Jurnal Ilmu dan Teknologi Kelautan Tropis

p-ISSN: 2087-9423
e-ISSN: 2620-309X

Article Info:
Received 02 December 2025
Revised 23 January 2026
Accepted 20 March 2026

Corresponding Author:
Achmad Farajallah
Department of Biology
IPB University
E-mail: achamad@apps.ipb.ac.id

Reanalysis of Mitochondrial and Nuclear Markers (COI, 16S, 28S) in Coenobitidae, Diogenidae, and Paguridae for Species Delimitation

Tazkiah Auliaputri[a], Tyani Fitrian[b], Achmad Alfiyan[b], Intan Ratu Pratiwi[a], Vendi Eko Susilo[a, d], Ali Mashar[c], Achmad Farajallah[a*]

[a] Department of Biology, Faculty of Mathematics and Natural Sciences, IPB University, Jl. Agatis Kampus IPB Dramaga, Bogor 16680, Indonesia

[b] Research Centre for Biosystematics and Evolution, National Research and Innovation Agency (BRIN), Jl. Raya Jakarta Bogor Km 46, Cibinong, Bogor, West Java, Indonesia

[c] Department of Aquatic Resources Management, Faculty of Fisheries and Marine Sciences, IPB University, Dramaga, Bogor 16680, Indonesia

[d] Biology study program, Faculty of Mathematics and Natural Sciences, University of Jember, Jember, East Java 68121, Indonesia

Abstract

Public sequence repositories, such as GenBank, contain an extensive collection of molecular data that are widely used for phylogenetic inference and species identification. However, inconsistent annotation and potential misidentifications can obscure true genetic patterns, particularly in groups with complex morphology such as Paguroidea. This study reanalyzed 4,871 COI, 16S, and 28S sequences belonging to the families Coenobitidae, Diogenidae, and Paguridae retrieved from GenBank. Following rigorous quality filtering, 2,450 sequences were retained for downstream analyses to quantify intra and interspecific genetic distances for each family and each molecular marker. The results reveal substantial heterogeneity in genetic divergence across families, with COI and 16S exhibiting broad ranges and frequent overlap between intra and interspecific distances, whereas 28S displayed very low variation consistent with its slower evolutionary rate. Elevated intraspecific distances in several taxa suggest potential cryptic lineages, while unusually shallow interspecific values indicate possible synonymy or recent divergence. Rather than proposing taxonomic changes, this study provides a curated reference framework that establishes empirical distance thresholds for molecular species delimitation in Paguroidea. These findings offer a practical baseline for future research employing COI, 16S, and 28S markers in species identification, validation of public sequence data, and the refinement of hermit crab systematics.

Keywords: Genetic distance, Paguroidea, Public data review, Species delimitation

1. Introduction

Public DNA sequence repositories, such as GenBank, serve as major gateways to global genomic information. They are interconnected with international sequence databases, including the International Nucleotide Sequence Database Collaboration (INSDC), which comprises DDBJ (Japan), ENA/EMBL (Europe), GenBank (USA), and other specialized repositories, such as ENBL, ENOL, and EBOL. Through these coordinated networks, GenBank provides unified access to curated genomic, mitochondrial, and ribosomal sequence data submitted by laboratories worldwide, ensuring broad coverage and cross-referenced annotation across multiple platforms.

These databases allow researchers to rapidly retrieve and examine molecular data, most commonly through sequence similarity search tools such as the Basic Local Alignment Tool (BLAST) (Altschul et al., 1990). BLAST identifies homologous sequences by comparing a query sequence against millions of deposited sequences using a heuristic algorithm designed for speed and efficiency. Its advantages include high computational speed, the ability to detect both close and distant sequence similarities, statistical scoring of alignments, and broad applicability across genes, genomes, and taxonomic groups. BLAST enables researchers to infer putative species identity, assess sequence quality, detect contaminants, and explore phylogenetic relationships with high reliability. As the volume of genomic data continues to

expand, these interconnected public databases and search tools have become indispensable for resolving taxonomic uncertainties, delineating species boundaries, and conducting comparative analyses across mitochondrial and nuclear genetic markers (Hebert et al., 2003; Ratnasingham and Hebert, 2007).

Within the superfamily Paguroidea, which includes the families Coenobitidae, Diogenidae, and Paguridae, species delimitation has long been complicated by overlapping morphological traits and plasticity in diagnostic characters (Malay and Paulay, 2010; Mclaughlin et al., 2010). These three families also represent the most taxonomically challenging and ecologically diverse hermit crab lineages: Coenobitidae includes terrestrial and semi-terrestrial species that have undergone extensive adaptation to land environments (Rahayu et al., 2016; Hamasaki et al., 2017). Diogenidae comprises one of the largest hermit crab families, exhibiting considerable morphological variation and a complex of morphologically similar species (Rahayu and Komai, 2000; Lemaitre et al., 2018). This challenge is also found in Paguridae, where subtle differences between its numerous marine species often lead to misidentification. However, advancements in taxonomy have led to a significant increase in the number of newly described species within Paguridae over the past decade (Asakura, 2001; McLaughlin and Rahayu, 2008; Rahayu and Komai, 2013; Komai and Rahayu, 2013, 2014; Komai et al., 2021; Lemaitre and Felder, 2023). Consequently, molecular data have become essential for verifying species identity, evaluating the consistency of publicly deposited sequences, and establishing more reliable genetic thresholds to support accurate species delimitation within this group.

Despite the growing availability of sequence data in public repositories, many records remain insufficiently validated, inconsistently annotated, or derived from misidentified specimens, which may propagate taxonomic inaccuracies. Studies on other taxa have demonstrated that GenBank sequences assigned to single species can represent multiple divergent lineages or entirely different species (Vivien and Martin, 2025), underscoring the need for careful re-evaluation of publicly available datasets, particularly for morphologically variable taxa such as Paguroidea. In classical taxonomy, species delimitation is governed by nomenclatural rules that require the designation of a holotype, typically an adult specimen with stable and fully developed morphological characters, to anchor the species name. However, in practical applications, especially for molecular traits or those exhibiting substantial developmental variability. Consequently, genomic data deposited in public repositories may originate from specimens that cannot be confidently matched to described species based solely on morphology. This contrast between nomenclatural standards and the provenance of molecular data contributes to inconsistent taxonomic assignments and further complicates efforts to validate sequences in groups with strong ontogenetic plasticity, such as Paguroidea (ICZN, 1999; Bokulich, 2020).

In molecular taxonomy, the combined use of mitochondrial and nuclear markers provides a robust framework for resolving evolutionary relationships across taxonomic levels. The COI gene serves as the standard animal DNA barcode and is highly informative for distinguishing closely related species (Hebert et al., 2003). Meanwhile, the 16S rRNA gene, which mutates more slowly, is effective for assessing interspecific divergence in decapods (Schubart, 2009; Hadadi et al., 2023). In contrast, the nuclear 28S rRNA gene, which exhibits an even slower rate of evolution, provides a robust phylogenetic signal for reconstructing relationships at the genus and family levels (Tsang et al., 2011; Bracken-Grissom et al., 2013). These markers enable a comprehensive assessment of genetic variation within both recent and ancient lineages.

In this study, we conducted a review analysis of 4,871 cytochrome coxidase subunit I (COI), 16S, and 28S sequences retrieved from GenBank, representing the families Coenobitidae, Diogenidae, and Paguridae. Our objective was to estimate intra-and interspecific genetic distance thresholds for each family and gene, providing an empirical molecular basis for species delimitation in Paguroidea. By standardizing the reassessment of publicly accessible sequences, this work aims to enhance the stability of molecular taxonomy and clarify evolutionary relationships among terrestrial and marine hermit crabs.

2. Materials and Methods

2.1. Data Collection

Molecular data of the families Coenobitidae, Diogenidae, and Paguridae within the superfamily Paguroidea were obtained from the GenBank database (https://www.ncbi.nlm.nih.gov/nuccore). All COI, 16S, and 28S sequences of Coenobitidae, Diogenidae, and Paguridae were retrieved through BLASTn searches using representative reference sequences from each family as query sequences. The BLAST algorithm (Altschul et al., 1990) was set to a maximum target of 5,000 sequences. All retrieved sequences were downloaded in the FASTA format using the Multiple Sequence Alignment (MSA) viewer.

2.2. Data Filtering

Sequence quality control was performed prior to analysis. Sequences containing ambiguous IUPAC bases, other than A, T, G, and C (e.g., R, Y, S, W, K, M, B, D, H, V, and N), were removed using TextPad 8. Sequences shorter than 500 bp were excluded from the dataset. In addition, sequences labelled with uncertain identifications such as *sp., cf.,* or *nr.* were eliminated. To ensure homology, only sequences covering the same target gene region were considered. Furthermore, all sequences included in this study originated from published papers to ensure data reliability and validated taxonomic identifications.

2.3. Sequence Trimming and Alignment

All filtered sequences were trimmed to remove non-homologous regions. Sequence alignment was conducted using the ClustalW algorithm (Thompson et al., 1994) implemented in MEGA v11 (Tamura et al., 2021) with default parameters. The resulting aligned dataset was visually inspected to confirm alignment accuracy across taxa.

2.4. Genetic Distance Analysis

The aligned dataset was analyzed to estimate intra-and interspecific genetic distances using the Kimura 2-parameter (K2P) substitution model in MEGA v11 (Kimura, 1980). The final datasets were subsequently imported into RStudio v2025.09.0 for further statistical analysis, including the calculation of minimum, maximum, and mean intraspecific distances. The results were used to evaluate species-level divergence and potential barcode gaps among taxa.

2.5. Phylogenetic reconstruction

Phylogenetic relationships among species were inferred using the maximum likelihood (ML) method implemented in IQ-Tree v3.0.1 (Wong et al., 2025). Model selection was performed in ModelFinder, which automatically identified the best-fit nucleotide substitution model based on the Akaike information criterion (AICc) scores, resulting in the GTR+I+G model. Tree robustness was assessed through 1,000 ultrafast bootstrap replicates (UFBoot2) and 1,000 SH-aLRT tests. Phylogenetic trees were visualized and edited using FigTree v1.4.4. Final figure editing and layout adjustments were performed using Inkscape v1.2.2.

3. Results and Discussion

3.1. Sequence Retrieval, Data Curation, and Representation Across Families

A total of 4,871 DNA sequences were retrieved from GenBank across the three target genes (COI, 16S, and 28S) representing members of Coenobitidae, Diogenidae, and Paguridae. After sequence quality filtering, ambiguous bases, short fragments, and records with uncertain taxonomic annotations were removed; the final curated dataset consisted of 2,450 sequences **(Table 1)**.

Table 1. Summary of sequences datasets used in this study retrieved from GenBank and retained after quality filtering.

Gene	Families	Entry Data (sequence)	Final Data Amount (sequence)	Number of species	Ratio (Final data : Number of species)
COI	Coenobitidae	268	99	11	9 : 1
	Diogenidae	2,123	915	162	≈6 : 1
	Paguridae	1,164	573	53	≈11 : 1
16S	Coenobitidae	44	25	11	≈2 : 1
	Diogenidae	769	413	129	≈3 : 1
	Paguridae	428	365	80	≈4 : 1
28S	Coenobitidae	4	4	3	≈1 : 1
	Diogenidae	27	19	16	≈1 : 1
	Paguridae	44	37	23	≈2 : 1
Total		4,871	2,450		

Across all genes, Diogenidae exhibited the highest reduction from initial to final sequence counts, indicating a substantial number of low-quality or inconsistently annotated records within this family. This pattern is consistent with the well-documented taxonomic complexity and morphological overlap within Diogenidae, which may contribute to ambiguous labeling in public databases. In contrast, Coenobitidae showed the highest retention rate, reflecting generally better sequence consistency and lower levels of taxonomic ambiguity.

The number of species represented in the final dataset also differed among families and markers, with Diogenidae exhibiting the greatest species richness, followed by Paguridae and Coenobitidae. This uneven distribution reflects both biological diversity and research bias across Paguroidea lineages. Overall, the curated dataset provides a robust foundation for downstream analyses of genetic divergence and phylogenetic relationships, ensuring that subsequent interpretations are based on high-quality, taxonomically reliable molecular data.

3.2. Genetic Distance Patterns Across Families and Markers

A reanalysis of genetic distances across COI, 16S, and 28S revealed consistent trends among the three families, in which intraspecific divergence was generally lower than interspecific divergence, although notable overlap persisted in several markers **(Table 2)**. The mitochondrial COI gene showed the clearest separation between intra-and interspecific values, especially between Coenobitidae, whereas Diogenidae exhibited the broadest intraspecific ranges (0 – 0.075), reflecting its higher taxonomic complexity and possible species complex structure. The 16S gene displayed a more conservative divergence pattern overall, yet still produced elevated interspecific distances in Diogenidae and Paguridae. In contrast, the nuclear 28S marker exhibited extremely low variation, with intraspecific divergence nearly absent and interspecific values remaining modest, consistent with the slow evolutionary dynamics of nuclear ribosomal genes.

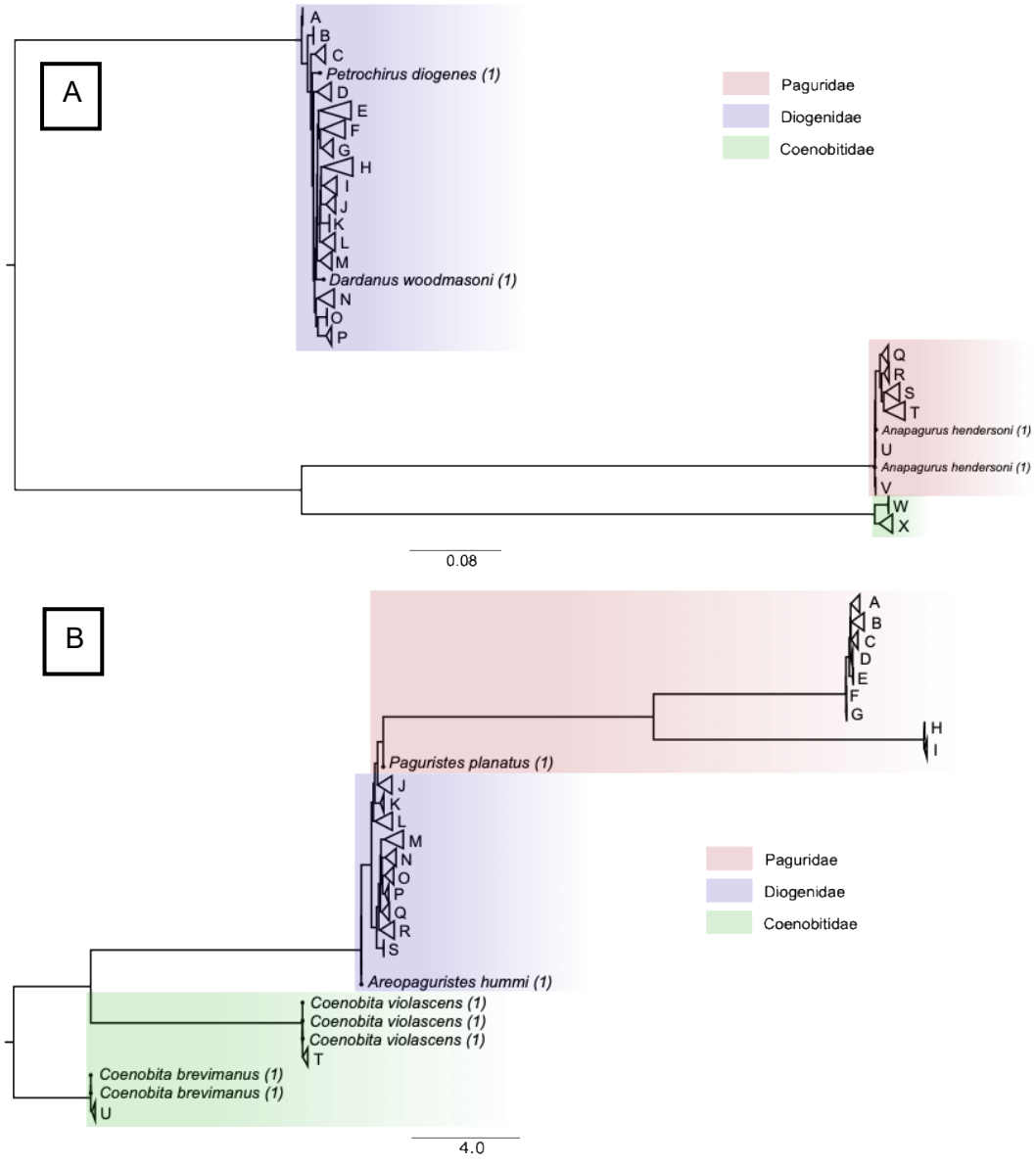
Table 2. Summary of intra and interspecific genetic distances across three families and genes allowing comparison of divergence patterns.

Gene	Families	Intraspecies Distance (Range)	Intra Average	Interspecies Distance (Range)	Inter Average
COI	Coenobitidae	0 – 0.051	0.016	0 – 0.374	0.195
	Diogenidae	0 – 0.075	0.009	0 – 0.046	0.017
	Paguridae	0 – 0.025	0.005	0 – 0.031	0.012
16S	Coenobitidae	0 – 0.015	0.0055	0.005 – 0.020	0.012
	Diogenidae	0 – 0.066	0.0095	0.010 – 0.070	0.025
	Paguridae	0 – 0.045	0.009	0.004 – 0.104	0.026
28S	Coenobitidae	0.006	0.006	0.006	0.006
	Diogenidae	0	0	0 – 0.054	0.017
	Paguridae	0 – 0.007	0.001	0 – 0.090	0.022

These divergence patterns are also influenced by the unequal representation of species and sequence counts across families. Families with higher retained sequence numbers, such as Diogenidae in COI (≈6:1 ratio) and Paguridae in COI (≈11:1 ratio), showed greater diversity and wider divergence ranges, suggesting that larger datasets capture more complete intraspecific variation along with potential cryptic structures. Conversely, low ratios in markers such as 28S (≈1–2:1) indicate limited sampling and reduced detectable variation, which restricts the marker's resolution for species delimitation. Overall, COI remains the most informative marker for distinguishing species-level boundaries, whereas 16S and 28S are more suitable for higher-level phylogenetic contexts.

3.3. Phylogenetic Reconstruction

Phylogenetic relationships among members of Coenobitidae, Diogenidae, and Paguridae were reconstructed using maximum likelihood analyses of the COI, 16S, and 28S datasets. Each gene produced distinct topological patterns consistent with their evolutionary rates, allowing comparison of phylogenetic resolution across mitochondrial and nuclear markers. To enhance interpretability, branches with weak support were collapsed, and only nodes with robust bootstrap or SH-aLRT values were retained in the final trees (**Figure 1A–C**), all species within the clades mentioned in **Table 3**. Overall, the reconstructed phylogenies reveal varying levels of clade stability among families and markers, providing insights into both higher-level relationships and potential species-level inconsistencies within Paguroidea.



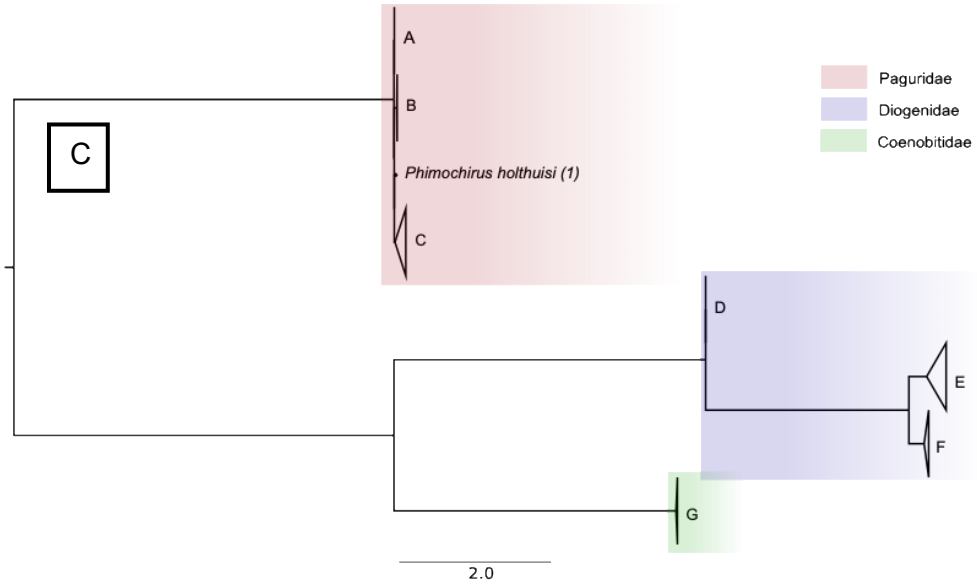


Figure 1. Phylogeny tree constructed in Maximum Likelihood from A. COI, B. 16S, and C. 28S sequences, illustrating the clustering patterns and evolutionary relationships inferred from each gene markers.

Table 3. Collapsed composition for the phylogenetic trees in Figure 1A-C. List all species grouped within each collapsed node from COI, 16S, and 28S phylogenetic trees, providing a reference for clade membership.

A.1 COI

A.1.1 Diogenidae

A = *Aniculus retipes* (4)
B = *Dardanus hessii* (1)
C = *Dardanus guttatus* (2), *D. venosus* (1)
D = *Dardanus arrosor* (18), *D. insignis* (1), *D. calidus* (8), *D. deformis* (7), *D. pectinatus* (1)
E = *Areopaguristes cyanops* (1), *Diogenes pallescens* (5), *D. deflectomanus* (1), *D. singaporensis* (1), *D. alias* (2), *D. violaceus* (1), *D. dubius* (1), *D. miles* (1), *D. manaarensis* (1), *D. planimanus* (1), *D. holthuisi* (1), *D. arguinensis* (4), *D. costatus* (6), *D. armatus* (22), *D. ponticus* (18), *D. pugilator* (26), D. *curvimanus* (18), *D. klaasi* (1), *D. rectimanus* (2), *D. takedai* (1), *D. nitidimanus* (1), *D. edwardsii* (3), *D. laevicarpus* (1), *D. avarus* (2), *D. foresti* (1), *D. moosai* (1), *D. goniochirus* (1), D. *brevirostris* (16), *Strigopagurus boreonotus* (1), *Paguristes albimaculatus* (1), *Paguropsina inermis* (2), *P. pistillata* (3), *Paguropsis andersoni* (3), *P. confusa* (2), *P. typica* (3), *P. gigas* (2), *P. lacinia* (3); *Areopaguristes japonicus* (2), *A. nigroapiculus* (4), *Paguristes digitalis* (3), *P. ortmanni* (6), *P. turgidus* (4), *P. barnardi* (5), *P. gamianus* (5), *P. doederleini* (1), *P. miyakei* (1), *P. jalur* (3), *P. tortugae* (6), *Pseudopaguristes calliopsis* (1)
F = *Clibanarius aequabilis* (1), *C. erythropus* (17), *C. corallinus* (11), *C. rhabdodactylus* (4), *C. zebra* (2), *C. signatus* (2); *Clibanarius antillensis* (15), *C. tricolor* (4); *Clibanarius arethusa* (5), *C. rutilus* (4), *C. infraspinatus* (10), *C. eurysternus* (14), *C. clibanarius* (11), *C. rutilus* (4), *C. taeniatus* (3), *C. demani* (2), *C. longitarsus* (15), *C. striolatus* (12), *C. padavensis* (12), *C. sclopetarius* (18), *C vittatus* (20), *C. englaucus* (11), *C. merguiensis* (13), *C. humilis* (10), *C. virescens* (21); *Clibanarius cruentatus* (1), *C. snelliusi* (4)
G = *Isocheles pacificus* (1), *I. sawayai* (5), *Loxopagurus loxochelis* (4)
H = *Calcinus albengai* (1), *C. anani* (2), *C. fuscus* (10), *C. argus* (9), *C. imperialis* (1), *C. isabellae* (5), *C. vanninii* (2), *C. dapsiles* (3), *C. inconspicuous* (1), *Ciliopagurus caparti* (2), *Calcinus pascuensis* (1), *C. californiensis* (3), *C. obscurus* (3), *C. mclaughlinae* (1), *C. explorator* (2), *C. tibicen* (19), *C. guamensis* (9), *C. vachoni* (16), *C. elegans* (18), *C. pictus* (10), *C. orchidae* (1), *C. morgani* (16), *C. gaimardii* (5), *C. laevimanus* (13), *C. seurati* (18), *C. gouti* (4), *C. hakahau* (2), *C. laurentae* (4), *C. lineapropodus* (7), *C. pulcher* (20); *Calcinus tubularis* (21), *C. verrillii* (1); *Calcinus latens* (15)
I = *Dardanus gemmatus* (2), *D. pedunculatus* (4), *D. hessii* (1), *D. impressus* (1), *D. jacquesi* (1)
J = *Calcinus haigae* (16), *C. minutus* (7), *C. rosaceus* (10), *C. hazletti* (21)

K = *Diogenes ovatus* (2)
L = *Ciliopagurus galzini* (4), *C. grandis* (1), *C. strigatus* (11), *C. vakovako* (1), *C. krempfi* (2), *C. tricolor* (7), *C. shebae* (1), *C. hawaiiensis* (1), *Tetralobistes weddellii* (1)
M = *Areopaguristes engyops* (1), *Paguristes eremita* (1), *P. syrtensis* (4), *Paguristes rubropictus* (4), *P. seminudus* (1)
N = *Dardanus crassimanus* (8), *D. setifer* (1), *D. fucosus* (2), *D. lagopodes* (24), *D. sanguinocarpus* (6), *Dardanus megistos* (1)
O = *Dardanus aspersus* (1), *D. callichela* (2)
P = *Aniculus aniculus* (1), *A. erythraeus* (2)

A.1.2 Paguridae

Q = *Anapagurus alboranensis* (3), *A. chiroacanthus* (1)
R = *Anapagurus bicorniger* (1), *A. petiti* (1)
S = *Discorsopagurus schmitti* (1), *Pagurus arcuatus* (6), *P. kennerlyi* (3), *P. pectinatus* (3), *P. constans* (1), *P. pubescens* (40), *P. undosus* (1), *P. rathbuni* (2), *P. trigonocheirus* (1), *P. granosimanus* (9); *Elassochirus cavimanus* (4), *E. gilli* (4), *E. tenuimanus* (6), *Pagurus acadianus* (4), *P. bernhardus* (169), *P. aleuticus* (1), *P. middendorffii* (2); *Pagurus beringanus* (5), *P. brachiomastus* (26), *P. nigrofascia* (2), *P. lanuginosus* (3), *P. simulans* (8), *P. maculosus* (2), *P. proximus* (10), *P. quinquelineatus* (1), *P. rectidactylus* (1), *P. caurinus* (1); *Pagurus filholi* (1), *P. japonicus* (1), *P. gracilipes* (1), *P. minutus* (72); *Pagurus hirsutiusculus* (33), *P. samuelis* (53); *Pagurus criniticornis* (6)
T = *Goreopagurus poorei* (1), *Pagurus pollicaris* (4), *Raripagurus roseangelae* (1), *Pagurus longicarpus* (46), *P. boriaustraliensis* (5), *P. pitagsaleei* (4), *P. fraserorum* (3), *Pagurixus festinus* (1), *P. ruber* (1), *Propagurus deprofundis* (4), *Pagurus hirtimanus* (1), *P. pseudosculptimanus* (3)
U = *Anapagurus hendersoni* (2)
V = *Anapagurus hendersoni* (2)

A.1.3 Coenobitidae

W = *Birgus latro* (22)
X = *Coenobita brevimanus* (16), *C. violascens* (4), *C. cavipes* (1), *C. clypeatus* (1), *C. perlatus* (2), *C. lila* (6), *C. purpureus* (16), *C. longitarsis* (4), *C. rugosus* (12), *C. pseudorugosus* (15)

A.2 16S

A.2.1 Paguridae

A = *Anapagurus alboranensis* (3), *A. breviaculeatus* (2), *A. pusillus* (6), *A. laevis* (6), *A. bicorniger* (3), *A. petiti* (1), *A. chiroacanthus* (1), *A. curvidactylus* (1), *A. hyndmanni* (2), *A. longispina* (3); *Spiropagurus elegans* (2); *Paguridium minimum* (2), *Pagurus pseudosculptimanus* (4), *P. cuanensis* (4), *P. prideaux* (7), *P. alatus* (3), *P. pubescentulus* (5), *Pagurus mbizi* (3), *P. excavatus* (3), *P. forbesii* (1); *Iridopagurus caribbensis* (1), *Phimochirus formani* (2), *P. holthuisi* (2), *P. tunnelli* (3), *P. operculatus* (2), *P. holthuisi* (2)
B = *Catapaguroides microps* (1), *Catapagurus cracens* (1), *C. tenuilamina* (1); *Pagurus brevidactylus* (3), *P. provenzanoi* (1), *P. criniticornis* (4), *P. maclaughlinae* (2), *P. leptonyx* (2), *P. villosus* (1); *Pagurus heblingi* (1); *Goreopagurus piercei* (1); *Manucomplanus ungulatus* (2), *Rhodochirus rosaceus* (1), *Pagurus bullisi* (2), *Tomopagurus merimaculosus* (1), *Protoniopagurus bioperculatus* (1), *Pylopagurus discoidalis* (2); *Pagurus anachoretus* (3), *P. carneus* (2), *P. comptus* (3), *P. forceps* (3), *Pylopaguridium markhami* (1)
C = *Elassochirus cavimanus* (2), *E. gilli* (1), *E. tenuimanus* (1), *Pagurus bernhardus* (125), *P. ochotensis* (8), *P. brachiomastus* (27), *P. simulans* (8), *P. caurinus* (1), *P. middendorffii* (6), *P. minutus* (14), *P. granosimanus* (1), *P. proximus* (15), *P. rectidactylus* (1), *P. kennerlyi* (1), *P. pectinatus* (11), *P. pubescens* (3), *P. aleuticus* (1), *Pagurixus eminens* (1), *Nematopagurus longicornis* (2), *N. marianicus* (1), *Propagurus gaudichaudii* (1)
D = *Pagurus exilis* (1), *P. gladius* (1), *P. perlatus* (1), *P. pollicaris* (3)
E = *Pagurus chevreuxi* (4)
F = *Cestopagurus timidus* (2)
G = *Cestopagurus timidus* (2)
H = *Pagurus hirsutiusculus* (1), *P. japonicus* (1)
I = *Pagurus lanuginosus* (2), *P. maculosus* (2), *P. longicarpus* (1)

A.2.2 Diogenidae

J = *Areopaguristes hewatti* (1), *Paguristes anomalus* (1), *P. tortugae* (5), *P. hernancortezi* (1), *P. robustus* (1), *P. doederleini* (1), *P. miyakei* (1), *Pseudopaguristes calliopsis* (1), *P. marocanus* (1), *Areopaguristes tudgei* (1)
K = *Areopaguristes cyanops* (1), *A. mauritanicus* (2)
L = *Paguristes albimaculatus* (1), *P. erythrops* (2), *P. sericeus* (2), *P. wassi* (2), *P. grayi* (1), *P. inconstans* (1), *P. spinipes* (2), *P. moorei* (2), *P. triangulatus* (2), *Paguropsina inermis* (5), *P. pistillata* (3), *Paguropsis gigas* (1), *P. lacinia* (3), *P. andersoni* (1), *P. typica* (5), *P. confusa* (3), *Paguristes eremita* (1), *P. syrtensis* (1), *P. rubropictus* (3), *P. calvus* (1), *P. seminudus* (1)
M = *Aniculus retipes* (1), *Dardanus venosus* (2), *D. fucosus* (2), *D. arrosor* (4), *D. insignis* (4), *D. pectinatus* (1), *D. crassimanus* (1), *D. lagopodes* (3), *D. sanguinocarpus* (1), *D. setifer* (1), *D. guttatus* (1), *D. aspersus* (1), *D. hessii* (2), *D. deformis* (2), *D. gemmatus* (1), *D. impressus* (1); *Petrochirus Diogenes* (2), *P. pustulatus* (2)
N = *Clibanarius aequabilis* (2), *C. erythropus* (3), *C. antillensis* (41), *C. tricolor* (28), *C. corallinus* (6), *C. merguiensis* (5), *C. virescens* (3), *C. infraspinatus* (6), *C. rutilus* (2), *C. snelliusi* (2), *C. longitarsus* (6), *C. sclopetarius* (2), *C. vittatus* (15), *C. striolatus* (1), *C. padavensis* (6), *C. taeniatus* (1), *C. eurysternus* (4)
O = *Diogenes arguinensis* (8), *D. erythromanus* (2), *D. minimus* (1), *D. goniochirus* (1), *D. nitidimanus* (1), *D. rectimanus* (1), *D. armatus* (11), *D. pugilator* (12), *D. curvimanus* (18), *Isocheles sawayai* (4), *I. wurdemanni* (2), *Loxopagurus loxochelis* (2)
P = *Ciliopagurus caparti* (2), *C. hawaiiensis* (1), *C. galzini* (2), *C. strigatus* (4), *C. tricolor* (3), *C. vakovako* (2), *Pseudopagurus granulimanus* (1), *Diogenes ovatus* (3)
Q = *Calcinus albengai* (1), *C. inconspicuous* (1), *C. dapsiles* (1), *C. argus* (2), *C. anani* (1), *C. fuscus* (1), *C. laevimanus* (3), *C. morgani* (3), *C. seurati* (1), *C. elegans* (2), *C. orchidae* (2), *C. gaimardii* (3), *C. vanninii* (1), *C. isabellae* (1), *C. imperialis* (1), *C. exploratory* (1), *C. californiensis* (2), *C. obscurus* (3), *C. tibicen* (1), *C. talismani* (1), *C. tibicen* (30), *C. pascuensis* (1), *C. spicatus* (2), *C. haigae* (2), *C. minutus* (1), *C. rosaceus* (2), *C. nitidus* (1), *C. hazletti* (1), *C. vachoni* (3), *C. gouti* (2), *C. hakahau* (2), *C. laurentae* (1), *C. lineapropodus* (2), *C. pulcher* (3), *C. tubularis* (4), *C. verrillii* (1), *C. latens* (3)
R = *Calcinus guamensis* (2)

A.2.3 Coenobitidae

S = *Coenobita violascens* (1), *C. purpureus* (3), *C. rugosus* (2), *C. cavipes* (3), *C. perlatus* (2), *Birgus latro* (2)
T = *Coenobita clypeatus* (1), *C. compressus* (1), *C. lila* (1), *C. variabilis* (1)

A.3 28S

A.3.1 Paguridae

A = *Phimochirus randalli* (2)
B = *Iridopagurus caribbensis* (1), *I. reticulatus* (1)
C = *Bythiopagurus macrocolus* (1), *Manucomplanus ungulatus* (1), *Pylopagurus discoidalis* (1), *Porcellanopagurus filholi* (1), *Pylopaguridium markhami* (1), *Pagurus alatus* (4), *P. cuanensis* (2), *P. prideaux* (4), *P. excavatus* (3), *Xylopagurus cancellarius* (1), *Pagurus hirsutiusculus* (1), *P. bernhardus* (2*), P. pubescents* (2), *Labidochirus splendescens* (1), *Pagurus longicarpus* (1), *Pagurus pollicaris* (3), *Agaricochirus alexandri* (1), *Pagurus bullisi* (1), *Tomopagurus merimaculosus* (1)

A.3.2 Diogenidae

D = *Areopaguristes hewatti* (2)
E = *Areopaguristes hummi* (2), *A. pilosus* (1), *Paguristes turgidus* (1); *Paguristes puncticeps* (1), *P. sericeus* (1), *P. triangulates* (1)
F = *Calcinus laevimanus* (1), *C. obscurus* (1); *Clibanarius albidigitus* (1), *C. antillensis* (1), *C. corallinus* (1), *C. vittatus* (1), *Dardanus arrosor* (1), *D. insignis* (1), *D. fucosus* (2)

A.3.3 Coenobitidae

G = *Birgus latro* (2), *Coenobita compressus* (1), *C. perlatus* (1)

Phylogenetic analysis based on three gene markers, COI, 16S, and 28S, reconstructed the phylogenetic relationships of the Diogenidae, Paguridae, and Coenobitidae families. To improve visual clarity and analysis, clades with low support values (bootstrap <50) were collapsed using Figtree. Analysis of the phylogenetic tree based on the COI gene reveals that the Coenobitidae family, including the genera *Coenobita* and *Birgus latro,* exhibits strong monophyleticity with high support values, highlighting the reliability of the COI marker in reconstructing evolutionary relationships within this group. Conversely, the Diogenidae and Paguridae families display notable clade instability, particularly in the genera *Calcinus, Clibanarius, Dardanus* and *Pagurus*, which are divided into multiple small groups with low support **(Figure 1A)**.

Phylogenetic analysis based on the 16S gene revealed a more stable phylogenetic pattern than that of the COI tree, particularly in reconstructing relationships at the family level **(Figure 1B)**. The Coenobitidae family showed strong monophyleticity. Meanwhile, the Diogenidae and Paguridae families showed increased stability compared with the COI tree, although some genera, such as *Dardanus* and *Clibanarius*, still exhibited clade structure instability, especially in groups with sequences from different geographic regions. The 28S tree showed a more stable and well-defined structure for the Diogenidae and Paguridae **families (Figure 1C)**. The genera within Diogenidae, such as *Calcinus, Clibanarius,* and *Dardanus* tend to form consistent monophyletic groups, although relationships between genera still showed complexity. Within the Paguridae grouping of genera, such as *Pagurus, Phimochirus,* and *Iridopagurus* show a more stable pattern was observed than that in the COI and 16S trees.

3.4. Discussion

Across all three examined families, the ratio between available sequences and the number of species varies substantially among genetic markers, reflecting differences between these genes and their representation in public repositories **(Table 1)**. COI exhibits the highest ratios, consistent with its role as the primary barcoding marker for animals (Hebert et al., 2003). Elevated ratios typically indicate broader sampling across geographic and populations, which can reveal deeper intraspecific structure or cryptic diversity within species. However, high ratios may also elevate the possibility of problematic entries, as many GenBank submissions originate from various studies and ontogenetic stages that may not meet consistent taxonomic standards (Schlick-Steiner et al., 2010; Collins and Cruickshank, 2013). In contrast, 16S generally shows moderate ratios, whereas 28S shows the lowest, reflecting a more limited application for species-level diagnostics and reduced capacity to capture genetic differences in variation.

However, a larger number of available sequences does not equate to higher reliable data. Public databases often contain sequences that are untrimmed, have low-quality bases, or have been edited inconsistently from the chromatograms, and entries lacking sufficient morphological validation, particularly in taxa with high morphological plasticity, such as hermit crabs (McLaughlin et al., 2007; Schubart, 2009; Mclaughlin et al., 2010; Collins and Cruickshank, 2013). Several studies have shown that misidentification is not rare; sequences labeled to a single species often represent distinct genetic lineages or even different species (Vivien and Martin, 2025).

Our review analysis of COI, 16S, and 28S sequences from Diogenidae, Pagurodae, and Coenobitidae highlights consistent patterns related to species delimitation and phylogenetic inference within Paguroidea. Across all three families, the datasets reveal substantial heterogeneity in genetic divergence, with wide and often overlapping ranges of intraspecific and interspecific distances. This pattern is especially evident in the mitochondrial markers COI and 16S, undermining the applicability of a universal genetic threshold for species delimitation.

For COI, intraspecific distances span from 0–0.051 in Coenobitidae (mean 0.016), 0–0.075 in Diogenidae (mean 0.009), and 0–0.025 in Paguridae (mean 0.005). However, interspecific

distances show extensive overlap with these ranges, particularly in Coenobitidae (0–0.374; mean 0.195) and Diogenidae (0–0.046; mean 0.017). Similarly, in 16S, overlap between intra- and interspecific distances is evident, with intraspecific values spanning 0–0.015 in Coenobitidae, 0–0.066 in Diogenidae, and 0–0.045 in Paguridae, while interspecific distances reach up to 0.020, 0.070, and 0.104, respectively **(Table 2)**. These findings align with observations in other decapod species, where variation in mitochondrial evolutionary rates and historical demography frequently blur the expected barcode gaps (Schubart, 2009; Antu et al., 2024).

The degree of overlap between intra-and interspecific patterns strongly suggests that several nominal species in Paguroidea may represent species complexes rather than single, cohesive evolutionary lineages. Comparable patterns have been repeatedly documented in freshwater crabs, where markedly divergent mitochondrial clades occur within traditionally defined species, producing COI distances that fall within interspecific ranges (Daniels et al., 2006, 2025). Such patterns underscore how geographic isolation, historical fragmentation, and limited dispersal capabilities can generate cryptic diversity, resulting in intraspecific divergence values comparable to the upper values observed in Coenobitidae and Diogenidae in our dataset.

Similar issues have been widely recognized in marine decapods, including hermit crabs. COI may produce unexpectedly low interspecific distances due to incomplete lineage sorting or introgression, particularly in species with broad distributions or recent diversification (Schubart, 2009). Furthermore, hermit crabs are known for high levels of morphological plasticity, convergence, and homoplasy (McLaughlin et al., 2007), making morphological diagnoses unreliable and increasing the likelihood that species boundaries mask multiple divergent mitochondrial lineages. Recent findings in Paguridae (Sultana et al., 2022) have also revealed deep mitochondrial divergence within a single species and unusually low divergence between morphologically distinct species, paralleling the patterns found here.

Conversely, unusually high intraspecific values, such as those approaching 0.05 in COI for some Coenobitidae and Diogenidae species, and values up to 0.045 in 16S for several Paguridae taxa, suggest the presence of cryptic diversity and lineage structuring. In particular, intraspecific COI values exceeding 0.02–0.03 and 16S values exceeding 0.01–0.02 are widely regarded as indicative of species complexes in crustaceans. Several taxa in this study exceed these thresholds, warranting further taxonomic reassessment.

In contrast to mitochondrial markers, the nuclear 28S gene shows very low intraspecific variation (0.007 across families) and modest but distinguishable interspecific divergence (up to 0.090 in Paguridae). These patterns are consistent with its slow evolutionary rate and previously reported utility for distinguishing deeper phylogenetic splits rather than closely related species (McLaughlin et al., 2007; Schubart, 2009; Tsang et al., 2011; Bracken-Grissom et al., 2013). Nevertheless, the detectable levels of interspecific divergence support its continued relevance when incorporated into multilocus phylogenetic frameworks.

Overall, this study provides a systematic reassessment of publicly available COI, 16S, and 28S sequences for Coenobitidae, Diogenidae, and Paguridae, offering the most comprehensive evaluation of genetic divergence patterns across these families. By quantifying instances of overlap or unusually high divergence, our results established an empirical reference framework that can be used to guide species identification and molecular diagnosis in future hermit crab studies. Rather than proposing taxonomic changes, our primary aim was to provide a validated and curated baseline against which newly generated sequences can be compared. These genetic distance thresholds derived from a large, cleaned, and standardized dataset offer practical benchmarks for molecular species delimitation, particularly for researchers working with COI, 16S, and 28S markers in Paguroidea. We anticipate that the divergence values presented here will serve as a foundation for future taxonomic, phylogenetic, and biodiversity assessments involving the species and families included in this study.

4. Conclusions

This study reviewed 4,871 COI, 16S, and 28S sequences from GenBank representing the three main families of Paguroidea: Coenobitidae, Diogenidae, and Paguridae. The results revealed

an extensive overlap between intra-and interspecies variation in the COI and 16S genes, whereas the 28S gene exhibited a much lower level of divergence, reflecting the slower evolutionary rate of nuclear markers. High intraspecific divergence in several species suggests the presence of species complexes or cryptic lineages. This study provides a curated empirical baseline of genetic distances to support molecular species identification, delimitation, and validation of public sequence data in these families.

Conflicts of Interest

There are no conflicts to declare.

Acknowledgements

We would like to express sincere gratitude to Prof. Dr. Ir. Dwi Listyo Rahayu for her invaluable insights and guidance throughout the discussions, which greatly contributed to the development of this work, particularly in shaping the writing, refining the ideas, and deepening our understanding of hermit crab taxonomy.

AI Writing Statement

During the preparation of this work, the authors used Grammarly and Google Translate to correct the grammar and helped in writing this manuscript. After using those tools, the authors reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2

Antu, D.R., Jahan, I., Islam, T.T., Mahjabin, M., Datta, S.K., Ahmed, S., & Ahmed, M.S. (2024). Morphology and DNA barcoding reveal four new species of hermit crab (Crustacea: Decapoda: Anomura) for the fauna of Bangladesh. *Zoology and Ecology*, *34*(2), 144-153. https://doi.org/10.35513/21658005.2024.2.8

Asakura, A. (2001). A revision of the hermit crabs of the genera Catapagurus A. Milne-Edwards and Hemipagurus Smith from the Indo-West Pacific (Crustacea : Decapoda : Anomura : Paguridae). *Invertebrate Systematics*, *15*(6), 823. https://doi.org/10.1071/IT00034

Bokulich, A. (2020). Understanding scientific types: holotypes, stratotypes, and measurement prototypes. *Biology & Philosophy*, *35*(5), 54. https://doi.org/10.1007/s10539-020-09771-1

Bracken-Grissom, H.D., Cannon, M.E., Cabezas, P., Feldmann, R.M., Schweitzer, C.E., Ahyong, S.T., Felder, D.L., Lemaitre, R., & Crandall, K.A. (2013). A comprehensive and integrative reconstruction of evolutionary history for Anomura (Crustacea: Decapoda). *BMC Evolutionary Biology*, *13*(128), 1-28. https://doi.org/10.1186/1471-2148-13-128

Collins, R.A., & Cruickshank, R.H. (2013). The seven deadly sins of DNA barcoding. *Molecular Ecology Resources*, *13*(6), 969-975. https://doi.org/10.1111/1755-0998.12046

Daniels, S.R., Cumberlidge, N., Pérez-Losada, M., Marijnissen, S.A.E., & Crandall, K.A. (2006). Evolution of Afrotropical freshwater crab lineages obscured by morphological convergence. *Molecular Phylogenetics and Evolution*, *40*(1), 227–235. https://doi.org/10.1016/j.ympev.2006.02.022

Daniels, S.R., Marais, H.J., Da Fonseca, M., & Peer, N. (2025). Mind the gap: DNA barcoding as a tool to delineate evolutionary units in Afrotropical freshwater crabs (Decapoda: Brachyura: Potamoidea: Deckeniidae and Potamonautidae). *Journal of Crustacean Biology*, *45*(4), 1-11. https://doi.org/10.1093/jcbiol/ruaf061

Hadadi, N., Madduppa, H., Shita Azaraly, I., & Anggraini, N.P. (2023). DNA Barcoding, Identifikasi Morfologi dan Kepadatan Populasi Genus Tridacna di Perairan Maluku Utara. *Jurnal Ilmu Dan Teknologi Kelautan Tropis*, *15*(3), 349-368. https://doi.org/10.29244/jitkt.v15i3.47110

Hamasaki, K., Iizuka, C., Sanda, T., Imai, H., & Kitada, S. (2017). Phylogeny and phylogeography of the land hermit crab *Coenobita purpureus* (Decapoda: Anomura: Coenobitidae) in the Northwestern Pacific Region. *Marine Ecology*, *38*(1), 1-17. https://doi.org/10.1111/maec.12369

Hebert, P.D.N., Cywinska, A., Ball, S.L., & DeWaard, J.R. (2003). Biological identifications through DNA barcodes. *Proceeding of Royal Society London*, *270*(1512), 313-321. https://doi.org/10.1098/rspb.2002.2218

ICZN. (1999). *Article 72. General provisions*. International Code of Zoological Nomenclature (4th Ed.). London, United Kingdom: The International Trust for Zoological Nomenclature.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*(2), 111-120. https://doi.org/10.1007/BF01731581

Komai, T., & Rahayu, D.L. (2013). Six new species of the hermit crab genus Decaphyllus de Saint Laurent, 1968 (Crustacea: Decapoda: Anomura: Paguridae) from the Boholo Sea, The Philippines, and the Ryukyu Islands, Japan. *Raffles Bulletin of Zoology*, *61*(2), 589-620.

Komai, T., & Rahayu, D.L. (2014). New records and new species of the hermit crab genus Pagurus Fabricius, 1775 (Crustacea: Decapoda: Anomura: Paguridae) from the Philippines. *Raffles Bulletin of Zoology*, *62*, 620-646.

Komai, T., Rahayu, D.L., & Fujita, Y. (2021). Hermit crabs of the family paguridae (Crustacea: Decapoda: Anomura) from christmas and cocos (keeling) islands, indian ocean territory of australia, with descriptions of three new species and reappraisal of the genus kumepagurus komai & osawa, 2012. *Raffles Bulletin of Zoology*, *69*(1), 378-402. https://doi.org/10.26107/RBZ-2021-0059

Lemaitre, R., & Felder, D.L. (2023). A new genus and species of micro-pagurid hermit crab (Crustacea, Decapoda, Paguridae) from the rariphotic reef zone in Curaçao, Lesser Antilles, Caribbean Sea. *Zootaxa*, *5277*(2), 374-380. https://doi.org/10.11646/ZOOTAXA.5277.2.7

Lemaitre, R., Rahayu, D.L., & Komai, T. (2018). A revision of “blanket-hermit crabs” of the genus paguropsis henderson, 1888, with the description of a new genus and five new species (Crustacea, anomura, diogenidae). *ZooKeys*, *2018*(752), 17-97. https://doi.org/10.3897/zookeys.752.23712

Malay, M.C.M.D., & Paulay, G. (2010). Peripatric speciation drives diversification and distributional pattern of reef hermit crabs (Decapoda: Diogenidae: Calcinus). *Evolution*, *64*(3), 634-662. https://doi.org/10.1111/j.1558-5646.2009.00848.x

Mclaughlin, P.A., Komai, T., Lemaitre, R., & Rahayu, D.L. (2010). Annotated Checklist of Anomuran Decapod Crustaceans of the World (Exclusive of the Kiwaoidea and Families Chirostylidae and Galatheidae of the Galatheoidea) Part I – Lithodoidea, Lomisoidea and Paguroidea. *Raffles Bulletin of Zoology*, *23*, 5-107.

McLaughlin, P.A., Lemaitre, R., & Sorhannus, U. (2007). Hermit crab phylogeny: A reappraisal and its “fall-out.” *Journal of Crustacean Biology*, *27*(1), 97-115. https://doi.org/10.1651/S-2675.1

McLaughlin, P.A., & Rahayu, D.L. (2008). A new genus and species of hermit crab of the family Paguridae (Crustacea: Anomura: Paguroidea) from the Vanuatu Archipelago. *Proceedings of the Biological Society of Washington*, *121*(3), 365-373. https://doi.org/10.2988/08-07.1

Rahayu, D.L., & Komai, T. (2000). Shallow-Water Hermit Crabs (Crustacea: Decapoda: Anomura: Diogenidae and Paguridae) of Phuket, Thailand. *Phuket Mar. Biol. Cent. Res. Bull.*, *63*, 21-44.

Rahayu, D.L., & Komai, T. (2013). Two new species of Pylopaguropsis alcock Crustacea: Decapoda: Anomura: Paguridae) from the Philippines. *Raffles Bulletin of Zoology*, *61*(2), 621-631.

Rahayu, D.L., Shih, H.Te, & Ng, P.K.L. (2016). A new species of land hermit crab in the genus Coenobita Latreille, 1829 from Singapore, Malaysia and Indonesia, previously confused with C. cavipes Stimpson, 1858 (Crustacea: Decapoda: Anomura: Coenobitidae). *Raffles Bulletin of Zoology*, *34*, 470-488.

Ratnasingham, S., & Hebert, P.D.N. (2007). BOLD: The Barcode of Life Data System: Barcoding. *Molecular Ecology Notes*, *7*(3), 355-364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E., & Crozier, R.H. (2010). Integrative taxonomy: A multisource approach to exploring biodiversity. In *Annual Review of Entomology* (Vol. 55, pp. 421-438). https://doi.org/10.1146/annurev-ento-112408-085432

Schubart, C.D. (2009). Mitochondrial DNA and Decapod Phylogenies. In *Decapod Crustacean Phylogenetics* (pp. 47-65). CRC Press. https://doi.org/10.1201/9781420092592-c4

Sultana, Z., Babarinde, I.A., & Asakura, A. (2022). Diversity and Molecular Phylogeny of Pagurid Hermit Crabs (Anomura: Paguridae: Pagurus). *Diversity*, *14*(2). https://doi.org/10.3390/d14020141

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, *38*(7), 3022-3027. https://doi.org/10.1093/molbev/msab120

Thompson, J.D., Higgins, D.G., & Gibson, T.J. (1994). *CLUSTAL W : improving the sensitivity of progressive multiple sequence alignment through sequence weighting , position-specific gap penalties and weight matrix choice*. *22*(22), 4673-4680. https://doi.org/10.1093/nar/22.22.4673

Tsang, L. M., Chan, T.Y., Ahyong, S.T., & Chu, K.H. (2011). Hermit to king, or hermit to all: Multiple transitions to crab-like forms from hermit crab ancestors. *Systematic Biology*, *60*(5), 616-629. https://doi.org/10.1093/sysbio/syr063

Vivien, R., & Martin, P. (2025). Maintaining taxonomic accuracy in genetic databases: A duty for taxonomists—Reanalysis of the DNA sequences from Mercan et al. (2024) on the genus Potamothrix (Annelida, Clitellata) in Turkish lakes. *Zootaxa*, *5575*(4), 555-562. https://doi.org/10.11646/zootaxa.5575.4.5

Wong, T., Ly-Trong, N., Ren, H., Baños, H., Roger, A., Susko, E., Bielow, C., De Maio, N., Goldman, N., Hahn, M., Huttley, G., Lanfear, R., & Minh, B. Q. (2025). *IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models*. https://doi.org/10.32942/X2P62N