Integrating geometric morphometrics and DNA barcoding: A consolidated taxonomic tool in identifying selected cryptic Carangid species encountered in the Indian Ocean.

Lahiru Pandi Gamage¹, Naveen Ranasinghe², Dona Munasinghe¹, and Tsung Lee²

¹University of Ruhuna
²National Chung Hsing University

May 22, 2023

Abstract
Geometric morphometrics and mitochondrial cytochrome oxidase I (COI) DNA barcoding are crucial for identifying closely related cryptic carangid species. We integrated both taxonomic methods for promising identification within selected carangid groups, trevallies (Turrum coeruleopinnatum, Platycaranx malabaricus, and Atropus hedlandensis) and scads (Selar crumenophthalmus, Selar boops, and Atule mate). Despite a plethora of carangid barcode data, the knowledge bridge on carangid evolutionary footprints provides limited information on their origin, evolution, and distribution. Procrustes-defined data derived from shape differences between species and between morphs within species were analyzed by principal component analysis (PCA) and canonical variate analysis (CVA) (P < 0.0001), and were independent of intraspecific variation. Geometric morphometric clustering was evaluated using mtDNA COI barcoding, and each morph/species cluster was found to be compatible with the corresponding species. Average Kimura 2-Parameter (K2P) divergences were obtained in accordance with taxonomic hierarchy and were consistent with the 2% species delimitation: conspecific, congeneric, confamilial divergences were 0.28%, 4.50%, and 11.90% respectively and intraspecific and interspecific divergences were in the ranges (0.00-0.60)% and (2.10-18.70)% respectively. The greatest divergence was observed between the Indian and Indo-Australian Archipelago (IAA) individuals, whereas the lowest divergence was observed between the common ancestral cluster and IAA individuals. However, both consolidated taxonomic approaches provided a clear resolution of the selected carangid species over cryptic speciation. The origin of the carangid ancestor and its centered distribution in the IAA region are well described by regional characteristic divergences and are further explained by the center of origin and overlap hypotheses.
Department of Zoology,
University of Ruhuna, Sri Lanka
mail address: madushanlahirugamage@gmail.com

ABSTRACT

Geometric morphometrics and mitochondrial cytochrome oxidase I (COI) DNA barcoding are crucial for identifying closely related cryptic carangid species. We integrated both taxonomic methods for promising identification within selected carangid groups, trevallies (Turrum coerulopinnatum, Platycaranx malabaricus, and Atropus hedlandensis) and scads (Selar crumenophthalmus, Selar boops, and Atule mate). Despite a plethora of carangid barcode data, the knowledge bridge on carangid evolutionary footprints provides limited information on their origin, evolution, and distribution. Procrustes-defined data derived from shape differences between species and between morphs within species were analyzed by principal component analysis (PCA) and canonical variate analysis (CVA) (P < 0.0001), and were independent of intraspecific variation. Geometric morphometric clustering was evaluated using mtDNA COI barcoding, and each morph/species cluster was found to be compatible with the corresponding species. Average Kimura 2-Parameter (K2P) divergences were obtained in accordance with taxonomic hierarchy and were consistent with the 2% species delimitation: conspecific, congeneric, confamilial divergences were 0.28%, 4.50%, and 11.90% respectively and intraspecific and interspecific divergences were in the ranges (0.00-0.60)% and (2.10-18.70)% respectively. The greatest divergence was observed between the Indian and Indo-Australian Archipelago (IAA) individuals, whereas the lowest divergence was observed between the common ancestral cluster and IAA individuals. However, both consolidated taxonomic approaches provided a clear resolution of the selected carangid species over cryptic speciation. The origin of the carangid ancestor and its centered distribution in the IAA region are well described by regional characteristic divergences and are further explained by the center of origin and overlap hypotheses.

Keywords: Carangid ancestor, COI divergence, Evolutionary footprints, Indo-Australian Archipelago (IAA), Kimura 2-Parameter (K2P), Land marking

1. INTRODUCTION

The family Carangidae encompasses four subfamilies with 30-32 genera and approximately 140 species, which are commonly identified as jacks, pompanos, and scads (Dameru et al., 2018). Owing to their high commercial demand, carangid fish are considered potential candidates in the field of aquaculture, including the food and ornamental fish industries, and there is now global attention toward large-scale culturing practices (Kappen et al., 2018; Rombenso et al., 2016).

Biodiversity monitoring is based on taxonomic studies that rely on accurate species identification. Biological and ecological data collected from precise sources have been linked to aquaculture, conservation and population management practices (Jordán et al., 2008; Joshi & Deshpande, 2011). Ambiguous taxonomic studies derived from misidentified species would affect not only the relevant species but also the ecosystem through improper conservation efforts and resource applications (Garcia-Vazquez et al., 2012). Phenotypic (morphometric and meristic) characteristics are used for conventional species identification (Cadрин, 2000). Previous studies have restricted themselves to morphological features such as osteology and, scale arrangements for carangid species delimitation (Suzuki, 1962). However, these measures are not always reliable under certain conditions, such as cryptic speciation (Murta, 2000), which implies that monophyletic autapomorphies have not been observed among individuals previously described in the genus Carangoides (Kimura et al., 2022).

The characteristic intraspecific and interspecific cryptic diversity within the family Carangidae creates taxonomic ambiguities over phenotypic characters (Mat Jaafar et al., 2012). Cryptic speciation within the family arises from either developmental (genetic) or environmental backgrounds. The effects of environmental factors on the phenotypic plasticity of natural populations of several carangid fish species have been documented; for example, hydrological conditions, such as water velocity, may affect body shape variation, including head depth, head length, caudal peduncle depth, and body depth (Pakkasmaa & Piironen, 2000;
Saleh et al., 2017). Their wide distribution over microhabitats, including estuaries, and high salinity tolerance increases the possibility of cryptic speciation through phenotypic variations (Tran Vinh Phuong et al., 2015).

Due to the seasonal abundance and lack of knowledge regarding carangid species identification over taxonomic ambiguities, commercial fraud has been observed in local fish landing sites in Sri Lanka. Because the phenotypic plasticity is pronounced among three species of scads; Purse-eye scad: *Selar crumenophthalmus* (Bloch, 1793), Oxeye scad:*Selar boops* (Cuvier, 1833), and Yellow-tail scad: *Atule mate* (Cuvier, 1833) and trevallies; Bumpnose trevally:*Carangoides hedlandensis* (Whitley, 1934), Malabar trevally:*Carangoides malabaricus* (Bloch and Schneider, 1801), and Coastal trevally: *Carangoides coerulopinnatus* (Rüppell, 1830), they are often sold by mixing them together in local fish markets. However, species previously studied under the Genus *Carangoides* have been revised because of their paraphyletic behavior in recent molecular phylogenetic studies. Hence, the three *Carangoides* species in the current study must be recognized under isolated species as *Atropus hedlandensis*, *Platycaranx malabaricus*, and *Turrum coeruleopinnatum* respectively (Kimura et al., 2022). Therefore, cryptic speciation among distant taxa was a matter of concern for taxonomic clarification in this study.

As an alternative solution to overcome limitations in closely related species identification such as higher phenotypic variability from larvae to adults or phenotypic alterations under fragmented, damaged, processed conditions and other conditions, the “DNA Barcode Concept” was introduced (Hebert et al., 2003; Pegg et al., 2006). The partial mitochondrial cytochrome oxidase I (COI) gene region (~650 bp) fulfills three criteria for a global standard barcode gene: species-level sequence variability and divergence even in cryptic species differentiation, a sequence short enough to be extracted and amplified, and conserved flanking sites for universal primer amplification in a wide range of taxa (Gómez et al., 2007; Hubert et al., 2012; Pfenninger et al., 2007; Zemlak et al., 2009). The contribution of the COI gene in reflecting phylogenetic/geographic relationships is more informative for carangid species identification than phenotypic plasticity (Kimura et al., 2022; Lu et al., 2011; Ptaszynska et al., 2012).

The shape of the carangid body is highly variable, ranging from elongated and fusiform to deep and strongly compressed (Mansor et al., 1998; Smith-Vaniz, 1999; Honebrink, 2000). The selected carangid fish species in the current study represent two distinct body shapes: deeper trevallies and shallower scads. However, within these two major clusters, sub-clusters of lower taxonomic levels were expected based on the geometric morphometric analysis. Landmark-based image analysis is an effective tool for differentiating species with similar shapes, including carangids (Costa & Cataudella, 2007; Gamage et al., 2022; Richtsmeier et al., 2002; Santos & Quilang, 2012; Zelditch et al., 2004).

The current study aimed to measure the validity and the possibility of considering two consolidated methods, geometric morphometric and COI barcoding of selected carangid fish species against their phenotypic confusion and provide supportive background information for the generic revision of the genus *Carangoides*. Furthermore, the phylogeographic information elicited from this study will be used to describe the evolutionary footprints of carangid fishes.

2. MATERIALS AND METHODS

2.1 Sample collection and morphological identification

Putative carangid fish species were collected from several fish markets related to their landing sites and along the southern coastal belt of Sri Lanka, representing the three districts of Galle, Matara, and Hambantota, from October to December 2021 (Figure 1). All fish specimens were identified as species based on their morphological characteristics according to the FAO-Fisheries Identification Guidelines (Fischer and Whitehead, 1974) and other published keys for carangid fish (Annie and Albert, 2009; Mansor et al., 1998). To avoid bias caused by body asymmetry, the left side of each specimen was photographed immediately and subjected to preliminary morphometric analysis.

2.2 Morphometric data analysis

Total weight (in grams) and length (in centimeters), including total length, fork length, standard length, scute length, and body depth, were measured to assess the well-being and degree of fatness of the fish.
by calculating the condition factor (Zelditch et al., 2004). Sixteen (16) homologous landmarks (Figure 2) were selected for comparison, and each landmark was digitized using the computer programs: tpsUtil v1.4 and tpsDig2 v1.1 for which the x and y coordinates were generated and saved as a ‘tps’ file (Rohlf, 2015). Non-shape variations, including errors associated with rotation, scale, and translation were removed using the generalized procrustes analysis (GPA) in the programs CoordGen6f and MorphoJ (Klingenberg, 2011). A multivariate analysis of variance (MANOVA), regression was performed to determine whether there was a significant influence of fish body size on body shape variation, considering the dependent variable (body shape) and independent variable (centroid body size: mean position of all coordinates) imported into MorphoJ software. Statistical significance of the regression was tested using permutation tests (10,000 replicates) against the null hypothesis of independence. Procrustes superimposed/standardized coordinate data were imported into both SPSS v17.0 (SPSS Inc., 2008) statistical software and MorphoJ to summarize the canonical variates analysis (CVA) data, illustrating the carangid body shape variations.

2.3 Sample processing and DNA extraction

The epaxial muscle behind the operculum on the right side of the fish was considered for DNA extraction and samples were submerged in absolute ethanol and stored at -20 °C. After morphometric categorization of the summarized CVA, a minimum of three individuals from each morph/cluster, regardless of the site, were randomly selected for molecular analysis. Muscle tissues of about 25 mg were utilized for DNA extraction using the QIAGEN® DNeasy® Blood and Tissue Kit, (Cat. ID 69504, North Rhine Westphalia, Germany) according to the manufacturer’s protocol.

2.4 PCR amplification and DNA sequencing

Partial amplification of an approximately 652 bp fragment of the mitochondrial COI (cytochrome oxidase subunit I) gene region was performed using two universal fish primer pairs, FishF1/R1 and FishF2/R2 (Table 1). Polymerase chain reaction (PCR) was carried out in 25 μl reaction volumes comprising of 10 μl of 2× GoTaq® Green Master Mix; Promega, (Cat. ID MT723, Woods Hollow, Wisconsin, USA) (contained PCR buffer, MgCl₂ Solution, dNTPs, and Taq Polymerase all together), 0.5 μl of each primer (10 μmol), 11 μl of nuclease-free water and 3 μl of DNA template (approximately 200 ng/μl). The thermal cycle for the reaction consisted of 2 minutes of the initial denaturing phase at 94 °C, further 30 seconds at 94 °C, 30 seconds of primer annealing phase at 56 °C-58 degC, 3 minutes of elongation phase at 72 degC and a final extension phase at 72 degC for 7 minutes. PCR amplicons that produced clear bands on agarose gel electrophoresis were sent to Macrogen Pvt. Ltd, South Korea, and the Department of Molecular Biology & Biotechnology, University of Peradeniya, Sri Lanka for purification and next-generation/sanger sequencing.

2.5 Sequence analysis

2.5.1 Determination of species status of selected specimens

Consensus sequences with an average of 652 bp amplicon size were obtained through the bidirectional alignment of the acquired sequences using the ClustalW option available in the BioEdit: sequence alignment editor software package (Hall, 1999). The maximum compatibility and similarity of the consensus sequences of each morph were compared with the previously published reference sequences using the basic local alignment search tool (BLASTn) option in the National Center for Biotechnology Information (NCBI) database (www.ncbi.nlm.nih.gov/Genbank) (Table 2).

2.5.2 Publication of DNA barcoding data

After clarifying the species status of the collected samples, aligned sequences were converted into amino acid sequences using MEGA v11 software (Tamura et al., 2021) to check for any stop codons. Consensus DNA sequences were annotated using the GenBank database and the barcode of life data system (BOLD) (www.barcodinglife.com) and then deposited under the relevant accession numbers and BOLD IDs (Table 2).

2.5.3 Determination of phylogenetic/geographic status of identified species
Additional COI sequences for each species (representing all geographic localities) were acquired from the GenBank and BOLD databases for genetic distance comparison using the MEGA v11 software. Using the pairwise Kimura-2-Parameter (K2P) as the substitution distance model (Kimura, 1980), maximum likelihood trees (Saitou, 1987) were constructed with 1000 bootstrap replicates to estimate the strength of the phylogenetic/geographic relationships of the sample analyzed. Skipjack tuna (Katsuwonus pelamis, Scombridae) (Linnaeus, 1758) and milkfish (Chanos chanos, Chanidae) (Forsskal, 1775) were considered the out-group taxa in the divergence analysis for better resolution of the results. For better visualization, all constructed trees were modified using iTOL v5 software (https://itol.embl.de) (Letunic & Bork, 2021). Based on the nucleotide divergence of the aligned COI sequences of the study sample, a median-joining network was constructed by the PopART v1.7 software to determine the interrelationships among the haplotypes of the corresponding species (Leigh & Bryant, 2015) associated with the Indo-Australian Archipelago (IAA). Tajima’s D statistic was performed on the COI gene to determine the pattern of mutations in the IAA region.

3. RESULTS
3.1 Initial phenotypic identification and biometric analysis
From the taxonomic guide-based assessment, 360 individuals were collected as scads (150 from A. mate, 90 from S. crumenophthalmus, 116 from S. boops and four unconfirmed individuals: SUNK. The four unconfirmed individuals with phenotypic confusion between S. boops and S. crumenophthalmus were labelled SUNK1, SUNK2, SUNK3, and SUNK4 for further clarification. Ninety individuals were collected as trevallies and initially identified as T. coeruleopinnatum (27), P. malabaricus (30), and A. hedlandensis (30), whereas three unconfirmed individuals (TUNK) with phenotypic entanglement between P. malabaricus and T. coeruleopinnatum were labelled as TUNK1, TUNK2, and TUNK3. The average morphometric values are indicated in Table 3; both biometric approaches, regression of length-weight relationship and approximately the value around 1.000 on relative condition factor (Kn), eliminated the sample bias of body shape variation analysis in the current study (P > 0.05).

3.2 Geometric morphometric analysis
Generalized procrustes superimposition generated a new set of coordinates for each individual in the sample. Principal component analysis (PCA) of the procrustes-transformed data subsequently resulted in 28 significant (α = 0.05) principal components (PCs), of which the first two explained 89.3% of the total variation (PC1 = 73.1%, PC2 = 16.2%). The results of canonical variate analysis (CVA) yielded two significant (α = 0.05) canonical variates (CVs) out of five CVs, accounting for CV1 = 84.9% and CV2 = 13% respectively and 97.9% of the total variation (P < 0.0001). According to the graphical representation of the above major PCs and CVs, the samples clustered into species groups and were independent of intraspecific variation (Figure 3). However, the phenotypic plasticity among individuals of S. boops and S. crumenophthalmus tends to make characteristically close associations when clustering. Procrustes distances followed by the Mahalanobis grouping indicated that the lowest resolution distances were observed among individuals of the trevallies (0.043), and between individuals of S. crumenophthalmus and S. boops (0.035) (Table 4). Even if this provides statistical information on their phenotypic mutualism, both CVA and PCA graphically summarized individuals into relevant species groups (P < 0.0001). Individuals previously labelled as SUNK1, SUNK2, and SUNK3 belonged to the species group S. crumenophthalmus whereas SUNK4 belonged to the species group S. boops. All individuals labelled as TUNK1, TUNK2, TUNK3, and TUNK4 were clustered with individuals of T. coeruleopinnatum.

3.3 Mitochondrial COI gene sequence analysis
3.3.1 Confirmation of the species designation through DNA barcoding
Partial mitochondrial cytochrome oxidase I gene (COI) barcodes were recovered from randomly selected individuals from each species group/morph cluster, followed by geometric morphometrics and phenotypically unconfirmed individuals under SUNK, TUNK which were especially considered for barcoding. Thus, 22 COI barcodes with an average length of 652 bp were obtained without stop codons, insertions/deletions or
heterozygous sites, removing the bias that all amplified sequences constitute functional mitochondrial COI barcoding sequences (Mat Jaafar et al., 2012). All the barcode sequences were compatible with the previously published reference sequences available in GenBank (https://blast.ncbi.nlm.nih.gov) and confirmed the maximum compatibility (more than 99%) with the corresponding species status (Table 2). Both consolidated morphometric and molecular approaches undoubtedly explain why individuals previously labelled as SUNK1, SUNK2, and SUNK3 were confirmed to be S. crumenophthalmus and SUNK4 as S. boops. Correspondingly, individuals under TUNK were confirmed to be T. coeruleopinnatum.

3.3.2 Nucleotide divergence assessment

A total of 110 partial mitochondrial COI barcoding sequences of the corresponding species, 22 from the current study and 88 from the BOLD database and GenBank, were aligned and nucleotide divergences (K2P) were calculated. The average nucleotide divergence in the relevant taxonomic strata is listed in Table 5. As expected, the average interspecific genetic divergence was greater than the intraspecific divergence. Both the average intraspecific divergence range of (0.00-0.60)% (<2% speciation criteria) and interspecific divergence range of (2.10-18.70)% (>2% speciation criteria) are within the standards of species delimitation (Pereira et al., 2013). In the phylogenetic analyses, the maximum likelihood tree formed monophyletic clusters of trevallies and scads (Figure 5) illustrating the utility of COI barcoding in providing species-level resolution.

Individuals concentrated in the Indo-Australian Archipelago (IAA) region had the greatest genetic divergence compared with those in the peripheral Indian region (Table 7). The phylogeography of each species of both trevallies (Figure 6) and scads (Figure 7) clearly shows the isolation of the regional IAA and Indian clusters; further, the genetic divergence of IAA is lower with the common ancestral node and greater with the peripheral Indian group and vice versa (Table 7). The information collected from localized genomic variations provides clues regarding the evolutionary lineage of carangid fish. The distinctive mutation rates among the haplotypes of the selected carangid species encountered in the IAA region were clearly visualized using region-specific haplotype networking (Figure 8). Tajima’s D statistics for the COI data extracted from the IAA region were negative for all six carangid species studied (P > 0.05) and were obtained for A. mate (-0.5192), S. boops (-2.6663), S. crumenophthalmus (-2.7124), A. hedlandensis (-0.5550), P. malabaricus (-1.5444), and T. coeruleopinnatum (-0.9458).

4. DISCUSSION

4.1 Criticality of phenotypic separation of carangids used in the study

Field identification of S. crumenophthalmus and S. boops was not entirely satisfactory based on morphological features alone. This taxonomic ambiguity was caused by the infeasibility of separating these two morph species from a group of fish using only the following characteristics. Both arched and straight-line sections are equal in S. boops but the arched sections are longer and the curvatures are lower in S. crumenophthalmus. The scutes start at the 1st ray of 2nd dorsal fin and there are approximately 43-46 scutes in S. boops, whereas 29-42 scutes start at the 8th ray of 2nd dorsal fin in S. crumenophthalmus, as explained by (Smith-Vaniz, 1999). The average interspecific divergence between them ranged from 2.1% to 4.7% (approximately associated with the 2% conceptual speciation value) and explains the possibility of phenotypic hybridization followed by successful interbreeding, increasing the phenotypic ambiguity further. Therefore, subtle observations were also required for field observations of the other studied carangids. However, integrating both geometric morphometrics and COI barcoding emphasizes their potential as taxonomic tools for the identification of closely related carangid species, including the phenotypic ambiguity observed among S. crumenophthalmus, S. boops and T. coeruleopinnatum and P. malabaricus.

4.2 Geometric morphometrics in carangid body shape

Geometric morphometrics analysis yielded two sets of principal components (PCs) and canonical variates (CVs). PC1 and CV1 fluctuated on the X-axis, representing carangid body shape variations, whereas PC2 and CV2 fluctuated on the Y-axis, representing their body shape differences. Variations in the anterior
insertion of the dorsal fin were demonstrated by PC1 and CV1; for PC1, the more negative extremity was associated with shallower body depth, whereas a more positive extremity was associated with deeper body depth and for CV1 vice versa. Furthermore, disparities in the anterior tip of the snout were demonstrated by PC2 and CV2 (Figure 4); for PC2, closer to the anterior extreme of the orbit, more negative extremities were expected, whereas the farther the anterior extreme correlated with more positive extremities and for CV2 and vice versa.

PC1 and CV1 separated the three species of both scads and trevallies. This separation implied that *A. mate*, *S. crumenophthalmus*, *S. boops*, *P. malabaricus*, *A. hedlandensis*, and *T. coeruleopinnatum* achieved greater body depths. In contrast, PC2 and CV2 separated *A. mate* from the other two *Selar* species. This implies that the anterior tip of the snout of *A. mate* is farther from the anterior extreme of the eye orbit than in the other two species. Correspondingly, the anterior tip of the snout of *T. coeruleopinnatum* was closer to the anterior extreme of the orbit of the eye than that in the other two species. The steeper head profile of the carangid fish contributed to the increased distance between the anterior tip of the snout and the anterior extreme of the eye orbit. This is more clearly observed in *A. hedlandensis* than in the other two trevallies, and the characteristic head bump is another taxonomic feature that confirms the species status (Santos & Torres, 2020; Smith-Vaniz, 1999).

4.3 Utility of COI barcoding in carangid species identification

The mean Kimura-2-Parametric (K2P) intraspecific divergence (0.00–0.60%) obtained was compatible with the previously reported data for marine fish including carangid species (0.24–0.39)% (Gamage et al., 2022; Mat Jaafar et al., 2012; Zhang & Hanner, 2011) and freshwater fish species (0.30 – 0.45)% (Hubert et al., 2008). As described by Mat Jaafar et al. (2012), the average interspecific genetic divergence within the family Carangidae increased to 18.70% and this implies that increasing divergence was obtained with conspecific, congeneric and confamilial boundaries (Ward et al., 2005). Both mitochondrial COI barcoding divergences and phylogenetic relationships were compatible with the 2% species delimitation, which is broadly consistent with previous studies and further revealed its utility as a taxonomic tool consolidating geometric morphometrics.

Phylogenetic isolation of the three trevallies of the study into completely different subclades further described their paraphyletic behavior from the genus *Carangoides*. The average congeneric divergences are summarized in Table 6, the three genera *Atropus*, *Turrum*, and *Platycaranx*, which were previously described under *Carangoides*, demonstrated significant divergence from each other (*P* < 0.05). This provides background information for the generic revisions described by Kimura et al. (2022). Hence, cryptic speciation observed in distant taxa beyond the congeneric level is a matter of concern.

4.4 Phylogeography revealing the origin of carangids and regional genetic differences

According to the phylogeography of the respective species, revealed that the common ancestral cluster was encountered in the Indo-Malay Archipelago (IMA). The gene flow of carangid fish from the common ancestor and the rise in cryptic species can be explained by past geological events. This is evidenced by the separation of two distinct geographically isolated branches from the common ancestral node and the remarkable genetic divergence between them. The common ancestral cluster has the lowest genetic divergence with the individuals studied in the IAA region, revealing that they are the closest descendants of the common ancestor, and the greatest divergence with the individuals in the Indian Ocean, revealing that they are vicariants resulting from past geological isolation.

4.5 Evolution of carangid fish along with prehistoric events

The evolutionary footprints of their fossilized remains extended into the early Paleocene epoch that lasted from approximately 60 million years ago (Rabosky et al., 2018; Uyeno & Suda, 1991) and it is assumed that carangid ancestors evidently originated in the IAA region during this period, where the present centered distribution occurred (http://www.fishbase.org) (Figure 9) (Mat Jaafar et al., 2012). Land mass rearrangements have had an astonishing effect on animal dispersal in the IAA region. At the origin of the carangid
ancestor, Sundaland was a continental promontory and the origin of the center allowed ancestral descendants to disperse along with the water current from the Pacific to the Indian region. From the Late Eocene to the Early Pliocene, the gradual emergence of the volcanic island chain and the northward movement of Australia led to a gradual increase in shallow sea areas and the closure of the deep-sea bridge between peripheral areas. The center of the overlap was pronounced during the Late Miocene and Early Pliocene (10–5 MYA), resulting in geologically isolated gene pools that evolved independently in the IAA and peripheral Indian regions (Hall, 2002; Lohman et al., 2011).

4.6 Origin of regional cryptic speciation (Indo-Malay Archipelago: IMA)

Cryptic speciation has become entrenched to refer to the zone of IMA and the shreds of evidence are also traced by both climatic and geological influence on habitat reconstructions of the region (Hall, 2002; Lohman et al., 2011). Rising and falling sea levels during the Pleistocene led to the emergence of diverse breeding microhabitats for carangid fish, such as shallow reef areas and estuaries (Lohman et al., 2011). More recent evidence has shown that carangid fish have remarkable diversity and distribution in the IMA region and are evidently diversified in optimum resources and habitat flexibility. During the spawning season, most are euryhaline and their salinity tolerance ranges from euhaline oceanic salinity to estuarine salinity (Smith & Parrish, 2002). The average intraspecific genetic divergence of the IMA region (range 0.00–4.30)% is considerably higher as explained by Mat Jaafar et al. (2012) and exceeds the marginal 2% speciation delimitation. Furthermore, even in the congeneric strata of cosmopolitan localities, the characteristically higher mutation rate would change their evolution towards sub-speciation and increase the potential for the rise of cryptic speciation, followed by hybridization between closely related species (Murakami et al., 2007). This is hypotheimically explained by the Tajima’s D statistic calculated for the IMA region, which is significantly negative for each regional carangid species, similar to the observations made by Jaafar (2013). This is evident from the bottleneck effect, followed by rapid population expansion due to past geological events and a subsequent severe reduction in the regional effective size (Jaafar, 2013; Omori & Wu, 2017). Therefore, the emergence of microhabitats increased the possibility of rapid mutations and low-frequency regional haplotypes.

4.7 Relationship between carangid body shape and regional nucleotide divergences

Moreover, this phenomenon observed in the current study is consistent with the deep divergences in inshore carangid species (average COI divergence: 5.10%) compared with the offshore species (average COI divergence: 2.60%), and revealed a higher abundance of putative inshore cryptic species in the IMA region (Mat Jaafar et al., 2012; Zemlak et al., 2009). Variations in genetic divergences correlated indirectly with carangid body shapes; this is clearly evident in the carangids examined in the current study. The three corresponding species of trevallies tend to have compressiform body types and are highly diverse in inshore reef areas (Smith-Vaniz, 1999) within isolated geographical localities. This was further confirmed by the greater genetic divergence (1.56%) of trevallies between the Indian and IAA regions, compared to that of pelagic encounters of scads (1.03%).

5. CONCLUSION

Integrating both taxonomic approaches, the landmark-based geometric morphometrics method and DNA barcoding (GMM: COI barcoding) contributed to the promising and consolidated results of carangid species identification over the pronounced phenotypic impediments within the family. Therefore, such intensified taxonomic data are helpful for the effective monitoring of carangid biodiversity and population conservation management practices. Characteristic Indo-Malay genetic divergence and remarkable nucleotide divergence gap between the Indian Ocean provide clues to the origin of carangid cryptic speciation and its evolutionary distribution along with prehistoric phenomena, which coincide with the center of origin and overlap hypothesis. Furthermore, modern paleontological approaches including their migratory behaviors, are recommended for a better resolution of the research findings of the study.

AUTHOR CONTRIBUTIONS
Lahiru Gamage and Naveen Ranasinghe were involved in the investigation, conceptualization, methodology, formal analysis, validation, visualization, and manuscript writing. The whole study was supervised and reviewed by Dona Munasinghe and Tsung-Han Lee. All authors of this paper have read and approved the final version submitted.

ACKNOWLEDGMENTS

We would like to acknowledge the financial support granted by the Vice-Chancellor (VC) Fund of the University of Ruhuna, Sri Lanka for ongoing research (2021) for their contribution to PCR product sequencing. This paper is also a south-bound collaboration of the Department of Life Sciences, National Chung Hsing University, Taiwan, under the support of the iEGG and Animal Biotechnology Center from The Feature Area Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE), Taiwan (MOE 109-S-0023-A) to T.H. Lee.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare regarding the content of the manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from either the corresponding author or the first author upon a reasonable request. Individual genotype data are available on DataDryad: https://datadryad.org/stash.

ORCID

Lahiru Gamage https://orcid.org/0000-0002-7325-9078
Naveen Ranasinghe https://orcid.org/0000-0001-5412-6483
Dona Munasinghe https://orcid.org/0000-0001-9627-314X
Tsung-Han Lee https://orcid.org/0000-0002-3542-4335

REFERENCES


Fischer, W., & Whitehead, P. J. P. (1974). *FAO species identification sheets for fishery purposes Eastern Indian Ocean (fishing area 57) and Western Central Pacific (fishing area 71)*. http://www.fao.org/docrep/009/e9163e/e9163e00.htm


### TABLE 1
Universal fish primer sequences for cytochrome c oxidase subunit I (COI) used in PCR.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer sequence 5’- 3’</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fish F1</td>
<td>TCAACCAACCACAAAGACATTGGCAC</td>
<td>(Ward et al., 2005)</td>
</tr>
<tr>
<td>Fish R1</td>
<td>TAGACTTCTGGGTGGCAAGAATCA</td>
<td>(Ward et al., 2005)</td>
</tr>
<tr>
<td>Fish F2</td>
<td>TCGACTAATCATAAAGATACGCAC</td>
<td>(Ward et al., 2005)</td>
</tr>
<tr>
<td>Fish R2</td>
<td>ACTTCAGGGTGACCGAAGAATCAGAA</td>
<td>(Ward et al., 2005)</td>
</tr>
</tbody>
</table>

### TABLE 3
Summary of different morphometric values of the carangids studied in the study.

<table>
<thead>
<tr>
<th>Species</th>
<th>Species Parameter</th>
<th>Morphometric Parameter</th>
<th>Morphometric Parameter</th>
<th>Morphometric Parameter</th>
<th>Morphometric Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A. mate</em> (n = 150)</td>
<td>Total Weight (g)</td>
<td>82.46</td>
<td>82.46</td>
<td>4.50 (4.00-5.00)</td>
<td>1.001 ± 0.05</td>
</tr>
<tr>
<td></td>
<td>Total Length (cm)</td>
<td>(79.00-89.00)</td>
<td>(79.00-89.00)</td>
<td>(14.30-15.50)</td>
<td>1.000 ± 0.03</td>
</tr>
<tr>
<td></td>
<td>Body Depth (cm)</td>
<td>15.00</td>
<td>18.47</td>
<td>5.46 (5.00-5.50)</td>
<td>1.001 ± 0.05</td>
</tr>
<tr>
<td></td>
<td>Condition factor</td>
<td></td>
<td></td>
<td></td>
<td>1.004 ± 0.03</td>
</tr>
<tr>
<td><em>S. crumenophthalmus</em> (n = 90)</td>
<td>Total Weight (g)</td>
<td>86.83</td>
<td>82.00</td>
<td>8.27 (8.00-8.60)</td>
<td>1.004 ± 0.03</td>
</tr>
<tr>
<td></td>
<td>Total Length (cm)</td>
<td>(79.00-89.00)</td>
<td>(82.00-89.00)</td>
<td>(8.00-18.80)</td>
<td>1.004 ± 0.03</td>
</tr>
<tr>
<td></td>
<td>Body Depth (cm)</td>
<td>17.15</td>
<td>18.53</td>
<td>5.67 (5.00-5.50)</td>
<td>1.001 ± 0.05</td>
</tr>
<tr>
<td></td>
<td>Condition factor</td>
<td></td>
<td></td>
<td></td>
<td>1.004 ± 0.03</td>
</tr>
</tbody>
</table>
### Morphometric Parameters

<table>
<thead>
<tr>
<th>Species</th>
<th>Species</th>
<th>Morphometric Parameter</th>
<th>Morphometric Parameter</th>
<th>Morphometric Parameter</th>
<th>Morphometric Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. hedlandensis</td>
<td>A. hedlandensis</td>
<td>156.32 (150.00-160.00)</td>
<td>18.22 (17.9-18.5)</td>
<td>9.18 (8.50-10.00)</td>
<td>1.000 ± 0.03</td>
</tr>
<tr>
<td>(n = 30)</td>
<td>(n = 30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T. coeruleopinnatum</td>
<td>T. coeruleopinnatum</td>
<td>151.38 (148.00-155.00)</td>
<td>17.48 (17.00-17.80)</td>
<td>9.45 (8.50-9.50)</td>
<td>1.000 ± 0.05</td>
</tr>
<tr>
<td>(n = 27)</td>
<td>(n = 27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUNK (n = 4)</td>
<td>SUNK (n = 4)</td>
<td>85.00 (79.00-88.00)</td>
<td>18.38 (17.50-18.80)</td>
<td>5.63 (5.00-6.50)</td>
<td>1.013 ± 0.03</td>
</tr>
<tr>
<td>TUNK (n = 3)</td>
<td>TUNK (n = 3)</td>
<td>130.67 (124.00-135.00)</td>
<td>18.30 (17.50-18.80)</td>
<td>9.10 (8.50-9.30)</td>
<td>1.001 ± 0.05</td>
</tr>
</tbody>
</table>

**Note**: $K_n$ represents relative condition factor

### TABLE 4 Procrustes distances between morph groups followed by the Mahalanobis distance analysis (10000 permutation rounds: < 0.0001).

### TABLE 5 Kimura 2-parameter (K2P) distances between different taxonomic levels of carangids analyzed.
TABLE 6 Congeneric nucleotide K2P distances for five genera of carangids studied in the study.

<table>
<thead>
<tr>
<th>Genus</th>
<th>Genus</th>
<th>No. of sequences (n)</th>
<th>Mean K2P distance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Carangoides</em></td>
<td><em>Atropus</em></td>
<td>30</td>
<td>0.32</td>
</tr>
<tr>
<td>(Early designation before the generic revision)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Turrum</em></td>
<td></td>
<td>30</td>
<td>0.40</td>
</tr>
<tr>
<td><em>Platycaranx</em></td>
<td></td>
<td>40</td>
<td>0.60</td>
</tr>
<tr>
<td><em>Selar</em></td>
<td></td>
<td>60</td>
<td>0.37</td>
</tr>
<tr>
<td><em>Atule</em></td>
<td></td>
<td>40</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Note: aGeneric revision of the species formerly belonging to the genus *Carangoides* and its related genera (Carangiformes: Carangidae) (Kimura et al., 2022)

TABLE 7 Summary of regional intraspecific genetic distances (K2P) and the corresponding divergence with the closest ancestor of the species.

<table>
<thead>
<tr>
<th>Species</th>
<th>IAA region</th>
<th>Indian region</th>
<th>Between IAA and the Indian region</th>
<th>Between IAA and the closest ancestor</th>
<th>Between Indian and the closest ancestor</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A. mate</em></td>
<td>0.45</td>
<td>0.00</td>
<td>0.80</td>
<td>0.13</td>
<td>0.70</td>
</tr>
<tr>
<td><em>S. crumenophthalmus</em></td>
<td>0.40</td>
<td>0.17</td>
<td>0.90</td>
<td>0.25</td>
<td>0.50</td>
</tr>
<tr>
<td><em>S. boops</em></td>
<td>0.38</td>
<td>0.15</td>
<td>1.40</td>
<td>0.18</td>
<td>1.30</td>
</tr>
<tr>
<td><em>P. malabaricus</em></td>
<td>0.60</td>
<td>0.10</td>
<td>1.10</td>
<td>0.19</td>
<td>1.00</td>
</tr>
<tr>
<td><em>A. hedlandensis</em></td>
<td>0.34</td>
<td>0.10</td>
<td>1.90</td>
<td>0.16</td>
<td>1.70</td>
</tr>
<tr>
<td><em>T. coeruleopinnatum</em></td>
<td>0.48</td>
<td>0.18</td>
<td>1.70</td>
<td>0.28</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note: IAA: Indo–Australian Archipelago

FIGURE LEGENDS

FIGURE 1 Morphological characteristics of cryptic carangid species studied in the study. (a) *S. crumenophthalmus*: longer arched section and lower curvature of the lateral line, (b) *S. boops*: both arched and straight-line sections of the lateral line are equal, (c) *A. mate*, (d) *A. hedlandensis*, (e) *P. malabaricus*, (f) *T. coeruleopinnatum*.

FIGURE 2 The 16 defined landmark points for extracting the body shape data in carangid species studied: (1) anterior tip of upper jaw, (2) midpoint indication of eye, (3) forehead (ending of frontal bone), (4) origin of first dorsal fin, (5) origin of second dorsal fin, (6) anterior origin of caudal peduncle, (7) anterior origin of caudal fin, (8) forked point of caudal fin, (9) posterior origin of caudal fin, (10) posterior origin of caudal peduncle, (11) origin of anal fin, (12) origin of pelvic fin, (13) down point indication of operculum, (14) and (15) linear extremities of eye, and (16) superior insertion of pectoral fin.

FIGURE 3 Graphical representation of principal component analysis (PCA) and canonical variate analysis (CVA) of the carangid body shape variation (n = 360). (a) Principal component analysis on the shape variables PC1 and PC2 axes based on the left body side account for 73.1% and 16.2% respectively of the
total variance. (b) Canonical variate analysis on the shape variables CV1 and CV2 axes based on the left body side account for 84.9% and 13% respectively of the total variance.

**FIGURE 4** The procrustes deformation grids based on the first two canonical variates (CV1 and CV2) and principal components (PC1 and PC2) are shown for the left body side. (a) CV1 and (c) PC1 represent the carangid body shape variations in the anterior insertion of the dorsal fin. (b) CV2 and (d) PC2 represent the carangid body shape differences accounting for disparities in the anterior tip of the snout.

**FIGURE 5** Maximum likelihood tree (K2P distances) indicating the monophyletic clusters of the selected six carangid species for the study and the paraphyletic behaviour of the three genera, *Atropus*, *Turrum* and *Platycaranx* with their previously assumed genus *Carangoides*.

**FIGURE 6** Maximum likelihood trees (K2P distances) indicate the isolation of the regional clusters, Indian and Indo-Australian Archipelago (IAA) in relation to the ancestral cluster. (a) *Atropus hedlandensis* (b) *Platycaranx malabaricus*, and (c) *Turrum coeruleopinnatum*.

**FIGURE 7** Maximum likelihood trees (K2P distances) indicate the isolation of the regional clusters, Indian and Indo-Australian Archipelago (IAA) in relation to the ancestral cluster. (a) *Selar crumenophthalmus*, (b) *Selar boops*, and (c) *Atule mate*.

**FIGURE 8** Median-joining network constructed for haplotypes concentrated in the Indo-Australian region (IAA) based on the intraspecific COI genetic divergence. Each circle represents one unique haplotype, with the area proportional to the haplotype frequency in all populations. (a) *Atropus hedlandensis*, (b) *Atule mate*, (c) *Platycaranx malabaricus*, (d) *Selar boops*, (e) *Turrum coeruleopinnatum*, and (f) *Selar crumenophthalmus*.

**FIGURE 9** Cenozoic reconstructions of land and sea in the Indo-Australian Archipelago (IAA). (a) Origin of the carangid ancestor along with the center of origin (60 MYA), (b) Gradual emergence of shallow sea areas in the IAA region as a result of northward movement of the Indo-Australian plate (40 MYA), (c) Accumulation of the ancestral descendants in the IAA region and separation of the peripheral Indian clusters (5 MYA) (d) Recent distribution trend of carangids represent their centered distribution in the IAA region.
FIGURE 1 Morphological characteristics of cryptic carangid species studied in the study. (a) *S. crumenophthalmus*: longer arched section and lower curvature of the lateral line, (b) *S. boops*: both arched and straight-line sections of the lateral line are equal, (c) *A. mate*, (d) *A. hedlandensis*, (e) *P. malabaricus*, (f) *T. coeruleopinnatum*. 
FIGURE 2 The 16 defined landmark points for extracting the body shape data in carangid species studied: (1) anterior tip of upper jaw, (2) midpoint indication of eye, (3) forehead (ending of frontal bone), (4) origin of first dorsal fin, (5) origin of second dorsal fin, (6) anterior origin of caudal peduncle, (7) anterior origin of caudal fin, (8) forked point of caudal fin, (9) posterior origin of caudal fin, (10) posterior origin of caudal peduncle, (11) origin of anal fin, (12) origin of pelvic fin, (13) down point indication of operculum, (14) and (15) linear extremities of eye, and (16) superior insertion of pectoral fin.
FIGURE 3 Graphical representation of principal component analysis (PCA) and canonical variate analysis (CVA) of the carangid body shape variation ($n = 360$). (a) Principal component analysis on the shape variables PC1 and PC2 axes based on the left body side account for 73.1% and 16.2% respectively of the total variance. (b) Canonical variate analysis on the shape variables CV1 and CV2 axes based on the left body side account for 84.9% and 13% respectively of the total variance.
FIGURE 4 The procrustes deformation grids based on the first two canonical variates (CV1 and CV2) and principal components (PC1 and PC2) are shown for the left body side. (a) CV1 and (c) PC1 represent the carangid body shape variations in the anterior insertion of the dorsal fin. (b) CV2 and (d) PC2 represent the carangid body shape differences accounting for disparities in the anterior tip of the snout.
FIGURE 5 Maximum likelihood tree (K2P distances) indicating the monophyletic clusters of the selected six carangid species for the study and the paraphyletic behaviour of the three genera, *Atropus*, *Turrum* and *Platycaranx* with their previously assumed genus *Carangoides*. 
FIGURE 6 Maximum likelihood trees (K2P distances) indicate the isolation of the regional clusters, Indian and Indo-Australian Archipelago (IAA) in relation to the ancestral cluster. (a) *Atropus hedlandensis* (b) *Platycaranx malabaricus*, and (c) *Turrum coeruleopinnatum*. 

- common ancestral node
- Peripheral Indian Ocean
- Indo-Australian Archipelago
- Ancestral Archipelago
FIGURE 7 Maximum likelihood trees (K2P distances) indicate the isolation of the regional clusters, Indian and Indo-Australian Archipelago (IAA) in relation to the ancestral cluster. (a) Selar crumenophthalmus, (b) Selar boops, and (c) Atule mate.
Figure 8 Median-joining network constructed for haplotypes concentrated in the Indo-Australian region (IAA) based on the intraspecific COI genetic divergence. Each circle represents one unique haplotype, with the area proportional to the haplotype frequency in all populations. (a) *Atropus hedlandensis*, (b) *Atule mate*, (c) *Platycaranx malabaricus*, (d) *Selar boops*, (e) *Turrum coeruleopinnatum*, and (f) *Selar crumenophthalmus*. 
FIGURE 9 Cenozoic reconstructions of land and sea in the Indo-Australian Archipelago (IAA). (a) Origin of the carangid ancestor along with the center of origin (60 MYA), (b) Gradual emergence of shallow sea areas in the IAA region as a result of northward movement of the Indo-Australian plate (40 MYA), (c) Accumulation of the ancestral descendants in the IAA region and separation of the peripheral Indian clusters (5 MYA) (d) Recent distribution trend of carangids represent their centered distribution in the IAA region.