up of the POT1c locus to illustrate the types of mutations. Red box on the right panel shows ATG site of accessions with deletions and nucleotide substitutions (letters with no color). Middle and left panel shows some accessions with nucleotide substitutions. Left panel shows example deletions within coding region.

On the other hand, the POT1c locus exhibits multiple nucleotide substitutions that are not silent and multiple random deletions within the coding region (Figure 2-8A &B).These deletions range from as little as 7nts to over 400 nts. Taken together, these findings suggest that coding region deletions are better tolerated at the POT1c locus than the dS17.

The POT1c locus is not actively silenced

Since gene silencing pathways play a critical role in modulating transcription in Arabidopsis[379,380], we used publicly available databases to investigate whether POT1c exhibits signatures of gene silencing, namely DNA methylation and small interfering RNA[370–373]. Arabidopsis displays three types of DNA methylation, CG, CHG, and CHH[381,382]. Notably, only CG methylation is associated with sequences immediately upstream of the POT1a and POT1c coding regions (Figure 2-4A). To investigate whether CG methylation is regulated in POT1a and POT1c loci, and if methylation affects the level of POT1a and POT1c transcripts, we analyzed the methylation profile in plants lacking the chromatin remodeler *DDM*1[383]. In hypomethylated *ddm1* mutants we found no change in the methylation profile of POT1c. However, CG methylation was decreased at the POT1a locus (Figure 2-4A). Biochemical analysis by ChopPCR confirmed these results (Figure 2-9A). Although DNA methylation was affected by the loss of DDM1 at the POT1a locus, *in silico* data showed no change in the transcript levels of POT1a or POT1c in hypomethylated plants (Figure 2-4A). In contrast, RT-PCR reactions using the AtMu1 TSA2 transposable element as a positive control showed a marked change in transcription (Figure 2-9B). These findings argue that DNA methylation does not play a significant

role in the transcriptional regulation of POT1a or POT1c.

We found no evidence that POT1c transcription is regulated by the small RNA pathway. The published sRNA sequencing data do not identify sRNAs near or within the POT1c locus[373]. To further investigate possible sRNA-mediated regulation, we assessed POT1c transcript levels in plants lacking three Dicer-like proteins, DCL2, 3, 4[380]. As expected, miR173, a TAS1 transcript, increased in the triple mutant. However, neither the POT1a nor POT1c transcript level was altered in this background (Figure 2-9B). In addition, there was no change in the DNA methylation profile for either POT1a or POT1c in the *dcl2,3,4* triple mutant, as determined from *in silico* data and our ChopPCR experiments (Figure 2-4A and Figure 2-9A).

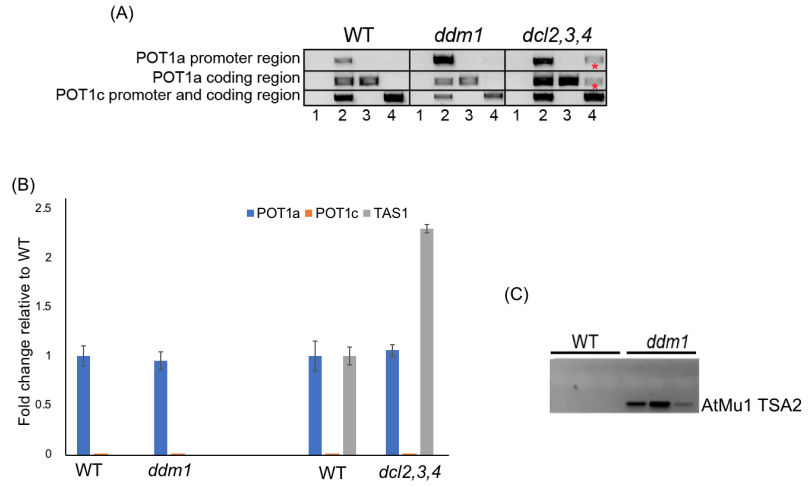


Figure 2-9. The POT1c transcript is not actively silenced. (A) Results for Chop PCR. No DNA (1), untreated (2), treated with DNA methylation sensitive restriction enzyme SnaBI, HypCH4IV, and HhaI (3), and treated with restriction enzyme McrBC that cuts at sites of CG methylation (4). Red asterisks indicate amplification of DNA from incomplete digestion by McrBC. (B) qPCR analysis of POT1c and POT1a transcripts in ddm1 and dcr2,3,4 mutants. POT1a, POT1c, and TAS1 transcripts were normalized to the GAPDH transcript level. The fold change in POT1c was adjusted relative to WT POT1a. The mir173 target TAS1 was used as a control in the dcl2,3,4 mutant. (C) ddm1 mutant RT-PCR. PCR data for WT and ddm1 were compared to the AtMu1 TSA2 transposable element to confirm the ddm1 mutation.

POT1c locus lacks a functional promoter

We used both *in silico* and molecular genetics approaches to investigate why POT1c expression is so low. Analysis of POT1a and POT1c loci revealed two transposable elements (TE), At2TE07050 and At2TE07055, immediately upstream of the POT1c ATG start codon (Figure 2-4A). No TEs are predicted within the POT1a promoter region. In addition, a single, strong promoter with a transcription start site (TSS) peak is predicted for the POT1a locus, but not for POT1c (Figure 2-4B). To directly test whether a functional promoter is associated with the POT1c locus, we created a transcriptional reporter construct bearing 845 bp of sequence upstream of the POT1c start codon fused to a GUS-reporter. As a positive control, we used a GUS-reporter under the control of CYCB1 in a *pot1a* mutant background. In these plants strong GUS signals are observed in all of tissues analyzed, as opposed to the negative control (WT), where not GUS signal was detected (Figure 2-10). Expression of the POT1c-GUS reporter construct was monitored in root, shoot apical meristem, embryo, pollen, and seeds. However, similar to the negative control, there was no GUS signal in any tissues analyzed. (Figure 2-10). These data are consistent with *in silico* and qPCR data indicating that under standard growth conditions, POT1c is not actively transcribed.

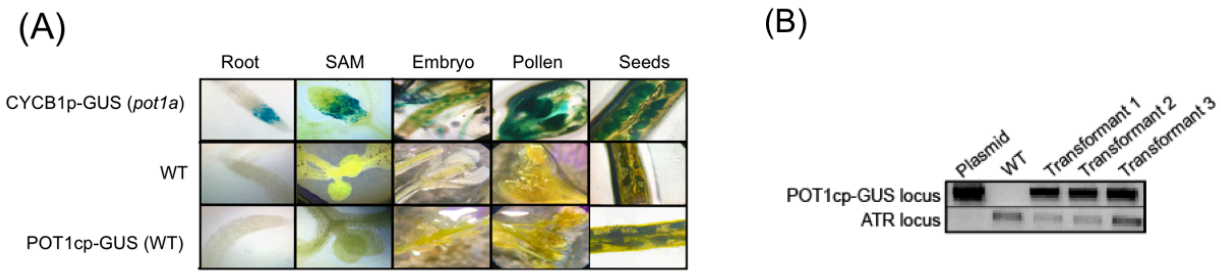


Figure 2-10. The POT1c locus lacks a functional promoter. (A) Positive control: GUS reporter under the control of the CYCB1 promoter in a pot1a mutant background. Negative control: untransformed WT. GUS reporter expressed from the POT1c promoter. SAM: shoot apical meristem. (B) PCR amplification of the POT1cp-GUS construct in three transformants. PCR amplification of ATR locus is shown at the bottom for DNA quality.

TE insertion occurred after the POT1a locus duplication

A striking feature of the POT1c locus is the presence of two TEs immediately upstream of the coding region. The TEs show low sequence conservation of 78.8% pairwise ID with 0% identical sequence across the 855 accessions. This observation suggests 1) TEs were ancestral to the locus or 2) were inserted very soon after the speciation event that separated *A. thaliana* from *A. lyrata*. Hypothesis 1 predicts that POT1c was essentially silenced immediately after its creation and is now subject to genetic drift. Hypothesis 2 argues that expression of POT1c was deleterious, leading to rapid silencing by the TE insertion.

To explore these two possibilities, we compared the *A. thaliana* POT1a and POT1c loci to the POT1a locus from *A. lyrata* (Figure 2-11). We identified a 200 nt sequence in *A. thaliana* positioned between the TrpC gene and the POT1c start codon that is conserved with E- value of 3.57152e-45 in *A. lyrata*. Surprisingly, in *A. lyrata* this 200 nt sequence is not located adjacent to TrpC, but rather is embedded in the p450 gene some 700 nts

upstream of the POT1a start codon (Figure 2-11). Because the two TEs associated with POT1c reside between the POT1c and the 200 nt conserved element, this observation indicates that the TEs were inserted after the duplication event that gave rise to the POT1c/dS17 locus (Figure 2-11).

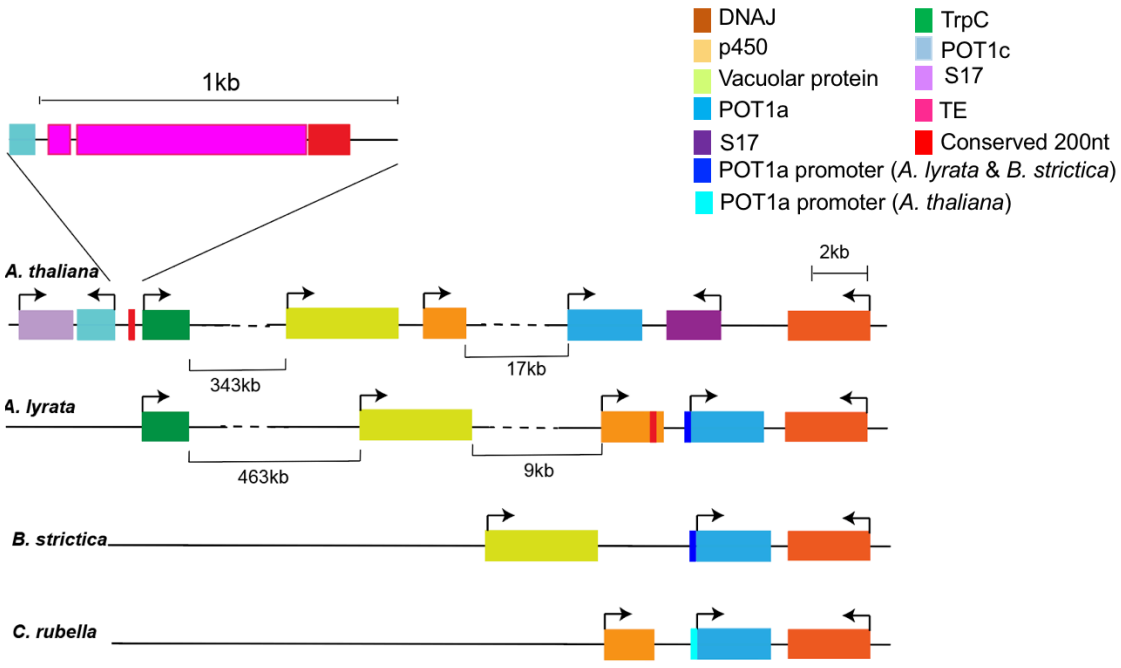


Figure 2-11. Organization of the POT1a and POT1c loci in Brassicaceae. A schematic diagram of the POT1a and POT1c loci shows gene rearrangement. POT1c and dS17 (change in the diagram) are unique to *A. thaliana*, as are the TEs located upstream of AtPOT1c.

Discussion

Gene duplication drives genetic diversity, allowing speciation[384,385]. In this study we examined how gene duplication has impacted POT1, one of the most highly conserved components of the telomere complex. We report the rise and fall of a new POT1 gene in *A. thaliana*. POT1c harbors single OB-fold, compared to POT1a and POT1b that has two OB-folds, and there are several amino acid residues that are POT1c specific , including the key amino acid in nucleic acid binding pocket where the phenylalanine is substituted to tyrosine. While *POT1c* locus is present is all *A. thaliana* accessions, the sequence is not well conserved. In addition, some accessions have deletions in the ORF that would prohibit the production of a polypeptide.

Although a POT1c transcript was reported in the TAIR database, we could find no evidence of POT1c transcription under standard growth conditions. Further, *in silico* promoter analysis failed to identify a transcriptional start site for POT1c, and experiments with a GUS reporter construct showed no promoter activity for the locus. On the chance that POT1c is expressed in a tiny fraction of critical cells, we created a CRISPR line to disrupt POT1c function *in vivo*. Plants with POT1c knockout show no telomere defect. In the "out of pollen" hypothesis, new genes of flowering plants often have expression bias to mature pollen[386]. Such case was not observed in *pot1c* mutants: no pollen defects or developmental defects of any kind were observed.

A striking feature of the POT1c locus is the presence of two TEs upstream of the *ORF*. We asked whether TEs were inserted before or after the POT1c locus was derived using an evolutionary reconstruction approach to evaluate the *POT1a* and *POT1c* loci from species closely related to *A. thaliana.* We discovered that the original *POT1c/dS17*

duplication product contained additional ~ 1 kb sequence upstream of *POT1a* gene. However, only 200 nt of the upstream sequence is retained upstream of the POT1c gene, indicating TEs were inserted upstream of the POT1c start codon. We present a model for the evolution of the *A. thaliana POT1a* and *POT1c* loci (). After *A. thaliana* speciation, *S17* was integrated between *POT1a* and *DNAJ*. This genome segment, corresponding to a portion of *p450*, *POT1a*, and *S17*, was then duplicated and inserted 255 kb upstream of *POT1a*. In the process, some sequences were deleted, creating *POT1c* from the *POT1a* locus. A functional promoter may have been associated with the early *POT1c* locus, but it was soon disrupted by the insertion of TEs. An additional 17 kb sequence was inserted upstream of the *POT1a* coding region, giving rise to a new promoter. Thus, POT1c locus seems to have become non-functionalized through the TE insertion.

The local gene duplication event giving rise to the *POT1c* locus included the neighboring ribosomal protein-encoding gene *S17*, allowing us an opportunity to compare the fate of the two nascent genes. The *POT1c* locus is highly divergent across different *A. thaliana* accessions, but the *dS17* is highly conserved. Thus, the *dS17* locus, but not *POT1c*, was subjected to purifying selection. The fate of retained duplicate gene pairs is thought to be determined early with positive selection playing a significant role[387]. The *AtPOT1a* lineage provides an instructive example in this regard[319]. On the other hand, pseudogenes and redundant genes are fixed through neutral genetic drift, while sub-functionalized and neo-functionalized genes are fixed through genetic drift and selective advantage[388,389].

The half-life of *A. thaliana* genes derived from whole genome duplication (WGD) is estimated to be 17 million years[390]. *AtPOT1a* and *AtPOT1b* fall into this category[319]. In

contrast, the average half-life for genes like *POT1c* and *dS17* that arose from small-scale duplication (SSD) is typically only a few million years[391]. *A. thaliana* harbors approximately 187 species-specific duplicate gene pairs; 40% of them including *POT1c* are not expressed and have become non-functionalized[392]. Other such gene pairs are expressed and have redundant function[392]. The ribosomal protein genes like *dS17* are known to evolve slowly, and are "over-retained" following gene duplication[393–395]. Because reduced expression of one ribosomal subunit can negatively impact ribosome assembly and lead to profound defects in physiology and development[396,397], the gene balance hypothesis[64] posits that ribosome components such as *S17* would be subject to critical dosage balance[398,399].

Was *POT1c* non-functionalized simply due to chance, or was there evolutionary pressure to inactivate the locus? Truncated and lowly expressed duplicate genes are preferred targets for non-functionalization due to their rapid divergence[400]. However, it is not known whether truncation of the *POT1c* ORF or its poor expression resulted from selective pressure. What is clear is that the POT1c locus was inactivated very soon after it formed. TEs are embedded in the POT1c promoter region of every *A. thaliana* accession, indicating that TE invasion was ancestral to diversification of the POT1c locus. In addition, *POT1c* is not subjected to active gene silencing through DNA methylation or the sRNA pathway, consistent with gene inactivation prior to establishing this type of regulatory mechanism.

Duplication of POT1 genes is rare across eukarya, and in all of the organisms where the *POT1* gene family has expanded, non-overlapping functions are reported for the paralogs[309,311,315,316]. Thus, proper dosage of POT1 may be critical for viability. Shakirov et

al. reported that over-expression of the C-terminus of AtPOT1a or the N-terminus of AtPOT1b resulted in telomere shortening and chromosome end deprotection, respectively[367], supporting the conclusion that a truncated form of POT1 could be highly deleterious. We were unable to test the hypothesis that expression of the *POT1c* had a dominant negative impact, since over-expression of the POT1c locus failed to produce a properly spliced transcript. However, if the newly born *POT1c* was detrimental to Arabidopsis, selective pressure would have been strong to quickly render the locus non-functional.

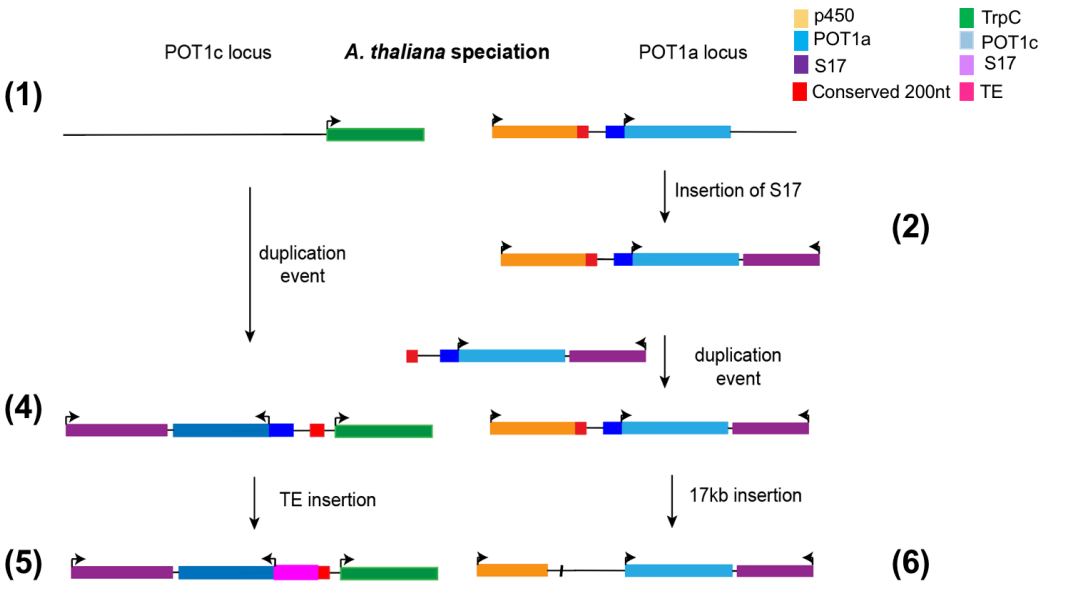


Figure 2-12. Model of POT1c and POT1a loci evolution in *A. thaliana*. (1) Original locus after *A. thaliana* speciation. (2) Insertion of S17 (purple) downstream of POT1a. (3) Local duplication of 3' p450 (orange), POT1a (light blue), and S17. (4) Inversion and insertion of duplicate gene (POT1c: blue). dS17 is shown as same purple as original locus S17. POT1c is shown as blue due to its partial deletion within the POT1c locus giving rise to unique POT1 gene compared to POT1a. (5) Transposable element insertion between POT1c ATG and 3' p450. (6) 17kb insertion between p450 and POT1a ATG, providing POT1a with a new promoter.

CHAPTER III

ARABIDOPSIS NOP2A DETERMINES TELOMERE LENGTH SET POINT

Summary

The maintenance of telomeres at chromosome ends is a fundamental cellular process, remarkably conserved across eukaryotic kingdoms. Telomere length is likely under strong stabilizing selection in nature, since accelerated shortening of telomere tracts can cause early onset of aging diseases[19,21,22,25,44]. Interestingly, mean telomere length shows genetic (heritable) trait variation among individuals in a number of species, including humans[106,401,402]; however, the genetic factors that establish telomere length set point remain elusive. Multi-parent Advanced Generation Inter-Cross (MAGIC) mapping populations provide an ideal resource to define candidate loci that underlie quantitative traits such as telomere length. By taking a systems genetic approach we were able to leverage this genetic diversity to identify novel gene, NOP2A, establishing this important element of natural variation. We found that loss-of-function NOP2A/OLI2 mutants displayed rapid telomere shortening, but which in later generations stabilized at a new shorter telomere length set point. Same phenotype is observed in other genes that are involved in OLI2 genetic pathway, specifically, OLI5 and OLI7. These genes are involved in ribosome biogenesis and rRNA processing as the major regulators of telomere length in Arabidopsis, directly connecting these two crucial cellular pathways. Since NOP2 is a known proliferation-associated antigen that drives cancer progression in human and mouse cells[403–406] and was previously shown to directly associate with human telomerase

enzyme[407], our findings in *A. thaliana* establish NOP2A as a major regulator of telomere length that connects telomere biology, translation and cell proliferation pathway.

Introduction

The length of the telomeric DNA is species-specific (i.e. 5-15 kb at birth in humans and 2-9 kb in the model plant *Arabidopsis thaliana*)[70–74] but considerable variation in average telomere length exist within mouse, yeast and plant populations and between human individuals[106,401,402]. This telomere length polymorphism is known to be under strong genetic control[109,408], but only a handful of genetic factors have been shown to cause population-specific telomere length diversity between or within species. The model plant, *Arabidopsis thaliana*, offers an ideal system to advance these goals due to its small genome size, massive genetic resources and well documented heritable telomere length variation[106].

Multi-parent Advanced Generation Inter-Cross (MAGIC)[341] and other complex mapping populations provide an ideal resource to define candidate loci that underlie quantitative traits such as telomere length. Unlike Genome Wide Association Study (GWAS), MAGIC populations have equal power to detect rare and common variants. Furthermore, in *Arabidopsis thaliana* the MAGIC population samples the diversity of 19 genotypes, providing far more raw genetic variation than typical bi-parental designs. By taking a systems genetic approach — combining DNA, RNA and quantitative genetic polymorphism — we were able to leverage this genetic diversity to identify novel genes establishing this important element of natural variation.

Here we show that analysis of natural DNA sequence polymorphism in *Arabidopsis thaliana* MAGIC population identifies important genes controlling telomere length variation. Specifically, we identified a 848kb major effect QTL peak on distal end of chromosome 5 in a multi-parent mapping population. Allelic variation at this QTL explained 47% of telomere length variation. To search for candidate genes, we employed a systems-genetic approach by combining DNA variant, gene expression polymorphism, T-DNA mutation knock-out and quantitative complementation comparative analyses. We found that loss-of-function NOP2A/OLI2 (At5gg55920) mutants displayed rapid telomere shortening in the first 3 plant generation, but which in later generations stabilized at a new shorter telomere length set point. The same phenotype is observed in other genes that are involved in OLI2 genetic pathway, specifically, OLI5 and OLI7.

Arabidopsis At5g55920 gene is annotated as OLIGOCELLULA 2 (OLI2) and encodes a homolog of the human and *S. cerevisiae* NOP2, a putative S-adenosyl-L-methionine-dependent methyltransferase superfamily protein involved in ribosome biogenesis, cell proliferation and cancer progression[403,409–411]. NOP2 gene family is highly conserved and is present in yeast, plant and mammalian genomes. Human NOP2 is a proliferation-associated antigen that is highly expressed in a wide range of malignant tumors.[403] Mouse NOP2 has been characterized as a cancer-causing gene that drives cancer progression[412]. More recently, human NOP2 was purified as telomerase accessory protein and, through contacts with hTERT, NOP2 was shown to stimulate transcription of cyclin D1[407]. Similarly, *A. thaliana* NOP2A/OLI2 gene plays a role in organ size control by promoting cell proliferation and preventing compensation in normal leaf development[413]. OLI5 and OLI7 are two other genes identified with OLI2, and these two genes encodes

ribosomal protein L5.

Therefore, genes involved in ribosome biogenesis and rRNA processing in Arabidopsis are directly connected to the two crucial cellular pathways: cell proliferation and ribosomal biogenesis. This findings in *A. thaliana* establish NOP2A as a major regulator of telomere length that connects telomere biology, translation and cell proliferation pathway.

Materials and methods

Multi-parent QTL mapping

We conducted QTL mapping in the *A. thaliana* multi-parent advanced generation intercross (MAGIC) population[341]. The genotype matrix was imputed using shallow resequencing data[414] via the construct.haplotype function within the HAPPY software package[341]. We employed the accompanying happy.hbrem algorithm available in the R environment for statistical computing to infer QTL peaks. Confidence intervals were constructed around QTL peaks using a $-\log_{10}$ *P*-value 'drop' of 5. That is, we assumed the causal locus of a QTL peak existed within an interval where the P-value was no more than 5-orders of magnitude greater than the peak minimum *P*-value. This provided a very conservative genomic region with which to screen candidate genes.

Computational screen for candidate genes

We screened for candidate genes within the QTL interval on chromosome 5 using a combination of gene expression and DNA sequence polymorphism. Allelic effect distributions at the QTL peak demonstrated that the SF-2 parent contributes a private allele

that was responsible for causing telomere length variation. To infer candidate genes, we counted all SNPs that were private to the SF-2 parental genome for each gene in the interval. We also compared parental gene expression via a t-test comparing normalized expression of the SF-2 parent to the other parental expression values. Genes with >1 SF-2 private SNP in the CDS and FDR-corrected P-value of differential expression ≤ 0.05 were considered candidates.

Telomerase activity assay (TRAP)

Protein extract was prepared from 6 day-old seedlings. 1 mL Buffer W (50 mM Tris Acetate pH7.5, 5 mM $MgCl_2$, 100 mM potassium glutamate, 20 mM EGTA, 10% glycerol, and 30% PVP) was added to ground tissue sample and incubated for 15 min at 4°C. Samples were centrifuged for 10 min at 15,000 rpm and the protein concentration was measured using the Bradford assay. 0.5μM of substrate telomeric DNA was extended by 150 ng of protein extract together with 12.5μL Dynamo SYBR master mix for 45 min at 37°C. Extended products were then measured by qPCR as output to gauge telomerase activity, and normalized to wild type (WT) activity for the fold-change difference.

Telomere length analysis (TRF)

Bulk telomere length was determined using Terminal Restriction Fragment (TRF) analysis. Genomic DNA was digested overnight with MseI enzyme at 37°C. Digested DNA was resolved on a 0.8% agarose gel overnight at 55V. DNA on the gel was denatured and transferred to a nitrocellulose or PVDF membrane. The membrane was hybridized with 5' [^{32}P] labeled oligonucleotide probe $(TTTAGGG)_4$, and exposed with phosphor-

screen and imaged. The image was analyzed using QuantityOne software.

Results

Genetic mapping of telomere length

We surveyed telomere length in the *A. thaliana* MAGIC population using telomere restriction fragment (TRF) assay, which employs Southern blots to obtain standardized mean telomere length values for each MAGIC line. Telomere length (mean TRF) was normally distributed across the MAGIC population, ranging from ~1.7 kb to 8 kb, suggesting a polygenic genetic architecture (Figure 3-1A). Quantitative Trait Locus (QTL) analysis of mean TRF was performed using the mpMap software package, which takes advantage of the Hidden Markov Model for a multipoint probabilistic reconstruction of the genome of each MAGIC line, permitting statistical inference at the haplotype level. This analysis identified a major effect ~848 kb QTL interval on chromosome 5 and several minor QTL on chromosome 1 (Figure 3-1B), which together account for 55% of telomere length variation in the MAGIC population. When taken in isolation, allelic differences at the major effect chromosome 5 QTL explain 47.1% of the total mean telomere length variation present in the MAGIC population. Interestingly, the QTL interval did not harbor any known genes involved in telomere biology; therefore, the causal DNA polymorphism was associated with a novel gene.

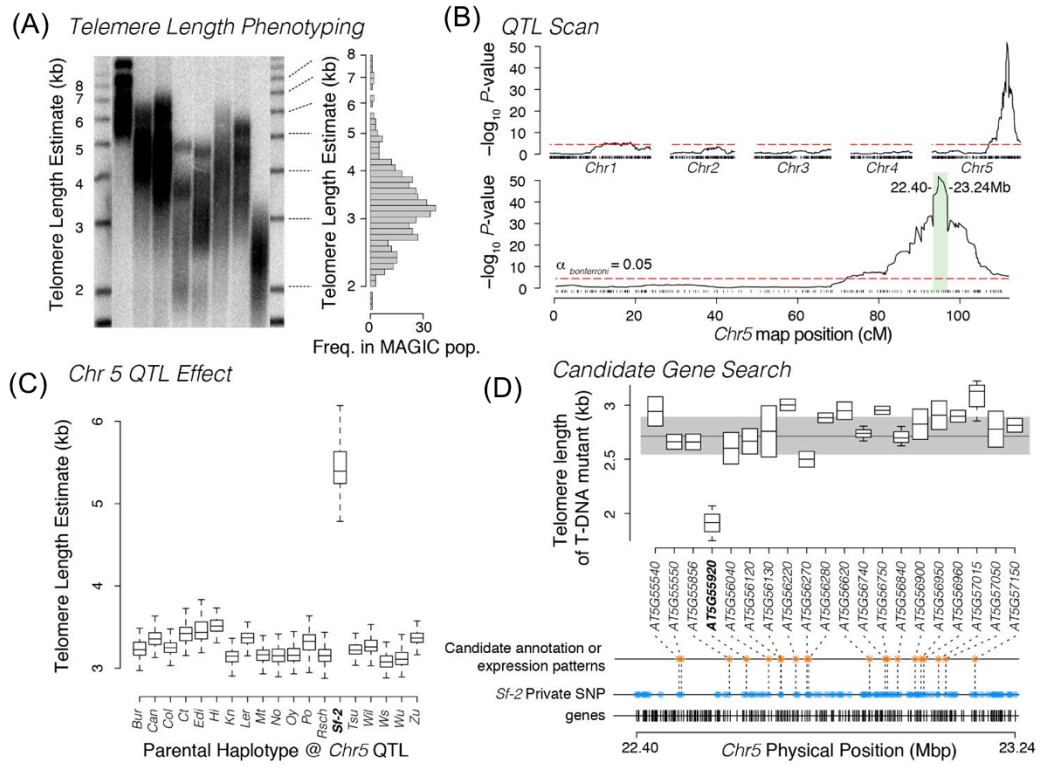


Figure 3-1. Genetic mapping of telomere length variation in Arabidopsis. (A) Representative picture of telomere length variation in *A. thaliana* MAGIC lines. Overall pattern of mean telomere length distribution (in kb) in the MAGIC lines is shown on the right. (B) QTL scan of mean telomere length in the MAGIC population. The majority of natural genetic variation in telomere length of Arabidopsis MAGIC lines is explained by a large effect QTL on chromosome 5 and several minor effect peaks on chromosome 1 (upper panel). Genome-wide permutation threshold (alpha=0.05). Major effect QTL is localized to ~848 kb on the right arm of chromosome 5 (lower panel). (C) Estimates of contribution of 19 parental haplotypes at Chromosome 5 QTL to telomere length (in kb) in *Arabidopsis thaliana* MAGIC lines. Plants that inherited the rare Sf-2 allele show up to 50% longer telomeres. (D) The high confidence ~848 kb QTL interval on chromosome 5 contains 249 genes (black bars, bottom), of which 112 genes harbor one or more Sf-2 specific polymorphisms (blue dots). Of these, 20 genes have been selected for transgenic analysis due to the presence of private Sf-2 alleles, statistical differences in expression between SF-2 and other parental accessions or due to interesting annotation (orange asterisks). Knockouts of 19 of these genes harbor telomeres in the normal range for the wild type Col-0 accession (grey shaded area). However, transgenic T-DNA knockouts of At5g55920 gene showed the strongest effect on telomere length, indicating that variation at this locus may be responsible for the observed QTL.

To define a list of potential candidate genes, we examined the parental allelic effect at the chromosome 5 QTL. While there was little variation among 18 of the 19 alleles, MAGIC progeny that inherited the Sf-2 allele exhibited a 50% increase in telomere length (Figure 3-1C). The identified chromosome 5 QTL interval contained 249 candidate genes based on the TAIR 10 Columbia (Col-0) reference genome (Figure 3-1D, black bars at the bottom). We scanned this interval in re-sequenced genomes of all 19 parental accessions for the presence of candidate genes with known Sf-2 specific DNA or gene expression polymorphism. Of the 249 genes, 112 genes harbored one or more Sf-2 specific DNA variants within coding or promoter (2kb upstream) regions (Figure 3-1D, blue dots). We then employed an analogous approach to gene expression, and found 33 genes with statistically significant expression differences between Sf-2 and all other parental MAGIC accessions. To further rank candidate genes, we examined annotations of all 112 genes with Sf-2 specific polymorphisms, excluding very unlikely candidates (e.g. photosystem II genes) and specifically focusing on loci possibly involved in DNA/RNA binding or metabolism, development, signal transduction or chromosome biology. Overall, 20 candidate genes with Sf-2 specific polymorphisms that either showed unique expression pattern in Sf-2 or interesting annotations or both were selected for further genetic analysis (Figure 3-1D, orange dots).

Identification of NOP2A/OLI2 as a candidate gene for telomere length

To analyze the potential role of selected genes, we screened ≥3 biological replicates of homozygous T-DNA insertion mutant plants for each of the 20 candidate genes inside the chromosome 5 QTL interval. T-DNA mutants of 19 of the 20 genes displayed

telomeres in the wild type range of the Columbia (Col-0) background in which they were generated (Figure 3-1D, grey shaded area in the upper panel). Strikingly, homozygous T-DNA *nop2a-2* mutants of *NOP2A/OLI2* (*OLIGOCELLULA 2*, At5g55920) gene harbored telomeres that were ~30% shorter than the wild type (Figure 3-1D). *Arabidopsis NOP2A/OLI2* encodes a homolog of the human and *S. cerevisiae* NOP2, a putative S-adenosyl-L-methionine-dependent methyltransferase superfamily protein thought to reside in nucleolus and involved in ribosome biogenesis, cell proliferation and cancer progression[403,409–411]. Similarly, in *Arabidopsis* the *NOP2A/OLI2* gene controls organ size by promoting cell proliferation and preventing compensation in normal leaf development, and *nop2a-2* mutant plants that are null for *NOP2A* function are characterized by up to 2-fold reduction in the total number of cells in the first true leaves[413].

To further characterize telomere length phenotypes observed in *nop2a* mutants, we analyzed a total of three mutant lines, *nop2a-2* (SALK_129648), *nop2a-3* (SALK_082871) and *nop2a-4* (SAIL_1279_H03), with T-DNA insertions in different locations inside the coding region of *AtNOP2A* gene (Figure 3-2A). Indeed, all individual homozygous plants with *NOP2A* gene disruptions displayed similar telomere length shortening (~30% reduction) as compared to Columbia wild type (Figure 3-2B), regardless of the T-DNA insertion site. Since multiple loss-of-function mutants display dramatically shortened telomeres, we conclude that *NOP2A* is a positive regulator of telomere length in *Arabidopsis thaliana*.

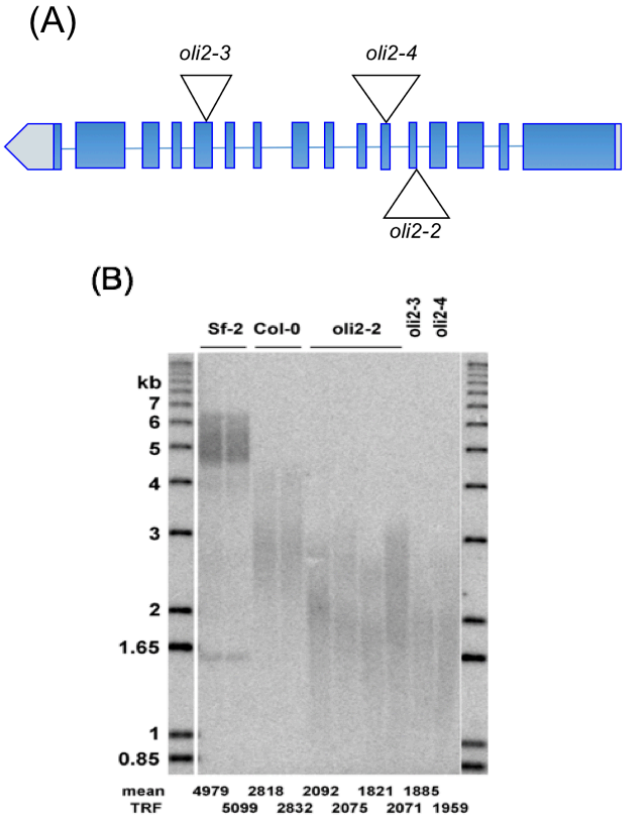


Figure 3-2. *AtNOP2A* gene has short telomeres. (A) Map of the *AtOLI2* gene. Open blue boxes represent exons; black lines, introns; and gray boxes, untranslated regions. Positions of T-DNA insertion sites and the corresponding T-DNA numbers are indicated. (B) T-DNA mutant lines of the OLI2 gene (At5g55920), *oli2-2*, *oli2-3* and *oli2-4* show shorter telomere length than the corresponding wild type, accession Col-0. Mean telomere length (mean TRF) for each individual analyzed plant is indicated at the bottom. Two additional mutant lines with T-DNA insertions inside the *OLI2* gene display shorter telomere length.

NOP2A gene disruptions cause re-establishment of telomere length at a shorter set point, resulting in an alternative telomere-length homeostasis

An important aspect of telomere length characterization is to determine if a particular gene affects telomere length maintenance or establishment. If telomeres continuously shorten from one plant generation of the mutant to the next (the so-called *ever-shorter-telomere* phenotype), this outcome indicates that the gene contributes to

telomere length maintenance. For instance, in most mutants of the genes involved in the telomerase pathway telomere length progressively shortens across generations until capacity of such cells to proliferate is depleted[5,136,363,415]. Alternatively, if shorter telomere length becomes fixed in the first few generations and is later maintained at this shorter level, this phenotype implies that a new equilibrium has been established and thus set point is affected in the mutant.

To test if Arabidopsis *NOP2A* mutations affect telomere length maintenance or establishment, we analyzed several consecutive generations of *NOP2A* T-DNA mutants. Surprisingly, *nop2a-2* and *nop2a-3* mutants did not display additional telomere length shortening across generations (Figure 3-3A) beyond what we observed in plants from the initial batch of seeds received from the ABRC stock center. Since both the *nop2a-2* and *nop2a-3* mutants were already homozygous for the *NOP2A* mutations when we identified them, we suspect that these T-DNA lines had already been propagated at the Arabidopsis stock center for several generations in the absence of *NOP2A* prior to our analysis. Therefore, we performed multigenerational analysis of homozygous *nop2a-4*$^{-/-}$, wild type *nop2a-4*$^{+/+}$ and heterozygous *nop2a-4*$^{+/-}$ siblings, which were freshly segregated by self-pollination from the heterozygous *nop2a-4*$^{+/-}$ parent.

As expected, telomeres in the three consecutive generations of wild-type *nop2a-4*$^{+/+}$ sergeants appeared as a homogeneous smear of products with mean telomere length in the normal Col-0 range of 2.8-3.2 kb (Figure 3-3B). Interestingly, as for many other Arabidopsis telomere length mutants[172,363], *At NOP2A* is not haplo-insufficient for telomere maintenance, as plants heterozygous for the *nop2a-4* T-DNA insertion (*nop2a-4*$^{+/-}$) exhibited a wild-type telomere profile (Figure 3-3B). In contrast, the first three

generations of homozygous *nop2a-4*$^{-/-}$ plants displayed progressive telomere shortening from one generation to another (Figure 3-3B). Notably, the rate of telomere shortening was the fastest in the first generation G1 (497 bp) but slowed down to 394 bp in G2 and further to 250 bp in G3.

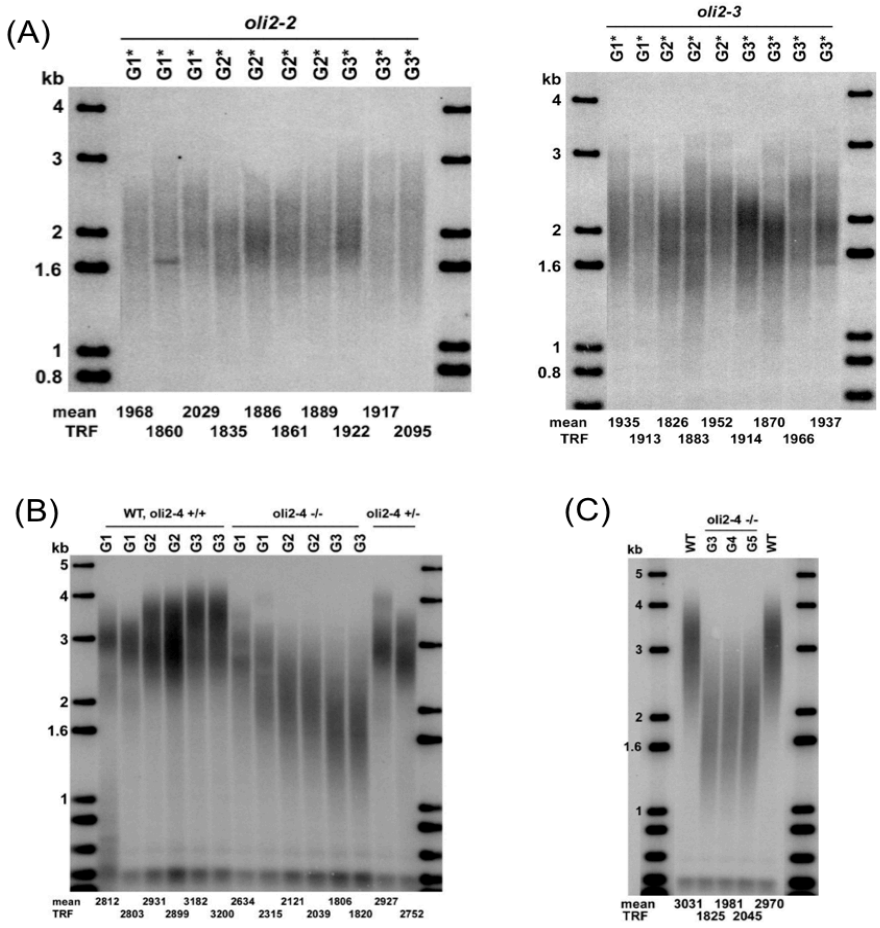


Figure 3-3. Inactivation of *AtNOP2A* gene leads to the establishment of the new shorter telomere length set point. (A) Telomere length is stable at a new shorter set point in several generations of *oli2-2* and *oli2-3* mutant plants. Asterisks indicate generations propagated in the lab, with G1* being homozygous for *oli2-2* mutants, whose seeds were received directly from the ABRC stock center. (B) Telomere length analysis of several consecutive generations of *oli2-4* mutants. G1 plants are wild type, heterozygous and homozygous progeny of a single self-pollinated heterozygous *oli2-4*$^{+/-}$ plant. (C) Telomere length analysis of consecutive G3-G5 generations of *oli2-4*$^{-/-}$ mutants.

Remarkably, upon reaching the new stable level at 1.8-2.0 kb, telomere length stabilized in G3 and was maintained at this lower set point in the consecutive G4 and G5 generations of *nop2a-4* knockout plants (Figure 3-3C). Taken together with results on *nop2a-2* and *nop2a-3* mutants, these data indicate that plants deficient in *NOP2A* gene function display progressively shorter telomeric DNA tracts until a new telomere length homeostasis is reached at the level ~1 kb shorter than in the wild type Col-0 plants. Such phenotype is novel in Arabidopsis and indicates that *NOP2A* gene functions in the establishment of species-specific telomere length set point.

Arabidopsis NOP2A is necessary for full telomerase activity in vitro

Similar to the situation in the first 3 generations of *nop2a-4* mutants, plants with inactivated telomerase subunits TERT and POT1a also lose 200–500 bp of telomeric DNA per plant generation[172,363,416]. However, these mutants are unable to stabilize their telomeric tracts at a new shorter set point, which in later generations ultimately leads to massive genome instability and a complete loss of cellular proliferation capacity[416]. These observations prompted us to biochemically and genetically evaluate the possible connection between NOP2A and telomerase in Arabidopsis. First, we considered the possibility that NOP2A is required for telomerase enzyme activity. Using quantitative telomerase activity Q-TRAP assay, we measured *in vitro* telomerase activity levels in extracts from *nop2a-2* mutants, wild type plants and *pot1a* seedlings, which were previously shown to have a 10-fold reduction in telomerase activity. Remarkably, *nop2a-2* and *pot1a* mutant plants harbored similar levels of telomerase activity reduction (Figure 3-4A). These data implicate NOP2A in the promotion of maximum telomerase activity *in*

vitro, though the reduced level of telomerase activity observed in *nop2a-2* mutants is still sufficient to stably maintain telomere length at the new shorter set point of 1.8-2.0 kb.

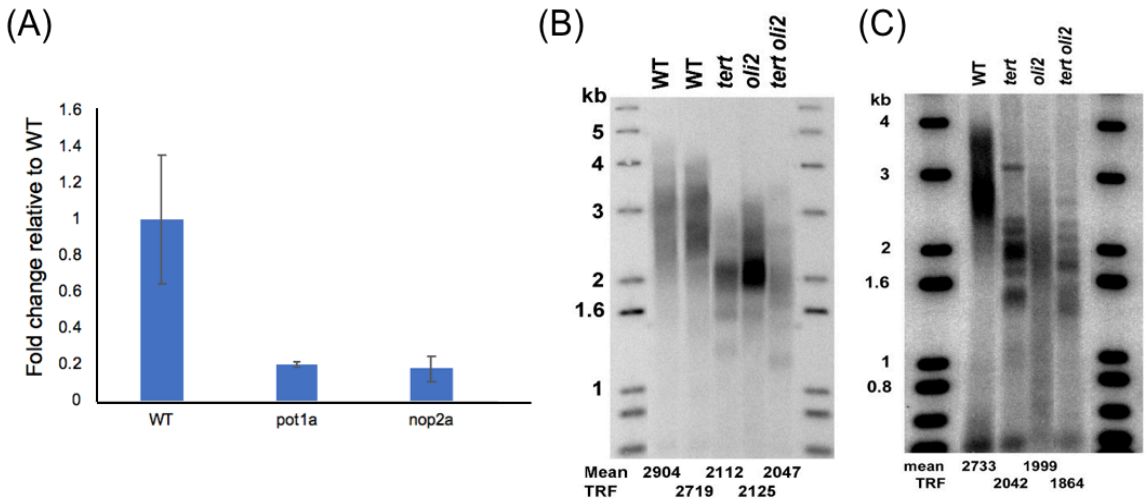


Figure 3-4. NOP2A association with telomerase. (A) Quantitative TRAP assay for telomerase activity. Telomerase enzyme activity levels vary between natural Arabidopsis accessions but do not significantly change in *oli2-2* mutants. (B&C) Telomeres shorten in *oli2-2* mutant plants to the similar degree as in Arabidopsis tert mutants. *oli2-2* mutant plants were crossed with mutant plants heterozygous for T-DNA insertion in the catalytic telomerase subunit *tert*. Double heterozygous for *oli2-2* and tert mutations F1 plants were self-pollinated to generate *oli2-2* and *tert* single homozygous F2 plants (B) and F3 plants (C) and plants double homozygous for both mutations, as well as wild-type. Telomeres shorten in double mutants to the similar degree as in either single mutant.

To assay for possible genetic interactions between *NOP2A* and telomerase *in vivo*, individual *tert* and *nop2a-2* T-DNA mutants were crossed, and double heterozygous F1 mutants were allowed to self-pollinate to generate F2 progeny. Homozygous *nop2a-2, tert*

and *nop2a-2 tert* F2 mutant plants were identified by genotyping and assayed for telomere phenotypes. Both individual F2 *nop2a-2* and *tert* mutant siblings showed a similar degree of telomere shortening (Figure 3-4B). Furthermore, telomeres in *nop2a-2 tert* double mutants shortened at approximately the same rate as in their single mutant siblings (Figure 3-4B). Similar results were obtained in F3 plants (second generation of single and double mutants) (Figure 3-4C). Overall, the available genetic and biochemical data indicate that *NOP2A* and telomerase mutants may function in the same genetic pathway.

Although NOP2A mutants are viable and fertile, they harbor up to 50% fewer leaf cells, consistent with major defects in cell proliferation[413]. We next asked if such severe morphological abnormalities and deficiency in telomere length maintenance are associated with cytogenetic defects and genome instability, such as chromosome fusions. As a gauge of genome stability, we determined the frequency of anaphase bridges in mitotic cells of *nop2a-3* pistils. No anaphase bridges were observed in *nop2a-3* cells, indicating that chromosome ends remain protected and thus, *NOP2A* is not involved in chromosome protection *per se* (Table 3-1). In a parallel approach, the more sensitive telomere fusion PCR assay using primers directed outward from the unique subtelomeric sequences found on Arabidopsis chromosomes[415] also did not detect fusion PCR products, reflecting the absence of covalent telomere-to-telomere chromosome fusions in *NOP2A* mutants (Figure 3-5). Taken together, these data argue that, like telomerase subunit genes *TERT* and *POT1a*, Arabidopsis *NOP2a* is dispensable for chromosome end protection.

Plant line	Number of anaphases analyzed	Number of anaphase bridges	% of anaphases bridges
Col - 0	92	0	0
Oli2-3 G3*	95	0	0
Oli5 G3*	58	3	5
Oli7 G4*	76	3	4

Table 3-1. Anaphase bridge analysis of *oli2*, *oli5*, and *oli7* mutants.

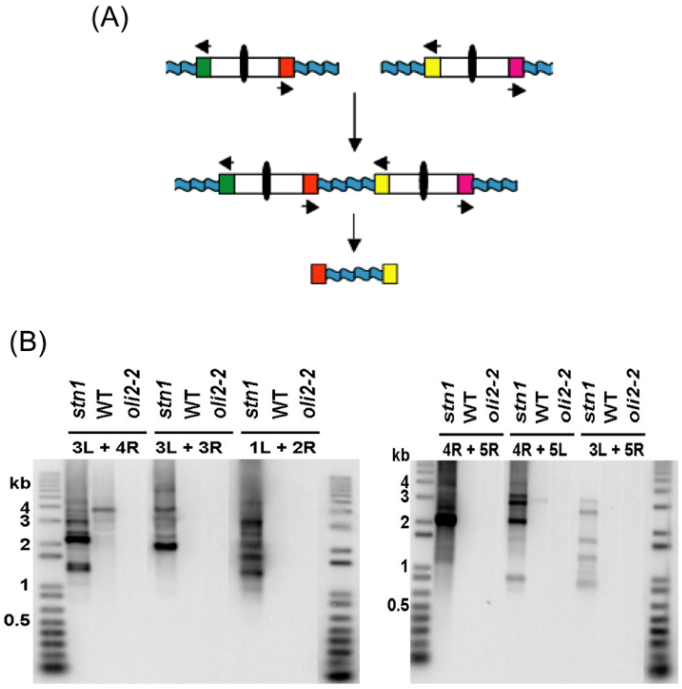


Figure 3-5. PCR amplification of chromosome fusion junctions. (A) Diagram of the PCR strategy to amplify chromosome fusion junctions[415]. Arrows denote unique subtelomeric primers directed toward the chromosome terminus. PCR amplification occurs only when two subtelomeric regions are joined end-to-end (oval: centromere; wavy line: telomere). (B) Southern blot analysis of fusion PCR products using a telomeric probe. Fusion PCR results for 6 different subtelomeric primer combinations are shown using DNA from stn1 mutants (positive control) with well-characterized genome instability and massive chromosome fusions[238], wild-type template DNA (negative control) and DNA from *oli2-2* plants. Results for a single plant are shown in each lane.

Arabidopsis OLI genes connect telomere length homeostasis and ribosome biogenesis.

Arabidopsis NOP2A/OLI2 was originally identified in a forward genetic screen for regulators of cell proliferation[413]. Several other Arabidopsis genes were identified in the same screen, including partially redundant OLI5 (At3g25520 locus) and OLI7 (At5g39740 locus), which encode ribosomal proteins RPL5A and RPL5B, respectively. In humans and other eukaryotes, the highly conserved RPL5 protein binds the 5S rRNA to form a ribonucleoprotein particle that participates in the formation of the large ribosomal subunit in the nucleolus and is subsequently exported to the cytoplasm[417]. Since all three OLI genes were originally identified in the same genetic screen, localize to the nucleolus and display similar cell proliferation defects, we tested if T-DNA insertional mutants of *AtOLI5* and *AtOLI7* genes will also display telomere length defects. Indeed, both *oli5* and *oli7* homozygous T-DNA mutants displayed shorter telomeres (Figure 3-6A &B) compared to Col-0 wild type plants. Furthermore, mean telomere length in homozygous *oli5-2* mutants remained shorter than in the wild type across at least 3 generations that they were propagated in the lab and was clearly stabilized at exactly the same level (~1.8-2.0 kb) as in all *nop2a/oli2* mutants (Figure 3-6A). Homozygous *oli5-3*$^{-/-}$ mutants segregated from a heterozygous *oli5-3*$^{+/-}$ parent also displayed progressive telomere shortening from one generation to another (Figure 3-6C) until stabilization at ~1.7-1.8 kb. The *oli7-2* T-DNA mutants also displayed telomere shortening, though its degree was smaller (Figure 3-6B), possibly reflecting the nature of *oli7-2* mutation, which is likely not a null mutant, as its corresponding T-DNA was inserted at the end of the OLI7 gene.

Altogether, these genetic data clearly indicate that all three rRNA-interacting OLI genes are important for telomere length control, thus strongly supporting the notion that

the entire network of evolutionarily conserved ribosome biogenesis genes encoding rRNA-binding proteins participates in setting the threshold for appropriate species-specific telomere length set point.

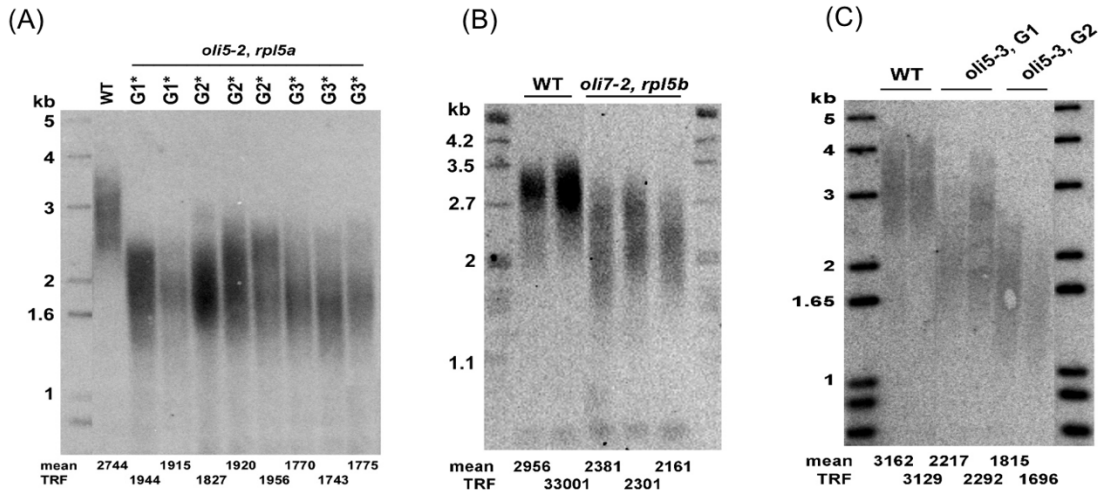


Figure 3-6. Genetic analysis of OLI2 interaction network identifies several proteins with nucleolar function and implicates rRNA biogenesis and ribosome assembly as a conserved mechanism influencing telomere length set point. Arabidopsis OLI5 (A & C) and OLI7 (B) T-DNA knockout plants harbor stably shorter telomere phenotypes with a new length set point. G1, generation 1 mutants; G2, generation 2 mutants.

Discussion

There is considerable variation in average telomere length that exist within mouse, yeast and plant populations and between human individuals[106,401,402]. However, only a handful of factors have been shown to cause population-specific telomere length diversity

between or within species. The model plant, *Arabidopsis thaliana*, offers an ideal system to advance these goals due to its small genome size, massive genetic resources and well documented heritable telomere length variation[106]. Here, we show how MAGIC lines represent a very powerful new tool and great genomic resource for discovering novel genetic factors controlling natural telomere length variation.

NOP2 may connect telomere biology, cell cycle and cancer progression

The NOP2 gene family is highly conserved and is present in yeast, plant and mammalian genomes. Human and mouse NOP2 proteins exhibit a striking 96% identity at the amino acid level, while AtNOP2 and human NOP2 proteins are 61% identical and 73% similar. The high degree of evolutionary conservation suggests a critical role in all eukaryotes. Human NOP2 is a proliferation-associated antigen highly expressed in a wide range of malignant tumors[418]. Similarly, mouse NOP2 has been characterized as a cancer-causing gene that drives cancer progression[412] and is implicated in the development of colorectal, gastric and liver cancers[419–421]. More recently, human NOP2 was purified as telomerase accessory protein and, through contacts with hTERT, NOP2 was shown to stimulate transcription of cyclin D1[407]. Similarly, AtNOP2 is involved in determination of plant organ size and the regulation of cell proliferation[413].

Our findings in *A. thaliana* directly implicate AtNOP2 in telomere length maintenance, and thus lend strong support to the hypothesis that NOP2 may connect telomere biology, cell cycle and cancer progression. Overall, our data in Arabidopsis raise important possibilities that human NOP2 may promote oncotransformation through interactions with telomere maintenance pathways.

nop2a contributes to a new telomere length set point

Despite the strong connection between NOP2 and cellular proliferation and tumorigenesis in eukaryotes, surprisingly little is known about the molecular interactions of NOP2 and its mechanism of action. Interestingly, although telomere length shortens to a similar degree in *nop2a-2* and *tert* mutants, their phenotypes are not identical. Inactivation of both *tert* and *pot1a* (a telomerase accessory protein) leads to a unique banding pattern of telomere profile on TRF gels[172,363], with each band representing an individual telomere. In all three of the *nop2a* mutants, however, telomere profile on TRF gels remains smeary. One interpretation of this observation (consistent with Q-TRAP results *in vitro*) is that telomerase is still partially active in these mutants and can thus generate the observed heterogeneity in individual telomere lengths. Nevertheless, telomere shortening observed in the first several generations of *nop2a-4* mutants indicates that telomerase, though active, is unable to fully compensate for overall telomere loss until a new homeostasis at the shorter set point is established. Whether this represents a defect in telomerase recruitment to the chromosome end or its elevated dissociation rate from the telomeric DNA will need to be determined in the future.

Connecting ribosome biogenesis and telomere length

In Arabidopsis, OLI5 (RPL5A) and OLI7 (RPL5B) proteins are extremely conserved at the amino acid level (only five amino acid substitutions). In addition, both genes have a very similar expression pattern and are considered duplicate genes with redundant function[422]. Nevertheless, all 3 identified OLI genes function as rRNA/ribosome biogenesis factors. The only known example of RNA modifying telomere protein that has

previously been shown is the dyskerin protein, but it was thought to be more specific for the telomerase RNA biogenesis rather than for general ribosome biogenesis. Therefore, for the first time we are connecting ribosome biogenesis as a major determinant of species-specific telomere length.

Functional conservation of telomere regulation

In summary, we highlight the functional conservation of telomerase subunits and functions between plant and animal organisms. In addition, we have identified another conserved protein, NOP2A, that plays a role in telomere length regulation. Although telomere length in human and yeast NOP2 mutants have never been previously reported, our conclusions in Arabidopsis correlate well with available data from other species. NOP2A mutants in Arabidopsis are viable and fertile, but this is not the case for human and yeast orthologues, further providing *A. thaliana* as a good model to further study the connection of NOP2A in species-specific telomere length determination.

CHAPTER IV

MOLECULAR CHARACTERIZATION OF NOP2A LOCUS IN *A. THALIANA* ACCESSIONS

Summary

The length of the telomeric DNA is species-specific and varies between 5-15 kb at birth in humans, 2-9 kb in the model plant *Arabidopsis thaliana*, 40-160 kb in *Nicotiana tabacum*, 250-300 bp in yeast and 50-100 kb in mouse[72–74,106–108]. Most association studies between telomere length and SNPs in known telomere maintenance genes have so far been inconclusive. To date, common genetic variants at only seven loci have shown a replicated association, and each of these variants alone explains less than 1% of telomere length polymorphism[423–426]. Arabidopsis natural variation studies, fueled by the availability of hundreds of accessions (natural genetically diverse populations) collected from geographically distinct locations offers a unique opportunity to analyze natural intra-species variation in telomere length in a multicellular eukaryote[427,428]. Our initial analysis has resulted in the identification of *NOP2A* gene as a major regulator of plant telomere length. The rare haplotype behind *NOP2A* gene is donated by the Sf-2 accessions. In this study, we investigated whether *NOP2A* is the causal gene, and whether allelic variations in *NOP2A* locus are responsible for the remarkable increase in telomere length observed in Sf-2.

Introduction

Telomere length maintenance is a fundamental cellular process, which is remarkably conserved across eukaryotic evolution. However, the length of the telomeric DNA is species-specific and varies between 5-15 kb at birth in humans, 2-9 kb in the model plant *Arabidopsis thaliana*, 40-160 kb in *Nicotiana tabacum*, 250-300 bp in yeast and 50-100 kb in mouse[72–74,106–108].

Earlier family and twin studies indicated that up to 80% of human telomere length variation between individuals measured at birth is determined genetically[429,430], but the nature of the causal genetic loci remains elusive. Difficulty in identifying these factors can be in part attributed to the complex quantitative genetic nature of telomere length set point phenotype[109,110].

Most association studies between telomere length and SNPs in known telomere maintenance genes have so far been inconclusive. To date, common genetic variants at only seven loci have shown a replicated association with mean telomere length in genome-wide association studies[423–426], and each of these variants alone explains less than 1% of telomere length polymorphism. In addition to studies in humans, research in model eukaryotes continues to yield valuable information on genes involved in telomere length control. For example, a genome-wide screen of F1 cross of two different mouse strains identified a helicase gene with an essential role in telomere maintenance[431,432]. A study in yeast using a cross between a vineyard and a laboratory strains identified two loci responsible for telomere length variation, including a Ubiquitin ligase complex BUL2, which controls not only telomere length but also cellular life span[109,433].

Because of its superior biological and genomic resources, Arabidopsis offers a unique opportunity for the analysis of natural variation in telomere length. There is now an explosion of Arabidopsis natural variation studies, fueled by the availability of hundreds of accessions (natural genetically diverse populations) collected from geographically distinct locations[427,428]. The availability of these vast resources for genomic research offers a unique opportunity to analyze natural intra-species variation in telomere length in a multicellular eukaryote. Because Arabidopsis is self-pollinating, these accessions are inbred and mostly homozygous, greatly facilitating genetic analysis and mapping. Significant natural variation in Arabidopsis was reported for every phenotypic trait investigated[428], including remarkable variation in telomere length set point[106,434]. Since its genetics lacks most of the technical and biological limitations associated with the human system, Arabidopsis is the model of choice for the identification of evolutionarily conserved genetic factors establishing population-specific telomere length polymorphism.

Our initial analysis of genetic architecture of telomere length control performed in *Arabidopsis thaliana* has resulted in the identification of *NOP2A* gene as a major regulator of plant telomere length. NOP2A homologue in humans is NOP2 (also known as NOL1 and p120, UniProtKB-P46087), which was originally discovered as a tumor-associated marker whose abundance in many cancer types directly correlates with tumor progression and poor patient prognosis[435,436]. Expression of *NOP2* mRNA and protein parallels the rate of cell proliferation[437] with *NOP2* mRNA present in growing fibroblasts, but absent with quiescence[438]. NOP2 protein peaks in late G1-and early S-phase of synchronized fibroblasts[439] and CHO-Ki cells[404]. Strikingly, NOP2 mRNA levels in many tumor cell types are 15-60 fold higher than in normal cells[438]. Insertional mutagenesis experiments in

mice provided additional support for NOP2 as a cancer-causing gene that drives cancer progression[412]. Specifically, mouse NOP2 has been implicated in the development of colorectal, gastric an d liver cancers[419–421]. Thus, our preliminary experiments in the model plant *Arabidopsis thaliana* have already resulted in the identification of an important evolutionarily conserved gene involved in telomere biology.

The most intriguing aspect of our preliminary mapping result is that the rare haplotype behind chromosome 5 QTL, which we tentatively mapped to *NOP2A* gene, is apparently donated by the Sf-2 accession. Since most genetic resources in *A. thaliana* are developed for the reference Col-0 accession, very little information is available on the specific molecular differences between *NOP2A* loci in other Arabidopsis accessions, including Sf-2.

In this study, we investigated whether *NOP2A* is the causal gene, and whether allelic variations in *NOP2A* locus are responsible for the remarkable increase in telomere length observed in Sf-2. Telomere length and expression differences of *NOP2A* alleles in all 19 MAGIC parental accessions show no obvious correlations between NOP2A gene, suggesting that differences in NOP2A gene expression may not be causal for the chromosome 5 QTL. Analysis of amino acid sequence shows six accessions that harbor a total of 8 unique amino acid substitutions that are located outside of the highly conserved rRNA methyl-transferase domain. Most of the amino acid substitutions are seen in both Sf-2 and Bur-0, accessions that harbor long telomeres. These correlations suggest that one or more Sf-2 and Bur-0 specific amino acid changes may be causal for the longer telomere phenotype observed in MAGIC plants with Sf-2 specific haplotype in chromosome 5 QTL region. Telomeres of F2 generation wild type Col-0 and Sf-2 cross plants, genotyped for

the accession-specific *NOP2A* locus, shows range of telomere length. This indicats that allelic differences in the *NOP2A* locus may play a role in the establishment of accession-specific telomere length in Arabidopsis. Study of *A. thaliana* NOP2A among 19 MAGIC parental accessions revealed importance of NOP2A in determining telomere length. This study shows the presence of potential amino acids that may play a regulatory role in telomere length determination, as well as the presence of other genes that play a role in telomere length determination.

Materials and methods

Construct generation

NOP2A locus with its native promoter (1500 bp upstream of ATG and 800 bp down stream of TAA) was cloned into, pCBK05 destination vector using SbfI and BamHI (pCBK05:NOP2A).

Plant line and Plant transfection

SALK_129648 line was used as oli2-2.

Small scale Agrobacterium culture was grown for 24hours at 30C°. 500mL LB was then inoculated with 0.5mL of small scale culture, and was grown at 28 C° until the OD_{595} reached 1.5. Agrobacterium culture was spun down for 10 minutes at 4,000 rpm, and was re-suspended in 15% sucrose 0.2% Silwet66 solution. Plants with flowers were dipped in Agrobacterium solution for one minute, and kept in dark overnight. T1 seeds were selected on BASTA for pCBK05 vector, and Kanamycin for pDs vector on ½ MS media.

Telomere length analysis (TRF)

Bulk telomere length was analyzed using Terminal Restriction Fragment (TRF) analysis. Genomic DNA was digest overnight with MseI enzyme at 37C°. Digested DNA was then ran on a 0.8% agarose gel overnight at 55V. DNA on the gel was denatured and transferred to a membrane. The membrane was probed with 5'-radio-labeled oligo $(TTTAGGG)_4$, and exposed with phosphor-screen.

Results

Difference in NOP2A between A. thaliana accessions.

Since telomere length is known to vary between Arabidopsis accessions[106], we first measured telomere length in all 19 parental accessions used to generate the MAGIC RIL population (Figure 4-1A). As expected, these accessions showed substantial variation in telomere length, with the shortest mean telomere length (2,531 bp) present in the Kn-0 accession, and the longest (4,947 bp) – in the Sf-2 accession (Figure 4-1B). The reference Col-0 accession harbored second shortest mean telomeres at 2,727 bp. We next asked if levels of NOP2A gene expression also differ among different parental accessions. Significant expression differences for NOP2A gene were observed by Q-PCR in Arabidopsis accessions, with the highest levels detected in Ct-1 accession and the lowest in Ws-0. (Figure 4-1B). Overall, no obvious correlations between NOP2A gene expression levels and mean telomere length were detected, suggesting that differences in NOP2A gene expression may not be causal for the chromosome 5 QTL.

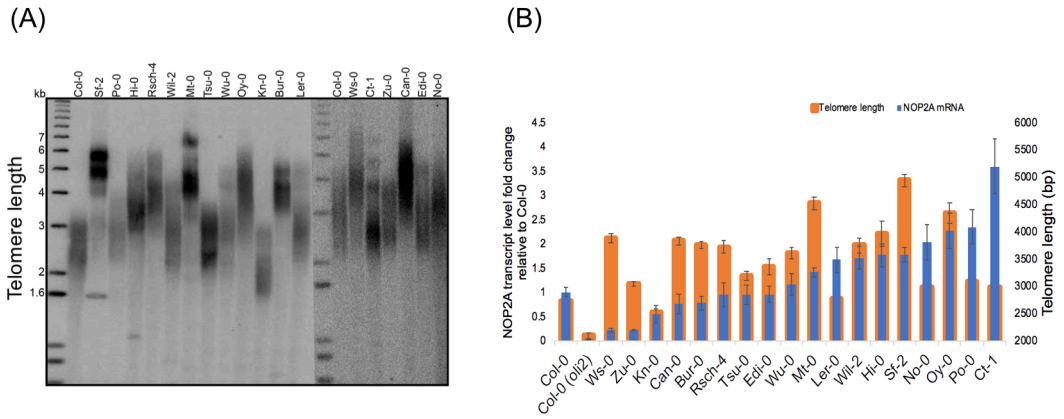
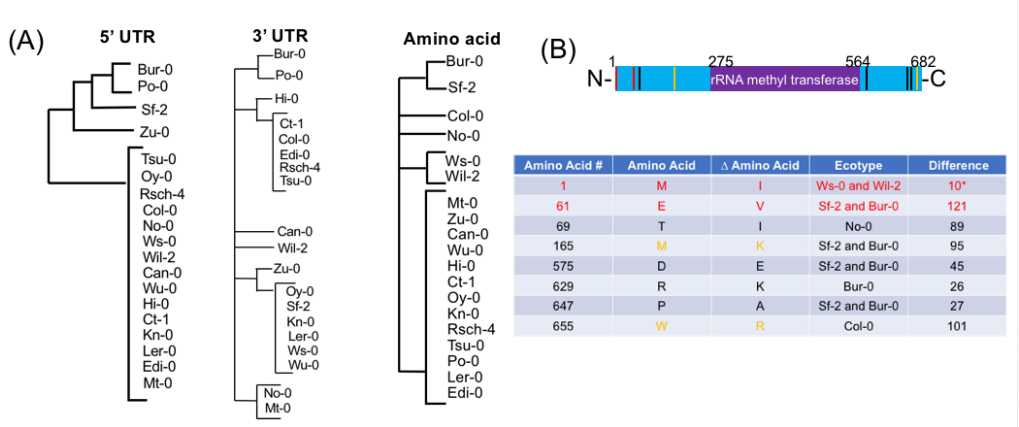


Figure 4-1. Telomere length and NOP2A expression level in different A. thaliana accessions. (A) Bulk telomere length (TRF) of 19 MAGIC parental accessions. (B) NOP2A transcript level and mean telomere length of 19 MAGIC parental accessions. NOP2A transcript level was normalized to wild-type Col-0 level.

The observed differences in transcript abundance (Figure 4-1B) can arise from variations in either transcription efficiency or mRNA stability, which, in turn, may be correlated with sequence polymorphism in 5' and 3' UTR regions of the *NOP2A* gene. We thus looked at the nucleotide sequence variation in the *NOP2A* locus among Arabidopsis accessions, focusing initially on the 5' (1000 bp upstream of the translation start codon ATG) and 3' (1,500 bp downstream of the stop codon) sequences of *NOP2A* alleles in all 19 parental accessions of the MAGIC population. While 15 out of 19 accessions have identical 5' (upstream) sequence, 4 accessions (Bur-0, Po-0, Sf-2, and Zu-0) harbor nucleotide polymorphism within this region (Figure 4-2A). Substantially more sequence polymorphism is present in the 3' (downstream) region of the NOP2A gene, and at least several phylogenetically close groups can be identified among the alleles. Despite this observation, no obvious correlation between *accession*-specific telomere length, NOP2A transcript level and sequence variation in the 5' and 3' regions of the gene can be

identified. One interpretation of this result is that sequence variation in these regions does not affect the steady state transcript level of NOP2A gene. However, the influence of sequence variation upstream and downstream of the NOP2A coding region still cannot be completely ruled out, as transcription rates and mRNA stability can also be modulated by other cis- or trans-acting *accession*-specific factors. Furthermore, longer stretches of the nucleotide sequences upstream and downstream of the NOP2A coding region may need to be analyzed in the future to test their effect on NOP2A gene expression.

We next asked if NOP2A alleles harbor accession-specific amino acid changes. While most NOP2A proteins have identical amino acid sequence, six accessions harbor a total of 8 unique amino acid substitutions (Figure 4-2A).



Amino Acid #	Amino Acid	Δ Amino Acid	Ecotype	Difference
1	M	I	Ws-0 and Wil-2	10*
61	E	V	Sf-2 and Bur-0	121
69	T	I	No-0	89
165	M	K	Sf-2 and Bur-0	95
575	D	E	Sf-2 and Bur-0	45
629	R	K	Bur-0	26
647	P	A	Sf-2 and Bur-0	27
655	W	R	Col-0	101

Figure 4-2. DNA and Amino acid sequence comparison in different A. thaliana accessions. (A) Phylogenetic tree of NOP2A sequence between 19 MAGIC parental accessions. 5' UTR (1000 kb upstream of ATG) (left), 3' UTR (1,500 kb downstream of stop codon) (middle), and amino acid (right). (B) Table of difference in amino acid. Amino acid #: amino acid position. Amino acid: amino acid majority of accessions have. Δ amino acid: difference amino acid the unique accession has. Accession: accession with the unique amino acid. Difference: Amino acid difference calculated using composition, polarity, and molecular volume[440]. Red amino acids denotes major difference, and orange denotes change that can affect post translational modification. Diagram above the table shows location of amino acids that are different. Purple box denotes rRNA methyl transferase domain.

Interestingly, all substitutions are located outside of the highly conserved rRNA methyl-transferase domain (Figure 4-2B, top). For example, a unique T69I substitution is present in No-0 accession (Figure 4-2B, top). Interestingly, the reference Col-0 accession harbors one amino acid change in position 655 (W to R substitution) that is not present in other accessions. While the exact impact of this W655R substitution is unclear, such change can potentially affect NOP2A protein ubiquitination profile.

A very unusual NOP2A substitution is present in Ws-0 and Wil-2 accessions. An AT**G** to AT**C** codon change abolishes proper NOP2A protein translation start codon. Interestingly, another in-frame ATG codon 90 nucleotides downstream is present in all NOP2A sequences. Whether this alternative start codon is indeed utilized in Ws-0 and Wil-2 accessions is currently unclear, but this scenario can be tested in genetic complementation assays in the future (see below). Indeed, both Ws-0 and Wil-2 accessions belong to the group of accessions with relatively longer telomeres (upper half of the spectrum), and the influence of such substantial coding region change will clearly need to be tested *in vivo*.

The most remarkable finding from the analysis of the putative NOP2A protein sequences in the parental MAGIC accessions is the observation that accessions Sf-2 and Bur-0 have the most amino acid changes (4 and 5, respectively) in comparison to all other accessions. Although in two cases a very similar amino acid is present in the protein sequence (R629K change in the Bur-0 allele and D575E change in both Bur-0 and Sf-2 alleles), the effects of the other 3 amino acid changes may be more substantial. Specifically, the M165K change can potentially allow ubiquitination of Sf-2 and Bur-0 sequences, while P647A can potentially affect protein secondary structure if this amino

acid is located within a α-helix. In addition, both E61V and M165K amino acid substitutions introduce changes in protein hydrophobicity by adding or removing a charged side chain. Interestingly, in addition to Sf-2 accession harboring the longest telomeres among all 19 MAGIC parent, telomeres in Bur-0 accession are also the seventh longest in the spectrum. These correlations suggest that one or more Sf-2 and Bur-0 specific amino acid changes, either alone or in combination with nucleotide substitutions in other regions of the *NOP2A* gene, may indeed be causal for the longer telomere phenotype observed in MAGIC plants with Sf-2 specific haplotype in chromosome 5 QTL region.

Genetic complementation assay to compare complementation efficiency of NOP2A alleles from Col-0 and Sf-2 accessions

As discussed earlier, all recombinant inbred MAGIC plants that inherited a rare Sf-2 specific allele inside the chromosome 5 QTL region display up to 50% longer telomeres than the rest of the MAGIC population. To test our working hypothesis that polymorphism in the *NOP2A* locus is causal for this phenotype, we performed a series of transgenic experiments aiming to quantitatively and qualitatively complement *NOP2A* deficiency by transforming *NOP2A* knockout plants (in Col-0 background) with *NOP2A* alleles from Sf-2 (longer telomeres) and Col-0 (shorter telomeres) accessions under the control of their natural promoters. We expected that the presence of a functional *NOP2A* allele from either accession will complement telomere deficiency observed in the *NOP2A* knockout. However, if our hypothesis is correct, a difference in the degree of complementation between the two alleles may be observed. Specifically, plants with Sf-2 *NOP2A* allele would have longer telomeres compared to the plants with *NOP2A* allele from Col-0.

We transformed *nop2a-2* knockout plants with the wild-type *AtNOP2A* alleles from the Col-0 and Sf-2 genetic backgrounds harboring the entire coding regions and containing 1 kb each of sequences upstream of the translation start codon and downstream of the stop codon. Transformants were selected on the medium containing herbicide BASTA and analyzed to determine their telomere length. As expected, transformation with Col-0 allele of *NOP2A* was able to rescue shorter telomere phenotype of the *oli2-2* mutant, increasing telomere length back to the wild type Col-0 level (Figure 4-3B). Similarly, transformants harboring the Sf-2 allele of *NOP2A* also harbored longer telomeres. However, the degree of telomere lengthening in transformants with Sf-2 specific *NOP2A* allele was the same as in transformants with Col-0 specific allele (Figure 4-3A). To test if the transformants with Sf-2 specific *NOP2A* allele required several additional plant generations to reach telomere length above the Col-0 level, primary T1 transformants were self-pollinated to generate T2 and then T3 generation plants. Both T2 and T3 generation transgenic plants harbored telomeres in the same range as T1 generation or the Col-0 wild type plants, indicating that once the *oli2-2* mutation is complemented with either Col-0 or Sf-2 *NOP2A* alleles, telomere length remains stable in the consecutive plant generations (Figure 4-3C). Overall, these data indicate that with the experimental approach employed here no functional differences between Col-0 and Sf-2 alleles of *NOP2A* gene can be established. Although further experiments will be necessary to confirm or reject the hypothesis that variations in *NOP2A* alleles are causal for the chromosome 5 QTL, our data provide direct evidence that both *NOP2A* alleles are functional, and further confirm that Arabidopsis *NOP2A* is a novel major player in the regulation of telomere length in *A. thaliana*.

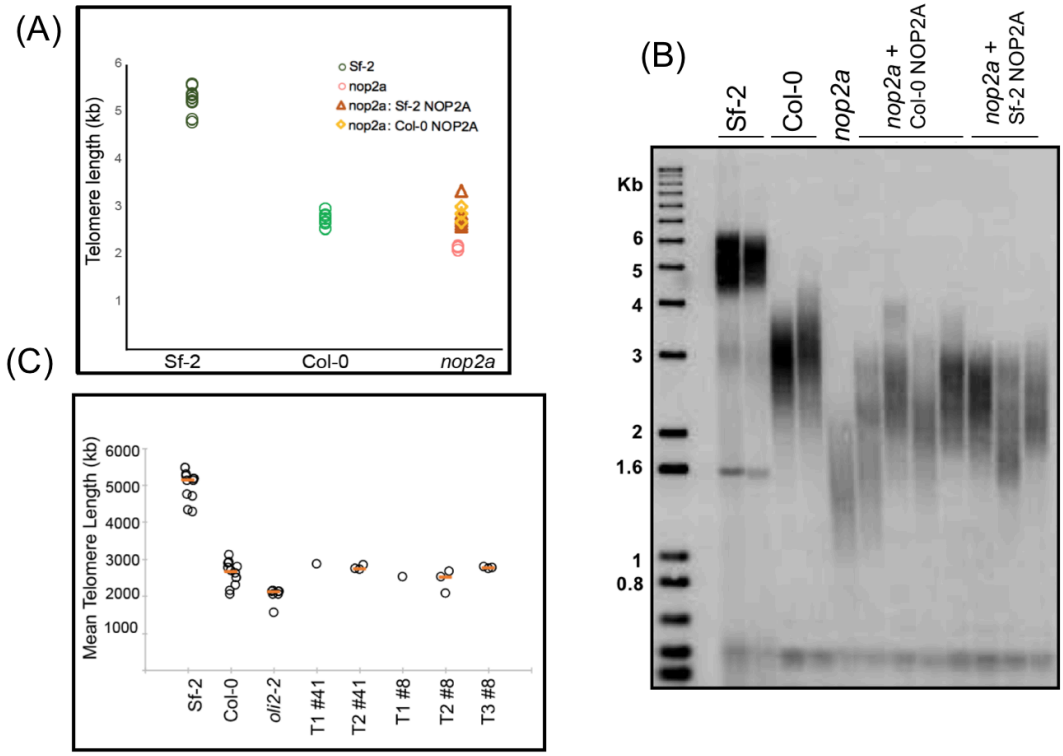


Figure 4-3. nop2a genetic complementation analysis. (A) Telomere length of Sf-2 (Dark green), Col-0 (Green), and Col-0 nop2a mutant (Pink). Col-0 *nop2a* mutant was complement by either Col-0 NOP2A locus (orange diamond) or Sf-2 NOP2A locus (brown triangle). (B) Parent progeny telomere length of Col-0 *nop2a* mutant complemented by Sf-2 NOP2A (#41 and #8).

Alternative strategies to test for NOP2A causality

Although we were unable to verify the causality of *NOP2A* locus for chromosome 5 QTL with the functional complementation experiments presented in Figure 4-3, establishing causality can often be difficult and complicated by the presence of multiple significant SNPs and epistasis, or redundancy of *NOP2A* homeologs. Alternative strategies may include inactivating *NOP2A* gene in Sf-2 and other Arabidopsis accessions with the

Crispr-Cas9 strategy, or crossing different accessions and monitoring telomere length in the segregating progeny with alternative *NOP2A* alleles.

To initiate the latter approach, wild type Col-0 and Sf-2 plants were crossed, genotyped for the accession-specific *NOP2A* locus, and the bulk telomere length of F2 plants was analyzed (Figure 4-4). Plants that have acquired both Col-0 specific *NOP2A* alleles harbored telomere length of 2-5 kb, which is typical for Col-0 accession (Figure 4-4, Col-0/Col-0 plants). Similarly, heterozygous plants with one Sf-2 specific and one Col-0 specific alleles harbored telomeres in the same range, 2-5 kb.

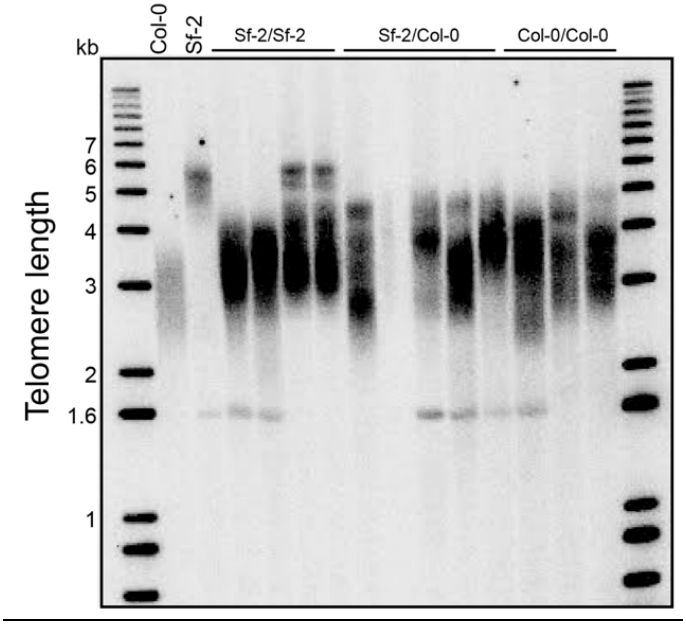


Figure 4-4. Telomere analysis of F2 Sf-2 x Col-0 line. Wild type Col-0 and Sf-2 were crossed, and NOP2A locus of F2 generation was genotype for Col-0 or Sf-2. F2 plants were segregated into plants with both NOP2A locus from Sf-2 (Sf-2/Sf-2), both NOP2A locus from Col-0 (Col-0/Col-0), and heterozygote for Col-0 and Sf-2 (Sf-2/Col-0).

However, plants that acquired both *NOP2A* alleles from Sf-2 accessions displayed two distinct profiles: several plants harbored short telomeres in the Col-0 range (2-4 kb), while other plants harbored a much wider range between 2 and 6 kb, with the longer telomere tracts approaching the telomere length maximum seen in the Sf-2 control (Figure 4-4). Although additional experiments are clearly necessary, these data indicate that allelic differences in the *NOP2A* locus may indeed play a role in the establishment of accession-specific telomere length in Arabidopsis. Further experiments with the progeny of F2 plants analyzed in Figure 4-4 may shed additional light on the importance of accession-specific *NOP2A* alleles in telomere length determination.

Discussion

There are set telomere range Between species, and there are also significant variations in average telomere length within species. While it is difficult to identify gene(s) that have phenotype of a quantitative trait, our lab has previously identified NOP2A. NOP2A has been shown to play a role in telomere length control in Col-0, but how much NOP2A contributes to *A. thaliana* telomere length has not been studied in detail. Here we have investigated whether NOP2A is the major gene responsible for Sf-2's long telomeres, and how NOP2A differ between 19 MAGIC parental accessions. Through genetic complementation, we have shown that NOP2A is necessary for Col-0 to have telomere length of 2-4kb, but it is not sufficient for Col-0 to have long telomeres (5-6 kb). We have also shown through cross-accession analysis, that there may be another gene(s) that play(s) a role in telomere length control together with NOP2A. Finally, we show difference in

transcript level as well as amino acid sequence of NOP2A within 19 MAGIC parental accessions.

Complementation analysis of Col-0 *nop2a* with Sf-2 NOP2A locus does not give Col-0 a long telomeres (5-6kb). In addition, long telomeres (5-6kb) are only seen in F2 Sf-2 x Col-0 lines that have NOP2A loci from Sf-2, but not all plants that have NOP2A loci from Sf-2 have long telomeres. These data indicate that NOP2A is necessary, and it is one of the major gene that plays a role in determining telomere length in *A. thaliana*. However, it is not sufficient to have long telomeres (5-6 kb). This is not surprising because telomere length phenotype is a quantitative trait[109,110]. Presence of plants that have two telomere length in both Sf-2/Sf-2 and Col-0/Col-0 indicates the existence of another genes that function in telomer length determination together and independently with NOP2A, respectively. It is possible, that one of these genes lie within the QTL peak from chromosome 1. Further analysis of these plants at this QTL locus may allow identification of other genes that are involved in telomere length determination.

While no correlations are seen between NOP2A transcript level and telomere length, difference in NOP2A transcript levels are seen between all 19 MAGIC parental accessions. There are also difference in DNA and amino acid Sequence. Although it is still unclear how NOP2A transcript levels are affected in these accessions, difference in transcript level suggests that amount of NOP2A affects telomere length. This hypothesis can be supported by the result from Sf-2 x Col-0 analysis, where Sf-2/Col-0 does not have long telomeres. Difference in amino acid sequence can lead to change in protein regulation and stability through post-translational modification. Col-0, Sf-2, and Bur-0 have single amino acid substitution that may affect ubiquitination. However, the major difference unique to Sf-2

and Bur-0 is the glutamine to valine substation. This substitution can potentially prevent phosphorylation of Sf-2 and Bur-0 NOP2A, if this residue is phosphorylated in other accessions.

In summary, study of *A. thaliana* NOP2A among 19 MAGIC parental accessions revealed importance of NOP2A in determining telomere length. Furthermore, the presence of other genes that plays telomere length determination, together and independent of NOP2A, opens a door to allow further study of how telomere length set point are determined. In addition, difference in amino acid sequence between these accessions will allow further biochemical studies to tease apart the mechanism of how NOP2A is involved in telomere length determination.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

Telomeres function to solve the end-protection and end-replication problems. In most species, the end-protection problem is addressed with the help of the telomere capping protein complexes shelterin and CST, while the end-replication problem is resolved by telomerase reverse transcriptase. Moreover, through complex interactions between telomere capping proteins and telomerase accessory proteins, proper regulation of telomerase and telomere length homeostasis is achieved. In this dissertation, I characterized two novel telomere-related genes: *POT1c,* one of three paralogs of POT1 previously implicated in telomerase regulation and the DNA damage response, and *NOP2A* gene identified in a screen for quantitative trait loci that impact telomere length set point.

It is common for plants to have multiple gene paralogs (gene families). Among the many examples in *A. thaliana* are genes encoding single-strand DNA binding proteins RPA[441–443], the RNA methyltransferase NOP2[444], the TRFL double-strand telomere binding proteins and POT1[367,445].

POT1 is one of the most highly conserved telomere proteins, and its duplication is rare[254,271,364]. AtPOT1a has the ancestral function, and is important for telomerase activity by stimulating telomere repeat addition processivity[172,178]. Although POT1a does not show a telomere capping phenotype, POT1a has maintained a critical amino acid in OB1 (Phe65) that is important for telomeric DNA recognition[305,307]. In contrast, POT1b has diverged.

POT1b has lost the critical amino acid that allows telomeric DNA binding, and it does not associate with the canonical telomerase RNP or regulate telomerase activity. Instead, POT1b has acquired a new function in plant development. Thus, the duplicated POT1 genes in *A. thaliana* do fully embody the conserved functions of the single-copy POT1 genes from human or fission yeast, but instead provide examples of sub-functionalization and neo-functionalization[308,309,311–316]. Intriguingly, *A. thaliana* carries a third POT1 locus, POT1c, which is unique to *A. thaliana*. In Chapter II, I examined the fate of this new duplicated gene by analyzing its expression, function and evolution.

The second major focus of this dissertation is the characterization of novel gene that impact telomere length set point in *A. thaliana*. Species-specific telomere length set point is a quantitative trait determined by many genes. Yet, only a handful of factors that play a role in determining population-specific telomere length between or within species has been characterized.

POT1c is a non-functional gene

The POT1 protein is highly conserved across eukaryotes and is essential for telomere maintenance[254,271,364]. In some organisms, the POT1 protein family has undergone an expansion, and in each case there is evidence for functional divergence in the duplicated copies. For example, first OB-fold of the mouse POT1a is a negative regulator of telomere length and the C-terminal domain of the mouse POT1b is required to control the resection of the 5'-end of the chromosome [309,310,308,310,311]. These findings indicate the mouse POT1 genes have sub-functionalized[309,311]. In the ciliate Tetrahymena, gene divergence is more extreme. Tetrahymena Pot1is important for telomere length regulation, prevention of DNA

damage response, and telomere end protection[315]. In contrast, Tetrahymena Pot2 plays a role in sexual reproduction, as well as recruitment of telomerase and/or endonuclease to micronuclear chromosome breakage sites for DNA cleavage[316].
In this case, the ciliate POT1 proteins show evidence of neo-functionalization[315,316].

Similar to Tetrahymena, *A. thaliana* POT1b exhibits evidence of neo-functionalization. Unlike AtPOT1a, AtPOT1b accumulates in the cytoplasm instead of the nucleus. In addition, preliminary data indicate that POT1b functions in modulating the DDR and in promoting early plant development (C. Castillo-Gonzales, B. Barbero, X. Xie, and D. Shippen, unpublished data).

C. elegans and *A. thaliana* and are the only known species with more than two copies of *POT1*. Three out of the four *C. elegans* POT1 genes appear to have undergone sub-functionalization, while the role of the fourth POT1 is still unknown[312–314]. Notably, the domain structure of the *C. elegans* POT1 genes is distinct from most other POT1 proteins, as the worm POT1-like genes encode only a single OB-fold. The third *A. thaliana POT1* gene, POT1c, is also predicted to harbor only a single OB-fold. In addition, POT1c is a newly duplicated gene derived from the POT1a, and completely unique to *A. thaliana*. This observation raises several important questions. Does POT1c express a sub-function of POT1a? Alternatively, is POT1c like POT1b gaining a novel function? Finally, is POT1c a functional gene at all? In Chapter II, evidence was presented to show that the POT1c gene became non-functionalized, a pseudogene,. Transposon invasion into a previously functional promoter, and accumulation of deletions throughout the gene body, led to POT1c gene silencing.

What does POT1c do?

POT1c arose from a proximal duplication and inversion event of At2g05210 (POT1a) and At2g05220 (S17), which created At2g04395 (POT1c) and At2g04390 (dS17). Previous studies indicated that POT1a positively regulates telomerase through promoting repeat addition processivity[178]. Therefore, the first goal was to test the hypothesis that POT1c retained some or all of the functions of POT1a. Several questions needed to be answered. First, the expression profile of POT1c needed to be determined. POT1a expression is ubiquitous, while the telomerase catalytic protein TERT is expressed almost exclusively in reproductive organs and meristematic tissues[367,446]. However, no POT1c expression was detected in any of the tissues analyzed.

A second goal is to determine if POT1c has a biological function by examining the consequences of a null mutation in the POT1c gene. To this end, a homozygous CRISPR mutation was engineered in the POT1c gene (*pot1c-1*). Molecular and developmental phenotypes were assessed. However, no obvious phenotypes in telomere maintenance, telomerase regulation, plant growth, development or reproduction, or the response to DNA damage were detected. These data indicate that POT1c does not have a major function in *A. thaliana* under normal growth conditions.

Why is POT1c not expressed?

The absence of POT1c transcripts and the lack of phenotype in *pot1c-1* mutants suggest several possibilities. First, the lack of POT1c transcripts is that the gene is actively silenced. Genetic silencing can be active or permanent. In Chapter II, the POT1c locus is shown to be unaffected by the canonical active gene silencing pathways. POT1c is not

expressed in *ddm1* (which relieves DNA methylation) and *dcl2,3,4* mutants (which disrupt the sRNA pathway). Furthermore, DNA methylation status is not changed in *ddm1* and *dcl2,3,4* mutants and no small RNAs are identified within the POT1c locus. Chapter II also shows the absence of a functional promoter for POT1c. Several promoter prediction websites were not able to predict a POT1c promoter with high certainty. In support of this prediction, a GUS reporter expressed under the control of the POT1c promoter region, was not detected in any of the tissues analyzed.

The alternative possibility is that POT1c is a non-functional pseudogene gene. Loss of gene function can arise from accumulation of mutations in the locus or from permanent genetic silencing. This question is answered in Chapter II from the *A. thaliana* accession analysis and the *A. thaliana-A. lyrata* POT1a-POT1c locus analysis, respectively. Conservation of POT1a, POT1c, S17, and dS17 loci within 855 *A. thaliana* accessions revealed that POT1a, S17, and dS17 are well conserved in DNA and amino acid sequence, while POT1c is not. The conservation of dS17 can be attributed to its important function in ribosome biogenesis. What is interesting is the complete lack of conservation at the POT1c locus even though the POT1c and the dS17 duplication event happened together. The huge difference in the level of conservation between POT1c and dS17 shows the different fate of these duplicated genes. Evidence for permanent genetic silencing is seen at the POT1c locus. Cross-species analysis shows that POT1c sustained an ancestral POT1a promoter during the POT1a-S17 local duplication event, which was then disrupted by the insertion of two transposable elements. In addition, the gene body of POT1c accumulated extensive deletions, ranging from a few nucleotides to over 400 nucleotides).

Thus, the lack of a functional promoter and poor conservation at the POT1c locus points to POT1c as a non-functional gene.

Why is POT1c silenced?

As discussed above, duplicate genes may be subjected to several fates. For example, *A. thaliana* POT1b has diverged from POT1a to have a unique function outside of telomere biology, without being silenced. Why then was POT1c silenced, not having the chance to evolve like POT1b? The question that needs to be addressed is whether expression of POT1c causes a deleterious effect on *A. thaliana*. Is there a dosage threshold on the level of the POT1 gene that can be expressed? Or is the protein product from the POT1c locus aberrant?

TERT expression levels are tightly regulated in humans, and mutations that cause mis-regulation of the TERT expression can lead to immortal phenotype[447]. In addition, overexpression of human Est1 and TRF2 lead to a telomere uncapping phenotype[448,449]. These observations suggest that the dosage of telomere-related proteins needs to be controlled, and when this regulation is disrupted unwanted effects arise. On the other hand, Shakirov et al. has shown that overexpression of full-length POT1a or POT1b in *A. thaliana* does not lead to telomere defects[367]. However, overexpression of the OB-fold domains or the C-terminus could have detrimental consequences for telomere maintenance or chromosome end protection. These findings argue while the dosage of full-length POT1 proteins may not be critical, dominant negative effects can occur when the balance of individual domains is altered. A plant line overexpressing POT1c under the promoter of ubiquitin has been generated to test this hypothesis. We expect two possible outcomes

from expressing POT1c in *A. thaliana*, either phenotypes with negative effect or no phenotypic effect.

The POT1c OB-fold is a chimera of POT1a OB1 and OB2. Does POT1c behave like POT1b OB1 overexpression? POT1a OB1 can bind telomeric DNA and is sufficient to stimulate telomerase processivity, but POT1b OB1 does not bind telomeric DNA and somewhat inhibits telomerase processivity *in vitro*[307]. This is due to a change in the amino acid between POT1a (F65) and POT1b (V63), a residue that is important for DNA binding[305,307]. Phenylalanine at this position provides base stacking between the amino acid and the DNA[305]. The amino acid residue at the same location in POT1c is tyrosine (Y65), which is an aromatic amino acid similar to phenylalanine with an additional hydroxy-group. However, a crude *in vitro* binding assay showed no telomeric DNA binding for POT1c. This observation suggests that POT1c could potentially affect POT1a in a manner similar to the POT1b OB1 overexpression line.

Alternatively, the POT1c locus may produce a peptide that does not fold properly. Tertiary structure prediction for the POT1c peptide does not indicate a complete OB-fold, because the C-terminus lacks an alpha-helix to cap the beta-barrel. In addition, POT1a OB1 is a stable protein *in vitro* but POT1c is not, which may be due to the presence of a C-terminal tail. POT1c may not be stable *in vivo* either, as no transgenic protein can be detected in tobacco leaves or in *A. thaliana* protoplasts. In contrast, POT1a has been successfully expressed in both systems.

Evolution of POT1c and POT1a loci

While the evolutionary forces at work on the POT1a and POT1b lineages in the Brassicaceae has been studied[319], the evolutionary path of genetic loci encoding these proteins as well as the newly derived POT1c locus is not clear. There are several unanswered questions: 1) what happened to the loci surrounding POT1a, POT1b, and POT1c after *A. thaliana-A. lyrata* speciation? 2) are loci neighboring POT1a and POT1b conserved, or did these loci evolve? and 3) how did POT1c become a non-functional gene? In Chapter II, evidence was presented for the creation of the POT1c locus, and the divergence of the POT1a locus.

A. thaliana has approximately 187 species-specific duplicate gene pairs, and 60% of these duplicate genes are expressed[392]. Most of these duplicate genes (over 80%) are inserted in the same chromosome of the original gene, including POT1c-dS17 duplicate. Species-specific duplication events have the highest conservation rate, while neo-functionalization and specialization rates are lower, but approximately equivalent to each other. On the other hand, genus-specific duplication events have the highest specialization rate, followed by neofunctionalization and conservation rates. Thus, conservation cases decrease and specialization cases increase over the evolutionary time scale[392]. Tandem duplications contribute to genetic redundancy, while dispersed duplications, such as the POT1b locus, contribute to evolutionary novelty[394]. Proximal duplications, such as POT1c and ds17, are balanced in their contributions to genetic novelty and redundancy[394]. In Chapter II, an *A. thaliana-A. lyrata* cross species analysis at the POT1a-POT1c loci revealed that POT1c originally had an ancestral POT1a promoter, and that this promoter was disrupted by the insertion of two transposable elements. Since *A. thaliana* POT1a also

lacks this ancestral POT1a promoter, the local duplication event at the POT1a locus can be inferred to have occurred soon after *A. thaliana-A. lyrata* speciation. If POT1c-dS17 followed *A. thaliana's* genetic trend, a proximal duplication, POT1c and ds17 would have had an equal chance of contributing to either novelty or redundancy. Indeed, dS17 appears to have maintained its conserved function as a ribosomal protein, but POT1c took the path of non-functionalization. Silencing of the POT1c locus has led to the accumulation of DNA mutations leading to non-conservation between accessions.

What happened to the POT1a locus after speciation?

After the *A. thaliana-A. lyrata* speciation event, the S17 gene was inserted downstream of *A. thaliana* POT1a, and then the POT1a-S17 locus was duplicated. Sometime after the duplication, *A. thaliana* POT1a lost its original promoter and acquired a new one. Why did *A. thaliana* POT1a acquire a new promoter? *A. thaliana* has gone through massive genome rearrangement since speciation[450], with a large number of transposable element insertions[451,452]. There are multiple locations in the *A. thaliana* genome, including the POT1a locus, where a large DNA sequence has been inserted. *A. thaliana* potentially acquired a new POT1a promoter due to the loss of its original promoter after a massive insertion of a DNA sequence. It is interesting to note, however, that the *A. thaliana* POT1b locus as well as its neighboring locus has not been disrupted. Was the acquisition of this new promoter a random event, or was there pressure on the *A. thaliana* POT1a gene that required further regulation? There is a difference between the *A. thaliana* and *A. lyrata* POT1a transcript level, and *A. lyrata* shows higher POT1a RPKM (reads per kilobase of transcript per million mapped reads) level, translating to a 2-fold

higher level of transcripts. The difference in POT1a expression between the two species may be due to the difference in chromosome number, where the number of chromosomes in *A. lyrata* is higher than in *A. thaliana*. Therefore, *A. thaliana* may not require as much active telomerase compared to *A. lyrata*. Another possibility is that in addition to its telomeric function, POT1a may have an additional non-telomeric function that requires more POT1a protein. The first question can potentially be answered by comparing the level of TERT molecules and telomerase activity present in *A. thaliana* and *A. lyrata*. Furthermore, with the growing evidence of telomere proteins having roles outside telomere function[453,454], it will be interesting to investigate whether *A. lyrata* POT1a also has a separate function obtained independently from *A. thaliana* POT1a that does not involve telomeres.

NOP2A plays a role in the telomere length set point

Each species maintains a fixed average telomere length. The set point is established by a balance of mechanisms that promote elongation of short telomeres or shortening of excessively long telomeres[225]. Regulation of telomere length is modulated by trans-acting factors, such as telomerase and homologous recombination machinery. Cis-acting telomere binding proteins also play a critical role in establishing and maintaining length homeostasis. A protein counting model has been proposed to control telomere length where short telomeres are associated with fewer telomere binding proteins. Conversely, long telomeres bind more proteins, which ultimately inhibits telomere extension. Despite these studies of cis-acting telomere binding proteins, relatively little is known about the fundamental mechanisms that control the telomere length set point. This is due to the

nature of phenotypes associated with quantitative traits, such as telomere length [109]. Phenotypes that are associated with quantitative traits can be identified through Quantitative Trait Loci (QTL) mapping, using organisms with the trait of interest and a genetically variable population. Different species have different telomere length set points, and there are significant variations in average telomere length that exist within many different organisms, including *A. thaliana* accessions[106]. Utilizing MAGIC line and QTL, Chapter III shows evidence for the identification of a new gene, NOP2A, that contributes to the telomere length set point in *A. thaliana*.

How is NOP2A involved in telomere biology?

Chapter III shows data for the role of NOP2A, also known as OLI2, in establishing telomere length set point. Unlike *pot1a* or *tert* mutants, which show an ever shorter telomere phenotype, and CST mutants, which show massive telomere degradation from telomere deprotection [172,363,416], the *nop2a* mutant reaches a shorter but stable telomere length. NOP2A/OLI2 is part of a gene family that includes OLI5 and OLI7, and all three genes have been shown to play a role in the induction of cell compensation, a cell proliferation defect in developing leaf primordia that triggers excessive cell expansion[413]. Interestingly, NOP2A/OLI2 is, an rRNA methyltransferase, while the OLI5 and OLI7 genes encode ribosomal L5 proteins. Although all of the OLI pathway mutants show the same telomere length phenotype as the *nop2a* mutant, the other NOP2A paralogues (NOP2B, NOP2C, NSUN5, and NOL1) do not show this phenotype. Thus, the telomere phenotype observed in *nop2a* is somehow related to both cell proliferation and ribosomal biogenesis.

To investigate the molecular basis of short telomeres in *nop2a* mutants, genetic crosses were performed to create *nop2a/pot1a* and *nop2a/tert* double mutants. Neither of these lines show any synergistic phenotypes with respect to telomere length, and instead have a telomere profile identical to *pot1a* and *tert* single mutants. These results indicate that NOP2A functions downstream of the telomerase pathway. Notably, the *nop2a* mutant shows decreased telomerase activity similar to the *pot1a* mutant. However, because the telomere length profile of the *nop2a* is distinct from the pot1a mutant, the data indicate that NOP2A does not impact telomerase in the same way as POT1a. This raises an important question: does NOP2A affect telomerase activity directly or indirectly?

This question can be tested by first examining whether NOP2A associates with telomerase *in vivo*. While there is no available NOP2A antibody for *A. thaliana*, the human NOP2 antibody targets a conserved amino acid sequence that can recognize *A. thaliana* NOP2A and NOP2B. Immunoprecipitation with this antibody did not pull down telomerase, indicating that NOP2A does not associate with active telomerase. It is possible that NOP2B and NOP2A compete for the binding of the NOP2 antibody, diluting the signal for NOP2A association with active telomerase. To counteract this problem, immunoprecipitation of telomerase activity can be tested in the *nop2b* mutant background. Furthermore, association of NOP2A with TERT needs to be analyzed. Human NOP2 associates with TERT protein[407]; therefore, *A. thaliana* NOP2A may also associate with TERT protein even though it is not associated with the active telomerase. In addition, more detailed biochemical analysis of the NOP2A protein may help shed light on how it works in telomere length determination.

Finally, human NOP2 has been reported to associate with TERT, but this association does not affect telomerase activity[407]. Instead the NOP2A-TERT interaction is important for CYCD1 transcription. NOP2-telomerase association in the context of telomere length regulation has not been explored[407], making the discovery of NOP2A function in Arabidopsis telomere length set point a novel finding. Human NOP2 is well studied in regard to cell proliferation and is positively correlated with proliferation rate[403,409–411]. The conservation of function between hNOP2 and AtNOP2A in the cell proliferation pathway argues that hNOP2 may well play a role in the telomere length set point.

NOP2A in other *A. thaliana* accessions

MAGIC progeny that inherited Sf-2 specific haplotype at NOP2A locus showed longer telomere length in the QTL analysis. In Chapter IV, genetic complementation data are presented showing that a Col-0 *nop2a* mutant is complemented with the Sf-2 NOP2A locus. The Sf-2 NOP2A construct was sufficient to restore mutant telomeres back to wild type Col-0 (2-4 kb) level, but not to the wild type Sf-2 (4-6 kb) length. This result is not entirely surprising, since telomere length set point is a quantitative trait[109]. The observation demonstrates that while NOP2A is not sufficient to elongate Col-0 telomere length to the Sf-2 level, it is necessary for Col-0 to have telomere length of 2-4kb. From this result, an important question arises: is NOP2A one of the reasons Sf-2 has long telomeres? To answer this question, a *nop2a* mutant in the Sf-2 background needs to be examined. Generation of *nop2a* in Sf-2 using CRISPR-Cas9 is currently underway. It will be interesting to see how the absence of NOP2A impacts telomere length set point in Sf-2.

Is NOP2A a major gene in determining telomere length?

Although NOP2A is not the only factor involved in determining telomere length, Chapter III shows a large QTL peak on chromosome 5 where the NOP2A locus is located. In addition, Chapter IV shows that NOP2A is necessary for Col-0 to have its wild type telomere length. Together these findings imply that NOP2A may be one of the major determinants for telomere length set point.. To test this hypothesis Chapter IV shows data from cross-accession, Col-0 x Sf-2, telomere length analysis where F2 plants were genotyped at the NOP2A locus and segregated into three genotypes (Col-0/Col-0, Col-0/Sf-2, and Sf-2/Sf-2). Three key conclusions can be made from the results of these experiments: 1) only the Sf-2/Sf-2 genotype has long telomeres, 2) not all the Sf-2/ Sf-2 genotypes have long telomeres, and 3) some Col-0/ Col-0 genotypes have telomeres longer than wild type Col-0. The first observation indicates that NOP2A is necessary for long telomeres, the second observation indicates that there is another factor that functions with NOP2A, and the third observation suggests that there is another factor independent of NOP2A that plays a role in telomere length determination.

A smaller QTL peak on chromosome 1 was also identified during QTL mapping of telomere length among MAGIC lines. Does the other factor reside within this QTL interval peak? This question can be answered by analyzing the telomeres of F2 progeny from cross of Sf-2/Sf-2 with long telomeres by Col-0/ Col-0. The segregation pattern of telomeres in these F3 plants can be analyzed to determine if the phenotype segregates in a standard Mendelian manner. Furthermore, F3 plants with long telomeres can be genotyped at the QTL peak from chromosome 1 to see whether these plants acquired this loci from Sf-2.

It has been shown that plants with Sf-2/ Col-0 genotype have a similar telomere profile to Col-0/ Col-0. There are plants with telomeres longer than Col-0, but none of the plants have telomeres in the range of 4-6kb. This suggests that plants need to acquire both copies of NOP2A loci from Sf-2 to have long telomeres. However, only a small number of plants have been tested; therefore, more samples need to be analyzed to confirm this hypothesis.

Is there a difference in NOP2A locus among different accessions?

Is there a difference in the NOP2A locus that can explain telomere length set point variation among *A. thaliana* accessions? In Chapter IV, the NOP2A transcript level as well as NOP2A DNA and amino acid sequence of 19 MAGIC parental accessions are analyzed to address this question.

Transcript level analysis of NOP2A using data sets from the GEO repository (GSE53198) shows a range in NOP2A transcript levels; however, there is no correlation between transcript abundance and telomere length. Although it is difficult to draw any conclusions about the effect of NOP2A transcript level with telomere length, over-expression of NOP2A in Col-0 can be tested to see whether the level of NOP2A protein correlates with telomere length.

Amino acid alignment of NOP2A shows several amino acid differences among 19 MAGIC parental accessions. Interestingly, these differences are located outside the conserved rRNA methyltransferase domain, suggesting that telomere length regulation may not be conveyed by the catalytic domain of NOP2A. A role for rRNA methyltransferase activity in telomere length regulation cannot be completely excluded

however, since amino acid differences outside the catalytic active site might alter post-translational modification or even allosteric regulation of the active site. Three major differences identified are E61V and M165K in both Sf-2 and Bur-0, and W655R in Col-0. These changes can potentially affect binding affinity to yet unknown binding partner(s) involved in telomere length set point. Experiments that could shed light on how natural variation in NOP2A contributes to variations in telomere length regulation include1) analysis of NOP2A phosphorylation and ubiquitination and 2) mass-spectroscopy on NOP2A immunoprecipitated samples to identify binding partners.

Conclusions and Summary

In summary, this dissertation has provided new insight for the consequences of a newly duplicated gene that was implicated in telomere biology, POT1c. The data demonstrate a very different path this duplicate has taken relative to its two other paralogs. POT1c has non-functionalized through permanent silencing, while POT1b has neo-functionalized and POT1a has retained several of the ancestral functions. The data presented show 1) POT1c does not have a functional promoter; 2) transposable elements were inserted upstream of the locus after duplication, and 3) the locus is not conserved among 855 *A. thaliana* accessions. In marked contrast, the progenitor of POT1c, POT1a, is an essential gene that functions in telomerase regulation, and its sequence is highly conserved among 855 *A. thaliana* accessions. Study of *A. thaliana* POT1c gene locus has revealed that dosage balance was placed on POT1c locus through non-functionalization, consistent with the hypothesis of dosage balance of POT1 proteins[307].

Another important gene analyzed in this dissertation is NOP2A, a novel determinant in telomere length homeostasis and set point. Data presented in this dissertation demonstrated that 1) NOP2 may connect telomere biology, cell cycle and cancer progression, 2) *nop2a* contributes to a new telomere length set point, 3) ribosome biogenesis and telomere length can be connected, 4) NOP2A is one of the major factors involved in establishing telomere length in *A. thaliana*, 5) there are other factors that function in telomere length set point together with NOP2A, and 6) independent of NOP2A, and there are differences in NOP2A among accessions that potentially affect telomere length among these accessions. We have identified another conserved protein, NOP2A, that plays a role in telomere length regulation. Telomere length defects in human or yeast NOP2 mutants have not been previously reported, making this a novel finding. Furthermore, our data suggest that human NOP2 could promote oncotransformation through interactions with telomere maintenance pathways. In addition, for the first time we are connecting ribosome biogenesis as a major determinant of species-specific telomere length. Finally, although the NOP2A locus from Sf-2 is not sufficient, it is necessary for long telomeres in *A. thaliana*.

REFERENCES

1. Hayflick, L. & Moorhead, P. S. The serial cultivation of human diploid cell strains. *Exp. Cell Res.* **25,** 585–621 (1961).

2. Martens, U. M., Chavez, E. A., Poon, S. S. S., Schmoor, C. & Lansdorp, P. M. Accumulation of short telomeres in human fibroblasts prior to replicative senescence. *Exp. Cell Res.* **256,** 291–299 (2000).

3. Olovnikov, A. M. 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).

4. Olovnikov, A. M. Telomeres, telomerase, and aging: Origin of the theory. *Exp. Gerontol.* **31,** 443–448 (1996).

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

6. Harley, C. B., Vaziri, H., Counter, C. M. & Allsopp, R. C. The telomere hypothesis of cellular aging. *Exp. Gerontol.* **27,** 375–382 (1992).

7. Sharpless, N. E. & DePinho, R. A. How stem cells age and why this makes us grow old. *Nature Reviews Molecular Cell Biology* **8,** 703–713 (2007).

8. de Lange, T. How Telomeres Solve the End-Protection Problem. *Science.* **326,** 948-952 (2009).

9. O'Sullivan, R. J. & Karlseder, J. Telomeres: Protecting chromosomes against genome instability. *Nature Reviews Molecular Cell Biology* **11,** 171–181 (2010).

10. Lazzerini-Denchi, E. & Sfeir, A. Stop pulling my strings — what telomeres taught us about the DNA damage response. *Nat. Rev. Mol. Cell Biol.* **17,** 364–378 (2016).

11. Maestroni, L., Matmati, S. & Coulon, S. Solving the telomere replication problem. *Genes* **8***,* 1-16 (2017). doi:10.3390/genes8020055

12. Blackburn, E. H. The molecular structure of centromeres and telomres. *Annu. Rev. Biochem.* **53,** 163–194 (1984).

13. McClintock, B. Thestability ofbroken ends of chromosomes in *Zea mays*. *Genetics* **26,** 234–282 (1941).

14. Orr-Weaver, T. L., Szostak, J. W. & Rothstein, R. J. Yeast transformation: a model system for the study of recombination. *Proc. Natl. Acad. Sci.* **78,** 6354–6358 (1981).

15. Weinert, T. & Hartwell, L. The RAD9 gene controls the cell cycle response to DNA damage in *Saccharomyces cerevisiae*. *Science.* **241,** 317–322 (1988).

16. Sandell, L. L. & Zakian, V. A. Loss of a yeast telomere: Arrest, recovery, and chromosome loss. *Cell* **75,** 729–739 (1993).

17. Van Steensel, B., Smogorzewska, A. & De Lange, T. TRF2 protects human telomeres from end-to-end fusions. *Cell* **92,** 401–413 (1998).

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

19. Shay, J. W. & Wright, W. E. Ageing and cancer: the telomere and telomerase connection. *Novartis Found. Symp.* **235,** 146–149 (2001).

20. Shay, J. W. & Bacchetti, S. A survey of telomerase activity in human cancer. *Eur. J. Cancer* **33,** 787–791 (1997).

21. Ju, Z. & Rudolph, K. L. Telomeres and telomerase in stem cells during aging and disease. *Genome Dyn.* **1,** 84–103 (2006).

22. Blasco, M. A. Telomeres and human disease: Ageing, cancer and beyond. *Nature Reviews Genetics* **6,** 611–622 (2005).

23. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **144,** 646–674 (2011).

24. Rodrigo, T., Calado, M. D. & Dumitriu, B. Telomere dynamics in mice and humans.

Semin. Hematol. **50,** 165–174 (2013).

25. Calado, R. T. & Young, N. S. Telomere diseases. *N Engl J Med* **361,** 2353–2365 (2009).

26. Kim, N. W. *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science.* **226,** 2011–2015 (1994).

27. Bryan, T. M., Englezou, A., Dalla-Pozza, L., Dunham, M. A. & Reddel, R. R. Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat. Med.* **3,** 1271–1274 (1997).

28. Henson, J. D. *et al.* A robust assay for alternative lengthening of telomeres (ALT) in tumors demonstrates the significance of ALT in sarcomas and astrocytomas. *Clin. Cancer Res.* **11,** 217–225 (2005).

29. Subhawong, A. P. *et al.* The alternative lengthening of telomeres phenotype in breast carcinoma is associated with HER-2 overexpression. *Mod. Pathol.* **22,** 1423–1431 (2009).

30. Jeyapalan, J. N., Mendez-Bermudez, A., Zaffaroni, N., Dubrova, Y. E. & Royle, N. J. Evidence for alternative lengthening of telomeres in liposarcomas in the absence of ALT-associated PML bodies. *Int. J. Cancer* **122,** 2414–2421 (2008).

31. Hakin-Smith, V. *et al.* Alternative lengthening of telomeres and survival in patients with glioblastoma multiforme. *Lancet* **361,** 836–838 (2003).

32. Costa, A. *et al.* Telomere maintenance mechanisms in liposarcomas: Association with histologic subtypes and disease progression. *Cancer Res.* **66,** 8918–8924 (2006).

33. Dunham, M. a, Neumann, A. a, Fasching, C. L. & Reddel, R. R. Telomere maintenance by recombination in human cells. *Nat. Genet.* **26,** 447–450 (2000).

34. Ramsay, A. J. *et al.* POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nat. Genet.* **45,** 526–530 (2013).

35. Robles-Espinoza, C. D. *et al.* POT1 loss-of-function variants predispose to familial melanoma. *Nat. Genet.* **46,** 478–481 (2014).

36. Speedy, H. E. *et al.* A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. *Nat. Genet.* **46,** 56–60 (2014).

37. Shi, J. *et al.* Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. *Nat. Genet.* **46,** 482–486 (2014).

38. Diaz de Leon, A. *et al.* Telomere lengths, pulmonary fibrosis and telomerase (TERT) Mutations. *PLoS One* **5,** 1-12 (2010).

39. De Lange, T. Telomeres and senescence: Ending the debate. *Science.* **279,** 334–335 (1998).

40. Stewart, S. A. & Weinberg, R. A. Senescence: Does it all happen at the ends? *Oncogene* **21,** 627–630 (2002).

41. Smith, S. K. & Kipling, D. The role of replicative senescence in cancer and human ageing: Utility (or otherwise) of murine models. *Cytogenet. Genome Res.* **105,** 455–463 (2004).

42. Planas-Silva, M. D. & Weinberg, R. A. The restriction point and control of cell proliferation. *Curr. Opin. Cell Biol.* **9,** 768–772 (1997).

43. Kalan, S. & Loayza, D. Shelterin complex in telomere protection: Recent insights and pathological significance. *Cell Health Cytoskelet.* **6,** 11–26 (2014). doi:10.2147/CHC.S49687

44. Armanios, M. & Blackburn, E. H. The telomere syndromes. *Nat. Rev. Genet.* **13,** 693–704 (2012).

45. Lee, H. W. *et al.* Essential role of mouse telomerase in highly proliferative organs. *Nature* **392,** 569–574 (1998).

46. Herrera, E. *et al.* Disease states associated with telomerase deficiency appear earlier in mice with short telomeres. *EMBO J.* **18,** 2950–2960 (1999).

47. Goytisolo, B. F. A. *et al.* Shorttelomeres result in organismal hypersensitivity to ionizing radiation inmammals. *JEM* **192,** 1625–1636 (2000).

48. Ferrón, S. *et al.* Telomere shortening and chromosomal instability abrogates proliferation of adult but not embryonic neural stem cells. *Development* **131,** 4059–4070 (2004).

49. Leri, A. *et al.* Ablation of telomerase and telomere loss leads to cardiac dilatation and heart failure associated with p53 upregulation. *EMBO J.* **22,** 131–139 (2003).

50. Oh, H. *et al.* Telomere attrition and Chk2 activation in human heart failure. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 5378–5383 (2003).

51. O'Sullivan, J. N. *et al.* Chromosomal instability in ulcerative colitis is related to telomere shortening. *Nat. Genet.* **32,** 280–284 (2002).

52. Wiemann, S. U. *et al.* Hepatocyte telomere shortening and senescence are general markers of human liver cirrhosis. *FASEB J.* **16,** 935–942 (2002).

53. Samani, N. J., Boultby, R., Butler, R., Thompson, J. R. & Goodall, A. H. Telomere shortening in atherosclerosis. *Lancet* **358,** 472–473 (2001).

54. Ball, S. E. *et al.* Progressive telomere shortening in aplastic anemia. *Blood* **91,** 3582–3592 (1998).

55. Yamaguchi, H. *et al.* Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia. *N. Engl. J. Med.* **352,** 1413-1424 (2005).

56. Marrone, A., Stevens, D., Vulliamy, T., Dokal, I. & Mason, P. J. Heterozygous telomerase RNA mutations found in dyskeratosis congenita and aplastic anemia reduce telomerase activity via haploinsufficiency. *Blood* **104,** 3936–3942 (2004).

57. Walne, A. J. *et al.* TINF2 mutations result in very short telomeres: analysis of a large cohort of patients with dyskeratosis congenita and related bone marrow failure syndromes. *Blood* **112,** 3594–3600 (2008).

58. Anderson, B. H. *et al.* Mutations in CTC1, encoding conserved telomere maintenance component 1, cause Coats plus. *Nat. Genet.* **44,** 338–342 (2012).

59. Mitchell, J. R., Wood, E. & Collins, K. A telomerase component is defective in the human disease dyskeratosis congenita. *Lett. to Nat.* **402,** 551–555 (1999).

60. Vulliamy, T. *et al.* The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Lett. to Nat.* **413,** 432–435 (2001).

61. Sasa, G. S., Ribes-Zamora, A., Nelson, N. D. & Bertuch, A. A. Three novel truncating TINF2 mutations causing severe dyskeratosis congenita in early childhood. *Clin. Genet.* **81,** (2012).

62. Zhong, F. *et al.* Disruption of telomerase trafficking by TCAB1 mutation causes dyskeratosis congenita service disruption of telomerase trafficking by TCAB1 mutation causes dyskeratosis congenita. **25,** 11–16 (2011). doi:10.1101/gad.2006411

63. Walne, A. J. *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).

64. Vulliamy, T. *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).

65. Armanios, M. Y. *et al.* Telomerase mutations in families with idiopathic pulmonary fibrosis. *N. Engl. J. Med.* **356,** 1317–1326 (2007).

66. Dokal, I. Dyskeratosis congenita in all its forms. *Br. J. Haematol.* **110,** 768–779 (2000).

67. Jonassaint, N. L., Guo, N., Califano, J. A., Montgomery, E. A. & Armanios, M. The gastrointestinal manifestations of telomere-mediated disease. *Aging Cell* **12,** 319–323 (2013).

68. Martinez, P. *et al.* Mammalian Rap1 controls telomere function and gene expression through binding to telomeric and extratelomeric sites. *Nat. Cell Biol.* **12,** 768–780

(2010).

69. Zakian, V. A. Telomeres: beginning to understand the end. *Science.* **270,** 1601–1607 (1995).

70. Moyzis, R. K. *et al.* A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci.* **85,** 6622–6626 (1988).

71. Brown, W. R. Molecular cloning of human telomeres in yeast. *Nature* **338,** 774–776 (1989).

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

73. Runge, K. W. & Zakian, V. A. Introduction of extra telomeric DNA sequences into *Saccharomyces cerevisiae* results in telomere elongation. *Mol. Cell. Biol.* **9,** 1488–1497 (1989).

74. Fajkus, J., Kovařík, A., MKrálovics, R. & Bezděk, M. Organization of telomeric and subtelomeric chromatin in the higher plant *Nicotiana tabacum*. *Mol. Gen. Genet.* **247,** 633–638 (1995).

75. Deng, Z., Wang, Z. & Lieberman, P. M. Telomeres and viruses: common themes of genome maintenance. *Front. Oncol.* **2,** 1-11 (2012).

76. Kobryn, K. & Chaconas, G. The circle is broken: Telomere resolution in linear replicons. *Current Opinion in Microbiology* **4,** 558–564 (2001).

77. Hinnebusch, J. & Tilly, K. Linear plasmid and chromosomes in bacteria. *Mol. Microbiol.* **10,** 917–922 (1993).

78. Zhang, L. & Rong, Y. S. Retrotransposons at *Drosophila* telomeres: Host domestication of a selfish element for the maintenance of genome integrity. *Biochimica et Biophysica Acta 1891* **7,** 771–775 (2012). doi:10.1016/j.bbagrm.2012.01.018

79. Pardue, M.-L. & DeBaryshe, P. G. Retrotransposons that maintain chromosome ends. *PNAS* **108,** 20317–20324 (2011).

80. Louis, E. J. Are *Drosophila* telomeres an exception or the rule? *Genome Biol.* **3,** 1-6 (2002).

81. Kraus, E., Leung, W.-Y. & Haber, J. E. Break-induced replication: A review and an example in budding yeast. *Proc. Natl. Acad. Sci.* **98,** 8255–8262 (2001).

82. Nosek, J., Kosa, P. & Tomaska, L. On the origin of telomeres: A glimpse at the pre-telomerase world. *BioEssays* **28,** 182–190 (2006).

83. Deneke, J., Ziegelin, G., Lurz, R. & Lanka, E. The protelomerase of temperate *Escherichia coli* phage N15 has cleaving-joining activity. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 7721–7726 (2000).

84. Hertwig, S., Klein, I., Lurz, R., Lanka, E. & Appel, B. PY54, a linear plasmid prophage of *Yersinia enterocolitica* with covalently closed ends. *Mol. Microbiol.* **48,** 989–1003 (2003).

85. Huang, W. M. *et al.* Linear chromosome-generating system of*Agrobacterium tumefaciens* C58: Protelomerase generates and protects hairpin ends. *J. Biol. Chem.* **287,** 25551–25563 (2012).

86. Casjens, S. Evolution of the linear DNA replicons of the *Borrelia spirochetes. Current Opinion in Microbiology* **2,** 529–534 (1999).

87. Huang, W. M., Robertson, M., Aron, J. & Casjens, S. Telomere exchange between linear replicons of *Borrelia burgdorferi*. *J. Bacteriol.* **186,** 4134-4141 (2004). doi:10.1128/JB.186.13.4134-4141.2004

88. Traverse, K. L. & Pardue, M. L. A spontaneously opened ring chromosome of *Drosophila melanogaster* has acquired He-T DNA sequences at both new telomeres. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 8116–20 (1988).

89. Levis, R. W., Ganesan, R., Houtchens, K., Tolar, L. A. & Sheen, F. miin. Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* **75,** 1083–1093 (1993).

90. Abad, J. P. *et al.* TAHRE, a novel telomeric retrotransposon from *Drosophila melanogaster*, reveals the origin of *Drosophila* telomeres. *Mol. Biol. Evol.* **21,** 1620–1624 (2004).

91. Mikhailovsky, S., Belenkaya, T. & Georgiev, P. Broken chromosomal ends can be elongated by conversion in *Drosophila melanogaster*. *Chromosoma* **108,** 114–120 (1999).

92. Gilson, E. & Géli, V. How telomeres are replicated. *Nat. Rev. Mol. Cell Biol.* **8,** 825–838 (2007).

93. Henderson, E. R. & Blackburn, E. H. An overhanging 3' terminus is a conserved feature of telomeres. *Mol. Cell. Biol.* **9,** 345–348 (1989).

94. Lingner, J. & Cech, T. R. Purification of telomerase from *Euplotes aediculatus*: requirement of a primer 3' overhang. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 10712–10717 (1996).

95. Makarov, V. L., Hirose, Y. & Langmore, J. P. Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. *Cell* **88,** 657–666 (1997).

96. Dionne, I. & Wellinger, R. J. Cell cycle-regulated generation of single-stranded G-rich DNA in the absence of telomerase. *Proc. Natl. Acad. Sci.* **93,** 13902–13907 (1996).

97. Vodenicharov, M. D. & Wellinger, R. J. DNA Degradation at unprotected telomeres in yeast is regulated by the CDK1 (Cdc28/Clb) cell-cycle kinase. *Mol. Cell* **24,** 127-137 (2006). doi:10.1016/j.molcel.2006.07.035

98. Wu, P., Hiroyuki Takai & Lange, T. de. Telomeric 3′ overhangs derive from resection by Exo1 and Apollo and fill-in by POT1b-associated CST. *Cell* **150,** 39–52 (2012).

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

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

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

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

103. Wang, R. C., Smogorzewska, A. & De Lange, T. Homologous recombination generates t-loop-sized deletions at human telomeres. *Cell* **119,** 355–368 (2004).

104. Hug, N. & Lingner, J. Telomere length homeostasis. *Chromosoma* **115,** 413–425 (2006).

105. Shore, D. & Bianchi, A. Telomere length regulation: Coupling DNA end processing to feedback regulation of telomerase. *EMBO J.* **28,** 2309–2322 (2009).

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

107. Kipling, D. & Cooke, H. J. Hypervariable ultra-long telomeres in mice. *Lett. to Nat.* **347,** 400–402 (1990).

108. Allshire, R. C., Dempster, M. & Hastie, N. D. Human telomeres contain at least three types of G-rich repeat distributed non-randomly. *Nucleic Acids Res.* **17,** 4611–4627 (1989).

109. Gatbonton, T. *et al.* Telomere length as a quantitative trait: Genome-wide survey and genetic mapping of telomere length-control genes in yeast. *PLoS Genet.* **2,** 0304–0315 (2006).

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

111. Stewart, J. a, Chaiken, M. F., Wang, F. & Price, C. M. Maintaining the end: roles of

telomere proteins in end- protection, telomererereplication andlength regulation. *Mutat Res.* **730,** 12–19 (2012).

112. Lustig, A. J. Clues to catastrophic telomere loss in mammals from yeast telomere rapid deletion. *Nat. Rev. Genet.* **4,** 916–923 (2003).

113. Nora, G. J., Buncher, N. A. & Opresko, P. L. Telomeric protein TRF2 protects Holliday junctions with telomeric arms from displacement by the werner syndrome helicase. *Nucleic Acids Res.* **38,** 3984–3998 (2010).

114. Schmutz, I., Timashev, L., Xie, W., Patel, D. J. & De Lange, T. TRF2 binds branched DNA to safeguard telomere integrity. *Nat. Struct. Mol. Biol.* **24,** 734–742 (2017).

115. Poulet, A. *et al.* TRF2 promotes, remodels and protects telomeric Holliday junctions. *EMBO J.* **28,** 641–651 (2009).

116. Karlseder, J., Smogorzewska, A. & De Lange, T. Senescence induced by altered telomere state, not telomere loss. *Science.* **295,** 2446–2449 (2002).

117. Pickett, H. A., Cesare, A. J., Johnston, R. L., Neumann, A. A. & Reddel, R. R. Control of telomere length by a trimming mechanism that involves generation of t-circles. *EMBO J.* **28,** 799–809 (2009).

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

119. Zellinger, B., Akimcheva, S., Puizina, J., Schirato, M. & Riha, K. Ku Suppresses formation of telomeric circles and alternative telomere lengthening in *Arabidopsis*. *Mol. Cell* **27,** 163–169 (2007).

120. Lundblad, V. Telomere maintenance without telomerase. *Oncogene* **21,** 522–531 (2002).

121. Natarajan, S. & McEachern, M. J. Recombinational telomere elongation promoted by DNA circles. *Mol. Cell. Biol.* **22,** 4512–4521 (2002).

122. Cesare, A. J. & Griffith, J. D. Telomeric DNA in ALT cells is characterized by free telomeric circles and heterogeneous t-loops. *Mol Cell Biol* **24,** 9948–9957 (2004).

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

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

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

126. Shippen-lentz, D. & Blackburn, E. H. Telomere terminal transferase activity from *Euplotes crassus* adds large numbers of TTTTGGGG repeats onto telomeric primers. *Mol. Cell. Biol.* **9,** 2761–2764 (1989).

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

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

129. Nakamura, T. M. *et al.* Telomerase catalytic subunit homologs from yeast and human. *Science.* **277,** 955–959 (1997).

130. Singer, M. S. & Gottschling, D. E. TLC1: template RNA component of Saccharomyces cerevisiae telomerase. *Science.* **266,** 404–409 (1994).

131. Leonardi, J., Box, J. A., Bunch, J. T. & Baumann, P. TER1, the RNA subunit of fission yeast telomerase. *Nat. Struct. Mol. Biol.* **15,** 26–33 (2008).

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

133. Guttman, M. *et al.* Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* **28,** 503–510 (2010).

134. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25,** 1915–1927 (2012).

135. Banfai, B. *et al.* Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* **22,** 1646–1657 (2012).

136. Lendvay, T. S., Morris, D. K., Sah, J., Balasubramanian, B. & Lundblad, V. Senescence mutants of *Saccharomyces cerevisiae* with a defect in telomere replication identify three additional EST genes. *Genetics* **144,** 1399–1412 (1996).

137. Collins, K. The biogenesis and regulation of telomerase holoenzymes. *Nat. Rev. Mol. Cell Biol.* **7,** 484–494 (2010).

138. Blackburn, E. H. & Collins, K. Telomerase: An RNP enzyme synthesizes DNA. *Cold Spring Harb. Perspect. Biol.* **3,** 1-9 (2011). doi:10.1101/cshperspect.a003558

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

140. Podlevsky, J. D. & Chen, J. J. L. Evolutionary perspectives of telomerase RNA structure and function. *RNA Biol.* **13,** 720–732 (2016).

141. Wu, R. A., Upton, H. E., Vogan, J. M. & Collins, K. Telomerase mechanism of telomere synthesis. *Annu. Rev. Biochem.* **86,** 439–460 (2017).

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

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

144. McCormick-Graham, M. & Romero, D. P. A single telomerase RNA is sufficient for the synthesis of variable telomeric DNA repeats in ciliates of the genus *Paramecium*. *Mol. Cell. Biol.* **16,** 1871–1879 (1996).

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

146. Schmidt, J. C. & Cech, T. R. Human telomerase : biogenesis , trafficking, recruitment, and activation. *Genes Dev.* **29,** 1095–1105 (2015).

147. Gallardo, F. & Chartrand, P. Telomerase biogenesis: The long road before getting to the end. *RNA Biol.* **5,** 212–215 (2008).

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

149. Tucey, T. M. & Lundblad, V. Regulated assembly and disassembly of the yeast telomerase quaternary complex. *Genes Dev.* **28,** 2077–2089 (2014).

150. Gallardo, F., Olivier, C., Dandjinou, A. T., Wellinger, R. J. & Chartrand, P. TLC1 RNA nucleo-cytoplasmic trafficking links telomerase biogenesis to its recruitment to telomeres. *EMBO J.* **27,** 748–757 (2008).

151. Milne, G. T., Jin, S., Shannon, K. B. & Weaver, D. T. Mutations in two Ku homologs define a DNA end-joining repair pathway in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16,** 4189–4198 (1996).

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

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

154. Seto, A. G., Zaug, A. J., Sobel, S. G., Wolin, S. L. & Cech, T. R. *Saccharomyces cerevisiae* telomerase is an Sm small nuclear ribonucleoprotein particle. *Lett. to Nat.*

401, 177–180 (1999).

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

156. Vogan, J. M. *et al.* Minimized human telomerase maintains telomeres and resolves endogenous roles of H/ACA proteins, TCAB1, and Cajal bodies. *Elife* **5,** 1-21 (2016).

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

158. Tycowski, K. T., Shu, M. Di, Kukoyi, A. & Steitz, J. A. A conserved WD40 protein binds the cajal body localization signal of scaRNP particles. *Mol. Cell* **34,** 47–57 (2009).

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

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

161. BeataE.Jady, Patricia Richard, Bertrand, E. & Kiss, T. Cell cycle-dependent recruitment of telomerase RNA and cajal bodies to human telomeres. *Mol. Biol. Cell* **17,** 944–954 (2006).

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

163. Dragon, F., Pogacić, V. & Filipowicz, W. In vitro assembly of human H/ACA small nucleolar RNPs reveals unique features of U17 and telomerase RNAs. *Mol. Cell. Biol.* **20,** 3037–3048 (2000).

164. Dez, C. *et al.* Stable expression in yeast of the mature form of human telomerase RNA depends on its association with the box H/ACA small nucleolar RNP proteins

Cbf5p, Nhp2p and Nop10p. *Nucleic Acids Res.* **29,** 598–603 (2001).

165. Darzacq, X. *et al.* Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *J. Cell Biol.* **173,** 207–218 (2006).

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

167. Cohen, S. B. *et al.* Protein composition of catalytically active human telomerase from immortal cells. *Science.* **1185,** 1850–1853 (2007).

168. Venteicher, A. S. & Artandi, S. E. TCAB1: Driving telomerase to Cajal bodies. *Cell Cycle* **8,** 1329–1331 (2009).

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

170. Lai, C. K., Miller, M. C. & Collins, K. Roles for RNA in telomerase nucleotide and repeat addition processivity. *Mol. Cell* **11,** 1673–1683 (2003).

171. Greider, C. W. Telomerase is processive. *Mol. Cell. Biol.* **11,** 4572–4580 (1991).

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

173. Lingner, J., Cech, T. R., Hughes, T. R. & Lundblad, V. 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).

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

175. Steiner, B. R., Hidaka, K. & Futcher, B. Association of the Est1 protein with telomerase activity in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 2817–2821 (1996).

176. Zhou, J., Hidaka, K. & Futcher, B. The Est1 Subunit of Yeast Telomerase binds the Tlc1 telomerase RNA. *Mol. Cell. Biol.* **20,** 1947–1955 (2000).

177. DeZwaan, D. C. & Freeman, B. C. The conserved Est1 protein stimulates telomerase DNA extension activity. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 17337–17342 (2009).

178. Renfrew, K. B., Song, X., Lee, J. R., Arora, A. & Shippen, D. E. POT1a and components of CST engage telomerase and regulate its activity in *Arabidopsis*. *PLoS Genet.* **10,** 1-12 (2014).

179. Talley, J. M., DeZwaan, D. C., Maness, L. D., Freeman, B. C. & Friedman, K. L. 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).

180. Tuzon, C. T., Wu, Y., Chan, A. & Zakian, V. A. The *saccharomyces cerevisiae* telomerase subunit Est3 binds telomeres in a cell cycle- and Est1-dependent manner and interacts directly with Est1 *in vitro. PLoS Genet.* **7,** 1-12 (2011).

181. Lee, J., Mandell, E. K., Rao, T., Wuttke, D. S. & Lundblad, V. Investigating the role of the Est3 protein in yeast telomere replication. *Nucleic Acids Res.* **38,** 2279–2290 (2010).

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

183. Yen, W.-F., Chico, L., Lei, M. & Lue, N. F. Telomerase regulatory subunit Est3 in two *Candida* species physically interacts with the TEN domain of TERT and telomeric DNA. *Proc. Natl. Acad. Sci.* **108,** 20370–20375 (2011).

184. Rao, T. *et al.* Structure of Est3 reveals a bimodal surface with differential roles in telomere replication. *Proc. Natl. Acad. Sci.* **111,** 214–218 (2014).

185. Lee, J. S., Edward K. Mandell, Tucey, T. M., Morris, D. K. & Victoria, L. The Est3 protein associates with yeast telomerase through an OB-fold domain. *Nat. Struct. Mol. Biol.* **15,** 990–997 (2008).

186. Fell, V. L. & Schild-Poulter, C. The Ku heterodimer: Function in DNA repair and beyond. *Mutat. Res.* **763,** 15–29 (2015).

187. Downs, J. A. & Jackson, S. P. A means to a DNA end: The many roles of Ku. *Nat. Rev. Mol. Cell Biol.* **5,** 367–378 (2004).

188. Nugent, C. I. *et al.* Telomere maintenance is dependent on activities required for end repair of double-strand breaks. *Curr. Biol.* **8,** 657–662 (1998).

189. Boulton, S. J. & Jackson, S. P. Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. *EMBO J.* **17,** 1819–1828 (1998).

190. Polotnianka, R. M., Li, J. & Lustig, A. J. The yeast Ku heterodimer is essential for protection of the telomere against nucleolytic and recombinational activities. *Curr. Biol.* **8,** 831–834 (1998).

191. Bertuch, A. A. & Lundblad, V. The Ku heterodimer performs separable activities at double-strand breaks and chromosome termini. *Mol. Cell. Biol.* **23,** 8202–8215 (2003).

192. Riha, K., Watson, J. M., Parkey, J. & Shippen, D. E. Telomere length deregulation and enhanced sensitivity to genotoxic stress in Arabidopsis mutants deficient in Ku70. *EMBO J.* **21,** 2819–2826 (2002).

193. Riha, K. & Shippen, D. E. Ku is required for telomeric C-rich strand maintenance but not for end-to-end chromosome fusions in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 611–615 (2003).

194. Gallego, M. E. Telomerase Dependence of Telomere Lengthening in ku80 Mutant *Arabidopsis*. *Plant Cell* **15,** 782–789 (2003).

195. Fagagna, D. di F. *et al.* Effects of DNA nonhomologous end-joining factors on telomere length and chromosomal stability in mammalian cells. *Curr. Biol.* **11,** 1192–1196 (2001).

196. Samper, E., Flores, J. M. & Blasco, M. A. Restoration of telomerase activity rescues chromosomal instability and premature aging in Terc-/- mice with short telomeres. *EMBO Rep.* **2,** 800–807 (2001).

197. Wyman, C. & Kanaar, R. DNA Double-Strand Break Repair: All's Well that Ends Well. *Annu. Rev. Genet.* **40,** 363–383 (2006).

198. Van Gent, D. C., Hoeijmakers, J. H. J. & Kanaar, R. Chromosomal stability and the DNA double-stranded break connection. *Nat. Rev. Genet.* **2,** 196–206 (2001).

199. Khanna, K. K. & Jackson, S. P. DNA double-strand breaks: Signaling, repair and the cancer connection. *Nat. Genet.* **27,** 247–254 (2001).

200. Forney, J. D. & Blackburn, E. H. Developmentally controlled telomere addition in wild-type and mutant paramecia. *Mol. Cell. Biol.* **8,** 251–258 (1988).

201. Magnenat, L., Tobler, H. & Müller, F. Developmentally regulated telomerase activity is correlated with chromosomal healing during chromatin diminution in Ascaris suum. *Mol. Cell. Biol.* **19,** 3457–3465 (1999).

202. Kramer, K. M. & Haber, J. E. New telomeres in yeast are initiated with a highly selected subset of TG1-3 repeats. *Genes Dev.* **7,** 2345–2356 (1993).

203. Sprung, C. N., Reynolds, G. E., Jasin, M. & Murnane, J. P. Chromosome healing in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 6781–6786 (1999).

204. Flint, J. *et al.* Healing of broken human chromosomes by the addition of telomeric repeats. *Am. J. Hum. Genet.* **55,** 505–512 (1994).

205. Fitzgerald, M. S., Shakirov, E. V., Hood, E. E., McKnight, T. D. & Shippen, D. E. Different modes of de novo telomere formation by plant telomerases. *Plant J.* **26,** 77–87 (2001).

206. Varley, H., Di, S., Scherer, S. W. & Royle, N. J. Characterization of terminal deletions at 7q32 and 22q13.3 healed by *de novo* telomere addition. *Am. J. Hum. Genet.* **67,** 610–622 (2000).

207. Wong, a C. *et al.* Molecular characterization of a 130-kb terminal microdeletion at 22q in a child with mild mental retardation. *Am. J. Hum. Genet.* **60,** 113–120 (1997).

208. Schulz, V. P. & Zakian, V. A. The *Saccharomyces* PIF1 DNA helicase inhibits telomere elongation and *de novo* telomere formation. *Cell* **76,** 145–155 (1994).

209. Myung, K., Pennaneach, V., Kats, E. S. & Kolodner, R. D. *Saccharomyces cerevisiae* chromatin-assembly factors that act during DNA replication function in the maintenance of genome stability. *Proc. Natl. Acad. Sci.* **100,** 6640–6645 (2003).

210. Zhang, W. & Durocher, D. De novo telomere formation is suppressed by the Mec1-dependent inhibition of Cdc13 accumulation at DNA breaks. *Genes Dev.* **24,** 502–515 (2010).

211. Phillips, J. A., Chan, A., Paeschke, K. & Zakian, V. A. The Pif1 helicase, a negative regulator of telomerase, acts preferentially at long telomeres. *PLoS Genet.* **11,** 1–19 (2015).

212. Makovets, S. & Blackburn, E. H. DNA damage signalling prevents deleterious telomere addition at DNA breaks. *Nat. Cell Biol.* **11,** 1383–1386 (2009).

213. Boulé, J. B., Vega, L. R. & Zakian, V. A. The yeast Pif1p helicase removes telomerase from telomeric DNA. *Nature* **438,** 57–61 (2005).

214. Jia, P., Chastain, M., Zou, Y., Her, C. & Chai, W. Human MLH1 suppresses the insertion of telomeric sequences at intra-chromosomal sites in telomerase-expressing cells. *Nucleic Acids Res.* **45,** 1219–1232 (2017).

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

216. Xu, H., Nelson, A. D. L. & Shippen, D. E. A Transposable element within the non-canonical telomerase RNA of *Arabidopsis thaliana* modulates telomerase in response to DNA damage. *PLoS Genet.* 1-20 (2015). doi:10.1371/journal.pgen.1005281

217. Wang, K. C. & Chang, H. Y. Molecular mechanisms of long noncoding RNAs. *Mol Cell* **43,** 904–914 (2012).

218. Kung, J. T. Y., Colognori, D. & Lee, J. T. Long Noncoding RNAs: Past, present, and future. *Genetics* **193,** 651–669 (2013).

219. Linger, B. R. & Price, C. M. Conservation of Telomere protein complexes: Shuffling through Evolution. *Crit Rev Biochem Mol Biol* **44,** 434–446 (2009).

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

221. Palm, W. & de Lange, T. How shelterin protects mammalian telomeres. *Annu. Rev. Genet.* **42,** 301–334 (2008).

222. Gottschling, D. E. & Zakian, V. A. Telomere proteins: Specific recognition and protection of the natural termini of *Oxytricha macronuclear* DNA. *Cell* **47,** 195–205 (1986).

223. Horvath, M. P., Schweiker, V. L., Bevilacqua, J. M., Ruggles, J. A. & Schultz, S. C. Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA. *Cell* **95,** 963–974 (1998).

224. De Lange, T. Shelterin: The protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19,** 2100–2110 (2005). doi:10.1101/gad.1346005

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

226. Wellinger, R. J. The CST complex and telomere maintenance: The exception becomes the rule. *Mol. Cell* **36,** 168–169 (2009).

227. Price, C. M. *et al.* Evolution of CST function in telomere maintenance. *Cell Cycle* **9,** 3157–3165 (2010).

228. Hockemeyer, D. & Collins, K. Control of telomerase action at human telomeres. *Nat. Struct. Mol. Biol.* **22,** 848–852 (2015).

229. Bianchi, A. & Shore, D. How telomerase reaches its end: Mechanism of telomerase regulation by the telomeric complex. *Mol. Cell Rev.* **31,** 153–165 (2008).

230. Surovtseva, Y. V. *et al.* Conserved Tlomere mintenance Component 1 interacts with STN1 and maintains chromosome ends in higher eukaryotes. *Mol. Cell* **36,** 207–218 (2009).

231. Gelinas, A. D. *et al.* Telomere capping proteins are structurally related to RPA with an additional telomere-specific domain. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 19298–303 (2009).

232. Garvik, B., Carson, M. & Hartwell, L. Single-stranded DNA arising at telomeres in cdc13 mutants may constitute a specific signal for the RAD9 checkpoint. *Mol. Cell. Biol.* **15,** 6128–6138 (1995).

233. Nugent, C. I., Hughes, T. R., Lue, N. F. & Lundblad, V. Cdc13p: A single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance. *Science.* **274,** 249–252 (1996).

234. Pennock, E., Buckley, K. & Lundblad, V. Cdc13 delivers separate complexes to the telomere for end protection and replication. *Cell* **104,** 387–396 (2001).

235. Grandin, N., Reed, S. I. & Charbonneau, M. Stn1, a new *Saccharomyces cerevisiae* protein, is implicated in telomere size regulation in association with CDc13. *Genes Dev.* **11,** 512–527 (1997).

236. Puglisi, A., Bianchi, A., Lemmens, L., Damay, P. & Shore, D. Distinct roles for yeast Stn1 in telomere capping and telomerase inhibition. *EMBO J.* **27,** 2328–2339 (2008).

237. Grandin, N., Damon, C. & Charbonneau, M. Ten1 functions in telomere end protection and length regulation in association with Stn1 and Cdc13. *EMBO J.* **20,** 1173–1183 (2001).

238. Song, X. *et al.* STN1 protects chromosome ends in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 19815–20 (2008).

239. Leehy, K. A., Lee, J. R., Song, X., Renfrew, K. B. & Shippen, D. E. MERISTEM

DISORGANIZATION1 encodes TEN1, an essential telomere protein that modulates telomerase processivity in *Arabidopsis*. *Plant Cell* **25,** 1343–1354 (2013).

240. Chandra, A., Hughes, T. R., Nugent, C. I. & Lundblad, V. Cdc13 both positively and negatively regulates telomere replication Cdc13 both positively and negatively regulates telomere replication. *Genes Dev.* **15,** 404–414 (2001). doi:10.1101/gad.861001

241. Evans, S. K. & Lundblad, V. Est1 and Cdc13 as comediators of telomerase access. *Science.* **286,** 117–120 (1999).

242. Li, S. *et al.* Cdk1-dependent phosphorylation of Cdc13 coordinates telomere elongation during cell cycle progression. *Cell* **136,** 50–61 (2009).

243. Casteel, D. E. *et al.* A DNA polymerase-α·primase cofactor with homology to replication protein A-32 regulates DNA replication in mammalian cells. *J. Biol. Chem.* **284,** 5807–5818 (2009).

244. Goulian, M. & Heard, C. J. The mechanism of action of an accessory protein for DNA polymerase alpha/primase. *J. Biol. Chem.* **265,** 13231–13239 (1990).

245. Grossi, S., Puglisi, A., Dmitriev, P. V., Lopes, M. & Shore, D. Pol12, the B subunit of DNA polymerase α, functions in both telomere capping and length regulation. *Genes Dev.* **18,** 992–1006 (2004).

246. Goulian, M., Heard, C. J. & Grimm, S. L. Purification and properties of an accessory protein for DNA polymerase a/primase. *J. Biol. Chem.* **265,** 13221–13230 (1990).

247. Wang, F. *et al.* Human CST has independent functions during telomere duplex replication and C-strand fill-in. *Cell Rep.* **2,** 1096–1103 (2012).

248. Huang, C., Dai, X. & Chai, W. Human Stn1 protects telomere integrity by promoting efficient lagging-strand synthesis at telomeres and mediating C-strand fill-in. *Cell Res.* **22,** 1681–1695 (2012).

249. Qi, H. & Zakian, V. A. The S*accharomyces* telomere-binding protein Cdc13p interacts with both the catalytic subunit of DNA polymerase and the telomerase-

associated Est1 protein. *Genes Dev.* **14,** 1777–1788 (2000).

250. Derboven, E., Ekker, H., Kusenda, B., Bulankova, P. & Riha, K. Role of STN1 and DNA polymerase α in telomere stability and genome-wide replication in *Arabidopsis*. *PLoS Genet.* **10,** 1-13 (2014).

251. Barrero, J. M., Gonzalez-Bayon, R., del Pozo, J. C., Ponce, M. R. & Micol, J. L. INCURVATA2 encodes the catalytic ubunit of DNA Ppolymerase and interacts with genes involved in chromatin-mediated cellular memory in *Arabidopsis thaliana*. *Plant Cell Online* **19,** 2822–2838 (2007).

252. Zhong, Z., Shiue, L., Kaplan, S. & de Lange, T. A mammalian factor that binds telomeric TTAGGG repeats in vitro. *Mol. Cell. Biol.* **12,** 4834–4843 (1992).

253. Broccoli, D., Smogorzewska, A., Chong, L. & de Lange, T. Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nat. Genet.* **17,** 231-235 (1997).

254. Baumann, P. & Cech, T. R. Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science.* **292,** 1171–1175 (2001).

255. Van Steensel, B. & De Lange, T. Control of telomere length by the human telomeric protein TRF1. *Letters to Nature* **385,** 740–743 (1997).

256. Ancelin, K. *et al.* Targeting assay to study the cis functions of human telomeric proteins: evidence for inhibition of telomerase by TRF1 and for activation of telomere degradation by TRF2. *Mol. Cell. Biol.* **22,** 3474–3487 (2002).

257. Karlseder, J. *et al.* The telomeric protein TRF2 binds the ATM kinase and can inhibit the ATM-dependent DNA damage response. *PLoS Biol.* **2,** 1150-1156 (2004).

258. Celli, G. B. & de Lange, T. DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. *Nat. Cell Biol.* **7,** 712–718 (2005).

259. Denchi, E. L. & De Lange, T. Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Lett. to Nat.* **448,** 1068–1071 (2007).

260. Amiard, S. *et al.* A topological mechanism for TRF2-enhanced strand invasion. *Nat. Struct. Mol. Biol.* **14,** 147–154 (2007).

261. Baker, A. M., Fu, Q., Hayward, W., Lindsay, S. M. & Fletcher, T. M. The Myb/SANT domain of the telomere-binding protein TRF2 alters chromatin structure. *Nucleic Acids Res.* **37,** 5019–5031 (2009).

262. Smogorzewska, A. *et al.* Control of Human Telomere Length by TRF1 and TRF2. *Mol. Cell. Biol.* **20,** 1659–1668 (2000).

263. Li, B. & Lange, T. de. Rap1 affects thelength and heterogeneity ofhuman telomeres. *Mol. Biol. Cell* **14,** 5060–5068 (2003).

264. O'Connor, M. S., Safari, A., Liu, D., Qin, J. & Songyang, Z. The human Rap1 protein complex and modulation of telomere length. *J. Biol. Chem.* **279,** 28585–28591 (2004).

265. Janoušková, E. *et al.* Human Rap1 modulates TRF2 attraction to telomeric DNA. *Nucleic Acids Res.* **43,** 2691–2700 (2015).

266. Ye, J. Z. S. & De Lange, T. TIN2 is a tankyrase 1 PARP modulator in the TRF1 telomere length control complex. *Nat. Genet.* **36,** 618–623 (2004).

267. Kim, S. H. *et al.* TIN2 mediates functions of TRF2 at human telomeres. *J. Biol. Chem.* **279,** 43799–43804 (2004).

268. Takai, K. K., Kibe, T., Donigian, J. R., Frescas, D. & Lange, T. de. Telomere protection by TPP1/POT1 requires tethering to TIN2. *Mol Cell* **44,** 647–659 (2011).

269. Gong, Y. &de Lange, T. A Shld1-controlled POT1a provides support for repression of ATR signaling at telomeres through RPA exclusion. *Mol. Cell* **40,** 377–387 (2010). doi:10.3174/ajnr.A1256.Functional

270. Loayza, D. & de Lange, T. POT1 as a terminal transducer of TRF1 telomere length control. *Lett. to Nature* **423,** 1013–1018 (2003). doi:10.1038/nature01688

271. Kelleher, C., Kurth, I. & Lingner, J. Human Protection of Telomeres 1 (POT1) is a negative regulator of telomerase activity *in vitro*. *Mol. Cell. Biol.* **25,** 808–818 (2005).

272. Xin, H. *et al.* TPP1 is a homologue of ciliate TEBP-β and interacts with POT1 to recruit telomerase. *Nature* **445,** 559–562 (2007).

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

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

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

276. Murzin, a G. OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.* **12,** 861–867 (1993).

277. Flynn, R. L. & Zou, L. Oligonucleotide/Oligosaccharide-Binding (OB) fold proteins: A growing family of genome guardians. *Crit Rev Bioche* **45,** 266–275 (2010).

278. Arcus, V. OB-fold domains: A snapshot of the evolution of sequence, structure and function. *Curr. Opin. Struct. Biol.* **12,** 794–801 (2002).

279. Theobald, D. L., Mitton-Fry, R. M. & Wuttke, D. S. Nucleic acid recognition by OB-fold proteins. *Annu Rev Biophys Biomol Struct.* **32,** 115–133 (2003).

280. Agrawal, V. & Kishan, K. V. R. OB-fold: growing bigger with functional consistency. *Curr. Protein Pept. Sci.* **4,** 195–206 (2003).

281. Zou, Y., Liu, Y., Wu, X. & Shell, S. M. Functions of human Replication Protein A (RPA): From DNA replication to DNA damage and stress responses. *J Cell Physiol* **208,** 267–273 (2006).

282. Chen, R. & Wold, M. S. Replication Protein A: Single-stranded DNA's first responder : Dynamic DNA-interactions allow Replication Protein A to direct single-strand DNA intermediates into different pathways for synthesis or repair. **36,** 1156–1161 (2015).

283. Wold, M. S. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem.* **66,** 61–92 (1997).

284. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2. *Science.* **289,** 905–920 (2000).

285. Brodersen, D. E., Clemons, W. M., Carter, A. P., Wimberly, B. T. & Ramakrishnan, V. Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus:* Structure of the proteins and their interactions with 16 S RNA. *J. Mol. Biol.* **316,** 725–768 (2002).

286. Wimberly, B. T. *et al.* Structure of the 30S ribosomal subunit. *Nature* **407,** 327–339 (2000).

287. Pellegrini, L. & Venkitaraman, A. Emerging functions of BRCA2 in DNA recombination. *Trends in Biochemical Sciences* **29,** 310–316 (2004).

288. Yang, H. J. BRCA2 function in DNA binding and recombination from a BRCA2[ndash]DSS1-ssDNA structure . *Sci.* **297,** 1837–1848 (2002).

289. Ellenberger, T. & Tomkinson, A. E. Eukaryotic DNA ligases: structural and functional insights. *Annu. Rev. Biochem.* **77,** 313–338 (2008).

290. Froelich-Ammon, S. J., Dickinson, B. A., Bevilacqua, J. M., Schultz, S. C. & Cech, T. R. Modulation of telomerase activity by telomere DNA-binding proteins in Oxytricha. *Genes Dev.* **12,** 1504–1514 (1998).

291. Miyoshi, T., Kanoh, J., Saito, M. & Ishikawa, F. Fission yeast pot1-Tpp1 protects telomeres and regulates telomere length. *Science.* **320,** 1341–1344 (2008).

292. Raices, M. *et al. C. elegans* telomeres contain G-Strand and C-Strand overhangs that are bound by distinct proteins. *Cell* **132,** 745–757 (2008).

293. Theobald, D. L. & Wuttke, D. S. Prediction of multiple tandem OB-fold domains in telomere end-binding proteins Pot1 and Cdc13. *Structure* **12,** 1877–1879 (2004).

294. Hughes, T. R., Weilbaecher, R. G., Walterscheid, M. & Lundblad, V. Identification of the single-strand telomeric DNA binding domain of the Saccharomyces cerevisiae Cdc13 protein. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6457–62 (2000).

295. Mitton-Fry, R. M., Anderson, E. M., Hughes, T. R., Lundblad, V. & Wuttke, D. S. Conserved structure for single-stranded telomeric DNA recognition. *Science.* **296,** 145–147 (2002).

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

297. Loayza, D., Parsons, H., Donigian, J., Hoke, K. & De Lange, T. 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).

298. Lei, M., Podell, E. R., Baumann, P. & Cech, T. R. DNA self-recognition in the structure of Pot1 bound to telomeric single-strand DNA. *Nature* **426,** 198-203 (2003).

299. Trujillo, K. M., Bunch, J. T. & Baumann, P. Extended DNA binding site in Pot1 broadens sequence specificity to allow recognition of heterogeneous fission yeast telomeres. *J. Biol. Chem.* **280,** 9119–9128 (2005).

300. Shakirov, E. V., Song, X., Joseph, J. A. & Shippen, D. E. POT1 proteins in green algae and land plants: DNA-binding properties and evidence of co-evolution with telomeric DNA. *Nucleic Acids Res.* **37,** 7455–7467 (2009).

301. Wikstrom, N., Savolainen, V. & Chase, M. W. Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. B Biol. Sci.* **268,** 2211–2220 (2001).

302. Shakirov, E. V. *et al.* Protection of Telomeres 1 is required for telomere integrity in the moss *Physcomitrella patens*. *Plant Cell* **22,** 1838–1848 (2010).

303. Baumann, P. & Price, C. Pot1 and telomere maintenance. *FEBS Lett.* **584,** 3779–3784 (2010).

304. Baumann, P., Podell, E. & Cech, T. R. Human Pot1 (protection of telomeres) protein: cytolocalization, gene structure, and alternative splicing. *Mol. Cell. Biol.* **22,** 8079–8087 (2002).

305. Nandakumar, J., Podell, E. R. & Cech, T. R. How telomeric protein POT1 avoids RNA to achieve specificity for single-stranded DNA. *Proc. Natl. Acad. Sci.* **107,** 651–656 (2010).

306. Shakirov, E. V., McKnight, T. D. & Shippen, D. E. POT1-independent single-strand telomeric DNA binding activities in *Brassicaceae*. *Plant J.* **58,** 1004–1015 (2009).

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

308. Hockemeyer, D., Daniels, J. P., Takai, H. & de Lange, T. Recent expansion of the telomeric complex in rodents: Two distinct POT1 proteins protect mouse telomeres. *Cell* **126,** 63–77 (2006).

309. Wu, L. *et al.* Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres. *Cell* **126,** 49–62 (2006).

310. Palm, W., Hockemeyer, D., Kibe, T. & de Lange, T. Functional dissection of human and mouse POT1 proteins. *Mol. Cell. Biol.* **29,** 471–82 (2009).

311. He, H. *et al.* POT1b protects telomeres from end-to-end chromosomal fusions and aberrant homologous recombination. *EMBO J.* **25,** 5180–5190 (2006).

312. Shtessel, L. *et al. Caenorhabditis elegans* POT-1 and POT-2 repress telomere maintenance pathways. *Genetics* **3,** 305–313 (2013).

313. Cheng, C., Shtessel, L., Brady, M. M. & Ahmed, S. *Caenorhabditis elegans* POT-2 telomere protein represses a mode of alternative lengthening of telomeres with

normal telomere lengths. *Proc. Natl. Acad. Sci.* **109,** 7805–7810 (2012).

314. Meier, B. *et al.* The MRT-1 nuclease is required for DNA crosslink repair and telomerase activity in vivo in *Caenorhabditis elegans*. *EMBO J.* **28,** 3549–3563 (2009).

315. Naduparambil, J. K., Lescasse, R., Linger, B. R. & Price, C. M. *Tetrahymena* POT1a regulates telomere length and prevents activation of a cell cycle checkpoint. *Mol. Cell. Biol.* **27,** 1592–1601 (2007).

316. Cranert, S., Heyse, S., Linger, B. R., Lescasse, R. & Price, C. *Tetrahymena* Pot2 is a developmentally regulated paralog of Pot1 that localizes to chromosome breakage sites but not to telomeres. *Eukaryot. Cell* **13,** 1519–1529 (2014).

317. Kellogg, E. a. Relationships of cereal crops and other grasses. *Proc. Natl. Acad. Sci.* **95,** 2005–2010 (1998).

318. Paterson, A. H., Bowers, J. E. & Chapman, B. A. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci.* **101,** 9903–9908 (2004).

319. Beilstein, M. A. *et al.* Evolution of the telomere-associated protein POT1a in *Arabidopsis thaliana* is characterized by positive selection to reinforce protein-protein interaction. *Mol. Biol. Evol.* **32,** 1329–1341 (2015).

320. Schranz, M. E., Kantama, L., De Jong, H. & Mitchell-Olds, T. Asexual reproduction in a close relative of *Arabidopsis*: A genetic investigation of apomixis in *Boechera* (Brassicaceae). *New Phytol.* **171,** 425–438 (2006).

321. Shakirov, E. V., Salzberg, S. L., Alam, M. & Shippen, D. E. Analysis of *Carica papaya* telomeres and telomere-associated proteins: insights into the evolution of telomere maintenance in *Brassicales*. *Trop Plant Biol.* **3-4,** 202–215 (2008). doi:10.1109/TMI.2012.2196707.Separate

322. Lei, M., Podell, E. R. & Cech, T. R. 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).

323. Marcand, S., Gilson, E. & Shore, D. A protein-counting mechanism for telomere length regulation in yeast. *Science.* **275,** 986-990 (1997).

324. Ray, a & Runge, K. W. The yeast telomere length counting machinery is sensitive to sequences at the telomere-nontelomere junction. *Mol. Cell. Biol.* **19,** 31–45 (1999).

325. Marcand, S., Brevet, V. & Gilson, E. Progressive cis-inhibition of telomerase upon telomere elongation. *EMBO J.* **18,** 3509–3519 (1999).

326. Levy, D. L. & Blackburn, E. H. Counting of Rif1p and Rif2p on *Saccharomyces cerevisiae* telomeres regulates telomere length. *Mol. Cell. Biol.* **24,** 10857–10867 (2004).

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

328. Mendel, G. Experiments in plant-hybridisation. *Proc. Nat. Hist. Soc. Brünn* **IV,** 3–47 (1865).

329. East, E. M. Studies on size inheritance in *Nicotiana*. *Genetics* **1,** 164–176 (1916).

330. Falconer, D. S. & Mackay, T. F. C. Introduction to quantitative genetics. *Murnt Mil.* **England: Longman,** 4th edition (1996). ISBN: 0582243025

331. Kearsey, M. J. The principles of QTL analysis (aminmal mathematics approach). *J. Exp. Bot.* **49,** 1619–1623 (1998).

332. Lynch, M. & Walsh, B. Genetics and analysis of quantitative traits. *Am. J. Hum. Genet* **68(2),** 548–549 (2001).

333. Miles, C. M. & Wayne, M. Quantitative Trait Locus (QTL) analysis. *Nat. Educ.* **1,** 1–7 (2008).

334. Veeresha, B. A., Rudra Naik, V., Chetti, M. B., Desai, S. A. & Biradar, S. S. QTL mapping in crop plants: Principles and application. *Int. J. Dev. Res.* **5,** 2961–2965 (2015).

335. Darvasi, A. & Soller, M. Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* **141,** 1199–1207 (1995).

336. Talbot, C. J. *et al.* High-resolution mapping of quantitative trait loci in outbred mice. *Nat. Genet.* **21,** 305–308 (1999).

337. Mott, R., Talbot, C. J., Turri, M. G., Collins, A. C. & Flint, J. A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc. Natl. Acad. Sci.* **97,** 12649–12654 (2000).

338. Valdar, W. *et al.* Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat. Genet.* **38,** 879–887 (2006).

339. Macdonald, S. J. & Long, A. D. Joint estimates of quantitative trait locus effect and frequency using synthetic recombinant populations of Drosophila melanogaster. *Genetics* **176,** 1261–1281 (2007).

340. Beyer, P., Morell, M., Mackay, I. & Powell, W. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr. Opin. Plant Biol.* **11,** 215–221 (2008).

341. Kover, P. X. *et al.* A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.* **5,** 1-15 (2009).

342. Dell'Acqua, M. *et al.* Genetic properties of the MAGIC maize population: A new platform for high definition QTL mapping in *Zea mays*. *Genome Biol.* **16,** 1–23 (2015).

343. Bandillo, N. *et al.* Multi-parent advanced generation inter-cross (MAGIC) populations in rice: Progress and potential for genetics research and breeding. *Rice* **6,** 1–15 (2013).

344. Vignal, A., Milan, D., SanCristobal, M. & Eggen, A. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.* **34,** 275–305 (2002).

345. Kordrostami, M. & Rahimi, M. Molecular markers in plants: Concepts and applications. *Genet.* **13,** 4022–4029 (2015).

346. Mammadov, J., Aggarwal, R., Buyyarapu, R. & Kumpatla, S. SNP markers and their impact on plant breeding. *Int. J. Plant Genomics* **2012,** 1-11 (2012).

347. Gupta, P. K. & Rustgi, S. Molecular markers from the transcribed/expressed region of the genome in higher plants. *Funct. Integr. Genomics* **4,** 139–162 (2004).

348. Vos, P. *et al.* A new technique for DNA fingerprinting. *Nucleic Acids Res.* **44,** 388–396 (1995).

349. Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. & Tingey, S. V. DNA polymorphisms amplified by arbitrary primers are usuful as genetic markers. *Nucleic Acids Res.* **18,** 6531–6535 (1990).

350. Bryan, G. J., Stephenson, P., Kirby, A. C. J., Smith, J. B. & Gale, M. D. Low levels of DNA sequence variation among adapted genotypes of hexaploid wheat. *Theor Appl Genet* **99,** 192–198 (1999).

351. Kochert, G. Restriction fragment length polymorphism in plants and its implications. *Subcell Biochem* **17,** 167–190 (1991).

352. Powell, W. *et al.* The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* **2,** 225–238 (1996).

353. Vieira, M. L. C., Santini, L., Diniz, A. L. & Munhoz, C. de F. Microsatellite markers: What they mean and why they are so useful. *Genet. Mol. Biol.* **39,** 312–328 (2016).

354. Bevan, M. *et al.* Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature* **391,** 485–488 (1998).

355. The *Arabidopsis* Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408,** 796–815 (2000).

356. Lin, X. *et al.* Sequence and analysis of chromosome 2 of the plant *Arabidopsis*

thaliana. Nature **402,** 761–8 (1999).

357. Theologis, A. *et al.* Sequence and analysis of chromosome 1 of the plant *Arabidopsis thalian*a. *Nature* **408,** 816–820 (2000).

358. European Union Chromosome 3 *Arabidopsis* Sequence Consortium, T. I. for G. R. & K. D. R. I. et al. Sequence and Analysis of chromosome 3 of the plant *Arabidopsis thaliana*. *Lett. to Nat.* **408,** 820–822 (2000).

359. Kazusa DNA Research Institute, The Cold Spring Harbor and Washington University Sequencing Consortium, The European Union Arabidopsis Genome Sequencing Consortium & Institute of Plant Genetics and Crop Plant Research (IPK). Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana. Nature* **408,** 823–826 (2000).

360. Rhee, S. Y. *et al.* The *Arabidopsis* Information Resource (TAIR): A model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucleic Acids Res.* **31,** 224–228 (2003).

361. Clark, R. M. *et al.* Common Sequence Polymorphisms. *Science.* **317,** 338–342 (2007).

362. Alonso-Blanco, C. *et al.* 1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana*. *Cell* **166,** 481–491 (2016).

363. Fitzgerald, M. S. *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,** 14813–8 (1999).

364. Bunch, J. T., Bae, N. S., Leonardi, J. & Baumann, P. Distinct requirements for Pot1 in limiting telomere length and maintaining chromosome stability. *Mol. Cell. Biol.* **25,** 5567–5578 (2005).

365. Hockemeyer, D., Sfeir, A. J., Shay, J. W., Wright, W. E. & De Lange, T. POT1 protects telomeres from a transient DNA damage response and determines how human chromosomes end. *EMBO J.* **24,** 2667–2678 (2005).

366. Yu, Y. *et al.* POT1 inhibits the efficiency but promotes the fidelity of

nonhomologous end joining at non-telomeric DNA regions. *Aging* **9,** 2529–2543 (2017).

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

368. Bremer, K. Gondwanan evolution of the grass alliance of families (Poales). *Evolution (N. Y)*. **56,** 1374-1387 (2002). doi:10.1111/j.0014-3820.2002.tb01451.x

369. Rossignol, P., Collier, S., Bush, M., Shaw, P. & Doonan, J. H. *Arabidopsis* POT1A interacts with TERT-V(I8), an N-terminal splicing variant of telomerase. *J. Cell Sci.* **120,** 3678–3687 (2007). doi:10.1242/jcs.004119

370. Stroud, H. *et al.* DNA methyltransferases are required to induce heterochromatic re-replication in *Arabidopsis*. *PLoS Genet.* **8,** 1–7 (2012).

371. Stroud, H., Greenberg, M. & Feng, S. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* **152,** 352–364 (2013).

372. Cokus, S. J. *et al.* Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452,** 215–219 (2008).

373. Mosher, R. A., Schwach, F., Studholme, D. & Baulcombe, D. C. PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc. Natl. Acad. Sci.* **105,** 3145–3150 (2008).

374. Zhang, X. *et al.* Genome-wide High-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126,** 1189–1201 (2006).

375. Beilstein, M. A., Nagalingum, N. S., Clements, M. D., Manchester, S. R. & Mathews, S. Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* **107,** 18724–18728 (2010).

376. Cao, J. *et al.* Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat. Genet.* **43,** 956–965 (2011).

377. Pignatta, D. *et al.* Natural epigenetic polymorphisms lead to intraspecific variation in *Arabidopsis* gene imprinting. *Elife* **3,** 1–24 (2014).

378. Zhiponova, M. K. *et al.* Helix-loop-helix/basic helix-loop-helix transcription factor network represses cell elongation in *Arabidopsis* through an apparent incoherent feed-forward loop. *Proc. Natl. Acad. Sci.* **111,** 2824–2829 (2014).

379. Bewick, A. J. *et al.* On the origin and evolutionary consequences of gene body DNA methylation. *Proc. Natl. Acad. Sci.* **113,** 9111–9116 (2016).

380. Molnar, A., Melnyk, C. & Baulcombe, D. C. Silencing signals in plants: A long journey for small RNAs. *Genome Biol.* **12,** 1–8 (2011).

381. Henderson, I. R. & Jacobsen, S. E. Epigenetic inheritance in plants. *Nature* **447,** 418–424 (2007).

382. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* **11,** 204–220 (2010).

383. Lippman, Z. *et al.* Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430,** 471–476 (2004).

384. Ohno, S. Duplication for the sake of producing more of the same. *Evol. by Gene Duplic.* **Springer, Berlin, Heidelberg,** 59–65 (1970). doi:10.1007/978-3-642-86659-3_11

385. Gu, Z. *et al.* Role of duplicate genes in genetic robustness against null mutations. *Nature* **421,** 63-66 (2003). doi:10.1038/nature01198

386. Wu, D. D. *et al.* Out of pollen" hypothesis for origin of new genes in flowering plants: Study from *Arabidopsis thaliana*. *Genome Biol. Evol.* **6,** 2822-2829 (2014). doi:10.1093/gbe/evu206

387. Moore, R. C. & Purugganan, M. D. The early stages of duplicate gene evolution. *Proc. Natl. Acad. Sci.* **100,** 15682-15687 (2003). doi:10.1073/pnas.2535513100

388. Walsh, J. B. How often do duplicated genes evolve new functions? *Genetics* **139,** 421-428 (1995).

389. Lynch, M., O'Hely, M., Walsh, B. & Force, A. The probability of preservation of a newly arisen gene duplicate. *Genetics* **159,** 1789-1804 (2001). doi:10.1006/scdb.1996.0068

390. Lynch, M. & Conery, J. S. The evolutionary demography of duplicate genes. *Journal of Structural and Functional Genomics* **3,** 35-44 (2003). doi:10.1023/A:1022696612931

391. Lynch, M. The Evolutionary fate and consequences of duplicate genes. *Science.* **290,** 1151-1155 (2000). doi:10.1126/science.290.5494.1151

392. Wang, J., Tao, F., Marowsky, N. C. & Fan, C. Evolutionary fates and dynamic functionalization of young duplicate genes in *Arabidopsis* genomes. *Plant Physiol.* **172,** 427–440 (2016).

393. De Bodt, S., Maere, S., 'Van de Peer', Y., Bodt, S. De & Van De Peer, Y. Genome duplication and the origin of angiosperms. *Trends Ecol Evol* **20,** 591-597 (2005). doi:10.1016/j.tree.2005.07.008

394. Wang, Y. *et al.* Modes of gene duplication contribute differently to genetic novelty and redundancy, but show parallels across divergent angiosperms. *PLoS One* **6,** 1-17 (2011).

395. Blanc, G. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell Online* **16,** 1679-1691 (2004). doi:10.1105/tpc.021410

396. Van Lijsebettens, M. *et al.* An S18 ribosomal protein gene copy at the *Arabidopsis* PFL locus affects plant development by its specific expression in meristems. *Embo J* **13,** 3378-3388 (1994). doi:10.1111/j.0022-3646.1985.00072.x

397. Tsugeki, R., Kochieva, E. Z. & Fedoroff, N. V. A transposon insertion in the *Arabidopsis* SSR16 gene causes an embryo-defective lethal mutation. *Plant J.* **10,** 479-489 (1996). doi:10.1046/j.1365-313X.1996.10030479.x

398. Horiguchi, G., Van Lijsebettens, M., Candela, H., Micol, J. L. & Tsukaya, H. Ribosomes and translation in plant developmental control. *Plant Science* **191-192,** 24-34 (2012). doi:10.1016/j.plantsci.2012.04.008

399. Xue, S. & Barna, M. Specialized ribosomes: A new frontier in gene regulation and organismal biology. *Nature Reviews Molecular Cell Biology* **13,** 355-369 (2012). doi:10.1038/nrm3359

400. Yang, L., Takuno, S., Waters, E. R. & Gaut, B. S. Lowly expressed genes in arabidopsis thaliana bear the signature of possible pseudogenization by promoter degradation. *Mol. Biol. Evol.* **28,** 1193-1203 (2011). doi:10.1093/molbev/msq298

401. Cook, D. E. *et al.* The genetic basis of natural variation in *Caenorhabditis elegans* telomere length. *Genetics* **204,** 371-383 (2016). doi:10.1534/genetics.116.191148

402. Gilson, E. & Londoño-Vallejo, A. Telomere length profiles in humans: All ends are not equal. *Cell Cycle* **6,** 2486-2494 (2007). doi:10.4161/cc.6.20.4798

403. McGrath, P. C. *et al.* Proliferation-associated nucleolar antigen P120: a prognostic marker in node-negative breast cancer. *Surgery* **116,** 616-620 (1994).

404. Bolton, W. E. *et al.* Expression of proliferation-associated antigens (PCNA, p120, p145) during the reentry of G0 cells into the cell cycle. *Cytometry* **17,** 66-74 (1994). doi:10.1002/cyto.990170109

405. Kourtidis, A., Ngok, S. P. & Anastasiadis, P. Z. P120 catenin: An essential regulator of cadherin stability, adhesion-induced signaling, and cancer progression. *Prog. Mol. Biol. Transl. Sci.* (2013). doi:10.1016/B978-0-12-394311-8.00018-2

406. Hendley, A. M. *et al.* p120 catenin suppresses basal epithelial cell extrusion in invasive *Pancreatic neoplasia*. *Cancer Res.* **76,** 3351-3363 (2016). doi:10.1158/0008-5472.CAN-15-2268

407. Hong, J., Lee, J. H. & Chung, I. K. Telomerase activates transcription of cyclin D1 gene through the interaction with NOL1. *J. Cell Sci.* **129,** 1566–1579 (2016). doi:10.1242/jcs.181040

408. Shachar, R., Ungar, L., Kupiec, M., Ruppin, E. & Sharan, R. A systems-level approach to mapping the telomere length maintenance gene circuitry. *Mol. Syst. Biol.* **172,** 1-8 (2008). doi:10.1038/msb.2008.13

409. Hong, B., Brockenbrough, J. S., Wu, P. & Aris, J. P. Nop2p is required for pre-rRNA processing and 60S ribosome subunit synthesis in yeast. *Mol. Cell. Biol.* **17,** 378–388 (1997).

410. Woolford, J. L. & Baserga, S. J. Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics* **195,** 643-681 (2013). doi:10.1534/genetics.113.153197

411. Sharma, S., Yang, J., Watzinger, P., Kötter, P. & Entian, K. D. Yeast Nop2 and Rcm1 methylate C2870 and C2278 of the 25S rRNA, respectively. *Nucleic Acids Res.* **41,** 9062-9076(2013). doi:10.1093/nar/gkt679

412. Ranzani, M. *et al.* Lentiviral vector-based insertional mutagenesis identifies genes associated with liver cancer. *Nat. Methods* **10,** 155-161 (2013). doi:10.1038/nmeth.2331

413. Fujikura, U., Horiguchi, G., Ponce, M. R., Micol, J. L. & Tsukaya, H. Coordination of cell proliferation and cell expansion mediated by ribosome-related processes in the leaves of *Arabidopsis thaliana. Plant J.* **59,** 499–508 (2009).

414. Gan, X. *et al.* Multiple reference genomes and transcriptomes for *Arabidopsis thaliana. Nature* **477,** 419–423 (2011).

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

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

417. Mathieu, O. Changes in 5S rDNA chromatin organization and transcription during heterochromatin establishment in *Arabidopsis*. *Plant Cell Online* **15,** 2929-2939 (2003). doi:10.1105/tpc.017467

418. Peglion, F. & Etienne-Manneville, S. p120catenin alteration in cancer and its role in tumour invasion. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368,** 1-13 (2013). doi:10.1098/rstb.2013.0015

419. Starr, T. K. *et al.* A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. *Science.* **323,** 1747-1750 (2009). doi:10.1126/science.1163040

420. Bard-Chapeau, E. A. *et al.* Transposon mutagenesis identifies genes driving hepatocellular carcinoma in a chronic hepatitis B mouse model. *Nat. Genet.* **46,** 24-32 (2014). doi:10.1038/ng.2847

421. Takeda, H. *et al.* Sleeping Beauty transposon mutagenesis identifies genes that cooperate with mutant Smad4 in gastric cancer development. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 2057-2065 (2016). doi:10.1073/pnas.1603223113

422. Moore, R. C. & Purugganan, M. D. The evolutionary dynamics of plant duplicate genes. *Current Opinion in Plant Biology* **8,** 122-128 (2005). doi:10.1016/j.pbi.2004.12.001

423. Codd, V. *et al.* Common variants near TERC are associated with mean telomere length. *Nat. Genet.***42,** 197-199 (2010). doi:10.1038/ng.532

424. Codd, V. *et al.* Identification of seven loci affecting mean telomere length and their association with disease. *Nat. Genet.* **45,** 422-427 (2013). doi:10.1038/ng.2528

425. Bojesen, S. E. *et al.* Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat. Genet.* **45,** 371-384 (2013). doi:10.1038/ng.2566

426. Levy, D. *et al.* Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. *Proc. Natl. Acad. Sci.* **107,** 9293-9298 (2010). doi:10.1073/pnas.0911494107

427. Shindo, C., Bernasconi, G. & Hardtke, C. S. Natural genetic variation in *Arabidopsis*: Tools, traits and prospects for evolutionary ecology. *Annals of Botany* **99,** 1043-1054 (2007). doi:10.1093/aob/mcl281

428. Koornneef, M., Alonso-Blanco, C. & Vreugdenhil, D. Naturally occurring genetic

variation in *Arabidopsis thaliana*. *Annu. Rev. Plant Biol.* **55,** 141-172 (2004). doi:10.1146/annurev.arplant.55.031903.141605

429. Slagboom, P. E., Droog, S. & Boomsma, D. I. Genetic determination of telomere size in humans: a twin study of three age groups. *Am. J. Hum. Genet.* **55,** 876-882 (1994). doi:10.1086/426734

430. Njajou, O. T. *et al.* Telomere length is paternally inherited and is associated with parental lifespan. *Proc. Natl. Acad. Sci.* **104,** 12135-12139 (2007). doi:10.1073/pnas.0702703104

431. Zhu, L. *et al.* Telomere length regulation in mice is linked to a novel chromosome locus. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 8648-8653 (1998). doi:10.1073/pnas.95.15.8648

432. Ding, H. *et al.* Regulation of murine telomere length by Rtel: An essential gene encoding a helicase-like protein. *Cell* **117,** 873-886 (2004). doi:10.1016/j.cell.2004.05.026

433. Kwan, E. X., Foss, E., Kruglyak, L. & Bedalov, A. Natural polymorphism in bul2 links cellular amino acid availability with chronological aging and telomere maintenance in yeast. *PLoS Genet.* **7,** 1-15 (2011). doi:10.1371/journal.pgen.1002250

434. Maillet, G., White, C. I. & Gallego, M. E. Telomere-length regulation in inter-ecotype crosses of *Arabidopsis*. *Plant Mol. Biol.* **62,** 859-866 (2006). doi:10.1007/s11103-006-9061-7

435. Freeman, J. W., Hazlewood, J. E., Auerbach, P. & Busch, H. Optimal loading of scraped HeLa cells with monoclonal antibodies to the proliferation-associated Mr 120,000 nucleolar antigen. *Cancer Res.* **48,** 5246-5250 (1988).

436. Freeman, J. W. *et al.* Prognostic significance of proliferation associated nucleolar antigen P120 in human breast carcinoma. *Cancer Res.* **51,** 1973-1978 (1991).

437. M.V., B. The mitogen-activated protein kinase pathway mediates growth arrest or E1A-dependent apoptosis in SKBr3 human breast cancer cells. *Int. J. Cancer* **78,** 511-517 (1998).

438. Hazlewood, J. *et al.* mRNA levels for human nucleolar protein P120 in tumor and nontumor cells. *Cancer Commun.* **1,** 29-34 (1989).

439. Fonagy, A., Swiderski, C., Dunn, M. & Freeman, J. W. Antisense-mediated specific inhibition of P120 protein expression prevents G1- to S-phase transition. *Cancer Res.* **52,** 5250-5256 (1992).

440. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science.* **185,** 862–864 (1974).

441. Shultz, R. W., Tatineni, V. M., Hanley-Bowdoin, L. & Thompson, W. F. Genome-wide analysis of the core DNA replication machinery in the higher plants *Arabidopsis* and rice. *Plant Physiol.* **144,** 1697-1714 (2007). doi:10.1104/pp.107.101105

442. Ishibashi, T. *et al.* Two types of replication protein A in seed plants. *FEBS J.* **272,** 3270-3281 (2005). doi:10.1111/j.1742-4658.2005.04719.x

443. Osman, K., Sanchez-Moran, E., Mann, S. C., Jones, G. H. & Franklin, F. C. H. Replication protein A (AtRPA1a) is required for class I crossover formation but is dispensable for meiotic DNA break repair. *EMBO J.* **28,** 394-404 (2009). doi:10.1038/emboj.2008.295

444. Burgess, A. L., David, R. & Searle, I. R. Conservation of tRNA and rRNA 5-methylcytosine in the kingdom Plantae. *BMC Plant Biol.* **15,** 1–17 (2015).

445. Karamysheva, Z. N., Surovtseva, Y. V., Vespa, L., Shakirov, E. V. & Shippen, D. E. A C-terminal Myb extension domain defines a novel family of double-strand telomeric DNA-binding proteins in *Arabidopsis*. *J. Biol. Chem.* **279,** 47799–47807 (2004).

446. McKnight, T. D., Fitzgerald, M. S. & Shippen, D. E. Plant telomeres and telomerases. A review. *Biochem.* **62,** 1224–1231 (1997).

447. Bell, R. J. A. *et al.* Understanding TERT promoter mutations: A common path to immortality. *Mol. Cancer Res.* **14,** 315–323 (2016).

448. Reichenbach, P. *et al.* A Human homolog of yeast Est1 associates with telomerase and uncaps chromosome ends when overexpressed. *Curr. Biol.* **13,** 568–574 (2003).

449. Nera, B., Huang, H. S., Lai, T. & Xu, L. Elevated levels of TRF2 induce telomeric ultrafine anaphase bridges and rapid telomere deletions. *Nat. Commun.* **6,** 1-11 (2015).

450. Yogeeswaran, K. *et al.* Comparative genome analyses of *Arabidopsis* spp.: Inferring chromosomal rearrangement events in the evolutionary history of *A. thaliana*. *Genome Res.* **15,** 505–515 (2005). doi:10.1101/gr.3436305.The

451. Lockton, S. & Gaut, B. The evolution of transposable elements in natural populations of self-fertilizing *Arabidopsis thaliana* and its outcrossing relative *Arabidopsis lyrata*. *BMC Evol. Biol.* **10,** 1–11 (2014).

452. Hollister, J. D. & Gaut, B. S. Population and evolutionary dynamics of helitron transposable elements in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **24,** 2515–2524 (2007).

453. Haendeler, J. *et al.* Mitochondrial telomerase reverse transcriptase binds to and protects mitochondrial DNA and function from damage. *Arterioscler. Thromb. Vasc. Biol.* **29,** 929-935 (2009). doi:10.1161/ATVBAHA.109.185546

454. Lee, J. R. *et al.* Heat-shock dependent oligomeric status alters the function of a plant-specific thioredoxin-like protein, AtTDX. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 5978-5983 (2009). doi:10.1073/pnas.0811231106