CHAPTER 6

A COMPREHENSIVE PHYLOGENETIC ANALYSIS OF PRAWN AND SHRIMP BASED ON MITOCHONDRIAL COI SEQUENCE

CONTENTS

6.1 Introduction

The molecular phylogeny concept is a modern technique, which is a combination of molecular and statistical methodologies to infer evolutionary relationships among organisms using molecular data. This empirical approach uses the structure and role of molecules and how these evolutionary relationships develop over time. It's a beneficial technique to detect the uniqueness or difference of individuals, populations, or species based on their molecular sequences. It can give precise data on the genetic relationship of the individual species, avoiding taxonomic ambiguity. The nucleotide sequences different within a gene reflect the evolutionary relationship between two organisms and species.

Mitochondrial DNA (mtDNA) is frequently used for the polygenetic study of a wide range of marine taxa. Many mtDNA regions that code for protein genes or involve in regulation as the control regions are used as genetic markers to investigate intra-species and interspecies diversity. Mitochondrial DNA can assemble multiple base substitutions over a long-time duration, offering a comparative method for taxonomic, evolutionary, and phylogenetic analysis. The COI and 16s rDNA genes are very popular for such research, which is best suited to study the taxonomic and

phylogenetic analyses from the species up to the family level (Johns and Avise, 1998; Hebert et al., 2004).

Molecular phylogenetic analysis studies on prawns and shrimps are rare in India, with only a few studies investigating phylogeny of such a large and diverse group of Decapoda. These studies also include genetic diversity and the population structure of prawns and shrimps. Mishra et al. (2009) studied the genetic diversity of *Metapenaeus dobsoni* from three locations of India using five random primers for PCR amplification. The polymorphic loci varied from 22.3-40.9 % in the three stock populations. The molecular phylogeny is also used in differentiating and studying the intra and interspecies patterns of the species population. In 2012, Kumar and his team studied the efficiency of DNA barcoding in delineating *P. monodon* in its different life stages (Mysis, post larvae, and adult). This study clearly showed that the partial COI gene provides an accurate delineation of species irrespective of their different life stages. In 2013, Khedkar et al. studied the population genetics of the Giant tiger prawn (*Penaeus monodon*) from Andhra Pradesh, East coast of India. The genetic diversity of the Kakinada population appears high compared to the other populations. This kind of study might be useful as genetic indicators for an aquaculture field, like planning for selective breeding, stock diversity, and hatchery stock assessment. In 2014, Jose et al. established the phylogenetic relationship of *Glyphocrangon investigatoris* with other species of *genus Glyphocrangon*u sing the mtCOI gene.

Rajakumaran et al. (2014) studied the phylogenetic relationship among thirteen species of penaeidae shrimps based on the morphometric (UPGMA cluster analysis) and molecular (RAPD) data. The morphological and molecular data are a reliable method for confirming the phylogenetic relationship of these economically important species. In 2015, Rajkumar and his team identified the six species of Penaeidae prawns using DNA barcoding of (COI gene). They also established the phylogenetic evolution of the penaeoidae species, which formed the polyphyletic clade. Chakraborty et al. (2015) analyzed the molecular phylogeny of *Aristeus*

alcocki from the South-West coast of India using two nuclear-protein coding genes (PEPCK and NaK) and two mitochondrial genes (16S rDNA and COI). The sequence identity, genetic distance, and phylogenetic tree investigations show that *A. alcocki* is closely similar to *A. antennatus.*

To date, several studies have been carried out to establish the diversity of shrimps from the southern part of India through DNA barcoding approaches (Mamatha et al., 2016; Subbaiya et al., 2017). Chowdhury et al. (2018) studied the phylogenetic relationship among five species of the genus *Parapenaeopsis* using partial sequences of mt 16S rDNA and mt COI genes. They resolved the taxonomic ambiguity between *P. stylifera* and *P. coromandelica*, a closely similar species. Kundu and his team (2018) developed DNA barcode data for morphologically identified six species of penaeid shrimps collected from Chilika Lake, India. The Mitochondrial (COI) gene is a successful partial gene segment for establishing the evolutionary relationship among crustacean species (Saad and El-Sadek, 2017). Purushothaman and his team (2019) investigated the taxonomy and phylogenetic relationships of the deep-water penaeoid shrimps (families: Solenoceridae, Penaeidae, and Aristeidae) using three mitochondrial (COI, Cytb, 16S rRNA) genes from the Indian coast. The analysis shows that all the species in these families are monophyletic. Karuppasamy et al. (2020) studied the molecular and systematic identification of food marine shrimps using the mt COI gene from the Southeast Coast of India. All the generated sequences were showed 99−100% similarities with the referral database sequences.

In this chapter, the molecular phylogeny of prawns and shrimps in Gujarat waters was investigated using COI gene sequences, with a focus on the complex formed by morphologically similar species. Relevant type specimens were examined and sequences to link the species name to the molecular clades recovered in the COI phylogenies. We also developed a standard DNA barcode of a few species of prawns and shrimps, which is previously not available on the NCBI database.

6.2 Material and Methods

6.2.1 Sampling of Molecular data

A total of 112 specimens of marine prawn and shrimps were collected for DNA examination. The samples represented 52 species of shrimps and prawns from different habitats and localities along the coastal area of Gujarat, including Union territory Daman and Diu.

6.2.2 DNA extraction

Genomic DNA was extracted from abdominal tissue or pleopods of prawns and shrimps. The initial weight was approximately 20 mg, and extraction was carried out using DNeasy Blood and Tissue kit (Qiagen) (Annexure 6.1).

6.2.3 Quantification of genomic DNA

After extraction of genomic DNA, quantification was carried out using QIAxpert (QIAGEN). The purity and concentration of DNA were measured using gel electrophoresis (Annexure 6.2).

6.2.4 PCR amplification

The region of the mtCOI gene was amplified using primers given in (Table 6.1). The final volume of 20 µl containing 10 µl Taq PCR master mix (HiMedia), 10 pmol forward primer, 10 pmol reverse primer, 2 µl template DNA, and nuclease-free water (to make up the final volume). The gene amplification was carried out in a thermal cycler (Applied Biosystems Veriti®) using the PCR condition given in tables 6.2 and 6.3.

Table 6.1 Oligonucleotide primers used in the present study to amplify the COI gene.

Primer	Primer sequence 5'-3'	Reference
LCO 1490	GGTCAACAAATCATAAAGATATTGG	Folmer et al., 1994
HCO 2198	TAAACTTCAGGGTGACCAAAAAATCA	
C1-J-1718	GGAGGATTTGGAAATTGATTAG	Simon et al., 1994

C1-N-2191	CCCGGTAAAATTAAAATATAAACTTC	

Table 6.2 Thermal cycling conditions used in the present study to amplify the COI gene using LCO 1490 and HCO2198.

Stage 1	**Stage 2 (35 Cycles)**			**Stage 3 (5 Cycles)**			**Stage 4**	
94°C 1 min.	94°C 1min.	45°C 1 min. 30 sec.	68°C 1 min. 30 sec.	94°C 1 min.	50°C 1 min. 30 sec.	68°C 1 min. 30 sec.	68°C 10 min.	4°C ∞

Table 6.3 Thermal cycling conditions used in the present study to amplify the COI gene using C1-J-1718 and C1-N-2191.

Stage 1	**Stage 2 (35 Cycles)**			**Stage 3 (1cycle)**	**Stage 4**
95°C 3 min.	95°C 30 sec.	50°C 30 sec.	72°C 1 min. 30 sec.	72 °C 7 min	4°C ∞

6.2.5 Qualification and Quantification of PCR

After PCR amplification, the amplified regions were observed using gel electrophoresis (Annexure 6.2).

6.2.6 PCR Purification

Positive PCR amplified were purified, using EXOSAP IT (affymetrix) according to the manufacturer's instructions. In a new PCR tube, 10 µl of positive PCR product and 4 µl of EXOSAP were added. A thermal cycler was run.

Table 6.4 PCR purification condition used in the present study.

Stage I	**Stage II**	**Stage III**
37˚C for 15 minutes	80˚C for 15 minutes	4˚C

6.2.7 Sequencing

The PCR amplified product was bi-directionally sequenced using the same set of the universal primers (LCO 1490 HCO 2198 and COI 1718F COI 2191R) on the ABI 3730x196 capillary DNA analyzer using Big Dye Terminator v 3.1 sequencing kit at Eurofins, Bangalore.

6.2.8 Sequence and Phylogenetic analysis

The qualities of the bi-directional chromatogram of the generated sequences were checked, and noisy parts were trimmed at both ends to avoid the noisy part using sequences nucleotide sequence DNA baser assembler 5.15 version. Further, each sequence was a checked for sequence similarity through BLASTn analysis with Gene bank sequences (NCBI database). All the sequences were submitted on the NCBI portal with a unique accession code. The multiple sequences alignment was done for all the species of order Dendrobranchiata and Pleocyemata by multiple sequences alignment was opting using the MEGA X version. All missing data and gaps were eliminated from the data set. The phylogenetic was performed using the ML (Maximum Likelihood) method in the Kimura 2-parameter model with 1000 bootstrap replicates using the MEGA X version (Kumar et al., 2018).

6.3 Data analysis and Results

DNA was extracted from 47 different individuals of prawns and shrimps collected from different locations and habitats along the Gujarat coast. Only 42 PCR products were obtained and then sequenced for the COI gene. DNA extractions from the samples are shown in (fig. 6.1), and PCR products are shown in (fig. 6.2). Most of the generated or developed sequences show 90-100% similarity with the conspecies database sequences in GeneBank. In the present study, a total of 42 homologous sequences (27 sequences of Dendrobranchiata and 15 sequences of Pleocyemata) were developed, and 30 sequences were successfully uploaded on NCBI. In addition, 6 sequences were obtained from the NCBI for the phylogenetic analyses (Table 6.5).

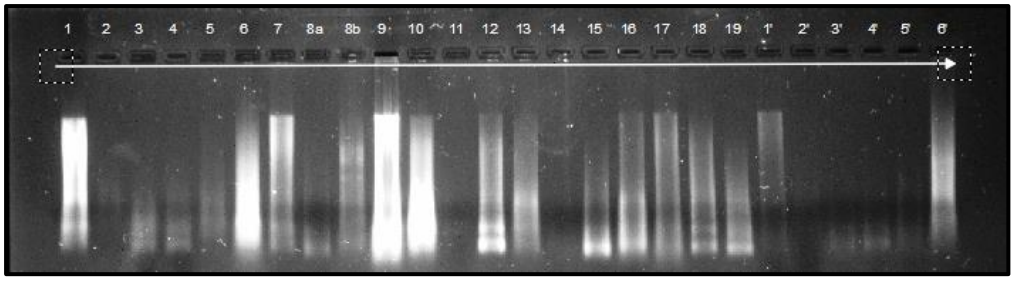

Figure 6.1 DNA extraction products visualized on an agarose gel and stained with ethidium bromide.

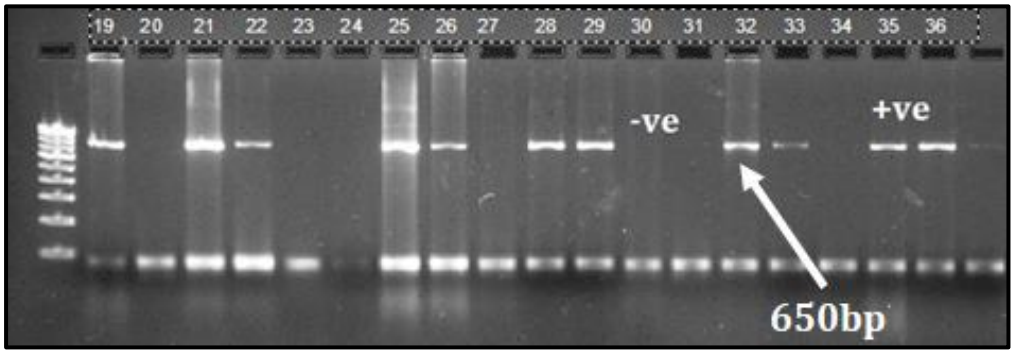

Figure 6.2 PCR final products visualized on an agarose gel stained with ethidium bromide.

The phylogenetic analysis was performed using the ML method (maximum likelihood), which is every site was considered for the calculation, and the likelihood of the replacement of a nucleotide from the pools of nucleotide was calculated. The phylogenetic analysis was performed using the bootstrapping method to ensure the statistical validity of each branch point in a tree. The bootstrap values higher than 90 out of 100 were considered statistically significant, whereas the value is below 50 was essentially random change. The phylogenetic relationship of Dendrobranchiata and Pleocyemata species is discussed below.

Table 6.5 List of species, locality, GeneBank accession number, and the number of base pairs.

S. No.	Family	Species	Locality	Accession Number	Base-pairs
1	Penaeidae	*Ganjampenaeopsis uncta*	Porbandar	processed	662
2		*Megokris granulosus*	Porbandar	MT677665	435
3		*Megokris sedili*	Porbandar	MT677666	262
4		*Metapenaeopsis barbata*	Porbandar	MT986019	375
5		*Metapenaeopsis stridulans*	Porbandar	Processed	673
6		*Metapenaeus affinis*	Veraval	MT986021	396
7		*Metapenaeus brevicornis*	Okha	MT677667	531
8		*Metapenaeus dobsoni*	Okha	Processed	651
9		*Metapenaeus kutchensis*	Veraval	MT986015	548
10		*Metapenaeus monoceros*	Okha	MT986020	427
11		*Metapenaeus moyebi*	Veraval	FJ435653*	847
12		*Mierspenaeopsis hardwickii*	Okha	MN205308	702
13		*Mierspenaeopsis sculptilis*	Verval	MT677668	431
14		*Parapenaeopsis stylifera*	Okha	MT986016	374
15		*Parapenaeus fissuroides indicus*	Porbandar	Processed	630
16		*Parapenaeus longipes*	Porbandar	MT677669	544
17		*Penaeus canaliculatus*	Porbandar	Processed	650
18		*Penaeus indicus*	Porbandar	MT677670	463

19		*Penaeus japonicus*	Okha	MT986017	338
20		*Penaeus latisulcatus*	Okha	MT677671	554
21		*Penaeus merguiensis*	Veraval	MT677672	552
22		*Penaeus monodon*	Veraval	MT677673	553
23		*Penaeus penicillatus*	Porbandar	MT986018	396
24		*Penaeus semisulcatus*	Okha	MT677674	554
25		*Trachysalambria curvirostris*	Porbandar	MT677675	630
26	Solenoceridae	*Solenocera choprai*	Porbandar	MN228496	646
27		*Solenocera crassicornis*	Veraval	MN340982	629
28		*Solenocera koelbeli*	Porbandar	MN228495	546
29	Callianassidae	*Gilvossius rotundicaudatus*	Shivrajpur	MT611419	458
30		*Neocallichirus jousseaumei*	Shivrajpur	Processed	375
31	Alpheidae	*Alpheus chiragricus*	Okha	Processed	597
32		*Alpheus edwardsii*	Shivrajpur	-	-
33		*Alpheus lobidens*	Shivrajpur	Partial sequence	678
34		*Alpheus malabaricus*	Daman	-	-
35		*Alpheus pacificus*	Veraval	MW008767	338
36		*Athanas dimorphus*	Okha	-	-
37		*Athanas parvus*	Pirotan Island	-	-
38		*Synalpheus coutierei*	Shivrajpur	KJ625030*	444

39	Hippolytidae	*Exhippolysmata ensirostris ensirostris*	Dandi	MT986022	488
40		*Latreutes anoplonyx*	Jakhau	Processed	641
41		*Saron marmoratus*	Shivrajpur	MW008766	468
42	Lysmatidae	*Lysmata vittata*	Shivrajpur	Processed	492
43	Thoridae	*Thor amboinensis*	Pirrotan Island	Processed	718
44	Palaemonidae	*Ancylocaris brevicarpalis*	Narara	Processed	665
45		*Cuapetes grandis*	Shivrajpur	KU064963*	489
46		*Nematopalaemon tenuipes*	Veraval	MT611064	409
47		*Palaemon pacificus*	Gopnath	MT611061	409
48		*Palaemon serrifer*	Shivrajpur	MT611063	513
49		*Palaemon styliferus*	Okha	MT611062	522
50	Pandalidae	*Procletes levicarina*	Porbandar	MK470824*	483
51	Upogebiidae	*Upogebia carinicauda*	Pirotan Island	AB771735*	429
52	Spongicolidae	*Microprosthema validum*	Pirotan Island	KJ722807*	491

***Downloaded from the NCBI website** (for the single copy specimens or where we could not get the sequencing results).

6.3.1 Phylogenetic analysis of Dendrobranchiata species

The analysis involved 28 nucleotide sequences. Out of 28 COI gene sequences, and 1 sequence was obtained from the NCBI GeneBank. All positions containing missing data and gaps were eliminated. The evolutionary history of Dendrobranchiata species was inferred using the Maximum Likelihood method based on the Kimura 2-parameter model with uniform rate distribution. There was a total of 67 positions in the final dataset.

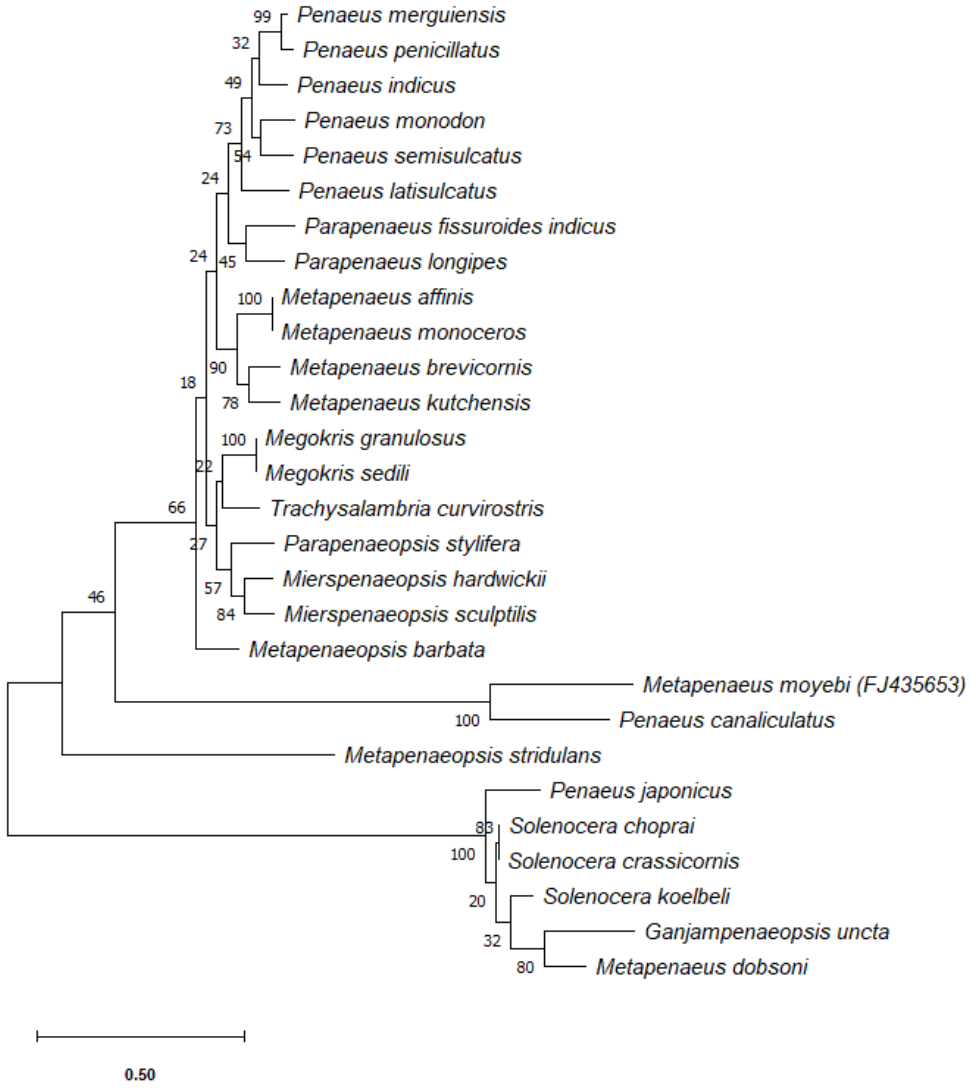

Figure 6.3 Maximum likelihood tree deciphering phylogeny of Dendrobranchiata species collected from the Gujarat coast based on the COI gene.

The evolutionary history of prawns was inferred using the Maximum Likelihood method based on the Kimura 2-parameter model (Kimura, 1980) with a uniform rate. The tree with drawn to scale, with branch lengths measured in the number of substitutions per site, and the highest log-likelihood was -1076.51. The optimal maximum likelihood tree is shown in fig. 6.3.

6.3.2 Phylogenetics analysis of Pleocymeta species

The analysis involved 17 nucleotide sequences. Out of 17 COI gene sequences, 5 sequences were obtained from the NCBI GeneBank. All positions containing missing data and gaps were eliminated. The evolutionary history of Pleocymeta species was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model with uniform rate distribution.

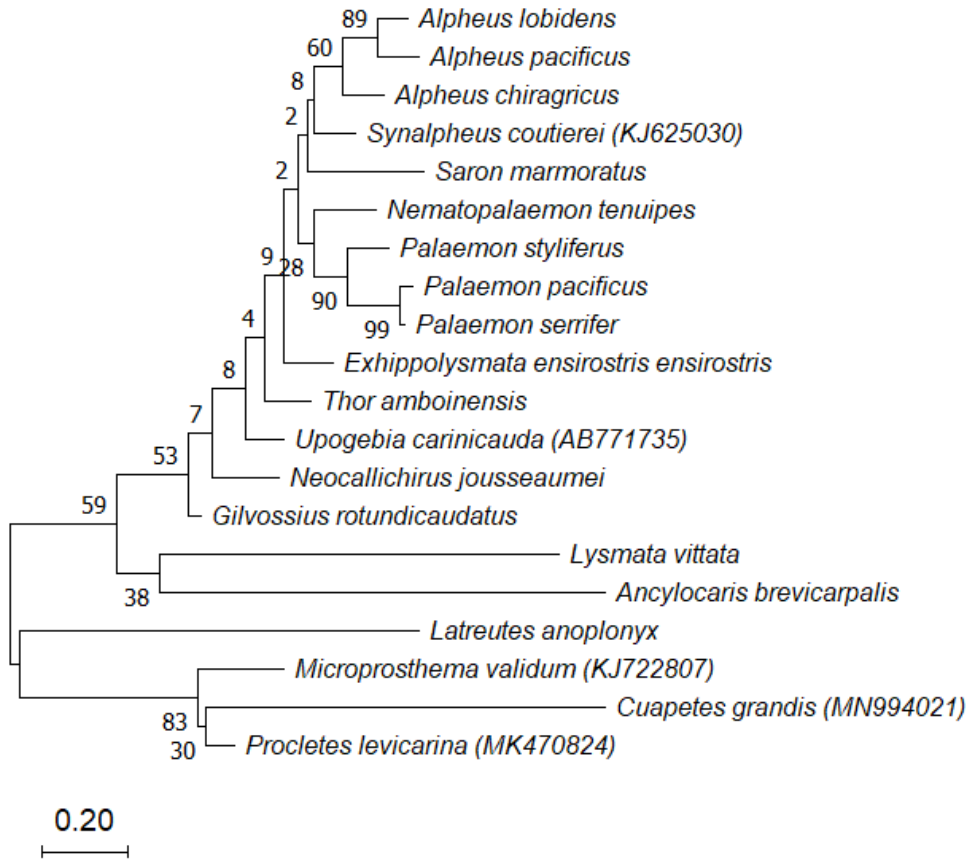

Figure 6.4 Maximum likelihood tree deciphering phylogeny of Pleycometa species collected from the Gujarat coast based on the COI gene.

There was a total of 250 positions in the final dataset. The evolutionary history was inferred using the Maximum Likelihood method based on the Kimura 2-parameter model (Kimura, 1980) with a uniform rate. The tree with drawn to scale, with branch lengths measured in the number of substitutions per site, and the highest log-likelihood was -4177.34. The optimal maximum likelihood tree is shown in fig. 6.4.

6.4 Discussion

The Penaeidae has a long evolutionary history (Dall et al., 1990), high species diversity, and some genera have shown large genetic distances between each other. Nevertheless, traditional classification schemes (Kubo 1949; Glaessner, 1969; Burkenroad, 1983; Dall et al., 1990; Pérez Farfante and Kensley 1997; Martin and Davis 2001) have treated the Penaeidae as a single entity. In 2008, Chan et al. studied the molecular phylogeny of 30 species from 20 genera of Penaeidae by mitochondrial 16S rRNA gene. The observation generally supports the three-tribe scheme proposed by Burkenroad (1983), but it is not consistent with the five-group classification of Kubo (1949). The mitochondrial COI gene is an informative tool for studying phylogenetic analysis among a wide array of penaeid genera, although it is applicable in studies at a lower systematic level (Lavery et al., 2004). In 1991, Palumbi and Benzie proposed that even though the morphology of penaeid shrimps evolved very slowly, changes in the mt DNA could still facilitate substantial genetic diversity. Most of the studies, however, have focused on the phylogenetic issues of individuals.

Our study presents the most comprehensive and robust molecular phylogenetic survey of Penaeoidea to date from the Gujarat coast. It is also the first molecular phylogenetic study. The resulting phylogenetic tree is very different from those obtained from mitochondrial markers, which suggest a close relationship between genus *Penaeus*, *Parapenaeus*, *Metapenaeus*, *Megokris*, and *Mierspenaeopsis*. The only exception is the second clade containing the two species *M. moyebi*, *P. canalicatus,* with 100% bootstrap support. P. *merguiensis* and *P. penicillatus* were the most

closely related species, with *P. indicus* as the sister taxon. Lavery et al. (2004) also recorded the same kind of similar observation. They found out the *Penaeus penicillatus* and *Penaeus silasi* were most closely related to *P. indicus* as the sister taxon. The present study was also supported by morphological taxonomy (Holthuis, 1980). All species of the genus *Penaeus* are arranged with the help of sister clades except *P. canaliculatus* and *P. japonicus*. As the same, all the species of the genus *Metapenaeus* shared the sister clades except the *M. moyebi* and *M. dobsoni*. Both species of the genus *Megokris* are closely similar and formed sister clades with high bootstrap support (100%), which is ample their genetic congruence. The taxonomic revision supported this genus from the sister taxon *T. curvirostris*. Some species like *M. sedili*, *M. granulosus* initially grouped under the genus *Trachypenaeus* are revised and assigned under a new genus *Megokris* erected by Pérez Farfante and Kensley (1997). *Mierspenaeopsis hardwickii* and *Mierspenaeopsis sculptilis* are closely resalted species with high bootstrap support (84%), with the sister taxon viz. *Parapenaeopsis stylifera* with bootstrap support of 57%. The present study is also identical to the taxonomy, where they have observed a significant difference in petasma structure. Based on this Sakai and Shinomiya (2011) separated these two species (*M. hardwickii* and *M. sculptilis*) from the genus *Parapenaeopsis* and assigned them to a new genus. *Metapenaeopsis stridulans* stood alone, as immediate next clade to the later with 46% bootstrap support, indicating its common ancestry. All three species of the family Solenoceridae arranged in a single sister clade. *S. choprai* are closely similar species to and *S. crassicornis*, shared clads with 83% bootstrap value. *Metapenaeopsis* and *Solenocera* were closely related but far diverged from their ancestor genus. The single species of the genus *Ganjampenaeopsis* were recorded during the present study was shared a sister clade with *M. dobsoni*, with an 80% bootstrap value.

This study, representing the first molecular investigation of systematic relationships within the suborder Pleocyemata and encompassing 3 infraorder, 20 species across 16 different genera, illustrates several

important results. The consistent results were obtained from the ML tree using the mt COI gene were used to understand the phylogenetic relationships among major lineages of pleocyemata species identified during the present study (fig. 6.4). During this study, 4 sequences were first time submitted on the NCBI database. The results strongly supported the formation of three distinct clades. All the species level grouping within the genus *Alpheus* formed a subsequent clade that was strongly supported with 89% bootstrap value. Our molecular phylogeny proposes a close relationship (sister position) between *A. lobidens* and *A. pacificus* (fig. 6.4). Genus *Alpheus* is the most diverse group under the infraorder caridea (Shen et al., 2012). The position of the family among pleocyemata in all analyses obtained here showed Alpheidae closer to Hippolytidae (as in clade 5); they belong to the same superfamily Alpheoidea (Christoffersen, 1990). The first clade (herein referred to as 'Palaemonidae' clade) comprises *Palaemon serrifer* as a sister species with Palaemon *pacificus*, with a high bootstrap value (99%), share a sister taxon with other genera of this family. Carvalho et al. (2017) established an integrative approach to the evolution among 60 species of *Palaemon* and 15 species from other genera of Palaemonida. *Ancylocaris brevicarpalis* and *Cuapetes grandis* were arranged with the other taxon, supported by very low bootstrap value.

The phylogenetic analysis constituting only one species of for the genus *Exhippolysmata* (*E. ensirostris ensirostris*) was directly formed in the different clades. It showed a great divergence to another genus that is highly separated from the root clade. It highly diverges from the ancestor clade caridean. The Callianassidaea constituted the first large clade. Both the species of Ghost shrimp (*Gilvossius rotundicaudatus* and *Neocallichirus jousseaumei*) and mud shrimp (*Upogebia carinicauda*) were arranged nearby, supported by a low bootstrap value (53%). Thalassinidea divided it into two groups (namely Gebiidea and Axiidea). The monophyly of Thalassinidea has been supported by some morphological cladistic analyses as well as molecular data or combined morphological and

molecular analysis (Lin et al., 2012). *Microprosthema validum* species belonging to the family Spongicolidae shared a sister clad with the *Procletes levicarina* (family Pandalidae), with good bootstrap value. Species *Procletes levicarina* arranged with *Cuapetes grandis,* with low bootstrap value (30%).

The present study reveals the utility of mt COI genes in the reconstruction of Dendrobranchiata and Pleycometa phylogeny. Although a limited number of species were examined (each genus and family) in this study, studies involving the complete mitochondrial genomes of these groups are necessary to understand the comprehensive phylogeny of the Dendrobranchiata and Pleycometa.