

Identification and characterization of sea squirt telomerase reverse transcriptase

Yang Li^a, Joel A. Yates^a, Julian J.-L. Chen^{a,b,*}

^a Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287, USA

^b School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA

Received 7 November 2006; received in revised form 15 May 2007; accepted 24 May 2007

Available online 5 June 2007

Abstract

Telomerase is essential for maintaining telomere length and chromosome stability in most eukaryotic organisms. The telomerase ribonucleoprotein complex consists of two essential components, the catalytic telomerase reverse transcriptase protein (TERT) and the intrinsic telomerase RNA. The sea squirts, as urochordates, occupy a key position in the phylogenetic tree of the chordates: they diverged from the other chordates just before the lineage of vertebrates, and thus provide special insight into the origin and evolution of vertebrate genes. Here, we report the cloning and characterization of TERT genes from two sea squirts, *Ciona intestinalis* and *Ciona savignyi*. The *C. intestinalis* TERT (*CinTERT*) gene encodes 907 amino acids and consists of 17 exons, which are similar to vertebrate TERT genes. The *C. savignyi* TERT (*CsaTERT*) gene encodes 843 amino acids, but surprisingly does not contain any introns. Both *Ciona* TERTs contain all of the reverse transcriptase (RT) motifs (1, 2, A, B, C, D, and E) that are typically present in telomerase and viral RTs. Interestingly, the alignment of *Ciona* and vertebrate TERT sequences reveals a previously unknown motif, named motif 3, located between motifs 2 and A. The *Ciona* TERT gene is expressed in all tissues analyzed except the brain and heart. This is the first report of the TERT gene in invertebrate chordates.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Telomere; Invertebrate; Chordate; TERT

1. Introduction

Telomeres are DNA–protein complexes that cap the ends of linear chromosomes to prevent end-to-end fusion or apoptosis (Blackburn, 2001; deLange, 2002). The telomeric DNA in vertebrates consists of tandem repeats of TTAGGG sequence, and associates with a number of telomere-binding proteins. In addition to their protective role in chromosome stability and genomic integrity, telomeres also play a critical role in cellular aging. In normal somatic cells, telomeres shorten progressively, which eventually leads to cell senescence or death. Immortal

cells, such as stem cells and tumor cells, utilize an enzyme called telomerase to maintain telomere length and cellular immortality.

Telomerase is a specialized reverse transcriptase (RT) that adds telomeric DNA repeats to chromosome ends (Greider and Blackburn, 1985, 1987). The telomerase enzyme consists of two essential components, the catalytic telomerase reverse transcriptase (TERT) protein, and the telomerase RNA (TR) that provides a template for telomeric DNA synthesis (Greider and Blackburn, 1989). The TERT protein was cloned initially from the ciliate *Euplotes aediculatus* and the yeast *Saccharomyces cerevisiae* (Lingner et al., 1997). Homologs of TERT have since been identified in a variety of eukaryotic species, including vertebrates — human, dog, mouse, rat, hamster, chicken, *Xenopus* and Fugu fish (Meyerson et al., 1997; Nakamura et al., 1997; Greenberg et al., 1998; Guo et al., 2001; Kuramoto et al., 2001; Wong et al., 2003; Delany and Daniels, 2004; Nasir et al., 2004; Yap et al., 2005); plants — rice, *Asparagales* and *Arabidopsis* (Fitzgerald et al., 1999; Oguchi et al., 1999; Heller-Uszynska et al., 2002; Sykorova et al., 2006); yeasts — *Schizosaccharomyces*

Abbreviations: aa, amino acids; bp, base pairs; cDNA, complementary DNA; TERT, telomerase reverse transcriptase; kb, kilobase(s); kDa, kilodalton (s); PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; RT, reverse transcriptase; RT-PCR, reverse transcriptase-polymerase chain reaction; TR, telomerase RNA; UTR, untranslated region.

* Corresponding author. Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287, USA. Tel.: +1 480 965 3650; fax: +1 480 965 2747.

E-mail address: JLChen@asu.edu (J.J.-L. Chen).

pombe and *Candida albicans* (Nakamura et al., 1997; Metz et al., 2001); Insects — *Bombyx mori* and *Tribolium castaneum* (Osanai et al., 2006); nematodes (Meier et al., 2006); and protozoa — *Oxytricha*, *Tetrahymena*, *Paramecium*, *Leishmania* and *Plasmodium* (Bryan et al., 1998; Collins and Gandhi, 1998; Takenaka et al., 2001; Figueiredo et al., 2005; Giardini et al., 2006).

The motifs conserved in TERT protein sequences can be grouped into three distinct domains: (1) an N-terminal domain, (2) a central-RT domain and (3) a C-terminal domain. The RT domain contains motifs that are universally conserved among all known TERT proteins and other RTs (Poch et al., 1989; Xiong and Eickbush, 1990; Nakamura and Cech, 1998). The RT motifs include seven highly conserved sequence motifs, named motifs 1, 2, A–E, that are essential for catalytic activity of telomerase (Lingner et al., 1997; Nakamura et al., 1997). Motifs in the amino-terminal and carboxyl-terminal domains of TERT are shared by many species, with the exception of *Giardia lamblia* and *Caenorhabditis elegans* (Malik et al., 2000). Each domain of TERT has distinct functional roles that are critical for recruitment to the telomeres, catalysis, and RNA binding (Autexier and Lue, 2006).

Sea squirts (or *Ciona*) belong to the subphylum Urochordata within the class Ascidiacea. The urochordates are close relatives to the vertebrate chordates, and considered to be the most basal branch of the phylum Chordata. *Ciona* has a notochord similar to vertebrates during its free-swimming tadpole life stage. However, when it progresses into its next surface-attached sessile stage, the notochord and the neural tube regress, and a simple neural complex is formed. Therefore, *Ciona* lack a backbone or vertebral column, and are considered to be invertebrate Chordates or protochordates. Because of their unique position in the evolutionary tree, *Ciona* have attracted much attention as a model organism in studying comparative biology and evolutionary origin of vertebrate organs such as the brain and heart. Here, we report the identification and characterization of TERT genes from two sea squirts, *C. intestinalis* and *C. savignyi*.

2. Materials and methods

2.1. Total RNA isolation

Total RNA was isolated from various tissues (brain, heart, intestine, stomach, ovary, muscle and testis) of *Ciona intestinalis* using TriZol reagent (Life Technologies, USA). Live animals of *C. intestinalis* were obtained from Marine Biological Laboratories (Woods Hole, Massachusetts, USA). In brief, 30–50 mg of fresh tissue from *C. intestinalis* was homogenized in 1 ml TriZol reagent and the resulting solution was mixed with 0.2 ml chloroform, followed by centrifugation at 12,000 ×g for 15 min at 4 °C. The aqueous phase containing total RNA was transferred to a new tube and extracted once or twice with acid-phenol/chloroform, pH 4.5 (Ambion, USA), followed by chloroform extraction and isopropanol precipitation. RNA concentrations were determined by optical density (OD₂₆₀) measurement using a Nanodrop ND-1000 spectrophotometer (NanoDrop Technologies, USA).

2.2. Identification and cloning of *Ciona* TERT genes

The *C. intestinalis* genome database (Assembly version 2.0 at a coverage of 11×) at <http://genome.jgi-psf.org> was searched using the TBLASTN algorithm with the vertebrate TERT protein sequences as query. A scaffold (scaffold #497) containing putative exons that are homologous to vertebrate TERT genes was identified. Gene-specific primers were designed to PCR amplify partial cDNA fragments of *Ciona* TERT. The exact 5' and 3'-ends of the full-length cDNA sequence of the *C. intestinalis* TERT gene (named *CinTERT*) were determined by 5'- and 3'-rapid amplification of cDNA ends (RACE) using a SMART RACE cDNA Amplification Kit (Clontech, USA) following manufacturer's instructions. To confirm the cDNA sequence, primers annealed to the 5' and 3'-UTR regions were used for RT-PCR, and the DNA product of RT-PCR was sequenced directly. The exon and intron boundaries of the *CinTERT* were determined by aligning the cDNA sequence to the genomic sequence of the scaffold #497. The cloned *CinTERT* cDNA sequence has been deposited in the GenBank database under the Accession No. EF077623.

The *Ciona savignyi* genome database (13× coverage) at <http://www.broad.mit.edu/annotation/Ciona> was searched using the *CinTERT* sequence as query. A single ORF that encodes a putative TERT sequence (843 amino acids) was found in the scaffold #427. The *C. savignyi* genomic DNA and cDNA samples were obtained from Dr. William Smith at U.C. Santa Barbara. The *CsaTERT* gene was PCR amplified from genomic DNA and cDNA samples using a pair of primers, a forward primer (5'-ACG ACC TGA ATA ATC TGT TGG ATT G-3') and a reverse primer (5'-ACC TTC TCA ACC TTG CAT GGC AAT CAC-3'), that anneal to the 5'-UTR and the 3'-end of the gene. The PCR products were gel purified and sequenced directly. The cloned *CsaTERT* cDNA sequence has been deposited in the GenBank database under the Accession No. EF514225.

2.3. Detection of *Ciona* TERT expression by RT-PCR

Approximately 1 µg of total RNA was reverse transcribed into complementary DNA (cDNA) using Thermoscript (Invitrogen, USA) and an oligo dT18 primer. The cDNA was then used as template in PCR reactions with Ex-Taq DNA polymerase (Takara Bio Inc. Japan) and specific primers. A pair of primers complementary to *CinTERT* exon 15 (5'-CAA CCA TGG GTG ATT GGG AAT CCA-3') and exon 17 (5'-GGA ACA TTC CAA CTG CCC TTT GTA G-3') was used to amplify the TERT cDNA from exons 15 to 17. Another pair of primers complementary to exon 5 (5'-TGG GCA GAG ATA GTC GCA TAC-3') and exon 15 (5'-TCG CAC CCA ACA TGG CAA GTT C-3') was used to amplify the TERT cDNA from exons 5 to 15 that covers the whole RT domain of the TERT gene. The primers were designed to anneal specifically to different exons to ensure that the PCR products were amplified from cDNA but not genomic DNA. The expected sizes of the two PCR products amplified from the TERT cDNA are 292 bp and 1.6 kb, respectively. As an internal control for RT-PCR, a 207 bp fragment of actin cDNA was amplified using a pair of

primers, a forward primer (5'-AGA GCT ACG AAC TTC CTG ACG GAC AG-3') and a reverse primer (5'-GAA CAT AGT GGA ACC TCC AGA GAG AAC-3'), complementary to exon 4 and exon 5, respectively. The PCR was carried out with 95 °C for 2 min, followed by 30–40 cycles of 95 °C for 15 s, 55 °C for 30 s, and 72 °C for 1–2 min, followed by a final elongation at 72 °C for 3 min. The more abundant actin cDNA was amplified with only 30 cycles of PCR, whereas the TERT cDNA was amplified with 40 or 45 cycles of PCR.

2.4. Sequence analysis of *Ciona* TERT

Multiple sequence alignment of TERT proteins was performed using ProbCons for the RT domain and C-terminal region. ProbCons is an improved tool for multiple sequence alignment of proteins using a combination of probabilistic modeling and consistency-based alignment techniques to achieve a higher accuracy (Do et al., 2005). ClustalX was used to align the N-terminal region in which the sequence and length vary significantly (Chenna et al., 2003). A phylogenetic tree was constructed from the aligned sequences of the RT domain using MEGA version 3.1 obtained from <http://www.megasoftware.net> (Kumar et al., 2004).

3. Results and discussion

3.1. Identification and characterization of the sea squirt TERT (*Ciona* TERT) genes

To find the TERT gene in the *C. intestinalis* genome database, we performed a TBLASTN search using the vertebrate TERT protein sequences as query. The search successfully identified a putative *Ciona* TERT sequence that encodes a conserved RT motif B (GIPQG). Gene prediction analysis of this sequence as well as its flanking sequences obtained from the genome database exhibited a number of small and scattered open reading frames (ORFs) separated presumably by introns. To obtain the full-length cDNA sequence of this gene, we carried out 5'- and 3'-RACE using a cDNA library prepared from *C. intestinalis* ovary mRNA and gene-specific primers that target the predicted exon sequences of the putative *Ciona* TERT gene.

The 5'- and 3'-RACE revealed a cDNA sequence of 3588 bp that contains a 5'-UTR of 402 bp, a 3'-UTR of 462 bp and an ORF of 2724 bp assuming that translation initiates at the first ATG codon (position 403–405). This ORF is predicted to encode a protein of 907 amino acids. On the basis of sequence similarity with other TERT proteins, we have named this protein gene the *C. intestinalis* TERT (*Cin*TERT). To characterize the promoter region of the TERT gene, we performed PCR to amplify the upstream flanking region of the gene. About 500 bp of the *Cin*-TERT promoter sequence was obtained. A putative TATA box, 5'-TATTAA-3', is located 36 bp upstream of the transcription initiation site (Fig. 1B, underlined). Based on the result of the 3'-RACE, polyadenylation of the *Cin*TERT mRNA occurs at the A triplet. A putative polyadenylation signal, 5'-AAAATAA-3', is located 20 bp upstream of the A triplet (Fig. 1C, underlined).

To predict more precisely the translation initiation site, the cDNA sequence was analyzed using a recently developed algorithm that predicts translation initiation sites with high sensitivity and specificity (Li and Leong, 2005). Using this algorithm, the same ATG codon at position 403–405 was predicted with the highest probability to be the translation initiation site (data not shown). Mapping of the cloned cDNA sequence to the genomic sequence from the database indicated that the *Cin*TERT is composed of 17 exons and 16 introns (Fig. 1A), similar to vertebrate TERT genes (Leem et al., 2002; Yap et al., 2005). All introns conformed to the GT/AG rule at the splicing junctions and have sizes ranging from 239 to 897 bp, similar to the sizes of introns in vertebrate TERT genes. From the transcription start site to the polyadenylation site, the genomic sequence of the *Cin*TERT gene spans 10.5 kb, smaller than the 41 kb human TERT gene (Leem et al., 2002) but larger than the 6.3 kb fugu gene (Yap et al., 2005). The *Cin*TERT gene was mapped to chromosome 8q based on a BLAST search of the *C. intestinalis* genomic and cDNA database (<http://ghost.zool.kyoto-u.ac.jp>) (Satou et al., 2005). Overall, the similarity in protein sequence and gene structure suggests that the *Cin*TERT identified here is an ortholog of vertebrate TERTs.

Two types of *C. intestinalis* have recently been reported (Suzuki et al., 2005; Nydam and Harrison, 2007). One type inhabits the Pacific Ocean, while the other type inhabits the northwest Atlantic Ocean. The *Cin*TERT cDNA cloned in this study was from an Atlantic *C. intestinalis* (see Materials and

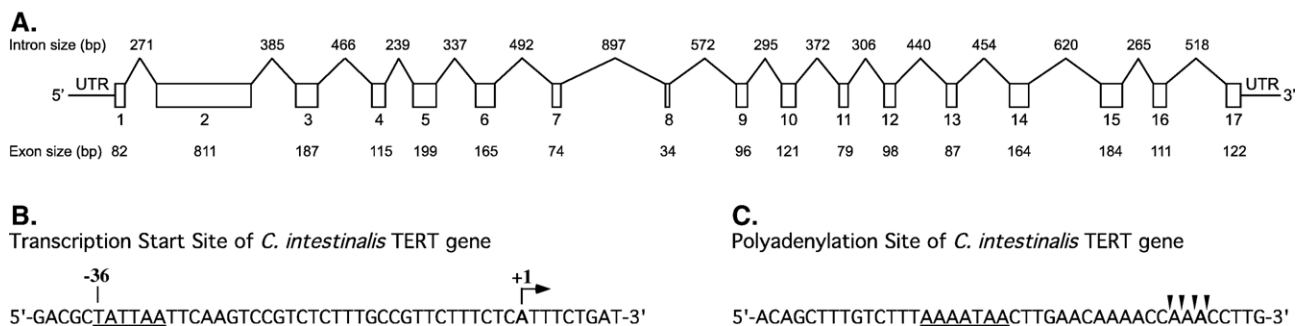


Fig. 1. Schematic of *C. intestinalis* TERT gene structure. (A) Splicing pattern of *C. intestinalis* TERT gene. Exons are numbered from 1 to 17, and shown as open boxes, whereas introns are shown as lines. The 5' and 3'-untranslated regions (UTRs) are shown. The sizes (bp) of each intron and exon are shown. (B) Transcriptional start site sequence of *Ciona* TERT gene. The transcription start site is shown in bold letters and indicated with +1. The putative TATA box, 36 bp upstream of the start site, is underlined. (C) Polyadenylation site sequence of *Ciona* TERT gene. Polyadenylation sites are shown in bold letters with the cleavage sites indicated by arrowheads. A putative polyadenylation signal is underlined.

Fig. 2. (A) Multiple amino acid sequence alignment of the N-terminal regions of TERT proteins. The alignment includes TERT sequences from human (*Homo sapiens* AF015950), dog (*Canis familiaris* AF80351), mouse (*Mus musculus* NM-009354), hamster (*Mesocricetus auratus* AF149012), *Xenopus* (*Xenopus laevis* AF212299), fugu fish (*Fugu rubripes* AY861384), sea urchin (*Strongylocentrotus purpuratus* EF611988), sea squirts (*C. intestinalis* and *C. savignyi*) and nematodes (*Caenorhabditis elegans* NP-492374, *C. remanei* DQ178631 and *C. briggsae* CAE66970). Shaded residues represent the conserved amino acids. The positions of amino acids are shown at both ends of each line. Hyphens indicate gaps introduced to optimize alignment. Identical residues are shown above the alignment. Chicken (*Gallus gallus* AY502592) TERT sequence is omitted from this alignment due to its unusual length variation in the N-terminal region. (B) Multiple amino acid sequence alignment of the central and C-terminal regions. Conserved residues in the T motif and RT motifs are shaded with specific colors and indicated above the alignment. Residues absolutely conserved in all RTs are indicated with red asterisks below the alignment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

D.

T

1 **2**

L F V FY T K Y W **RF PK G RP**

Human EE L LAKFLHWM -SVYVVE LRSFFYV TETTFQKNRLFFYKSVWSQLSGICIRQLHKLVRQLRE LSEA EVRQHREARPAL LTSRLRFIPKPDGLRPI VNM 636
Dog EE L LARFLVLDVGH IYVVKLLRSFFYV TETTFQKNRLFFYKSVWSQLSGICIRQLFNSVHLRE LSEA EVRHREARPAL LTSRLRFIPKPSGLRPI VNM 626
Mouse ER L L LAF LFWLM -DTYV VQ LRSFFYV TETTFQKNRLFFYKSVWSQLSGICIRQLHRLVRLE LSQEEVRHHQDTWLAMP ICALRRFIPKPNGLRPI VNM 626
Hamster ER L LAF LFWLM -DA YVVE LRSFFYV TETTFQKNRLFFYKSVWSRLCSICVRHHLERVLRLQELSQEEVRQGEAWPAMP ICALRRFIPKPSGLRPI VNM 630
Chicken EE L LAKFLYWL M-DSYV IELKLSFFYV TETTFQKNRLFFYKFIWYKGLONIGIRDFHFAKVLHRLALSS EEMEVIRQKYFP IASRLRFIPKMNGLRPFVRL 845
Xenopus EK L LAKFVFWLM -DTYV IQLKLSFFYV TETTFQKNRLFFYKSVWSKLLONIGIRLKHLEKVLKRLSSDELENMQQWKNV LVSRLRFFIPKTNGLRPI SKI 691
Fugu TQ L LGGFLAWLL -DGKVVGLVVRACFYATSMGQKNALIRFYRQEVAKLGLDLA FRSHISKGQMVLETPDQVAA LPK ---STIISRLRFIPKTDGMRPI TRV 583
C. intestinalis SS N L LRRFQWMI -EDYV FVVLSGSFFYV TETTFQKNRLFFYKHKMAEIVAYGVALL LKQNRWAVPPIADVTIRP ----YFSVRFCKPKVNGLRPI MRL 454
C. savignyi KDT L L LQFKWII -EEYLFV LIRGSFFYV TETTFQKNRLFFYKPKIKISEAVAYGVALL LKQKRWA FHSS EKHVTVP ---YFAVRFCKPLNGLRPI MKL 393

3 **A**

L I **FKV D** **P L**

Human DYVVGART ---FRREKRAERLTSRVKALFSV LNYERARRPGLLGASV LGLDDI HRAWRTFVLVRVAQD -PPPELYFVKVDVTGAYDTIPQDR LTEV IAS I 732
Dog SYVIMGART ---FHRDKKVQHLTSQ LTKLFSV LNYERARRPGLLGASMLGMDI HRAWRTFVLIRQAQ -PAPQLYFVKVDVTGAYDALIPQDR LVEIVAN 722
Mouse DYSMGTRA ---GLRRKQAQHTQRLKTLFSMLN YERTKHPLHMLSSVLGMDIYRTWRA FVLVRVALD -QTPRMYFVKADV TGA YDAIPQDK LVEIVANM 722
Hamster SY-MGTRA ---FDKGKQAQHTQCLKTLFSV LNYELTKHTNLLGASV LGLNDIYRTWRTFVLVRVRLD -PAPRMYFVKADV TGA YDAIPQDK LVEIVANM 725
Chicken SRVVEGQKLSKSEREKKIQRNYTQLKNLFSV LNYERTVNTS IIGSSVFGRDDIYRKWKEFVTKV FESGGEMPHFYFVKVDGSRADFIPHKHKLVEISQV 945
Xenopus SSTLSQQ -SKENQEKIHHFSSQIRNLFSV LNYEWNRLS IIGSSVFGMDIYKWKKFVLD FEKPQVEKLQFYFVKTDVGKAYDTIPHSK LVEISKV 790
Fugu I-----GADAKTRLYQSHVRDL DMLRACVCS TPSLLGS TVVGMDIHKV LSS IAPAQKE ---KPQPLYFVKMDVSGAYESIPHNK LIEVINQV 669
C. intestinalis R-----NADKMP ---DMVPFEV IASV LKLVCSKVPQATGFATQHHAH IYNGWKL FVLSLAGIP -GRKRL FVKMDIHRFCDSIPVVK LILIEVGL 540
C. savignyi R-----NPKVP -SSKMSFN IISV LK LALQRFPSLGF A I KDRSDIYCKWRQFL LK LKE INAKRWQLYFVKLDICRCFDSIPITKLRLRIVQDL 480

B **C** **D**

G PQG S LC G ME **R DD L T** **A FL** **YG** **K NF**

Human IQKPQ--NTYCVRRYAVVQKAAHGHRVAKFKSHVSTLTDLQ--PYMRQFVAHLQE--TSLRDAVVI EQSSSLNEASSGLFDVFLRFMCHHAVRIRGKSYV 840
Dog IRPQ-ESTYCVRRYAVVQRTARGHVRKAFKRHVSTFADLQ--PYMRQFVERLQE--TSLRDAVVI EQSSSLNEAGSSLFHLRLRVHNVHVRIRGKSYI 817
Mouse IRHS-ESTYCIQRQYAVVRDSQQGQVHKSFRQV TSLD LQ--PYMGQFLKHLQDSDASALRNSVIEQS ISMNESSSLDFDLFLHFRHSVVKIGDROYT 819
Hamster IRHP-DNSYCIGHQYAVVQRDQGIHKSFRQV TSLD LQ--PHMGQFLKHLQDSDASALRNSVIEQS LNEASSSLDFDLFLFRVNSVVKIGGROYV 822
Chicken LKPESTQVYGIWRWAVIMITPTGARKLYKRHVSTFEDFI--PDNMQFVSKLQE--RTSLRNAIVVEQCLTFNENSTLFTFLQMLHNNILEIGHRYVI 1041
Xenopus INPNA NEVYCIIRYAVITVDPTGRIKSKFRHVS LADVL--PNMKQFVSNQOE--KNL RNTI LVEQNLLNESSVKVLA FQOIRSHI RLKIDRYVM 886
Fugu LTPVLNEVFTIRRFKAIWADSEHGLKKA FIRQAD FLEANMGSINMKQELTSLQK--KGK LHSV LVEQIFSSDLEGKDALQFFTQILKGSV IQFGKKT Y 767
C. intestinalis LD RS-GEVD--DGFVGLVVRACFYATSMGQKNALIRFYRQEVAKLGLDLA FRSHISKGQMVLETPDQVAA LPK ---STIISRLRFIPKTDGMRPI TRV 583
C. savignyi LDRA-GTLD-----VSIVTKGKLKRICIHHCSQYIVDQLKFLFRDVTLSVLSKFFI 529

E **Region c-1** **Region c-2** **Region c-3**

W G **D Y** **D** **N Y**

Human QMFAHGLFPWCGLLLDTRT-----LEVQSDYSSYARTS-IRASLTFNRGFKAGNRMRRLFGVLR LKCHSLF L D LQVNS LQTVCNTNIYK I 1004
Dog QLPFAHGLFPWCGLLLDTRT-----LEVSCDYSSYARTS-IRASLTFSGQAGKPRNRMRRLFAVLR LKCCALF L D LQVNG IHTVYMNYYK I 995
Mouse QLPFAHGLFPWCGLLLDTRT-----LEVFCDYSSYARTS-IKTLSTFQSVFKAGKTRMRKL LSVLR LKCHGLF L D LQVNS LQTVCNTNIYK I 997
Hamster QLPFAHGLFPWCGLLLDTRT-----LEVLC DYTYGARTS-IKASLTFQRTFKAGNRMRRLFAVLR LKCHSLF L D LQVNS LQTVCNTNIYK I 1000
Chicken QLPDCLRLSWCGLLLDVQT-----LEVYCDYSSYARTS-IRSSLSFNSSRIAGKNMKCKLTA VLR LKCHPL L D LKINS LQTVALNIYK I 1218
Xenopus QLPFHGLFRWCGLLLDTRT-----LDVYDYSSYARTS-IRSSMTFCHSSAAGKYMKQKLIRVLR LKCHSLF L D LK VNS LRTVCNTNIYK I 1063
Fugu VLP LQDLFPWCGLLLDTRT-----LDIYKDYSSYADLS-IRYSLTLGCSGASHQMKRKLGMIRL LKCHALF L D LK TNS LEA IYKNIYK L 946
C. intestinalis L D QEQDTFPWFYGLYLPFC LNMGRKQLKQKGTNFFGGSVDLSAYQSTHCIRDFIVSQSP--GTNYNFPQIIRNLTRRTVICFPDQIPKGMIRYRNLYE I 783
C. savignyi L L NNEEPFHWFGYIFPTC LNMNKKQIKQTGTNFFGGISMDLSSYQSTRYRDFMSTRSS--VTIYRQCSIQKNMTRRVDHMLFDPQICTRKRIRFNLYE I 723

<--- Region c-2 **Region c-3**

L A **I** **F H** **L**

Human LL LQAYRFHACV LQLPFHQ-QVWKNPTFFLRV I SDTASLCYSILKAKNAGMSLGAKGAAGLPSEAVQWLCHOAFLKLTRHRTVYVPL LGS LR-----1097
Dog FL LQAYRFHACV LQLPFHQ-PVRKNPSFFLRV I ADTASCCYSLLKARNAGLSLGAKGASGLPSEAA RWLCHLHAFLLKLAHSGTYRCLLGA LQ-----1088
Mouse FL LQAYRFHACV IQLPFDQ-RVRKNLTFFLGI I SSQAQSCYAILKVKNPMGMLK--ASGSFPPEAAHWLYCQAFLLKLAHSHYIKCLLGLPLR-----1087
Hamster FL LQAYRFHACV LQLPFDQ-HVRKNPAFFLS I SN IASCCYSILKVKNAAGMLTAKAGSAGSFPPEAAHWLYCQAFLLKLAHSHVYIKCLLGLPLR-----1093
Chicken FL LQAYRFHACV LQLPFDQ-KVRNNPDPFLRI I SDTASCCYF LKAKNPGVSLGSKDASGMFPFEAAELCYHAFIVKLNSHNKYIKCLLKP LK-----1311
Xenopus FL LQAYRFHACV VQLPFDQ-RVMNNPPFFLT V I SDMA PCFYTTFAKAKND LTRGSKDVSCQNFVEAVQWLSYQAF LTLKLNHNKVLYKCLIGPLQ-----1156
Fugu LL LHALRHFHYCAQSLPFGQ-SVAKNPAYFL LMIWDMVEYTNYLIR LSNNG L I SGTSSQTGSQYQAEV LFLCFLSLV LSKHRRYLKOLLHLH-----1039
C. intestinalis AM LGA I QMVA--HQLHGEQPW I GNPWOFRSN I QMANAVRNIQRVFKS-----KHFLTRPLVSLVYKFAKFTKMKRHSN--NKAALPVLQMGINOC 872
C. savignyi ALLGA IHTVA--YDLH-RK-DIYKDL IJN I FNLVAHVRV I IORFKG-----KHFI SNISLVKWIIFKAEYV LK RHHY-YKKV I PSLHROV L I 810

Fig. 2 (continued).

gene duplication or RNA-mediated retrotransposition, the mechanistic basis for the origin of this intron-less TERT gene remains to be elucidated.

3.2. Comparison of *Ciona* and vertebrate TERT protein sequences

Because of the different degrees of sequence and length conservation in different domains of the TERT protein, a given sequence alignment algorithm sometimes fails to align the protein sequences in different regions with the same accuracy. Different domains of the TERT protein might have evolved at different rates along different lineages of species. For example, the sequences and length of the N-terminal and C-terminal domains are usually more variable and are thus more difficult to be aligned correctly. To achieve a more accurate sequence alignment to reveal important motifs, we employed a novel approach with which sequences of individual domains of the protein from closely related species were aligned separately.

The sequences of the N-terminal domain of TERTs from vertebrates, sea squirts, sea urchin and nematodes were aligned using the ClustalX program. The N-terminal domain of vertebrate TERT proteins contains five telomerase-specific motifs, named region v-I, v-II, v-III, v-IV and motif T (Kuramoto et al., 2001). Similar to the *Ciona* TERT, nematode TERTs also have a truncated N-terminal domain (Malik et al., 2000; Meier et al., 2006). By aligning the sequences of the N-terminal domain from sea squirts, nematodes and sea urchin TERTs, we found that both sea squirt and nematode TERTs lack region v-I but contain region v-II near the N-termini of the proteins (Fig. 2A).

Region v-I has been previously defined as motif GQ (Xia et al., 2000; Bosoy et al., 2003), and is conserved in most species with the exception of insects (Osanai et al., 2006) and nematodes (Malik et al., 2000). The GQ motif has been shown to be important for repeat addition processivity in human and yeast TERTs (Moriarty et al., 2004; Lue, 2005). It also shows a strong affinity to the single-stranded telomeric DNA primer (Jacobs et al., 2006). Within the region v-I/GQ, there is another motif called the “dissociate activities of telomerase” (DAT) motif that has been shown to be required for telomerase activity *in vivo* (Armbruster et al., 2001). It would be interesting to know how *Ciona* telomerase functions without these important motifs in the region v-I.

While lacking region v-I, the *Ciona* TERT contains most residues conserved in region v-II, v-III and v-IV (Fig. 2A). Region v-III includes a previously identified motif CP that is

conserved in ciliates (Bryan et al., 1998). Region v-IV has been previously defined as motif QFP, and is also conserved in *Ciona* TERT (Fig. 2A). The nematode TERTs are much more divergent in sequence and length, and contain only parts of the region v-III, v-IV and motif T (data not shown) (Meier et al., 2006).

The sequence alignment of chordate TERTs, in addition to identifying chordate-specific motifs, also refines the boundaries of known RT motifs of TERT proteins. The universal RT motifs (1, 2, A, B, C, D and E) were originally identified based on RT sequences in viruses or retrotransposons (Poch et al., 1989; Xiong and Eickbush, 1990), and later on TERT sequences (Lingner et al., 1997; Nakamura et al., 1997). However, the boundaries of these motifs in TERT are not well defined and can vary between different groups of species. With the appropriate evolutionary distance between *Ciona* and human, the alignment of chordate TERT sequences allows for a better boundary definition of the conserved motifs in human TERT. In this study, the motif boundaries are defined by the first and last conserved residues within the motif (Fig. 2).

To obtain a more accurate alignment of the sequences from the RT and C-terminal domains, we employed a novel sequence alignment algorithm, ProbCons, that is based on probabilistic consistency and has been shown to have significant improvement over other methods for multiple sequence alignment of proteins (Do et al., 2005). The central domain of TERT contains all previously known RT motifs (i.e. 1, 2, A, B, C, D, and E) (Fig. 2B). In *Ciona* TERTs, the distance between motifs A and B is significantly shorter than other TERTs. *Ciona* TERTs contain all invariant amino acids that are conserved in RTs, including a lysine (K) in motif 1, an arginine (R) in motif 2, an aspartic acid (D) in motif A, a glutamine (Q) and a glycine (G) in motif B, two aspartic acids (DD) in motif C, a lysine (K) in motif D and a glycine (G) in motif E (Fig. 2B). These invariant residues are crucial for enzyme activity (Lingner et al., 1997). In addition to the previously known RT motifs, we have also identified a novel motif, named motif 3, located between motifs 2 and A (Fig. 2B). This motif 3 is also present in other groups of species outside of chordates such as sea urchin, plants and ciliates (data not shown). The sequences of motif 3 from human and *Ciona* share 28.6% identity and 45.7% similarity in sequence. This sequence conservation suggests that motif 3 plays an important role in telomerase function. Mutations of the conserved residues in the motif 3 altered activity or processivity of human telomerase (Xie and Chen, unpublished data).

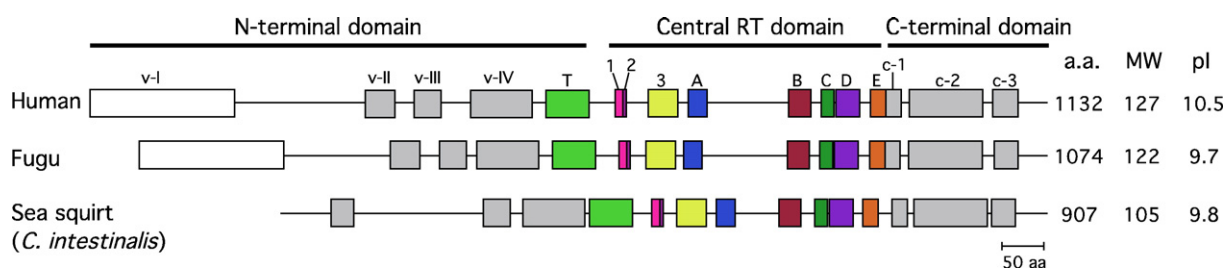


Fig. 3. Schematic representation of TERT protein motifs. The structures of human (*Homo sapiens*), fugu (*Fugu rubripes*) and Sea squirt (*Ciona intestinalis*) TERT proteins are shown. Motifs are shown as boxes. Three domains, the N-terminal, the central-RT and the C-terminal domains are indicated above the schematic of the motif structures. The length (a.a.), predicted molecular weight (MW, in kDa), and isoelectric point (pI) are indicated at the right.

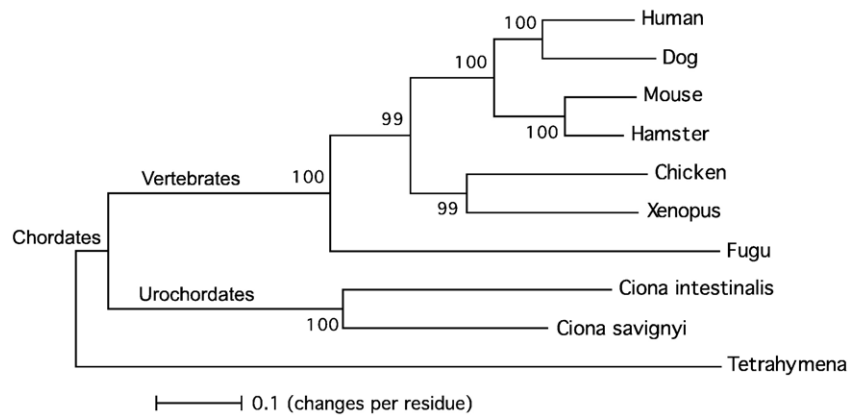


Fig. 4. Phylogenetic tree based on the TERT protein sequences. The tree was derived using the neighbor-joining method from the aligned sequences of motif T and the RT domain (1, 2 and A–E) of ten TERTs. *Tetrahymena* TERT sequence is included as an outgroup to root the tree. The number next to each node indicates a value as a percentage of 1000 bootstrap replicates.

The sequence alignment of the C-terminal region revealed a number of insertions and deletions between conserved motifs in the *Ciona* TERT (Fig. 2B). These gaps in the alignment allow better definition of the boundaries of each conserved region. We have thus redefined three conserved chordate motifs, named region c-1, c-2, and c-3, in the C-terminal region (Fig. 2B). These three regions contain most of the conserved residues that were previously shown to be critical for human telomerase activity *in vivo* (Banik et al., 2002).

Overall, motifs in the N-terminal region are the least conserved, while motifs in the central region are the most conserved (Fig. 3). Many of these motifs, such as the telomerase-specific “T” motif, are conserved only in telomerase RT, and thus serve as sequence signatures to distinguish telomerase RT from other RTs (Nakamura et al., 1997). The presence of these telomerase-specific signatures supports that the *Ciona* TERTs identified in this study are orthologs of vertebrate TERT. Furthermore, the phylogenetic relationship of the chordate TERT sequences reveals the evolutionary position of the identified *Ciona* TERT (Fig. 4), and is consistent with the phylogeny of eukaryotes based on sequences of other molecules, such as hemoglobin (Czelusniak et al., 1990).

3.3. Expression pattern of *Ciona* TERT mRNA

In humans, telomerase activity is usually regulated at the transcriptional level of the TERT gene. The expression of human TERT is restricted to those tissues of high turnover such as testis, intestine and bone marrow (Kolquist et al., 1998; Ulaner et al., 1998). To examine the pattern of TERT expression in *Ciona*, we examined the presence of the *CinTERT* transcript in a variety of tissues by RT-PCR. Two primer pairs were used to amplify the cDNA between the exons 5 and 15 or between the exons 15 and 17. PCR products of expected sizes were observed for both TERT and actin genes (Fig. 5). Alternative splicing products of TERT mRNA were not observed (data not shown). *Ciona* TERT mRNAs were detected in the intestine, stomach, ovary, muscle and testis (Fig. 5, lanes 3–7). The pattern is largely consistent with the known requirement of telomerase in tissues of high turnover, except the stomach and muscle (Fig. 5, lanes 4 and 6). In

vertebrates, TERT is not expressed in stomach and muscle tissues (Yap et al., 2005). Since TERT is normally expressed in the tissues with significant regenerative capacity such as liver, one explanation would be that *Ciona* stomach and muscle tissue possess higher levels of regenerative capacity than that of human.

In normal human somatic cells, the TERT mRNA levels and telomerase activity are generally undetectable (Kim et al., 1994; Meyerson et al., 1997; Nakamura et al., 1997; Kolquist et al., 1998; Ulaner et al., 1998). In a teleost fish, *Fugu rubripes*, TERT is down-regulated in the brain and heart (Yap et al., 2005). Similarly, *Ciona* TERT mRNA also is not detectable in the brain and heart (Fig. 5, lanes 1 and 2). The specific down-regulation of the TERT gene in brain and heart tissue might make sea squirts an attractive model system for studying telomerase regulation during cell differentiation and organ development. However, in some organisms such as mouse, telomerase activity is not well regulated in adult tissues, and does not correlate with the expression of TERT mRNA (Martin-Rivera et al., 1998). To fully understand the regulation of telomerase in *Ciona*, both TERT protein level, telomerase RNA expression and telomerase activity will need to be measured

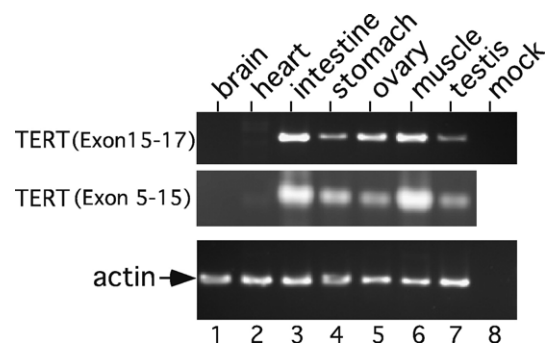


Fig. 5. Expression of *Ciona* TERT in various tissues. *Ciona* total RNA was reverse transcribed using oligo (dT18) primer to generate cDNA library. Specific *Ciona* TERT cDNA fragments were amplified by PCR using primer pairs annealing to either exons 5 and 15, or exons 15 and 17, respectively. Cytoplasmic actin gene was amplified as a control for the quality of cDNA. Mock: reverse transcriptase enzyme omitted.

from various tissues. This will then provide a foundation to a better understanding of the origin and evolution of telomerase regulation in vertebrates.

In conclusion, this work describes the identification and characterization of TERT protein from invertebrate chordates, *C. intestinalis* and *C. savignyi*. The *Ciona* telomerase provides an attractive system with an ideal evolutionary distance that is close enough to humans to identify commonality, but divergent enough to reveal conserved and critical features of the gene. To thoroughly study the biochemistry of *Ciona* telomerase, we would need to identify the telomerase RNA in *Ciona*. It would then allow us to reconstitute active *Ciona* telomerase enzyme for functional analysis. The cloning of *Ciona* TERT therefore offers a biochemical and structural basis for these future studies on *Ciona* telomerase.

Acknowledgements

We are grateful for Guoliang Li's assistance on predicting the translation initiation site of the *C. intestinalis* TERT gene. We also thank Dr. William Smith (U.C. Santa Barbara Ascidian Stock Center) for providing genomic DNA and cDNA of *C. savignyi*. We are grateful to Drs. Mary Armanios and Tien-Hsien Chang, and lab members for critical reading of and helpful comments on the manuscript. This work was supported by an NSF CAREER Award to JLC.

References

- Armbruster, B.N., Banik, S.S., Guo, C., Smith, A.C., Counter, C.M., 2001. N-terminal domains of the human telomerase catalytic subunit required for enzyme activity in vivo. *Mol. Cell. Biol.* 21, 7775–7786.
- Autexier, C., Lue, N.F., 2006. The structure and function of telomerase reverse transcriptase. *Annu. Rev. Biochem.* 75, 493–517.
- Banik, S.S., et al., 2002. C-terminal regions of the human telomerase catalytic subunit essential for in vivo enzyme activity. *Mol. Cell. Biol.* 22, 6234–6246.
- Blackburn, E.H., 2001. Switching and signaling at the telomere. *Cell* 106, 661–673.
- Bosoy, D., Peng, Y., Mian, I.S., Lue, N.F., 2003. Conserved N-terminal motifs of telomerase reverse transcriptase required for ribonucleoprotein assembly in vivo. *J. Biol. Chem.* 278, 3882–3890.
- 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. U. S. A.* 95, 8479–8484.
- Chenna, R., et al., 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* 31, 3497–3500.
- Collins, K., Gandhi, L., 1998. The reverse transcriptase component of the *Tetrahymena telomerase* ribonucleoprotein complex. *Proc. Natl. Acad. Sci. U. S. A.* 95, 8485–8490.
- Czelusniak, J., Goodman, M., Moncrief, N.D., Kehoe, S.M., 1990. Maximum parsimony approach to construction of evolutionary trees from aligned homologous sequences. *Methods Enzymol.* 183, 601–615.
- deLange, T., 2002. Protection of mammalian telomeres. *Oncogene* 21, 532–540.
- Dehal, P., et al., 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298, 2157–2167.
- Delany, M.E., Daniels, L.M., 2004. The chicken telomerase reverse transcriptase (chTERT): molecular and cytogenetic characterization with a comparative analysis. *Gene* 339, 61–69.
- Do, C.B., Mahabhashyam, M.S., Brudno, M., Batzoglou, S., 2005. ProbCons: probabilistic consistency-based multiple sequence alignment. *Genome Res.* 15, 330–340.
- Figueiredo, L.M., Rocha, E.P., Mancio-Silva, L., Prevost, C., Hernandez-Verdun, D., Scherf, A., 2005. The unusually large *Plasmodium* telomerase reverse-transcriptase localizes in a discrete compartment associated with the nucleolus. *Nucleic Acids Res.* 33, 1111–1122.
- Fitzgerald, M.S., Riha, K., Gao, F., Ren, S., McKnight, T.D., Shippen, D.E., 1999. Disruption of the telomerase catalytic subunit gene from arabidopsis inactivates telomerase and leads to a slow loss of telomeric DNA. *Proc. Natl. Acad. Sci. U. S. A.* 96, 14813–14818.
- Giardini, M.A., et al., 2006. The putative telomerase reverse transcriptase component of *Leishmania amazonensis*: gene cloning and characterization. *Parasitol. Res.* 98, 447–454.
- Greenberg, R.A., Allsopp, R.C., Chin, L., Morin, G.B., DePinho, R.A., 1998. Expression of mouse telomerase reverse transcriptase during development, differentiation and proliferation. *Oncogene* 16, 1723–1730.
- Greider, C.W., Blackburn, E.H., 1985. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* 43, 405–413.
- Greider, C.W., Blackburn, E.H., 1987. The telomere terminal transferase of *Tetrahymena* is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* 51, 887–898.
- Greider, C.W., Blackburn, E.H., 1989. A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. *Nature* 337, 331–337.
- Guo, W., Okamoto, M., Lee, Y.M., Baluda, M.A., Park, N.H., 2001. Enhanced activity of cloned hamster TERT gene promoter in transformed cells. *Biochim. Biophys. Acta* 1517, 398–409.
- Heller-Uszynska, K., Schnippenkoetter, W., Kilian, A., 2002. Cloning and characterization of rice (*Oryza sativa* L) telomerase reverse transcriptase, which reveals complex splicing patterns. *Plant J.* 31, 75–86.
- Jacobs, S.A., Podell, E.R., Cech, T.R., 2006. Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. *Nat. Struct. Mol. Biol.* 13, 218–225.
- Kim, N.W., et al., 1994. Specific association of human telomerase activity with immortal cells and cancer. *Science* 266, 2011–2015.
- Kolquist, K.A., et al., 1998. Expression of TERT in early premalignant lesions and a subset of cells in normal tissues. *Nat. Genet.* 19, 182–186.
- Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* 5, 150–163.
- Kuramoto, M., Ohsumi, K., Kishimoto, T., Ishikawa, F., 2001. Identification and analyses of the *Xenopus* TERT gene that encodes the catalytic subunit of telomerase. *Gene* 277, 101–110.
- Leem, S.H., et al., 2002. The human telomerase gene: complete genomic sequence and analysis of tandem repeat polymorphisms in intronic regions. *Oncogene* 21, 769–777.
- Li, G.L., Leong, T.Y., 2005. Feature selection for the prediction of translation initiation sites. *Geno. Prot. Bioinfo.* 3, 73–83.
- 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.
- Lue, N.F., 2005. A physical and functional constituent of telomerase anchor site. *J. Biol. Chem.* 280, 26586–26591.
- Malik, H.S., Burke, W.D., Eickbush, T.H., 2000. Putative telomerase catalytic subunits from *Giardia lamblia* and *Caenorhabditis elegans*. *Gene* 251, 101–108.
- Martin-Rivera, L., Herrera, E., Albar, J.P., Blasco, M.A., 1998. Expression of mouse telomerase catalytic subunit in embryos and adult tissues. *Proc Natl Acad Sci U S A* 95, 10471–10476.
- Meier, B., et al., 2006. trt-1 is the *Caenorhabditis elegans* catalytic subunit of telomerase. *PLoS Genet.* 2, e18.
- Metz, A.M., Love, R.A., Strobel, G.A., Long, D.M., 2001. Two telomerase reverse transcriptases (TERTs) expressed in *Candida albicans*. *Biotechnol. Appl. Biochem.* 34, 47–54.
- Meyerson, M., et al., 1997. hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization. *Cell* 90, 785–795.
- Moriarty, T.J., Marie-Egyptienne, D.T., Autexier, C., 2004. Functional organization of repeat addition processivity and DNA synthesis determinants in the human telomerase multimer. *Mol. Cell. Biol.* 24, 3720–3733.

- Nakamura, T.M., Cech, T.R., 1998. Reversing time: origin of telomerase. *Cell* 92, 587–590.
- Nakamura, T.M., et al., 1997. Telomerase catalytic subunit homologs from fission yeast and human. *Science* 277, 955–959.
- Nasir, L., et al., 2004. Isolation and expression of the reverse transcriptase component of the *Canis familiaris* telomerase ribonucleoprotein (dog-TERT). *Gene* 336, 105–113.
- Nydam, M.L., Harrison, R.G., 2007. Genealogical relationships within and among shallow-water *Ciona* species (Ascidacea). *Mar Biol.* 151, 1839–1847.
- 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.
- Osanai, M., Kojima, K.K., Futahashi, R., Yaguchi, S., Fujiwara, H., 2006. Identification and characterization of the telomerase reverse transcriptase of *Bombyx mori* (silkworm) and *Tribolium castaneum* (flour beetle). *Gene* 376, 281–289.
- Poch, O., Sauvaget, I., Delarue, M., Tordo, N., 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* 8, 3867–3874.
- Satou, Y., Kawashima, T., Shoguchi, E., Nakayama, A., Satoh, N., 2005. An integrated database of the ascidian, *Ciona intestinalis*: towards functional genomics. *Zool. Sci.* 22, 837–843.
- Suzuki, M.M., Nishikawa, T., Bird, A., 2005. Genomic approaches reveal unexpected genetic divergence within *Ciona intestinalis*. *J. Mol. Evol.* 61, 627–635.
- Sykorova, E., Leitch, A.R., Fajkus, J., 2006. Asparagales telomerases which synthesize the human type of telomeres. *Plant Mol. Biol.* 60, 633–646.
- Takenaka, Y., Matsuura, T., Haga, N., Mitsui, Y., 2001. Expression of telomerase reverse transcriptase and telomere elongation during sexual maturation in *Paramecium caudatum*. *Gene* 264, 153–161.
- Ulaner, G.A., Hu, J.F., Vu, T.H., Giudice, L.C., Hoffman, A.R., 1998. Telomerase activity in human development is regulated by human telomerase reverse transcriptase (hTERT) transcription and by alternate splicing of hTERT transcripts. *Cancer Res.* 58, 4168–4172.
- Vinson, J.P., et al., 2005. Assembly of polymorphic genomes: algorithms and application to *Ciona savignyi*. *Genome Res.* 15, 1127–1135.
- Wong, S.C., Ong, L.L., Er, C.P., Gao, S., Yu, H., So, J.B., 2003. Cloning of rat telomerase catalytic subunit functional domains, reconstitution of telomerase activity and enzymatic profile of pig and chicken tissues. *Life Sci.* 73, 2749–2760.
- Xia, J., Peng, Y., Mian, I.S., Lue, N.F., 2000. Identification of functionally important domains in the N-terminal region of telomerase reverse transcriptase. *Mol. Cell. Biol.* 20, 5196–5207.
- Xiong, Y., Eickbush, T.H., 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 9, 3353–3362.
- Yap, W.H., Yeoh, E., Brenner, S., Venkatesh, B., 2005. Cloning and expression of the reverse transcriptase component of pufferfish (*Fugu rubripes*) telomerase. *Gene* 353, 207–217.