Genomics 115 (2023) 110582

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

A chromosome-level genome assembly of *Ostrea denselamellosa* provides initial insights into its evolution

Zhen Dong [a], Yitian Bai [a], Shikai Liu [a], Hong Yu [a], Lingfeng Kong [a], Shaojun Du [c], Qi Li [a,b,*]

[a] Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao 266003, China
[b] Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China
[c] Institute of Marine and Environmental Technology, Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, Baltimore, MD, United States

ARTICLE INFO

Keywords:
Ostrea denselamellosa
Oysters
Chromosome-level assembly
Hi-C
Phylogeny
Evolution

ABSTRACT

The oyster *Ostrea denselamellosa* is a live-bearing species with a sharp decline in the natural population. Despite recent breakthroughs in long-read sequencing, high quality genomic data are very limited in *O. denselamellosa*. Here, we carried out the first whole genome sequencing at the chromosome-level in *O. denselamellosa*. Our studies yielded a 636 Mb assembly with scaffold N50 around 71.80 Mb. 608.3 Mb (95.6% of the assembly) were anchored to 10 chromosomes. A total of 26,412 protein-coding genes were predicted, of which 22,636 (85.7%) were functionally annotated. By comparative genomics, we found that long interspersed nuclear element (LINE) and short interspersed nuclear element (SINE) made up a larger proportion in *O. denselamellosa* genome than in other oysters'. Moreover, gene family analysis showed some initial insight into its evolution. This high-quality genome of *O. denselamellosa* provides a valuable genomic resource for studies of evolution, adaption and conservation in oysters.

1. Introduction

The flat oyster *Ostrea denselamellosa* (NCBI: txid74434) is a potential economically important species, naturally distributed in subtidal zone along the coasts of China, Japan and Korea [1]. The *O. denselamellosa* population has declined during the past years due to climate change, reclamation, habitat destruction and over-exploitation [2]. Like other oysters in the genus *Ostrea*, *O. denselamellosa* is a live-bearing species, which differs significantly from oysters in genus *Crassostrea*, *Saccostrea* and most other bivalves. During reproduction, the eggs are pressed into the pallial cavity of female individuals, and then are fertilized and grow to D-shaped larvae within 1–3 days there. The D-shaped larvae go through a planktonic period, and then attach to a hard substrate to become juvenile [3–5]. To date, *O. denselamellosa* can not be cultivated due to its unique pattern of reproduction and maladaptation of larvae to the high salinity environment of living adults [2]. However, live-bearing is poorly researched in invertebrates. Previous studies in *Ostrea* were mainly focused on its mitochondrial genome, seed production and biological characteristics [2,6,7].

With the rapid development of next-generation sequencing technology, more and more oyster genome data have become available, such as *Crassostrea gigas* [8,9], *Crassostrea hongkongensis* [10,11], *Crassostrea virginica* [12], *Crassostrea ariakensis* [13,14], and *Saccostrea glomerata* [15]. However, the genomic features and evolutionary characteristics of *Ostrea* remain poorly understood. High-quality genomic resources for *O. denselamellosa* are particularly important. It can not only provide genetic information for researchers to study their manage sustainable conservation and economic development, but also provide a solid foundation for subsequent research which makes *O. denselamellosa* a promising model for live-bearing oysters.

In this study, we report the first chromosome-level genome sequence for *O. denselamellosa* by utilizing PacBio long sequencing, Illumina short

Abbreviations: BLAST, Basic Local Alignment Search Tool; bp, base pairs; BUSCO, Benchmarking Universal Single-Copy Orthologs; BWA, Burrows-Wheeler Aligner; cDNA, complementary DNA; cds, coding gene sequence; DEG, differentially expressed genes; Gb, Gigabase pairs; GO, Gene Ontology; kb, kilobase pairs; KEGG, Kyoto Encyclopedia of Genes and Genomes; Mb, Megabase pairs; NCBI, National Center for Biotechnology Information; PacBio, Pacific Biosciences; PASA, Program to Assemble Spliced Alignments; RAxML, Randomized Axelerated Maximum Likelihood; RNA-seq, RNA sequencing; SRA, Sequence Read Archive; TE, tandem repeat; TMM, Trimmed Mean of M-values; TPM, Transcripts Per Million.

* Corresponding author at: Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao 266003, China.
E-mail address: qili66@ouc.edu.cn (Q. Li).

https://doi.org/10.1016/j.ygeno.2023.110582
Received 23 September 2022; Received in revised form 9 February 2023; Accepted 11 February 2023
Available online 14 February 2023

reads and high-throughput chromosome conformation capture (Hi-C) technology. The genome sequence was annotated with full-length transcriptome data and RNA-seq data. Comparative genomics and comparative transcriptomics were used to explain the evolution and phylogeny of *O. denselamellosa*. Together, these results will not only be helpful to understand the evolution of *O. denselamellosa* but also generate valuable genomic and genetic resources for breeding and protection of live-bearing oysters.

2. Materials and methods

2.1. Sample collection and genome sequencing

Adult *O. denselamellosa* were collected in July 2020 from Jiaonan (35.88°N, 119.97°E), Shandong Province, China. The *O. denselamellosa* had an average shell height of 109.09 mm and shell length of 97.83 mm (Fig. 1). Adductors, gonads, gills, labial palps and mantles were dissected and frozen immediately in liquid nitrogen. The samples were stored at −80 °C for later use of DNA and RNA extraction. Genomic DNA from an adductor of a female oyster (Fig. S1) was used to build several libraries with short and long DNA inserts. These include a 150 bp insert Illumina paired-end library, 15 kb insert Pacbio hifi library and Hi-C library for genome assembly by Illumina HiSeq X Ten platform (Illumina HiSeqX Ten, RRID:SCR_016385), Pacbio Sequel II platform (PacBio Sequel II System, RRID:SCR_017990) and Illumina HiSeq X platform (Illumina HiSeq X Ten, RRID:SCR_016385), respectively. In addition, total RNA was extracted from adductor (smooth muscle), adductor (striated muscle), gonads, gills, labial palps and mantle tissues of the female individual used in whole genome sequencing and a male gonad using TRIzol reagent. Seven libraries of the seven tissues mentioned above were constructed for genome annotation. RNAs from all these tissues were pooled together for full-length transcriptome sequencing (Iso-Seq) library construction on a PacBio Sequel platform. The clean data was generated by SMRTlink (v6.0) (https://www.pacb.com/support/software-downloads/) with minLength = 50. Another twelve RNA-seq libraries were constructed for different gene expression analysis, including libraries of adductor, gonads, gills and mantle tissues from each of the three female individuals in stage of ovulation. All RNA-seq libraries were built by NovaSeq 6000 platform (Illumina NovaSeq 6000 Sequencing System, RRID:SCR_016387) in Novogene Company (Beijing, China).

2.2. Genome assembly and assessment

Illumina paired-end reads were used as the first quality control by FASTP (fastp, RRID:SCR_016962) [16] with the default parameter. JELLYFISH (v2.2.10) (Jellyfish, RRID:SCR_005491) [17] and PLATANUS (v2.2.2) (Platanus, RRID:SCR_015531) [18] were employed to estimate the genome size based on a 19-mer distribution with high quality Illumina paired-end reads mentioned above. Heterozygosity was inferred using GENOMESCOPE (v2.0) (GenomeScope, RRID:SCR 017014) [19]. MASURCA (v3.3.2) (MaSuRCA, RRID:SCR_010691) [20] was also used to estimate the genome size.

PacBio hifi reads were used to assemble the contigs of *O. denselamellosa* genome using HIFIASM (v0.14-r313) (Hifiasm, RRID:SCR_021069) [21] with parameters "-l3 -s 0.55". Sequence redundancy was removed with PURGE_DUPS (v1.2.5) (purge dups, RRID:SCR_021173) [22] based on the Illumina paired-end reads. After assembling the draft genome of *O. denselamellosa*, we generated a chromosome-level genome assembly by applying Hi-C technology to anchor the scaffolds into the chromosomes. First, Hi-C reads of *O. denselamellosa* were mapped to draft assembly by BWA (v0.7.8) (BWA, RRID:SCR_010910) [23]. Then the contact matrix was constructed by JUICER (v1.6) (Juicer, RRID:SCR_017226) [24]. 3D-DNA (v180922) (3D de novo assembly, RRID:SCR_017227) [25] was used to detect and correct the assembly errors. Finally, JUICEBOX (Juicebox, RRID:SCR_021172) (v1.11.08) [24] was used to visualize the results generated by 3D-DNA and correct errors manually in order to optimize the quality of chromosome-level genome sequence (Fig. S2).

Genome quality assessment was conducted using Benchmarking Universal Single-Copy Orthologs (BUSCO) (v5.0.0) (BUSCO, RRID:SCR_015008) [26] with the parameters of "-m genome". We searched the genome against 954 metazoan single-copy orthologues from metazoa_odb10 (https://busco-data.ezlab.org/v4/data/lineages/). The genome size and N50 were tested by QUAST (v5.0.2) (QUAST, RRID:SCR 001228) [27].

2.3. Genome annotation

The chromosome-level genome assembly was used to build a de novo database by REPEATMODELER (v2.0.1) (RepeatModeler, RRID:SCR_015027) [28]. The data of *Ostrea* and *Crassostrea* extracted from RepBase (v20181026) (Repbase, RRID:SCR_021169) [29] were referred as the homologous database. The two databases were combined into one as a combined database. REPEATMASKER (v4.1.2) (RepeatMasker, RRID:SCR_012954) [30] was used with the parameters "-nolow -no_is -norna -html -gff -dir" based on the combined database to detect de novo and homology-based transposable elements (TEs). REPEATPROTEINMASK was used with parameters "-engine ncbi -noLowSimple -pvalue 0.0001" to predict protein masker. Finally, TANDEM REPEATS FINDER (TRF) (v4.09.1) (Tandem Repeats Finder, RRID:SCR_022193) [31] was used to detect simple or tandem repeats. BEDTOOLS (v2.30.0)

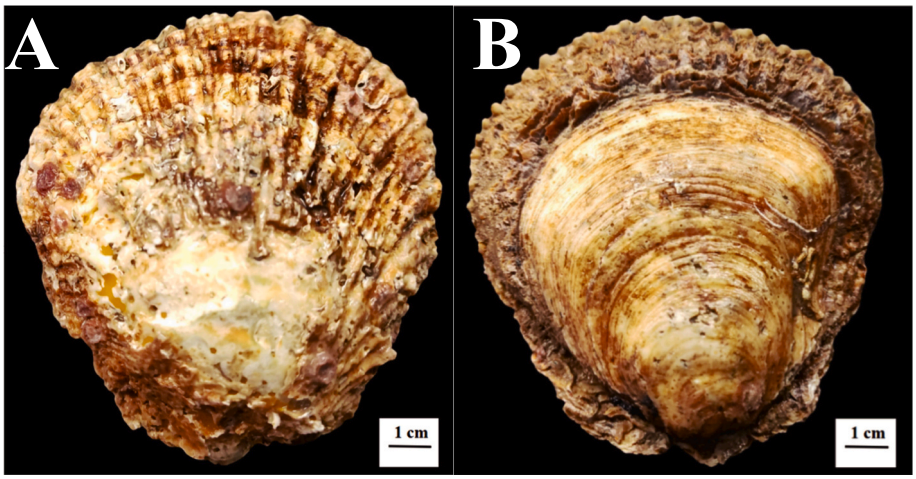

Fig. 1. Photographs of *Ostrea denselamellosa*. A. Left shell. B. Right shell.

(BEDTools, RRID:SCR_006646) [32] was used with parameters "maskfasta -soft" to transform the repeat prediction result from hardmask to softmask. The proportional distribution of DNA transposons, LTR, Helitron, LINE and SINE retrotransposons in the genomes of 9 representative Mollusca was generated. The path where we downloaded the genome of other 8 Mollusca was shown in Table S1 and the repeat annotations were calculated in the same way as done for *O. denselamellosa*. It is worth noting that the data extracted from the RepBase corresponded to the genus in which the species belong to. The repeat landscapes were built by using a R script which was modified from https://github.com/ValentinaBoP/TransposableElements.

LTRHARVEST (v6.3.0) (LTRharvest, RRID:SCR_018970) [33] was initially used to identify long terminal repeat (LTR) fragment in *O. denselamellosa* genome (unmask version) with patameters "-minlenltr 80 -maxlenltr 7000 -mintsd 4 -maxtsd 6 -motif TGCA -seqids yes". LTR was finally detected by LTRRETRIEVER (v2.9.0) (https://github.com/oushujun/LTR_retriever) using the result from LTRHARVEST. Kimura 2-Paramete was used to calculate the rate of nucleic acid mutation which is necessary parameter for LTR_retriever.

Genome structure and gene model prediction were constructed based on MAKER (v3.01.03) (MAKER, RRID:SCR 005309) [34] by integrating three approaches (homology-based prediction, ab initio annotation and transcriptome-based prediction). Homology-based prediction was performed using homologous protein sequences from *C. virginica*, *Mizuhopecten yessoensis*, *C. gigas*, *Danio rerio* and Mollusca data from uniport (Table S1). RNA-seq data and Iso-seq data were assembled using TRINITY (v2.12.0) (Trinity, RRID:SCR 013048) [35] and SMRTLINK package (v9.0.0.92188) respectively. The assembled transcripts from RNA-seq were aligned against the *O. denselamellosa* genome using Program to Assemble Spliced Alignment (PASA) (v2.4.1) [36]. Then the transcripts from PASA and SMRTLINK were combined and optimized by CD-HIT (v4.8.1) (CD-HIT, RRID:SCR_007105) [37] with parameters "-c 0.97 -n 10 -d 0 -M 16000". For ab initio annotation, BRAKER (v2.1.5) (BRAKER, RRID:SCR_018964) [38] was used to create and train the species model with softmasked assembly. AUGUSTUS (v3.4.0) (Augustus, RRID:SCR_008417) [39] was used to predict genes based on the species model trained by BRAKER. BUSCO was used to assess the quality of structural annotation with the parameters of "-l metazoa_odb10 -m prot".

The functional annotation was predicted by EGGNOG-MAPPER (v2.1.0–1) [40] based on eggNOG orthology data (eggNOG, RRID:SCR_002456) [41], Swiss-prot (I UniProtKB/Swiss-Prot, RRID:SCR_021164) [42] database, NR, Interproscan (InterProScan, RRID:SCR_005829) based on Pfam (Pfam, RRID:SCR_004726), PRINTS, PANTHER, ProSiteProfiles and SMART database. Sequence searches were performed using DIAMOND (DIAMOND, RRID:SCR_016071) [43] with *E*-value $1e^{-5}$.

2.4. Genome phylogeny and gene family analysis

To better understand the genomic differences between *O. denselamellosa* and other Mollusca, we obtained 14 published genomes including six Bivalves (pacific oyster *C. gigas*; eastern oyster *C. virginica*; Sydney oyster *S. glomerata*; pearl oyster *Pinctada fucata martensii*; scallops *M. yessoensis* and cold-seep mussel *Bathymodiolus platifrons*), five Gastropods (owl limpet *Lottia gigantea*; snails *Biomphalaria glabrata*, *Pomacea canaliculate, Chrysomallon squamiferum* and California sea hare *Aplysia californica*), two Cephalopod (octopus *Octopus bimaculoides* and *Nautilus pompilius*), and one Brachiopod (*Lingula anatina*) designated as the outer group. Only the longest transcript was selected for each gene with alternative splicing variants.

Gene family clusters of all 15 species were analyzed by ORTHOFINDER (v2.5.2) (OrthoFinder, RRID:SCR_017118) [44] with the inflation parameter I set to 1.5. Phylogenetic tree was inferred by single-copy orthologues. The single-copy orthologues were selected for alignment using MUSCLE (v3.6) [45] and then were first concatenated into supergenes for each species and optimized by TRIMAL (v7.221) (trimAl, RRID:SCR_017334) [46] with "-gt 0.6 -cons 60". They were then subsequently analyzed by maximum likelihood (ML) using RAXMLHPC-PTHREADS (v8.2.12) (RAxML, RRID:SCR_006086) [47] with 1000 bootstrap replicates, and PROTGAMMAJTT was selected as amino acid substitution model.

Protein sequences of all 15 species were used to build the divergence time tree. First, CODEML from PAML package (v4.9j) (PAML, RRID:SCR_014932) [48] was used to estimate substitution rate in per time unit. Then MCMCTREE from PAML package was used to estimate the divergence time. Five reference divergence time points (Table S2) were retrieved from TimeTree database (http://timetree.org) (TimeTree, RRID:SCR_021162) [49]. The burn-in, sample frequency, and number of samples were set as 1×10^7, 1×10^3 and 1×10^4, respectively. Output was visualized by FIGTREE (FigTree, RRID:SCR_008515) (v1.4.4) (http://tree.bio.ed.ac.uk/software/figtree/).

CAFE (v5) (CAFE, RRID:SCR_005983) [50] was performed to analyze the expansion and contraction of gene families with the estimated phylogenetic tree information. Gene families with *P*-value ≤ 0.05 were referred as significantly expanded or contracted ones. The significantly changed gene families were then used for GO and KEGG enrichment analysis by CLUSTERPROFILER (v4.0.5) (clusterProfiler, RRID:SCR_016884) [51]. The significantly overrepresented GO terms and KEGG pathways were identified with FDR cutoff <0.05.

2.5. Synteny analysis

To identify the synteny relationships between *O. denselamellosa, C. gigas, C. ariakensis and C. virginica,* we applied the method used in the research of European flat oyster genome [52], the longest coding DNA sequences (CDS) for each gene were identified in these four species based on MCscanX in the JCVI toolkit (jcvi, RRID:SCR_021641) [53]. The corresponding chromosome was determined by the fraction of genes in a block of approximate 25 genes.

2.6. Analysis of positively selected genes by using branch-site model

A total of 6737 single-copy orthologues within a phylogenetic tree were identified among *O. denselamellosa*, *C. gigas*, *S. glomerata*, *C. ariakensis* and *C. virginica* using ORTHOFINDER. ParaAT [54] was used to align the protein sequences and translate them to coding sequence (cds). *O. denselamellosa* was designed as the "foreground" branch-site and the rest as the "background" (*C. gigas*, *S. glomerata*, *C. ariakensisand* and *C. virginica*). CODEML from PAML package was used to estimate the rate of synonymous and non-synonymous substitutions. The alternative hypothesis was set as follows: model = 2, NSsite = 2 and fix_omega = 0 and null hypothesis: model = 2, NSsite = 2, fix_omega = 1 and omega = 1. The wag model was employed as the substitution model. Likelihood values were calculated using Chi-square distribution in R, only genes with corrected P-value ≤ 0.01 were considered to be under positive selection or evolving rapidly.

2.7. Hox and ParaHox gene analysis

Hox and *ParaHox* genes are key regulators in body patterning and tissue segmentation during development [55,56]. To investigate their characteristics in *O. denselamellosa*, *Hox* and *ParaHox* gene clusters were systematically analyzed among ten molluscs. The *Hox* and *ParaHox* genes were identified in the *O. denselamellosa* genome using BLAST with an *E*-value threshold of $1e^{-10}$ against all *Hox* and *ParaHox* genes from the HomeoDB database (http://homeodb.zoo.ox.ac.uk/) (Table S3) [57]. The results were further confirmed by comparing to the Conserved Domains Database (http://www.ncbi.nlm.nih.gov/cdd). The same approach was also used to identify *Hox* and *ParaHox* genes in other Mollusca genomes.

2.8. Transcriptome analysis

The RNA-Seq results were first mapped to *O.denselamellosa* genome by HISAT2 (v2–2.2.1) (HISAT2, RRID:SCR_015530) [58]. SAMTOOLS (v1.9) (SAMTOOLS, RRID:SCR_002105) [59] was used to change the format and sort the results. The quantity of transcriptomes was detected by FEATURECOUNTS (v2.0.1) (featureCounts, RRID:SCR_012919) [60]. The result was standardized using TPM and TMM in order to balance the differences between tissues and individuals. Differential gene expression was analyzed using DESEQ2 R package (DESeq2, RRID:SCR_015687) [61] based on unstandardized gene count result. Gene expression with P. adjc <0.05 and absolute value of log2foldchange > 1 was considered as significantly different expressed genes.

3. Results and discussion

3.1. High-quality genome assembly

The genome of *O. denselamellosa* was generated via the combination of second-generation and third-generation sequencing technologies. As shown in Table S4, a total of 28.52 Gb PacBio hifi reads (~47.22×) with an average length of 14,458 bp were generated. This is above the general requirement for sequencing coverage for assembly of a common diploid genome. The clean data of Illumina pair-end reads and Hi-C reads were approximately 30.09 Gb (~49.82× coverage) and 70.2 Gb (~116.23× coverage), respectively. After error-correction and trimming, the high-quality Illumina pair-end reads were used to estimate the genome size. Based on JELLYFISH and PLATANUS analyses, the genome size was estimated to be ~576 Mb and ~ 606 Mb, respectively, with high heterozygosity of 0.801% (Fig. S3). In addition, we also used MASURCA to estimate genome size which gave an ~620 Mb estimation, slightly bigger than the estimate generated by the kmer analysis. An initial 636.09 Mb genome assembly was obtained by HIFIASM, and 95.6% of which were anchored onto 10 chromosomes (the same chromosomes numbers as other oysters) (Fig. 2A, Table 1). The chromosome size ranges from 33.26 to 81.40 Mb (Fig. 2B, Fig. S4, and Table S5). The scaffold N50 was 71.80 Mb, the density of CG fluctuates over a certain range with GC content 35.14% (Fig. 2B). This is similar to other molluscs (33.31%–36.58%) (Table S6). We used BUSCO to evaluate the accuracy and completeness of *O. denselamellosa* genome by searching against metazoan single-copy orthologues. Overall, 937 out of 954 (98.2%) complete genes were identified using BUSCO with 0.4% duplicate ratio represented in the assembled genome (Table S7). Together, these results indicate that the *O. denselamellosa* genome data were of high quality and sufficient for subsequent analyses.

Table 1
Basic information of the *O. denselamellosa* genome.

Characteristic	HIFIASM
Total length	636,086,909 bp
GC (%)	35.14%
Number of chromosomes	10
Number of scaffolds	218
Number of contigs	138
Scaffold N50	71,800,000 bp
Contig N50	13,952,798 bp
Anchored (%)	95.60%
N's per 100 k	30.71 bp

3.2. Genome annotation

Repetitive DNA elements including transposable elements (TEs), are one of the driving forces in shaping genomic architecture and evolution [62–64]. A total of 321.26 Mb repetitive sequences were identified in *O. denselamellosa* genome using ab initio prediction and homologous search. Togehter, they account for 50.51% of the *O. denselamellosa* genome (Table 2). Among the repetitive sequences, the main component was TEs, representing approximately 43.46% of the whole genome. Retroelements ranked the major type of TEs with 18.55%.

Long interspersed nuclear element (LINE) and short interspersed nuclear element (SINE) represented approximately 13.04% and 3.68% of *O. denselamellosa* genome, respectively, which was much more than those in other oysters' genomes (Fig. 3AB and Table S8). The divergence distribution of repeat sequences indicated that LINE went through a burst once in *O. denselamellosa* genome (Fig. 3C). This may be the reason why LINE made up a significantly higher proportion of *O. denselamellosa* genome than that of other oysters'. And such expansion of transposable elements might result from lineage-specific evolutionary events.

The Long Terminal Repeats (LTR) retrotransposons elements play

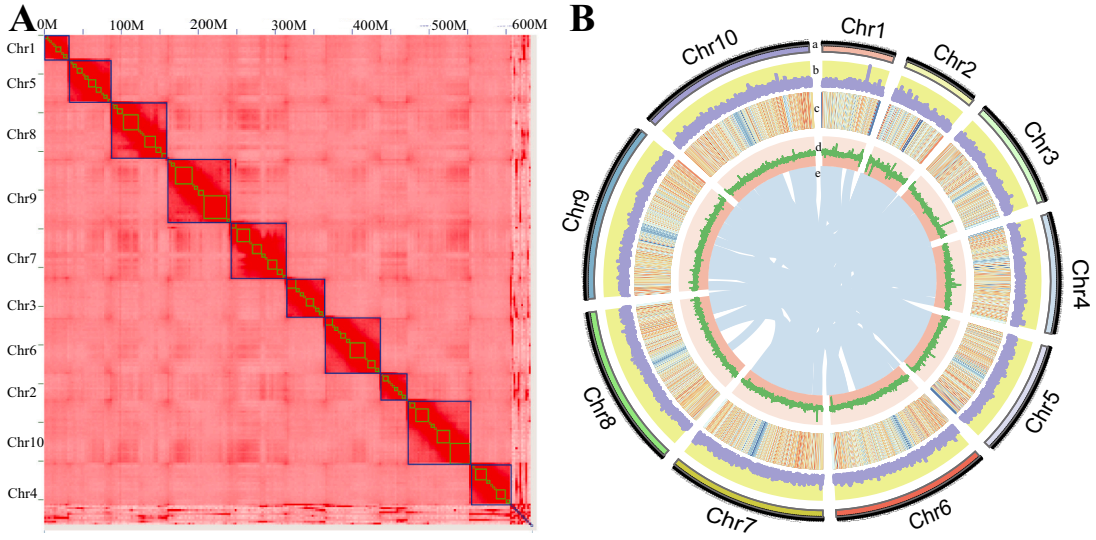

Fig. 2. Characterization and features of the *O. denselamellosa* genome.
A. A genome-wide contact matrix of *O. denselamellosa* from Hi-C data. The blue squares represent the ten chromosomes, and the green squares represented the scaffolds. It can be seen that the assembly was anchored to 10 chromosomes, and the fragment in the lower right represents the unanchored sequences. B. Characterization and features of the *O. denselamellosa* genome. a, the ten chromosomes; b, gene density; c, transposable element (TE) density; d, guanine-cytosine (GC) density and e, synteny block. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Repetitive elements in *O. denselamellos*a genome.

Repeat Class	No. elements	Total Length (Mb)	Percentage sequence (%)
Retroelemen	410,557	118.005389	18.55
SINE	138,557	23.395921	3.68
LINE	238,445	82.916912	13.04
LTR element	33,555	11.692556	1.84
DNA transposon	403,284	90.426744	14.22
Unclassified	368,460	68.031071	10.7
Total repeats	–	276.463204	43.46
Rolling-circle	172,664	43.5723	6.85
Other	2803	0.28345	0.04
Small RNA	0	0	0
Satellites	4323	1.226663	0.19
Simple repeats	0	0	0
Low complexity	0	0	0
Total	–	321.262167	50.51

important role in the genome evolution [65]. LINE function has been studied in human and mice, as well as flies. It has been implicated in chromosome inversions and evolution of new genes [66–68]. We speculate that LINE likely plays a role in oysters genome expansion and structure organization, but the correctness of the assumptions and the specific mechanism still need to be studied in the future. Sequence analysis detected 233, 701, 469, 516 and 419 LTRs in the genome of *O. denselamellosa*, *C. gigas*, *C. virginica*, *C. ariakensis* and *S. glomerata*, respectively. By using the divergence between a pair of LTRs at both ends, we were able to estimate the insertion time of LTRs in the four oysters. Unlike *Crassostrea*, the LTRs of the *O. denselamellosa* and *S. glomerata* had suddenly decrease at about 0.3 Mya (Fig. 3D). Due to the lack of ancient genome references, the insert locations of the expanded LTR in oysters and the effects on evolution and gene function have yet to be investigated.

A total of 26,412 protein-coding genes were predicted by comparing with the transcriptome data (RNA-Seq and Iso-Seq), (Table S9). The average size of the protein-coding gene is 11,845.5 bp. Based on the completeness assessment of the genome assembly, 95.7% complete BUSCOs were found in 954 metazoan gene set (Fig. S5 and Table S10). Functional annotation using EGGNOG databases showed that 62.45% (16,495) of the genes had homologues in the *O. denselamellosa* genome (Table S11). In total, 6794 KEGG pathways were annotated. Of which, signal transduction, infectious diseases and endocrine system were on the top (Fig. S6 and Table S12). In addition, 85.7% of the genes (22,636) could be annotated with at least one public database (Fig. S7).

3.3. Phylogenetic analyses and genetic differences during evolution

3.3.1. Phylogenetic analysis and divergence time estimation

To examine the evolutionary relationships between *O. denselamellosa* and other Lophotrochozoa, we analyzed protein coding sequences from 15 species. We identified 396,456 genes in 28,399 orthologs using ORTHOFINDER. 444 of which were single-copy orthologs (Fig. 4A). A phylogenetic and divergence time tree was constructed based on the single-copy orthologs. The data showed that *O. denselamellosa* and *S. glomerata* diverged from the recent common ancestor at around 86.28 Ma (in Cretaceous) (Fig. 4B). Our data also suggested that the divergence time of *Ostrea* and *Saccostrea* and *Crassostrea* was around 104.13 Ma (Fig. 4B), while Bivalvia and Gastropoda diverged around 538.44 Ma. Our estimation is consistent with findings from previous studies [69].

Three marine hypoxia events occurred in Cretaceous. One of which was the oceanic anoxic event 3 (OAE3) occurred at 84.6–87.3 Ma [70]. Interestingly, OAE3 was considered as an important turning point in the evolution of the Cretaceous climate system, representing a shift from Cretaceous greenhouse climate to Paleocene cooling [71]. This coincided with the appearance of mammals with placenta in Cretaceous. Thus, we hypothesize that the unique live-bearing reproductive style of *Ostrea* might be related to improving survival in an oxygen-starved and cold environment.

Next, we performed a gene family cluster analysis of *O. denselamellosa* and four other oyster species (*S. glomerata*, *C. gigas, C. virginica* and *C. ariakensis*). A core set of 12,119 gene families were identified (Fig. 4C). *S. glomerata* shared considerably more unique gene families (464) with *O. denselamellosa*. In contrast, *C. ariakensis* shared the least with them (87). In addition, we found that there were fewer lineage-specific gene families (344) in *O. denselamellosa* than in other oyster species. It may mean that *Ostrea* has a lower evolutionary rate in some genomic features compared with *Crassostrea* and *Saccostrea*.

3.3.2. Synteny analysis and genome evolution

The analysis revealed one to one correspondence between the chromosomes of each two species among *O. denselamellosa, C. gigas,* and *C. ariakensis.* Highly syntenic chromosomes were identified. These include *C. ar*chr9- *C. gi*chr4, *C. ar*chr8- *C. gi*chr1, *C. ar*chr7- *C. gi*chr10 and *C. ar*chr5- *C. gi*chr7, *O. de*chr6- *C. ar*chr8, *O. de*chr7- *C. ar*chr7, *O. de*chr10- *C. ar*chr5, *O. de*chr6- *C. gi*chr1, *O. de*chr7- *C. gi*chr10 and *O. de*chr10- *C. gi*chr7 (marked by red stars in Fig. 4D and Fig. S8). Interestingly, some intrachromosomal inversions were found in some of the homologous chromosomes of *C. gi*chr8 - *O. de*chr3, *C. gi*chr4 - *O. de*chr4, *C. gi*chr5 - *O. de*chr5, *C. gi*chr3 - *O. de*chr8, *C. gi*chr6 - *O. de*chr9 and *C. ar*chr9 - *O. de*chr4, *C. ar*chr6 - *O. de*chr8, *C. ar*chr4 - *O. de*chr9 (marked by black arrows in Fig. S8AB), which may be attributed to the chromosome inversion between *Ostrea* and *Crassostrea*. In addition, despite the high degree of collinearity of most chromosomes, the genomes of *O. denselamellosa* and *C. virginica* had undergone some complex interchromosomal rearrangements. This is mostly shown in *C. vi*chr5 – *O. de*chr4, *C. vi*chr6 – *O. de*chr4, *C. vi*chr10 – *O. de*chr2 and *C. vi*chr9 – *O. de*chr2. Similar rearrangements also occurred comparing *C. ariakensis* and *C. virginica* (marked by green boxes in Fig. 4D) chromosomes, which may suggest that chromosome rearrangement in *C. virginica* during evolution. However, this phenomenon could also be caused by the low quality of *C. virginica* genome sequence data.

Hox and *ParaHox* genes are a class of highly conserved transcription factors in the animal genomes, playing crucial roles in the early development and patterning of body axis and segment identity in metazoans [72–74]. The genomic organization of *Hox* and *ParaHox* genes varies considerably among Mollusca [75–77]. However, the *Hox* and *ParaHox* gene has been extensively studied in model animals, relatively little is known about these genes and their function in Mollusca. Mollusca have a high degree of morphological and evolutionary diversity, and preliminary information of molluscan *Hox* and *ParaHox* gene may be helpful for understanding and explaining the evolution of their morphology or behavior in some extent [78]. In this study, we identified 10 *Hox* genes and 3 *ParaHox* genes in the *O. denselamellosa* genome (Table S13). Similar findings were also observed in other oyster genomes. RNA-seq expression analysis of *Hox* and *ParaHox* genes reveals some tissue-specific expression patterns in *O. denselamellosa* (Fig. S9). *Post1* and *Post2* are extremely high-expressed in adductor. Meanwhile, *Hox* and *Lox* expressed more in mantle compared to other tissues.

All of the *Hox* genes in *O. denselamellosa* genome were clustered together on chromosome 6, which is the same as that in *C. virginica*, *C. gigas* and *M. yessoensis* [12,69]. However, these clusters were broken into three sections in the *S. glomerata* genome. In addition, we also found that *Antennapedia* (*Antp*) gene was not present in *O. denselamellosa* genome (Fig. 4F), which is consistent with previous findings in *C. hongkongensis*, *C. virginica* and *S. glomerata* [10,69,79]. This provides further support of *Antp* as a potential driver in byssus formation [80].

3.4. Gene family analysis

By comparing *O. denselamellosa* genome with other 14 Lophotrochozoa by CAFE, we found that 342 and 769 gene families were expanded or contracted respectively. Among them, 72 gene families

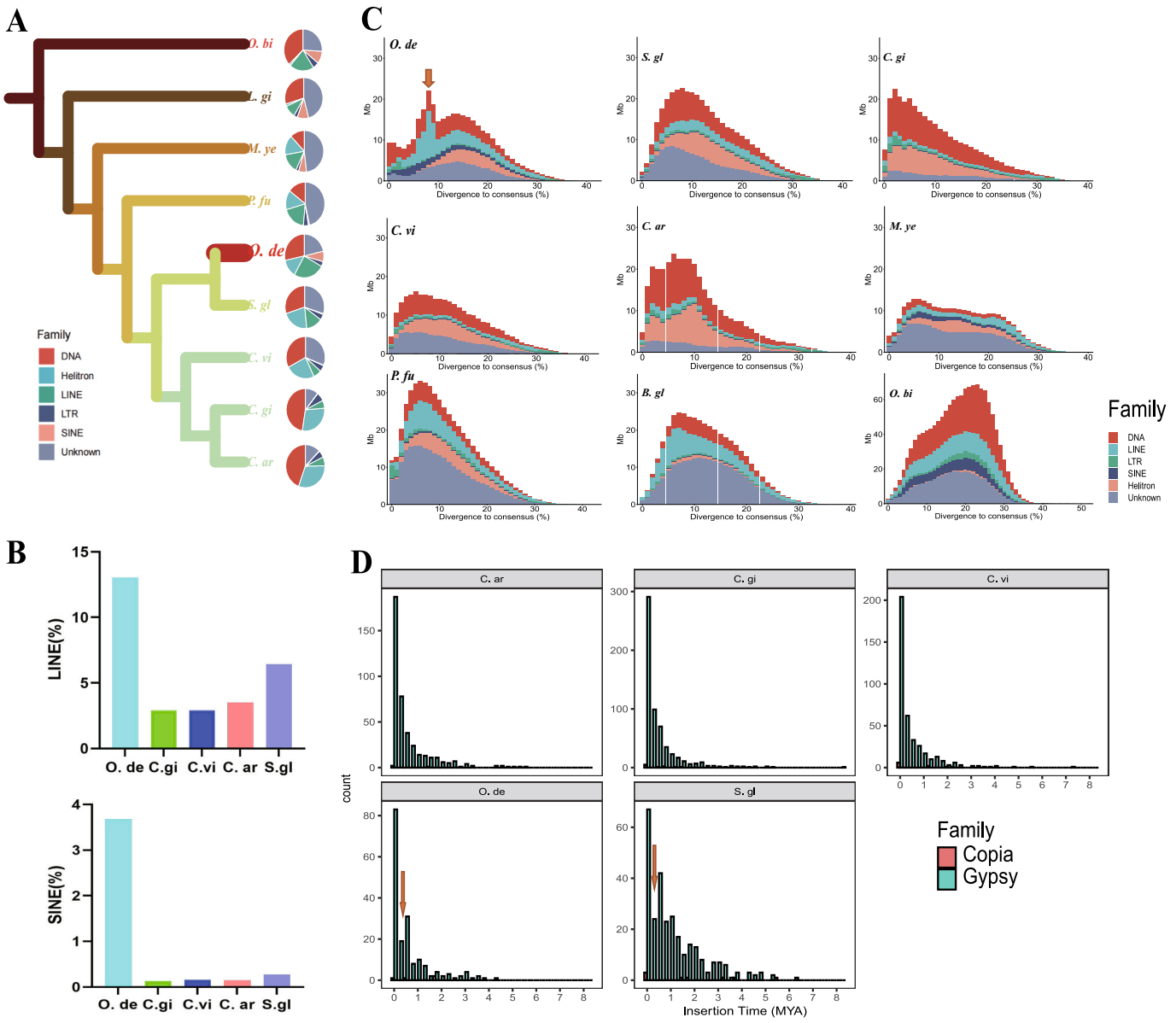

Fig. 3. Repeat element annotation of *O. denselamellosa* and other Mollusca.
A. Distribution of transposon elements (TEs) in the genome of ten oyster species. The different colors of the branches represent different genus. The results were calculated by REPEATMODELER and RepeatMasker (the detail data is available in Table S8). Phylogenetic tree was constructed by FASTTREE based on the protein sequences of each species. LINEs, long interspersed elements; SINEs, short interspersed elements. *C.ar*, *C. ariakensis*; *C. gi*, *C. gigas*; *C. vi*, *C. virginica*; *O. de*, *O. denselamellosa*; *S. glo*, *S. glomerata*; *P. fu*, *P. fucata martensii*; *M. ye*, *M. yessoensis*; *L. gi*, *L. gigantea*; *O. bi*, *O. bimaculoides*. B. Comparison of the proportions of SINE, LINE in the genomes of different species. C. Comparison of the transposable elements among Mollusca. The arrow indicates LINE burst once in *O. denselamellosa* genome. The x-axis indicates divergence to consensus of transposable elements in the genome, a higher number indicateed a longer period of time before present. The y-axis indicates the total and individual lengths (Mb) of transposable elements in each species. D. Count and insertion time of LTRs of five different oysters. The arrows indicate two significant declines of LTR insertion.

have undergone significant expansion, whereas 172 families were significant contracted. Such expansions and contractions are less than those in *S. glomerata* (1381 expanded and 805 contracted gene families) which is phylogenetically closest to *O. denselamellosa* compared with other Lophotrochozoa. Since gene family expansion and contraction events may accelerate gene evolution and elevate evolutionary rates [81], our results supported the idea that *O. denselamellosa* is slow evolving species in some extent.

GO analysis indicated that the expanded gene families in *O. denselamellosa* were significantly enriched (FDR cutoff <0.05) in multi-organism cellular process, dsRNA and caveola (Table S14, Table S15, and Fig. S10ab). In addition, the contracted gene families were mainly enriched in golgi cisterna membran , alpha-(1,2)-fucosyltransferase activity and carbohydrate binding (Table S16, S17, and Fig. S10cd). Moreover, among members of the expanded gene families, 10 GO terms and 1 KEGG pathway (Ko04510) correlated with adhesion function, 5 GO terms linked to cilia function and 13 GO terms in regulation of hormones during reproduction (Fig. 5A). It has been suggested that cilia may play a role in ovoviviparity [82]. Cilia on the surface of the outer gill flaps (where the larvae would be exposed) were found to be more developed than the inner gill flaps in the pregnant live-bearing Mollusca *Anodonta woodina* [82]. In addition, the evolution of live-bearing requires prolonged egg retention until embryogenesis is complete [83]. Steroid hormones (GO:0043401) were shown to be involved

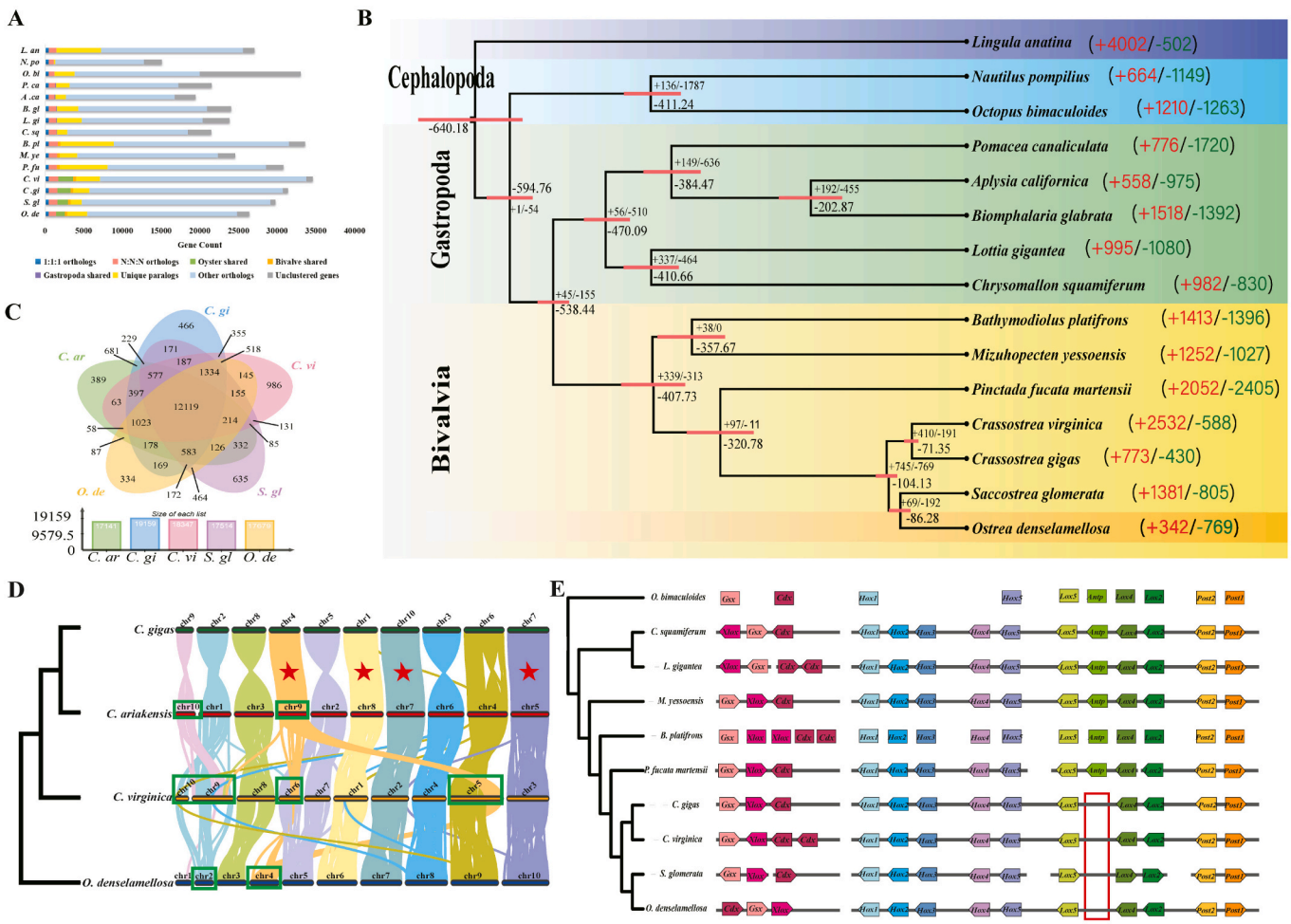

Fig. 4. Evolutionary analysis of *O. denselamellosa*.
A. Categories of orthologous and paralogous genes in 15 Lophotrochozoa based on their longest protein sequences. B. Phylogenetic tree of *Lophotrochozoa* based on 444 shared single-copy orthologous. Reference-calibrated time points were shown in Table S2. The estimated divergence time (MYA: million years) is shown at each branch point. Each lineage is shown in different colour blocks, the pink bars represent 95%HPD. *O. denselamellosa* is highlighted in dark yellow. Red and green indicate gene family expansion (+) and contraction (−), respectively. C. Venn diagram of 5 different oyster homologous genes via their longest protein sequences. D. Syntenic analysis via comparisons the longest CDS of *O. denselamellosa* with *C. gigas, C. ariakensis* and *C. virginica*. Different colors represent blocks on different chromosomes. Red stars indicate highly syntenic chromosomes, green boxes indicate interchromosomal rearrangements. E. *Hox* and *ParaHox* gene cluster locus in the genome of *O. denselamellosa* and other molluscs. The arrow direction indicates the transcription orientation, and the continuous gray line represents a chromosome. Note: The correspondence of chromosome numbers we used here of *C. gigas* and *C. virginica* to chromosome numbers in their the public genome: For *C. gigas*, chr1- NC47559.1, chr2- NC47560.1, chr3- NC47561.1, chr4- NC47562.1, chr5- NC47563.1, chr6- NC47564.1, chr7- NC47565.1, chr8- NC47566.1, chr9- NC47567.1, chr10- NC47568.1. For *C. virginica*, chr1- NC35780.1, chr2- NC35781.1, chr3- NC35782.1, chr4- NC35783.1, chr5- NC35784.1, chr6- NC35785.1, chr7- NC35786.1, chr8- NC35787.1, chr9- NC35788.1, chr10- NC35789.1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in regulating incubation chamber in lizards [84]. And studies have shown that hormones are constantly changing during sexual maturation and spawning in *C. gigas* [85]. Futhermore, some study in live-bearing lizard and syngnathid fishes highlighted the potential roles of adhesion (such as: GO:2000047) in embryo attachment [84,86]. Because studies on hormones and cell adhesion have very limited genetic basis in Molluscan, here, we are only inferring that genes related to cilia, hormones and cell adhesion described above may contribute to the evolution of reproductive patterns in the *O. denselamellosa*.

Beacuse the lineage-specific genes were thought might provide some potential candidates for further analysis in the process of evolution in *O. denselamellosa*, we carried out enrichment analysis of the 334 lineage-specific gene families and compared them with 4 other oyster species containing 3006 genes. GO enrichment analysis showed that most genes were enriched (FDR cutoff <0.05) in neuron (GO:0021884, GO:0021879, GO:0021872, GO:0021954), behavior (GO:0035641, GO:0035640, GO:0007626), response to light stimulus (GO:0050953, GO:0009416) (Fig. 5B, Fig. S11e and Table S18). All of these are associated with the quick response to the changeable environment. Moreover, we identified 20 enriched KEGG pathways (FDR cutoff <0.05). Three of them were primarily involved cell adhesion (ko04514), B cell receptor signaling (ko04662) and steroid hormone biosynthesis (ko00140) (Fig. 5C and Table S19), respectively. Immune-related gene families have been previously described in *Ostrea edulis* genomes [52,87,88] which play an important role in adaptation to harsh environments and defence against pathogens. Unlike most oysters, *O. denselamellosa* lives in the subtidal zone with high salinity, and these immune-related genes which are unique to the *O. denselamellosa*, may help the it adapt to its unique living environment.

The branch-site model was performed using CODEML to identify gene evolving under positive selection in *O. denselamellosa* compared with four other oyster species. A total of 1122 genes were identified as positive selected ones. However, only six GO terms were significantly enriched (FDR cutoff <0.05). Two of them are, GO:0005818 and

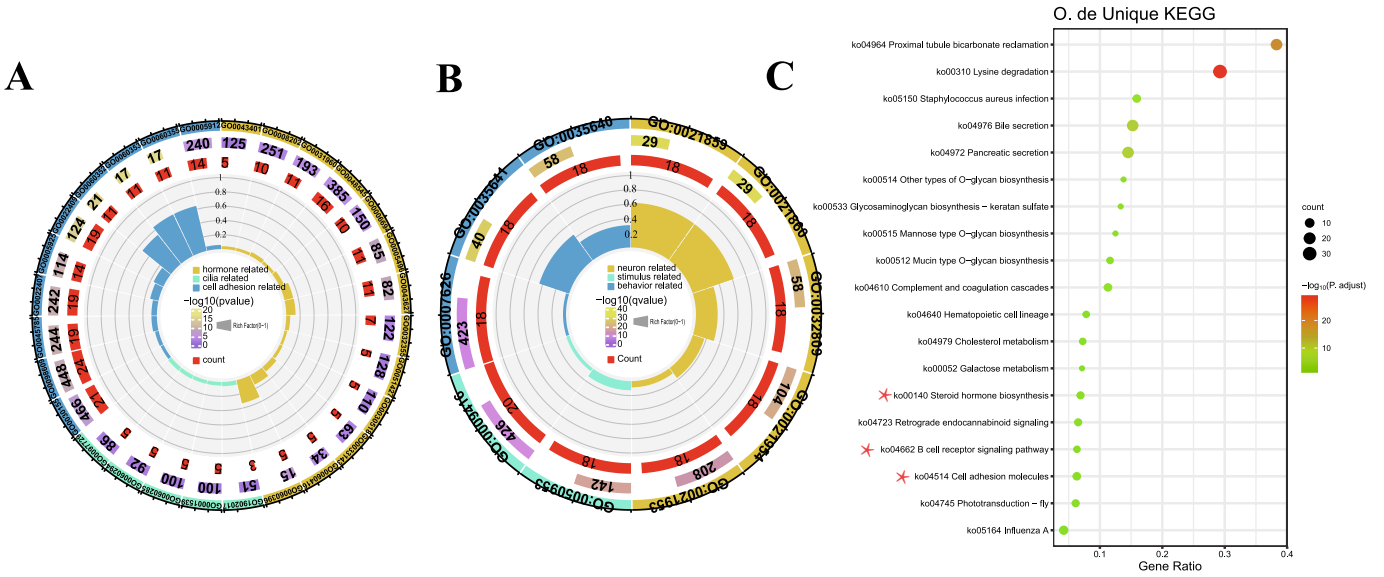

Fig. 5. Gene family enrichment in key protein groups.
A. Enriched GO terms related to hormone, cilia and cell adhesion of genes in significantly expanded gene families of *O. denselamellosa*. The circle, from the outer to the inner, represent: the id of each GO terms, −log10 (pvalue) and gene counts in each GO term, respectively. In the middle of the circle, the length of the sector represents Rich Factor. The detail data is available in Table S14. B. Enriched GO terms related to neuron, stimulus and behavior of genes in lineage-specific gene families of *O. denselamellosa*. The circle, from the outer to the inner, represent: the id of each GO terms, −log10 (pvalue) and gene counts in each GO term, respectively. In the middle of the circle, the length of the sector represents Rich Factor. The detail data is available in Table S18. C. Enriched KEGG Pathways of genes in lineage-specific gene families of *O. denselamellosa*. Red stars marked the pathways related to hormone, cell adhesion and immune system. The X axis represented Gene Ratio, colors represented −log10 (P. adjust), and the circle size represents gene count in each KEGG Pathways. The detail data is available in Table S19. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

GO:0090545 related to microtubules and SNF2-like ATPase domain, respectively (Table S20). *O. denselamellosa* keeps the larvae inside the female mantle cavity for 1–3 days to protect them from external conditions [89]. In addition, dead larvae will be found by the female oysters and ejected from the mantle cavity [90,91]. These two behaviors mentioned above tend to spend considerable energy. Genes related to these GO terms may help to improve the efficiency of metabolism and energy supply in order to meet the need of energy requirement for new larvae or adult to create a better place for larval hatching. Together, these expanded, unique and positive selected gene families may provide important candidates gene for future research on production pattern evolution and adaption in *Ostrea* even in oysters.

4. Conclusions

Here, by integrated analysis of third generation Pacbio long-read sequencing, Hi-C technology, full-length transcriptome and Illumina short reads, we provide a high-quality chromosomal-level genome assembly and annotation of *O. denselamellosa* with scaffold N50 length = 71.80 Mb and 95.7% complete metazoan BUSCOs. Comparative genomic analysis was performed to assess the evolutionary relationships of *O. denselamellosa* and other Lophotrochozoa. We speculated that reproduction evolution of the *Ostrea* may be related to OAE3. In addition, we found that the proportion of LINE and SINE expanded in the genome of *O. denselamellosa* compared with other oysters', which may contribute to its specific evolution. The divergence time of *O. denselamellosa* was calculated, and the geological event related to the adaptive evolution of *O. denselamellosa* was predicted. Furthermore, after analyzing the expanded, unique gene families and positive selected genes, GO and KEGG enrichment may provide *O. denselamellosa* with a genomic basis for the evolution of live-bearing production, energy metabolism, stress ability and immunity. The study will generate valuable genomic and genetic resources for studying the evolution in *Ostrea* and for breeding and protection of live-bearing oysters.

Author statement

All co-authors contributed to the design of the project and procedures. Z. D. prepared RNA for sequencing, performed genome assembly, annotation, comparative genomics and comparative transcriptome analysis. Z.D. also contributed to the initial manuscript draft and revision. Y. B. assisted in completing the analysis of the sequencing data. S. L. guided the sequencing data analysis. H. Y. and L. K. collected the oyster samples and identified their gonadal development. S. D. contributed to the manuscript revision. Q. L. guided the whole experiment process, edited the manuscript and offered financial support. All authors have read and approved the final manuscript.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Data availability

The Ostrea denselamellosa genome assembly and annotation data have been deposited in FigShare (https://doi.org/10.6084/m9.figshare.19801705) and Ostrea denselamellosa genome assembly could also be found at NCBI with SAMN28421488. Hi-C data, PacBio hifi data, Illumina paired-end data (SAMN28406680), full-length transcriptome data (SAMN28406681) and RNA-seq data (SAMN28406682-SAMN28406692) that support this study have been submitted to NCBI SRA database with project accession number PRJNA838121.

Acknowledgements

This work was supported by grants from the China Agriculture Research System Project (CARS-49), National Natural Science Foundation of China (31972789), and Earmarked Fund for Agriculture Seed Improvement Project of Shandong Province (2020LZGC016 and 2021LZGC027).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2023.110582.

References

[1] F. Xu, S. Zhang, An Illustrated Bivalvia Mollusca Fauna of China Seas, Science Press, Beijing, 2008. ISBN 978-7-03-021339-6. (in Chinese).
[2] L. Chen, Q. Li, Q.Z. Wang, L.F. Kong, X.D. Zheng, Techniques of artificial breeding of the oyster Ostrea denselamellosa, Periodical of Ocean Univiversity of China 41 (43), 2011, 43–6. (in Chinese, with English abstract).
[3] N.E. Buroker, Evolutionary patterns in the family Ostreidae: Larviparity vs. oviparity, J. Exp. Mar. Biol. Ecol. 90 (3) (1985) 233–247.
[4] D.O. Foighil, D.J. Taylor, Evolution of parental care and ovulation behavior in oysters, Mol. Phylogenet. Evol. 15 (2) (2000) 301–313.
[5] M.H. Yang, H.S. Kim, J.Y. Lee, C.H. Han, Artificial mass culture of flat oyster larvae, Ostrea denselamellosa, and collection rates according to various spat collection methods, Korean J. Malacol./ 17 (1) (2001) 35–44.
[6] H. Yu, L. Kong, Q. Li, Complete mitochondrial genome of Ostrea denselamellosa (Bivalvia, Ostreidae), Mitochondrial DNA Part A 27 (1) (2016) 711–712.
[7] A. Insua, C. Thiriot-Quievreux, The characterization of Ostrea denselamellosa (Mollusca, Bivalvia) chromosomes: karyotype, constitutive heterochromatin and nucleolus organizer regions, Aquaculture. 97 (4) (1991) 317–325.
[8] H. Qi, L. Li, G. Zhang, Construction of a chromosome-level genome and variation map for the Pacific oyster Crassostrea gigas, Mol. Ecol. Resour. 21 (5) (2021) 1670–1685.
[9] C. Penaloza, A. Gutierrez, L. Eory, S. Wang, X.M. Guo, A.L. Archibald, T.P. Bean, R. D. Houston, A chromosome-level genome assembly for the Pacific oyster Crassostrea gigas, GigaScience. 10 (3) (2021) 1–9.
[10] J. Peng, Q. Li, L. Xu, P. Wei, P. He, X. Zhang, L. Zhang, J. Guan, X. Zhang, Y. Lin, J. Gui, X. Chen, Chromosome-level analysis of the Crassostrea hongkongensis genome reveals extensive duplication of immune-related genes in bivalves, Mol. Ecol. Resour. 20 (4) (2020) 980–994.
[11] Y. Li, W.Y. Nong, T. Baril, H.Y. Yip, T. Swale, A. Hayward, D.E.K. Ferrier, J.H. L. Hui, Reconstruction of ancient homeobox gene linkages inferred from a new high-quality assembly of the Hong Kong oyster (Magallana hongkongensis) genome, BMC Genomics 21 (2020) 713.
[12] M. Gomez-Chiarri, W. Warren, X. Guo, D. Proestou, Developing tools for the study of molluscan immunity: the sequencing of the genome of the eastern oyster, *Crassostrea virginica*, Fish & Shellfish Immunol. 46 (1) (2015) 2–4.
[13] A. Li, H. Dai, X. Guo, Z. Zhang, K. Zhang, C. Wang, W. Wang, H. Chen, X. Li, H. Zheng, L. Li, G. Zhang, Genome of the estuarine oyster provides insights into climate impact and adaptive plasticity, Commun. Biol. 4 (2021) 1287.
[14] B. Wu, X. Chen, M. Yu, J. Ren, J. Hu, C. Shao, L. Zhou, X. Sun, T. Yu, Y. Zheng, Y. Wang, Z. Wang, H. Zhang, G. Fan, Z. Liu, Chromosome-level genome and population genomic analysis provide insights into the evolution and environmental adaptation of Jinjiang oyster Crassostrea ariakensis, Mol. Ecol. Resour. 22 (4) (2021) 1529–1544.
[15] D. Powell, S. Subramanian, S. Suwansa-ard, M. Zhao, W. O'Connor, D. Raftos, A. Elizur, The genome of the oyster Saccostrea offers insight into the environmental resilience of bivalves, DNA Res. 25 (6) (2018) 655–665.
[16] S.F. Chen, Y.Q. Zhou, Y.R. Chen, J. Gu, Fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics. 34 (17) (2018) 884–890.
[17] G. Marcais, C. Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers, Bioinformatics. 27 (6) (2011) 764–770.
[18] R. Kajitani, D. Yoshimura, M. Okuno, Y. Minakuchi, H. Kagoshima, A. Fujiyama, K. Kubokawa, Y. Kohara, A. Toyoda, T. Itoh, Platanus-allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions, Nat. Commun. 10 (2019) 1702.
[19] T.R. Ranallo-Benavidez, K.S. Jaron, M.C. Schatz, GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nature, Communications. 11 (1) (2020) 1432.
[20] A.V. Zimin, M. Guillaume, P. Daniela, R. Michael, S.L. Salzberg, J.A. Yorke, The MaSuRCA genome assembler, Bioinformatics. 29 (21) (2013) 2669–2677.
[21] H.Y. Cheng, G.T. Concepcion, X.W. Feng, H.W. Zhang, H. Li, Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm, Nat. Methods 18 (2) (2021) 170–175.
[22] D.F. Guan, S.A. McCarthy, J. Wood, K. Howe, Y.D. Wang, R. Durbin, Identifying and removing haplotypic duplication in primary genome assemblies, Bioinformatics. 36 (9) (2020), 2896–8.
[23] H. Li, R. Durbin, Fast and accurate short read alignment with burrows-wheeler transform, Bioinformatics. 25 (14) (2009) 1754–1760.
[24] N.C. Durand, M.S. Shamim, I. Machol, S.S.P. Rao, M.H. Huntley, E.S. Lander, E. L. Aiden, Juicer provides a one-click system for analyzing loop-resolution hi-C experiments, Cell Syst. 3 (2016) 95–98.
[25] O. Dudchenko, S.S. Batra, A.D. Omer, S.K. Nyquist, M. Hoeger, N.C. Durand, E. L. Aiden, De novo assembly of the Aedes aegypti genome using hi-C yields chromosome-length scaffolds, Science. 356 (6333) (2017) 92–95.
[26] F.A. Simao, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, Bioinformatics. 31 (19) (2015) 3210–3212.
[27] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies, Bioinformatics. 29 (8) (2020) 1072–1075.
[28] A.F. Smit, R. Hubley, RepeatModeler Open-1.0, Retrieved from, http://www.repeatmasker.org, 2008.
[29] W.D. Bao, K.K. Kojima, O. Kohany, Repbase update, a database of repetitive elements in eukaryotic genomes, Mob. DNA 6 (2015) 11.
[30] M. Tarailo-Graovac, N. Chen, Using RepeatMasker to identify repetitive elements in genomic sequences, Curr. Protoc. Bioinformatics 25 (1) (2009), 4.10.1–4.10.14.
[31] G. Benson, Tandem repeats finder: a program to analyze DNA sequences, Nucleic Acids Res. 27 (2) (1999) 573–580.
[32] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics. 26 (6) (2010) 841–842.
[33] D. Ellinghaus, S. Kurtz, U. Willhoeft, LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons, BMC Bioinforma. 9 (2008) 18.
[34] B.L. Cantarel, I. Korf, S.M.C. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. S. Alvarado, M. Yandell, MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes, Genome Res. 18 (1) (2008) 188–196.
[35] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-seq data without a reference genome, Nat. Biotechnol. 29 (7) (2011) 644–652.
[36] B.J. Haas, A.L. Delcher, S.M. Mount, J.R. Wortman, J.R.K. Smith, L. Hannick, M. Rama, C.M. Ronning, D.B. Rusch, C.D. Town, Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies, Nucleic Acids Res. 31 (19) (2003) 5654–5666.
[37] L.M. Fu, B.F. Niu, Z.W. Zhu, S.T. Wu, W.Z. Li, CD-HIT: accelerated for clustering the next-generation sequencing data, Bioinform. Oxford 28 (23) (2012) 3150–3152.
[38] K.J. Hoff, S. Lange, A. Lomsadze, M. Borodovsky, M. Stanke, BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS, Bioinformatics. 32 (5) (2015) 767–769.
[39] M. Stanke, A. Tzvetkova, B. Morgenstern, AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome, Genome Biol. 7 (2006) S11–S18.
[40] J. Huerta-Cepas, K. Forslund, L.P. Coelho, D. Szklarczyk, L.J. Jensen, C. Mering, P. Bork, Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper, Mol. Biol. Evol. 34 (8) (2017) 2115–2122.
[41] J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernandez-Plaza, S.K. Forslund, H. Cook, D.R. Mende, I. Letunic, T. Rattei, L.J. Jensen, C. Mering, P., Bork, eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses, Nucleic Acids Res. 47 (D1) (2019) D309–D314.
[42] B. Boeckmann, A. Bairoch, R. Apweiler, M.C. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout, M. Schneider, The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003, Nucleic Acids Res. 31 (1) (2003) 365–370.
[43] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, Nat. Methods 12 (1) (2015) 59–60.
[44] D.M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics, Genome Biol. 20 (1) (2019) 238.
[45] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (5) (2004) 1792–1797, https://doi.org/10.1093/nar/gkh340.
[46] S. Capella-Gutierrez, J.M. Silla-Martinez, T. Gabaldon, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses, Bioinformatics. 25 (15) (2009) 1972–1973.
[47] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics. 30 (9) (2014) 1312–1313.
[48] Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood, Mol. Biol. Evol. 24 (8) (2007) 1586–1591.
[49] S. Kumar, G. Stecher, M. Suleski, S.B. Hedges, Time tree: a resource for timelines, timetrees, and divergence times, Mol. Biol. Evol. 34 (7) (2017) 1812–1819.
[50] T. De Bie, N. Cristianini, J.P. Demuth, M.W. Hahn, CAFE: a computational tool for the study of gene family evolution, Bioinformatics. 22 (10) (2006) 1269–1271.
[51] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS- J. Integr. Biol. 16 (5) (2012) 284–287.
[52] X. Li, Y. Bai, Z. Dong, C. Xu, S. Liu, H. Yu, L. Kong, Q. Li, Chromosome-level genome assembly of the European flat oyster (*Ostrea edulis*) provides insights into its evolution and adaptation, in: Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics, 2022 (in press).
[53] Y. Wang, H. Tang, J.D. Debarry, X. Tan, J. Li, X. Wang, T. Lee, H. Jin, B. Marler, H. Guo, J.C. Kissinger, A.H. Paterson, MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity, Nucleic Acids Res. 40 (7) (2012), e49, https://doi.org/10.1093/nar/gkr1293.
[54] Z. Zhang, J.F. Xiao, J.Y. Wu, H.Y. Zhang, G.M. Liu, X.M. Wang, L. Dai, ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments, Biochem. Biophys. Res. Commun. 419 (4) (2012) 779–781.
[55] J.C. Pearson, D. Lemons, W. Mcginnis, Modulating Hox gene functions during animal body patterning, Nat. Rev. Genet. 6 (2005) 893–904.
[56] N.M. Brooke, J. Garcia-Fernandez, P.W. Holland, The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster, Nature. 392 (1998) (1998) 920–922.
[57] Y. Zhong, P.W.H. Holland, HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology, Evol. Dev. 13 (6) (2011) 567–568.
[58] D. Kim, B. Landmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, Nat. Methods 12 (4) (2015) 357–360.

[59] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map format and SAMtools, Bioinformatics. 25 (16) (2009) 2078–2079.
[60] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics. 30 (7) (2014) 923–930.
[61] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (12) (2014) 550.
[62] N.V. Fedorof, Transposable elements, epigenetics, and genome evolution, Science. 338 (6108) (2012) 758–767.
[63] T. Wicker, F. Sabot, Van A. Hua, J.L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. Sanmiguel, A.A. Schulman, A unifed classifcation system for eukaryotic transposable elements, Nat. Rev. Genet. 8 (2007) 973–982.
[64] E.A. Ritschard, B. Whitelaw, C.B. Albertin, I.R. Cooke, J.M. Strugnell, O. Simakov, Coupled genomic evolutionary histories as signatures of organismal innovations in cephalopods: co-evolutionary signatures across levels of genome organization may shed light on functional linkage and origin of cephalopod novelties, Bioessays. 41 (2) (2019), 1900073.
[65] H. Wang, Z. Xu, Identification of LTR retrotransposons in eukaryotic genomes: supports from structure and evolution, Int. J. Bioinforma. Res. Appl. 5 (4) (2009) 365–377.
[66] T.H. Eickbush, A.V. Furano, Fruit flies and humans respond differently to retrotransposons, Curr. Opin. Genet. Dev. 12 (6) (2002) 669–674.
[67] H. Kehrer-Sawatzki, B. Schreiner, S. Tänzer, M. Platzer, S. Müller, H. Hameister, Molecular characterization of the Pericentric inversion that causes differences between chimpanzee chromosome 19 and human chromosome 17, Am. J. Hum. Genet. 71 (2) (2002) 375–388.
[68] A.N. Carnell, J.I. Goodman, The long (LINEs) and the short (SINEs) of it: altered methylation as a precursor to toxicity, Toxicol. Sci. 75 (2) (2003) 229–235.
[69] S. Wang, J.B. Zhang, W.Q. Jiao, J. Li, X.G. Xun, Y. Sun, X.M. Guo, P. Huan, B. Dong, L.L. Zhang, X.L. Hu, X.Q. Sun, J. Wang, C.T. Zhao, Y.F. Wang, D.W. Wang, X. T. Huang, R.J. Wang, J. Lv, Z.M. Bao, Scallop genome provides insights into evolution of bilaterian karyotype and development, Nature Ecology & Evolution. 1 (5) (2017).
[70] Y. Huang, C. Wang, J. Gu, Cretaceous oceanic anoxic events: research progress and forthcoming prospect, Acta Geologica, Science. 82 (1) (2008) 1–10 (in Chinese, with English abstract).
[71] B.T. Huber, R.D. Norris, K.G. MacLeod, Deep-sea paleo-temperature record of extreme warmth during the cretaceous, Geology. 30 (2) (2002) 123–126.
[72] D.E.K. Ferrier, P.W.H. Holland, Ancient origin of the Hox gene cluster, Nat. Rev. Genet. 2 (1) (2001) 33–38, https://doi.org/10.1038/35047605.
[73] S.J. Gaunt, Hox cluster genes and collinearities throughout the tree of animal life, Int. J. Dev. Biol. 62 (11−12) (2018) 673–683, https://doi.org/10.1387/ijdb.180162sg.
[74] P.W.H. Holland, Evolution of homeobox genes, Wiley interdisciplinary reviews, Dev. Biol. 2 (1) (2013) 31–45, https://doi.org/10.1002/wdev.78.
[75] Z. Huang, W. Huang, X. Liu, Z. Han, G. Liu, G.A. Boamah, Y. Wang, F. Yu, Y. Gan, Q. Xiao, X. Luo, N. Chen, M. Liu, W. You, C. Ke, Genomic insights into the adaptation and evolution of the nautilus, an ancient but evolving “living fossil”, Mol. Ecol. Resour. 22 (1) (2022) 15–27, https://doi.org/10.1111/1755-0998.13439.
[76] R.M. Varney, C. Mcdougall Speiser, B.M. Degnan, K.M. Kocot, The iron-responsive genome of the chiton Acanthopleura granulata, Genome Biol. Evol. 13 (1) (2021), https://doi.org/10.1093/gbe/evaa263.
[77] A. Wanninger, T. Wollesen, The evolution of molluscs, Biol. Rev. 94 (1) (2019) 102–115, https://doi.org/10.1111/brv.12439.
[78] S. Wang, Z. Yang, Y. Li, T. Cheng, P. Liu, X. Mi, Z. Bao, Hot genes and their research progresses in Molluscs, Period. Ocean Univ. China 47 (8) (2017) 23–30 (in Chinese, with English abstract).
[79] G. Zhang, X. Fang, X. Guo, L. Li, R. Luo, F. Xu, P. Yang, L. Zhang, X. Wang, H. Qi, Z. Xiong, H. Que, Y. Xie, P.W.H. Holland, J. Paps, Y. Zhu, F. Wu, Y. Chen, J. Wang, J. Wang, The oyster genome reveals stress adaptation and complexity of shell formation, Nature. 490 (7418) (2012) 49–54.
[80] Y. Zhang, F. Mao, S. Xiao, H. Yu, Z. Xiang, F. Xu, J. Li, L. Wang, Y. Xiong, M. Chen, Y. Bao, Y. Deng, Q. Huo, L. Zhang, W. Liu, X. Li, H. Ma, Y. Zhang, X. Mu, Z. Yu, Comparative genomics reveals evolutionary drivers of sessile life and left-right shell asymmetry in bivalves, genomics proteomics bioinformatics, In press (2022), https://doi.org/10.1016/j.gpb.2021.10.005.
[81] F.C. Chen, C.J. Chen, W.H. Li, T.J. Chuang, Gene family size conservation is a good indicator of evolutionary rates, Mol. Biol. Evol. 27 (8) (2010) 1750–1758, https://doi.org/10.1093/molbev/msq055.
[82] N. Cong, L. Yuan, Preliminary scanning electron microscopic observations on mantle and gill lamella of Anodonta woodina Lea, Acta Hydrobiol. Sin. 17 (3) (1993) 288–289 (in Chinese, with English title).
[83] B.F. Murphy, M.B. Thompson, A review of the evolution of viviparity in squamate reptiles: the past, present and future role of molecular biology and genomics, J. Comp. Physiol. B. 181 (2011) 575–594.
[84] W. Gao, Y.-B. Sun, W.-W. Zhou, Z.-J. Xiong, L. Chen, H. Li, T.-T. Fu, K. Xu, W. Xu, L. Ma, Y.-J. Chen, X.-Y. Xiang, L. Zhou, T. Zeng, S. Zhang, J.-Q. Jin, H.-M. Chen, G. Zhang, D.M. Hillisi, X. Ji, Y.-P. Zhang, J. Che, Genomic and transcriptomic investigations of the evolutionary transition from oviparity to viviparity, Proc. Natl. Acad. Sci. U. S. A. 116 (9) (2019) 3646–3655, https://doi.org/10.1073/pnas.1816086116.
[85] K. Ono, M. Osada, T. Matsutani, K. Mori, T. Nomura, Gonadal prostaglandin f^{-2}-alpha profile during sexual maturation in the oyster Crassostrea gigas, marine, Biol. Lett. 3 (1982) 223–230.
[86] Y.H. Zhang, V. Ravi, G. Qin, H. Dai, H.X. Zhang, F.M. Han, X. Wang, Y.H. Liu, J. P. Yin, L.M. Huang, B. Venkatesh, Q. Lin, Comparative genomics reveal shared genomic changes in syngnathid fishes and signatures of genetic convergence with placental mammals, Natl. Sci. Rev. 7 (2020) 964–977, https://doi.org/10.1093/nsr/nwaa002.
[87] I. Boutet, H.J. Alves Monteiro, L. Baudry, T. Takeuchi, E. Bonnivard, B. Billoud, S. Farhat, R. Gonzales-Araya, B. Salaun, A. Andersen, J.-Y. Toullec, F. Lallier, J. F. Flot, N. Guiglielmoni, X. Guo, B. Allam, E. Pales-Espinoza, J. Hemmer-Hansen, M. Marbouty, R. Koszul, A. Tanguy, Chromosomal assembly of the flat oyster (*Ostrea edulis* L.) genome as a new genetic resource for aquaculture, Evol. Appl. 15 (2022) 1730–1748, https://doi.org/10.1111/eva.13462.
[88] M.K. Gundappa, C. Peñaloza, T. Regan, I. Boutet, A. Tanguy, R.D. Houston, T. B. Bean, D.J. Macqueen, A chromosome level reference genome for European flat oyster (*Ostrea edulis* L.), Evol. Appl. 15 (2022) 1713–1729, https://doi.org/10.1111/eva.13460.
[89] D.A. Mardones-Toledo, J.A. Montory, A. Joyce, R.J. Thompson, C.M. Diederich, J. A. Pechenik, M.L. Mardones, O.R. Chaparro, Brooding in the chilean oyster *Ostrea chilensis*: unexpected complexity in the movements of brooded offspring within the mantle cavity, PLoS One 10 (4) (2015), e122859, https://doi.org/10.1371/journal.pone.0122859.
[90] O.R. Chaparro, D.A. Mardones-Toledo, M.W. Gray, V.M. Cubillos, J.M. Navarro, L. P. Salas-Yanquin, Female–embryo relationships in *Ostrea chilensis*: brooding, embryo recognition, and larval hatching, Mar. Biol. 166 (1) (2019) 10, https://doi.org/10.1007/s00227-018-3457-1.
[91] M.W. Gray, O. Chaparro, K.B. Huebert, S.P. O’Neill, T. Couture, A. Moreira, D. C. Brady, Life history traits conferring larval resistance against ocean acidification: the case of brooding oysters of the genus *Ostrea*, J. Shellfish Res. 38 (3) (2019) 751–761, https://doi.org/10.2983/035.038.0326.