Turdoides affinis mitogenome reveals the translational efficiency and importance of NADH dehydrogenase complex-I in the Leiothrichidae family

Indrani Sarkar1,4, Prateek Dey1,4, Sanjeev Kumar Sharma1, Swapna Devi Ray1, Venkata Hanumat Sastry Kochiganti2, Rene Singh3, Admanabhan Pramod1 & Ram Pratap Singh1,3

Mitochondrial genome provides useful information about species concerning its evolution and phylogenetics. We have taken the advantage of high throughput next-generation sequencing technique to sequence the complete mitogenome of Yellow-billed babbler (Turdoides affinis), a species endemic to Peninsular India and Sri Lanka. Both, reference-based and de-novo assemblies of mitogenome were performed and observed that de-novo assembled mitogenome was most appropriate. The complete mitogenome of yellow-billed babbler (assembled de-novo) was 17,672 bp in length with 53.2% AT composition. Thirteen protein-coding genes along with two rRNAs and 22 tRNAs were detected. The arrangement pattern of these genes was found conserved among Leiothrichidae family mitogenomes. Duplicated control regions were found in the newly sequenced mitogenome. Downstream bioinformatics analysis revealed the effect of translational efficiency and purifying selection pressure over thirteen protein-coding genes in yellow-billed babbler mitogenome. Ka/Ks analysis indicated the highest synonymous substitution rate in the nad6 gene. Evolutionary analysis revealed the conserved nature of all the protein-coding genes across Leiothrichidae family mitogenomes. Our limited phylogeny results placed T. affinis in a separate group, a sister group of Garrulax. Overall, our results provide a useful information for future studies on the evolutionary and adaptive mechanisms of birds belong to the Leiothrichidae family.

Aves are one of the most diverse vertebrate classes with a huge number of species having a broad range of ecological behavior and complex morphology, all of which make it difficult to solve the riddles regarding their taxonomy along with phylogenetic and evolutionary relationship1–3. New and advanced scientific techniques have emerged to solve these riddles. In the last few years, genome sequencing has become more popular to obtain huge information on evolutionary history and revising the clustering pattern of traditional taxonomy4. Mitochondrial DNA with some of its inherent properties like small genome size, absence of extensive recombination frequency, simple structure of genome, maternal inheritance along with rapid evolutionary rate are now extensively utilized in taxonomic and phylogenetic studies of vertebrates5–10. Furthermore, it has been reported that complete mitogenomes retain more information than a single gene regarding the evolutionary history of the taxon and also provide consistent results compared to nuclear genes11. This also reduces the effect of homoplasy and frequent stochastic errors in phylogenetic studies11.

Yellow-billed babbler (Turdoides affinis) is one of the most common birds in peninsular India12, especially in the southern parts of Maharashtra, Chattisgarh and Andhra Pradesh12. The taxonomic classification of this bird

1National Avian Forensic Laboratory, Sálim Ali Centre for Ornithology and Natural History, Anaikatty, Coimbatore, Tamil Nadu 641108, India. 2Central Avian Research Institute, Bareilly, Uttar Pradesh 243122, India. 3Department of Life Science, Central University of South Bihar, Gaya, Bihar 824236, India. 4These authors contributed equally: Indrani Sarkar and Prateek Dey. *email: sksbreeder@gmail.com; rampratapsingh81@gmail.com
is quite dubious. Previously all babblers and allies were considered under Timaliidae family\textsuperscript{13}. Recent classification studies have split this family into five discrete families\textsuperscript{13}. Three of them, Leiothrichidae, Pellorneidae and Timaliidae consisted of traditional babblers while Zosteropidae that included mainly \textit{Yuhina} and some other minor species, and Sylviae grouped all the \textit{Sylvia} warblers\textsuperscript{14}. Among these five distinct families, Leiothrichidae is the largest consisting of 125 species distributed mainly in the Sino-Himalayan and South-Eastern parts of Asia\textsuperscript{14}. A study on the Leiothrichidae family suggested their origin prior to the Miocene–Pliocene boundary, which is well known for its noteworthy climatic turmoil in Asia\textsuperscript{15}. Hence, it is imperative to conduct in-depth molecular studies on this family to know some appealing unknown facts regarding their taxonomical and phylogenetic relationships.

The Leiothrichidae family mainly consisted of \textit{Grammatopelia}, \textit{Garrulax}, \textit{Trochalopteron}, \textit{Turdoides} and \textit{Argya}\textsuperscript{14}. Polyphyletic ambiguities were observed among members of \textit{Turdoides} due to which further taxonomic classification was done and some species were resurrected from this genus\textsuperscript{14–16}. Yellow-billed babbler which was considered to be under \textit{Turdoides} has been recently proposed to be placed under \textit{Argya} based on the revised taxonomy of Leiothrichidae\textsuperscript{14}; however, more investigations are required to confirm this. Yellow-billed babblers generally live in flocks of seven to ten members continuously squeaking, chattering and chirping. Reapers are seen generally assisting parents in nest building, chick feeding and maintaining nest sanitation as cooperative breeding character\textsuperscript{17–19}. Interestingly, the close relatives of \textit{T. affinis} for instance, \textit{Garrulax}, \textit{Sylvia}, \textit{Liocichla}, \textit{Minla} and \textit{Trochalopteron} have still not been reported to show such cooperative breeding behaviour indicating a divergence of \textit{T. affinis} at least from the behavioural ecology perspective\textsuperscript{20}.

Until now, the complete mitogenomes of various species (including \textit{Garrulax affinis}\textsuperscript{20}, \textit{G. albogularis}, \textit{G. canorus}, \textit{G. cineraceus}\textsuperscript{21}, \textit{G. elliotti}\textsuperscript{22}, \textit{G. formosus}\textsuperscript{23}, \textit{G. ocellatus}, \textit{G. perspicillatus}\textsuperscript{24}, \textit{G. poecilorhynchos}\textsuperscript{25}, \textit{G. sanno}\textsuperscript{26}, \textit{G. sanno}, \textit{Trochalopteron omnini}, \textit{Leiothrix argentauris}, \textit{Liocichla}, \textit{Liocichla onemian}\textsuperscript{27} and \textit{Minla ignotincta}\textsuperscript{28}) from Leiothrichidae family have been reported. However, the complete mitogenome of birds representing the genus \textit{Turdoides} is yet to be revealed.

In this study, we have sequenced and described the complete mitochondrial genome of \textit{T. affinis} obtained using two approaches, reference-based assembly, and de-novo assembly\textsuperscript{29,30}. We employed the two different approaches to align the same mitogenome to quantify the differences resulting from these two approaches and their effect on mitogenomic parameters. In addition, we have performed a detailed comparative analysis of mitogenomes of the available members from family Leiothrichidae to understand the overall species-specific differences.

**Materials and methods**

**Sample collection and DNA extraction.** A fresh road-killed specimen of \textit{T. affinis} was collected from Anaikatty Hills in the Coimbatore district of Tamil Nadu, India (76 47′43″ E and 11 5′51″ N), and transported immediately to the lab. Prior permission for collection of road-killed birds was obtained from Tamil Nadu Forest Department (Ref. No.WL5 (A)/2219/2018; Permit No. 14/2018). Muscle tissue was sampled from the specimen and stored in DESS buffer (20% DMSO, 0.25 M tetra-sodium EDTA, Sodium Chloride till saturation, pH 7.5) at −20 °C for further processes. About 50 mg of the muscle tissue in DESS buffer was taken and added to 500 μl of lysis buffer (10 mM Tris-pH 8.0, 10 mM EDTA-pH 8.0, 100 mM NaCl) and homogenized thoroughly. To the homogenate, 80 μl of 10% SDS along with 20 μl of Proteinase K (20 mg/ml) was added and incubated overnight at 55 °C. The DNA extraction was performed the following day using suitable volumes of Phenol, Chloroform and Isoamyl alcohol. The DNA pellet obtained was suspended in 100 μl of Tris–EDTA buffer (Sigma-Aldrich, USA) and quantified using spectrophotometer (DeNovix, USA) and Qubit 4 Fluorometer (ThermoFisher Scientific, USA). The quality of the DNA extracted was assessed by running it on 1% agarose gel stained with ethidium bromide intercalating dye.

**Library preparation.** For library preparation, 700 ng of extracted DNA was utilized as starting material in NEBNext Ultra II DNA Library Prep kit for Illumina (New England Biolabs, USA). The DNA was fragmented using a focused ultrasonicator (Covaris M220, USA) until the desired length of 270–300 base pairs was obtained. The fragmented DNA size was analyzed by running it in Fragment Analyzer (Agilent, USA) making sure that the size of most DNA fragments is between 270 and 300 base pairs. Adapter ligation was then carried out in a thermocycler following the “NEBNext Ultra II DNA Library Prep kit for Illumina” protocol using dual indexed primers present in the kit for creation of paired-end libraries. After the ligation reaction was completed, the size selection of Adapter-ligated DNA was carried out using NEBNext Sample Purification Beads (New England Biolabs, USA) followed by PCR enrichment of Adapter-ligated DNA following the manufacturer’s protocol. After the clean-up of the enriched DNA, it was again analysed for the required concentration and mean peak size in a Fragment Analyzer (Agilent, USA). The enriched DNA library fragments were subjected to sequencing in Illumina NextSeq 550 (Illumina, Inc., USA) using Illumina High Output Kit for NextSeq 500/550 (Illumina, Inc., USA). A PhiX control library (Illumina, Inc., USA) was also subjected to sequencing along with the sample DNA library as an internal control. At the end of the sequencing run, high-quality paired-end reads were obtained, and further bioinformatics analysis was performed.

**Assembly and annotation of the mitochondrial genome.** Illumina NextSeq 550 produced 8,852,137 raw reads from the whole-genome library. Cutadapt tool\textsuperscript{32} was used to trim the adapter and high-quality reads with a Phred (Q) score of 30 and above were selected for further analysis. Finally, we got 1,297,736 high quality reads after downsampling with Seqtk (https://github.com/lh3/seqtk) which were used for assembly. De-novo assembly was performed using SPAdes-3.11.1 software with default parameters. MITOS online server (https://
Phylogenetic analysis. Phylogenetic analysis was performed on the available whole mitogenomes of various species of the Leiothrichidae family. We prepared two complete mitogenome based phylogenetic trees, one with T. affinis mitogenome assembled through reference-based assembly and the other with de-novo assembly. The complete mitochondrial genomes of Pycnonotus sanctorialis (NC_031830.1), Paradoxornis guttata (NC_039536.1), Zosterops erythropleurus (NC_027942.1), Yuhina nigripennis (NC_040991.1) and Taeniopygia guttata (NC_007897.1) from Pycnonotidae, Sylviidae, Zosteropidae, Timaliidae and Estrildidae were used as the outgroup. Both Maximum Likelihood based phylogeny and Bayesian phylogeny were constructed using MegaX and MrBayes 3.2.6, respectively. A bootstrap value of 1000 was used in ML based tree. Generalized Time Reversible (GTR) Substitution applying partition by codon was applied for Bayesian tree generation. GTR model depicted 6 substitution rates for each of the 4 nucleotides. Bayesian phylogeny was build applying the integration of Markov Chain Monte Carlo (MCMC) algorithms. The number of generations in Markov Chain Monte Carlo parameters was set to be 1000.

Sequence analysis of mitogenome. The complete mitogenome of T. affinis was compared with other Leiothrichidae family avian species whose complete mitochondrial genomes were available at NCBI including Garrulax affinis\cite{20}, G. albogularis, G. canorus, G. cineraceus\cite{21}, G. elliotii, G. formosus\cite{22}, G. ocellatus\cite{23}, G. perspicillatus\cite{24}, G. poecilorhynchos\cite{25}, G. sannio, Trochaleapton melius, Leiothrix argentauris, L. lutea, Liocichla omeiensis\cite{26} and Minla ignitincta\cite{27}. These sequences were downloaded from the NCBI website and used for further comparative analysis. The protein-coding genes along with tRNAs and rRNAs were aligned to examine whether any rearrangements persist among these mitogenomes. The initiation and termination codons of all the protein-coding genes were curated through NCBI ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/). Circular genome views were obtained by CG view server\cite{28}. The boundary of the control region (CR) was determined following the method by Zhou et al.\cite{29}.

Detailed codon usage analysis of the select mitogenomes was performed using CodonW software\cite{30}. The studied codon usage parameters include Relative Synonymous Codon Usage (RSCU), Effective Number of Codons (ENC) and frequency of G + C at the third position of codons (GC3s). Different codon composition indices of individual genes, for example total GC content, as well as the frequency of each nucleotide at the third position of codons (A3, T3, G3 and C3) were also estimated. R-based heatmaps were generated based on the overall codon usage and amino acid usage analysis. Skew values for AT\((A-T)/(A+T)) and GC\((G-C)/(G+C))\) were calculated\cite{31} using DAMBE software (https://dambe.bio.outtawa.ca/DAMBE/dambe.aspx). Tandem Repeats Finder software (https://tandem.bu.edu/trf/trf.html) with its default settings was employed to detect any tandem repeats within the mitogenome. A BlastN based approach to find intra-genomic duplication of large fragments or interspersed repeats was employed\cite{32} where each mitogenome was searched against itself with an e-value of 1e-10. This analysis detected a negligible number of interspersed repeats, hence was not evaluated further.

Estimation of translational efficiency. This parameter measures the competence of codon-anticodon interactions indicating the accuracy of the translational machinery of genes in the absence of preferred codon set information. We calculated the translational efficiency according to the following equation\cite{33}:

$$P_2 = \frac{W_{WC} + SSU}{W_{WT} + SSY}$$

where \(P_2 > 0.5\) indicates the existence of translational selection.

RSCU-based cluster analysis and putative optimal codons. Generally, highly expressed genes utilize a specific set of codons termed as optimal codons. Due to the preferential use of this set of codons their Enc value lowers down in contrast to lowly expressed genes, which restrain more rare codons with higher Enc value\cite{34}. We identified the optimal codons of all investigated species from their RSCU values. RSCU = 1 indicated unbiased codon usage whereas RSCU > 1 and RSCU < 1 indicated a higher and lower usage frequency of that particular codon respectively\cite{35}.

Evolutionary analysis. The ratio \(\omega\) of non-synonymous substitution rate per synonymous site (Ka) to synonymous substitution rate per non-synonymous site (Ks) has been reported as an excellent estimator of evolutionary selection pressure or constrain on protein-coding genes. \(\omega > 1\) stands for positive Darwinian selection (diversifying pressure), on the contrary \(\omega < 1\) signifies purifying or refining selection. At neutral evolutionary state, the value of \(\omega\) becomes 1 symbolizing the equal rate of both synonymous and non-synonymous substitution\cite{36}. The mean genetic distance of the annotated protein-coding genes of the studied mitogenomes were calculated in terms of Kimura-2-parameter (K2P) substitution model by DnaSPver 6.12.03 software\cite{37}. Evolutionary analysis was done using PAML (https://abacus.gene.ucl.ac.uk/software/paml.html).

Results and discussion
Comparison of T. affinis mitogenome assembled using reference-based and de-novo assembly approach. In this study, we performed both, reference-based assembly and de-novo assembly of the newly sequenced mitogenome of T. affinis and found a considerable difference in the results between these two approaches. In reference-based assembly, the total size of the mitogenome was 17,882 bp with 47% GC and 53%
AT (Supplementary file 1) whereas the de-novo assembly resulted in 17,672 bp long mitogenome with 53.2% AT and 46.80% GC (Fig. 1, Table 1). AT and GC skewness were 0.13 and − 0.38, respectively for de-novo assembly. However, reference-based assembly resulted in lower AT (0.05) and GC skew (-0.14) for *T. affinis* mitogenome.

The Genbank accession number of complete *T. affinis* mitogenome (assembled de-novo) is MN848144.

Two rRNA (rrnS for small subunit and rrnL for large subunit), 13 protein-coding genes (PCGs) and 22 tRNAs specified for 20 amino acids (two tRNAs each for serine and lysine) were reported in both the mitogenomes. The total length of PCGs, tRNAs and rRNAs were 11,391 bp, 1534 bp and 2573 bp, respectively for de-novo assembly while these were 11,247 bp, 1535 bp and 2578 bp respectively for reference-based assembly.

The following results were identical for both the mitogenomes. For instance, most of tRNAs (16) were distributed on the positive (+) strand except tRNA Q(CAA), tRNA A(GCA), tRNA N(AAC), tRNA C(TGC), tRNA Y(TAC) and tRNA S2(TCA) that were distributed on the negative (−) strand. Both rRNAs along with all the PCGs, except nad6, were present on the negative (−) strand.

Two non-coding control regions were found and referred to as CR1 and CR2. The 5' boundary of CR1 was tRNA T(ACA) and the 3' boundary was tRNA P(CCA) while CR2 was present between tRNA E(GAA) and tRNA F(TTC). Lengths of CR1 and CR2 were 1138 bp and 976 bp, respectively in de-novo assembly (at par with the

---

**Figure 1.** The mitochondrial genome view of *Turdoides affinis*. Gene transcription direction is indicated by arrows. Colour codes are indicated at the right upper side of the figure. tRNAs are indicated with the single letter code of amino acids. Black sliding window indicated the GC content of all the regions. GC skew has been plotted through green and violet colour sliding windows. The figure was drawn by CGView Online server (https://s.tothard.afns.ualberta.ca/cgview_server/) using default parameters. The photograph of *Turdoides affinis* was taken by the first author and was edited with Paint.net.
other compared species) while for reference-based assembly CR1 was 825 bp (less than the average CR1 length of other compared species by 300 bp) and CR2 was 1528 bp long (extra 390 bp than the average CR2 length of compared species). The nucleotide composition of both CR1 and CR2 was calculated. AT of CR1 was 54.63% (45.37% GC) for de-novo assembly and 53.68% (46.32% GC) for reference-based assembly. CR2 showed 53.78% AT (46.22% GC) and 55.9% AT (44.1% GC) for de-novo and reference-based assembly respectively indicating a bias towards an AT for these regions. rRNAs, tRNAs and PCGs were arranged in the following manner in both the assemblies:

- tRNA F-rnS- tRNA V-rnL- tRNA L2-nad1- tRNA I- tRNA Q- tRNA M-nad2- tRNA W- tRNA A- tRNA N- tRNA C- tRNA Y-cox1- tRNA S2- tRNA D- tRNA K-atp8-atp6-cox3- tRNA G-nad3- tRNA R-nad4l-nad4- tRNA H- tRNA S1- tRNA L1-nad5-cob- tRNA T- tRNA P-nad6- tRNA E.

These results showed considerable differences between reference-based assembly and de-novo assembly.

### Nucleotide composition and translational efficiency

The newly sequenced complete mitogenome of *T. affinis* was compared with other available mitogenomes from the Leiothrichidae family (Table 2). We considered *Garrulax affinis* (KT182082.1)19, *G. albogularis* (KX082660.1), *G. canorus* (KT633399.1), *G. canorus* (JQ348398.1), *G. cineraceus* (KF926988.1), *G. elliotii* (KT272404.1)22, *G. formosus* (KR020504.1)23, *G. ocellatus* (KT233599.1).
Translational selection along with translational efficiency plays a pivotal role in natural selection escorting central dogma. The nucleotide at the third degenerating position of the codon is responsible for the superlative codon-anticodon binding energy. However, when the first two positions are taken uniquely when G or C is present in the first two positions. On the contrary, when the first two positions are taken calorically when A or U; C is the ‘right choice’ (at the third position)\(^38\). Thus, translational efficiency can be characterized by the P2 index, which allows us to choose between the pyrimidines in codons with UU, UA, AA, AU, GG, GC, CG and CC in the first and second position. Results from this analysis clearly (P2 > 0.5) validated the accountability of translational efficiency suggesting that those genes may have gone through several adaptations with rapid changes in their expression level, which might be enhancing the capability of mitochondrial metabolic processes substantiating boisterous nature of these birds. This result showed similarity with a previous study on Dragonflies where the increased mitochondrial capacity aided their elevated flight ability\(^42\).

### Table 2. List of complete mitochondrial genomes used for comparative mitogenomic study.

| Species                  | Accession no | Total length | AT skew | GC skew | PCG no | tRNA no |
|--------------------------|--------------|--------------|---------|---------|--------|---------|
| Turdoides affinis        | MN48144      | 17,672       | 0.13    | -0.38   | 13     | 22      |
| Garrulax affinis         | KT182082.1   | 17,856       | 0.10    | -0.39   | 13     | 22      |
| Garrulax albogularis     | KX082660.1   | 17,870       | 0.12    | -0.40   | 13     | 22      |
| Garrulax canorus         | KT633399.1   | 17,828       | 0.11    | -0.38   | 13     | 22      |
| Garrulax canorus         | JQ348938.1   | 17,785       | 0.11    | -0.38   | 13     | 22      |
| Garrulax ellioti         | KT272404.1   | 17,873       | 0.10    | -0.38   | 13     | 22      |
| Garrulax cineraceus      | KP926988.1   | 17,808       | 0.11    | -0.38   | 13     | 22      |
| Garrulax fornsos         | KR020504.1   | 17,869       | 0.10    | -0.38   | 13     | 22      |
| Garrulax oeicillatus     | KF995195.1   | 17,828       | 0.11    | -0.37   | 13     | 22      |
| Garrulax perspicillatus  | KF997865.1   | 17,873       | 0.10    | -0.38   | 13     | 22      |
| Garrulax poecilorhynchos | KR909134.1   | 17,814       | 0.11    | -0.39   | 13     | 22      |
| Garrulax sannio          | KR869824.1   | 17,840       | 0.11    | -0.37   | 13     | 22      |
| Garrulax sannio          | KT373847.1   | 17,848       | 0.11    | -0.37   | 14     | 22      |
| Leiothrix argentauris    | KJ690245.1   | 17,833       | 0.11    | -0.37   | 13     | 22      |
| Leiothrix lutea          | JQ423933.1   | 17,615       | 0.11    | -0.38   | 13     | 22      |
| Liocichla omiensis       | KU886092.1   | 17,830       | 0.11    | -0.37   | 13     | 22      |
| Minla ignotincta         | KT995474.1   | 17,868       | 0.11    | -0.38   | 13     | 22      |
| Trochalopteron milnei    | MH238447.1   | 17,871       | 0.10    | -0.38   | 13     | 22      |

Moreover, values of WWC, SSU, WWU and SSC calculated from the RSCU tables for detecting the translational efficiency clearly indicated the preference of WWC and SSC over WWU and SSU. This pointed towards codon selection of C between the pyrimidines (C or U) at the third position of the codon. Calculated P2 values were greater than 0.5 for all the protein-coding genes in the investigated Leiothrichidae mitogenomes (Table 1, Supplementary file 3) signifying the pivotal role of translational efficiency in dictating the codon usage pattern. The GC3-ENC plots (Fig. 2a) of protein-coding genes of all studied mitogenomes were below the curve designating the influence of translational selection over mutational pressure. Therefore, values of AT skew ranged from 0.09 (Trochalopteron milnei (MH238447.1)) to 0.13 (Turdoides affinis (MN848144)) among all species—GCC(A), UGC(C), UUC(F), GGA(G), AAA(K), CUA(L), UUA(L), AUA(M), CUC(P), CAA(Q), AGC(S), ACC(T), AAS(U), GUA(V), along with these GCU(A), GCA(A), GAC(D), GAA(E), CAC(H), AUA(I), CUC(L), CUU(L), AAU(N), AAA(N), CAC(P), CGC(R), UCU(S), AUC(T), GUC(V), UGA(W) and UAC(Y) were also frequently used in T. affinis. Heatmaps based on codon and amino acid usage (Supplementary file 4) analysis of compared mitochondrial genomes validated the codon preference.

GC3 vs. ENC plot analysis has been proved very efficient in predicting whether translational selection or mutational pressure persist over the genes of interest\(^35\). The GC3-ENC plots (Fig. 2a) of protein-coding genes of the compared mitogenomes were placed well below the curve indicating the predominance of selection pressure over mutational bias. RSCU analysis revealed a well degree of concordance among Leiothrichidae mitogenomes from the codon usage perspective (Fig. 2b, Supplementary file 3). To substantiate the factors governing this codon selection, GC3-ENC plot analysis was done. It has been proposed that, GC3-ENC plot of genes should be placed on or above the continuous ENC curve when only mutational pressure prevails. However, in the presence of translational selection, the plots should fall well below the aforementioned curve\(^35\). Here the GC3-ENC plots of all protein-coding genes of all studied mitogenomes were below the curve designating the influence of translational selection over those genes. Hence, we conclude that the codon usage pattern of T. affinis mitogenome along with other examined species is affected by the pervasiveness of translational selection over mutational pressure.

As41 enhancing their competence in the last phase of the mitochondrial respiratory chain and increasing the possibility of the superlative mitochondrial capacity is responsible for the superlative mitochondrial-anticodon binding energy\(^38\). Previous studies have found that U is preferred at the third position specifically when G or C is present in the first two positions. On the contrary, when the first two positions are taken by A or U; C is the ‘right choice’ (at the third position)\(^38\). Thus, translational efficiency can be characterized by the P2 index, which allows us to choose between the pyrimidines in codons with UU, UA, AA, AU, GG, GC, CG and CC in the first and second position. Results from this analysis clearly (P2 > 0.5) validated the accountability of translational efficiency suggesting that those genes may have gone through several adaptations with rapid changes in their expression level, which might be enhancing the capability of mitochondrial metabolic processes substantiating boisterous nature of these birds. This result showed similarity with a previous study on Dragonflies where the increased mitochondrial capacity aided their elevated flight ability\(^42\).
Comparative mitochondrial genomics. The select mitogenomes were compared with *T. affinis* mitogenome (*do- novo* assembled) for tRNA anticodons, start and stop codons, strand variability, intergenic and overlapping regions, GC/AT skew and RSCU (Supplementary file 3). The comparative anticodon analysis revealed an identical pattern of anticodon usages for all tRNAs among the investigated mitogenomes. Most of the protein-coding genes start with ATG in all the mitogenomes. AGG, AGA, TAG and TAA were identified as stop codons for most of the PCGs; however, in every mitogenome, there were some stop codons, which could not be perfectly
identified (Supplementary file 3). Strand variation property showed an exact pattern among the studied mitogenomes. Both rRNAs were on positive (+) strand. Except for nad6, all other PCGs were also on the positive (+) strand. While tRNA Q, tRNA N, tRNA C, tRNA Y, tRNA A, tRNA S2, tRNA P and tRNA E were located on the negative (−) strand, the other tRNAs were on positive (+) strand. RSCU based analysis revealed GCC(A), UGC(C), UUC(F), GGA(G), AAA(K), CUA(L), UUA(L), AUA(M), CCU(P), CAA(Q), AGC(S), UCC(S), ACC(T) and GUA(V) as optimal codons in the investigated mitogenomes. The comparative analysis of intergenic and overlapping regions also revealed an identical pattern among studied mitogenomes. Intergenic regions were found between trnL2 and nad1, nad1 and trnI, trnI and tRNA Q, tRNA A and tRNA N, tRNA N and tRNA C, tRNA Y and cox1, tRNA S2 and tRNA D, tRNA D and cox2, cox2 and tRNA K, atp6 and cox3, nad5 and cob, cob and tRNA T, tRNA T and tRNA P along with tRNA P(tgg) and nad6 (Supplementary file 3) for all select mitogenomes. The intergenic region between tRNA T(tgt) and tRNA P(tgg) were longest ranging from 939 bp (G. cineraceus) to 1139 bp (G. albogularis). Overlapping regions were much shorter, ranging from 2 to 7 bp (Supplementary file 3). The highest overlapping length was between nad4l and nad4 (4 bp for G. poecilorhynchus and T. affinis, and 7 bp for other mitogenomes; Supplementary file 3). These characters revealed identical patterns for the aforementioned characteristics in the compared mitogenomes.

Comparative tRNA structure analysis of T. affinis de-novo mitogenome. The wobble base pairing which does not follow the Watson–Crick base pairing rule, often substituting GC or AT base pairs, contributing to thermodynamic stability is of immense importance in studying the tRNA structure. Hence, while understanding the exact functional features of mitogenomes, tRNA acts as a pivotal tool. In T. affinis, all the tRNAs were folded into the classic secondary clover-leaf structure. In T. affinis mitogenome, though Watson–Crick base pairing dominated, wobble base pairing was also detected (Supplementary file 5). Three consecutive wobble base pairs were detected at the DHU loop of trnS2. In addition, mismatched base pairs were found in tRNA D, tRNA E, tRNA G, tRNA H, tRNA I, tRNA M and tRNA L2.

Control region of T. affinis de-novo mitogenome. Vertebrate mitochondrial control region (CR) is divided into three domains (I, II and III). Among all the investigated species of the Leiothrichidae family, a duplication of CR was observed. All the domains and conserved boxes were identified through sequence alignment. Conserved bird similarity box (BSB) was prominent in both CR1 and CR2 (Supplementary file 7). No tandem repeats were found among the CRs. Duplication of CR region is important in the regulation of replication and transcription within mitochondrial genome. Moreover, duplicated CR is also associated with extended longevity of bird species. Thus, the present study reported the genetic features of duplicated CR among select Leiothrichidae members including T. affinis, which will further be helpful in evolutionary analysis of this group.

Phylogenetic and evolutionary analysis. We performed a limited phylogenetic analysis with both the mitogenomes and observed that de-novo assembled mitogenome performed better. The phylogenetic analysis of reference-based assembled mitogenome placed T. affinis with Leiothrix lutea (Supplementary file 2), whereas in the case of de-novo assembled mitogenome based phylogeny, T. affinis formed a discrete but sister group of Garrulax (Fig. 3a,b). The latter finding is consistent with a recent report that revised the taxonomy of Leiothrichidae using a set of nuclear genes with mitochondrial protein-coding genes as phylogenetic marker. Though results of our phylogenetic analysis are limited because of the unavailability of complete mitogenome sequences

Figure 3. Phylogenetic tree based on complete mitogenome (de-novo assembled Turdoides affinis mitogenome). (a) ML based phylogeny with 1000 bootstrap values. (b) Bayesian phylogeny using GTR algorithm was constructed with MrBayes 3.2.6 version.
of other groups, it provides evidence that de-novo assembled mitogenome is more appropriate than a reference-based assembly approach.

Evolutionary distance estimation revealed higher K2P values in nad2, atp8 and nad6, whereas cox1 possessed the lowest value among the Leiothrichidae family (Fig. 4). Regarding the Ka/Ks analysis, the average synonymous substitution rate (Ks) for nad6 gene was highest whereas the non-synonymous substitution rate (Ka) was highest for atp8. The ω values of the protein-coding genes ranged from 0.015 to 0.188 and was in the following order: cox1 < cox3 < cob < nad1 < nad6 < atp6 < cox2 < nad2 < nad5 < nad4l < nad3 < atp8 (Supplementary file 3). Lowest K2P values for the cox1 gene indicated its conserved nature among the Leiothrichidae family, whereas the highest synonymous substitution rate in nad6 implicated that its protein sequence is conserved among the Leiothrichidae family. This gene codes NADH dehydrogenase 6 which is a subunit of the complex-I located in the inner mitochondrial membrane. Given the highest synonymous substitution rate in nad6 gene, suggestive of its importance in complex-I of mitochondria, it is reasonable to speculate that complex-I is largely used in ATP synthesis in the family Leiothrichidae; however, this needs further confirmation. Complex-I is responsible for exporting four H+ out of the mitochondrial matrix, generating H+ gradient across the mitochondrial membrane, which in effect speeds up ATP generation whereas this mechanism is absent when complex II is used

Highest Ka value of atp8 points to a highly variable nature of this protein indicating the erratic nature of mitochondrial atp8 among vertebrates. The ω (Ka/Ks) values of all the protein-coding genes were < 1 suggesting the persistence of purifying selection against deleterious mutations. Thus, the evolutionary analysis aids in understanding the influence of natural selection manipulating species evolution along with the interaction between selection and mutational pressure responsible for protein evolution which has been already suggested.

Conclusion
We report for the first time the complete mitogenome of Turdoides affinis (MN848144). Our results suggested that the de-novo assembly approach is more appropriate than a reference-based assembly approach. The comparative mitogenomics of the Leiothrichidae family reveals their preference towards AT-rich codons as well as the persistence of translational efficiency. Duplicated control regions are found among Leiothrichidae family mitogenomes that may be associated with their extended longevity. Evolutionary analysis confirms that protein-coding genes are under purifying selection pressure. Our limited phylogeny results place T. affinis in a separate group, a sister group of Garrulax.
46. Kundu, S. et al. Complete mitochondrial genome of Black Soft-shell Turtle (Nilssonia nigricans) and comparative analysis with other Trionychidae. *Sci. Rep.* 8, 17378–17389 (2018).
47. Ruokonen, M. & Kvist, L. Structure and evolution of the avian mitochondrial control region. *Mol. Phylogenet. Evol.* 23, 422–432 (2002).
48. Skujins, I., McMahon, R., Lenis, V. P. E., Gkoutos, G. V. & Hegarty, M. Duplication of the mitochondrial control region is associated with increased longevity in birds. *Aging* 8, 1781–1785 (2016).
49. Berg, M. J., Tymoczko, L. J. & Stryer, L. *Oxidative Phosphorylation in Eukaryotes Takes Place in Mitochondria. Biochemistry* 5th edn. (W H Freeman, New York, 2002).
50. Kumar, S. Patterns of nucleotide substitution in mitochondrial protein-coding genes of vertebrates. *Genetics* 143, 537–548 (1996).

**Acknowledgements**

We thank the Ministry of Environment, Forest and Climate Change, Govt. of India for financial support. We are also thankful to the Tamil Nadu Forest Department for providing the permissions and support to conduct the study.

**Author contributions**

R.P.S. and V.H.S.K. conceived the idea. R.P.S. and P.P. acquired funding. R.P.S., S.K.S., P.D. and S.D.R. collected samples, designed the experiments and generated DNA data. I.S. and R.S. analysed the data. R.P.S. and I.S. wrote the manuscript and generated all the figures. All authors reviewed the manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-72674-4.

Correspondence and requests for materials should be addressed to S.K.S. or R.P.S.

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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020.