New Phytologist

Research

Telomerase RNA gene paralogs in plants – the usual pathway to unusual telomeres

Michal Závodník[1,2]*, Petr Fajkus[2,3]*, Michal Franek[2], David Kopecký[4], Sònia Garcia[5], Steven Dodsworth[6], Andrés Orejuela[7], Agata Kilar[1,2], Jiří Ptáček[8], Martin Mátl[3], Anna Hýsková[1], Jiří Fajkus[1,2,3] and Vratislav Peška[3]

[1]Laboratory of Functional Genomics and Proteomics, NCBR, Faculty of Science, Masaryk University, Brno, CZ-61137, Czech Republic; [2]Mendel Centre for Plant Genomics and Proteomics, CEITEC Masaryk University, Brno, CZ-62500, Czech Republic; [3]Department of Cell Biology and Radiobiology, Institute of Biophysics of the Czech Academy of Sciences, Brno, CZ-61265, Czech Republic; [4]Centre of Plant Structural and Functional Genomics, Institute of Experimental Botany of the Czech Academy of Sciences, Olomouc, CZ-779 00, Czech Republic; [5]Institut Botànic de Barcelona (IBB-CSIC), Passeig del Migdia S/N, Barcelona, 08038, Catalonia, Spain; [6]School of Biological Sciences, University of Portsmouth, King Henry Building, King Henry I St., Portsmouth, PO1 2DY, UK; [7]Grupo de Investigación en Recursos Naturales Amazónicos – GRAM, Facultad de Ingenierías y Ciencias Básicas and Herbario Etnobotánico del Piedemonte Andino Amazónico (HEAA), Instituto Tecnológico del Putumayo – ITP, Mocoa, Putumayo, Colombia; [8]Potato Research Institute Havlíčkův Brod Ltd, Havlíčkův Brod, CZ-58001, Czech Republic

Authors for correspondence:
Jiří Fajkus
Email: fajkus@sci.muni.cz

Vratislav Peška
Email: vpeska@ibp.cz

Petr Fajkus
Email: fajkuspe@ibp.cz

Received: *5 April 2023*
Accepted: *6 June 2023*

New Phytologist (2023) **239:** 2353–2366
doi: 10.1111/nph.19110

Key words: evolution, land plants, mechanism of telomere transitions, telomerase RNA paralogs, unusual telomere DNA.

Summary

- Telomerase, telomeric DNA and associated proteins together represent a complex, finely tuned and functionally conserved mechanism that ensures genome integrity by protecting and maintaining chromosome ends. Changes in its components can threaten an organism's viability. Nevertheless, molecular innovation in telomere maintenance has occurred multiple times during eukaryote evolution, giving rise to species/taxa with unusual telomeric DNA sequences, telomerase components or telomerase-independent telomere maintenance. The central component of telomere maintenance machinery is telomerase RNA (TR) as it templates telomere DNA synthesis, its mutation can change telomere DNA and disrupt its recognition by telomere proteins, thereby leading to collapse of their end-protective and telomerase recruitment functions.
- Using a combination of bioinformatic and experimental approaches, we examine a plausible scenario of evolutionary changes in TR underlying telomere transitions.
- We identified plants harbouring multiple TR paralogs whose template regions could support the synthesis of diverse telomeres. In our hypothesis, formation of unusual telomeres is associated with the occurrence of TR paralogs that can accumulate mutations, and through their functional redundancy, allow for the adaptive evolution of the other telomere components.
- Experimental analyses of telomeres in the examined plants demonstrate evolutionary telomere transitions corresponding to TR paralogs with diverse template regions.

Introduction

Telomere DNA in plants is typically formed by (TTTAGGG/CCCTAAA)$_n$ stretches of DNA of variable length, which are maintained by a specialized ribonucleoprotein enzyme – the telomerase. This holoenzyme is composed of at least two essential parts, the telomerase RNA (TR) harbouring a C-rich template region that dictates the synthesis of the G-rich telomere DNA sequence by the second part, the telomerase reverse transcriptase (TERT) protein. In this way, the telomerase counteracts replicative telomere shortening (reviewed in Blackburn, 1991; Schrumpfova & Fajkus, 2020). Although the TTTAGGG repeat seems to be ancestral for land plants and is highly conserved throughout their evolution, individual exceptions have been identified with different telomere motifs, which are scattered across the land plant phylogeny. These include species from orders of Asparagales (motifs TTAGGG since the divergence of Iridaceae (Sykorova *et al.*, 2003) and CTCGGT TATGGG in *Allium* (Fajkus *et al.*, 2016)), Solanales (TTTTTTAGGG – in *Cestrum* (Peska *et al.*, 2015)), Lamiales ($T_{2\text{-}3}$CAGG – in *Genlisea hispidula* (Tran *et al.*, 2015)) and Alismatales (TTAGGG – in *Zostera marina* (Peska *et al.*, 2020)). Recent characterization of TR genes in plants (Fajkus *et al.*, 2019), extended to other early diverged groups from the Diaphoretickes (Fajkus *et al.*, 2021) or the Animalia order Hymenoptera (Insecta; Fajkus *et al.*, 2023) aided in explaining

*These authors contributed equally to this work.

molecular principle of changes in telomere DNA sequence, as the consequence of corresponding mutations in the template region of their telomerase RNAs.

Due to the interplay among telomere DNA, telomere binding proteins, telomerase and other accessory factors, that together ensure genome integrity, modification in any of the individual players can completely alter genome stability and potentially threaten the organism's survival. The evolutionary arms race between repetitive DNA and essential chromatin proteins is not unique to telomeres. For example, a recent study described the mechanism responsible for cross-species incompatibility in *Drosophila* where an abundant satellite DNA and its associated protein must co-evolve to preserve germline genome integrity (Brand & Levine, 2022).

Prevention of deleterious complications in telomere functions constraints telomere diversity. This leads to conservation of telomere sequence and maintenance. Emerging variants in the telomere maintenance machinery (including switches to telomerase-independent telomere maintenance, adaptive evolution of telomere binding proteins or acquisition of novel telomere binding proteins) remain as long as they are functional, while those that introduce deleterious complications do not persist (reviewed in Fajkus *et al.*, 2005; Louis & Vershinin, 2005; Podlevsky & Chen, 2016; Cervenak *et al.*, 2021). The most prominent substrate for telomere evolution is the TR, as a key eukaryotic noncoding RNA (ncRNA). On the one hand, TR has a very conserved function; on the other hand, it displays extreme divergence in sequence, structure and biogenesis pathways (reviewed in Podlevsky & Chen, 2016).

In this paper, we aim to find out which major events precede one of the most dramatic changes – the substitution of residues in the TR template. This affects all other key players in telomere maintenance. It results in a change in telomere DNA sequence and thus prompts the need for adaptation of telomere binding proteins to compensate for changes in their binding affinity and dynamics in order to maintain chromosome-end protective and telomerase recruitment functions. During previous work on plant TR characterization (Fajkus *et al.*, 2019), we noticed that TR genes frequently occur as multiple paralogs. Compared with single-copy genes, redundant paralogs and in particular rapidly evolving ncRNA, such as the TR, are prone to accumulate mutations. In this study, we exploited *c.* 1100 available Tracheophyta genome assemblies (at NCBI) to characterize their TR genes and the variability of telomere repeats. This allowed us to elucidate a previously hidden ongoing interference between TR paralogs. We hypothesize that such interference of multiple TR copies is an undesirable evolutionary situation for the organism's fitness, which can be further solved by natural selection (e.g. gene conversion leading to repair of a damaged TR copy, or pseudogenization leading to its elimination). The coexistence of multiple TR copies may not always proceed in favour of the organism's fitness (e.g. when genetic drive overcomes natural selection) and may result in the replacement of the ancestral telomere DNA sequence by another one. The different scenarios regarding the possible fate of TR genes after duplication and their consequences are illustrated in Fig. 1. Using a combination of advanced techniques (Benson, 1999; Nawrocki & Eddy, 2013; Peska *et al.*, 2017) to identify TR-like genes across plant genomes and to predict candidate telomere sequences from raw genomic data, we provide comprehensive data about telomerase RNA and telomere sequence evolution in vascular plants. Moreover, our predictions from genomic data are confirmed in selected examples using a set of experimental techniques, for example Bal31 sensitivity of terminal restriction fragments (BAL31-TRF; BAL31-NGS), fluorescence *in situ* hybridization (FISH), Telomere repeat amplification protocol (TRAP), RNA-Seq and RT-qPCR for specific TR paralog transcripts. In combination with an analysis of the evolutionary origin of duplicated TR genes, we demonstrate a plausible scenario of telomere transitions that have been described previously and also newly in this study. In summary, we propose a model explaining why a certain degree of telomere sequence variation is found in plants.

Materials and Methods

The workflow to characterize species in which unusual telomeres evolved has two major parts (Fig. 2). The first part represents data mining of genome assemblies and other genomic and transcriptomic data to extract information about TR subunits, their template regions and abundant tandem repeats. Combining this with information about putative telomere repeat motifs allows fast screening of genomes to predict those that have undergone telomere sequence change or have been dealing with it. The second part employs traditional wet-lab methods (FISH, BAL31-TRF, TRAP and sequencing of its products) to validate predictions. We used selected examples from families Lamiaceae, Papaveraceae, Poaceae, Polygonaceae, Ranunculaceae and Solanaceae with available plant material. In addition, mutual syntenic relationships of TRs were analysed among representatives from the family Brassicaceae, as they included many well-assembled and annotated genomes across their phylogeny. To trace the consequences of the coexistence of more TR variants with different templates, transcriptional contribution of TR paralogs was evaluated in *Fagopyrum esculentum* and *F. tataricum* (Polygonaceae) by RT-qPCR.

Plant material and preparation of DNA and RNA

A detailed overview of selected species from Lamiaceae (*Erytranthe guttata* (DC.) G.L. Nesom), Papaveraceae (*Papaver somniferum* L.), Poaceae (*Festuca pratensis* Huds.), Polygonaceae (*Fagopyrum esculentum* Moench, *Fagopyrum tataricum* (L.) Gaertn.), Ranunculaceae (*Aconitum napellus* L., *Delphinium consolida* L., *Delphinium cultorum* Voss., *Delphinium staphisagria* L., *Ranunculus acris* L., *Ranunculus repens* L.) and Solanaceae (*Atropa belladonna* L., *Capsicum annuum* L., *Capsicum chinense* Jacq., *Hyoscyamus niger* L., *Jaltomata sinuosa* (Miers.) Mione, *Physaliastrum chamaesarachoides* (Makino) Makino, *Salpiglossis sinuate* (Ruiz&Pav.), *Salpichroa organifolia* (Lam.) Baill., *Solanum melongena* L., *Vestia foetida* (Ruiz&Pav.) Hoffmanns.) is provided in

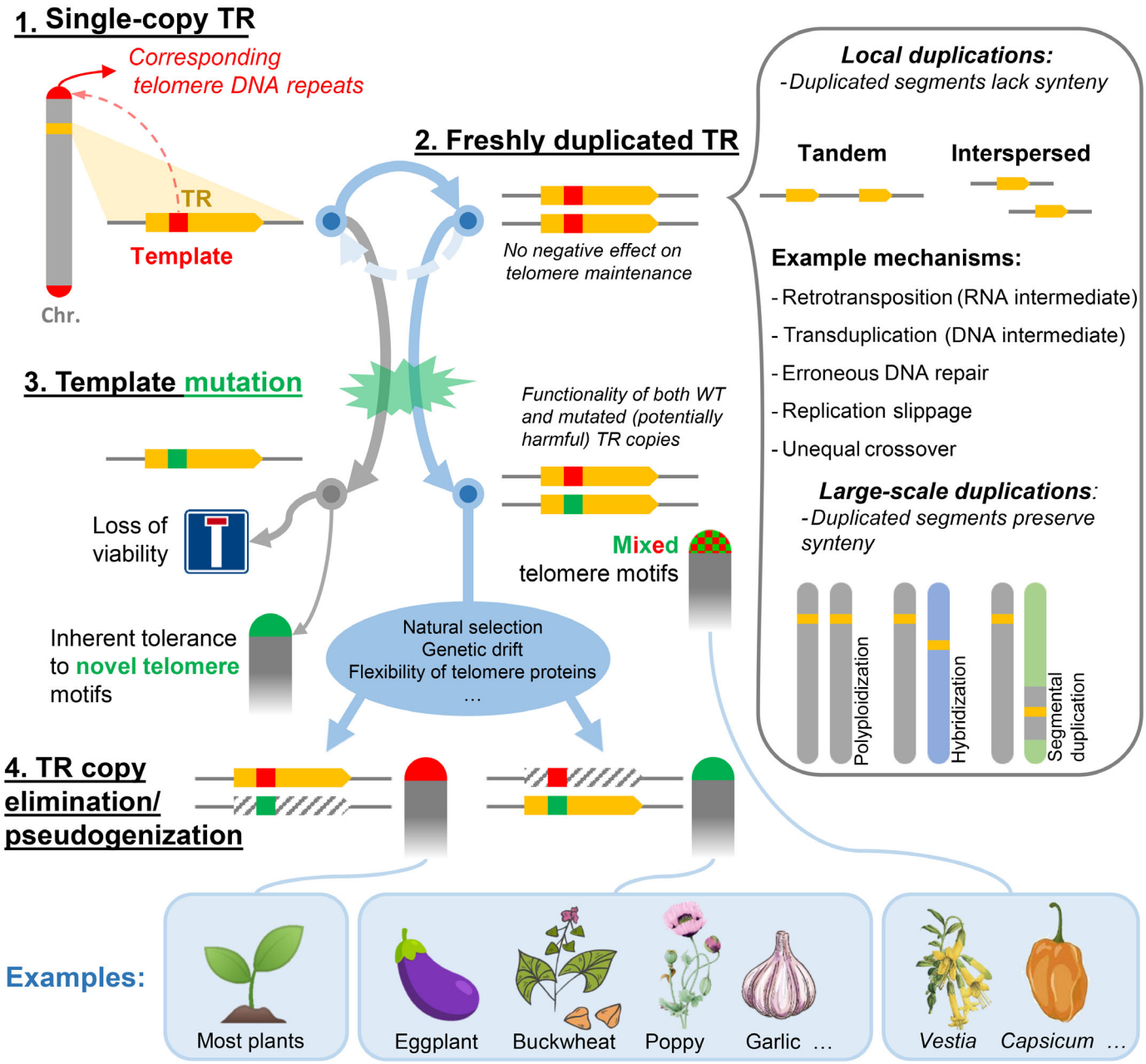

Fig. 1 Possible course of telomere DNA sequence changes. Two pathways are considered, either a mutation of the template region of a single-copy telomerase RNA (TR) gene (indicated with grey arrows) or such mutation after the previous TR duplication (indicated with blue arrows). While the first way (grey arrows) seems to be more straightforward, it does not allow for plant adaptation to a novel telomere sequence. This way might be evolutionary successful only in cases of specific template/telomere changes on which telomere and telomerase components are preadapted (e.g. recognition of the novel motif by telomere proteins is preserved). Frequent presence of multiple TR paralogs rather supports prevalence of the latter scenario (blue arrows) employing TR duplication (examples of possible duplication mechanisms are depicted on the right). Relaxed selection pressure on TR in more copies (compared with the single-copy gene) results in frequent mutability of redundant copies. This way provides adaptation window starting with presence of more TR copies coding for ancestral and mutated telomere motif (i.e. species with mixed telomeres). While natural selection tends to eliminate/ pseudogenize potentially harmful TR copies, genetic drift can act in the opposite direction, leading to elimination of ancestral TR and expansion of mutated copy (tolerated from the time when mutated and ancestral telomeres have been mixed). This cycle can repeat with each TR duplication and subsequent template mutation. Example outcomes of this process in specific plant species are shown at the bottom of the figure.

Supporting Information Methods S1. Methods used for nucleic acid preparations are described in Methods S2 (i.e. isolation of DNA, RNA and high-molecular-weight DNA, preparation of nuclei, DNA fibres and chromosome spreads for FISH and preparation of RNA-Seq and DNA-Seq libraries).

Fluorescence *in situ* hybridization and microscopy

Fluorescence *in situ* hybridization (FISH) was performed according to Goffova *et al.* (2019). Preparations of interphase nuclei, chromosome spreads and chromosome fibres are described in Methods S2.

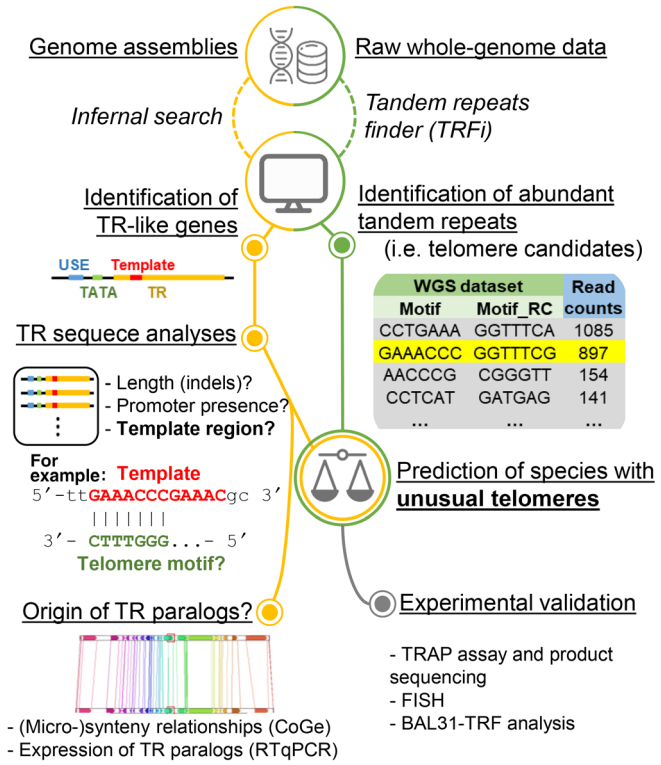

Fig. 2 Survey of species potentially harbouring noncanonical telomeres. The workflow combines obtaining information about telomerase RNA (TR) subunits and particular template regions (yellow pathway) with assessment of the abundance of tandem repeat motifs (green pathway). Examples of predicted species with unusual telomeres then undergo experimental validation (in grey). Origin of TR paralogs and their functionality is examined (bottom left). BAL31-TRF, terminal restriction fragments analysis after BAL31 digestion; CoGe, comparative genomics; FISH, fluorescence *in situ* hybridization; RTqPCR, reverse transcription quantitative real time polymerase chain reaction; TATA, TATA box of the promoter; TRAP, telomere repeat amplification protocol; USE, upstream sequence element of the type 3 promoter; WGS, whole genome sequencing.

Briefly, microscopic slides were rehydrated in 2× SSC and incubated with RNAse A. After a rinse in 2× SSC, slides were dehydrated in ethanol series. After air-drying, samples were hybridized with custom-locked nucleic acid (LNA) probes (see Methods S3) overnight at 37°C. For indirect detection, slides were blocked in 5% BSA in 2× SSC and then incubated with Alexa Fluor-488 conjugated streptavidin (#S11223; Thermo Fisher Scientific, Waltham, MA, USA). Samples were mounted in Vectashield (#H-1000-10; Vector laboratories (Burlingame, CA, USA) + DAPI (#D1306, Invitrogen)) and imaged on a Zeiss AxioImager Z2 (for details see Methods S3). The images (3D-stacks) were processed by a deconvolution algorithm in the ZEN BLACK software (v.3.0; Zeiss microscopy), telomere spot detection and visualization were performed in FIJI (Schindelin *et al.*, 2012) and IMARIS (v.10.0.0; Oxford Instruments, Oxfordshire, UK).

Analysis of terminal restriction fragments (TRF) and their sensitivity to BAL31 (BAL31-TRF)

To investigate the position of putative telomeric sequences, BAL31 nuclease digestion was performed according to Fojtová *et al.* (2015). Agarose blocks were equilibrated in BAL31 nuclease buffer and DNA was digested with BAL31 nuclease (New England Biolabs, Ipswich, MA, USA). Reactions were stopped by addition of an equal volume of 50 mM EGTA and BAL 31 nuclease was heat-inactivated at 60°C for 30 min. DNA in agarose blocks (*A. napellus*, *D. consolida*, *P. somniferum*, *R. repens* and *R. acris*) or in solution (*F. tataricum*) was digested by *Mse*I restriction enzyme (New England Biolabs). Samples were separated in Pulsed-Field Gel Electrophoresis (PFGE; CHEF Mapper instrument; Bio-Rad) according to Fojtová *et al.* (2015). MidRange PFG Marker (New England Biolabs) and Gene ruler 1 kb DNA Ladder (Thermo Fisher Scientific) were used as markers. PFGE gel was stained with EtBr, visualized in UV light and then alkali-blotted and hybridized with [^{32}P]-labelled probes. The telomere-specific probes were generated by nontemplate PCR using oligonucleotide primers (Methods S4; Ijdo *et al.*, 1991; Adamusova *et al.*, 2020). The nonspecifically bound probes were washed away, and hybridization signals were visualized using Typhoon FLA-7000 imaging system (Fujifilm, Minato, Japan). For detailed protocol, see Methods S4.

Telomere repeat amplification protocol (TRAP) and sequencing of products

Protein extracts were prepared from root tips of 7–14 d old seedlings of *F. esculentum*, *F. tataricum* and *P. somniferum*, root tips of mature plants of *D. consolida* according to (Fitzgerald *et al.*, 1996; Sykorova *et al.*, 2003). The protein concentration in extracts was determined according to Bradford (1976). Telomerase activity of prepared extracts was assayed under optimized conditions for *P. somniferum* (Fitzgerald *et al.*, 1996) or by modified TRAP protocol with Hot Start *Taq* DNA polymerase (New England Biolabs) for *F. esculentum*, *F. tataricum* and *D. consolida*. Primer combinations for TRAP designed to respective putative telomere repeats in plants were tested (Methods S5). Products of TRAP reactions were separated on a 12.5% polyacrylamide gel. Gel was stained with GelStar™ Nucleic acid gel stain (LONZA, Basel, Switzerland), and signals were visualized using FUSION FX (Vilber Lourmat, Collégien, France). TRAP products were purified by SPRIselect beads (Beckman Coulter, Brea, CA, USA). Purified TRAP products were cloned (TOPO TA Cloning® kit; Invitrogen) and sequenced. For detailed protocol, see Methods S5.

Transcriptional contribution of TR paralogs in Fagopyrum

Total RNAs (purified and quality checked as described in Methods S2c) of *F. esculentum* and *F. tataricum* were reverse transcribed to cDNA by M-MuLV reverse transcriptase (New England Biolabs). One µl of cDNA in mixture was used in qPCR. Reactions were performed in technical triplicates using FastStart SYBR Green Master (Roche, Basel, Switzerland) in cycler (Rotor-Gene 6000; Qiagen, Venlo, the Netherlands) and analysed by software (ROTOR-GENE Q Series Software, v.2.2.2). Efficiency of qPCR reactions with specific primer combinations (see Methods S6) was determined from calibration curves (Fig. S5) of reactions run on serial dilution of respective PCR

products obtained from genomic DNA. The CT values (i.e. the value of a 'threshold cycle', when qPCR curve intersects a threshold line) for respective reactions were transformed according to reaction efficiencies and normalized to 'FeVar1' in *F. esculentum* and 'FtVar1' in *F. tataricum* samples.

Identification of abundant tandem repeats in genomic data

Search for abundant tandem repeats to support functionality of particular motifs at telomeres was done similarly as in previous studies (e.g. Fajkus *et al.*, 2016, 2023; Peska *et al.*, 2020, 2021) by using the Tandem Repeats Finder (TRFi) tool (Benson, 1999) with custom-made scripts (Peska *et al.*, 2017). The analysed raw whole-genome sequencing data were sampled preferably from sequencing by Illumina platform and from one dataset per species with available genome assembly from the Sequence Read Archive (SRA). These datasets were supplemented with newly prepared genome datasets from *D. cultorum*, *D. staphisagria*, *D. consolida*, *S. sinuata* and four datasets from *F. tataricum* and *V. foetida* corresponding to genome sequencing of BAL31 treated and untreated high-molecular-weight DNAs and analysed (Peska *et al.*, 2017). Dataset accession numbers and the corresponding taxonomically sorted results are described in Table S2.

Identification of TR sequences

TRs were identified using our previously published covariance model for land plant TRs (Fajkus *et al.*, 2021) with the sequence-structure homology search tool Infernal (Nawrocki & Eddy, 2013). Genome assemblies subjected as the target database for TR homology search were obtained from the NCBI database using the Entrez (NCBI) tool (Gibney & Baxevanis, 2011) and the wget BASH utility. The cmsearch utility of Infernal was then used to perform searches in the target database of representative genome assemblies from 'Tracheophyta'. Standard parameters were applied to the Infernal tool. Significant Infernal hits were extended with 200 nucleotides of genomic context (using BLASTCMD tool; Camacho *et al.*, 2009) to assess the presence of the type 3 snRNA promoter. Results were sorted according to hit significance and taxonomy and processed to tabular output by using the TIDYVERSE package (Wickham *et al.*, 2019). TR sequences identified in genome assemblies were supplemented with TRs identified in total RNA-Seq data from *F. esculentum*, *F. tataricum*, *R. acris*, *S. sinuata and*, *V. foetida* assembled by TRINITY tool (Grabherr *et al.*, 2011) as described in Fajkus *et al.* (2021) and raw genomic DNA-Seq data from *Atropa belladonna* and *Warburgia ugandensis*.

Analysis of syntenic relationships among TR loci using the CoGe platform

Comparative analysis of gene collinearity between genomic regions containing TR genes and pseudogenes was performed by using the CoGe platform (Lyons & Freeling, 2008). Selected diploid Brassicales species with available annotated genomes at CoGe were subjected to this analysis (listed in Table S3). TR loci were identified in selected CoGe genomes by using CoGeBlast (Lyons *et al.*, 2008). The closest genomic feature (gene/exon) of each particular TR hit was used as a query to identify colinear regions in a selected set of genomes by using SynFind (algorithm: Last; gene window size = 40; Min. no. of genes: 4; scoring: 'collinear'; Tang *et al.*, 2015). Identified colinear regions between species and within species (in the case of more TR paralogs) were checked for presence of TR gene and sent to GEvo (Lyons *et al.*, 2008) for their visualization. All links to regenerate particular SynFind/GEvo analyses are provided in Table S3.

Results

Computational identification of TR genes and telomere motifs across land plants

The first step in understanding the evolution of TRs and telomere motifs across plants is the characterization of TR-like sequences and telomere repeats in genome assemblies. Using the Infernal tool with optimized covariance models for land plant telomerase RNAs (Fajkus *et al.*, 2021), 4829 hits were obtained of which 2219 corresponded to genuine TR sequences in 986 representative genomes (at NCBI). TR sequences were taxonomically ordered, filtered according to their hit significance (*e*-value) and extended with 200 nt of adjacent genomic sequence for further promoter analyses (summarized in Table S1). These data were supplemented with TRs identified in RNA-Seq and DNA-Seq data in species lacking genome assembly (e.g. *Delphinium*, *Ranunculus* and *Vestia*). Summary quantification of the TR search, presented in Fig. 3, indicates the frequent presence of more than one TR variant per genome. In parallel, whole-genome data (our own newly generated data together with data available in the Sequence Read Archive (SRA) at NCBI) from species selected across land plants were analysed by optimized TRFi to identify and quantify the genome proportion of tandem repeats. As we have previously shown, telomere repeats usually belong among the most highly abundant tandem repeats. The results for each analysed species are taxonomically ordered and presented in Table S2.

Selection of plant species indicating presence of an unusual telomere sequence and validation

As the aim of this study is to elucidate the evolutionary mechanisms leading to telomere sequence transitions, species, where such evolutionary changes had occurred or are currently in progress, were identified and examined first. The knowledge of TR, specifically of its template region, in combination with information about abundant tandem repeats in the respective genome, enables straightforward identification of the telomere sequence in such species (Fig. 2). Besides previously reported examples of unusual telomeres, we present additional species in Table 1, selected from across all the analysed genomes (Tables S1, S2), including important plant crops, carrying noncanonical telomeric repeats. Species potentially harbouring unusual telomere repeats (Table 1) are scattered throughout the land plant phylogeny and include early

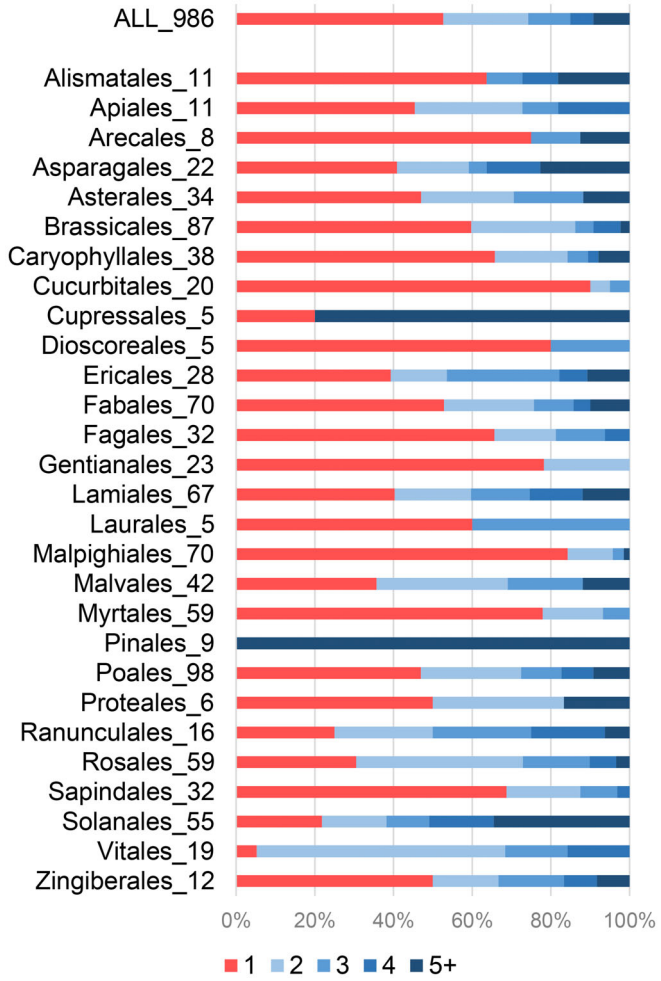

Fig. 3 Quantification of 2219 telomerase RNA (TR)-like sequences identified in 986 representative Tracheophyta (vascular plants) genomes (at NCBI, National Center for Biotechnology Information). Numbers after taxa names indicate numbers of corresponding representative species/genomes. The percentage shows the relative number of genomes with a single-copy TR (in red) or 2, 3, 4 and 5 and more TR copies (in shades of blue).

diverged angiosperms as well as monocot and eudicot clades in which species with noncanonical telomeres have been described previously or newly in this work (in bold letters). In representatives of Poales (*Festuca*, *Lolium*, *Zea*), TR template regions as well as high abundances of TTAGGG and TTTAGGG tandems in their genomic data (Tables S1, S2) suggested the presence of mixed telomeres in these species. However, FISH experiments in *Festuca* revealed terminal localization of only the plant canonical variant TTTAGGG, while the alternative TTAGGG variant was found in interstitial regions of some chromosomes (Fig. 4). This underlines a need for experimental validation of predicted species with noncanonical telomeres based on the above-mentioned clues (TR template + abundant tandem repeats).

Experimental validation in a subset of species

Selected species (overviewed in Methods S1) were experimentally investigated to test our predictions. We employed: Fluorescence *in situ* hybridization (FISH); Analysis of Terminal Restriction Fragments (TRF), including examination of their sensitivity to BAL31 nuclease; and Telomere Repeat Amplification Protocol (TRAP) with subsequent sequencing of *in vitro* synthesized telomerase products and quantitative real-time PCR analysis (RT-qPCR) of transcripts of TR variants.

FISH signals with candidate telomere probes Using custom, high-affinity probes for candidate telomeric sequences (Methods S3), the expected punctate pattern of telomere motifs in Caryophyllales was detected (Fig. 4a). *F. tataricum* shows the co-occurrence of two distinct telomeric motifs, TTTCGG and TTTCAGG, while *F. esculentum* displays only the TTTCGGG motif. The canonical motif is absent in both species (Figs 4a, S1). The TTTCGGG motif was present also in *E. guttata* from the order Lamiales (Fig. 4b). In Ranunculales, TTTCGGG (*D. consolida* and *A. napelus*) and TTCAGGG motifs (*P. somniferum*) were confirmed (Fig. 4d). In Solanales, *Solanum melongena* and *Salpiglossis sinuata* showed signals of the TTCAGGG motif only, indicating the loss of the ancestral TTTAGGG telomere in these species (Fig. 4e). Other Solanales species with more functional TR copies whose templates encode variant telomere motifs showed variability in telomeric signals. Different patterns were found in species possessing two distinct motifs, that is TTTTTTAGGG and TTTTTTCGGG in *V. foetida* or TTTAGGG and TTCAGGG in other species (Figs 4e,f, S1). Whereas almost complete colocalization of the signals was observed in *C. chinense* and *P. chamaesarachoides*, preferential signal localization of individual motifs at distinct telomeres was observed in *C. annuum*, *J. sinuosa*, *A. belladonna* and *V. foetida*. To gain a better view of the mutual distribution of TTTAGGG and TTCAGGG motifs, we performed 3D imaging followed by deconvolution, showing partial colocalization in *C. annuum*, *J. sinuosa* and *P. chamaesarachoides* (Figs 4e, S2). For all selected species with a telomere motif different from the canonical one, dual FISH labelling was used to confirm whether the TTTAGGG motif is absent (Fig. S1). Clear extended signals of the TTCAGGG probe were found on DNA fibres prepared from *S. melongena* (Fig. S1). This is in agreement with the probe detection on isolated nuclei (Fig. 4e). In contrast to Solanales, where FISH results confirmed the presumed coexistence of different motifs in a single species, FISH signals confirmed the telomere location of TTTAGGG, while TTAGGG was observed in interstitial regions in *F. pratensis* (Figs 4c, S1).

Sensitivity of TRF signals and telomere read counts to BAL31 nuclease digestion Terminal localization of the predicted telomere repeats was probed by two strategies employing treatment of high-molecular-weight genomic DNA by BAL31 nuclease. Terminal Restriction Fragments, generated with *MseI* endonuclease digestion and separated by PFGE, indicate progressive shortening and loss of signal intensity with increasing time of BAL31 treatment (Fig. 5b). TRF shortening was evaluated with the Walter tool (Lycka *et al.*, 2021; Fig. S3). An alternative approach using sequencing of whole-genome datasets from high-molecular-weight DNAs treated and untreated with BAL31 nuclease (BAL31-NGS; Peska *et al.*, 2017) confirmed depletion

Table 1 Overview of newly predicted and previously published species with unusual telomere sequence motifs.

Order	Species name (experimentally tested in this work in bold)	TR-like genes	TR templates (5′–3′)	Telomere expected (TRFi ⇔ Template) or (published)
Alismatales	*Zostera marina*	2	CTAACCCTAA	TTAGGG (Peska *et al*., 2020)
Asparagales	*Allium cepa*	2	AACCGAGCCCATAACCGA	CTCGGTTATGGG (Fajkus *et al*., 2016)
	Allium sativum	4	AACCGAGCCCATAACCG	CTCGGTTATGGG (Fajkus *et al*., 2016)
	Asparagus kiusianus	3	ACCCTAACCC	TTAGGG (Sykorova *et al*., 2003)
	Asparagus officinalis	4	ACCCTAACCC	TTAGGG (Sykorova *et al*., 2003)
	Asparagus setaceus	5	ACCCTAACCC	TTAGGG (Sykorova *et al*., 2003)
	Hemerocallis citrina	1	CCTAACCCTA	TTAGGG (Sykorova *et al*., 2003)
	Vanilla planifolia	1	ATCCCTAATCC	ATTAGGG
Canellales	*Warburgia ugandensis*	nd	CCCTGAACCCTAACCCT	TTAGGG TTCAGGG
Caryophyllales	***Fagopyrum esculentum***	4	AAAACCCGAACCC AACCCGAAACC AAACCCGAACCC	T(2-4)CGGG
	Fagopyrum tataricum	6	TGAACCTGAAC AACCGAACCGAAC	T(2-3)CAGG T(2-3)CGG
Fabales	*Quillaja saponaria*	1	CCTAACCCTAACC	TTAGGG
Lamiales	***Erythranthe guttata***	1	AACCCGAAACC	TTTCGGG
	Genlisea hispidula	nd	GAACCTGAACC	T(2-3)CAGG (Tran *et al*., 2015)
Malpighiales	*Euphorbia esculа*	5	ACCCTAACC	TTAGGG
Poales	*Lolium perenne*	3	CTAACCCTAAACC	T(2-3)AGGG
	Festuca pratensis	3	CTAACCCTAAACC	T(2-3)AGGG
	Zea mays	1	CTAACCCTAAACCCTA	T(2-3)AGGG
Ranunculales	***Delphinium consolida***	2	AAACCCGAAAAC	T(3-4)CGGG
	Ranunculus acris	2	GAACCCTGAAC	TTCAGGG
	Ranunculus repens	2	AACCCTGAACC	TTCAGGG
	Kingdonia uniflora	6	CTGATACCCGCCTGA	TATCAGGCGGG
	Papaver armeniacum	4	CCCTGAACCCTGA	TTCAGGG
	Papaver atlanticum	2	CCCTGAACCCTGA	TTCAGGG
	Papaver bracteatum	2	CCCTGAACCCTGA	TTCAGGG
	Papaver californicum	1	CCTGAACCCTGA	TTCAGGG
	Papaver nudicaule	2	CTGAACCCTGA	TTCAGGG
	Papaver somniferum	3	CCTGAACCCTGAACC	TTCAGGG
Solanales	*Atropa belladona*	nd	TAAACCCTAAACC TGAACCCT	TTTAGGG TTCAGGG
	Capsicum annuum	3	CTAAACCCTAAACC TGAACCCTGAACC	TTTAGGG TTCAGGG
	Capsicum baccatum	4	CTAAACCCTAAACC TGAACCCTGAACC	TTTAGGG TTCAGGG
	Capsicum chinense	3	CTAAACCCTAAACC TGAACCCTGAACC	TTTAGGG TTCAGGG
	Cestrum elegans	2	TAAAAACCCTAAAAACT	TTTTTTAGGG (Peska *et al*., 2015)
	Datura stramonium	4	AAACCCTGAACC AAATCCTGAACT GAACCCTGAACT	TTCAGGG
	Jaltomata sinuosa	3	CTAAACCCTAAACC CCTGAACCCTGA	TTTAGGG TTCAGGG
	Salpiglossis sinuata	2	AACCCTAAACCC CTGAACCCTGAACCCT	TTTAGGG TTCAGGG
	Solanum clarkiae	6	CCTGAACCCTGA	TTCAGGG
	Solanum medicagineum	4	CCTGAACCCTGA	TTCAGGG
	Solanum melongena	5	CCTGAACCCTGA	TTCAGGG
	Solanum sejunctum	8	CCTGAACCCTGA	TTCAGGG
	Vestia foetida	2	GAAAAACCCGAAAAAC CTAAAACCCTAAAAACT	TTTTT(T)CGGG TTTTT(T)AGGG

These were predicted based on a combination of knowledge of the telomerase RNA (TR) template and abundant tandem repeats (summarized in Supporting Information Tables S1, S2, respectively).

of reads corresponding to predicted telomere repeats in BAL31 treated sequence libraries in both analysed species (*F. tataricum*, *V. foetida*; Fig. 5a; Table S2).

Synthesis of variant telomere repeats and differential contribution of TR paralogs To identify sequences genuinely synthesized by telomerase, TRAP assays and sequencing of the PCR

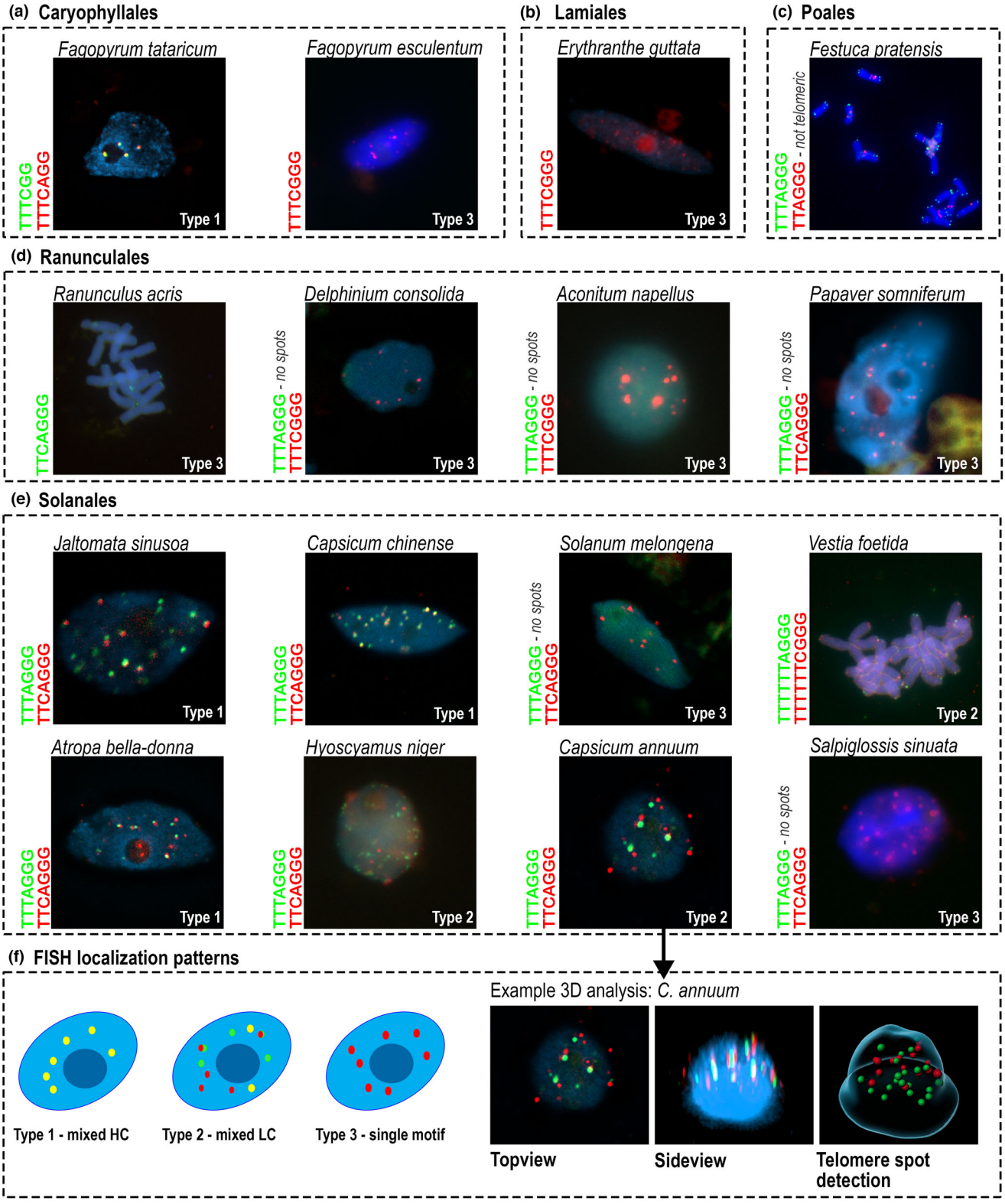

Fig. 4 Fluorescence *in situ* hybridization mapping of candidate telomeric sequences. Detection of probes targeting the predicted telomeric motifs in Caryophyllales (a), Lamiales (b), Poales (c), Ranunculales (d) and Solanales (e). Sequences detected by FISH, Fluorescene *in situ* hybridization, in different species are indicated on the left of the image. Distinct localization patterns of telomeric sequences (type 1 – mixed with highly colocalization of signals (HC); type 2 – with low colocalization of signals (LC); type 3 – single motif) are indicated at the bottom right for each species analysed, schematically depicted in (f). 3D imaging was performed for more precise signal colocalization is shown for *Capsicum annuum* and others in Supporting Information Fig. S2.

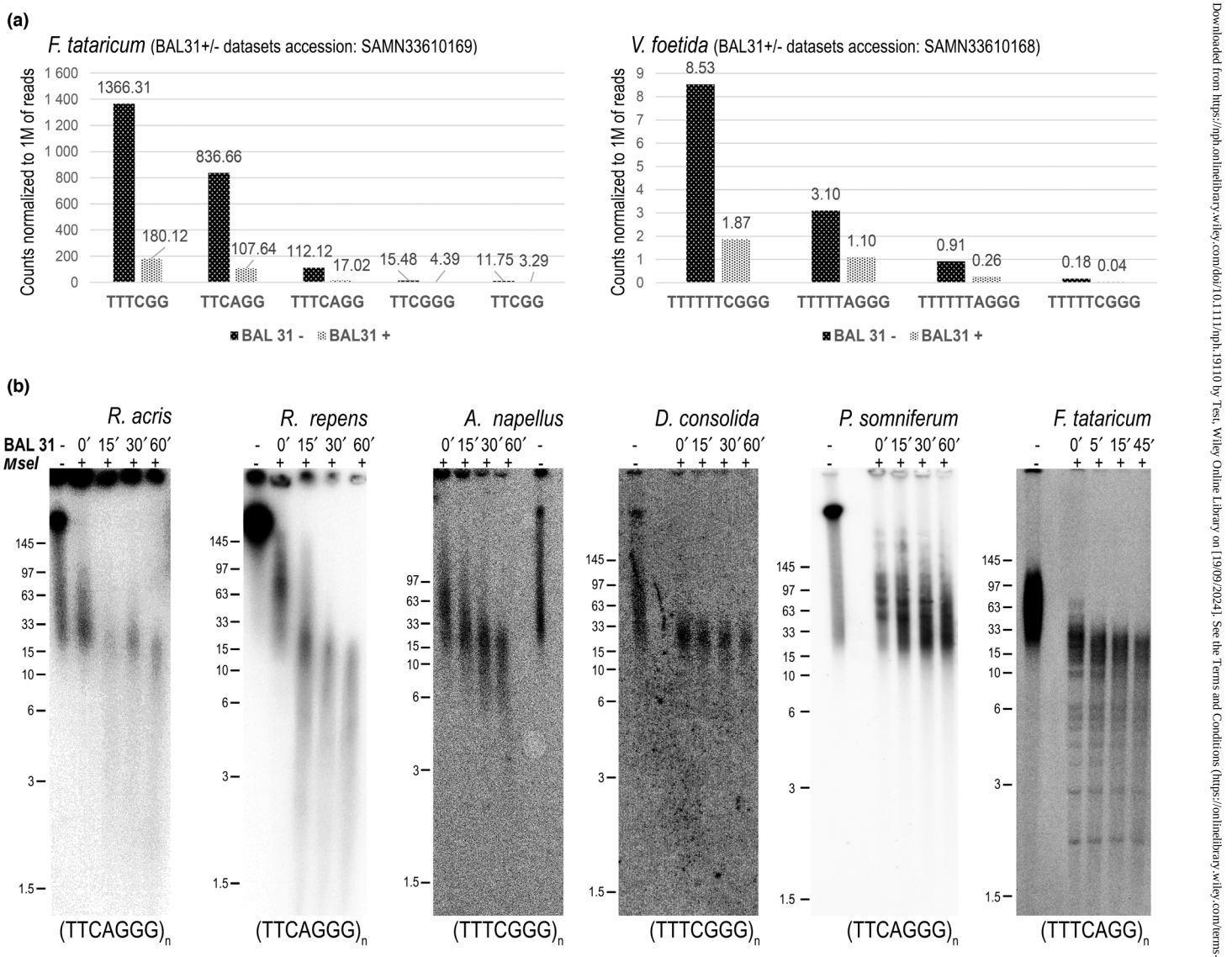

Fig. 5 Sensitivity of the candidate telomere sequences to BAL31 nuclease. (a) BAL31-NGS analysis of *Fagopyrum tataricum* and *Vestia foetida*. Normalized read counts (*y*-axis) and corresponding tandem motifs (*x*-axis) show relative abundance of these reads/motifs in BAL31 treated (+) and untreated (−) sequence libraries (generated from high-molecular-weight genomic DNAs). Only results for the candidate telomere motifs are shown, complete results, including all motifs identified by Tandem Repeats Finder (TRFi), are summarized in Supporting Information Table S2. (b) Terminal Restriction Fragment (TRF) analysis showing high-molecular-weight DNAs predigested with BAL31 for four-time intervals and then digested with *Mse*I and separated by pulse-field gel electrophoresis (PFGE). Hybridization signals of radioactively labelled concatemeric probes (indicated below images) are shown including molecular weight marker (on the left of each panel, kilobases). The signal changes with increasing time of BAL 31 treatment were assessed by WALTER tool (Lycka *et al*., 2021), see Fig. S3.

products in *F. esculentum*, *F. tataricum*, *D. consolida* and *P. somniferum* were employed (Fig. 6a,b). The sequenced TRAP products (Figs 6b, S4) show synthesis of TTCAGGG motif in *P. somniferum* and TTTCGGG motif in *D. consolida*. In *F. esculentum*, TRAP products indicate inaccurate synthesis of the repeats in TRAP (Fig. 6b,d), that is 61% of sequenced motifs corresponded to TTTTCGGG, 19% to TTTCGGG and 19% to other motifs. In *F. tataricum*, only motifs corresponding to T(2-4)CAGG variants were detected in TRAP reactions (Methods S5). However, *k*-mer analysis of long-read genome sequencing data (PacBio) generated in genome sequencing projects of *F. esculentum* and *F. tataricum* (Zhang *et al*., 2017; Penin *et al*., 2021) indicate a more realistic abundance of particular motif variants in their genomes (Fig. 6d; Table S2). Short T(2-3)CAGG and T(2-3)CGG arrays identified in *F. tataricum* appear to be intermingled at telomeres (no homogeneous T(2-3)CAGG or T(2-3)CGG arrays were observed in the PacBio data; Fig. 6d; Table S2). These mixed arrays together form long stretches (thousands of bp) as illustrated by reconstructed pseudo-chromosome ends from *de novo* assembled PacBio reads spanning chromosome-specific subterminal region and nonspecific telomere region (Fig. 6c). The inaccurate synthesis of telomere repeats in both *Fagopyrum* species may result from imprecise TR annealing to telomere, the usage of different TR

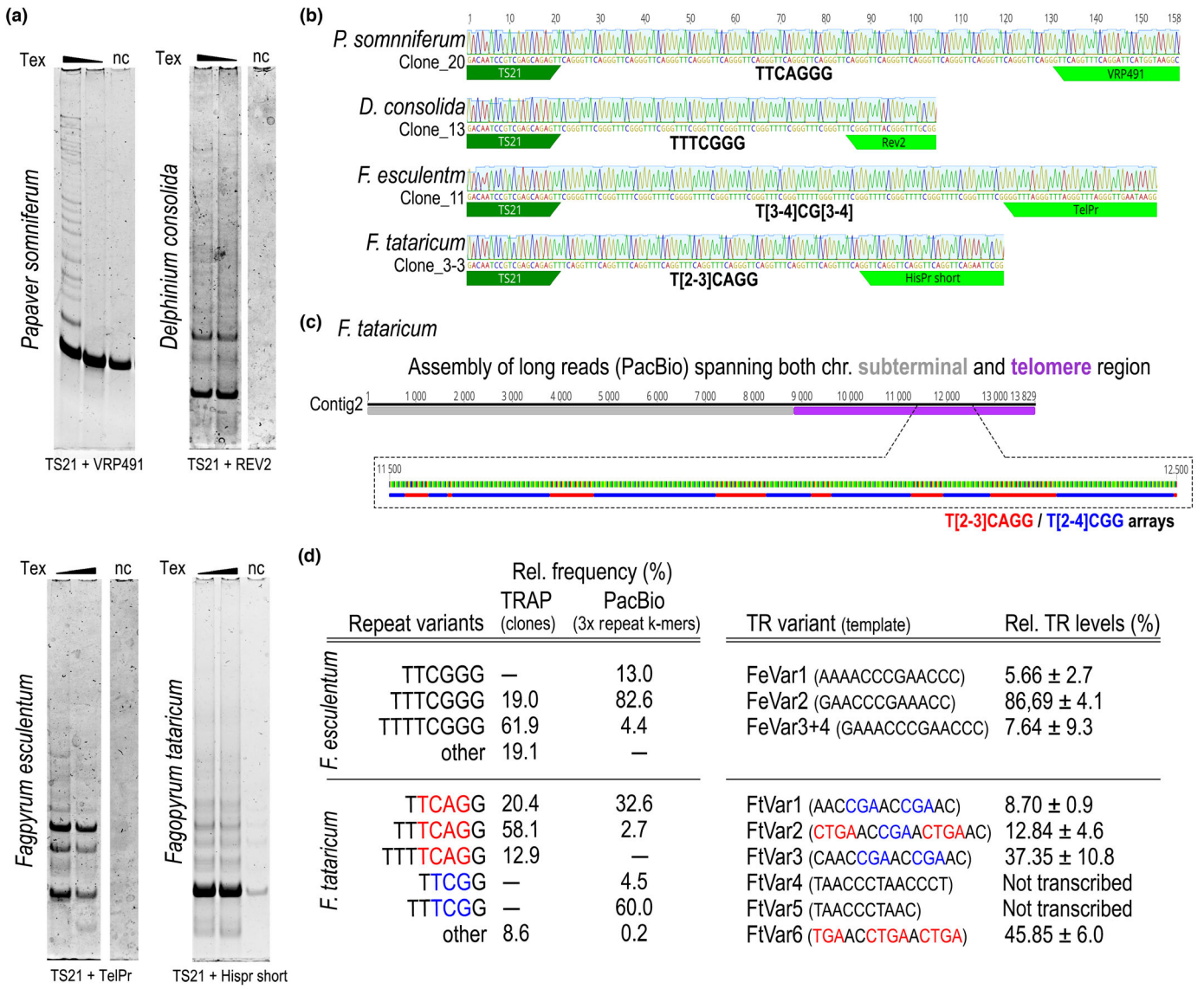

Fig. 6 (a) Analysis of telomerase activity by TRAP (telomere repeat amplification protocol). TRAP reactions were performed with two concentrations of crude telomerase extracts (Tex) from seedlings (*Papaver somniferum*, *Fagopyrum esculentum* and *Fagopyrum tataricum*) and root tips (*D. consolida*). Products were separated by polyacrylamide gel electrophoresis, stained and imaged. A reaction without extract was used as a negative control (nc). (b) Results of sequencing of telomerase products cloned from positive TRAP reactions. (c) Reconstruction of chromosome end from *F. tataricum* long reads (PacBio). Only reads spanning nonspecific telomere region and chromosome arm-specific subterminal region were used for the assembly. Telomeres showed intermingled arrays of T(2-3)CAGG and T(2-3)CGG motif variants. (d) Summary of the analysis of relative distribution of repeat variants in TRAP clones (Supporting Information Fig. S4) and *k*-mer analysis of PacBio data (Table S2). Relative transcript levels of TR (telomerase RNA) paralogs were examined by RT-qPCR (quantitative reverse transcription polymerase chain reaction). TR variants from *F. tataricum* that may be responsible for synthesis of corresponding repeat motifs are highlighted in colours (blue, red – similarly as in panel (c)).

paralogs, or a combination of both factors. To distinguish between different processes, relative transcription of TR paralogs was measured by RT-qPCR using variant-specific primers (Fig. 6d; Methods S6). While each of the TR paralogs from *F. esculentum* is capable of generating T(2-4)CGGG motif variants, the situation in *F. tataricum* is even more interesting. While FtVar1,2,3 variants (*c.* 55% of total TR levels) code for T(2-3)CGG motifs, FtVar6 (*c.* 45% of total TR levels) can generate the T(2-3)CAGG variant only. Moreover, transcripts of variants FtVar4,5 were not detected.

Origin and fate of TR paralogs

Plant genomes have undergone a series of polyploidization and diploidization events during their evolution, which gave rise to duplicated or even multiplicated large DNA segments (or whole chromosomes) within a single genome. The divergence level of such duplicated segments depends on how advanced postpolyploid diploidization has become (i.e. reversion of a structurally polyploid genome to a diploid-like genome), and how exposed the duplicated segment is to selection pressures. Duplicated genes

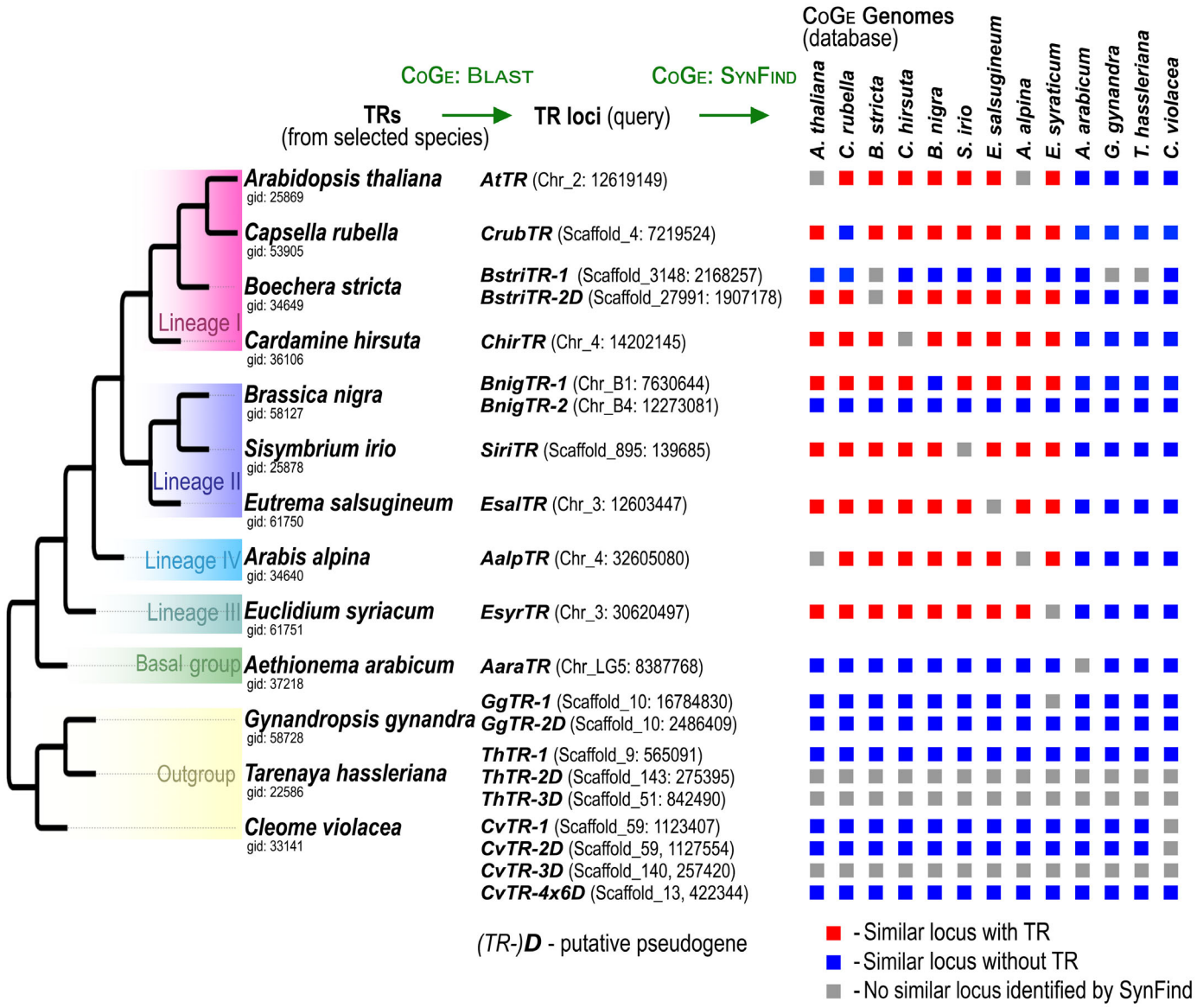

Fig. 7 Resolving the genomic context of telomerase RNA (TR) genes and paralogs using the CoGe (CoGe, comparative genomics) platform (Lyons & Freeling, 2008) in examples from Brassicales. TR loci were identified in the CoGe genomes in example diploid (ancestral polyploid) species by using CoGe : Blast with corresponding TRs identified in Supporting Information Table S1. Closest annotated feature (gene, exon) to the TR hits was used as a query for synteny analysis using the SynFind tool (Tang *et al*., 2015) against selected CoGe Genomes (for details, see Table S3). The absence of the region (in a respective genome) showing similar gene collinearity (as the TR query) is marked with grey boxes. Regions with conserved collinearity with either present or absent TR are indicated by red and blue boxes, respectively. Putative TR pseudogenes, that is truncated TR variants and/or variants lacking promoter are named as D in the suffix.

generated by polyploidization and large-scale duplications usually preserve gene synteny, that is duplicated genes co-localize in mutually colinear gene blocks. However, gene copies generated in a small-scale manner via tandem, retro- or trans-duplications are usually interspersed throughout the genome without any conserved gene collinearity of paralogs. To ask whether paralogs of the TR gene originated from polyploidization events or local duplications, they were analysed by using the SynFind tool (Tang *et al*., 2015) in species of Brassicaceae, involving diploid species (ancient polyploids) with a single copy or more TR copies (Fig. 7; Table S3). TR loci showed conserved synteny limited to species from lineages I to IV, whereas the corresponding locus without TR is evolutionarily conserved even outside of these lineages (e.g. in Cleomaceae). Lack of conserved TR synteny within early diverged clades (while loci are usually conserved; Fig. 7, blue boxes) indicates a high tolerance of the TR gene to moving into diverse genomic contexts. Moreover, TR copies in the diploids we analysed seem to result from small-scale duplications rather than having been created in ancient polyploidization events. In species with extra TR copies (*Boechera stricta*, *Cleome violacea*, *Gynandropsis gynandra* and *Tarenaya hassleriana*, except *Brassica nigra*), additional TR copies underwent pseudogenization

(truncated TR variants or variants lacking type 3 snRNA promoter). Interestingly, an ancestral variant in *B. stricta BstriTR-2D* (which shares synteny with other related Brassicaceae) was pseudogenized, while a newly emerged copy took over its function (*BstriTR-1*).

Discussion

Evolutionary changes in telomeres, previously reported based on scattered experimental findings only in individual plant species (Peska *et al.*, 2015, 2020; Tran *et al.*, 2015; Fajkus *et al.*, 2016) or specific phylogenetic taxa (Sykorova *et al.*, 2003) have been analysed here systematically across the entire land plant evolutionary tree based on available or newly generated genomic data. Due to the reverse approach applied in this work, in which the bioinformatic identification of putative unusual telomeres was the first step of the extensive evolutionary screen, the experimental workload has been substantially reduced. Subsequent validation experiments demonstrated terminal localization of candidate telomere repeats and their synthesis by telomerase in all tested species, except *Festuca* from Poales, where the putative TTAGGG tandem motif showed interstitial localization. Interestingly, species with more TRs whose templates code for different telomere motifs (listed in Table 1) showed a varied distribution of respective repeats at chromosome ends (or interphase nuclei). FISH signals on *F. tataricum*, *V. foetida* chromosomes, *Capsicum* and the other examined Solanaceae interphase nuclei displayed specific patterns corresponding either to mixed telomere motifs (e.g. TTTCAGG and TTTCGG in *F. tataricum*) or preferential motif distribution at specific chromosome arms and interphase nuclear foci (e.g. TTTAGGG and TTCAGGG in *C. annuum*). While mixed telomeres or an overall dominance of one of the telomere types could be explained as a result of concerted evolution of TRs or a differential function and expression of corresponding TRs, unequal (preferential) motif distribution at different chromosome ends within the same cell may suggest a possible functional significance in telomere clustering (e.g. in meiosis Harper *et al.*, 2004; Scherthan, 2007). Besides FISH experiments, the distribution of telomere variants was examined in detail in *F. tataricum* and *F. esculentum* in sequenced TRAP products and available genomic long reads (PacBio), including relative transcript levels of their TR paralogs.

This work points to a fundamental role of TR duplication in telomere evolution, in accordance with the generally accepted concept (Ohno, 1970; Louis, 2007). In view of structural RNAs with type 3 snRNA promoters (e.g. spliceosomal, 7SL, MRP RNAs), which frequently occur in multiple copies with overall low synteny conservation (Marz *et al.*, 2008), our similar observation in the case of plant TRs and their paralogs is not surprising. TRs were also shown to be prone to form pseudogenes independently in many species, consistent with the observation in, for example mammalian snoRNAs, which can be spread through the genome as nonautonomous mobile elements (Weber, 2006). Prominent examples are *Alu* elements, which evolved from 7SL RNA and represent *c.* 10% of the human genome. Overall, the variation among snRNA paralogs and their relative expression may thus represent a very powerful evolutionary tool (e.g. variant spliceosomes contribute to gene regulation in tissue/cell-specific manner; Mabin *et al.*, 2021).

This work highlights the role of frequent TR duplication and subsequent change in establishment of noncanonical telomeres in plants. Telomeres, as well as centromeres, represent key chromosome structures ensuring stable chromosome transmission over generations. Centromere DNA and protein variation is a common phenomenon in plants, which is regarded as a critical factor responsible for the formation of new species. Alterations in centromere location, function or gene transcription and epigenetic status in pericentromeric regions promote genetic divergence of new species (reviewed, e.g. in Malik & Henikoff, 2009; Comai *et al.*, 2017). In contrast to centromeres, telomeres have appeared to be highly uniform which – in this view – did not provide much space for their active contribution to speciation. Consequently, telomere studies in plants have usually been purely analytic, mostly lacking an adequate evolutionary context (with rare exceptions as, e.g. Fulneckova *et al.*, 2013; Cervenak *et al.*, 2021). The recent identification of TRs across the plant phylogeny allowed us to address this issue with an unprecedented evolutionary representation of taxa. Previous studies described specific telomere changes as 'exceptions' lacking the *Arabidopsis*-type TTTAGGG telomere motif, based on *ad hoc* experimental data. In this work, we were able to apply, for the first time, a systematic approach, in which we first predicted telomeric DNA and the corresponding TRs across a broad phylogenetic span, based on available or newly generated genomic data (Table 1). Subsequently, we demonstrated experimentally the presence of predicted telomeres and TRs in selected examples. Our results provide a new interpretation of telomere evolutionary changes. Rather than a static list of exceptions from a common plant telomere DNA motif, we offer the view of an ongoing battle among frequently duplicated TR genes. These tend to be frequently mutated and evolve, while minimizing the risk of the loss of viability, as proposed in our working hypothesis (Fig. 1). This updated view of the telomere evolution process highlights the role of the 'telomere factor' in the processes of speciation, cross compatibility of species (e.g. in Solanaceae or buckwheat breeding) or meiotic segregation. In this view, telomere DNA is considered as a product of the intricate and dynamic evolution of TRs.

Acknowledgements

This research was supported by the Czech Science Foundation project 20-01331X. VP, and Bal31-NGS data generation was supported by the Czech Science Foundation project 23-06643S. Collaboration with Potato Research Institute, v.v.i., was supported by the European Regional Development Fund Project 'SINGING PLANT' (CZ.02.01/0.0/0.0/16_026/0008446). CF Plant Sciences and CF Genomics of CEITEC, Masaryk University, are gratefully acknowledged for the technical support and obtaining of the scientific data presented in this paper. Computational resources were provided by the e-INFRA CZ project

(ID:90140), supported by the Czech MEYS and by the ELIXIR-CZ project (ID:90255), part of the international ELIXIR infrastructure. Samuel C. Beer (Angelica Mill, USA), Barbora Švecová (Central Institute for Supervising and Testing in Agriculture, Brno, CZ) and Magdaléna Chytrá (The Botanical Garden of the Faculty of Science, Masaryk University, CZ) kindly provided *Fagopyrum* and Solanaceae plant material.

Competing interests

None declared.

Author contributions

PF, JF and VP planned and designed the research. MZ, MF, MM, AH and DK performed experiments. Bioinformatic analyses were performed by PF, AK and VP. SD, AO, SG and JP provided plant material and sequencing data. MZ, PF and JF wrote the manuscript. All authors contributed to the manuscript preparation with comments and approved the final manuscript. MZ and PF contributed equally to this work.

ORCID

Steven Dodsworth https://orcid.org/0000-0001-6531-3540
Jiří Fajkus https://orcid.org/0000-0002-3112-1716
Petr Fajkus https://orcid.org/0000-0001-8891-7574
Michal Franek https://orcid.org/0000-0002-9627-1301
Sònia Garcia https://orcid.org/0000-0002-3143-0527
Agata Kilar https://orcid.org/0000-0003-3876-6581
David Kopecký https://orcid.org/0000-0002-2834-1734
Martin Mátl https://orcid.org/0000-0002-7215-3386
Andrés Orejuela https://orcid.org/0000-0002-3511-1478
Vratislav Peška https://orcid.org/0000-0002-0526-2988

Data availability

All data are included in the article and Supporting Information. Sequencing data generated in this study are available at NCBI SRA (Bioproject: PRJNA941781).

References

Adamusova K, Khosravi S, Fujimoto S, Houben A, Matsunaga S, Fajkus J, Fojtova M. 2020. Two combinatorial patterns of telomere histone marks in plants with canonical and non-canonical telomere repeats. *The Plant Journal* **102**: 678–687.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27**: 573–580.

Blackburn EH. 1991. Telomeres. *Trends in Biochemical Sciences* **16**: 378–381.

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Analytical Biochemistry* **72**: 248–254.

Brand CL, Levine MT. 2022. Cross-species incompatibility between a DNA satellite and the Drosophila Spartan homolog poisons germline genome integrity. *Current Biology* **32**: 2962–2971.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Cervenak F, Sepsiova R, Nosek J, Tomaska L. 2021. Step-by-step evolution of telomeres: lessons from yeasts. *Genome Biology and Evolution* **13**: evaa268.

Comai L, Maheshwari S, Marimuthu MPA. 2017. Plant centromeres. *Current Opinion in Plant Biology* **36**: 158–167.

Fajkus J, Sykorova E, Leitch AR. 2005. Telomeres in evolution and evolution of telomeres. *Chromosome Research* **13**: 469–479.

Fajkus P, Adamik M, Nelson ADL, Kilar AM, Franek M, Bubenik M, Frydrychova RC, Votavova A, Sykorova E, Fajkus J *et al.* 2023. Telomerase RNA in Hymenoptera (Insecta) switched to plant/ciliate-like biogenesis. *Nucleic Acids Research* **51**: 420–433.

Fajkus P, Kilar A, Nelson ADL, Hola M, Peska V, Goffova I, Fojtova M, Zachova D, Fulneckova J, Fajkus J. 2021. Evolution of plant telomerase RNAs: farther to the past, deeper to the roots. *Nucleic Acids Research* **49**: 7680–7694.

Fajkus P, Peska V, Sitova Z, Fulneckova J, Dvorackova M, Gogela R, Sykorova E, Hapala J, Fajkus J. 2016. Allium telomeres unmasked: the unusual telomeric sequence (CTCGGTTATGGG)n is synthesized by telomerase. *The Plant Journal* **85**: 337–347.

Fajkus P, Peska V, Zavodnik M, Fojtova M, Fulneckova J, Dobias S, Kilar A, Dvorackova M, Zachova D, Necasova I *et al.* 2019. Telomerase RNAs in land plants. *Nucleic Acids Research* **47**: 9842–9856.

Fitzgerald MS, McKnight TD, Shippen DE. 1996. Characterization and developmental patterns of telomerase expression in plants. *Proceedings of the National Academy of Sciences, USA* **93**: 14422–14427.

Fojtová M, Fajkus P, Polanská P, Fajkus J. 2015. Terminal restriction fragments (TRF) method to analyze telomere lengths. *Bio-Protocol* **5**: e1671.

Fulneckova J, Sevcikova T, Fajkus J, Lukesova A, Lukes M, Vlcek C, Lang BF, Kim E, Elias M, Sykorova E. 2013. A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biology and Evolution* **5**: 468–483.

Gibney G, Baxevanis AD. 2011. Searching NCBI databases using ENTREZ. *Current Protocols in Bioinformatics* **Chapter 1**: 1.3.1–1.3.25.

Goffova I, Vagnerova R, Peska V, Franek M, Havlova K, Hola M, Zachova D, Fojtova M, Cuming A, Kamisugi Y *et al.* 2019. Roles of RAD51 and RTEL1 in telomere and rDNA stability in *Physcomitrella patens*. *The Plant Journal* **98**: 1090–1105.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**: 644–652.

Harper L, Golubovskaya I, Cande WZ. 2004. A bouquet of chromosomes. *Journal of Cell Science* **117**(Pt 18): 4025–4032.

Ijdo JW, Wells RA, Baldini A, Reeders ST. 1991. Improved telomere detection using a telomere repeat probe (TTAGGG)n generated by PCR. *Nucleic Acids Research* **19**: 4780.

Louis EJ. 2007. Evolutionary genetics: making the most of redundancy. *Nature* **449**: 673–674.

Louis EJ, Vershinin AV. 2005. Chromosome ends: different sequences may provide conserved functions. *BioEssays* **27**: 685–697.

Lycka M, Peska V, Demko M, Spyroglou I, Kilar A, Fajkus J, Fojtova M. 2021. WALTER: an easy way to online evaluate telomere lengths from terminal restriction fragment analysis. *BMC Bioinformatics* **22**: 145.

Lyons E, Freeling M. 2008. How to usefully compare homologous plant genes and chromosomes as DNA sequences. *The Plant Journal* **53**: 661–673.

Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang HB, Wang XY, 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 Physiology* **148**: 1772–1781.

Mabin JW, Lewis PW, Brow DA, Dvinge H. 2021. Human spliceosomal snRNA sequence variants generate variant spliceosomes. *RNA* **27**: 1186–1203.

Malik HS, Henikoff S. 2009. Major evolutionary transitions in centromere complexity. *Cell* **138**: 1067–1082.

Marz M, Kirsten T, Stadler PF. 2008. Evolution of spliceosomal snRNA genes in metazoan animals. *Journal of Molecular Evolution* **67**: 594–607.

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**: 2933–2935.

Ohno S. 1970. *Evolution by gene duplication*. New York, NY, USA: Springer Science & Business Media.

Penin AA, Kasianov AS, Klepikova AV, Kirov IV, Gerasimov ES, Fesenko AN, Logacheva MD. 2021. High-resolution transcriptome atlas and improved genome assembly of common buckwheat, *Fagopyrum esculentum*. *Frontiers in Plant Science* **12**: 612382.

Peska V, Fajkus P, Bubenik M, Brazda V, Bohalova N, Dvoracek V, Fajkus J, Garcia S. 2021. Extraordinary diversity of telomeres, telomerase RNAs and their template regions in Saccharomycetaceae. *Scientific Reports* **11**: 12784.

Peska V, Fajkus P, Fojtova M, Dvorackova M, Hapala J, Dvoracek V, Polanska P, Leitch AR, Sykorova E, Fajkus J. 2015. Characterisation of an unusual telomere motif (TTTTTTAGGG)n in the plant *Cestrum elegans* (Solanaceae), a species with a large genome. *The Plant Journal* **82**: 644–654.

Peska V, Matl M, Mandakova T, Vitales D, Fajkus P, Fajkus J, Garcia S. 2020. Human-like telomeres in *Zostera marina* reveal a mode of transition from the plant to the human telomeric sequences. *Journal of Experimental Botany* **71**: 5786–5793.

Peska V, Sitova Z, Fajkus P, Fajkus J. 2017. BAL31-NGS approach for identification of telomeres *de novo* in large genomes. *Methods* **114**: 16–27.

Podlevsky JD, Chen JJ. 2016. Evolutionary perspectives of telomerase RNA structure and function. *RNA Biology* **13**: 720–732.

Scherthan H. 2007. Telomere attachment and clustering during meiosis. *Cellular and Molecular Life Sciences* **64**: 117–124.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B *et al.* 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**: 676–682.

Schrumpfova PP, Fajkus J. 2020. Composition and function of telomerase – a polymerase associated with the origin of eukaryotes. *Biomolecules* **10**: 1425.

Sykorova E, Lim KY, Kunicka Z, Chase MW, Bennett MD, Fajkus J, Leitch AR. 2003. Telomere variability in the monocotyledonous plant order Asparagales. *Proceedings of the Biological Sciences* **270**: 1893–1904.

Tang HB, Bomhoff MD, Briones E, Zhang LS, Schnable JC, Lyons E. 2015. SynFind: compiling syntenic regions across any set of genomes on demand. *Genome Biology and Evolution* **7**: 3286–3298.

Tran TD, Cao HX, Jovtchev G, Neumann P, Novak P, Fojtova M, Vu GT, Macas J, Fajkus J, Schubert I *et al.* 2015. Centromere and telomere sequence alterations reflect the rapid genome evolution within the carnivorous plant genus Genlisea. *The Plant Journal* **84**: 1087–1099.

Weber MJ. 2006. Mammalian small nucleolar RNAs are mobile genetic elements. *PLoS Genetics* **2**: e205.

Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R, Grolemund G, Hayes A, Henry L, Hester J. 2019. Welcome to the Tidyverse. *Journal of Open Source Software* **4**: 1686.

Zhang L, Li X, Ma B, Gao Q, Du H, Han Y, Li Y, Cao Y, Qi M, Zhu Y *et al.* 2017. The tartary buckwheat genome provides insights into rutin biosynthesis and abiotic stress tolerance. *Molecular Plant* **10**: 1224–1237.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 FISH mapping of candidate telomeric probes.

Fig. S2 3D imaging of selected Solanaceae species.

Fig. S3 Analysis of TRF signals corresponding to Fig. 5(b) by WALTER tool.

Fig. S4 Sequencing of cloned TRAP products.

Fig. S5 RT-qPCR calibration curves.

Methods S1 Overview of experimental species.

Methods S2 Preparation of gDNA, hmwDNA, RNA, nuclei, DNA fibres, chromosomes and NGS libraries.

Methods S3 FISH.

Methods S4 TRF analysis.

Methods S5 TRAP assay.

Methods S6 RT-qPCR.

Table S1 Identification of plant TRs.

Table S2 Identification of abundant tandem repeats.

Table S3 Synteny analysis.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.