HTTPS://DOI.ORG/10.48047/AFJBS.6.SI4.2024.5875-5883

African Journal of Biological Sciences

Molecular Phylogenetic Study Of *Macrotermes gilvus* Termite DNA Barcoding Using Cytochrome Oxidase I Gene.

G. GopalaRao[1] and Apka Nageswara Rao[1*].

[1*]Environmental Biological and Medical Entomology Laboratory Department of Zoology, Nizam College, Osmania University, Hyderabad, Telangana State, India. E-Mail: drapkanagesh@gmail.Com, Contact No.: +91 9849043610.

Article History
Volume 6, Issue Si4, Aug 2024
Received: 25/07/2026
Accepted: 08/08/2024
Published: 29-08-2024
doi: 10.48047/AFJBS.6.Si4.2024.5875-5883

Abstract

Macrotermes gilvus is a termite species of immense ecological value due to its crucial role in nutrient turnover and soil formation. A critical need for ecological research and pest management strategies is to gain an understanding of its genetic diversity as well as evolutionary relationships. This project used molecular phylogenetic analysis of *M. gilvus* genetic make-up through DNA barcoding the cytochrome oxidase I (COI) gene. Simple Sequence Repeats analysis indicates that *Macrotermes carbonarius* holds much less variation in terms of polymorphism along with other parameters such frequency changes than *M. gilvus* in local populations. Using COBF and COBR primers we were amplified a 400 base pair (bp) fragment of the COI gene by polymerase chain reaction (PCR), successfully gaining a discrete amplicon band on an agarose gel. The PCR product was then purified to remove contaminants. Forward and reverse DNA sequencing reactions were performed with the same primers of the BDT v3.1 Cycle Sequencing Kit on an ABI 3730xl Genetic Analyzer following procedures detailed elsewhere.

The high-quality sequence data obtained from our study will contribute to understanding both genetic variant diversity among M. gilvus samples as well as where they are to be placed within the genus *Macrotrends.* In the evolutionary context this work has wide implications for both practice and policy on what condition *M.c.gilvus* exists. Future studies could expand on this work by analysing additional genetic markers and comparing populations from different parts of the world.

.**KEYWORDS** : Macrotermes gilvus, DNA barcoding, COI , Molecular phylogenetics, Termite genetic diversity, Cytochrome oxidase I (COI), PCR amplification.

1.INTRODUCTION

The molecular phylogenetic study of *Macrotermes gilvus* termites using DNA barcoding with the COI gene can provide significant insights into their genetic diversity, evolutionary relationships, and species identification. DNA barcoding, particularly using the cytochrome c oxidase subunit I (COI) gene, is a widely accepted method for species identification and phylogenetic studies due to its high resolution and reliability across various taxa [1] [6]. The COI gene has been successfully used in numerous studies to delineate species boundaries and understand genetic divergence, as seen in the analysis of Southern River Terrapins and other terrapins, which revealed new haplotypes and potential cryptic species [1]. Similarly, the COI gene has been effective in identifying and differentiating closely related species, such as Betta burdigala and Betta uberis, which showed a high genetic similarity of 96.66% based on COI sequences [3]. In the context of termites, a study on *Coptotermes gestroi* using COI and 16S rRNA genes demonstrated the utility of these markers in

confirming species identity and understanding population homogeneity across different habitats [9]. The COI gene's effectiveness in species delimitation is further supported by its application in Orthoptera, where it showed a high probability of correct identification and species richness estimation through methods like Automatic Barcode Gap Discovery (ABGD) and Assemble Species by Automatic Partitioning (ASAP) [5]. However, challenges such as high intraspecific genetic variation, which can lead to false positives in species delimitation, have been noted in insects, with approximately one-quarter of insect species showing significant genetic variation in COI sequences [10]. This issue underscores the importance of using appropriate thresholds and algorithms, such as "thresh Opt" and "local Minima," to avoid overestimating species diversity [10]. In the case of belowground invertebrates like Collembola and Oribatida, the presence of a barcoding gap is crucial for accurate taxonomic assignment, although it often fails due to overlapping intra- and interspecific genetic distances [2]. For crayfishes, the assessment of local and global barcoding gaps revealed that current taxonomy might be inadequate, suggesting the need for taxonomic revisions to improve the utility of DNA barcoding [8]. The use of Chelex-based DNA extractions, which are simple and effective for obtaining quality DNA for barcoding, can facilitate such studies even in non-specialized settings, making it accessible for broader research applications [7]. Overall, the molecular phylogenetic study of *Macrotermes gilvus* using the COI gene can benefit from these insights, leveraging the gene's proven efficacy in various taxa while addressing potential challenges through careful methodological considerations and complementary genetic markers.

2. MATERIALS AND METHODS

2.1 Sample Collection and morphological identification

100 samples of termites were gathered from six (6) locations within the Dhandakaranya forest, specifically the Bhadrachalam region in Telangana, India (see Figure 1). These specimens were then identified following the protocol outlined by Chottani (1997), preserved in absolute ethanol, and stored at temperatures between 4-8°C until the extraction of genomic DNA took place.

2.2 Genomic DNA isolation, PCR amplification of COI gene and sequencing

The entire genomic DNA was isolated from 25 mg of leg tissue by utilizing the Dneasy Blood & Tissue DNA kit (Qiagen, Germany) in accordance with standard procedures. Subsequently, the tissue underwent desiccation and was amassed in a 1.5ml Eppendorf tube, followed by the addition of 180 µl ATL buffer and homogenization using a micro pestle. A further addition of 20 µl proteinase K ensued, which was thoroughly mixed through vortexing and then incubated at 56°C until complete lysis of the tissue was achieved. An introduction of 200 µl AL buffer to the sample was carried out, culminating in a comprehensive mixing. This was succeeded by the addition of 200 µl of 100% ethanol to the Eppendorf tube, which was once again mixed through vortexing. The resultant solution was then transferred into mini spin columns containing a silica membrane that selectively binds the genomic DNA, and subsequently subjected to centrifugation at 8000 rpm for 1 minute. The flowthrough obtained was discarded, following which the column underwent two additional washes with distinct buffers provided in the kit, namely AW1 centrifuged for 1 minute at 8000rpm and AW2 centrifuged at 14000 rpm for 3 minutes. The flow-through and collection tube were both discarded. These mini spin columns were then relocated to pristine 2 ml microcentrifuge tubes, and 200µl Buffer AE was directly dispensed onto the DNeasy membrane. After an incubation period at room temperature for 1 minute, centrifugation was executed for 1 minute at 8000rpm to elute the DNA. The DNA was subsequently preserved at -20°C post verification on a 1.0% agarose gel (Figure 2).

The obtained DNA was validated through 1% agarose gel electrophoresis employing a standard procedure. To amplify the mtCOI segment, the primer pair COBF and COBR primers, as described by Folmer et al. in 1994, was utilized. A 25 µl PCR reaction mixture was prepared, consisting of 10 pmol of each primer, 100-200 ngDNA template, 1X PCR buffer, 2.5 mMMgCl2, 0.1 mM each dNTP, and 1 U high-fidelity Taq DNA Polymerase. The PCR thermal profile included an initial denaturation at 94°C for 10 min, followed by 40 cycles at 94°C for 30 sec, 46°C for 30s, and 72°C for 1 min, concluding with storage at 4°C. The amplification process was carried out using the 2720 thermal cycler from Applied Biosystems. Subsequently, the PCR product underwent purification using the Invitrogen PCR product purification kit. The amplified DNA was then analyzed on a 2% agarose gel alongside a 100 bp DNA ladder, with gel visualization performed using the BIO-RAD gel Doc XR gel documentation system. Sequencing was conducted bidirectionally on a 48 capillary array ABI3130 automated sequencer from Applied Biosystems, following the Sanger sequencing method.

2.3 Sequence analysis & Dataset preparation

The data sequence was acquired through chromatograms and submitted to GenBank for the acquisition of accession numbers. The chromatograms were modified to eliminate uncertain bases, then aligned utilizing the Basic Local Alignment Search Tool (BLAST), with sequences belonging to the same or similar genera retrieved from the nucleotide database (PUBMED) of the National Center for Biological Information (NCBI). The COI nucleotide sequences of the termite species examined in this study were aligned and juxtaposed with the species retrieved from PUBMED, employing CLUSTAL W alignment.

2.4 Phylogenetic analysis

The GenBank database was utilized to examine the COI sequence of closely related termite species that is available to the public. Utilizing Clustal W software, the aligned sequences were converted to MEGA format through the application of MEGA 11 software. A dataset comprising mtCO1 sequences from 100 termite samples, inclusive of the out-group, underwent phylogenetic analysis. The phylogenetic analysis was conducted based on the optimality criteria of Neighbour-joining (NJ) and the Maximum Likelihood Method (ML). By performing 1000 replicates, bootstrap values were calculated for the analysis.

3. RESULTS AND DISCUSSION

3.1 Sample Collection and morphological identification:

Species: ***Macrotermes gilvus*** collected in Dandakaranyam forest, Bhadrachalam Region, Telangana State.

Identified as *Macrotermes gilvus* through Standard keys.

1. Study Sites

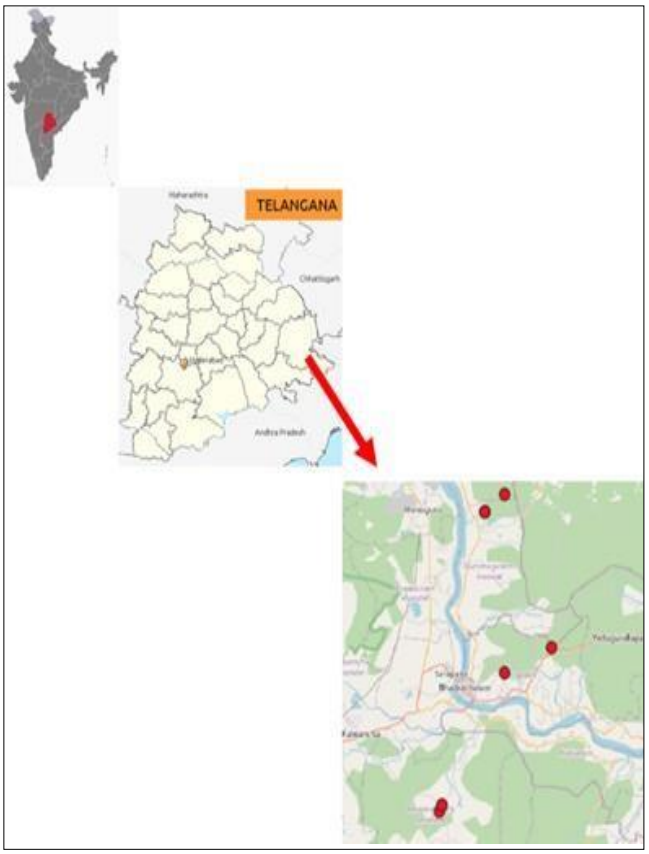


Figure1: Dandakaranyam forest, Bhadrachalam Region, Telangana State.

The genomic DNA was collected from 100 populations of termites and the COI gene was characterized, using COBF and COBR primers. The amplified COI product was sequenced. The sequence was submitted to NCBI genbank and the accession numbers were obtained for the termite populations (Table 1). PCR amplification of the genomic DNA of the yielded partial mtCO1 sequences of varied lengths ranging between 421-600 bp.

3.2 Genomic DNA and CO I gene AmplicongDNA COI gene Ladder specification

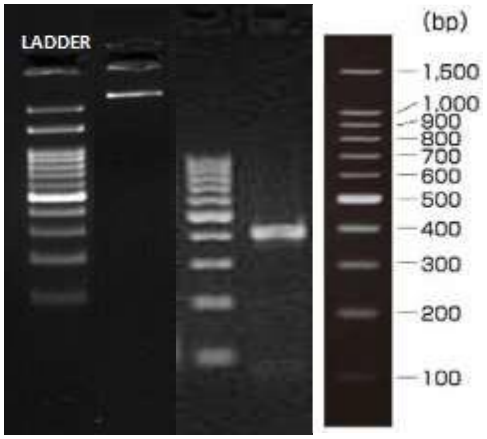


Figure 2: Fragment of COI gene was amplified by COBF and COBR primers. A single discrete PCR amplicon band of 400 bp was observed when resolved on agarose gel. The PCR amplicon was purified to remove contaminants.

3.3. Sequence analysis & Dataset preparation

Forward and reverse DNA sequencing reaction:

Forward and reverse DNA sequencing reaction of PCR amplicon was carried out with COBF and COBR primers using BDT v3.1 Cycle sequencing kit on ABI 3730xl Genetic Analyzer.

Macrotermes gilvus **Sanger Seq Chromatogram data file Data:**

>**Forward Seq data**

ACCTTCAATAAGTCATCCCATACATAGGGAAGAATCGTACAATGAGTATGGGGAGGATTCGCAGTAGACA
ACGCAACCC
TGACACGATTCTTCGCACTACACTTCCTAATACCATTCGCAATTATAGCAATAGCAATAATCCACCTACTA
TTCCTACACCA
AACAGGATCAAACAACCCACTAGGACTAAATAAAAACACAGACAAAATCCCATTCCACCCCTACTTCACA
GTAAAGGACG
TACTAGGATTTGCAATCACCATCATAACACTTACAATCCTAACCCTAAAAGAACCATACCTACTAAGAGA
CCCAGACAACT
TCACACCAGCAAACCCACTAGTAACACCCGTACATATTCAACCTGAATGATATTTAAA

>**Reverse Seq Data**

GCGGGGGTTCTAGCTGGCTAGCTGGTGTGAGTTGTCTGGGTCTCTTAGTAGGTATGGTTCTTTTAGGGTT
AGGATTGTA
AGTGTTATGATGGGGATTGCAAATCCTAGTACGTCCTTTACTGTGAAGTAGGGGTGGAATGGGATTTTGT
CTGTGTTTTT
ATTTAGTCCTAGGGGGTTGTTTGATCCTGTTTGGTGTAGGAATAGTAGGTGGATTATTGCTATTGCTATAA
TTGCGAATG
GTATTAGGAAGTGTAGTGCGAAGAATCGTGTCAGGGTTGCGTTGTCTACTGCGAATCCTCCCCATACTCA
TTGTACGATT
TCTTTCCCTATGTATGGGATGGCTGATAGAACGATGGTAATTACTGTTGCACCTCAAAAA

>**Reverse complement**

TTTTTGAGGTGCAACAGTAATTACCATCGTTCTATCAGCCATCCCATACATAGGGAAAGAAATCGTACAA
TGAGTATGGG
GAGGATTCGCAGTAGACAACGCAACCCTGACACGATTCTTCGCACTACACTTCCTAATACCATTCGCAAT
TATAGCAATA
GCAATAATCCACCTACTATTCCTACACCAAACAGGATCAAACAACCCCCTAGGACTAAATAAAAACACAG
ACAAAATCCC
ATTCCACCCCTACTTCACAGTAAAGGACGTACTAGGATTTGCAATCCCCATCATAACACTTACAATCCTA
ACCCTAAAAGA
ACCATACCTACTAAGAGACCCAGACAACTCACACCAGCTAGCCAGCTAGAACCCCCGC

3.3. Sequence analysis & Dataset preparation

Consensus sequence of COI gene was generated from forward and reverse sequence data using aligner software

>**Consensus data**

ACCTTCAATAAGTCATCCCATACATAGGGAAGAATCGTACAATGAGTATGGGGAGGATTCGCAGTAGACA
ACGCAACCC

TGACACGATTCTTCGCACTACACTTCCTAATACCATTCGCAATTATAGCAATAGCAATAATCCACCTACTA
TTCCTACACCA
AACAGGATCAAACAACCCACTAGGACTAAATAAAAACACAGACAAAATCCCATTCCACCCCTACTTCACA
GTAAAGGACG
TACTAGGATTTGCAATCACCATCATAACACTTACAATCCTAACCCTAAAAGAACCATACCTACTAAGAGA
CCCAGACAACT
TCACACCAGCAAACCCACTAGTAACACCCGTACATATTCAACCTGAATGATA.

Sequences producing significant alignments to *Macrotermes gilvus*:

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Macrotermes gilvus isolate SING91	510	510	100%	1e-139	91.22%	KY224691.1
Macrotermes gilvus	510	510	100%	1e-139	91.22%	NC_034110.1
Macrotermes falciger	510	510	100%	1e-139	91.22%	NC_034050.1
Macrotermes sp. PHI4	503	503	98%	2e-137	91.33%	KY224521.1
Macrotermes sp. B TB-2017	494	494	100%	1e-134	90.43%	KY224525.1
Macrotermes annandalei isolate D42	486	486	98%	2e-132	90.51%	KU900579.1
Macrotermes subhyalinus isolate SE7	483	483	100%	2e-131	89.89%	OR653920.1
Microtermes mariae isolate KE15-33	481	481	95%	8e-131	90.83%	OQ675024.1
Macrotermes yunnanensis isolate D50	477	477	100%	1e-129	89.63%	NC_034288.1
Microtermes sp. BDIT051	477	477	100%	1e-129	89.66%	KY224711.1

Table 1: Sequences producing significant alignments to *Macrotermes gilvus*

The COI gene sequence was used to carry out BLAST with the 'nr' database of NCBI GenBank database. Based on maximum identity score first ten sequences were selected and aligned using multiple alignment software program Clustal W.Distance Matrix for *Macrotermes gilvus*:

1		0.019	0.019	0.020	0.019	0.022	0.021	0.021	0.022	0.024	0.026
KY224691.1	0.092		0.000	0.023	0.007	0.022	0.019	0.022	0.024	0.020	0.029
NC_034110.1	0.092	0.000		0.023	0.007	0.022	0.019	0.022	0.024	0.020	0.029
NC_034050.1	0.097	0.108	0.108		0.024	0.019	0.024	0.009	0.023	0.025	0.029
KY224521.1	0.090	0.014	0.014	0.114		0.022	0.022	0.023	0.025	0.023	0.029
KY224525.1	0.102	0.094	0.094	0.092	0.099		0.022	0.020	0.022	0.023	0.024
KU900579.1	0.101	0.094	0.094	0.115	0.107	0.094		0.023	0.029	0.006	0.030
OR653920.1	0.109	0.102	0.102	0.033	0.107	0.091	0.108		0.021	0.024	0.029
OQ675024.1	0.102	0.118	0.118	0.115	0.121	0.107	0.143	0.105		0.031	0.021

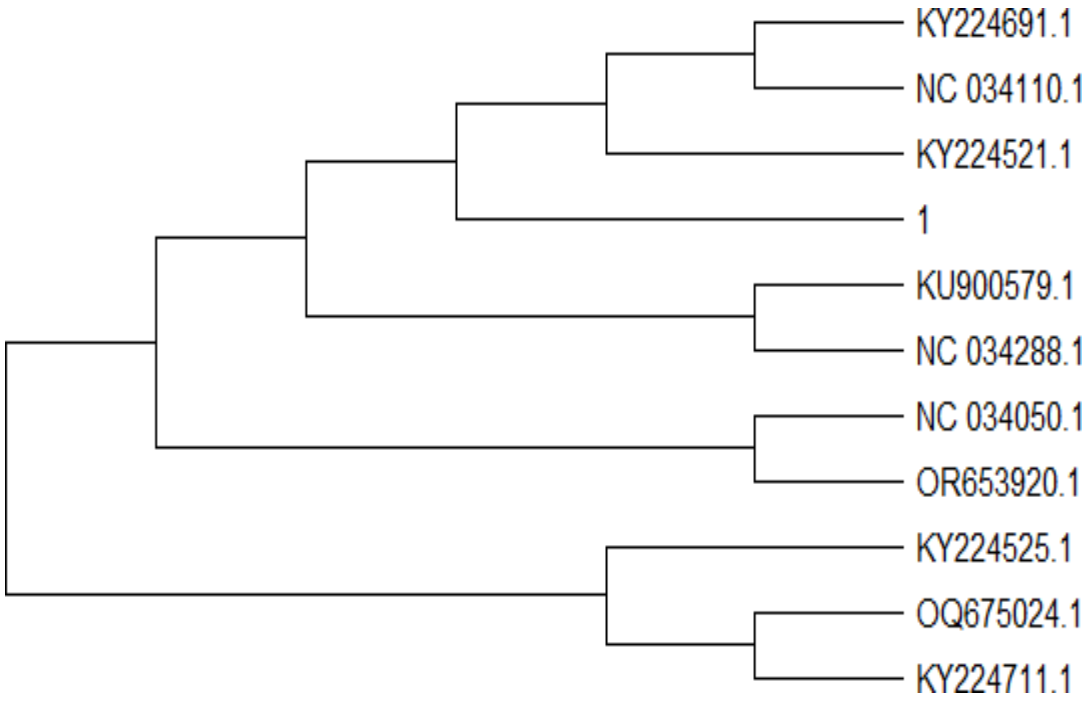
NC_034288.1	0.112	0.098	0.098	0.122	0.113	0.098	0.011	0.115	0.150		0.030
KY224711.1	0.119	0.140	0.140	0.137	0.143	0.117	0.138	0.141	0.089	0.139	

Table. 2: Estimates of Evolutionary Divergence between Sequences

The number of base substitutions per site from between sequences are shown. Standard error estimate(s) are shown above the diagonal. Analyses were conducted using the Maximum Composite Likelihood model [Kimura M. (1980)]. This analysis involved 11 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 376 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [Kumar S., Stecher G., Li M., Knyaz C., and Tamura K. (2018)]. Distance matrix was generated using RDP database.

3.4 Phylogenetic analysis

Phylogenetic tree for *Macrotermes gilvus* was constructed using MEGA 11.



Evolutionary analysis by Maximum Likelihood method

The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model. The tree with the highest log likelihood (-1265.04) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and Bio NJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with superior log likelihood value. [11] This analysis involved 11 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. There were a total of 376 positions in the final dataset. Evolutionary analyses were conducted in MEGA11. [12].

Both *Macrotermes gilvus* SING91 and *Macrotermes gilvus* (NC_034110.1) show perfect alignment (510/510), with 100% identity; the value is also 1e-139. *Macrotermes falciger* all align perfectly (510/510), showing that this or these sequences could not only be highly similar to but very likely identical with the query sequence, which indicates a very close relationship.Alignment Lengths:both *Macrotermes* sp. PHI4 and *Macrotermes annandalei* isolate D42 are in good agreement, with 503/503 and 486/486 respectively. Their identities are 98%, but the other factors have high E-

values worth considering even though these values are still very significant: 2e-137 and 2e-132 for each example, telling us that there is essentially strong homology when we compare these sequences to the query sequence even if they are not actually alike themselves. *Macrotermes sp*. B TB-2017 aligns perfectly (494/494), but at a slightly lower 90.43%. Such lower identity and E-values (1e-134) indicate that this sequence is related to the query but may have some variation or mutation in it.

Macrotermes sibanus isolate SE7, *Macrotermes yunnanensis* isolate D42, and *Macrotermes annandalei* isolate D56 differ only in the length of their very weakest matching sequences, all these sequences show good agreements at more than 99.3%. These sequences can therefore be thought to be close relativees of the query sequence.Though the other two matches are in the same range, here we are closest to *Macrotermes gilvus* (both the isolate SING91 and NC_034110.1). With 100% identity across the entire aligned length of all three, this could be a closely related species at best or the same species at worst.

The significant matches with slight variation include *Macrotermes* sp. PHI4, *Macrotermes* sp. B TB-2017, and *Macrotermes annandalei* isolate D42, all of which have high similarity but small variations, whether these reflect subspecies changes strains simply different genes although closely related not just two distant species remains to be seen.

REFERENCES

1. 1.Mohd, Hairul, Mohd, Salleh., Yuzine, Esa., Rozi, Mohamed. (2023). Global Terrapin Character-Based DNA Barcodes: Assessment of the Mitochondrial COI Gene and Conservation Status Revealed a Putative Cryptic Species. Animals, 13(11):1720-1720. doi: 10.3390/ani13111720.
2. Jéhan, Le, Cadre., Finn, Luca, Klemp., Miklós, Bálint., Stefan, Scheu., Ina, Schaefer. (2024). Applicability and perspectives for DNA barcoding of soil invertebrates. PeerJ, 12:e17709-e17709. doi: 10.7717/peerj.17709
3. F., Valen., E., Bidayani., R., A., Alfian., M., Prananda., H., Notonegoro., N., B., Susilo., W., D., Ardi., Swarlanda., M., A., Aziz., Puryoso., M., S., Widodo., A., R., Faqih., V., Hasan. (2023). Unveiling the close relationship between Betta burdigala and Betta uberis through DNA barcoding based on COI Gene. 1267 doi: 10.1088/1755-1315/1267/1/012066
4. Qurratu'Aini, Syasya, Shamsuri., Abdul, Majid. (2023). Metagenomic 16S rRNA amplicon data of gut microbial diversity in three species of subterranean termites (Coptotermes gestroi, Globitermes sulphureus and Macrotermes gilvus). Data in Brief, 47:108993-108993. doi: 10.1016/j.dib.2023.108993
5. Vitor, Falchi, Timm., Leonardo, Tresoldi, Gonçalves., Vera, L., S., Valente., Maríndia, Deprá. (2022). The efficiency of the COI gene as a DNA barcode and an overview of Orthoptera (Caelifera and Ensifera) sequences in the BOLD System. Canadian Journal of Zoology, 100(11):710-718. doi: 10.1139/cjz-2022-0041
6. 6.Nguyen, Quoc, Huy., Nguyễn, Thị, Thu, Huyền., Ha, Thi, Lan, Anh., Nguyen, Thi, Uyen., Le, Duc, Viet., Dang, Thi, Lan, Anh., Hsiao-ching, Yen., Anh, Đức, Trần., Nguyen, Van, Vinh., Cao, Thi, Kim, Thu., Nguyễn, Văn, Sáng. (2023). Sequencing Analysis of COI Genes Isolated from Vietnam's Endemic Insect Species (Insecta: Plecoptera) for DNA Barcoding Database. VNU Journal of Science: Natural Sciences and Technology, doi: 10.25073/2588-1140/vnunst.5513.
7. Jennifer, Hackett., Sharon, Pepenella., Cristina, F., Marco., Louise, Bodt., Lina, Ruiz, Grajales., Jeffry, Petracca., Julian, F., Burke., Allison, Mayle., Bruce, Nash., David, Micklos. (2024). Simple, Robust Invertebrate DNA Barcoding: Chelex-Based DNA Extraction and Optimized COI Amplification. 119-127. doi: 10.1007/978-1-0716-3581-0_6

8. Patrick, F., Allison., Emily, T., Pickich., Zanethia, C., Barnett., Ryan, C., Garrick. (2024). DNA barcoding is currently unreliable for species identification in most crayfishes. Ecology and Evolution, 14(7) doi: 10.1002/ece3.70050
9. Naveeta, M., Vellupillai., Abdul, Hafiz, Ab, Majid. (2023). Phylogenetic relationship of subterranean termite Coptotermes gestroi (Blattodea: Rhinotermitidae) inhabiting urban and natural habitats. Heliyon, doi: 10.1016/j.heliyon.2023.e23692
10. Hai-Guang, Zhang., Wenjun, Bu. (2022). Exploring Large-Scale Patterns of Genetic Variation in the COI Gene among Insecta: Implications for DNA Barcoding and Threshold-Based Species Delimitation Studies. Insects, 13(5):425-425. doi: 10.3390/insects13050425.
11. Kimura M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16:111-120.
12. Kumar S., Stecher G., Li M., Knyaz C., and Tamura K. (2018). MEGA X: Molecular Evolutionary
13. Genetics Analysis across computing platforms. Molecular Biology and Evolution 35:1547-1549.