VOL.66 NO.02 JULY-DECEMNER 2024 • PRINT ISSN 0025-3146 • ONLINE ISSN 2321-7898

JMBAI

JOURNAL OF THE MARINE BIOLOGICAL ASSOCIATION OF INDIA

MBAI
Marine Biological Association of India
ESTD.1958

सी एम एफ आर आइ
CMFRI

Available online at: www.mbai.org.in doi:10.6024/jmbai.2024.66.2.2491-01

Mitogenome mapping of the Green Tiger Shrimp *Penaeus semisulcatus* De Haan, 1844 in Indian waters illuminates lineage diversity from the Indo-West Pacific

N. S. Jeena*, G. Tamilmani, M. Sakthivel, S. Rahuman, K. A. Sajeela, V. Vetrivel, K. K. Anikuttan, P. Rameshkumar, B. Johnson, R. Bavithra and A. Gopalakrishnan
ICAR-Central Marine Fisheries Research Institute, Kochi- 682018, Kerala, India.

*Correspondence e-mail: jeena.ns@icar.gov.in
ORCiD: https://orcid.org/0000-0002-8142-0474

Received: 03 May 2024 Revised: 04 Oct 2024
Accepted: 04 Oct 2024 Published: 19 Dec 2024

Original Article

Abstract

Penaeus semisulcatus is a key crustacean species with a patchy distribution along the Indian coast. It constitutes a monospecific fishery along the southeastern coast of India, underscoring its pivotal significance in the economic dynamics of the region. Currently, the species contains two morphotypes with banded and non-banded antennae, presumably representing two species. The former is widespread in the Indo-West Pacific while the latter appears endemic to the Persian Gulf. The present study focuses on the mitogenome characterization of the banded antennae morphotype of *P. semisulcatus* for the first time from Indian waters to bridge the knowledge gap in genomics in the Indo-West Pacific. The 16,004 base-pair length mitogenome was assembled using genome skimming. It contained 13 protein-coding genes (PCGs), two rRNA genes, 22 tRNA genes, and a noncoding control region, with a basic pan-crustacean gene content and arrangement pattern. The genetic divergence (K2P distance) values using 13 PCGs between the two morphotypes was 16.6% and between the Indian and Pacific mitochondrial lineages was 6.9%. The phylogenetic analysis also inferred the polyphyletic nature of *Penaeus sensu stricto* and the grouping of both morphotypes with high probability values. This research contributes to the understanding of the genetic diversity of *P. semisulcatus* and sheds light on its evolutionary processes. The availability of the complete mitogenome serves as a valuable resource for further comparative genomic studies and contributes to proper taxonomic identification, sustainable management, and conservation efforts of this commercially important species.

Keywords: *Green tiger shrimp, genome skimming, morphotype, phylogeny.*

Introduction

Penaeus semisulcatus De Haan, 1844 commonly called the green tiger shrimp or grooved tiger prawn is one of the commercially important penaeid shrimp widely distributed in the Indo-West Pacific from the Red Sea, east and southeast Africa to Japan, Korea, in the seas of the Indo-Malay Archipelago and northern Australia (Holthuis, 1980; Halim *et al.*, 2021). Along the coast of India, it is more common along the east than on the west (Kurian and Sebastian, 1976) and constitutes an important fishery along the coast of Tamil Nadu (Rajkumar *et al.*, 2023) which is traded as 'Mandapam flower shrimp' in the export market. It accounts for 4.68% of the total Indian marine penaeid landings, 58% of shrimp landings from Tamil Nadu with more than 70% of the total shrimp landings in Ramanathapuram District that forms a major livelihood fishery of the region (CMFRI, 2019). Green tiger shrimp is a promising species in capture fisheries and culture practices due to its large size (range from 41 mm to 240 mm), high demand, and premium market price. Hatchery production of seeds and larviculture of *P. semisulcatus* was developed in the late 1980s by the Mandapam Regional Centre of ICAR-CMFRI (CMFRI, 2009). Sea ranching of hatchery-produced seeds (post-larvae–PLs) is carried out at Mandapam Regional Centre to replenish the natural stock to support sustainable fishery of the species in the Gulf of Mannar and Palk Bay (CMFRI, 2022).

Genetic and morphometric investigations of *P. semisulcatus* from the Indo-Pacific region have been conducted at regional and broader scales. Two morphotypes, namely the banded

antennae type and the non-banded antennae type, have been described in the species which are considered to represent two sympatric species (Jahromi, 2011; Jahromi *et al.*, 2019), the latter of which is awaiting formal description. The banded antennae type is the widespread morphotype (Halim *et al.,* 2021). The other type is endemic to the Persian Gulf and the Sea of Oman, but the two morphotypes co-occur in this region (Jahromi *et al.,* 2021; Alizadeh *et al.,* 2022). Sharawy *et al.* (2016, 2017) carried out population characterization for the species using molecular markers like cytochrome c oxidase subunit 1 (COI), 16S ribosomal RNA (16S rRNA), and 18S rRNA in the Gulf of Suez and through the Suez Canal. Hassanien and Al-Rashada (2019) assessed genetic variation in the species using randomly amplified polymorphic DNA, and simple sequence repeat from two geographically distant locations in the Arabian and Suez Gulf. Another study was performed using microsatellite loci in the Persian Gulf and Oman Sea (Jahromi *et al.,* 2021). Assessment of genetic diversity and phylogeographic criteria in the northwest of the Red Sea using a hypervariable 5′ barcode area of the COI gene has been carried out by Mohammed-Geba and Yousif (2022). Genetic structure analysis of the species using COI, 16S rRNA, and mitochondrial control region (CR) disclosed two separate lineages originating from the Indian Ocean and the Western Pacific (Alam *et al.,* 2017). Also, two lineages were identified using COI and CR in the Indo-Pacific Ocean with a focus on the Indo-Malay Archipelago (Halim *et al.,* 2021).

In the past years, molecular studies using single gene markers have been replaced by complete mitogenome. Owing to its advantages including maternal inheritance, lack of recombination, high copy number, and high nucleotide base substitution, mitogenomes have been used extensively in crustacean research for species delimitation, classification, phylogenetics, and phylogeographic studies (Shen *et al.,* 2013; Gonzalez-Castellano *et al.,* 2020). Mitogenome data of the green tiger shrimp is available in GenBank from the Pacific (Accession No. MG821354; Zhong *et al.*, 2019) and the Persian Gulf (Accession No. LC493087, LC497027). Currently, no genetic studies on this species have been reported from Indian waters, and the present study aimed to map the mitogenome of its banded morphotype distributed along the Indian coast to bridge the genetic information gap existing in the Indo-West Pacific region.

Material and methods

A single specimen of the banded antennae morphotype of *P. semisulcatus* (Fig. 1) was obtained from the coastal regions of Madapam, Tamil Nadu, India (9° 23.930′ N, 79°16.696′ E). The specimen was morphologically identified (Sudhakara Rao *et al.,* 2013), and the tissue was preserved in 90% ethanol for subsequent molecular analysis.

Genomic DNA extraction followed the conventional phenol-chloroform method (Sambrook and Russell, 2006), and RNase A (Sigma-Aldrich, USA) was used to degrade any potential RNA contamination. Subsequently, to eliminate RNase A proteins, the genomic DNA was purified using the NucleoSpin gDNA Clean-up Kit (Macherey-Nagel, Germany). The purified DNA was quantified using a Qubit 3.0 Fluorometer with the dsDNA HS assay (ThermoFisher Scientific) and visualized on 2% agarose gels. Sequencing libraries were generated using the NEBNext Ultra DNA Library Prep Kit for Illumina (NEB, USA), following the manufacturer's instructions. Sequencing was carried out on the Illumina NovaSeq platform, employing 150 bp paired-end sequencing. We employed Low Coverage-Whole Genome Sequencing (LC-WGS) reads from whole-cell extraction that includes a high copy number of extranuclear sequences, making it an efficient and cost-effective approach for assembling complete mitochondrial genomes (Chak *et al.,* 2020).

Raw reads with a Phred score lower than 20 and adapters were removed using Trimmomatic v. 0.39 (Bolger *et al.,* 2014). Reference-based mitochondrial genome assembly of Green Tiger Shrimp was performed using NOVOPlasty 4.3.1 with 43 million PE reads (Dierckxsens *et al.,* 2017). The complete mitochondrion of *P. semisulcatus* with NCBI GenBank accession number LC493087 was used as a bait reference for the NOVOPlasty software. Genes were annotated using DOGMA and MITOS web servers, following the mitochondrial code for invertebrates (Bernt *et al.,* 2013; Wyman *et al.,* 2004). To refine the boundaries, the annotated genes were manually edited in Geneious (Kearse *et al.,* 2012) whereas tRNAs were annotated using Arwen v.1.2.3 (Laslett and Canbäck, 2008). The verification of ribosomal RNA (rRNA) was carried out using RNAmmer 1.2 (Lagesen *et al.,* 2007). Additionally, the non-coding control region (NCR) was examined through Tandem Repeats Finder (Benson, 1999).

Fig. 1. Banded antennae morphotype of *Penaeus semisulcatus* collected from Mandapam, Tamil Nadu.

Table 1. Genetic data for *P. semisulcatus*, highlighting its morphotypes and sampling locations

Morphotype identified	Location	GenBank Accession No. (COI)	Reference
Banded type	Hormoz, Iran	KC525192, KC525195	Jahromi *et al*., 2019
Banded type	Jask, Iran	KC525193, KC525194	
Banded type	Penang, Malaysia	KC525188– KC525191	
non-Banded type	Bushehr, Iran	KC525184, KC525185	
non-Banded type	Jask, Iran	KC525187	
non-Banded type	Hormoz, Iran	KC525186	
Banded type	GuangXi Province, China	MG821354	Zhong *et al*, 2019
non-Banded type	Iran	LC497027	NCBI
Banded type	Persian Gulf, Iran	LC493087	NCBI
Banded type	Mandapam, Tamil Nadu, India	OR906295	This study

The partial COI sequences of the banded and non-banded types of the species (Jahromi *et al.,* 2019), the COI region of the three mitogenomes available in the NCBI database, and the mitogenome generated in this study (Table 1) were used to construct a phylogenetic tree to validate the sequences with their morphological types. The genetic divergence for the COI gene was calculated according to the Kimura 2-Parameter (K2P) model of sequence evolution (Kimura, 1980) under uniform rates in MEGA X (Kumar *et al.,* 2018), and the Neighbour-joining (NJ) tree (Saitou and Nei, 1987) was constructed in the software. The same model was used to estimate divergence values for mitochondrial protein-coding genes (13PCGs: 11085 bp) of the mitogenomes from different waters. The Relative Synonymous Codon Usage (RSCU) of thirteen PCGs (excluding incomplete codons) was also computed using MEGA. The compositional skewness of each genome component was determined using the following formulas: AT-skew = (A-T)/(A+T); GC-skew = (G-C)/(G+C) (Perna and Kocher, 1995). Genome visualization was performed using CGView Server (Grant and Stothard, 2008).

The complete mitochondrial DNA genomes of *P. semisulcatus*, including 27 shrimp species within the genus *Penaeus sensu lato* (*s.l.*), were utilized for phylogenetic analysis. Phylogenetic trees were constructed using concatenated datasets of protein-coding and rRNA genes (13115 bp) with the application of both Maximum Likelihood (ML) and Bayesian Inference (BI) methods. The software tools IQ TREE v1. 6.12 and MrBayes 3.2.7 were employed for ML and BI, respectively incorporating the best substitution models and partition schemes for each dataset through PartitionFinder version 2.1.1 (Huelsenbeck and Ronquist, 2001; Nguyen *et al*., 2015; Lanfear *et al.,* 2017;). To assess the robustness of the ML tree topology, 1,000 bootstrap replications of the observed data were conducted. For the BI tree, two simultaneous runs were executed with four chains over 10 million generations, with tree sampling at intervals of 100 generations. The initial 25% of samples were discarded and not considered for summary statistic calculations. The maximum standard deviation of split frequencies was used to confirm the convergence of the runs with Tracer v1.7 (Rambaut *et al.,* 2018). All phylogenetic results were visualized in Fig Tree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Results and discussion

General features of the mitogenome

LC-WGS resulted in the generation of 7 Gb paired-end reads. The newly assembled mitochondrial genome of *P. semisulcatus* was 16004 base-pair (bp) in length and was submitted under GenBank accession number OR906295. It comprised 13 protein-coding genes (PCGs), two ribosomal RNA genes (rRNAs: *rrnS* and *rrnL),* 22 transfer RNA genes (tRNA), and a non-coding region (NCR). The PCGs range from 159 bp (*ATP8*) to 1724 bp *(ND5*) in size, with a total length of 11157 bp. tRNAs varied from 66 bp in tRNA-Tyr, tRNA-Leu, tRNA-Ala, tRNA-Asn and, tRNA-Pro to 72 bp in tRNA-Val, with a total length of 1493 bp. The length of the small encoding ribosomal subunit, *rrnS* (12S rRNA), and large subunit, *rrnL* (16S rRNA) were 856 bp and 1370 bp, respectively. Twenty-three of the thirty-seven genes were encoded by the plus strand, with the others encoded by the minus strand (Table 2).

In addition, a single long intergenic space of 997 bp was assumed to be the mitochondrial control region and was encoded in plus strand. Most of the PCGs and tRNA genes were encoded in heavy stand/plus stand while only four PCGs (*ND I, ND IV, ND IVL*, and *ND V*) and eight tRNA genes (tRNA-Gln, tRNA-Cys, tRNA-Tyr, tRNA-Phe, tRNA-His, tRNA-Pro, tRNA-Leu, and, tRNA-Val) were encoded in light strand/minus strand. Furthermore, both the ribosomal RNAs were encoded in the minus strand. The gene order observed in *P. semisulcatus* is similar to the typical pan-crustacean ground pattern (Lavrov

Table 2. Mitochondrial gene annotation of *P. semisulcatus*

Gene	Strand	Position		Size (bp)	Intergenic/ overlapping (nt)	Codon (Start/Stop)
		From	To			
tRNA-Ile		1	67	67	15	
tRNA-Gln	Minus	83	152	70	30	
tRNA-Met	Plus	183	251	69	0	
ND 2	Plus	252	1253	1002	-2	ATT/TAA
tRNA-Trp	Plus	1252	1320	69	15	
tRNA-Cys	Minus	1336	1403	68	1	
tRNA-Tyr	Minus	1405	1470	66	2	
COI	Plus	1473	3006	1534	0	ACG/T—
tRNA-Leu	Plus	3007	3072	66	5	
COII	Plus	3078	3765	688	0	ATG/T—
tRNA-Lys	Plus	3766	3835	70	3	
tRNA-Asp	Plus	3839	3907	69	0	
ATP 8	Plus	3908	4066	159	-7	ATC/TAA
ATP 6	Plus	4060	4734	675	14	ATG/TAA
COIII	Plus	4749	5538	790	0	ATG/T—
tRNA-Gly	Plus	5539	5605	67	0	
ND 3	Plus	5606	5957	352	0	ATG/T—
tRNA-Ala	Plus	5958	6023	66	2	
tRNA-Arg	Plus	6026	6090	65	0	
tRNA-Asn	Plus	6091	6156	66	4	
tRNA-Ser	Plus	6161	6227	67	0	
tRNA-Glu	Plus	6228	6297	70	19	
tRNA-Phe	Minus	6317	6384	68	-1	
ND5	Minus	6384	8107	1724	9	ATG/TA-
tRNA-His	Minus	8117	8183	67	0	
ND4	Minus	8184	9524	1341	-7	ATG/TAA
ND 4L	Minus	9518	9817	300	2	ATG/TAA
tRNA-Thr	Plus	9820	9887	68	0	
tRNA-Pro	Minus	9888	9953	66	1	
ND6	Plus	9955	10470	516	3	ATT/TAA
cytochrome b (cytb)	Plus	10474	11610	1137	-2	ATG/TAG
tRNA-Ser	Plus	11609	11678	70	17	
ND1	Minus	11696	12634	939	5	ATA/TAA
tRNA-Leu	Minus	12640	12706	67	0	
rrnL	Minus	12707	14076	1370	3	
tRNA-Val	Minus	14080	14151	72	0	
rrnS	Minus	14152	15007	856	0	
CR	Plus	15008	16004	997		

et al., 2000; Tan *et al.,* 2017). The circular representation of the mitogenome of *P. semisulcatus* is shown in Fig. 2.

Upon comparison of the partial sequences of the COI gene, it was observed that the mitogenome from this study (OR906295) and the Persian Gulf, Iran (LC493087) cluster with banded antennae morphotype of *P. semisulcatus* (Fig. 3). Conversely, the second mitogenome submitted from Iran

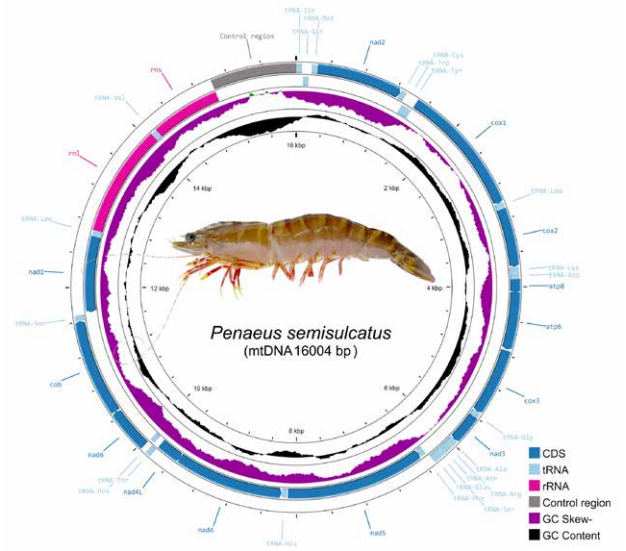

Fig. 2. Schematic representation of circular mitogenome map of banded antennae morphotype of *P. semisulcatus* from the Bay of Bengal, India. The annotated map depicts 13 protein-coding genes (PCGs), two ribosomal RNA genes (rrnS: 12 S ribosomal RNA and rrnL: 16 S ribosomal RNA), 22 transfer RNA (tRNA) genes, and the putative control region. The inner circle depicts GC content

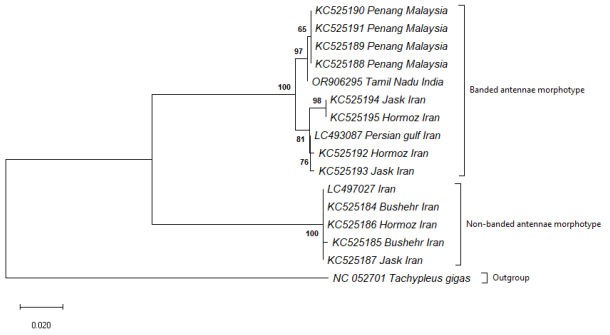

Fig. 3. NJ tree based on genetic distance analysis of COI sequences showing the genetic relationships of two morphotypes of *P. semisulcatus*. Numbers at branching points are bootstrap support. The branch names have accession numbers followed by the collection location

(LC497027) clustered with the COI sequences of the non-banded antennae morphotype with high bootstrap support.

Compositional analysis

The complete mitochondrial genome of *P. semisulcatus* had the following nucleotide compositions: 34.84% thymine, 15.47% cytosine, 33.51% adenine, and 12.57% guanine. The AT/ (A+T) content observed in the whole mitogenome of *P. semisulcatus* was 68.35 % (Table 3) which is similar to the previous studies (Zhong *et al.,* 2019). The average AT content in the concatenated PCGs (with stop codons) was 66.85% which was higher than the GC content (33.15%). The *rrnS* and *rrnL* genes were highly AT biased with 71.39%. The identified control region/D-loop region was also heavily inclined toward AT nucleotides with an average content of 82.25% (Table 3). Upon examination of this region, numerous mononucleotide adenine and thymine repeats were identified along the stretch. Among the thirteen PCGs of *P. semisulcatus*, the lowest AT content was 61.40% in *COX3* gene, whereas the highest AT content was 75.21% in *ATP8* gene.

The AT skew and GC skew in the mitogenome were -0.02 and -0.21 respectively (Table 3). In addition, all the PCGs showed negative AT and GC skews, indicating that there is a skew away from A in favour of T, and G in favour of C similar to the reports in penaeid shrimps (Guo *et al.,* 2021; Collins *et al.,* 2022).

Codon usage in PCGs and genetic divergence

The initiation of twelve of the thirteen PCGs in the mitochondrial genome of *P. semisulcatus* was with traditional start codons of crustaceans (ATG, ATC, and ATT). However, ACG, an alternative putative start codon reported in decapod crustaceans was the start codon in *COX1* (Zhang *et al.*, 2016; Chak *et al.,* 2020)

Eight out of thirteen PCGs terminated with complete or

Table 3. Nucleotide composition and skew analysis of the mitogenome of *P. semisulcatus* in this study

Features	Total	A%	A	T%	T	G%	G	C%	C	A+T%	G+C%	AT-skew	GC-skew
mtDNA (including intergenic spacers)	16004	33.51	5363	34.84	5575	12.57	2012	19.08	3054	68.35	31.65	-0.02	-0.21
tRNAs	1493	32.89	491	32.08	479	19.56	292	15.47	231	64.97	35.03	0.01	0.12
rRNAS	2226	37.38	832	34.01	757	18.1	403	10.51	234	71.39	28.61	0.05	0.27
PCGs (with stop codons)	11157	27.72	3093	39.13	4366	15.92	1776	17.23	1922	66.85	33.15	-0.17	-0.04
1st codon sites	3719	31.30	1164	35.22	1310	18.18	676	15.30	569	66.52	33.48	-0.06	0.09
2nd codon sites	3719	22.45	835	40.28	1498	17.37	646	19.90	740	62.73	37.27	-0.28	-0.07
3rd codon sites	3719	29.42	1094	41.89	1558	12.21	454	16.48	613	71.31	28.69	-0.17	-0.15
NCR	997	40.42	403	41.83	417	7.92	79	9.83	98	82.25	17.75	-0.02	-0.11

Table 4. Codon usage for the thirteen protein-coding genes of *P. semisulcatus*

Amino acid	Codon	Number	RSCU	Amino acid	Codon	Number	RSCU	Amino acid	Codon	Number	RSCU	Amino acid	Codon	Number	RSCU
Phenyl alanine	UUU(F)	198	1.35	Serine	UCU (S)	120	2.72	Tyrosine	UAU(Y)	106	1.49	Tryptophan	UGA (W)	81	1.6
	UUC(F)	96	0.65		UCC (S)	31	0.7		UAC(Y)	36	0.51		UGG (W)	20	0.4
Leucine	UUA(L)	301	3.08		UCA (S)	62	1.41	Histidine	CAU(H)	62	1.41	Arginine	CGU (R)	19	1.27
	UUG(L)	38	0.39		UCG (S)	12	0.27		CAC(H)	26	0.59		CGC (R)	3	0.2
Leucine2	CUU(L)	106	1.09	Proline	CCU (P)	97	2.6	Glutamine	CAA(Q)	52	1.41		CGA (R)	31	2.07
	CUC(L)	25	0.26		CCC (P)	14	0.38		CAG(Q)	22	0.59		CGG (R)	7	0.47
	CUA(L)	89	0.91		CCA (P)	32	0.86	Asparagine	AAU(N)	102	1.48	Serine2	AGU (S)	46	1.04
	CUG(L)	27	0.28		CCG (P)	6	0.16		AAC(N)	36	0.52		AGC (S)	17	0.39
Isoleucine	AUU(I)	235	1.64	Threonine	ACU (T)	96	1.9	Lysine	AAA(K)	63	1.59		AGA (S)	65	1.47
	AUC(I)	51	0.36		ACC (T)	28	0.55		AAG(K)	16	0.41		AGG (S)	0	0
Methionine	AUA(M)	155	1.6		ACA (T)	73	1.45	Aspartic acid	GAU(D)	40	1.08	Glycine	GGU (G)	66	1.1
	AUG(M)	39	0.4		ACG (T)	5	0.1		GAC(D)	34	0.92		GGC (G)	9	0.15
Valine	GUU(V)	92	1.36	Alanine	GCU (A)	144	2.34	Glutamic acid	GAA(E)	59	1.51		GGA (G)	114	1.89
	GUC(V)	26	0.39		GCC (A)	41	0.67		GAG(E)	19	0.49		GGG (G)	52	0.86
	GUA(V)	135	2		GCA (A)	53	0.86	Cysteine	UGU(C)	35	1.75	Stop	UAA (*)	0	0
	GUG(V)	17	0.25		GCG (A)	8	0.13		UGC(C)	5	0.25		UAG (*)	0	0

conventional stop codons (TAA, TAG), and the remaining genes were terminated with truncated stop codon T. Truncated stop codons are frequently found in mitochondrial PCGs, and it is believed that post-transcriptional poly-adenylation completes them to full stop codon (Ojala *et al.,* 1981; Rahuman *et al.,* 2020). ATG stands out as the most frequently utilized start codon, while TAA holds the accounts of being the most commonly employed stop codon (Table 2).

Table 4 illustrates the prevalence of codon families in the Protein-Coding Genes (PCGs) of the mitochondrial genome. A total of 3695 amino acids were encoded in the *P. semisulcatus* mitogenome, among which the three predominant codon families were UUA (leucine), AUU (isoleucine) and UUU (phenylalanine), which are AT-rich. The least chosen codons were from GC-rich codon families such as CGC (arginine), ACG (threonine), and UGC (cysteine). The frequencies of AT-rich codon families are more than amino acids coded by GC-rich codon families, similar to other species in Penaeidae. These findings are also consistent with the nucleotide usage analysis, as the least used codons are those composed mainly of C and G.

Table 5 provides a comprehensive overview of the collective genetic divergence exhibited in thirteen protein-coding genes of *P. semisulcatus* sourced from different seas. The genetic divergence between the banded morphotype from Indian waters (OR906295) and the non-banded antennae morphotype from the Persian Gulf (LC497027) is 16.62%. This notable genetic variation is also consistent with the single-gene divergence (COI) observed in this study (15.61%) and report from the Persian Gulf (17%) (Jahromi *et al.,* 2019). Additionally, a genetic divergence of 6.92% is noted between specimens from Indian waters (OR906295) and those from Chinese waters (MG821354). This indicates the presence of another lineage from the South China Sea. However, a divergence of 1.49% is noted between the banded form from the Persian Gulf (LC493087) and Indian waters (OR906295) which is much less compared to Chinese waters.

Table 5. K2P genetic distance between mitogenomes (13 PCGs) of *P. semisulcatus* computed in MEGA X. The NCBI accession numbers of mitogenomes and locations are given in titles. The asterisk (*) represents the non-banded antennae morphotype of *P. semisulcatus*

Accession No.	OR906295 (Tamil Nadu, India)	LC493087 (Persian Gulf, Iran)	MG821354 (China)	*LC497027 (Iran)
OR906295	———			
LC493087	1.49			
MG821354	6.92	7.25		
*LC497027	16.62	16.94	16.13	———

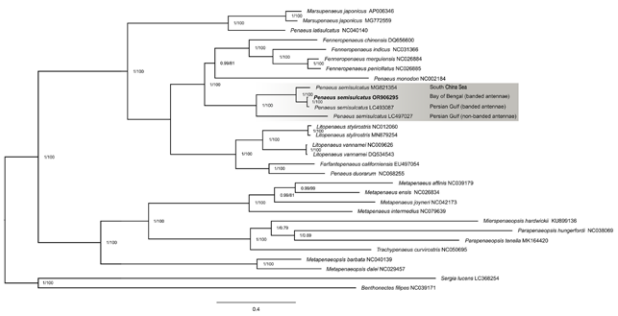

Fig. 4. Consensus phylogenetic tree with BI and ML methods based on the nucleotide alignments of 13 PCGs and 2 rRNA genes in the mitochondrial genome of *Penaeus semisulcatus* and 27 penaeid shrimps. Labels indicate the generic names followed by GenBank accession numbers. The mitogenome data of *P. semisulcatus* is highlighted in grey and the sequence from this study is in bold. Outgroups included are *Benthonectes filipes* and *Sergia lucens*. The number at each node represents posterior probability and bootstrap values

Phylogenetic analysis

Concatenation of PCGs (w/o stop codons) and rRNAs resulted in a sequence of 13115 bp length. A partition scheme was developed that takes into account the position of the nucleotide within the codon, since these sites can evolve at different rates. The optimal partitioning method thus yielded ten sub-sets, each of which contained an estimated nucleotide substitution model.

Phylograms reconstructed for the family Penaeidae with protein-coding and rRNA gene datasets have been depicted in Fig. 4. Both the phylograms generated using Maximum likelihood (ML) and Bayesian Inference (BI) shared the same topology. The species from the family Penaeidae in this study clustered into two major clades, clade 1 contains shrimps from genus *Marsupenaeus*, *Fenneropenaeus*, *Penaeus, Litopenaeus,* and *Farfantepenaeus,* while species from genus *Metapenaeus*, *Mierspenaeopsis*, *Parapenaeopsis*, *Trachypenaeus*, and *Metapenaeopsis* clusters in the second clade.

The first major clade grouped the species into two subclades, one subclade included species of *Marsupenaeus* and *Penaeus*. The subclade 2 divided *Fenneropenaeus* and *Penaeus* into one clade and *Litopenaeus + Farfantepenaeus+ Penaeus* into another, corroborating previous studies with high bootstrap and posterior probability values (Hurzaid *et al.,* 2020; Katneni *et al.,* 2021).

In addition, the species from *Penaeus sensu stricto* (s.s) showed a polyphyletic nature. *Penaeus monodon* and *Penaeus semisulcatus* nested with shrimps from the genus *Fenneropenaeus* validating the polyphyletic nature of *Penaeus s.s* (Hurzaid *et al.,* 2020). The entirety of the banded antennae morphotype in *P. semisulcatus* formed a single, distinguishable major clade, accompanied by two minor clades, indicating two different lineages (Alam *et al.,* 2017; Halim *et al.,* 2021). In contrast, the non-banded antennae morphotype, having diverged from the group mentioned above, acts as an outgroup to the banded morphotype. The genetic divergence and phylogenetic analysis collectively indicated evident differentiation between the two morphotypes within *P. semisulcatus*, suggesting that it may represent two distinct species (Jahromi *et al.,* 2019).

The phylogenetic analysis supported the monophyly of *Metapenaeus*, *Metapenaeopsis*, *Fenneropenaeus*, *Litopenaeus,* and *Farfantepenaeus* (Zhang *et al.,* 2016). The major clade 2, *Metapenaeus* grouped into three clades. All the species from *Metapenaeus* and *Metapenaeopsis* were grouped into two distinct clades. However, the shrimps from *Trachypenaeus*, *Parapenaeopsis*, *Mierspenaeopsis* formed the third sub-clade with *Trachypenaeus* diverging initially from others (Ramirez *et al.,* 2021; Ferreira *et al.,* 2023).

Conclusion

This research addresses a critical gap by furnishing genomic data on green tiger shrimp from Indian waters. Our investigation has yielded valuable insights into the mitogenomic profile of the banded antennae morphotype of *P. semisulcatus* from Indian waters, a widespread morphotype with Indo-West Pacific distribution. Consequently, this study unveils distinctions in the genetic makeup of *P. semisulcatus* mitogenomes across diverse oceanic regions, presenting new avenues for research. Given the ongoing scrutiny of *P. semisulcatus* morphotypes, suggesting the potential existence of distinct species, our findings may contribute further clarification, particularly relevant to Indian waters. The availability of the entire mitogenome not only facilitates comparative genome studies within the *Penaeus* genus but also enhances our comprehension of the evolutionary processes shaping this group.

Acknowledgements

The authors are thankful to the Director of CMFRI and the Head of MBFNHD, CMFRI for providing facilities for this work.

Author contributions

Conceptualization: JNS; Methodology: JNS; Writing Original Draft: JNS, SR; Data Analysis: SR, JNS, SKA; Supervision: TG; Data Collection: SM, VV, KKA, PR, JR, BJ, RB; Supervision: AG.

Data availability

The data supporting this study are publicly available at the repository [NCBI GenBank: Accession number OR906295]

Conflict of interest

The authors declare that they have no conflict of financial or non-financial interests that could have influenced the outcome or interpretation of the results.

Ethical statement

No ethical approval is required as the study does not include activities that require ethical approval or involve protected organisms/ human subjects/ collection of sensitive samples/ protected environments.

Funding

This research was supported by the PMMSY project titled "Sea ranching of Green Tiger Shrimp (*Penaeus semisulcatus*) Post Larvae (PL) in Palk Bay and Gulf of Mannar, Tamil Nadu, for sustainable shrimp production."

References

Alam, M. M., M. D. S. T. De Croos and S. Pálsson. 2017. Mitochondrial DNA variation reveals distinct lineages in *Penaeus semisulcatus* (Decapoda, Penaeidae) from the Indo-West Pacific Ocean. *Mar. Ecol*, 38: e12406.

Alizadeh, E., M. Safaie, M. Momeni and E. Kamrani. 2022. Population dynamics of two morphotypes of green tiger shrimp *Penaeus semisulcatus* De Haan, 1844 in the northern coast of Iran. *Indian J. Fish*, 69: 7-18.

Benson, G. 1999. Tandem Repeats Finder: A Program to Analyze DNA Sequences. *Nucleic Acids Res.,* 27: 573-80.

Bernt, M., A. Donath, F. Jühling, F. Externbrink, C. Florentz, G. Fritzsc, J. Pütz, M. Middendorf and P. F. Stadler. 2013. MITOS: improved *de novo* metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.*, 69: 313-319.

Bolger, A. M., M. Lohse and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinform.,* 30: 2114-2120.

Chak, S. T., P. Barden and J. A. Baeza. 2020. The complete mitochondrial genome of the eusocial sponge-dwelling snapping shrimp *Synalpheus microneptunus. Sci. Rep.*, 10: 7744.

CMFRI. 2009. Annual Report 2008–2009, 58 pp.

CMFRI. 2019. Annual Report 2019, p. 74-90

CMFRI. 2022. Annual Report 2022, 108 pp.

Collins, S. B., H. D. Bracken-Grissom and J. A. Baeza. 2022. The complete mitochondrial genome of the pink shrimp *Farfantepenaeus duorarum* (Burkenroad, 1939) (Decapoda: Dendrobranchiata: Penaeidae). *J. Crust. Biol.*, 42: ruac007.

Dierckxsens, N., P. Mardulyn and G. Smits. 2017. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res.*, 45: e18-e18.

Ferreira, C. S. M., D. C. de Mesquita, Í. A. de Freitas Lutz, I. B. Veneza, T. S. Martins, P. D. C. P. Santana, J. A. B. Miranda, J. M. de Sousa, S. C. D. N. Matos, F. C. A. F. Holanda, M. I. da Cunha Sampaio and G. F. Evangelista-Gomes. 2023. First record of rainbow shrimp, exotic species *Mierspenaeopsis sculptilis* (Heller, 1862), in the Brazilian coastal amazon, validated by DNA barcode. *BMC Zool.*, 8: 11.

Gonzalez-Castellano, I., J. Pons, E. González-Ortegón and A. Martínez-Lage. 2020. Mitogenome phylogenetics in the genus Palaemon (Crustacea: Decapoda) sheds light on species crypticism in the rockpool shrimp *P. elegans. PLoS ONE*, 15(8): p.e0237037.

Grant, J. R and P. Stothard. 2008. The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res.*, 36: 181-184.

Guo, Y., H. Liu, J. Feng, J. Li, Y. Ye, B. Guo and C. Qu. 2021. Characterization of the complete mitochondrial genomes of two species of Penaeidae (Decapoda: Dendrobranchiata) and the phylogenetic implications for Penaeoidea. *Genomics,* 113: 1054-1063.

Halim, S. A. A. A., A. S. Othman, N. A. M. Akib, N. A. Jamaludin, Y. Esa and S. A. M. Nor. 2021. Mitochondrial markers identify a genetic boundary of the Green Tiger Prawn (*Penaeus semisulcatus*) in the Indo-Pacific Ocean. Zoological studies, 60 pp.

Hassanien, H. A and Y. Al-Rashada. 2019. Estimation of genetic variation in green tiger prawn, *Penaeus semisulcatus* by using random amplified polymorphic DNA, inter simple sequence repeat and simple sequence repeat markers. *Aquac. Res.,* 50: 1626-1633.

Holthuis, L.B. 1980. FAO species catalogue. Shrimps and prawns of the world. An annotated catalogue of species of interest to fisheries Vol. 1, Food and Agriculture Organization of the United Nations, Rome. 271 pp.

Huelsenbeck, J. P and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinform.,* 17: 754-755.

Hurzaid, A., T. Y. Chan, S. A. Mohd Nor, Z. A. Muchlisin and W. J. Chen. 2020. Molecular phylogeny and diversity of penaeid shrimps (Crustacea: Decapoda) from South-East Asian waters. *Zool. Scri.,* 49: 596-613.

Jahromi, S. T. 2011. Morphological And Molecular Analysis of *Penaeus semisulcatus* from the Persian Gulf and Oman Sea. Doctoral Dissertation, Universiti Sains Malaysia.

Jahromi, S. T., A. S. Othman and R. Rosazlina. 2019. Morphometrics and Mitochondrial DNA Genes Analysis Suggest a New Species of Penaeus (Crustacea: Penaeidae) from the Persian Gulf. *Biochem. Genet*, 57 (2): 193-213.

Jahromi, S. T., A. S. Othman, R. Rosazlina, S. Pourmozaffar and M. Gozari. 2021. Population genetics of *Penaeus semisulcatus* from Persian Gulf and Oman Sea using newly developed DNA microsatellite markers. *Iran. J. Fish. Sci.*, 20 (1): 157-178.

Katneni, V. K., M. S. Shekhar, A. K. Jangam, B. C. Paran, A. Selvaraj, K. K. Krishnan, N. Kaikkolante, S. K. Prabhudas, G. Gopalapillai and V. K. Koyadan. 2021. Phylogenetic relations and mitogenome-wide similarity metrics reveal monophyly of *Penaeus* sensu lato. *Ecol. Evol.*, 11: 2040-2049.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran and T. Thierer. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinform.,* 28: 1647-1649.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, 16 (2):111-120.

Kurian, C. V. and V. O. Sebastian. 1976. Shrimps and Shrimp Fisheries of India. (revised fifth ed.), Hindustan Publishing Corporation, Delhi, India, 307 pp.

Kumar, S., G. Stecher., M. Li., C. Knyaz and K. Tamura. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.*, 35: 1547-1549.

Lagesen, K., P. Hallin, E. A. Rødland, H. H. Stærfeldt, T. Rognes and D.W. Ussery. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.,* 35: 3100-3108.

Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld and B. Calcott. 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.*, 34: 772-773.

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

Lavrov, D. V., J. L. Boore and W. M. Brown. 2000. The Complete Mitochondrial DNA Sequence of the Horseshoe Crab *Limulus polyphemus. Mol. Biol. Evol.*, 17: 813-24.

Mohammed-Geba, K. and A. Yousif. 2022. COI gene sequencing reveals genetic population structuring and a separate lineage of the green tiger prawn *Penaeus semisulcatus* in the Gulf of Suez and the Bitter Lakes, Egypt. *Aquat. Living Resour.*, 35 pp.

Nguyen, L. T., H. A. Schmidt, A. Von Haesele and B. Q. Minh. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.*, 32: 268-274.

Ojala, D., J. Montoya and G. Attardi. 1981. tRNA Punctuation Model of RNA Processing in Human Mitochondria. *Nature*, 290: 470-74.

Perna, N. T and T. D. Kocher. 1995. Patterns of Nucleotide Composition at Fourfold Degenerate Sites of Animal Mitochondrial Genomes. *J. Mol. Evol.*, 41: 353-58.

Rahuman, S., N. S. Jeena, P. K. Asokan, R. Vidya and P. Vijayagopal. 2020. Mitogenomic architecture of the multivalent endemic black clam (*Villorita cyprinoides*) and its phylogenetic implications. *Sci. Rep.,* 10: 15438.

Rajkumar, M., S. L. Pillai, P. Rameshkumar, R. Saravanan, S. Thirumalaiselvan, J. Jose, K. S. Sobhana and M. R. George. 2023. Reproductive biology of green tiger shrimp *Penaeus semisulcatus* De Haan, 1844 in Palk Bay, southeast coast of India. *Reg. Stud. Mar. Sci.*, 66: 103161.

Rambaut, A., A. J. Drummond, D. Xie, G. Baele and M. A. Suchard. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.*, 67: 901-904.

Ramirez, J. L., L. Simbine, C. G. Marques, E. Zelada-Mázmela, L. E. Reyes-Flores, A. S. López, J. Gusmão, C. Tavares, Jr. P. M. Galetti and P. D. Freitas. 2021. DNA Barcoding of Penaeidae (Decapoda; Crustacea): Non-Distance-Based Species Delimitation of the Most Economically Important Shrimp Family. Diversity, 13: 460.

Saitou, N. and M.. Nei.1987. The neighbor-joining method: a new method for reconstruction of phylogenetic trees, *Mol. Biol. Evol.*, 4: 406-425.

Sambrook, J. and D.W. Russell. 2006. Purification of nucleic acids by extraction with phenol: chloroform. Cold Spring Harbor Protocols. https://doi.org/10.1101/pdb.prot4455.

Sharawy, Z., E. M. Abbas, M. G. Desouky and M. Kato. 2016. Descriptive analysis and molecular identification of the green tiger shrimp *Penaeus semisulcatus* (De Haan, 1844) in Suez Gulf, Red Sea. *Int. J. Fish. Aquat. Stud.*, 4: 426-432.

Sharawy, Z. Z., E. M. Abbas, A. R. Khafage, A. Galal-Khallaf, R. F. Ismail, H. O. Ahmed, K. Mohammed-Geba and M. Kato. 2017. Descriptive analysis, DNA barcoding and condition index of *Penaeids* (Crustacea: Decapoda) from the Egyptian Mediterranean coast. *Fish. Res.*, 188: 6-16.

Shen, H., A. Braband and G. Scholtz. 2013. Mitogenomic analysis of decapod crustacean phylogeny corroborates traditional views on their relationships. *Mol. Phylogenet. Evol.,* 66 (3): 776-789.

Sudhakara Rao, G., E. V. Radhakrishnan and J. Jose. 2013. *Handbook of Marine Prawns of India.* ICAR-Central Marine Fisheries Research Institute. p. 1-29.

Tan, M. H., H. M. Gan, Y. P. Lee, G. C. Poore and C. M. Austin. 2017. Digging deeper: new gene order rearrangements and distinct patterns of codons usage in mitochondrial genomes among shrimps from the Axiidea, Gebiidea and Caridea (Crustacea: Decapoda). Peer J 5: e2982.

Wyman, S. K., R. K. Jansen and J. L. Boore. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinform.*, 20: 3252-3255.

Zhang, D., J. Huang, F. Zhou, F. Gong and S. Jiang. 2016. The complete mitochondrial genome of banana shrimp *Fenneropenaeus merguiensis* with phylogenetic consideration. *Mitochondrial DNA Part A,* 27: 2606-2607.

Zhong, S., Y. Zhao, G. Huang and W. Xu. 2019. The complete mitochondrial genome of *Penaeus semisulcatus* (Decapoda: Penaeidae). Mitochondrial DNA B Resour, 27: 3274-3275.