Alignment of cDNA sequence with genome:

- Red: cDNA sequence without match in WGS trace depository (BlastN, E=10)
- Red + Light Red: cDNA sequence without high quality match in assembled genome (BlastN, E=0.01)
- Alignment of cDNA and genomic sequence with frequent gaps and mismatches

Low complexity regions

GC content

- >85%
- >75%
- >70%
- >65%
Figure S1  Platypus TERT cDNA clone.

(a) (a) Cloning strategy. A schematic representation of the assembled OanTERT cDNA sequence with the 16 regions predicted to be encoded by separate exons is shown at the top. Nineteen different types of RT-PCR clones obtained from platypus mRNA using 15 distinct primer pairs (p1-p15) are depicted below. Cloning primers are listed in Table S1c (Additional file 1). Forty-four clones were sequenced and the number of clones of each type and the source tissue are shown on the right of every clone schema. Continuous consensus sequence was assembled by DNA Baser (Heracle Biosoft S.R.L., Pitești, Romania). The full-length TERT sequence is based on six overlapping DNA segments (without primer matching sequences) shown in purple ink. Ten types of the clones were alternatively spliced. Insertion of the intronic sequence was detected from alternative splicing are indicated. The clones which introduced sequence and deletions resulting from alternative splicing are indicated. Continuous consensus sequence was assembled by DNA Baser (Heracle Biosoft S.R.L., Pitești, Romania). The full-length TERT sequence is based on six overlapping DNA segments (without primer matching sequences) shown in purple ink. Ten types of the clones were alternatively spliced. Insertion of the intronic sequence was detected from alternative splicing are indicated.
Figure S1  Platypus TERT cDNA clone - continued.

(b) The sequence comparison of the genomic contig [GenBank:NW_001794359.1] with the platypus TERT cDNA [GenBank:JF441071]. The major areas of misalignments and sequence gaps in genomic sequence are indicated. The low sequence complexity regions in cDNA determined by BlastN build in filter and the areas of cDNA with high GC content determined by GC-Profile program (http://tubic.tju.edu.cn/GC-Profile/) are indicated below the cDNA schematic representation. The genomic assembly also contains a duplication of the exons 13 and 14. Because the corresponding areas of the cDNA do not have any abnormalities, the duplication in the genomic assembly is likely the assembling error.

(c) The assembled composite sequence of OanTERT [GenBank:JF441071]. Start and stop codons are indicated by white letters on the blue background. The boundaries of regions predicted to be encoded by 16 different exons are shown (the boundary nucleotides are on the green background) as well as two alternative splice donors and one alternative splice acceptor (the boundary nucleotide of the alternatively spliced exon are on green background with two neighboring nucleotides creating the beginning or end of the alternatively spliced intron on the yellow background).

(d) The region of genomic sequence [GenBank:NW_001794359.1] flanking at 3’ side the sequence matching OanTERT cDNA. The end of the cDNA sequence and two closest possible polyadenylation signals are shown.

(e) The region of genomic sequence [GenBank:NW_001794359.1] flanking at 5’ side the sequence matching OanTERT cDNA. The beginning of cDNA, ATG start codon, and the first upstream in-frame stop codon are indicated. The platypus genomic sequence does not contain any ATG codon between the start codon and the closest upstream in-frame stop codon supporting the correct determination of the ORF start codon. The encoded N-terminal amino acid sequence of the platypus TERT (MASAAPFFAVHA sequence preceding the TEN (telomerase essential N-terminal) domain) has the same size as that of human and mouse TERT.