

Noncanonical chromosomal-end-specific telomeric repeat arrays in naturally telomerase-negative yeasts

Broňa Brejová^{1,*}, Viktória Hodorová^{2,†}, Hana Lichancová², Askar Gafurov¹, Dominik Bujna¹, Filip Brázdovič^{2,3}, Filip Červenák⁴, Tomáš Petřík⁴, Eva Hegedúsová^{2,5}, Michaela Forgáčová Jakúbková², Martina Neboháčová², Ľubomír Tomáška⁴, Matthias Sipiczki⁶, Tomáš Vinar⁷, Jozef Nosek^{2,*}

¹Department of Computer Science, Faculty of Mathematics, Physics, and Informatics, Comenius University Bratislava, Mlynská dolina, 842 48 Bratislava, Slovakia

²Department of Biochemistry, Faculty of Natural Sciences, Comenius University Bratislava, Ilkovičova 6, 842 15 Bratislava, Slovakia

³Present address: Institute of Microbiology of the Czech Academy of Sciences, Vídeňská 1083, 142 00 Prague, Czech Republic

⁴Department of Genetics, Faculty of Natural Sciences, Comenius University Bratislava, Ilkovičova 6, 842 15 Bratislava, Slovakia

⁵Present address: Central European Institute of Technology (CEITEC), Masaryk University, Kamenice 753/5, 625 00 Brno, Czech Republic

⁶Department of Genetics and Applied Microbiology, University of Debrecen, Egyetem-tér 1, H-4032 Debrecen, Hungary

⁷Department of Applied Informatics, Faculty of Mathematics, Physics, and Informatics, Comenius University Bratislava, Mlynská dolina, 842 48 Bratislava, Slovakia

*To whom correspondence should be addressed. Email: brejova@fmph.uniba.sk

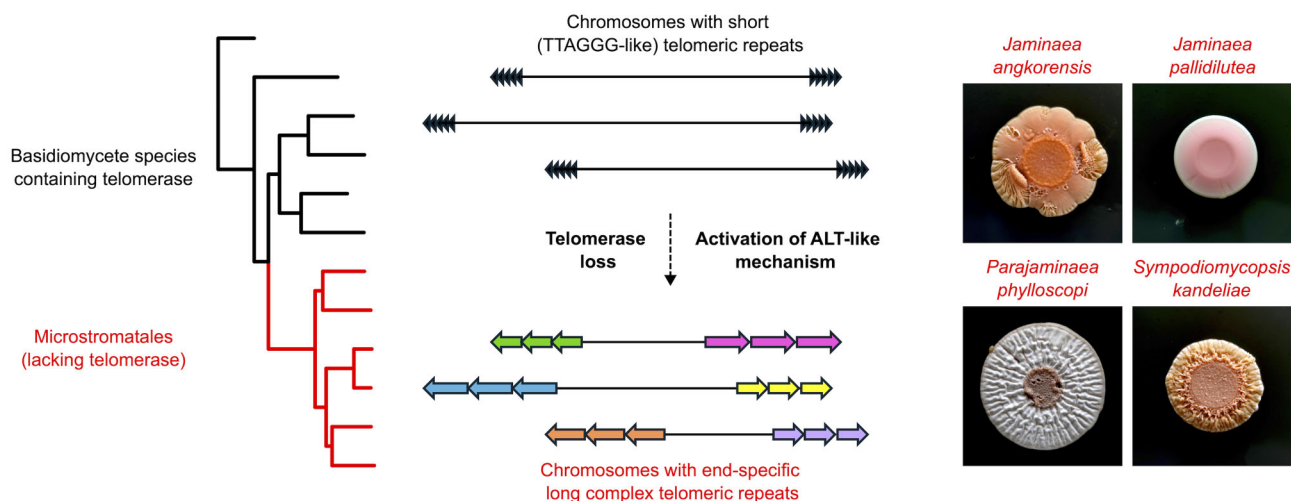
Correspondence may also be addressed to Jozef Nosek. Email: jozef.nosek@uniba.sk

[†]The first two authors should be regarded as Joint First Authors.

Abstract

In most eukaryotes, chromosomal DNA terminates with tandem repeats of a short G-rich motif, such as the canonical TTAGGG sequence. The arrays of telomeric repeats are maintained by telomerase or by alternative lengthening of telomeres (ALT). Here, we report that nuclear chromosomes of several basidiomycetous yeasts classified into the order Microstromatales carry unusual telomeres. We demonstrate that instead of TTAGGG-like repeats these telomeres are composed of unique, tandemly repeated motifs, which are in most cases specific to a particular chromosomal end. In contrast to other basidiomycetes, the Microstromatales genomes lack orthologs coding for the telomerase catalytic subunit Est2 and a shelterin component Tpp1 indicating that noncanonical telomeric repeats are maintained by a telomerase-independent mechanism. We hypothesize that in a common ancestor of Microstromatales the loss of telomerase and Tpp1 was compensated by activation of an ALT-like mechanism, which promoted amplification of various motifs and formation of distinct telomeric repeat arrays at most chromosomal ends.

Graphical abstract



Received: September 24, 2025. Revised: April 26, 2026. Accepted: May 21, 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Introduction

Eukaryotic chromosomes terminate with specialized nucleoprotein structures termed telomeres that protect the genome integrity by preventing nucleolytic degradation and unwanted DNA repair activity [1, 2]. Telomeric DNA typically consists of arrays of a short G-rich motif repeated in tandem. In all major eukaryotic lineages, the motif is represented by the hexanucleotide TTAGGG. Its minor variations (e.g. TTAGG and TTTAGGG) were identified in protists, fungi, plants, and animals. Yet, noncanonical telomeric motifs occur in many species. Particularly puzzling variability was described in ascomycetous yeasts, whose telomeric sequences may be unusually long and/or irregular and may differ even between closely related species [3–9]. The length of telomeric repeats varies between organisms, from tens of base-pairs (bp) in ciliates [10] up to above 150 000 bp in mice [11] or even more, as has been reported in barley callus culture [12]. The very ends of chromosomes usually terminate with a 3′ single-stranded (ss) G-rich overhang of telomeric DNA [10, 13], which has a propensity to fold into quadruplex (G4) structures [14]. Mammalian telomeric DNA is bound by a six-subunit complex dubbed shelterin (Pot1–Tpp1–Tin2–Rap1–Trf1–Trf2) and a heterotrimeric protein complex CST (Ctc1–Stn1–Ten1), which mediate the chromosomal end protection and telomere length regulation [2, 15, 16]. Telomeric repeats are commonly synthesized by the enzyme telomerase reverse transcriptase (TERT), whose core components are the catalytic subunit Est2 and telomerase RNA (TR) providing both a structural scaffold for protein subunits and a template for the synthesis of telomeric repeats [17–19]. The telomerase provides a robust means of telomere maintenance. However, cells lacking its activity, including a significant subset of human cancers and immortalized cell lines, employ ALT (alternative lengthening of telomeres) mechanisms dependent on homologous recombination and DNA repair processes via break-induced replication [1, 20–24]. The ALT pathways may operate both in telomerase-positive and telomerase-negative cells and represent a back-up mechanism that can be selectively induced in cells depleted of telomerase activity either due to a mutation or chemical inhibition. For example, surviving clones arising in a senescent population of mutant yeast *Saccharomyces cerevisiae* lacking functional telomerase amplify either subterminal Y′ elements (type I survivors requiring a strand invasion protein Rad51) or the telomeric repeats (type II survivors dependent on the MRX complex subunit Rad50 and DNA helicase Sgs1) eventually generating hybrid (type I/type II) telomeres. Maintenance of these telomeres relies on recombinational protein Rad52 [25–28]. Yet, the telomerase-deficient mutants with long telomeric repeat arrays may remain viable even in the absence of Rad52 employing the “inherited-long-telomere” pathway [29]. In addition, yeast mutants defective for telomerase, Rad52, and exonuclease Exo1 generate survivors that stabilize their chromosomal termini by formation of palindromes [30]. Analyses of *Schizosaccharomyces pombe* mutants revealed that the loss of telomerase can also be overcome by chromosome circularization [31] or by the “heterochromatin amplification-mediated and telomerase-independent” mechanism [32].

The ALT mechanisms represent the primary mode of telomere maintenance in eukaryotes that naturally lack telomerase. This can be exemplified by dipterans in which vari-

ous noncanonical telomeres have evolved [33]. Studies of chironomids (e.g. *Chironomus pallidivittatus*) and the mosquito *Anopheles gambiae* revealed that their chromosomal ends are composed of complex repeats elongated by a recombinational mechanism [34–36]. In contrast, the fruit fly *Drosophila melanogaster* maintains its chromosomal termini by transposition of non-LTR retrotransposons [37]. Noncanonical telomeric repeats presumably maintained by an ALT mechanism were recently reported also in several nematode species from the genus *Meloidogyne* [38] and *Strongyloides stercoralis* [39]. Interestingly, the first case of a vertebrate naturally lacking the telomerase has been reported in highly regenerative newt *Pleurodeles waltli*, whose chromosomes terminate with interspersed variant repeats, which are maintained by an ALT mechanism [40].

In this study we show that *Jaminala angkorensis*, a basidiomycetous yeast taxonomically classified into the order Microstromatales (class Exobasidiomycetes, subphylum Ustilaginomycotina), lacks canonical TTAGGG-like telomeric repeats maintained by telomerase, which are typical for basidiomycetes [41–43], including the model species, corn smut *Mycosarcoma maydis* (formerly *Ustilago maydis* [44]). By combination of nuclease mapping, long read sequencing, and bioinformatic analysis, we demonstrate that 20 chromosome-sized contigs in the genome assembly of *J. angkorensis* terminate with noncanonical telomeric sequences. The very ends in 36 out of 40 chromosomal termini are composed of distinct motifs varying both in size and nucleotide composition. Some chromosomes terminate with complex arrays, composed of distinct proximal and distal repeats. For comparative purposes, we generated the chromosome-level genome assemblies of additional three Microstromatales species, *Jaminala pallidilutea*, *Parajaminala (Jaminala) phylloscopi*, and *Symptodomyces kandeliae*. These species also lack TTAGGG-like telomeres, and their chromosomes terminate with various noncanonical repeats distinct from those identified in *J. angkorensis*. We also show that all investigated species lack the genes for the catalytic subunit of telomerase Est2 (TERT) and a component of the shelterin complex Tpp1, which forms a structured interface with N-terminal domain of Est2 and is crucial for chromosomal end protection as well as telomerase recruitment to telomeric DNA [45, 46]. Except for Est2 and Tpp1, the orthologs coding for other key proteins involved in telomere maintenance are encoded in the Microstromatales genomes. As both the telomerase and TTAGGG-like telomeric repeats commonly occur in basidiomycetes, we hypothesize that noncanonical telomeres in Microstromatales emerged by activation of an ALT-like mechanism, which compensated for the loss of telomerase-dependent telomere synthesis and the complete erosion of ancestral telomeric repeats.

Materials and methods

Genome and transcriptome sequencing

Yeast strains and cultivation are described in [Supplementary Data \(Supplementary Methods and Supplementary Table S1\)](#). Yeasts from overnight cultures were harvested by centrifugation, resuspended in 100 mM ethylenediaminetetraacetic acid (EDTA, pH 8.0) and 2% (v/v) 2-mercaptoethanol, and incubated for 30 min at room temperature. The cells were pel-

leted, washed, and resuspended in CPES (40 mM citric acid, 120 mM Na₂HPO₄, 20 mM EDTA, 1.2 M sorbitol) supplemented with 5 µg/ml lysing enzymes (Sigma), 3.3 µg/ml chitinase (Sigma), and 4 µl/ml Viscozyme (Sigma) and then incubated for 4.5 h at 37°C with occasional shaking. High molecular weight (HMW) genomic DNA was then isolated using series of phenol and phenol:chloroform:isoamyl alcohol (25:24:1) extractions, precipitated with an equal volume of 96% (v/v) ethanol, washed with 70% (v/v) ethanol, and air-dried. The precipitate was dissolved in TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and RNA was digested by RNase A (150 µg/ml) for 30 min at 37°C. DNA was then extracted by phenol:chloroform:isoamylalcohol (25:24:1), precipitated with 0.1 M NaCl and two volumes of 96% (v/v) ethanol, washed with 70% (v/v) ethanol, air-dried, dissolved in TE, and further purified by anion-exchange chromatography on a Genomic-tip 100/G (Qiagen).

For nanopore sequencing, the libraries were prepared from ~2–3 µg HMW DNA using Ligation Sequencing kits (SQK-LSK108, SQK-LSK109) and sequenced on a MinION Mk1B with a FLO-MIN106 (R9.4.1) flow cell (Oxford Nanopore Technologies) according to manufacturer's instructions, except the fragmentation step was omitted.

Short read sequencing was performed using TruSeq DNA PCR-free (350) paired-end (2 × 101, 2 × 151-nt) libraries on Illumina platforms (HiSeq2000, NovaSeq6000) in MacroGen Europe. For RNA-seq, yeast cultures were grown in YPD media till OD₆₀₀ ~1. Total cellular RNA was extracted with hot acidic phenol [47] and purified using an RNeasy Mini Kit (Qiagen). TruSeq mRNA paired-end (2 × 101) and TruSeq stranded mRNA paired-end (2 × 151-nt) libraries were sequenced on NovaSeq6000 or NovaSeqX platforms in MacroGen Europe. *Jaminala angkorensis* RNA was sequenced also using a Direct RNA Sequencing kit (SQK-RNA004) on a MinION Mk1D and with a FLO-MIN004RA (R10.4.1 RNA) flow cell (Oxford Nanopore Technologies) according to manufacturer's instructions.

For telomere mapping by long read sequencing, two experiments with BAL-31 nuclease were performed. In the first experiment, two samples of *J. angkorensis* HMW DNA (~5 µg) were digested with 2 U BAL-31 nuclease (New England Biolabs) in 50 µl of 1× BAL-31 reaction buffer at 30°C for 15 and 30 min. In the second experiment, *J. angkorensis* HMW DNA samples (~5 µg) were digested with 1 and 3 U BAL-31 in 50 µl of 1× BAL-31 reaction buffer at 30°C for 30 min. In both experiments, the undigested control samples were incubated for 30 min in a 1× BAL-31 reaction buffer without the enzyme. The nuclease reactions were stopped by addition of 100 mM ethylenediaminetetraacetic acid (EGTA, pH 7.0) to a final concentration of 20 mM and the samples were incubated for 20 min at 65°C. DNA was then purified using Agencourt AMPure XP magnetic beads (Beckman Coulter) and the tagmentation library was prepared using a Rapid Barcoding kit (SQK-RBK004) and sequenced on a MinION Mk1B with a FLO-MIN106 (R9.4.1).

All nanopore reads were basecalled by Guppy 4.4.1. Later they were basecalled by newer Dorado base caller v.0.7.2, with dna_r9.4.1_e8_sup@v3.6 model. These newer basecalls were used for estimating telomere array lengths. RNA-seq nanopore data was basecalled with Dorado v.1.4.0 using rna004_130bps_sup@v5.3.0 model.

Genome assemblies

For *J. angkorensis*, we assembled nanopore reads by Miniasm v.0.3-r179 [48] and Flye v.2.8.3 [49] genome assemblers, and Illumina reads by SPAdes v.3.12.0 [50]. The overall assembly is mostly based on the Miniasm assembly, which was polished by two iterations of Racon v1.4.3 [51] using nanopore reads. Based on manual examination of read alignments to all three assemblies and the assemblies to each other, two contigs were connected together, some local misassemblies caused by nanopore sequencing errors were replaced by SPAdes versions and some telomeres missing in the Miniasm assembly or having wrong length were replaced by Flye versions. The resulting assembly was polished using Illumina reads by three iterations of Pilon v.1.24 [52]. A single ribosomal DNA (rDNA) repeat was polished separately and replaced in the assembly because repetitive regions without uniquely mapping reads are difficult to polish by short reads.

The other three species (*J. pallidilutea*, *P. phylloscopi*, and *S. kandeliae*) were assembled by similar methods. *Jaminala pallidilutea* is based mostly on Miniasm assembly, whereas *P. phylloscopi* and *S. kandeliae* are based mostly on Flye assemblies. SPAdes assemblies were not used in these cases. Adjustments were again made by comparing assemblies with long reads and with each other and replacing problematic pieces. Low coverage sequences were trimmed from contig ends. In *J. pallidilutea* and *S. kandeliae*, two contigs were connected manually. In *S. kandeliae*, chromosome 6 was expanded with two 20-kb inverted repeats based on detailed read analysis. Again, Miniasm assemblies were polished by two iterations of Racon v1.4.3 using nanopore reads, the resulting assembly was polished by three iterations of Pilon using short reads and a single ribosomal DNA repeat was polished separately.

Bioinformatic analyses of BAL-31 data

For the BAL-31 experiments, our goal was to compare sequencing coverage in the nuclease-treated and control samples to determine which regions were depleted by the enzyme treatment. As read alignments are not reliable in non-unique repetitive regions, we used *k*-mer analysis (for *k* = 21). First, *k*-mer occurrences were counted in nanopore reads using Jellyfish [53]. For each *k*-mer in the assembly, we tested the depletion of the *k*-mer in the treated sample compared to the control sample using the one-sided Fisher exact test. We selected *k*-mers with *P*-value <2.5% after the Bonferroni multiple testing correction. Next, we computed the coverage of 1 kb windows with step 100 by these significantly depleted *k*-mers. Windows covered >50% we considered as depleted regions. Nearby windows (distance <5 kb) were merged to a single region. The window coverage by significant *k*-mers and centers of depleted regions were visualized along chromosomes using the ggplot2 library in R [54].

Sensitivity of telomeric DNA to BAL-31 nuclease

Total cellular DNA of *J. angkorensis* (~4.5 µg) was digested with 0, 3, 6, 12, and 24 U BAL-31 nuclease (New England Biolabs) for 30 min at 30°C in a final volume 120 µl of 1× BAL-31 reaction buffer. The reaction was stopped by addition of EGTA to 30 mM and heated for 10 min at 65°C. DNA was then purified using Agencourt AMPure XP beads (Beckman Coulter) and digested

with restriction endonuclease EcoRV (New England Biolabs). Agarose gel electrophoresis and Southern blot hybridization were performed essentially as described in the [Supplementary Data](#).

Analysis of single-stranded telomeric DNA

In the first experiment, total cellular DNA samples (75 ng each) of *J. angkorensis* were digested in 50- μ l reactions using 26.7 U of exonuclease I (ExoI, New England Biolabs), 12 U of RecJ_f (New England Biolabs), and 13.3 U of mung bean nuclease (MBN, New England Biolabs) at 37°C for 3 h (ExoI, RecJ_f) or at 30°C for 45 min (MBN). The volumes of native nuclease-treated as well as undigested control samples were adjusted to 100 μ l with 1 $\times$ SSC. For denatured samples, DNA in 50 μ l of 0.2 M NaOH was incubated at 65°C for 15 min and the final volume adjusted to 100 μ l with 1 $\times$ SSC. In the second experiment, *J. angkorensis* DNA (~9 μ g) was digested using 20 U MBN at 30°C for 30 min in a total volume of 80 μ l. The reaction was stopped by addition of 9 μ l of 0.1% (w/v) sodium dodecyl sulfate (SDS) and then diluted with deionised water to 990 μ l. For a control undigested sample, genomic DNA (~9 μ g) was diluted with deionised water to a final volume of 990 μ l. For denatured samples, the 90- μ l aliquots of both the nuclease-digested and control DNAs were mixed with 10 μ l of 2 M NaOH and incubated for 30 min at 65°C. Dot blotting was performed using Immobilon Ny+ membrane (Millipore) and DHM-48 plastic manifold (Scie-Plas) under 80 mbar vacuum. The crosslinking was performed in a CX-2000 UV crosslinker (UVP) for 18 s at 5000 μ J/cm². The membrane strips were pre-hybridized in 5 $\times$ SSC, 5 $\times$ Denhardt solution, and 0.5% (w/v) SDS at 47.5°C for 1 h, and then hybridized overnight with labeled oligonucleotide probes ([Supplementary Table S2](#)) in the same buffer at 47.5°C. The membranes were washed for 15 min with 5 $\times$ SSC; 0.1% (w/v) SDS at 25°C, followed by two more washes at 42°C. Following the exposure to a phosphor storage screen, image acquisition was performed with Amersham Typhoon scanner.

Rolling-circle amplification

To identify telomeric circular molecules such as C-circles, we digested genomic DNA samples (~25 ng per reaction) of *J. angkorensis* with 5 U exonuclease V (RecBCD) (New England Biolabs) at 37°C for 30 min to remove linear molecules. The enzyme was inactivated at 80°C for 20 min. The sample was then amplified using 5 U phi29 DNA polymerase (New England Biolabs) at 30°C for 24 h. The amplification was self-primed, i.e. carried out without the addition of primers. The reaction was stopped by heating at 65°C for 20 min and the samples were analyzed by electrophoresis on a 0.7% (w/v) agarose gel ([Supplementary Fig. S1](#)). About 2 μ g of amplified DNA was purified using Agencourt AMPure XP magnetic beads (Beckman Coulter), digested with 20 U T7 endonuclease I (New England Biolabs) for 15 min at 37°C to remove branched structures, and purified using AMPure XP beads. The samples were then sequenced using a SQK-LSK109 kit on a MinION Mk1B with a FLO-MIN106 (R9.4.1) flow cell (Oxford Nanopore Technologies). The experiment was performed in duplicates. Amplification products from the samples nondigested with exonuclease V (RecBCD) were sequenced as controls. The data were basecalled by Dorado base caller v.0.7.2 with dna_r9.4.1_e8_sup@v3.6 model

and analyzed using methods for detection of telomeric motifs described below.

Dot blot analysis of transcripts derived from chromosomal ends

The samples of *J. angkorensis* total RNA (0.25–1 μ g) were diluted with three volumes of denaturation buffer [60% (v/v) formamide, 7.77% (w/v) formaldehyde, 7.7 $\times$ MOPS buffer (pH 7.0)], incubated for 5 min at 65°C, cooled on ice, and mixed with an equal volume of 20 $\times$ SSC (3 M NaCl, 0.3 M sodium citrate, pH 7.0). The blotting onto an Immobilon Ny+ membrane (Millipore), crosslinking, and hybridization were performed as described above, except that the pre-hybridization and hybridization were done at 42°C and the membranes were washed three times for 15 min (once at 25°C and twice at 37°C). The nuclease-treated control samples were digested with 2 U DNase I (New England Biolabs) for 10 min at 37°C or 20 μ g of RNase A (Invitrogen) for 30 min at 37°C prior blotting.

Telomeric motifs

Telomeric and subtelomeric motifs were found by Tantan [55] with maximum motif length set to 250 bp in the terminal 50 kbp at each chromosomal contig end. The found motifs and their occurrences were then examined manually to exclude other tandem repeats located further from telomeres and to avoid redundant overlapping motifs. All motifs are listed in the 5' to 3' direction leading out of the chromosome (that is, motifs from left chromosomal ends are reversed compared to their appearance in the assembly).

To estimate the lengths of telomeric repeats, we selected nanopore reads that span the unique sequence region nearest to the chromosome end and counted the number of bases covered by telomeric and subtelomeric motifs. To lessen the impact of the reads that end before the chromosome end, we selected only reads longer than a threshold selected based on read length and coverage in a particular species (30 kbp for *J. angkorensis* and *J. pallidilutea*, 20 kbp for *P. phylloscopi*, 25 kbp for *S. kandeliae*). To search for telomeric motifs in reads, we created a hidden Markov model (HMM) with a single background state representing non-repetitive sequence and a simple profile HMM [56] for each distal and proximal motif found at a particular chromosomal end (see details in Supplementary Methods and [Supplementary Fig. S2](#)). HMMs are well suited for modeling tandem repeats [55, 57, 58]. The most probable sequence of states for a given read segment was estimated using the standard Viterbi algorithm [56]. A typical read consists of a subtelomeric segment emitted by the background state, followed by one or several telomeric segments emitted by the profile HMMs and a short background segment located at the very end corresponding to adapters or low-quality sequence. Spans of lower quality sequence emitted by the background state can interrupt telomeric repeats. We report the total number of bases emitted by the profile HMM states as the length of the telomeric array in a given read. We show the violin plot of these lengths, and we use the third quartile as our estimate of the length for a given chromosomal end. We use the third quartile instead of the median, as some of the reads do not extend to the full length of the chromosome.

To assess similarity of motif sequences, we enumerated all 7-mers in each motif (including 7-mers spanning the bound-

ary between successive motif copies) and compared 7-mer multisets by the containment distance defined as $1 - N_i/N_s$, where N_i is the size of the intersection of the 7-mer multisets for the two motifs and N_s is the size of the smaller multiset. The distance is always between 0 and 1, with value 0 if the multisets are identical or one is a subset of the other and value 1 if they have no k -mer in common. We applied the single-linkage agglomerative clustering, where we merge clusters until a given distance threshold, with distance between clusters being the distance between nearest cluster representatives. We show clusters at distance thresholds 0.1, 0.7, and 0.8.

To detect telomeric motifs in Illumina and nanopore RNA-seq data, we first trimmed adapters and low-quality ends from Illumina data with fastp v.0.23.4 [59] with default settings. We then searched all reads against a database of telomeric motifs using blastn v.2.12.0+ [60] with parameters tuned for finding shorter and repetitive matches, namely, -task blastn-short -dust no -soft_masking false -word_size 10 -evalue 1e-5. The database contained concatenation of two motif copies for each motif to allow detection of sequences spanning the motif end. For groups of identical motifs, only one instance was included in the database. Only matches of length at least 30 bp and minimum sequence identity 95% were retained. Since some motifs are similar, some reads matched multiple motifs. In these cases, only the motif with the highest score was retained. For further analysis, we kept whole Illumina read pairs and for nanopore we selected at most 500 bp upstream and downstream for telomeric matches. These sequences were then analyzed by the HMMs used to detect motifs in genomic nanopore reads. Finally, we counted nanopore reads and Illumina read pairs in which an HMM detected the corresponding telomeric motif. We confirmed that all found reads were in TERRA orientation (G-rich strand) when using the stranded Illumina library or direct RNA nanopore sequencing. Investigation of parts of reads upstream of telomeric signals confirmed that when these were sufficiently long to be mapped to the genome, they predominantly map to the sub-telomeric regions corresponding to the telomeric motif found.

To detect telomeric motifs in nanopore sequencing data from rolling-circle amplification (RCA) experiments, we used the same technique as for RNA-seq data, namely first preselecting reads using blast, followed by analysis by the HMMs of telomeric motifs. We required that the length of the region identified by the HMM as containing the repeats of a telomeric motif was at least 1.5 times the length of the corresponding motif. We considered only a single longest telomeric repeat in each read and counted the number of reads identified for each motif. We normalized the counts of discovered reads by the total read count in each experiment to produce the number of motif-containing-reads per million reads. Finally, we aligned all reads to the genome assembly by minimap2 and computed the fraction of the genome covered by at least one read.

Electrophoretic karyotyping

Chromosomal DNA samples were prepared from yeast cells embedded in agarose plugs as follows. About 8×10^8 cells from an overnight culture grown in YPD medium were washed successively with water, 50 mM EDTA (pH 8.0), and CPES buffer (1.2 M sorbitol, 40 mM citric acid, 120

mM Na_2HPO_4 , 20 mM EDTA, pH 8.0). The cells were then resuspended in 1 ml of CPES buffer supplemented with 5 mM dithiothreitol, 15 μg chitinase (Sigma), 20 μl Viscozyme (Sigma), and 24 μg lysing enzyme (Sigma), and incubated for 60 min at 37°C. Spheroplasts were pelleted by centrifugation ($1700 \times g$ for 1 min), resuspended in 1 ml of 1% (w/v) low melting point agarose (UltraPure, Life Technologies) prepared in 50 mM EDTA (pH 8.5), and incubated in 10 mM Tris-HCl (pH 7.5), 0.45 M EDTA (pH 8.5) for 30 min at 37°C. The buffer was then replaced by 0.45 M EDTA (pH 8.5), 1% (w/v) N-lauroyl sarcosine, and 1 mg/ml proteinase K (Applichem), and incubated for 48 h at 50°C and the plugs were washed with 0.45 M EDTA (pH 8.5) and 1% (w/v) N-lauroyl sarcosine. Chromosomal DNAs were separated by pulsed-field gel electrophoresis (PFGE) in a Pulsaphor unit (LKB) with a contour-clamped homogeneous electric field configuration as specified in legend to Fig. 1.

Phylogenetic analysis

The phylogeny was inferred from the genome sequences (see above). Single-copy orthologs were identified in the genome sequences by BUSCO v.5.1.2 in the genome mode with basidiomycota_odb10 lineage [61] and the supermatrix was calculated from 1764 orthologs present in at least 50% of the genomes using the BUSCO_phylogenomics pipeline (options: -psc 50 -supermatrix [62]). The tree was built using FastTree 2.1.11 [63] and the ggtree package in R 4.3.1 [64].

Analysis of protein inventories

The reference genome and predicted proteome sequences of *J. angkorensis*, *J. pallidilutea*, *P. phylloscopi*, and *S. kandeliae* generated in this study and the seven Exobasidiomycete species [*Acaromyces ingoldii* (GCF_003144295.1), *Ceraceosorus guamensis* (GCF_003144195.1), *Jaminaea rosea* (GCF_003144245.1), *Meira miltonrushii* (GCF_003144205.1), *Pseudomicrostroma glucosiphilum* (GCF_003144135.1), *Tilletiaria anomala* (GCF_000711695.1), and *Tilletiopsis washingtonensis* (GCF_003144115.1)] downloaded from the GenBank database (<https://www.ncbi.nlm.nih.gov/datasets/genome/>; [65, 66]) were used in sequence analyses. Amino acid sequences of *M. maydis* proteins downloaded from the Uniprot database (<https://www.uniprot.org/> [67]) were used as queries in BlastP and tBlastN searches (2.15.0+ [60]) of the predicted proteomes and reference genome assemblies, respectively, with an expect value (e -value) threshold 1e-1. Identified proteins were next used as queries in reciprocal Blast searches against the *M. maydis* proteome. In addition, the searches of HMM domain profiles were performed using hmmsearch (HMMER 3.4; <http://hmmer.org/> [68]). Note that the homologs of Ten1 (*J. angkorensis*, *J. pallidilutea*, *P. phylloscopi*) and Trf1/Tay1 (*P. phylloscopi*) are weakly conserved and were not detected in the Blast searches using the *M. maydis* queries. Ten1 in Microstromatales was identified by the presence of the Ten1_2 domain as well as by the BlastP searches with the *S. kandeliae* Ten1 query. Trf1/Tay1 in *P. phylloscopi* was searched using the *J. angkorensis* Trf1/Tay1 query and confirmed by the presence of two SANT/Myb domains.

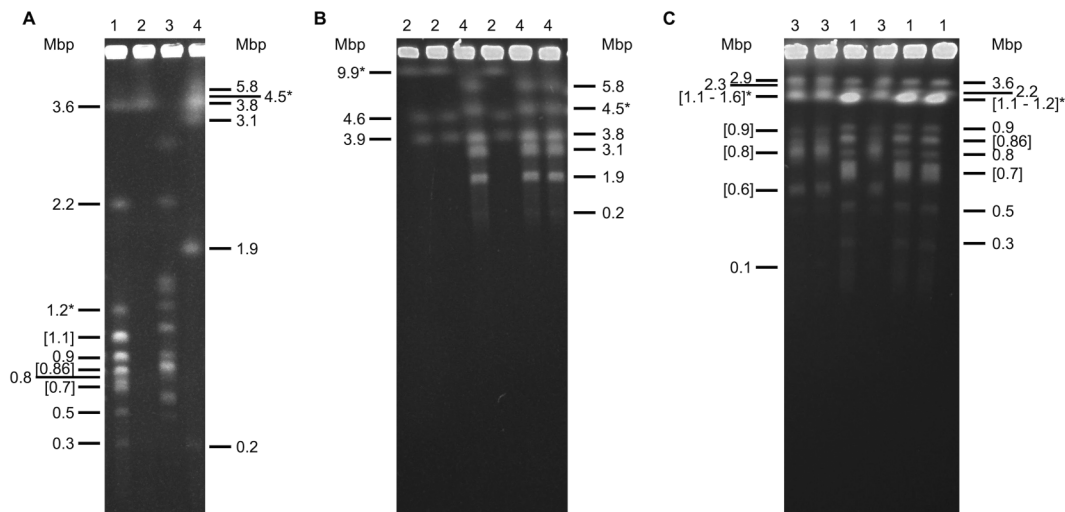


Figure 1. Electrophoretic karyotypes determined by PFGE indicate that nuclear contigs in the Microstromatales genome assemblies likely represent the whole chromosomes. DNA samples of *J. pallidilutea* (lane 1), *P. phylloscopi* (lane 2), *J. angkorensis* (lane 3), and *S. kandeliae* (lane 4) were prepared in agarose plugs and electrophoretically separated (see the “Materials and methods” section for details). The separations were performed in 0.5× TBE at 9°C and following settings: **(A)** a 0.8% (w/v) agarose gel; pulses from 60 to 600 s (linear interpolation) for 72 h at 100 V; **(B)** a 0.8% (w/v) agarose gel; pulses from 800 to 3600 s (linear interpolation) for 96 h at 54 V; and **(C)** a 0.9% (w/v) agarose gel; pulses from 60 to 135 s (linear interpolation) for 26 h at 145 V. The sizes of bands containing multiple co-migrating chromosomes are shown in square brackets; an asterisk indicates the chromosome containing the rDNA cluster repeats in the corresponding species.

Results

Nuclear chromosomes of *J. angkorensis* terminate with noncanonical telomeric repeats that are in most cases chromosomal-end-specific

To investigate the genetic make-up of *J. angkorensis*, we first assembled its genome sequence using long reads generated by nanopore technology on a MinION device (Oxford Nanopore Technologies) and polished it using short Illumina reads (Supplementary Table S1). The resulting genome assembly (20.75 Mb, 60.3% G + C) consists of 20 nuclear contigs with sizes ranging from 0.1 to 2.9 Mb and a 30-kb long contig representing a circular mitochondrial DNA (Table 1 and Supplementary Table S1).

The nuclear contigs correspond to the profile of the electrophoretic karyotype determined by PFGE, although several chromosomes with similar sizes co-migrate as groups and could not be resolved into individual bands (Fig. 1). The PFGE results indicate that the nuclear contigs likely represent the whole chromosomes.

Surprisingly, these contigs lack the telomeric repeats commonly occurring in basidiomycetes (e.g. Guzmán *et al.* [41]). Instead, we found that they terminate with tandem repeats of various noncanonical motifs. Since our attempts to identify TTAGGG-like repeats both in the genome assembly and the raw sequencing reads failed, we decided to map the sequences at the chromosomal termini experimentally. Our approach was modified from the original method developed by Peška *et al.* [69, 70]. It is based on the depletion of telomeric repeats in sequencing reads generated by next-generation sequencing of HMW genomic DNA digested with BAL-31 nuclease. This enzyme specifically shortens the ends of linear DNA molecules, so, compared with the untreated control, the nuclease-treated DNA samples are depleted of the terminal sequences. Specifically, we isolated the genomic DNA from *J. angkorensis* and digested it with BAL-31 nuclease. We performed two experiments, each with two different conditions

for the nuclease reaction. In the first experiment, we used ~0.2 U BAL-31 per µg of DNA and the samples were digested for 15 and 30 min. In the second experiment, the enzymatic digestion was performed for 30 min with ~0.2 U and ~0.6 U BAL-31 per µg of DNA. The DNA fragments from the digested and control samples were then barcoded and sequenced using a tagmentation kit on a nanopore sequencer. In the obtained nanopore reads (Supplementary Table S1), we counted the occurrences for each 21-mer present in the genome assembly. We then highlighted 1-kb windows that were >50% covered by 21-mers that were significantly depleted in the treated sample, compared with the control (Fig. 2).

The sequence near each contig end was depleted in at least one experiment, and as expected, the samples with a higher dose of BAL-31 or longer treatment time generally show more depleted terminal sequences. Only a single region distant from chromosome ends was depleted in one of the experiments. This region is located on chromosome 4 and contains long repeats of rDNA, which are known to be unstable and eventually can generate long linear DNA fragments [71]. The BAL-31 experiments confirmed that the chromosomal ends of all 20 nuclear chromosomes of *J. angkorensis* terminate with tandem repeats of various noncanonical motifs. Overall, we identified 36 unique terminal repeat motifs. Their sizes and G + C content vary from 52 to 178 bp and 41.4% to 66.3%, respectively (Table 1 and Supplementary Tables S3 and S4). While most chromosomal ends terminate with an array composed of a single distinct motif, chromosome 20 contains repeats of the same 130-bp long motif at both ends. In addition, several chromosomal ends are more complex and consist of the proximal and distal repeats. The right ends of chromosomes 2, 8, and 13 possess distal repeats of the same 52-bp long motif, yet these termini differ in the proximal repeats consisting of motifs with the sizes of 80, 126, and 72 bp, respectively. Similarly, a 178-bp long motif occurs at the distal repeats of the right telomere of chromosomes 4 and 18, whose proximal repeats are composed of 73- and 63-bp long motifs, respectively.

Table 1. Characteristics of telomeric repeats in *J. angkorensis*

Chromosome				Left telomere			Right telomere		
No.	Size		Median coverage	Proximal repeat motif	Distal repeat motif	Array size estimate [kb]	Proximal repeat motif	Distal repeat motif	Array size estimate [kb]
	[bp]	G + C%		[bp]	[bp]		[bp]	[bp]	
01	2 873 230	60.4	303		72	2.0		73	8.6
02	2 288 400	60.4	303		167	8.6	80	52 ^a	5.6
03	1 591 733	60.4	304		73	18.0		72	13.5
04	1 373 946	59.8	301		72	4.8	73	178 ^b	12.8
05	1 282 165	60.4	303		73	18.1		59	20.4
06	1 121 287	59.9	298		73	7.3		73	3.0
07	1 106 543	60.6	299		63	15.2		73	12.2
08	907 346	60.6	300		62	8.4	126	52 ^a	19.7
09	883 201	60.7	293		100	6.0		156	4.4
10	847 713	60.3	294		72	17.1	155	137	18.1
11	831 147	60.4	296		152	3.8		73	9.5
12	822 261	60.4	291		153	4.3		83	16.4
13	799 268	60.2	300		73	4.4	72	52 ^a	10.6
14	770 973	60.6	298		63	5.9		72	3.7
15	759 781	60.6	300		73	1.8		72	1.6
16	620 785	60.3	295		72	17.1		59	21.6
17	596 954	59.9	301		151	5.6		72	2.9
18	563 986	60.6	289		83	5.7	63	178 ^b	20.5
19	545 337	60.4	298		74	19.0		164	6.2
20	135 699	59.2	291		130 ^c	4.2		130 ^c	7.0

a, b, or c—these motifs are identical.

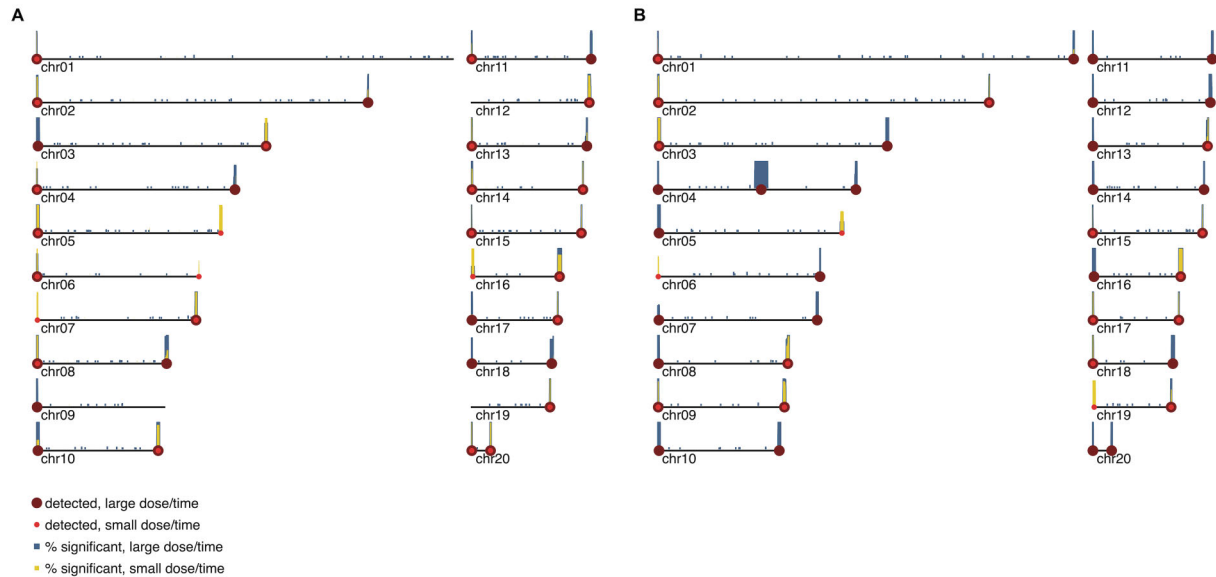


Figure 2. Genomic regions of *J. angkorensis* depleted after BAL-31 treatment coincide with tandem repeats of non-canonical motifs located at the ends of chromosomal contigs. The experiments compare untreated control with samples treated with BAL-31 nuclease with increasing time of digestion (first experiment) or increasing enzyme dose (second experiment). Yellow and blue bars show the coverage of 1-kb windows by 21-mers significantly depleted by the BAL-31 treatment, red circles highlight the regions where depleted 21-mers cover >50% of window length. **(A)** The first experiment, with ~0.2 U BAL-31 per µg of DNA and the treatment times are 15 and 30 min. **(B)** The second experiment, with ~0.2 U and ~0.6 U BAL-31 per µg of DNA and the treatment time is 30 min. In both plots, darker colors and larger circles are used for longer treatment or a larger enzyme dose.

Taken together, the terminal regions of 19 out of 20 chromosomes appear to be unique, represented either by an array of a single distinct telomeric repeat or by a complex of proximal and distal repeats as described above.

To further confirm the identified telomeric motifs, we chose the sequences present at the termini of chromosomes 15 and

20 and tested their sensitivity to BAL-31 nuclease using a conventional terminal restriction enzyme fragment (TRF) analysis based on Southern blot hybridization of restriction enzyme-digested genomic DNA with oligonucleotide probes derived from the motifs identified at the ends of both chromosomes. Our results show that these sequences are sensitive to increas-

ing amounts of the nuclease, thus confirming their location at the chromosomal ends (Fig. 3).

Next, we estimated the sizes of repetitive arrays at the ends of individual chromosomes by analysis of nanopore sequencing reads and found that the approximate lengths of telomeric repeats differ by >10-fold (from ~1.6 to ~21.6 kb; Fig. 4; Table 1; and Supplementary Table S3). We also performed a TRF analysis based on Southern blot hybridization of restriction enzyme-digested genomic DNA with oligonucleotide probes derived from telomeric repeats of chromosomes 1, 3, 15, and 20. This experiment confirmed that individual chromosomal ends differ in the size of TRFs produced by the employed restriction enzymes (Supplementary Fig. S3).

To investigate whether noncanonical telomeres of *J. angkorensis* terminate with a 3' ss overhang, which is considered a conserved feature of telomeres in most eukaryotes [13], we digested its genomic DNA with exonuclease I (ExoI), RecJ_f, and MBN. These enzymes exhibit 3' to 5' (ExoI) and 5' to 3' (RecJ_f) ss-specific exonuclease and ss-specific endonuclease (MBN) activity. Digested DNA samples, together with undigested native and denatured DNA controls, were blotted and hybridized with radiolabeled ss-oligonucleotides derived from telomeric motifs present at both ends of chromosome 15 (G-rich probe) as well as their complementary counterparts (C-rich probe). Our results revealed the telomeric ssDNA regions detectable in native controls, which are almost completely degraded by MBN. In contrast to MBN, the treatment with the exonucleases ExoI and RecJ_f did not remove the signal detected in native controls. Moreover, in contrast to canonical telomeres, the ssDNA regions were detected by both the G-rich and C-rich probes (Fig. 5). These results indicate that chromosomal termini lack ss-overhangs, although we cannot exclude the presence of very short protruding strands that are not detectable by the probes used in the experiment. Rather than overhangs at the very ends of *J. angkorensis* chromosomes, the identified ssDNA regions may represent gaps within the internal parts of the telomeric repeats. Such gaps could result from incomplete DNA synthesis and/or unrepaired recombination events occurring within the telomeric DNA.

Noncanonical telomeres are common in the order Microstromatales

Surprising molecular organization of *J. angkorensis* telomeres prompted us to analyze the chromosomal termini of three additional species *J. pallidilutea*, *P. phylloscopi*, and *S. kandeliae*, which also belong to the order Microstromatales. As in the case of *J. angkorensis*, we assembled their genome sequences using nanopore and Illumina reads (Supplementary Table S1). The nuclear genome assemblies of these species contain 16, 3, and 6 contigs, respectively (Supplementary Table S1), corresponding to their electrophoretic karyotypes (Fig. 1), and therefore they likely represent complete chromosomes. The four Microstromatales genomes differ by the overall size, the G + C content, and the number of chromosomes, which vary from 17.2 Mb (*J. pallidilutea*) to 20.7 Mb (*J. angkorensis*), from 49.8% (*S. kandeliae*) to 60.3% (*J. angkorensis*), and from 3 (*P. phylloscopi*) to 20 (*J. angkorensis*), respectively. The chromosome sizes in these species range from ~0.1 to 9.9 Mb (Supplementary Table S1 and Fig. 1).

Similarly to *J. angkorensis*, the chromosomal contigs of these species lack TTAGGG-like repeats and their ends

are composed of repeats of a single noncanonical motif, which is in most cases chromosomal-end-specific, or they contain a combination of proximal and distal repeats (Supplementary Tables S3 and S4). In the distal repeats of *J. pallidilutea*, *P. phylloscopi*, and *S. kandeliae*, we identified 18, 5, and 7 unique motifs, respectively, with sizes ranging from 41 to 100 bp, 101 to 132 bp, and 55 to 122 bp. The subsequent analysis of nanopore sequencing reads indicated that also the lengths of individual telomeric repeats are highly variable (Supplementary Fig. S5). We found that in each species, some motifs are repeated at different chromosomal ends. To characterize the similarities between individual motifs more broadly, we compared their 7-mer compositions (Fig. 6 and Supplementary Table S5).

As a result, we found out that some groups of motifs within a single species are highly similar (7-mer containment distance <0.1, see the “Materials and methods” section). The similarity of motifs among different species is limited (closest pairs are at a distance of 0.71, involving a group of three *J. pallidilutea* and three *S. kandeliae* motifs) and even within a single species, we see multiple groups of very distant motifs. Furthermore, no interspecific homology was found by Blast with *e*-value threshold 10^{-3} (Supplementary Table S5). This suggests that telomeric motifs do not share a single common evolutionary origin or they mutate at a very high rate.

Inventory of proteins involved in telomere maintenance

The diversity of noncanonical telomeric repeats identified in all four Microstromatales species also raises the question on the mode of telomere maintenance. To provide an insight into its molecular nature, we investigated the inventory of proteins potentially associated with telomeres or mediating telomeric functions. We annotated protein-coding genes in the genome assemblies using Augustus gene prediction software [72] supported by RNA-seq, followed by manual curation of selected genes. In total, we identified 8263, 6317, 6583, and 7053 protein-coding genes in *J. angkorensis*, *J. pallidilutea*, *P. phylloscopi*, and *S. kandeliae*, respectively. We searched the predicted proteomes as well as the genome sequences using the BlastP and tBlastN tools [60], and the queries derived from amino acid sequences of 28 proteins from *M. maydis*, which represents the best characterized basidiomycete model organism in telomere biology and belongs to the same subphylum Ustilaginomycotina as the order Microstromatales. The queries included protein subunits of telomerase (Est1, Est2), components of the shelterin complex (Pot1, Rap1, Tpp1, Trf1/Tay1, Trf2), the CST complex (Stn1, Ten1), the Ku70/80 complex (Ku70, Ku80), the MRN/MRX complex (Mre11, Rad50, Nbs1), ATR/ATM kinases (Mec1, Tel1), DNA helicases (Blm/Sgs1, Dna2, Pif1, Rad54, Srs2), recombinational protein Rad52 and the members of Rad51 family (Rad51, Rad57/Rec2), a homolog of BRCA2 Brh2, exonuclease Exo1, a flap endonuclease Rad27/Fen1, and the ASTRA complex subunit (Tel2). The predicted proteomes were also searched using hhmsearch and HMM profiles of the corresponding protein domains [68]. Note that homologs of the shelterin component Tin2 and Ctc1 subunit of the CST complex were not identified in *M. maydis* [73] and the searches with human protein queries or the corresponding protein domains did not reveal any significant hit in any examined species. For comparative purposes, we also performed these searches in two

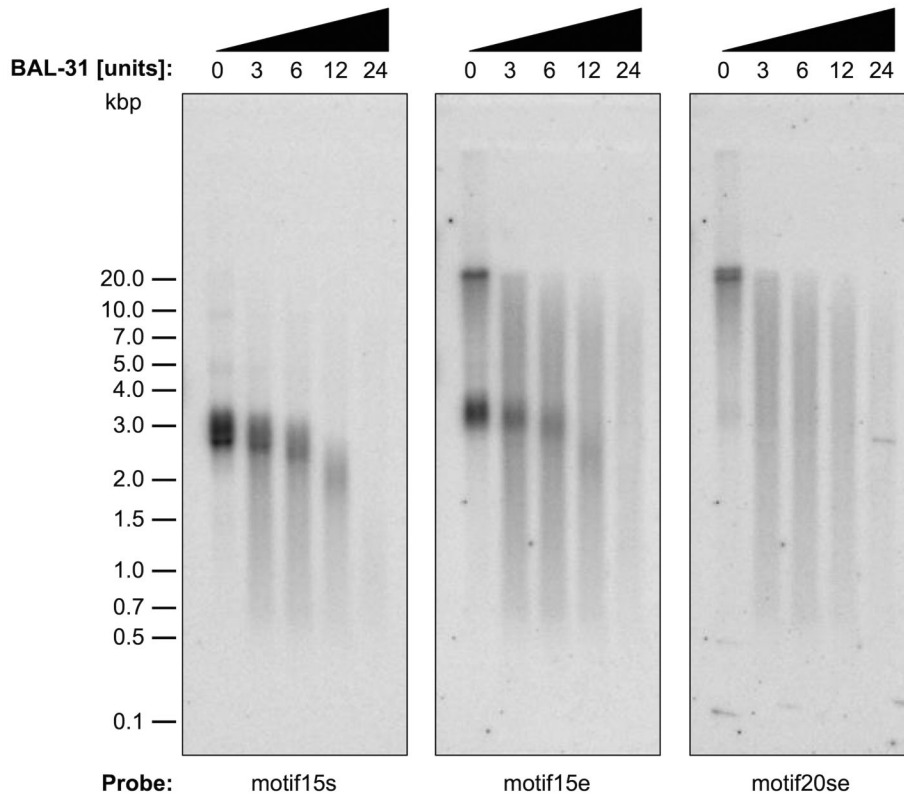


Figure 3. Noncanonical telomeric repeats are sensitive to BAL-31 nuclease. Total cellular DNA of *J. angkorensis* was first digested with 0, 3, 6, 12, and 24 units of BAL-31 nuclease and then cleaved by restriction endonuclease EcoRV. The resulting DNA fragments were electrophoretically separated in a 0.7% (w/v) agarose gel, blotted onto a nylon membrane, and hybridized with radioactively labeled oligonucleotide probes derived from telomeric repeats of chromosome 15 (motif15s, motif15e) and chromosome 20 (motif20se) (Supplementary Table S2; see the “Materials and methods” section for details).

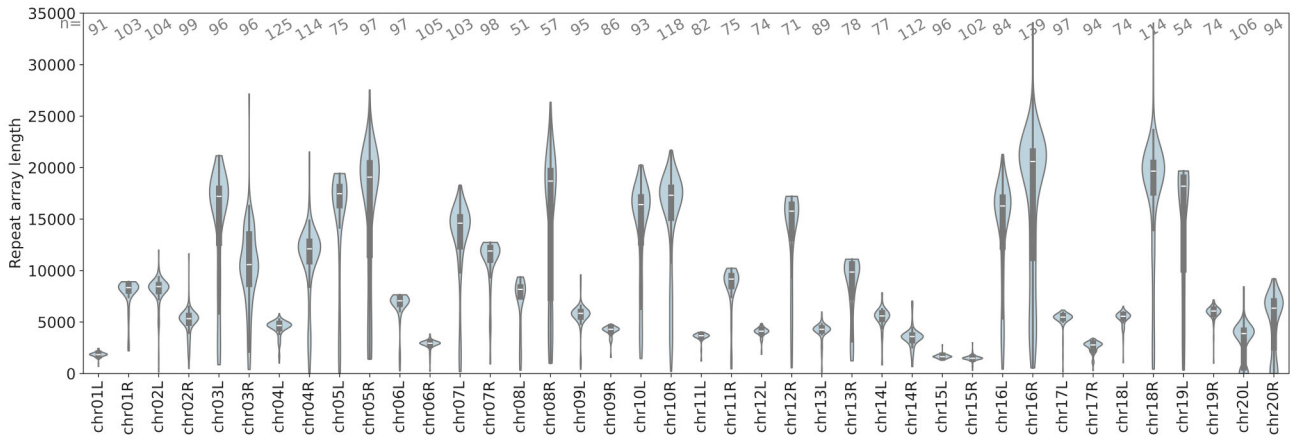


Figure 4. Violin plots of the approximate lengths of telomeric repeats estimated from individual nanopore sequencing reads for each chromosomal end in *J. angkorensis*. Both distal and proximal motifs are considered in the length. Only reads spanning a unique sequence region closest to chromosome end and longer than 30 kb were considered. Nonetheless some reads may represent fragments not reaching the actual chromosome length. The numbers at the top of the plot show the read count used in each violin plot.

additional species—*Jaminaea rosea* and *Pseudomicrostroma glucosiphilum* from the order Microstromatales, as well as in five species (*Acaromyces ingoldii*, *Ceraceosorus guamensis*, *Meira miltomrushi*, *Tilletiaria anomala*, *Tilletiopsis washingtonensis*) classified into other orders of the class Exobasidiomycetes that possess canonical telomeric repeats. Our searches revealed that except for Est2 and Tpp1, the Mi-

crostromatales genomes code for the homologs of all searched protein queries (Fig. 7). The absence of the TERT catalytic subunit Est2 and the shelterin component Tpp1, which is involved in telomerase recruitment to telomeres, indicates that noncanonical telomeres in Microstromatales are maintained by a telomerase-independent mechanism. Moreover, the presence of homologs of other telomeric proteins including those

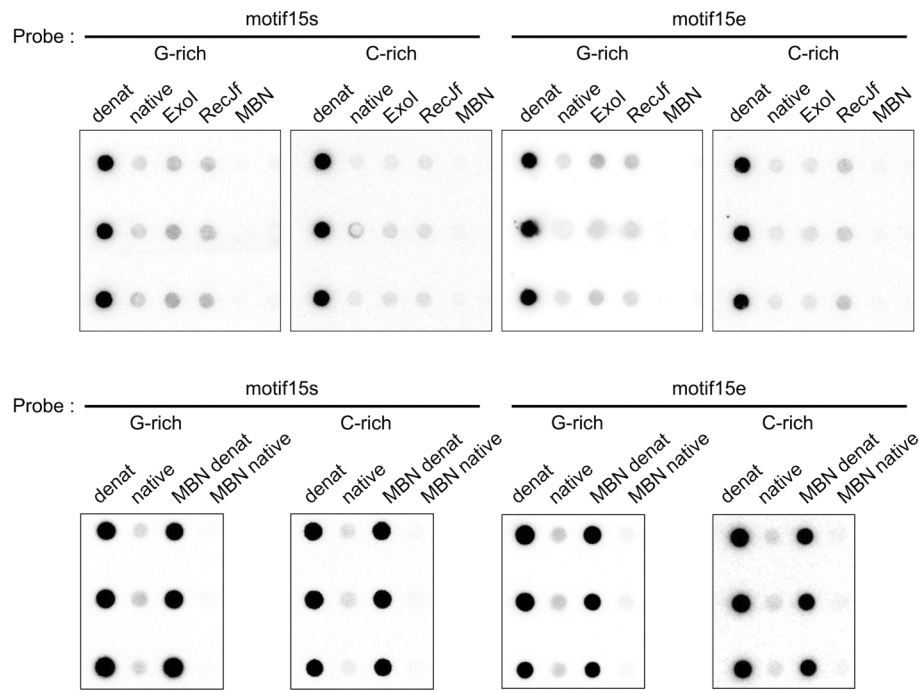


Figure 5. Telomeric regions of *J. angkorensis* contain ssDNA. DNA samples were analyzed by dot-blot hybridization with radioactively labeled oligonucleotide probes derived from telomeric repeats of chromosome 15; G-rich probes: motif15s, motif15e; C-rich probes: motif15s_anti, motif15e_anti (Supplementary Table S2; see the “Materials and methods” section for details). Upper panel: the samples containing denatured DNA, native DNA control, and native DNA samples digested with Exol, RecJf, and MBN. Lower panel: the samples containing denatured DNA, native DNA control, MBN-digested and denatured DNA, and MBN-digested native DNA. The experiments were performed in three technical replicates.

involved in DNA recombination and repair indicates that *J. angkorensis* and other closely related species contain the inventory of molecular tools required for processes involved in telomere maintenance in telomerase-negative ALT cells. The absence of TTAGGG-like repeats further points to similarity with the type I survivors emerging in cultures of *M. maydis* mutants lacking telomerase, that carry telomeres with amplified subterminal repeats and require the activities of Rad51 and Brh2 [74]. We found that Est2 and Tpp1 homologs are missing also in the proteomes of *J. rosea* and *P. glucosiphilum*, whose genome assemblies lack the TTAGGG repeats [65]. Although it would be interesting to investigate whether these two species also possess noncanonical telomeres, currently available genome assemblies were obtained by short read sequencing technology and do not contain complete chromosomes.

Telomeric circles and TERRA-like transcripts in *J. angkorensis*

To investigate whether Microstromatales species exhibit features commonly associated with ALT, we analyzed the presence of telomeric circular DNA and telomeric repeat-containing RNA (TERRA) in *J. angkorensis*.

Telomeric circles are extrachromosomal ss- or ds-DNAs composed of telomeric repeats. These molecules may eventually serve as templates for elongation of telomeric repeats in ALT cells [75]. First, we performed Southern blot hybridization of electrophoretically separated undigested genomic DNA with both G-rich and C-rich ss-telomeric probes derived from chromosome 15. However, we did not detect any low molecular weight fraction that would correspond to circular DNA (Supplementary Fig. S4). Next, we employed

an RCA to increase the detection sensitivity. In this experiment, a sample of genomic DNA was digested using exonuclease V (RecBCD) to remove linear molecules. Next, circular DNAs were amplified using phi29 DNA polymerase (Supplementary Fig. S1) and the RCA products were sequenced using nanopore technology (Supplementary Table S1). The experiment was repeated two times together with the control containing genomic DNA, which was not treated with the nuclease prior amplification. By comparing the resulting nanopore reads with the telomeric motifs and the whole genome assembly, we identified circular molecules containing 5 and 34 out of 6 proximal and 36 distal telomeric motifs of *J. angkorensis*, respectively (Supplementary Table S6).

TERRA molecules are long non-coding RNAs transcribed from the subtelomeric loci toward chromosomal ends. These transcripts invade telomeric ds-DNA resulting in the formation of RNA:DNA structures dubbed R-loops, which trigger homologous recombination-mediated telomere elongation [76]. To detect TERRA-like molecules in *J. angkorensis*, we employed two approaches. First, we performed a dot-blot hybridization of total RNA with oligonucleotide probes derived from the telomeric motifs of chromosome 15. Using the C-rich probes, which hybridize to the G-rich RNA strand, we identified molecules that are resistant to DNase I and sensitive to RNase A indicating that they represent TERRA-like transcripts (Fig. 8). Since the majority of TERRA transcripts in yeast are polyadenylated [77], we searched RNA-seq data for the presence of reads derived from the telomeric repeats. Two RNA-seq experiments were carried out using TruSeq mRNA and TruSeq mRNA stranded libraries on Illumina NovaSeq6000 and NovaSeqX platform, respectively. In addition, we also performed a direct RNA sequencing on a nanopore se-

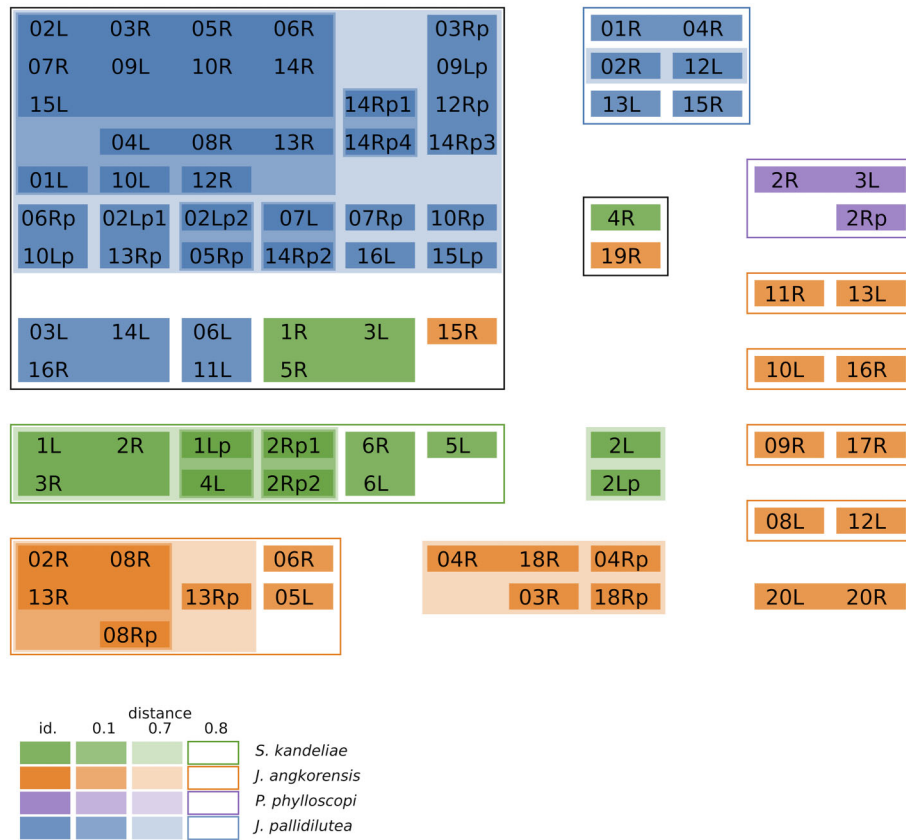


Figure 6. Clustering of noncanonical telomere motifs based on 7-mer containment distance shows multiple distant groups of similar motifs within species, and little similarity of motifs between species. Each text label corresponds to a single motif specified by the chromosome number followed by left (L) or right (R) chromosomal end and letter “p” denoting proximal pattern. For example, 01R is the distal motif at the right end of chromosome 01. The species is distinguished using the background hue according to the legend. The darkest tint of each hue connects motifs that are identical, slightly lighter are those clustering at distance 0.1 (highly similar motifs). Conversely, the lightest tints connect motifs clustering at distance 0.7 and frames enclose clusters at distance 0.8. At this distance, two multi-species clusters use black frames. Motifs that do not cluster with others even at distance 0.8 are not shown (there are 22 in *J. angkorensis*, 5 in *J. pallidilutea*, and 5 in *P. phylloscopi*). Note that maximum possible distance is 1, achieved when the two motifs do not share even a single 7-mer.

quencer. In data from both technologies, we identified a small fraction of sequencing reads corresponding to TERRA-like transcripts (Supplementary Table S7). Overall, 17 out of 42 distinct motifs were detected in at least one sequencing library. In both libraries preserving transcript strand information, all motifs were found on the G-rich strand.

Discussion

Although ascomycetous yeasts *S. cerevisiae* and *S. pombe* have proven to be excellent model organisms in telomere biology, telomeres in basidiomycetes appear more similar to the ends of mammalian chromosomes. This can be exemplified by *M. maydis*, whose telomeric repeats are composed of canonical, human-like motif TTAGGG [41]. Other similarities with human cells include the composition of the shelterin complex and the role of *M. maydis* orthologs of recombination protein Rad51, a BRCA2 homolog Brh2, and a DNA helicase Blm as well as the Ku70/80 and MRN/MRX complexes in telomere maintenance [73, 74, 78–81].

In this study, we show that basidiomycetous yeasts classified into the order Microstromatales are naturally occurring telomerase-deficient organisms. These species lack the orthologs encoding the catalytic subunit of telomerase and the

shelterin component Tpp1, as well as canonical, TTAGGG-like telomeric repeats typical for basidiomycetes. We found that chromosome-sized contigs in the *J. angkorensis* nuclear genome assembly terminate with a plethora of noncanonical repeats that, in most cases, are specific to a particular chromosomal end. The sensitivity of these repeats to BAL-31 nuclease (Figs 2 and 3) indicates that these repeats represent genuine telomeric DNA. This conclusion is further supported by the analysis of chromosome-level genome assemblies of the other three species, *J. pallidilutea*, *P. phylloscopi*, and *S. kandeliae*, whose chromosomal termini also contain various non-canonical repeats. Similar to *J. angkorensis*, all three species lack homologs of the *EST2* and *TPP1* genes. We also found that these two genes as well as the canonical telomeric repeats are missing in the published genome sequences of additional two species from this phylogenetic lineage, *J. rosea* and *P. glucosiphilum* [65]. Although some contigs in the genome assemblies of these two species terminate with tandem repeats of relatively long motifs, the nature of their telomeres remains unknown as these assemblies have not yet been finalized to the level of complete chromosomes. Nevertheless, our results indicate that the absence of Est2, Tpp1, and TTAGGG-like telomeric repeats is a common feature of the order Microstromatales, making these yeasts particularly suitable for the in-

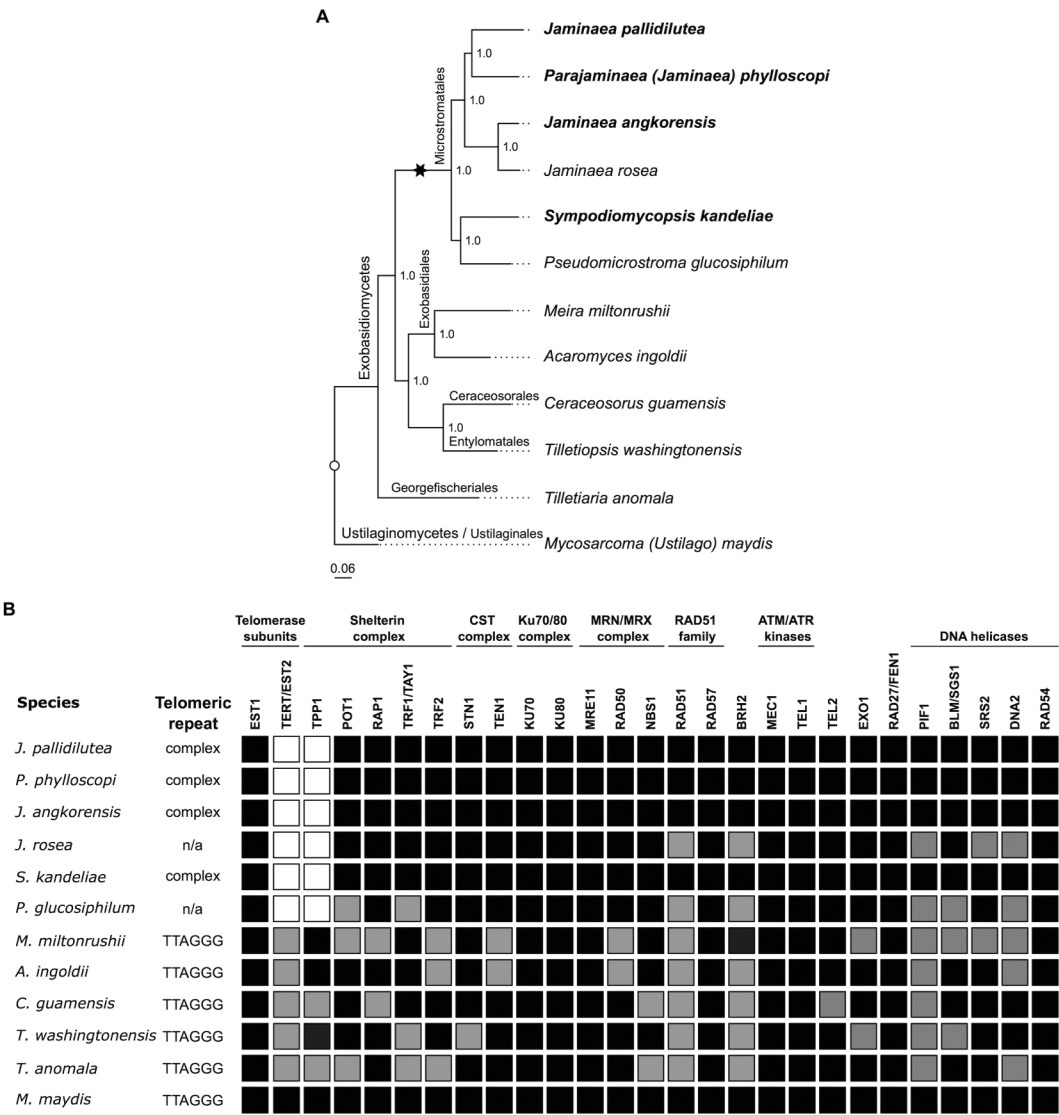


Figure 7. Phylogenetic context of *J. angkorensis* and related Microstromatales species whose chromosomes terminate with noncanonical repeats. The phylogenetic tree was built from 1764 single-copy orthologs present in at least 50% of the reference genome assemblies of indicated species; *M. maydis* was used as an outgroup (see the “Materials and methods” section for details). Analyses of the genome sequences revealed that species classified in the order Microstromatales (lineage is indicated by an asterisk) lack TTAGGG telomere motif as well as orthologs coding for the catalytic subunit of telomerase Est2 and a shelterin component Tpp1. Black and gray squares represent the presence of orthologs of *M. maydis* proteins identified in the predicted proteomes or the genome assemblies by BlastP and tBlastN searches, respectively. The absence of an ortholog is indicated by an empty square; n/a—telomeric repeats are missing in published genome assemblies of *J. rosea* and *P. glucosiphilum*.

vestigation of telomerase-independent mechanisms of telomere maintenance and their evolutionary origin. The Microstromatales species belong to the class Exobasidiomycetes, whose other lineages possess Est2 and Tpp1 as well as the canonical telomeric repeats (Fig. 7). It is therefore likely that the ancestor of Microstromatales contained TTAGGG-like telomeric repeats, and their complete erasure was a direct consequence of telomere shortening in the absence of telomerase activity. Interestingly, except the rDNA repeats, the proximal and distal telomeric repeats represent the most prominent tandem repeats in the examined Microstromatales genomes (Fig. 9).

The sudden change of telomeric DNA sequence from TTAGGG-like to noncanonical repeats may provide a burden on the stability of nucleoprotein complexes at the chromosomal ends. Conversely, structural alterations of these complexes may facilitate the sequence alterations. For example,

the diversification of telomeric DNA motifs in distinct lineages of ascomycetous yeasts was accompanied by recruitment of different telomeric DNA-binding proteins (TBPs), such as Rap1, Tay1, and Taz1, whose orthologs are loosely conserved indicating they underwent rapid evolution. In addition, experimental studies of DNA-binding properties revealed that these proteins recognize a broader spectrum of target motifs, which opens a space for changes of telomeric sequences [8, 82–84]. This can be exemplified by Tay1 from the ascomycetous yeast *Yarrowia lipolytica*, which binds both *Y. lipolytica*-specific and human-like telomeric motifs [85]. Rapid diversification of the Tay1 protein family in fungi further underscores possible alterations of its DNA-binding properties [84]. Similarly, limited DNA-binding specificity of conserved TBP CEH-37 to various telomeric motifs is thought to underlie the diversification of telomeric repeat motifs in ne-

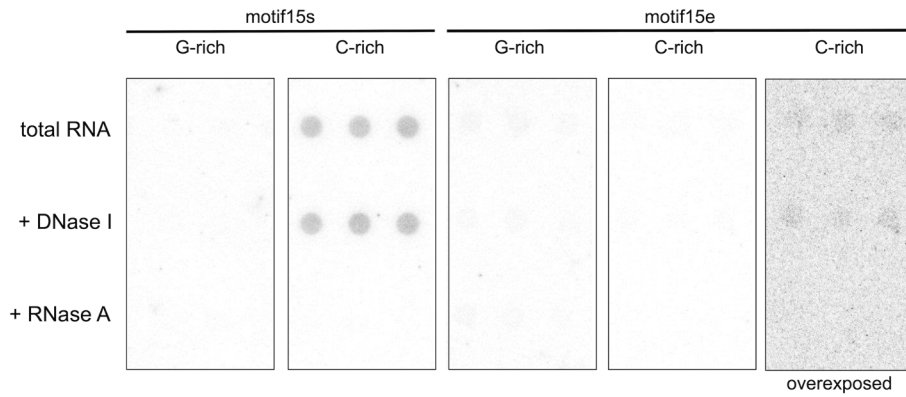


Figure 8. TERRA-like transcripts identified by dot-blot hybridization. Total RNA samples from *J. angkorensis* and the controls digested with DNase I or RNase A were blotted onto a nylon membrane and hybridized with radioactively labeled oligonucleotide probes derived from the telomeric repeats of chromosome 15; G-rich probes: motif15s, motif15e; C-rich probes: motif15s_anti, motif15e_anti (Supplementary Table S2; see the “Materials and methods” section for details).

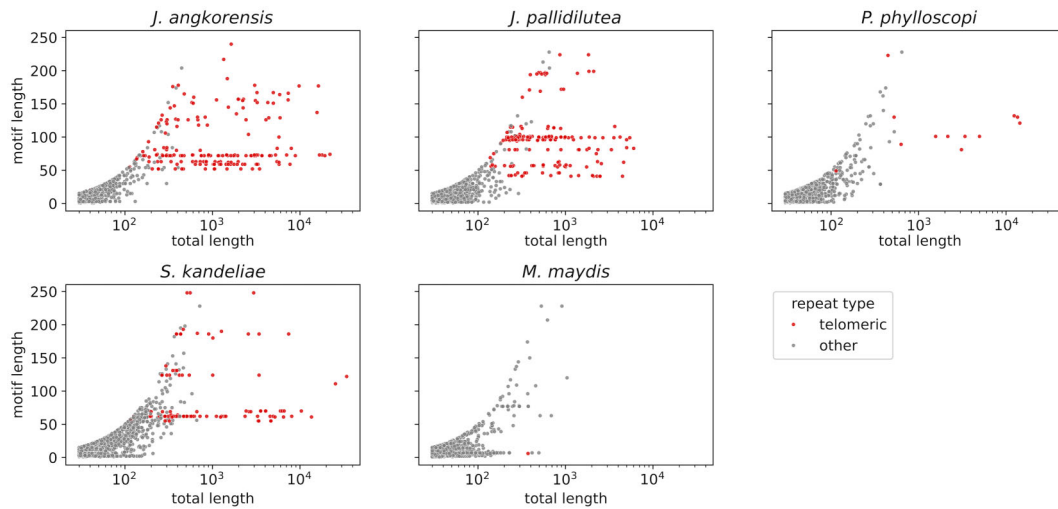


Figure 9. Tandem repeats present in the Microstromatales genomes. All tandem repeats of motif length up to 250 bp with at least two full copies of the motif and total length at least 30 bp found by Tantan [55] in the Microstromatales genomes and the genome of *M. maydis* (NCBI RefSeq assembly GCF_000328475.2). Each point represents one tandem repeat motif, with the total tandem array length as the x coordinate (in log scale) and the motif length as the y coordinate. Color distinguishes the repeats overlapping annotated telomeric repeats from others. In each species, the longest repeat arrays are formed by telomeric motifs. Note that some found tandem repeats overlapping annotated telomeres are relatively short; these are often fragments separated by errors in sequence assembly or shorter stretches of proximal repeats. *M. maydis* is shown for comparison. Only one telomeric repeat array of the TTAGGG motif is present in its assembly.

matodes [86]. The amino acid sequences of Microstromatales Trf1/Tay1 and Trf2 proteins are also weakly conserved. For example, the overall sequence identity of *J. angkorensis* and *M. maydis* orthologs is only 35.5 and 16.9%, respectively. The most conserved regions in their sequences correspond to the MYB domains involved in binding to DNA. The two MYB domains present in Trf1/Tay1 and the MYB domains in Trf2 of these two species exhibit the pairwise identity of 66.2, 65.2, and 48.9%, respectively (Supplementary Fig. S6A). Since the predicted structures of the MYB domains appear to be conserved (Supplementary Fig. S6B), the Microstromatales telomeric DNA-binding proteins may have similar roles as their counterparts in *M. maydis*, though they may bind to different DNA sequence motifs. However, the identification of binding sites in relatively long sequences of divergent noncanonical telomeric motifs is challenging and will require detailed biochemical characterization of both proteins.

In addition to the orthologs of Trf1/Tay1 and Trf2, we identified other two components of the shelterin complex (Pot1 and Rap1) encoded in the genome sequences of Microstromatales, but this lineage lacks Tpp1 (Fig. 7). In mammalian cells, Tpp1 does not bind directly to the telomeric DNA, it forms a stable heterodimer with ssDNA-binding protein Pot1 [87, 88], interacts with Tin2 [89], recruits telomerase to telomeric DNA, and stimulates its processivity [90, 91]. Mice with Tpp1 null mutation die perinatally and their cells exhibit accelerated telomere shortening as well as increased frequency of chromosome end fusions and fragility demonstrating that Tpp1 also prevents the activation of DNA damage response (DDR) and maintains the stability of chromosomal ends [92]. Notably, the selective degradation of Tpp1 during HSV-1 virus infection results in a rapid loss of telomeric repeats, dissociation of shelterin components from telomeres, DDR activation, and promotion of the viral DNA replication, presumably by

recruiting telomere-associated proteins to the virus replication compartments [93]. The loss of Tpp1 thus results in uncapped chromosomes, which tend to undergo various rearrangements eventually leading to karyotype alterations. It is therefore of note that the number of chromosomes in the examined Microstromatales species vary from 3 (*P. phylloscopi*) up to 20 (*J. angkorensis*). It is also conceivable that the loss of both Est2 and Tpp1 in an ancestor of Microstromatales led to the complete removal of ancestral TTAGGG-like repeats along with remodeling of telomeric nucleoprotein complexes that could be prerequisites for the emergence of noncanonical repeats, which in turn stabilized the chromosomal ends.

While most telomeric repeats in the studied Microstromatales species consist of a unique motif, the ends of several chromosomes are complex and contain combinations of proximal and distal repeats. In a subset of telomeres, the identified distal repeats are composed of the same motif, which indicates that the motif could expand to the ends of multiple chromosomes. This points to two intriguing evolutionary scenarios. In the first, the proximal repeats may represent remnants of former telomeric repeats that were capped by repeats of a new motif. The occurrence of several types of proximal repeats on some chromosomal ends of *J. pallidilutea* (i.e. the left and right telomere of chromosomes 2 and 14, respectively) and *S. kandeliae* (i.e. the right telomere of chromosome 2) is suggestive of multiple rounds of such replacements. This idea is also supported by *P. phylloscopi* chromosome 2, where the 101-bp long motifs present at the distal array of the left telomere and the proximal array of the right telomere are highly similar. In some cases, different telomeres carry the identical motifs (e.g. both telomeres of chromosome 20 in *J. angkorensis*) indicating that the original repeats could be erased and completely replaced by a motif spread from another telomere. Such propagation of telomeric motifs can be facilitated by extrachromosomal telomeric circles through a recombination-based mechanism. Alternatively, novel motifs could first emerge and expand as proximal repeats, and if they provide functional solutions to the end-protection and end-replication problems, could eventually become functional telomeres. The latter idea is consistent with the hypothesis that rapidly evolving subtelomeric regions can act as drivers of divergence in both telomeric DNA sequences and associated DNA-binding proteins [94]. For example, in the yeast *Nauvovozyma castellii*, the loss of telomerase yields amplification and spreading of TelKO elements derived from subtelomeric regions of the chromosomes [95] indicating that subtelomeres can indeed actively participate in the protection of chromosomal ends in case of a sudden loss of telomerase.

It has been demonstrated that reintroduction of telomerase activity into *S. cerevisiae* type I and type II survivors restores the telomerase-dependent mode of telomere maintenance as well as the wild-type length of telomeric repeats [96]. Although telomerase can extend some non-canonical primers [97, 98], it seems unlikely that it would be able to recognize diverse telomeric motifs present at the chromosomal termini of Microstromatales for addition of canonical telomeric repeats. However, the lack of tools for genetic manipulation of Microstromatales currently does not allow testing if reintroduction of telomerase and TPP1 would restore ancestral canonical telomeres in these species.

The molecular organization of the chromosomal-end-specific telomeric repeat arrays in Microstromatales is reminiscent of distinct telomeric repeats discovered at the ends

of linear mitochondrial DNA molecules in some *Tetrahymena* species [99] and composite telomeres recently reported at the ends of nuclear chromosomes in nematodes of the genus *Meloidogyne* [38]. Importantly, the presence of multiple types of telomeric repeats within the same cell raises questions on the mode of maintenance of individual chromosomal ends. Our searches for proteins mediating the telomere functions in Microstromatales genome sequences identified orthologs of key players participating in these processes, including those essential for ALT (Fig. 7). Therefore, we hypothesize that noncanonical telomeres of Microstromatales are maintained via an ALT-like pathway similar to those operating in telomerase-deficient mutants. This idea is also supported by variable lengths of individual telomeric repeats (Fig. 4 and Supplementary Fig. S5) as well as by identification of telomeric circular DNAs (Supplementary Table S6) and TERRA-like transcripts (Fig. 8 and Supplementary Table S7), which represent hallmarks of ALT cells [75, 76, 100]. Moreover, the absence of ss-overhangs at the very ends of *J. angkorensis* chromosomes is reminiscent of blunt-ended chromosomes in angiosperm plants [101] in which the telomere protection is mediated by the Ku70/80 complex [102]. Based on the telomeric features identified in Microstromatales, we posit that chromosomal ends of these fungi are maintained by a recombination-dependent process. However, the exact molecular mechanism underlying this process remains to be elucidated.

It has been shown that the telomerase-deficient mutants of *M. maydis* progressively lose telomeric repeats and surviving clones emerge in senescent cultures. Such cells resemble the type I and, albeit rarely, type II survivors of *S. cerevisiae est2* mutants [74, 103]. Since the genome sequences of *J. angkorensis* and related species lack TTAGGG-like repeats, they remind the type I survivors, which maintain the chromosomal termini by recombination-mediated amplification of subtelomeric repeats such as UTASa in *M. maydis* [104] or Y' elements in *S. cerevisiae* [105]. By analogy, the telomeric repeats of Microstromatales could arise from a similar source. Surprising variability of telomeric motifs indicates that either the ancestral chromosomes differed by subterminal repeats or these repeats underwent a rapid diversification. This goes in line with the studies showing that subtelomeres are rapidly evolving regions of the eukaryotic genomes, which represent hotspots for genetic innovations [106].

Acknowledgements

Author contributions: Broňa Brejová (Conceptualization [equal], Data curation [lead], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Methodology [equal], Project administration [supporting], Resources [equal], Software [lead], Supervision [equal], Validation [equal], Visualization [equal], Writing – original draft [equal], Writing – review & editing [equal]), Viktória Hodorová (Investigation [equal], Methodology [equal], Visualization [equal], Writing – review & editing [equal]), Hana Lichanová (Investigation [equal], Methodology [equal], Writing – review & editing [supporting]), Askar Gafurov (Investigation [equal], Methodology [equal], Software [supporting], Writing – review & editing [supporting]), Dominik Bujna (Investigation [equal], Methodology [equal], Writing – review & editing [supporting]), Filip Brázdovič (Investigation [equal], Methodology [equal], Visualization [supporting], Writing – review & editing [equal]), Filip Červenák (Investigation [equal],

Methodology [equal], Visualization [supporting], Writing – review & editing [equal]), Tomáš Petrík (Investigation [equal], Methodology [equal], Visualization [supporting], Writing – review & editing [supporting]), Eva Hegedúsová (Investigation [supporting], Methodology [supporting], Writing – review & editing [supporting]), Michaela Forgáčová Jakúbková (Investigation [supporting], Methodology [supporting], Writing – review & editing [supporting]), Martina Neboháčová (Investigation [supporting], Methodology [supporting], Writing – review & editing [supporting]), Ľubomír Tomáška (Conceptualization [equal], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Methodology [equal], Resources [equal], Supervision [equal], Validation [equal], Writing – review & editing [equal]), Matthias Sipiczki (Investigation [supporting], Methodology [supporting], Resources [equal], Writing – review & editing [supporting]), Tomáš Vinař (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Methodology [equal], Resources [equal], Software [equal], Supervision [equal], Validation [equal], Writing – review & editing [equal]), and Jozef Nosek (Conceptualization [equal], Data curation [supporting], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Methodology [equal], Project administration [lead], Resources [equal], Supervision [equal], Validation [equal], Visualization [equal], Writing – original draft [equal], Writing – review & editing [equal])

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This research was supported by grants from the Slovak Research and Development Agency [18-0239 and 22-0144 (to J.N.), 23-0056 (to L.T.)] and the Scientific Grant Agency of the Ministry of Education, Science and Sport of the Slovak Republic [1/0567/26 (to T.V.), 1/0234/23 (to J.N.), 1/0031/24 (to L.T.), and 1/0140/25 (to B.B.)]. Additional support was provided by the Advancing University Capacity and Competence in Research, Development and Innovation (ACCORD) project and the European Union NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V06-00079 (to J.N.). Funding to pay the Open Access publication charges for this article was provided by the Slovak Research and Development Agency (22-0144, 23-0056) and/or the Scientific Grant Agency of the Ministry of Education, Science and Sport of the Slovak Republic (1/0234/23, 1/0031/24).

Data availability

The genome assemblies, Illumina, and nanopore reads have been deposited in the European Nucleotide Archive (ENA) and GenBank databases. The accession numbers are shown in [Supplementary Table S1](#). The genomes can be interactively visualized in the genome browser at <http://genome.compbio.fmph.uniba.sk/>.

Code availability: Software developed for the analysis in this article is available at <https://github.com/fmf-compbio/jamang-code> (<https://doi.org/10.5281/zenodo.20329256>). Existing software tools are listed in the “Materials and methods” section.

References

- McEachern MJ, Krauskopf A, Blackburn EH. Telomeres and their control. *Annu Rev Genet* 2000;34:331–58. <https://doi.org/10.1146/annurev.genet.34.1.331>
- de Lange T. Shelterin-mediated telomere protection. *Annu Rev Genet* 2018;52:223–47. <https://doi.org/10.1146/annurev-genet-032918-021921>
- Moyzis RK, Buckingham JM, Cram LS *et al*. A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes. *Proc Natl Acad Sci USA* 1988;85:6622–6. <https://doi.org/10.1073/pnas.85.18.6622>
- Meyne J, Ratliff RL, Moyzis RK. Conservation of the human telomere sequence (TTAGGG)_n among vertebrates. *Proc Natl Acad Sci USA* 1989;86:7049–53. <https://doi.org/10.1073/pnas.86.18.7049>
- Gunišová S, Elboher E, Nosek J *et al*. Identification and comparative analysis of telomerase RNAs from *Candida* species reveal conservation of functional elements. *RNA* 2009;15:546–59. <https://doi.org/10.1261/rna.1194009>
- Fulnecková J, Ševčíková T, Fajkus J *et al*. A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biol Evol* 2013;5:468–83. <https://doi.org/10.1093/gbe/evr019>
- Fajkus P, Peška V, Šitová Z *et al*. Allium telomeres unmasked: the unusual telomeric sequence (CTCGGTTATGGG)_n is synthesized by telomerase. *Plant J* 2016;85:337–47. <https://doi.org/10.1111/tpl.13115>
- Červenák F, Šepšiová R, Nosek J *et al*. Step-by-step evolution of telomeres: lessons from yeasts. *Genome Biol Evol* 2021;13:evaa268. <https://doi.org/10.1093/gbe/evaa268>
- Peška V, Fajkus P, Bubeník M *et al*. Extraordinary diversity of telomeres, telomerase RNAs and their template regions in Saccharomycetaceae. *Sci Rep* 2021;11:12784. <https://doi.org/10.1038/s41598-021-92126-x>
- Klobutcher LA, Swanton MT, Donini P *et al*. All gene-sized DNA molecules in four species of hypotrachs have the same terminal sequence and an unusual 3' terminus. *Proc Natl Acad Sci USA* 1981;78:3015–9. <https://doi.org/10.1073/pnas.78.5.3015>
- Kipling D, Cooke HJ. Hypervariable ultra-long telomeres in mice. *Nature* 1990;347:400–2. <https://doi.org/10.1038/347400a0>
- Kilian A, Stiff C, Kleinhofs A. Barley telomeres shorten during differentiation but grow in callus culture. *Proc Natl Acad Sci USA* 1995;92:9555–9. <https://doi.org/10.1073/pnas.92.21.9555>
- Henderson ER, Blackburn EH. An overhanging 3' terminus is a conserved feature of telomeres. *Mol Cell Biol* 1989;9:345–8. <https://doi.org/10.1128/mcb.9.1.345-348.1989>
- Tang J, Kan ZY, Yao Y *et al*. G-quadruplex preferentially forms at the 3' end of vertebrate telomeric DNA. *Nucleic Acids Res* 2008;36:1200–8. <https://doi.org/10.1093/nar/gkm1137>
- Lim CJ, Cech TR. Shaping human telomeres: from shelterin and CST complexes to telomeric chromatin organization. *Nat Rev Mol Cell Biol* 2021;22:283–98. <https://doi.org/10.1038/s41580-021-00328-y>
- Takai H, Aria V, Borges P *et al*. CST-polymerase α -primase solves a second telomere end-replication problem. *Nature* 2024;627:664–70. <https://doi.org/10.1038/s41586-024-07137-1>
- Greider CW, Blackburn EH. The telomere terminal transferase of *Tetrahymena* is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* 1987;51:887–98. [https://doi.org/10.1016/0092-8674\(87\)90576-9](https://doi.org/10.1016/0092-8674(87)90576-9)

18. Blackburn EH, Collins K. Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harb Perspect Biol* 2011;3:a003558. <https://doi.org/10.1101/cshperspect.a003558>
19. Lue NF, Autexier C. Orchestrating nucleic acid–protein interactions at chromosome ends: telomerase mechanisms come into focus. *Nat Struct Mol Biol* 2023;30:878–90. <https://doi.org/10.1038/s41594-023-01022-7>
20. McEachern MJ, Haber JE. Break-induced replication and recombinational telomere elongation in yeast. *Annu Rev Biochem* 2006;75:111–35. <https://doi.org/10.1146/annurev.biochem.74.082803.133234>
21. Cesare AJ, Reddel RR. Alternative lengthening of telomeres: models, mechanisms and implications. *Nat Rev Genet* 2010;11:319–30. <https://doi.org/10.1038/nrg2763>
22. Roumelioti FM, Sotiriou SK, Katsini V *et al.* Alternative lengthening of human telomeres is a conservative DNA replication process with features of break-induced replication. *EMBO Rep* 2016;17:1731–7. <https://doi.org/10.15252/embr.201643169>
23. Apte MS, Cooper JP. Life and cancer without telomerase: ALT and other strategies for making sure ends (don't) meet. *Crit Rev Biochem Mol Biol* 2017;52:57–73. <https://doi.org/10.1080/10409238.2016.1260090>
24. Hoang SM, O'Sullivan RJ. Alternative lengthening of telomeres: building bridges to connect chromosome ends. *Trends Cancer* 2020;6:247–60. <https://doi.org/10.1016/j.trecan.2019.12.009>
25. Lundblad V, Blackburn EH. An alternative pathway for yeast telomere maintenance rescues est1- senescence. *Cell* 1993;73:347–60. [https://doi.org/10.1016/0092-8674\(93\)90234-h](https://doi.org/10.1016/0092-8674(93)90234-h)
26. Chen Q, Ijima A, Greider CW. Two survivor pathways that allow growth in the absence of telomerase are generated by distinct telomere recombination events. *Mol Cell Biol* 2001;21:1819–27. <https://doi.org/10.1128/MCB.21.5.1819-1827.2001>
27. Huang P, Pryde FE, Lester D *et al.* SGS1 is required for telomere elongation in the absence of telomerase. *Curr Biol* 2001;11:125–9. [https://doi.org/10.1016/s0960-9822\(01\)00021-5](https://doi.org/10.1016/s0960-9822(01)00021-5)
28. Kockler ZW, Comeron JM, Malkova A. A unified alternative telomere-lengthening pathway in yeast survivor cells. *Mol Cell* 2021;81:1816–29. <https://doi.org/10.1016/j.molcel.2021.02.004>
29. Grandin N, Charbonneau M. Telomerase- and Rad52-independent immortalization of budding yeast by an inherited-long-telomere pathway of telomeric repeat amplification. *Mol Cell Biol* 2009;29:965–85. <https://doi.org/10.1128/MCB.00817-08>
30. Maringe L, Lydall D. Telomerase- and recombination-independent immortalization of budding yeast. *Genes Dev* 2004;18:2663–75. <https://doi.org/10.1101/gad.316504>
31. Nakamura TM, Cooper JP, Cech TR. Two modes of survival of fission yeast without telomerase. *Science* 1998;282:493–6. <https://doi.org/10.1126/science.282.5388.493>
32. Jain D, Hebden AK, Nakamura TM *et al.* HAATI survivors replace canonical telomeres with blocks of generic heterochromatin. *Nature* 2010;467:223–7. <https://doi.org/10.1038/nature09374>
33. Mason JM, Randall TA, Čapková Frydrychová R. Telomerase lost? *Chromosoma* 2016;125:65–73. <https://doi.org/10.1007/s00412-015-0528-7>
34. Saiga H, Edström JE. Long tandem arrays of complex repeat units in *Chironomus* telomeres. *EMBO J* 1985;4:799–804. <https://doi.org/10.1002/j.1460-2075.1985.tb03700.x>
35. Nielsen L, Edström JE. Complex telomere-associated repeat units in members of the genus *Chironomus* evolve from sequences similar to simple telomeric repeats. *Mol Cell Biol* 1993;13:1583–9. <https://doi.org/10.1128/mcb.13.3.1583-1589.1993>
36. Roth CW, Kobeski F, Walter MF *et al.* Chromosome end elongation by recombination in the mosquito *Anopheles gambiae*. *Mol Cell Biol* 1997;17:5176–83. <https://doi.org/10.1128/MCB.17.9.5176>
37. Levis RW, Ganesan R, Houtchens K *et al.* Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* 1993;75:1083–93. [https://doi.org/10.1016/0092-8674\(93\)90318-k](https://doi.org/10.1016/0092-8674(93)90318-k)
38. Mota APZ, Koutsovoulos GD, Perfus-Barbeoch L *et al.* Unzipped genome assemblies of polyploid root-knot nematodes reveal unusual and clade-specific telomeric repeats. *Nat Commun* 2024;15:773. <https://doi.org/10.1038/s41467-024-44914-y>
39. Chung G, Piano F, Gunsalus KC. TeloSearchLR: an algorithm to detect novel telomere repeat motifs using long sequencing reads. *G3* 2025;15:jka062. <https://doi.org/10.1093/g3journal/jka062>
40. Yu Q, Gates PB, Rogers S *et al.* Telomerase-independent maintenance of telomere length in a vertebrate. *bioRxiv*, <https://doi.org/10.1101/2022.03.25.485759>, 26 March 2022, preprint; not peer reviewed.
41. Guzmán PA, Sánchez JG. Characterization of telomeric regions from *Ustilago maydis*. *Microbiology* 1994;140:551–7. <https://doi.org/10.1099/00221287-140-3-551>
42. Sanpedro-Luna JA, Jacinto-Vázquez JJ, Anastacio-Marcelino E *et al.* Telomerase RNA plays a major role in the completion of the life cycle in *Ustilago maydis* and shares conserved domains with other Ustilaginales. *PLoS One* 2023;18:e0281251. <https://doi.org/10.1371/journal.pone.0281251>
43. Sanpedro Luna JA, Sánchez Alonso P. Identification of the putative gene encoding the RNA subunit of telomerase in *Malassezia* clade species through comparative genomic analysis. *G3* 2025;15:jka185. <https://doi.org/10.1093/g3journal/jka185>
44. McTaggart AR, Shivas RG, Boekhout T *et al.* *Mycosarcoma* (Ustilaginaceae), a resurrected generic name for corn smut (*Ustilago maydis*) and its close relatives with hypertrophied, tubular sori. *IMA Fungus* 2016;7:309–15. <https://doi.org/10.5598/imafungus.2016.07.02.10>
45. Liu B, He Y, Wang Y *et al.* Structure of active human telomerase with telomere shelterin protein TPP1. *Nature* 2022;604:578–83. <https://doi.org/10.1038/s41586-022-04582-8>
46. Sekne Z, Ghanim GE, van Roon AM *et al.* Structural basis of human telomerase recruitment by TPP1-POT1. *Science* 2022;375:1173–6. <https://doi.org/10.1126/science.abn6840>
47. Collart MA, Oliviero S. Preparation of yeast RNA. *Curr Protoc Mol Biol* 2001;23:13.12.1–13.12.5. <https://doi.org/10.1002/0471142727.mb1312s23>
48. Li H. Minimap and miniasm: fast mapping and *de novo* assembly for noisy long sequences. *Bioinformatics* 2016;32:2103–10. <https://doi.org/10.1093/bioinformatics/btw152>
49. Kolmogorov M, Yuan J, Lin Y *et al.* Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–6. <https://doi.org/10.1038/s41587-019-0072-8>
50. Bankevich A, Nurk S, Antipov D *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–77. <https://doi.org/10.1089/cmb.2012.0021>
51. Vaser R, Sović I, Nagarajan N *et al.* Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res* 2017;27:737–46. <https://doi.org/10.1101/gr.214270.116>
52. Walker BJ, Abeel T, Shea T *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 2014;9:e112963. <https://doi.org/10.1371/journal.pone.0112963>
53. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* 2011;27:764–70. <https://doi.org/10.1093/bioinformatics/btr011>

54. Wickham H. ggplot2. *Wiley Interdiscip Rev Comput Stat* 2011;3:180–5. <https://doi.org/10.1002/wics.147>
55. Frith MC. A new repeat-masking method enables specific detection of homologous sequences. *Nucleic Acids Res* 2011;39:e23. <https://doi.org/10.1093/nar/gkq1212>
56. Durbin R, Eddy SR, Krogh A *et al.* *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, 1st ed.. Cambridge: Cambridge University Press, 1998. <https://doi.org/10.1017/CBO9780511790492>
57. Nánási M, Vinař T, Brejová B. Probabilistic approaches to alignment with tandem repeats. *Algorithms Mol Biol* 2014;9:3. <https://doi.org/10.1186/1748-7188-9-3>
58. Olson DR, Wheeler TJ. ULTRA-effective labeling of tandem repeats in genomic sequence. *Bioinform Adv* 2024;4:vbae149. <https://doi.org/10.1093/bioadv/vbae149>
59. Chen S, Zhou Y, Chen Y *et al.* fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–90. <https://doi.org/10.1093/bioinformatics/bty560>
60. Altschul SF, Madden TL, Schäffer AA *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–402. <https://doi.org/10.1093/nar/25.17.3389>
61. Manni M, Berkeley MR, Seppey M *et al.* BUSCO: assessing genomic data quality and beyond. *Curr Protoc* 2021;1:e323. <https://doi.org/10.1002/cpz1.323>
62. McGowan J. BUSCO Phylogenomics. https://github.com/jamiemcg/BUSCO_phylogenomics (7 July 2024, date last accessed).
63. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;5:e9490. <https://doi.org/10.1371/journal.pone.0009490>
64. Yu G, Smith DK, Zhu H *et al.* ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2017;8:28–36. <https://doi.org/10.1111/2041-210X.12628>
65. Kijpornyongpan T, Mondo SJ, Barry K *et al.* Broad genomic sampling reveals a smut pathogenic ancestry of the fungal clade Ustilaginomycotina. *Mol Biol Evol* 2018;35:1840–54. <https://doi.org/10.1093/molbev/msy072>
66. Toome M, Kuo A, Henrissat B *et al.* Draft genome sequence of a rare smut relative, *Tilletiaria anomala* UBC 951. *Genome Announc* 2014;2:e00539–14. <https://doi.org/10.1128/genomeA.00539-14>
67. UniProt Consortium. UniProt: the Universal Protein knowledgebase in 2023. *Nucleic Acids Res* 2023;51:D523–31. <https://doi.org/10.1093/nar/gkac1052>
68. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol* 2011;7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
69. Peška V, Fajkus P, Fojtová M *et al.* Characterisation of an unusual telomere motif (TTTTTAGGG)n in the plant *Cestrum elegans* (Solanaceae), a species with a large genome. *Plant J* 2015;82:644–54. <https://doi.org/10.1111/tpj.12839>
70. Peška V, Sitová Z, Fajkus P *et al.* BAL31-NGS approach for identification of telomeres *de novo* in large genomes. *Methods* 2017;114:16–27. <https://doi.org/10.1016/j.ymeth.2016.08.017>
71. Zylstra A, Hadj-Moussa H, Horkai D *et al.* Senescence in yeast is associated with amplified linear fragments of chromosome XII rather than ribosomal DNA circle accumulation. *PLoS Biol* 2023;21:e3002250. <https://doi.org/10.1371/journal.pbio.3002250>
72. Stanke M, Schöffmann O, Morgenstern B *et al.* Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 2006;7:62. <https://doi.org/10.1186/1471-2105-7-62>
73. Yu EY, Kojic M, Holloman WK *et al.* Brh2 and Rad51 promote telomere maintenance in *Ustilago maydis*, a new model system of DNA repair proteins at telomeres. *DNA Repair* 2013;12:472–9. <https://doi.org/10.1016/j.dnarep.2013.04.027>
74. Yu EY, Hsu M, Holloman WK *et al.* Contributions of recombination and repair proteins to telomere maintenance in telomerase-positive and negative *Ustilago maydis*. *Mol Microbiol* 2018;107:81–93. <https://doi.org/10.1111/mmi.13866>
75. Cesare AJ, Griffith JD. Telomeric DNA in ALT cells is characterized by free telomeric circles and heterogeneous t-loops. *Mol Cell Biol* 2004;24:9948–57. <https://doi.org/10.1128/MCB.24.22.9948-9957.2004>
76. Feretzaki M, Pospíšilová M, Valador Fernandes R *et al.* RAD51-dependent recruitment of TERRA lncRNA to telomeres through R-loops. *Nature* 2020;587:303–8. <https://doi.org/10.1038/s41586-020-2815-6>
77. Luke B, Panza A, Redon S *et al.* The Rat1p 5' to 3' exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in *Saccharomyces cerevisiae*. *Mol Cell* 2008;32:465–77. <https://doi.org/10.1016/j.molcel.2008.10.019>
78. Yu EY, Pérez-Martín J, Holloman WK *et al.* Mre11 and Blm-dependent formation of ALT-like telomeres in Ku-deficient *Ustilago maydis*. *PLoS Genet* 2015;11:e1005570. <https://doi.org/10.1371/journal.pgen.1005570>
79. Yu EY, Zahid SS, Ganduri S *et al.* Structurally distinct telomere-binding proteins in *Ustilago maydis* execute non-overlapping functions in telomere replication, recombination, and protection. *Commun Biol* 2020;3:777. <https://doi.org/10.1038/s42003-020-01505-z>
80. Zahid S, Aloe S, Sutherland JH *et al.* *Ustilago maydis* telomere protein Pot1 harbors an extra N-terminal OB fold and regulates homology-directed DNA repair factors in a dichotomous and context-dependent manner. *PLoS Genet* 2022;18:e1010182. <https://doi.org/10.1371/journal.pgen.1010182>
81. Syed S, Aloe S, Sutherland JH *et al.* *Ustilago maydis* Trf2 ensures genome stability by antagonizing Blm-mediated telomere recombination: fine-tuning DNA repair factor activity at telomeres through opposing regulations. *PLoS Genet* 2024;20:e1011515. <https://doi.org/10.1371/journal.pgen.1011515>
82. Steinberg-Neifach O, Lue NF. Telomere DNA recognition in *Saccharomycotina* yeast: potential lessons for the co-evolution of ssDNA and dsDNA-binding proteins and their target sites. *Front Genet* 2015;6:162. <https://doi.org/10.3389/fgene.2015.00162>
83. Sepšiová R, Nečasová I, Willcox S *et al.* Evolution of telomeres in *Schizosaccharomyces pombe* and its possible relationship to the diversification of telomere binding proteins. *PLoS One* 2016;11:e0154225. <https://doi.org/10.1371/journal.pone.0154225>
84. Lue NF. Duplex telomere-binding proteins in fungi with canonical telomere repeats: new lessons in the rapid evolution of telomere proteins. *Front Genet* 2021;12:638790. <https://doi.org/10.3389/fgene.2021.638790>
85. Višacká K, Hofr C, Willcox S *et al.* Synergism of the two Myb domains of Tay1 protein results in high affinity binding to telomeres. *J Biol Chem* 2012;287:32206–15. <https://doi.org/10.1074/jbc.M112.385591>
86. Song H, Kim S, Lim DS *et al.* Robust binding capability and occasional gene loss of telomere-binding proteins underlying telomere evolution in Nematoda. *Genome Biol Evol* 2025;17:evaf085. <https://doi.org/10.1093/gbe/evaf085>
87. Chen C, Gu P, Wu J *et al.* Structural insights into POT1-TPP1 interaction and POT1 C-terminal mutations in human cancer. *Nat Commun* 2017;8:14929. <https://doi.org/10.1038/ncomms14929>
88. Rice C, Shastrula PK, Kossenkova AV *et al.* Structural and functional analysis of the human POT1-TPP1 telomeric complex. *Nat Commun* 2017;8:14928. <https://doi.org/10.1038/ncomms14928>
89. Takai KK, Kibe T, Donigian JR *et al.* Telomere protection by TPP1/POT1 requires tethering to TIN2. *Mol Cell* 2011;44:647–59. <https://doi.org/10.1016/j.molcel.2011.08.043>

90. Zaug AJ, Podell ER, Nandakumar J *et al.* Functional interaction between telomere protein TPP1 and telomerase. *Genes Dev* 2010;24:613–22. <https://doi.org/10.1101/gad.1881810>
91. Sandhu R, Sharma M, Wei D *et al.* The structurally conserved TELR region on shelterin protein TPP1 is essential for telomerase processivity but not recruitment. *Proc Natl Acad Sci USA* 2021;118:e2024889118. <https://doi.org/10.1073/pnas.2024889118>
92. Tejera AM, Stagno d'Alcontres M, Thanasoula M *et al.* TPP1 is required for TERT recruitment, telomere elongation during nuclear reprogramming, and normal skin development in mice. *Dev Cell* 2010;18:775–89. <https://doi.org/10.1016/j.devcel.2010.03.011>
93. Deng Z, Kim ET, Vladimirova O *et al.* HSV-1 remodels host telomeres to facilitate viral replication. *Cell Rep* 2014;9:2263–78. <https://doi.org/10.1016/j.celrep.2014.11.019>
94. Saint-Leandre B, Levine MT. The telomere paradox: stable genome preservation with rapidly evolving proteins. *Trends Genet* 2020;36:232–42. <https://doi.org/10.1016/j.tig.2020.01.007>
95. Jaiswal RK, Garibo Domingo T, Grunchev H *et al.* Subtelomeric elements provide stability to short telomeres in telomerase-negative cells of the budding yeast *Naumovozyma castellii*. *Curr Genet* 2025;71:19. <https://doi.org/10.1007/s00294-025-01325-w>
96. Teng SC, Zakian VA. Telomere-telomere recombination is an efficient bypass pathway for telomere maintenance in *Saccharomyces cerevisiae*. *Mol Cell Biol* 1999;19:8083–93. <https://doi.org/10.1128/MCB.19.12.8083>
97. Wang H, Gilley D, Blackburn EH. A novel specificity for the primer-template pairing requirement in *Tetrahymena telomerase*. *EMBO J* 1998;17:1152–60. <https://doi.org/10.1093/emboj/17.4.1152>
98. Lue NF, Xia J. Species-specific and sequence-specific recognition of the dG-rich strand of telomeres by yeast telomerase. *Nucleic Acids Res* 1998;26:1495–502. <https://doi.org/10.1093/nar/26.6.1495>
99. Morin GB, Cech TR. Mitochondrial telomeres: surprising diversity of repeated telomeric DNA sequences among six species of *Tetrahymena*. *Cell* 1988;52:367–74. [https://doi.org/10.1016/s0092-8674\(88\)80029-1](https://doi.org/10.1016/s0092-8674(88)80029-1)
100. Bryan TM, Englezou A, Gupta J *et al.* Telomere elongation in immortal human cells without detectable telomerase activity. *EMBO J* 1995;14:4240–8. <https://doi.org/10.1002/j.1460-2075.1995.tb00098.x>
101. Kazda A, Zellinger B, Rössler M *et al.* Chromosome end protection by blunt-ended telomeres. *Genes Dev* 2012;26:1703–13. <https://doi.org/10.1101/gad.194944.112>
102. Valuchová S, Fulneček J, Prokop Z *et al.* Protection of *Arabidopsis* blunt-ended telomeres is mediated by a physical association with the Ku heterodimer. *Plant Cell* 2017;29:1533–45. <https://doi.org/10.1105/tpc.17.00064>
103. Bautista-España D, Anastacio-Marcelino E, Horta-Valerdi G *et al.* The telomerase reverse transcriptase subunit from the dimorphic fungus *Ustilago maydis*. *PLoS One* 2014;9:e109981. <https://doi.org/10.1371/journal.pone.0109981>
104. Sánchez-Alonso P, Guzmán P. Organization of chromosome ends in *Ustilago maydis*. RecQ-like helicase motifs at telomeric regions. *Genetics* 1998;148:1043–54. <https://doi.org/10.1093/genetics/148.3.1043>
105. Louis EJ, Haber JE. The structure and evolution of subtelomeric Y' repeats in *Saccharomyces cerevisiae*. *Genetics* 1992;131:559–74. <https://doi.org/10.1093/genetics/131.3.559>
106. Brown CA, Murray AW, Verstrepen KJ. Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr Biol* 2010;20:895–903. <https://doi.org/10.1016/j.cub.2010.04.027>