Comparative mitogenomics and phylogenetics of the family Carangidae with special emphasis on the mitogenome of the Indian Scad *Decapterus russelli*

Anjaly Jose1,2*, Sandhya Sukumaran1, Lakshmi P. Mukundan1, Neenu Raj1, Sujitha Mary1, K. Nisha1 & A. Gopalakrishnan1

Carangids are abundant and commercially important marine fish that contribute to a significant portion of the fisheries in many parts of the world. In the present study, we characterized the complete mitogenome of the Indian scad, *Decapterus russelli* and performed a comprehensive comparative mitogenomic analysis of the family Carangidae. The comparative mitogenomics provided valuable insights into the structure, variability, and features of the coding and non-coding regions that evolved across species over millions of years. The structural features of tRNAs revealed changes in the frequency of mismatched and wobble base pairs, which is reflected in the base composition of H and L strands. The highly conserved sequence motif of the mTERF binding site in carangids over the ~400 MYA of their divergence demonstrated the functional importance of these sites. The control region of carangids was characterized by the presence of discontinuous repeat units with a high rate of sequence divergence in the form of base substitutions, insertions, and deletions. The maintenance of secondary structures in the control region independent of the rapid evolution of primary structure suggested the effect of selective constraints on their maintenance. Maximum likelihood (ML) and Bayesian inference (BI) phylogeny revealed a similar topology consistent with previous taxonomic studies. The extant carangids diverged through the evolutionary events experienced during the Cretaceous, Paleogene, and Neogene periods.

The carangids represent an important group of exploited pelagic coastal resources of the world with diverse morphological features comprising approximately 147 species1. They occupy a variety of habitats consisting of pelagic, benthopelagic and reef-associated regions in all tropical and subtropical seas2,3. Currently, 4 subfamilies are recognized within the Carangidae4. Annual global capture production of Carangidae was estimated at 5.13 million tons in 2018, 65% of which came from Asia (https://www.fao.org/fishery/en/statistics). Among the carangids, one of the commercially most important species is the Indian scad, *D. russelli* (Ruppell, 1830), providing a cheap source of animal protein. It is also commonly used as live fish bait5. It is widespread in the Indo-Pacific and forms an important fishery resource in India and Southeast Asia6. Most studies on carangids have focused on biology, ecology, and taxonomy7–18, and it is important to study the genetic and genomic traits to resolve taxonomic ambiguities, understand evolutionary traits, and implement conservation strategies.

The vertebrate mitochondrial genome, which plays a crucial role in cell metabolism and organism function, is ~15–20 kb in size and consists of 13 protein-coding genes (PCGs), 22 tRNA genes, 2 rRNA genes and a non-coding region (control region)19. The PCGs, tRNA and rRNA have a relatively low mutation rate because they are subject to strong purifying selection20. In contrast, the control region evolves at a high rate compared to the other parts of the mitogenome, and these high rates of sequence divergence are mostly localized in the variable flanking region, which is adjacent to the conserved sequence motifs21–23. The extensive comparative studies in fish24 have demonstrated the ability of the control region to form stable secondary structures and the existence

1Marine Biotechnology Division, ICAR-Central Marine Fisheries Research Institute, Ernakulam North P O, Kochi, Kerala 682018, India. 2Mangalore University, Mangalagangotri, Mangalore, Karnataka 574199, India. *email: anjalijose2@gmail.com
of selective constraints for their maintenance. Although Zhu et al. analyzed the phylogenetic relationships of carangids using the control region, there is a gap in knowledge regarding the mutational pattern and the non-random distribution of these mutations between the conserved motifs and their flanking regions. There is also a lack of knowledge about the structure and variability of rRNA genes in carangids and the frequency of occurrence of G–U wobble base pairs, the fundamental building block of RNA structure critical for RNA function in various biological systems.

The phylogenetic relatedness between the carangids has been extensively studied using comparative morphological and molecular tools. Although there has been a discourse on monophyly vs. paraphyly of carangids, the congruent topologies of phylogenetic trees constructed using DNA sequences suggested paraphyly of carangids if Echeneoids were not included. Carangids can be considered as promising candidates for evolutionary studies, as they have a wide diversity in terms of habitat (marine, brackish water, freshwater), morphology (having variable body size; less than 20 cm to ~ 2 m), and shape (deep-bodied, streamlined). Some species of remoras live even as commensals on large pelagic vertebrates. The phylogenetic study inferred from nuclear gene supermatrix suggested a terminal cretaceous crown age for carangids with major lineages diversifying during the Paleogene. A late Eocene age has also been proposed for the crown carangids. But densely sampled molecular studies indicated a Cenomanian stem age of the Carangoids which had undergone a significant diversification during the late cretaceous. The subfamily Naucratinae originated early in the Paleocene, Trachinotinae date back to the early Oligocene and the remaining Carangoid lineages including the morphologically heterogeneous group, Caranginae comprising the genus Decapterus appear to have originated during the middle Eocene. The genus Decapterus originated in the upper Miocene. The genetic distance between the genus Decapterus was significantly greater than that between Trachurus, suggesting that Decapterus arose long before its sister group Trachurus, making them the most divergent species group. The morphology of the branched eels of the genus Decapterus includes multiple forms such as long-bodied, cylindrical, mackerel-like with a small mouth, reduced teeth and reduced scutes on the straight part of the lateral line, suggesting adaptations to the wide ranging off-shore life.

In the present study, we report the complete mitogenome of Decapterus for the first time elucidating the special characteristics. Although recent studies have examined the evolutionary history of carangids, we reanalyzed the patterns by including mitogenomes from 37 species and describing the relationships and evolutionary patterns among species with a robust phylogenetic tree. We also performed a comprehensive comparative analysis of carangid mitogenomes especially the mutational patterns of coding and non-coding regions, control region secondary structure, and structural features of tRNA to gain insights into the evolutionary dynamics of mitochondrial DNA.

Results and discussion
Mitochondrial genome organization and composition. The total length of the mitochondrial genome of Decapterus was 16,542 bp. The genome organization is congruent with that observed in other teleost fishes with a standard set of 13 putative PCGs, 22 tRNA genes, 2 tRNA genes, and a control region. All genes were encoded on the heavy strand (H) except for the ND6 and 8 tRNA genes which were encoded on the light strand (L) (Supplementary Fig. S1). A similar gene arrangement was observed in other Carangidae species studied (Supplementary Table S1). The overall nucleotide composition of Decapterus mitochondrial genome was 27.5% A, 30.2% C, 25.4% T, and 16.9% G. The AT and GC skew values of the whole mitogenome were 0.039 and − 0.028 respectively.

The results suggested that A and C are predominant base constituents than T and G, with the overall nucleotide composition being severely C-skewed. These results agree well with their respective counterparts in other Decapterus spp. The bias against G was presented following the results reported previously in other teleosts. The AT/GC- skew of other Carangidae species were also determined (Supplementary Table S2). The sequence of the complete mitogenome of Decapterus has been deposited with NCBI, GenBank with the accession number MN711693.

Overlapping and Intergenic spacer regions. Seven gene overlapping regions ranging from 1–7 bp and 11 spacer regions varying in length from 1 to 38 bp were detected in the Decapterus mitogenome with the longest intergenic spacer region observed between tRNA Asn and tRNA Cys. Significant differences in the number and length of both the intergenic spacer region and the overlapping gene regions were observed in carangids (Supplementary Table S3). The number of spacer and overlapping gene regions varied between 10–12 and 6–15, respectively, between species. The intergenic spacer is distributed over 11 regions in almost all carangid species studied so far. However, an additional spacer region was observed in C. armatus (tRNA-His-tRNA-Ser (GCT)), E. bipinnulata (ND6-tRNA-Glu), G. speciosus (tRNA-Val-16SrRNA), P. niger (16SrRNA-tRNA-Leu (TAA)), S. crumenophthalminus (ND6-tRNA Glu) and S. dumerili (ND6-tRNA Glu). In addition, the spacer region found between tRNA-Arg and ND4L in all species was absent in Seriola spp, S. nigrofasciata, T. carolinus and T. ovatus. Besides, C. equula and P. dentex lacked a spacer region found between ND1 and tRNA-Ile which was present in all other taxa. A comparative analysis among 37 carangids revealed that the longest spacer (42 bp) occurred between tRNA-Asn and tRNA-Cys of T. carolinus. Seven regions of overlapping gene regions were observed in most carangids, with the longest overlapping region being between ATP8 and ATP6 in some species and between ND4L and ND4 in others.

Protein Coding Genes (PCGs). The total length of PCGs in the Decapterus mitogenome was 11,425 bp, ranging in size from 165 bp (ATP8) to 1839 bp (ND5) accounting for 69% of the entire genome. GC skew values of all PCGs were negative except for the case with the L-strand encoded gene, ND6, which was greater than zero.
Comparison among carangids revealed that most PCGs started with the ATN initiation codon (Supplementary Table S4). Besides, 6 genes exhibited incomplete termination codons, either TA (COIII) or T (ND2, COII, ND3, ND4, and Cyb-b). The remaining genes revealed complete stop codons in open reading frames either TAA (ND1, COI, ATP6, and ND4L) or TAG (ATP8, ND5, and ND6). The incomplete codons can be completed by polyadenylation of the messenger RNA after cleavage. Comparative analysis among carangids revealed that most PCGs started with the ATN initiation codon (Supplementary Table S4). COI of all the species started with TAG instead of ATC in G. speciosus. Other PCGs start with ATG except ATA for D. russelli, T. blochii, and D. troschelii. Codon usage and skew values of all the species started with GTG except ATC in A. djedaba.

Table 1. Sequence characteristics of Decapterus russelli mitochondrial genome.
We observed that the value of Ka was significantly smaller than Ks in all 13 PCGs on analyzing the evolutionary rate of PCGs. This indicates that there is a much lesser chance for a mutation to be different between species that can modify the amino acid sequences of a protein sufficiently to change its function than one which is silent\(^{52}\). The average Ka/Ks ratios substantially differed among PCGs indicating varying functional constraints experienced by individual genes (Supplementary Fig. S2; Supplementary Table S5). The fundamental features of genome evolution were largely determined by the influence of two non-adaptive forces, mutation pressure and genetic drift\(^{53}\). Nevertheless, the deleterious effect of mutation is very difficult to establish under purifying selection\(^{54}\). In PCGs, the average Ka/Ks ratio varied from 0.008 in COI to 0.469 in ATP8. Altogether the results indicated that the functional genes of carangids have undergone strong purifying selection to eliminate deleterious mutations thus maintaining the protein structure.

We estimated the degree of mitochondrial gene conservation by assessing the total p-genetic distance between 37 Carangidae species (Supplementary Fig. S3). In our analysis, we obtained the highest overall p genetic distance for the ATP8 (0.301) and the lowest value for the COI gene (0.019) based on the data of the first and second nucleotide position of codons. The same trend was also observed in the values discerned from the full-length comparison of each gene, with ATP8 and COI represented by a total p-genetic distance value of 0.390 and 0.134, respectively. Based on these results it could be inferred that gene, ATP8 has the fastest evolutionary rate among Carangidae and COI the slowest. Besides, all genes had overall high p-genetic distance values for wobble nucleotide position, ranging from 0.393 to 0.564. This result was consistent with previous reports in other fish species that most of the differences occurred at the 3rd codon position of PCGs in mtDNA\(^{55}\).

**Codon usage analysis.** A total of 3800 amino acids of PCGs are encoded in the mitogenome of *D. russelli*. The codon usage patterns are shown in Supplementary Fig. S4A. The amino acid Leu was represented by six different codons and the rest of them by either four or two. The amino acids Leu (16.46), Ala (8.75), and Thr (7.96) were most abundant in the *D. russelli* mitogenome. Among these, the frequency of Leu (CUC, CUA), Ala (GCC), and Thr (ACC) was highest. The frequencies of stop (UAG), Cys (UGU), and Arg (CGG) in PCGs were lowest. A comparative analysis showed that the codons ending in A or C occurred most frequently in amino acids with sixfold and fourfold degenerate third positions in Carangidae. In a two-fold degenerate codon, C was used more frequently than T. Consistent with the overall bias against G\(^{56}\), it was the least common wobble position nucleotide in all categories. However, analysis of relative synonymous codon usage (RSCU) showed that the codons UCC for Ser, CUC and CUA for Leu and GCC for Ala occurred most frequently, while UUG for Leu, AGU for Ser and GGC for Ala were rare in the mitogenome of Carangidae (Supplementary Fig. S4B). It was also inferred from the RSCU value that the synonymous codons NNA and NNC were in the majority than the codons NNT and NNG (Supplementary Table S6).

**Transfer RNAs and ribosomal RNAs.** The mitogenome of *D. russelli* contained 22 tRNA genes typical of the metazoan mitogenome. In total, the tRNA genes contributed 1555 bp to the entire genome, which ranged from 67 bp (tRNA-Cys) to 74 bp (tRNA-Lys (TAA)). The secondary structure of tRNA genes predicted by ARWEN\(^{57}\) makes it clear that all tRNAs could be folded into a canonical cloverleaf structure, with the exception of tRNA-Ser (GCT), in which no recognizable dihydrouridine (DHU) stem was found (Supplementary Fig. S5). Comparative analysis with other carangids revealed that the total length of tRNAs varied from 1548 (*S. rivoliana* and *E. bipinnulata*) to 1564 (*T. blochii*). The overall A + T content was estimated at 54.2%. Base skew analysis of 22 concatenated tRNA genes (Supplementary Table S2) revealed a positive AT skew (0.125) and a negative GC skew (−0.098). The comparisons showed that the AT skew varied from 0.105 (*S. nigrofasciata*) to 0.224 (*G. speciosus*). The AT content of the 12S rRNA gene was 52.2% with a positive AT skew of 0.172 (*G. speciosus*). The presence of a higher proportion of A and C bases in rRNA genes indicates an AC-rich trend, consistent with previous reports from other teleosts\(^{50,59,60}\). Consistent with previously reported results from other mitochondrial genomes\(^{41,63}\), these two rRNA genes were positioned between tRNA-Phe and tRNA-Leu (TAA) separated by tRNA-Val. Comparative analysis among carangids (Supplementary Fig. S6) indicated that the length of two concatenated rRNA genes varied from 2655 bp (*G. speciosus*) to 2683 bp (*A. kleinii*). The AT skew varied from 0.1615 to 0.2256 and the GC skew varied from −0.124 to −0.077 (Supplementary Table S2).

**Features of tRNAs.** 

- **Secondary structure.** The comparative analysis of 37 carangids revealed that most of the tRNAs are structurally different in both the stem and the loop. However, 2 tRNA genes, tRNA-Trp, and tRNA-Leu (TAG) exhibited similarities in their secondary structure (Fig. 1). As can be seen from Table 2, length...
variation was detected in all four stems. The anticodon stem (AC stem) of all the tRNAs was 5 bp, except for tRNA-Ser (GCT), which displayed length variation (5 or 6 bp). Though the pseudouridine (ΨU) stem of most tRNAs was 5 bp, tRNA-Phe (4 bp), tRNA-Met (4 or 5 bp), tRNA-Lys (4 bp), and tRNA-Ser (GCT) (6 bp) indicated length variation. In the acceptor stem (AA stem), tRNA-Ser (TGA) displayed a length of 6 bp, and the rest of the tRNAs showed more of a length variation of either 7 bp or a range of 7 to 8 bp. Although the DHU stem was mostly 4 bp in length with a range of 3 to 5 bp, tRNA-Ser (GCT) of all carangids examined had only a small loop with no stem.

Of the four loops, the length of the AC loop was fixed (4 nucleotides), whereas the remaining three loops showed considerable variation in length. Of these, the DHU loop was extremely variable, ranging from 3–13
nucleotides. Conversely, the variable loop (V loop), mostly 4 nucleotides with a range of 3–5, and TΨU loop, mostly 7 nucleotides with a range of 6–9 displayed the least variation.

Wobble base pairs in the stem regions. The wobble base pair is a unique feature of RNA structure and their often high evolutionary conservation underscores their functional and structural importance. Due to its thermodynamic stability, the G-U often substitutes G-C or A-U base pairs and it differs from Watson–Crick and other mismatched base pairs owing to its structural, conformational, and functional properties. Diverse proteins and other ligands specifically targeting tRNA exploit these features and thus the comparative analysis of tRNAs is very crucial for understanding the structural and functional features of the mitogenome. Supplementary Table S8 illustrates the frequencies of Watson–Crick and Wobble pairing at each of the 25 base pair sites of each tRNA in 37 carangids (8 in AA stem + 5 in DHU stem + 6 in AC stem + 6 in TΨU stem). Seven sites from each stem (1–87, 2–86 and 8–80 base pair sites in the AA stem, 12–32 and 15–30 in the DHU stem, 40–48 in the AC stem, and 64–74 in the TΨU stem) revealed a high rate of (100%) Watson–Crick and Wobble base pairing (hydrogen bonding) and these sites may play a pivotal role in maintaining stem structure. Among the tRNAs, tRNA-Ala, tRNA-Asn, tRNA-Cys, tRNA-Ser (TGA), tRNA-Asp, and tRNA-Pro had a higher frequency (≥ 98% overall average) of hydrogen bonding than other tRNAs. Moreover, L strand-encoding tRNAs exhibited, on average, a higher frequency of hydrogen bonding (97.48%) than H strand encoding tRNAs (94.78%).

Table 3 shows the number of G-U wobble base pairs in each stem of 22 tRNAs for the 37 carangid species. Individual stems and tRNAs rather displayed considerable variations in the frequency of GU base pairs, 6% of the base pairs comprised of wobble base pairs on average in the four stems of mitochondrial RNA with a large difference between stems (4.57% in the TΨU stem and 12.26% in the DHU stem). A substantial difference was also discerned between the tRNAs (0.72% in tRNA-Ile versus 19% in tRNA-Glu) with no wobble base pairs observed in the case of tRNA-Lys and tRNA-Leu (TAG). Furthermore, the tRNAs encoded by different strands exhibited a significant difference in the frequency of wobble base pairs, as the L strand encoded tRNAs had a higher frequency of wobble base pairs (13.18%) than those of H strand encoded tRNAs (3%) (Mann–Whitney U test, p<0.01).

Mitochondrial transcription termination factor (mTERF) binding site. We compared tRNA-Leu (TAA) which has been reported to contain the target binding site for mTERF in humans to identify its presence in Carangidae. An important tridecameric sequence (5’TGCCAGAGCCTGG3’) reported in the human mitochondrial tRNA-Leu (TAA) gene was also identified in Carangidae in the same region as the corresponding part of the tRNA-Leu (TAA) gene.
their DHU arm. The mTERF binds to this site with high affinity and regulates the level of transcription from the rRNA genes into the remaining genes of the major coding strand.

All bases in the 13 nucleotide positions identified in 37 carangid taxa were completely identical to those in humans (Supplementary Fig. S7), suggesting that this site in carangids may follow the same mechanism of function as proposed for humans.

Control region: mutational patterns and secondary structure. The control region of *D. russelli* was shorter than other carangids, consisting of 840 bp flanked by tRNA-Pro and tRNA-Phe on either side. The control region comprised the highest A + T content (63.6%) compared to other regions of the mitogenome (Supplementary Table S2). AT skew was positive (0.038) and GC skew was negative (−0.104). Considerable variation was observed in the AT and GC skew values of other carangids, ranging from −0.047 (S. crumenophthalmus) to 0.065 (T. blochii) and −0.278 (C. ignobilis) to −0.093 (D. maruadsi), respectively. The composition of the nucleotide adenine (A) is higher than that of thymine (T) in *D. russelli* compared to other carangids (Supplementary Table S2). However, the control region of *A. mate*, *A. djedaba*, *C. ignobilis*, *G. speciosus*, and *T. blochii* possess a slightly higher percentage of adenine (A) than that of *D. russelli*. Besides, the composition of thymine (T) is higher than adenine (A) in *C. armatus*, *C. malabaricus*, *C. tille*, *C. melanpygas*, *T. japonicas*, *S. crumenophthalmus*, *S. dumerili*, and *S. lalandi*. The GC skew of all the carangids was negative. It was found that the maximum length variation for the control region was observed in the range of 840 bp (*D. russelli*) to 904 bp (*S. crumenophthalmus*) in the carangids examined. Among the carangids, *S. crumenophthalmus* possessed the longest mitogenome, and *S. rivoliana* and *S. dumerili* were determined to have the shortest mitogenome. It was evident that these species do not exhibit any notable variation with respect to their PCGs, tRNA and rRNA (Supplementary Fig. S6), hence the 61–62 bp variation in the control region alone contributed to the difference in their overall mitogenome length. The majority of the primary sequences of the control region are presumed to be not involved in regulatory function as this region shows great variation even among closely related species.

Comparative mitogenomics of carangids revealed high variability in the mitochondrial control region due to length variations and the accumulation of base substitutions and indels. Consistent with the recognition sites in many marine fish, we have detected the 3 central conserved sequence blocks (CSB- F, CSB- E, and CSB- D) and 3 conserved sequence blocks (CSB- 1, CSB- 2, and CSB- 3) in all taxa. Palindromic sequence motifs—"TACAT"

### Table 3. Occurrence of wobble pairings in the stem regions of 22 tRNA genes in the mitogenome of 37 species of Carangidae.

| tRNA AA stem | H strand coded | L strand coded | Total base pairs | Percentage |
|--------------|----------------|----------------|------------------|------------|
|              | Wobble bp      | Wobble bp      | Wobble bp      | Wobble bp  | Total |
| Phe          | 10             | 6              | 0               | 1          | 17   | 2.5  |
| Val          | 2              | 2              | 0               | 3          | 7    | 0.93 |
| Leu(TAA)     | 0              | 37             | 0               | 3          | 40   | 4.6  |
| Ile          | 6              | 0              | 0               | 0          | 6    | 0.72 |
| Met          | 33             | 8              | 0               | 11         | 44   | 5.50 |
| Trp          | 2              | 35             | 0               | 0          | 37   | 4.80 |
| Asp          | 2              | 0              | 0               | 17         | 19   | 2.50 |
| Lys          | 0              | 0              | 0               | 0          | 0    | 0    |
| Gly          | 0              | 65             | 0               | 0          | 65   | 8.40 |
| Arg          | 8              | 6              | 2               | 18         | 34   | 4.34 |
| His          | 2              | 7              | 0               | 23         | 32   | 4.20 |
| Ser(GGT)     | 0              | 5              | 7               | 12         | 1.64 |
| Leu(TAG)     | 0              | 0              | 0               | 0          | 0    | 0    |
| Thr          | 0              | 0              | 6               | 0          | 6    | 0.80 |
| Average (%)  | 1.72           | 8.27           | 0.50            | 3.29       | 3.00 |
| Glu          | 37             | 37             | 0               | 0          | 74   | 9.46 |
| Ala          | 42             | 30             | 41              | 34         | 150  | 18.82 |
| Asn          | 24             | 37             | 0               | 18         | 79   | 10.03 |
| Cys          | 52             | 34             | 1               | 11         | 88   | 10.81 |
| Tyr          | 37             | 0              | 37              | 0          | 74   | 9.52 |
| Ser(TGA)     | 3              | 6              | 29              | 34         | 66   | 9.40 |
| Gla          | 68             | 9              | 71              | 9          | 148  | 19.00 |
| Pro          | 35             | 74             | 32              | 1          | 142  | 18.13 |
| Average      | 14.13          | 18.84          | 14.25           | 6.75       | 13.18 |
| Coefficient of variation | 1.30 | 1.29 | 1.84 | 1.32 | 0.88 |
| Grand average | 6.18            | 12.26          | 5.50            | 4.57       | 6.70 |

Control region: mutational patterns and secondary structure. The control region of *D. russelli* was shorter than other carangids, consisting of 840 bp flanked by tRNA-Pro and tRNA-Phe on either side. The control region comprised the highest A + T content (63.6%) compared to other regions of the mitogenome (Supplementary Table S2). AT skew was positive (0.038) and GC skew was negative (−0.104). Considerable variation was observed in the AT and GC skew values of other carangids, ranging from −0.047 (*S. crumenophthalmus*) to 0.065 (*T. blochii*) and −0.278 (*C. ignobilis*) to −0.093 (*D. maruadsi*), respectively. The composition of the nucleotide adenine (A) is higher than that of thymine (T) in *D. russelli* compared to other carangids (Supplementary Table S2). However, the control region of *A. mate*, *A. djedaba*, *C. ignobilis*, *G. speciosus*, and *T. blochii* possess a slightly higher percentage of adenine (A) than that of *D. russelli*. Besides, the composition of thymine (T) is higher than adenine (A) in *C. armatus*, *C. malabaricus*, *C. tille*, *C. melanpygas*, *T. japonicas*, *S. crumenophthalmus*, *S. dumerili*, and *S. lalandi*. The GC skew of all the carangids was negative. It was found that the maximum length variation for the control region was observed in the range of 840 bp (*D. russelli*) to 904 bp (*S. crumenophthalmus*) in the carangids examined. Among the carangids, *S. crumenophthalmus* possessed the longest mitogenome, and *S. rivoliana* and *S. dumerili* were determined to have the shortest mitogenome. It was evident that these species do not exhibit any notable variation with respect to their PCGs, tRNA and rRNA (Supplementary Fig. S6), hence the 61–62 bp variation in the control region alone contributed to the difference in their overall mitogenome length. The majority of the primary sequences of the control region are presumed to be not involved in regulatory function as this region shows great variation even among closely related species.
and ‘ATGTA’ were found in multiple copies distributed towards the 5’ end of the control region [Fig. 2; Supplementary Table S9] and are believed to act as the termination site of heavy strand replication69. The segment corresponding to Extended Termination Associated Sequence (ETAS) previously reported in *Scomberoides commersonianus*, a member of the *Carangidae*25, could not be aligned to the corresponding segment in the control region of either species.

No tandem repeat sequences were detected in the control region of *Carangidae*, but the T homopolymer of 5–12 nucleotides between CSB-D and CSB-1 was found in all taxa. Along with these discontinuous multiple interval repeats of ATA TTA and TATAAT were also observed in the control region of *Carangidae*. All species except *D. maruadsi* displayed substantial differences from the key sequences of CSB-E25 and all were substitutions ranging from 1 to 9 nucleotides. Furthermore, *M. cordyla*, *P. dentex*, and *Caranx* spp. lacked a ‘GTGGG’- box which is a typical feature of CSB-E in teleosts69. The CSB-D was highly conserved among carangids and exhibited high sequence similarity between species (Table 4). Although the functions of the CSBs remain unclear, CSB-D is highly conserved in fish families and is involved in regulation of H-strand replication and initiation of D-loop structure70 and could play a potential role in mitochondrial metabolism66. Conserved sequence blocks (CSB-1,-2, and -3) exhibit a higher level of divergence in carangids with interspecies substitutions, insertions, and deletions (Table 4). Except for *P. dentex*, *S. nigrofasciata*, *S. leptolepis* and *Seriola* spp. other species showed a characteristic sequence at the 3’ end (Table 4). These variations imply rapid evolution of the primary structure in the *Carangidae* control region, which can provide information for understanding their structural and functional relationship. Furthermore, these conserved regions are involved in the formation of secondary structures in the control region (Fig. 2) that serve as the target sites for proteins or enzymes69. These conserved sequence motifs formed similar secondary structures in *Carangidae* although their primary sequences are not conserved. Therefore, the data reinforce that although there is sequence divergence in the conserved sequence motif, their structural elements and locations appeared to be conserved in *Carangidae* and some selective constraints on these regions acted to maintain their structure and function.

**Carangidae mitogenome sequence comparison.** We compared the complete mitochondrial genome of *D. russelli* with other available mitogenomes of the *Carangidae* family. The DNA coding sequences showed high interspecies identity (Supplementary Figs. S8, S9). The control region contributed to a larger percentage difference between the nucleotide sequences. It has been found that ND1, COI, COII and COIII are highly conserved among the DNA coding sequences. In addition, ATP8 was the least conservative of all species, in good agreement with the results obtained from analysis of Ka/Ks and overall mean p genetic distance. With the exception of ND6 and Cyt-b, other PCGs expressed orthologous group (COG) clusters. A total of 12 COGs of proteins were detected from 11 PCGs. ND5 was determined to have 2 COGs, pCOG, and cCOG, and the remaining 10 PCGs each expressed a single COG each (cCOG). COGs are a group of proteins inferred to be orthologous i.e. they are direct evolutionary counterparts71. Lineage-specific duplications (orthologous relationships) exist mainly between gene (protein) families rather than between individual proteins72. The identification of COGs associated with specific PCGs could indicate the presence of a specific protein in a specific species, and it is crucial to determine the functional relevance of these proteins in that specific species73. The results indicated that the COGs present in carangids are involved in inorganic ion transport and metabolism (pCOG) and energy production and conservation (cCOG)73. The ND5 gene has been found to perform these two functions, while energy metabolism is the sole function of the remaining 10 PCGs.
| Species | Central conserved blocks | CSB- F | CSB- E | CSB- D |
|---------|-------------------------|--------|--------|--------|
| Reference sequence | ATGTAAGTAGAACCTACC | AGGAGAAGCTATTTGGGTTTT | TATTTCAATTCACCCCAAAC |
| Dr | ****** | ****** | ****** |
| Ac | ******G* | ******GAA* | ****** |
| Ai | ****** | ******G | ****** |
| Ad | ****** | ******TA* | ****** |
| Ak | ****** | ******AA* | ****** |
| Am | ****** | ******GAA* | ****** |
| Ca | ****** | ******AA* | ****** |
|Cb | ****** | ******GAA* | ****** |
| Ce | ****** | ******AA**CC | ****** |
| Cm | ****** | ******AA**A*** | ****** |
| Cp | GCCCA*A* | ******GAA* | ****** |
| Ci | **A***G*AGG*TC*** | **A***GACATTTC* | ****** |
| Cj | **C***GA***GG | **A***AA***C** | ****** |
| Ck | **C***G**CC*GG | **A***AA***A** | ****** |
| Cn | **C***G**CC*GG | **A***AA***A** | ****** |
| Ct | **C***G**CC*GG | **A***AA***A** | ****** |
| Cme | **C***G**CC*GG | **A***AA***A** | ****** |
| Dmar | ****** | ******AA* | ****** |
| Dm | ****** | ******AA* | ****** |
| Dt | ****** | ******AA* | ****** |
| Eb | **C***AAAAAA | ******GTA* | ****** |
| Es | ******G* | ******GAA* | ****** |
| Mc | **C***AAAAAA | ******GAA* | ****** |
| Nm | ****** | ******AA* | ****** |
| Np | ****** | ******AA* | ****** |
| Nq | ****** | ******AA* | ****** |
| Sr | ****** | ******GAA* | ****** |
| Tt | ****** | ******AA* | ****** |
| Tj | ****** | ******AA* | ****** |
| Sr | ****** | ******GAA* | ****** |
| Sn | ****** | ******AA* | ****** |
| Tb | ****** | ******AA* | ****** |
| To | ****** | ******AA* | ****** |
| Uc | ****** | ******AA* | ****** |
| Us | ****** | ******AA* | ****** |

| Species | Conserved sequence blocks | CSB- 1 | CSB- 2 | CSB- 3 |
|---------|---------------------------|--------|--------|--------|
| Reference sequence | CATATTAGGATATCAAGAGCATAA | GTAAAACCCCCCTACCCCA | TCCTGAAAAACCCCCCAAAG |
| Ca | ****** | ****** | ****** |
| Cb | ****** | ****** | ****** |
| Cc | ****** | ****** | ****** |
| Cd | ****** | ****** | ****** |
| Ce | ****** | ****** | ****** |
| Cm | ****** | ****** | ****** |
| Cp | ****** | ****** | ****** |
| Cj | ****** | ****** | ****** |
| Ck | ****** | ****** | ****** |
| Cl | ****** | ****** | ****** |
| Cn | ****** | ****** | ****** |
| Ct | ****** | ****** | ****** |
| Cme | ****** | ****** | ****** |

Continued
| Species | Conserved sequence blocks | CSB- 1 | CSB- 2 | CSB- 3 |
|---------|---------------------------|--------|--------|--------|
| Dmac    |                           |        |        |        |
| Dm      |                           |        |        |        |
| Dmar    |                           |        |        |        |
| Dt      |                           |        |        |        |
| Eb      | ****CT**********         |        |        |        |
| Gs      | ""C"***AT**********      |        |        |        |
| Mc      | ****TA**********         |        |        |        |
| Pn      | ****AT**********         |        |        |        |
| Pd      | ""C"***G**********       |        |        |        |
| Tt      |                           |        |        |        |
| Tj      |                           |        |        |        |
| Sc      | ****CGT**********        |        |        |        |
| Sd      | ****G**********          |        |        |        |
| Si      |                           |        |        |        |
| Sq      | ****A**********          |        |        |        |
| Sle     | ""C"***CT**********      |        |        |        |
| Sr      | ****A**********          |        |        |        |
| Sn      | ****AGA**********        |        |        |        |
| Tb      | ****AT**********         |        |        |        |
| To      | ****AT**********         |        |        |        |
| Tc      | G"C"***ATA**********     |        |        |        |
| Uh      |                           |        |        |        |
| Us      |                           |        |        |        |
| Species | 3' end terminal sequence |        |        |        |
| Reference sequence | TAT TATAAT TTT CACT |        |        |        |
| Dr      | ****C**********          |        |        |        |
| Ac      |                           |        |        |        |
| Ai      |                           |        |        |        |
| Ad      |                           |        |        |        |
| Ak      |                           |        |        |        |
| Am      |                           |        |        |        |
| Ca      |                           |        |        |        |
| Ch      |                           |        |        |        |
| Ce      |                           |        |        |        |
| Cp      |                           |        |        |        |
| Cm      |                           |        |        |        |
| Dmac    | ****C**********          |        |        |        |
| Dm      | ****C**********          |        |        |        |
| Dmar    | ****C**********          |        |        |        |
| Dt      | ****C**********          |        |        |        |
| Eb      | ****A**********          |        |        |        |
| Gs      |                           |        |        |        |
| Mc      |                           |        |        |        |
| Pn      |                           |        |        |        |
| Pd      | absent                    |        |        |        |
| Ci      |                           |        |        |        |
| Cj      |                           |        |        |        |
| Cme     |                           |        |        |        |
| Ti      |                           |        |        |        |
| Tj      |                           |        |        |        |
| Sc      | ****C**********          |        |        |        |
| Sd      | Absent                    |        |        |        |
| Sl      | Absent                    |        |        |        |
| Sq      | Absent                    |        |        |        |
| Sle     | Absent                    |        |        |        |
| Continued |                  |        |        |        |
The phylogenetic position of *D. russelli* is shown in Fig. 3 relative to 36 species of carangids based on 1380 bp sequences. Analogous topologies were obtained for both maximum likelihood analysis (ML) and Bayesian inference (BI) with almost 100% bootstrap support and strong posterior probability values, respectively (Fig. 3). The result implied that 37 species formed 3 lineages corresponding to the subfamilies Caranginae, Nauicratinae, and Trachinotinae. The reconstructed phylogenetic tree places *D. russelli* in the same branch with other *Decapterus* spp. and forms a sister group to the genus *Trachurus* and the findings support the previous classification of carangids14,77. The phylogenetic tree recovered the three carangid subfamilies as monophyletic and it also supported the groupings (*Trachinotinae + (Naucratinae + Caranginae)*). This result confirmed the phylogenetic relationship of carangids described by previous findings based on morphological synapomorphic characters31, and partial sequences of mitochondrial genes32,75. Within the Nauicratinae, the monotypic *E. bipinnulata* branched individually early from *Seriolina* and *Seriola* spp. After diverging from *Seriolina*, all *Seriola* species formed a monophyletic clade. Furthermore, both phylogenetic trees indicated two well-separated lineages in Caranginae. The first lineage included the genus *Salar*, *Trachurus*, and *Decapterus*. This lineage also encompassed the species *C. equula* on the same branch with *P. dentex* in contrast to other *Carangoides* spp. which were grouped into a different branch (Fig. 3). This result confirms the findings of previous studies14,77 and thus provides compelling evidence for the assignment of *C. equula* to its previously described genus *Kaiwarinus*. The second lineage was the most taxonomically diverse group of the family Carangidae delineated in the present study as it consisted of 10 genera divided into two larger groups. The first lineage comprised the genera: *Caranx* + *Megalasps* + *Gnathodon* as well as *Atule* + *Alepes*. The latter also included the monotypic *S. leptolepis* which individually branched off from this lineage early. The second lineage included *Uraspis* + *Parastromateus* and *Alectis* + *Carangoides*. In addition, a paraphyletic relationship between *Caranx* + *M. cordyla*, *Alepes* + *A. mate*, and *Uraspis* + *P. niger* was also observed.

**Divergence time estimation.** The RelTime ML analysis revealed that the differentiation of *D. russelli* could be dated back to about 4.69 Mya in the early Pliocene. The family Carangidae had diverged from other lineages when the Cretaceous–Paleogene extinction event occurred and it was estimated to be around 79 Mya towards the end of the Cretaceous (Supplementary Fig. S10). It was observed that the separation of the two groups, *Trachinotinae* and (*Caranginae + Nauicratinae*) occurred at 74 Mya at the end of the Cretaceous. Further, the split between the subfamilies *Caranginae* and Nauicratinae occurred at 62.20 Mya (Early Paleocene). Although the diversification of most species after the Cretaceous mass extinction occurred in the Paleogene7, several more recent Neogene radiations can also be observed (Fig. 4). The most recent split was observed at 0.34 Mya between the two *Uraspis* species. Our analyzes of the diversification of the genus *Trachurus* (18.40 Mya, early Miocene) were consistent with previous studies based on the data of morphological characters and ecological aspects75 and cyt b gene sequence analysis78. The split between *Seriola* + *Seriolina* and *Elgatis* occurred during the early Paleocene, consistent with Swart et al.46 but much younger than Le et al.44. The radiation of *Seriola* appears to be 20.87 Mya which is in direct agreement with Le et al.44, but slightly predates the estimates by Santini and Carnevale38.

**Conclusion**

The present investigation focused on the characterization of the complete mitochondrial genome of the Indian Scad, *D. russelli*, and a comprehensive comparative analysis of Carangidae with special emphasis on the structure and variability of tRNA and control region. The carangid tRNA was structurally different in both the stem and loop, with the exception of tRNA-Trp, and tRNA-Leu (TAG). These structural features resulted in mismatched or wobble base pairs in tRNAs. The high rate of Watson–Crick and Wobble base pairing in the stem regions of tRNA suggested their role in stem structure maintenance. The classical cloverleaf secondary structure was retained in all tRNAs except tRNA-Ser (GCT). The presence of an mTERF binding site and a very high evolutionary

| Species | 3' end terminal sequence |
|---------|--------------------------|
| Sr      | Absent                   |
| Sn      | Absent                   |
| Tb      | ***************G*T**     |
| To      | ***************T**       |
| Tc      | ***************G*T**     |
| Uh      | ****************         |
| Us      | ****************         |

**Table 4.** The comparison of Central conserved blocks, Conserved sequence blocks and 3’ terminal region in 37 species of Carangidae. *Dr* = *D. russelli*, *Ac* = *A. ciliaris*, *Ai* = *A. indica*, *Ad* = *A. dybowska*, *Ak* = *A. kleinii*, *Am* = *A. mate*, *Cc* = *C. carpio*, *Cb* = *C. bairdi*, *Ce* = *C. equula*, *Cn* = *C. malabaricus*, *Cp* = *C. plagiotrema*, *Ci* = *C. ignobilis*, *Ct* = *C. tille*, *Cm* = *C. melanopterus*, *Dm* = *D. macrourus*, *Dn* = *D. marui*, *Dt* = *D. tabl*, *Eb* = *E. bipinnulata*, *Gs* = *G. specious*, *Mc* = *M. cordyla*, *Pn* = *P. niger*, *Pd* = *P. dentex*, *Tt* = *T. trachurus*, *Tp* = *T. japonicas*, *Sc* = *S. crumenophthalmus*, *Sd* = *S. dumerilii*, *Sl* = *S. lalandi*, *Sq* = *S. quinquerguadiata*, *Sle* = *S. leptolepis*, *Sr* = *S. rivoliana*, *Sn* = *S. nigrofasciata*, *Tb* = *T. blochii*, *To* = *T. ovatus*, *Tc* = *T. carolinus*, *Uh* = *U. helvola*, *Us* = *U. secunda*. Asterisks indicate the same nucleotide as the reference sequence. Insertions are represented as bold and deletions as underlined. Other nucleotides denote base substitutions.

Phylogenetic analysis. The phylogenetic position of *D. russelli* is shown in Fig. 3 relative to 36 species of carangids based on 1380 bp sequences. Analogous topologies were obtained for both maximum likelihood analysis (ML) and Bayesian inference (BI) with almost 100% bootstrap support and strong posterior probability values, respectively (Fig. 3). The result implied that 37 species formed 3 lineages corresponding to the subfamilies Caranginae, Nauicratinae, and Trachinotinae. The reconstructed phylogenetic tree places *D. russelli* in the same branch with other *Decapterus* spp. and forms a sister group to the genus *Trachurus* and the findings support the previous classification of carangids14,77. The phylogenetic tree recovered the three carangid subfamilies as monophyletic and it also supported the groupings (*Trachinotinae + (Naucratinae + Caranginae*)). This result confirmed the phylogenetic relationship of carangids described by previous findings based on morphological synapomorphic characters31, and partial sequences of mitochondrial genes32,75. Within the Nauicratinae, the monotypic *E. bipinnulata* branched individually early from *Seriolina* and *Seriola* spp. After diverging from *Seriolina*, all *Seriola* species formed a monophyletic clade. Furthermore, both phylogenetic trees indicated two well-separated lineages in Caranginae. The first lineage included the genus *Salar*, *Trachurus*, and *Decapterus*. This lineage also encompassed the species *C. equula* on the same branch with *P. dentex* in contrast to other *Carangoides* spp. which were grouped into a different branch (Fig. 3). This result confirms the findings of previous studies14,77 and thus provides compelling evidence for the assignment of *C. equula* to its previously described genus *Kaiwarinus*. The second lineage was the most taxonomically diverse group of the family Carangidae delineated in the present study as it consisted of 10 genera divided into two larger groups. The first lineage comprised the genera: *Caranx* + *Megalasps* + *Gnathodon* as well as *Atule* + *Alepes*. The latter also included the monotypic *S. leptolepis* which individually branched off from this lineage early. The second lineage included *Uraspis* + *Parastromateus* and *Alectis* + *Carangoides*. In addition, a paraphyletic relationship between *Caranx* + *M. cordyla*, *Alepes* + *A. mate*, and *Uraspis* + *P. niger* was also observed.

Divergence time estimation. The RelTime ML analysis revealed that the differentiation of *D. russelli* could be dated back to about 4.69 Mya in the early Pliocene. The family Carangidae had diverged from other lineages when the Cretaceous–Paleogene extinction event occurred and it was estimated to be around 79 Mya towards the end of the Cretaceous (Supplementary Fig. S10). It was observed that the separation of the two groups, *Trachinotinae* and (*Caranginae + Nauicratinae*) occurred at 74 Mya at the end of the Cretaceous. Further, the split between the subfamilies *Caranginae* and Nauicratinae occurred at 62.20 Mya (Early Paleocene). Although the diversification of most species after the Cretaceous mass extinction occurred in the Paleogene7, several more recent Neogene radiations can also be observed (Fig. 4). The most recent split was observed at 0.34 Mya between the two *Uraspis* species. Our analyzes of the diversification of the genus *Trachurus* (18.40 Mya, early Miocene) were consistent with previous studies based on the data of morphological characters and ecological aspects76 and cyt b gene sequence analysis78. The split between *Seriola* + *Seriolina* and *Elgatis* occurred during the early Paleocene, consistent with Swart et al.46 but much younger than Le et al.44. The radiation of *Seriola* appears to be 20.87 Mya which is in direct agreement with Le et al.44, but slightly predates the estimates by Santini and Carnevale38.
conservation of its sequence suggest the functional significance of these sites. The high rate of sequence divergence in the flanking regions of conserved sequence motifs indicated the absence of typical mitochondrial control region coding constraints. The presence of evolutionarily conserved secondary structures of the conserved sequence motifs despite sequence divergence revealed the effect of selective constraints on their maintenance. The length variations in protein-coding, tRNA, and rRNA genes among Carangidae were negligible.

The phylogeny inferred from 13 gene concatenated data set resolved the monophyletic lineage of each subfamily and robustly supported the grouping (Trachinotinae + (Caranginae + Naucratinae)) with the caveat that the subfamily Scomberoidinae could not be included as the mitogenome sequences were not available. Further, the speciation of the genus Decapterus occurred before the separation of Tethys Sea and D. russelli diverged before the mid-Pliocene warm period. Future studies involving the mitogenomic data of Scomberoidinae are required for a better understanding of the phylogenetic and evolutionary relationship between major groups.
within the Carangidae. Mitogenomic phylogeny is also influenced by the habitat characteristics of each group, as habitat-specific selection and adaptation in mitochondrial coding genes has recently been reported\(^9\). The subfamilies Trachinotinae, Naucriatae, and Caranginae consist of species with differential habitat preferences (pelagic, benthopelagic, and reef-associated) and schooling behavior. The mutational patterns in the protein-coding regions of these fish need to be further analyzed for adaptive variation and selection and correlated with habitat characteristics to understand the forces of selection, if any.

**Materials and methods**

**Sampling.** The wild specimen of *D. russelli* was collected from Kochi, Kerala (9°56′19″N 76°15′45″E) from trawl and purse seines. The species was confirmed based on morphological characters. Tissue samples were preserved in 95% ethanol for DNA extraction and further analysis. The fish sample used in this study was handled according to the guidelines for the care and use of fish in research by De Tolla et al.\(^80\). The protocols were approved by the ethical committee of the ICAR- Central Marine Research Institute, Kochi. These methods are also reported following ARRIVE guidelines (http://arriveguidelines.org).

**DNA extraction and PCR amplification.** Genomic DNA was isolated as previously described\(^94\). The extracted DNA was then eluted in T.E buffer (1X) and stored at −20 °C for later use as a PCR template. The quality and quantity (concentration) of the isolated total DNA was checked by gel electrophoresis and NanoDrop One Spectrophotometer (Thermo Scientific, USA), respectively. The complete mitochondrial genome of *D. russelli* was determined by combining the sequences of contiguous and overlapping fragments amplified by PCR using both universal as well as custom-designed primers. Initial amplifications were performed using mitochondrial universal primer sequences (FishF1 + FishR1, 16SaRL + 16SbRH, 12S1 + 12S2, Dloop-Thr-F + Dloop-Phe-R, ATP 8.2 L8831 + CO3.2 H9236)\(^85–86\). The sequenced products of this reaction then served as a template for subsequent amplification. Fifteen sets of new primer pairs were designed based on the aligned complete mitogenome sequences of 36 species of carangids using Primer 3 software (http://frodo.wi.mit.edu/primer3)\(^87\). The above primers were screened for potential secondary targets using Primer Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast)\(^88\). PCR amplifications were performed in a Biorad T100 thermal cycler (Bio-Rad, USA). The reaction mixture included 15.6 µl sterile deionized water, 2 µl buffer (10X), 0.4 µl dNTP mix (10 mM each), 0.4 µl of each oligonucleotide (10 µM), 1 µl of template DNA (50 ng/µl) and 0.2 µl (1 unit) of Taq DNA polymerase with a total reaction volume of 20 µl. For confirmation, PCR products were electrophoresed on 1.5% agarose gel (1X TBE) and visualized using a gel documentation system (Bio-Rad, USA). The products were then sequenced completely in both directions using sanger sequencing.

**Sequence analysis.** Target sequences were assembled with the software MEGA X\(^89\). The identity of the *D. russelli* mitogenome was verified with other carangid mitogenomes (Supplementary Table S2) by BLAST search. These mitogenome sequences served as the reference for assembly and manual annotations. The organization and the order of the genes were initially determined using the MitoFish database\(^80\) and later verified manually by comparison with the reference sequences. The 13 PCGs and their codon usage were summarized with MEGA X\(^57\). The products were then sequenced by comparison with the reference sequences. The 13 PCGs and their codon usage were summarized with MEGA X\(^89\). The rate of Ka/Ks substitutions in the mitogenome of Carangidae was calculated by DnaSP 6.12.03\(^91\). The secondary structure of conserved sequence motifs of the control region was inferred using Mfold web server\(^92\). The complete mitogenome map of the species was graphically visualized by OrganellarGenomeDRAW \(^93\). The 37 mitogenomes of Carangidae were compared by using CG view comparison tool (CCT)\(^94\) with *D. russelli* as the reference sequence. Genes were assigned by Clusters of Orthologous Groups (COG) using the CG view comparison tool (CCT) and BLAST was used to align other genomes to that of *D. russelli*.

**Phylogenetic analysis.** All available mitogenome sequences of other carangid species were retrieved from GenBank to find out the relative position of *D. russelli* among carangids and to delineate its phylogenetic relationship. *L. crocea* (GenBank accession number NC011710) was taken as an outgroup. The 13 PCGs were aligned using Clustal W and translated into amino acid sequences in MEGA X\(^80\). For the subsequent phylogenetic analysis, multiple alignments were generated for the 13 concatenated protein-coding sequences of *D. russelli* and 36 other Carangidae species. The phylogenetic relationship was inferred by constructing a tree based on Bayesian inference (BI) and maximum likelihood (ML) methods. The best model, GTR + G + I, was selected using Akaike Information Criterion (AIC). Bayesian phylogenetic analysis was performed using MrBayes\(^96\) with the Markov Chain Monte Carlo Method (MCMC). MCMC chains were run for 10 million generations and obtained a standard deviation of split frequency less than 0.01. The ML phylogenetic tree was inferred using RaxML\(^97\) and a consensus tree was obtained with 1000 bootstrap replicates using default parameters. The final phylogenetic tree was displayed, annotated, and embellished in FigTree software (http://tree.bio.ed.ac.uk/software/figtree/)\(^98\).

**Divergence time estimation.** Molecular dating of 37 Carangidae species was performed with the RelTimeML method in MEGA X\(^99\) using GTR + G + I modeling. The phylogenetic tree inferred from Bayesian analysis was imported to MEGA X to estimate divergence times. The time tree web resource (http://www.timetree.org)\(^100\) was used to infer calibration boundaries for a pair of taxa (minimum and maximum) and the same database was used to establish the time tree of 27 genera of carangids.
Figure 4. Timetree chronogram of 37 species of Carangidae for the 13 PCGs inferred from Bayesian topology. Divergence times were estimated using 32 calibrations. The red text indicates the species sequenced in this study. The Timetree was generated by using MEGA X®.
Data availability

All data generated or analyzed during this study is included in this published article (and its Supplementary information files).

Received: 17 December 2021; Accepted: 23 March 2022

Published online: 04 April 2022

References

1. Lin, P. L. & Shao, K. T. A review of the carangid fishes (Family Carangidae) from Taiwan with descriptions of four new records. Zool. Stud. 38, 33–68 (1999).

2. Honebrink, R. A review of the biology of the family Carangidae, with emphasis on species found in Hawaiian waters. Division of Aquatic Resources Technical report 20–01, Department of Land and Natural Resources: Honolulu, HI, USA. (2000).

3. Laroche, W. A., Smith-Vaniz, W. F. & Richardson, S. L. Carangidae: Development in Ontogeny and Systematics of Fishes (ed. Moser, H.G. et al.) 510–522 (American Society of Ichthyologists and Herpetologists Special Publications, 1984).

4. Gushiken, S. Phylogenetic relationships of the perciform genera of the family Carangidae. Ichthyol. Res. 34, 443–461 (1988).

5. Froese, R. & Pauly, D. FishBase. World Wide Web electronic publication. http://www.fishbase.org, (2021).

6. Rohlfritsch, A. & Borsa, P. Genetic structure of Indian scad mackerel Decapterus russelli: Pleistocene vicariance and secondary contact in the Central Indo-West Pacific Seas. Heredity 95, 315–326 (2005).

7. Balasubramanian, N. K. & Natarajan, P. Studies on the biology of the scads, Decapterus russelli and Decapterus macrostomus off Vizhinjam, southwest coast of India. Indian. J. Fish. 47, 291–300 (2000).

8. Murty, V. S. Observations on some aspects of biology and population dynamics of the scad Decapterus russelli (Ruppell) (Carangidae) in the trawling grounds off Kakinada. J. Mar. Biol. Assoc. India. 33, 396–408 (1991).

9. Raje, S. G. On some aspects of biology of the mackerel scad Decapterus russelli (Ruppell), Indian. J. Fish. 44, 97–99 (1997).

10. Jaiswar, A. K., George, I. P., Gulati, D. K. & Swamy, R. P. A study on length-weight relationship, food and feeding habits of Indian scad, Decapterus russelli (Ruppell, 1830) along the northwest coast of India. J. Indian Fish. Assoc. 23, 1–6 (1993).

11. Poojary, N., Tiwari, L. R. & Jaiswar, A. K. Food and feeding habits of the Indian scad, Decapterus russelli (Ruppell, 1830) from Mumbai waters, north-west coast of India. Indian. J. Fish. 57, 93–99 (2010).

12. Sunil, V. & Suryanarayan, H. The reproductive biology of Ambassis gymnocephalus and Decapterus russelli in the Neendakara Zone, Kerala. J. Anim. Morphol. Physiol. 41, 119–123 (1994).

13. Moiseyeva, Y. B. & Zhuk, N. N. Oogenesis in some species of carangids in the west Indian Ocean. J. Ichthyol. 35, 137–149 (1995).

14. Manojkumar, P. P. Maturation and spawning of Decapterus russelli (Ruppell, 1830) along the Malabar Coast. Indian. J. Fish. 52, 171–178 (2005).

15. Kalhoro, M. T. et al. Stock assessment of Indian Scad, Decapterus russelli in Pakistani marine waters and its impact on the national economy. Fish Aqua. J. 8, 11–11 (2017).

16. Reuben, S. et al. Fishery, biology and stock assessment of carangid resources from the Indian seas. Indian. J. Fish. 39, 195–234 (1992).

17. Manojkumar, P. P. Stock assessment of Indian scad, Decapterus russelli (Ruppell, 1830) off Malabar. J. Mar. Biol. Assoc. India. 49, 79–80 (2007).

18. Sen, S. et al. Stock structure analysis of Decapterus russelli (Ruppell, 1830) from east and west coast of India using truss network analysis. Fish. Res. 112, 38–43 (2011).

19. Boore, J. L. Animal mitochondrial genomes. Nucl. Acids Res. 27, 1767–1780 (1999).

20. Jacobsen, M. W., Da Fonseca, R. R., Bernatchez, L. & Hansen, M. M. Comparative analysis of complete mitochondrial genomes suggests that relaxed purifying selection is driving high nonsynonymous evolutionary rate of the NADH2 gene in whitefish (Coregonus sp.). Mol. Phylogenet. Evol. 95, 161–170 (2016).

21. Brown, W. M., Prager, E. M., Wang, A. & Wilson, A. C. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. 18, 225–239 (1982).

22. Saccone, C., Pesole, G. & Shibata, T. The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. J. Mol. Evol. 33, 83–91 (1991).

23. Foran, D. R., Hixon, J. E. & Brown, W. M. Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. Nucleic Acids Res. 16, 5841–5861 (1988).

24. Sebastian, W., Sukumaran, S. & Gopalakrishnan, A. The signals of selective constraints on the mitochondrial non-coding control region: insights from comparative mitogenomes of Clupeid fishes. Genetica 149, 1–11 (2021).

25. Zhu, S. H., Zheng, W. J., Zou, J. X., Yang, Y. C. & Shen, X. Q. Mitochondrial DNA control region structure and phylogenetic relationship of Carangidae. Zool. Res. 28, 606–614 (2007).

26. Suzuki, K. Anatomical and taxonomical studies on the carangid fishes of Japan. Rep. Fac. Fish. Prefect. Un. Mie. 19, 225–232 (1962).

27. Smith-Vaniz, W. F. & Slaight, J. C. Comparative revision of Scomberoides, Oligopilates, Parona, and Hypacinus with comments on the phylogenetic position of Campogammarus (Pisces: Carangidae). Proc. Cal. Acad. Sci. 39, 185–256 (1973).

28. Smith-Vaniz, W. F. Carangidae in Sea Fishes (ed. Smith, M. M & Heemstra, P. C.) 638–660 (Springer, 1986).

29. Springer, V. G. & Smith-Vaniz, W. F. Supranuclear and pterygiophore insertion patterns in carangid fishes, with description of a new Eocene carangid tribe, Paratracchinotini, and a survey of anterior anal-fin pterygiophore insertion patterns in Acanthocephali. J. Mol. Evol. 16, 1–73 (2008).

30. Hilton, E. J., Johnson, G. D. & Smith-Vaniz, W. F. Ontology and Systematics of Parastromateus niger (Perciformes: Carangidae), with Comments on the Carangid Dorsal Gill-Arch Skeleton. Copeia 2, 312–333 (2010).

31. Smith-Vaniz, W. F. Carangidae: relationships in Ontogeny and Systematics of Fishes (ed. Moser, H.G. et al.) 522–530 (American Society of Ichthyologists and Herpetologists Special Publications, 1984).

32. Reed, D. L., Carpenter, K. E. & Martin, J. D. Molecular systematics of the Jacks (Perciformes: Carangidae) based on mitochondrial cytochrome b sequences using parsimony, likelihood, and Bayesian approaches. Mol. Phylogenet. Evol. 23, 513–524 (2002).

33. Reed, D. L., deGravelle, M. J. & Carpenter, K. E. Molecular systematics of Selene (Perciformes: Carangidae) based on cytochrome b sequences. Mol. Phylogenet. Evol. 21, 468–475 (2001).

34. Cárdenas, L. et al. Origin, diversification, and historical biogeography of the genus Trachurus 385 (Perciformes: Carangidae). Mol. Phylogenet. Evol. 35, 496–507 (2005).

35. Betancur-R, R., Li, C., Munroe, T. A., Ballesteros, J. A. & Ortí, G. Addressing gene tree discordance and non-stationarity to resolve a multi-locus phylogeny of the flatfishes (Teleostei: Pleuronectiformes). Syst. Biol. 62, 763–785 (2013).

36. Near, T. J. et al. Phylogeny and tempo of diversification in the superfamilies of spinning-rayed fishes. Proc. Natl. Acad. Sci. 110, 12738–12743 (2013).

37. Rabosky, D. L. et al. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nat. Commun. 4, 1–8 (2013).
38. Santini, F. & Carnevale, G. First multilocus and densely sampled timetree of trevallies, pompanos and allies (Carangoidei, Perciformes) suggests a Cretaceous origin and Eocene radiation of a major clade of piscivores. *Mol. Phylogenet. Evol.* **83**, 33–39 (2015).
39. Britz, R. & Johnson, G. D. Ontogeny and homology of the skeletal elements that form the sucking disc of remoras (*Teleostei, Echeneoioidi, Echeneidae*). *J. Morphol.** 273, 1353–1366 (2012).
40. Swart, B. L., von der Heyden, S., Bester-van der Merwe, A. & Roets-Wilding, R. Molecular systematics and biogeography of the circumglobal distributed genus Seriola (*Pisces: Carangidae*). *Mol. Phylogenet. Evol.* **93**, 274–280 (2015).
41. Bannikov, A. F. On the taxonomy, composition and origin of the Family Carangidae. *J. Ichthyol.* **27**, 1–8 (1987).
42. Kijima, A., Taniguchi, N. & Ochiai, A. Genetic divergence and relationship among fifteen species of genera Trachurus, Decapterus Selar and Salaroides. *Ipn. J. Ichthyol.* **35**, 167–175 (1988).
43. Nichols, J. T. Notes on carangin fishes. 6. East Indian mackerel scads (Decapterus) described and differentiated. *Am. Mus. Novit.* **1196**, 1–8 (1942).
44. Li, Z. et al. Complete mitogenomes of three carangidae (perciformes) fishes: genome description and phylogenetic considerations. *Int. J. Mol. Sci.* **21**, 4685 (2020).
45. Li, M. et al. Whole mitogenome of the Japanese scad *Decapterus marsaudis* (Perciformes: Carangidae). *Mitochondrial DNA Part A* **27**, 306–307 (2016).
46. Li, M. et al. Characterization of the mitochondrial genome of the Shortfin mako shark *Isurus paucus* (Perciformes: Carangidae). *Mitochondrial DNA Part A* **27**, 82–83 (2016).
47. Zou, K., Chen, Z., Zhang, P. & Li, M. Mitochondrial genome of the Mackerel scad *Decapterus macarellus* (Perciformes: Carangidae). *Mitochondrial DNA Part A* **27**, 2151–2152 (2016).
48. Norfatmah, M. Y., Teh, L. K., Salleh, M. Z., Isa, M. M. & SitiAzzah, M. N. Complete mitochondrial genome of Malaysian Mahseer (*Tor tambroides*). *Gene* **548**, 263–269 (2014).
49. Xie, X. Y., Huang, G. F., Li, Y. T., Zhang, Y. T. & Chen, S. X. Complete mitochondrial genome of *Acrosalocheilus parallens* (Cypriniformes, Barbinae). *Mitochondrial DNA Part A* **27**, 3339–3340 (2016).
50. Nardi, F., Carapelli, A., Fanciulli, P. P., Dallai, R. & Frati, F. The complete mitochondrial DNA sequence of the basal hexapod *Tetradontothorpa bilanensis*: evidence for heteroplasy and tRNA translocations. *Mol. Biol. Evol.* **18**, 1293–1304 (2001).
51. Wang, X. et al. Complete mitochondrial genome of jaweline goby (*Synchogobius husta*). *Mitochondrial DNA Part A* **27**, 4586–4587 (2016).
52. Hurst, L. D. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* **18**, 486–486 (2002).
53. Shi, X., Tian, P., Lin, R., Huang, D. & Wang, J. Characterization of the complete mitochondrial genome structure of the gobiose head whiptail *Gobiosoma globiceps* (Gadiformes: Macruridae) and its phylogenetic analysis. *PLoS ONE* **11**, e0153666. https://doi.org/10.1371/journal.pone.0153666 (2016).
54. Yang, H. et al. Characterization of the complete mitochondrial genome sequences of three croakers (*perciformes, sciaenidae*) and novel insights into the phylogenetics. *Int. J. Mol. Sci.* **19**, 1741 (2018).
55. Zhuang, X., Qu, M., Zhang, X. & Ding, S. A comprehensive description and evolutional analysis of 22 groupers (*Perciformes, Epinephelidae*) mitochondrial genomes with emphasis on two novel genome organizations. *PLoS ONE* **8**, e73561. https://doi.org/10.1371/journal.pone.0073561 (2013).
56. Stepień, C. A. & Kocher, T. D. Molecules and morphology in studies of fish evolution in *Molecular systematics of fishes* (ed. Kocher, T. D. & Stepień, C. A.) 1–19 (Academic press, 1997).
57. Laslett, D. & Canébäck, B. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* **24**, 172–175 (2008).
58. Janke, A. & Pääbo, S. Editing of a tRNA anticodon in marsupial mitochondria changes its codon recognition. *Nucleic acids Res.* **21**, 1523–1525 (1993).
59. Zhu, K. C. et al. Sequencing and characterization of the complete mitochondrial genome of Japanese Swellshark (*Cephaloscyllium umbratile*). *Sci. Rep.* **7**, 1–14 (2017).
60. Zardoya, R. & Meyer, A. The complete DNA sequence of the mitochondrial genome of a "living fossil", the coelacanth (*Latimeria chalumnae*). *Trends Genet.* **22**, 595–600 (2006).
61. Clayton, D. A. A tridecamer DNA sequence supports human mitochondrial RNA 3'-end formation in vitro. *Mol. Cell. Biol.* **18**, 4502–4509 (1988).
62. Broughton, R. E., Milam, J. E. & Roe, B. A. The complete sequence of the zebrafish (*Danio rerio*) mitochondrial DNA and evolutionary patterns in vertebrate mitochondrial DNA. *Genome Res.* **11**, 1958–1967 (2001).
63. Lee, W. J., Conroy, J., Howell, W. H. & Kocher, T. D. Structure and evolution of teleost mitochondrial control regions. *J. Mol. Evol.* **41**, 54–66 (1995).
64. Clayton, D. A. Replication of animal mitochondrial DNA. *Cell* **28**, 693–705 (1982).
65. Koonin, E. V. The Clusters of Orthologous Groups (COGs) Database: phylogenetic classification of proteins from complete genomes in *The NCBI handbook* (ed. McEntyre, J & Ostell, J) 21–27 (The National Library of Medicine Press, 2010).
66. Natale, D. A. et al. Towards understanding the first genome sequence of a crenarchaeon by genome annotation using clusters of orthologous groups of proteins (COGs). *Genome Biol.* **1**, 693–705 (1998).
67. Remm, M., Storm, C. E. & Sonnhammer, E. L. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**, 1041–1052 (2001).
68. Zinsmeister, W. J. Discovery of fish mortality horizon at the KT boundary on Seymour Island: re-evaluation of events at the end of the Cretaceous. *J. Paleontol.* **72**, 556–571 (1998).
69. Li, B., Wang, H., Yang, L., Liu, S. & Zhuang, Z. Complete mitochondrial genome of pseudocaranx dentex (*Perciformes, Perco- formes*) provides insight into phylogenetic and evolutionary relationship among carangidae family. *Genes* **12**, 1234 (2021).
70. Shabonevey, I. Y. Systematics, morpho-ecological characteristics and origin of Carangids of the genus Trachurus. *J. Ichthyol.* **20**, 15–24 (1981).
Acknowledgements
We would like to thank Dr. P. Vijayagopal, Head in charge, Marine Biotechnology Division (MBTD) and Director, ICAR- Central Marine Fisheries Research Institute (CMFRI), Kochi for providing the necessary facilities for conducting the research. The first author is grateful to the Council of Scientific and Industrial Research (CSIR) for the financial assistance in the form of a research fellowship.

Author contributions
A.J. performed the experiments, analyzed and interpreted the data and prepared the main manuscript text. S.S. conceptualized and coordinated the study and reviewed and edited the manuscript. L.P.M. and S.M. participated in the data analysis. N.R. and N.K. prepared Fig. 1. A.G. supervised the study.

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/s41598-022-09636-5.

Correspondence and requests for materials should be addressed to A.J.

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) 2022