Complete mitogenome of the endangered and endemic Nicobar treeshrew (*Tupaia nicobarica*) and comparison with other Scandentians

Shantanu Kundu 1, Avas Pakrashi 1, Manokaran Kamalakannan 2, Devkant Singha 1, Kaomud Tyagi 1, Dhiruti Banerjee 1,2, Chinnadurai Venkatraman 2 & Vikas Kumar 1*

The Nicobar treeshrew (*Tupaia nicobarica*) is an endangered small mammal endemic to the Nicobar Island of the Andaman Sea, India regarded as an alternative experimental animal model in biomedical research. The present study aimed to assemble the first mitochondrial genome of *T. nicobarica* to elucidate its phylogenetic position with respect to other Scandentians. The structure and variation of the novel mitochondrial genome were analyzed and compared with other Scandentians. The complete mitogenome (17,164 bp) encodes 37 genes, including 13 protein-coding genes (PCGs), 22 transfer RNA (tRNAs), two ribosomal RNA (rRNAs), and one control region (CR). Most of the genes were encoded on majority strand, except *nad6* and eight tRNAs. The nonsynonymous/synonymous ratio in all PCGs indicates strong negative selection among all Tupaiidae species. The comparative study of CRs revealed the occurrence of tandem repeats (CGTACA) found in *T. nicobarica*. The phylogenetic analyses (Maximum Likelihood and Bayesian Inference) showed distinct clustering of *T. nicobarica* with high branch supports and depict a substantial divergence time (12–19 MYA) from the ancestor lineage of Tupaiidae. The 16S rRNA dataset corroborates the taxonomic rank of two subspecies of *T. nicobarica* from the Great and Little Nicobar Islands. In the future, whole nuclear genome sequencing is necessary to further improve our understanding of evolutionary relationships among treeshrews, and will have implications for biomedical research.

The world treeshrew account for 23 species under four genera (*Anathana*, *Dendrogale*, *Ptilocercus* and *Tupaia*) of two families (*Tupaiidae* and *Ptilocercidae*) which are distributed in South Asia, Southeast Asia, and Southwest China 1,2. The mainland of India is known by two species, the Madras treeshrew, *Anathana ellioti* and the Northern treeshrew, *Tupaia belangeri*. However, the Nicobar treeshrew, *Tupaia nicobarica* is endemic to the Nicobar Islands 1. This species is categorized as 'Endangered' species in the IUCN Red List of Threatened Species and listed under ‘Appendix II’ in the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) 2. The treeshrews, including *T. nicobarica* are arboreal in nature and live in different forest types including scrub jungle, moist deciduous forests, and montane sholas. The members of *Anathana*, *Dendrogale*, and *Tupaia* are diurnal and solitary in habit; however the *Ptilocercus* species is nocturnal and live in family groups 4.

The biogeography of India is mainly classified into two categories, the mainland and groups of islands (Lakshadweep and Andaman-Nicobar). The Andaman and Nicobar (AN) archipelago comprises 572 islands located on the Bay of Bengal 5. The AN archipelago was formed due to collision between the Indian Plate and Eurasian Plate which commenced about 50 million years ago and continues today 6. Due to the distance of the archipelago from the mainland and different diachronic processes, this group of islands accommodates numerous unique elements of biodiversity 7,8. This archipelago is a trenchant biogeographic entity that can be a model for evolutionary studies 9,10. The study on faunal diversity of AN archipelago has been started since nineteenth century, considering the large size or charismatic vertebrate fauna, including mammals, birds, and herpetofauna 9,10,11. Due to the

1Centre for DNA Taxonomy, Molecular Systematics Division, Zoological Survey of India, Calcutta 700053, India. 2Mammal and Osteology Section, Zoological Survey of India, Calcutta 700053, India. Email: vikaszi77@gmail.com
remoteness and inaccessibility throughout the year, most of the regions of AN archipelago are sparsely explored. Remarkably, these oceanic islands also provide a suitable habitat for many smaller mammals like, tree shrews (order Scandentia), shrews (order Eulipotyphla), and rodents (order Rodentia) due to their preferable spatial niche and carrying capacity13. Among them, the Great and Little Nicobar Islands are located about 1800 km east from mainland India, 600 km west of Thailand, 950 km north of Myanmar, and 180 km of Sumatra. On the basis of geographical distribution, two subspecies have been recognized, viz., *T. nicobarica nicobarica* from the Great Nicobar Island and *T. nicobarica surda* from the Little Nicobar Island14. However, the status of these subspecies has been endorsed pending further molecular studies3,15,16.

Remarkably, the zoonotic disease has been originated or transmitted through different mammalian species including tree shrews and cause life threatening to human beings throughout the globe17,18. Hence, the tree shrews species is evidenced to be a significant model for various human disorders like, depression, myopia, hepatitis B and C virus infections, and hepatocellular carcinoma19,20. The molecular study and genome sequencing has been demonstrated the genetic basis of signaling pathways in nervous and immune systems of the Chinese tree shrew and evidenced as a potential model for biomedical research21.

The characteristics of complete mitogenomes are widely used genomics approaches for systematics studies and evolutionary research of wider group of taxa including mammals22-24. So far, the complete mitogenome of five species, *Tupaia belangeri*, *Tupaia minor*, *Tupaia montana*, *Tupaia splendidula*, and *Tupaia tana* were generated from different geographical regions. The molecular studies were previously aimed to infer the phylogenetic and evolutionary relationship, genetic structure, and possible gene flow of Scandentia across their range distribution in Southeast Asia25-28. In addition, the partial mitochondrial genes (12S rRNA, 16S rRNA, and Cyb) were also utilized to elucidate the phylogenetic position and diversification of Tupaiidae species including *T. nicobarica*15,27,28. However, the in-depth genetic information and structural motifs of *T. nicobarica* mitogenome is still anonymous to the scientific communities. To fill the gap of knowledge, the present study aimed to determine the complete mitogenome of *T. nicobarica* from AN archipelago, India. The comparative analyses were confronted to check the structure and variation within the Tupaiidae mitogenomes. The phylogenetic analyses and divergence time were estimated to infer the evolutionary relationship of *T. nicobarica* comparing with other Tupaiidae species. Further, an additional dataset of mitochondrial 16S rRNA was also constructed to clarify the taxonomic rank of the extant subspecies of *T. nicobarica* in the Great and Little Nicobar Island.

Materials and methods

**Sample collection, mitochondria separation, and DNA extraction.** The museum sample of *T. nicobarica* is vouchered in 70% ethanol at the National Zoological Collections of Andaman and Nicobar Regional Centre, Zoological Survey of India, which was collected from the Galathia, Campbell bay (06.83 N 93.87 E) on 30th November 2018 (contact person: Govindaraja Gokulakrishnan, Email: gokul7701@gmail.com). No tree shrew was killed as the collected specimen was a natural kill; hence no prior permission was required in the present study. Before preserving, the muscle tissue sample was aseptically collected from the hind leg of the specimen with ample attention and stored in 70% ethanol at Mammal and Osteology section, Zoological Survey of India, Kolkata under voucher ID (Reg. No. 28532) for downstream molecular investigation. The collection locality map was prepared by the online platform (https://noaa.maps.arcgis.com). We used whole mitochondria for the extraction of mitochondrial DNA as per standard protocol29. The tissue sample was homogenized with 1 ml buffer containing 0.32 M Sucrose, 1 mM EDTA, and 10 mM Tris•HCl by the WiseTis HG-15 homogenizer. To remove the nuclei and cell debris, the working mixture was centrifuged at 700 g for 5 min at 4 °C. The supernatant was collected in 1.5 ml centrifuge tube and centrifuged at 12,000 g for 10 min at 4 °C to precipitate the mitochondrial pellet. The pellet was re-suspended in 500 µl of buffer (50 mM Tris•HCl, 25 mM of EDTA, 150 mM NaCl) and incubated at 56 °C for 1-2 h along with 20 µl of proteinase K (20 mg/ml). The mitochondrial DNA was extracted by using QIAamp DNA Investigator Kit (QIAGEN Inc.) and the eluted volume was reduced to 100 µl to increase the mtDNA concentration. The quality of the extracted mtDNA was checked through 1% agarose gel electrophoresis, and the concentration was quantified with a NANO DROP 2000 spectrophotometer (Thermo Scientific).

**Sequencing, assembly and annotation.** The complete mitogenome sequence and assembly were carried out at PHIXGEN Pvt. Ltd, Gurugram, India (http://www.phixgen.com). The mitochondrial DNA (> 100 ng) was used in Illumina TruSeq Nano DNA HT library preparation kit for library assembly (Illumina, Inc, USA). The mtDNA was fragmented by ultra-sonication (Covaris M220, Covaris Inc., Woburn, MA, USA) and the A-tailed fragments were joined with the sequencing indexed adapters done by the Illumina kit. The mtDNA fragments (450 bp) were selected through sample purification. The amplified PCR library was examined using Bioanalyzer 2100 (Agilent Technologies, Inc., Waldbronn, Germany) with high sensitivity DNA chips. Total > 4 million raw reads were generated through Illumina NextSeq500 (150 × 2 chemistry) (Illumina, Inc, USA). The high-quality reads were downsampled to 2 million using Seqtk (https://github.com/lh3/seqtk) and iterative assembly was performed by using NOVOPlasty v2.6.7 using default parameters30. The mitogenome of *T. belangeri* (accession no. NC_002521) was used as a reference seed to start the assembly. The typical circular representation of the generated mitogenome of *T. nicobarica* was plotted by CGView Server (http://stothard.afns.ualberta.ca/cgview_server/) with default parameters31. Further, the contig was subjected to confirmation by the MITOS v806 online webserver (http://mitos.bioinf.uni-leipzig.de). The direction and arrangements of PCGs, tRNAs, and rRNAs were confirmed through MITOS online server (http://mitos.bioinf.uni-leipzig.de)32. The start and stop codons of each PCG were assured through the Open Reading Frame Finder web tool (https://www.ncbi.nlm.nih.gov/orffinder/) on the basis of vertebrate mitochondrial genetic code and other publicly
available reference sequences of Tupaiidae. The mitogenome was submitted to the GenBank database (Accession No. MW751815) using the NCBI Bankit submission tool.

**Dataset construction and comparative analyses.** On the basis of taxonomic classification, the mitogenomes of five Tupaiidae species were downloaded from GenBank and merged in the dataset for comparative analysis (Supplementary Table S1). The genome sizes and nucleotide compositions of all the studied species were calculated using MEGA X. To calculate the base composition skew, we utilized previously known formulas: AT skew = (A − T)/(A + T), GC skew = (G − C)/(G + C)[34]. The overlapping regions and intergenic spacers of *T. nicobarica* and other Tupaiidae species mitogenomes were calculated manually. The pairwise test of the synonymous (Ks) and non-synonymous (Ka) substitutions were calculated between *T. nicobarica* and other Tupaiidae species using DnaSPv6.0[35]. The comparative analysis of Relative Synonymous Codon Usage (RSCU) and relative abundance of amino acids were also calculated using MEGA X. The secondary structures of tRNA genes were affirmed by tRNAscan-SE Search Server 2.0 (http://lowelab.ucsc.edu/tRNAscan-SE/) and ARWEN 1.2[36, 37]. To speculate the putative domains and motif, the CR of *T. nicobarica* and other Tupaiidae species was screened from the database. The tandem repeats within the CR were predicted by the online Tandem Repeats Finder web tool (https://tandem.bu.edu/tra/TRA.html)[38].

**Phylogenetic analysis and divergence time estimation.** To assess the phylogenetic relationship, the dataset was constructed with the representatives of Scandentia, Dermoptera, Primates, Lagomorphs, and Rodents based on the previous literatures[15, 25, 39, 40]. The 13 PCGs of 14 mitogenomes were aligned and concatenated using TranslatorX (with MAFFT algorithm with L-INS-i strategy and GBlocks parameters) and SequenceMatrix v1.7.8457[41]. The best fit model (GTR + I + G) was estimated by PartitionFinder 2 using lowest Bayesian information criterion (BIC) criterion[42] and the maximum-likelihood (ML) tree was constructed using the IQ-Tree web server with 1000 bootstrap support[43]. The estimation of divergence times among Tupaiidae species were calculated by Bayesian relaxed clock method in BEAST v2.4.7[45]. The GTR + I + G substitution model, empirical base frequencies, and relaxed uncorrelated log-normal clock with the Yule speciation model was applied as Tree prior. A total of four fossil calibration points were applied in the phylogeny to constraint the analysis as described in the previous study[46-48]. (1) node A, 18 million years ago (MYA) log-normal prior as the minimum age of *Tupaiata* based on the fossil of *T. miocenica*, (2) node B, 23 MYA log-normal prior as the minimum age for the split between *Pygathris* and *Hylobates*, (3) node C, a normal prior for the primate outgroups with a mean of 77.5 MYA, and (4) node D, the total tree height is considered as a normal prior with a mean of 90 MYA for Scandentia, Dermoptera, and Primates. Two independent Markov chain Monte Carlo (MCMC) runs were performed for 1,000,000 generations with 25% burn-in and trees sampled every 2000 generations. To combine these runs, log files were imported into LogCombiner of the BEAST package. The adequate MCMC mixing and convergence were estimated using the effective sample size (ESS) values (> 200 in all parameters) in Tracer v1.7.1[49]. To reconstruct the Bayesian phylogeny, TreeAnnotator was used which is included in the BEAST package. The consensus tree was further visualized by FigTree 1.4.4 with 95% higher probability density (HPD) values on divergence time[50].

**Analyses of 16S rRNA dataset.** To check the taxonomic rank of two named subspecies, *T. nicobarica nicobarica* from the Great Nicobar Island and *T. nicobarica surda* from the Little Nicobar Island, an additional dataset of 16S rRNA gene was constructed, including 18 database sequences of 12 Tupaiidae species known from South and Southeast Asian countries (Supplementary Table S2). The 16S rRNA sequence of the Sunda flying lemur, Galeopterus variegatus (Accession No. NC_004031) was used as an out-group in the second dataset. The genetic distances were calculated using MEGA X and the best fit model for this dataset was estimated using Mr. MODELTEST v2 with lowest BIC (Bayesian Information Criterion) score[51]. The Bayesian tree was constructed in Mr. Bayes 3.1.2 by selecting nst = 6 for GTR + I + G model and four (one cold and three hot) metropolis-coupled Markov chain Monte Carlo (MCMC) runs were performed for 10,000,000 generations with 25% burn-in and trees sampled every 2000 generations. To combine these runs, log files were imported into LogCombiner of the BEAST package. The adequate MCMC mixing and convergence were estimated using the effective sample size (ESS) values (>200 in all parameters) in Tracer v1.7.1[49]. To reconstruct the Bayesian phylogeny, TreeAnnotator was used which is included in the BEAST package. The consensus tree was further visualized by FigTree 1.4.4 with 95% higher probability density (HPD) values on divergence time[50].

**Results and discussion**

**Mitogenome structure and organization.** The mitogenome (17,164 bp) of the endangered Nicobar treeshrew, *T. nicobarica* was determined in the present study (GenBank accession no. MW751815). The mitogenome contained 37 genes, comprising 13 PCGs, 22 tRNAs, 2 rRNAs, and a major non-coding CR. Among them, nine genes (*nad5* and eight tRNAs) were placed on the negative strand, while the remaining 28 genes were placed on the positive strand (Table 1, Fig. 1). In the order Scandentia, the length of the Tupaiidae mitogenome varied from 16,183 bp (*T. montana*) to 17,164 bp (*T. nicobarica*). All Tupaiidae species showed the same gene arrangement as observed in typical vertebrate’s mitogenome[54]. The nucleotide composition of the *T. nicobarica* mitogenome was A + T biased (58.3%), as in all Tupaiidae species ranging from 58.3% (*T. nicobarica*) to 59.72% (*T. tana*) (Table 2). The AT skew and GC skew were 0.11 and −0.30 in the mitogenome of *T. nicobarica*. The comparative analysis showed that the AT skew ranged from 0.08 to 0.11 and the GC skew from −0.28 to −0.30 (Table 2). A total of 14 overlapping regions with a total length of 87 bp were identified in *T. nicobarica* mitogenome. The longest overlapping region (43 bp) was observed between the ATP synthase F0 subunit 8 (*atp8*) and ATP synthase F0 subunit 6 (*atp6*). Further, a total of 14 intergenic spacer regions with a total length of 68 bp
were observed in *T. nicobarica* mitogenome with the longest region (33 bp) between tRNA-Asparagine (*trnN*) and tRNA-Cysteine (*trnC*) (Supplementary Table S3).

### Protein-coding genes

The total length of PCGs was 11,410 bp in *T. nicobarica*, which represents 66.47% of the complete mitogenome. The nucleotide composition of the *T. nicobarica* PCGs was A + T biased (57.95%), as in all Tupaiidae species ranging from 57.95% (*T. nicobarica*) to 59.61% (*T. tana*) (Table 2). The AT skew and GC skew were 0.09 and −0.36 in the PCGs of *T. nicobarica* (Table 2). Most of the PCGs of *T. nicobarica* initiated with an ATG start codon; however, the ATC initiation codon was found in the NADH dehydrogenase subunit 2 (*nad2*), ATT in NADH dehydrogenase subunit 3 (*nad3*), ATA in NADH dehydrogenase subunit 5 (*nad5*).

The TAG termination codon was used by six PCGs, TAA by four PCGs, AGA by Cytochrome oxidase subunit 1 (*cox1*), AGG by NADH dehydrogenase subunit 6 (*nad6*), and incomplete T(AA) by NADH dehydrogenase subunit 4 (*nad4*), respectively. The comparative study revealed that, most of the PCGs in other Tupaiidae species were initiated by ATG start codon and terminated by TAA stop codon (Supplementary Table S4).

The analysis of mitogenome for detecting positive selection of PCGs assists to understand the influences of natural selection in evolution and protein function. The comparison of synonymous (Ks) and nonsynonymous (Ka) substitution rates in PCGs, witnessed for Darwinian selection and adaptive molecular evolution. It is reported that, for positive selection Ka/Ks > 1, for neutrality Ka/Ks = 1, and for negative selection Ka/Ks < 1.

### Table 1. List of annotated mitochondrial genes of the Nicobar treeshrew *T. nicobarica*.

| Gene  | Direction | Location | Size (bp) | Anti- codon | Start codon | Stop codon | Intergenic Nucleotides |
|-------|------------|----------|-----------|-------------|-------------|------------|------------------------|
| *trnF* | +          | 1–66     | 66        | GAA         | −           | −          | −1                     |
| *trnS* | +          | 66–1013  | 948       | −           | −           | −          | 4                     |
| *trnY* | +          | 1018–1084| 67        | TAC         | −           | −          | 0                     |
| *trnL* | +          | 1085–2655| 1571      | −           | −           | −          | 0                     |
| *trnL2*| +          | 2656–2730| 75        | TAA         | −           | −          | 2                     |
| *nad1* | +          | 2733–3689| 957       | −           | ATG         | TAG        | −2                    |
| *trnI* | +          | 3688–3756| 69        | GAT         | −           | −          | −3                    |
| *trnQ* | −          | 3754–3824| 71        | TTG         | −           | −          | −1                    |
| *trnM* | +          | 3824–3892| 69        | CAT         | −           | −          | 0                     |
| *nad2* | +          | 3893–4936| 1044      | −           | ATC         | TAG        | −2                    |
| *trnW* | +          | 4935–5001| 67        | TCA         | −           | −          | 4                     |
| *trnA* | −          | 5006–5075| 70        | TGC         | −           | −          | 1                     |
| *trnN* | −          | 5077–5149| 73        | GTT         | −           | −          | 33                    |
| *trnC* | −          | 5183–5249| 67        | GCA         | −           | −          | 0                     |
| *trnT* | −          | 5250–5315| 66        | GTA         | −           | −          | 1                     |
| *cox1* | +          | 5317–6864| 1548      | −           | ATG         | AGA        | −5                    |
| *trnS2*| −          | 6860–6928| 69        | TGA         | −           | −          | 4                     |
| *trnD* | +          | 6933–7001| 69        | GTC         | −           | −          | 0                     |
| *cox2* | +          | 7002–7685| 684       | −           | ATG         | TAG        | 1                     |
| *trnK* | +          | 7687–7750| 64        | TTT         | −           | −          | 2                     |
| *atp8* | +          | 7753–7956| 204       | −           | ATG         | TAA        | −43                   |
| *atp6* | +          | 7914–8594| 681       | −           | ATG         | TAA        | −1                    |
| *cox3* | +          | 8594–9379| 786       | −           | ATG         | TAA        | −1                    |
| *trnG* | +          | 9379–9445| 67        | TCC         | −           | −          | 9                     |
| *nad3* | +          | 9455–9802| 348       | −           | ATT         | TAG        | −10                   |
| *trnR* | +          | 9793–9858| 66        | TCG         | −           | −          | 1                     |
| *nad4* | +          | 9860–10,156| 297   | −           | ATG         | TAA        | −7                    |
| *nad4* | +          | 10,150–11,257| 1378 | −          | ATG         | T(AA)  | 0                     |
| *trnI* | +          | 11,528–11,596| 69  | −          | GTG         | −           | 0                     |
| *trnS1*| +          | 11,597–11,655| 59   | −          | GCT         | −           | −1                    |
| *trnH* | +          | 11,655–11,724| 70   | −          | TAG         | −           | −9                    |
| *nad5* | +          | 11,716–13,536| 1821 | −          | ATA         | TAG        | 2                     |
| *nad6* | −          | 13,539–14,060| 522  | −          | ATG         | AGG        | 0                     |
| *trnE* | −          | 14,061–14,128| 68   | −          | TTC         | −           | 3                     |
| *cob*  | +          | 14,132–15,271| 1140 | −          | ATG         | TAG        | −1                    |
| *trnT* | +          | 15,271–15,338| 68   | −          | TGT         | −           | 1                     |
| *trnP* | −          | 15,340–15,407| 68   | −          | TGG         | −           | 0                     |
| CR     | −          | 15,408–17,164| 1757 | −          | −           | −          | −                     |

Table 1. List of annotated mitochondrial genes of the Nicobar treeshrew *T. nicobarica*. 
This approach has the benefit to reveal the natural selection acting on PCGs. Thus, to investigate the evolutionary rates between homologous gene pairs, Ka/Ks substitutions were calculated and compared with six Tupaiidae species. The average Ka/Ks values of 13 PCGs varied from 0.006 (cox1) to 0.153 (atp8) and resulted in the following order: cox1 < cox3 < cox2 < atp6 < nad3 < cyt b < nad4 < nad6 < nad4l < nad5 < nad2 < atp8 (Supplementary Table S5, Supplementary Fig. S1). Most of the PCGs show Ka/Ks values of < 1, which indicated a strong negative selection among the studied Tupaiidae species, that reflects natural selection works against deleterious mutations with negative selective coefficients as highlighted general patterns in other vertebrates60. The comparative RSCU analysis indicated a significant fall in the frequency of GCG codon in Alanine (Ala) was observed in T. nicobarica, T. montana, T. minor, T. splendidula, and T. tana, except in T. belangeri with CCG in Proline (Pro) (Supplementary Fig. S2).

Ribosomal RNA and transfer RNA genes. The total length of two rRNA genes of T. nicobarica was 2,519 bp, compared to a range from 2,508 bp (T. montana) to 2,520 bp (T. belangeri) among other Tupaiidae species in the present dataset. The AT content within rRNA genes was 58.56%, while the AT and GC skew were 0.22 and −0.09 respectively observed in T. nicobarica rRNAs (Table 2). A total of 22 tRNAs were found in the T. nicobarica mitogenome with a total length of 1,497 bp. In other Tupaiidae species, the length of tRNAs varied from 1,493 bp (T. minor) to 1,564 bp (T. belangeri). The AT content within tRNA genes was 60.86%, while the AT and GC skew were 0.11 and −0.12, respectively observed in T. nicobarica tRNAs (Table 2). Most of the tRNA genes
were predicted to be folded into classical cloverleaf structures, except trnS1 (without DHU stem and loop) and trnK (without DHU loop) (Supplementary Fig. S3). The conventional pairings (A=T and G≡C) were observed in most of the tRNAs bases61; however, wobble base pairing was observed in the stem of 14 tRNAs (trnA, trnN, trnQ, trnE, trnC, trnG, trnL1, trnK, trnL2, trnP, trnS2, trnT, trnY, and trnW) (Supplementary Fig. S3). The wobble base pairing is a key feature of RNA structure and often substitutes the conventional base pairs due to thermodynamic stability. These characteristics play crucial functional roles in a wide range of phenomena62. Thus, the comparisons of tRNAs secondary structures are crucial for inferring the structural and functional features of the mitogenomes63.

**Control regions.** The CR of *T. nicobarica* was typically distributed with three functional domains: extended termination associated sequences (ETAS), central domain (CD), and the conserved sequence block (CSB), as observed in other mammalian mitochondrial CRs64. Although, the ETAS and CSB domains contain varying numbers of tandem repeats, the CD domain consists with highly conserved sequences. Hence, the pattern of CR was varied among different mammals, including Tupaiidae (Scandentia). The total length of *T. nicobarica* CR was 1,757 bp, compared to a range of 778 bp (*T. splendidula*) to 1,757 bp (*T. nicobarica*) in the present dataset. In the *T. nicobarica* CR, the AT and GC skew was 0.12 and −0.30 (Table 2). The CR is also involved in the initiation of replication and is positioned between trnP and trnF for most of the Tupaiidae including *T. nicobarica*. The ETAS domain was divided into two regions: ETAS1 (60 bp) and ETAS2 (67 bp), while the CSB domain was

Table 2. Nucleotide composition of the mitochondrial genomes of different treeshrew mtDNA. The A + T biases of the complete mitogenome, PCGs, tRNAs, rRNA, and CRs were calculated by AT-skew = (A − T)/(A + T) and GC-skew = (G − C)/(G + C), respectively.
further divided into three regions: CSB1 (25 bp), CSB2 (17 bp), and CSB3 (18 bp). After CSB3, a six base pair (CGTACA) tandem repeats were found 60.3 times in *T. nicobarica*, while eight base pair (CACACA TA) were found 23.8 times in *T. belangeri* (Fig. 2). Due to the short nucleotide length, no tandem repeats were found in other Tupaiidae species CRs. The structural features of CR play an important function in influencing transcription and replication in the mitochondrial genome. The present study evaluated the genetic features of CR among the studied Tupaiidae species mitogenomes including *T. nicobarica* that will be helpful to speculate the evolutionary pattern of this group.

**Phylogenetic inference.** The phylogenetic position of Scandentia is repetitively argued and examined within the eutherian tree. The treeshrews are widely considered as living fossils due to their approximating ancestral lineages with primates. Based on the anatomical evidence, Primates, Chiroptera, Dermoptera, and Scandentia were hypothetically within the superordinal clade Archonta without considering the paleontological or molecular evidence. Later on, the phylogenetic position of Scandentia has been studied based on the complete mitochondrial DNA sequences of wider group of taxa and corroborated a closer relationship with Lagomorpha. Further, multiple loci of mitochondrial genes has been assessed to check the phylogeny of treeshrews and diversification and the timescale of diversification in Southeast Asia. The present ML and BA phylogenies clearly discriminate *T. nicobarica* from other congeners and are congruent with earlier evolutionary hypotheses of Scandentia (Fig. 3, Supplementary Fig. S4). Further, using four calibration points from earlier studies, the present mitogenome-based dating analysis indicates that, the Tupaiidae species (Scandentia) were diverged from Primates and Dermoptera during the Cretaceous period (81–101 MYA). However, the basal node of Scandentia, Primates and Dermoptera was diverged from the Lagomorphs and Rodents during the same Cretaceous period (82–125 MYA). As a whole the divergence time estimations are little deviated due to the exclusion of Lagomorphs and Rodents in earlier analysis. However, the representative of Scandentia family members (Ptilocercidae and Tupaiidae) in earlier analysis revealed that, they were diverged during Neogene to Paleogene period.

**Biogeographic connection and conservation implication.** The tectonic drifts allowed multiple possibilities for dispersal and colonization events of many animals into the same or distant geographical distribution. Due to the adjacent biogeographic realms, the biological affinities between the Indian mainland and Southeast Asia have been well documented. However, the faunal diversity of the AN archipelago and their biotic
networks is still anonymous in spectacular aspects. The bathymetric study evidenced that the well-developed seamounts have been detected on the Andaman seafloor, which extended up to Sumatra and Java Islands72,73. Considering the skeletal variation, the treeshrew species showed intraspecific variation depending upon their distribution in mainland and island ecosystems74. A recent molecular study also elucidates the biogeographic connection of smaller mammals in the AN archipelago with the Indo-Malayan and Sundaic realms7. The present mitogenome based phylogeny also manifested the close relationship of $T. \text{nicobarica}$ with $T. \text{minor}$, $T. \text{tana}$, $T. \text{splendidula}$, and $T. \text{montana}$ (distributed in Thailand, Peninsular and East Malaysia, Brunei Darussalam, Sumatra, and Indonesia) as compared with the widespread species $T. \text{belangeri}$ known from South and Southeast Asia. Further, the single gene (16S rRNA) based phylogeny showed sister relationship of $T. \text{nicobarica}$ with $T. \text{javanica}$ (distributed in Sumatra and Java) as compared with other congeners. The two subspecies of $T. \text{nicobarica}$ were discriminated only by their different distributional pattern in two distinct islands. Prior to this study, they were neither examined morphologically nor tested genetically to assure their taxonomic status. The present molecular based assessment clearly distinguished two subspecies with 0.7% genetic distance by 16S rRNA gene. This preliminary molecular information will help for rapid and reliable identification of this highly threatened and endemic species from the Great and Little Nicobar Islands. However, further research can be done with the extensive sampling and generation of microsatellite data to substantiate their population genetic structure to formulate precise conservation action plans.

Considering the conservation implication, the previous studies reported that, this arboreal mammal species confronted several threats due to the forest loss and fragmentation, and ongoing road construction from Galathia to Indira Point at the Great Nicobar Island75. Although the species is listed on the IUCN with decreasing population trend, it has not yet listed in the Indian wildlife (Protection) Act, 1972. Other than a single ecology and behavior study and a nest record, no ample assessment has been approached so far3,76.

Besides, the treeshrews species were considered as a significant model for studying hepatitis and influenza H1N1 viral infections19,20. A recent study characterized the genome sequence to demonstrate the genetic basis of signaling pathways in nervous and immune systems of the Chinese treeshrew ($Tupaioides belangeri \text{chinenis}$) and evidenced as a potential model for biomedical research21. The $T. \text{belangeri \text{chinenis}}$ also maintained sufficient intraspecific variation (5.4–9.5%) with the Northern Treeshrew, $T. \text{belangeri}$ in all 13 PCGs. Hence, the generation of molecular information from different geographical region is crucial for elucidating the actual evolutionary history of this small mammal species.

We propose the whole genome sequencing of $T. \text{nicobarica}$ is essential as a genetic resource for conservation purposes. The genome sequence will also assist to predict the signaling pathways linked with many pathogenic microorganisms as well as able to develop potential mitigations programs in advance. As the population of this treeshrew is confined to the insular habitats in the Nicobar Islands, we propose an integrated approach with taxonomy, ecology, and further molecular studies to save this endemic species before it reaches to the brink of extinction.

**Data availability**

The following information was supplied regarding the accessibility of DNA sequences: The generated complete mitochondrial genome sequences of $Tupaioides \text{nicobarica}$ are deposited in GenBank of NCBI under accession number MW751815.
Received: 6 September 2021; Accepted: 29 December 2021
Published online: 18 January 2022

References
1. Burgin, C. J., Colella, J. P., Kahn, P. L. & Upham, N. S. How many species of mammals are there?. J. Mamm. 99, 1–14 (2018).
2. Wilson, D. E. & Mittermeier, R. A. Handbook of the Mammals of the World, Vol 8: Insectivores, Sloths and Colugos Vol 709 (Lynx Edicions, 2018).
3. Oommen, M. A. & Shanker, K. Ecology and behaviour of endemic tree-shrew Tupaia nicobarica Zeledon 1869 on Great Nicobar Island, India. J. Bom. Nat. Hist. Soc. 105, 53–63 (2008).
4. IUCN. The IUCN Red List of Threatened Species, Version 2021–2 (IUCN, 2021).
5. Yahya, H. S. A. & Zarri, A. A. Status, ecology and behaviour of Narcondam Hornbill (Aceros narcondami) in Narcondam Island, Andaman and Nicobar Islands, India. J. Bom. Nat. His. Soc. 99, 434–445 (2002).
6. Datta-Roy, A. & Karanth, K. P. The Out-of-India hypothesis: What do molecules suggest?. J. Biosci. 34, 687–697 (2009).
7. Kamalakannan, M. et al. Discovery of a new mammal species (Soricidae: Eulipotyphla) from Narcondam volcanic island, India. Sci. Rep. 11, 9416. https://doi.org/10.1038/s41598-021-88859-4 (2021).
8. Matthews, T. J., Rigal, F., Triantis, K. A. & Whittaker, R. J. A global model of island species–area relationships. Proc. Natl. Acad. Sci. USA 116, 12337–12342 (2019).
9. Upham, N. S., Essedtyn, J. A. & Jetz, W. Ecological causes of uneven diversification and richness in the mammal tree of life. BioRxiv https://10.1101/504803 (2019).
10. Davidar, P., Yoganand, K. & Ganesh, T. Distribution of forest birds in the Andaman Islands: Importance of key habitats. J. Biogeogr. 28, 663–671 (2001).
11. Harikrishnan, S. et al. Macroeology of Terrestrial Herpetofauna in Andaman & Nicobar Archipelago. Wildlife Institute of India, Uttarakhand, India. 1–49 (2014).
12. Kamalakannan, M. & Venkatraman, C. A Checklist of Mammals of India (Zoological Survey of India, 2017).
13. Menon, V. Indian Mammals—A Field Guide Vol 528 (Hachette Book Publishing India Pvt Limited, 2014).

Figure 4. Genetic status of two known subspecies of T. nicobarica based on 16S rRNA sequences. (A) Map showing the distribution of other comparative Tupaiidae species in the present phylogeny. The first author (S.K.) prepared the map by using software QGIS 2.6.1 (http://www.qgis.org), the artwork of T. nicobarica subspecies and edited manually in Adobe Photoshop CS 8.0. (B) BA Phylogeny showed distinct clustering of T. nicobarica subspecies and other Tupaiidae species. Numbers on the nodes are posterior probabilities. (C) Distribution pattern of T. nicobarica nicobarica and T. nicobarica surda in the Great and Little Nicobar Island, respectively.

Figure 4. Genetic status of two known subspecies of T. nicobarica based on 16S rRNA sequences. (A) Map showing the distribution of other comparative Tupaiidae species in the present phylogeny. The first author (S.K.) prepared the map by using software QGIS 2.6.1 (http://www.qgis.org), the artwork of T. nicobarica subspecies and edited manually in Adobe Photoshop CS 8.0. (B) BA Phylogeny showed distinct clustering of T. nicobarica subspecies and other Tupaiidae species. Numbers on the nodes are posterior probabilities. (C) Distribution pattern of T. nicobarica nicobarica and T. nicobarica surda in the Great and Little Nicobar Island, respectively.
14. Miller, G. S. Mammals of the Andaman and Nicobar Islands. Proc. U.S. Nat. Mus. 24, 751–795 (1902).
15. Roberts, T. E., Lanier, H. C., Sargs, E. J. & Olson, L. E. Molecular phylogeny of tree-shrews (Mammalia: Scandentia) and the time-scale of diversification in Southeast Asia. Mol. Phylogenet. Evol. 60, 358–372 (2011).
16. Oomen, M. A. Tree-shrews. In Mammals of South Asia (eds Johnsingh, A. J. T. & Manirakib, N.) 52–67 (University Press, 2013).
17. Damas, J. et al. Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. Proc. Natl. Acad. Sci. USA 117, 22311–22322 (2020).
18. Wardell, M., Baylis, M. & Blagrove, M. S. C. Predicting mammalian hosts in which novel coronaviruses can be generated. Nat. Commun. 12, 780 (2021).
19. Caio, J., Yang, E.-B., Su, J.-J., Li, Y. & Chow, P. The tree shrews: Adjuncts and alternatives to primates as models for biomedical research. J. Med. Primatol. 32, 123–130 (2003).
20. Yang, Z. F. et al. The tree shrew provides a useful alternative model for the study of influenza H1N1 virus. Virol. J. 10, 111 (2013).
21. Yu, E. et al. Genome of the Chinese tree shrew. Nat. Commun. 4, 1426 (2013).
22. Pacheco, M. A. et al. Escalante, Evolution of modern birds revealed by mitogenomes: Timing the radiation and origin of major orders. Mol. Biol. Evol. 28, 1927–1942 (2011).
23. Finstermeier, K. et al. A mitogenomic phylogeny of living primates. PLoS One 8, e69504 (2013).
24. Arnason, U. et al. Mammalian mitogenomic relationships and the root of the eutherian tree. Proc. Natl. Acad. Sci. USA 99, 8151–8156 (2002).
25. Schmitz, J., Ohme, M. & Zischler, H. The complete mitochondrial genome of Tupai a belangeri and the phylogenetic affiliation of scandentia to other eutherian orders. Mol. Biol. Evol. 17, 1334–1343 (2000).
26. Parker, D. et al. Little genetic structure in a Bornean endemic small mammal across a steep ecological gradient. Mol. Ecol. 29, 4074–4090 (2020).
27. Olson, I. E., Sargs, E. J. & Martin, R. D. Intraordinal phylogenetics of tree-shrews (Mammalia: Scandentia) based on evidence from the mitochondrial 12S rRNA gene. Mol. Phylogenet. Evol. 35, 656–673 (2005).
28. Kundu, S. et al. Molecular investigation of non-volant endemic mammals through mitochondrial cytochrome b gene from Andaman and Nicobar archipelago. Mitochondrial DNA B 5, 1447–1452 (2020).
29. Kundu, S. et al. The complete mitochondrial genome of the endangered Assam Roofed Turtle, Pangshura sylhetensis (Testudines: Geoemydidae): Genomic features and phylogeny. PLoS One 15, e222523 (2020).
30. Dierckxsens, N., Mardulyn, P. & Smits, G. NOVOPLAST: De novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45, e18 (2017).
31. Grant, J. R. & Stothard, P. The CGView Server: A comparative genomics tool for circular genomes. Nucleic Acids Res. 36, W181–184 (2008).
32. Bernt, M. et al. MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69, 313–319 (2013).
33. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549 (2018).
34. Perna, N. T. & Kocher, T. D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41, 353–359 (1995).
35. Rojas, I. et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. 34, 3299–3302 (2017).
36. Laslett, D. & Canback, B. ARWEN, a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 22, 172–175 (2008).
37. Lowe, T. M. & Chan, P. P. tRNAscan-SE on-line: Search and contextual analysis of transfer RNAGenes. Nucleic Acids Res. 44, W54–57 (2016).
38. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580 (1999).
39. Xu, L., Chen, S. Y. & Nie, W. H. Evaluating the phylogenetic position of Chinese tree shrew (Tupai a belangeri chinensis) based on complete mitochondrial genome: Implication for using tree shrew as an alternative experimental animal to primates in biomedical research. J. Genet. Genom. 39, 131–137 (2012).
40. Zhou, X., Sun, F., Xu, S., Yang, G. & Li, M. The position of tree shrews in the mammalian tree: Comparing multi-gene analyses with phylogenomic results leaves monophyly of Euar chonta doubtful. Integr. Zool. 10, 186–198 (2015).
41. Abascal, F., Zardoya, R. & Telford, M. J. Transl X: Multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 38, W7–13 (2010).
42. Vaidya, G., Lohman, D. J. & Meier, R. J. SequenceMatrix: Concatenation software for the fast assembly of multigene datasets with character set and codon information. Cladistics 27, 171–180 (2010).
43. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34, 772–773 (2016).
44. Trifinopoulos, J., Nguyen, L.-T., van Hae selier, A. & Minh, B. Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 44, W232–W235 (2016).
45. Bouchkaert, R. et al. BEAST2: A software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537 (2014).
46. Mein, F. & Ginsburg, L. Les mammifères du gisement miocène inférieur de Li Mae Long, Thaïlande: Systématique, biostratigraphie et paléoenvironnement. Geodiversitas 19, 783–844 (1997).
47. Eizirik, E., Murphy, W. J., Springer, M. S. & O’Brien, S. J. Molecular phylogeny and dating of early primate divergences. In Anthropoid Origins. New visions (eds Ross, C. F. &Kay, R. F.) (Kluwer Academic, 2004).
48. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: A public knowledge-base of divergence times among organisms. Bioinformatics 22, 2971–2972 (2006).
49. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67, 901–904 (2018).
50. Rambaut, A. FigTree. Version 1.4.4 Institute of Evolutionary Biology (University of Edinburgh, 2014).
51. Nylander, J. A. A. MrModeltest v2 (University of Copenhagen, 2004).
52. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574 (2003).
53. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. Bioinformatics 23, 127–128 (2007).
54. Anderson, S. et al. Complete sequence of bovine mitochondrial DNA conserved features of the mammalian mitochondrial genome. J. Mol. Evol. 36, 883–717 (1982).
55. Hirsh, A. E. & Fraser, H. B. Protein dispensability and rate of evolution. Nature 411, 1046–1049 (2001).
56. Bloom, J. D., Láb háválu, S. T. & O’tey, C. R. Protein stability promotes evolvability. Proc. Natl. Acad. Sci. USA 103, 5869–5874 (2006).
57. Yang, Z. & Bielawski, J. P. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. 15, 496–503 (2000).
58. Yang, Z. H. & Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 17, 32–43 (2000).
59. Nei, M. & Kumar, S. Molecular Evolution and Phylogenetics (Oxford University Press, 2000).
60. Meiklejohn, C. D., Montooth, K. L. & Rand, D. M. Positive and negative selection on the mitochondrial genome. Trends Genet. 23, 259–263 (2007).
61. Varani, G. & McClain, W. H. The G-U wobble base pair: A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.* 1, 18–23 (2000).
62. Crick, F. H. C. Codon-anticodon pairing: The wobble hypothesis. *J. Mol. Biol.* 19, 548–555 (1966).
63. Takashi, P. S., Miya, M. & Mabuchi, K. Structure and variation of the mitochondrial genome of fishes. *BMC Genom.* 17, 719 (2016).
64. Sbisa, E., Tanziello, F., Reyes, A., Pesole, G. & Saccone, C. Mammalian mitochondrial D-loop region structural analysis: Identification of new conserved sequences and their functional and evolutionary implications. *Gene* 205, 125–140 (1997).
65. Tannaham, J. W. The mitochondrial genome: Structure, transcription, translation and replication. *Biochim. Biophys. Acta* 1410, 103–123 (1999).
66. Shao, R., Barker, S. C., Mitten, H., Aoki, Y. & Fukunaga, M. Evolution of duplicate control regions in the mitochondrial genomes of Metazoa: A case study with Australasian Ixodes ticks. *Mol. Biol. Evol.* 22, 629–630 (2005).
67. Li, Q. & Ni, X. An early Oligocene fossil demonstrates treeshrews are slowly evolving “living fossils”. *Sci. Rep.* 6, 18627 (2016).
68. Novacek, M. J. Mammalian phylogeny: Shaking the tree. *Nature* 356, 121–125 (1992).
69. McKenna, M. C. & Bell, S. K. *Classification of Mammals Above the Species Level* (Columbia University Press, 1997).
70. Roberts, T. E., Sargis, E. J. & Olson, L. E. Networks, trees, and treeshrews: Assessing support and identifying conflict with multiple loci and a problematic root. *Syst. Biol.* 58, 257–270 (2009).
71. Garg, S. & Biju, S. D. New microhyloid frog genus from Peninsular India with Southeast Asian affinity suggests multiple Cenozoic biotic exchanges between India and Eurasia. *Sci. Rep.* 9, 1906 (2019).
72. Rodolfo, K. S. Bathymetry and marine geology of the Andaman basin, and tectonic implications for Southeast Asia. *Geol. Soc. Am. Bull.* 80, 1203–1230 (1969).
73. Tripathi, S. K. *et al.* Morphology of submarine volcanic seamounts from inner volcanic arc of Andaman Sea. *Indian J. Geosci.* 71, 451–470 (2017).
74. Sargis, E. J., Woodman, N., Morningstar, N. C., Bell, T. N. & Olson, L. E. Skeletal variation and taxonomic boundaries among mainland and island populations of the common treeshrew (*Mammalia: Scandentia: Tupaiidae*). *Biol. J. Linn. Soc.* 120, 286–312 (2017).
75. Molur, S. *et al.* Status of Non-volant Small Mammals: Conservation Assessment and Management Plan (CAMP) Workshop Report 618 (Zoo Outreach Organisation, 2005).
76. Kamalakannan, M., Gokulakrishnan, G., Venkatraman, C., Sivaperuman, C. & Chandra, K. First record of a Nicobar treeshrew nest in a fallen palm tree. *Mammalia* 85, 159–160 (2021).

**Acknowledgements**

The authors thankful to the Director of Zoological Survey of India (ZSI), Ministry of Environment, Forests and Climate Change (MoEF&CC), Govt. of India. This work was financially 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)’ to V.K. The first author (S.K) acknowledges a fellowship Grant received from the Council of Scientific and Industrial Research (CSIR) Senior Research Associateship (Scientists’ Pool Scheme) Pool No. 9072-A.

**Author contributions**

Conceptualization: S.K., V.K.; data curation: S.K. M.K.; formal analysis: S.K., A.P., D.S.; funding acquisition: D.B., V.K.; investigation: S.K., K.T.; methodology: S.K. M.K., A.P., D.S.; project administration: D.B., V.K.; resources: D.B., C.V., V.K.; software: S.K. K.T., A.P.; supervision: D.B., V.K.; visualization: S.K., M.K., V.K.; writing—original draft: S.K., M.K., A.P.; writing—review and editing: S.K., M.K., V.K.

**Funding**

Ministry of Environment, Forest and Climate Change (MoEF&CC), Zoological Survey of India (ZSI) in-house project, ‘National Faunal Genome Resources (NFGR).’

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-04907-7.

**Correspondence** and requests for materials should be addressed to V.K.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2022