Chromosome Res (2014) 22:153–166
DOI 10.1007/s10577-014-9423-y

RESEARCH ARTICLE

Extending the model of Arabidopsis telomere length and composition across Brassicaceae

Andrew D. L. Nelson · Evan S. Forsythe · Xiangchao Gan · Miltos Tsiantis · Mark A. Beilstein

Published online: 21 May 2014

Abstract Telomeres are repetitive TG-rich DNA elements essential for maintaining the stability of genomes and replicative capacity of cells in almost all eukaryotes. Most of what is known about telomeres in plants comes from the angiosperm *Arabidopsis thaliana*, which has become an important comparative model for telomere biology. Arabidopsis tolerates numerous insults to its genome, many of which are catastrophic or lethal in other eukaryotic systems such as yeast and vertebrates. Despite the importance of Arabidopsis in establishing a model for the structure and regulation of plant telomeres, only a handful of studies have used this information to assay components of telomeres from across land plants, or even among the closest relatives of Arabidopsis in the plant family Brassicaceae. Here, we determined how well Arabidopsis represents Brassicaceae by comparing multiple aspects of telomere biology in species that represent major clades in the family tree. Specifically, we determined the telomeric repeat sequence, measured bulk telomere length, and analyzed variation in telomere length on syntenic chromosome arms. In addition, we used a phylogenetic approach to infer the evolutionary history of putative telomere-binding proteins, CTC1, STN1, TEN1 (CST), telomere repeat-binding factor like (TRFL), and single Myb histone (SMH). Our analyses revealed conservation of the telomeric DNA repeat sequence, but considerable variation in telomere length among the sampled species, even in comparisons of syntenic chromosome arms. We also found that the single-stranded and double-stranded telomeric DNA-binding complexes CST and TRFL, respectively, differ in their pattern of gene duplication and loss. The TRFL and SMH gene families have undergone numerous duplication events, and these duplicate copies are often retained in the genome. In contrast, CST components occur as single-copy genes in all sampled genomes, even in species that experienced recent whole genome duplication events. Taken together, our results place the Arabidopsis model in the context of other species in Brassicaceae, making the family the best characterized plant group in regard to telomere architecture.

Keywords Telomere · Brassicaceae · Telomere repeat · Telomere length · Telomere-binding proteins · CST · TRFL

Responsible Editor: Martin Lysak and Paul Fransz.

Electronic supplementary material The online version of this article (doi:10.1007/s10577-014-9423-y) contains supplementary material, which is available to authorized users.

A. D. L. Nelson · E. S. Forsythe · M. A. Beilstein (✉)
School of Plant Sciences, University of Arizona,
Tucson, AZ 85721, USA
e-mail: mbeilstein@email.arizona.edu

X. Gan · M. Tsiantis
Department of Comparative Development and Genetics, Max Planck Institute for Plant Breeding Research,
Carl-von-Linné-Weg 10, 50829 Cologne, Germany

Abbreviations

BLAST	Basic local alignment search tool
CDS	Coding sequences

CST	CTC1, STN1, TEN1
CTC1	Conserved telomere maintenance component 1
DS	Double stranded
DSB	Double-strand break
GCR	Gross chromosomal rearrangement
PETRA	Primer extension telomere repeat amplification
SS	Single stranded
TBP1	Telomere-binding protein 1
TRF	Terminal restriction fragment
TRFL	Telomere repeat-binding factor like
WGD	Whole genome duplication

Introduction

Linear chromosomes present two unique challenges for plant and other eukaryotic genomes. First, chromosome ends must be hidden from the DNA repair machinery that would confuse them for a DNA break. Second, a mechanism must exist for the complete replication of the ends of chromosomes, a feat impossible via normal semi-conservative DNA replication. In almost all eukaryotes, these issues are resolved by telomeres. Telomeric DNA repeat arrays and associated proteins form a highly dynamic complex that prevents chromosome termini from being recognized as a double-strand break. Telomere associated proteins also serve as gatekeepers to regulate the length of the telomere tract by modulating access of conventional DNA replication machinery and telomerase, the enzyme responsible for elongating telomere tracts at the chromosome terminus. Together, telomeres, telomere-binding proteins, and telomerase promote genome integrity and complete transfer of genetic material to offspring.

Given the essential nature and deep conservation of telomeres across eukaryotes, it is no surprise that the manner in which Arabidopsis protects and maintains its chromosome ends is, in many ways, similar to strategies reported in vertebrates and yeast. Aspects of telomere architecture, telomere protein composition, telomere extension, and replication are conserved between Arabidopsis and other model systems (reviewed in Nelson and Shippen 2012; Watson and Riha 2010). Nevertheless, several exciting and interesting variations in telomere maintenance have been reported for Arabidopsis. For example, in contrast to other eukaryotic model organisms, chromosomes in Arabidopsis can be blunt ended (Kazda et al. 2012). There is also evidence that some Arabidopsis telomere components, while being conserved in structure, serve functions different from those reported in other model systems (Shakirov et al. 2005; Surovtseva et al. 2007). In addition, telomerase-interacting RNAs have been found to regulate telomerase activity during DNA damage (Cifuentes-Rojas et al. 2012), a finding currently unique to Arabidopsis. Moreover, research on Arabidopsis suggests that it displays an extraordinary tolerance to genome instability (Riha et al. 2001) and a capacity to survive even in the absence of core telomere components shown to be crucial in other eukaryotes (Surovtseva et al. 2009). While these features of Arabidopsis telomere biology have often been hypothesized to occur in other plants, only a few studies have focused on assaying specific components of telomeres from across land plants or even among the closest relatives of Arabidopsis in the plant family Brassicaceae (Shakirov et al. 2008, 2009, 2010).

In Arabidopsis, the telomere is composed of TTTA GGG repeats (Richards and Ausubel 1988). This repeat, first observed in Arabidopsis, is found across land plants, including the moss *Physcomitrella patens* (Table 1; Shakirov et al. 2010). Interestingly, the only known exceptions among sampled land plants are species in the order Asparagales, which share the human telomeric repeat (TTAGGG) (Sykorova et al. 2003; Weiss-Schneeweiss et al. 2004). Unlike the telomeric repeat sequence, telomere length varies. Arabidopsis telomeres range from 1.5 to 9 Kb, depending on the ecotype (Shakirov and Shippen 2004). Among land plants, telomeres ranging from 0.5 to 3.5 Kb in *P. patens* and up to 160 Kb in *Nicotiana tabacum* have been observed (Table 1; Shakirov et al. 2010; Fajkus et al. 1995).

The protein complexes required for telomere end protection consist of both single-strand (SS) and double-strand (DS) DNA-binding proteins. These form a proteinaceous barrier against DNA repair machinery and exonucleolytic degradation of telomeric DNA. Two major telomere-specific-binding complexes have been described, shelterin and CST (de Lange 2009; Giraud-Panis et al. 2010). The six-membered shelterin complex was originally discovered in vertebrates, although various components have also been identified in yeast and plants (Palm and de Lange 2008; Watson and Riha 2010). Vertebrate shelterin consists of TRF1, TRF2, RAP1, TIN2, TPP1, and POT1. In Arabidopsis, only TRF-like proteins (TRFLs) and POT1 orthologs are present.

Table 1 Telomere length variation and sequence conservation

	Organism	Telomere Repeat	Telomere length (Kb)	Lifespan	Reference
Brassicaceae	*Arabidopsis thaliana*	TTTAGGG	1.5–9	Annual	Figure 1, Shakirov and Shippen 2004; Richards and Ausubel 1988
	Camelina hispida	NA	2–7	Annual	Figure 1
	Capsella rubella	TTTAGGG	.85–2.5	Annual	Figure 1
	Cardamine hirsuta	TTTAGGG	1–3	Annual	Figure 1
	Cardamine cordifolia	NA	2–7	Biennial	Figure 1
	Eutrema salsugineum	TTTAGGG	.85–2.5	Annual	Figure 1
	Schrenkiella parvula	TTTAGGG	2.5–6	Annual	Figure 1
	Brassica oleracea	TTTAGGG	2–7	Annual	Figure 1, Shakirov et al. 2008
	Brassica rapa	TTTAGGG	1.5–5	Annual	Figure 1
	Aethionema arabicum	TTTAGGG	2–5.5	Perennial	Figure 1
	Carica papaya	TTTAGGG	25–50+	Perennial	Shakirov et al. 2008
	Nicotiana tabacum	TTTAGGG	40–160	Annual	Fajkus et al. 1995
	Pisum sativum	TTTAGGG	10–50	Annual	Cesare et al. 2003
	Zea mays	TTTAGGG	2–40	Annual	Richards and Ausubel 1988
	Oryza sativa	TTTAGGG	5–11	Annual	Mizuno et al. 2006
	Othocallis siberica	TTAGGG	>10	Biennial	Weiss-Schneeweiss et al. 2004
	Asparagales	TTAGGG/TTTAGGG	NA	NA	Adams et al. 2001; Sýkorová et al. 2003
	Pinus palustris	TTTAGGG	0.5–30	Perennial[a]	Flanary and Kletetschka 2005
	Pinus longaeva	TTTAGGG	2–25	Perennial[b]	Flanary and Kletetschka 2005
	Selaginella moellendorffii	TTTAGGG	1–5.5	Perennial	Shakirov and Shippen 2012
	Selaginella martensii	TTTAGGG	NA	Perennial	Fuchs and Schubert 1996
	Physcomitrella patens	TTTAGGG	0.5–3.5	NA	Shakirov et al. 2010

Telomere sequence and average length are shown from a variety of plant species. Telomere repeat sequences were obtained from genomic data or, where possible, from previously published work. TTTAGGG is the dominant repeat sequence. NA implies either data unavailable, as in the case of *C. hispida* and *C. cordifolia*, or that the dataset is represented by a large number of species, as in the case of Aspargales

[a] "Medium-lived" perennial, with an average life span of 100–200 years

[b] "Long-lived" perennial with an average life span of 2,000–5,000 years

Plant TRFL proteins contain a Myb-related DNA-binding domain similar to the human DS telomere-binding proteins TRF1 and TRF2. TRFLs from various land plant lineages display a binding specificity towards DS telomeric DNA (Karamysheva et al. 2004; Shakirov and Shippen 2012). Arabidopsis encodes six TRFLs that specifically bind DS telomeric DNA in vitro and a sister group of six TRFLs that lack the motif necessary to bind telomeric DNA (Chen et al. 2001; Karamysheva et al. 2004). The six DS telomeric DNA-binding TRFLs contain a plant-specific domain called the Myb extension that is critical for telomere-binding specificity (Sue et al. 2006). Possibly due to redundancy of function in Arabidopsis TRFLs, it has been difficult to pinpoint the specific role of any single gene. However, Arabidopsis deficient for AtTBP1, one of the DS telomere-binding TRFLs, have elongated telomeres, consistent with a loss in telomere length regulation, a phenotype shared with vertebrate TRF mutants (Hwang and Cho 2007).

POT1 is the other Arabidopsis shelterin-like component identified by sequence similarity with its vertebrate counterparts. In vertebrates and in fission yeast, POT1 is an essential SS telomeric DNA-binding protein (Baumann and Cech 2001; reviewed in de Lange 2009). It binds and masks the single-stranded terminus of the telomere, termed the G-overhang, from DNA damage repair mechanisms. In addition, it regulates telomere extension by telomerase (Colgin et al. 2003; Wang et al. 2007). An examination of POT1 in *P. patens* indicates that it performs a very similar chromosome end-protection function to that of its orthologs in vertebrates and fission yeast (Shakirov et al. 2010).

POT1 appears to have undergone a dramatic shift in function during the evolution of land plants. For example, neither of the two reported POT1 proteins in *Arabidopsis thaliana* (AtPOT1a and AtPOT1b) bind telomeric DNA. Moreover, a survey of representative species from across land plant lineages by Shakirov et al. (2009) identified DNA binding by a POT1 ortholog in only three of the 11 species tested. In Arabidopsis, AtPOT1a co-immunoprecipitates with telomerase activity and has a profound impact on telomerase regulation (Surovtseva et al. 2007). In addition, *Atpot1a* null

mutants phenocopy mutants null for the telomerase reverse transcriptase (TERT) component of telomerase; the phenotype of either mutation is progressively shorter telomeres over subsequent generations (Surovtseva et al. 2007). Overexpression of the N-terminus of AtPOT1b results in rapid telomere shortening and chromosome end de-protection (Shakirov et al. 2005). Whether AtPOT1b plays a role in telomere length maintenance remains an open line of inquiry.

There are a variety of other proteins that have been shown to bind DS telomeric DNA in Arabidopsis. Among these is a family of plant-specific proteins harboring a single Myb histone (SMH; Schrumpfova et al. 2004; Marian et al. 2003) domain. Referred to interchangeably as both SMH proteins and telomere repeat-binding (TRB) factors, an in vivo telomere protection phenotype has not been established for this five-membered protein family. However, SMH family members demonstrate telomere-sequence specificity in vitro (Mozgova et al. 2008; Schrumpfova et al. 2004).

The second major telomere complex, CST, is a heterotrimer consisting of CTC1/STN1/TEN1 in plants and vertebrates (Giraud-Panis et al. 2010). Although CST was only recently reported in plants and vertebrates, two of its components, STN1 and TEN1, are orthologous to the predominant telomere-binding complex found in budding yeast, while the third member of the heterotrimer is analogous to Cdc13 of budding yeast (Price et al. 2010). Vertebrate and plant CST is believed to bind the g-overhang, where it is necessary for chromosome end protection and telomere replication (Price et al. 2010). Complex integrity is crucial for the role of CST in protecting the chromosome end; loss of any member results in rapid telomere shortening and chromosomal end-to-end fusions (Song et al. 2008; Surovtseva et al. 2009; Leehy et al. 2013; reviewed in Nelson and Shippen 2012). Ultimately, telomeric binding proteins work hand in hand with telomerase and DNA replication machinery to facilitate the complete replication of telomeres.

The recent accumulation of whole genome sequence for various members of the plant family Brassicaceae permits a more thorough evaluation of the Arabidopsis model of telomere architecture. Brassicaceae contains species that have undergone relatively recent whole genome duplication (WGD) events, as in *Leavenworthia alabamica* and *Brassica rapa* as well as species that have experienced massive genome rearrangements, such as *A. thaliana* (Haudry et al. 2013; Cheng et al. 2013). We adapted assays developed in Arabidopsis to other species of Brassicaceae to better understand whether these large-scale genomic differences are correlated with changes in telomere length and/or protein composition. Using these assays, we determined the telomeric repeat sequence and telomere length from representative species across the family. In addition, we identified components of both the DS and SS telomeric DNA-binding complexes from all available genomic data for species in Brassicaceae. We inferred phylogeny for the TRFL and SMH proteins along with components of the CST complex. Here, we present the most comprehensive view of conservation and variation in telomere length and telomere-binding proteins in Brassicaceae, effectively extending and refining the Arabidopsis model of plant telomere biology. Finally, there are a variety of other proteins that function at the chromosome end, such as those involved in the sensing and repair of DNA damage. These have been reviewed extensively elsewhere and thus are not the focus of this paper (Watson and Riha 2010).

Materials and methods

Plant material and DNA extraction

A. thaliana (Col-0) seeds were a gift from Dr. Dorothy Shippen, Texas A&M University, while *Arabidopsis lyrata* and *Arabidopsis arenosa* tissue was received from Dr. Ravi Palanivelu (University of Arizona). *Capsella rubella* was donated by Dr. Steven Wright, University of Toronto. *Cardamine hirsuta* seeds were obtained from Dr. Angela Hay, Max Planck Institute for Plant Breeding Research. *Eutrema salsugineum* seeds were a gift from Dr. Karen Schumaker, University of Arizona; Drs. Maheshi Dassanayake and Dong-Ha Oh (Louisiana State University) donated *Schrenkiella parvula* seeds. *Cardamine cordifolia* flower and leaf tissue was obtained from Drs Noah Whiteman and Anna Nelson-Dittrich (University of Arizona). *B. rapa* and *Brassica oleracea* tissue were given by Dr. Rebecca Mosher (University of Arizona). *Camelina sativa* (accession 18034) are available from the Beilstein lab on request. *Aethionema arabicum* seeds (accession no. 309) were obtained from the Kew Millenium Seed Bank project (Kew Royal Botanical Gardens). Standard Arabidopsis

growth conditions were used for all species. DNA was extracted from at least three individuals for each species at similar growth stages according to the CTAB method described previously (Weigel and Glazebrook 2002).

Analysis of the telomeric repeat sequence and telomere length

Bulk telomere length was measured using the terminal restriction fragment (TRF) length analysis. Genomic DNA isolated from each species was digested with the restriction enzyme Tru1I (ThermoFisher). Following digestion and ethanol precipitation, samples were separated on a 1 % agarose gel, followed by Southern blot using a [^{32}P]-radiolabeled $(T_3AG_3)_4$ oligonucleotide probe. To measure length differences among homologous chromosome arms from species in the family, we adapted the primer extension telomere repeat amplification (PETRA) assay, first developed in Arabidopsis (Heacock et al. 2004), to six other Brassicaceae species. PETRA was performed as described in Heacock et al. (2004) with modifications to extension time depending on expected length of PETRA products—e.g., for expected products longer than 6 kb, extension time was increased to 4 min. We chose target chromosome arms for PETRA using the karyotype analyses of Mandakova and Lysak (2008) and Lysak et al. (2010). Subtelomeric primers were then designed based on micro-synteny using CoGe (Lyons et al. 2008) and BLAST (Altschul et al. 1990). Primers are listed in Online Resource 1. For *C. hirsuta* and *A. arabicum*, BLAST was performed to identify genes found adjacent to telomere regions in Arabidopsis, and then the scaffold was scanned for the presence of telomeric repeats as an indication that these genes were located in the subtelomeric region. PETRA products were cloned into PGEM-T Easy (Promega) and sequenced to verify primer binding site and presence of the telomeric repeat. The distance from the telomeric repeat to the primer binding site was subtracted from the overall length to calculate the length of the telomere.

Alignment and phylogenetic analysis

CTC1, STN1, TEN1, SMH, and TRFL coding DNA and protein sequence was gathered from publicly available sequences through TAIR (Lamesch et al. 2012), NCBI, Phytozome (Goodstein et al. 2012), and CoGe (Lyons et al. 2008) using a combination of BLASTp and tBLASTn (Gish and States 1993). Accession numbers for CST components are listed in Online Resource 2. TRFL accession numbers appear in Online Resource 3. For *A. arabicum*, *L. alabamica*, *Sisymbrium irio*, and *Tarenaya hassleriana*, no open-reading frame annotations were available, and thus, splice sites were inferred in silico based on alignment with previously annotated orthologs. CTC1 and STN1 amino acid alignments as well as the Ten1 CDS nucleotide alignment were inferred using the multiple alignment tool under default parameters in the program Geneious v6 (Biomatters, Auckland, New Zealand). TRFL genes were aligned by amino acid using ClustalW (Larkin et al. 2007) under default parameters. All sites in the alignment represented by fewer than 50 % of the taxa were trimmed from the alignment. Ten1 phylogeny was inferred from aligned nucleotide data in RAxML v7.04 (Stamatakis and Alachiotis 2010) using a general time reversible (GTR) model with gamma distributed rate heterogeneity. CTC1, STN1, and TRFL phylogenies were inferred from amino acid alignments using the WAG model of amino acid transitions (Whelan and Goldman 2001) with gamma distributed rate heterogeneity in RAxML. In all cases, bootstrap support was calculated by inferring phylogeny from 100 bootstrap datasets in RAxML. All alignments are available from the authors upon request.

Results

Telomeric repeat and length variation in Brassicaceae

The conservation of the Arabidopsis telomeric repeat (TTTAGGG) throughout Brassicaceae was determined by analyzing publically available genomic data, sequencing Telomere Repeat Amplification Protocol (TRAP) products and compiling these data with previously published results (Table 1). The telomeric repeat observed in Arabidopsis 25 years ago (TTTAGGG; Richards and Ausubel 1988) is conserved within the sampled species of the family and throughout many distant lineages of land plants, with the exception of the order Asparagales (Adams et al. 2001; Sykorova et al. 2003). Given its widespread incidence among land plants, the Arabidopsis telomeric repeat appears to be characteristic of plants more generally.

To determine if Arabidopsis telomere length is representative of other Brassicaceae species, we analyzed bulk telomere length using TRF length analysis on 12

other species within the family. The species were chosen because they represent the phylogenetic diversity within the family (Fig. 1), including the earliest diverging lineage, *A. arabicum*. Although telomere length varies within the family (Fig. 2), all species analyzed harbored telomeres close in length to those found in Arabidopsis. *A. lyrata* telomeres were the longest, ranging from 3 to 10 kb, while *E. salsugineum* and *C. rubella* were the shortest with telomeres ranging from 1 to 2.5 kb (Fig. 2). We observed no significant difference in telomere length based on recent genome duplication events (*B. rapa* vs. *S. parvula*; Figs. 1 and 2) or genome rearrangements (*A. thaliana* vs *A. lyrata*; Figs. 1 and 2). In addition, *A. arabicum*, a perennial, also harbors telomeres within the Arabidopsis range, although they may be longer than observed by TRF (described below). Next, we analyzed the impact of gross chromosomal rearrangement (GCR) on telomere length for specific chromosome arms. Brassicaceae has undergone multiple such events, as evidenced by chromosome number and genome organization (Figs. 1 and 3a). Arabidopsis has a greatly reduced genome, with five homologous chromosome pairs. This is in contrast to its close relatives *A. lyrata* and *C. rubella*, both of which contain eight chromosome pairs (Lysak et al. 2010). To better understand the impact of these GCRs on telomere length, we asked whether telomere length was correlated with GCR (Figs. 1 and 3a). Using available genome sequences, we designed subtelomeric primers for four chromosome arms that remain syntenic throughout Brassicaceae. For each new set of primers, PETRA products were cloned and sequenced to verify that the subtelomeric primer bound to the desired locus and that the expected product was amplified. A few primer sets failed repeatedly and therefore, these arms were not included in the final data set.

PETRA products correlated with TRF results, with the exception of *A. arabicum*. In *A. arabicum*, PETRA products were 2–3-kb longer than the observed telomere length by TRF (Figs. 2 and 3b). Subtelomeric PETRA primers are designed to bind adjacent to the start of the TTTAGGG repeat. However, for many species, the telomere repeat sequence proximal to the subtelomeric region contains numerous mismatches, likely due to errors during DNA replication. Since the TRF assay makes use of an endonuclease with a 4-bp recognition sequence (TTAA), the discrepancy between PETRA and TRF is likely due to the occurrence of mismatched repeat sequence (TTAA) within the telomere tract, causing telomeric DNA to be a substrate for digestion by the endonuclease. Due to the discrepancy between our TRF and PETRA results and the possibility of false positives, staggered primers were designed for each syntenic chromosome arm in *A. arabicum*. Consistent with the primer design, the first product is slightly smaller than the second product for each arm (Online Resource 4). Taking the PETRA results into account, *A. arabicum* telomeres are

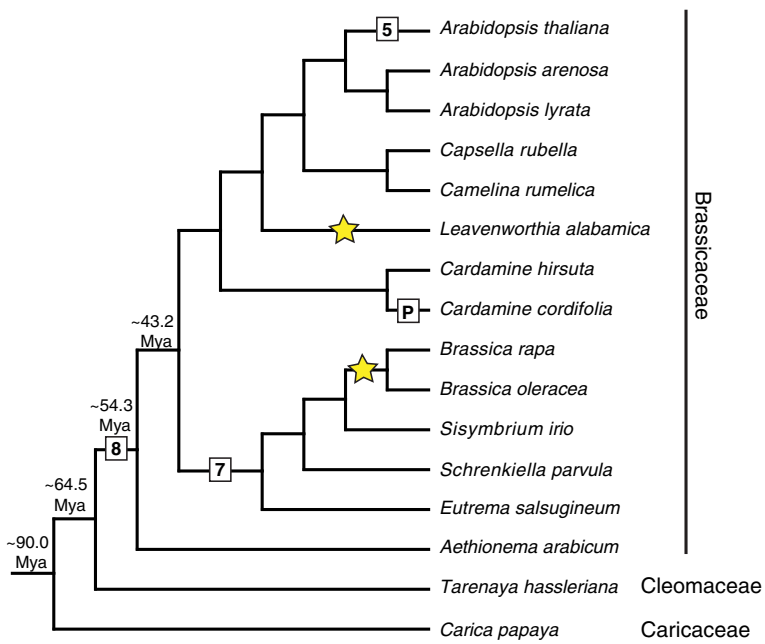

Fig. 1 Phylogenetic tree of species sampled in this study. Adapted from Beilstein et al. 2010. Approximate times of divergence are listed. Base chromosome number is indicated (*white box*) for all species derived from that node. *Yellow stars* indicate a genome duplication or triplication event. "*P*" indicates the plant is perennial

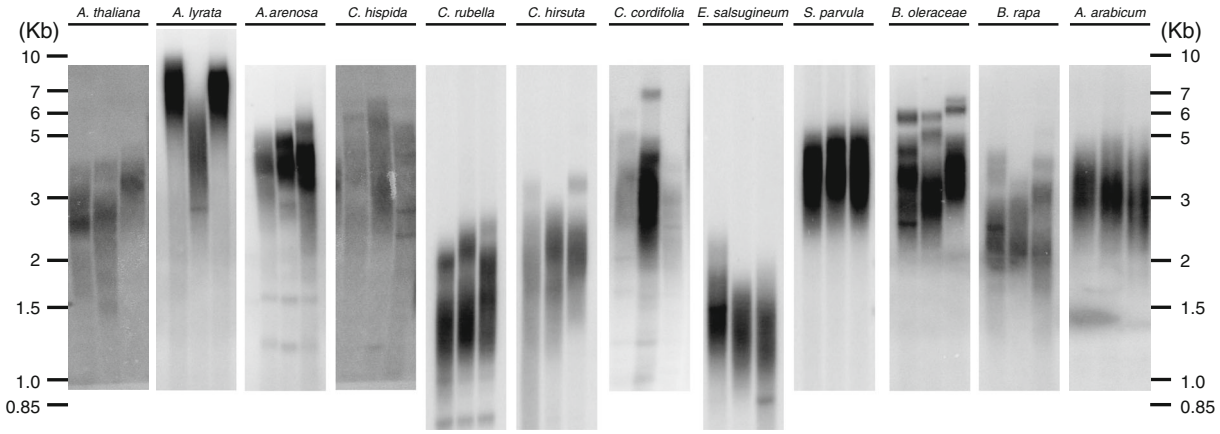

Fig. 2 Bulk telomere length was determined by terminal restriction length analysis (TRF). Southern blots are organized based on the phylogenetic relationships within the family. Blots were probed with the canonical plant telomere repeat $(TTTAGGG)^3$. Three independent biological replicates were performed for each species. Length is shown to the side in kilobases

approximately 6.5–7-kb long (Fig. 3b) well within the range seen in Arabidopsis ecotypes (Shakirov and Shippen 2004).

Despite the large-scale synteny shared between these chromosome arms, significant differences were seen between even closely related species such as Arabidopsis and *A. lyrata*. While the PCR reaction only worked for two of the syntenic chromosome arms for *A. lyrata*, the telomeres on chromosome arm 1L and 6R were approximately twice as long as their counterparts in Arabidopsis (Fig. 3b). Interestingly, despite its relative intactness in terms of gene synteny throughout the species examined, chromosome arm 1L shows a higher degree of telomere length variation than seen with other chromosome arms with less overall synteny, such as 2R and 3L (Fig. 3b). A more detailed analysis will be required to identify potential factors influencing the difference in telomere length in Brassicaceae.

Phylogenetic analysis of telomere binding proteins in Brassicaceae

To address whether Arabidopsis is unique in retaining 12 TRFLs or if this gene family shares a similar retention history in other plants, we examined the duplication pattern of TRFLs in Brassicaceae, using *Carica papaya* (order Brassicales) and *Theobroma cacao* (order Malvales) as out-groups.

Phylogeny for TRFL was inferred from an amino acid alignment of 585 residues and included 82 accessions retrieved from BLAST searches, representing homologs from eight species. The resulting topology revealed two major lineages: TRFLs with the Myb-domain+Myb extension (class 1, encoding telomere binding proteins in Arabidopsis) and TRFLs with only the Myb-domain (class 2, non-telomere-binding proteins in Arabidopsis) (Fig. 4a; Karamysheva et al. 2004). The duplication giving rise to class 1 and 2 appears to predate the evolution of eurosids II (Online Resource 5), but the exact origin of the duplication cannot be inferred from the tree included here. The topology of the tree within the class 1 TRFL clade revealed three duplication events that pre-date the emergence of Brassicaceae, and two duplication events which may have occurred at the base of the family (Fig. 4a; red stars indicate Brassicaceae-specific duplications). Outside Brassicaceae, class 1 genes were also retrieved from *C. papaya* and *T. cacao*. We recovered two class 1 members from *C. papaya* and three from *T. cacao*. Although CpTRFL1 was formerly annotated as a TRFL1, CpTRFL1, along with TcTRFL1, are actually more closely related to the TBP1/TRFL9 clade (Shakirov et al. 2008). Thus, together, TRFL9 and TBP1 represent one distinct clade of TRFL class 1. CpTRFL2 and TcTRFL2 are sister to the Brassicaceae TRFL2 clade. From these data, we can infer that Brassicaceae TRFL1 and TRP1 duplicated after the divergence between *C. papaya* and Brassicaceae. The third TcTRFL is sister to the Brassicaceae TRFL4 clade, indicating that the common ancestor of Brassicales+Malvales contained a TRFL4 copy.

Interestingly, it appears that harboring six TRFL class 1 genes is common to most Brassicaceae. The branch

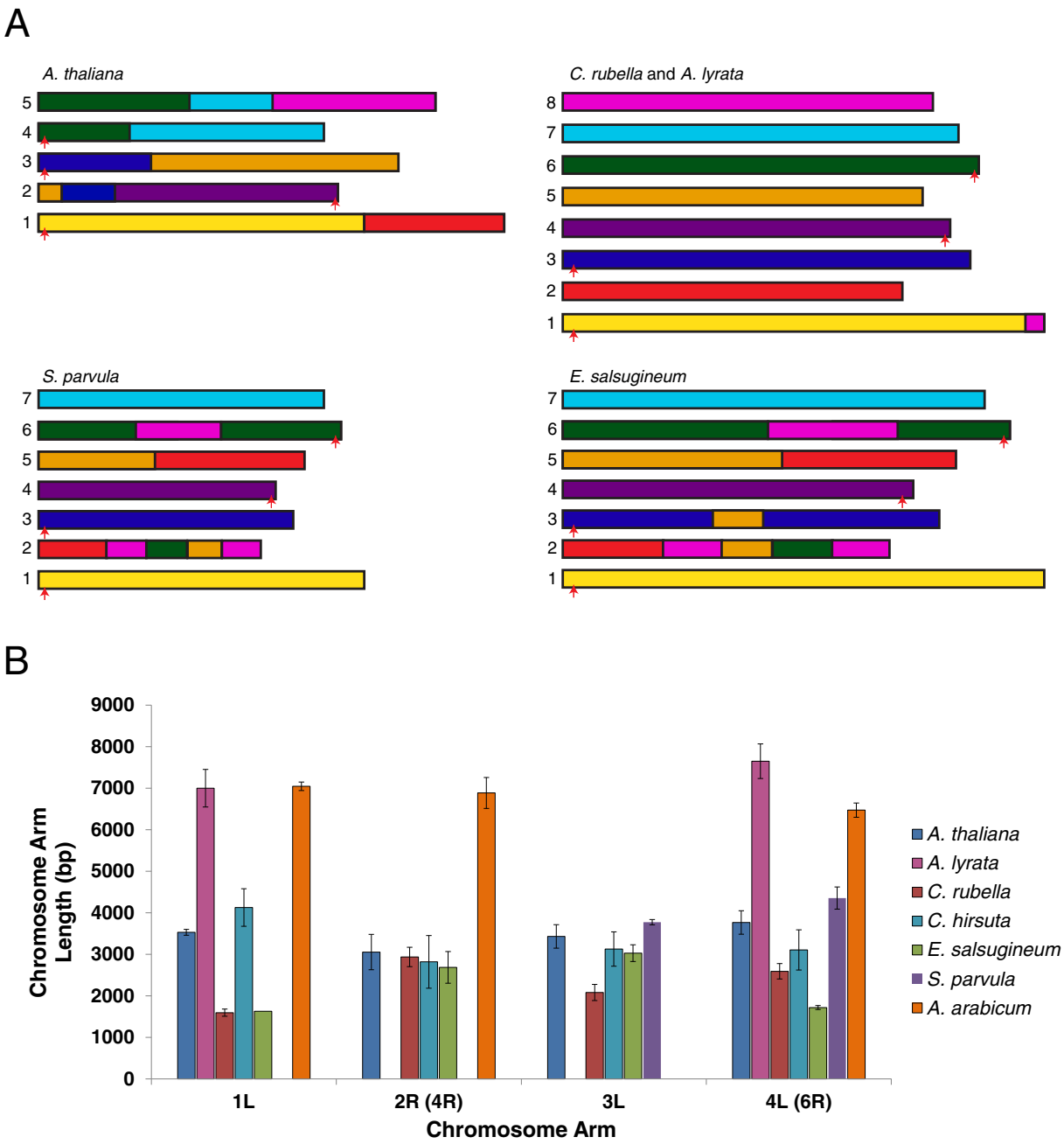

Fig. 3 Analysis of telomere length of syntenic chromosome arms by PETRA. **a** Schematic representing the genome arrangement for five Brassicaceae members with sequenced genomes. PETRA primers were designed for chromosome arms displaying synteny among all species examined (*red arrows*). For *C. hirsuta* and A. *arabicum*, synteny was confirmed by BLAST and CoGe (see Materials and methods for more information). **b** Average chromosome arm length calculated from PETRA products. Some arms are missing for particular species due to multiple failed PCR reactions. *Error bars* represent at minimum three independent biological replicates. Only one data point was available for *E. salsugineum* arm 1L (sample #4). *1L* left arm of chromosome 1, *2R* right arm of chromosome 2, etc. *A. thaliana* 2R is equivalent to 4R in all other species. *A. thaliana* 4L is equivalent to 6R in all other species. Adapted from: Mandakova and Lysak 2008; Yang et al. 2013

uniting TRFL9 and TBP1 is well supported (Fig. 4a). These two genes have undergone further, more phylogenetically restricted duplication events since in *B. rapa*, there are two copies of TRFL9 and three copies

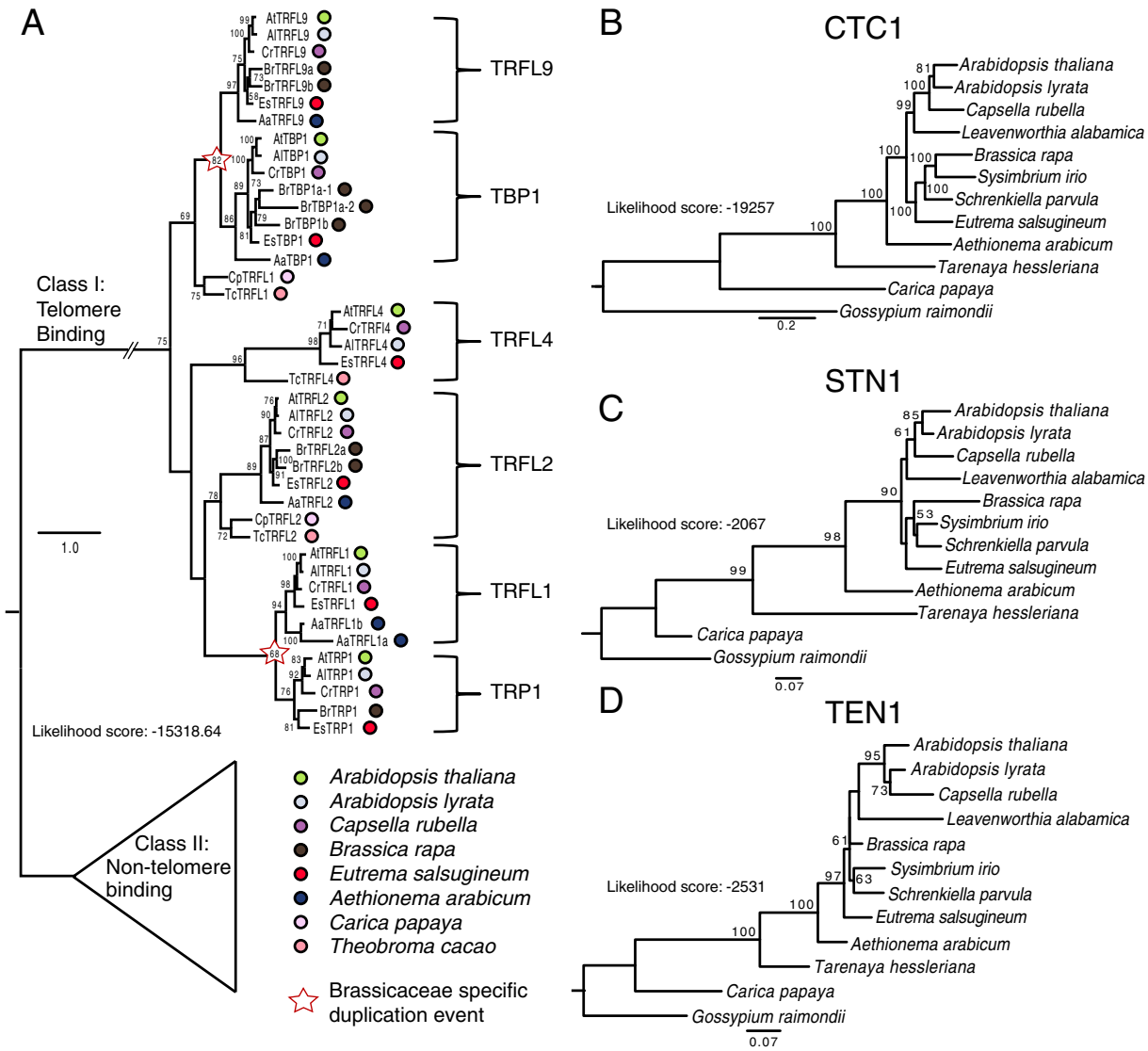

Fig. 4 Maximum likelihood trees of coding DNA sequence (Ten1) and protein sequence (CTC1, STN1, and TRFLs) from members of Brassicaceae plant family and close relatives. CTC1 and Ten1 alignments were generated using Geneious multiple alignment and trimmed manually to exclude large unaligned regions. TRFLs were aligned with ClustalW and trimmed of any sites represented by less than half of the taxa. Bootstrap support (*node labels*) was inferred using 100 bootstrap replicates. Bootstrap values of <50 are omitted from the trees. The TRFL tree (**a**) is rooted along the branch leading to non-telomere-binding TRFL family 2 proteins. Orthology of TRFL genes is inferred by reciprocal BLAST and confirmed by phylogenetic grouping. *Stars* indicate gene duplications specific to Brassicaceae plants. Non-telomere-binding TRFL family tree is shown in Online Resource 5. CTC1 (**b**), STN1 (**c**), and Ten1 (**d**) trees were rooted along the known out-group species (*G. raimondii*, Malvales)

of TBP1 present. Inspection of the TRFL4/2/1/TRP1 clade reveals an even greater level of gene duplication and loss. Despite extremely low expression levels in *A. thaliana* (Karamysheva et al. 2004), TRFL2 is retained in each species examined, including a recent duplication event represented by paralogous copies in *B. rapa*. In contrast, TRFL4, which is also expressed at low levels, shows no pattern of further duplication and may have been independently lost from *A. arabicum*, *B. rapa*, and *C. papaya*. TRFL1 is a single copy within all Brassicaceae except *A. arabicum*, where the two paralogs may be due to a tandem duplication event since they are adjacent to one another in the *A. arabicum* genome (data not shown). This duplication of TRFL1

in *A. arabicum* is coincident with the loss of the TRFL1 paralog, TRP1.

SMH phylogeny was inferred in a similar manner to the TRFLs (Online Resource 6). In contrast to the TRFLs, there has been no major expansion of the five-membered SMH family within Brassicaceae. Within the Arabidopsis SMH family, three proteins have been characterized to bind telomeric DNA in vitro (TRB1, TRB2, and TRB3), whereas two other genes share high sequence similarity and are referred to as TRB-like genes (TRBL). All five SMH genes are retained as a single copy in Brassicaceae, except in the case of species that experienced relatively recent whole genome duplication events such as *L. alabamica*, *B. rapa*, and *A. arabicum*. These species have duplicates of either the TRBs or TRBL genes.

We also analyzed the evolutionary history of members of the CST complex (CTC1/STN1/TEN1) from 11 members of Brassicaceae plus *C. papaya* and *Gossypium raimondii*. CST putatively serves as a capping complex by binding SS telomeric DNA. For CTC1, phylogeny was inferred from an alignment 1,409 residues in length (Fig. 4b). The STN1 tree was generated from an alignment of 164 residues (Fig. 4c). Finally, the TEN1 tree was produced from an alignment of 387 nucleotides. In contrast to the TRFL double-strand telomere-binding proteins, all sampled species retain only a single copy of CTC1, STN1, and TEN1. In each case, phylogenies inferred from the data are consistent with the accepted organismal phylogeny (Beilstein et al. 2010).

Discussion

Given observed differences in Arabidopsis telomere biology, we aimed to clarify which components of telomere architecture are well conserved among the close relatives of Arabidopsis in the plant family Brassicaceae. We present evidence to support the conclusion that the telomeric repeat observed in Arabidopsis is conserved across the family. The strong conservation of the TTTAGGG repeat motif within Brassicaceae and more broadly in plants is likely the result of evolutionary pressure imposed by the requirement of telomere-binding proteins to recognize the chromosome terminus and protect it from DSB repair machinery or attack by nucleases (Linger and Price 2009). Interestingly, among sampled taxa from across land plants, the only exceptions to the repeat are observed in the order Asparagales (Table 1), indicating that the shift to an alternative repeat may have occurred once at the base of the group (Sykorova et al. 2003). Whether a correlated change in telomere-binding proteins also occurred in this group is not known.

The increase in sampling within Brassicaceae for telomere length and the identity of the repeat motif makes it the most well-characterized group among plants. While we observed no variation in the telomeric repeat, there was considerable variation in telomere length. However, there appears to be no correlation between phylogeny and telomere length variation within Brassicaceae. In addition, neither ploidy level nor genome rearrangement events appear to affect telomere length. Longevity and lifestyle of the species also does not appear to have a profound effect on telomere length, as telomeres in the biennial *C. cordifolia* were only slightly longer than its close relative *C. hirsuta* (2–5 vs 1.2–3.5 kb). Interestingly, the telomere length variation we observed in Brassicaceae all fell within the range previously seen for different *A. thaliana* ecotypes (Shakirov and Shippen 2004). However, if *A. thaliana* is representative of Brassicaceae, then there may be intra-species variation that we did not uncover here.

Telomere length at homologous chromosome ends was also measured and compared among five Brassicaceae species. Interestingly, telomeres in *C. rubella* and *E. parvula* are close to the length believed to be the critical threshold governing recognition of the chromosome terminus as a double-strand break in Arabidopsis. Telomeres shorter than 1 kb in Arabidopsis are recruited into telomere-to-telomere chromosomal fusions, a hallmark of deprotected chromosome ends (Heacock et al. 2007). Variation in telomere length among species has been observed in other eukaryotic clades, indicating critical threshold values can differ even among closely related species (Kipling and Cooke 1990; Hemann and Greider 2000). In sum, our results show that considerable length variation also exists within the sampled Brassicaceae, critical threshold values likely differ among them, and genomic stability and viability can be maintained by a wide variety of telomere lengths.

Our analyses of the TRFL and CST gene families revealed that these DS and SS DNA-binding complexes differ in their patterns of duplication and retention in Brassicaceae. The DS binding proteins have undergone at least two relatively recent duplication events that are reflected across the species of the family. In part, these

duplications are responsible for the 12 TRFL genes present in Arabidopsis, six of which contain a Myb+ Myb extension domain that allows for telomere-specific DNA binding (class 1), whereas the other six lack this domain and do not bind telomeric DNA (class 2) (Karamysheva et al. 2004). It is intriguing that all sampled Brassicaceae retain a minimum of five family 1 TRFL genes, while other plants do not appear to share in this abundance. *P. patens* and *C. papaya* contain two, whereas *Selaginella moellendorffii* contains three (Shakirov et al. 2008; Shakirov and Shippen 2012). In addition, we recovered only three class 1 TRFLs from *T. cacao*. Members of the Brassicaceae last shared a common ancestor with *C. papaya* approximately 90 million years ago (Beilstein et al. 2010), and the exact timing of TRFL duplications following the divergence of these lineages is not known. Thus, whether the proliferation of TRFLs is unique to Brassicaceae or if additional TRFLs are present in species sharing a more recent common ancestor with the group cannot be determined with the current sampling. One possibility is that the proliferation is correlated with the alpha whole genome duplication event that is believed to have occurred at the base of Brassicaceae (Bowers et al. 2003). One intriguing question presented by these data is as follows: Why do species in Brassicaceae retain these additional copies post duplication?

In stark contrast to the duplication events observed in TRFL proteins, members of the SS telomeric DNA-binding complex, CST, are single copy in Brassicaceae. This is particularly interesting given the alpha genome duplication event known to have occurred near the base of the family as well as numerous other lineage-specific duplication events. An example of the latter is *B. rapa*, which has undergone a relatively recent whole genome triplication event (Wang et al. 2011) but retains a single copy of each CST component. In addition, known polyploids *Zea mays* and *Sorghum bicolor* have a single copy of each member of the CST complex (Data not shown). It is possible that redundant copies are disadvantageous and that there is a requirement to maintain dosage balance among the different components. Such balance may be much easier to maintain when genes occur as only a single copy in the genome, thereby favoring loss of duplicates rather than retention. Interestingly, this phenomenon extends well beyond plants. With the exception of the fungal genus *Candida* (Ascomycota), all species in which CST has been described harbor a single copy of each component (Lue and Chan 2013). It is possible that copy number is also affected by alternative functions for CST. Such functions have not yet been described in plants, but evidence in human cells suggests there may be additional roles for CST that are not associated with telomere end protection (Stewart et al. 2012).

The SMH/TRB gene family displays an evolutionary history intermediate to that of either the TRFLs or the CST complex. There has been no major Brassicaceae expansion as is the case with the TRFLs. Neither is there evidence of gene loss following WGD. To date, there is no in vivo evidence to suggest the SMH/TRB proteins are necessary for telomere length regulation. While one SMH protein, AtTRB1, has been shown to interact with AtPOT1b in vitro (Kuchar and Fajkus 2004), three of the SMH genes (At1g17520, At1g49950, and At5g67850) have been identified as potential transcription factors and are upregulated in response to plant hormones such as auxin or gibberellic acid (Yanhui et al. 2006; Koroleva et al. 2005). Given the well-characterized telomeric-DNA binding of the SMH/TRB protein family, it is possible that these proteins bind to the telomere repeat containing promoter element referred to as the telo box (Tremousaygue et al. 2003). This element is found in the promoter of numerous genes expressed in meristematic tissue (Manevski et al. 2000; Gaspin et al. 2010). Alternatively, these proteins could be involved in transcription of telomeric RNA (TERRA) molecules at the chromosome end (Vrbsky et al. 2010).

As exploration of telomere biology continues in plants, the Arabidopsis model of telomere architecture and end protection will serve as a critical comparison points for other plants. Our data indicate that numerous aspects of the model hold across the closest relatives of Arabidopsis. Interestingly, the lack of correlation between telomere length, plant lifestyle, and TRFL copy number implies that there are other genetic factors regulating the length of telomeres that may be more highly conserved across Brassicaceae. However, variation in the number of telomere-binding proteins, length of bulk telomeres, and telomere length at homologous chromosome arms also exists. The impact of this variation on how plants maintain genome integrity is currently not appreciated, but future studies to address this question are likely to deepen our understanding of the essential role of telomeres for genome viability.

Acknowledgments Sequence data for *C. hirsuta* lab was funded by grant BB/H006974/1 from the BBSRC, a core grant from the Max Planck Society, and the Gatsby Charitable foundation to M. Tsiantis.

References

Adams SP, Hartman TP, Lim KY, Chase MW, Bennett MD, Leitch IJ, Leitch AR (2001) Loss and recovery of Arabidopsis-type telomere repeat sequences 5′-(TTTAGGG)(n)-3′ in the evolution of a major radiation of flowering plants. Proc Biol Sci 268:1541–1546

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Baumann P, Cech TR (2001) Pot1, the putative telomere end-binding protein in fission yeast and humans. Science 292: 1171–1175

Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S (2010) Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. Proc Natl Acad Sci U S A 107:18724–18728

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422: 433–438

Cesare AJ, Quinney N, Willcox S, Subramanian D, Griffith JD (2003) Telomere looping in *P. sativum* (common garden pea). Plant J 36:271–279

Chen CM, Wang CT, Ho CH (2001) A plant gene encoding a Myb-like protein that binds telomeric GGTTTAG repeats in vitro. J Biol Chem 276:16511–16519

Cheng F, Mandakova T, Wu J, Xie Q, Lysak MA, Wang X (2013) Deciphering the diploid ancestral genome of the Mesohexaploid *Brassica rapa*. Plant Cell 25:1541–1554

Cifuentes-Rojas C, Nelson AD, Boltz KA, Kannan K, She X, Shippen DE (2012) An alternative telomerase RNA in Arabidopsis modulates enzyme activity in response to DNA damage. Genes Dev 26:2512–2523

Colgin LM, Baran K, Baumann P, Cech TR, Reddel RR (2003) Human POT1 facilitates telomere elongation by telomerase. Curr Biol 13:942–946

de Lange T (2009) How telomeres solve the end-protection problem. Science 326:948–952

Fajkus J, Kovarik A, Kralovics R, Bezdek M (1995) Organization of telomeric and subtelomeric chromatin in the higher plant *Nicotiana tabacum*. Mol Gen Genet 247:633–638

Flanary BE, Kletetschka G (2005) Analysis of telomere length and telomerase activity in tree species of various life-spans, and with age in the bristlecone pine *Pinus longaeva*. Biogerontology 6: 101–111

Fuchs J, Schubert I (1996) Arabidopsis-type telomere sequences on chromosome termini of *Selaginella martensii* Spring (Pteridophyta). Biologisches Zentralbl 115:260–265

Gaspin C, Rami JF, Lescure B (2010) Distribution of short interstitial telomere motifs in two plant genomes: putative origin and function. BMC Plant Biol 10:283

Giraud-Panis MJ, Teixeira MT, Geli V, Gilson E (2010) CST meets shelterin to keep telomeres in check. Mol Cell 39:665–676

Gish W, States DJ (1993) Identification of protein coding regions by database similarity search. Nat Genet 3:266–272

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178–D1186

Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-Lopez Z, Steffen JG, Hazzouri KM et al (2013) An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. Nat Genet 45:891–898

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

Heacock ML, Idol RA, Friesner JD, Britt AB, Shippen DE (2007) Telomere dynamics and fusion of critically shortened telomeres in plants lacking DNA ligase IV. Nucleic Acids Res 35:6490–6500

Hemann MT, Greider CW (2000) Wild-derived inbred mouse strains have short telomeres. Nucleic Acids Res 28:4474–4478

Hwang MG, Cho MH (2007) *Arabidopsis thaliana* telomeric DNA-binding protein 1 is required for telomere length homeostasis and its Myb-extension domain stabilizes plant telomeric DNA binding. Nucleic Acids Res 35:1333–1342

Karamysheva ZN, Surovtseva YV, Vespa L, Shakirov EV, Shippen DE (2004) 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

Kazda A, Zellinger B, Rossler M, Derboven E, Kusenda B, Riha K (2012) Chromosome end protection by blunt-ended telomeres. Genes Dev 26:1703–1713

Kipling D, Cooke HJ (1990) Hypervariable ultra-long telomeres in mice. Nature 347:400–402

Koroleva OA, Tomlinson ML, Leader D, Shaw P, Doonan JH (2005) High-throughput protein localization in Arabidopsis using Agrobacterium-mediated transient expression of GFP-ORF fusions. Plant J 41:162–174

Kuchar M, Fajkus J (2004) Interactions of putative telomere-binding proteins in *Arabidopsis thaliana*: identification of functional TRF2 homolog in plants. FEBS Lett 578: 311–315

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M et al (2012) The Arabidopsis information resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 40:D1202–D1210

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948

Leehy KA, Lee JR, Song X, Renfrew KB, Shippen DE (2013) MERISTEM DISORGANIZATION1 Encodes TEN1, an essential telomere protein that modulates telomerase processivity in Arabidopsis. Plant Cell 25:1343–1354

Linger BR, Price CM (2009) Conservation of telomere protein complexes: shuffling through evolution. Crit Rev Biochem Mol Biol 44:434–446

Lue NF, Chan J (2013) Duplication and functional specialization of the telomere-capping protein Cdc13 in Candida species. J Biol Chem 288:29115–29123

Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang H, Wang X, Bowers J, Paterson A, Lisch D et al (2008) Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. Plant Physiol 148:1772–1781

Lysak MA, Mandakova T, Lacombe E (2010) Reciprocal and multi-species chromosome BAC painting in crucifers (Brassicaceae). Cytogenet Genome Res 129:184–189

Mandakova T, Lysak MA (2008) Chromosomal phylogeny and karyotype evolution in x=7 crucifer species (Brassicaceae). Plant Cell 20:2559–2570

Manevski A, Bertoni G, Bardet C, Tremousaygue D, Lescure B (2000) In synergy with various cis-acting elements, plant insterstitial telomere motifs regulate gene expression in Arabidopsis root meristems. FEBS Lett 483:43–46

Marian CO, Bordoli SJ, Goltz M, Santarella RA, Jackson LP, Danilevskaya O, Beckstette M, Meeley R, Bass HW (2003) The maize Single myb histone 1 gene, Smh1, belongs to a novel gene family and encodes a protein that binds telomere DNA repeats in vitro. Plant Physiol 133:1336–1350

Mizuno H, Wu J, Kanamori H, Fujisawa M, Namiki N, Saji S, Katagiri S, Katayose Y, Sasaki T, Matsumoto T (2006) Sequencing and characterization of telomere and subtelomere regions on rice chromosomes 1S, 2S, 2L, 6L, 7S, 7L and 8S. Plant J 46:206–217

Mozgova I, Schrumpfova PP, Hofr C, Fajkus J (2008) Functional characterization of domains in AtTRB1, a putative telomere-binding protein in *Arabidopsis thaliana*. Phytochemistry 69: 1814–1819

Nelson AD, Shippen DE (2012) Surprises from the chromosome front: lessons from Arabidopsis on telomeres and telomerase. Cold Spring Harb Symp Quant Biol 77:7–15

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

Price CM, Boltz KA, Chaiken MF, Stewart JA, Beilstein MA, Shippen DE (2010) Evolution of CST function in telomere maintenance. Cell Cycle 9:3157–3165

Richards EJ, Ausubel FM (1988) Isolation of a higher eukaryotic telomere from Arabidopsis Thaliana. Cell 53:127–136

Riha K, McKnight TD, Griffing LR, Shippen DE (2001) Living with genome instability: plant responses to telomere dysfunction. Science 291:1797–1800

Schrumpfova P, Kuchar M, Mikova G, Skrisovska L, Kubicarova T, Fajkus J (2004) Characterization of two *Arabidopsis thaliana* myb-like proteins showing affinity to telomeric DNA sequence. Genome 47:316–324

Shakirov EV, Shippen DE (2004) Length regulation and dynamics of individual telomere tracts in wild-type Arabidopsis. Plant Cell 16:1959–1967

Shakirov EV, Shippen DE (2012) *Selaginella moellendorffii* telomeres: conserved and unique features in an ancient land plant lineage. Front Plant Sci 3:161

Shakirov EV, Surovtseva YV, Osbun N, Shippen DE (2005) The Arabidopsis Pot1 and Pot2 proteins function in telomere length homeostasis and chromosome end protection. Mol Cell Biol 25:7725–7733

Shakirov EV, Salzberg SL, Alam M, Shippen DE (2008) Analysis of *Carica papaya* telomeres and telomere-associated proteins: insights into the evolution of telomere maintenance in Brassicales. Trop Plant Biol 1:202–215

Shakirov EV, Song X, Joseph JA, Shippen DE (2009) 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

Shakirov EV, Perroud PF, Nelson AD, Cannell ME, Quatrano RS, Shippen DE (2010) Protection of Telomeres 1 is required for telomere integrity in the moss Physcomitrella patens. Plant Cell 22:1838–1848

Song X, Leehy K, Warrington RT, Lamb JC, Surovtseva YV, Shippen DE (2008) STN1 protects chromosome ends in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A 105:19815–19820

Stamatakis A, Alachiotis N (2010) Time and memory efficient likelihood-based tree searches on phylogenomic alignments with missing data. Bioinformatics 26:i132–i139

Stewart JA, Wang F, Chaiken MF, Kasbek C, Chastain PD 2nd, Wright WE, Price CM (2012) Human CST promotes telomere duplex replication and general replication restart after fork stalling. EMBO J 31:3537–3549

Sue SC, Hsiao HH, Chung BC, Cheng YH, Hsueh KL, Chen CM, Ho CH, Huang TH (2006) Solution structure of the *Arabidopsis thaliana* telomeric repeat-binding protein DNA binding domain: a new fold with an additional C-terminal helix. J Mol Biol 356:72–85

Surovtseva YV, Shakirov EV, Vespa L, Osbun N, Song X, Shippen DE (2007) Arabidopsis POT1 associates with the telomerase RNP and is required for telomere maintenance. EMBO J 26: 3653–3661

Surovtseva YV, Churikov D, Boltz KA, Song X, Lamb JC, Warrington R, Leehy K, Heacock M, Price CM, Shippen DE (2009) Conserved telomere maintenance component 1 interacts with STN1 and maintains chromosome ends in higher eukaryotes. Mol Cell 36:207–218

Sykorova E, Lim KY, Kunicka Z, Chase MW, Bennett MD, Fajkus J, Leitch AR (2003) Telomere variability in the monocotyledonous plant order Asparagales. Proc Biol Sci 270:1893–1904

Tremousaygue D, Garnier L, Bardet C, Dabos P, Herve C, Lescure B (2003) Internal telomeric repeats and 'TCP domain' protein-binding sites co-operate to regulate gene expression in *Arabidopsis thaliana* cycling cells. Plant J 33:957–966

Vrbsky J, Akimcheva S, Watson JM, Turner TL, Daxinger L, Vyskot B, Aufsatz W, Riha K (2010) siRNA-mediated methylation of Arabidopsis telomeres. PLoS Genet 6:e1000986

Wang F, Podell ER, Zaug AJ, Yang Y, Baciu P, Cech TR, Lei M (2007) The POT1-TPP1 telomere complex is a telomerase processivity factor. Nature 445:506–510

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F et al (2011) The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet 43: 1035–1039

Watson JM, Riha K (2010) Comparative biology of telomeres: where plants stand. FEBS Lett 584:3752–3759

Weigel D, Glazebrook J (2002) Arabidopsis: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Weiss-Schneeweiss H, Riha K, Jang CG, Puizina J, Scherthan H, Schweizer D (2004) Chromosome termini of the monocot plant Othocallis siberica are maintained by telomerase, which

specifically synthesises vertebrate-type telomere sequences. Plant J 37:484–493

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18: 691–699

Yang R, Jarvis DE, Chen H, Beilstein MA, Grimwood J, Jenkins J, Shu S, Prochnik S, Xin M, Ma C et al (2013) The reference genome of the halophytic plant *Eutrema salsugineum*. Front Plant Sci 4:46

Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q et al (2006) The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant Mol Biol 60: 107–124