BMC Biology

RESEARCH

Open Access

High-resolution chromosome-level genome of *Scylla paramamosain* provides molecular insights into adaptive evolution in crabs

Yin Zhang[1,2,3†], Ye Yuan[1,2,3†], Mengqian Zhang[1,2,3†], Xiaoyan Yu[1,2,3], Bixun Qiu[1,2,3], Fangchun Wu[1,2,3], Douglas R. Tocher[1], Jiajia Zhang[1,2,3], Shaopan Ye[1,2,3], Wenxiao Cui[1,2,3], Jonathan Y. S. Leung[1,2,3], Mhd Ikhwanuddin[2,3,4], Waqas Waqas[1,2,3], Tariq Dildar[1,2,3] and Hongyu Ma[1,2,3*]

Abstract

Background Evolutionary adaptation drives organismal adjustments to environmental pressures, exemplified in the diverse morphological and ecological adaptations seen in Decapoda crustaceans, particularly brachyuran crabs. Crabs thrive in diverse ecosystems, from coral reefs to hydrothermal vents and terrestrial habitats. Despite their ecological importance, the genetic mechanisms underpinning their developmental processes, reproductive strategies, and nutrient acquisition remain poorly understood.

Results Here, we report a comprehensive genomic analysis of the green mud crab *Scylla paramamosain* using ultralong sequencing technologies, achieving a high-quality chromosome-level assembly. The refined 1.21 Gb genome, with an impressive contig N50 of 11.45 Mb, offers a valuable genomic resource. The genome exhibits 33,662 protein-coding genes, enriched in various pathways related to development and environmental adaptation. Gene family analysis shows expansion in development-related pathways and contraction in metabolic pathways, indicating niche adaptations. Notably, investigation into Hox gene regulation sheds light on their role in pleopod development, with the *Abd-A* gene identified as a linchpin. Post-transcriptional regulation involving novel-miR1317 negatively regulates *Abd-A* levels. Furthermore, the potential role of *fru* gene in ovarian development and the identification of novel-miR35 as a regulator of *Spfru2* add complexity to gene regulatory networks. Comparative functional analysis across Decapoda species reveals neo-functionalization of the *elovl6* gene in the synthesis of long-chain polyunsaturated fatty acids (LC-PUFA), suggesting its importance in environmental adaptation.

Conclusions Our findings shed light on various aspects of crab biology, including genome sequencing, assembly, and annotation, as well as gene family expansion, contraction, and regulatory mechanisms governing crucial developmental processes such as metamorphosis, reproductive strategies, and fatty acid metabolism.

Keywords Whole genome, Crab, Early development, Ovary maturation, LC-PUFA

†Yin Zhang, Ye Yuan and Mengqian Zhang contributed equally to this work.

*Correspondence:
Hongyu Ma
mahy@stu.edu.cn
Full list of author information is available at the end of the article

Background

Evolutionary adaptation results in diverse modifications including changes to body plans, development, and growth patterns [1]. These adaptations are evident among arthropods, particularly Decapoda crustaceans, which employ diverse adaptive strategies and include several economically valuable species (e.g., shrimp, lobsters, and crabs). True crabs (Brachyura) can be traced back to the early mid-Jurassic period [2, 3] and have achieved high levels of diversity and ecological functions, ranging across nearly all marine ecosystems (e.g., shallow coral reefs to hydrothermal vents), as well as freshwater and terrestrial types. The success of brachyuran crabs and their ecological diversification suggests that acquiring a crab-like morphology may have served as an important innovation and, thus, they are an ideal group for studying evolutionary diversification [4, 5].

The majority of animals have complex life cycles that include larval and adult stages with distinct morphologies and ecological niches. Crustaceans have the most diverse body plan among arthropods [6]. In Decapoda crabs, larvae undergo metamorphosis from a laterally compressed zoea to a dorsoventrally compressed megalopa with crab-like features [5], recapitulating the morphological evolution of Decapoda. However, the evolutionary and genetic mechanisms underpinning crab metamorphosis remain inadequately understood [7–9], despite indications of a correlation between body form and ecology [10]. At the same time, understanding the genetic mechanisms of reproductive strategies is important for elucidating evolutionary history [11]. While mammals have complex reproductive behaviors and sophisticated physiological mechanisms [12], in fish, gametes mature in the gonad before mating occurs [13]. While shrimp mate after gonadal maturation [14], crabs mate after females undergo reproductive molting when their ovaries are immature, thus triggering ovary maturation.

In the domain of life history, fatty acids, specifically long-chain polyunsaturated fatty acids (LC-PUFA), fulfill several essential functions including the reproduction, growth, and life history [15–17]. The allocation of capacities entails a balance between acquiring diverse fatty acids from diet and synthesizing LC-PUFA through multiple pathways regulated by fatty acid desaturase (Fads) and elongase (Elovl) [18, 19]. While it was previously thought that marine animals had limited ability to synthesize LC-PUFA due to the high availability of these fatty acids in marine ecosystems [20], more recent evidence suggests that euryhaline herbivorous fish and some invertebrates can produce LC-PUFA endogenously [21, 22]. Interestingly, the genes encoding different types of Elovl are more diverse in invertebrates than in vertebrates [20, 23, 24], although this diversity remains poorly understood.

Genome sequencing provides crucial information for understanding organism evolution and identifying genes involved in adaptation [14, 25, 26]. Despite the availability of several genome maps from some decapods (e.g., shrimps and crabs) [14, 27–32], the current genomic data for this group are insufficient to facilitate comprehensive studies on development, reproduction, and nutrient acquisition within this ancient, successful lineage. *Scylla paramamosain*, commonly known as green mud crab, is highly valued not only for its economic importance but also for its role in coastal ecosystems where it contributes to the balance of the aquatic food web. Understanding the molecular mechanisms underlying its development, reproduction, and nutrient acquisition is crucial for both evolutionary adaptations and ecological studies. However, previous genomic studies often suffer from limitations in genome completeness and resolution. Here, we present a comprehensive analysis of the genome of *S. paramamosain*, providing a robust model system to investigate the molecular foundations of adaptive evolution within Decapoda. We specifically ask (1) What are the genomic features that contribute to the adaptability of *S. paramamosain* to diverse environments? (2) How do specific genes and their regulatory mechanisms contribute to developmental processes and environmental adaptation? In doing so, we aim to provide a comprehensive understanding of the genetic and molecular basis of adaptive evolution in this species. The insights gained from this research offer a valuable resource for comparative and functional genomic studies in crustaceans, with broader implications for evolutionary biology and environmental adaptation research.

Results

Genome sequencing, assembly, and annotation

We obtained a total of 98.60 Gb pass reads using Nanopore ultralong-read sequencing. The *S. paramamosain* (Fig. 1A) genome was 1.21 Gb with a contig N50 of 11.45 Mb. In total, we obtained 421 contigs with an average length of 11,740,790 bp, which we assembled into 233 scaffolds with a scaffold N50 of 23.61 Mb (Table 1, Additional file 1: Table S1). The N50 of the sequences we obtained is approximately five times that achieved in the previous genome assembly of the species (N50 < 200 kb) [30]. Overall, we identified 95.26% of the 1013 Arthropoda Benchmarking Universal Single-Copy Orthologs (BUSCOs) in the assembled genome by searching against arthropoda_odb_10 gene set. The contigs were further anchored to 49 pseudochromosomes with a total length of 1,106,215,978 bp, representing 99.47% of the

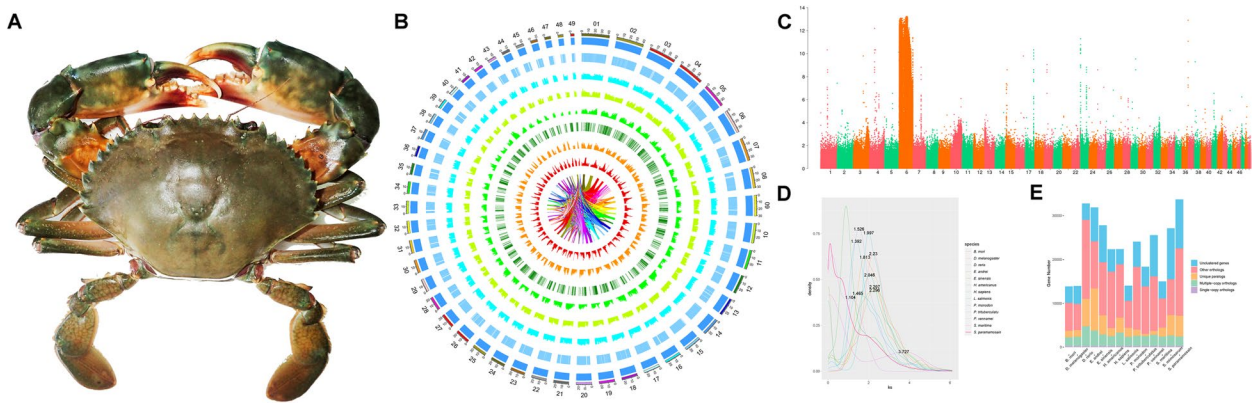

Fig. 1 Schematic representation of the genomic characteristics of *Scylla paramamosain*. **A** Image of *S. paramamosain*. **B** Schematic representation of the genomic characteristics of *S. paramamosain*. From the outer to inner, Track 1: the 49 chromosomes of *S. paramamosain*; Track 2: scaffolds anchored to each chromosome; Track 3: protein-coding genes; Track 4: distribution of gene density across the genome; Track 5: distribution of GC content in the genome; Track 6: distribution of five significantly expanded gene families in the genome; Track 7: distribution of SSRs in the genome; Track 8: distribution of transposable elements in the genome; Track 9: schematic presentation of major interchromosomal relationships in *S. paramamosain*. **C** Sex-linked region in the chromosomes of *S. paramamosain*. **D** Whole genome duplicates event in 12 species including *S. paramamosain*, *Portunus trituberculatus*, *Eriocheir sinensis*, *Penaeus vannamei*, *Homarus americanus*, *Penaeus monodon*, *Drosophila melanogaster*, *Bombyx mori*, *Eisenia andrei*, *Stegodyphus mimosarum*, *Strigamia maritima*, and *Danio rerio*. **E** Single-copy orthologs are defined as orthologs that were present as a single-copy gene in all 14 species (*S. paramamosain*, *P. trituberculatus*, *E. sinensis*, *P. vannamei*, *H. americanus*, *P. monodon*, *D. melanogaster*, *B. mori*, *E. andrei*, *Lepeophtheirus salmonis*, *S. mimosarum*, *S. maritima*, *D. rerio*, and *Homo sapiens*). Multiple-copy orthologs represent the gene groups present in all species with gene numbers > 1. Species-specific paralogs represent genes uniquely present in only one species. Other types of orthologs represent the gene groups that are absent in some species and not species-specific paralogs.

Table 1 Summary of *Scylla paramamosain* genome assembly

Stat type	Contig length (bp)	Contig number	Scaffold length (bp)	Scaffold number	Gap length (bp)	Gap number
N50	11,448,802	37	23,607,973	20	100	113
N60	9,208,292	49	21,816,234	25	100	135
N70	6,865,899	64	19,336,055	31	100	158
N80	3,900,000	87	15,570,585	38	100	180
N90	1,000,000	151	9,316,856	47	100	203
Longest	30,957,417	1	42,181,690	1	100	225
Total	1,210,737,920	458	1,210,760,420	233	22,500	225
Length ≥ 1 kb	1,210,737,920	458	1,210,760,420	233	0	0
Length ≥ 2 kb	1,210,737,920	458	1,210,760,420	233	0	0
Length ≥ 5 kb	1,210,737,920	458	1,210,760,420	233	0	0

genome (Fig. 1B). Chromosome 01 (chr01) is the largest, with a length of 42.18 Mb, whereas chr49 has the shortest length (5.67 Mb). In addition, chr06 has a strongly sex-linked signal using genome-wide association studies (Fig. 1C and Additional file 1: Fig. S1). Previously, we identified 13 sex-specific SNPs [33] located on chr06 in *S. paramamosain*, indicating that chr06 is the sex chromosome of this species. We observed a peak of Ks ranging from 0 to 1 and a summit at 0.2, representing a potential ancient whole-genome duplication (Fig. 1D).

Using ultralong-read sequencing, we detected 587.19 Mb of repetitive sequences, accounting for 58.97% of the genome. The genome contains 206,695 simple sequence repeats (SSRs) which are a specific type of tandem repeat characterized by short, repeating units, representing 0.25% of the total genome. Dinucleotide repeats were the dominant type of SSRs, accounting for 58.85% of the total SSRs. Moreover, there were 345,610 tandem repeats with a length of 24,183,525 bp, representing 2% of the whole genome. As a predominant force driving genome expansion and evolution [34, 35], transposable elements make up 48.50% of the *S. paramamosain* genome and contain 17.40% long interspersed elements (LINEs), 8.22% long terminal repeats (LTRs), 22.05% DNA transposons, 0.33%

rolling circle (RC), and 0.35% miniature inverted-repeat transposable elements (MITEs) (Additional file 1: Fig. S2 and Additional file 1: Table S2).

In *S. paramamosain*, there are 33,662 protein-coding genes with a predicted average gene length of 12,535.06 bp, CDS length of 1208.52 bp, exon length of 225.29 bp, and intron length of 2595.33 bp (Additional file 1: Table S4 and Additional file 1: Fig. S3). In total, 92.31% of the genes were annotated to public databases, including SwissProt (14,738 genes), InterPro (28,920 genes), Kyoto Encyclopedia of Genes and Genomes (KEGG; 11,128 genes), Gene Ontology (GO; 12,328 genes), EuKaryotic Orthologous Groups (KOG; 12,381 genes), and Non-Redundant Protein Sequence Database (Nr, 20,927 genes) (Additional file 1: Fig. S4). To explore the evolutionary forces shaping the genome of *S. paramamosain*, we conducted a selection pressure analysis by calculating the Ka/Ks ratio for protein-coding genes. Our analysis identified eight genes under positive selection (Additional file 1: Table S4), which are likely to play crucial roles in environmental adaptation. These genes are involved in key biological processes such as stress response and metabolism. We also annotated noncoding RNAs, including 449 miRNAs, 1055 rRNAs, 36 snRNAs, and 3153 tRNAs (Additional file 1: Table S5).

Gene family expansion, contraction, and genome evolution

Most expansion events are associated with the adaptive evolution of specific phenotypes. To evaluate potential adaptive evolution via gene family expansion and contraction events in *S. paramamosain*, we used gene family clustering analysis on 14 species, including 5 crustaceans (*Portunus trituberculatus, Eriocheir sinensis, Penaeus vannamei, Homarus americanus*, and *Penaeus monodon*), 6 insects (*Drosophila melanogaster, Bombyx mori, Eisenia andrei, Lepeophtheirus salmonis, Stegodyphus mimosarum*, and *Strigamia maritima*), and 2 vertebrates (*Danio rerio* and *Homo sapiens*). *S. paramamosain* had 297 single-copy genes, 4775 unique genes, and 11,097 unclustered genes. Furthermore, the unclustered and unique genes combined to yield a total of 15,872 species-specific genes (Fig. 1E and Additional file 1: Fig. S5). These species-specific genes were significantly enriched in aminoacyl-tRNA biosynthesis, the TNF signaling pathway, RNA degradation, ubiquitin-mediated proteolysis, the longevity regulating pathway, circadian entrainment, autophagy–animal, neurotrophin signaling pathway, adrenergic signaling in cardiomyocytes, MAPK signaling pathway, NF-kappa B signaling pathway, mTOR signaling pathway, and c-type lectin receptor signaling pathway (Additional file 1: Fig. S6). Most of these pathways are related to the development and lifestyle of *S. paramamosain*. Phylogenetic analysis based on 297 single-copy orthologous genes suggested that crabs (represented by *S. paramamosain*, *E. sinensis*, and *P. trituberculatus*) diverged from shrimp (represented by *L. vannamei*) ~ 481.54 million years ago (Mya), that marine crabs diverged from freshwater crabs (represented by *E. sinensis*) ~ 280.48 Mya, and that *S. paramamosain* diverged from *P. trituberculatus* ~ 141.1 Mya (Additional file 1: Fig. S7).

We identified a total of 12,274 gene families through family clustering, with 725 specific to *S. paramamosain*. Of the 33,662 genes, 22,565 could be classified into distinct gene families, with an average of 1.84 genes per family. As variations in gene copy number might support adaptive evolution, we examined the expansion and contraction of gene families in the *S. paramamosain* genome to explore the potential mechanisms underlying their adaptability. There were 1545 expanded gene families and 2671 contracted families in *S. paramamosain* (Fig. 2). Among them, the expanded genes were significantly enriched in development-related pathways. The

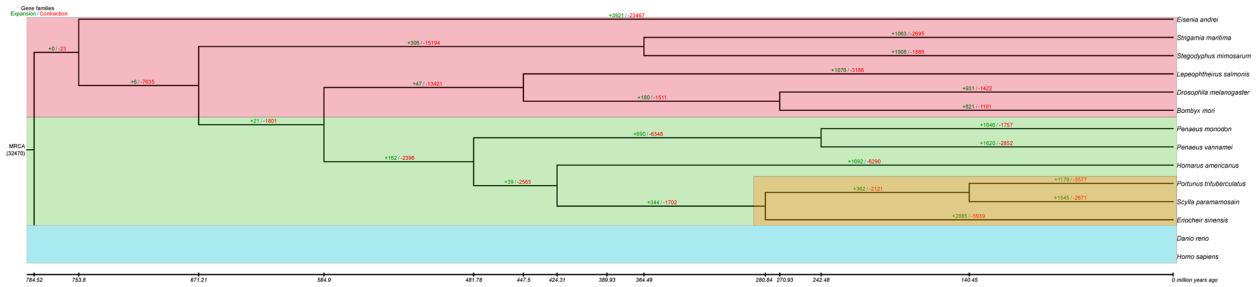

Fig. 2 Phylogenetic tree and gene family expansion/contraction of *Scylla paramamosain*. Phylogenetic tree inferred from 338 single-copy orthologs among 14 selected species (*S. paramamosain, Portunus trituberculatus, Eriocheir sinensis, Penaeus vannamei, Homarus americanus, Penaeus monodon, Drosophila melanogaster, Bombyx mori, Eisenia andrei, Lepeophtheirus salmonis, Stegodyphus mimosarum, Strigamia maritima, Danio rerio,* and *Homo sapiens*). The numbers of expanded gene families are indicated by the symbol "+" and the numbers of contracted gene families are indicated by the symbol "–". The number below the MRCA (most recent common ancestor) represents the total number of orthologs

expanded gene families in *S. paramamosain* were also significantly associated with these pathways (Additional file 1: Table S6) compared to shrimp and other crabs, suggesting that the specific developmental pattern of *S. paramamosain* may be attributed to the expansion of these gene families. On the other hand, the contracted gene families were mainly annotated to those associated with nutritional metabolism (Additional file 1: Fig. S8), indicating that there have been alterations in nutrient intake during evolution.

The function and regulatory mechanisms of Hox genes in the development of pleopods

Hox genes encode homeodomain-containing transcription factors that help direct tissue differentiation and morphological development [36]. We found that the *S. paramamosain* genome contained all ten canonical Hox gene clusters (*lab, pb, Hox3, Dfd, Scr, ftz, Antp, Ubx, Abd-A*, and *Abd-B*) (Fig. 3 A, B) found in the arthropod ancestor [37, 38]. The Hox gene clusters were located on the same chromosome (Fig. 3A), similar to the spatial collinearity in most bilaterians [39]. However, because of low-quality sequencing data, most Hox gene sequences contain breaks that prevent the verification of the transcribed direction of the Hox gene cluster [40]. The collinearity location of the Hox gene cluster in the same chromosome strongly supports the high integrity of our genome assembly of *S. paramamosain*.

During the early stages of development in *S. paramamosain*, we found that *SpUbx*, *SpAntp,* and *SpAbd-A* were upregulated during zoea stages III and V (Fig. 3C), suggesting their potential involvement in regulating the zoeal development process. Additionally, there was a significant decrease in the expression levels of both *SpUbx* and *SpAntp* within the abdomen of zoea stage III to V (Fig. 3D) but an increase in *SpAbd-A* (Fig. 4A). Given that zoea stages III to V are critical periods for abdominal and maxillary development (Fig. 3E), we speculate that these three genes primarily function on the abdomen and maxillae of zoea larvae. Furthermore, there was an increase followed by a decrease in *Sp*Abd-A protein expressions in the abdomen of zoea I to zoea V, with a peak observed in zoea stage III (Fig. 4B). The *Sp*Abd-A protein was primarily expressed in maxillopods, pereiopods, and abdominal segments in the zoea (Fig. 4C). Interference with the expression level of *SpAbd-A* during zoea stage IV resulted in the absence of pleopods in larvae (Fig. 4D).

Post-transcriptional regulation also fine-tunes *Abd-A* levels, wherein microRNAs such as miR-10 target *Abd-A* transcripts for degradation [41]. We found that RNAhybrid: novel-miR1317 was a potential target miRNA of *SpAbd-A* through the transcriptome of *S. paramamosain* larvae (Fig. 4E). Furthermore, we found a negative relationship between the expression level of novel-miR1317 and *SpAbd-A* (Fig. 4F), verifying that novel-miR1317 negatively regulated the expression of the *Abd-A* gene, possibly facilitating the development of pleopods. In most insects, miR-iab-4 has a negative regulatory effect on *Abd-A* and *Ubx* [42], while miR-iab-8 regulates *Abd-A* and *Abd-B* [43]. Herein, we find a significant role of

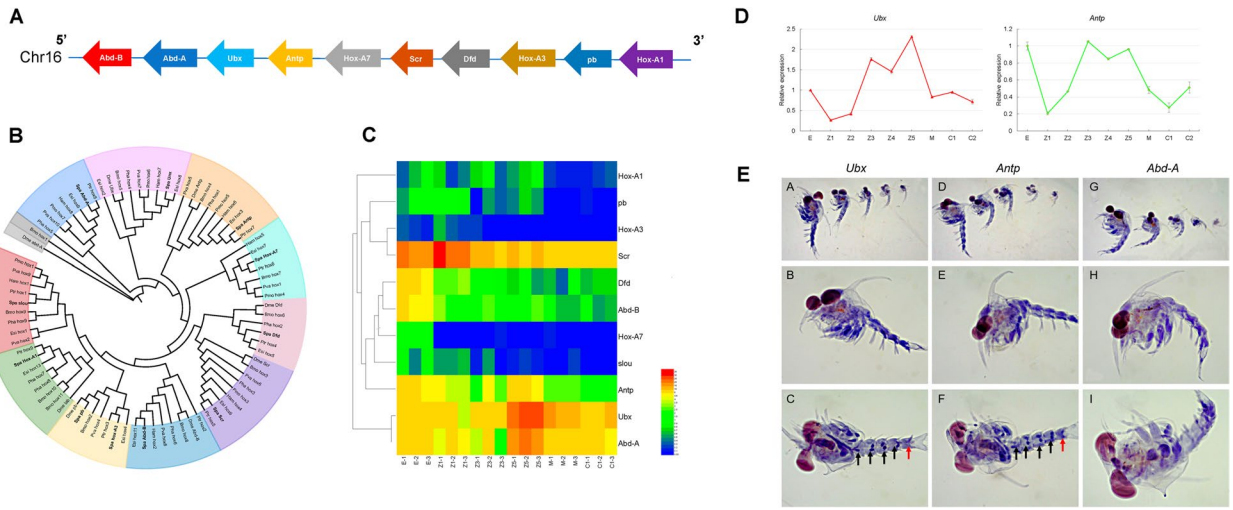

Fig. 3 The conserved hox gene clusters and their expressions in *Scylla paramamosain*. **A** Location and transcript direction in the genome of *S. paramamosain* Hox genes. **B** Distribution and relationships of Hox gene cluster in *S. paramamosain*, *Bombyx mori* (Bmo), *Drosophila melanogaster* (Dme), *Eriocheir sinensis* (Esi), *Homarus americanus* (Ham), *Parhyale hawaiensis* (Pha), *Penaeus monodon* (Pmo), *Penaeus vannamei* (Pva), and *Portunus trituberculatus* (Ptr). **C** Temporal expression of *S. paramamosain* Hox cluster genes. E embryo, Z1 zoea I, Z3 zoea III, Z5 zoea V, M megalopa stage, and C1 crablet I. **D** Expressions of *Ubx* and *Antp* in early different developmental stages of *S. paramamosain*. **E** The expression locations of *Ubx, Antp* and *Abd-A* in *S. paramamosain* larvae

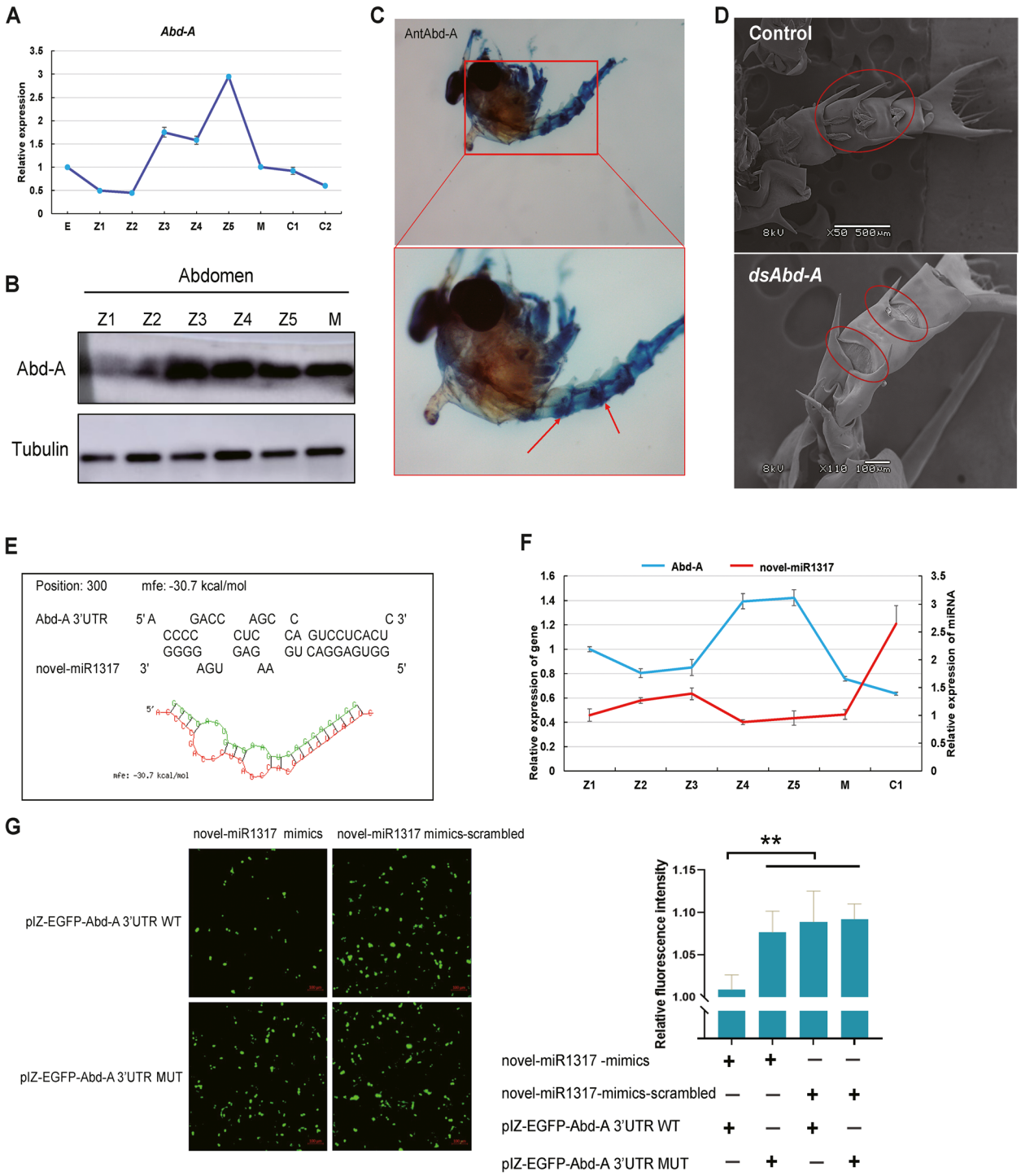

Fig. 4 The function and regulation of *Abd-A* gene in *Scylla paramamosain* larvae. **A** Expressions of *Abd-A* in different stages of *S. paramamosain* larvae. **B** Expression of *Sp*Abd-A protein in the abdomen of *S. paramamosain* larvae. **C** Expression locations of *Sp*Abd-A protein in *S. paramamosain* larvae. **D** Phenotype changes of *S. paramamosain* at zoea stage V after RNAi at zoea stage IV. **E** Predicted binding sites of *Abd-A* and novel-miR1317. **F** Negative expression patterns between novel-miR1317 and *Abd-A* in *S. paramamosain* larvae. **G** The S2 cells were co-transfected with WT *Abd-A* 3´-UTR, and the mutated-type of *Abd-A* 3´-UTR (MT), together with novel-miR1317 mimics or negative control mimics. The data were expressed as the relative fluorescence intensity. Z1-5 zoea I-V, M megalopa stage, C1 crablet I

SpAbd-A in the development of *S. paramamosain* larval abdominal pleopods, with its expression is negatively regulated by novel-miR1317 (Fig. 4G). These findings provide a foundation for further understanding the mechanisms of pleopod development of *S. paramamosain* larvae.

Gene regulation of ovarian development

Reproductive molting is crucial for the reproductive cycle in crustaceans. In other arthropods (*D. melanogaster*), the *fruitless* (*fru*) gene plays a role in the sex determination pathway [44, 45], while we obtained the transcript fragments of *fruitles2* (*Spfru2*) according to the gonadal transcriptome data of *S. paramamosain* [46]. We found higher expression levels of *Spfru2* in the gonad, thoracic ganglion, gill, and muscle in females compared to males (Additional file 1: Fig. S9A). Moreover, the expression level of *Spfru2* was the highest in ovarian stage III (Fig. 5A), mainly localized in the cytoplasm of oocytes; in males, it was detected mainly in the epithelia of seminiferous tubules and spermatids (Fig. 5B). This implies that *Spfru2* may function in ovarian maturation in *S. paramamosain*.

The expression level of *Spfru2* was highest on day 3 post mating (Additional file 1: Fig. S9B). After the reproductive molt, females start ovarian development, including oogenesis [47], a metabolically demanding reproductive process. The in vitro tissue culture experiments demonstrated a significant decrease in *Spfru2* expression at 8 h (Additional file 1: Fig. S9C), indicating the effectiveness of RNA interference. The *fru2* gene interference exhibited incomplete and fragmented egg cell morphology at 8 h (Fig. 5C). These findings suggest that the *fru2* gene also plays an important role in maintaining egg cell morphology, thereby influencing female gonadal development and maturation.

We found opposite expression patterns of novel-miR35 and *Spfru2* in different gonad development stages of females and males (Additional file 1: Fig. S9). Within 48 h after injection of miRNA mimics/inhibitor, the expression levels of novel-miR35 and *Spfru2* had opposite variant trends in the gonads (Fig. 5D). We further validated

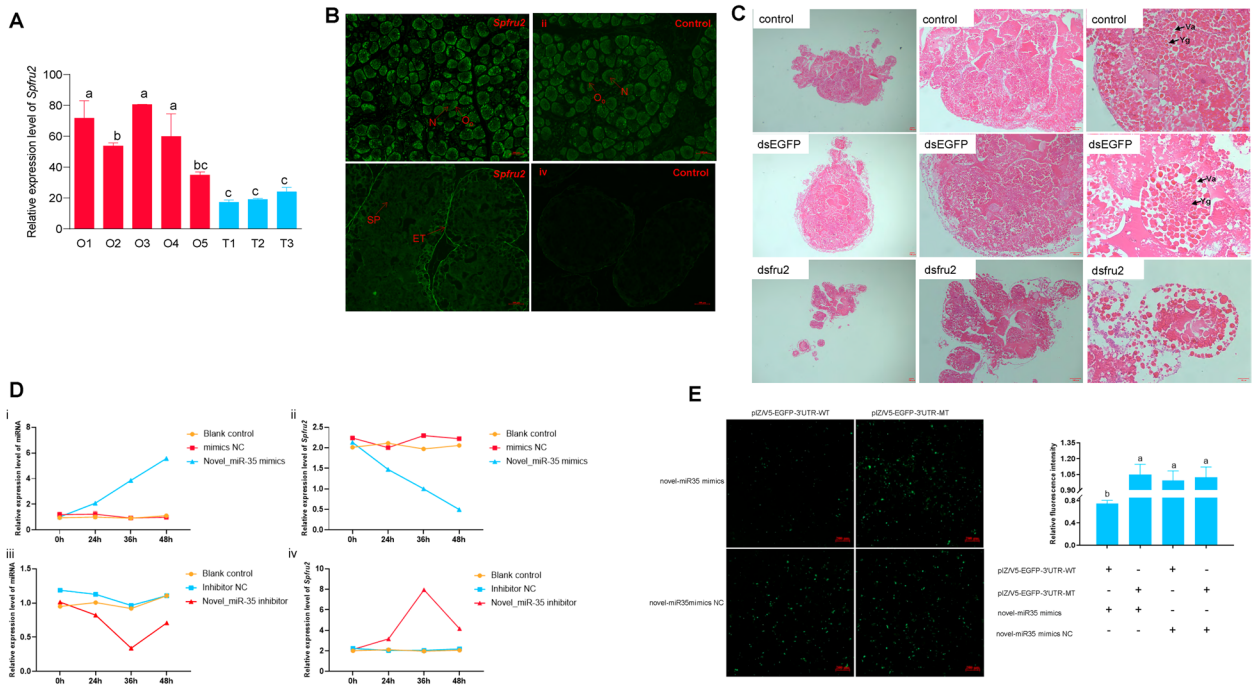

Fig. 5 The expression and regulation of *fru2* gene in gonads of *Scylla paramamosain*. **A** Relative expression of *Spfru2* in different gonad developmental stages. O ovary, T testis, O1–O5 stage OI to stage OV, T1–T3 stage TI to stage TIII. Bars with different lowercase letters indicate significant differences ($p < 0.05$). **B** Distribution of *Spfru2* in ovary (i, ii) and testis (iii, iv). N nucleus, Oo oocyte, ET epithelia, SP spermatid. Scale bars: 100 μm. **C** Phenotypic change of the ovary tissue after the interference of *fru2* expression. Yg yolk granule, Va vacuole. **D** Relative expression of novel-miR35 and *Spfru2* of the ovary after overexpression and silencing of novel-miR35. i Relative expression of novel-miR35 in the ovary after overexpression of novel-miR35. ii Relative expression of *Spfru2* in the ovary after overexpression of novel-miR35. iii Relative expression of novel-miR35 in the ovary after silencing of novel-miR35. iv Relative expression of *Spfru2* in the ovary after silencing of novel-miR35. Bars with different lowercase letters indicate significant difference ($p < 0.05$). **E** The S2 cells were co-transfected with WT *fru2* 3′-UTR, and the mutated-type of *fru2* 3′-UTR (MT), together with novel-miR35 mimics or negative control mimics. The data were expressed as the relative fluorescence intensity

the interaction between novel-miR35 and *Spfru2* using in vitro green fluorescent reporter assays. Both the novel-miR35 mimics and pre-novel-miR35 plasmid effectively reduced fluorescence intensity when co-transfected with the WT 3′-UTR reporter plasmid into *Drosophila* S2 cells. However, this effect was largely restored for the co-transfected plasmid containing the mutant type (MT) 3′-UTR (Fig. 5E, 5). These findings suggest that *Spfru2* might be a target of novel-miR35 in *S. paramamosain*.

Neo-functionalization of the *elovl6* gene in the LC-PUFA synthesis pathway

The LC-PUFA are defined as fatty acids with 20-24 carbon atoms and ≥ 2 double bonds. Further, depending upon the position of the last double bond relative to the methyl end, they are categorized as n-6 or n-3 series LC-PUFA, providing critical nutrients for animal reproduction and metamorphosis [22]. Among them, several n-3 LC-PUFA, especially eicosapentaenoic acid (EPA, 20:5n-3) and docosahexaenoic acid (DHA, 22:6n-3), contribute to an organism's health via multiple pathways that play roles in cell membrane lipid functions, neurodevelopment, and immune responses, as well as mitigating several diseases [22]. Generally, LC-PUFA can be enriched in body tissues either by dietary acquisition or, in some animal species, through endogenous biosynthesis [22]. The fatty acid elongase, Elovl6, is an enzyme involved in the endogenous synthesis of fatty acids, participates in fatty acid metabolism, energy balance, insulin resistance, and metabolic diseases [48, 49]. To illuminate the neo-functionalization of the *elovl6* gene and its involvement in LC-PUFA biosynthesis in *S. paramamosain*, we first analyzed the molecular characteristics of the *Spelovl6* gene. The variable promoter of the *elovl6* gene in *S. paramamosain* resulted in the generation of three transcripts (*elovl6a*, *elovl6b*, and *elovl6c*) through splicing (Fig. 6A). These exhibited the typical structural features of the Elovl protein family (Fig. 6B). The expression levels of the *elovl6* genes increased during embryo development, reaching their highest levels in the prehatching period, and then decreasing (Additional file 1: Fig. S10 A, B, C). The expression levels of *elovl6* genes in the larval stages were generally stable but declined in the megalopa stage (Additional file 1: Fig. S10 D, E, F).

We characterized *S. paramamosain* Elovl6a, Elovl6b, and Elovl6c functionally using a yeast heterologous expression system (Additional file 1: Table S6). Upon addition of 16:0 fatty acid substrate, the Elovl6-transformed yeast exhibited the production of saturated fatty acids (SFA) ranging from C_{18} to C_{22}. Similarly, incubation with 18:1n-9 resulted in the generation of several elongation products of monounsaturated fatty acids (MUFA). Endogenous 18:1n-9 was elongated to form 20:1n-9 and then 22:1n-9. We found that all C_{18} PUFA substrates were converted to their respective C_{20} products, with elongation observed for 18:3n-3, 18:4n-3, 18:2n-6, and 18:3n-6 to their corresponding C_{20} PUFA. Notably, *S. paramamosain* Elovl6 had a higher affinity toward n-3 PUFA than n-6 PUFA when utilizing C_{18} PUFA substrates. However, neither of the C_{20} LC-PUFA, arachidonic acid (ARA; 20:4n-6), or eicosapentaenoic acid (EPA; 20:5n-3), elongate. We also observed a significant decrease in the expression of the *elovl6* gene following 18 h and 36 h interference in isolated hepatopancreas of *S. paramamosain* (Additional file 1: Fig. S10G), resulting in a reduction in 16:0, 16:1n-7, 18:1n-9, 18:2n-6, 18:3n-3, 20:5n-3, and 22:6n-3. This validated the capability of Elovl6 for elongation of C_{18} PUFA to C_{20} PUFA (Additional file 1: Fig. S10H). Thus, we confirmed the capacity of Elovl6 in the elongation of C_{18} PUFA, in addition to SFA and MUFA. In contrast, Elovl6 proteins in mammals [48, 50] and teleosts [51, 52] are only involved in the elongation of SFA and MUFA.

The above results suggest that Elovl6 may undergo neo-functionalization in *S. paramamosain*. Indeed, decapod crabs inhabit a diverse range of environments, including seawater (e.g., *P. trituberculatus*), freshwater (e.g., *E. sinensis*), and brackish water (e.g., *S. paramamosain*). Given the observation that the multi-functionalization and neo-functionalization of duplicated genes has occurred in fish that vary in their use of fresh and saltwater [53, 54], we speculated this might also be true for *S. paramamosain*. To investigate the influence of the environment on the evolutionary adaptation of key enzymes of LC-PUFA synthesis in Decapod crabs, we functionally characterized three transcripts of the *E. sinensis and P. trituberculatus elovl6* genes. We demonstrated that the *Ptelovl6* and *Eselovl6* genes exhibited comparable fatty acid elongation activities to that of *Spelovl6* (Fig. 6C), implying the potential neo-functionalization of Elovl6. Furthermore, the phylogenetic analysis of *elovl6* in *E. sinensis*, *P. trituberculatus*, and *S. paramamosain* showed the evolutionary difference between *elovl6* and *elovl2/4/5/8* (Fig. 6D), which indicated that neo-functionalization of the *elovl6* gene in the LC-PUFA synthesis pathway occurred in these crustaceans.

Discussion

Here, we present a comprehensive analysis of the *S. paramamosain* genome and achieved contig and scaffold N50 values that substantially surpass previous assemblies [14, 27–32]. By anchoring contigs to 49 pseudochromosomes, we attained a more precise depiction of the genome architecture, with Chr06 possessing a prominent sex-linked signal, offering important insights into sex determination mechanisms. Furthermore, our detection of a potential ancient whole-genome duplication event

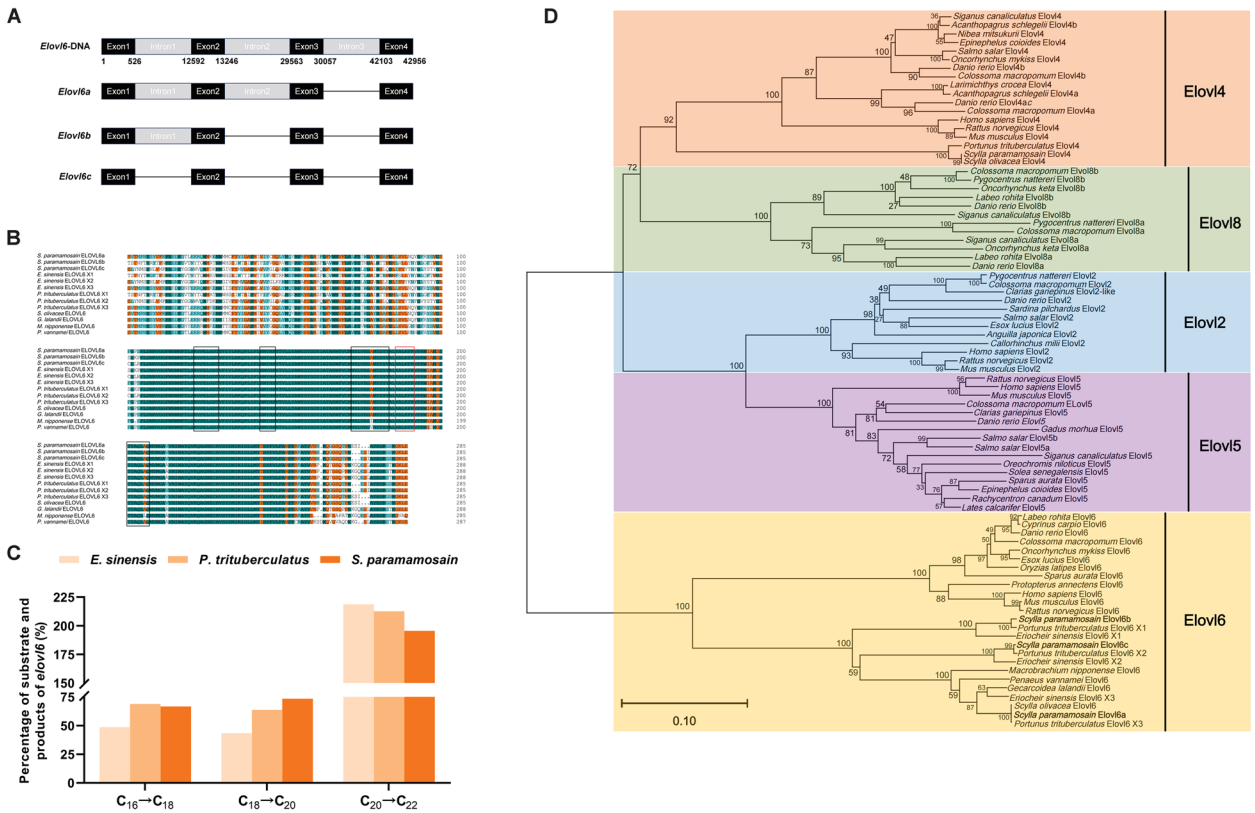

Fig. 6 Neo-functionalization of the *elovl6* gene in the LC-PUFA synthesis pathway in crustacean species. **A** Schematic diagram of the alternative splicing in *elovl6*. *elovl6*-DNA, the DNA sequence of *elovl6*; *elovl6a* contains the first and second introns; *elovl6b* contains the first intron; *elovl6c*, the *elovl6* transcript of introns normal splicing. Black boxes represent exons. Gray boxes represent introns. **B** Alignment of the amino acid sequences of Elovl6 from several crustacean species (*Scylla olivacea* (QEV88898.1), *Gecarcoidea lalandii* (QKG32709.1), *Macrobrachium nipponense* (ANM86278.1), and *Penaeus vannamei* (AKJ77890.1)). Identical amino acid residues were indicated by dark green, while similar residues were indicated by blue and orange. The conserved HXXHH histidine motif was in the box, and highly conserved motifs (KXXEXXDT, NXXXHXXMYSYY, KXXKXX, and TXXQXXQ) were framed in a black box. **C** Phylogenetic tree comparing *S. paramamosain* Elovl6 with elongase proteins from other organisms. The tree was constructed using the neighbour-joining method. The horizontal branch length is proportional to the amino acid substitution rate per site. The numbers represent the frequencies with which the tree topology was replicated after 1000 iterations. **D** Comparison of Elovl6 extended carbon chain conversion efficiency in *E. sinensis*, *P. trituberculatus*, and *S. paramamosain*

suggests a dynamic evolutionary history underlying the genomic organization of *S. paramamosain*. The differences in genome size between the present *S. paramamosain* genome and previous assemblies can be attributed to several factors. First, the current study employed advanced sequencing technologies, such as Nanopore ultra-long reads, HiFi sequencing, and Hi-C scaffolding, which provide higher accuracy and longer read lengths compared to earlier methods [30]. This advancement reduces assembly gaps and improves the resolution of repetitive regions, leading to a more complete genome assembly (Additional file 1: Table S1). Additionally, our assembly includes a higher proportion of repetitive sequences, such as transposable elements and simple sequence repeats, which was more complete assembled than the published genome [30]. The improved contiguity of the current assembly, characterized by a higher N50 value and fewer fragmented contigs, allows for the inclusion of previously fragmented or missing genomic regions. Moreover, the comprehensive annotation process employed in present study enhances the identification of additional coding and noncoding regions, contributing to the overall genome size.

The assembly of the genome can be affected by several factors, including individual heterozygosity, repetitive sequences, and GC bias [55, 56]. In addition, long reads can span complex or repetitive regions with a single continuous read, thus eliminating ambiguity and contributing to a more complete genome assembly [57]. Our investigation into genome assembly emphasized the utility of ultralong-read sequencing in resolving repetitive sequences, culminating in the identification of 58.97% of the genome as repetitive. This is much higher than that in *Eriocheir sinensis* [31] and *L. vannamei* [14], but lower

than that of other crustaceans for which genomic data are available [26, 30]. Repetitive sequences and TEs likely have an important contribution to genome expansion and evolution [55], with a higher proportion observed in *S. paramamosain* compared to several other crustacean species [14, 27–32]. The identification of numerous SSRs and TRs here underscores the importance of these elements in shaping the genomic landscape of *S. paramamosain*, which were also reflected by the complex crustacean genome [14, 31]. Thus, the analysis of SSRs, TRs, and TEs provided valuable resources for the analysis of genome expansion and evolution. *S. paramamosain* has 33,662 protein-coding genes, which is high compared to other crustaceans [14, 27–32, 58]. This number of coding genes was relatively higher than other crustaceans (*E. sinensis* [31] and *Procambarus virginalis* [26]). The average CDS length of *S. paramamosain* protein-coding genes was similar to that of *E. sinensis* [27] and *L. vannamei* [14], but longer than that of *P. trituberculatus* [29] (Additional file 1: Table S1). The number of genes annotated to various databases, including SwissProt, InterProt, KEGG, GO, KOG, and Nr, accounted for 92.31% of the total genes in *S. paramamosain*, which was significantly higher compared to other crustaceans (*E. sinensis* [31], *P. virginalis* [26], *L. vannamei* [14], and *P. trituberculatus* [29]). This comprehensive annotation greatly enhances our understanding of the genomics of *S. paramamosain*. Gene family analysis revealed species-specific genes enriched in pathways related to development, indicating potential adaptations. Phylogenetic analysis positioned *S. paramamosain* among crustaceans, highlighting their evolutionary divergence [3]. Namely, the divergence of marine crabs from freshwater crabs occurred approximately 280.48 Mya, while the divergence of *S. paramamosain* from *P. trituberculatus* occurred around 141.1 Mya. What is more, the high number of species-specific genes and the total gene count observed in *S. paramamosain* can be attributed to several factors associated with the high-quality of the present genome assembly and annotation. The use of advanced sequencing technologies, including Nanopore ultra-long reads and Hi-C scaffolding, has yielded a highly contiguous and complete genome assembly, characterized by an N50 value that markedly surpasses previous assembly [30]. This increased level of genome completeness has facilitated the accurate identification and annotation of a greater number of protein-coding genes (Additional file 1: Table S1), particularly those that are species-specific. Furthermore, the rigorous and comprehensive annotation process applied in this study, which employed multiple gene prediction models and leveraged extensive functional databases, has significantly enhanced gene identification. This approach, coupled with the high-quality genome assembly, has contributed to the discovery of a large number of genes unique to *S. paramamosain*, reflecting both the technical advancements and the thoroughness of the annotation process.

We found that the expression of the Hox genes did not show whole-cluster temporal collinearity (Fig. 3C) like vertebrates, but rather exhibited subcluster-level temporal collinearity, similar to other crustaceans (*E. sinensis* [31], *P. monodon* [46]), mollusks (*Patinopecten yessoensis* [59], *Crassostrea gigas* [55]), and tunicates (*Ciona intestinalis* [60]). In addition, we observed that *S. paramamosain* Hox genes played a role in directing tissue differentiation and morphological development during larval stages, particularly in the development of pleopods. The expression patterns of Hox genes during larval development suggest their involvement in the regulation of key developmental processes, with potential implications for understanding the evolution of crustacean body plans. For example, we investigated the regulatory mechanisms of Hox genes in pleopod development, demonstrating the crucial role of *Abd-A* in abdominal limb development. In crustaceans, *Abd-A* establishes segmental identity and boundaries [61], the loss of which been associated with a lack of an abdomen in the Cirripedia [49]. Because morphological changes are mainly driven by alterations in gene expression patterns, we can speculate that abdominal degeneration may be linked to low levels of *Abd-A* during brachyurization metamorphosis. Post-transcriptional regulation involving novel-miR1317 fine-tunes *Abd-A* levels, unveiling a nuanced layer of gene regulation.

Our exploration of the *fru* gene in ovarian development underscored its potential role in female gonadal development and maturation in *S. paramamosain*. We found that the expression of *fru2* was negatively regulated by novel-miR35, which adds complexity to the gene regulatory network governing reproduction. The divergent expression patterns of *fru* in male and female crabs, alongside its localization in gonadal tissues, highlight its potential involvement in ovarian maturation and gonadal development. Furthermore, the identification of novel microRNA targeting *fru* underscores the intricate post-transcriptional regulatory mechanisms governing reproductive processes in crustaceans.

The neo-functionalization of the *elovl6* gene in the LC-PUFA synthesis pathway is an important adaptation in *S. paramamosain*. Here, we elucidate the neo-functionalization of the *elovl6* gene in the synthesis of LC-PUFA, shedding light on its multiple transcripts and tissue-specific expression. Comparative analysis among *S. paramamosain*, *P. trituberculatus*, and *E. sinensis* suggests

evolutionary divergence in LC-PUFA biosynthesis pathways, potentially driven by environmental adaptation.

Conclusions

Overall, our findings shed light into several aspects of crab biology. These include genome sequencing, assembly, and annotation, as well as gene family expansion, contraction, and regulatory mechanisms governing crucial developmental processes such as metamorphosis, reproductive strategies, and fatty acid metabolism. We explore gene family expansions and contractions, particularly in pathways related to development and metabolism, indicating how these genomic features support the crab's ability to thrive in different niches. The study delves into the roles of Hox genes, such as *Abd-A*, in pleopod development and their regulation by novel-miR1317. Additionally, we investigate the *fru* gene's involvement in ovarian development and regulation by novel-miR35. We perform functional characterization of the *elovl6* gene, suggesting its importance in the synthesis of LC-PUFA and its potential neo-functionalization in LC-PUFA synthesis across Decapoda species. As such, our findings significantly advance our understanding of the *S. paramamosain* genome, providing a valuable resource for future studies in comparative genomics, evolutionary biology, and functional genomics in crustaceans. Future studies leveraging this genomic resource will further elucidate the genetic basis of key adaptive traits and inform conservation and management strategies for economically important crab species.

Methods

Sample preparation and genomic DNA isolation

A wild adult male *S. paramamosain* with stage III testis was collected off the coast of Shantou, China, and genomic DNA was extracted from its testis for sequencing. A Grandomics Genomic BAC-long DNA Kit was used to isolate ultralong DNA according to the manufacturer's guidelines. The total DNA quantity and quality were evaluated using a NanoDrop One UV–Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and a Qubit 3.0 Fluorometer (Invitrogen Life Technologies, Carlsbad, CA). Large DNA fragments were obtained through gel cutting with the Blue Pippin system (Sage Science, Beverly, MA). Qualification of DNA involved visual inspection for foreign matter, assessment of degradation and size via 0.75% agarose gel electrophoresis, a check on purity (OD260/280 between 1.8 and 2.0; OD260/230 between 2.0 and 2.2) using Nanodrop 2000, and precise quantification with a Qubit 3.0 Fluorometer (Invitrogen, USA).

Library construction and sequencing

Approximately 8–10 μg of genomic DNA (> 50 kb) was selected using the SageHLS HMW library system (Sage Science, Beverley MA, USA) and processed with the Ligation Sequencing 1D Kit (Oxford Nanopore Technologies, Shanghai, China) following the manufacturer's instructions. Library construction and sequencing were conducted on the PromethION (Oxford Nanopore Technologies) at the Genome Center of Grandomics (Wuhan, China). After quality inspection, large DNA fragments were recovered using the BluePippin automatic nucleic acid recovery instrument. Terminal repair, A-tailing, and ligation were performed using the LSK109 connection kit. Qubit was used to assess the constructed DNA library precisely. The library was then loaded into a flow cell, transferred to the PromethION sequencer, and subjected to real-time single-molecule sequencing. In the ONT sequencing platform, base calling and the conversion of nanopore-generated signals to base sequences [62] were executed using the Guppy toolkit (Oxford Nanopore Technologies). Pass reads with a mean qscore_template value ≥ 7 were obtained and used for subsequent assembly (https://github.com/nanoporetech/taiyaki).

Genome assembly, evaluation, and correction

A pure three-generation assembly was employed using filtered reads post quality control in NextDenovo software (reads_cutoff:1 k, seed_cutoff:28 k) (https://github.com/Nextomics/NextDenovo.git). The NextCorrect module was employed to correct the original data, yielding a consistency sequence after 13 Gb of error correction. De novo assembly using the NextGraph module produced a preliminary assembly of the genome. ONT three-generation data and Pb HiFi three-generation data (Additional file 1: Table S7) were then utilized with Nextpolish software (https://github.com/Nextomics/NextPolish.git) for genome correction. The corrected genome (Polish Genome) was obtained after three rounds of correction for both ONT and Pb HiFi data. Bwa mem default parameters were used to compare RNA sequencing data to the genome, and Pilon was iteratively calibrated three times to derive the final genome sequence. GC depth analysis and BUSCO prediction (https://busco.ezlab.org/) were used to assess genome quality and completeness.

Chromosome anchoring by Hi-C sequencing

To anchor hybrid scaffolds onto chromosomes, genomic DNA was extracted from crab testes for Hi-C library construction. The process involved cross-linking cells with formaldehyde, lysing cells, resuspending nuclei, and subsequent steps leading to proximity ligation. After

overnight ligation, cross-linking was reversed, and chromatin DNA manipulations were performed. DNA purification and shearing to 400 bp lengths were followed by Hi-C library preparation using the NEBNext Ultra II DNA Library Prep Kit for Illumina, which was then sequenced on the Illumina NovaSeq/MGI-2000 platform. Cell samples were fixed with formaldehyde and subjected to lysis and extraction for sample quality assessment. After passing the quality test, the Hi-C fragment preparation involved chromatin digestion, biotin labeling, end ligation, DNA purification, and library construction. The library was sequenced on the MGI-2000 platform, and data were processed to extract high-quality reads. The analysis included filtering for adapters, removing low-quality reads, and eliminating reads with an N content exceeding five. Reads were aligned using Bowtie2 [63], and contig clustering was performed using LACHESIS software [64].

Gene annotation

Tandem repeats were annotated using GMATA (https://sourceforge.net/projects/gmata/?source=navbar) and Tandem Repeats Finder (TRF) (http://tandem.bu.edu/trf/trf.html), identifying simple repeat sequences (SSRs) and all tandem repeat elements. Transposable elements (TEs) were identified through an ab initio and homology-based approach, with RepeatMasker (https://github.com/rmhubley/RepeatMasker) used for searching known and novel TEs. Gene prediction employed three methods: GeMoMa (http://www.jstacs.de/index.php/GeMoMa) for homolog prediction, PASA (https://github.com/PASApipeline/PASApipeline) for RNAseq-based prediction, and Augustus (https://github.com/Gaius-Augustus/Augustus) for de novo prediction. EVidenceModeler (EVM) (http://evidencemodeler.github.io/) integrated gene sets, which underwent further filtering for transposons and erroneous genes. Untranslated regions and alternative splicing regions were determined using PASA based on RNA-seq assemblies which were from our lab and downloaded from NCBI database using the keyword "*Scylla paramamosain*" (Additional file 1: Table S8). Functional annotation involved comparisons with several public databases, including SwissProt, NR, KEGG, KOG, and GO. InterProScan identified putative domains and GO terms. BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used against public protein databases to assess gene function. The prediction of noncoding RNA (ncRNA) entailed using tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/) and Infernal cmscan (http://eddylab.org/infernal/) for tRNAs and other noncoding RNAs. BUSCO was used to evaluate gene predictions, aligning annotated protein sequences to evolution-specific BUSCO databases.

Evolutionary analysis

To ascertain homologous relationships between *S. paramamosain* and other animal species, protein sequences were acquired and aligned using OrthMCL (https://orthomcl.org/orthomcl/). Initially, protein sets were gathered from 14 sequenced animal species, selecting the longest transcripts for each gene after excluding miscoded and prematurely terminated genes. Subsequently, pairwise alignment of these extracted protein sequences was conducted to identify conserved orthologs, employing Blastp with an E-value threshold of $\leq 1 \times 10^{-5}$. Further identification of orthologous intergenome gene pairs, paralogous intragenome gene pairs, and single-copy gene pairs was achieved using OrthMCL. Species-specific genes, including *S. paramamosain*-specific genes and unclustered genes, were extracted. Functional annotation and enrichment tests of species-specific genes were performed utilizing information from homologs in the online GO (http://www.geneontology.org/) and KEGG (https://www.genome.jp/kegg/) databases.

Building upon the orthologous gene sets identified with OrthMCL, molecular phylogenetic analysis was executed using shared single-copy genes. Coding sequences were extracted from single-copy families, followed by multiple alignments of each ortholog group using MAFFT (https://mafft.cbrc.jp/alignment/software/). To eliminate poorly aligned sequences, Gblocks were applied, and the GTRGAMMA substitution model of RAxML (https://cme.h-its.org/exelixis/web/software/raxml/hands_on.html) was employed to construct phylogenetic trees with 1000 bootstrap replicates. The resulting tree file was visualized using Figtree (http://tree.bio.ed.ac.uk/software/figtree/), after which the RelTime tool (https://www.megasoftware.net/) of MEGA-CC was used to compute mean substitution rates along each branch and estimate species divergence times. Three fossil calibration times obtained from the TimeTree (http://www.timetree.org/) database served as temporal controls, including the divergence times of *S. paramamosain*.

The detection of significant expansions or contractions in specific gene families, often indicative of adaptive divergence in closely related species, was carried out based on the OrthoMCL results. To examine this, a birth and death process to model gene gain and loss over a phylogeny was used (https://github.com/hahnlab/CAFE). Furthermore, the ratio of the nonsynonymous (Ka) to the synonymous (Ks) substitution rates of protein-coding genes was calculated in accordance with the neutral theory of molecular evolution [65]. The average Ka/Ks values were determined to identify positively selected genes within the *S. paramamosain* lineage with a branch-site likelihood ratio test using Codeml (http://abacus.gene.

ucl.ac.uk/software/) from the PAML package. Genes with a p value < 0.05 under the branch-site model were considered to be positively selected.

Whole-genome duplication events were investigated using fourfold synonymous third-codon transversion (4DTv) and synonymous substitution rate (Ks) estimation. Initial steps involved extracting protein sequences and conducting all-*vs*.-all paralog analysis through self-Blastp in these species. After filtering by identity and coverage, the Blastp results were analyzed with MCScanX [66], and the respective collinear blocks were identified. Subsequently, Ks and 4DTv were calculated for the syntenic block gene pairs using the KaKs-Calculator (https://sourceforge.net/projects/kakscalculator2/), and potential whole-genome duplication events were evaluated based on their Ks and 4DTv distribution.

Genome-wide association studies (GWAS) analysis

GWAS was conducted to identify genomic regions associated with sex traits in *S. paramamosain*. Genomic DNA was extracted from a population of 146 individuals collected from a crab culture farm in Niutianyang (Shantou, Guangdong, China) located at N23°22′23″, E116°42′16″. Based on whole-genome sequence data, a GWAS analysis of the sex trait was conducted using the PLINK software with logistic regression. Following completion of the GWAS analysis, significant SNP loci were identified by applying the Bonferroni correction method, considering loci with p-values less than 0.05 divided by the number of SNPs tested. The visualization of the results was performed using the qqman package in R to generate Manhattan and QQ plots. Finally, genomic structural annotation of important candidate SNP loci was carried out using SnpEff software, and gene function annotation was performed based on protein sequences from candidate genes using DIAMOND software with the Nr protein database.

Microinjection and RNA interference

To identify the function of the genes, RNA interference combined with microinjection were used. Based on the gene sequences, primers were designed for the amplification of the *Abd-A*, *fru2*, and *elovl6* gene fragments and subsequent dsRNA synthesis. Specific primers were designed incorporating protective bases and T7 promoter sequences at the 5' end of both the positive and negative primers to facilitate dsRNA synthesis. The plasmid containing the *Abd-A, fru2* and *elovl6* gene sequence fragments served as the template for PCR amplification, respectively, and the resulting product underwent purification. The PCR product equipped with the T7 promoter was used as the template for the synthesis of dsRNA, a crucial step for gene interference through in vitro transcription. Next, larvae (zoeal stage IV) of *S. paramamosain* were positioned on a custom-made agarose gel plate. Then, dsRNA-Abd-A and injection indicator mixture was injected between the cuticle and abdominal space of the larvae using a microinjector. The development of the abdomen and limbs was observed through electron microscopy as the larvae progressed to the zoeal V stage.

Hepatopancreas treatment in vitro

To detect the function of Elovl6 in synthesizing fatty acid, the hepatopancreas of *S. paramamosain* were cultured in vitro. Crabs were anesthetized on ice for 10 min, followed by sterilization in 75% ethanol for 5 min. Hepatopancreas tissues were subsequently dissected and infiltrated with phosphate-buffered saline (PBS) containing 1% penicillin and streptomycin for 30 min, then washed eight times for 5 min each time using the PBS solution as above. Next, hepatopancreas tissues were cut into approximately 20 mg fragments using scissors. The fragments were then precultured at room temperature (25 ℃) in a 24-well plate with 150 μL of Leibovitz L15 medium (containing 1% penicillin and streptomycin). The 24-well culture plates were supplemented with dsRNA-Elovl6 and dsRNA-EGFP (control group) and placed in a 28 ℃ incubator for cultivation. After 18 h, 24 h, and 36 h of culture, tissue fragments were collected from each treatment for total RNA extraction and subsequent qRT-PCR analysis, with corresponding parallel samples of hepatopancreas tissue also collected for fatty acid analysis.

Sequence and phylogenetic analysis

To provide clues for predicting functions or evolution of the *elovl6* genes, multiple sequence alignments were performed with the DNAMAN software (v6.0.3.99, Lynnon Corporation, USA). The phylogenetic tree was mapped using the neighbor-joining method with MEGA software (v11.0.13, Arizona State University, USA), based on deduced amino acid sequences of Elovls from *S. paramamosain*, as well as representative mammals, teleosts, and crustacean species downloaded from NCBI Genbank. Confidence in the resulting phylogenetic tree branch topology was measured by bootstrapping 1000 replicates.

Quantitative real-time PCR (qRT-PCR)

To detect the expressions of genes and miRNAs, qRT-PCR was used. Firstly, miRNAs were isolated from distinct developmental stages (embryo, zoea, megalopa, and crablet) and tissues (ovary I–V and testis I–III) of *S. paramamosain* using the miRcute miRNA Isolation Kit. Subsequently, the miRcute enhanced miRNA cDNA First Strand Synthesis Kit was deployed for the reverse

transcription process. Finally, the miRcute enhanced miRNA Fluorescence Quantitative Detection Kit was applied during the qPCR analysis to elucidate the expression dynamics of both miRNAs and the genes. The total RNA was extracted using the RNAiso Plus kit (Takara Co. Ltd., Japan). Preceding the quantitative real-time polymerase chain reaction (qRT-PCR), the RNA samples were treated with RQ1 RNase-Free DNase (Takara Co. Ltd.) to eradicate genomic DNA contamination. Next, cDNA synthesis was conducted using 500 ng of DNase-treated RNA and the Talent qPCR Premix (SYBR Green) kit (TIANGEN Biotech Co., Ltd., Beijing), in accordance with the manufacturer's instructions. The qPCR primers were designed using Primer 6.0 software (Additional file 1: Table S9), and the reactions were performed with a Mini Option real-time detector (Roche LightCycle@480). Each reaction mixture contained 10 μL of Talent qPCR Premix (2×), 0.6 μL of PCR forward primer (10 μM), 0.6 μL of PCR reverse primer (10 μM), 2.0 μL of RT reaction solution containing cDNA at an amount of 20 ng, and 6.8 μL of RNase-free water. The amplification protocol involved an initial denaturation step at 95 ℃ for 3 min, followed by 40 cycles of denaturation at 95 ℃ for 5 s, annealing at 60 ℃ for 10 s, and extension at 72 ℃ for 15 s. Next, there was a melting curve analysis by sustaining the reaction at 72 ℃ for 6 s, followed by a denaturation step at 95 ℃ for 5 s. All products were initially resolved through agarose gel electrophoresis to verify amplicon sizes. Gene expression levels were normalized to the reference gene (*18 s rRNA*) and calculated using the optimized comparative Ct method ($2^{-\Delta\Delta Ct}$).

Protein expression detection using western blotting

To detect the Abd-A protein expression in the cephalothorax and abdomen, western blotting was performed. Proteins were extracted from the cephalothorax and abdomen of *S. paramamosain* larvae using a cryogenic protein lysis solution following flash freezing with liquid nitrogen. Total protein was then extracted from each sample. After quantifying the protein concentration using the BCA protein assay kit (Sangon Biotech Co. Ltd., Shanghai, China), an equal number of proteins were loaded for SDS-PAGE electrophoresis. Following separation based on molecular weight, the proteins were transferred onto a PVDF membrane from the polyacrylamide gel and probed with specific antibodies. The next steps included incubation with HRP-conjugated secondary antibodies and detection using an ECL hypersensitive chemiluminescence kit (Sangon Biotech).

Fluorescence in situ hybridization (FISH)

To localize the expression of *fru2* gene in the gonads of *S. paramamosain*, fluorescence in situ hybridization (FISH) was utilized. The FISH analysis was performed on *S. paramamosain* gonads that were fixed in 4% paraformaldehyde prepared with DEPC-treated water 2 h prior to sectioning at a thickness of 10–12 μm using a cryostat set at − 20 to − 25 ℃. Then, no more than three sections per slide were thaw-mounted onto charged Superfrost Plus slides. The *fru2* probe, modified with FITC fluorescence at the 3′ end, was synthesized commercially (Sangon Biotech). The prehybridization and hybridization procedures followed the method described by Pinaud et al. [67]. Initially, a prehybridization solution diluted with high-grade formamide in a 1:1 ratio was added to each slide (50 μL per sample). Then, the slides were placed in a humid chamber for 30 min at room temperature before removing the coverslips in a 2×SSC solution. After prehybridization, the probe was added to the hybridization buffer at a concentration of 1 ng/μL. A 25-mL hybridization mix was applied to each slide before placing them in a metal slide holder immersed in a mineral oil bath maintained at 65 ℃ for 14–16 h. Following hybridization, the slide holder was removed from the mineral oil bath and slides were rinsed twice (30 s each) in chloroform and immersed in a fresh batch of 2×SSC solution and washed at room temperature for 1 h; they were then washed again in an SSC solution containing 50% formamide at 65 ℃ for 1 h; finally, they were washed twice (each time lasting 30 min) in tenfold diluted (0.1×) SSC solution while maintaining the temperature at 65 ℃. The signals resulting from hybridization on each slide were ultimately detected using fluorescence microscopy.

Yeast expression and fatty acid detection

To verify and compare the functions of *elovl6* genes in different species of crabs (*S. paramamosain*, *E. sinensis*, and *P. trituberculatus*), the yeast expression system was used. PCR amplification of open reading frames (ORFs) corresponding to *Eselovl6*, *Ptelovl6*, and *Spelovl6* was performed using a high-fidelity DNA polymerase, KOD-Plus-Neo (Toyobo, Japan), and cDNA templates. Specific primers that incorporated BamHI (GGATCC) and EcoRI (GAATTC) restriction sites were used according to the manufacturer's instructions. The primers designed for ORF cloning, along with their respective restriction sites, are detailed in Additional file 1: Table S9. For *Eselovl6* and *Ptelovl6*, the amplification process involved an initial denaturation step at 96 ℃ for 3 min, followed by 23 cycles of denaturation at 95 ℃ for 30 s, annealing at 58 ℃ for 15 s, and extension at 72 ℃ for 20 s, culminating in a final extension at 72 ℃ for 1 min. The PCR conditions for *Spelovl6* included an initial denaturing step at 94 ℃ for 2 min, followed by 35 cycles of denaturation at 98 ℃ for 30 s, annealing at 94 ℃ for 30 s, extension at 68 ℃ for 30 s, and a final extension at 72 ℃ for 7 min.

The resulting DNA fragments were purified using the TIAN quick mini purification kit (Tiangen Biotech), after which the fragments were digested with the corresponding BamHI and EcoRI restriction endonucleases (Thermo Scientific, USA). These were ligated into pYES2 yeast expression vector (Invitrogen, USA) using the DNA Ligation Kit Mighty Mix (Takara, Japan). The resulting plasmid constructs, *pYEs-elovl6*, *pYPtelovl6*, and *pYSpelovl6*, were then introduced into *E. coli* DH5α competent cells (Takara, Japan) screened for the presence of recombinants using PCR.

The transformation and selection of yeast with recombinant plasmids and yeast culture followed modified protocols from Monroig et al. [68]. Purified plasmids inserted with either one of the *elovl6* ORFs or no insert were used to transform yeast (*Saccharomyces cerevisiae* strain INVSc1) competent cells through the polyethylene glycol/lithium acetate (PEG/LiAc) method. The transformed yeast containing the corresponding recombinants were selected on *S. cerevisiae* minimal solid media plates without uracil (SC-Ura, containing 2% glucose and 2% yeast agar) and cultured at 30 ℃ for 3 days. Single colonies were then picked and transferred into liquid SC-Ura medium with 2% glucose and then allowed to grow again in a shaking incubator at 30 ℃ under 250 rpm for 2 days. Yeast genomic DNA was extracted from the bacterial solution using a yeast genomic DNA rapid extraction kit (Solarbio, Beijing, China), and the presence of the resulting plasmid constructs in *S. cerevisiae* was identified and confirmed by PCR screening.

For heterologous expression, successfully transformed yeast were cultured for 24 h in SC-Ura broth containing 2% raffinose at 30 ℃ and 250 rpm, diluted to an OD600 of 0.4 in SC-Ura broth and allowed to grow until the OD600 reached 1. The suspensions were centrifuged to obtain precipitated yeast, which was then resuspended in SC-Ura broth (containing 2% galactose) at an OD600 of 0.4. Next, a mixture of the suspensions, 10% tergitol-type (NP-40, Beyotime Biotechnology, Shanghai, China), and corresponding fatty acid substrates (including 16:0, 18:1n-9, 18:2n-6, 18:3n-3, 18:3n-6, 18:4n-3, 20:4n-6, and 20:5n-3) at final concentrations of 0.50 mM for C_{16}, 0.50 mM for C_{18}, and 0.75 mM for C_{20} were added to the centrifuge tubes. The tubes were placed into a shaker and incubated at 30 ℃ for 2 days at 250 rpm. Subsequently, the yeast was collected via centrifugation, washed with Hanks's balanced salt solution (Solarbio), and processed for fatty acid analysis. This experiment was replicated with three separate recombinant colonies for each recombinant yeast strain.

Fatty acid analysis by gas chromatography–mass spectrometry

After the yeast collection, the fatty acid contents were detected by gas chromatography–mass spectrometry (GC–MS) to analyze the ability of *elovl6* genes to synthesize fatty acid in different crabs. The yeast obtained above underwent a 48-h freeze-drying process and were ground into powder. Approximately 200 mg of this yeast powder was then thawed at 4 ℃ and placed in a 12-mL screw-top glass tube with a Teflon-sealed lid and 0.25 mg/mL of a methyl esterification solution, which was composed of 99 mL methanol, 1 mL sulfuric acid, and 0.025 g *2,6-Di-tert-butyl-4-methylphenol* as an antioxidant. After vortexing for 2 min and ultrasonic disruption for 30 min, the samples were incubated in a water bath at 80 ℃ for 4 h and cooled. Next, 1 mL of *n*-hexane was added to the mixture and shaken vigorously for 1 min, after which 1 mL of ultrapure water was added to facilitate layer separation. The resulting supernatant was then filtered through a nylon syringe filter with a 0.22-μm ultrafiltration membrane (SCAA-104, ANPEL, China) and collected into a clean ampoule. The fatty acid methyl ester (FAME) solution was concentrated under a stream of nitrogen gas in a Termovap sample concentrator and resuspended in 600 μL of *n*-hexane and stored at − 20 ℃ until analysis. The FAME samples were separated and analyzed by GC–MS using an Agilent 7890B-5977A GC–MS (Agilent Technologies, Santa Clara, CA, USA) equipped with a fused-silica ultra inert capillary column (DB-WAX, 30 m × 250 μm internal diameter, film thickness 0.25 μm; Agilent J & W Scientific, CA, USA). The oven temperature was increased from 100 ℃ to 200 ℃ at a rate of 10 ℃/min, with a hold time of 5 min at 200 ℃. Next, the temperature was increased to 230 ℃ at 2 ℃/min with a hold time of 10 min at 230 ℃, followed by a final increase from 230 to 240 at 10 ℃/min. The injection, interface, and ion source temperatures were adjusted to 250, 240, and 230 ℃, respectively. High-purity helium (99.999%) served as the carrier gas with a constant flow rate of 1 mL/min. A 0.5 μL sample was injected at a 1:20 split ratio by an autosampler. The collision energy was set at 70 eV, and mass spectra data were acquired in full scan mode (scanning range 40–500 *m/z*). Fatty acids were identified through mass spectrometry using a commercially available standard library (National Institute of Standards and Technology Mass Spectral Library 2011) and the relative retention times of standards. The elongation conversion efficiencies of the fatty acid substrates were determined by calculating the proportion of exogenously added fatty acids (FAs) to the elongated FA products, given by the following equation.

$$\text{Elongation conversion efficiency}(\%) = \frac{\text{Product area}}{\text{Product area} + \text{substrate area}} \times 100$$

Abbreviations

LC-PUFA	Long-chain polyunsaturated fatty acid
DNA	Deoxyribonucleic acid
ONT	Oxford Nanopore Technologies
BUSCO	Benchmarking Universal Single-Copy Orthologs
SNP	Single-nucleotide polymorphism
SSR	Simple sequence repeat
LINEs	Long interspersed elements
LTRs	Long terminal repeats
RC	Rolling circle
CDS	Coding sequence
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene Ontology
KOG	EuKaryotic Orthologous Groups
Nr	Non-redundant protein sequence database
RNA	Ribonucleic acid
miRNA	MicroRNA
rRNA	Ribosomal RNA
snRNA	Small nuclear RNA
tRNA	Transfer RNA
Mya	Million years ago
lab	Labial
pb	Proboscipedia
Hox3	Homeobox gene cluster 3
Dfd	Deformed
Scr	Sex combs reduced
ftz	Fushi taraz
Antp	Antennapedia
Ubx	Ultrabithorax
Abd-A	Abdominal-A
Abd-B	Abdominal-B
fru	Fruitless
EPA	Eicosapentaenoic acid
DHA	Docosahexaenoic acid
elovl6	Elongase of very long chain fatty acids 6
SFA	Saturated fatty acids
MUFA	Monounsaturated fatty acids

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-024-02054-1.

Additional file 1: Figs. S1–S10 and Tables S1–S9. Fig. S1. Genome-wide association studies (GWAS) analysis of *Scylla paramamosain* genome using different generations of populations. Fig. S2. Divergence rate of transposons in *Scylla paramamosain* genome. Fig. S3. Characterization of different genic regions in *Scylla paramamosain* and five other arthropods. Fig. S4. Functional annotation of predicted genes in *Scylla paramamosain*. Fig. S5 Ortholog gene analysis and phylogenetic trees of 14 selected species. Fig. S6. GO (A) and KEGG (B) enrichment of unique genes in *Scylla paramamosain*. Fig. S7. Phylogenetic trees of 14 selected species. Fig. S8. KEGG enrichment of expansion (A) and contraction (B) genes in *Scylla paramamosain*. Fig. S9 The expression and regulation of *fru2* gene in gonads of *Scylla paramamosain*. Fig. S10 Neo-functionalization of the *elovl6* gene in the LC-PUFA synthesis pathway in crustacean species. Table S1 Comparison of published crustaceans genome. Table S2 Details of repeat sequences in *Scylla paramamosain*. Table S3. Comparison of protein-coding genes between *Scylla paramamosain* and other species. Table S4. Positive selection genes in *Scylla paramamosain*. Table S5. Summary of non-coding genes in *Scylla paramamosain*. Table S6. Functional characterization of *Eriocheir sinensis*, *Portunus trituberculatus*, and *Scylla paramamosain* Elovl6 via heterologous expression in yeast Saccharomyces cerevisiaea. Table S7. Statistics of genome sequencing data of *Scylla paramamosain*. Table S8. RNA assemblies used in gene annotation of *Scylla paramamosain*. Table S9. Primers used for *Eselovl6*, *Ptelovl6*, and *Spelovl6* functional characterization.

Acknowledgements

We express our gratitude to Dr. Noah Esmaeili for his valuable input in enhancing the linguistic quality of the manuscript.

Authors' contributions

Hongyu Ma provided the funding of the research and conceptualized the whole experiment. Yin Zhang, Ye Yuan, Mengqian Zhang, Xiaoyan Yu, Bixun Qiu and Fangchun Wu collected the specimens. Yin Zhang, Ye Yuan, Xiaoyan Yu, Bixun Qiu, Fangchun Wu, Jiajia Zhang and Mengqian Zhang performed the experiment. Yin Zhang, Ye Yuan and Mengqian Zhang analyzed the data and prepared the required figures. Yin Zhang, Ye Yuan and Mengqian Zhang wrote the manuscript. Shaopan Ye and Yin Zhang upload the genome data. Hongyu Ma, Shaopan Ye, Wenxiao Cui, Jonathan Y. S. Leung, Mhd Ikhwanuddin, Tariq Dildar, Waqas Waqas and Douglas R. Tocher revised the manuscript and contributed to additional discussion. All authors read and approved the final manuscript.

Funding

This study was supported by the National Key Research & Development Program of China (2018YFD0900201), the National Natural Science Foundation of China (42076133, 42306126, 42306125), the National Plan for the Special Support for Top-notch Talents (ZUTINGZI201548), the Leading Talent Project of Special Support Plan of Guangdong Province (2019TX05N067), the Program of Agricultural and Rural Department of Guangdong Province (2022-SPY-00–014), the Guangdong Natural Science Foundation (2022A1515110488, 2022A1515111151), and the STU Scientific Research Foundation for Talents (NTF21016, NTF21020).

Data availability

The whole genome sequences are deposited at DDBJ/ENA/GenBank under the accession JAYKKS000000000. The version described in this paper is version JAYKKS010000000. Raw sequencing data are deposited in NCBI under BioProject PRJNA1059155.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Guangdong Provincial Key Laboratory of Marine Biotechnology, Shantou University, 243 Daxue Road, Shantou 515063, China. [2]International Joint Research Center for the Development and Utilization of Important Mariculture Varieties Surrounding the South China Sea Region, Shantou University, Shantou, China. [3]STU-UMT Joint Shellfish Research Laboratory, Shantou University, Shantou, China. [4]Higher Institute Centre of Excellence (HICoE), Institute of Tropical Aquaculture and Fisheries, Universiti Malaysia Terengganu, Kuala Nerus, Terengganu, Malaysia.

Received: 26 August 2024 Accepted: 25 October 2024
Published online: 07 November 2024

References

1. Sanger TJ, Rajakumar R. How a growing organismal perspective is adding new depth to integrative studies of morphological evolution. Biol Rev. 2019;94:184–98. https://doi.org/10.1111/brv.12442.
2. Schweitzer CE, Feldmann RM. The oldest Brachyura (Decapoda: Homolodromioidea: Glaessneropsoidea) known to date (Jurassic). J Crustacean Biol. 2010;30:251–6. https://doi.org/10.1651/09-3231.1.
3. Ma KY, Qin J, Lin C, Chan T, Ng PKL, Chu KH, et al. Phylogenomic analyses of brachyuran crabs support early divergence of primary freshwater

crabs. Mol Phylogenet Evol. 2019;135:62–6. https://doi.org/10.1016/j.ympev.2019.02.001.
4. Morrison CL, Harvey AW, Lavery S, Tieu K, Huang Y, Cunningham CW. Mitochondrial gene rearrangements confirm the parallel evolution of the crab-like form. Proc R Soc B. 2002;269:345–50. https://doi.org/10.1098/rspb.2001.1886.
5. Wolfe JM, Luque J, Bracken-Grissom HD. How to become a crab: phenotypic constraints on a recurring body plan. BioEssays. 2021;43:2100020. https://doi.org/10.1002/bies.202100020.
6. Deutsch JS, Mouchel-Vielh E. Hox genes and the crustacean body plan. BioEssays. 2003;25:878–87. https://doi.org/10.1002/bies.10319.
7. McNamara JC, Faria SC. Evolution of osmoregulatory patterns and gill ion transport mechanisms in the decapod Crustacea: a review. J Comp Physiol B. 2012;182:997–1014. https://doi.org/10.1007/s00360-012-0665-8.
8. Bracken-Grissom HD, Cannon ME, Cabezas P, Feldmann RM, Schweitzer CE, Ahyong ST, et al. A comprehensive and integrative reconstruction of evolutionary history for Anomura (Crustacea: Decapoda). BMC Evol Biol. 2013;13:128 https://doi.org/10.1186/1471-2148-13-128.
9. Scholtz G. Evolution of crabs - history and deconstruction of a prime example of convergence. Contrib Zool. 2014;83:87–105. https://doi.org/10.1163/18759866-08302001.
10. Luque J, Feldmann RM, Vernygora O, Schweitzer CE, Cameron CB, Kerr KA, et al. Exceptional preservation of mid-Cretaceous marine arthropods and the evolution of novel forms via heterochrony. Sci Adv. 2019;5: 5. https://doi.org/10.1126/sciadv.aav3875.
11. Flatt T, Heyland A, editors. Mechanisms of life history evolution: the genetics and physiology of life history traits and trade-offs. Oxford University Press; 2011.
12. Gong H, Wang T, Wu M, Chu Q, Lan H, Lang W, et al. Maternal effects drive intestinal development beginning in the embryonic period on the basis of maternal immune and microbial transfer in chickens. Microbiome. 2023;11:41. https://doi.org/10.1186/s40168-023-01490-5.
13. Turner K, Solanki N, Salouha HO, Avidor-Reiss T. Atypical centriolar composition correlates with internal fertilization in fish. Cells. 2022;11: 758. https://doi.org/10.3390/cells11050758.
14. Zhang X, Yuan J, Sun Y, Li S, Gao Y, Yu Y, et al. Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. Nat Commun. 2019;10:356. https://doi.org/10.1038/s41467-018-08197-4.
15. Tocher DR. Fatty acid requirements in ontogeny of marine and freshwater fish. Aquacult Res. 2010;41:717–32. https://doi.org/10.1111/j.1365-2109.2008.02150.x.
16. Ting SY, Janaranjani M, Merosha P, Sam K, Wong SC, Goh P, et al. Two elongases, Elovl4 and Elovl6, fulfill the elongation routes of the LC-PUFA biosynthesis pathway in the orange mud crab (*Scylla olivacea*). J Agric Food Chem. 2020;68:4116–30. https://doi.org/10.1021/acs.jafc.9b06692.
17. Závorka L, Blanco A, Chaguaceda F, Cucherousset J, Killen SS, Liénart C, et al. The role of vital dietary biomolecules in eco-evo-devo dynamics. Trends Ecol Evol. 2023;38:72–84. https://doi.org/10.1016/j.tree.2022.08.010.
18. Twining CW, Bernhardt JR, Derry AM, Hudson CM, Ishikawa A, Kabeya N, et al. The evolutionary ecology of fatty-acid variation: implications for consumer adaptation and diversification. Ecol Lett. 2021;24:1709–31. https://doi.org/10.1111/ele.13771.
19. Castro LFC, Tocher DR, Monroig O. Long-chain polyunsaturated fatty acid biosynthesis in chordates: insights into the evolution of Fads and Elovl gene repertoire. Prog Lipid Res. 2016;62:25–40. https://doi.org/10.1016/j.plipres.2016.01.001.
20. Kabeya N, Fonsec MM, Ferrie DEK, Navarr JC, Ba LK, Francis DS, et al. Genes for de novo biosynthesis of omega-3 polyunsaturated fatty acids are widespread in animals. Sci Adv. 2018;4:4. https://doi.org/10.1126/sciadv.aar6849.
21. Monroig Ó, Kabeya N. Desaturases and elongases involved in polyunsaturated fatty acid biosynthesis in aquatic invertebrates: a comprehensive review. Fish Sci. 2018;84:911–28. https://doi.org/10.1007/s12562-018-1254-x.
22. Xie D, Chen C, Dong Y, You C, Wang S, Monroig Ó, et al. Regulation of long-chain polyunsaturated fatty acid biosynthesis in teleost fish. Prog Lipid Res. 2021;82: 101095. https://doi.org/10.1016/j.plipres.2021.101095.
23. Monroig Ó, Shu-Chien AC, Kabeya N, Tocher DR, Castro LFC. Desaturases and elongases involved in long-chain polyunsaturated fatty acid biosynthesis in aquatic animals: from genes to functions. Prog Lipid Res. 2022;86: 101157. https://doi.org/10.1016/j.plipres.2022.101157.
24. Ramos-Llorens M, Hontoria F, Navarro JC, Ferrier DEK, Monroig Ó. Functionally diverse front-end desaturases are widespread in the phylum Annelida. Biochim Biophys Acta Mol Cell Biol Lipids. 2023;1868: 159377. https://doi.org/10.1016/j.bbalip.2023.159377.
25. Yim H, Cho YS, Guang X, Kang SG, Jeong J, Cha S, et al. Minke whale genome and aquatic adaptation in cetaceans. Nat Genet. 2014;46:88–92. https://doi.org/10.1038/ng.2835.
26. Gutekunst J, Andriantsoa R, Falckenhayn C, Hanna K, Stein W, Rasamy J, et al. Clonal genome evolution and rapid invasive spread of the marbled crayfish. Nat Ecol Evol. 2018;2:567–73. https://doi.org/10.1038/s41559-018-0467-9.
27. Song L, Bian C, Luo Y, Wang L, You X, Li J, et al. Draft genome of the Chinese mitten crab, *Eriocheir sinensis*. GigaScience. 2016;5:5. https://doi.org/10.1186/s13742-016-0112-y.
28. Tang B, Wang Z, Liu Q, Wang Z, Ren Y, Guo H, et al. Chromosome-level genome assembly of *Paralithodes platypus* provides insights into evolution and adaptation of king crabs. Mol Ecol Resour. 2021;21:511–25. https://doi.org/10.1111/1755-0998.13266.
29. Tang B, Zhang D, Li H, Jiang S, Zhang H, Xuan F, et al. Chromosome-level genome assembly reveals the unique genome evolution of the swimming crab (*Portunus trituberculatus*). GigaScience. 2020;9:1–10. https://doi.org/10.1093/gigascience/giz161.
30. Zhao M, Wang W, Zhang F, Ma C, Liu Z, Yang M, et al. A chromosome-level genome of the mud crab (*Scylla paramamosain*) provides insights into the evolution of chemical and light perception in this crustacean. Mol Ecol Resour. 2021;21:1299–317. https://doi.org/10.1111/1755-0998.13332.
31. Cui Z, Liu Y, Yuan J, Zhang X, Ventura T, Ma KY, et al. The Chinese mitten crab genome provides insights into adaptive plasticity and developmental regulation. Nat Commun. 2021;12:2395. https://doi.org/10.1038/s41467-021-22604-3.
32. Chebbi MA, Becking T, Moumen B, Giraud I, Gilbert C, Peccoud J, et al. The genome of *Armadillidium vulgare* (Crustacea, Isopoda) provides insights into sex chromosome evolution in the context of cytoplasmic sex determination. Mol Biol Evol. 2019;36:727–41. https://doi.org/10.1093/molbev/msz010.
33. Waiho K, Shi X, Fazhan H, Li S, Zhang Y, Zheng H, et al. High-density genetic linkage maps provide novel insights into ZW/ZZ sex determination system and growth performance in mud crab (*Scylla paramamosain*). Front Genet. 2019;10: 298. https://doi.org/10.3389/fgene.2019.00298.
34. Bire S, Rouleux-Bonnin F. Transposable elements as tools for reshaping the genome: it is a huge world after all! Methods Mol Biol. 2012;859:1–28. https://doi.org/10.1007/978-1-61779-603-6_1.
35. Elliott TA, Gregory TR. Do larger genomes contain more diverse transposable elements? BMC Evol Biol. 2015;15:69. https://doi.org/10.1186/s12862-015-0339-8.
36. Yan X, Nie H, Huo Z, Ding J, Li Z, Yan L, et al. Clam genome sequence clarifies the molecular basis of its benthic adaptation and extraordinary shell color diversity. iScience. 2019;19:1225–37. https://doi.org/10.1016/j.isci.2019.08.049.
37. Grenier JK, Garber TL, Warren R, Whitington PM, Carroll S. Evolution of the entire arthropod Hox gene set predated the origin and radiation of the onychophoran/arthropod clade. Curr Biol. 1997;7:547–53. https://doi.org/10.1016/S0960-9822(06)00253-3.
38. Hughes CL, Kaufman TC. Hox genes and the evolution of the arthropod body plan. Evol Dev. 2002;4:459–99. https://doi.org/10.1046/j.1525-142X.2002.02034.x.
39. Uengwetwanit T, Pootakham W, Nookaew I, Sonthirod C, Angthong P, Sittikankaew K, et al. A chromosome-level assembly of the black tiger shrimp (*Penaeus monodon*) genome facilitates the identification of growth-associated genes. Mol Ecol Resour. 2021;21:1620–40. https://doi.org/10.1111/1755-0998.13357.
40. Kim D, Lee B, Kim H, Jeong C, Hwang D, Kim I, et al. Identification and characterization of homeobox (Hox) genes and conservation of the single Hox cluster (324.6kb) in the water flea *Daphnia magna*. J Exp Zool Part B Mol Dev Evol. 2018;330:76–82. https://doi.org/10.1002/jez.b.22793.
41. Miura S, Nozawa M, Nei M. Evolutionary changes of the target sites of two microRNAs encoded in the Hox gene cluster of *Drosophila* and other insect species. Genome Biol Evol. 2011;3:129–39. https://doi.org/10.1093/gbe/evq088.

42. Mouchel-Vielh E, Rigolot C, Gibert JM, Deutsch JS. Molecules and the body plan: the Hox genes of cirripedes (Crustacea). Mol Phylogenet Evol. 1998;9:382–9. https://doi.org/10.1006/mpev.1998.0498.
43. Garaulet DL, Lai EC. Hox miRNA regulation within the *Drosophila* Bithorax complex: Patterning behavior. Mech Dev. 2015;138:151–9. https://doi.org/10.1016/j.mod.2015.08.006.
44. Manoli DS, Foss M, Villella A, Taylor BJ, Hall JC, Baker BS. Male-specific fruitless specifies the neural substrates of *Drosophila* courtship behaviour. Nature. 2005;436:395–400. https://doi.org/10.1038/nature03859.
45. Cachero S, Ostrovsky AD, Yu JY, Dickson BJ, Jefferis GS. Sexual dimorphism in the fly brain. Curr Biol. 2010;20:1589–601. https://doi.org/10.1016/j.cub.2010.07.045.
46. Yang X, Ikhwanuddin M, Li X, Lin F, Wu Q, Zhang Y, et al. Comparative transcriptome analysis provides insights into differentially expressed genes and long non-coding RNAs between ovary and testis of the mud crab (*Scylla paramamosain*). Mar Biotechnol. 2018;20:20–34. https://doi.org/10.1007/s10126-017-9784-2.
47. Yang Y, Chen F, Qiao K, Zhang H, Chen HY, Wang KJ. Two male-specific antimicrobial peptides SCY2 and Scyreprocin as crucial molecules participated in the sperm acrosome reaction of mud crab *Scylla paramamosain*. Int J Mol Sci. 2022;23: 3373. https://doi.org/10.3390/ijms23063373.
48. Matsuzaka T, Shimano H, Yahagi N, Kato T, Atsumi A, Yamamoto T, et al. Crucial role of a long-chain fatty acid elongase, Elovl6, in obesity-induced insulin resistance. Nat Med. 2007;13:1193–202. https://doi.org/10.1038/nm1662.
49. Shimano H. Novel qualitative aspects of tissue fatty acids related to metabolic regulation: Lessons from Elovl6 knockout. Prog Lipid Res. 2012;51:267–71. https://doi.org/10.1016/j.plipres.2011.12.004.
50. Junjvlieke Z, Mei C, Khan R, Zhang W, Hong J, Wang L, et al. Transcriptional regulation of bovine elongation of very long chain fatty acids protein 6 in lipid metabolism and adipocyte proliferation. J Cell Biochem. 2019;120:13932–43. https://doi.org/10.1002/jcb.28667.
51. Li Y, Wen Z, You C, Xie Z, Tocher DR, Zhang Y, et al. Genome-wide identification and functional characterization of two LC-PUFA biosynthesis elongase (elovl8) genes in rabbitfish (*Siganus canaliculatus*). Aquaculture. 2020;522:735127. https://doi.org/10.1016/j.aquaculture.2020.735127.
52. Wang X, Sun S, Cao X, Gao J. Quantitative phosphoproteomic analysis reveals the regulatory networks of Elovl6 on lipid and glucose metabolism in zebrafish. Int J Mol Sci. 2020;21: 2860. https://doi.org/10.3390/ijms21082860.
53. Matsushita Y, Miyoshi K, Kabeya N, Sanada S, Yazawa R, Haga Y, et al. Flatfishes colonised freshwater environments by acquisition of various DHA biosynthetic pathways. Commun Biol. 2020;3:516. https://doi.org/10.1038/s42003-020-01242-3.
54. Ishikawa A, Stuart YE, Bolnick DI, Kitano J. Copy number variation of a fatty acid desaturase gene Fads2 associated with ecological divergence in freshwater stickleback populations. Biol Lett. 2021;17:20210204. https://doi.org/10.1098/rsbl.2021.0204.
55. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, et al. The oyster genome reveals stress adaptation and complexity of shell formation. Nature. 2012;490:49–54. https://doi.org/10.1038/nature11413.
56. Li Y, Sun X, Hu X, Xun X, Zhang J, Guo X, et al. Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins. Nat Commun. 2017;8:1721. https://doi.org/10.1038/s41467-017-01927-0.
57. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol. 2018;36:338–45. https://doi.org/10.1038/nbt.4060.
58. Polinski JM, Zimin AV, Clark KF, Kohn AB, Sadowski N, Timp W, et al. The American lobster genome reveals insights on longevity, neural, and immune adaptations. Sci Adv. 2021;7(26):eabe8290. https://doi.org/10.1126/sciadv.abe8290.
59. Wang S, Zhang J, Jiao W, Li J, Xun X, Sun Y, et al. Scallop genome provides insights into evolution of bilaterian karyotype and development. Nat Ecol Evol. 2017;1:1–12. https://doi.org/10.1038/s41559-017-0120.
60. Ikuta T, Yoshida N, Satoh N, Saiga H. *Ciona intestinalis* Hox gene cluster: its dispersed structure and residual colinear expression in development. Proc Natl Acad Sci USA. 2004;101:15118–23. https://doi.org/10.1073/pnas.0401389101.
61. Martin A, Serano JM, Jarvis E, Bruce HS, Wang J, Ray S, et al. CRISPR/Cas9 mutagenesis reveals versatile roles of Hox genes in crustacean limb specification and evolution. Curr Biol. 2016;26:14–26. https://doi.org/10.1016/j.cub.2015.11.021.
62. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. 2019;20:129. https://doi.org/10.1186/s13059-019-1727-y.
63. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9. https://doi.org/10.1038/nmeth.1923.
64. Joshua N, Burton AA. Chromosome-scale contiging of de novo genome assemblies based on chromatin interactions. Nat Biotechnol. 2013;31:1119–25. https://doi.org/10.1038/nbt.2727.
65. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. GPB. 2010;8:77–80. https://doi.org/10.1016/S1672-0229(10)60008-3.
66. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40:40. https://doi.org/10.1093/nar/gkr1293.
67. Pinaud R, Mello CV, Velho TA, Wynne RD, Tremere LA. Detection of two mRNA species at single-cell resolution by double-fluorescence in situ hybridization. Nat Protoc. 2008;3:1370–9. https://doi.org/10.1038/nprot.2008.115.
68. Monroig Ó, Wang S, Zhang L, You C, Tocher DR, Li Y. Elongation of long-chain fatty acids in rabbitfish *Siganus canaliculatus*: cloning, functional characterisation and tissue distribution of Elovl5-and Elovl4-like elongases. Aquaculture. 2012;350:63–70. https://doi.org/10.1016/j.aquaculture.2012.04.017.

Publisher's Note