



OPEN

Extraordinary diversity of telomeres, telomerase RNAs and their template regions in Saccharomycetaceae

Vratislav Peska¹✉, Petr Fajkus^{1,3}, Michal Bubeník^{1,3}, Václav Brázda¹, Natália Bohálová^{1,2}, Vojtěch Dvořáček¹, Jiří Fajkus^{1,3} & Sònia Garcia⁴✉

Telomerase RNA (TR) carries the template for synthesis of telomere DNA and provides a scaffold for telomerase assembly. Fungal TRs are long and have been compared to higher eukaryotes, where they show considerable diversity within phylogenetically close groups. TRs of several Saccharomycetaceae were recently identified, however, many of these remained uncharacterised in the template region. Here we show that this is mainly due to high variability in telomere sequence. We predicted the telomere sequences using Tandem Repeats Finder and then we identified corresponding putative template regions in TR candidates. Remarkably long telomere units and the corresponding putative TRs were found in *Tetrapisispora* species. Notably, variable lengths of the annealing sequence of the template region (1–10 nt) were found. Consequently, species with the same telomere sequence may not harbour identical TR templates. Thus, TR sequence alone can be used to predict a template region and telomere sequence, but not to determine these exactly. A conserved feature of telomere sequences, tracts of adjacent Gs, led us to test the propensity of individual telomere sequences to form G4. The results show highly diverse values of G4-propensity, indicating the lack of ubiquitous conservation of this feature across Saccharomycetaceae.

Chromosomes are capped by nucleoprotein structures called telomeres, which are involved in several important functions such as the regulation of gene expression, recombination, mitosis and meiosis^{1–4}. Telomeres protect chromosome termini from eliciting the DNA damage response and play a fundamental role in longevity and cell proliferation⁵. Most telomeres carry 3' overhangs that can be generated by incomplete synthesis at the lagging strand after degradation of the RNA primer of the distal-most Okazaki fragment and resection, and/or by resection of a blunt-end telomere intermediate resulting from leading strand synthesis^{6–8}. This end-replication problem is solved by an enzyme called telomerase⁹. While a deficit in telomerase would translate into shorter telomeres over each replication, triggering senescence, its continued or dysregulated expression could unbalance growth control, potentially leading to oncogenesis, if control mechanisms do not direct the cell e.g., to apoptosis¹⁰. Telomerase is responsible for telomere maintenance in most eukaryotes and had originated early in eukaryotic evolution in association with linearization of the chromosomes^{11–14}. Telomerase consists of two main parts; a catalytic protein component—telomerase reverse transcriptase (TERT), and the telomerase RNA (TR) subunit. TR contains a template region directly responsible for encoding the telomere sequence, from which the enzyme repeatedly copies (via reverse transcription) a sequence motif that forms the tandem DNA repeat arrays of telomeres. Telomerase is not completely universal¹⁵, since, for example, in *Drosophila*, telomerase had been lost and telomeres composed of short tandem repeats were replaced with terminal retrotransposons that use targeted retrotransposition, thus engaging in a mechanism similar to reverse transcription^{16,17}. Nevertheless, telomerase remains the most common mechanism to maintain telomeres across eukaryotes, and without any known exceptions among plants^{14,18–20}.

The distribution of nucleotides is uneven between and also along the telomere DNA strands as it is obvious from telomere motif consensus in many plants, vertebrates, insects, some fungi and many others—(T_xA_yG_z)_n. The

¹Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno 61265, Czech Republic. ²Department of Experimental Biology, Faculty of Science, Masaryk University, Brno 62500, Czech Republic. ³Mendel Centre for Plant Genomics and Proteomics, CEITEC, Masaryk University, Brno 62500, Czech Republic. ⁴Institut Botànic de Barcelona (IBB-CSIC, Ajuntament de Barcelona), Passeig del Migdia s/n, 08038 Barcelona, Catalonia, Spain. ✉email: vpeska@ibp.cz; soniagarcia@ibb.csic.es

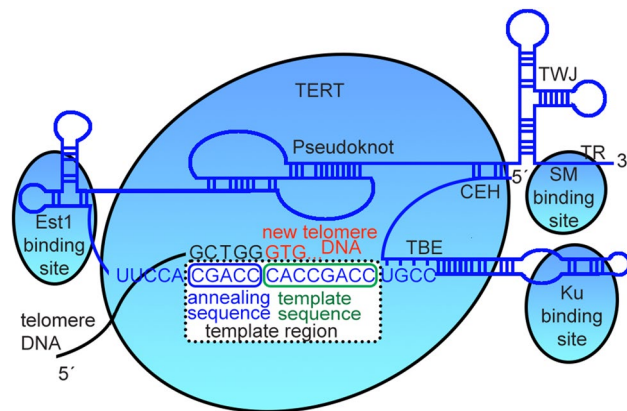


Figure 1. Telomerase RNA organization for *Lachancea* sp. The most relevant binding sites for protein-RNA interactions are indicated by blue ellipses (Est1, SM and KU), and the bigger one is the contact region with the reverse transcriptase (TERT). TBE template boundary element, TWJ three-way junction, CEH core enclosed helix. Adapted from Waldl et al.³⁶ The template region is located between the EST1 binding site and Ku binding hairpins. (Figure created using Adobe Photoshop 22.2.0).

yeast telomeric repeat sequence (TG₁₋₃ where more than 80% is TGTGGG) was first described in *Saccharomyces cerevisiae* in two back-to-back papers in Nature in 1984^{21,22}. Later works showed unprecedented variability in the telomere sequences in yeasts, and made the eukaryotic consensus obsolete. The step-by-step telomere sequence evolution in yeasts was currently proposed²³. However, any obvious feature of the telomere sequence common for all yeasts is difficult to determine, except the clusters of at least two adjacent guanines in the telomere motifs, although the significance of it has not been yet clarified. Several hypotheses explain the role of this clustering e.g., in binding of conserved telomere proteins or forming specific secondary structures, such as G-quadruplexes (G4), which were assumed to occur in telomere regions of several species, including human and *S. cerevisiae*^{24,25}. The asymmetric distribution of guanines and cytosines between the complementary strands arise from the fact that the template region of TR is C-rich and, consequently, the telomere DNA strand produced by the telomerase is G-rich. The template region is longer than the telomere repeat unit, which allows a complete repeat to be translocated to an annealing part of the template region for addition of the next repeat, according to the template sequence^{19,26} (Fig. 1). It has been shown in various plants that the same telomere sequence can originate from template regions with different lengths and annealing sequences. Moreover, in *Arabidopsis thaliana*, the annealing sequence can be as short as two nucleotides^{27,28}.

The length of the whole telomerase RNA is highly variable across organisms, ranging from 150 to more than 2000 nucleotides^{29,30}, with the human sequence of 451 nt³¹ or baker's yeast of 1158 nt^{32,33} as examples. The reasons for this diversity are unknown, although it may be related to structural rather than encoding functions of TRs. In the case of yeasts, several structural features were identified, such as the Ku binding hairpin, the Est1 binding site, the pseudoknot, the three-way junction and the SM binding site³⁴ (Fig. 1). The template region is, in some cases, located between the Ku and the Est1 binding sites. The longest possible template region, which includes short direct repeats at both ends, can be easily identified if the telomere sequence and telomerase RNA for a species are available³⁵. For example, in *Saccharomyces cerevisiae*, the real template is formed by 5'-CACCACACCACACA³², while the predicted template region obtained in our approach (i.e., corresponding to the tandem repeats found by Tandem Repeats Finder) is narrower (in lowercase). Nevertheless, predicting promising TRs by using BLAST searches, as recently applied in land plants²⁷, is a challenging task even among relatively closely related taxa. In the case of considerable evolutionary divergence among species, such as in yeasts, the BLAST strategy is not feasible to start with, given the low sequence conservation among TRs³⁶. Although TRs are variable in the primary structure and are known only in a limited number of organisms, innovative bioinformatics tools using phylogenetics and covariance models^{33,37} and conserved secondary structures in TRs³⁸⁻⁴⁰ have led to discoveries even in more distant eukaryotic kingdoms such as protostomes⁴¹.

In yeasts, TR sequences are particularly diverse in length and complexity, and only a few are available. A recent survey aimed to provide yeast TR sequences, identifying 46 cases, from which 27 telomerase RNAs were found for the first time in the family Saccharomycetaceae. Several strategies were used, including the comparison of common sequence features of known TRs in this family, synteny-based homology searches, the use of known telomere template sequences, or employing a set of known TR regions for covariance model-based searches and BLAST queries (including all template regions available from Saccharomycetaceae at that time)³⁶. However, these authors failed to find template regions in 13 of the TRs that they attempted to analyse, mostly due to the lack of knowledge of specific telomere sequences. Here we complement and expand these results by predicting telomerase RNAs in several closely related species, describing their telomere motifs and predicted template regions. Interestingly, we predict one of the longest known telomere units across eukaryotes (28 bp) in the genus *Tetrapisispora*, as well as its corresponding TR, including the template region. We also compare propensities of the identified telomere candidates to form a guanine quadruplex (G4; G-quadruplex), and finally, we discuss the significance of motif length in relation to telomerase processivity.

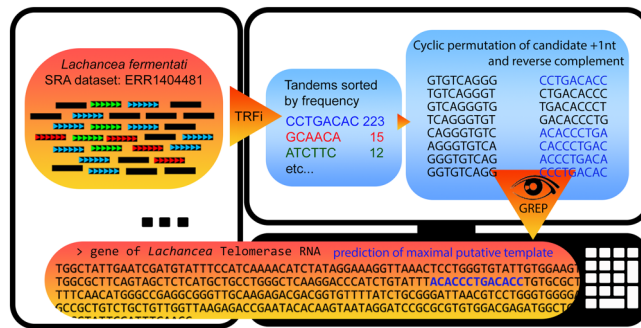


Figure 2. Example of the pipeline for finding telomere motifs and predicting template regions in *Lachancea fermentati*. NGS data in short reads were processed by Tandem Repeats Finder (TRFi) to obtain a list of tandem motifs sorted according to their frequency in the dataset (indicated by different colours). Permuted sequences and their reverse complements were checked using a full text search based on the grep command in a Linux terminal for their presence between the Ku binding hairpin and the Est1 binding site. (Figure created using Adobe Photoshop 22.2.0).

Results

Validation of the approach.

We analysed 38 genomic datasets from Saccharomycetaceae. These datasets represented species from the genera *Candida*, *Kluyveromyces*, *Lachancea*, *Saccharomyces*, *Tetrapisispora*, *Torulasporea* and *Zygosaccharomyces*. We successfully validated our approach on the raw data from SRA database (NCBI) in 19 species from the first four genera (Supplementary Table S1). Therefore, we omitted trimming and prefiltering as the downloaded data exhibited sufficient quality for the prediction of telomere sequence as one of the most abundant tandems. Thus, we neglected a little probability that low-quality reads can produce specific telomere candidates with regular pattern corresponding to the telomerase RNA counterparts. The in silico approach we used in this work is summarized in Fig. 2. Briefly, the telomere candidates were usually among the most prevalent tandems from Tandem Repeats Finder analysis (TRFi) and showed conserved clusters of two or more adjacent guanines (or cytosines in reverse complement in genomic data). TRFi is a reference free alignment-based algorithm which can be set by several parameters, including score for matches, mismatches, and gaps⁴², more details about our setup are described in Methods. We use (for simplicity) the term “maximum putative template” for the longest possible region from TR that corresponds to any (mostly only one) tandem repeat consensus obtained in TRFi. It does not mean that we delimit the maximum for a real template. Rather, the maximum putative template always forms at least a substantial core part of the experimentally verified template region. Our predictions of basic telomere repeats and maximum putative template regions were not identical, but consistent with the repeats and template regions identified previously in biochemical experiments^{29,32–36,43–47}. Thus, for most of them, we put the new coordinates of the maximum putative template region according to the consensus telomere from TRFi results in the Supplementary Table S2. In all cases, the telomere sequence candidate from TRFi shared 100% homology with the maximum putative template regions.

Complete predictions.

We have predicted new TRs, their template regions, and telomere sequences in the species *Lachancea cidri*, three species of *Tetrapisispora* (*T. fletii*, *T. iriomotensis*, and *T. namnaoensis*), three species of *Torulasporea* (*T. franciscae*, *T. maleeae* and *T. pretoriensis*), and *Zygosaccharomyces sapae*. These results were based on a combination of the outcome of synteny-based homology search for telomerase RNA, and the TRFi analysis.

Additional predictions.

We identified telomere candidates and template regions in four species and one strain of *Lachancea* (*L. fermentati*, *L. meyersii*, *L. mirantina*, *L. nothofagi* and *L. sp.* CBS 6924), two species of *Torulasporea* (*T. delbrueckii* and *T. microellipsoides*) and two species of *Zygosaccharomyces* (*Z. bailii* and *Z. rouxii*) (Supplementary Table S1).

Modified TRFi in case of *K. lactis*, *L. kluyveri*, and *T. blattae*.

A modified setting of Tandem Repeats Finder parameters with final reduction in number from five to two units of detected motifs of at least 20 nt in tandem uncovered the proper telomere sequence in *L. kluyveri* but not in *K. lactis*. For the latter, we were not able to detect the telomere sequences in the available datasets despite the fact that the motif (25 nt), TR, and template region had been reported previously^{34,36,43}. In comparison, the reads of *L. kluyveri* (100 nt) were not long enough to detect at least five adjacent tandem repeats, which illustrated how this parameter could act as a bottleneck of the method. If we set this parameter to a lower stringency e.g., to two adjacent repeats, then we should be able to detect the telomere sequence such as in *L. kluyveri* and *T. blattae* (Supplementary Table S2). Nevertheless, setting the default parameter to five makes the method selective for telomere repeats that are usually of very low complexity as compared to e.g., short interstitial telomere sequences type II and III⁴⁸ and teloboxes present in promoters⁴⁹, where less than five regular motifs are expected. On the other hand, the dataset analysed from

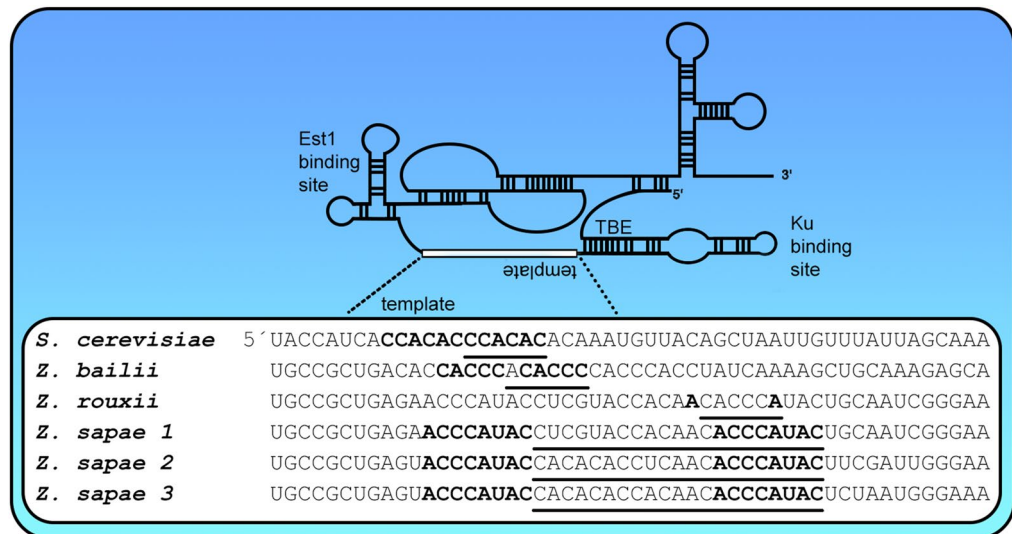


Figure 3. Schematic picture of telomerase RNA and template regions in *Saccharomyces cerevisiae*, *Zygosaccharomyces bailii*, *Z. rouxii*, and *Z. sapae*. We identified the same putative consensus telomere sequence as 5'-ACACCC (or its permutations) in *Saccharomyces cerevisiae*, *Zygosaccharomyces bailii* and *Z. rouxii*, although the template regions, annealing and telomere synthesis during reverse transcription differed in all three species. The corresponding template regions are indicated in bold in the sequences (direct repeats), and underlined fonts (template sequence), respectively. (1) In *S. cerevisiae* the maximum template region consists of two adjacent CCACAC repeats. (2) In *Z. bailii* the template region is formed by a shorter sequence than in *S. cerevisiae* and enables the annealing of five nucleotides and the addition of six nucleotides complementary to ACACCC. (3) The template region in *Z. rouxii* is a bit paradoxical. The synthesis is probably running on the template sequence CACCCA (a permuted version of ACACCC) and annealing occurs through a single base pair (the first A in the template). However, the template region from *Z. rouxii* corresponds exactly to a much longer telomere sequence (21 nt) from the species *Z. sapae*, with a potentially longer annealing (up to 8 bp). Three TR paralogs were found in *Z. sapae*. (Figure created using Adobe Photoshop 22.2.0).

K. lactis was produced on the HiSeq X Ten platform, which systemically under-represents (and in some cases completely misses) telomere sequences, as reported previously⁵⁰.

All predictions and validations are summarized in Supplementary Table S1. The details and results from TRFi are shown in Supplementary Table S2. Several additional key results are: (1) long telomere motifs were found in *Kluyveromyces* sp. (25 nt) and *L. kluyveri* (26 nt); (2) *L. cidri* and *L. fermentati* exhibited the telomere motif, (CTGACACC)_n. Their candidate TRs shared 56% identity and the maximum putative template was identical for both of them. (3) *L. mirantina*, *L. waltii*, and *L. thermotolerans* shared the telomere motif (CCAACACC)_n, but their template regions differed. Similarly, (4) some *Lachancea* sp. analysed shared the (CACCCAGC)_n but their template regions differed. (5) We detected short (6 nt—the same as in *S. cerevisiae*) and long (21 nt) telomere motifs in *Zygosaccharomyces*. Remarkably, the annealing part of the template region in *Z. rouxii* was formed by a single nucleotide. *Z. sapae* possessed three TR paralogs with distinct template regions (Fig. 3, Supplementary Table S1 and Supplementary Table S2). (6) The telomere motif in *T. microellipsoides* was 8 nt long, and the rest of the genus showed motifs of 13 nt. (7) In the *Tetrapisispora* sp. analysed two telomere motifs per species were detected (Supplementary Table S1).

Prediction of G4 propensity in telomere sequences. We analysed the set of 20 consensus sequences identified by Tandem Repeats Finder (Supplementary Table S3) for the presence of G4 forming sequences (GFS). The predicted telomere sequences varied in the length of their repeat units (6 to 28 nt) and in GC content (34.5–69.1%). The QGRS mapper identified 19 out of the 20 sequences as potential GFS with an average score of 27 (from 8 to 63). The G4Hunter score was found to be above the threshold of 1.2 in only six sequences. The highest G4Hunter score was 2.2 for the telomere sequences from *L. mirantina*, *L. thermotolerans* and *L. waltii* (TGT TGGGG), followed by the sequence repeat from *C. glabrata* (GTCTGGGTGCTGTGGGGTC) with a G4Hunter score of 1.84. Species from the *Kluyveromyces* genus with scores under 0.6 represented the other extreme of the scale. The simulated sequence from the *Saccharomyces* species tested, *Z. bailii*, and *Z. rouxii*, (TGTGGG), was of G4Hunter score 1.72. These G4Hunter scores were comparable to those of real telomeres from *S. cerevisiae*, which varied from 1.56 to 1.68. For more details and a comparison between the QGRS versus G4Hunter results, see Supplementary Table S3.

In vitro validation of G4 formation. The predicted propensity of selected telomeric sequences to form G4 was experimentally tested by a combination of two methods, circular dichroism (CD) spectroscopy, and the Thioflavin T (ThT) fluorescence assay (Supplementary Table S4) to assess G4 formation. Taken together these

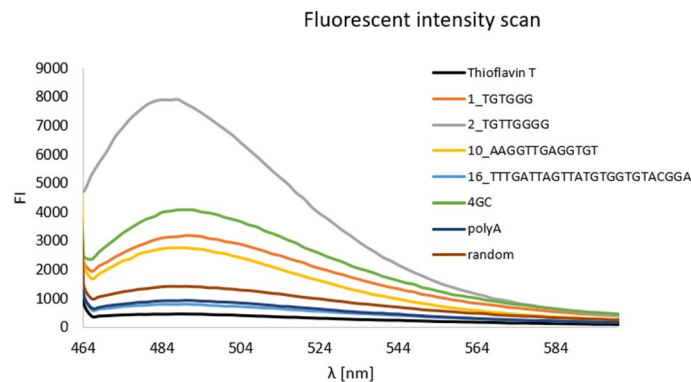


Figure 4. Thioflavin T (ThT) fluorescence assay. We tested four oligonucleotides representing selected telomere sequences from *Sacharomyces* sp., *Zygosaccharomyces* sp., *Lachancea* sp., *Torulaspota* sp., and *Kluyveromyces marxianus*; three controls—one G4-forming sequence (called 4GC) and two negative controls (polyA and random). The G4 propensity was confirmed in vitro by both methods in three out of four telomere sequences tested as well as in positive control (in this figure, Supplementary table S4). The only sample sequence which did not form G4 in vitro was the one that was negative by prediction also. (Figure created using Adobe Photoshop 22.2.0).

methods provide more reliable answer than any of them separately⁵¹. We tested four oligonucleotides representing selected telomere sequences from *Sacharomyces* sp., *Zygosaccharomyces* sp., *Lachancea* sp., *Torulaspota* sp., and *Kluyveromyces marxianus*; three controls—one G4-forming sequence (called 4GC) and two negative controls (polyA and random). The G4 formation was confirmed by both methods in three out of four telomere sequences tested as well as in the positive control (Fig. 4, Supplementary table S4). The only sample sequence which did not form G4 in vitro was the one that was also predicted to be negative. As evidence of G4 formation in vitro, we considered the shift of the CD spectroscopy peak to the maximum typical for the particular type of G4 (parallel G4 = maximum \approx 264 nm, minimum \approx 245 nm, antiparallel G4 = maximum \approx 295 nm, minimum \approx 260 nm, hybrid = maximum \approx 295 and 260 nm, minimum \approx 245 nm) and a stronger signal when G4 was stabilized in the presence of K⁺ ions. ThT is a light-up probe with an increased fluorescence intensity specific for G4 detection in vitro (Fig. 4, Supplementary Table S4).

Discussion

The approach used here combines a prediction of telomere sequence in silico and its support by congruence with the maximum putative template region in independently identified TR candidates. Our findings are consistent with those previously reported in 18 cases, including the already published data from species of the genera *Candida*, *Kluyveromyces*, *Lachancea* and *Saccharomyces*^{33–36,43,45,47,52}.

The consensus telomere sequence detected in all *Saccharomyces* analysed was predominantly (CCACAC)_n, already described as the basic telomere motif in *S. cerevisiae*^{21,43,53}. All species tested showed congruence between our telomere sequence prediction from genomic data, and prediction of the template region from previous reports^{34,36,43}. We also detected some other motifs that could represent telomere sequence variability (telomerase slippage, abortive synthesis, and sequence degeneration) but they were not considered representative (Supplementary Table S2). *Candida glabrata*'s relatively long telomere motif (16 nt) and its corresponding template region was also detected, although the relative length of the template region with respect to the telomere motif was rather short (1.19)²⁹. Other telomere-like sequences were found for this species at an even higher frequency (Supplementary Table S2), although they did not correspond to either the published template regions or our maximum putative template region.

Considerable variability in telomere units was described at the level of genome assembly in numerous species of Saccharomycetaceae, e.g., *T. iriomotensis*, *S. cerevisiae*, and *Kazachstania exigua*²³. We show here that even telomeres with complex sequences have a clear consensus with its maximum putative template counterpart. We have predicted telomere sequences for 18 species, with up to three possible variants in some of them (e.g., *Z. sapae*). In 10 of these, an approach based on the analysis of predicted TRs, without knowing the telomere sequence, had failed to recover the template region³⁶. Here, we report, for the first time, the telomere sequence and the maximum putative template regions in *Torulaspota*, using the previously predicted TRs from *T. delbrueckii* and *T. microellipsoides* and our own TR predictions for other species of *Torulaspota*. Our results from Saccharomycetaceae also support some of the putative telomere sequences in comprehensive telomere searches across Ascomycota²³. *Torulaspota delbrueckii* and *T. microellipsoides* are examples of the intrageneric variability of the telomere sequence. Several telomere-like motifs were also detected in *T. delbrueckii* (Supplementary Table S2), although they did not match the template region and their frequency was much lower. They may represent, e.g., variability in the telomere sequence or remnants of ancestral motifs. The analysis of genomic data from three additional *Torulaspota* species (*T. franciscae*, *T. maleeae* and *T. pretoriensis*) showed that they presented the same telomere sequence candidate as *T. delbrueckii*, however the template region of *T. maleeae* was shorter by

6 nucleotides and annealing was only possible with 2 nucleotides. This illustrates that even the same telomere motif can be encoded by a dramatically different template region (see Supplementary Table S1).

As with *Torulopsis*, Waldl et al.³⁶ predicted TRs in *Zygosaccharomyces bailii* and *Z. rouxii*, but they did not show template regions in any of these sequences. We reanalysed these species datasets, and one more from the same genus (*Z. sapae*), with heterogeneous results. In the case of *Z. bailii* and *Z. rouxii*, we predicted CCACAC as a telomere sequence candidate and the corresponding template regions in TRs. We obtained an unexpected outcome however, for *Z. sapae*. First of all, we found three homologs of TR, each with a unique template region in this species and its counterpart in the TRFi output. Second, the three telomere motifs were relatively long (21 nt) and their frequency in the dataset differed by orders of magnitude (from 7.2×10^{-6} to 3.0×10^{-5}). Third, one of the *Z. sapae* template regions overlapped with the template in *Z. rouxii*, which, paradoxically, had a much shorter prevalent telomere motif, 6 instead of 21 nt. The paradox of a short telomere motif (CACCCA)_n in *Z. rouxii* and a long telomere candidate from *Z. sapae* (CCCATACCTCGTACCACAACA)_n may have a simple explanation. It was shown that the base pairing between the template RNA and the telomere DNA is limited not only by the structure of the telomerase RNA but also by the protein subunit TERT⁵⁴. Thus, we speculate that the two species may differ in their prevalent telomere sequences despite having the same template and surrounding regions because the change is caused by the preferred annealing mode. This explanation was also recently proposed for the transition between plant- and human-type telomere sequences in marine plants from the *Zostera* genus²⁸. In summary, *Z. sapae* is an extreme example of intraspecific telomere sequence variability ensured by multiple TR paralogs across all eukaryotes.

We further analysed the *Lachancea* species, which also illustrate telomere sequence variability at the genus level, both in nucleotide composition and in sequence length (from 8 to 26 nt). We described new telomere motifs for 6 taxa (Supplementary Table S1) and new TR, including the template region, in *L. cidri*. We found possible signs of telomerase errors in *L. thermotolerans* (and in *L. cidri* to a lesser extent), in which several similar motifs were proposed (Supplementary Table S2).

In the genus *Tetrapispora*, until now, the TR was known only in *T. blattae* (with an unspecified template region) and telomere sequences were found only very recently in three other species of the genus²³; we also detected the three published telomere candidates, with 25/28, and 25 nt long motifs, respectively, *T. fleetii*, *T. iriomotensis*, and *T. namnaoensis*. Cervenak et al.²³ also showed that *T. iriomotensis* had telomere motifs 28/29 nucleotides long at the ends of genomic scaffolds. This would make these species record holders in the motif length among Saccharomycetaceae. Except for *T. blattae*, all *Tetrapispora* analyzed TRs contain direct repeats of template regions with dimeric sub-repeats (ATC)₂, which probably allow a dual annealing register and the occurrence of the two motifs, 28/25 nt long. *T. fleetii* and *T. namnaoensis* share identical template regions and telomere motifs, while *T. iriomotensis* differs from the previous two by three nucleotide polymorphisms in the telomere sequence (Supplementary Table S2). *T. blattae* has a shorter telomere sequence (25/21 nt) and a TR that does not contain the sub-repeats. The enormous telomere repeat unit size, in fact one of the longest units reported in Saccharomycetaceae, led us to hypothesize that yeast may compensate for a possible low telomerase processivity by expanding the telomere motif to keep the telomere length stable. The hypothetical ancestral telomere sequence TTAGGG^{23,55} is here substituted by motifs of multiple sizes, e.g., by a fourfold increase in *Tetrapispora*. On the other hand, the genus *Lachancea* exhibits the opposite trend, a shortening of the telomere motif despite the sequence variability in descendant taxa (Fig. 5).

Telomere sequences are relatively well conserved across large groups of higher eukaryotes, where a single telomere type (e.g., TTTAGGG among plants) is the most common and others appear as outliers (e.g. CTCGGTTATGGG in *Allium* sp.)^{19,28,56,57}. The apparent extraordinary molecular diversity in yeasts can be related to their short generation times, diversity and early evolutionary divergence: e.g., *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, both Ascomycota, are similarly diverse from each other as they are from animals, their ancestors having separated about 420–330 Mya⁵⁸.

Despite the divergence, telomere sequences keep track of guanines (Gs) as one of the most conserved features in all species tested. Certain G-rich sequences form G4 structures involved in gene regulation, telomere protection, and other biological processes^{59,60}. It raises questions about the propensity to form G4 as a common telomere sequence feature. G4 formation in vitro was previously confirmed in the telomere repeats *S. pombe* and *S. cerevisiae*^{61,62}. Biological relevance of G4 in *S. cerevisiae* was also studied in vivo. The telomere G4 stabilization restored compromised capping and was proposed as a rudimentary and possible back-up mechanism in the telomere protection⁶³. Thus, we decided to predict if all sequences tested are prone to form G4 and verify this prediction by in vitro measurements. We selected four sequences for the in vitro assays from both predicted extremes and an average range. G4Hunter, QGRS mapper, as well as experimental measurements in vitro (Fig. 4, Supplementary table S4), demonstrated a notable heterogeneity of motifs tested in their propensity to form G4 secondary structures that range from very weak prediction (no G4 in vitro) to very strong ones. The strong prediction was obtained in only five sequences from the list. The best predictions were in *L. mirantina*, *L. thermotolerans*, and *L. waltii*. While the lowest scores were in *Kluyveromyces* sp., the G4 prediction in the real telomere DNA from *S. cerevisiae* is comparable to the scores in hypothetical sequences from only the basic motif (CCACAC)_n or to the scores of human telomeres, which is consistent with the previous finding of an accumulation of G4-forming sequences in telomeres of *S. cerevisiae*²⁴. Our results imply that the ability to form G4 in vivo may be different in the telomere sequences tested despite the presence of tracks of Gs in all of them.

Another interesting finding is that several telomere-like sequences supported by TR candidates can be found in a single species. Heterogeneity in telomere repeats within a species had been already detected in yeast species^{26,64} and other organisms^{30,65,66}. Besides the existence of three error mechanisms proposed to explain intraspecific telomere sequence heterogeneity (telomerase stalling, stuttering and misincorporation)⁶⁷, telomerase RNA paralogs with different template regions were published²⁸. These results raise questions about possible pseudogenes of TR candidates, which can complicate the prediction conclusions. If needed, in further experimental

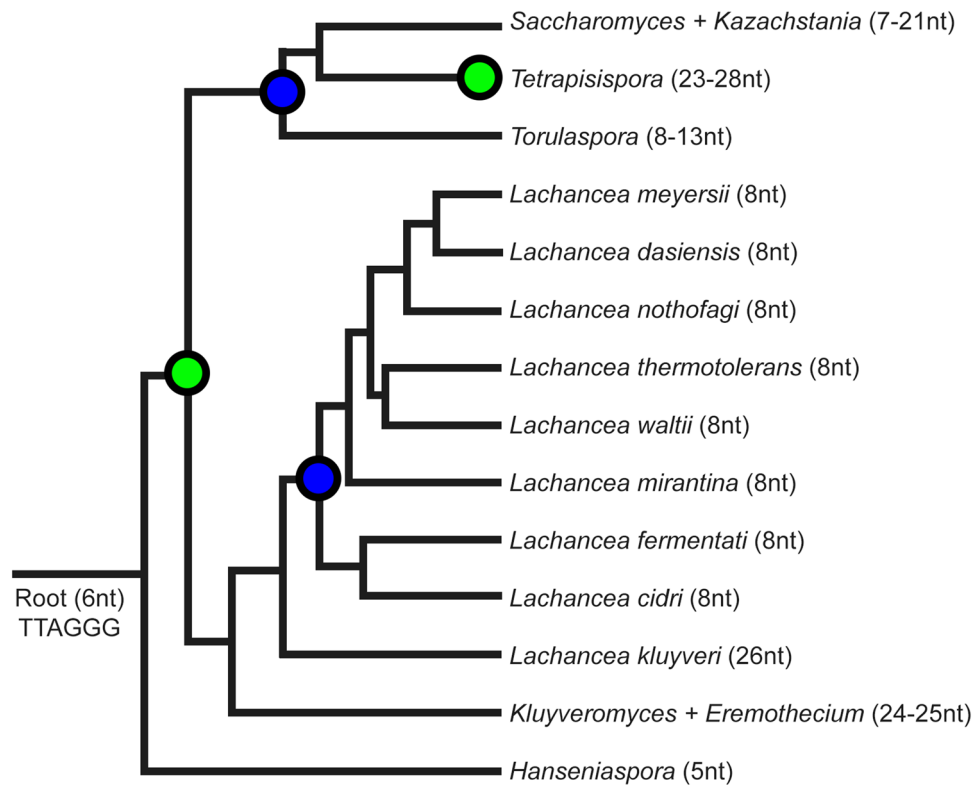


Figure 5. Schematic evolution of telomere motif length in Saccharomycetaceae. The telomere motif length undergoes a series of stretchings (green nodes) and shrinkages (blue nodes). The root represents a common ancestor with Ascomycota and other eukaryotes. Recent publication hypothesizes TTAGGG as an ancestral sequence for all fungi and all eukaryotes, respectively⁵⁵. The scheme of the tree is based on a recent genus-level phylogeny in Saccharomycotina⁷³. The motif lengths are based either on results in this work or previous articles^{23,74}. The branch lengths in this scheme do not represent evolutionary distances. (Figure created using Adobe Photoshop 22.2.0).

progress for selected species, the predictions should be verified by gene deletion or template region modification to conclude for sure that the telomerase RNA is identified.

Materials and methods

Genomic data. Datasets were downloaded from public sequence read archives (SRA, NCBI, <https://www.ncbi.nlm.nih.gov/sra/>). Illumina HiSeq/MiSeq/Illumina Genome Analyzer data from Whole Genome Shotgun (WGS) strategies were preferred for their uniform read length and genome representativeness. Optimal datasets were not available in the case of *Saccharomyces boulardii* (454 GS FLX Titanium), *Kluyveromyces lactis* and *Saccharomyces pastorianus* (HiSeq X Ten), although in general we have avoided HiSeq X Ten as much as possible because this platform systemically under-represents telomere sequences⁵⁰. The read length was correlated with the chances of detecting long tandem repeats, however, for nine species, only 40–50 nt long reads were available. Read length, number of reads used per species and other detailed information can be found in Supplementary Table S1 and S2. TR sequences were downloaded from GenBank (NCBI) according to previously published data^{29,33,34,36,44–47}. Template regions were considered, according to coordinates in Waldl et al.³⁶. For a complete list of the species analysed and accession numbers, see Supplementary Table S1.

Tandem repeats finder. The analysis of short tandem repeats was performed using Tandem Repeats Finder (TRFi)⁴² with custom made scripts as described previously⁶⁸, with some modifications, which we call the default setup in this work: the detected motif length was 5–50 nt long and the minimum number of such repeats in tandem was set to five units. This option of five units makes the method stricter than previously used and makes it more selective for reads with lower complexity, less sensitive to sequencing errors at single positions because the consensus is calculated from more repeat units. We omitted prefiltering as the downloaded data exhibited sufficient quality for TRFi analysis. TRFi parameters for the alignment and matching/indel probabilities were set as recommended by default⁴². The vast majority of telomere candidates were actually consensus sequences built from hundreds and thousands of repeat units each. TRFi was repeated with the minimum number of repeats set to two units in three cases (*Kluyveromyces lactis*, *Lachancea kluyveri*, and *Tetrapisispora blattae*). Empirically, the telomere candidates emerged among the first fifty most abundant tandems and only exceptionally was the candidate identified as a much less abundant sequence, or was not identified at all⁶⁸ (Supplementary Table S2).

The motifs in the TRFi results are consensual patterns from the detected tandem arrays, possibly containing a certain level of variability and/or degeneracy.

Prediction of telomere motifs, TRs, and template regions. Novel TRs were predicted in synteny-based homology searches according to a previous study³⁶ (Supplementary Table S1). We used tblastx search with a known TR locus + surrounding regions (usually ± 10 Kb) from related organisms as a query to identify a collinear genomic region in the species of interest. Subsequently, a sequence corresponding to the telomerase RNA gene was searched in dot plot analysis and alignment. This workflow provided a very limited number of TR candidates per species, usually only a single possible candidate. Independently, telomere motifs were predicted from TRFi results with respect to the frequency of detected repeats and typical features like unequal distribution of G/C in the complementary strands, presence of two or more adjacent Gs etc. The typical TRFi output was only a short list (in the order of units) of several abundant repeats followed by many others with significantly lower abundance. The telomere candidate was usually one of the abundant repeats, however not exclusively. Then the revision of corresponding telomerase RNA was done as follows: a set of permuted tandems and their reverse complements (always prolonged by one, two etc., up to ten nucleotides from the following unit, labelled as +1, +2, up to +10 in Supplementary Figure S1–S4) were checked using the plain text search (grep—command-line utility in Unix) for their presence either in the whole gene or (if available) between the Ku binding hairpin and Est1 binding site in the telomerase RNA, the positions of which were estimated by local alignment of conserved regions with annotated data in previous work³⁶. When the permutation (+1 etc.) matched, it was considered as a telomere sequence candidate and the telomerase RNA counterpart as either a result of approach validation or a newly predicted template region in the TR. The final selection of candidates was done manually considering the frequency of the motif and the length of the annealing part. The candidates with very low G/C content (e.g., ATATA in *C. glabrata*) were excluded even if there was a match with overlap +1 in TR (see Supplementary Table S2). All motifs described above, including selected candidate templates are depicted in Supplementary Figure S1–S4 and marked in Supplementary Table S2. The maximum sequence in the putative TR matching the telomere candidate was then taken as a maximum putative template region. The length estimation of the relative template region was calculated as the ratio between the length of the maximum putative template region and the length of the repeat unit in the predicted telomere sequence (Supplementary Table S1). A pipeline of the procedure used to obtain new telomere motifs and template regions is shown in Fig. 2. Overall, we either: (1) detected the published telomere sequences and template regions when there was congruence between them and our TRFi results, (2) predicted a telomere sequence and a maximum putative template region when these were not available for a given species, but TRFi detected a promising candidate consistent with the region in TR (between Ku and Est1 binding sites if information about them was available), or (3) predicted the telomere sequence, TR, and maximum putative template region for a given species. Only in one case (*K. lactis*) could we not draw any conclusion because of inconsistency between the published TR and the outcome of TRFi results (Supplementary Table S1).

Prediction of G4 forming sequences in telomere DNAs. Candidate telomere sequences (Supplementary Table S3) were analysed for G4 forming sequences (GFS) using G4Hunter Web Tool⁶⁹ and QGRS mapper⁷⁰. Both tools are used for G4 prediction and count on G-richness and G-skewness of a DNA or RNA sequence. They provided quadruplex propensity scores, full sequences of window peak scores in the case of G4Hunter and peak-sequence scores in QGRS. We simulated telomeres by simple repeats of a candidate unit. In the case of *S. cerevisiae*, we compared GFS in simulated and real sequences. The window size for G4Hunter was set to 25. The analysis by QGRS mapper was performed with default parameters (max length of the GFS = 30, min number of guanidines in G-track = 2, max loop size = 36 nt). In general, the higher the score, the higher the probability of the sequence forming a G4 is. The results of the analyses are presented in (Supplementary Table S3).

Circular Dichroism (CD) spectroscopy. Synthetic oligonucleotides were purchased from Integrated DNA Technologies (Coralville, Iowa, United States) and diluted in water to a concentration of 100 μ M. The oligonucleotides were then heated at 95 °C for 5 min in either 1 mM sodium phosphate buffer (pH 7) and 0.3 mM EDTA or 1 mM sodium phosphate (pH 7), 10 mM potassium phosphate (pH 7) and 90 mM KCl and slowly cooled to room temperature. CD measurements were carried out in a Jasco 815 (Jasco International Co., Ltd., Tokyo, Japan) dichrograph in 1 cm path-length microcells at 23 °C. A set of four scans with a data pitch of 0.5 nm and 200 nm min⁻¹ scan speed was averaged for each sample. CD signals were expressed as a difference in the molar absorptions, $\Delta\epsilon$ of the left- and right-handed circularly polarized light⁷¹.

Thioflavin T fluorescence assay. Oligonucleotides were further diluted to a 2 μ M concentration in 100 mM Tris-HCl (pH 7.5) and 100 mM KCl buffer, heated at 95 °C for 5 min and slowly cooled to room temperature. Thioflavin T (ThT) was diluted in water with addition of KCl to 1 μ M final concentration. Experiments were performed in 384-well microplates from CORNING (Flat Bottom Black Polyester). Each condition was tested in triplicate at room temperature. Oligonucleotides and ThT were mixed at 1:0.5 molar ratio to a final volume of 20 μ l. Fluorescence emission was collected at 460–700 nm, every 2 nm after excitation at 425 nm in a microplate reader (Spark, Tecan)⁷².

Received: 15 January 2021; Accepted: 3 June 2021

Published online: 17 June 2021

References

- Lim, C. J. & Cech, T. R. Shaping human telomeres: From shelterin and CST complexes to telomeric chromatin organization. *Nat. Rev. Mol. Cell Biol.* <https://doi.org/10.1038/s41580-021-00328-y> (2021).
- Lu, W., Zhang, Y., Liu, D., Songyang, Z. & Wan, M. Telomeres-structure, function, and regulation. *Exp. Cell Res.* **319**, 133–141. <https://doi.org/10.1016/j.yexcr.2012.09.005> (2013).
- Shay, J. W. Telomeres and aging. *Curr. Opin. Cell Biol.* **52**, 1–7. <https://doi.org/10.1016/j.ccb.2017.12.001> (2018).
- Victorelli, S. & Passos, J. F. Telomeres and cell senescence—Size matters not. *EBioMedicine* **21**, 14–20. <https://doi.org/10.1016/j.ebiom.2017.03.027> (2017).
- Blackburn, E. H. Telomere states and cell fates. *Nature* **408**, 53–56. <https://doi.org/10.1038/35040500> (2000).
- Lingner, J., Cooper, J. P. & Cech, T. R. Telomerase and DNA end replication: No longer a lagging strand problem?. *Science* **269**, 1533–1534. <https://doi.org/10.1126/science.7545310> (1995).
- Soudet, J., Jolivet, P. & Teixeira, M. T. Elucidation of the DNA end-replication problem in *Saccharomyces cerevisiae*. *Mol. Cell* **53**, 954–964. <https://doi.org/10.1016/j.molcel.2014.02.030> (2014).
- Wellinger, R. J. In the end, what's the problem?. *Mol. Cell* **53**, 855–856. <https://doi.org/10.1016/j.molcel.2014.03.008> (2014).
- Greider, C. W. & Blackburn, E. H. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* **43**, 405–413. [https://doi.org/10.1016/0092-8674\(85\)90170-9](https://doi.org/10.1016/0092-8674(85)90170-9) (1985).
- Shay, J. W. Role of telomeres and telomerase in aging and cancer. *Cancer Discov.* **6**, 584–593. <https://doi.org/10.1158/2159-8290.CD-16-0062> (2016).
- Blasco, M. A. Telomerase beyond telomeres. *Nat. Rev. Cancer* **2**, 627–633. <https://doi.org/10.1038/nrc862> (2002).
- Koonin, E. V. The origin of introns and their role in eukaryogenesis: A compromise solution to the introns-early versus introns-late debate?. *Biol. Direct.* **1**, 22. <https://doi.org/10.1186/1745-6150-1-22> (2006).
- Nosek, J., Kosa, P. & Tomaska, L. On the origin of telomeres: A glimpse at the pre-telomerase world. *BioEssays* **28**, 182–190. <https://doi.org/10.1002/bies.20355> (2006).
- Schrumpfova, P. P. & Fajkus, J. Composition and function of telomerase—a polymerase associated with the origin of eukaryotes. *Biomolecules* **10**, 1425. <https://doi.org/10.3390/biom10101425> (2020).
- Chan, S. R. W. L. & Blackburn, E. H. Telomeres and telomerase. *Philos. Trans. R. Soc. B* **359**, 109–121. <https://doi.org/10.1098/rstb.2003.1370> (2004).
- Rubin, G. M. Isolation of a telomeric DNA sequence from *Drosophila melanogaster*. *Cold Spring Harb. Symp. Quant. Biol.* **42**(Pt 2), 1041–1046. <https://doi.org/10.1101/sqb.1978.042.01.104> (1978).
- Traverse, K. L. & Pardue, M. L. A spontaneously opened ring chromosome of *Drosophila melanogaster* has acquired He-T DNA sequences at both new telomeres. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8116–8120. <https://doi.org/10.1073/pnas.85.21.8116> (1988).
- Fajkus, P., Peska, V., Fajkus, J. & Sykorova, E. Origin and fates of TERT gene copies in polyploid plants. *Int. J. Mol. Sci.* **22**, 1783. <https://doi.org/10.3390/ijms22041783> (2021).
- Peska, V. & Garcia, S. Origin, diversity, and evolution of telomere sequences in plants. *Front. Plant Sci.* **11**, 117. <https://doi.org/10.3389/fpls.2020.00117> (2020).
- Schrumpfova, P. P., Fojtova, M. & Fajkus, J. Telomeres in plants and humans: Not so different, not so similar. *Cells-Basel* **8**, 58. <https://doi.org/10.3390/cells8010058> (2019).
- Shampay, J., Szostak, J. W. & Blackburn, E. H. DNA sequences of telomeres maintained in yeast. *Nature* **310**, 154–157. <https://doi.org/10.1038/310154a0> (1984).
- Walmsley, R. W., Chan, C. S., Tye, B. K. & Petes, T. D. Unusual DNA sequences associated with the ends of yeast chromosomes. *Nature* **310**, 157–160. <https://doi.org/10.1038/310157a0> (1984).
- Cervenak, F., Sepsiova, R., Nosek, J. & Tomaska, L. Step-by-step evolution of telomeres: Lessons from yeasts. *Genome Biol. Evol.* <https://doi.org/10.1093/gbe/evaa268> (2020).
- Cutova, M. *et al.* Divergent distributions of inverted repeats and G-quadruplex forming sequences in *Saccharomyces cerevisiae*. *Genomics* **112**, 1897–1901. <https://doi.org/10.1016/j.ygeno.2019.11.002> (2020).
- Moye, A. L. *et al.* Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.* **6**, 7643. <https://doi.org/10.1038/ncomms8643> (2015).
- Ares, M. Jr. & Chakrabarti, K. Stuttering against marginotomy. *Nat. Struct. Mol. Biol.* **15**, 18–19. <https://doi.org/10.1038/nsmbo108-18> (2008).
- Fajkus, P. *et al.* Telomerase RNAs in land plants. *Nucleic Acids Res.* **47**, 9842–9856. <https://doi.org/10.1093/nar/gkz695> (2019).
- Peska, V. *et al.* Human-like telomeres in *Zostera marina* reveal a mode of transition from the plant to the human telomeric sequences. *J. Exp. Bot.* **71**, 5786–5793. <https://doi.org/10.1093/jxb/eraa293> (2020).
- Kachouri-Lafond, R. *et al.* Large telomerase RNA, telomere length heterogeneity and escape from senescence in *Candida glabrata*. *FEBS Lett.* **583**, 3605–3610. <https://doi.org/10.1016/j.febslet.2009.10.034> (2009).
- McCormick-Graham, M. & Romero, D. P. A single telomerase RNA is sufficient for the synthesis of variable telomeric DNA repeats in ciliates of the genus *Paramecium*. *Mol. Cell Biol.* **16**, 1871–1879. <https://doi.org/10.1128/mcb.16.4.1871> (1996).
- Feng, J. *et al.* The RNA component of human telomerase. *Science* **269**, 1236–1241. <https://doi.org/10.1126/science.7544491> (1995).
- Singer, M. S. & Gottschling, D. E. TLC1: Template RNA component of *Saccharomyces cerevisiae* telomerase. *Science* **266**, 404–409. <https://doi.org/10.1126/science.7545955> (1994).
- Dandjinou, A. T. *et al.* A phylogenetically based secondary structure for the yeast telomerase RNA. *Curr. Biol.* **14**, 1148–1158. <https://doi.org/10.1016/j.cub.2004.05.054> (2004).
- Gunisova, S. *et al.* Identification and comparative analysis of telomerase RNAs from *Candida* species reveal conservation of functional elements. *RNA* **15**, 546–559. <https://doi.org/10.1261/rna.1194009> (2009).
- Tzfati, Y., Fulton, T. B., Roy, J. & Blackburn, E. H. Template boundary in a yeast telomerase specified by RNA structure. *Science* **288**, 863–867. <https://doi.org/10.1126/science.288.5467.863> (2000).
- Walld, M. *et al.* TERribly difficult: Searching for telomerase RNAs in Saccharomycetes. *Genes (Basel)* **9**, 372. <https://doi.org/10.3390/genes9080372> (2018).
- Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935. <https://doi.org/10.1093/bioinformatics/btt509> (2013).
- Chen, J. L., Blasco, M. A. & Greider, C. W. Secondary structure of vertebrate telomerase RNA. *Cell* **100**, 503–514. [https://doi.org/10.1016/s0092-8674\(00\)80687-x](https://doi.org/10.1016/s0092-8674(00)80687-x) (2000).
- Qi, X. *et al.* The common ancestral core of vertebrate and fungal telomerase RNAs. *Nucleic Acids Res.* **41**, 450–462. <https://doi.org/10.1093/nar/gks980> (2013).
- Xie, M. *et al.* Structure and function of the smallest vertebrate telomerase RNA from teleost fish. *J. Biol. Chem.* **283**, 2049–2059. <https://doi.org/10.1074/jbc.M708032200> (2008).
- Logeswaran, D., Li, Y., Podlevsky, J. D. & Chen, J. J. Monophyletic origin and divergent evolution of animal telomerase RNA. *Mol. Biol. Evol.* <https://doi.org/10.1093/molbev/msaa203> (2020).

42. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580. <https://doi.org/10.1093/nar/27.2.573> (1999).
43. Cohn, M., McEachern, M. J. & Blackburn, E. H. Telomeric sequence diversity within the genus *Saccharomyces*. *Curr. Genet.* **33**, 83–91. <https://doi.org/10.1007/s002940050312> (1998).
44. Dietrich, F. S. *et al.* The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* **304**, 304–307. <https://doi.org/10.1126/science.1095781> (2004).
45. McEachern, M. J. & Blackburn, E. H. Runaway telomere elongation caused by telomerase RNA gene mutations. *Nature* **376**, 403–409. <https://doi.org/10.1038/376403a0> (1995).
46. Ortiz-Merino, R. A. *et al.* Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch. *PLoS Biol.* **15**, e2002128. <https://doi.org/10.1371/journal.pbio.2002128> (2017).
47. Seto, A. G., Livengood, A. J., Tzfaty, Y., Blackburn, E. H. & Cech, T. R. A bulged stem tethers Est1p to telomerase RNA in budding yeast. *Genes Dev.* **16**, 2800–2812. <https://doi.org/10.1101/gad.1029302> (2002).
48. Uchida, W., Matsunaga, S., Sugiyama, R. & Kawano, S. Interstitial telomere-like repeats in the *Arabidopsis thaliana* genome. *Genes Genet. Syst.* **77**, 63–67. <https://doi.org/10.1266/ggs.77.63> (2002).
49. Tremousaygue, D. *et al.* 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. <https://doi.org/10.1046/j.1365-3113x.2003.01682.x> (2003).
50. Farmery, J. H. R., Smith, M. L., Diseases, N.B.-R. & Lynch, A. G. Telomerecat: A ploidy-agnostic method for estimating telomere length from whole genome sequencing data. *Sci. Rep.* **8**, 1300. <https://doi.org/10.1038/s41598-017-14403-y> (2018).
51. Bedrat, A., Lacroix, L. & Mergny, J. L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44**, 1746–1759. <https://doi.org/10.1093/nar/gkw006> (2016).
52. Bosoy, D., Peng, Y., Mian, I. S. & Lue, N. F. Conserved N-terminal motifs of telomerase reverse transcriptase required for ribonucleoprotein assembly in vivo. *J. Biol. Chem.* **278**, 3882–3890. <https://doi.org/10.1074/jbc.M210645200> (2003).
53. Wang, S. S. & Zakian, V. A. Sequencing of *Saccharomyces telomeres* cloned using T4 DNA polymerase reveals two domains. *Mol. Cell Biol.* **10**, 4415–4419. <https://doi.org/10.1128/mcb.10.8.4415> (1990).
54. Forstemann, K. & Lingner, J. Telomerase limits the extent of base pairing between template RNA and telomeric DNA. *EMBO Rep.* **6**, 361–366. <https://doi.org/10.1038/sj.embor.7400374> (2005).
55. Fulneckova, J. *et al.* A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biol. Evol.* **5**, 468–483. <https://doi.org/10.1093/gbe/evt019> (2013).
56. Fajkus, P. *et al.* Allium telomeres unmasked: The unusual telomeric sequence (CTCGGTTATGGG)_n is synthesized by telomerase. *Plant J.* **85**, 337–347. <https://doi.org/10.1111/tj.13115> (2016).
57. Peska, V. *et al.* Characterisation of an unusual telomere motif (TTTTTTAGGG)_n in the plant *Cestrum elegans* (Solanaceae), a species with a large genome. *Plant J.* **82**, 644–654. <https://doi.org/10.1111/tj.12839> (2015).
58. Teixeira, M. T. & Gilson, E. Telomere maintenance, function and evolution: The yeast paradigm. *Chromosome Res.* **13**, 535–548. <https://doi.org/10.1007/s10577-005-0999-0> (2005).
59. Matsumoto, S. & Sugimoto, N. New insights into the functions of nucleic acids controlled by cellular microenvironments. *Top. Curr. Chem. (Cham)* **379**, 17. <https://doi.org/10.1007/s41061-021-00329-7> (2021).
60. Wallgren, M. *et al.* G-rich telomeric and ribosomal DNA sequences from the fission yeast genome form stable G-quadruplex DNA structures in vitro and are unwound by the Pfh1 DNA helicase. *Nucleic Acids Res.* **44**, 6213–6231. <https://doi.org/10.1093/nar/gkw349> (2016).
61. Jurikova, K. *et al.* Role of folding kinetics of secondary structures in telomeric G-overhangs in the regulation of telomere maintenance in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **295**, 8958–8971. <https://doi.org/10.1074/jbc.RA120.012914> (2020).
62. Tran, P. L., Mergny, J. L. & Alberti, P. Stability of telomeric G-quadruplexes. *Nucleic Acids Res.* **39**, 3282–3294. <https://doi.org/10.1093/nar/gkq1292> (2011).
63. Smith, J. S. *et al.* Rudimentary G-quadruplex-based telomere capping in *Saccharomyces cerevisiae*. *Nat. Struct. Mol. Biol.* **18**, 478–485. <https://doi.org/10.1038/nsmb.2033> (2011).
64. Cohn, M. & Blackburn, E. H. Telomerase in yeast. *Science* **269**, 396–400. <https://doi.org/10.1126/science.7618104> (1995).
65. Tran, T. D. *et al.* Centromere and telomere sequence alterations reflect the rapid genome evolution within the carnivorous plant genus *Genlisea*. *Plant J.* **84**, 1087–1099. <https://doi.org/10.1111/tj.13058> (2015).
66. Sykorova, E. *et al.* Telomere variability in the monocotyledonous plant order Asparagales. *Proc. Biol. Sci.* **270**, 1893–1904. <https://doi.org/10.1098/rspb.2003.2446> (2003).
67. McCormick-Graham, M., Haynes, W. J. & Romero, D. P. Variable telomeric repeat synthesis in *Paramecium tetraurelia* is consistent with misincorporation by telomerase. *EMBO J.* **16**, 3233–3242. <https://doi.org/10.1093/emboj/16.11.3233> (1997).
68. Peska, V., Sitova, Z., Fajkus, P. & Fajkus, J. BAL31-NGS approach for identification of telomeres de novo in large genomes. *Methods* **114**, 16–27. <https://doi.org/10.1016/j.ymeth.2016.08.017> (2017).
69. Brazda, V. *et al.* G4Hunter web application: A web server for G-quadruplex prediction. *Bioinformatics* **35**, 3493–3495. <https://doi.org/10.1093/bioinformatics/btz087> (2019).
70. Kikin, O., D’Antonio, L. & Bagga, P. S. QGRS Mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* **34**, W676–682. <https://doi.org/10.1093/nar/gkl253> (2006).
71. Vorlickova, M. *et al.* Circular dichroism and guanine quadruplexes. *Methods* **57**, 64–75. <https://doi.org/10.1016/j.ymeth.2012.03.011> (2012).
72. Renaud de la Faverie, A., Guedin, A., Bedrat, A., Yatsunyk, L. A. & Mergny, J. L. Thioflavin T as a fluorescence light-up probe for G4 formation. *Nucleic Acids Res.* **42**, e65. <https://doi.org/10.1093/nar/gku111> (2014).
73. Shen, X. X. *et al.* Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* **175**, 1533–1545. <https://doi.org/10.1016/j.cell.2018.10.023> (2018).
74. Cervenak, F. *et al.* Identification of telomerase RNAs in species of the *Yarrowia* clade provides insights into the co-evolution of telomerase, telomeric repeats and telomere-binding proteins. *Sci. Rep.* **9**, 13365. <https://doi.org/10.1038/s41598-019-49628-6> (2019).

Acknowledgements

This work was supported by Projects SYMBIT CZ.02.1.01/0.0/0.0/15_003/0000477, EXPRO (Czech Science Foundation 20-01331X), and the Spanish Government [CGL2016-75694-P (AEI/FEDER, UE)]. S.G. is a holder of a Ramón y Cajal contract (RYC-2014-16608). Computational resources were supplied by the project “e-Infraestructura CZ” (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

Author contributions

V.P.—design of the work, telomere and telomerase sequence analysis, interpretation of data, draft of the work, P.F.—telomerase sequence analysis, interpretation of data, V.B. and N.B.—G4 analysis, M.B. and V.D.—telomere

sequence analysis, J.F.—interpretation of data, substantively revised the manuscript, S.G.—draft of the work, interpretation of data, substantively revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-92126-x>.

Correspondence and requests for materials should be addressed to V.P. or S.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021