Assessment of Phylogenetic Inter-Relationships in Mud Crab Genus Scylla (Portunidae) Based on Mitochondrial DNA Sequence

Darlina Md. Naim[1,*], Hurul Adila-Aida Mohamad Rosly[2] and Siti Azizah Mohd. Nor[1,2]

[1] School of Biological Sciences, Universiti Sains Malaysia, Penang, Malaysia

[2] Centre for Marine and Coastal Studies, Universiti Sains Malaysia, Penang, Malaysia

Abstract

Mud crab of the genus *Scylla* possessed almost similar characteristics despite the established key identification which contributed to a slight confusion when inspecting them at a morphological level. Therefore, this phylogenetic study may give an insight to differentiate between species at a molecular level. This study examined 520 base pairs (bp) of the mitochondrial cytochrome *c* oxidase I (COI) gene from 60 individuals belonging to the genus *Scylla*, a group of mud crabs that inhabit over vast geographic areas ranging from south eastern and eastern Africa to South-east Asia and Indo-Pacific regions. The samples were taken from various locations throughout peninsular Malaysia, Sabah and Sarawak. The nucleotide sequences were subjected to phylogenetic analyses by using the neighbour – joining (NJ) and maximum parsimony (MP) methods. All two methods revealed the reciprocally monophyletic relationship between *Scylla olivacea*, *S. paramamosain*, *S. tranquebarica* and *S. serrata*. The clustering of all four species of *Scylla* samples into four different clades suggested that their genetic identity belongs to their respective species, thus strongly supporting their status as different species available throughout Malaysia. Overall, the present study was able to be a reference on the phylogenetic relationships and assessment of genetic structure of *Scylla* sp. in Malaysia.

Keywords: mud crab, Scylla, phylogenetics, COI

1. Introduction

Taxonomy is the foundation of traditional conservation practices [1, 2] and understanding the taxonomic details of a species is central to the development of successful management strategies for sustainable fisheries resources. Mitochondrial DNA (mtDNA) has been one of the most widely used molecular markers for phylogenetic and taxonomic studies in animals [3] due to its maternal mode of inheritance and mainly non-recombining nature [4]. Though mtDNA sequence data have proved valuable in determining phylogenetic relationships, there are considerable differences in the characteristics of different types of gene and it is crucial that the choice of gene is appropriate to the problem being tackled [5]. Cytochrome oxidase subunit I (COI) is the largest of the three mitochondria-encoded Cytochrome Oxidase subunits, and is one of the largest protein-coding genes in the metazoan mitochondrial genome which has been used as a target gene for a number of molecular phylogenetic and identification studies [6].

Mud crab of the genus Scylla belongs to the Portunidae family. They are also known as swimming crabs which characterized by their broad paddles of the flattened fifth pair of legs. The discoverer recognized the existence of only one species, Scylla serrata (Forskål, 1775), until Estampador's [7] revision made clear that there are more than one species exists. Numerous convincing arguments for the recognition of the species involved were made to minimize the confusion of the taxonomic nomenclature [8] until genetic studies have come up to support and correct the findings [9].

Several genetic methods using allozyme electrophoresis, mitochondrial genes and nuclear genes [10] havebeen carried out in attempt to justify species in this genus and to date, there are four distinct species wererecognized namely Scylla serrata, S. paramamosain, S. tranquebarica and S. olivacea [10]. However, the relationships among these species are still not clear. Therefore, it is imperative that data collected from COI gene sequence be employed for defining the phylogenetic relationships among species within the Scylla genus. In this current study, the sequence of COI gene has been employed to reappraise the taxonomic status and unravel the phylogenetic relationships among species within genus Scylla.

2. Materials and Methods

2.1. Sample collection

A total of 50 individuals were caught from eight different mangrove swamps across Malaysia, two within the east coast Peninsular (Terengganu, $n = 10$; Kelantan, $n = 11$), three within the west coast of Peninsular Malaysia (Kuala Perlis, $n = 5$; Perak, $n = 5$; Penang, $n = 5$), one within the south Peninsular (Johor, $n = 4$) and the other two within Sabah and Sarawak (Sarawak, $n = 6$; Sabah, $n = 4$) [Figure 1]. Mud crabs were caught using five to seven crab pots baited with fish laid at suitable mangrove area (spaced approximately 100 m apart) and lifted and re-baited every 24 hours for three successive days. More details on sampling method were described in Rosly et al. (in press). From each specimen, approximately 5-10 g muscle tissue was removed from a single claw and placed in 95% alcohol.

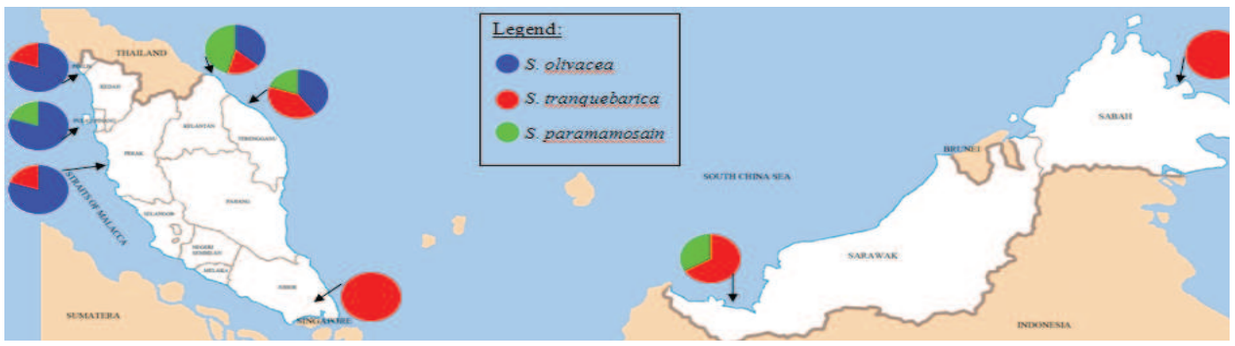

Figure 1. Sampling sites and species distribution in Malaysia. Each colour represents each species as indicated in the legend and the size of fractions corresponds to the frequency of samples obtained.

2.2. Total DNA extraction, polymerase chain reaction (PCR) and sequencing

Total genomic DNA from muscle tissue was extracted following the protocols from AquaGenomic Solution Kit (BioSyntech, USA). Polymerase chain reaction (PCR) was used to amplify the

target region of the COI gene in the mtDNA genome of all sampled individuals. Heavy strand primer Mtd-10 5′- T TGA TTT TTT GGT CAT CCA GAA GT - 3′ [11] and light strand primer C/N 2769 5′- TT AAG TCC TAG AAA ATG TTG RGG GA - 3′ [16] were used to amplify a 542 bp region of the COI region. Each PCR amplification was performed in a total volume of 20 µl of PCR mixture consisting of 10x PCR buffer, 2.5 mM dNTP mixture, 5U of *i-Taq* DNA polymerase, 25mM $MgCl_2$, 0.5 pmol of each forward and reverse primer and 1.6 µl (20 ng) of DNA from each sample. Thermal cycling conditions (on a G-STORM; Gene Technologies Ltd.) were 35 x [94°C for 30 s, 50°C for 30 s, 72°C for 1 min, 94°C for 3 min] and a final incubation at 72°C for 5 min.PCR products were then purified and sent to First Base Laboratories Sdn Bhd (1st BASE) for sequencing.

2.3. Alignment and sequence properties

The program MEGA ver. 4.0 [12] was used to visualize and align all sequences, including COI partial sequences of *S. serrata* obtained from Genbank (Genbank accession no: GU055497.1 – GU055506.1). The resulting sequences for each individual were then aligned using CLUSTAL W ver. 1.6 with default settings and were manually checked and trimmed in the BIOEDIT ver 7.0.9 sequence editing program; alignments were subsequently revised by eye in an effort to maximize positional homology. All aligned sequences were then imported into Basic Local Alignment Search Tool (BLAST) software to ensure the identity of the samples.

2.4. Phylogenetic analysis

The sequence characteristics of the COI region were calculated using MEGA ver. 4.0 [12]. The phylogenetic relationships among haplotypes were reconstructed using maximum parsimony (MP) and neighbour joining (NJ; [13]). We conducted MP and NJ phylogenetic analysis in MEGA ver. 4.0 [12]. Neighbour joining analysis was executed using non-parametric re-sampling procedure with 10,000 replicates. We performed MP analysis under the HKY+I model using the heuristic search algorithm with tree bisection reconnection (TBR) branch swapping and 100 random sequence addition replicates. Both trees reconstructions included 5 sequences of indigenous portunid from Genbank; *Portunus pelagicus* (Genebank accession no: GQ272560.1 – GQ272564.1). The genetic distance, G_{st} between populations was calculated using the Tamura-Nei distance [14] based on unequal base frequencies and unequal ratios of transitions to transversions (Ti:Tv) implemented in MEGA ver. 4.0 [12]. The frequency of each haplotype, haplotype diversity (*h*) (i.e., the probability that two randomly selected haplotypes are present in the sample) and the nucleotide diversity (π) within populations and geographical regions was estimated using ARLEQUIN ver. 3.11 [15].

3. Results and Discussions

Overall 50 partial sequences (see Supplementary Material for Genbank accession number for each sequence analysed) of 520 base pairs (bp) each of the mtDNA COI gene correspond to 50 individuals representing three *Scylla* species (*S. olivacea, S. paramamosain* and *S. tranquebarica*) and 10 COI partial sequences of *S. serrata* obtained from Genbank (Genbank accession no: GU055497.1 – GU055506.1) were analyzed. All the sequences show high sequence similarity (99-100%) to their respective species sequence in Genbank, indicate that misidentification of species does not occur in this study. The 520 bp *Scylla* sequences comprised 109 variable sites, of which 102 are parsimony informative sites (Table 1). This high number of parsimony-informative sites indicates that COI mtDNA is capable to be an informative locus candidate for phylogenetic studies [16] as well as for population differentiation and population genetic structure (Rosly et al., in press).

Taxa	n	*N*hap	G+C (%)	Number of sites				Ti	Tv
				Variable	Conserved	Parsim-Info	Multiple substitution		
S. olivacea	20	16	42.44	20	500	11	21	19	2
S. tranquebarica	20	4	19.40	6	514	5	6	5	1
S. paramamosain	10	6	27.04	6	514	0	6	6	0
S. serrata	10	10	24.16	16	504	0	16	14	2

Table 1. Summary of number of haplotypes, nucleotide diversity (π) and haplotype diversity (h) with standard deviation (SD), nucleotide composition, transition, transversion and multiple substitution dataset for each species implemented in ARLEQUIN ver. 3.11 software [25]. The number of sites was generated by using MEGA ver. 4 [17]. n = sample size; Ti = transition; Tv = transversion.

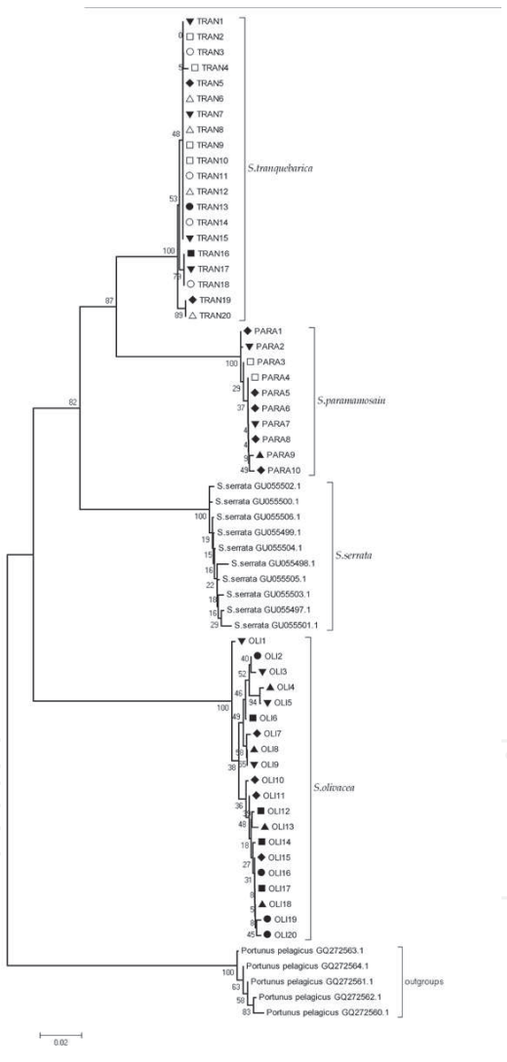

Figure 2. Neighbour-joining (NJ) phylogram showing the relationships among cytochrome c oxidase I (COI) sequences of Scylla species and 5 outgroups of genus Portunus analyzed in the present study. The number at each node represents the bootstrap percentage value based on 10000 pseudoreplications for the neighbor-joining (NJ) analysis generated using MEGA ver. 4.0 [17] software.

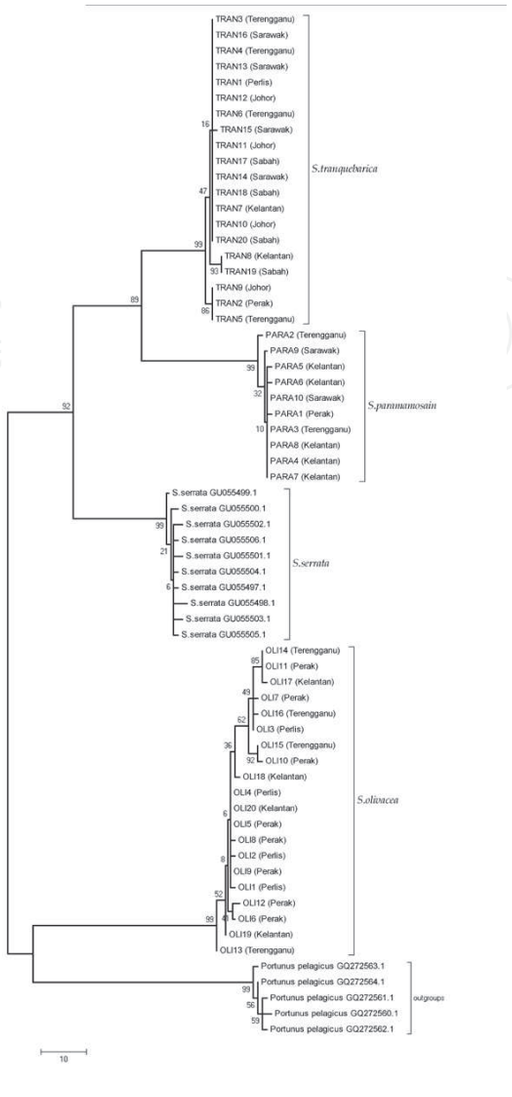

Figure 3. Maximum Parsimony (MP) phylogram showing the relationships among cytochrome c oxidase I (COI) sequences of Scylla species and 5 outgroups of genus Portunus analyzed in the present study. The number at each node represents the bootstrap percentage value based on 1000 replicates and applies the close-neighbour interchange (CNI) using the initial tree by random addition of 100 replicates for the maximum parsimony (MP) analysis generated using MEGA ver. 4.0 [17] software.

This study has highlighted the monophyly pattern of mud crab genus *Scylla* as evident by the consistent of four major clades in both phylogenetic trees (Figures 2 and 3). All samples were clustered according to their respective species; *S. tranquebarica, S. paramamosain, S. serrata* and *S. olivacea* with both trees showed that *S. tranquebarica* branched the earliest from all the species analysed in this study (Figures 2 and 3).The generic relationships revealed by our neighbour-joining (NJ) and maximum parsimony (MP) tree are highly in agreement with the previous molecular phylogeny study of *Scylla* based on COI sequence data [10]. However, our current results provide new insights into the classifications of the genus *Scylla* and presents additional evident regarding their inter-relationships that were not included and discussed in previous studies.

The construction of phylogenetic trees using NJ and MP analyses produced identical tree topologies with strong levels of support for all nodes (Figures 2 and 3). Additionally, clear relationships among mud crab species was observed in both trees and supports the placement of outgroup taxa, *Portunus pelagicus* at the base of both trees (Figures 2 and 3). The relative efficiencies of the NJ and MP in obtaining the correct topology for phylogenetic inference were studied by computer simulation. The NJ method gives a correct topology even when the distance measures used are not unbiased estimators of nucleotide substitutions, while for the MP method, both the weighted and unweighted parsimony are generally less efficient than the NJ method even in the case where the MP method gives a consistent tree. However, the NJ and MP analysis in our study returned a similar tree topology (see Figures 1 and 2), as demonstrated by several other studies at different taxa. Thus, the same tree topology (although different bootstrap values) that demonstrated by both of the phylogenetic trees has highly support the inter-taxa relationship among mud crabs genus *Scylla* sampled from Malaysia.

The relatively deep gene trees clearly show a large divergence between mud crab genus *Scylla* analysed in this study (Figures 1 and 2). This was further corroborated by the pairwise genetic distance among the four species of *Scylla* which ranges from *Gst* = *0*.086-0.198 (mean±SD; 0.146±0.261) for COI gene (Table 2). Other studies, particularly in marine fishes also show a large genetic distance between species. The intraspecific and interspecific differentiation in various crustacean taxa has been reported to be ranges at *Gst* = 0.00-1.47 and *Gst* = 5.2-31.6 respectively. Accepting this threshold as valid, the genetic distances between four species of mud crab in our study falls within the range of interspecific differentiation of crustacean taxa, thus further supported the study by Fuseya and Watanabe [8] which reported that there exist at least three distinct species within the genus *Scylla* (see e.g. [10, 12]).

Taxa	S. olivacea	S. tranquebarica	S. paramamosain	S. serrata
S. olivacea				
S. tranquebarica	0.165			
S. paramamosain	0.198	0.086		
S. serrata	0.181	0.114	0.131	

Table 2. Pairwise Tamura-Nei genetic distances (Gst) for COI gene sequence among Scylla.

In conclusion, in this study we were able to provide useful insights into phylogenetic relationships and the genetic identity of *Scylla* species from mangrove areas in Malaysia. However, further studies using larger samples from other areas of its geographical distribution, sequence data from other mtDNA regions, and information based on nuclear DNA markers are required to support our findings.

4. Acknowledgements

We thank the Fisheries Research Institute of Penang for their assistance in sample collection. This project was funded by Universiti Sains Malaysia through research grant 304/PBIOLOGI/6311003.

5. References

[1] Avise, J.C., 1989. A role for molecular genetics in the recognition and conservation of endangered species. Trend in Ecology and Evolution, 4: p. 279-281.

[2] Brophy, T.R., 2004. Geographic variation and systematic in the South-east Asian turtles of the genus (Testudines: Bataguridae). Faculty Pulications and Presentations. Paper 58.

[3] Avise, J.C., 2004. Molecular markers, natural history, and evolution. 2nd Edition, Sinauer, Sunderland, MA.

[4] Avise, J.C., 1994. Molecular Markers, Natural History and Evolution. Chapman and Hall, Inc., New York.

[5] Sunnucks, P., 2000. Efficient genetic markers for population biology. Trend in Ecology and Evolution 15: p. 199-203

[6] Hebert, P.D.N., A, Cywinska, S.L. Ball, et al. 2003. Biological identifications through DNA barcodes. Proceding of Royal Society London Series B, 270: p. 313-321.

[7] Estampador, E.P., 1949. Studies on Scylla (Crustacea: Portunidae). I. A revision of the genus. Philippine Journal of Science 78: p. 95–108

[8] Fuseya, R, S. Watanabe, 1996. Genetic variability in the mud crab genus *Scylla* (Brachyura: Portunidae). Fisheries Science 62(5): p. 705-709

[9] Keenan, C. P., D. L. Mann, S. Lavery, P. Davie, 1995. Genetic relationship, morphological identification and taxonomy of mangrove crabs, genus *Scylla*, from throughout the Indo-Pacific. ACIAR Project Report, QDPI, Brisbane.

[10] Keenan, C.P., 1999. Aquaculture of the Mud Crab, Genus Scylla – Past, present and future. In: Keenan CP, Blackshaw A (eds) Mud crab aquaculture and biology. Proceedings of an international scientific forum held in Darwin, Australia, 21-24 April 1997. ACIAR Proceedings No. 78, Ferguson and Co., Brisbane Australia, p. 9-13

[11] Tamura, K, J. Dudley, M. Nei, S. Kumar, 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution, 24: p. 1596-1599.

[12] Thompson, J.D., D.G. Higgins, T.J. Gibson, 1994. "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice". Nucleic Acids Research, 22: p. 4673-4680

[13] Hall, T.A., 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Ser 41: p. 95-98

[14] Tamura, K., M. Nei, 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution, 10: p. 512-526

[15] Excoffier, L., G. Laval, S. Schneider, 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary and Bioinformatics Online, 1: p. 47-50

[16] Kamarudin, K.R., R.M. Aisyah, H. U. Ridzwan, U. Gires, A. Hajar Fauzan, A. Mohd Hanafi, I. Mohd Yaman, 2011. Molecular phylogeny of *Holothuria (Mertensiothuria) leucopilota* (Brandt 1835) as inferred from cytochrome C oxidase I mitochondrial DNA gene sequences. Sains Malaysiana, 40: p. 125-133.