OPEN ACCESS

REPORT

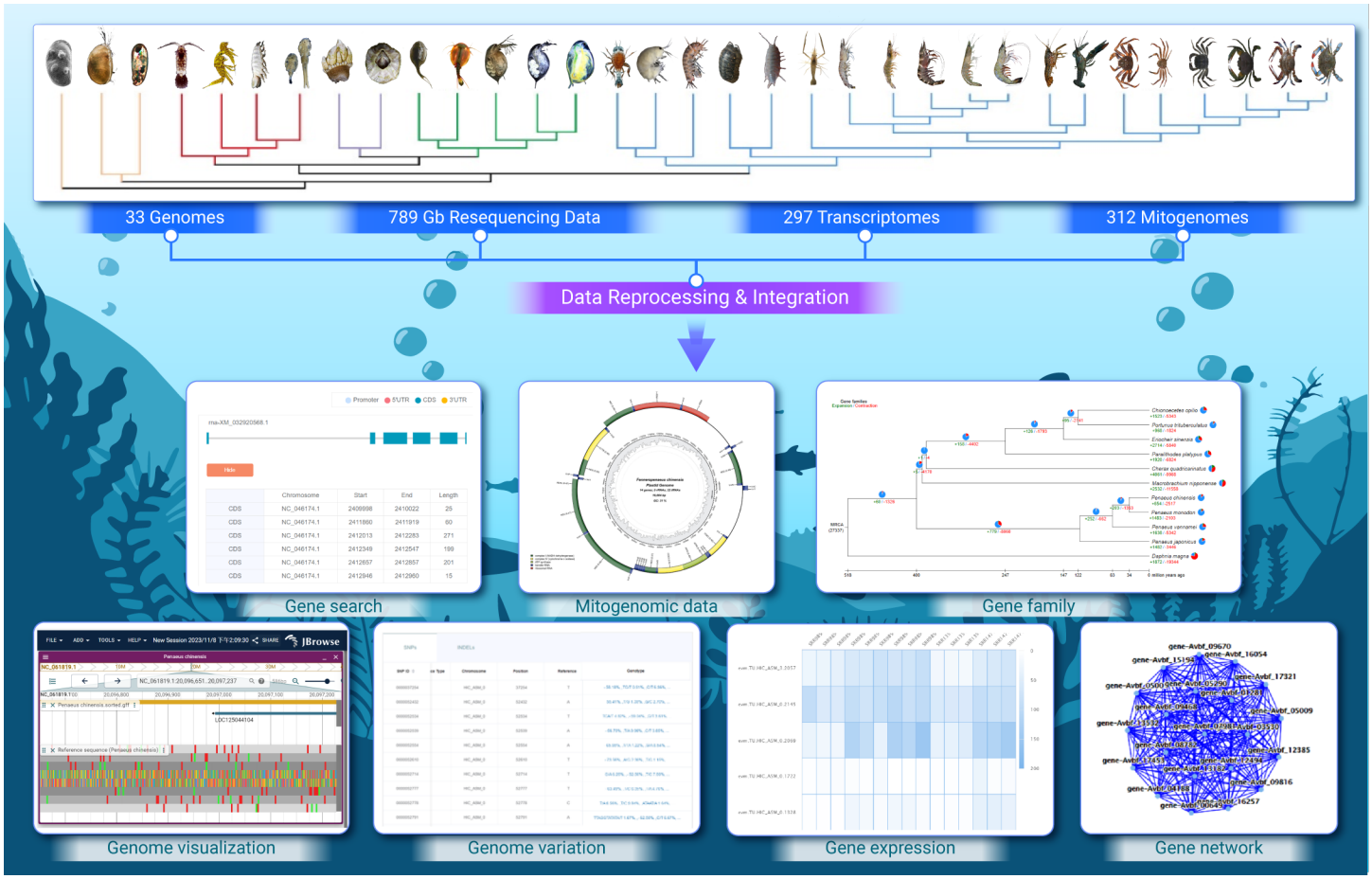
CRUSTADB: An integrated genomics platform for crustaceans

Qiong Wang,[1,3,4] Jianjian Lv,[1,3,4] Ping Liu,[1,3] Xianyun Ren,[1,3] Jitao Li,[1,3] Yuanning Li,[2,*] and Jian Li[1,3,*]

*Correspondence: lijian@ysfri.ac.cn (J.L.); yuanning.li@email.sdu.edu.cn (Y.L.)

Received: June 3, 2024; Accepted: January 4, 2025; Published Online: January 20, 2025; https://doi.org/10.59717/j.xinn-life.2024.100116

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- CRUSTADB is an open-access comprehensive genomic database specifically for crustaceans.
- CRUSTADB provides foundational resources for variant detection and gene expression evaluation within genomes.
- CRUSTADB offers a suite of user-friendly tools designed for multifaceted integrative and comparative analyses.

REPORT

www.the-innovation.org/life

CRUSTADB: An integrated genomics platform for crustaceans

Qiong Wang,[1,3,4] Jianjian Lv,[1,3,4] Ping Liu,[1,3] Xianyun Ren,[1,3] Jitao Li,[1,3] Yuanning Li,[2,*] and Jian Li[1,3,*]

[1]National Key Laboratory of Mariculture Biobreeding and Sustainable Goods, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China
[2]Institute of Marine Science and Technology, Shandong University, Qingdao 266237, China
[3]Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao Marine Science and Technology Centre, Qingdao 266237, China
[4]These authors contributed equally
*Correspondence: lijian@ysfri.ac.cn (J.L.); yuanning.li@email.sdu.edu.cn (Y.L.)

Received: June 3, 2024; Accepted: January 4, 2025; Published Online: January 20, 2025; https://doi.org/10.59717/j.xinn-life.2024.100116

Citation: Wang Q., Lv J., Liu P., et al. (2025). CRUSTADB: An integrated genomics platform for crustaceans. *The Innovation Life* **3**:100116.

Crustacea is a species-rich subphylum of Arthropoda that contains mostly aquatic species. Here, we established a comprehensive genomic database dedicated to the Crustacea. CRUSTADB represents a significant advancement in crustacean genomics, integrating an extensive array of genomic resources. The database encompasses 33 latest genomes, providing a foundational resource for genetic and genomic studies within this clade. Additionally, it includes an impressive 789 Gb of resequencing data across 10 species within the order Decapoda, 297 transcriptomes from 28 diverse species, and 312 mitochondrial genomes, offering a broad spectrum of genetic information. CRUSTADB is not merely a repository of genomic data; it also offers a suite of user-friendly tools designed for multifaceted integrative and comparative analyses. These tools enable users to visually browse the genomes and access a wide range of detailed information, including genome assembly statistics, genomic variations, gene annotations, expression profiles, gene family dynamics, and mitochondrial genomic data. Thus, CRUSTADB will facilitate a comprehensive understanding of crustacean genomics, serving as an invaluable resource for researchers in the field. CRUSTADB can be accessed at http://crustacean_ysfri.qnlm.ac/#/home.

INTRODUCTION

Crustacea is a species-rich subphylum of Arthropoda, predominantly comprising aquatic species. With over 40,000 known species, crustaceans inhabit various regions of the Earth.[1] Crustaceans are significant aquatic source worldwide, particularly within the order Decapoda, which includes economically important species such as shrimp, crabs, crayfish, and lobsters. Additionally, certain crustacean species, such as water fleas and brine shrimp, serve as valuable food sources in aquaculture.[2] According to *The State of World Fisheries and Aquaculture 2024* published by FAO,[3] total global fisheries and aquaculture production in 2022 reached 223.2 million tons, with a total value of approximately $471.8 billion. The global aquaculture and marine fisheries production of crustaceans accounted for 18.4 million tons in 2022, and the traded aquatic animal products of crustacean was 23 percent of the total value, ranking only second to finfish. Thus, crustaceans hold considerable ecological and economic value, making their preservation and sustainable management of utmost importance.

Genomic research on crustaceans has encountered significant hurdles previously due to the intricate structural complexity and high heterozygosity of their genomes.[4] Five years ago, high-quality genomes were only available for three crustaceans, the water flea *Daphnia pulex*,[5] the amphipod *Parhyale hawaiensis*[6] and the marbled crayfish *Procambarus virginalis*.[7] Due to the absence of genomic data of most crustacean species, the mechanisms underlying the formation n of various phenotypes remained poorly characterized, such as the migration between different environments, specific lifestyles, the evolution of specific tissues or organs, and reproduction. Recent advancements in third-generation sequencing technologies have revolutionized genomic research on crustaceans. Whole-genome sequencing is now extensively employed in studying the biology, physiology, ecology, evolution, and selective breeding of crustaceans.[8] Over the past five years, with the publication of genome of the pacific white shrimp *Litopenaeus vannamei*,[9] there has been a remarkable increase in the number of crustacean genomes sequenced.[2] This expanding genomic dataset, coupled with the growing utilization of non-model organisms in molecular biology, offers researchers a plethora of opportunities to test numerous hypotheses and explore various aspects of crustacean biology. As the demand for genomic information continues to grow and accumulate omics data, the development of an integrated genomic database becomes indispensable.

There have been some crustacean related databases published, like CrusTome,[10] a transcriptome database resource for large-scale analyses across Crustacea (https://zenodo.org/records/7730440), Crustybase,[11] an interactive online database for crustacean transcriptomes (https://crustybase.org/), CrusTF,[12] a comprehensive resource of transcriptomes for evolutionary and functional studies of crustacean transcription factors (http://qinlab.sls.cuhk.edu.hk/CrusTF) and CAT,[13] a crustacean annotated transcriptome database (http://cat.sls.cuhk.edu.hk/). However, a comprehensive platform for multi-omics resources of crustacean is absent. Here, we established CRUSTADB (http://crustacean_ysfri.qnlm.ac/#/home), an open-access comprehensive genomic database specifically for crustaceans. CRUSTADB integrates large amounts of genomic resources and provides convenient tools for multi-level integrative and comparative analyses. It offers highly valuable customized datasets or resources. With these resources, visitors could access detailed information on genome assembly statistics, genomic variations, gene annotation, expressional profiles, gene families and mitogenomic data. Moreover, it also provides several useful and convenient tools for BLAST-based sequence comparison and PCR primer design.

METHODS

Database frame construction

CRUSTADB collected 33 high-quality genomes of crustaceans from National Center for Biotechnology Information (NCBI), China National Center for Bioinformation (CNCB), Figshare and our own team, covering 5 Classes, including Malacostraca, Ostracoda, Thecostraca, Hexanauplia and Branchiopoda, and containing 14 Decapoda species. Since the genomic annotation files of *Paralithodes platypus*, *Procambarus virginalis* and *Daphnia pulex* were unavailable before the database construction, we annotated them by ourselves basing on the publicly available databases, including ProDom,[14] PRINTS,[15] Pfam,[16] SMART,[17] PANTHER[18] and PROSITE.[19] The genomes can be browsed visually in the JBrowse module, which provides a dynamic web platform for genomic visualization and analysis.[20]

To detect variants on genomes, we collected 789 Gb resequencing data from NCBI SRA, which covered 100 samples, 10 Decapoda species (Table S2). We also collected RNA-seq data from NCBI SRA to evaluate the expression of each gene in different species, containing 297 transcriptomes from 28 reference genome-profiled species (Table S2). Furthermore, the database collected 312 mitochondrial genomes from NCBI (Table S3). The circle map of each mitochondrial genome was performed on Tutools platform (https://www.cloudtutu.com), a free online data analysis website.

CRUSTADB is implemented with the Linux operating system, using J2EE as the framework, MySQL as the back-end database and nginx as the web server. Web user interfaces were developed by Vue (v. 2.0) JavaScript and JDK Java framework.

Phylogenetic analysis

Ten Decapoda species, *Portunus trituberculatus, Eriocheir sinensis, Paralithodes platypus, Chionoecetes opilio, Penaeus chinensis, Penaeus monodon, Penaeus japonicas, Penaeus vannamei, Macrobrachium nipponense* and *Cherax quadricarinatus* were chosen to perform comparative genomics analysis. *Daphnia magna* was used as out group species. Single-

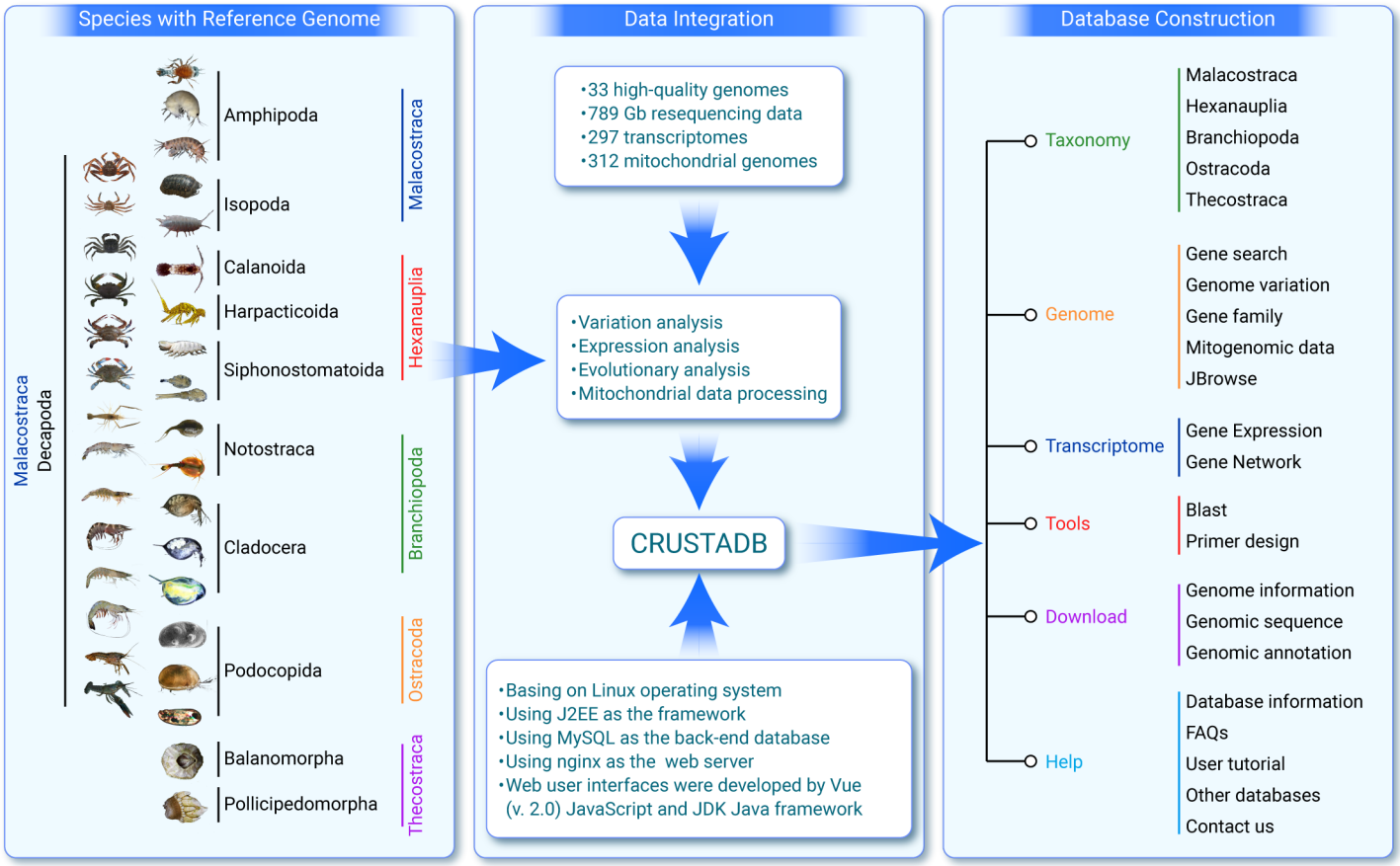


Figure 1. Overview of the CRUSTADB and summary of data composition and module setting.

copy homologous genes of the 11 species were used for phylogenetic tree construction. Briefly, MUSCLE [21,22] was used for multiple sequence alignment on each single-copy homologous gene with the default settings. The alignment results were merged to form a super alignment matrix, and was used to construct the phylogenetic tree with RAxML,[23,24] maximum likelihood (ML) method was used under the JTT matrix-based model. The best tree was used as an input tree for divergence time estimation using MCMCTREE in the PAML package.[25] In the gene family expansion and contraction analysis, we filtered the gene families with the results of the clustering analysis of gene families using CAFE software.[26]

Resequencing data reprocessing

The database integrates 789 Gb resequencing data from 100 samples, covering 10 Decapoda species (table S4). All sequencing data was filtered by NGS QC Toolkit[27] with default parameters. The genome resequencing data were mapped to reference genomes using BWA[28] with default parameters. Then the BAM files were sorted and duplicate reads were removed using Picard toolkit (https://github.com/broadinstitute/picard). Finally, the Genome Analysis Toolkit (GATK)[29] was used for SNP and indel calling.

Transcriptome data reprocessing

The database collected 297 transcriptomes from 28 reference genome-profiled species (table S5). Clean reads of each RNA-seq library were aligned to the reference genome to obtain unique mapped reads by using the tool of STAR[30] with default parameters. Only reads with a perfect match or one mismatch were further analyzed and annotated. The read counts were adjusted by DEseq2[31,32] and edgeR program package.[33] Expression level of each transcript for each tissue was calculated and normalized into FPKM (fragments per kilobase of transcript per million fragments mapped) values by RSEM software.[34] The Gene Network module constructs gene network basing on gene expression data. Pearson's correlation coefficient (PCC) between two genes was used to evaluate the correlation.

RESULTS

Genomic data presentation

CRUSTADB is a comprehensive genomic database for crustaceans. It has collected 33 high-quality genomes of the crustaceans, including Malacostraca, Ostracoda, Thecostraca, Hexanauplia, and Branchiopoda, and containing 14 Decapoda species, covering most of the high-quality genomes of crustaceans currently published (Figure 1 & Table 1). The database provides essential information, genomic assembly details, and images of these species; it also offers an interactive module for gene search and browsing specific gene details. The inclusion of these genomes will significantly contribute to the research on the evolutionary history of crustaceans.

The home page displays the phylogenetic tree and photograph slideshow of the 33 crustacean species that have high-quality reference genomes. Users can view the species information of Taxonomy module by click on the species name on phylogenetic tree or photograph in slideshow. The Taxonomy module displays the essential information and genome assembly statistic of the 33 crustacean species (Figure 2A). Users can choose class to search species, and choose species to browse the essential information. The genomic sequences and annotation information could be browse visually in the JBrowse module, which provides a dynamic web platform for genomic visualization and analysis (Figure 2B).

Basic information of protein-coding genes in 33 crustacean species was integrated in Gene Search module, including gene location, gene size, transcription direction and functional annotations (Figure 2C). Choose class and species to search the genes of this species. Three types of keywords, genomic region, gene ID and gene name, can be used to retrieve gene information. Users can click the gene ID to browse the details of searched gene, like basic information, structure information and functional annotation, and click on the download button to download the gene information.

The Mitogenomic Data module integrates the mitogenomic data of 312 crustacean species (Figure 2D). Users can view the mitochondrial map and annotation information by retrieving species in the list, or type in keyword, and download related resources.

Table 1. Summary of the genomic data of crustaceans

Species Name	Assembly Version	Assembly Level	Genome Size	Publication
Callinectes sapidus	Csap_IMET_v1	Chromosome	997,995,872	Bachvaroff, et al.[35]
Scylla paramamosain	Sp_Genome_20200329	Chromosome	1,546,139,864	Zhao, et al.[36]
Portunus trituberculatus	ASM2074055v1	Chromosome	864,443,858	Lv, et al.[37]
Eriocheir sinensis	ASM333651v1	Chromosome	1,567,615,418	Cui, et al.[38]
Paralithodes platypus	ASM1328300v1	Chromosome	4,805,175,703	Tang, et al.[39]
Chionoecetes opilio	ASM1658430v1	Scaffold	2,002,919,378	-
Penaeus chinensis	ASM1920278v2	Chromosome	1,466,079,641	Wang, et al.[40]
Penaeus monodon	NSTDA_Pmon_1	Chromosome	2,394,347,767	Uengwetwanit, et al.[41]
Penaeus japonicus	-	Chromosome	1,535,262,098	Ren, et al.[42]
Penaeus vannamei	ASM378908v1	Scaffold	1,663,581,301	Zhang, et al.[9]
Penaeus indicus	CIBA_Pind_1.1	Scaffold	1,935,640,391	-
Macrobrachium nipponense	ASM1510439v1	Chromosome	1,985,027,095	Jin, et al.[43]
Cherax quadricarinatus	-	Chromosome	4,486,779,883	Article in Press
Procambarus virginalis	DKFZ_Pvir_1.0	Scaffold	3,700,845,369	Gutekunst, et al.[44]
Armadillidium nasatum	CNRS_Arma_nasa_1.0	Scaffold	1,223,175,971	Becking, et al.[45]
Armadillidium vulgare	Arma_vul_BF2787	Scaffold	1,725,108,002	Chebbi, et al.[46]
Hyalella azteca	Hazt_2.0.1	Scaffold	550,885,727	Poynton, et al.[47]
Trinorchestia longiramus	ASM678305v1	Scaffold	886,359,443	Patra, et al.[48]
Parhyale hawaiensis	Phaw_5.0	Scaffold	2,752,560,740	Kao, et al.[6]
Darwinula stevensoni	Dst_b1v02_max_arth_ b2g_genome	Scaffold	382,098,207	Tran Van, et al.[49]
Notodromas monacha	Nmo_b1v02_max_arth_ b2g_genome	Scaffold	376,665,953	Tran Van, et al.[49]
Cyprideis torosa	Cto_b1v02_max_arth_ b2g_genome	Scaffold	334,833,579	Tran Van, et al.[49]
Amphibalanus amphitrite	SNU_Aamp_1	Scaffold	613,414,131	-
Pollicipes pollicipes	Ppol_2.1	Scaffold	770,104,822	Bernot, et al.[50]
Caligus rogercresseyi	ASM1338718v1	Chromosome	478,201,101	Gallardo-Escarate, et al.[51]
Lepeophtheirus salmonis	UVic_Lsal_1.1	Chromosome	650,284,782	Skern-Mauritzen, et al.[52]
Tigriopus californicus	Tcal_SD_v2.1	Chromosome	191,142,546	Barreto, et al.[53]
Eurytemora affinis	Eaff_2.0	Scaffold	389,032,277	Eyun, et al.[54]
Daphnia pulicaria	SC_F0-13Bv2	Chromosome	185,057,732	Wersebe, et al.[55]
Daphnia magna	ASM399081v1	Chromosome	122,937,721	Lee, et al.[56]
Daphnia pulex	ASM2113471v1	Chromosome	133,196,385	-
Lepidurus arcticus	ASM372404v1	Scaffold	73,105,941	Savojardo, et al.[57]
Lepidurus apus	Lubb2018	Scaffold	87,970,075	Savojardo, et al.[57]

Application in phylogenetic analysis

Comparative genomics analysis was performed on 10 Decapoda species, *Portunus trituberculatus, Eriocheir sinensis, Paralithodes platypus, Chionoecetes opilio, Penaeus chinensis, Penaeus monodon, Penaeus japonicas, Penaeus vannamei, Macrobrachium nipponense* and *Cherax quadricarinatus*. Through identification of homologous genes and clustering analysis of gene families, single-copy gene families and multi-copy gene families were obtained (Figure 3A). These families are relatively conserved among species. There are 1,827 common families among these species, 504 of them are single-copy genes. Species-specific gene families also were obtained, which may be related to species-specific (Figure 3B & Table S1). The single-copy homologous genes of the 11 species were used for phylogenetic tree construction (Figure 3C).

The expanded and contracted gene family information of each species is provided in the Gene Family module. In the phylogenetic tree, the numbers of expanded and contracted gene families are labelled with red and green, respectively. A total of 301,275 gene families, 22,526 expanded and 98,640 contracted, were detected among these species. Users can acquire the divergence time and expanded/contracted gene family numbers of these species, and download details of these expanded/contracted gene families. Phylogenetic studies based on these genomes play a crucial role in addressing important biological questions, such as freshwater adaptation, the formation of short-tailed crabs, body size variations, and more.

Variant detection and gene expression evaluation within genomes

By collecting and processing 789 Gb of resequencing data from 100

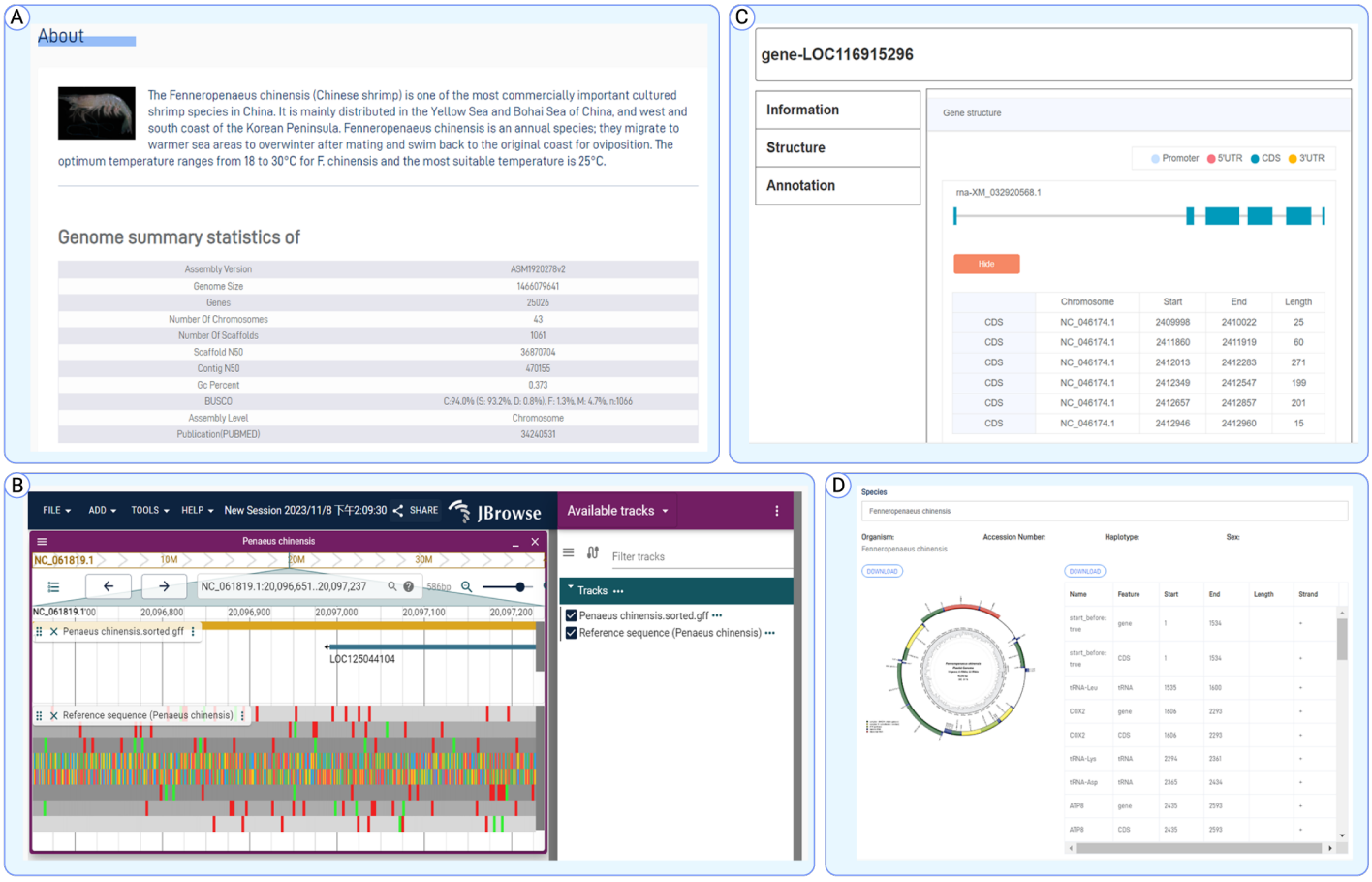


Figure 2. Screenshots of the web interfaces of the genomic modules (A) The species information page with details of genome assembly. (B) The JBrowse module. (C) The Gene search page. (D) The mitogenomic information page.

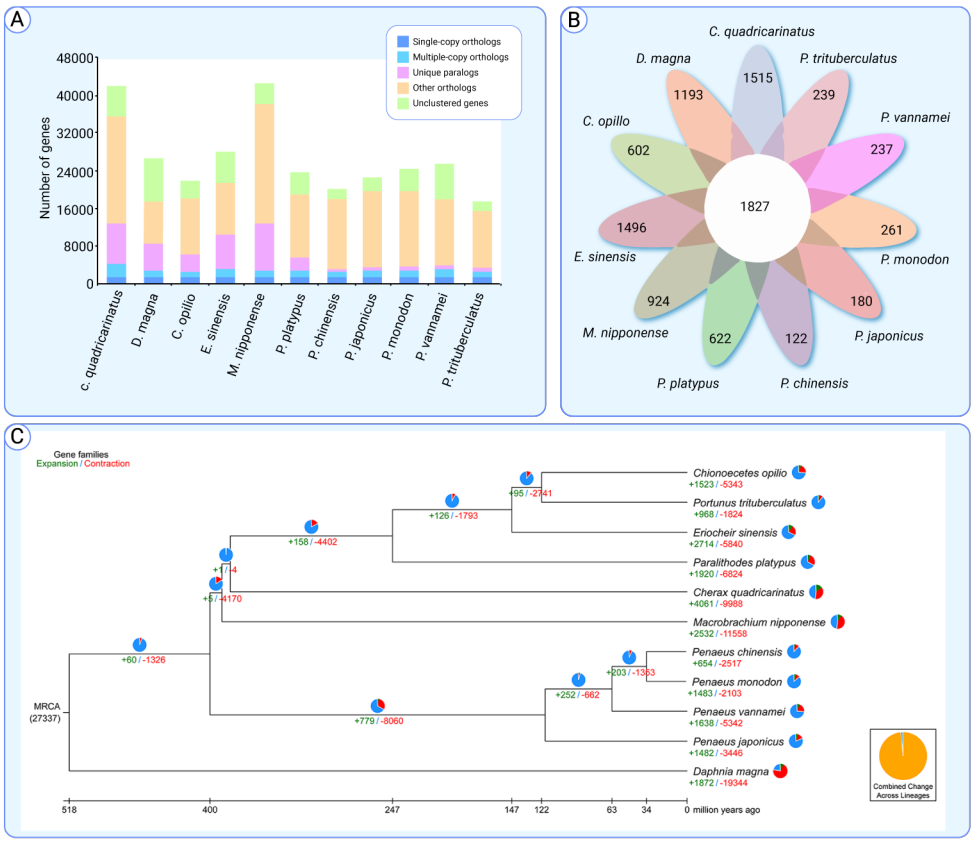


Figure 3. Phylogenetic analysis of the Decapoda species (A) Statistic of the homologous gene number. (B) The common and unique gene family number of each species. (C) Phylogenetic tree construction and expanded (green)/contracted (red) gene family detection.

samples across 10 Decapoda species, 65,796,396 SNPs and 19,216,334 indels were identified and included in the database (Figure 4A). Users can search variations and view the mutation details in the Genome Variation module by choosing species and inputting genomic region or gene name.

Abundant transcriptomes were integrated by the Gene Expression module to evaluate the expression level of each gene in different tissues (Figure 4B). Users can select the name in the list to retrieve species, and type in keyword to retrieve interested gene or click the search icon directly to retrieve all genes, and click the interested genes to view their expression profile.

The Gene Network module constructs gene network basing on gene expression data (Figure 4C). Select the name in the list to retrieve species, and enter the gene ID and set the PCC value threshold to visualize the top 20 co-expressed genes with the highest PCC value. A summary table of these co-expressed

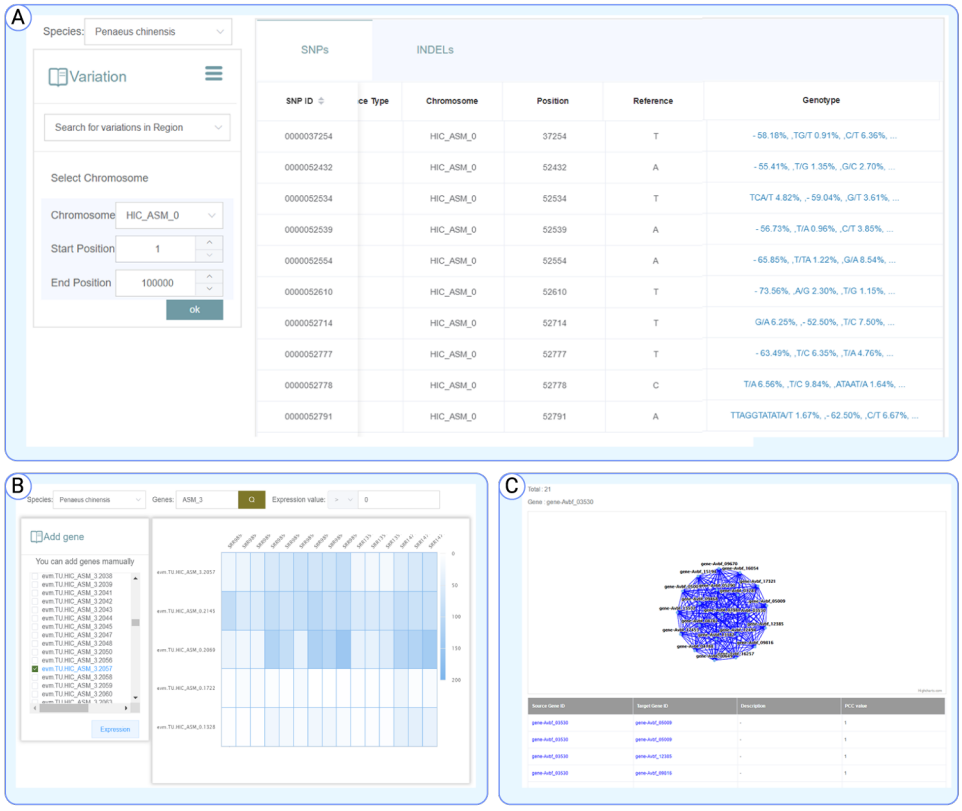


Figure 4. Screenshots of the resequencing and transcriptomic related modules (A) The Genome variation module. (B) The Gene expression module. (C) The Gene network module.

genes are provided. Users can click the geneID to browse the basic information of target gene in the summary table.

Convenient tools

CRUSTADB provides several useful and convenient tools. The BLAST tool is used for sequence search and comparison of supplied sequence data. The background databases of nucleotide and protein are from the 33 crustacean species that have well-annotated genomes. Enter a sequence or file in FASTA format and select the nucleotide or protein database. Users can view the details of the BLAST result. The Primer tool is used to design primers by inputting nucleotide sequences. Users can set the PCR parameters and click the search button to make a query.

All genomic sequence and annotation data are publicly available. The genomic information is tabulated in a summary table of Download module. Click download button to obtain genomic sequence and annotation files.

DISCUSSION

Crustaceans thrive in every type of aquatic ecosystem and demonstrate remarkable adaptability, even in extreme conditions such as temperature, pressure, salinity, and anoxia.[58] Crustaceans are covered with chitinous exoskeletons, and must be shed for growth.[59] Generally, the body of adult crustacean is constituted by four primary tagmata, head, pereon, pleon and telson.[58,60] But, this body plan is not universal, some species evolved special forms, such as barnacles, depart radically from the typical crustacean that they are easily confused with species of other phyla. However, due to structural complexity and high heterozygosity of the genomes of most crustacean species, development of genomic researches is slow.[4] There has been little progress in the characterization of many phenotypes of crustaceans, such as the migration between different environments, specific lifestyles, the evolution of specific tissues or organs and reproduction.[2]

In recent years, as the nuclear genomes of a large number of species have been successfully assembled, whole genome sequencing (WGS) and SNPs detection have also been widely used.[61] Some SNPs are strongly correlated with different traits,[62-67] which can usually provide some new clues for species evolution and environmental adaptation research. Whole transcriptome analysis provides the first step toward functional characterization and annotation of genes previously revealed by DNA sequencing.[68] The genomic and transcriptomic sequencing data play essential roles in decoding genomic structure and function, identifying genetic networks underlying cellular, physiological, biochemical and biological systems and discovering molecular biomarkers that response to diseases, pathogens and environmental challenges.[69] Molecular phylogenetic studies have categorized the extant crustacean lineages into several groups, contains Malacostraca, Ostracoda, Branchiopoda, Thecostraca, Hexanauplia and Remipedia, etc.[70-74] Some studies considered Pancrustacea with Remipedia as the sister group to Hexapoda.[75] The fossil evidences showed that crustaceans were present and quite widely diversified by the Cambrian period over 500 million years ago, and became relatively common in the Permian a little less than 300 million years ago.[58,60] There is increasing evidences suggest that crustaceans and insects are very closely related. The pancrustacean hypothesis considered that the insects arise from the crustacean lineage.[76,77] One theory is that crustaceans firstly evolved in the young Earth's seas, a branchiopod-like ancestral crustacean later colonized freshwater habitats, and finally invaded the land as insects.[78] However, the relationship between the major classes of crustaceans remains contentious.[58] Testing these evolutionary hypotheses need to rely on comparative genomics studies. The genomes collected by this database will contribute to research the evolutionary history of crustaceans. Phylogenetic studies or other comparative analyses on these genomes may help to solve a series of important biological issues, like freshwater adaptation, short-tailing formation of crabs, body size variations, and so on.

In the future, CRUSTADB will be periodically updated. The development team will continue to extend the collection of public high-quality genomes and to further enhance the identification of gene families. Increasing availability of new high-throughput sequencing data will be used to characterize the genomic variation, gene expression and gene network. These additions are anticipated to enhance the efficiency of the applications of CRUSTADB in aquatic breeding and phylogenetic study.

CONCLUSIONS

CRUSTADB is an open-access comprehensive genomic database specifically for crustaceans. Compared to the existing crustacean-related databases, which mostly focus on specific types of resources, it covers more types of data resources, containing high-quality genome assemblies, resequencing data, transcriptomes data, and mitochondrial genomes. CRUSTADB is not merely a repository of genomic data, it also offers a suite of user-friendly tools designed for multifaceted integrative and comparative analyses. These tools enable users to visually browse the genomes and access a wide range of detailed information, including genome assembly statistics, genomic variations, gene annotations, expression profiles, gene family dynamics, and mitochondrial genomic data. With this database, users could obtain the downstream information of interested genes which were explored from omics analysis, and compared their expression level with the samples collected by our database.

REFERENCES

1. Soderhall I. (2016). Crustacean hematopoiesis. *Dev. Comp. Immunol.* **58**:129–141. DOI:10.1016/j.dci.2015.12.009
2. Yuan J., Yu Y., Zhang X., et al. (2023). Recent advances in crustacean genomics and

their potential application in aquaculture. *Rev. Aquac.* **15**:1–21. DOI:10.1111/raq.12791
3. FAO (2024). The State of World Fisheries and Aquaculture 2024. Blue Transformation in action.
4. Stillman J. H., Colbourne J. K., Lee C. E., et al. (2008). Recent advances in crustacean genomics. *Integr. Comp. Biol.* **48**:852–868. DOI:10.1093/icb/icn096
5. Colbourne J. K., Pfrender M. E., Gilbert D., et al. (2011). The ecoresponsive genome of Daphnia pulex. *Science* **331**:555–561. DOI:10.1126/science.1197761
6. Kao D., Lai A. G., Stamataki E., et al. (2016). The genome of the crustacean Parhyale hawaiensis, a model for animal development, regeneration, immunity and lignocellulose digestion. *eLife* **5**:e20062. DOI:10.7554/eLife.20062
7. Gutekunst J., Andriantsoa R., Falckenhayn C., et al. (2018). Clonal genome evolution and rapid invasive spread of the marbled crayfish. *Nat. Ecol. Evol.* **2**:567–573. DOI:10.1038/s41559-018-0467-9
8. Aquaculture Genomics, Genetics and Breeding Workshop, Abdelrahman H., ElHady M., et al. (2017). Aquaculture genomics, genetics and breeding in the United States: Current status, challenges, and priorities for future research. *BMC Genomics* **18**:191. DOI:10.1186/s12864-017-3557-1
9. Zhang X., Yuan J., Sun Y., et al. (2019). Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nat. Commun.* **10**:356. DOI:10.1038/s41467-018-08197-4
10. Perez-Moreno J. L., Kozma M. T., DeLeo D. M., et al. (2023). CrusTome: A transcriptome database resource for large-scale analyses across Crustacea. *G3 (Bethesda)* **13**:jkad098. DOI:10.1093/g3journal/jkad098
11. Hyde C. J., Fitzgibbon Q. P., Elizur A., et al. (2020). CrustyBase: An interactive online database for crustacean transcriptomes. *BMC Genomics* **21**:637. DOI:10.1186/s12864-020-07063-2
12. Qin J., Hu Y., Ma K. Y., et al. (2017). CrusTF: A comprehensive resource of transcriptomes for evolutionary and functional studies of crustacean transcription factors. *BMC Genomics* **18**:908. DOI:10.1186/s12864-017-4305-2
13. Nong W., Chai Z. Y. H., Jiang X., et al. (2020). A crustacean annotated transcriptome (CAT) database. *BMC Genomics* **21**:32. DOI:10.1186/s12864-019-6433-3
14. Servant F., Bru C., Carrere S., et al. (2002). ProDom: automated clustering of homologous domains. *Brief Bioinform* **3**:246–251. DOI:10.1093/bib/3.3.246
15. Attwood T. K., Beck M. E., Bleasby A. J., et al. (1994). PRINTS--a database of protein motif fingerprints. *Nucleic Acids Res.* **22**:3590-3596.
16. Mistry J., Chuguransky S., Williams L., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**:D412–D419. DOI:10.1093/nar/gkaa913
17. Letunic I., Khedkar S. and Bork P. (2021). SMART: Recent updates, new developments and status in 2020. *Nucleic Acids Res.* **49**:D458–D460. DOI:10.1093/nar/gkaa937
18. Thomas P. D., Ebert D., Muruganujan A., et al. (2022). PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci.* **31**:8–22. DOI:10.1002/pro.4218
19. Hulo N., Bairoch A., Bulliard V., et al. (2006). The PROSITE database. *Nucleic Acids Res.* **34**:D227–230. DOI:10.1093/nar/gkj063
20. Buels R., Yao E., Diesh C. M., et al. (2016). JBrowse: A dynamic web platform for genome visualization and analysis. *Genome Biol.* **17**:66. DOI:10.1186/s13059-016-0924-1
21. Edgar R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**:1792–1797. DOI:10.1093/nar/gkh340
22. Edgar R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **5**:113. DOI:10.1186/1471-2105-5-113
23. Rokas A. (2011). Phylogenetic analysis of protein sequence data using the Randomized Axelerated Maximum Likelihood (RAXML) Program. *Curr Protoc. Mol. Biol.* **Chapter 19**:Unit19 11. DOI:10.1002/0471142727.mb1911s96.
24. Stamatakis A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313. DOI:10.1093/bioinformatics/btu033
25. Yang Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**:1586–1591. DOI:10.1093/molbev/msm088
26. De Bie T., Cristianini N., Demuth J. P., et al. (2006). CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **22**:1269–1271. DOI:10.1093/bioinformatics/btl097
27. Patel R. K. and Jain M. (2012). NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS One* **7**:e30619. DOI:10.1371/journal.pone.0030619
28. Li H. and Durbin R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**:589–595. DOI:10.1093/bioinformatics/btp698
29. McKenna A., Hanna M., Banks E., et al. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**:1297–1303. DOI:10.1101/gr.107524.110
30. Dobin A., Davis C. A., Schlesinger F., et al. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. DOI:10.1093/bioinformatics/bts635
31. Anders S. and Huber W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* **11**:R106. DOI:10.1186/gb-2010-11-10-r106
32. Love M. I., Huber W. and Anders S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**:550. DOI:10.1186/s13059-014-0550-8
33. Robinson M. D., McCarthy D. J. and Smyth G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139–140. DOI:10.1093/bioinformatics/btp616
34. Li B. and Dewey C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**:323. DOI:10.1186/1471-2105-12-323
35. Bachvaroff T. R., McDonald R. C., Plough L. V., et al. (2021). Chromosome-level genome assembly of the blue crab, Callinectes sapidus. *G3 (Bethesda)* **11**:jkab212. DOI:10.1093/g3journal/jkab212.
36. Zhao M., Wang W., Zhang F., et al. (2021). A chromosome-level genome of the mud crab (Scylla paramamosain estampador) provides insights into the evolution of chemical and light perception in this crustacean. *Mol. Ecol. Resour.* **21**:1299–1317. DOI:10.1111/1755-0998.13332
37. Lv J., Li R., Su Z., et al. (2022). A chromosome-level genome of Portunus trituberculatus provides insights into its evolution, salinity adaptation and sex determination. *Mol. Ecol. Resour.* **22**:1606–1625. DOI:10.1111/1755-0998.13564
38. Cui Z., Liu Y., Yuan J., et al. (2021). The Chinese mitten crab genome provides insights into adaptive plasticity and developmental regulation. *Nat. Commun.* **12**:2395. DOI:10.1038/s41467-021-22604-3
39. Tang B., Wang Z., Liu Q., et al. (2021). Chromosome-level genome assembly of Paralithodes platypus provides insights into evolution and adaptation of king crabs. *Mol. Ecol. Resour.* **21**:511–525. DOI:10.1111/1755-0998.13266
40. Wang Q., Ren X., Liu P., et al. (2022). Improved genome assembly of Chinese shrimp (Fenneropenaeus chinensis) suggests adaptation to the environment during evolution and domestication. *Mol. Ecol. Resour.* **22**:334–344. DOI:10.1111/1755-0998.13463
41. Uengwetwanit T., Pootakham W., Nookaew I., et al. (2021). A chromosome-level assembly of the black tiger shrimp (Penaeus monodon) genome facilitates the identification of growth-associated genes. *Mol. Ecol. Resour.* **21**:1620–1640. DOI:10.1111/1755-0998.13357
42. Ren X., Lv J., Liu M., et al. (2022). A chromosome-level genome of the kuruma shrimp (Marsupenaeus japonicus) provides insights into its evolution and cold-resistance mechanism. *Genomics* **114**:110373. DOI:10.1016/j.ygeno.2022.110373
43. Jin S., Bian C., Jiang S., et al. (2021). A chromosome-level genome assembly of the oriental river prawn, Macrobrachium nipponense. *Gigascience* **10**:giaa160. DOI:10.1093/gigascience/giaa160
44. Gutekunst J., Maiakovska O., Hanna K., et al. (2021). Phylogeographic reconstruction of the marbled crayfish origin. *Commun. Biol.* **4**:1096. DOI:10.1038/s42003-021-02609-w
45. Becking T., Chebbi M. A., Giraud I., et al. (2019). Sex chromosomes control vertical transmission of feminizing Wolbachia symbionts in an isopod. *PLoS Biol.* **17**:e3000438. DOI:10.1371/journal.pbio.3000438
46. Chebbi M. A., Becking T., Moumen B., et al. (2019). The genome of Armadillidium vulgare (crustacea, isopoda) provides insights into sex chromosome evolution in the context of cytoplasmic sex determination. *Mol. Biol. Evol.* **36**:727–741. DOI:10.1093/molbev/msz010
47. Poynton H. C., Hasenbein S., Benoit J. B., et al. (2018). The toxicogenome of Hyalella azteca: A model for sediment ecotoxicology and evolutionary toxicology. *Environ. Sci. Technol.* **52**:6009–6022. DOI:10.1021/acs.est.8b00837
48. Patra A. K., Chung O., Yoo J. Y., et al. (2020). First draft genome for the sand-hopper Trinorchestia longiramus. *Sci. Data* **7**:85. DOI:10.1038/s41597-020-0424-8
49. Tran Van P., Anselmetti Y., Bast J., et al. (2021). First annotated draft genomes of nonmarine ostracods (Ostracoda, Crustacea) with different reproductive modes. *G3 (Bethesda)* **11**:jkab043. DOI:10.1093/g3journal/jkab043.
50. Bernot J. P., Avdeyev P., Zamyatin A., et al. (2022). Chromosome-level genome assembly, annotation, and phylogenomics of the gooseneck barnacle Pollicipes pollicipes. *Gigascience* **11**:.giac021. DOI:10.1093/gigascience/giac021
51. Gallardo-Escarate C., Valenzuela-Munoz V., Nunez-Acuna G., et al. (2021). Chromosome-scale genome assembly of the sea louse Caligus rogercresseyi by SMRT sequencing and Hi-C analysis. *Sci. Data* **8**:60. DOI:10.1038/s41597-021-00842-w
52. Skern-Mauritzen R., Malde K., Eichner C., et al. (2021). The salmon louse genome: Copepod features and parasitic adaptations. *Genomics* **113**:3666–3680. DOI:10.1016/j.ygeno.2021.08.002
53. Barreto F. S., Watson E. T., Lima T. G., et al. (2018). Genomic signatures of mitonuclear coevolution across populations of Tigriopus californicus. *Nat. Ecol. Evol.* **2**:1250–1257. DOI:10.1038/s41559-018-0588-1
54. Eyun S. I., Soh H. Y., Posavi M., et al. (2017). Evolutionary history of chemosensory-related gene families across the Arthropoda. *Mol. Biol. Evol.* **34**:1838–1862. DOI:10.1093/molbev/msx147
55. Wersebe M. J., Sherman R. E., Jeyasingh P. D., et al. (2023). The roles of recombination and selection in shaping genomic divergence in an incipient ecological species complex. *Mol. Ecol.* **32**:1478–1496. DOI:10.1111/mec.16383
56. Lee B. Y., Choi B. S., Kim M. S., et al. (2019). The genome of the freshwater water flea Daphnia magna: A potential use for freshwater molecular ecotoxicology. *Aquat. Toxicol.* **210**:69–84. DOI:10.1016/j.aquatox.2019.02.009
57. Savojardo C., Luchetti A., Martelli P. L., et al. (2019). Draft genomes and genomic divergence of two Lepidurus tadpole shrimp species (Crustacea, Branchiopoda, Notostraca). *Mol. Ecol. Resour.* **19**:235–244. DOI:10.1111/1755-0998.12952
58. VanHook A. M. and Patel N. H. (2008). Crustaceans. *Curr. Biol.* **18**:R547–550. DOI:10.

1016/j.cub.2008.05.021
59. Hyde C. J., Elizur A. and Ventura T. (2019). The crustacean ecdysone cassette: A gatekeeper for molt and metamorphosis. *J. Steroid. Biochem. Mol. Biol.* **185**:172–183. DOI:10.1016/j.jsbmb.2018.08.012
60. Giribet G. and Edgecombe G. D. (2012). Reevaluating the arthropod tree of life. *Annu. Rev. Entomol.* **57**:167–186. DOI:10.1146/annurev-ento-120710-100659
61. Yao Z., Long S., Wang C., et al. (2022). Population genetic characteristics of Hainan medaka with whole-genome resequencing. *Front. Genet.* **13**:946006. DOI:10.3389/fgene.2022.946006
62. Jones F. C., Grabherr M. G., Chan Y. F., et al. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**:55–61. DOI:10.1038/nature10944
63. Gaither M. R., Gkafas G. A., de Jong M., et al. (2018). Genomics of habitat choice and adaptive evolution in a deep-sea fish. *Nat. Ecol. Evol.* **2**:680–687. DOI:10.1038/s41559-018-0482-x
64. Narum S. R., Di Genova A., Micheletti S. J., et al. (2018). Genomic variation underlying complex life-history traits revealed by genome sequencing in Chinook salmon. *Proc. Biol. Sci.* **285**. DOI:10.1098/rspb.2018.0935.
65. Brookes B., Jeon H. B., Derry A. M., et al. (2022). Neutral and adaptive drivers of genomic change in introduced brook trout (Salvelinus fontinalis) populations revealed by pooled sequencing. *Ecol. Evol.* **12**:e8584. DOI:10.1002/ece3.8584
66. Ezzeddini R., Somi M. H., Taghikhani M., et al. (2021). Association of Foxp3 rs3761548 polymorphism with cytokines concentration in gastric adenocarcinoma patients. *Cytokine* **138**:155351. DOI:10.1016/j.cyto.2020.155351
67. Guo L., Gao Z. and Qian Q. (2014). Application of resequencing to rice genomics, functional genomics and evolutionary analysis. *Rice (N Y)* **7**:4. DOI:10.1186/s12284-014-0004-7
68. Altman R. B. and Raychaudhuri S. (2001). Whole-genome expression analysis: Challenges beyond clustering. *Curr. Opin. Struct. Biol.* **11**:340–347. DOI:10.1016/s0959-440x(00)00212-8
69. Jiang Z., Zhou X., Li R., et al. (2015). Whole transcriptome analysis with sequencing: Methods, challenges and potential solutions. *Cell Mol. Life Sci.* **72**:3425–3439. DOI:10.1007/s00018-015-1934-y
70. Toyota K., Miyakawa H., Hiruta C., et al. (2021). Sex determination and differentiation in decapod and Cladoceran Crustaceans: An overview of endocrine regulation. *Genes-Basel* **12**:305. DOI:10.3390/genes12020305
71. Schwentner M., Combosch D. J., Nelson J. P., et al. (2017). A phylogenomic solution to the origin of insects by resolving Crustacean-Hexapod relationships. *Curr. Biol.* **27**:1818–1824.e5. DOI:10.1016/j.cub.2017.05.040
72. Oakley T. H., Wolfe J. M., Lindgren A. R., et al. (2013). Phylotranscriptomics to bring the understudied into the fold: Monophyletic ostracoda, fossil placement, and pancrustacean Phylogeny. *Mol. Biol. Evol.* **30**:215–233. DOI:10.1093/molbev/mss216
73. Schoch C. L., Ciufo S., Domrachev M., et al. (2020). NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database (Oxford)* **2020**:baaa062. DOI:10.1093/database/baaa062
74. Bernot J. P., Owen C. L., Wolfe J. M., et al. (2023). Major revisions in pancrustacean phylogeny and evidence of sensitivity to taxon sampling. *Mol. Biol. Evol.* **40**. DOI:10.1093/molbev/msad175.
75. von Reumont B. M., Jenner R. A., Wills M. A., et al. (2012). Pancrustacean phylogeny in the light of new phylogenomic data: Support for Remipedia as the possible sister group of Hexapoda. *Mol. Biol. Evol.* **29**:1031–1045. DOI:10.1093/molbev/msr270
76. Richter S. (2002). The Tetraconata concept: Hexapod-crustacean relationships and the phylogeny of Crustacea. *Org. Divers Evol.* **2**:217–237. DOI:10.1078/1439-6092-00048
77. Jenner R. A. (2010). Higher-level crustacean phylogeny: Consensus and conflicting hypotheses. *Arthropod. Struct. Dev.* **39**:143–153. DOI:10.1016/j.asd.2009.11.001
78. Richter S., Loesel R., Purschke G., et al. (2010). Invertebrate neurophylogeny: Suggested terms and definitions for a neuroanatomical glossary. *Front. Zool.* **7**:29. DOI:10.1186/1742-9994-7-29

FUNDING AND ACKNOWLEDGMENTS

We acknowledge lab members in Jian Li and Ping Liu laboratory for discussion and critical comments on the manuscript; Dr Qi Liu and his team from Wuhan Onemoretech Co., Ltd for technique support on database construction and NGS data reprocessing. This work was supported by the National Key Research and Development Program of China (2022YFD2400104); the China Agriculture Research System of MOF and MARA (CARS-48); the Foundation of State Key Laboratory of Mariculture Biobreeding and Sustainable Goods (BRESG202306); the National Natural Science Foundation of China (42076116, 31902367); and the Central Public-interest Scientific Institution Basal Research Fund of CAFS (2023TD50, 2023XT0204). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Qiong Wang performed the data analyses and drafted the manuscript, Jianjian Lv performed data interpretation and revised the manuscript, Ping Liu, Xianyun Ren, and Jitao Li performed data acquisition and result presentation, Yuanning Li and Jian Li planned and designed the project. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

ETHICAL STATEMENT AND PATIENT CONSENT

Not applicable.

DATA AND CODE AVAILABILITY

All data generated or analysed during this study are included in the CRUSTADB database (http://crustacean_ysfri.qnlm.ac/#/home). The data are available on request from the corresponding authors.

SUPPLEMENTAL INFORMATION

It can be found online at https://doi.org/10.59717/j.xinn-life.2024.100116

LEAD CONTACT WEBSITE

Jian Li: http://ysfri.cafs.ac.cn/info/1093/13251.htm
Yuanning Li: https://faculty.sdu.edu.cn/liyuanning/zh_CN/