

Putative telomerase catalytic subunits from *Giardia lamblia* and *Caenorhabditis elegans*

Harmit S. Malik ^{a,b}, William D. Burke ^a, Thomas H. Eickbush ^{a,*}

^a Department of Biology, University of Rochester, Rochester, NY 14627-0211, USA

^b Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109-1024, USA

Received 28 January 2000; received in revised form 25 April 2000; accepted 28 April 2000

Received by D. Finnegan

Abstract

Eukaryotic chromosomes end in short nucleotide repeats that are added by the enzyme telomerase. The catalytic subunit of telomerase has been shown to be most closely related in sequence to reverse transcriptases encoded by eukaryotic retrotransposable elements. This raises the question as to whether the telomerase subunit was present in the first eukaryotes or was derived during early eukaryote evolution from the replication machinery of a retrotransposable element. We present the sequence of a putative telomerase catalytic subunit from the diplomonad parasite, *Giardia lamblia*. The *G. lamblia* subunit appears to have most of the characteristics of other sequenced telomerases, except that it lacks the conserved telomerase-specific ‘T’ motif previously identified in other eukaryotic genes. Searching genomic databases with the *G. lamblia* sequence, we also identified a potential telomerase catalytic subunit from *Caenorhabditis elegans*. The *C. elegans* subunit is uncharacteristically short, and lacks several motifs found in all other telomerases. The identification of a *G. lamblia* telomerase similar to that of most other eukaryotes suggests that telomerase dates back to the earliest extant marker of eukaryotic evolution. The atypical *C. elegans* telomerase, on the other hand, raises intriguing biochemical questions concerning sub-domains of the telomerase catalytic subunit previously considered indispensable. The enzymatic machinery for telomere formation in *C. elegans* is likely to differ substantially from that of other eukaryotes. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Non-LTR retrotransposons; Phylogeny; Reverse transcriptase; Telomeres

1. Introduction

Reverse transcriptase (RT) sequences of eukaryotic genomes can be divided into three classes. The first class includes retroviruses and retrotransposable elements that bear long terminal repeats (LTRs) in their DNA sequence (reviewed in Whitcomb and Hughes, 1992). The second class of eukaryotic RT sequences includes mobile retrotransposons that do not contain long terminal repeats. These non-LTR retrotransposons use the free 3′ end of a DNA strand released by cleavage of a target site to prime reverse transcription of the element’s RNA transcript directly onto the chromosome (Luan

et al., 1993). The final class of eukaryotic RT sequences comprises the catalytic subunit of the telomerase machinery that is responsible for maintaining the integrity of linear eukaryotic chromosomes (Lingner et al., 1997; Nakamura et al., 1997). Telomerases stabilize the ends of chromosomes by reverse transcribing a portion of a bound RNA onto the chromosome ends. This mechanism has similarities to that used by non-LTR elements to synthesize a new copy of the element at its target site. Further evidence of this similarity comes from *Drosophila melanogaster*, where the telomerase gene appears to have been lost and telomerase function taken over by the non-LTR elements, HeT-A and TART (Levis et al., 1993; Pardue et al., 1996).

The relative ancestry of telomerases and non-LTR elements has been debated (Nakamura et al., 1997; Eickbush, 1997; Nakamura and Cech, 1998). Did the telomerase machinery give rise to a new class of mobile elements, or was the non-LTR retrotransposition mechanism co-opted by cells to overcome replication problems

Abbreviations: EST, expressed sequence tag; LTR, long terminal repeat; ORF, open reading frame; PSSM, position-specific scoring matrix; RT, reverse transcriptase.

* Corresponding author. Tel.: +1-716-275-7247; fax: +1-716-275-20701.

E-mail address: eick@mail.rochester.edu (T.H. Eickbush)

associated with linear chromosomes? The origin of telomerases is predicted to correlate with or predate the origin of eukaryotes as linear chromosomes with telomeric repeats are found in the oldest known eukaryotes (reviewed in Nakamura and Cech, 1998). Here we report the sequence of a putative telomerase from the diplomonad parasite, *Giardia lamblia*. Due to the basal position of Diplomonads in the Eukarya, this sequence is expected to reflect one of the oldest branches of the telomerase reverse transcriptase. Using this telomerase sequence to search genomic databases, we also identified the putative telomerase catalytic subunit from *Caenorhabditis elegans*.

2. Materials and methods

2.1. DNA amplification, cloning and sequencing

Preliminary sequence data from the *Giardia lamblia* Genome Project were obtained from The Josephine Bay Paul Center WEB site at the Marine Biological Laboratory (www.bpc.mbl.edu). These preliminary sequencing runs identified a reverse transcriptase with homology to the telomerase reverse transcriptase. Using these sequences, oligonucleotide primers were constructed, and the entire telomerase gene was PCR amplified from *G. lamblia* genomic DNA (gift from Stephen Rich) and sequenced. The eight primers utilized (5'-TAGCCAGTAATAAGATTCATC-3', 5'-CTCCTCAGAGTTGGCTAGC-3', 5'-CGGTAATCCAGGAGCTAAGTC-3', 5'-AATGCTCGGATGTAGGGCCGG-3', 5'-CCGACGACACAGGGTCTCAC-3', 5'-TGGGTTCTCTAGAACAGTCT-3', 5'-TGCAACTACTAAATACAGGG-3', 5'-CTAAGTATAGGAGTATCCAG-3') were appropriately spaced (every 500 bp) and oriented to allow complete cloning of the telomerase gene on multiple overlapping fragments. PCR products were cloned into mp18T2 (Burke et al., 1999) and multiple clones organized into complementary pairs. Both strands of the region starting 250 bp upstream of the open reading frame to 460 bp downstream of the open reading frame were sequenced. The complete *G. lamblia* telomerase gene (GITERT) has been deposited in GenBank under the Accession No. AF195121.

The first eight introns in the *C. elegans* telomerase gene were confirmed by sequence comparison to sequenced ESTs (expressed sequence tags). To confirm the location of the two introns at the 3' end of the gene, primers 5'-GGAGGTCCACAGGGACATCCTATATCTTCA-3' (to conserved motif B in the RT domain) and 5'-GTGCCTCGATCGAATCGAATCGAAATAAATCT-3' (reverse complement of the presumed 3' untranslated region) were used to amplify the 3' end of the *C. elegans* gene both from a cDNA library and genomic DNA (gift from Mark Roth). The sequence

obtained from the PCR amplified product from the cDNA library confirmed the predicted exon–intron junctions as well as the termination codon of the *C. elegans* telomerase gene–CeTERT (Accession No. Z96047).

2.2. Sequence analysis

The *C. elegans* telomerase subunit could not be identified using gapped BLAST searches; however, an iterative PSI-BLAST search (Altschul et al., 1997) using the *G. lamblia* sequence as query did identify the sequence. To score for telomerase motifs outside the RT domain, we used all known telomerases except that from *S. cerevisiae* to construct position-specific scoring matrices (PSSMs) (Henikoff and Henikoff, 1994). These PSSMs were then used to search the ScTERT, GITERT and CeTERT sequences using both PSI-BLAST and MAST (Motif Sequence Alignment Tool) (Bailey et al., 1997). The RT domains (1 through E) of the various telomerases were aligned using previously defined motifs (Xiong and Eickbush, 1990; Nakamura et al., 1997) and CLUSTALX (Thompson et al., 1997). An alignment of the telomerases and non-LTR retrotransposons was used for phylogenetic analysis using the Neighbor Joining method (Saitou and Nei, 1987) and maximum parsimony heuristic options as implemented in PAUP* (Swofford, 1999) Version 4.0d64 (tree-bisection-reconnection branch swapping with maximum number of trees saved at each step limited to five). Bootstrapping was also carried out using PAUP* Version 4.0d64.

3. Results and discussion

3.1. Identification of the *G. lamblia* telomerase gene

The ends of *G. lamblia* linear chromosomes contain TAGGG repeats, suggesting that this organism would contain a telomerase (Adams et al., 1991). We identified a segment of the *G. lamblia* telomerase catalytic subunit in a public release of random genomic sequences provided by the *G. lamblia* Genome Project (The Josephine Bay Paul Center WEB site at the Marine Biological Laboratory). Using this preliminary data, we PCR amplified and sequenced the entire telomerase gene (see Section 2.1). The open reading frame (ORF) of the telomerase, termed GITERT in keeping with convention, is uninterrupted by introns as is the case for all *G. lamblia* genes identified to date (Smith et al., 1998). Starting at the first methionine codon, GITERT is 960 amino acid residues long, a length similar to that determined for other eukaryotic telomerases. As shown in Fig. 1, the location of the RT domain within the GITERT ORF is also in agreement with that found for the other eukaryotic genes. As will be discussed below, this RT contains all the conserved protein motifs pre-

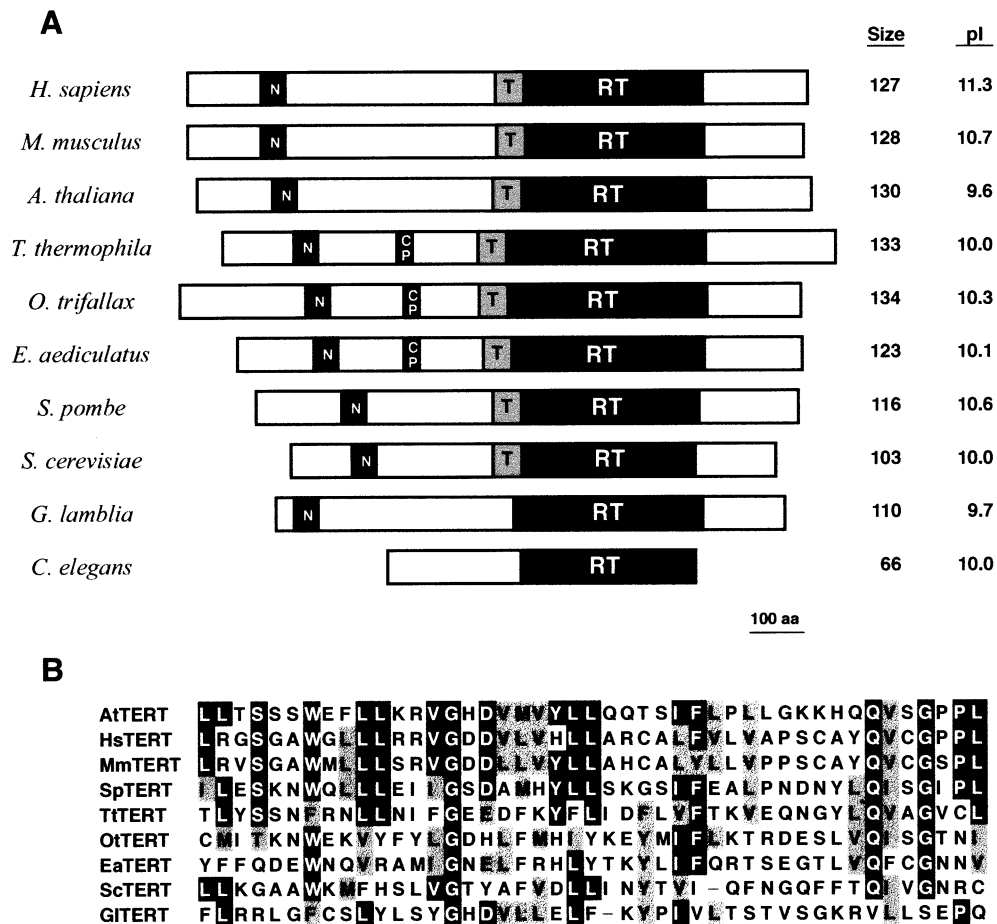


Fig. 1. (A) Schematic representation of all identified eukaryotic telomerases. The various telomerase ORFs are represented along with their deduced pI and molecular weights (from http://www.expasy.ch/tools/pi_tool.html). Shown within these ORFs are the core reverse transcriptase (RT) motifs (palm and fingers sub-domain) found in all telomerases (see Fig. 3), the 'N' motif conserved in all but *C. elegans* genes, and the T motif conserved in all but *G. lamblia* and *C. elegans*. The CP motif is conserved only in the ciliated protozoans (Bryan et al., 1998). The *C. elegans* telomerase is also the only protein that lacks the carboxy-terminal extension at the end of the core RT domain. (B) The 'N' motif identified as being conserved across most telomerases is presented in BOXSHADE format, with identical and similar amino acids present in five or more genes shown in reverse type and gray shading, respectively.

viously identified in other telomerases, with the exception of the 'T' motif.

3.2. Identification of the telomerase gene from *C. elegans*

C. elegans also has typical telomeric repeats of the sequence TTAGGC (Wicky et al., 1996). However, few components of the telomerase machinery have been identified in this model organism, despite its entire genome being essentially sequenced (The *C. elegans* Sequencing Consortium, 1998). When the *G. lamblia* telomerase sequence was used in a PSI-BLAST search (Altschul et al., 1997), all other previously described telomerases were recovered. To our surprise, this search also uncovered a *C. elegans* sequence at significant *E*-values. This gene had originally been annotated 'reverse-transcriptase like', but had gone unremarked upon in prior analyses of telomerase or reverse transcriptase sequences (see Nakamura et al., 1997; Bryan et al., 1998;

Oguchi et al., 1999). More recently, the *C. elegans* database (www.WormBase.org) had identified the gene as DY3.4 and attributed Neal Lue (Cornell University) to identifying it as the best match to telomerase in the database.

Features of the CeTERT (DY3.4) gene are presented in Fig. 2. Comparison of the CeTERT protein sequence to an EST database using TBLASTN searches confirmed both that this gene is expressed, and that the predicted splice sites for the first eight codons of the gene were correct (Fig. 2). Unfortunately, no ESTs were found that covered the last two predicted introns of the gene. To confirm the position of these final introns, we used PCR primers to the predicted 3' untranslated region and to an internal region of exon nine (arrows below the exon sequences in Fig. 2) to amplify the 3' end of CeTERT from a cDNA library. The sequence of this PCR product confirmed the predicted exon-intron junctions as well as the termination codon. In those regions

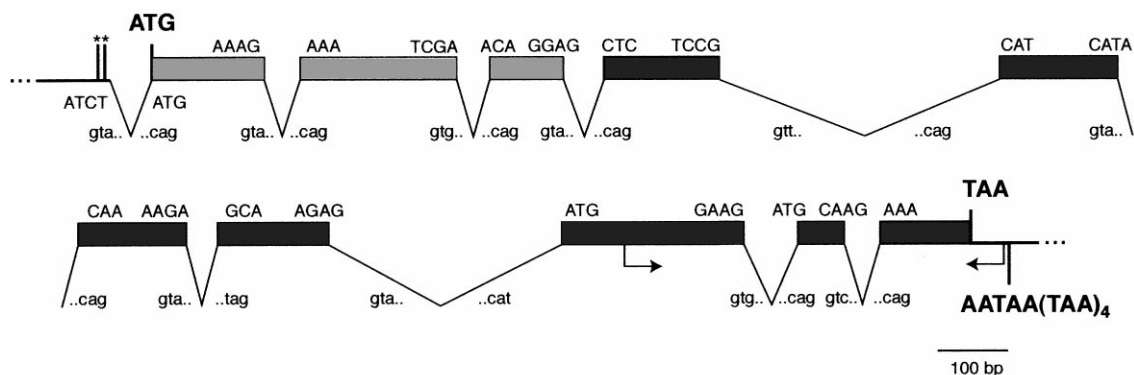


Fig. 2. Genomic features of the CeTERT gene. The exon–intron structure of the CeTERT gene is shown, with the exons represented by a bold line or boxes, and the introns as excised thin lines. The exon–intron junction sequences are shown at each splice site, with the intronic sequences in small letters. The first eight introns were confirmed by ESTs, while the last two were confirmed by PCR amplification of a cDNA library (primer locations indicated by arrows below the exon sequences — see Materials and methods). The predicted ORF is contained within the boxed regions, with the darker shading representing the RT domain. Upstream of the ATG codon (presumed translational startsite), there are two in-frame termination codons (represented by “**”). The presumed stop codon is indicated (TAA) as well as the putative cleavage/polyadenylation signal (AATAA) that is followed by four repeats of TAA. The transcriptional startsite has not been mapped.

of the RT domain where sequence alignments can be made, the CeTERT introns occur at locations different from those found in other telomerase genes (Bryan et al., 1998).

This putative *C. elegans* telomerase differs from other telomerases in several respects. First, although its *pI* is similar to that of other telomerases, it is only two-thirds to one-half the size of the telomerase from other eukaryotes (Fig. 1A). Second, the RT domain of the *C. elegans* sequence is not followed by a carboxyl-terminal domain of 100–200 amino acids, as is the case for all other telomerases. By analogy to the retroviral RT domain (Kohlstaedt et al., 1992) and indeed all DNA polymerases (reviewed in Steitz, 1999), one would expect a significant fraction of this carboxyl-terminal domain to encode the ‘thumb’ subdomain of the telomerase (also see Nakamura et al., 1997; Burke et al., 1999). As described above, the absence of this domain in *C. elegans* does not appear to be due to an incorrect assignment of the intron–exon boundaries. While we cannot rule out the possibility that alternative splicing gives rise to an active protein, both the presence of a ‘strong’ cleavage/polyadenylation site close to the termination codon and the presence of the adjacent gene only 250 nucleotides downstream of this codon makes such an alternative splicing model unlikely. We conclude that the putative telomerase of *C. elegans* is unusually truncated at both its N-terminal and C-terminal ends.

3.3. Unusual features of the GITERT and CeTERT sequences

The putative *G. lamblia* and *C. elegans* telomerase sequences encode identifiable RT domains which contain many of the previously described conserved motifs positioned appropriately (Fig. 3). However, there is a sig-

nificant difference between these two telomerases and those previously reported. The ‘T’ motif, which has been found in all telomerases prior to this report, was not found in either GITERT or CeTERT. In order to make this assertion, the T motif from all telomerases except *S. cerevisiae* were aligned and a position-specific scoring matrix (Henikoff and Henikoff, 1994) constructed which was then used as a PSI-BLAST/MAST query versus the *S. cerevisiae* telomerase sequence (ScTERT) as well as the GITERT and CeTERT sequences (see Materials and methods). (ScTERT was used to test this approach as it contains the poorest matches to the T motif). While ScTERT had a highly significant match (*E* value $<10^{-18}$, see Materials and methods), both GITERT and CeTERT had non-significant matches (approx. 1, where the search space was three proteins).

The CeTERT sequence also contains two unusual amino acid changes in the highly conserved motifs of the RT domain (Fig. 3, arrows). Previously sequenced telomerases all contain two aspartate residues in motif A. The first aspartate, which is also found in both CeTERT and GITERT, is conserved across all RT sequences and has been shown to be a critical component of the active site in retroviral RT. The second aspartate is common to all telomerases (there has been a conservative substitution of a glutamate for the aspartate in GITERT) and many non-LTR elements and group II introns (Lingner et al., 1997). In CeTERT, however, this position contains a serine residue. Mutagenesis studies in yeast have shown that substitution of an alanine for this aspartate residue in ScTERT results in telomere shortening (Lingner et al., 1997). The second unusual substitution in CeTERT is found in motif C. This motif contains the other active site residues of RT, the highly conserved paired aspartate residues found in all reverse transcriptases (see Xiong and Eickbush,

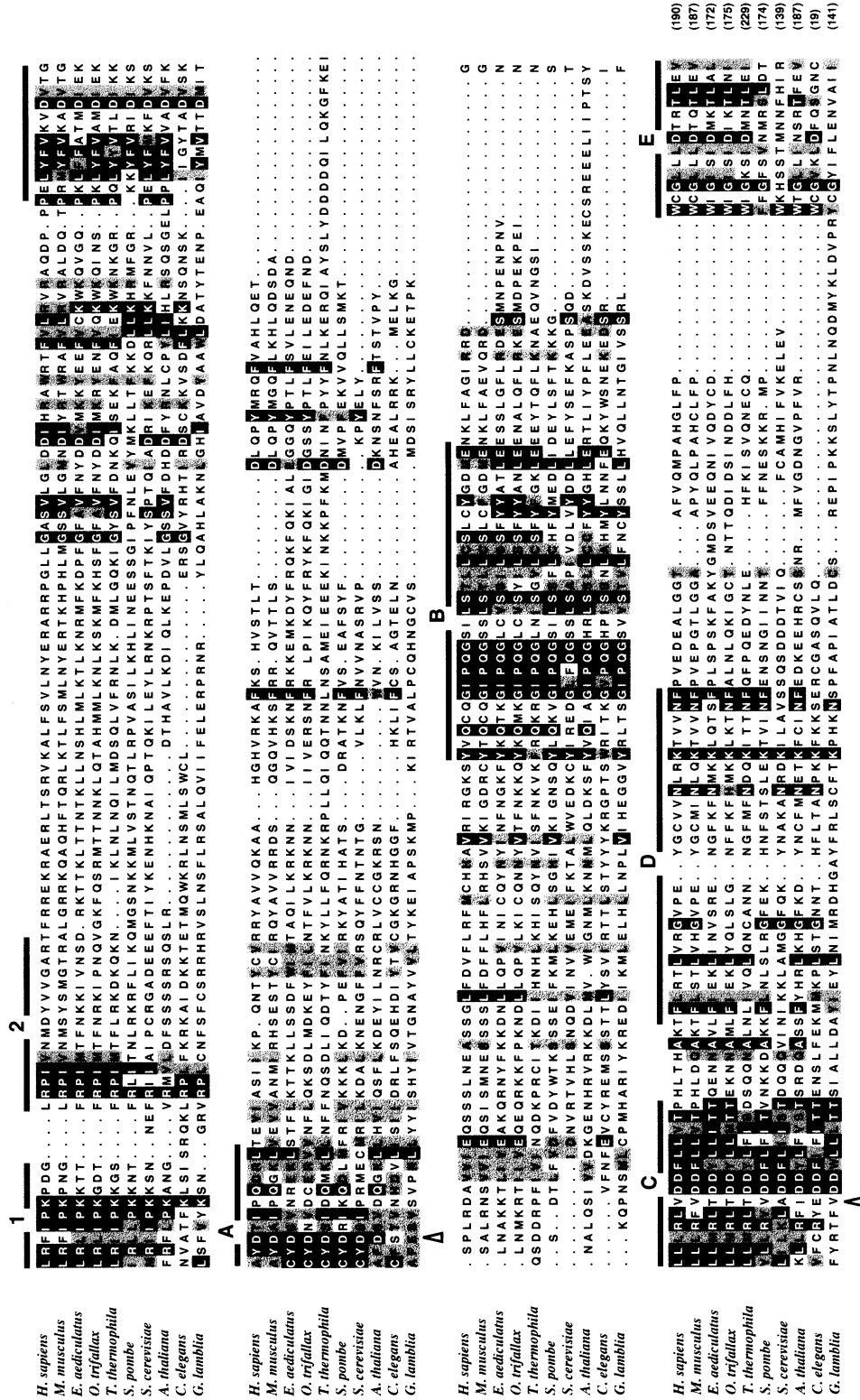


Fig. 3. Alignment of the various telomerase RT domains. The RT motifs were aligned using CLUSTAL-X. Residues conserved at a level of at least 60% are shaded with black and gray shading representing identical and similar amino acids, respectively. The previously identified conserved RT motifs (Nakamura et al., 1997) 1,2, A-E are indicated. The two open arrows indicate positions in which the *C. elegans* telomerase does not appear to match the consensus. Numbers in parentheses indicate the number of amino acids to the end of the ORFs.

1990). These paired aspartates are preceded by a hydrophobic residue or a small polar residue in all reverse transcriptases, except CeTERT, where they are preceded by a glutamate residue. These two differences near the active site of the enzyme are predicted to have a significant effect on the enzymatic properties of the encoded protein.

We also attempted to identify regions of the GITERT and CeTERT outside the RT domain that might be common to telomerases. Not surprisingly, the previously identified ciliated protozoan-specific motif, CP (Bryan et al., 1998), was not found in either GITERT or CeTERT. We did, however, identify another conserved motif nearer the amino-terminal (N-terminal) end (Fig. 1A). This hydrophobic region, which we term 'N' for N-terminal (Fig. 1B), is present in all previously identified telomerases. GITERT also contains a match to this motif at significant levels (E -values $< 10^{-3}$ using a MAST search, see Materials and methods). However, it lacks the conserved glutamine and glycine residues at one end. CeTERT on the other hand, again appears to be atypical as no significant match to the N motif could be found.

3.4. Phylogenetic analysis

The RT domains of the telomerase sequences as shown in Fig. 3 were aligned to representatives from the non-LTR class of retrotransposable elements and the group II intron-encoded ORFs. Presented in Fig. 4 is the resulting phylogeny using Group II introns and non-LTR elements as the outgroup. While GITERT is as predicted the earliest to diverge, CeTERT unexpectedly appears to be the second-most divergent. All other branching patterns within the telomerases are similar to those previously reported (Bryan et al., 1998; Oguchi et al., 1999).

The basal position of the CeTERT on the phylogeny (instead of its expected position on the branch leading to plant and vertebrate telomerases) has several possible explanations. First, the CeTERT gene may represent a lateral transfer from a protozoan or the differential propagation of a paralogous lineage not found in other eukaryotes to date. The analysis of telomerase genes from additional eukaryotic lineages will eventually support or refute this possibility. Second, the unusual position of CeTERT in the phylogenetic analysis could mean that the *C. elegans* sequence we have identified as a telomerase may represent a pseudogene. There is no evidence to support this assumption. Pseudogenes are rare in *C. elegans*, and all aspects of CeTERT including translation start and stop sites and RNA splice sites appear normal. CeTERT transcripts can be identified in cDNA libraries; thus we have every reason to expect that this gene encodes a functional protein. Third, it is possible that CeTERT represents a highly divergent,

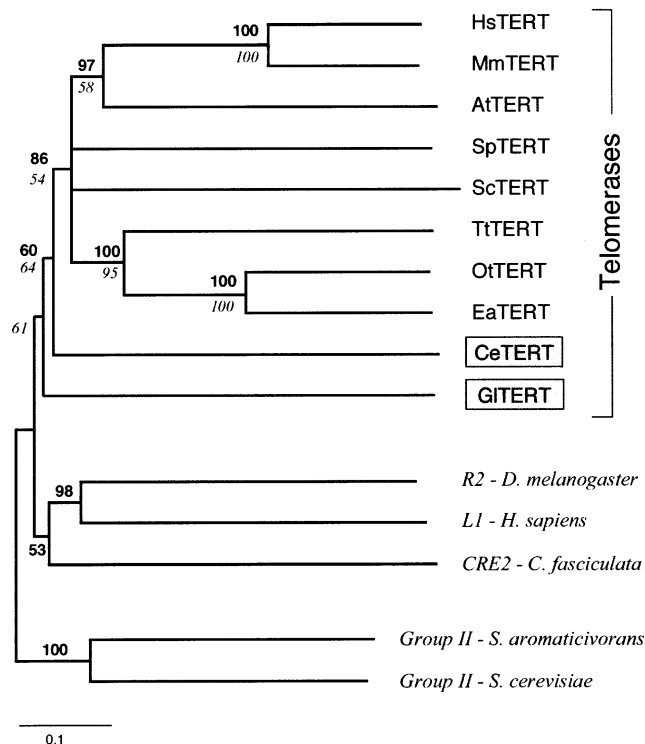


Fig. 4. Phylogeny of the telomerases based on their RT domains. The RT domains of the various telomerases (Fig. 3) were aligned to those of select non-LTR retrotransposable elements (see Malik et al., 1999) and Group II introns (*S. cerevisiae* mitochondrial — #V00694; *Sphingomonas* sp. eubacterial — AF079317). The neighbor-joining tree is arbitrarily rooted on the Group II introns and non-LTR elements and the bootstrap support is shown for the distance (bold numbers) and the parsimony (italics) methods whenever greater than 50%. Branches with less than 50% bootstrap support with both methods were collapsed. Both CeTERT and GITERT appear ancestral to the other telomerase subunits.

uncharacterized retroelement. Again there is no evidence to support this assumption. No retroelement characterized to date encodes such an abundance of introns. In addition, the CeTERT RT domain contains features exclusively associated with telomerases. For example, the large segment between motifs A and B in CeTERT is found only in telomerases. Finally, the CeTERT sequence falls within the telomerase branch in a total phylogeny of RT sequences (Fig. 4, and data not shown).

The most likely explanation for the unusual location of CeTERT within the telomerase phylogeny is a more rapid rate of evolution of CeTERT. Molecular phylogenetic analyses have revealed that many genes of *C. elegans* have undergone accelerated evolution relative to other animals (reviewed in Blaxter, 1998; Wang et al., 1999). This rate difference can result in the artefactual placement of *C. elegans* sequences early in eukaryotic gene phylogenies, a 'long-branch attraction' artefact. While accelerated rates as great as that suggested here have not been found, the acceleration could be com-

pounded in the case of CeTERT given the apparent relaxation of constraints imposed on some of the RT motifs (e.g., absence of a thumb and the unusual A and C motifs). Gene sequencing from other species of nematodes has revealed lower rates of sequence change than that found for *C. elegans* which has, in turn, allowed a more accurate positioning of the nematode lineage (Aguinaldo et al., 1997). Thus identification of telomerase sequences from additional nematodes would help establish the authenticity of this gene and better define what features, if any, of its sequence are unique to *C. elegans*.

We conclude that unless a telomerase gene exists in the small fraction of the *C. elegans* genome that remains unsequenced, CeTERT represents the only possible candidate for the catalytic subunit of telomerase. Direct evidence to support our contention that this gene represents the authentic telomerase could be obtained by inhibiting the activity of this gene and determining telomere and/or chromosome stability.

3.5. Concluding comments

With the identification of the *G. lamblia* and *C. elegans* telomerase catalytic subunits, significant structural variation in the eukaryotic telomerase catalytic subunit has been observed for the first time. The telomerase catalytic subunit has been suggested to interact with different binding partners in different organisms. For example, the *Tetrahymena thermophila* telomerase-associated subunits p80 and p95 (Collins et al., 1995) appear to have no homologs in the *S. cerevisiae* genome. The absence of the T motif in both GITERT and CeTERT suggests that even more radical peculiarities will be found in the telomeric machinery of *G. lamblia* and *C. elegans*. Because the *G. lamblia* telomerase is the most divergent, the absence of a T motif could simply mean that this motif evolved after *G. lamblia* split from the eukaryotic crown group. The highly unusual CeTERT, on the other hand, suggests that *C. elegans* has either evolved unique mechanisms for the formation of telomeres or that some functions of the catalytic subunit have been assumed by other protein components. The two non-conservative changes in the amino acid sequence in close proximity to its putative active sites (see Fig. 3) would predict changes in even the basic enzymatic activity of the protein. The absence of a thumb-analogous domain (Kohlstaedt et al., 1992; Nakamura et al., 1997) in CeTERT could indicate an impaired ability to recognize its cognate telomerase RNA. It is interesting to note in this regard that the telomeric RNA from *C. elegans* remains unidentified. Thus, the *C. elegans* telomerase presents points of interest both to the evolutionary biologist as well as the biochemist. *C. elegans* does have 'typical' telomeric repeats; thus, its divergent telomerase must still function

in spite of its reduced size and the absence of domains found in other telomerases. Given the genetic tools available in *C. elegans*, it will be possible to test what other components or pathways are associated with the formation of telomeres. Indeed, a genetic screen for genes required for germline immortality (mrt) has already identified repair genes that may identify shortened telomeres as a type of DNA damage (Ahmed and Hodgkin, 2000). This screen suggested that approx. 50 genes may be involved in determining germline immortality. Our analysis predicts that CeTERT would be one of these genes.

The identification of GITERT in this report and the identification of a retrotransposable element in the *G. lamblia* genome that represents the most divergent lineage of the non-LTR class to date (data in preparation), confirms that both telomerases and non-LTR elements date back to the earliest extant marker of eukaryote evolution (Sogin et al., 1989; Hashimoto et al., 1994, but see Embley and Hirt, 1998). Thus we are now better situated to re-analyze the origin and evolution of reverse transcriptase-like sequences in eukaryotic genomes (Eickbush, 1994; Malik et al., 1999).

Acknowledgements

Preliminary sequence data from the *Giardia lamblia* Genome Project were obtained from The Josephine Bay Paul Center WEB site at the Marine Biological Laboratory (www.bpc.mbl.edu). Sequencing was supported by the National Institute of Allergy and Infectious Diseases using equipment from LI-COR Biotechnology. We thank Stephen Rich and Mark Roth for the gift of *G. lamblia* genomic DNA and a *C. elegans* cDNA library, respectively. We thank Jorja Henikoff for her help with the MAST/PSI-BLAST motif searches, and Danna Eickbush and Steve Henikoff for comments on the manuscript. This work was supported by grants NSF MCB-9974606 to T.H.E. and NIH GM-29009 to Steve Henikoff.

References

- Adams, R.D., Nash, T.E., Wellems, T.E., 1991. Telomeric location of *Giardia* rDNA repeats. *Mol. Cell. Biol.* 11, 3326–3330.
- Aguinaldo, A.M.A., Turbeville, J.M., Linford, L.S., Rivera, M.C., Garey, J.R., Raff, R.A., Lake, J.A., 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387, 489–493.
- Ahmed, S., Hodgkin, J., 2000. MRT-2 checkpoint protein is required for germline immortality and telomere replication in *C. elegans*. *Nature* 403, 159–164.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST, a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.

- Bailey, T.L., Baker, M.E., Elkan, C.P., 1997. An artificial intelligence approach to motif discovery in protein sequences, application to steroid dehydrogenases. *J. Steroid Biochem. Mol. Biol.* 62, 29–44.
- Blaxter, M., 1998. *Caenorhabditis elegans* is a nematode. *Science* 282, 2041–2046.
- Burke, W.D., Malik, H.S., Jones, J.P., Eickbush, T.H., 1999. Conserved structure and mechanism of integration of the R2 retrotransposable element in all arthropods. *Mol. Biol. Evol.* 16, 502–511.
- Bryan, T.M., Sperger, J.M., Chapman, K.B., Cech, T.R., 1998. Telomerase reverse transcriptase genes identified in *Tetrahymena thermophila* and *Oxytricha trifallax*. *Proc. Natl. Acad. Sci. USA* 95, 8479–8484.
- Collins, K., Kobayashi, R., Greider, C.W., 1995. Purification of *Tetrahymena* telomerase and cloning of genes encoding the two protein components of the enzyme. *Cell* 81, 677–686.
- Eickbush, T.H., 1994. Origin and evolutionary relationships of retroelements. In: Morse, S.S. (Ed.), *The Evolutionary Biology of Viruses*. Raven Press, New York, pp. 121–157.
- Eickbush, T.H., 1997. Telomerase and retrotransposons, which came first? *Science* 277, 911–912.
- Embley, T.M., Hirt, R.P., 1998. Early branching eukaryotes? *Curr. Opin. Genet. Dev.* 8, 624–629.
- Hashimoto, T., Nakamura, Y., Nakamura, F., Shirakura, T., Adachi, J., Goto, N., et al., 1994. Protein phylogeny gives a robust estimation for early divergences of eukaryotes, phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*. *Mol. Biol. Evol.* 11, 65–71.
- Henikoff, S., Henikoff, J.G., 1994. Position-based sequence weights. *J. Mol. Biol.* 243, 574–578.
- Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A., Steitz, T.A., 1992. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* 256, 1783–1790.
- Levis, R.W., Ganesan, R., Houtchens, K., Tolar, L.A., Sheen, F.M., 1993. Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* 75, 1083–1093.
- Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V., Cech, T.R., 1997. Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276, 561–567.
- Luan, D.D., Korman, M.H., Jakubczak, J.L., Eickbush, T.H., 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site, a mechanism for non-LTR retrotransposition. *Cell* 72, 595–605.
- Malik, H.S., Burke, W.D., Eickbush, T.H., 1999. The age of evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* 16, 793–805.
- Nakamura, T.M., Cech, T.R., 1998. Reversing time, origin of telomerase. *Cell* 92, 587–590.
- Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W.H., Lingner, J., et al., 1997. Telomerase catalytic subunit homologs from fission yeast and human. *Science* 277, 955–959.
- Oguchi, K., Liu, H., Tamura, K., Takahashi, H., 1999. Molecular cloning and characterization of AtTERT, a telomerase reverse transcriptase homolog in *Arabidopsis thaliana*. *FEBS Lett.* 457, 465–469.
- Pardue, M.L., Danilevskaya, O.N., Lowenhaupt, K., Slot, F., Traverse, K.L., 1996. *Drosophila* telomeres, new views on chromosome evolution. *Trends Genet.* 12, 48–52.
- Saitou, N., Nei, M., 1987. The neighbor-joining method, a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Smith, M.W., Aley, S.B., Sogin, M., Gillin, F.D., Evans, G.A., 1998. Sequence survey of the *Giardia lamblia* genome. *Mol. Biochem. Parasitol.* 95, 267–280.
- Sogin, M.L., Gunderson, J.H., Elwood, H.J., Alonso, R.A., Peattie, D.A., 1989. Phylogenetic meaning of the kingdom concept, an unusual ribosomal RNA from *Giardia lamblia*. *Science* 243, 75–77.
- Steitz, T.A., 1999. DNA polymerases, structural diversity and common mechanisms. *J. Biol. Chem.* 274, 17395–17398.
- Swofford, D.L., 1999. PAUP 4.0. Laboratory of Molecular Systematics, Smithsonian Institution, Washington, DC.
- The *C. elegans* Sequencing Consortium 1998. Genome sequence of the nematode *C. elegans* a platform for investigating biology. *Science* 282, 2012–2018.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface, flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Wang, D.Y., Kumar, S., Hedges, S.B., 1999. Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. *Proc. R. Soc. Lond. B. Biol. Sci.* 266, 163–171.
- Whitcomb, J.M., Hughes, S.H., 1992. Retroviral reverse transcription and integration, process and problems. *Annu. Rev. Cell Biol.* 8, 275–306.
- Wicky, C., Villeneuve, A.M., Lauper, N., Codourey, L., Tobler, H., Muller, F., 1996. Telomeric repeats (TTAGGC)_n are sufficient for chromosome capping function in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* 93, 8983–8988.
- Xiong, Y., Eickbush, T.H., 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 9, 3353–3362.