scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Molecular Safeguards of Survival: *De novo* transcriptome assembly and tissue-specific transcriptomic profiling of the yellow-foot clam *Paphia malabarica*

Sandhya Sukumaran ✉, V. G. Vysakh & Wilson Sebastian

The yellow-foot clam, *Paphia malabarica* Chemnitz, 1782 is a regionally important fishery and aquaculture resource along the southwest coast of India. The fishery of this species from the Ashtamudi estuary has been certified as sustainable by the Marine Stewardship Council. We generated the first high-quality, comprehensive *de novo* transcriptome assembly of *Paphia malabarica* using Illumina paired-end sequencing. A total of 729.2 million raw reads were generated and further processed and assembled using the Trinity assembler. The final assembly yielded 55,980 predicted coding sequences (CDS) with an N50 of 1,605 bp. The BUSCO completeness of the transcriptome assembly was 97.8% against the Metazoa (metazoa_odb10) dataset and 88.52% against the Mollusca (mollusca_odb10) dataset. Functional annotations were performed using UniProt, KEGG, Reactome, InterPro, and eggNOG databases. Out of 56,713 transcripts, 20,624 sequences (36.37%) were annotated with Gene Ontology (GO) terms, yielding 360,305 annotations. Gene ontology and pathway analyses indicated that transcripts associated with key biological processes, such as replication, recombination and repair, transcription, signal transduction mechanisms, posttranslational modification, protein turnover, chaperones, carbohydrate transport and metabolism, and amino acid transport and metabolism, are important for the buried lifestyle of the yellow-foot clam, *Paphia malabarica*.

Background & Summary

Marine ecosystems are increasingly affected by climate change, with major consequences for biodiversity, ecosystem functioning, and the sustainability of living marine resources. Sedentary bivalves such as clams, oysters, and mussels are particularly vulnerable to climate-driven stressors, including rising seawater temperature, marine heatwaves, ocean acidification, hypoxia, and altered pathogen dynamics, as these factors directly influence their physiology, survival, reproduction, and distribution[1–3]. Venerid clams represent an ecologically and economically important group within marine bivalves, with several species supporting fisheries and aquaculture systems worldwide, including *Venerupis philippinarum*, *Mercenaria mercenaria*, and *Meretrix spp*[4]. Along the Indian coast, genera such as *Meretrix, Marcia*, and *Paphia* support regionally important fishery systems[5]. Habitat degradation, anthropogenic pollution, and environmental changes are major threats to the sustainability of clam populations and fisheries[6].

Clams exhibit wide variations in shape, size, colour, and body structure[7]. Their wedge-shaped muscular foot and long extendible siphons facilitate burrowing behaviour[8,9]. They inhabit estuarine and intertidal environments characterised by fluctuating salinity, temperature, and dissolved oxygen, making them relevant models for investigating molecular mechanisms of environmental adaptation and stress tolerance. They play important roles in biogeochemical cycling and sediment–water exchange processes[10]. These ecological and physiological characteristics make transcriptomic resources particularly valuable for understanding how benthic bivalves respond to rapidly changing coastal environments.

ICAR-Central Marine Fisheries Research Institute, Ernakulam North P.O., Kochi, Kerala, 682018, India. ✉e-mail: sandhyasukumarancmfri@gmail.com; sandhya.sukumaran@icar.org.in

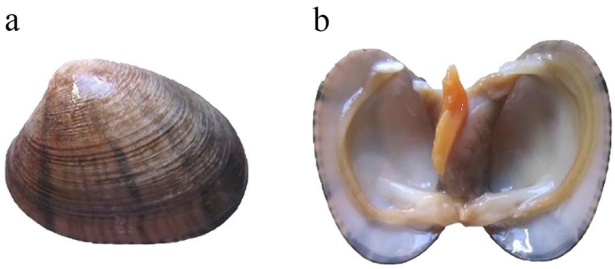


Fig. 1 Photograph of the yellow-foot clam *Paphia malabarica*. (**a**) External view of the shell. (**b**) Internal view of the shell.

Genomic and transcriptomic resources provide direct insights into the molecular mechanisms underlying environmental adaptation in bivalves[11–15]. Comprehensive identification of expressed protein-coding genes enables the characterisation of pathways associated with thermal tolerance, oxidative stress response, hypoxia adaptation, immune defence, and xenobiotic detoxification[16–18]. These molecular insights are particularly relevant for aquaculture, where they can support the identification of stress-resilient stocks, improve understanding of host-pathogen interactions, and inform strategies to mitigate the impacts of marine heatwaves and disease outbreaks. In addition, transcriptomic data can aid in elucidating genes involved in growth regulation, reproductive cycles, and larval development, thereby contributing to improved broodstock management and seed production[14,19].

The gene families and molecular underpinnings of bivalve stress responses have been extensively studied across multiple taxa, including clams, oysters, mussels, scallops, and cockles[11,20–25]. Bivalves have developed specialised tolerance mechanisms to counteract constant exposure to pathogens because they lack an adaptive immune system[26]. Innate immune pathways, including apoptosis regulation and stress-response signalling, play central roles in maintaining cellular homeostasis in bivalves under environmental stress[21,27].

Despite their ecological importance, genomic resources for many clam species remain limited. The genome and transcriptome of the Manila clam, *Ruditapes philippinarum*, have been characterised, providing insights into its unique metabolic mechanism and neuroregulatory system associated with its adaptation to infaunal lives[28]. A high-quality chromosome-level assembly of the Venus clam, *Cyclina sinensis*, provided insights into adaptation to a buried lifestyle[29]. Genomic investigations were also undertaken on the Asian clam *Corbicula fluminea*[15], the purple butter clam *Saxidomus purpuratus*[30], the Pismo clam, *Tivela stultorum*[31], the Snout otter clam, *Lutraria rhynchaena*[32] and the giant clam *Tridacna gigas*[33].

The yellow-foot clam, *Paphia malabarica* Chemnitz 1782, is widely distributed along the south-west and south-east coasts of India, and its fishery contributes to local consumption as well as exports[34]. Even though it is frequently referred to as *P. malabarica* in India, it has been synonymised as *Protapes gallus*[35,36]. The present authors have suggested retaining the name *Paphia malabarica* as the specimens collected from India were found to be different from *Protapes gallus* in morphological and molecular characteristics[37]. Some of the recent studies suggested reassignment of *P. malabarica* to *Marcia recens*, which is also yet to be accepted[38,39]. In view of the taxonomic ambiguity, the present study retains the name *Paphia malabarica*.

The fishery of yellow-foot clams from the Ashtamudi estuary, Kollam, Kerala, has been certified as sustainable by the Marine Stewardship Council and the fishery is now managed by a set of fishery management plans. The captive broodstock maturation and seed production of the yellow-foot clam have been standardised by temperature manipulation and enhanced feeding[40]. Hormone-induced spawning protocols have also been standardised in this species, contributing to efficient seed production[40].

Considering the ecological and economic importance of this species, transcriptomic analysis of *Paphia malabarica* was undertaken to generate a comprehensive gene expression resource. This dataset will facilitate the identification of molecular pathways associated with environmental adaptation, stress tolerance, and physiological regulation, and will support future studies aimed at improving breeding strategies and larval rearing.

Methods

Sample collection. *Paphia malabarica* specimens were collected from Ashtamudi estuary (Kollam, Kerala, India) and Dharmadam (Kannur, Kerala, India) following morphological descriptions[41] (Fig. 1). A total of six specimens were collected from each location. Tissues from adductor muscle, mantle, foot and gill were snap-frozen in liquid nitrogen and stored at −80 °C until further processing. The animal experimentation protocols were approved by the Institutional Animal Ethical Committee of ICAR-Central Marine Fisheries Research Institute (CMFRI), Kochi (Approval No: MBT/GEN/25-05). All methods reported in this manuscript were carried out following the ARRIVE guidelines (http://arriveguidelines.org).

RNA Extraction, library preparation and sequencing. RNA extraction from each individual tissue was undertaken using TRIzol reagent (Invitrogen, USA) following the manufacturer's instructions. Any genomic DNA was removed by treating the RNA samples using RNAse-Free DNase I (Qiagen, USA). RNA samples that had an RNA integrity number of at least 7 and an OD 260/280 $\geq$ 1.8 were chosen for additional analysis. We pooled an equal quantity of high-quality RNA from 6 samples for each tissue (tissue-specific samples), and these tissue-specific samples were then used for cDNA synthesis and subsequent tissue-specific transcriptome profiling.

Sample Name *P. malabarica*	% GC	Average Read Length	M Seqs
Adductor muscle R1	40%	150 bp	23.5 M
Adductor muscle R2	40%	150 bp	23.5 M
Foot R1	39%	150 bp	20.5 M
Foot R2	39%	150 bp	20.5 M
Gill R1	37%	150 bp	17.1 M
Gill R2	38%	150 bp	17.1 M
Mantle R1	38%	150 bp	27.0 M
Mantle R2	38%	150 bp	27.0 M
Pooled transcriptome 1R1	40%	100 bp	142.9 M
Pooled transcriptome 1R2	40%	100 bp	142.9 M
Pooled transcriptome 2 R1	41%	100 bp	59.1 M
Pooled transcriptome 2 R2	40%	100 bp	59.1 M
Pooled transcriptome 3 R1	41%	100 bp	53.5 M
Pooled transcriptome 3 R2	41%	100 bp	53.5 M

Table 1. Raw sequencing read statistics of *Paphia malabarica* transcriptome datasets.

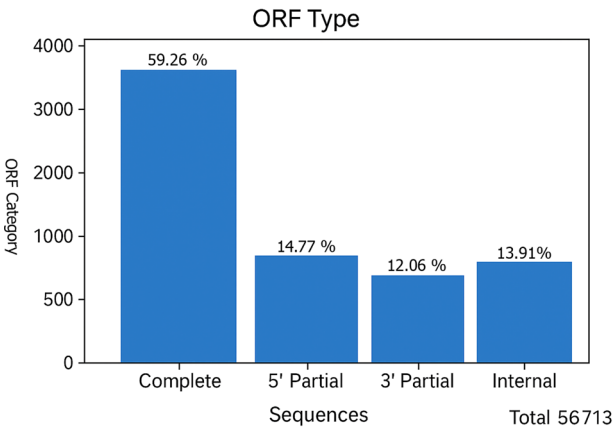


Fig. 2 Distribution of open reading frame (ORF) types in the *Paphia malabarica* transcriptome.

In addition, pooled RNA of the four tissues was sequenced in triplicate to generate a comprehensive transcriptomic assembly of *P. malabarica*. The RNA quality and quantity were checked using 1.5% denaturing agarose gel electrophoresis and the Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Library preparation was undertaken using the TruSeq Standard mRNA-seq library preparation kit. Quality of the library was assessed using the Agilent 2100 Bioanalyzer (Part No. G2939BA), and concentration was measured using the KAPA Library Quantification Kit (Cat. No. KK4824). Final sequencing was performed on an Illumina NovaSeq. 6000 platform in paired-end mode with 150-bp reads.

Pre-assembly processing. All bioinformatics analyses were performed using the high-performance computing system, FISH@CMFRI server facility at ICAR-CMFRI. A total of 729.2 million raw reads were generated through Illumina sequencing. Quality of the raw reads was assessed using FastQC v0.11.8[42]. The statistics of the raw data are shown in Table 1. Fastp v1.12[43] was used to remove adaptor sequences and low-quality reads with PHRED score < 20. A total of 674 million cleaned reads were retained for *de novo* transcriptome assembly, representing 98% of the raw reads.

***De novo* transcriptome assembly.** The transcriptome assembly was constructed using the Trinity v2.11 program with default parameters[44,45]. The final assembly comprised 274,880 transcripts, with an N50 of 2,164 bp. The Trinity assembly output was clustered with CD-HIT-est v. 4.8.1 using a hierarchical clustering algorithm to reduce redundancy. A total of 171,022 unigenes were finally identified. The output from CD-HIT-est[46] was then processed using TransDecoder v. 5.7.0[47] for detecting long open reading frames (ORFs) in assembled transcripts (Fig. 2). Transcriptome assembly was also screened with Kraken2 v2.1.3[48] and NCBI-FCS v0.5.0[49] to remove all potential contaminants. The quality and completeness of the assembly were checked using quality assessment tools like BUSCO v5.4.4 (Benchmarking Universal Single-Copy Orthologs)[50] and TransRate v1.0.3[51]. A comprehensive summary of all assemblies and corresponding statistics is provided in Tables 2, 3. BUSCO completeness analyses were conducted against two databases: Metazoa (metazoa_odb10) and Mollusca (mollusca_odb10). The genome module benchmark values against the Metazoa (metazoa_odb10) dataset were calculated as C: 97.8%, including [S: 72.12%, D: 25.68%], F: 0.73%, M: 1.47% (C: complete, S: single-copy, D: duplicated, F: fragmented,

Metric	Raw Assembly	Supertranscript Assembly	FCS Filtered Assembly	Clusters	CDS
Total Sequences (n seqs)	276,340	130,346	274,880	171,022	55,980
Smallest Contig Length	271 bp	301 bp	271 bp	275 bp	255 bp
Largest Contig Length	66,663 bp	66,727 bp	66,663 bp	66,663 bp	65,514 bp
Total Number of Bases (n bases)	377,601,325	152,616,097	376,563,653	193,802,086	65,726,733
Mean Contig Length	1,366.44 bp	1,170.85 bp	1,369.92 bp	1,133.20 bp	1,174.11 bp
Contigs <200 bp	0	0	0	0	0
Contigs with ORFs	84,452	26,132	84,084	39,735	42,866
Mean ORF Percentage	44.03%	34.22%	43.98%	41.98%	99.81%
N50	2,160 bp	2,062 bp	2,164 bp	1,823 bp	1,605 bp
GC Content	34%	34%	34%	34%	40%
Bases with 'N'	0	0	0	0	0

Table 2. Overview of combined assembly quality metrics.

Metrics	Gill	Muscle	Mantle	Foot
Total number of transcripts	142,612	89,143	167,926	122,303
Total number of genes (Trinity clusters)	102,768	68,285	119,230	89,240
GC content	34.97%	34.87%	34.92%	35.22%
Total assembled bases (all transcripts)	153,279,868 bp	91,731,467 bp	174,445,042 bp	133,344,912 bp
Total assembled bases (longest isoform per gene)	85,291,245 bp	55,353,422 bp	93,070,198 bp	74,623,202 bp
N50 (all transcripts)	2,178 bp	2,041 bp	2,182 bp	2,248 bp
N50 (longest isoform per gene)	1,649 bp	1,525 bp	1,583 bp	1,704 bp

Table 3. Overview of tissue-specific assembly quality metrics.

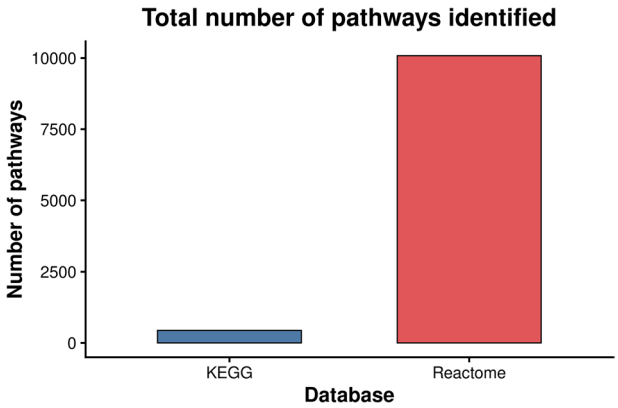


Fig. 3 Overview of functional pathway assignments derived from transcriptome annotation.

M: missing). The BUSCO completeness against the Mollusca Odb10 dataset indicated C: 88.52, including [S: 59.55%, D: 28.97%], F: 8.96%, M: 2.75% (C: complete, S: single-copy, D: duplicated, F: fragmented, M: missing). The read support was finally assessed by aligning reads back to the assembly using the sequence alignment tool STAR v2.7.8a[52].

Transcriptome annotation. Transcriptome annotation and protein-coding region prediction were performed using BLASTx and BLASTp against UniProtKB, which served as the primary reference database for functional annotation (E-value cutoff of 1e-10). Functional annotations and pathway analyses were performed using Reactome (Release V90)[53], KEGG (Release 117.0)[54], and EggNOG v2.1.12 (Figs. 3–5). A total of 32,190 predicted protein-coding sequences were annotated, which may include transcript fragments and isoforms. Gene Ontology (GO) analyses were performed using Blast2GO v5.2[55], classifying transcripts to cellular components, molecular functions and biological processes. From 56,713 transcripts, 20,624 sequences (36.37%) were annotated with Gene Ontology (GO) terms, yielding 360,305 individual annotations. In addition, functional protein domains were identified using InterProScan v4.0[56], which integrates prediction models from multiple databases, including PANTHER, Gene3D, and Pfam. Enzyme Code mapper[57] is employed for identification of enzyme classes across the assembly (Fig. 6).

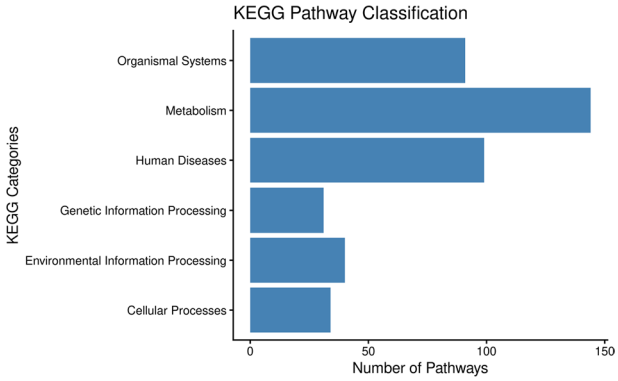


Fig. 4 Distribution of transcripts across KEGG pathway categories.

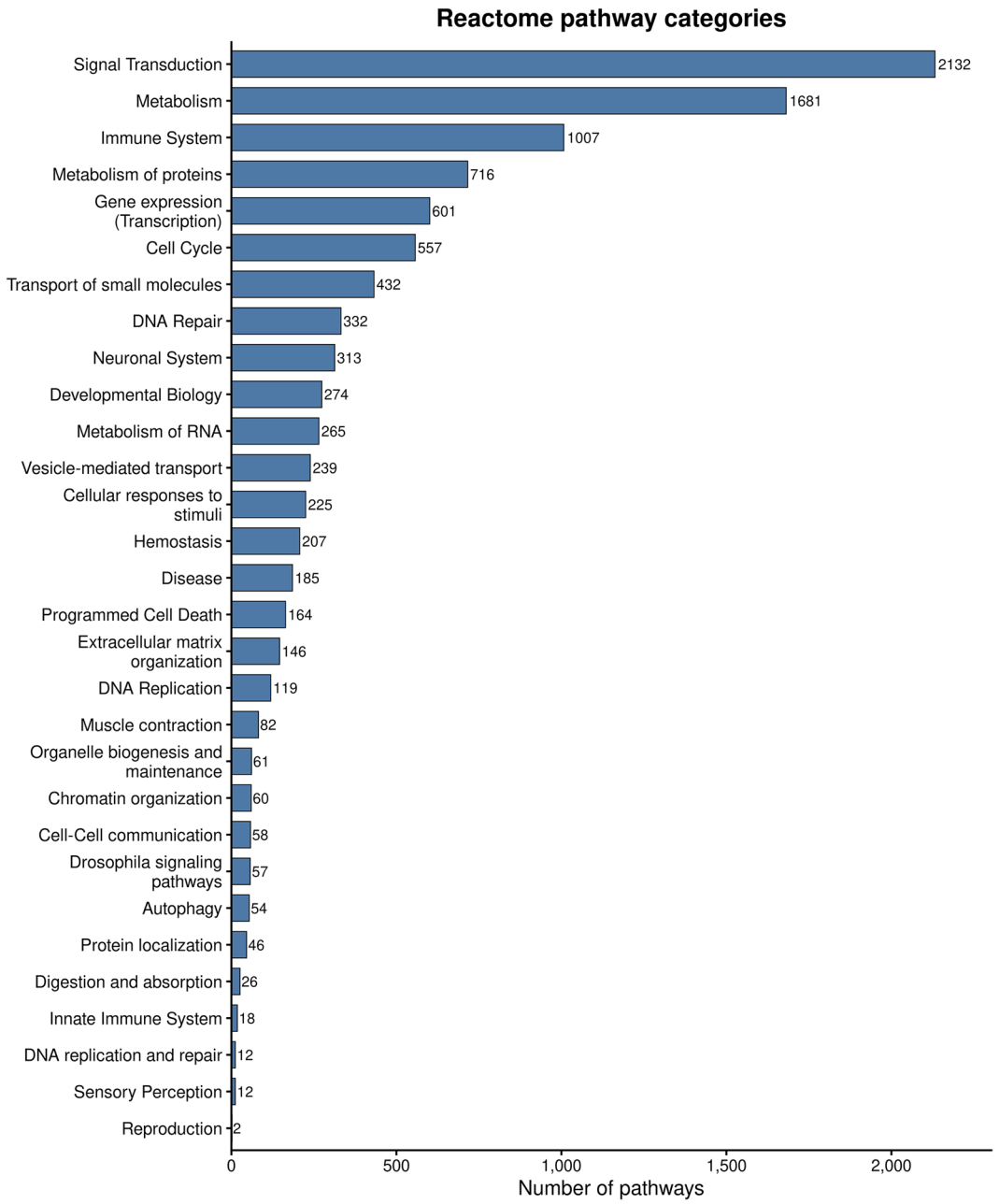


Fig. 5 Classification of transcripts based on Reactome pathway categories.

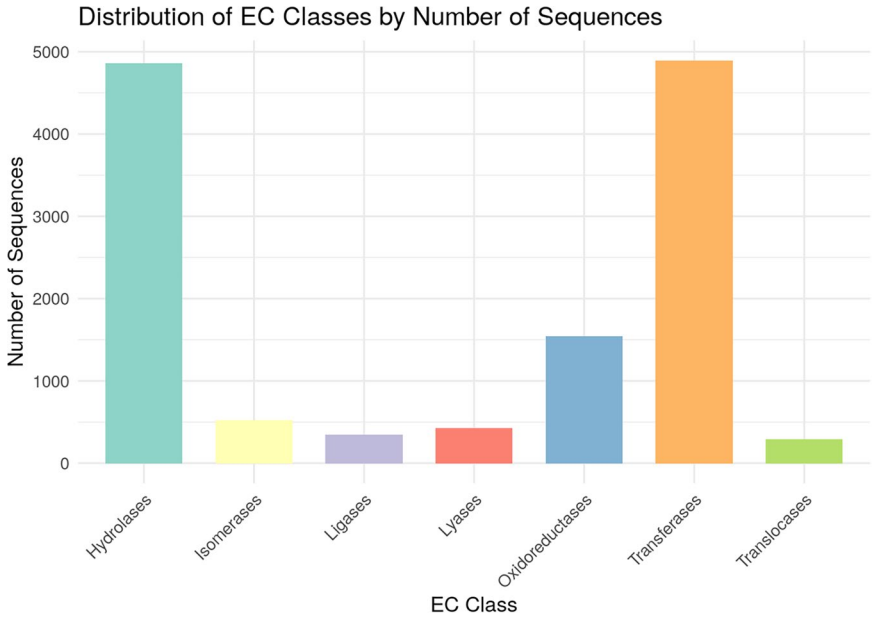


Fig. 6 Distribution of annotated transcripts across enzyme classes.

Data Records

The *de novo* transcriptome assembly has been submitted to NCBI GenBank under Bioproject PRJNA1177681 with accession number GKZK00000000.1[58]. The raw data generated through Illumina sequencing (tissue-specific and pooled) have been deposited in the NCBI Sequence Read Archive (SRA) under accession numbers SRR31127582-SRR31127588[59–65]. All the annotation files have been deposited in the Figshare database[66].

Technical Validation

Integrity and quality of RNA samples were validated using Qubit 4 Fluorometer (Invitrogen), NanoDrop One Spectrophotometer (ThermoScientific, USA), and Agilent 2200 TapeStation, and RNA samples with RNA integrity (RIN) values greater than or equal to 7 were used for pooling and sequencing. The completeness of the assembly was validated using BUSCO analysis.

Data availability

All the data reported in this manuscript are deposited in publicly accessible data repositories. The *de novo* transcriptome assembly has been submitted to NCBI GenBank under Bioproject PRJNA1177681 with accession number GKZK00000000.1(https://identifiers.org/ncbi/insdc:GKZK00000000.1). The raw data generated through Illumina sequencing (tissue-specific and pooled) have been deposited in the NCBI Sequence Read Archive (SRA) under accession numbers SRR31127582-SRR31127588 (https://identifiers.org/insdc.sra:SRR31127582; https://identifiers.org/insdc.sra:SRR31127583; https://identifiers.org/insdc.sra:SRR31127584; https://identifiers.org/insdc.sra:SRR31127585; https://identifiers.org/insdc.sra:SRR31127586; https://identifiers.org/insdc.sra:SRR31127587; https://identifiers.org/insdc.sra:SRR31127588). All the annotation files have been deposited in the Figshare database (https://doi.org/10.6084/m9.figshare.32752470).

Code availability

No custom code was generated in this study.

Received: 19 December 2025; Accepted: 25 June 2026;
Published online: 27 June 2026

References

1. Verdelhos, T., Marques, J. C. & Anastácio, P. Behavioral and mortality responses of the bivalves *Scrobicularia plana* and *Cerastoderma edule* to temperature, as indicator of climate change's potential impacts. *Ecol. Indic.* **58**, 95–103 (2015).
2. Andreyeva, A. Y. *et al*. Impacts of ocean acidification and hypoxia on cellular immunity, oxygen consumption and antioxidant status in Mediterranean mussel. *Fish Shellfish Immunol.* **154**, 109932 (2024).
3. Turner, L. M. *et al*. Pathogenic marine microbes influence the effects of climate change on a commercially important tropical bivalve. *Sci. Rep.* **6**, 32413 (2016).
4. Da Costa, F. *Clam Fisheries and Aquaculture*. (2012).
5. Kavitha, M. *et al*. Fishery and biology of *Paphia malabarica* (Dilwyn, 1817) off Karapad Bay, south-east coast of India. *Indian J. Fish*. **68**, (2021).
6. FAO. *The State of World Fisheries and Aquaculture 2018 (SOFIA). Meeting the Sustainable Development Goals*. (FAO, Rome, Italy, 2018).
7. Yang, Z. *et al*. The evo-devo of molluscs: Insights from a genomic perspective. *Evol. Dev.* **22**, 409–424 (2020).
8. Stanley, S. M. Adaptive Themes in the Evolution of the Bivalvia (mollusca). *Annu. Rev. Earth Planet. Sci.* **3**, 361 (1975).
9. Hochner, B. & Glanzman, D. L. Evolution of highly diverse forms of behavior in molluscs. *Curr. Biol. CB* **26**, R965–R971 (2016).

10. Vaughn, C. C. & Hakenkamp, C. C. The functional role of burrowing bivalves in freshwater ecosystems. *Freshw. Biol.* **46**, 1431–1446 (2001).
11. Sukumaran, S. *et al*. The chromosome level genome assembly of the Asian green mussel, Perna viridis. *Sci. Data* **11**, 930 (2024).
12. Vysakh, V. G., Sukumaran, S. & A, G. De novo transcriptome assembly of the *Perna viridis*: A novel invertebrate model for ecotoxicological studies. *Sci. Data* **12**, 147 (2025).
13. Vysakh, V. G., Sukumaran, S., Sebastian, W. & Gopalakrishnan, A. The transcriptomic footprint of *Mytella strigata*: de novo transcriptome assembly of a major invasive species. *Sci. Data* **12**, 201 (2025).
14. Zhang, Y., Nie, H., Yin, Z. & Yan, X. Comparative transcriptomic analysis revealed dynamic changes of distinct classes of genes during development of the Manila clam (*Ruditapes philippinarum*). *BMC Genomics* **23**, 676 (2022).
15. Zhang, T. *et al*. Dissecting the chromosome-level genome of the Asian Clam (*Corbicula fluminea*). *Sci. Rep.* **11**, 15021 (2021).
16. Costa, E. S. *et al*. Evaluation of metals and hydrocarbons in sediments from a tropical tidal flat estuary of Southern Brazil. *Mar. Pollut. Bull.* **92**, 259–268 (2015).
17. Wang, Y. *et al*. Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of illumina tags. *Appl. Environ. Microbiol.* **78**, 8264–8271 (2012).
18. Collins, A. L. *et al*. Sediment-associated organic matter sources and sediment oxygen demand in a Special Area of Conservation (SAC): A case study of the River Axe, UK. *River Res. Appl.* **33**, 1539–1552 (2017).
19. Xu, M. *et al*. Transcriptomic Analysis Provides Insights into Candidate Genes and Molecular Pathways Involved in Growth of *Mytilus coruscus* Larvae. *Int. J. Mol. Sci*. **25**, (2024).
20. Regan, T. *et al*. Ancestral Physical Stress and Later Immune Gene Family Expansions Shaped Bivalve Mollusc Evolution. *Genome Biol. Evol.* **13**, evab177 (2021).
21. Song, H. *et al*. The hard clam genome reveals massive expansion and diversification of inhibitors of apoptosis in Bivalvia. *BMC Biol.* **19**, 15 (2021).
22. Witkop, E. M., Proestou, D. A. & Gomez-Chiarri, M. The expanded inhibitor of apoptosis gene family in oysters possesses novel domain architectures and may play diverse roles in apoptosis following immune challenge. *BMC Genomics* **23**, 201 (2022).
23. Domínguez-Pérez, D. *et al*. Transcriptomic Profile of the Cockle *Cerastoderma edule* exposed to seasonal diarrhetic shellfish toxin contamination. *Toxins* **13**, 784 (2021).
24. Zhang, X. *et al*. Transcriptome Analysis Reveals the Genes Involved in Oxidative Stress Responses of Scallop to PST-Producing Algae and a Candidate Biomarker for PST Monitoring. *Antioxidants* **12**, 1150 (2023).
25. Vysakh, V. G., Sukumaran, S. & Gopalakrishnan, A. Evaluating the effects of zinc oxide nanoparticles on a sentinel aquatic invertebrate species: Transcriptomic analysis and potential implications for ecosystem health. *Mar. Pollut. Bull.* **212**, 117570 (2025).
26. Wang, L., Qiu, L., Zhou, Z. & Song, L. Research progress on the mollusc immunity in China. *Dev. Comp. Immunol.* **39**, 2–10 (2013).
27. Vogeler, S., Carboni, S., Li, X. & Joyce, A. Phylogenetic analysis of the caspase family in bivalves: implications for programmed cell death, immune response and development. *BMC Genomics* **22**, 80 (2021).
28. Sun, X. *et al*. Clam Genome and Transcriptomes Provide Insights into Molecular Basis of Morphological Novelties and Adaptations in Mollusks. *Biology* **13**, 870 (2024).
29. Wei, M. *et al*. Chromosome-Level Clam Genome Helps Elucidate the Molecular Basis of Adaptation to a Buried Lifestyle. *iScience* **23**, 101148 (2020).
30. Kim, J. *et al*. Chromosome-Level Genome Assembly of the Butter Clam Saxidomus purpuratus. *Genome Biol. Evol.* **14**, evac106 (2022).
31. Emery, K. A. *et al*. A chromosome-length genome assembly for the Pismo clam, Tivela stultorum, a long-lived bivalve species severely impacted by overfishing. *J. Hered.* **116**, 680–690 (2025).
32. Thai, B. T. *et al*. Whole Genome Assembly of the Snout Otter Clam, Lutraria rhynchaena, Using Nanopore and Illumina Data, Benchmarked Against Bivalve Genome Assemblies. *Front. Genet*. **10**, (2019).
33. Li, R. *et al*. The genome sequence of the giant clam, Tridacna gigas (Linnaeus, 1758). *Wellcome Open Res.* **9**, 145 (2024).
34. Thomas, S. Growth and population dynamics of short-neck clam Paphia malabarica from Dharmadom estuary, North Kerala, southwest coast of India. *J. Mar. Biol. Assoc. India* **51**, 87–92 (2009).
35. Rippingale, O. H. & McMichael, D. F. *Queensland and Great Barrier Reef Shells*. (Jacaranda Press, 1961).
36. Matsukuma, A. Paphia (Protapes) irrediviva Makiyama, an extinct tropical embayment element of Japanese cenozoic mollusca. and allied species from the Indo-West Pacific. *Saito Ho- Kai Spec Pub* **405**, pls.1-4 (1988).
37. Sukumaran, S. *et al*. Morphological and molecular investigations reveal that *Paphia malabarica* from Indian waters is not synonymous with *Paphia (Protapes) gallus*. *Reg. Stud. Mar. Sci.* **27**, 100549 (2019).
38. Arathi, A. R., Oliver, P. G., Ravinesh, R. & Kumar, A. B. The Ashtamudi Lake short-neck clam: re-assigned to the genus Marcia H. Adams & A. Adams, 1857 (Bivalvia, Veneridae). *ZooKeys* **799**, 1–20 (2018).
39. Rahuman, S., Jeena, N. S., Sebastian, W. & Sajeela, K. A. Mitogenomic profiling of Marcia recens (Holten, 1802): a step towards taxonomic resolution of India's Ashtamudi short-neck clam. *Molluscan Res.* 1–10 (2024).
40. Mk, A. *et al*. Recent Advances in Bivalve Seed Production in India: Salient Research Findings, *Technologies and Their Social Impact*. (2025).
41. Oliver, G. & Glover, E. Paphia (Protapes) (Bivalvia: Veneroidea) in the Arabian Sea, with the Description of a New Species. *J. Conchol.* **35**, 389–405 (1996).
42. Andrews, S. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data (2010).
43. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinforma. Oxf. Engl.* **34**, i884–i890 (2018).
44. Grabherr, M. G. *et al*. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
45. Haas, B. J. *et al*. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
46. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinforma. Oxf. Engl.* **22**, 1658–1659 (2006).
47. Tang, S., Lomsadze, A. & Borodovsky, M. Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res.* **43**, e78 (2015).
48. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
49. A, A. *et al*. Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biol.* **25**, (2024).
50. Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol. Biol. Clifton NJ* **1962**, 227–245 (2019).
51. Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M. & Kelly, S. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* **26**, 1134–1144 (2016).
52. Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).
53. Jassal, B. *et al*. The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).
54. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
55. Conesa, A. *et al*. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinforma. Oxf. Engl.* **21**, 3674–3676 (2005).

56. Jones, P. *et al*. InterProScan 5: genome-scale protein function classification. *Bioinforma. Oxf. Engl.* **30**, 1236–1240 (2014).
57. Heid, E., Probst, D., Green, W. H. & Madsen, G. K. H. EnzymeMap: curation, validation and data-driven prediction of enzymatic reactions. *Chem. Sci.* **14**, 14229–14242 (2023).
58. Sukumaran, S. *et al*. *Paphia malabarica*, transcriptome shotgun assembly. *GenBank*. https://identifiers.org/ncbi/insdc:GKZK00000000.1 (2024).
59. *NCBI Sequence Read Archive* https://identifiers.org/insdc.sra:SRR31127582 (2024).
60. *NCBI Sequence Read Archive* https://identifiers.org/insdc.sra:SRR31127583 (2024).
61. *NCBI Sequence Read Archive* https://identifiers.org/insdc.sra:SRR31127584 (2024).
62. *NCBI Sequence Read Archive* https://identifiers.org/insdc.sra:SRR31127585 (2024).
63. *NCBI Sequence Read Archive* https://identifiers.org/insdc.sra:SRR31127586 (2024).
64. *NCBI Sequence Read Archive* https://identifiers.org/insdc.sra:SRR31127587 (2024).
65. *NCBI Sequence Read Archive* https://identifiers.org/insdc.sra:SRR31127588 (2024).
66. Sukumaran, S. *et al*. De novo transcriptome assembly and tissue-specific transcriptomic profiling of the yellow-foot clam *Paphia malabarica*. *Figshare*. https://doi.org/10.6084/m9.figshare.32752470 (2026).

Acknowledgements

This research was funded by the Indian Council of Agricultural Research (ICAR-CMFRI institute project MBT/GNM/25). The authors would like to thank the Director, ICAR-Central Marine Fisheries Research Institute (ICAR-CMFRI), and Head, Marine Biotechnology, Fish Nutrition and Health Division, ICAR-CMFRI for providing facilities to carry out this work. The research was supported by the Indian Council of Agriculture Research, Department of Agriculture, Research and Education, Government of India.

Author contributions

S.S. and V.V.G. conceived the study. S.S., V.V.G. and W.S. carried out the lab work. S.S. and V.V.G. wrote the final manuscript.

Funding

This research was funded by the Indian Council of Agricultural Research (ICAR-CMFRI institute project MBT/GNM/25), Department of Agriculture, Research and Education, Government of India.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026