Original Article

Genetic divergence and phylogenetic analysis based on cytochrome c oxidase subunit-1 sequence of enope squid Abralia andamanica (Goodrich 1896) inhabiting Andaman Sea

Naveen Sathyen 1, Chaithanya E R 1, Anil Kumar P R 1, Sruthy K S 1, Rosamma Philip 1*

1Department of Marine Biology, Microbiology and Biochemistry, School of Marine Sciences, Cochin University of Science and Technology, Fine Arts Avenue, Kochi-682016, Kerala, India
*Corresponding author E-mail: rosammap@gmail.com, rose@cusat.ac.in
Phone: + 914842368120, Fax: + 914842381120

Received 26 June 2014; accepted 17 July 2014

Abstract

Cytochrome c oxidase subunit-1 (CO1) gene sequence of enope squid Abralia andamanica sampled from Andaman Sea was compared with the CO1 gene sequence of A. andamanica from China Sea. Assessment of CO1 nucleotide and protein sequence of enope squid Abralia andamanica from Andaman Sea and China Sea revealed that with respect to CO1 sequence genetic divergence exist between the two enope squids. A. andamanica inhabiting Andaman Sea exhibited more affinity towards A. veranyi than to its counterpart from China Sea. Pairwise distance calculated based on Kimura 2-parameter (K2P) model for the enope squids from Andaman Sea and China Sea was found to be almost half of the mean pairwise distance determined for order Teuthida indicating high genetic variation among the two enope squids with regard to CO1 gene. Phylogenetic analysis of the CO1 nucleotide sequence of A. andamanica was performed to determine its relationship with other squids belonging to the order Teuthida. A. andamanica aligned along with the clade formed by family Enoploteuthidae. Sixteen families of order Teuthida were considered for phylogenetic analysis, overlap was observed three families. The study endorses the proficiency of CO1 based DNA barcoding in determining the phylogeny and genetic divergence of a species.

Key words: DNA Barcoding; Phylogenetic analysis; Abralia andamanica; Squid; Cephalopods; Genetic divergence.

1. Introduction

Molluscs represent a group of organisms common to aquatic and terrestrial habitats. Cephalopods are extraordinary molluscs having vertebrate like intelligence, advanced visual perception and evolved complex behavior [1, 2, 3]. They are ecologically important and exclusively marine organisms inhabiting a range of marine environments from deep sea to shallow waters of rocky shores [4]. Cephalopods comprises more than 800 identified species [5], out of which approximately 450 known species represent squids, cuttle fishes and bobtail squids which form the super order Decapodiformes Leach, 1819 [6]. The rich diversity, worldwide distribution and rich fossil records of cephalopods have attracted the attention of biologists and geologists alike. Owing to morphological similarity, it is difficult to distinguish closely related cephalopods with conventional morphology based identification [7]. Under these circumstances, CO1 gene sequence could be effectively employed for taxonomic identification, phylogenetic analysis and determination of genetic variation in closely related cephalopods as it provides an option of classification based on intraspecific and interspecific genetic variations.

Our understanding of higher-level phylogenetic relationships of cephalopods is surprisingly rudimentary. Young and Vecchione [8] conducted a cladistic analysis of morphological characters which has been helpful in elucidating some of the relationships within the Coleoidea, particularly in regard to the Vampyromorpha and Octopoda. Phylogenetic analysis of relationships within the coleoids using molecular sequence data is a promising alternative to the tedious and time consuming morphology based approach. The initial examinations of coleoid relationships using molecular data [9, 10] focused primarily on members of the Sepioidea and Myopsida. Bonnaud et al. [10] examined approximately 500 base pairs of the mitochondrial cytochrome oxidase III gene and included two oegopsid and two octopod families and Vampyroteuthis infernalis. Phylogenetic analysis of the coleoid cephalopods based on molecular sequence data from the mitochondrial cytochrome c oxidase subunit I (COI) gene was attempted with success by Carlini and
Graves [11]. Enope squid, Abralia andamanica belongs to family Enoploteuthidae which comprises of mesopelagic squids inhabiting upper 200 m that undergo extensive diurnal vertical migration and have epipelagic paralarvae and juveniles. In the present study, we report the results of phylogenetic analysis of enope squid Abralia andamanica based on nucleotide sequence of Cytochrome oxidase subunit-1 (CO1) gene. CO1 gene sequence of Abralia andamanica from Andaman Sea was compared with that of A. andamanica from China Sea to examine the genetic variation between the two populations.

2. Materials and Methods

2.1 Sample collection
Enope squid, Abralia andamanica was caught from a depth of 500 m off Sound Island, Andaman Archipelago (India) during Cruise No. 292 of Fisheries and Oceanography Research Vessel Sagar Sampada (Ministry of Earth Sciences, Govt. of India). High Speed Demersal Trawl (HSDT) net operated on-board was employed for capturing the species. Mantle tissue of the squids was preserved in ethanol and stored at room temperature onboard Biological Laboratory facility of the research vessel.

2.2 Morphological Identification
Morphological identification of the specimen was carried out by comparing it with the description of the Holotype specimen deposited at Calcutta museum [12]. The holotype specimen was captured from a depth of 344 to 585 m from Andaman Sea.

2.3 DNA extraction, PCR amplification and sequencing
Total genomic DNA was extracted from the mantle tissue using TRI® reagent (Sigma) and following manufacturer’s instruction. Purity and quality of DNA was checked on 0.8% agarose gel. The concentration of dissolved DNA was estimated using UV spectrophotometer (Hitachi U-2900). DNA was diluted so as to obtain a final concentration of 100 ng/μL. The primers used for the amplification of CO1 gene were LCO1490 (5'-GGTCAACAAATCATATACTAAAA GATATTGGG-3') and HC02198 (5'-TAAACTTCAGGGTG G CCAAAATCA -3') [13]. Gene amplification was carried out in a 25 μl reaction volume containing 1x standard Taq buffer (10 mM Tris-HCl, 50 mM KCl, pH 8.3), 1.5 mM MgCl2, 200 mM dNTPs, 0.4 mM each primer and 1U Taq DNA polymerase (New England Biolabs). 1 μl of genomic DNA was used as template. The thermal regime consisted of an initial denaturation at 95°C for 5 minutes followed by 35 cycles of 95°C for 45 seconds, 50°C for 30 seconds and 72°C for 45 seconds and a final extension at 72°C for 10 minutes. Amplicons obtained were sequenced using ABI Prism Sequencing kit (BigDye Terminator Cycle) at SciGenom Sequencing Facility, India.

2.4 Sequence data analysis
The homologue searching of the nucleotide sequence (using blastn suite) and the deciphered amino acid sequence (using blastp suite) were performed with the Basic Local Alignment Search Tool (BLAST) through NCBI server (http://www.ncbi.nlm.nih.gov/blast). Nucleotide sequence was translated to protein with standard invertebrate mitochondrial code (codon 5) using DNA to Protein translation tool (http://insilico.ehu.es/translate/). Nucleotide and deduced protein sequences of COI genes of related organisms were retrieved from the GenBank (NCBI). The sequences were imported to BioEdit v.7.0.9. [14] and aligned using CLUSTALW [15]. Phylogenetic tree was constructed by the Neighbour-Joining (NJ) and Maximum Likelihood (ML) method based on nucleotide sequence of cytochrome oxidase subunit-1, using MEGA version 5.05 [16] and bootstrap analysis was carried out using 100 and 1,000 replicates with MEGA version 5.05. Kimura 2-parameter (K2P) model [17] was used to construct NJ and ML tree based on nucleotide sequence. Intraspecific and interspecific nucleotide sequence divergences under the K2P model were calculated in MEGA 5.05. Amino acid sequence divergence of CO1 protein based on Poisson model was calculated using MEGA 5.05. Further analysis of the nucleotide sequences of genus Abralia was carried out in DnaSP v.5 software [18].

3. Result
Morphological characters of A. andamanica include translucent body with reddish brown irregular spots and five pairs of appendages including a pair of tentacles (Fig. 1). Two rows of hooks are present on arms (12 to 14 hooks) for about three fourth of their length and on distal quarter two rows of suckers (up to 12 countable suckers) are present. Suckers are of compressed conical shape obliquely set on short stalks. Margin of the sucker is furnished with small teeth on the proximal side and large blunt teeth on the distal side. Ventro-lateral arms are provided with large membranous keel. Nearly conical mantle tapers gradually to the apex (Fig. 2a). Fins are rhomboidal in shape and reach about half way up the mantle. Lanceolate pen is provided with a strong median ridge (Fig. 2b). The
Partial COI amplicon of 582 bp encoding 194 amino acid residues could be recovered from *A. andamanica* (Fig. 3). Results of blastn and blastp confirmed the sequence to be of CO1. The obtained nucleotide and deduced amino acid sequences were deposited in GenBank database (GenBank ID: JX297202). Based on the similarity in BLAST results, nucleotide sequences of CO1 gene of squids belonging to various families of order Teuthida were downloaded from GenBank database. Phylogenetic relationship of *A. andamanica* to other squids of order Teuthida was established based on the nucleotide comparisons of CO1. Phylogenetic relationship of *A. andamanica* was virtually identical in NJ tree and ML. The NJ tree constructed to determine the relationship of *A. andamanica* within order Teuthida included 62 squids belonging to 16 families selected based on results of BLAST analysis and a group of three gastropods was selected to form the out-group (Fig. 4). Out of the 16 families considered for phylogenetic analysis, 13 formed single separate clade while cephalopods belonging to three families Cranchiidae, Pyroteuthiidae and Ommastrephidae were not aligned in a single clade indicating that overlap exists among Teuthida families with respect to CO1 nucleotide sequence. As expected *A. andamanica* was found to align along with the Enoploteuthidae family. Surprisingly, *A. andamanica* was found to be more closely related to *A. veranyi* (GenBank ID EU735394) [6] than to *A. andamanica* (GenBank ID HQ846076) sampled from China Sea [19].

Nucleotide frequencies of *A. andamanica* from Andaman Sea was calculated and compared with the nucleotide frequencies of *A. veranyi* and *A. andamanica* from China Sea (Table 1). In case of *A. andamanica* from Andaman Sea, C (%) was found to be slightly higher than the other two. A+T content of *A. andamanica* was found to be high (64.46%) which was in agreement with the A+T content of all squids considered for the phylogenetic analysis. Aminoacid frequencies were calculated for the CO1 protein sequences of *A. andamanica* from Andaman Sea, *A. andamanica* from China Sea and *A. veranyi*. The CO1 protein sequence of *A. veranyi* present in GenBank 168 database is the smallest (158 amino acid residues) in comparison to *A. andamanica* from Andaman Sea and China Sea and hence, in case of protein sequences, the region corresponding to the smallest sequence (158 amino acids) was considered for calculating amino acid frequencies of all three squids so as to get a uniform evaluation. Amino acid frequencies were found to be same for Asp (2.53%), Glu (2.53%), Phe (5.70%), His (1.90%), Leu (16.46%), Met (7.59%), Asn (5.06%), Pro (6.96%), Gln (1.90%), Arg (2.53%), Ser (9.50%), Thr (5.06%), Trp (2.53%) and Tyr (1.27%) in case of all three squids. In case of Ala and Gly, amino acid frequencies were found to be 6.96% and 9.49% respectively for both *A. andamanica* from Andaman Sea and China Sea, whereas for *A. veranyi* it was found to be 7.59% and 8.86% respectively. *A. andamanica* from Andaman Sea and *A. veranyi* exhibited frequencies of 6.39% and 5.69% for Ile and Val respectively which differ from Ile (6.96%) and Val (5.06%) frequencies of *A. andamanica* from China Sea.

K2P pairwise distance of *A. andamanica* from Andaman Sea, *A. andamanica* from China Sea and *A. veranyi* was calculated based on COI nucleotide sequences. Based on K2P distance values *A. andamanica* was found to be more closely related to *A. veranyi* (0.0815) than to *A. andamanica* from China Sea (0.1169) (Table 2). K2P distance for *A. veranyi* and *A. andamanica* from China Sea
Fig. 4: A bootstrapped neighbor-Joining tree based on K2P model, obtained using MEGA version 5.05 illustrating relationships between the nucleotide sequence of *A. andamanica* to the nucleotide sequences of previously reported CO1 from squids of order Teuthidae.
was calculated to be 0.1364. Overall mean K2P distance for all 62 CO1 nucleotide sequences considered for phylogenetic analysis was calculated to be 0.2242. Pairwise distances calculated for A. andamanica from Andaman Sea, A. andamanica from China Sea and A. vernayi based protein sequence using Poisson model is shown in Table 2. In case of protein sequence A. andamanica from Andaman Sea differed from A. vernayi and A. andamanica from China Sea by a distance value of 0.0063. A. vernayi was found to differ from A. andamanica from China Sea by a distance value of 0.0127. Nucleotide sequences of A. vernayi and A. andamanica from Andaman and China Sea were analyzed using DnaSP. The three sequences were identified as three haplotypes. Haplotype diversity, Hd was found to be 1.00 (SD: 0.272) and variance of haplotype diversity was found to be 0.0741. The number of polymorphic (segregating) sites, S was calculated to be 73. Total number of mutations, Eta was found to be 78. Nucleotide diversity, Pi for the three CO1 sequences was found to be 0.1057 (SD: 0.0308). Fu’s Fs statistics (2.807) [20] and Str [18] statistics (1.00) [21] based neutrality tests were conducted and probability that NHap = 3 was found to be 0.943 confirming the high divergence among the three nucleotide sequences.

4. Discussion:
Mitochondrial DNA based molecular markers are very informative when it comes to taxonomic identification and phylogenetic analysis of an organism. Sea surrounding Andaman Archipelago is largely unexplored and believed to hide vast biodiversity with enormous number of organisms waiting to be discovered. Molecular markers based on mitochondrial and genomic DNA could be used as effective tools for providing taxonomic and phylogenetic identities to these organisms. In the present study we considered CO1 gene for determining the phylogeny and genetic variation of A. andamanica captured from Andaman Sea. Identification of an organism based on barcoding requires the presence of very similar sequences in database [22]. Errors in identification of a specimen would lead to submission of wrong sequence and hence the role of traditional/morphological taxonomic studies becomes more important than ever. In the present study, specimen was identified by comparing with the description of original holotype submitted to Calcutta museum. A thorough morphological analysis was carried out and is briefly described in this paper.

BLAST analysis of nucleotide and deduced amino acid sequence of the amplicon obtained from A. andamanica confirmed it to be of CO1 superfamily. Phylogenetic tree based on nucleotide sequence was constructed by NJ method using K2P model. A. andamanica aligned with the other squids of family Enoploteuthidae to form a single clade. A. andamanica from Andaman Sea was found to be more closely related to A. vernayi than to A. andamanica from China Sea suggesting the existence of genetic divergence among the enope squids inhabiting Andaman Sea and China Sea. Out of the 16 families of order teuthidae considered for phylogenetic analysis overlapping was observed only for the members of family Cranchiidae, Pyroteuthidae and Ommastrephidae. Helicocranchia pfefferi and Cranchia scurba did not align with the main clade formed by family Cranchiidae. H. pfefferi showed close affinity with the clade formed by Illex argentinus, I. illecebrosus and I. Coinditi of family Ommastrephidae. C. scurba was found to be closely associated with Ancistrocheirus lesueurii (Ancistrocheiridae). Squids of family Ommastrephidae formed two clades, one formed by squids of genus Illex exhibiting affinity with H. pfefferi and the second clade formed of Todarodes pacificus and Ecleoteuthis luminosa displayed affinity to clade Loliginidae. Pyroteuthis margaritifera of family Pyroteuthidae was found to be closely associated with family Enoplotheuthidae while P. addolux (Pyroteuthidae) exhibited close association to Ancistrocheirus lesueurii. The close relationship of Pyroteuthidae to Ancistrocheiridae was also stated by Carlini and Graves [11]. One of the major observations in nucleotide based NJ tree, with regard to A. andamanica was its close affiliation with A. vernayi than to its counterpart from China. Results of the

Table 1. Comparison of nucleotide frequencies (CO1 gene) of A. andamanica from Andaman Sea to that of A. vernayi and A. andamanica from China Sea.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Place of sampling</th>
<th>T</th>
<th>C</th>
<th>A</th>
<th>G</th>
<th>G+C</th>
<th>A+T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abralia andamanica</td>
<td>Andaman Sea</td>
<td>36.4</td>
<td>19.4</td>
<td>29.0</td>
<td>15.1</td>
<td>34.54</td>
<td>65.46</td>
</tr>
<tr>
<td>Abralia andamanica</td>
<td>China Sea</td>
<td>36.9</td>
<td>17.3</td>
<td>29.7</td>
<td>16.1</td>
<td>33.96</td>
<td>66.04</td>
</tr>
<tr>
<td>Abralia andamanica</td>
<td>(Andaman Sea)</td>
<td>37.6</td>
<td>17.7</td>
<td>28.5</td>
<td>16.3</td>
<td>34.30</td>
<td>66.60</td>
</tr>
<tr>
<td>Abralia andamanica</td>
<td>(EU733594)</td>
<td>37.6</td>
<td>17.7</td>
<td>28.5</td>
<td>16.3</td>
<td>34.30</td>
<td>66.60</td>
</tr>
<tr>
<td>Abralia andamanica</td>
<td>(JX297202)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Pairwise distance computed for CO1 nucleotide sequences based on K2P parameter and protein sequences based on Poisson model.

<table>
<thead>
<tr>
<th>Species 1</th>
<th>Species 2</th>
<th>Nucleotide</th>
<th>SD</th>
<th>Protein</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abralia andamanica</td>
<td>Abralia andamanica</td>
<td>0.1169</td>
<td>0.0192</td>
<td>0.063</td>
<td>0.0062</td>
</tr>
<tr>
<td>(Andaman Sea)</td>
<td>(China Sea)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abralia andamanica</td>
<td>Abralia vernayi</td>
<td>0.0815</td>
<td>0.0148</td>
<td>0.063</td>
<td>0.0061</td>
</tr>
<tr>
<td>(Andaman Sea)</td>
<td>(China Sea)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abralia andamanica</td>
<td>Abralia vernayi</td>
<td>0.1364</td>
<td>0.0204</td>
<td>0.0127</td>
<td>0.0086</td>
</tr>
<tr>
<td>(China Sea)</td>
<td>(China Sea)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
K2P model based pairwise distance computed for CO1 nucleotide sequence of A. andamanica from Andaman Sea, A. andamanica from China Sea and A. veranyi confirmed that with respect to nucleotide sequence the squid in question is closely related to A. veranyi. Poisson model based pairwise distance computed for amino acid sequences of the three squids revealed that with regard to protein sequence A. andamanica from Andaman Sea is at equal distance from A. andamanica from China Sea and A. veranyi. Both nucleotide and protein based pairwise distances calculated revealed that the distance between A. andamanica from China Sea and A. veranyi was more than twice the distance between A. andamanica from Andaman Sea and A. veranyi specifying high genetic divergence between A. veranyi and A. andamanica from China Sea. The K2P pairwise distance value for Order Teuthida was just twice as large as the distance value for A. andamanica from Andaman Sea and China Sea was indicating that considerable genetic variation with regard to CO1 exists between the two. The COI gene has been shown to be among the most conserved protein coding genes in the mitochondrial genome of metazoans [23]. Difference in the protein sequence of A. andamanica from the two regions does signify the onset of divergent evolution in the species with respect to CO1 gene.

Statistical analysis of CO1 nucleotide sequences of A. veranyi, A. andamanica from Andaman Sea and A. andamanica from China Sea was carried out using DnaSP software. All three sequences were demarcated as different haplotypes by the program with very high probability. The results of neutrality tests indicate significant amount of variation between the three sequences. The most determining factor in population divergence should be attributed to geographic distance and environmental difference. These two factors can be considered responsible for the divergence in nucleotide sequence of A. andamanica representatives from Andaman Sea and China Sea. Geographically isolated populations are prone to higher rates of intraspecific divergence [24]. Different organisms have different rate of evolution with respect to mitochondrial and/or nuclear genome [25]. In comparison to other bilateral metazoans, molluscs have higher rate of evolution [26]. Taking all these assumptions into consideration, it could be suggested that gene flow restriction along with difference in environmental conditions have initiated genetic divergence among the two populations of A. andamanica inhabiting Andaman Sea and China Sea.

5. Conclusion

In the present study, mitochondrial CO1 gene based phylogenetic analysis of A. andamanica captured from Andaman Sea was carried out. Comparison of the enope squid in question with the CO1 nucleotide sequence of A. andamanica captured from China Sea revealed significant genetic variation between the two populations. An elaborate study with more specimens of A. andamanica obtained from larger geographical areas is required to get an exact picture of genetic divergence among their populations. NJ tree constructed based on nucleotide sequence of CO1 gene confirmed the phylogeny of A. andamanica as it grouped along with other squids of family Enoploteuthidae to form a single clade. Results of the study clearly illustrate the usefulness of CO1 gene in assessing the genetic and evolutionary history of a species.

6. Acknowledgments

The authors are grateful to the Director, Centre for Marine Living Resources and Ecology (CMLRE) and Ministry of Earth Sciences (MoES), Govt. of India for the research grant (MoES/10-MLR/2/2007) and scientific support for the work. Authors also thank Coordinator, National Centre for Aquatic and Animal Health (NCAAH) for scientific support.

7. Conflict of interest statement

We declare that we have no conflict of interest with any person, institute and/or any commercial identities regarding the manuscript.

8. References

Source of support: Nil; Conflict of interest: None declared