PNAS RESEARCH ARTICLE | BIOCHEMISTRY OPEN ACCESS

Biogenesis of telomerase RNA from a protein-coding mRNA precursor

Dhenugen Logeswaran[a,1], Yang Li[a,1], Khadiza Akhter[a], Joshua D. Podlevsky[a,2], Tamara L. Olson[a], Katherine Forsberg[a], and Julian J.-L. Chen[a,3]

Edited by Juli Feigon, University of California, Los Angeles, Los Angeles, CA; received March 16, 2022; accepted August 29, 2022

Telomerase is a eukaryotic ribonucleoprotein (RNP) enzyme that adds DNA repeats onto chromosome ends to maintain genomic stability and confer cellular immortality in cancer and stem cells. The telomerase RNA (TER) component is essential for telomerase catalytic activity and provides the template for telomeric DNA synthesis. The biogenesis of TERs is extremely divergent across eukaryotic kingdoms, employing distinct types of transcription machinery and processing pathways. In ciliates and plants, TERs are transcribed by RNA polymerase III (Pol III), while animal and ascomycete fungal TERs are transcribed by RNA Pol II and share biogenesis pathways with small nucleolar RNA (snoRNA) and small nuclear RNA (snRNA), respectively. Here, we report an unprecedented messenger RNA (mRNA)-derived biogenesis pathway for the 1,291 nucleotide TER from the basidiomycete fungus *Ustilago maydis*. The *U. maydis* TER (*Um*TER) contains a 5′-monophosphate, distinct from the 5′ 2,2,7-trimethylguanosine (TMG) cap common to animal and ascomycete fungal TERs. The mature *Um*TER is processed from the 3′-untranslated region (3′-UTR) of a larger RNA precursor that possesses characteristics of mRNA including a 5′ 7-methyl-guanosine (m^7G) cap, alternative splicing of introns, and a poly(A) tail. Moreover, this mRNA transcript encodes a protein called Early meiotic induction protein 1 (Emi1) that is conserved across dikaryotic fungi. A recombinant *Um*TER precursor expressed from an mRNA promoter is processed correctly to yield mature *Um*TER, confirming an mRNA-processing pathway for producing TER. Our findings expand the plethora of TER biogenesis mechanisms and demonstrate a pathway for producing a functional long noncoding RNA from a protein-coding mRNA precursor.

basidiomycete | ncRNA | RNA processing | telomere

Significance

Many noncoding RNA molecules are indispensable components of macromolecular complexes that govern critical cellular processes. Telomerase RNA is a long noncoding RNA essential for maintaining chromosome stability and cellular immortality in eukaryotes. In this study, we discover a basidiomycete fungal telomerase RNA that is processed from a messenger RNA (mRNA) transcript that encodes a conserved protein. Our finding demonstrates an unprecedented biogenesis pathway for generating a functional long noncoding RNA from a protein-coding mRNA precursor.

Many vital cellular processes that govern genetic information transfer from DNA to protein rely on a vast variety of RNA molecules. These RNAs include tens of thousands of messenger RNAs (mRNAs) that encode proteins as well as numerous noncoding RNAs (ncRNAs) that are not translated into proteins yet form crucial ribonucleoprotein complexes (1). The biogenesis of these diverse ncRNAs, such as ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small nuclear RNAs (snRNAs), and small nucleolar RNAs (snoRNAs), requires distinct RNA polymerases and diverse processing mechanisms. Among the diverse types of ncRNAs, long noncoding RNAs (lncRNAs) are generally defined as ncRNA molecules larger than 200 nucleotides and mainly regulate the expression of protein-coding genes at transcriptional, RNA processing, translational, and posttranslational levels (2).

Telomerase RNA (TER) is a distinct class of lncRNA that functions as an integral component of the telomerase ribonucleoprotein enzyme. TER provides a short template that is reiteratively used by the telomerase reverse transcriptase (TERT) catalytic component to perform de novo synthesis of telomeric DNA repeats onto telomeres at chromosome ends. This telomere replenishment is crucial for preserving genomic stability and maintaining cellular reproductive capacity (3). While the TERT proteins show broad evolutionary conservation (4), TERs are extremely divergent in sequence, length, structure, and biosynthesis pathway (5). The computational search for TERs in distinct groups of eukaryotes is a daunting challenge due to the lack of sequence conservation among TERs. Although advanced bioinformatics strategies have made significant progress for specific lineages of eukaryotes (6–8), many TERs await to be uncovered in some eukaryotic clades that are evolutionarily distant. For these eukaryotes, biochemical purification of the telomerase holoenzyme from cell lysates remains most effective for TER identification (9, 10).

Despite the poor conservation in primary sequence, TERs show secondary structure conservation in two domains that are required for telomerase activity (5). The 5′ proximal template-pseudoknot (T/PK) domain contains an RNA template that specifies the telomeric DNA repeat sequence to be synthesized, followed by a conserved pseudoknot structure that is indispensable for telomerase catalysis (11). A distal helical domain

Author affiliations: [a]School of Molecular Sciences, Arizona State University, Tempe, AZ 85281

Author contributions: D.L., Y.L., J.D.P., and J.J.-L.C. designed research; D.L., Y.L., K.A., and J.D.P. performed research; K.A., J.D.P., T.L.O., and K.F. contributed new reagents/analytic tools; D.L., Y.L., and J.J.-L.C. analyzed data; and D.L., Y.L., and J.J.-L.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]D.L. and Y.L. contributed equally to this work.

[2]Present address: Molecular and Microbiology Department, Sandia National Laboratories, Albuquerque, NM 87123.

[3]To whom correspondence may be addressed. Email: JLChen@asu.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2204636119/-/DCSupplemental.

Published October 5, 2022.

located downstream of the T/PK domain stimulates telomerase activity in the presence of T/PK and can independently bind the TERT protein (9, 12). This distal domain is known as conserved region 4/5 (CR4/5) in most animal taxa and filamentous fungi (6, 9, 13), eCR4/5 in other eukaryotic clades (14, 15), and helix IV in ciliates (16). The two-domain requirement for telomerase catalytic activity is universal and likely emerged during early evolution of eukaryotes (14).

The biogenesis of TER has evolutionarily diversified by employing different RNA polymerases along distinct eukaryotic lineages. Flagellate (781-993 nucleotides [nt]), fungal (900 to 2,400 nt), and animal TERs (239 to 609 nt) are transcribed by RNA polymerase II (Pol II) (5) and predominantly contain a 5′ 2,2,7-trimethylguanosine (TMG) cap (17, 18). However, ciliate (140 to 210 nt) and plant (231 to 350 nt) TERs are smaller and transcribed by RNA polymerase III (Pol III) with a characteristic 5′ triphosphate and 3′ uridine-tail (19–22). The intriguing diversity of TER is further exemplified by the presence of lineage-specific accessory proteins that bind the 3′ region of TER to regulate 3′-end maturation. These protein repertoires mainly protect the TER 3′ end from exonucleolytic degradation. The 3′ proximal region of animal TERs contains a box H/ACA snoRNA-like domain bound by a heterotetrameric complex of dyskerin, NHP2, NOP10, and GAR1 proteins (6, 23, 24). Additionally in mammalian TERs, the terminal stem-loop of this domain contains a Cajal body (CAB) box bound by telomerase Cajal body protein 1 (TCAB1) for Cajal body localization (25, 26). In contrast, fungal TERs from budding and fission yeasts employ the snRNA biogenesis pathway for TER maturation by harboring conserved Sm binding sites at 3′ regions for binding the heptameric Sm ring complex (27, 28).

The maturation of TER relies on a fine balance between RNA processing and decay. The 5′ end of fungal TER is typically protected by a TMG cap, and the 3′ end undergoes complex processing involving loading and unloading of protein complexes. In *Saccharomyces cerevisiae* TER (TLC1), the 3′-end maturation depends on the Nrd1-Nab3-Sen1 complex–mediated transcription termination and further stabilization by the Sm ring complex (29). Despite sharing the Sm binding site with *S. cerevisiae*, the 3′ end of *Schizosaccharomces pombe* TER (TER1) is generated by the first transesterification step of intron splicing without the subsequent religation of the exons (30). Filamentous fungal TERs undergo a similar 3′-end maturation by the spliceosomal cleavage of a terminal intron but use a distinct 5′-splice site sequence (31, 32).

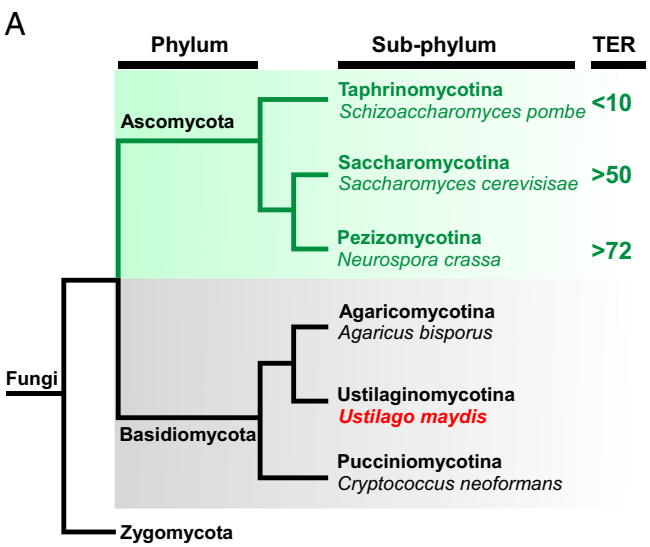
Here, we report a unique mRNA-derived lncRNA biogenesis pathway for the TER identified in the model fungus *Ustilago maydis*. The mature form of *U. maydis* TER (*Um*TER) possesses a 5′-monophosphate and is processed from an mRNA precursor that contains a 5′ 7-methylguanosine (m^7G) cap, a long 3′ untranslated region (3′ UTR), and a 3′ poly(A) tail. Moreover, this mRNA precursor undergoes alternative splicing and contains an open reading frame (ORF) that encodes a conserved protein. The processing of mature *Um*TER from the 3′-UTR of a protein-coding mRNA precursor is an unprecedented mechanism for TER biogenesis.

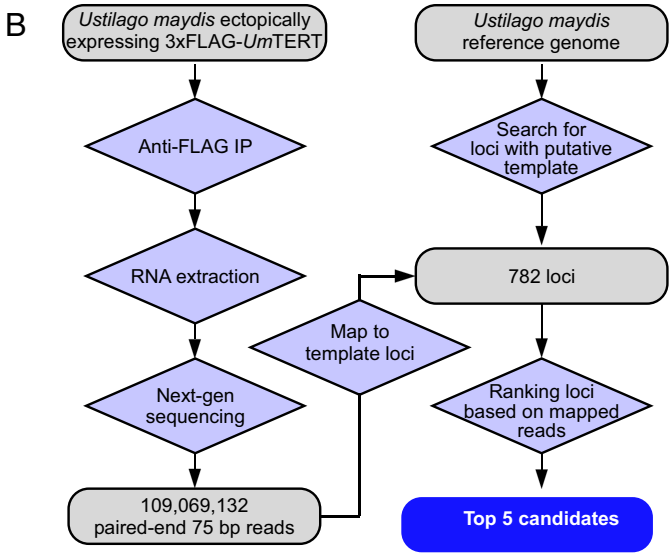
Results

Identification and Validation of *Um*TER. Within the fungal kingdom, TER has been identified and extensively studied in Ascomycota but not in the sister phylum Basidiomycota (Fig. 1*A*). Past studies of telomerase using ascomycete fungal models such as *S. cerevisiae* (33), *S. pombe* (30), and *Neurospora crassa* (9) have led to many important findings in telomerase biogenesis mechanisms. To further explore TER biogenesis across fungal phyla, we set out to identify basidiomycete TER from the corn smut fungus *U. maydis* (Fig. 1*A*). To that end, we generated a recombinant *U. maydis* strain expressing a 3xFLAG-tagged *U. maydis* TERT (3xFLAG-*Um*TERT) protein for affinity purification of *U. maydis* telomerase (*SI Appendix*, Fig. S1 *A* and *B*). The recombinant *U. maydis* telomerase holoenzyme purified by anti-FLAG immunoprecipitation (IP) showed significant telomerase activity detected by the telomere repeat amplification protocol (TRAP) assay (*SI Appendix*, Fig. S1*C*). The RNA molecules copurified with the active telomerase holoenzyme were extracted and analyzed by Illumina next-generation sequencing, which generated over 109 million short reads. By employing a proven bioinformatics strategy (9, 34), we searched the *U. maydis* genome for TER candidates and identified 782 genomic loci that contain putative template sequences for synthesizing the telomeric DNA repeats (TTAGGG)n (Fig. 1*B*). Mapping the Illumina sequencing reads onto these loci identified TER candidates that were then ranked by read coverage (Fig. 1*C*). The top candidate sequence with the highest read coverage was the 18S rRNA, presumably due to its abundance in the cell (Fig. 1*C*). The second TER candidate was an unannotated RNA with a conserved homolog found in the closely related species *Ustilago bromivora*. Notably, the *U. bromivora* homolog also contained a conserved TER template sequence, 5′-UAACCCUAA-3′ (Fig. 1*C*). The remaining candidates lacked template-containing homologs in *U. bromivora* and were not pursued any further (Fig. 1*C*). Northern blot analysis verified the presence of the second TER candidate in *U. maydis* with a size of ~1,300 nt (Fig. 1*D*), which is consistent with the length of the locus covered by the sequencing reads (Fig. 1*E*).

We then mapped the 5′ end of this TER candidate by a cap-independent 5′ Rapid Amplification of complementary DNA (cDNA) Ends (5′-RACE) procedure (*SI Appendix*, Fig. S2*A*-C). For mapping the 3′ end by 3′-RACE, the RNA was first added with a guanosine/inosine (G/I) tail at the 3′ end, followed by an RT reaction using an oligo-dC reverse primer for cDNA synthesis and nested PCR for cDNA amplification (*SI Appendix*, Fig. S2 *D–F*). The 5′-RACE generated a major cDNA product, and the 5′ end of the TER candidate was determined by sequencing cloned cDNA products (*SI Appendix*, Fig. S2 *B*, lane 1 and *SI Appendix*, Fig. S2*C*). The 3′ ends of this TER candidate appeared to be slightly heterogeneous, while the majority of the 3′-RACE cDNA products revealed a 3′ end at position 1291 (*SI Appendix*, Fig. S2 *E*, lanes 1 and 2, and *SI Appendix*, Fig. S2*F*). More importantly, the RACE analyses performed on the RNA copurified with the 3xFLAG-*Um*TERT also indicated the same 5′ and 3′ ends (*SI Appendix*, Fig. S2 *G–K*). Thus, our 5′- and 3′-RACE results concluded that *Um*TER candidate #2 is a *Um*TERT-bound RNA with a size of 1,291 nt (Fig. 1*E*) which is consistent with the northern blot result (Fig. 1*D*).

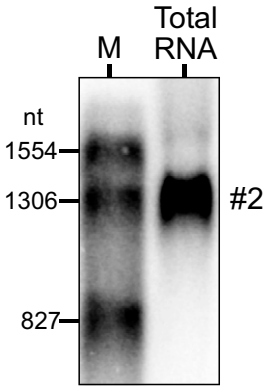
The homologs of the *Um*TER candidate #2 RNA were found by a bioinformatic search using the Basic Local Alignment Search Tool (BLAST) across three fungal taxonomy orders, namely, Ustilaginales, Urocystidales, and Violaceomycetales, and located between two protein-coding genes with a conserved gene synteny (*SI Appendix*, Fig. S3). To determine if *Um*TER candidate #2 is truly a telomerase component, we generated a *Um*TER gene knockout (Δ*Umter*) *U. maydis* strain with the 1,291-bp *Um*TER-encoding genomic region replaced by a hygromycin resistance gene cassette (*SI Appendix*, Fig. S4*A*) (35). Hygromycin-resistant





C

Rank	Coverage (per base)	No. of reads	Locus length	Putative Template (length)	Chromosome (Strand) Position	Annotated gene	Homolog with template
#1	1,211,816	11,485,950	716	ACCCUAACC (9)	4 (+) 563,363 - 564,078	18S rRNA	+
#2	585	9462	1200	UAACCCUAA (9)	8 (+) 126165 - 127364	—	+
#3	700	5355	322	ACCCUAAC (8)	1 (-) 1,726,838 - 1,727,159	5S rRNA	—
#4	152	1910	951	CCUAACCC (8)	10 (+) 464,873 - 465,823	—	—
#5	99	1754	1329	CCUAACCC (8)	9 (+) 400,696 - 402,024	Hypothetical protein	—



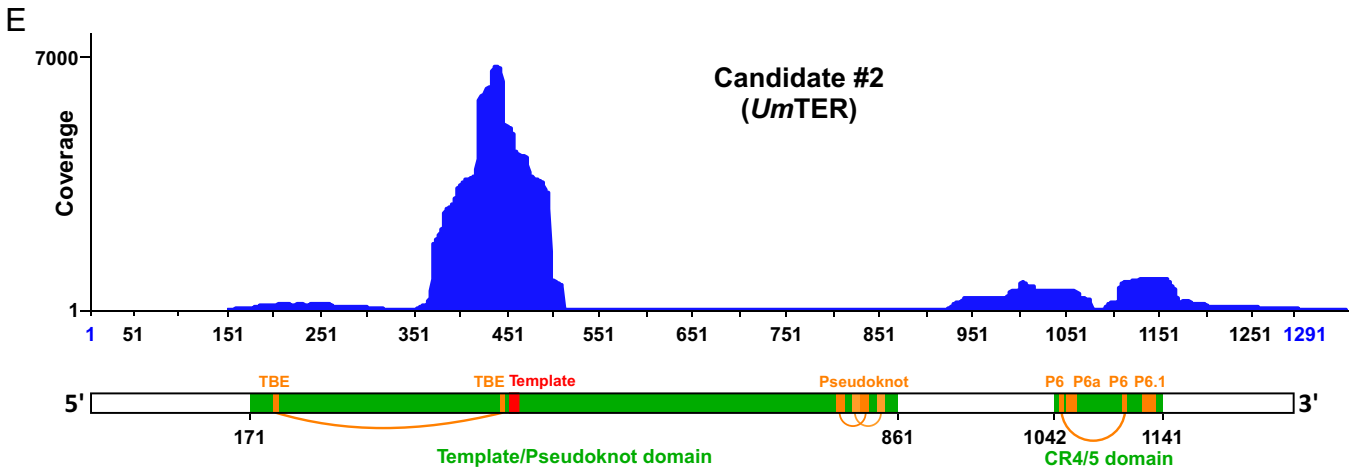


Fig. 1. Identification of *Um*TER. (*A*) Fungal TERs identified in phylum Ascomycota. A simplified tree of three fungal phyla, namely, Ascomycota, Basidiomycota, and Zygomycota, is shown. The numbers of TERs identified in each subphyla of the Ascomycota are shown. The basidiomycete model fugus *U. maydis* used in this study is indicated in red. Branch lengths are not proportional to evolutionary distance. (*B*) Screening strategy for the identification of *Um*TER candidates. RNA molecules bound to 3xFLAG-*Um*TERT after anti-FLAG IP were extracted and analyzed by Illumina next-generation sequencing. Sequencing reads were mapped to *U. maydis* genomic loci containing putative template sequences, and the top five loci with the highest numbers of mapped reads were selected for further analysis. (*C*) Sequence analysis and characterization of five top-ranking candidate loci. (*D*) Northern blot analysis of candidate #2 RNA. Total RNA (10 μg) of *U. maydis* was analyzed by Northern blot and probed with a radiolabeled riboprobe targeting candidate #2 sequence. RNA size markers (M) are shown to the *Left*. (*E*) Coverage map of Illumina short reads on candidate #2 (*Um*TER) locus. Positions of *Um*TER structural elements (template, TBE, pseudsoknot, P6, and P6.1) and essential regions (green) are shown below the coverage map.

Δ*Umter* clones were identified and verified by Northern blot analysis (*SI Appendix*, Fig. S4*B*). Two independent Δ*Umter* clones were analyzed by Terminal restriction fragment length (TRF) assay and showed progressive telomere shortening over 200 generations, while the wild-type (WT) strain was able to maintain telomere length during successive passages (*SI Appendix*, Fig. S4 *C* and *D*). Thus, *Um*TER candidate #2 is essential for telomere maintenance and indeed the authentic TER component of *U. maydis* telomerase.

***Um*TER Contains Two Structural Domains Essential for Telomerase Function.** The core telomerase enzyme consists of the catalytic TERT protein and the TER subunit that provides the template for DNA synthesis. As a prelude to performing detailed functional dissections of *Um*TER, we assembled *in vitro* transcribed *Um*TER with the *in vitro* synthesized recombinant *Um*TERT protein and successfully reconstituted telomerase activity (*SI Appendix*, Fig. S5*A*). *In vitro* reconstituted *U. maydis* telomerase was analyzed by direct primer extension assay and showed activity of template-directed DNA synthesis (*SI Appendix*, Fig. S5*A*, *Bottom*, lanes 1 and 6). The *in vitro* reconstituted *U. maydis* telomerase was not processive and capable of adding only a single repeat (*SI Appendix*, Fig. S5*A*, *Bottom*). This is in contrast to the ladder of bands observed using the TRAP activity assay with the telomerase holoenzyme from *U. maydis* cell lysate (*SI Appendix*, Fig. S1*C*, lane 1). We suspect that additional factors in the *U. maydis* telomerase holoenzyme may be responsible for the different telomerase processivity observed (36).

TERs from all major eukaryotic groups contain two structural domains essential or crucial for telomerase catalysis (14). To identify structural domains within *Um*TER necessary for reconstituting telomerase activity, we carried out a series of truncation analyses on *Um*TER (*SI Appendix*, Fig. S5 *B* and *C*) and identified two minimal *Um*TER fragments, nt 171 to 861 and nt 1,042 to 1,141, which can assemble in trans with *Um*TERT to reconstitute activity (*SI Appendix*, Fig. S5*C*). It was noted that the longer 5F2 RNA fragment reconstituted a lower activity than the 5F3 RNA fragment (*SI Appendix*, Fig. S5*C*, *Bottom Right*, lanes 5 and 6), which could be due to a suboptimal folding of fragment 5F2 with an extended 5′ region compared to fragment 5F3 (*SI Appendix*, Fig. S5*C*, *Left*). The two minimal *Um*TER domains that bind *Um*TERT independently were consistent with the two regions covered by most sequencing reads from the Illumina RNA sequencing analysis of the RNA extracted from the purified *U. maydis* telomerase holoenzyme (Fig. 1*E*). More importantly, based on a phylogenetic comparative analysis of 18 Basidiomycota fungal TER homolog sequences (*SI Appendix*, Fig. S6 *A* and *B*), these two structural domains of *Um*TER fold into TER-specific structural domains, i.e., the T/PK and the template-distal CR4/5, which are commonly present in both animal (6) and Ascomycota fungal (9) TERs (Fig. 2*A*). Deletion analysis on these two structural domains further minimized the T/PK and CR4/5 core domains required for telomerase activity (Fig. 2 *B*–*D*), which include functionally crucial elements, such as the putative template boundary element (TBE) in the T/PK domain and the P6.1 stem-loop in the CR4/5 domain (Fig. 2A) (6, 9). An activity assay of telomerase reconstituted from the truncated T/PK and CR4/5 fragments identified the minimal fragments, namely, T/PK-Δ4 with three regions deleted and CR4/5(1046-1135)-D3 with most of the P6b stem deleted (Fig. 2*B*, lane 7, Fig. 2C, lane 5, and Fig. 2*D*, lane 3). Notably, telomerase reconstituted from one of the truncated T/PK fragments, namely, T/PK-Δ2 with only the region 723 to 784 deleted, showed significantly higher activity than T/PK-Δ4 (Fig. 2*B*, lanes 5 and 7), which was likely due to

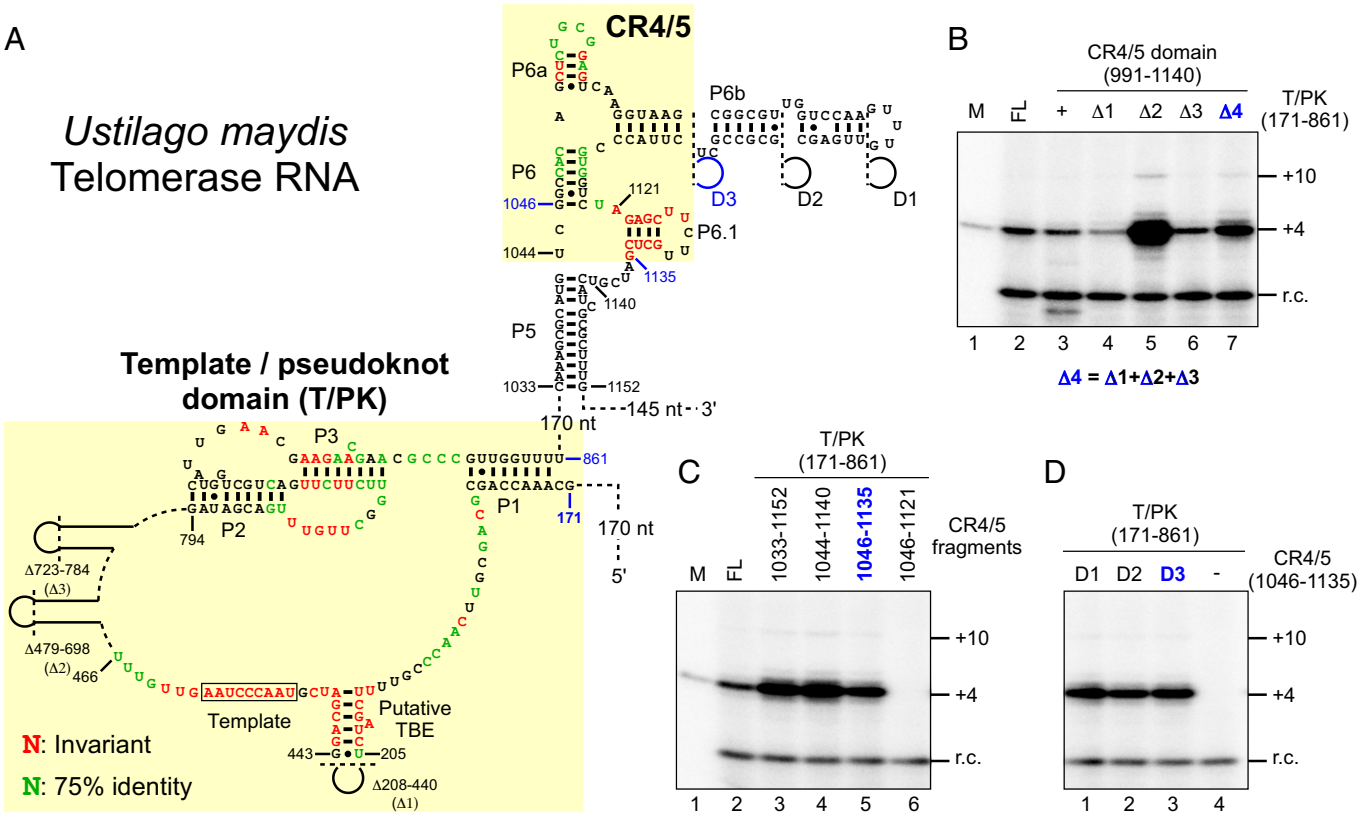


Fig. 2. Two minimal functional core domains of *Um*TER. (*A*) Secondary structure of the T/PK and the CR4/5 domains. The T/PK domain includes stems P1, P2, and P3; TBE; and template. The CR4/5 domain includes the P6, P6a, P6b, and P6.1 stems. Invariant nucleotides (red) or nucleotides with ≥75% identity (green) are indicated. The regions without a secondary structure determined are indicated by lines. Nucleotide positions corresponding to the 5′ and 3′ ends of minimal T/PK and CR4/5 fragments are shown in blue. Regions dispensable for in vitro telomerase activity are indicated (delta symbol). (*B*–*D*) Activity assay of telomerase reconstitution with *Um*TER T/PK and CR4/5 fragments. The length of the T/PK and CR4/5 fragments analyzed are shown to the *Top Right* of the gels. Truncations are indicated with a delta symbol (*B*), as length ranges (*C*), or with a D (*D*) with reference to annotations in *A*. Truncated stem-loops (Δ and D) are capped with 5′-GAAA-3′ tetraloop. The numbers of nucleotides (+4 or +10) added to the substrate primer are indicated to the *Right* of the gel. All two fragment activity assays are compared to that of full-length (FL) *Um*TER. A radiolabeled oligo identical to the +4 band is added as a marker (M). A ^{32}P end-labeled oligonucleotide is added to each reaction as the recovery control (r.c.) prior to ethanol precipitation of DNA products. The minimal fragments tested are indicated in blue.

higher enzyme turnover. However, the underlying mechanism for such higher activity from a truncated T/PK domain remains to be explored. Taken together, our results identified the minimal regions required for enzymatic activity, which is consistent with the structural domains essential for telomerase function and conserved between animal and fungal TERs (6, 9).

***Um*TER Contains a 5′ Monophosphate.** Animal and Ascomycota yeast TERs are transcribed by RNA Pol II followed by hypermethylation of the 5' m^7G cap by the TGS1 hypermethylase to generate a TMG cap (17, 18). The 5′ end of *Trypanosoma* TER is processed by transsplicing with a unique cap chemistry (37). To determine the 5′-end structure of *Um*TER, we first treated *U. maydis* total RNA with a decapping enzyme, namely, RNA 5′ pyrophosphohydrolase (RppH) (38), that removes the 5′-cap from the RNA Pol II transcripts leaving a 5′ monophosphate group that can then be ligated to an RNA adapter (Fig. 3*A*). The adapter-ligated RNA transcripts were then analyzed by a standard RNA ligase-mediated RACE (RLM-RACE) procedure using *Um*TER-specific reverse primers to amplify the 5′ cDNA of the *Um*TER transcripts (Fig. 3*A*). Intriguingly, the RppH decapping and RLM-RACE analysis detected multiple *Um*TER-related transcripts with various lengths. It appeared that the longer *Um*TER transcripts contain a 5′ cap and require RppH treatment

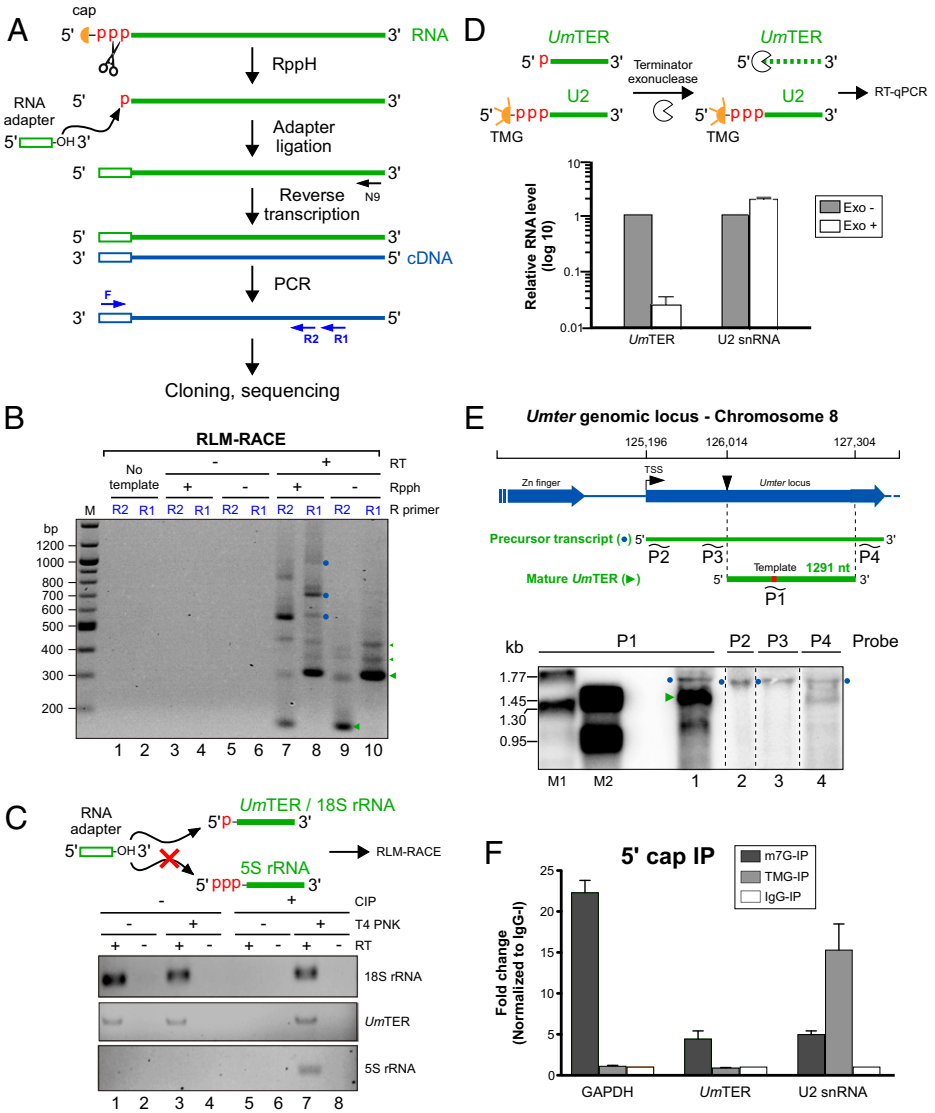


Fig. 3. The 5′-end structure of mature and precursor *Um*TER transcripts. (*A*) Schematic of RLM-RACE. Total RNA was treated with the decapping enzyme RppH to remove the 5′ cap and ligated to an RNA adapter, followed by RT-PCR to amplify the adapter-ligated *Um*TER transcripts. RACE cDNA products were cloned and sequenced. (*B*) Gel electrophoresis analysis of RLM-RACE PCR products. Bands that correspond to RppH-dependent (blue dots) and RppH-independent (green triangles) RACE cDNA products are indicated. (*C*) Detection of 5′ monophosphate in *Um*TER. The schematic shows that the RNA adapter for RLM-RACE can be ligated only to the RNA with a 5′-monophosphate (*Um*TER or 18S rRNA) but not to the 5S rRNA with a 5′ triphosphate. Gel electrophoresis analysis of RLM-RACE cDNA products distinguishes RNAs with 5′ monophosphate or triphosphate after treatments with CIP to remove all 5′ phosphates and/or T4 PNK to add 5′ monophosphate. (*D*) *Um*TER is sensitive to terminator exonuclease (Exo) that degrades RNA with 5′-monophosphate. U2 snRNA contains a 5′- TMG cap and is resistant to terminator exonuclease. *Um*TER and U2 snRNA treated with (+) or without (−) Exo are quantitated by RT-qPCR and normalized by the level of GAPDH mRNA in each total RNA sample. (*E*) Northern blot analysis of distinct *Um*TER transcripts. (*Top*) The schematic shows the genomic coordinates of the *Umter* gene and transcription products, precursor (blue dots), and mature *Um*TER (green triangle). Target sites of northern blot probes (P1 to P4) detecting specific transcripts are indicated. (*Bottom*) Northern blot of *Um*TER transcripts detected with specific probes (P1 to P4). Northern blot shows the detection of mature *Um*TER (green triangle) by probe P1 (lane 1) and precursor (blue dots) by probes P1-P4 (lanes 1 to 4). Positions of RNA size (Kb) markers are indicated on the *Left*. (*F*) IP analysis of RNAs with 5′ m^7G or TMG cap. RT-qPCR was performed to quantitate specific RNA transcripts enriched by IP using specific antibodies against either 5′ m^7G or TMG cap or IgG as a negative control. Enrichment of each transcript by each antibody is normalized to the RNA level in the IgG-IP sample.

to be detected by RLM-RACE (Fig. 3*B*, compare lanes 8 and 10 with R1 primer). In contrast, the short *Um*TER transcript contains a 5′ monophosphate and can thus be ligated to the RNA adapter for RACE detection with or without RppH treatment (Fig. 3*B*, compare lanes 8 and 10 with R1 primer). Moreover, the sequencing results of the PCR DNA product confirmed that the 5′ end of the short *Um*TER transcript is identical to the 5′ end of the mature *Um*TER determined previously by cap-independent template-switching RACE (*SI Appendix*, Fig. S2 *A*–*C*). We further confirmed the presence of a 5′ monophosphate in mature *Um*TER by treating the RNA with calf intestinal alkaline phosphatase (CIP) to remove 5′ phosphates. Upon CIP treatment, mature *Um*TER and 18S rRNA that also contains 5′-monophosphate cannot be detected by RLM-RACE (Fig. 3*C*, lane 5), which can then be rescued by T4 polynucleotide kinase (PNK) treatment to add back the 5′-monophosphate (Fig. 3*C*, lane 7). The 5S rRNA contained a 5′ triphosphate and was resistant to RLM-RACE (Fig. 3*C*, lanes 1 and 3), unless treated with CIP and T4 PNK sequentially to generate a 5′ monophosphate (Fig. 3*C*, lane 7). Furthermore, we confirmed the presence of a 5′ monophosphate in mature *Um*TER by treating the RNA with Terminator exonuclease that degrades specifically the RNAs with 5′ monophosphate such as rRNA but not the 5′-capped RNAs such as mRNA or snRNAs (Fig. 3*D*, *Top*). The RT-qPCR analysis of the Terminator exonuclease-treated RNA showed significant degradation of *Um*TER but not the U2 snRNA (Fig. 3*D*, *Bottom*). These results collectively confirm the presence of a 5′ monophosphate in the mature *Um*TER.

The Large *Um*TER Precursor Transcripts Contain a 5′-m^{7}G Cap. The presence of a 5′ monophosphate suggests that mature *Um*TER is posttranscriptionally processed at the 5′ end from larger precursors, likely from the 5′-capped *Um*TER transcripts detected by RLM-RACE in the RppH-treated sample (Fig. 3*B*, lanes 7 and 8). As expected, sequencing analysis of these longer cDNA products revealed a distinct 5′ end located further upstream of the 5′ end of mature TER (Fig. 3*E*). To directly detect the longer *Um*TER precursor transcripts, we performed Northern blot analyses with probes targeting specific regions in either the precursor or mature *Um*TER transcripts (Fig. 3*E*, *Top*). The radiolabeled riboprobe P1 detected the template region of *Um*TER, while the other three riboprobes P2, P3, and P4 targeted only the precursor transcripts at three different regions, i.e., near the transcription start site (TSS), upstream of the mature *Um*TER, and downstream of the mature *Um*TER, respectively (Fig. 3*E*, *Top*). The northern blot probed with the P1 riboprobe revealed a strong band of ∼1,300 nt RNA corresponding to the mature TER and a weak band of RNA with higher molecular weight (Fig. 3*E*, *Bottom*, lane 1). All three precursor-specific probes, namely, P2, P3 and P4, detected only the larger RNA, confirming the presence of the larger *Um*TER precursor transcripts (Fig. 3*E*, *Bottom*, lanes 2, 3, and 4). Interestingly, probe P4 targeting the region downstream of mature *Um*TER detected additional faint bands with smaller sizes (Fig. 3*E*, *Bottom*, lane 4), indicating that the 5′-end processing of the *Um*TER precursor likely precedes the 3′-end processing.

To determine the identity of the 5′-cap structure of the larger *Um*TER transcripts, we performed 5′-cap specific RNA IP using anti-m^{7}G, anti-TMG, or immunoglobulin G (IgG) antibodies, and we detected the affinity-purified RNA targets by RT-qPCR with primer sets specific to GAPDH mRNA, *Um*TER, or U2 snRNA. The results showed that the anti-m^{7}G IP significantly enriched the GAPDH mRNA and *Um*TER precursor transcripts over the IgG antibody negative control, supporting the presence of 5′ m^{7}G in the *Um*TER precursor transcripts, while the anti-TMG IP enriched only the 5′ TMG-capped U2 snRNA (Fig. 3*F*). A minor cross-reactivity of anti-m^{7}G IP to the 5′ TMG of U2 snRNA was observed as previously reported (Fig. 3*F*) (39). Our results indicated that the *Um*TER-encompassing long transcripts are 5′-m^{7}G capped.

In addition to the presence of 5′-m^{7}G cap, the promoter regions upstream of the *Um*TER precursor sequences were identified in 17 Ustilaginomycetes species and shared a conserved 5′-ACGCGAA-3′ sequence (*SI Appendix*, Fig. S7). This conserved promoter sequence was previously identified as a top-ranked binding site for the transcription factor Swi4 in a yeast chromatin IP study (40). The Swi4 transcription factor is part of a complex that activates mRNA transcription of multiple protein-coding genes involved in cell cycle regulation (41). Thus, the transcription of the *Um*TER precursor transcript may be regulated developmentally along the life cycle of *U. maydis*.

The *Um*TER Precursor Transcript Undergoes Alternative Intron Splicing. Sequencing analysis of the 5′-RACE DNA products indicated the presence of introns in the larger 5′ m^{7}G-capped *Um*TER transcripts of various sizes (Fig. 3*B*, lanes 7 and 8). To analyze these *Um*TER isoforms quantitatively, we performed Nanopore long-read sequencing on a cDNA library enriched with the *Um*TER precursor transcripts (Fig. 4*A*). Briefly, we performed the RT reaction using a reverse primer targeting a region immediately downstream of the mature *Um*TER sequence to generate cDNA products of the precursor transcripts. Two rounds of nested PCRs were performed to enrich specifically the *UmTER*-containing sequences (Fig. 4*A* and *SI Appendix, Materials and Methods*). Nanopore sequencing of the *Um*TER-enriched library successfully generated 2.6 million valid long reads. Mapping and isoform analysis of these long sequencing reads revealed five isoforms, namely, A to E, of the *Um*TER precursor transcripts (Fig. 4*B*). All five isoforms encompass the mature *Um*TER sequence (Fig. 4*B*, *Top*) and could potentially serve as precursors for producing mature *Um*TER. Isoform A represented the unspliced *Um*TER precursor transcript and accounted for 57.5% of the reads (Fig. 4*B*, *Bottom*). Isoforms B, C, D, and E represented products of alternative intron splicing at two 5′- and two 3′-splice sites, each of which accounted for 14.6%, 9.7%, 9.7%, and 8.4% of the total reads, respectively (Fig. 4*B*, *Bottom*). The introns removed from each of the four spliced isoforms contain the canonical 5′ splice site, branch site, and 3′ splice site sequences (*SI Appendix*, Fig. S8*A*) that are universally conserved in fungal introns (42–44). The branch-site sequence, 5′-UGCUAAGA-3′ (branch point underlined), in the intron is conserved (42) and could base pair with the *U. maydis* U2 snRNA to form an adenosine bulge (*SI Appendix*, Fig. S8*B*).

In addition to the presence of splicing isoforms, the *Um*TER precursor transcripts contain a 3′-poly(A) tail. The G/I tailing-mediated 3′-RACE located the 3′ end of the larger *Um*TER precursor transcripts at position 2295 from the TSS or 186 nt downstream of the 3′ end of the mature *Um*TER (*SI Appendix*, Fig. S9). Sequencing analysis of the cDNA product indicated the presence of a poly(A) tail with approximately 20 adenosine residues (*SI Appendix*, Fig. S9). In summary, the larger *Um*TER precursor transcripts showed hallmark characteristics of mRNA transcripts including a conserved mRNA promoter element, a 5′-m^{7}G cap, alternative intron splicing, and a poly(A) tail.

A Recombinant *Um*TER Precursor Transcript Yields Mature *Um*TER. To directly determine if the *Um*TER precursor is indeed processed to produce mature *Um*TER, we generated a

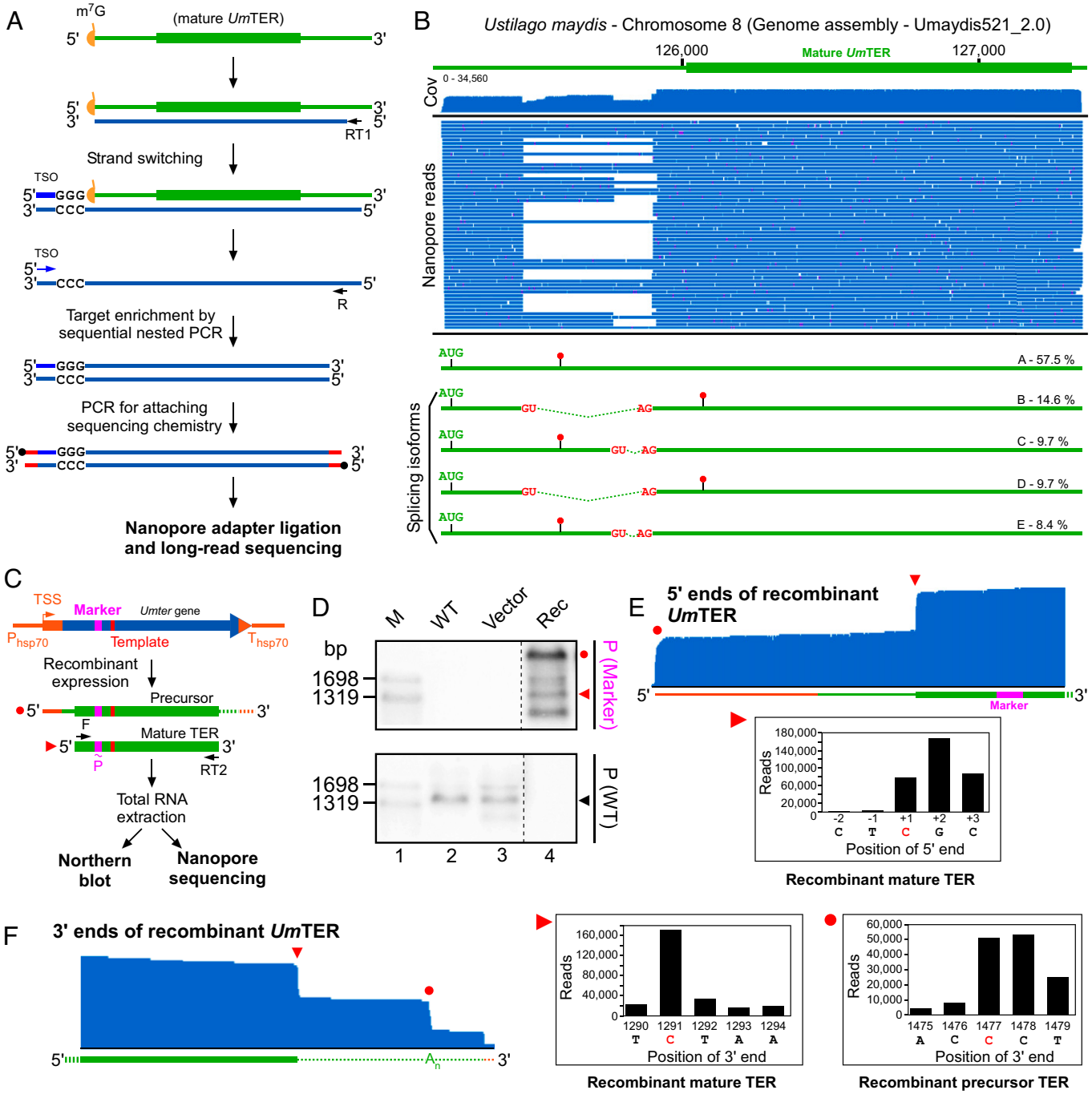


Fig. 4. Analysis of *Um*TER transcript isoforms by Nanopore long-read sequencing. (*A*) Schematic shows the procedure of *Um*TER transcript-enriched cDNA-PCR library preparation for Nanopore sequencing. The 5′-m^{7}G cap is depicted as a half circle (orange). The cap-independent cDNA synthesis was performed in a template-switching RT reaction using a template switch oligo (TSO). The nested-PCR–enriched cDNA library was sequenced using the FLO-MIN106D flow cell on a minION device. (*B*) Coverage map of *Um*TER sequencing reads. The Nanopore long reads were mapped to the *Um*TER genomic locus. A portion of the long-read map from the Integrative Genomics Viewer (IGV) shows the gaps in the intron-splicing isoforms. Five *Um*TER precursor isoforms (A to E) are depicted to show the locations of intron splice sites (GU for the 5′ splice site and AG for the 3′ splice site), start codon (AUG), and stop codons (red dots). The relative abundance (%) of each isoform is shown. The 3′ splice site of isoforms C and D is located 43 nt downstream of the 3′ splice site of isoforms B and E. (*C*) Schematic of an expression cassette for transcribing a recombinant *Um*TER precursor. The recombinant *Umter* gene (blue) transcription is driven by the hsp70 promoter (orange) and flanked by an hsp70 terminator sequence (orange) at the 3′-end. A short region in the recombinant *Umter* gene is replaced with a complementary sequence (purple) to serve as a marker in sequencing analysis. The template sequence is indicated in red. Total RNA extracted from this recombinant *U. maydis* strain was subject to Northern blot analysis and Nanopore sequencing. (*D*) Northern blot analysis of recombinant *Um*TER precursor and mature transcripts. RNA samples isolated from three *U. maydis* strains, WT, vector-only (Vector), and recombinant *Um*TER (Rec) were analyzed. Sizes of DNA markers are shown to the *Left*. On the blot probed for recombinant *Um*TER (Marker), bands of recombinant *Um*TER precursor and mature transcripts are indicated with a red dot and red triangle respectively. On the blot probed for WT *Um*TER (WT), the band of the WT mature *Um*TER is indicated with a black triangle. Probes corresponding to each blot are shown to the *Right*. (*E* and *F*) The 5′ and 3′ ends of precursor and processed recombinant *Um*TER transcripts. Coverage maps of *Um*TER sequencing reads are shown with the 5′ and 3′ ends of the recombinant *Um*TER precursor (red dot) and processed (red triangle) transcripts indicated. The numbers of reads with 5′ or 3′ ends at each specific position are shown in the bar graphs. The expected 5′ or 3′ end of the respective transcripts are indicated in red nucleotides.

U. maydis strain that expresses a recombinant *Um*TER precursor transcript from a Pol II mRNA promoter (Fig. 4*C*). This recombinant gene construct contains the hsp70 promoter to transcribe the first 390 nt of the *U. maydis* GAPDH protein-coding sequence, followed by the mature *Um*TER sequence with 200 bp upstream and 264 bp downstream flanking sequences, and the hsp70 terminator (*SI Appendix*, Fig. S10*A*). To differentiate the recombinant *Um*TER from WT *Um*TER, a 112-bp marker sequence was engineered to replace the region (257 to 368) (*SI Appendix*, Fig. S10*A*), which is located within a functionally dispensable region (208 to 440) (Fig. 2*A*). Following PCR validation of desired transformants (*SI Appendix*, Fig. S10 *B* and *C*), Northern blot analyses were performed using radioactive probes that detect either the engineered marker or WT sequence to confirm the expression of the recombinant *Um*TER transcript (Fig. 4*C*). The Northern blot showed the presence of the recombinant *Um*TER precursor transcript (Fig. 4 *D*, lane 4, red circle) and the processed mature *Um*TER (Fig. 4 *D*, lane 4, red triangle), suggesting that the expressed recombinant *Um*TER precursor can be processed to produce mature *Um*TER. The bands above the processed recombinant *Um*TER were likely transcripts partially processed at either the 5′ or 3′ end, but not both (Fig. 4*D*, lane 4).

To quantitate the distribution of the recombinant precursor and mature *Um*TER transcripts, we performed Nanopore long-read sequencing using a similar strategy described in Fig. 4*A* but with a reverse primer targeting both the mature and precursor transcripts (Fig. 3*C*). More than one million (1,042,646) valid reads were obtained and 99% of the reads were mapped to the recombinant *Um*TER sequence with less than 0.5% of the reads mapped to the WT sequence. This dramatic mapping disparity is due to the overexpression of the recombinant *Um*TER compared to the WT *Um*TER gene (Fig. 4*D*, lane 4, black triangle). The mapping of the recombinant *Um*TER sequencing reads confirmed that the 5′-end position of the recombinant mature *Um*TER (Fig. 4*E*, red triangle) is identical to the 5′ end of the WT mature *Um*TER (*SI Appendix*, Fig. S2*I*).

Similarly, we performed Nanopore sequencing to determine the 3′ ends of processed recombinant *Um*TER (Fig. 4*F*). Data analysis from over one million (1,175,036) valid reads confirmed the correct 3′-end processing of the recombinant mature *Um*TER (Fig. 4*F*, red triangle and *SI Appendix*, Fig. S2*K*). Furthermore, the 3′ end of the recombinant precursor transcript has the poly(A) tail at the same position as the WT precursor (Fig. 4*F*, *Right* and *SI Appendix*, Fig. S7). However, it was noticed that some RNA molecules smaller than the processed recombinant *Um*TER were detected in the Northern blot analysis (Fig. 4*D*, lane 4) but absent in the Nanopore sequencing analysis. We suspected these smaller RNAs were products of RNase degradation that possess a 3′ phosphate and thus were resistant to enzymatic G/I-tailing during Nanopore library preparation (Fig. 4*D*, lane 4). Overall, our result showed that a recombinant *Um*TER precursor expressed from a Pol II mRNA promoter can be correctly processed to generate mature *Um*TER.

The *Um*TER Precursor Is a Protein-Coding mRNA. Isoforms A, C, and E of the *Um*TER precursor transcripts contain an ORF that encodes a hypothetical protein, UMAG_03168 (NCBI RefSeq - XP_011389625.1). Protein BLAST searches using UMAG_03168 as a query identified homologs in both Basidiomycota (*SI Appendix*, Table S1) and Ascomycota (*SI Appendix*, Table S2) phyla. The homolog identified in *S. cerevisiae* with a BLAST E-value of 10^{-5} is a protein called Early meiotic induction protein 1 (Emi1). This Emi1 protein appears to induce the expression of a transcription factor called Inducer of meiosis (Ime1) during meiosis initiation in yeast (45, 46).

To determine if UMAG_03168 is truly a homolog of yeast Emi1, we performed multiple sequence alignments of Basidiomycete homologs of UMAG_03168 (*SI Appendix*, Fig. S11*A*) and Ascomycete homologs of yeast Emi1 (*SI Appendix*, Fig. S11*B*) independently, which revealed two similar consensus sequences. A phylogenetic analysis of the aligned sequences for both UMAG_03168 and Emi1 homologs inferred a phylogenetic tree consistent with the established relationships between the two fungal phyla (*SI Appendix*, Fig. S11*C*), supporting the sequence homology between *U. maydis* UMAG_03168 and yeast Emi1 protein. In addition, we found the presence of highly conserved twin Cx_9C motifs in both *U. maydis* UMAG_03168 and the yeast Emi1 protein sequence (Fig. 5*A*). Each Cx_9C motif harbors a pair of cysteine residues spaced by 9 amino acids with positions 4 and 7 being predominantly hydrophobic (Fig. 5*A*) (47).

In addition to the sequence analysis, we performed structure predictions of UMAG_03168 and yeast Emi1 protein sequences using Alphafold2 (48), which inferred two similar structures (*SI Appendix*, Fig. S11 *D* and *E*). These two structures clearly showed a high degree of structural homology with a similar positioning of the universally conserved residues in 5 alpha helices (α1 to α5) (*SI Appendix*, Fig. S11 *D* and *E*). In both predicted structures, the twin Cx_9C motifs stabilize two antiparallel α-helices, namely, α2 and α4, through forming intramolecular disulfide bridges (Fig. 5 *B* and *C*). This type of hairpin-like arrangement has been observed in other proteins with the same twin Cx_9C motifs (47). The conserved positions of the twin Cx_9C motifs and the similar hairpin structure support structural homology between UMAG_03168 and the yeast Emi1 protein.

Based on sequence and structural homology, we renamed UMAG_03168 as the *U. maydis* Emi1 protein (*Um*Emi1). However, *Um*Emi1 is likely to be encoded only in isoform A. Other alternatively spliced isoforms have much lower abundance (Fig. 4*B*) and would produce proteins with altered sequence and structure in helix α5 (*SI Appendix*, Fig. S11*F*). Although we suspect that *Um*Emi1 may play a role in meiosis induction in *U. maydis* (49), its exact function would require further functional studies.

To experimentally confirm the expression of the 15.2-kDa *Um*Emi1 protein in *U. maydis* cells, liquid chromatography-tandem mass spectrometry analysis was performed. Briefly, *U. maydis* cell lysate was resolved on a sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel, and three gel slices covering three protein size ranges, namely, 14 to 16, 16 to 18, and 18 to 20 kDa, were excised and analyzed separately. Following in-gel trypsinization and mass spectrometry analysis, two tryptic peptides unique to *Um*Emi1, namely, ESAENVWELR and ESAENVWELRR, were detected with 100% and 92% probabilities, respectively (Fig. 5*D* and *SI Appendix*, Fig. S12*A* and Table S3). Protein BLAST analysis against the *U. maydis* proteome independently confirmed that both peptides are unique to *Um*Emi1. These two peptides differed by only one arginine at the C terminus, which was likely due to partial trypsin digestion. Importantly, both peptides were detected from only the SDS-PAGE gel slice covering protein sizes between 14 and 16 kDa and not from the other two gel slices, supporting that they were derived from the 15.2-kDa *Um*Emi1 protein.

To further confirm the mass spectrometry results, we expressed a recombinant 3xFLAG-*Um*Emi1 protein in *U. maydis* cells

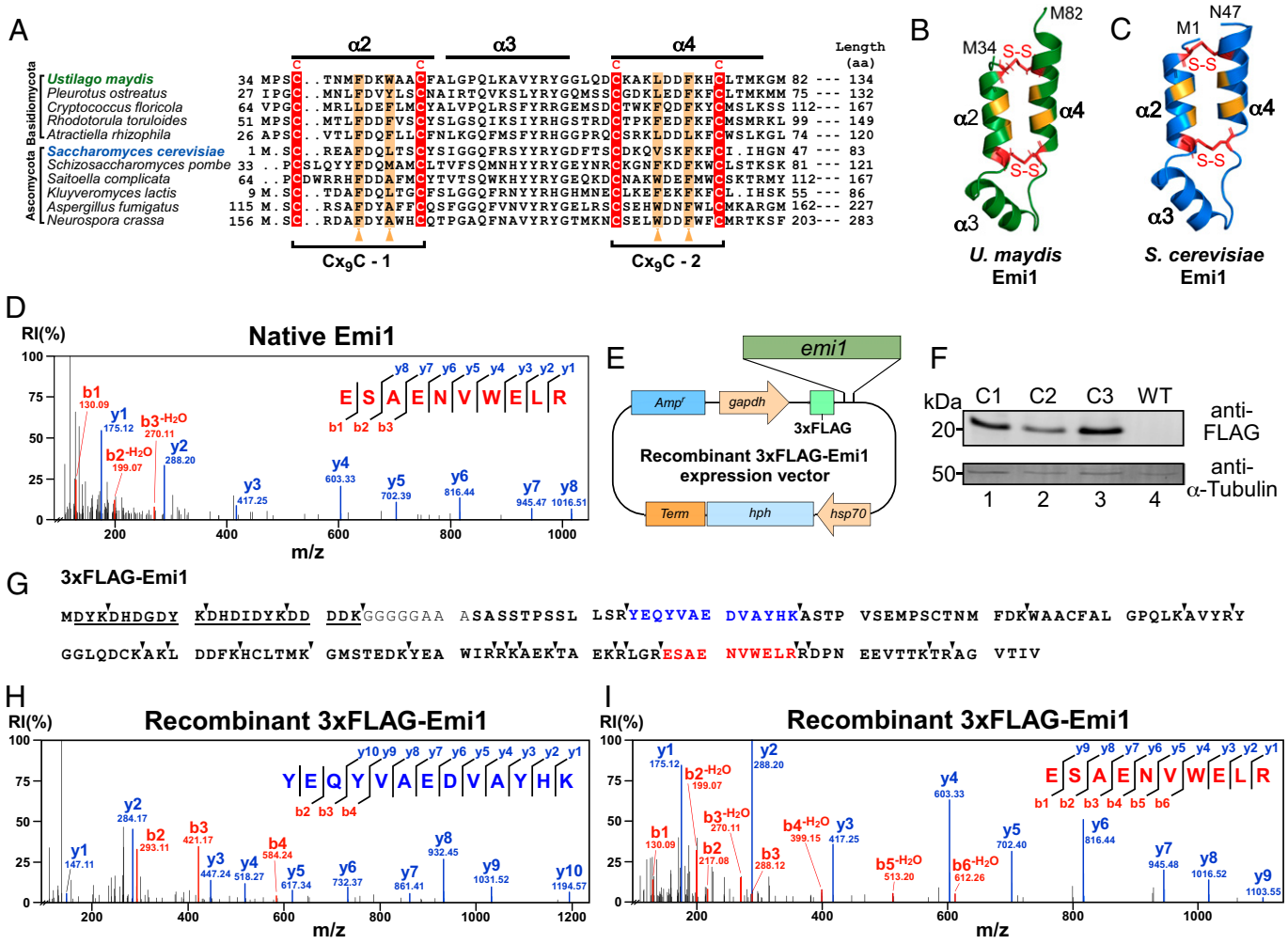


Fig. 5. Detection and homology analysis of the *U. maydis* Emi1 protein. (*A*) Multiple sequence alignment of the twin Cx$_9$C motifs in Emi proteins from representative basidiomycete and ascomycete fungal species. Secondary structure elements, alpha helices α2 to 4, are predicted and indicated above the alignment. The first and last amino acid positions of each sequence are shown to the *Left* and *Right* of the sequence, respectively. The length of each protein is shown to the *Right*. Universally conserved cysteine residues are shaded in red. The 9 amino acid spacer (x$_9$) within each twin Cx$_9$C motif is indicated below the alignment, and the two conserved hydrophobic residues at positions 4 and 7 are shaded in orange. (*B* and *C*) Cartoon representation of Alphafold 2-predicted tertiary structural models of twin Cx$_9$C motifs of *Um*Emi1 (*B*) and *S. cerevisiae* Emi1 (*C*). Universally conserved cysteine residues and disulfide bridges predicted by Alphafold 2 are highlighted in red. (*D*) Tandem mass spectrum of a *Um*Emi1-specific tryptic peptide fragment ESAENVWELR labeled with b- and y-ion fragments. The values of mass to charge ratio (*m/z*) are indicated for the detected *Um*Emi1-derived peptides. RI refers to relative intensity. (*E*) Schematic representation of the extrachromosomal plasmid used to express the 3xFLAG-*Um*Emi1 protein in *U. maydis* cells. Promoters are shown as arrows with the direction of transcription indicated by the arrow. (*F*) Western blot analysis of constitutively expressed 3xFLAG-*Um*Emi1 protein. Three independent *U. maydis* transformant clones (C1, C2, and C3) were analyzed with the WT cells (WT) as a control. The α-tubulin protein was probed by an anti-α-tubulin antibody to show equal loading of the protein samples. (*G*) Primary amino acid sequence of the recombinant 3xFLAG-*Um*Emi1 protein. All predicted Trypsin cleavage sites are shown with arrowheads. Tryptic peptides detected by MS/MS are colored. The 3xFLAG tag is underlined. (*H* and *I*) Tandem mass spectrum of two unique *Um*Emi1 tryptic peptides, namely, YEQYVAEHVAYHK (*H*) and ESAENVWELR (*I*), are shown with b- and y-ion fragments labeled.

(Fig. 5*E*). The expression of the recombinant protein was confirmed by anti-FLAG Western blot analysis (Fig. 5*F*). The recombinant 3xFLAG-*Um*Emi1 protein was then affinity purified by anti-FLAG IP and analyzed by the same mass spectrometry approach. From the purified 3xFLAG-*Um*Emi1 protein sample, three unique tryptic peptides, namely, YEQYVAEDVAYHK (×7), ESAENVWELRR (×3), and ESAENVWELR (×2), covering two separate regions of the *Um*Emi1 protein were detected multiple times (Fig. 5 *G–I* and *SI Appendix*, Fig. S12*B* and Table S3). Reproducible detection of the two peptides, ESAENVWELR and ESAENVWELRR, in both native *Um*Emi1 and the purified recombinant 3xFLAG-*Um*Emi1 protein samples validated the methodology employed and confirmed the expression of the *Um*Emi1 protein in *U. maydis* cells. Noticeably, spectra of the peptide fragment ESAENVWELRR from both native and recombinant *Um*Emi1 predominantly showed doubly protonated y ions (*SI Appendix*, Fig. S12), which was likely due to the protonation of both the arginine residues in the peptide (50). Collectively, the mass spectrometry-based detection of *Um*Emi1 in *U. maydis* cells supports the expression of the *Um*Emi1 protein and that the *Um*TER precursor is indeed a protein-coding mRNA.

Discussion

The remarkable mRNA-derived biogenesis pathway employed by *Um*TER further demonstrates the incredible diversity of TER biogenesis across major eukaryotic kingdoms (Fig. 6). In most eukaryotic lineages including the basal branching trypanosome, TERs are predominantly transcribed by RNA Pol II (5), while ciliate and plant TERs are transcribed by RNA Pol III (7, 19–22). This dramatic Pol II-Pol III switch of TER

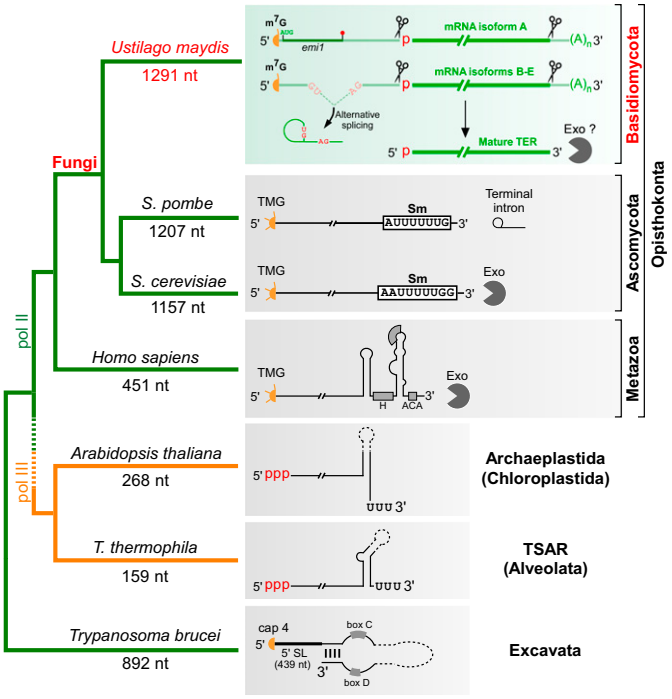


Fig. 6. Distinct mechanisms of TER biogenesis. The phylogenetic tree of major eukaryotic lineages shown is based on recent phylogenomic analyses (51). The representative species and its corresponding TER length are shown above and below the branch, respectively. Branch lengths do not correspond to evolutionary distances. Branches and nodes are colored according to the transcription machinery, namely, Pol II (green) and Pol III (orange), utilized for TER biogenesis in different clades. *Right Panels* show schematics of TER biogenesis pathways from representative species with distinct 5′ chemistry, 3′ biogenesis domains, and processing events. For *U. maydis*, the putative nucleolytic processing of Emi1-encoding mRNA precursor is indicated. The start and stop codons of the *emi1* ORF are shown. TER domains crucial for maturation are shown as sequences within open boxes or rectangles in the respective TER schematic. Exo, exosome-associated nucleases. For *T. brucei*, a spliced leader (SL) is added to the mature TER during a transsplicing event. Respective eukaryotic phyla, kingdoms and supergroups are indicated on the *Right*.

transcription machinery presumably occurred early in eukaryotic evolution prior to the branching of two supergroups, namely, TSAR (Telonemia, Stramenopila, Alveolata, and Rhizaria) that includes ciliates and Archaeplastida (Chloroplastida) that include land plants (51, 52). Metazoa and fungi belong to the same supergroup Opisthokonta and use RNA Pol II for TER transcription. While it remains speculative which apparatus for TER transcription is employed in early eukaryotes, the *Trypanosoma brucei* TER being a Pol II transcript suggests that the RNA Pol II is the ancestral machinery for TER transcription (Fig. 6).

The use of different transcription machinery inevitably led to distinct TER biogenesis mechanisms (Fig. 6). For example, the Pol III–transcribed TERs are relatively smaller in size and possess a 5′ triphosphate and a 3′ poly(U) tail that is typically bound and protected by an accessory protein with a La motif (53). In contrast, the Pol II–transcribed TERs from metazoa and fungi are significantly larger and possess a hypermethylated 5′-TMG cap and a processed 3′ end (Fig. 6). The 3′ ends of metazoan and ascomycete TERs are defined through snoRNA- and snRNA-like structural elements, respectively, and are processed by exosome-mediated exonucleases (Fig. 6). The *Um*TER precursor is transcribed by RNA Pol II as a typical mRNA transcript but then undergoes an unusual maturation process presumably through endonucleolytic cleavages in the 3′ UTR region to release the mature *Um*TER with a 5′ monophosphate (Fig. 6). In the absence of a protective 5′ cap, the mature *Um*TER would rely on stable RNA structural elements and/or protein binding (54–56) to protect the 5′ end from 5′-to-3′ degradation by exonucleases such as XRN1 (57).

The 3′ end of ascomycete yeast TER contains an Sm binding site that is bound by the heptameric Sm ring protein complex to protect the 3′ end from 3′-to-5′ exonuclease degradation (27, 30). The 3′ region of mature *Um*TER also contains a putative Sm-binding site, 5′-AUUUUU-3′, located 9 nt upstream of the 3′ end (*SI Appendix*, Fig. S13), which suggests *Um*TER may undergo a 3′-end maturation process similar to the yeast TER. The *S. pombe* and *N. crassa* TERs employ a unique spliceosomal cleavage reaction to generate the 3′ end (30–32). We did not find any potential 5′ splice site at or near the 3′ end of the mature *Um*TER. Thus, the *Um*TER 3′-end maturation may not employ a terminal intron splicing but rather rely on an endonucleolytic mechanism that is yet to be identified.

This mRNA-derived biogenesis of *Um*TER may have originated through either chromosomal translocations, inversions, or deletions that merged the *Umter* gene with the 3′ UTR of an mRNA gene, creating a unique polycistronic mRNA-lncRNA fusion gene (58). The conserved ORF in the *Um*TER precursor mRNA encodes a protein homologous to the yeast Emi1 protein (Fig. 5*A*–*C* and *SI Appendix*, Figs. S10 and S11). Yeast Emi1 is required for up-regulation of Ime1 that is the master regulator of yeast meiosis and is activated during early meiosis (45, 46). As *U. maydis* does not appear to harbor an *Ime1* ortholog (49), the target genes of the *Um*Emi1 protein are yet to be identified. Since the *Um*Emi1 protein and *Um*TER are produced from the same polycistronic gene, the processing of *Um*TER from the mRNA precursor would presumably result in mRNA degradation and affect *Um*Emi1 protein expression. Thus, it remains to be explored if the *Um*TER biogenesis plays a role in *U. maydis* meiotic induction. Furthermore, disrupting the *Um*TERT coding gene (*trt1*) was shown to impair *U. maydis* teliospore production (59), a crucial step in the meiotic phase of *U. maydis* life cycle, which suggests telomerase may play a role in *U. maydis* meiosis. Elucidation of the detailed mechanism and regulation of this mRNA-derived *Um*TER biogenesis would bring more insights to RNA and telomere biology.

Materials and Methods

The identification of *Um*TER was carried out through affinity purification of recombinant *U. maydis* telomerase, TRAP telomerase activity assay, Illumina next-generation sequencing of TERT-bound RNA, and bioinformatics search of *Um*TER candidates. The validation of *UmTER* was performed via RNA secondary structure determination by phylogenetic comparative analysis, in vitro reconstitution of *U. maydis* telomerase activity, and TRF telomere length analysis of Δ*Um*TER strains. The analysis of *UmTER* biogenesis employed Northern blot analysis, specific enzymatic treatments, 5′-cap-specific IP, RT-qPCR analysis, and Nanopore long-read sequencing of both WT and recombinant *Um*TER precursor. The analysis of *Um*Emi1 protein expression employed in-gel trypsinization of native or recombinant *Um*Emi1 proteins and mass spectrometry analysis of tryptic peptides. Details of all materials and methods are available in *SI Appendix, Materials and Methods*.

Data, Materials, and Software Availability. The *Um*TER gene sequence has been deposited in GenBank (accession no. ON015423) (60), and all other TERs identified in this work have been deposited in GenBank Third Party Annotation (TPA) database (accession no. BK061232–BK061250) (61, 62). Next-generation sequencing data generated in this study have been deposited in the Sequence Reads Archive (SRA) database, https://www.ncbi.nlm.nih.gov/sra (BioProject identifier PRJNA817681) (63).

ACKNOWLEDGMENTS. We thank Drs. Eun Young Yu and Neal Lue for sharing the *U. maydis* strains and Dr. William Holloman for helpful advice on

U. maydis transformation. We thank Dr. Renu Goel from the Northwestern University proteomics core facility for critical discussions on mass spectrometry experiments and data analysis. This work was supported by NSF grants to J.J.-L.C. (MCB1616078 and MCB2046798). Proteomics services were performed by the Northwestern Proteomics Core Facility, generously supported by National Cancer Institute (NCI) Cancer Center Support Grant (CCSG) P30 CA060553 awarded to the Robert H Lurie Comprehensive Cancer Center, instrumentation award (S10OD025194) from NIH Office of Director and the National Resource for Translational and Developmental Proteomics supported by P41 GM108569.

1. T. R. Cech, J. A. Steitz, The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
2. J. J. Quinn, H. Y. Chang, Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* **17**, 47–62 (2016).
3. J. W. Shay, W. E. Wright, Telomeres and telomerase: Three decades of progress. *Nat. Rev. Genet.* **20**, 299–309 (2019).
4. A. G. Lai *et al.*, The protein subunit of telomerase displays patterns of dynamic evolution and conservation across different metazoan taxa. *BMC Evol. Biol.* **17**, 107 (2017).
5. J. D. Podlevsky, J. J.-L. Chen, Evolutionary perspectives of telomerase RNA structure and function. *RNA Biol.* **13**, 720–732 (2016).
6. D. Logeswaran, Y. Li, J. D. Podlevsky, J. J.-L. Chen, Monophyletic origin and divergent evolution of animal telomerase RNA. *Mol. Biol. Evol.* **38**, 215–228 (2021).
7. P. Fajkus *et al.*, Telomerase RNAs in land plants. *Nucleic Acids Res.* **47**, 9842–9856 (2019).
8. A. Mosig, J. J.-L. Chen, P. F. Stadler, Homology search with fragmented nucleic acid sequence patterns. *Lect. Notes Comput. Sci.* **4645**, 335–345 (2007).
9. X. Qi *et al.*, The common ancestral core of vertebrate and fungal telomerase RNAs. *Nucleic Acids Res.* **41**, 450–462 (2013).
10. J. Leonardi, J. A. Box, J. T. Bunch, P. Baumann, TER1, the RNA subunit of fission yeast telomerase. *Nat. Struct. Mol. Biol.* **15**, 26–33 (2008).
11. J.-L. Chen, C. W. Greider, Functional analysis of the pseudoknot structure in human telomerase RNA. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8080–8085, discussion 8077–8079 (2005).
12. C. J. Bley *et al.*, RNA-protein binding interface in the telomerase ribonucleoprotein. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20333–20338 (2011).
13. J.-L. Chen, M. A. Blasco, C. W. Greider, Secondary structure of vertebrate telomerase RNA. *Cell* **100**, 503–514 (2000).
14. J. D. Podlevsky, Y. Li, J. J.-L. Chen, The functional requirement of two structural domains within telomerase RNA emerged early in eukaryotes. *Nucleic Acids Res.* **44**, 9891–9901 (2016).
15. J. D. Podlevsky, Y. Li, J. J.-L. Chen, Structure and function of echinoderm telomerase RNA. *RNA* **22**, 204–215 (2016).
16. D. X. Mason, E. Goneska, C. W. Greider, Stem-loop IV of tetrahymena telomerase RNA stimulates processivity in trans. *Mol. Cell. Biol.* **23**, 5606–5613 (2003).
17. J. Franke, J. Gehlen, A. E. Ehrenhofer-Murray, Hypermethylation of yeast telomerase RNA by the snRNA and snoRNA methyltransferase Tgs1. *J. Cell Sci.* **121**, 3553–3560 (2008).
18. L. Chen *et al.*, Loss of human TGS1 hypermethylase promotes increased telomerase RNA and telomere elongation. *Cell Rep.* **30**, 1358–1372.e5 (2020).
19. C. W. Greider, E. H. Blackburn, A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature* **337**, 331–337 (1989).
20. J. Lingner, L. L. Hendrick, T. R. Cech, Telomerase RNAs of different ciliates have a common secondary structure and a permuted template. *Genes Dev.* **8**, 1984–1998 (1994).
21. J. Song *et al.*, The conserved structure of plant telomerase RNA provides the missing link for an evolutionary pathway from ciliates to humans. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 24542–24550 (2019).
22. P. Fajkus *et al.*, Evolution of plant telomerase RNAs: Farther to the past, deeper to the roots. *Nucleic Acids Res.* **49**, 7680–7694 (2021).
23. T. H. D. Nguyen *et al.*, Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* **557**, 190–195 (2018).
24. C. K. Tseng, H. F. Wang, M. R. Schroeder, P. Baumann, The H/ACA complex disrupts triplex in hTR precursor to permit processing by RRP6 and PARN. *Nat. Commun.* **9**, 5430 (2018).
25. B. E. Jády, E. Bertrand, T. Kiss, Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. *J. Cell Biol.* **164**, 647–652 (2004).
26. A. S. Venteicher *et al.*, A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science* **323**, 644–648 (2009).
27. A. G. Seto, A. J. Zaug, S. G. Sobel, S. L. Wolin, T. R. Cech, *Saccharomyces cerevisiae* telomerase is an Sm small nuclear ribonucleoprotein particle. *Nature* **401**, 177–180 (1999).
28. W. Tang, R. Kannan, M. Blanchette, P. Baumann, Telomerase RNA biogenesis involves sequential binding by Sm and Lsm complexes. *Nature* **484**, 260–264 (2012).
29. J. F. Noël, S. Larose, S. Abou Elela, R. J. Wellinger, Budding yeast telomerase RNA transcription termination is dictated by the Nrd1/Nab3 non-coding RNA termination pathway. *Nucleic Acids Res.* **40**, 5625–5636 (2012).
30. J. A. Box, J. T. Bunch, W. Tang, P. Baumann, Spliceosomal cleavage generates the 3′ end of telomerase RNA. *Nature* **456**, 910–914 (2008).
31. X. Qi *et al.*, Prevalent and distinct spliceosomal 3′-end processing mechanisms for fungal telomerase RNA. *Nat. Commun.* **6**, 6105 (2015).
32. R. Kannan, R. M. Helston, R. O. Dannebaum, P. Baumann, Diverse mechanisms for spliceosome-mediated 3′ end processing of telomerase RNA. *Nat. Commun.* **6**, 6104 (2015).
33. B. Lemieux *et al.*, Active yeast telomerase shares subunits with ribonucleoproteins RNase P and RNase MRP. *Cell* **165**, 1171–1181 (2016).
34. Y. Li *et al.*, Identification of purple sea urchin telomerase RNA using a next-generation sequencing based approach. *RNA* **19**, 852–860 (2013).
35. K. Bösch *et al.*, Genetic manipulation of the plant pathogen ustilago maydis to study fungal biology and plant microbe interactions. *J. Vis. Exp.* **115**, 54522 (2016).
36. B. Min, K. Collins, An RPA-related sequence-specific DNA-binding subunit of telomerase holoenzyme is required for elongation processivity and telomere maintenance. *Mol. Cell* **36**, 609–619 (2009).
37. R. Sandhu *et al.*, A trans-spliced telomerase RNA dictates telomere synthesis in *Trypanosoma brucei*. *Cell Res.* **23**, 537–551 (2013).
38. A. Deana, H. Celesnik, J. G. Belasco, The bacterial enzyme RppH triggers messenger RNA degradation by 5′ pyrophosphate removal. *Nature* **451**, 355–358 (2008).
39. K. Borden, B. Culjkovic-Kraljacic, V. H. Cowling, To cap it all off, again: Dynamic capping and recapping of coding and non-coding RNAs to control transcript fate and biological activity. *Cell Cycle* **20**, 1347–1360 (2021).
40. X. S. Liu, D. L. Brutlag, J. S. Liu, An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nat. Biotechnol.* **20**, 835–839 (2002).
41. B. J. Andrews, I. Herskowitz, The yeast SWI4 protein contains a motif present in developmental regulators and is part of a complex involved in cell-cycle-dependent transcription. *Nature* **342**, 830–833 (1989).
42. M. E. Wilkinson, C. Charenton, K. Nagai, RNA splicing by the spliceosome. *Annu. Rev. Biochem.* **89**, 359–388 (2020).
43. D. M. Kupfer *et al.*, Introns and splicing elements of five diverse fungi. *Eukaryot. Cell* **3**, 1088–1100 (2004).
44. N. Sheth *et al.*, Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res.* **34**, 3955–3967 (2006).
45. A. H. Enyenihi, W. S. Saunders, Large-scale functional genomic analysis of sporulation and meiosis in *Saccharomyces cerevisiae*. *Genetics* **163**, 47–54 (2003).
46. Y. Kassir, D. Granot, G. Simchen, IME1, a positive regulator gene of meiosis in *S. cerevisiae*. *Cell* **52**, 853–862 (1988).
47. S. Longen *et al.*, Systematic analysis of the twin cx(9)c protein family. *J. Mol. Biol.* **393**, 356–368 (2009).
48. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
49. M. E. Donaldson, B. J. Saville, Bioinformatic identification of Ustilago maydis meiosis genes. *Fungal Genet. Biol.* **45** (suppl. 1), S47–S53 (2008).
50. G. Zhang, R. S. Annan, S. A. Carr, T. A. Neubert, Overview of peptide and protein analysis by mass spectrometry. *Curr. Protoc. Mol. Biol.* **108**, Unit16.1 (2014).
51. F. Burki, A. J. Roger, M. W. Brown, A. G. B. Simpson, The new tree of eukaryotes. *Trends Ecol. Evol.* **35**, 43–55 (2020).
52. T. Cavalier-Smith *et al.*, Multigene eukaryote phylogeny reveals the likely protozoan ancestors of opisthokonts (animals, fungi, choanozoans) and Amoebozoa. *Mol. Phylogenet. Evol.* **81**, 71–85 (2014).
53. K. L. Witkin, K. Collins, Holoenzyme proteins required for the physiological assembly and activity of telomerase. *Genes Dev.* **18**, 1107–1118 (2004).
54. M. P. Deutscher, Degradation of stable RNA in bacteria. *J. Biol. Chem.* **278**, 45041–45044 (2003).
55. J. Houseley, D. Tollervey, The many pathways of RNA degradation. *Cell* **136**, 763–776 (2009).
56. X. Hu *et al.*, Quality-control mechanism for telomerase RNA folding in the cell. *Cell Rep.* **33**, 108568 (2020).
57. N. L. Garneau, J. Wilusz, C. J. Wilusz, The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* **8**, 113–126 (2007).
58. G. Leonard, T. A. Richards, Genome-scale comparative analysis of gene fusions, gene fissions, and the fungal tree of life. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21402–21407 (2012).
59. D. Bautista-España *et al.*, The telomerase reverse transcriptase subunit from the dimorphic fungus Ustilago maydis. *PLoS One* **9**, e109981 (2014).
60. D. Logeswaran *et al.*, Ustilago maydis telomerase RNA, complete sequence. GenBank. https://www.ncbi.nlm.nih.gov/nuccore/ON015423. Deposited 16 March 2022.
61. D. Logeswaran *et al.*, The telomerase RNA. https://www.ncbi.nlm.nih.gov/nuccore?term=BK061232. GenBank Third Party Annotation. Deposited 24 March 2022.
62. D. Logeswaran *et al.*, The telomerase RNA. https://www.ncbi.nlm.nih.gov/nuccore?term=BK061250. GenBank Third Party Annotation. Deposited 24 March 2022.
63. Arizona State University. Identification of Ustilago maydis telomerase RNA (TR) and characterization of TR precursors. NCBI BioProject. https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA817681. Deposited 19 March 2022.