

De novo assembly and characterization of the Chinese three-keeled pond turtle (*Mauremys reevesii*) transcriptome: presence of longevity-related genes

Huazong Yin, Liuwang Nie, Feifei Zhao, Huaxing Zhou, Haifeng Li, Xianmei Dong, Huanhuan Zhang, Yuqin Wang, Qiong Shi and Jun Li

College of Life Science, Anhui Normal University, Provincial Key Lab of the Conservation and Exploitation Research of Biological Resources in Anhui, Wuhu, Anhui, China

ABSTRACT

Mauremys reevesii (Geoemydidae) is one of the most common and widespread semi-aquatic turtles in East Asia. The unusually long lifespan of some individuals makes this turtle species a potentially useful model organism for studying the molecular basis of longevity. In this study, pooled total RNA extracted from liver, spleen and skeletal-muscle of three adult individuals were sequenced using Illumina Hiseq 2500 platform. A set of telomere-related genes were found in the transcriptome, including *tert*, *tep1*, and six shelterin complex proteins coding genes (*trf1*, *trf2*, *tpp1*, *pot1*, *tin2* and *rap1*). These genes products protect chromosome ends from deterioration and therefore significantly contribute to turtle longevity. The transcriptome data generated in this study provides a comprehensive reference for future molecular studies in the turtle.

Subjects Genetics, Genomics, Zoology

Keywords *Mauremys reevesii*, *de novo* assembly, RNA-seq, Telomere, Telomerase activity, Shelterin complex, Longevity

Submitted 14 December 2015

Accepted 1 May 2016

Published 24 May 2016

Corresponding author

Liuwang Nie,
lwnie@mail.ahnu.edu.cn

Academic editor

Abhishek Kumar

Additional Information and
Declarations can be found on
page 11

DOI 10.7717/peerj.2062

© Copyright
2016 Yin et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

INTRODUCTION

Longevity has always been a trait of great interest to researchers, and numerous studies have been performed on humans and model organisms to better understand its molecular mechanisms, such as the maintenance of telomeric structure. Synthesized by telomerase (*Blackburn, Greider & Szostak, 2006; Bodnar et al., 1998*), telomeres are specialized structures at the ends of eukaryotic chromosomes that help to maintain genome integrity by protecting chromosomes from rearrangements or fusion to each other (*McClintock, 1939; Muller, 1938*). Introducing telomerase into normal human retinal pigment epithelial cells and foreskin fibroblasts significantly extends the lifespan of the cells (*Bodnar et al., 1998*). Several protein complexes have also been implicated in longevity. For example, the shelterin complex specifically recognizes and binds to telomeric DNA, preventing the chromosome end from being detected as a DNA double-strand break (*Palm & de Lange, 2008*). Research on the naked mole rat (*Heterocephalus glaber*) suggests that shelterin complex-encoding genes are related to the species' longevity (*Kim et al., 2011*). Synthesized telomeres are shaped and safeguarded by the shelterin complex, an essential component

of telomere function (De Lange, 2005) that consists of six proteins: TRF1, TRF2, POT1, RAP1, TIN2, and TPP1 (Bilaud et al., 1997; Broccoli et al., 1997; Chong et al., 1995; De Lange, 2005; Houghtaling et al., 2004; Kim, Kaminker & Campisi, 1999; Li, Oestreich & De Lange, 2000; Liu et al., 2004; Xin, Liu & Songyang, 2008; Ye et al., 2004; Zhong et al., 1992).

Turtles are an ideal model organism for research on the molecular basis of longevity. They are the most morphologically distinct order to have originated from the late Permian and early Triassic period (about 220 million years ago) (Li et al., 2008). In addition to their characteristic shell, turtles also have many remarkable physiological traits, such as anoxia (Lutz, Prentice & Milton, 2003) and cold tolerance (Packard et al., 1997), temperature-determined sex differentiation (De Souza & Vogt, 1994; Mrosovsky, Dutton & Whitmore, 1984), and of particular note, a long lifespan. Many individuals have been recorded as living more than 100 years (Gibbons, 1987; Shaffer et al., 2013). For example, the *Geochelone gigantea* known as Marion's tortoise lived for more than 150 years (Schmidt & Inger, 1957), and Lonesome George, a *Geochelone nigra*, was reported to have lived more than 100 years.

In addition, previous reports have described many physical characters that are associated with turtle longevity, such as being reproductively active at very advanced ages and negligible functional impairment with age (Miller, 2001). Some authors have linked longevity in turtles to enhanced mechanisms of reoxygenation for surviving brain anoxia (Lutz, Prentice & Milton, 2003). A study on European freshwater turtles (*Emys orbicularis*) found that telomere length was only shortened negligibly in adults compared to circulating embryonic blood cells (Girondot & Garcia, 1999). However, currently, we do not know whether turtle longevity is associated with telomerase activity or the insulin/IGF-1 signaling pathway.

Mauremys reevesii (Geoemydidae) is widespread and common in China, the Korean Peninsula, and Japan (Van Dijk et al., 2014). Due to its longevity, it has substantial cultural significance in China as an auspicious omen. In this study, total RNA extracted from liver, spleen, and skeletal muscle of three adult females were used to generate a pooled cDNA library, which was subsequently sequenced on an Illumina HiSeq 2500 platform. Next generation sequencing (NGS) technologies have been broadly applied in genome and transcriptome sequencing (Reis-Filho, 2009), due to their greater sensitivity, which supplies accurate results that can detect previously unknown and/or rare genes. The two primary aims of our study were as follows: (1) to better understand the molecular mechanism behind the long lifespan of turtles by identifying longevity-associated genes, and (2) to generate transcriptome data as a useful resource for future studies of turtle longevity and other traits.

MATERIALS & METHODS

Ethical approval

Procedures involving animals and their care were approved by the Animal Care and Use Committee of Anhui Normal University under approval number #20140111.

Sample collection and RNA extraction

Three adult female turtles were collected from our plant at Wuhu, Anhui, China. The liver, spleen, and skeletal muscle were collected and dissected. Tissue samples were stored at

–80 °C. Total RNA was extracted from the tissues separately using a Trizol kit (Invitrogen, CA, USA) according to the manufacturer’s protocol. Extracted RNA was quantified with Nanodrop (Thermo, CA, USA) and the integrity and size distribution were checked with agarose gel electrophoresis. High-quality RNA from all three tissues was pooled for cDNA synthesis and sequencing.

cDNA library construction and sequencing

Two micrograms of pooled total RNA was used for cDNA library construction using TruSeq® RNA LT Sample Prep Kit v2 (Illumina, CA, USA) according to the manufacturer’s protocol. We then prepared the synthesized cDNA for sequencing library construction by performing end-repair, 3’-end adenylation, as well as adapter-ligation and enrichment. Sequencing was performed using an Illumina Hiseq 2500 platform (Quail *et al.*, 2008) at Genergy Bio-technology Co., Ltd. (Shanghai, China).

Sequence data processing and *de novo* assembly

The raw reads generated by the Illumina sequencer were saved as fastq format files. Adapter sequences were trimmed and low-quality reads were removed from the raw reads using Trim Galore (version 0.3.5) software. FastQC (version 0.10.1) was used to check the quality of pretreated data; reads that achieved a high Phred score (>28) were used for the assembly. We used Trinity (Grabherr *et al.*, 2011), with a k value = 25, to perform *de novo* assembly on the trimmed and quality-checked reads. Sequence data were partitioned into many individual de Bruijn graphs, representing the transcriptional complexity at a given gene or locus. Each graph was processed independently to extract full-length splicing isoforms and to tease apart transcripts derived from paralogous genes. The final Trinity output was analyzed. Gene expression level was calculated using RSEM software (version 1.2.3) (Li & Dewey, 2011).

Blast against turtle’s reference proteomes

We used Blastx to query all *M. reevesii* transcripts to the proteomes of the green sea turtle, the western painted turtle, and the soft shell turtle, which were downloaded from the GenBank. The e-value cutoff was set to $1 \times e^{-10}$ and the maximum target number was set to 20. The top results were selected as the annotation of the gene (termed as “unique protein”). A Venn diagram of the homologous genes across the three turtle proteomes was generated with VENNY (Oliveros, 2007).

Functional annotation

Sense and component strands of potential protein coding sequences (CDS) were predicted using Transdecoder in Trinity software v2.0.2 package (<http://transdecoder.github.io/>), based on a Markov model with default parameters. CDS were then translated into amino acid sequences with reference to a standard codon table. We used Blastp to search the Swissprot/Uniprot database (Balakrishnan *et al.*, 2005) with our potential protein sequences as queries. The e-value cutoff was set to $1 \times e^{-3}$. A gene name was assigned to each contig based on the top Blastp hit.

Gene ontology (GO) analyses (Ashburner *et al.*, 2000) on all predicted protein sequences were conducted using InterProScan, set to default parameters (Zdobnov & Apweiler, 2001).

The GO terms associated with each assembled sequence of the turtle transcriptome were then classified into biological processes, molecular functions, and cellular components. InterProScan was also used to predict the functional domains, signal peptides, and other protein characters by blasting to the Conserved Domain Database Interpro.

We then employed the KEGG Automatic Annotation Server (KAAS) with default settings ([Moriya et al., 2007](#)) to perform a KEGG pathway analysis ([Kanehisa & Goto, 2000](#)) on each contig. Telomere-related genes were filtered from the annotation results.

Analysis of candidate longevity-associated genes

Candidate genes associated with turtle longevity were screened from the logfile generated by Blastp. Two shelterin genes (i.e., *pot1* and *tin2*) were not detected in this step. To search the lost but should be existed genes in *M. reevesii* transcriptome, we used released *pot1* (XM_007063192.1) and *tin2* (XM_005290151.2) genes to blastn against the primary assembled *M. reevesii* transcriptome data. And the *pot1* and *tin2* genes were finally screened out that escaped from functional annotation.

The filtered longevity-associated genes were then blasted to the NR database using Blastn, to verify the accuracy of previous annotations. Candidate gene fragments were aligned, overlapping fragments were assembled into one sequence, and separated fragments were concatenated into a single sequence. FPKM values of each fragment were also checked using the results from the RSEM software.

Screened genes were translated into amino acids, using MEGA 6.06 ([Tamura et al., 2013](#)) and according to a standard codon table. Amino acid sequences were then aligned with published data using the online alignment tool MAFFT (version 7) (<http://mafft.cbrc.jp/alignment/software/>) ([Kato et al., 2002](#)), with default settings. BioEdit 7.2.3 ([Hall, 1999](#)) was used to display the alignment results.

RESULTS

Sequencing and *de novo* assembly

Total RNA extracted from liver, spleen, and skeletal muscle were used to generate a pooled cDNA sample and subsequently sequenced. A total of 160,998,396 paired-end raw reads of 100 bp length were generated and stored in fastq format (GenBank accession number: [SRX1469958](#)). We obtained 152,214,434 (94.5%) high-quality reads with an average length of 98.8 bp. The results of *de novo* assembly yielded 459,911 isoforms, which clustered into 230,085 transcripts with average length of 660 bp and median length of 342 bp. 190 transcripts encompassed over 100 isoforms and 6,051 transcripts encompassed over 10 isoforms with an average of two isoforms per transcript overall. Additionally, the highest number of isoforms found in one transcript was 1,048. These results may indicate the widespread existence of alternative splicing in *M. reevesii*. GC content for the entire final assembly and all protein coding sequences were 46.67% and 51.16%, respectively. Repetitive elements and microsatellites were also analyzed ([File S1](#)).

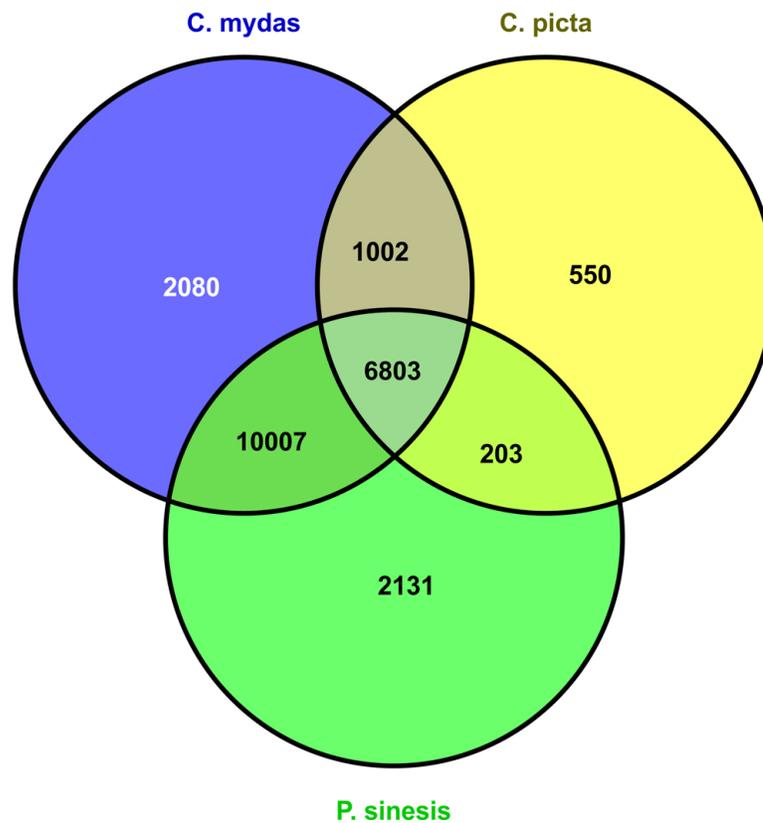


Figure 1 *M. reevesii* homologous gene detection in diverse turtle proteomes.

Table 1 Summary of Blastx search results of *M. reevesii* transcriptome.

Database	Unigene hits	Unique protein
<i>Chelonia mydas</i>	19,892	18,263
<i>Pelodiscus sinensis</i>	19,144	27,267
<i>Chrysemys picta bellii</i>	8,558	27,198

Comparison with turtle's reference proteomes

The assembly quality of the *M. reevesii* transcriptome was assessed with a Blastx comparison to the reference proteomes of three turtles: the green sea turtle (*Chelonia mydas*), the Chinese soft-shell turtle (*Pelodiscus sinensis*), and the western painted turtle (*Chrysemys picta bellii*) (Fig. 1 and Table 1). *Chelonia mydas* exhibited the highest degree of similarity. The comparison to *P. sinensis* and *C. picta* yielded slightly more unique proteins than transcripts hits, but overall, we obtained a total of 22,776 positive transcripts hits.

Transcripts expression level

The expression of each transcript was quantified using RSEM software, set to default parameters. Statistical results are shown in Table 2. The transcripts with the highest levels of expression were related to metabolism and translation activity.

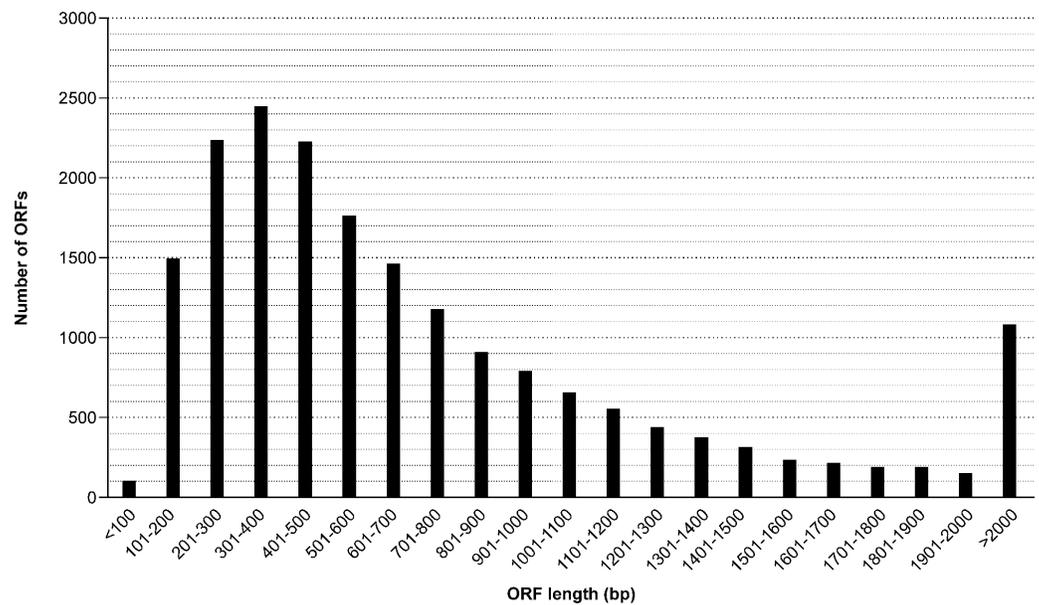


Figure 2 Length distribution of identified ORF (open reading frames) from the *M. reevesii* transcriptome assembly.

Table 2 Statistics of FPKM distribution of assembled unigenes.

	Unigenes
>1,000 FPKM	101
>100 FPKM	809
<10 FPKM	218,717 (95.1%)
Max FPKM	12367.8
Min FPKM	0
Total	230,085

Functional annotation

The results of our functional annotation revealed 42,918 (18.7%) transcripts that were predicted to potentially code for proteins. These were then translated into amino acid sequences by referring to a standard codon table. After a blastp search on the resultant amino acid sequences, we obtained a total of 534,077 hits, the best of which corresponded to 18,846 unique protein accessions in the Swissprot/Uniprot database. Length distribution for all opening reading frames ranged from 40 bp to 34,967 bp, with an average length of 874 bp (Fig. 2). The Blastp top-hit species distribution of gene annotations showed the highest homology to *Chelonia mydas* (6,654 annotation results) and *Pelodiscus sinensis* (5,747 annotation results) (Fig. S1). Those two species supplied 65.8% of all annotation information due to their close phylogenetic relationships to *M. reevesii*.

Table 3 Statistics of predicted protein domains and site characters.

Domains	Counts
Zinc finger, C2H2-like	1,943
Immunoglobulin-like fold	1,734
Fibronectin, type III	1,605
Zinc finger C2H2-type/integrase DNA-binding domain	1,543
Ankyrin repeat	1,517
Cadherin	1,217
P-loop containing nucleoside triphosphate hydrolase	1,106
G protein-coupled receptor, rhodopsin-like	1,037
Src homology-3 domain	856
EGF-like, conserved site	802
PDZ domain	788
Immunoglobulin subtype	744
Protein kinase domain	720
Leucine-rich repeat, typical subtype	720
Sushi/SCR/CCP	654
Low-density lipoprotein (LDL) receptor class A repeat	593
Thrombospondin, type 1 repeat	584
Pleckstrin homology domain	569
EF-hand domain	538
Small GTPase superfamily	494

The results of a gene ontology (GO) analysis assigned 11,695 unique proteins to 4,031 terms for biological processes, molecular functions, and cellular components. Within biological processes, metabolic (26%) and cellular processes (26%) were the most well-represented. Next, the majority of the proteins assigned to molecular functions were associated with binding (85%). Finally, within cellular components, cell (33%) and membrane (31%) proteins were the most well-represented (Fig. 3).

The most abundant conserved protein domain found in our data was the zinc finger C2H2 domain, followed by the immunoglobulin-like fold and fibronectin-III. Zinc finger-associated conserved protein domains are 8% of all conserved domains, and 62.7% of those are associated with zinc finger C2H2 (Table 3).

The KEGG pathway analysis annotated 4,486 transcripts into 338 pathways, with 3.5 KEGG pathways per transcript on average. Of all the annotated sequences, a large portion (1,581, or 35.2% of 4,486) were related to metabolism, specifically of carbohydrates (332 sequences), lipids (247 sequences), and amino acids (291 sequences). Next, 855 sequences (19.1%) were involved in signal transduction. The most well-represented was the PI3K-Akt signaling pathway (ko04151; 224 sequences), which could be activated by IGF-1, followed by the MAPK signaling pathway (ko04010; 159 sequences). Finally, 483 sequences (10.8%) were associated with the immune system, including the T cell (ko04660) and B cell (ko04662) receptor signaling pathways.

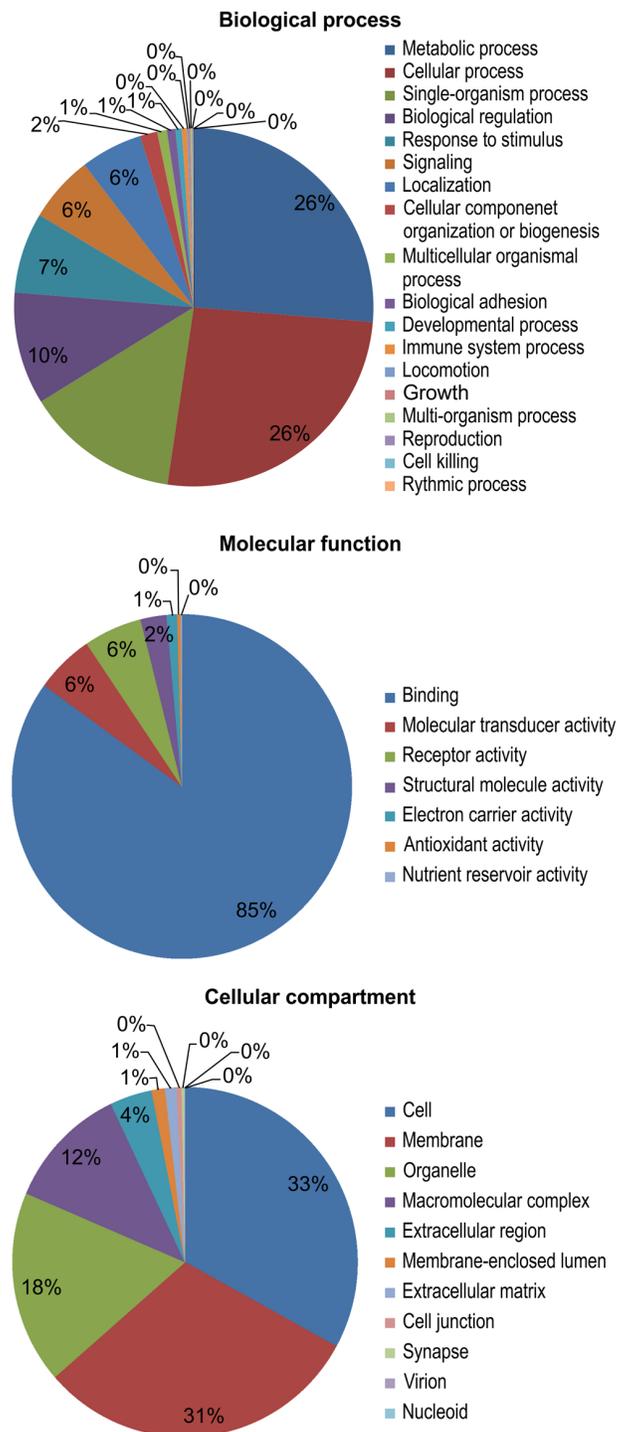


Figure 3 Gene ontology (GO) functional categories of the *M. reevesii* assembly.

Table 4 Sequence length and comparison of *tert* and TERT between *M. reevesii* and *C. picta*, *C. mydas*, *P. sinensis* and *H. sapiens*.

	<i>M. reevesii</i>	<i>C. picta</i>	<i>C. mydas</i>	<i>P. sinensis</i>	<i>H. sapiens</i>
<i>Tert</i> (nt)	3,693	4,095(97%)	4,065(96%)	3,522(88%)	3,399(70%)
TERT (aa)	1,230	1,352(96%)	1,354(93%)	1,173(81%)	1,132(60%)

Longevity-related genes

Candidate genes related to longevity were screened out from the total pool of annotated genes. Specifically, fragments of *tert*, *tep1* and six (*trf1*, *trf2*, *tpp1*, *pot1*, *tin2* and *rap1*) shelterin proteins were found (File S2). Genes were translated referring to standard codon table and aligned with public database.

Tert gene, encoding the catalytic subunit of telomerase enzyme, has been drawn a lot of attention by biologists. We obtained a 3,817 bp of *tert* gene fragment including two conserve domains in 3' terminal: RNA binding domain of telomerase (TRBD) and reverse transcriptase (RT) domain are essential for the gene function. The *tert* gene sequence length and identity were compared with other turtles and human (Table 4). Nucleotide conservation is interestingly higher than amino acid conservation, the similar results were also found in *Nothobranchius furzeri* (Hartmann et al., 2009).

FPKM values are approximately 10 of all aforementioned genes and genes that were closely related in function (File S3). The similar expression levels indicated that the candidate genes were accurately identified.

DISCUSSION

Profile of the *M. reevesii* transcriptome

More than 15-Gb high-quality data were generated with Illumina sequencing in this study. A total length of 16,478,555 bp (0.75% of 2.2 Gb) for 42,918 assembled transcripts were annotated that encoded for mRNA in *M. reevesii* transcriptome, functional annotation of the mRNAs provided fully information of transcripts and whole profile of the transcriptome that could be used in further study. The most abundant domain, C2H2 zinc finger, was consistent with the results found in mammals and humans, the fact that the motif is the most prevalent and the largest sequence-specific DNA-binding protein family (Lander et al., 2001; Tupler, Perini & Green, 2001). The conservation and evolution of this domain could be further studied (Englbrecht, Schoof & Böhm, 2004).

Lastly, our functional annotation results also revealed a small amount of contaminated transcripts associated with viruses. While such data are negligible, they should nonetheless be excluded in future studies.

The active expression of telomere-associated genes may contribute to turtle longevity

Telomere shortening is now considered the molecular clock that triggers cell and organismal senescence (Harley & Goldstein, 1978). To prevent premature shortening, telomere length is maintained by telomerase, previously considered to be inactive in human somatic

cells (Kim *et al.*, 1994; Shay & Bacchetti, 1997). Using RNA-seq technology, we were able to find two coding genes of the telomerase complex—TERT and TEP1—in the *M. reevesii* transcriptome. Both genes are essential for proper telomerase function: transient expression of TERT has been found to reconstitutes telomerase activity (Weinrich *et al.*, 1997), while TEP1 is a component of the ribonucleoprotein complex (Poderycki *et al.*, 2005). While TEP1 has been closely linked to telomerase activity (Nakayama *et al.*, 1997), other research has thrown doubt on this connection (Uchida *et al.*, 1999). Thus, the exact function of TEP1 is a subject for further research. The presence of these two genes might implies that activated telomerase is present in turtle somatic tissues to protect the turtle from senescence.

We were able to detect all the reported six genes that encode shelterin proteins in our data. Expression of these genes could prevent the telomeres from shortening during cell division and thus were predicted having a significant effect to the longevity of this species.

CONCLUSION

In this study, the transcriptome of the Chinese three-keeled pond turtle was sequenced using the Illumina Hiseq 2500 platform. A *de novo* assembly was then evaluated to uncover longevity-related candidate genes, mainly associated with telomere function, which offer a clue to the mechanisms behind turtle longevity. Future studies will incorporate RT-PCR and immunohistochemical techniques to test the gene expression levels and telomerase activity in different tissues. Finally, we believe the transcriptome data generated here can serve as a valuable resource for any investigations of turtle longevity and other notable characters in this order.

Abbreviations

<i>tert</i>	telomerase reverse transcriptase
<i>tep1</i>	telomerase protein component 1
<i>trf1</i>	Telomeric Repeat Binding Factor 1
<i>tpp1</i>	Tripeptidyl peptidase 1
<i>rap1</i>	Ras-related protein 1
<i>pot1</i>	Protection of telomeres 1
<i>tin2</i>	TRF1-interacting nuclear factor 2
FPKM	Fragments per Kilobase of transcript per Million mapped reads

ACKNOWLEDGEMENTS

We would like to thank Genergy for providing sequencing service and primary data analysis and Editage for providing editorial assistance. We also thank Xuming Zhou and PhD Xianzhao Kan for professional advice.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The National Natural Science Foundation of China (NSFC, No. 31372198 and 30970351) and the Research Fund of the Key Laboratory of Biotic Environment and Ecological Safety of Anhui province funded the work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

The National Natural Science Foundation of China: 31372198, 30970351.

Research Fund of the Key Laboratory of Biotic Environment: Ecological Safety of Anhui.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Huazong Yin conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables.
- Liuwang Nie conceived and designed the experiments, reviewed drafts of the paper.
- Feifei Zhao performed the experiments, contributed reagents/materials/analysis tools.
- Huaxing Zhou, Xianmei Dong, Huanhuan Zhang, Yuqin Wang, Qiong Shi and Jun Li assistance.
- Haifeng Li prepared figures and/or tables.

Animal Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

Procedures involving animals and their care were approved by the Animal Care and Use Committee of Anhui Normal University under approval number #20140111.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

GenBank: [SRX1469958](https://www.ncbi.nlm.nih.gov/nuclseq/SRX1469958).

Data Availability

The following information was supplied regarding data availability:

Figshare: <https://figshare.com/s/aa9d1a8dc99ce8ace2bb>

DOI: [10.6084/m9.figshare.2008326](https://doi.org/10.6084/m9.figshare.2008326).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.2062#supplemental-information>.

REFERENCES

- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT. 2000. Gene ontology: tool for the unification of biology. *Nature Genetics* 25:25–29 DOI 10.1038/75556.
- Balakrishnan R, Christie KR, Costanzo MC, Dolinski K, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hong EL, Nash R. 2005. Fungal BLAST and model organism BLASTP best hits: new comparison resources at the Saccharomyces Genome Database (SGD). *Nucleic Acids Research* 33:D374–D377.
- Bilaud T, Brun C, Ancelin K, Koering CE, Laroche T, Gilson E. 1997. Telomeric localization of TRF2, a novel human telobox protein. *Nature Genetics* 17:236–239 DOI 10.1038/ng1097-236.
- Blackburn EH, Greider CW, Szostak JW. 2006. Telomeres and telomerase: the path from maize, tetrahymena and yeast to human cancer and aging. *Nature Medicine* 12:1133–1138 DOI 10.1038/nm1006-1133.
- Bodnar AG, Ouellette M, Frolkis M, Holt SE, Chiu C-P, Morin GB, Harley CB, Shay JW, Lichtsteiner S, Wright WE. 1998. Extension of life-span by introduction of telomerase into normal human cells. *Science* 279:349–352 DOI 10.1126/science.279.5349.349.
- Broccoli D, Smogorzewska A, Chong L, De Lange T. 1997. Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nature Genetics* 17:231–235 DOI 10.1038/ng1097-231.
- Chong L, Van Steensel B, Broccoli D, Erdjument-Bromage H, Hanish J, Tempst P, De Lange T. 1995. A human telomeric protein. *Science* 270:1663–1667 DOI 10.1126/science.270.5242.1663.
- De Lange T. 2005. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes and Development* 19:2100–2110 DOI 10.1101/gad.1346005.
- De Souza RR, Vogt RC. 1994. Incubation temperature influences sex and hatchling size in the neotropical turtle *Podocnemis unifilis*. *Journal of Herpetology* 28:453–464.
- Englbrecht CC, Schoof H, Böhm S. 2004. Conservation, diversification and expansion of C2H2 zinc finger proteins in the Arabidopsis thaliana genome. *BMC Genomics* 5:39 DOI 10.1186/1471-2164-5-39.
- Gibbons J. 1987. Why do turtles live so long? *BioScience* 37:262–269.
- Girondot M, Garcia J. 1999. Senescence and longevity in turtles: what telomeres tell us. In: *Proceedings of the 9th extraordinary meeting of the Europea Societas herpetologica*, vol. 1. 25–29.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Trinity: full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29:644–652 DOI 10.1038/nbt.1883.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: *Nucleic acids symposium series*, vol. 41, 95–98.

- Harley CB, Goldstein S. 1978.** Cultured human fibroblasts: distribution of cell generations and a critical limit. *Journal of Cellular Physiology* **97**:509–515 DOI [10.1002/jcp.1040970326](https://doi.org/10.1002/jcp.1040970326).
- Hartmann N, Reichwald K, Lechel A, Graf M, Kirschner J, Dorn A, Terzibasi E, Wellner J, Platzer M, Rudolph KL. 2009.** Telomeres shorten while Tert expression increases during ageing of the short-lived fish *Nothobranchius furzeri*. *Mechanisms of Ageing and Development* **130**:290–296 DOI [10.1016/j.mad.2009.01.003](https://doi.org/10.1016/j.mad.2009.01.003).
- Houghtaling BR, Cuttonaro L, Chang W, Smith S. 2004.** A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2. *Current Biology* **14**:1621–1631 DOI [10.1016/j.cub.2004.08.052](https://doi.org/10.1016/j.cub.2004.08.052).
- Kanehisa M, Goto S. 2000.** KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**:27–30 DOI [10.1093/nar/28.1.27](https://doi.org/10.1093/nar/28.1.27).
- Katoh K, Misawa K, Kuma Ki, Miyata T. 2002.** MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**:3059–3066 DOI [10.1093/nar/gkf436](https://doi.org/10.1093/nar/gkf436).
- Kim EB, Fang X, Fushan AA, Huang Z, Lobanov AV, Han L, Marino SM, Sun X, Turanov AA, Yang P, Yim SH, Zhao X, Kasaikina MV, Stoletzki N, Peng C, Polak P, Xiong Z, Kiezun A, Zhu Y, Chen Y, Kryukov GV, Zhang Q, Peshkin L, Yang L, Bronson RT, Buffenstein R, Wang B, Han C, Li Q, Chen L, Zhao W, Sunyaev SR, Park TJ, Zhang G, Wang J, Gladyshev VN. 2011.** Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature* **479**:223–227 DOI [10.1038/nature10533](https://doi.org/10.1038/nature10533).
- Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PdL, Coviello GM, Wright WE, Weinrich SL, Shay JW. 1994.** Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**:2011–2015 DOI [10.1126/science.7605428](https://doi.org/10.1126/science.7605428).
- Kim S-h, Kaminker P, Campisi J. 1999.** TIN2, a new regulator of telomere length in human cells. *Nature Genetics* **23**:405–412 DOI [10.1038/70508](https://doi.org/10.1038/70508).
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W. 2001.** Initial sequencing and analysis of the human genome. *Nature* **409**:860–921 DOI [10.1038/35057062](https://doi.org/10.1038/35057062).
- Li B, Dewey CN. 2011.** RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**:323 DOI [10.1186/1471-2105-12-323](https://doi.org/10.1186/1471-2105-12-323).
- Li B, Oestreich S, De Lange T. 2000.** Identification of human Rap1: implications for telomere evolution. *Cell* **101**:471–483 DOI [10.1016/S0092-8674\(00\)80858-2](https://doi.org/10.1016/S0092-8674(00)80858-2).
- Li C, Wu XC, Rieppel O, Wang LT, Zhao LJ. 2008.** An ancestral turtle from the Late Triassic of southwestern China. *Nature* **456**:497–501 DOI [10.1038/nature07533](https://doi.org/10.1038/nature07533).
- Liu D, Safari A, O'Connor MS, Chan DW, Laegeler A, Qin J, Songyang Z. 2004.** PTOP interacts with POT1 and regulates its localization to telomeres. *Nature Cell Biology* **6**:673–680 DOI [10.1038/ncb1142](https://doi.org/10.1038/ncb1142).

- Lutz PL, Prentice HM, Milton SL. 2003.** Is turtle longevity linked to enhanced mechanisms for surviving brain anoxia and reoxygenation? *Experimental Gerontology* **38**:797–800 DOI [10.1016/S0531-5565\(03\)00111-6](https://doi.org/10.1016/S0531-5565(03)00111-6).
- McClintock B. 1939.** The behavior in successive nuclear divisions of a chromosome broken at meiosis. *Proceedings of the National Academy of Sciences of the United States of America* **25**:405–416 DOI [10.1073/pnas.25.8.405](https://doi.org/10.1073/pnas.25.8.405).
- Miller J. 2001.** Escaping senescence: demographic data from the three-toed box turtle (*Terrapene carolina triunguis*). *Experimental Gerontology* **36**:829–832 DOI [10.1016/S0531-5565\(00\)00243-6](https://doi.org/10.1016/S0531-5565(00)00243-6).
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007.** KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* **35**:W182–W185 DOI [10.1093/nar/gkm321](https://doi.org/10.1093/nar/gkm321).
- Mrosovsky N, Dutton P, Whitmore C. 1984.** Sex ratios of two species of sea turtle nesting in Suriname. *Canadian Journal of Zoology* **62**:2227–2239 DOI [10.1139/z84-324](https://doi.org/10.1139/z84-324).
- Muller H. 1938.** The remaking of chromosomes. *Collecting Net* **13**:181–195.
- Nakayama J-i, Saito M, Nakamura H, Matsuura A, Ishikawa F. 1997.** TLP1: a gene encoding a protein component of mammalian telomerase is a novel member of WD repeats family. *Cell* **88**:875–884 DOI [10.1016/S0092-8674\(00\)81933-9](https://doi.org/10.1016/S0092-8674(00)81933-9).
- Oliveros J. 2007.** VENNY. An interactive tool for comparing lists with Venn Diagrams. Available at <http://bioinfogpcnbcscics/tools/venny/indexhtml>.
- Packard GC, Lohmiller LD, Packard MJ, Lang JW. 1997.** Cold tolerance in hatchling painted turtles (*Chrysemys picta*): supercooling or tolerance for freezing? *Physiological and Biochemical Zoology* **70**:670–678.
- Palm W, De Lange T. 2008.** How shelterin protects mammalian telomeres. *Annual Review of Genetics* **42**:301–334 DOI [10.1146/annurev.genet.41.110306.130350](https://doi.org/10.1146/annurev.genet.41.110306.130350).
- Poderycki MJ, Rome LH, Harrington L, Kickhoefer VA. 2005.** The p80 homology region of TEP1 is sufficient for its association with the telomerase and vault RNAs, and the vault particle. *Nucleic Acids Research* **33**:893–902 DOI [10.1093/nar/gki234](https://doi.org/10.1093/nar/gki234).
- Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, Swerdlow H, Turner DJ. 2008.** A large genome center's improvements to the Illumina sequencing system. *Nature Methods* **5**:1005–1010 DOI [10.1038/nmeth.1270](https://doi.org/10.1038/nmeth.1270).
- Reis-Filho JS. 2009.** Next-generation sequencing. *Breast Cancer Research* **11**:1–7.
- Schmidt KP, Inger RF. 1957.** *Living reptiles of the world*. New York: Doubleday.
- Shaffer HB, Minx P, Warren DE, Shedlock AM, Thomson RC, Valenzuela N, Abramyan J, Amemiya CT, Badenhorst D, Biggar KK. 2013.** The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biology* **14**:1–23 DOI [10.1186/gb-2013-14-3-r28](https://doi.org/10.1186/gb-2013-14-3-r28).
- Shay J, Bacchetti S. 1997.** A survey of telomerase activity in human cancer. *European Journal of Cancer* **33**:787–791 DOI [10.1016/S0959-8049\(97\)00062-2](https://doi.org/10.1016/S0959-8049(97)00062-2).
- Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S. 2013.** MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**:2725–2729 DOI [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197).

- Tupler R, Perini G, Green MR. 2001.** Expressing the human genome. *Nature* **409**:832–833 DOI [10.1038/35057011](https://doi.org/10.1038/35057011).
- Uchida N, Otsuka T, Shigematsu H, Maeda M, Sugio Y, Itoh Y, Niho Y. 1999.** Differential gene expression of human telomerase-associated protein hTERT and TEP1 in human hematopoietic cells. *Leukemia Research* **23**:1127–1132 DOI [10.1016/S0145-2126\(99\)00149-6](https://doi.org/10.1016/S0145-2126(99)00149-6).
- Van Dijk P, Iverson J, Rhodin A, Shaffer H, Bour R. 2014.** Turtles of the world: annotated checklist of taxonomy, synonymy, distribution with maps, and conservation status. *Chelonian Research Monographs* **5**:329–479.
- Weinrich SL, Pruzan R, Ma L, Ouellette M, Tesmer VM, Holt SE, Bodnar AG, Lichtsteiner S, Kim NW, Trager JB. 1997.** Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTERT. *Nature Genetics* **17**:498–502 DOI [10.1038/ng1297-498](https://doi.org/10.1038/ng1297-498).
- Xin H, Liu D, Songyang Z. 2007.** The telosome/shelterin complex and its functions. *Genome Biology* **9**:232–232 DOI [10.1186/gb-2008-9-9-232](https://doi.org/10.1186/gb-2008-9-9-232).
- Ye JZ-S, Hockemeyer D, Krutchinsky AN, Loayza D, Hooper SM, Chait BT, De Lange T. 2004.** POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. *Genes and Development* **18**:1649–1654 DOI [10.1101/gad.1215404](https://doi.org/10.1101/gad.1215404).
- Zdobnov EM, Apweiler R. 2001.** InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**:847–848 DOI [10.1093/bioinformatics/17.9.847](https://doi.org/10.1093/bioinformatics/17.9.847).
- Zhong Z, Shiue L, Kaplan S, De Lange T. 1992.** A mammalian factor that binds telomeric TTAGGG repeats *in vitro*. *Molecular and Cellular Biology* **12**:4834–4843 DOI [10.1128/MCB.12.11.4834](https://doi.org/10.1128/MCB.12.11.4834).