Complete mitochondrial genome of Indian tent turtle, *Pangshura tentoria* and comparative mitochondriomics

Shantanu Kundu [1] , **Vikas Kumar** [Corresp., 1] , **Kaomud Tyagi** [1] , **Rajasree Chakraborty** [1] , **Iftikar Rahaman** [1] , **Kailash Chandra** [1]

[1] Centre for DNA Taxonomy, Molecular Systematics Division, Zoological Survey of India, Kolkata, West Bengal, India

Corresponding Author: Vikas Kumar
Email address: vikaszsi77@gmail.com

Complete mitochondrial genomes of Indian tent turtle, *Pangshura tentoria* was sequenced and annotated as 16,657 bp in length. This first assembly was encoded by 37 genes: 13 protein coding genes (PCGs), 22 transfer RNA (tRNAs), two ribosomal RNA (rRNAs) as similar to the typical vertebrate mitochondrial gene arrangement. The complete mitogenome has a base composition of A (33.30%), G (13.54%), C (27%), and T (26.13%). Most of the genes were encoded on major strand, except for the eight tRNAs and one PCG (*nad6*). Almost all PCGs were starting with an ATG initiation codon, except for cytochrome oxidase subunit 1 (*cox1*) with 'GTG' and NADH dehydrogenase subunit 5 (*nad5*) with 'ATA'. The typical termination codons, 'TAA' and 'AGA' has been observed in NADH dehydrogenase subunit 4l (*nad4l*) and NADH dehydrogenase subunit 6 (*nad6*) respectively; and others were used incomplete stop codons. The Relative Synonymous Codon Usage (RSCU) analysis revealed the maximum abundance of Alanine, Isoleucine, Leucine, and Threonine in *P. tentoria*. Codon distribution per thousand codon (CDsPT) values for all the amino acids showed the maximum values were present for Leucine in all geoemydid turtles. Further, the PCGs showed non-synonymous (Ka)/synonymous (Ks) values were <1 that indicated a strong negative selection among the studied species. The tRNAs were folded into classic clover-leaf secondary structures, except for *trnS* (GCT), lacking of the conventional DHU arm or stem. Further, the 10 tRNAs showed G-T mismatches and forming weak bonds. In the control region (CR) of *P. tentoria*, a single tandem repeat of eight base pairs (TTCTCTTT) was resulted with two copy numbers. The comparative study of CR with other geoemydid turtles revealed the numbers of tandem repeats were frequent in the 3´ end and structural characteristic were species-specific. The Maximum Likelihood (ML) phylogeny showed 32 geoemydid turtles were clustered distinctly with high bootstrap support and congruent with the previous phylogenetic hypothesis. Further, the representative mitogenome sequences of other family/suborder were depicted discrete clades in the ML tree. The study argued the complete mitochondrial genome sequence of

P. tentoria and comparative mitochondriomics of geoemydid turtles would be useful for further phylogenetic reconciliation and evolutionary research.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27202v1 | CC BY 4.0 Open Access | rec: 14 Sep 2018, publ: 14 Sep 2018

Complete mitochondrial genome of Indian tent turtle, *Pangshura tentoria* and comparative mitochondriomics

Shantanu Kundu, Vikas Kumar, Kaomud Tyagi, Rajasree Chakraborty, Iftikar Rahaman, Kailash Chandra

Centre for DNA Taxonomy, Molecular Systematics Division, Zoological Survey of India, M Block, New Alipore, Kolkata-700053, India.

* Corresponding author:

Dr. Vikas Kumar, Scientist D
Centre for DNA Taxonomy, Molecular Systematics Division,
Zoological Survey of India, M Block, New Alipore, Kolkata-700053
Email: vikaszsi77@gmail.com
Mob. No. +91-9674944233

Running Head:

Mitochondriomics of Geoemydid turtles

ABSTRACT

Complete mitochondrial genomes of Indian tent turtle, *Pangshura tentoria* was sequenced and annotated as 16,657 bp in length. This first assembly was encoded by 37 genes: 13 protein coding genes (PCGs), 22 transfer RNA (tRNAs), two ribosomal RNA (rRNAs) as similar to the typical vertebrate mitochondrial gene arrangement. The complete mitogenome has a base composition of A (33.30%), G (13.54%), C (27%), and T (26.13%). Most of the genes were encoded on major strand, except for the eight tRNAs and one PCG (*nad6*). Almost all PCGs were starting with an ATG initiation codon, except for cytochrome oxidase subunit 1 (*cox1*) with 'GTG' and NADH dehydrogenase subunit 5 (*nad5*) with 'ATA'. The typical termination codons, 'TAA' and 'AGA' has been observed in NADH dehydrogenase subunit 4l (*nad4l*) and NADH dehydrogenase subunit 6 (*nad6*) respectively; and others were used incomplete stop codons. The Relative Synonymous Codon Usage (RSCU) analysis revealed the maximum abundance of Alanine, Isoleucine, Leucine, and Threonine in *P. tentoria*. Codon distribution per thousand codon (CDsPT) values for all the amino acids showed the maximum values were present for Leucine in all geoemydid turtles. Further, the PCGs showed non-synonymous (Ka)/synonymous (Ks) values were <1 that indicated a strong negative selection among the studied species. The tRNAs were folded into classic clover-leaf secondary structures, except for *trnS* (GCT), lacking of the conventional DHU arm or stem. Further, the 10 tRNAs showed G-T mismatches and forming weak bonds. In the control region (CR) of *P. tentoria*, a single tandem repeat of eight base pairs (TTCTCTTT) was resulted with two copy numbers. The comparative study of CR with other geoemydid turtles revealed the numbers of tandem repeats were frequent in the 3′ end and structural characteristic were species-specific. The Maximum Likelihood (ML) phylogeny showed 32 geoemydid turtles were clustered distinctly with high bootstrap support and congruent with the previous phylogenetic hypothesis. Further, the representative mitogenome sequences of other family/suborder were depicted discrete clades in the ML tree. The study argued the complete mitochondrial genome sequence of *P. tentoria* and comparative mitochondriomics of geoemydid turtles would be useful for further phylogenetic reconciliation and evolutionary research.

Subjects Zoology, Genomics

Keywords Freshwater turtles, *Pangshura*, Mitogenome, Genomics, Phylogeny, Evolution

INTRODUCTION

The geoemydid turtles are the most ornamental and highly threatened living fauna in the world (*Fritz & Havaš, 2007*). The genus *Pangshura Gray, 1856* belongs to the family Geoemydidae is one of the endangered South and Southeast Asian turtle species (*IUCN, 2018*). The group is known as a sister group of land tortoises (Family Testudinidae) and comprises about 71 species excepting the new world genera (*TTWG, 2017*). Most of the geoemydid species are adapted in freshwater eco-system, however, some prefer estuarine or terrestrial habitats (*Ernst et al., 2000*). In India, 16 geoemydid turtles were distributed in the northern, eastern and northeastern region, and *Vijayachelys silvatica* is endemic to Western Ghats in Southern part (*Buhlmann et al., 2009*; *Kundu et al., 2018a*). The Indian tent turtle, *Pangshura tentoria* (*Gray, 1834*) with three living congeners, previously placed with the three large-sized *Kachuga* species for more than a century (*Ernst & Barbour, 1989*; *Das, 1991, 1995*) and further resurrected based on morphology and molecular studies (*Das, 2001*; *Praschag et al., 2007*). Further, in the recent past the fossils record suggested a new *Pangshura* species, *P. tatrotia* from the Pliocene Siwaliks of Pakistan (*Walter & Tyler, 2010*). Fossil records of *Pangshura* in deposition of the Siwalik Hills and Narmada Valley, demonstrated their existence in India from Pleistocene periods (*Baruah et al., 2016*). *P. tentoria* is abundantly distributed in the riverine systems in Bangladesh and northeast India and categorized as 'Lower Risk/least concern' in the International Union for Conservation of Nature (IUCN) red data list and 'Appendix II' in the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). However, due to the anthropogenic threats, the population of *P. tentoria* has dramatically declined in the northeastern region and other parts in India and thus the species is recommended to be listed in Indian Wildlife (Protection) Act, 1972 (*van Dijk et al., 2000*).

The mitochondriomics is evidenced effective in systematics and molecular phylogenetics studies by comparing the unique features of genes inside the mitochondrial genome (*Mindell et al., 1999*; *Parham et al., 2006*). The genetic motifs in complete mitogenomes demonstrated the phylogenetic position of turtles compare to other amniotes and suggested the sister relationship of archosaurs (*Zardoya & Meyer, 1998*; *Kumazawa & Nishida, 1999*). Till date the complete mitogenomes of 31 geoemydid turtles under seven genera (Family: Geoemydidae) has been assembled throughout the world and are available in GenBank database. However, the complete mitochondrial genome of genus *Pangshura* has not been assembled so far and unknown to the

world. Further, the structural features of genes and their arrangements and strand asymmetry, and inversion of the CR in mitogenomes are worthwhile to demonstrate the genomic characteristics and adjudicated the phylogenetic relationship (*San Mauro et al., 2006*; *Wei et al., 2010*; *Shi et al., 2013*; *Fonseca et al., 2014*). Thus, the present study was aimed to elucidate the detailing of *P. tentoria* mitogenomes and performed the comparative analysis with other geoemydid turtles for in-depth insights in phylogeny and evolutionary relationship.

MATERIALS AND METHODS

Sample collection and mitochondrial DNA extraction

The survey was conducted in northeast India, and *P. tentoria* sample was collected from Arunachal Pradesh state (Longitude 27°30' N and latitude 95°59' E) (Fig. S1). Prior permission was acquired from the wildlife authority, Arunachal Pradesh Biodiversity Board (Letter No. SFRI/APBB/09-2011-1221-1228 dated 22.07.2016) and Zoological Survey of India, Kolkata (Letter No. ZSI/MSD/CDT/2016-17 dated 29.07.2016) for handling the threatened animal. The blood sample was collected aseptically from the hind limbs by using micro-syringe and subsequently stored in EDTA containing vial at 4°C. The specimen was released in the same eco-system with adequate attention after collecting the biological sample. A drop of blood sample was centrifuged at 700X g for 5 min at 4ºC in 1 ml buffer (0.32 M Sucrose, 1 mM EDTA, 10 mM TrisHCl) to remove nuclei and cell debris. The supernatant was collected in 1.5 ml eppendorf tubes and centrifuged at 12000X g for 10 min at 4ºC to precipitate the mitochondria. The mitochondrial pellet was resuspended in 200 µl of buffer (50mM TrisHCl, 25mM of EDTA, 150mM NaCl), with the addition of 20 µl of proteinase K (20 mg/ml) followed by incubation at 56 ºC for 1 hour. Lastly, the mitochondrial DNA was extracted by Qiagen DNeasy Blood & Tissue Kit (QIAGEN Inc., Germantown MD). The DNA quality was checked in a 1% agarose gel electrophoresis and the concentration of mitochondrial DNA was quantified by NANODROP 2000 spectrophotometer (Thermo Scientific, USA).

Genome sequencing, assembly and annotation

Complete mitochondrial genome sequencing, *denovo* assembly and annotation were carried out at Genotypic Technology Pvt. Ltd. Bangalore, India (http://www.genotypic.co.in/). 200ng of DNA was used in Illumina TruSeq Nano DNA HT library preparation kit for library assembly. The purified A-tailed fragments were ligated with the sequencing indexed adapters after the fragmentation of mitochondrial DNA by ultra-sonication (Covaris M220, Covaris Inc., Woburn,

MA, USA). 450bp of fragments were selected using sample purification beads and amplified by polymerase chain reaction (PCR) to enrich it. The amplified PCR library was analyzed by Bioanalyzer 2100 (Agilent Technologies, Inc., Waldbronn, Germany) using High Sensitivity DNA chips. After obtaining the required concentration and mean peak size, total >4 million raw reads were generated through Illumina NextSeq500 (150 X 2 chemistry) (Illumina, Inc, USA). The raw reads were processed using cutadapt tool (http://code.google.com/p/cutadapt/) for adapters and low quality bases trimming with cutoff of phread quality score of Q20 and sequencing depth was > ~71000X. The high quality reads were down sampled to 2 million reads using seqtk program (https://github.com/lh3/seqtk) and down sampled high quality reads were denovo assembled using SPAdes-3.7.1. using default parameters (*Bankevich et al., 2012*). The fasta-formatted mitochondrial genome assembly was performed by aligning the contigs against the non-redundant nucleotide database of GenBank using the BLASTn search algorithm (http://blast.ncbi.nlm.nih.gov/Blast). The assembled contigs were homology searched against the Refseq database (https://www.ncbi.nlm.nih.gov/refseq/). The generated sequence annotation was also checked in MITOS online server (http://mitos.bioinf.uni-leipzig.de). The DNA sequences of PCGs were initially translated into the putative amino acid sequences on the basis of the genetic code of vertebrate mitochondrial genome (Fig. S2). The generated mitogenome was submitted in the GenBank of National Center for Biotechnology Information (NCBI) through the Sequin submission tool.

Genome characterization and comparative analysis

The circular representation of the generated complete mitochondrial genome of *P. tentoria* was mapped by CGView Server (http://stothard.afns.ualberta.ca/cgview_server/) with default parameters (*Grant & Stothard, 2008*). The representative of 31 geoemydid turtle's mitogenomes were downloaded on 9th June 2018 from GenBank and incorporated in the dataset for comparative analysis (Table S1). The six mitogenome sequences from different taxonomic hierarchy of Testudines were also acquired from GenBank for in-depth phylogenetic analysis. The genome size and comparative analysis of nucleotide composition were calculated using MEGA6.0 (*Tamura et al., 2013*). The direction and arrangements of each gene were also checked through MITOS online server (http://mitos.bioinf.uni-leipzig.de). The overlapping regions and intergenic spacers between genes were counted manually in Microsoft Excel. The start and stop codons of PCGs were checked through Open Reading Frame Finder

(https://www.ncbi.nlm.nih.gov/orffinder/) web tool. The comparative analysis of relative synonymous codon usage (RSCU), relative abundance of amino acids, and codons distribution were calculated using MEGA6.0. The synonymous (Ks) and non-synonymous (Ka) substitutions were calculated by DnaSPv5.0 (*Librado & Rozas, 2009*). The tRNA genes were verified in MITOS online server (http://mitos.bioinf.uni-leipzig.de), tRNAscan-SE Search Server 2.0 (http://lowelab.ucsc.edu/tRNAscan-SE/) and ARWEN 1.2 with the default settings (*Laslett & Canbäck, 2008*; *Lowe & Chan, 2016*). The large and small subunits of RNA (rrnL and rrnS) were annotated by the MITOS online server (http://mitos.bioinf.uni-leipzig.de). The tandem repeats in the control regions (CR) were predicted by the online Tandem Repeats Finder web tool (https://tandem.bu.edu/trf/trf.html) (*Benson, 1999*). The base composition skewness were calculated as described earlier: AT skew = [A−T]/[A+T], GC skew = [G−C]/[G+C] (*Perna & Kocher, 1995*). To assess the phylogenetic relationship, the PCGs were aligned individually by codons using MAFFT algorithm and concatenated in TranslatorX with L-INS-i strategy with default settings (*Abascal et al., 2010*). The best model for Maximum Likelihood (ML) analysis was calculated through MEGA6.0 program and 'GTR+G+I' model was selected with the lowest Bayesian information criterion (BIC) scores in MEGA6.0. The database sequence of *Chelus fimbriata*, family Chelidae (suborder: Pleurodira) was incorporated as an out-group in the phylogenetic analysis (Fig. S2).

RESULTS AND DISCUSSION

General genome features and organization

In this study, the complete mitogenomes, (16657 bp) of Indian tent turtle, *P. tentoria* was determined. The mitogenome encode by 37 genes, among them 13 were PCGs, 22 were tRNAs, two were rRNAs, and a major non-coding CR. Among these, 28 genes (12 PCGs, 14 tRNAs and two rRNAs) were located on the heavy strand and the remaining genes (*nad6* and eight tRNAs) were located on the light strand (Table 1, Fig. 1). The study depicted the gene rearrangements of *P. tentoria* were same as in the typical vertebrate gene arrangement described earlier (*Anderson et al., 1982*). The total length of PCGs was 11295 bp, however 1551 bp for tRNAs and 2562 bp for rRNA genes. The nucleotide composition of complete *P. tentoria* mitogenome was biased toward A+T (59.44%). The A+T composition of PCGs, tRNAs, rRNAs and CR were 58.52%, 60.28%, 58.86% and 66.06% respectively. In other geoemydid species, the A+T composition was also similar to *P. tentoria* and biased towards A+T with a variable frequency ranging from

58.12% (*B. trivittata*) to 62.52% (*H. depressa*). The AT skewness was 0.120 and GC skewness was -0.331 in the complete mitogenome of *P. tentoria*. The comparative analysis revealed that, the AT skewness was varied from 0.100 (*C. aurocapitata* and *N. platynota*) to 0.156 (*B. trivittata*) and GC skewness varied from -0.366 (*B. trivittata*) to -0.320 (*C. dentata*) in other geoemydid species (Table 2).

Overlapping and intergenic spacer regions

Six overlapping sequences with a total length of 12 bp were identified in *P. tentoria* complete mitogenome. These sequences varied in length from 1 bp to 4 bp with the longest overlapping region present between NADH dehydrogenase subunit 4L (*nad4l*) and NADH dehydrogenase subunit 4 (*nad4*) as well as in between ATP synthase F0 subunit 8 (*atp8*) and ATP synthase F0 subunit 6 (*atp6*). The intergenic spacers within this mitogenomes, spread over 19 regions and ranged in size from 1 bp to 27 bp with a total length of 115 bp. The longest spacer (27 bp) was observed between transfer RNA-Asparagine (*trnN*) and transfer RNA-Cysteine (*trnC*) (Table 1). In other geoemydid species, the number of overlapping sequences varied from four to six with a length variation 18 bp (*C. dentata*) to 94 bp (*M. leprosa*) with longest overlapping region present between transfer RNA Proline (*trnP*) and CR of *M. leprosa* (67 bp). The longest intergenic spacer of 29 bp was present between transfer RNA-Asparagine (*trnN*) and transfer RNA-Cysteine (*trnC*) of *N. platynota* (Table S2).

Protein coding genes

The total length of PCGs was 11295 bp in *P. tentoria*, which shared 67.8% of complete mitogenome. Furthermore, a contrast of nucleotide composition, AT skew and GC skew of PCGs in geoemydid species was exhibited in Table 2. The A+T composition was 58.52% in PCGs of *P. tentoria*. In other species, The A+T composition varied from 57.44% (*B. trivittata*) to 61.96% (*H. grandis*). The AT skew value of PCGs was 0.052 and GC skew value was -0.348 in *P. tentoria*. The AT skew value in other geoemydid species varied from 0.038 (*C. flavomarginata*) to 0.092 (*B. trivittata*) and GC skew value varied from -0.386 (*B. trivittata*) to -0.327 (*C. tcheponensis*). All PCGs of *P. tentoria* were starting with an ATG initiation codon, except for Cytochrome oxidase subunit 1 (*cox1*) gene with 'GTG' and NADH dehydrogenase subunit 5 (*nad5*) with 'ATA'. Out of 13 PCGs, typical termination codons, 'TAA' and 'AGA' has been observed in *nad4l* and *nad6* respectively; and rests of the PCGs were used incomplete termination codons (Table S2). In other geoemydid turtles, almost all of the PCGs have an

initiation codon 'ATG', except 'ATA' (*nad2* in *B. trivittata* and *C. amboinensis*; *cytb* in *C. aurocapitata* and *M. reevesii*), 'ATT' (*nad6* in *M. annamensis*) and 'GTG' (*cox1* in all the species except *C. dentata*). In the context of termination codon, almost all PCGs were used typical 'TAA', six PCGs have shown 'TAG' termination codons among these geoemydid species (*cox2*, *nad1*, *nad2*, *nad3*, *nad4*, and *nad6*), two PCGs have shown 'AGA' termination codon (*nad3* in *B. trivittata*, *C. amboinensis* and *C. aurocapitata*; *nad6* in *B. trivittata* and *M. sinensis*), 'AGG' termination codon (*cox1* of all the species and *nad6* in most of the species except, *B. trivittata, M. reevesii, M. sinensis* and *S. quadriocellata*), incomplete TA(G) termination codon was seen in *nad1* of *B. trivittata, C. amboinensis, C. bourreti, C. picturata, C. trifasciata, H. depressa, H. grandis, M. caspica, M. japonica, M. nigricans, M. megalocephala, M. rivulata, M. sinensis* and *S. bealei; cox3* of *M. mutica; nad5* of *S. quadriocellata* and *cytb* of *C. picturata*. Single, incomplete termination codon 'T' was seen in *nad2* and *cox3* of almost all studied geoemydid species; *nad6* in *S. quadriocellata* and *cytb* in almost all the species except *B. trivittata, C. amboinensis, C. aurocapitata, C. flavomarginata, C. dentata, C. oldhamii* and *M. reevesii* (Table S3).

Relative synonymous codon usage

The Relative Synonymous Codon Usage (RSCU) analysis revealed the maximum abundance of Alanine, Isoleucine, Leucine, Threonine in *P. tentoria*, whereas, Arginine, Aspartic Acids, Cysteine, Lysine were less abundant (Fig. 2). In other geoemydid species, the maximum abundance of Alanine, Asparagine, Isoleucine, Leucine, Serine, Threonine was observed and Arginine, Aspartic Acids, Cysteine, Lysine were less abundant (Table S4). The RSCU analyses of *P. tentoria* PCGs also indicated the major proportion of codons bearing Cytosine (C) or Guanine (G) in the third position rather than Adenine (A) and Thymine (T). The relative usage of the AAC and GAC were more, compared to the AAT and GAT in case of Asparagine and Aspartic Acids respectively. This same usage was more or less observed in other geoemydid species (Fig. 3). Codon distribution per thousand codon (CDsPT) values for all the amino acids showed the same result and the maximum values were present for Leucine in all the geoemydid species (Table S5). The maximum CDsPT value for Leucine was observed in *P. tentoria* (165.5) and minimum value was observed in *S. quadriocellata* (117.7) (Fig. S3).

Synonymous and nonsynonymous substitutions

The ratios of synonymous and nonsynonymous substitutions are generally known as an indicator of selective pressure and evolutionary relationship among different species (*Zhu et al., 2017*). It was stated that, the Ka/Ks>1 evidenced for positive selection, Ka/Ks=1 for neutral mutation, and Ka/Ks<1 for negative selection (*Nei & Kumar, 2000*; *Yang & Bielawski, 2000*). To investigate the evolutionary rates and comparative analysis with the complete mitogenome of *P. tentoria*, 32 geoemydid species representing eight genera (*Pangshura, Batagur, Cuora, Cyclemys, Heosemys, Mauremys, Notochelys, Sacalia*), sequence divergences by counting Ka/Ks substitutions were calculated. The Ka/Ks values of 13 PCGs varied from 0.006 (between *P. tentoria* and *B. trivittata* in *cox1*) to 0.549 (between *P. tentoria* and *H. annandalii* in *nad6*). All PCGs showed Ka/Ks values <1 which indicated a strong negative selection among all geoemydid species which reflects natural selection works against profitless mutations with negative selective coefficients. The percentages of synonymous and non-synonymous substitution variation was highest in *nad6* with an average value of 0.279, which indicated the least selective pressure in *nad6* gene. As Ka/Ks ratio is least in *cox1* with an average value of 0.012, this PCG is under most selective pressure. Among all the species pair, *P. tentoria* and *B. trivittata* showed least Ka/Ks value (0.006 in *cox1*) as compared to other pairs, implying a closer phylogenetic relationship between these two species. The Ka/Ks ratio of all the PCGs follows the order: *cox1<cox3<cox2<cytb<atp6<nad3<nad5<atp8<nad4 <nad4l<nad2<nad1<nad6* (Fig. 4).

Transfer RNAs and ribosomal RNAs

The representative secondary structures of 22 tRNAs were identified for *P. tentoria* complete mitogenome, total 1551 bp ranging from 67 bp to 76 bp with A+T content is 60.28%. In other geoemydid species, tRNA genes varied from 1496 bp (*M. annamensis*) to 1796 bp (*C. aurocapitata*). The A+T content of other geoemydid species varied from 60.02% (*B. trivittata*) to 63.94% (*C. tcheponensis*). The AT skewness and GC skewness of tRNA genes for *P. tentoria* were 0.026 and 0.055, respectively. The AT skewness of other species varied from 0.024 (*B. trivittata*) to 0.051 (*M. leprosa*) and GC skewness varied from 0.008 (*C. aurocapitata*) to 0.052 (*S. bealei* and *S. quadriocellata*). Among all 22 tRNA genes, 14 were in major strand and remaining eight (*trnQ, trnA, trnN, trnC, trnY, trnS2, trnE, and trnP*) were in minor strand and the anticodons of all tRNAs genes were resulted in Table S6. All of the tRNAs are meant to be folded into classic clover-leaf secondary structures in *P. tentoria,* except for *trnS* (GCT), which lacked the conventional DHU arm or stem; instead, a larger loop was observed there. The G-C

and A-T bonding were normally observed in the secondary structures of tRNA genes, however the secondary structures of 10 tRNA genes; $tRNA^{Ala}$(TGC), $tRNA^{Asn}$(GTT), $tRNA^{Cys}$(GCA), $tRNA^{Gln}$(TTG), $tRNA^{Glu}$(TTC), $tRNA^{Gly}$(TCC), $tRNA^{Leu}$(TAA), $tRNA^{Pro}$(TGG), $tRNA^{Ser}$(TGA), $tRNA^{Tyr}$(GTA) in *P. tentoria* shows G-T mismatches and forming weak bonds. The acceptor stem of $tRNA^{Asp}$(GTC), $tRNA^{Gln}$(TTG), $tRNA^{His}$(GTG), $tRNA^{Ile}$(GAT), $tRNA^{Met}$(CAT), $tRNA^{Phe}$(GAA), $tRNA^{Ser}$(GCT), $tRNA^{Val}$(TAC) had a small unconventional loop. An exceptional loop on TψC arm were present in $tRNA^{Asp}$(GTC) and $tRNA^{Met}$(CAT) and also an unconventional small loop on DHU arm was observed in $tRNA^{Trp}$(TCA) (Fig. S4). The length of rRNA genes in *P. tentoria* is 2562 bp. In other geoemydid species, the length of rRNA varied from 2553 bp (*C. picturata*) to 2859 bp (*S. quadriocellata*). The A+T composition of rRNA genes in *P. tentoria* is 58.86% and varied from 57.78% (*B. trivittata*) to 61.60% (*H. annandalii*) in other geoemydid species. The AT skewness and GC skewness of *P. tentoria* rRNA genes were 0.265 and -0.155, respectively. In other geoemydid species, The AT skewness varied from 0.257 (*H. depressa*) to 0.289 (*B. trivittata*) and GC skewness varied from -0.186 (*M. nigricans*) to -0.156 (*H. depressa*).

Control region

The control region (CR) known for the initiation of replication in vertebrates (*Bing et al., 2006*), is located between *trnP* and *trnF* in all the studied geoemydid species with a varying size. The length of the CR of *P. tentoria* was 949 bp and A+T composition was 66.06% (Table 2). A single tandem repeat of eight base pairs (TTCTCTTT) with two copy numbers was observed in the CR of *P. tentoria* (Fig. 5). The AT and GC skew values were negative, -0.030 and -0.236 respectively in *P. tentoria*. In the other geoemydid species the length of CR varied from 914 bp (*M. japonica*) to 1722 bp (*C. galbinifrons*). The AT and GC skew of other geoemydid species ranges from -0.142 (*C. pani*) to 0.044 (*C. oldhamii*) and -0.432 (*M. leprosa*) to -0.188 (*C. dentata*) respectively. It is resulted that Adenine (A) composition is equal to Thymine (T) composition in *C. atripons*, *C. pulchristriata* and *M. reevesii*. Adenine (A) composition is more as compared to Thymine (T) in *C. oldhamii, C. tcheponensis, H. depressa* and *M. rivulata*. The other geoemydid species have less Adenine (A) composition as compared to Thymine (T). The numbers of tandem repeats are higher at the 3′ end of the CR in most of the studied geoemydid species (Fig. 5). The AT composition ranges from 64.94% (*B. trivittata*) to 77.31% (*C. pani*). Among all, in 15 geoemydid species, a single tandem repeat was observed in the CR with repeat

length varied from 5 bp (*M. megalocephala*) to 10 bp (*C. trifasciata*, *M. annamensis*, *M. nigricans*). Overall, the CR in geoemydid mitogenomes showed a specific sequence and structural characteristic, which were species-specific and can be potentially used as a genetic marker for evolutionary and population genetics study.

Phylogeny of geoemydid mitogenomes

Mitochondrial and nuclear genes has been largely used for effective species identification and delimitation. However, to understand the robust evolutionary relationship, reconcile of phylogeny by in-depth molecular data is necessitated. Comparative mitochondriomics has demonstrated its utility in elucidating the phylogenetic relationships of many taxa, including turtles (*Amer & Kumazawa, 2009*; *Chen et al., 2013*; *Jiang et al., 2016*; *Li et al., 2017*; *Kundu et al., 2018b*). Originally, the genus *Pangshura* was erected by *Gray, 1856* within *Batagur*, a genus contained *Pangshura*, *Batagur*, *Kachuga* and *Morenia* species. Later on *Pangshura* was elevated as a distinct genus through morphological characteristics by *Günther, 1864* and *Moll, 1986*. Nevertheless, *Das, 2001* and *Schleich & Kästle, 2002* also supported *Pangshura* as a discrete monophyletic genus with four species. The first phylogenetic hypothesis based on molecular data was established by *Spinks et al., 2004* and evidenced *Batagur*, *Callagur*, *Hardella*, *Pangshura* as monophyletic genera. Further, *Praschag et al., 2007* with extensive taxon sampling of all species/subspecies and using of mitochondrial DNA, corroborated the well-supported monophyly of *Pangshura*. The present study constructed the phylogenetic relationship through concatenated PCGs derived from 32 geoemydid species mitogenomes (Fig. 6). The constructed ML phylogeny revealed all geoemydid species, clustered together with high bootstrap support and congruent with the previous evolutionary hypothesis (*Shaffer et al., 1997*; *Spinks et al., 2004*, *Praschag et al., 2007*; *Le et al., 2007*; *Guillon et al., 2012*). *P. tentoria* under genus *Pangshura* shows sister clade with the *Batagur trivittata* as described earlier. The congeners of *Mauremys*, *Cuora*, *Cyclemys*, *Heosemys*, *Sacalia*, and *Notochelys* are also clustered separately with significant bootstrap supports in ML tree. Further, the representative mitogenome sequences of other family/suborder, Testudinidae, Emydidae, Platysternidae, Cheloniidae, Trionychidae under suborder Cryptodira (hidden-necked turtles) and Chelidae under suborder Pleurodira (side-necked turtles) were clustered distinctly in ML phylogeny (Fig. 6). Thus, using complete mitochondrial genome, the present study was able to generate a robust phylogeny with high statistical values for each node and elucidate the uncertainty in the relationship between

Pangshura and other geoemydid species. Furthermore, based on the jawed morphology, the family Geoemydidae was divided into two subfamilies Geoemydinae and Batagurinae (*Gaffney & Meylan, 1988*). The studied genus *Pangshura* was transferred to Batagurinae subfamily with other living genera (*Batagur*, *Geoclemys*, *Hardella*, *Malayemys*, *Morenia*, *Ocadia*, and *Orlitia*). However, complete mitochondrial genomes of Batagurinae taxa is limited in global database. Thus, additional taxon sampling of geoemydid species under Batagurinae subfamily from different geographical locations and their mitogenome data would be useful to reconcile the phylogenetic and evolutionary relationship hereafter.

CONCLUSIONS

The sequencing, annotation of the mitochondrial genome of a member of genus *Pangshura* was accomplished in this study. The mitogenome of *P. tentoria* was found to be 16,657 bp in length and showed similar gene order in other geoemydid turtle species. The PCGs of *P. tentoria* showed the ratio of non-synonymous and synonymous substitution were <1 in compare to other geoemydid species that indicates a strong negative selection. Further, the comparative mitochondriomics study revealed the numbers and structural characteristic of tandem repeats were species-specific in geoemydid turtles. The maximum likelihood phylogeny showed distinct clustered of 32 geoemydid turtles and the relationships were congruent with the previous phylogeny and taxonomic classification. The complete mitogenome reported in the present study is expected to allow for further genomics studies of the extant *Pangshura* species and would be useful for estimating genetic differences within and between populations and conservation genetics.

ACKNOWLEDGEMENTS

We thank the Director of Zoological Survey of India (ZSI), Ministry of Environment, Forests and Climate Change (MoEF&CC), Govt. of India for providing necessary permissions and facilities.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Ministry of Environment, Forest and Climate Change (MoEF&CC), Zoological Survey of India (ZSI) in-house project, 'National Faunal Genome Resources (NFGR)'. 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:

MoEF&CC, ZSI in-house grant.

Fellowship to the first author (SK) from NMHS large grant 'Conservation of Threatened Vertebrate Fauna in Indian Himalayan Region through Long-Term Monitoring and Capacity Building.

Competing Interests

The authors declare that they have no competing interests.

Field Study Permissions

The following information was supplied relating to field study approvals:

Prior permission was acquired from the wildlife authority, Arunachal Pradesh Biodiversity Board (Letter No. SFRI/APBB/09-2011-1221-1228 dated 22.07.2016) and Zoological Survey of India, Kolkata (Letter No. ZSI/MSD/CDT/2016-17 dated 29.07.2016) for handling the threatened animal.

DNA Deposition

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

The genomic DNA of *Pangshura tentoria* were deposited in Centre for DNA Taxonomy, Molecular Systematics Division, Zoological Survey of India, Kolkata under IDs 'ZSI_NFGR-TT7'.

Data Availability

The following information was supplied regarding the availability of DNA sequences: The complete mitogenome of *Pangshura tentoria* is deposited in GenBank of NCBI under accession number MH795989.

REFERENCES

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Research **38**: W7-13.

Amer SAM, Kumazawa Y. 2009. Complete sequence of the mitochondrial genome of the endangered nile soft-shelled turtle *Trionyx triunguis*. Egyptian Journal of Experimental Biology (Zoology) **5**: 43- 50.

386 **Anderson S, Bruijn M, Coulson A, Eperon I, Sanger F, Young I. 1982.** Complete sequence
387 of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome.
388 Journal of Molecular Biology **156**: 683-717.
389 **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,**
390 **Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G,**
391 **Alekseyev MA, Pevzner PA. 2012**. SPAdes: a new genome assembly algorithm and its
392 applications to single-cell sequencing. The Journal of Computational Biology **19**: 455-477.
393 **Baruah C, Devi P, Sharma DK. 2016.** Comparative morphometry and biogeography of the
394 freshwater turtles of genus *Pangshura* (Testudines: Geoemydidae: *Pangshura*). International
395 Journal of Pure and Applied Zoology **4**: 107-123.
396 **Benson G. 1999.** Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids
397 Research **27**, 573-580.
398 **Bing X, Fei M, Yi S, Qing-Wei L. 2006.** Comparative Analysis of Complete Mitochondrial
399 DNA Control Region of Four Species of Strigiformes. Acta Genetica Sinica **33**, 965-974.
400 **Buhlmann KA, Akre TSB, Iverson JB, Karapatakis D, Mittermeier RA, Georges A,**
401 **Rhodin AGJ, Van Dijk PP, Gibbons JW. 2009.** A global analysis of tortoise and
402 freshwater turtle distributions with identiication of priority conservation areas. Chelonian
403 Conservation and Biology **8**:116-149.
404 **Chen X, Zhou Z, Peng X, Huang X, Chen Z. 2013.** Complete mitochondrial genome of the
405 endangered Asian giant softshell turtle *Pelochelys cantorii* (Testudinata: Trionychidae).
406 Mitochondrial DNA 24: 111-113.
407 **Das I. 1991.** Colour Guide to the Turtles and Tortoises of the Indian Subcontinent. Portishead,
408 Avon: R & A Publishing Limited.
409 **Das I. 1995.** Turtles and Tortoises of India. Bombay, Delhi, Calcutta and Madras: WWF India
410 and Oxford University Press.
411 **Das I. 2001.** Die Schildkröten des Indischen Subkontinents. Frankfurt am Main: Chimaira.
412 **Ernst CH, Altenburg RGM, Barbour RW. 2000.** Turtles of the World. World Biodiversity
413 Database, CD-ROM Series, Windows, Version 1.2. Amsterdam: Biodiversity Center of ETI.
414 **Ernst CH, Barbour RW. 1989.** Turtles of the World. Washington, DC: Smithsonian Institution
415 Press.

Fonseca MM, Harris DJ, Posada D. 2014. The Inversion of the Control Region in Three Mitogenomes Provides Further Evidence for an Asymmetric Model of Vertebrate mtDNA Replication. PLoS ONE **9**: e106654. doi:10.1371/journal.pone.0106654.

Fritz U, Havaš P. 2007. Checklist of chelonians of the world. Vertebrate Zoology. **57**:149-368.

Gaffney ES, Meylan PA. 1988. A phylogeny of turtles. In: Benton, M.J. (Ed.), The Phylogeny and Classification of Tetrapods. Clarendon Press, Oxford, England, pp. 157-219.

Grant JR, Stothard P. 2008. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Research **36**: W181-184.

Gray JE. 1834. Characters of several new species of freshwater tortoises (*Emys*) from India and China. Proceedings of the Zoological Society of London 1834: 53-54.

Gray JE. 1856. Catalogue of Shield Reptiles in the Collection of the British Museum. Part I. Testudinata (Tortoises). London: British Museum, 79 pp.

Guillon JM, Guéry L, Hulin V, Girondot M. 2012. A large phylogeny of turtles (Testudines) using molecular data. Contributions to Zoology 81: 147-158.

Günther A. 1864. Reptiles of British India. London: Ray Society.

IUCN. 2018. The IUCN red list of threatened species, Version 2018.1.

Jiang JJ, Xia E, Gao C. Gao L. 2016. The complete mitochondrial genome of western painted turtle, *Chrysemys picta bellii* (Chrysemys, Emydidae). Mitochondrial DNA. **27**: 787-788.

Kumazawa Y, Nishida M. 1999. Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: Statistical evidence for Archosaurian affinity of turtles. Molecular Biology and Evolution **16**: 784-792.

Kundu S, Kumar V, Laskar BA, Tyagi K, Chandra K. 2018a. Pet and turtle: DNA barcoding identified twelve Geoemydid species in northeast India, Mitochondrial DNA **3**: 513-518. DOI: 10.1080/23802359.2018.1467215.

Kundu S, Kumar V, Tyagi K, Chakraborty R, Singha D, Rahaman I, Pakrashi A, Chandra K. 2018b. Sequencing and characterization of the complete mitochondrial genome of Critically Endangered Black Soft-shell Turtle (*Nilssonia nigricans*). BioRxiv. doi: https://doi.org/10.1101/361501.

Laslett D, Canbäck B. 2008. ARWEN, a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24: 172-175.

Le M, McCord WP, Iverson JB. 2007. On the paraphyly of the genus *Kachuga* (Testudines: Geoemydidae). Molecular Phylogenetics and Evolution **45**: 398-404.

Li H, Liu J, Xiong L, Zhang H, Zhou H, Yin H, Jing W, Li J, Shi Q, Wang Y, Liu J, Nie L. 2017. Phylogenetic relationships and divergence dates of softshell turtles (Testudines: Trionychidae) inferred from complete mitochondrial genomes. Journal of Evolutionary Biology **30**: 1011-1023.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. **25**: 1451-1452.

Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: Search and Contextual Analysis of Transfer RNA Genes. Nucleic Acids Research **44**: W54-57.

Mindell DP, Sorenson MD, Dimcheff DE, Hasegawa M, Ast JC, Yuri T. 1999. Interordinal Relationships of Birds and Other Reptiles Based on Whole Mitochondrial Genomes. Systematic Biology **48**: 138-152.

Moll EO. 1986. Survey of the freshwater turtles of India. Part I: The genus *Kachuga*. Journal of the Bombay Natural History Society **83**: 538-552.

Nei M, Kumar S. 2000. Molecular Evolution and Phylogenetics. New York: Oxford University Press.

Parham JF, Feldman CR, Boore JL. 2006. The complete mitochondrial genome of the enigmatic bigheaded turtle (Platysternon): description of unusual genomic features and the reconciliation of phylogenetic hypotheses based on mitochondrial and nuclear DNA. BMC Evolutionary Biology **6**: 1-11.

Perna NT, Kocher TD. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. Journal of Molecular Evolution **41**: 353-359.

Praschag P, Hundsdörfer AK, Fritz U. 2007. Phylogeny and taxonomy of endangered South and South-east Asian freshwater turtles elucidated by mtDNA sequence variation (Testudines: Geoemydidae: *Batagur*, *Callagur*, *Hardella*, *Kachuga*, *Pangshura*). Zoologica Scripta **36**: 429-442.

San Mauro D, Gower DJ, Zardoya R, Wilkinson M. 2006. A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome. Molecular Biology and Evolution **23**: 227-234.

Schleich HH, Kästle W. 2002. Amphibians and Reptiles of Nepal. Biology, Systematics, Field Guide. Koenigstein: Koeltz Scientifc Books, pp. 1201.

Shaffer HB, Meylan P, McKnight ML. 1997. Tests of turtle phylogeny: molecular, morphological, and paleontological approaches. Systematic Biology **46**: 235-268.

Shi W, Dong XL, Wang ZM, Miao XG, Wang SY, Kong XY. 2013. Complete mitogenome sequences of four flatfishes (Pleuronectiformes) reveal a novel gene arrangement of L-strand coding genes. BMC Evolutionary Biology **13**:13. DOI:10.1186/1471-2148-13-173.

Spinks PQ, Shaffer HB, Iverson JB, McCord WP. 2004. Phylogenetic hypotheses for the turtle family Geoemydidae. Molecular Phylogenetics and Evolution **32**:164-182.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular evolutionary genetics analysis Version 6.0. Molecular Biology and Evolution **30**: 2725-9.

TTWG (Turtle Taxonomy Working Group). 2017. [Rhodin, A.G.J., Iverson, J.B., Bour, R. Fritz, U., Georges, A., Shaffer, H.B., and van Dijk, P.P.] Turtles of the World: Annotated Checklist and Atlas of Taxonomy, Synonymy, Distribution, and Conservation Status (8th Ed.). In: Rhodin, A.G.J., Iverson, J.B., van Dijk, P.P., Saumure, R.A., Buhlmann, K.A., Pritchard, P.C.H., and Mittermeier, R.A. (Eds.). Conservation Biology of Freshwater Turtles and Tortoises: A Compilation Project of the IUCN/SSC Tortoise and Freshwater Turtle Specialist Group. Chelonian Research Monographs **7**: 1-292. doi: 10.3854/crm.7.checklist.atlas.v8.2017.

van Dijk PP. 2000. The status of turtles in Asian. In: van Dijk, P.P., Stuart, B.L., Rhodin, A.G.J. (Eds.), Asian Turtle Trade: Proceedings of a Workshop on Conservation and Trade of Freshwater Turtles and Tortoises in Asia. Chelonian Research Foundation, Lunenburg, MA, pp. 15-23.

Walter GJ, Tyler RL. 2010. *Pangshura tatrotia*, a new species of pond turtle (Testudinoidea) from the Pliocene Siwaliks of Pakistan. Journal of Systematic Palaeontology **8**: 449-458.

Wei SJ, Shi M, Chen XX, Sharkey MJ, van Achterberg C, Ye GY, He JH. 2010. New Views on Strand Asymmetry in Insect Mitochondrial Genomes. PLoSONE **5**: e12708. doi:10.1371/journal.pone.0012708.

Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. Trends in Ecology and Evolution **15**, 496-503.

Zardoya R, Meyer A. 1998. Complete mitochondrial genome suggests diapsid affinities of turtles. Proceedings of the National Academy of Sciences of the United States of America **95**, 14226-14231.

Zhu K, Liang Y, Wu N, Guo H, Zhang N, Jiang S, Zhang D. 2017. Sequencing and characterization of the complete mitochondrial genome of Japanese Swellshark (*Cephalloscyllium umbratile*). Scientific Reports **7**: 15299.

Figure 1

Figure 1 The mitochondrial genome of *P. tentoria* . Direction of gene transcription is indicated by arrows.

Protein-coding genes are shown as blue arrows, rRNA genes as orchid arrows, tRNA genes as coral arrows and non-coding region as grey rectangle. The GC content is plotted using a black sliding window, GC-skew is plotted using green and orange color sliding window as the deviation from the average in the complete mitogenome .

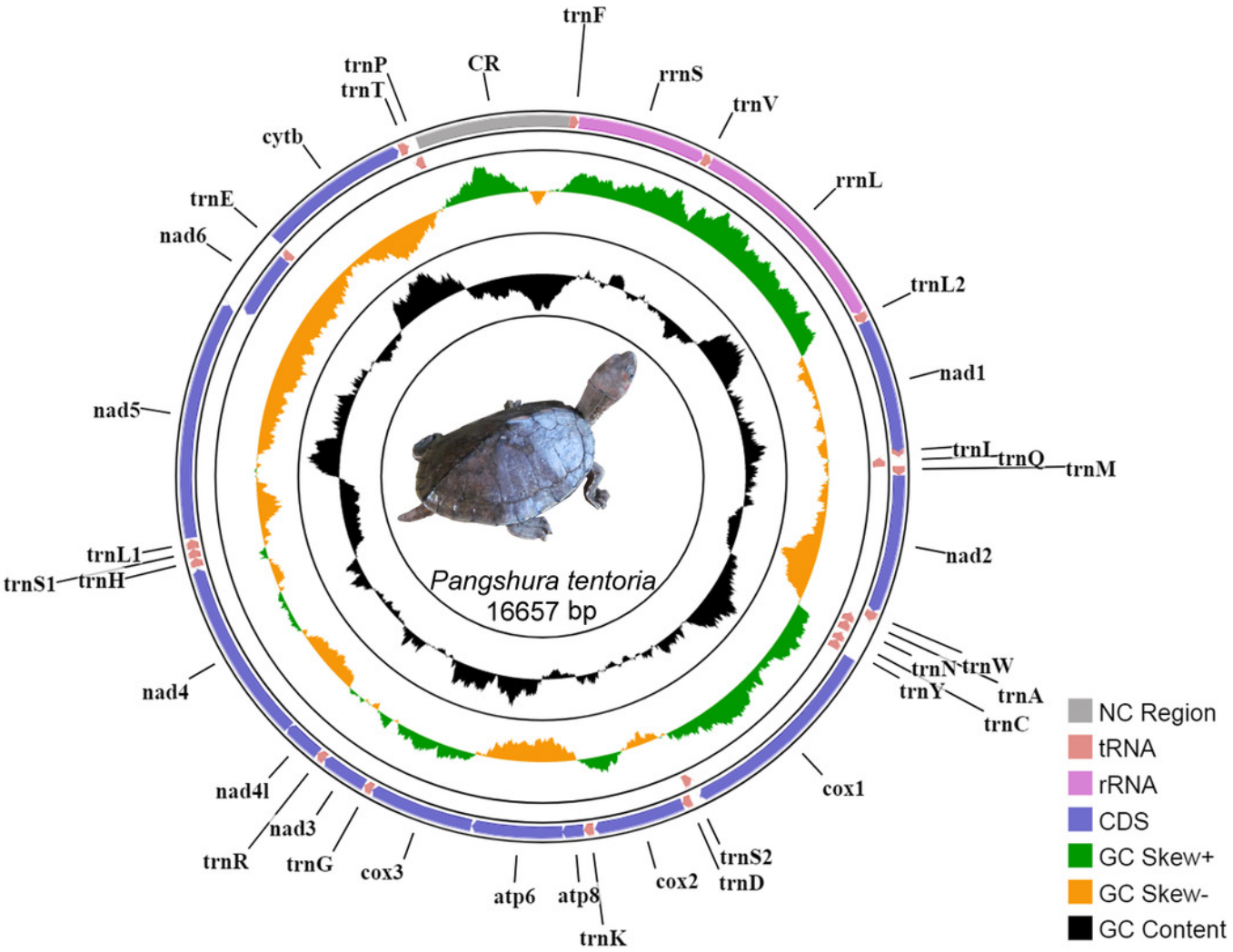

Figure 2

Figure 2 Comparative Relative synonymous codon usage (RSCU) in 32 geoemydid species including *P. tentoria* .

The cumulative RSCU values are represented on the y-axis while the codon families for each amino acid are represented on the x-axis.

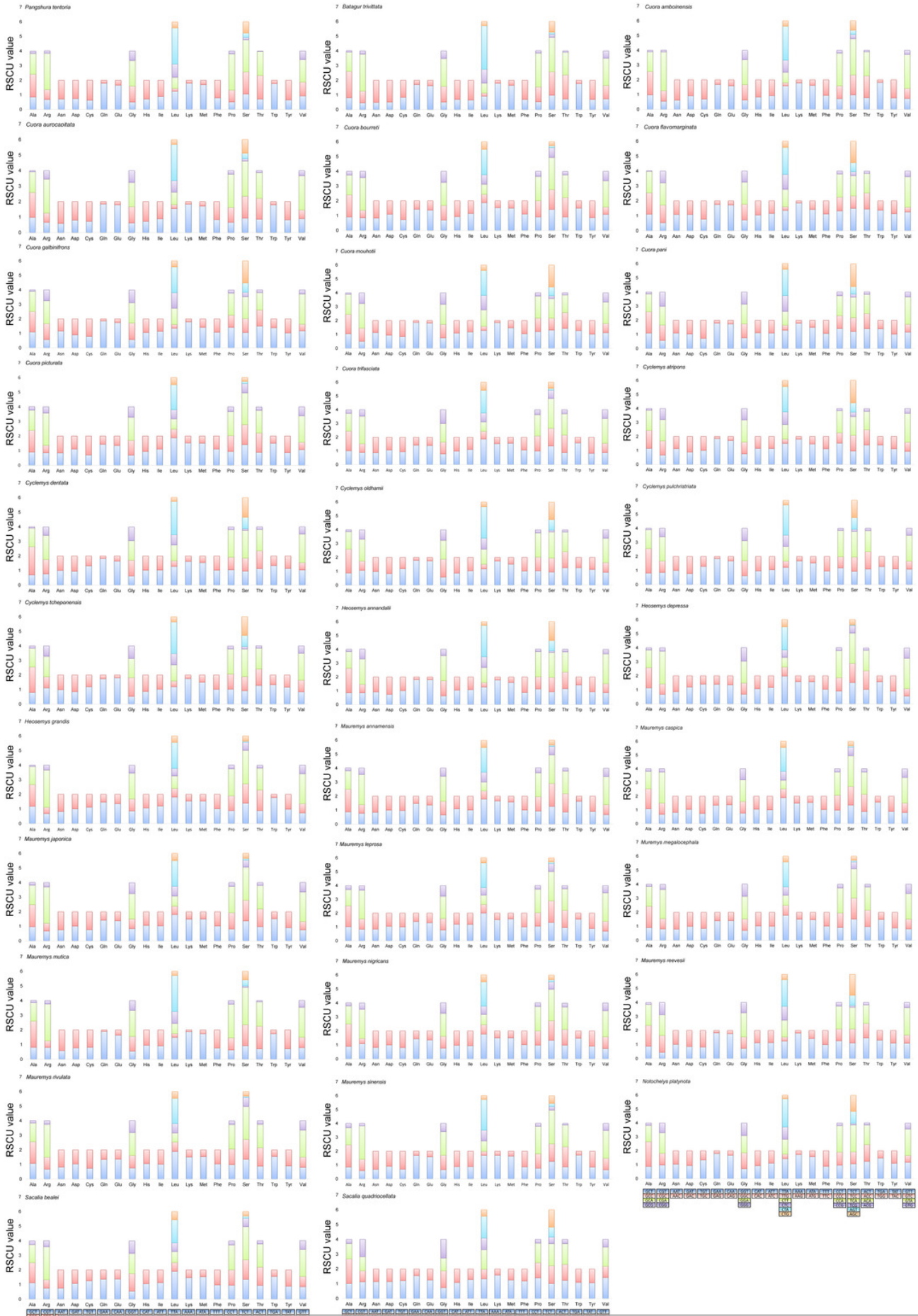

Figure 3

Figure 3 Comparison of codon usage within the 32 geoemydid species mitochondrial genome including *P. tentoria* .

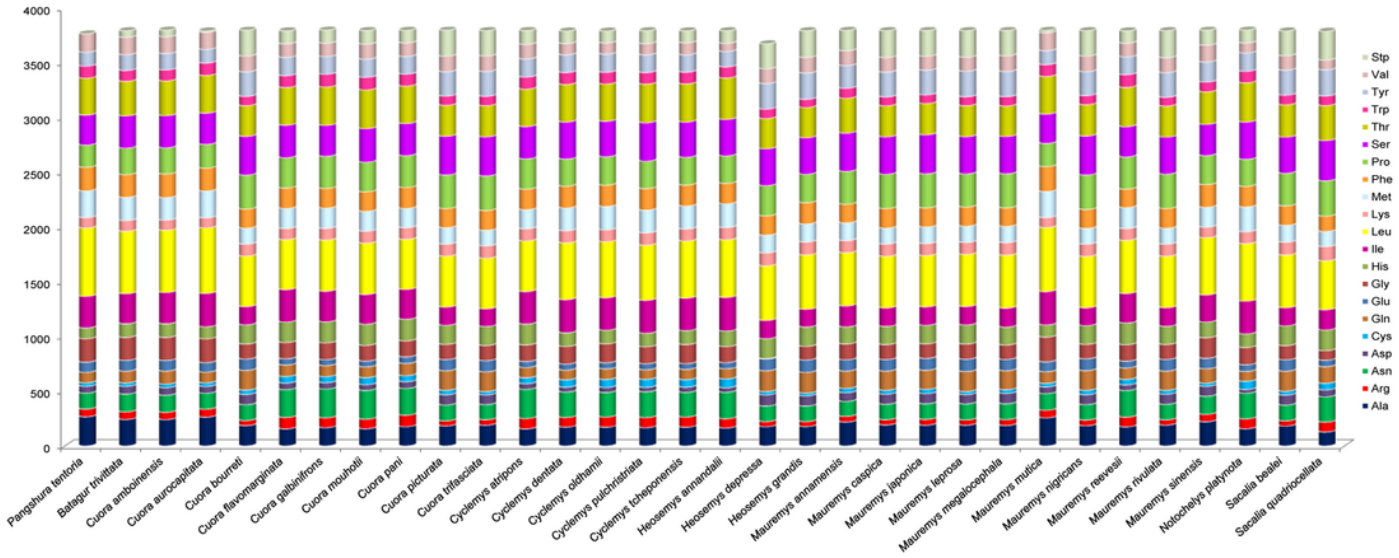

Figure 4

Figure 4 Ka/Ks ratios for the 13 mitochondrial protein-coding genes among *P. tentoria* and other geoemydid species representing seven genera.

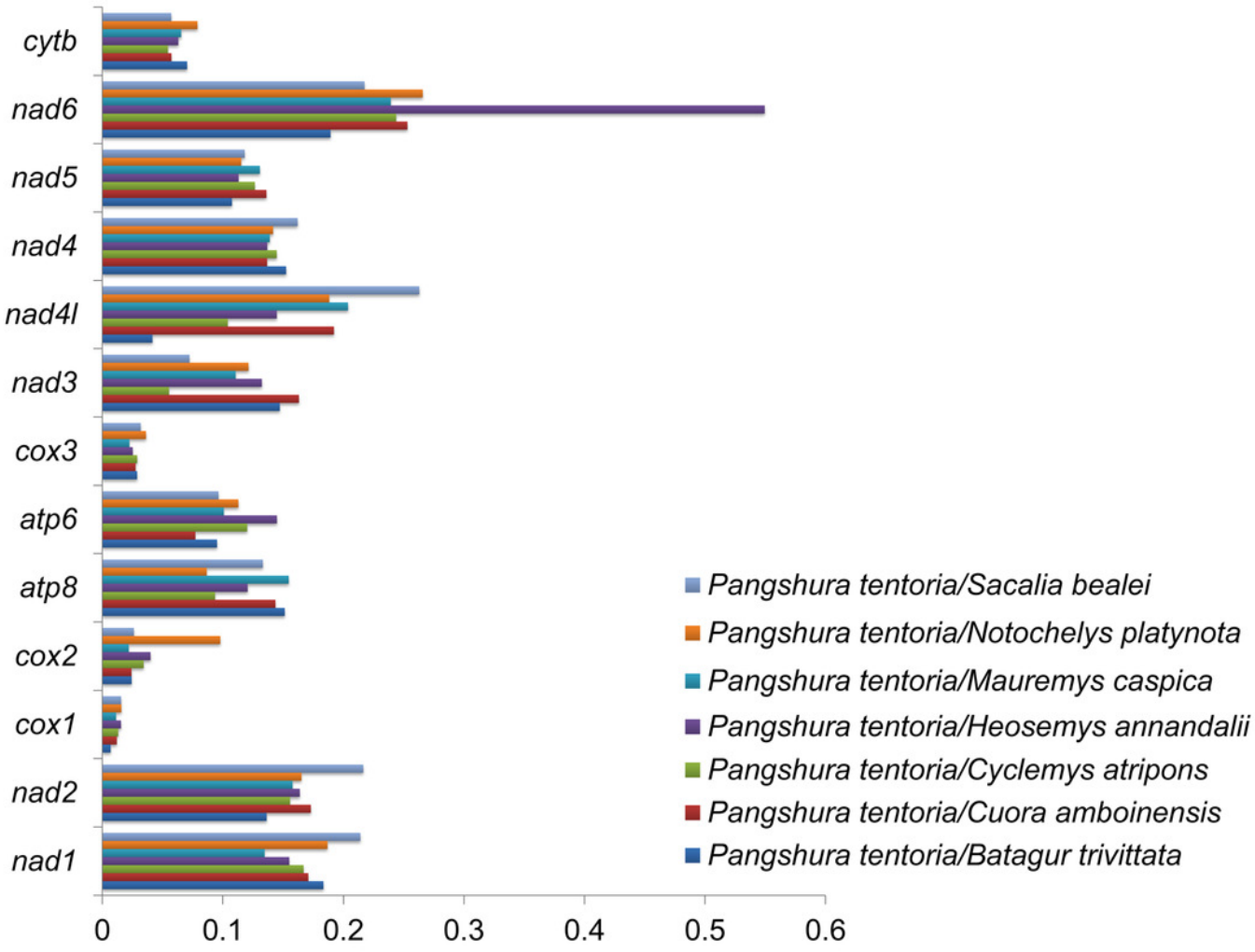

Figure 5

Figure 5 The structural organization of the control region of 32 geoemydid turtles mitogenomes.

The location and copy number of tandem repeats are shown by colored circles (Red, Green, and Violet). Non-repeat regions are indicated by blue colored box with sequence size inside.

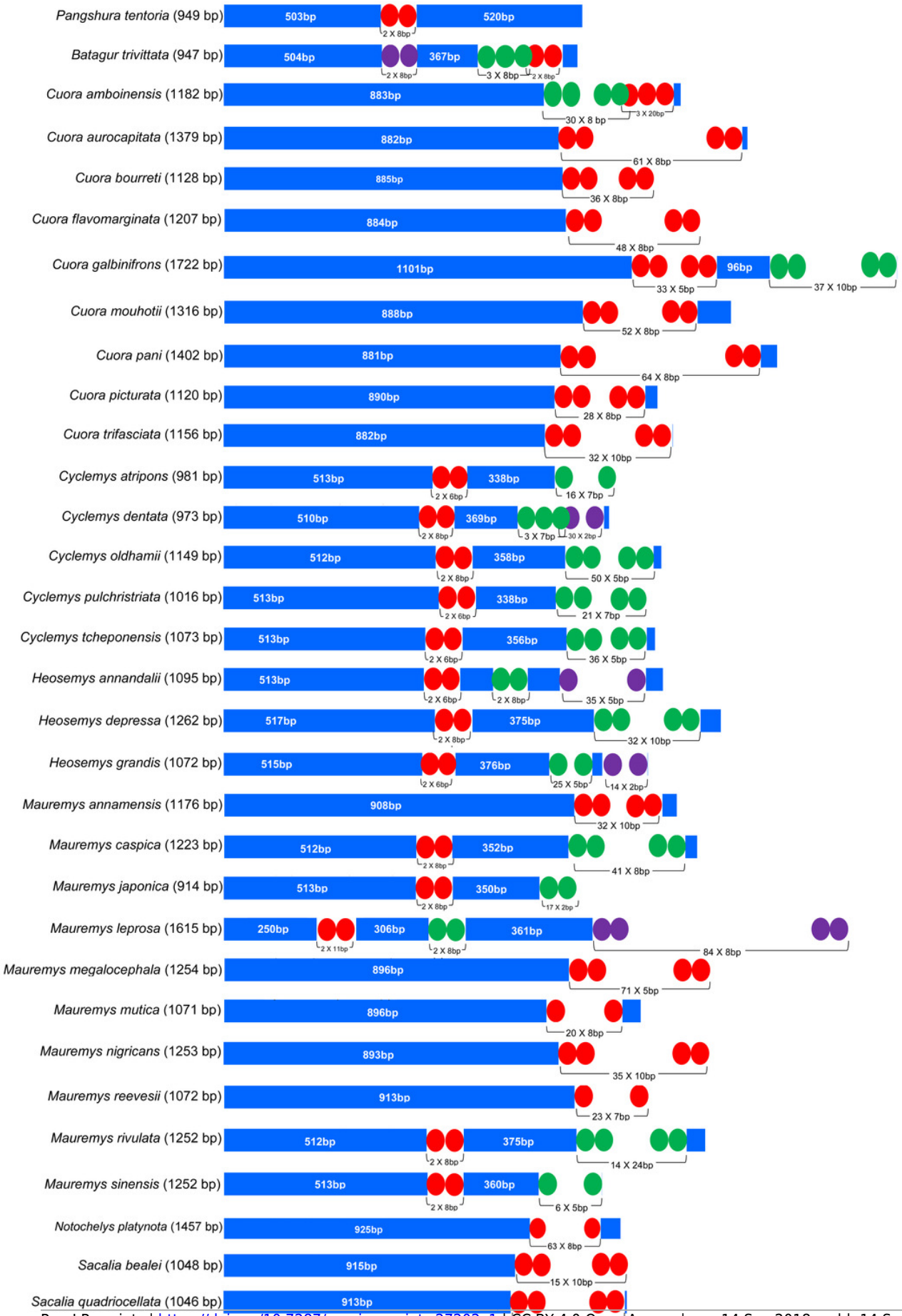
Pangshura tentoria (949 bp)
503bp
2 X 8bp
520bp
Batagur trivittata (947 bp)
504bp
2 X 8bp
367bp
3 X 8bp
2 X 8bp
Cuora amboinensis (1182 bp)
883bp
30 X 8 bp
3 X 20bp
Cuora aurocapitata (1379 bp)
882bp
61 X 8bp
Cuora bourreti (1128 bp)
885bp
36 X 8bp
Cuora flavomarginata (1207 bp)
884bp
48 X 8bp
Cuora galbinifrons (1722 bp)
1101bp
33 X 5bp
96bp
37 X 10bp
Cuora mouhotii (1316 bp)
888bp
52 X 8bp
Cuora pani (1402 bp)
881bp
64 X 8bp
Cuora picturata (1120 bp)
890bp
28 X 8bp
Cuora trifasciata (1156 bp)
882bp
32 X 10bp
Cyclemys atripons (981 bp)
513bp
2 X 6bp
338bp
16 X 7bp
Cyclemys dentata (973 bp)
510bp
2 X 8bp
369bp
3 X 7bp
30 X 2bp
Cyclemys oldhamii (1149 bp)
512bp
2 X 8bp
358bp
50 X 5bp
Cyclemys pulchristriata (1016 bp)
513bp
2 X 6bp
338bp
21 X 7bp
Cyclemys tcheponensis (1073 bp)
513bp
2 X 6bp
356bp
36 X 5bp
Heosemys annandalii (1095 bp)
513bp
2 X 6bp
2 X 8bp
35 X 5bp
Heosemys depressa (1262 bp)
517bp
2 X 8bp
375bp
32 X 10bp
Heosemys grandis (1072 bp)
515bp
2 X 6bp
376bp
25 X 5bp
14 X 2bp
Mauremys annamensis (1176 bp)
908bp
32 X 10bp
Mauremys caspica (1223 bp)
512bp
2 X 8bp
352bp
41 X 8bp
Mauremys japonica (914 bp)
513bp
2 X 8bp
350bp
17 X 2bp
Mauremys leprosa (1615 bp)
250bp
2 X 11bp
306bp
2 X 8bp
361bp
84 X 8bp
Mauremys megalocephala (1254 bp)
896bp
71 X 5bp
Mauremys mutica (1071 bp)
896bp
20 X 8bp
Mauremys nigricans (1253 bp)
893bp
35 X 10bp
Mauremys reevesii (1072 bp)
913bp
23 X 7bp
Mauremys rivulata (1252 bp)
512bp
2 X 8bp
375bp
14 X 24bp
Mauremys sinensis (1252 bp)
513bp
2 X 8bp
360bp
6 X 5bp
Notochelys platynota (1457 bp)
925bp
63 X 8bp
Sacalia bealei (1048 bp)
915bp
15 X 10bp
Sacalia quadriocellata (1046 bp)
913bp

Figure 6

Figure 6 Maximum Likelihood phylogenetic tree based on the concatenated nucleotide sequences of 13 PCGs of the 32 geoemydid turtles showing the evolutionary relationship of *P. tentoria* .

Species names, and GenBank accession numbers are indicated within parentheses with each node. Color boxes indicates the species clustering under respective families and clade of the studied *P. tentoria* and closely related *B. trivittata* was marked by green color box. The ML tree is drawn to scale with bootstrap support values were indicated along with the branches.

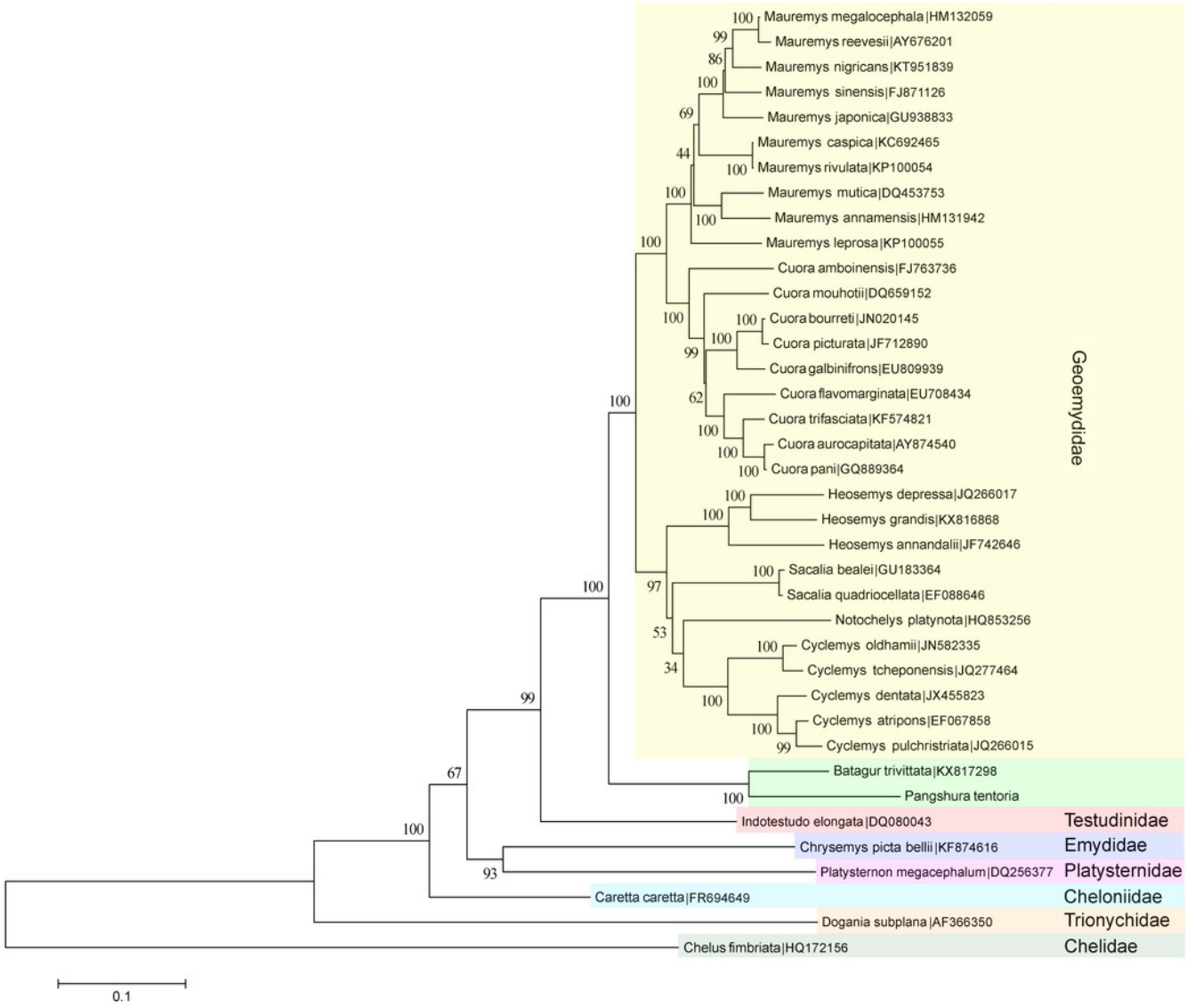

Table 1(on next page)

Table 1: List of annotated mitochondrial genes of *Pangshura tentoria*.

Table 1: List of annotated mitochondrial genes of *Pangshura tentoria*.

Gene	Direction	Location	Size (bp)	Anti codon	Start codon	Stop codon	Intergenic Nucleotides
trnF	+	203-271	69	GAA	.	.	0
rrnS	+	272-1235	964	.	.	.	0
trnV	+	1236-1304	69	TAC	.	.	-1
rrnL	+	1304-2901	1598	.	.	.	1
trnL2	+	2903-2978	76	TAA	.	.	0
nad1	+	2979-3938	960	.	ATG	(A)	8
trnI	+	3947-4017	71	GAT	.	.	-1
trnQ	-	4017-4087	71	TTG	.	.	-1
trnM	+	4087-4155	69	CAT	.	.	0
nad2	+	4156-5190	1035	.	ATG	(A)	4
trnW	+	5195-5267	73	TCA	.	.	1
trnA	-	5269-5337	69	TGC	.	.	1
trnN	-	5339-5412	74	GTT	.	.	27
trnC	-	5440-5505	66	GCA	.	.	0
trnY	-	5506-5577	72	GTA	.	.	1
cox1	+	5579-7114	1536	.	GTG	(A)	3
trnS2	-	7118-7188	71	GCT	.	.	0
trnD	+	7189-7258	70	GTC	.	.	0
cox2	+	7259-7936	678	.	ATG	(T)	10
trnK	+	7947-8020	74	TTT	.	.	1
atp8	+	8022-8183	162	.	ATG	(A)	-4
atp6	+	8180-8857	678	.	ATG	(A)	5
cox3	+	8863-9645	783	.	ATG	(TA)	1
trnG	+	9647-9714	68	TCC	.	.	1
nad3	+	9716-10064	354	.	ATG	(T)	1
trnR	+	10066-10134	69	TCG	.	.	0
nad4l	+	10135-10428	294	.	ATG	(TAA)	-4
nad4	+	10425-11795	1371	.	ATG	(A)	20
trnH	+	11816-11884	69	GTG	.	.	0
trnS1	+	11885-11951	67	TGA	.	.	-1
trnL1	+	11951-12023	73	TAG	.	.	12
nad5	+	12036-13823	1788	.	ATA	(A)	4
nad6	-	13828-14349	522	.	ATG	(AGA)	0
trnE	-	14350-14417	68	TTC	.	.	4
cytb	+	14422-15555	1134	.	ATG	(A)	10
trnT	+	15566-15637	72	TGT	.	.	0
trnP	-	15638-15708	71	TGG	.	.	0
A+T-rich Region		15709-16657 1-202	1151	.	.	-	.

Table 2(on next page)

Table 2. Nucleotide composition of the mitochondrial genome in different Geoemydid turtle's mtDNA .

The A+T biases of whole mitogenome, protein coding genes, tRNA, rRNA, and control regions were calculated by AT-skew = (A-T)/(A+T) and GC-skew= (G-C)/(G+C), respectively.

Table 2. Nucleotide composition of the mitochondrial genome in different Geoemydid turtle's mtDNA. The A+T biases of whole mitogenome, protein coding genes, tRNA, rRNA, and control regions were calculated by AT-skew = (A-T)/(A+T) and GC-skew= (G-C)/(G+C), respectively.

Species	Size	A%	T%	G%	C%	A+T%	AT-	GC-
Complete mitogenome								
Pangshura tentoria	16657	33.30	26.13	13.54	27.00	59.44	0.120	-0.331
Batagur trivittata	16463	33.60	24.52	13.25	28.62	58.12	0.156	-0.366
Cuora amboinensis	16708	33.82	26.74	13.05	26.36	60.57	0.116	-0.337
C. aurocapitata	16890	33.56	27.41	13.04	25.97	60.98	0.100	-0.331
C. bourreti	16649	33.90	26.84	13.05	26.19	60.75	0.116	-0.334
C. flavomarginata	16721	33.99	27.67	12.80	25.52	61.67	0.102	-0.331
C. galbinifrons	17244	34.12	27.58	12.47	25.81	61.70	0.106	-0.348
C. mouhotii	16837	34.03	27.33	12.81	25.81	61.37	0.109	-0.336
C. pani	16922	33.67	27.43	13.00	25.89	61.10	0.102	-0.331
C. picturata	16623	33.95	26.89	13.00	26.14	60.85	0.116	-0.335
C. trifasciata	16675	33.84	26.83	13.12	26.18	60.68	0.115	-0.332
Cyclemys atripons	16500	34.40	27.20	13.01	25.36	61.62	0.117	-0.321
C. dentata	16484	34.28	27.22	13.08	25.41	61.50	0.114	-0.320
C. oldhami	16656	34.35	26.83	13.10	25.71	61.18	0.122	-0.324
C. pulchristriata	16527	34.38	27.19	12.98	25.43	61.57	0.116	-0.324
C. tcheponensis	16593	34.20	26.77	13.19	25.83	60.97	0.121	-0.323
Heosemys annandalii	16604	35.14	26.71	12.27	25.87	61.85	0.136	-0.356
H. depressa	16773	35.00	27.52	12.53	24.93	62.52	0.119	-0.330
H. grandis	16581	34.70	27.67	12.52	25.09	62.38	0.112	-0.334
Mauremys	16844	33.70	26.85	13.04	26.38	60.56	0.113	-0.338
M. caspica	16741	34.04	27.17	12.91	25.87	61.21	0.112	-0.334
M. japonica	16443	34.02	26.45	13.01	26.50	60.48	0.125	-0.341
M. leprosa	17066	34.41	27.48	12.43	25.66	61.90	0.111	-0.347
M. megalocephala	16783	34.05	27.20	12.81	25.92	61.25	0.111	-0.338
M. mutica	16609	33.81	26.50	13.17	26.49	60.32	0.121	-0.335
M. nigricans	16779	34.07	26.85	12.96	26.09	60.93	0.118	-0.336
M. reevesii	16576	33.99	26.62	12.94	26.44	60.61	0.121	-0.342
M. rivulata	16766	34.31	26.91	12.94	25.83	61.22	0.120	-0.332
M. sinensis	16461	33.81	26.20	13.17	26.79	60.02	0.126	-0.340
Notochelys platynota	16981	34.39	28.10	12.24	25.25	62.49	0.100	-0.347
Sacalia bealei	16561	34.18	26.86	13.06	25.88	61.04	0.119	-0.329
S. quadriocellata	16816	34.13	26.75	13.16	25.94	60.88	0.121	-0.326
Protein Coding genes (PCGs)								
Pangshura tentoria	11295	30.78	27.73	13.51	27.96	58.52	0.052	-0.348
Batagur trivittata	11379	31.37	26.07	13.05	29.49	57.44	0.092	-0.386
Cuora amboinensis	11397	31.44	28.05	13.07	27.42	59.49	0.057	-0.354
C. aurocapitata	11373	31.18	28.11	13.35	27.33	59.30	0.051	-0.343
C. bourreti	11394	31.44	28.41	13.23	26.90	59.86	0.050	-0.340

C. flavomarginata	11377	31.41	29.10	13.03	26.44	60.51	0.038	-0.339
C. galbinifrons	11399	31.51	28.52	13.13	26.83	60.03	0.049	-0.342
C. mouhotii	11387	31.57	28.74	13.13	26.53	60.32	0.047	-0.337
C. pani	11393	31.10	28.28	13.29	27.30	59.39	0.047	-0.345
C. picturata	11395	31.51	28.52	13.17	26.79	60.03	0.049	-0.340
C. trifasciata	11382	31.28	28.31	13.31	27.08	59.60	0.049	-0.341
Cyclemys atripons	11387	31.88	29.16	13.05	25.88	61.05	0.044	-0.329
C. dentata	11376	31.82	29.16	13.09	25.91	60.98	0.043	-0.328
C. oldhami	11370	31.34	28.78	13.35	26.50	60.13	0.042	-0.329
C. pulchristriata	11380	31.76	29.04	13.06	26.12	60.80	0.044	-0.333
C. tcheponensis	11377	31.33	28.80	13.41	26.44	60.13	0.042	-0.327
Heosemys annandalii	11380	32.65	28.31	12.12	26.90	60.96	0.071	-0.378
H. depressa	11382	32.26	29.24	12.77	25.71	61.50	0.048	-0.336
H. grandis	11379	32.56	29.39	12.38	25.65	61.96	0.051	-0.348
Mauremys	11391	31.34	28.05	13.15	27.44	59.39	0.055	-0.351
M. caspica	11382	31.70	28.40	13.00	26.88	60.11	0.054	-0.348
M. japonica	11385	31.87	28.18	12.84	27.09	60.06	0.061	-0.356
M. leprosa	11382	31.92	28.73	12.89	26.43	60.66	0.052	-0.344
M. megalocephala	11385	31.62	28.49	13.02	26.85	60.11	0.052	-0.346
M. mutica	11392	31.43	27.94	13.29	27.33	59.37	0.058	-0.345
M. nigricans	11382	31.50	28.14	13.18	27.15	59.65	0.056	-0.346
M. reevesii	11377	31.81	28.02	13.04	27.11	59.84	0.063	-0.350
M. rivulata	11382	31.68	28.41	13.03	26.86	60.09	0.054	-0.346
M. sinensis	11395	31.68	27.86	13.04	27.40	59.54	0.064	-0.354
Notochelys platynota	11398	32.12	29.47	12.50	25.89	61.60	0.043	-0.348
Sacalia bealei	11373	31.82	28.57	12.92	26.67	60.39	0.053	-0.347
S. quadriocellata	11366	31.76	28.56	12.96	26.70	60.32	0.052	-0.346
tRNA genes								
Pangshura tentoria	1551	30.94	29.33	20.95	18.76	60.28	0.026	0.055
Batagur trivittata	1551	30.75	29.27	20.88	19.08	60.02	0.024	0.045
Cuora amboinensis	1608	32.46	30.09	19.21	18.22	62.56	0.037	0.026
C. aurocapitata	1796	32.01	30.23	19.04	18.70	62.24	0.028	0.008
C. bourreti	1553	32.13	29.62	19.63	18.60	61.75	0.040	0.026
C. flavomarginata	1553	32.38	30.00	19.51	18.09	62.39	0.038	0.037
C. galbinifrons	1552	32.02	29.25	19.78	18.94	61.27	0.045	0.021
C. mouhotii	1552	32.02	29.83	19.65	18.49	61.85	0.035	0.030
C. pani	1554	32.23	29.60	19.49	18.66	61.84	0.042	0.021
C. picturata	1553	32.19	29.55	19.63	18.60	61.75	0.042	0.026
C. trifasciata	1553	31.74	29.74	19.76	18.73	61.49	0.032	0.026
Cyclemys atripons	1551	32.10	30.10	19.66	18.11	62.21	0.032	0.040
C. dentata	1548	32.55	30.03	19.25	18.15	62.59	0.040	0.029
C. oldhami	1551	32.62	30.04	19.21	18.11	62.66	0.041	0.029
C. pulchristriata	1551	32.10	30.36	19.66	17.85	62.47	0.027	0.048
C. tcheponensis	1606	33.37	30.57	18.67	17.37	63.94	0.043	0.036

Heosemys annandalii	1550	32.25	29.87	19.35	18.51	62.12	0.038	0.022
H. depressa	1549	31.76	29.69	20.07	18.46	61.45	0.033	0.041
H. grandis	1549	32.27	29.82	19.49	18.39	62.10	0.039	0.028
Mauremys	1496	32.41	30.08	19.18	18.31	62.50	0.037	0.023
M. caspica	1554	32.17	29.66	19.62	18.53	61.84	0.040	0.028
M. japonica	1557	32.24	29.60	19.52	18.62	61.84	0.042	0.023
M. leprosa	1552	32.73	29.51	19.13	18.62	62.24	0.051	0.013
M. megalocephala	1554	32.36	29.72	19.49	18.40	62.09	0.042	0.028
M. mutica	1553	32.58	30.13	19.18	18.09	62.71	0.039	0.029
M. nigricans	1555	32.60	29.58	19.35	18.45	62.18	0.048	0.023
M. reevesii	1547	32.25	29.99	19.52	18.22	62.24	0.036	0.034
M. rivulata	1551	32.17	29.91	19.47	18.43	62.08	0.036	0.027
M. sinensis	1555	32.15	29.71	19.67	18.45	61.86	0.039	0.032
Notochelys platynota	1551	32.49	29.98	19.27	18.24	62.47	0.040	0.027
Sacalia bealei	1549	32.08	29.89	20.01	18.01	61.97	0.035	0.052
S. quadriocellata	1548	31.97	29.84	20.09	18.08	61.82	0.034	0.052
rRNA genes								
Pangshura tentoria	2562	37.23	21.62	17.36	23.77	58.86	0.265	-0.155
Batagur trivittata	2568	37.26	20.52	17.44	24.76	57.78	0.289	-0.173
Cuora amboinensis	2572	37.67	21.38	16.95	23.98	59.05	0.275	-0.171
C. aurocapitata	2577	37.91	21.57	16.99	23.51	59.48	0.274	-0.160
C. bourreti	2571	38.11	21.31	16.60	23.95	59.43	0.282	-0.181
C. flavomarginata	2562	38.09	22.24	16.66	22.98	60.34	0.262	-0.159
C. galbinifrons	2571	38.23	21.89	16.25	23.60	60.13	0.271	-0.184
C. mouhotii	2570	38.13	21.67	16.65	23.54	59.80	0.275	-0.171
C. pani	2568	37.96	21.53	16.97	23.52	59.50	0.276	-0.161
C. picturata	2553	38.22	21.38	16.56	23.81	59.61	0.282	-0.179
C. trifasciata	2568	38.04	21.30	16.78	23.87	59.34	0.282	-0.174
Cyclemys atripons	2561	38.73	21.98	16.39	22.88	60.71	0.275	-0.165
C. dentata	2565	38.55	21.94	16.56	22.92	60.50	0.274	-0.160
C. oldhami	2569	38.45	21.44	16.77	23.31	59.90	0.283	-0.163
C. pulchristriata	2564	38.72	21.95	16.41	22.89	60.68	0.276	-0.164
C. tcheponensis	2576	38.31	21.35	16.73	23.60	59.66	0.284	-0.170
Heosemys annandalii	2563	39.32	22.27	16.07	22.31	61.60	0.276	-0.162
H. depressa	2565	38.71	22.84	16.21	22.22	61.55	0.257	-0.156
H. grandis	2566	38.73	22.36	16.32	22.56	61.10	0.267	-0.160
Mauremys	2715	37.56	21.76	16.64	24.01	59.33	0.266	-0.181
M. caspica	2568	37.88	21.65	16.78	23.67	59.54	0.272	-0.170
M. japonica	2570	37.93	21.78	16.69	23.57	59.72	0.270	-0.171
M. leprosa	2567	38.05	21.97	16.67	23.29	60.03	0.268	-0.165
M. megalocephala	2574	37.91	21.91	16.55	23.62	59.82	0.267	-0.176
M. mutica	2568	37.65	21.80	16.93	23.59	59.46	0.266	-0.164
M. nigricans	2570	38.21	21.59	16.34	23.85	59.80	0.277	-0.186
M. reevesii	2573	37.89	21.99	16.51	23.59	59.89	0.265	-0.176

M. rivulata	2567	37.78	21.58	16.86	23.76	59.36	0.272	-0.169
M. sinensis	2570	37.50	21.43	16.88	24.16	58.94	0.272	-0.177
Notochelys platynota	2573	38.55	22.46	16.28	22.69	61.01	0.263	-0.164
Sacalia bealei	2574	38.11	21.87	16.70	23.31	59.98	0.270	-0.165
S. quadriocellata	2859	37.46	21.93	16.96	23.64	59.39	0.261	-0.164
Control regions								
Pangshura tentoria	949	32.03	34.03	12.96	20.96	66.06	-0.030	-0.236
Batagur trivittata	947	31.67	33.26	12.98	22.06	64.94	-0.024	-0.259
Cuora amboinensis	1182	33.16	40.27	10.74	15.82	73.43	-0.096	-0.191
C. aurocapitata	1379	33.06	43.65	9.35	13.92	76.72	-0.138	-0.196
C. bourreti	1128	32.53	39.53	10.72	17.19	72.07	-0.097	-0.231
C. flavomarginata	1207	33.88	40.84	9.61	15.65	74.73	-0.093	-0.239
C. galbinifrons	1722	34.90	41.28	6.79	17.01	76.19	-0.083	-0.429
C. mouhotii	1316	33.66	39.81	8.81	17.70	73.48	-0.083	-0.335
C. pani	1402	33.16	44.15	8.98	13.69	77.31	-0.142	-0.207
C. picturata	1120	32.50	39.10	10.71	17.67	71.60	-0.092	-0.245
C. trifasciata	1156	33.91	39.79	10.03	16.26	73.70	-0.079	-0.236
Cyclemys atripons	981	34.76	34.76	12.13	18.34	69.52	0	-0.204
C. dentata	973	33.09	35.25	12.84	18.80	68.34	-0.031	-0.188
C. oldhami	1149	37.94	34.72	10.79	16.53	72.67	0.044	-0.210
C. pulchristriata	1016	35.62	35.62	11.61	17.12	71.25	0	-0.191
C. tcheponensis	1073	36.25	34.20	11.64	17.89	70.45	0.029	-0.211
Heosemys annandalii	1095	34.79	36.43	10.59	18.17	71.23	-0.023	-0.263
H. depressa	1262	37.71	37.55	9.35	15.37	75.27	0.002	-0.243
H. grandis	1072	31.34	39.08	11.38	18.19	70.42	-0.109	-0.230
Mauremys	1176	32.48	40.39	10.54	16.58	72.87	-0.108	-0.222
M. caspica	1223	33.44	41.29	9.89	15.37	74.73	-0.105	-0.216
M. japonica	914	32.38	33.80	13.01	20.78	66.19	-0.021	-0.229
M. leprosa	1615	34.61	38.76	7.55	19.07	73.37	-0.056	-0.432
M. megalocephala	1254	33.97	40.19	9.56	16.26	74.16	-0.083	-0.259
M. mutica	1071	32.49	37.44	11.20	18.86	69.93	-0.070	-0.254
M. nigricans	1253	35.27	38.54	10.05	16.12	73.82	-0.044	-0.231
M. reevesii	1072	34.32	34.32	11.28	20.05	68.65	0	-0.279
M. rivulata	1252	37.53	37.22	10.06	15.17	74.76	0.004	-0.202
M. sinensis	935	31.65	34.75	12.40	21.17	66.41	-0.046	-0.261
Notochelys platynota	1457	32.18	40.28	8.30	19.21	72.47	-0.111	-0.396
Sacalia bealei	1048	32.72	36.54	11.92	18.79	69.27	-0.055	-0.223
S. quadriocellata	1046	33.26	36.23	11.56	18.92	69.50	-0.042	-0.241

4

5

6