7-19-2012

*Selaginella moellendorffii* telomeres: conserved and unique features in an ancient land plant lineage

Eugene V. Shakirov  
*Marshall University, shakirov@marshall.edu*

Dorothy E. Shippen

Follow this and additional works at: https://mds.marshall.edu/bio_sciences_faculty

Part of the Ecology and Evolutionary Biology Commons, Genetics and Genomics Commons, and the Molecular Biology Commons

**Recommended Citation**  

This Article is brought to you for free and open access by the Biological Sciences at Marshall Digital Scholar. It has been accepted for inclusion in Biological Sciences Faculty Research by an authorized administrator of Marshall Digital Scholar. For more information, please contact zhangj@marshall.edu, beachgr@marshall.edu.
Telomeres, the essential terminal regions of linear eukaryotic chromosomes, consist of G-rich DNA repeats bound by a plethora of associated proteins. While the general pathways of telomere maintenance are evolutionarily conserved, individual telomere complex components show remarkable variation between eukaryotic lineages and even within closely related species. The recent genome sequencing of the lycophyte Selaginella moellendorffii and the availability of an ever-increasing number of flowering plant genomes provides a unique opportunity to evaluate the molecular and functional evolution of telomere components from the early evolving non-seed plants to the more developmentally advanced angiosperms. Here we analyzed telomere sequence in S. moellendorffii and found it to consist of TTAGGG repeats, typical of most plants. Telomere tracts in S. moellendorffii range from 1 to 5.5 kb, closely resembling Arabidopsis thaliana. We identified several S. moellendorffii genes encoding sequence homologs of proteins involved in telomere maintenance in other organisms, including CST complex components and the telomere-binding proteins, POT1 and the TRFL family. Notable sequence similarities and differences were uncovered among the telomere-related genes in some of the plant lineages. Taken together, the data indicate that comparative analysis of the telomere complex in early diverging land plants such as S. moellendorffii and green algae will yield important insights into the evolution of telomeres and their protein constituents.

Keywords: telomere, Selaginella, POT1, TRF1-L, CST complex
in telomere biology. Our analysis indicates that *S. moellendorffii* harbors short telomere tracts consisting of canonical TTAGGG repeats. Furthermore, we find a full complement of the telomere-associated genes that have previously been described in other plants. Comparative studies of *S. moellendorffii* with other early diverging plants may be useful for studying the evolution of telomere proteins in plants.

RESULTS AND DISCUSSION

**SELAGINELLA MOELLENDORFFII TELOMERES**

Sequence analysis of terminal chromosomal scaffolds indicates that *S. moellendorffii* telomeres, like those of most other plants, are composed of tandem arrays of (TTTAGGG), repeats (Banks et al., 2011). To gauge the size of *S. moellendorffii* telomere tracts, we performed terminal restriction fragment (TRF) analysis using *T. viride* DNA. The blot was hybridized with probe corresponding to four repeats of TTAGGG. As shown in Figure 1A, *S. moellendorffii* telomere tracts migrated as a smear ranging from 1.5 to 5.5 kb, closely resembling telomere profile in many *A. thaliana* accessions (Richards and Ausubel, 1988; Shakirov and Shippen, 2004).

We verified that sequences detected by TRF analysis correspond to chromosome ends using the non-specific Bal31 exonuclease. Bal31 preferentially degrades DNA ends versus more internal genomic regions. DNA was pre-incubated with Bal31 prior to digestion with *T. viride* and a Southern blot was performed. After 15 min of Bal31 digestion, the hybridization products migrated faster on the gel and showed reduced intensity (Figure 1B, lane 2). With continued Bal31 incubation, the telomeric signal disappeared completely (Figure 1B, lanes 3–6). In contrast, several cross-hybridizing bands, corresponding to interstitial telomeric DNA were insensitive to Bal31 digestion for up to 90 min, supporting the conclusion that the Bal31-sensitive hybridization signal corresponds to terminal telomeric DNA. Thus, *S. moellendorffii* telomeres are comprised of 1.5–5.5 kb tracts of TTAGGG repeats.

**TELOMERE-RELATED GENES IN S. MOELLENDORFFII**

**POT1 proteins**

Single-strand (ss) telomere-binding proteins represent a key component of the telomere cap. Such proteins control telomerase access to the telomere and ensure chromosome end protection (de Lange, 2009). Overall, ss telomere binding proteins share limited sequence similarity, but they all bear signature N-terminal oligomannoside/oligosaccharide folds (OMF). One key ss telomere binding protein is Protection of telomeres (POT1; Baumann and Cech, 2001). In the moss *Physcomitrella patens*, a single-copy POT1 gene encodes a typical DNA binding protein that efficiently binds ss telomeric substrates in *vitro* (Shakirov et al., 2010). Furthermore, similar to its mammalian and fusion yeast counterparts, *P. patens* POT1 is involved in telomere end protection (capping). While *Py*POT1-deficient moss can survive long-term in culture, the mutant strain is sterile and shows end-to-end chromosome fusions, indicating that the over-all telomere protective function of POT1 is conserved between early diverging land plants and other eukaryotes (Shakirov et al., 2010).

Despite conservation of POT1 function in the earliest land plant lineages, several lines of biochemical and genetic evidence indicate that the functions of POT1 in vascular plants (starting with *S. moellendorffii*) may have changed substantially. First, biochemical analysis of POT1 proteins from 13 plants representing major evolutionary branches of plants has indicated that the ability to bind telomeric DNA has been lost for most plant POT1 proteins, including POT1 from *S. moellendorffii* (Shakirov et al., 2009a,b). In fact, besides the *P. patens* POT1 protein and its ortholog from the green alga *Ostreococcus lucimarinus*, only two other POT1 proteins (from *Asparagus officinalis* and *Z. mays*) out of a total of 16 surveyed have retained the capacity to bind telomeric DNA *in vitro* (Shakirov et al., 2009b). However, both *A. officinalis* and *Z. mays* are unusual plants with respect to telomere biology. *A. officinalis* possesses unconventional telomere repeats TTAGGG instead of the canonical TTAGGG, while *Z. mays* belongs to the only plant family surveyed other than Brassicaceae that harbors duplicated POT1 genes (Shakirov et al., 2009a,b).
Thus, the ability of POT1 proteins from *A. officinalis* and *Z. mays* to bind telomeric DNA may have been conserved due to unusual changes in organismal telomere biology (*A. officinalis*) or protein sub-functionalization (*Z. mays*). Alternatively, the ability of *A. officinalis* and *Z. mays* POT1 proteins to bind telomeric DNA may have evolved independently through parallel evolution.

The second line of evidence supporting unusually fast evolution of POT1 functions in vascular plants comes from the studies of *A. thaliana*. Unlike the situation in humans and most other organisms, *A. thaliana* and other members of the Brassicaceae family possess two full-length POT1 proteins (Shakirov et al., 2003, 2009b). Genetic and biochemical studies indicate that APOT1a is a positive regulator of telomere length working in the context of telomeres holoenzyme (Sunovtsева, 2007). In contrast, POT1b is implicated in chromosome end protection (Shakirov et al., 2005). Notably, APOT1a has a high binding specificity for the RNA subunit of telomerase (Cifuentes-Rojas et al., 2010), an unexpected mode of action for an OR-fold containing protein originally evolved to bind DNA. Unlike *A. thaliana*, but similar to the situation in *P. patens*, the *S. moellendorffii* genome encodes only a single POT1 protein. As expected from phylogenetic positions of their corresponding species, *S. moellendorffii* POT1 shares more amino acid similarity with *P. patens* POT1 (60%), than with *A. thaliana* POT1 proteins (46% to APOT1a and 47% to APOT1b; Shakirov et al., 2009b). Despite the overall higher amino acid conservation between *P. patens* and *S. moellendorffii* POT1 proteins, the loss of telomeric DNA binding capacity (Shakirov et al., 2009b) clearly suggests that the functional role of POT1 in *S. moellendorffii* telomere biology may in fact be more analogous to the situation in *A. thaliana* than in *P. patens*. Determining whether *P. patens* POT1 binds telomerase RNA must await the identification of this molecule in moss.

**TRF proteins**

The second class of telomeric DNA binding proteins associates with ds telomeric DNA. This family of telomere repeat binding factors (TRF) shares a conserved Myb-related DNA binding domain in the C-terminals and a central dimerization domain (Bilaud et al., 1996). Mammals and other vertebrates encode two ds telomere binding proteins, TRF1 and TRF2, with distinct functions in telomere homeostasis. TRF1 is thought to act primarily in telomere length control, while TRF2 is required for chromosome end protection through participation in T-loop formation (Broccoli et al., 1997; Griffith et al., 1999).

Unlike vertebrates, plants possess two related families of TRF-like (TRFL) proteins, class I and class II (Chen et al., 2001; Hwang et al. 2001; Karamysheva et al., 2004). In *A. thaliana*, there are 12 TRFL proteins, 6 in class I and 6 in class II. Members of class II do not bind ds TTAGGG repeats in vitro. In contrast, all six members of class I specifically bind ds telomeric DNA (Karamysheva et al., 2004). This interaction is dependent on the presence of a unique plant-specific Myb-extension motif, located at the extreme C-terminals of the TRFL protein (Figure 2). Overall, plants display remarkable variation in the number of class I TRFL genes. In dicot species, TRFL gene amplification appears to be a common theme, with three genes in grapes (*Vitis vinifera*), five genes in poplar (*Populus trichocarpa*), and six genes in *A. thaliana* (Table 1). In contrast, sequenced genomes of monocots and non-flowering plants, including *S. moellendorffii*, harbor 2 or 3 TRFL genes. While the precise role of individual TRFL proteins in plants remains unclear, amplification of TRFL gene family may provide a route for sub- and neo-functionalization with the potential for more dynamic control of telomere length or telomerase activity.

We also examined the evolutionary relationships of the available class I full-length plant TRFL proteins, using three *A. thaliana* class II proteins as the outgroup (Figure 3). As expected, class II proteins form a separate clade distinct from class I, consistent with the lack of the C-terminal Myb-extension motif. The evolutionary relationship of class I proteins correlates with the phylogenetic position of the corresponding plant species. Notably, the two TRFL proteins from green algae form a sister clade to all TRFL proteins from land plants, suggesting significant sequence divergence in this ancestral lineage.

**CST components**

A third group of evolutionarily conserved plant telomere proteins is a trimeric complex composed of CTC1, STN1, and TEN1, termed CST (Price et al., 2010). In the budding yeast *Saccharomyces cerevisiae*, a similar complex, composed of Cdc13, Stn1, and Ten1 proteins, was described over 20 years ago, but only recently has this complex come to light in multicellular eukaryotes (Wellinger, 2009). Individual CST components are highly divergent, with only 10–20% amino acid sequence identity between corresponding proteins from different eukaryotic lineages (Price et al., 2010). *A. thaliana* mutants deficient in STN1 and CTC1 are characterized by severe defects in telomere maintenance, massive end-to-end chromosome fusions, and elevated rates of telomere recombination (Song et al., 2008; Sunovtsева, 2009). As in vertebrates (Casted et al., 2009), the *A. thaliana* CTC1 subunit of CST physically interacts with the catalytic subunit of DNA polymerase α (Price et al., 2010), thus linking the CST complex to the telomere replication pathway.

To gain a better understanding of the evolution of CST complex in plants, we looked for the presence of genes encoding CST complex subunits in selected plant species with completely sequenced genomes. Single-copy STN1 and TEN1 genes were found in all organisms surveyed (Table 2). In addition, with the exception of the green alga *O. lucimarinus*, a single copy of CTC1 gene was also identified in all plant species analyzed, including *S. moellendorffii* (Table 2). The apparent absence of a clear CTC1 homolog in *O. lucimarinus* is intriguing. Strikingly, none of the CTC1 orthologs can be readily identified in several species of genera Chlamydomonas, Micromonas, and Chlorella, which belong to evolutionarily distinct lineages of green algae. These data indicate that either CTC1 sequence has diverged beyond recognition or that in green algae this protein has been functionally replaced by an unrelated polypeptide. This observation is in line with the observed sequence divergence of green algae TRFL proteins and suggests that many components of the telomere complex in green algae diverged substantially from their counterparts in land plants.
FIGURE 2 | Multiple alignment of C-terminal regions of plant group ITRFL proteins. Position of conserved plant-specific MYB-extension motif is indicated. Abbreviations: At, Arabidopsis thaliana; Cp, Carica papaya; Pt, Populus trichocarpa; Vv, Vitis vinifera; Os, Oryza sativa; Sb, Sorghum bicolor; Bd, Brachypodium distachyon; Pp, Physcomitrella patens; Sm, Selaginella moellendorffii; Ol, Ostreococcus lucimarinus. Accession IDs: AtTBP1, NP_196886; AtTRP1, NP_200751; AtTRFL1, NP_190243; AtTRFL2, NP_172234; AtTRFL4, NP_190947; AtTRFL9, NP_187862; CpTRFL1, EU909205; CpTRFL2, EU909206; PtTRFL1, XM_002316243; PtTRFL2, XM_002311318; PtTRFL4, XM_002299832; PtTRFL5, XM_002308126; VvTRFL1, CBI30542; VvTRFL2, EU909205; VvTRFL3, EU909206; OsTBP1, AF242298.1; OsTBP2, ABF95241; SmTRFL1, XP_00246682; SmTRFL2, XP_00246682; SmTRFL3, XP_002979224.1; SmTRFL4, XP_002984862.1; OlTRFL1, XP_001421395.1; OlTRFL2, XP_001416857.1. Protein alignment was generated using MEGA5 software (Tamura et al., 2011) and visualized in the BOXSHADE format.

Table 1 | Putative TRFL genes in selected plants with sequenced genomes.

<table>
<thead>
<tr>
<th>Species</th>
<th>Plant lineage</th>
<th>Number of TRFL genes</th>
<th>Sequence IDs</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitis vinifera</td>
<td>Dicot</td>
<td>3</td>
<td>CB130542, CB186113, CB131661</td>
<td></td>
</tr>
<tr>
<td>Populus trichocarpa</td>
<td>Dicot</td>
<td>5</td>
<td>XM_002316243, XM_002311138, XP_002313432, XM_002299832, XM_0022368126</td>
<td>Hwang et al. (2001), Chen et al. (2001)</td>
</tr>
<tr>
<td>Carica papaya</td>
<td>Dicot</td>
<td>2</td>
<td>EU090305, EU090306</td>
<td>Shakirov et al. (2008)</td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>Dicot</td>
<td>6</td>
<td>NP_196886, NP_200751, NP_190243, NP_172234, NP_190947, NP_187862, XP_002313432</td>
<td>Hwang et al. (2001), Chen et al. (2004)</td>
</tr>
<tr>
<td>Oryza sativa</td>
<td>Monocot</td>
<td>2</td>
<td>AF242298.1, ABF95241</td>
<td>Yu et al. (2000)</td>
</tr>
<tr>
<td>Brachypodium distachyon</td>
<td>Monocot</td>
<td>3</td>
<td>XP_003565159, XP_003565159, XP_003565159</td>
<td></td>
</tr>
<tr>
<td>Sorghum bicolor</td>
<td>Monocot</td>
<td>2</td>
<td>XP_00246682, XP_00246682, XP_00246682</td>
<td></td>
</tr>
<tr>
<td>Selaginella moellendorffii</td>
<td>Monocot</td>
<td>3</td>
<td>XP_002679234, XP_002679234, XP_002679234</td>
<td></td>
</tr>
<tr>
<td>Physcomitrella patens</td>
<td>Bryophyte</td>
<td>2</td>
<td>BP_00177300, BP_001767955</td>
<td></td>
</tr>
<tr>
<td>Ostreococcus lucimarinus</td>
<td>Green alga</td>
<td>2</td>
<td>BP_001421395, BP_001416857</td>
<td></td>
</tr>
</tbody>
</table>
Interestingly, it has been argued that the budding yeast *S. cerevisiae* replaced CTC1 with Cdc13 (Mitton-Fry et al., 2002). The two proteins share little sequence similarity, but possess structurally similar OB-fold DNA binding domains and interact with well-conserved protein binding partners (STN1 and TEN1). Our genome analysis indicates that single-copy genes encoding CST complex subunits are present in all land plants analyzed, from the earlier evolved non-seed plant lineages, represented by *P. patens* and *S. moellendorffii*, to the more developmentally advanced flowering plants. Thus, the important functions of CST complex in chromosome end protection and/or telomere replication are likely to be conserved throughout evolution of land plants.

**CONCLUSION AND OUTLOOK**

Lycophytes occupy a unique phylogenetic position in the evolution of land plants, as they are ancient representatives of vascular plants and sister to Euphylllophytes (which include flowering plants). Since most components of the telomere maintenance machinery have previously been analyzed only in Angiosperms, we examined the telomere repeat array and sequence homologs of telomere-related factors in *S. moellendorffii*. A sin A. thaliana

<table>
<thead>
<tr>
<th>Species</th>
<th>STN1</th>
<th>TEN1</th>
<th>CTC1</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Vitis vinifera</em></td>
<td>XP_003632321</td>
<td>CB_039529</td>
<td>CAN_0783971</td>
</tr>
<tr>
<td><em>Papaya</em></td>
<td>EEE94597</td>
<td>EEF08821</td>
<td>XM_002302411</td>
</tr>
<tr>
<td><em>Carica papaya</em></td>
<td>JX198688</td>
<td>JX198687</td>
<td>JX198685</td>
</tr>
<tr>
<td><em>Arabidopsis</em></td>
<td>NP_0597881.1</td>
<td>NP_175022.2</td>
<td>NP_00119860.1</td>
</tr>
<tr>
<td><em>thaliana</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Oryza sativa</em></td>
<td>EEE95102</td>
<td>EAZ38764.1</td>
<td>EEE69001.1</td>
</tr>
<tr>
<td><em>Brachypodium</em></td>
<td>XP_00357913</td>
<td>JX198686</td>
<td>XP_00357480.1</td>
</tr>
<tr>
<td><em>distachyon</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Sorghum bicolor</em></td>
<td>EER02221.1</td>
<td>EER95890.1</td>
<td>XP_002463031.1</td>
</tr>
<tr>
<td><em>Selaginella</em></td>
<td>E512876</td>
<td>E526886</td>
<td>E02963831.1</td>
</tr>
<tr>
<td><em>moellendorffii</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Physcomitrella</em></td>
<td>E0074536</td>
<td>XP_001782219.1</td>
<td>EDO49834</td>
</tr>
<tr>
<td><em>patens</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ostreococcus</em></td>
<td>ABO95476</td>
<td>ABO98342</td>
<td>ND*</td>
</tr>
<tr>
<td><em>lucimarinus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Ostreococcus lucimarinus* CTC1 sequence could not be discerned.

and *P. patens*, telomere tracts are short and consist of the canonical TTAGGG repeat. Furthermore, the structure of the tandemly repeated CTTAGGG repeat is characterized by TTAGG sequence elements (Coral and Maller, 1997). In addition, the importance of the TTAGGG repeat is further strengthened by the presence of a conserved DNA-binding domain, the RRE (recognition sequence for telomeric DNA) domain (Murphy et al., 1996). The TTAGGG repeat is also characteristic of the telomere region in other lower eukaryotes, such as fungi and yeasts. This suggests that the TTAGGG repeat is a key component in the telomere region in all eukaryotes.

**MATERIALS AND METHODS**

**TELOMERE LENGTH ANALYSIS AND Bal31 DIGESTION**

*S. moellendorffii* DNA was extracted as described by Coccioldone and Cone (1993). To detect telomeric DNA repeats, genomic DNA was digested with *Bal*31 (Fermentas; recognition sequence TTAGGG) and subjected to Southern blotting with 

\[ \text{[TTAGGG]}_{n} \]

probe (Fitzgerald et al., 1999). Radiolabeled signals were scanned by a STORM PhosphorImager (Molecular Dynamics), and the data were analyzed by IMAGEQUANT software (Molecular Dynamics). For the *Bal*31 exonuclease assay, 100 μg of *S. moellendorffii* genomic DNA was incubated with 50 units of *Bal*31 (New England Biolabs) in 1× *Bal*31 reaction buffer at 30°C. Equal measures of sample were removed at 15 or 30 min intervals for 90 min. Reactions were stopped by the addition of 20 mM EDTA and heating to 65°C for 15 min. DNA in each sample was precipitated with isopropanol and ammonium acetate, followed by *Bal*31 digestion. Digested DNA was separated on 0.8% agarose, blotted onto a nitrocellulose membrane and subjected to hybridization as described above.

www.frontlin.org
The telobox, a Myb-related telomeric DNA binding motif found in proteins from yeast, plants and human.

Nuclear Acid Res 24, 1294–1303.


A plant cDNA encoding a Myb-like protein that binds telomeric DNA and POT1s are components of Arabidopsis telomeres. Proc. Natl Acad. Sci. USA 103, 7455–7460.

Subject to an autoregulatory gene of murine leukemia virus that leads to rearrangements in provirus genomes. Cell 135, 575–586.


A yeast Tel1/Rad53 homologue, which interacts with the telomerase catalytic subunit gene and telomeres in the absence of telomerase activity in Arabidopsis thaliana. Plant Cell 13, 1717–1727.


Arabidopsis thaliana. 58, 1004–1015.


POT1 associates with the telomerase RNP and is required for telomere maintenance. EMBO J. 26, 3653–3661.


Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 May 2012; accepted: 29 June 2012; published online: 19 July 2012.


This article was submitted to Frontiers in Plant Evolution and Development, a specialty of Frontiers in Plant Science.

Copyright © 2012 Shakirov and Shippen. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and any copyright notices concerning any third-party graphics etc.

ADDENDUM