Genomics Proteomics Bioinformatics 20 (2022) 1066–1077

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com

ORIGINAL RESEARCH

Adaptive Bird-like Genome Miniaturization During the Evolution of Scallop Swimming Lifestyle

Yuli Li [1,2,#], Yaran Liu [1,#], Hongwei Yu [1], Fuyun Liu [1], Wentao Han [1], Qifan Zeng [1,3,4], Yuehuan Zhang [5], Lingling Zhang [1,2], Jingjie Hu [1,4], Zhenmin Bao [1,3,4], Shi Wang [1,2,4,*]

[1] *Sars-Fang Centre & MOE Key Laboratory of Marine Genetics and Breeding, College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China*
[2] *Laboratory for Marine Biology and Biotechnology, Pilot Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China*
[3] *Laboratory for Marine Fisheries Science and Food Production Processes, Pilot Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China*
[4] *Key Laboratory of Tropical Aquatic Germplasm of Hainan Province, Sanya Oceanographic Institution, Ocean University of China, Sanya 572000, China*
[5] *CAS Key Laboratory of Tropical Marine Bio-resources and Ecology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, China*

Received 17 February 2022; revised 8 June 2022; accepted 20 July 2022
Available online 26 July 2022

Handled by Peng Cui

KEYWORDS

Genome size;
Lifestyle evolution;
Genome sequencing;
Scallop;
Bird

Abstract Genome miniaturization drives key evolutionary innovations of adaptive traits in vertebrates, such as the flight evolution of **birds**. However, whether similar evolutionary processes exist in invertebrates remains poorly understood. Derived from the second-largest animal phylum, **scallops** are a special group of bivalve molluscs and acquire the evolutionary novelty of the swimming lifestyle, providing excellent models for investigating the coordinated genome and **lifestyle evolution**. Here, we show for the first time that **genome sizes** of scallops exhibit a generally negative correlation with locomotion activity. To elucidate the co-evolution of genome size and swimming lifestyle, we focus on the Asian moon scallop (*Amusium pleuronectes*) that possesses the smallest known scallop genome while being among scallops with the highest swimming activity. Whole-**genome sequencing** of *A. pleuronectes* reveals highly conserved chromosomal macrosynteny and microsynteny, suggestive of a highly contracted but not degenerated genome. Genome reduction of *A. pleuronectes* is facilitated by significant inactivation of transposable elements, leading to reduced gene length, elevated expression of genes

* Corresponding author.
E-mail: swang@ouc.edu.cn (Wang S).
# Equal contribution.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

https://doi.org/10.1016/j.gpb.2022.07.001

involved in energy-producing pathways, and decreased copy numbers and expression levels of biomineralization-related genes. Similar evolutionary changes of relevant pathways are also observed for bird genome reduction with flight evolution. The striking mimicry of genome miniaturization underlying the evolution of bird flight and scallop swimming unveils the potentially common, pivotal role of genome size fluctuation in the evolution of novel lifestyles in the animal kingdom.

Introduction

Eukaryotic genomes vary greatly in size by more than five orders of magnitude [1,2]. The variation of genome sizes caused by stochastic genetic processes has consequences for organismal fitness and, therefore, may act as the target of natural selection [2]. In vertebrates, genome size has been tentatively associated with life cycle complexity [3], metabolic rate [4,5], body size [6], longevity [7], and developmental rate [8]. Particularly, flying creatures tend to have smaller genomes, which has been recognized as evidence of natural selection acting on genome size [5,9]. The condition was construed as the pivotal adaptation of bird flight behavior by reducing the metabolic costs caused by large genome and cell size [9,10]. In contrast to several well-known cases in vertebrates, the effect of genome size variation on invertebrate adaptive evolution remains largely unexplored [11,12].

With more than 100,000 extant species, mollusca is the second-largest animal phylum, which globally distributes among diverse aquatic and terrestrial environments [13] and has succeeded in surviving through several mass extinction events [14]. The genome sizes of molluscs range from 293 Mb for *Neomenia permagna* in solenogastres to 7.68 Gb for *Diplommatina kiiensis* in gastropoda. Swimming is one of the special lifestyles among molluscs with calcified shells. Different from most bivalves with sessile and buried lifestyles, jet propulsion is produced by rapid and successive contraction of adductor muscle [15]. The swimming lifestyle of scallops is generally considered as an evolutionary novelty [16], representing excellent models for investigating the coordinated genome and lifestyle evolution. Distinct lifestyles from cementing, byssal attachment with sporadic swimming, recessing, and free-living to gliding (classified into ecomorphs A to E) have been well documented by measurement of muscle use in inducing escape responses [17,18]. Scallops with gliding behavior are recognized as the most active swimmers. Shell morphology and metabolic capacity of adductor muscle are closely related to the swimming endurance of scallop species [17,18]. However, genomic bases underlying the evolution of the scallop swimming lifestyle remain poorly understood. To date, the published scallop genomes are all derived from ecomorph B, C, or D [19–24], and none of the genomes have been published for scallops with ecomorph E (gliding). The Asian moon scallop (*Amusium pleuronectes*; Linnaeus, 1758), a tropical Indo-West Pacific bivalve, is naturally distributed along the coast from Myanmar to Dampier Archipelago, north to southern Japan, and east to Papua New Guinea [15]. Thin and circular shells and gigantic adductor muscle make *Amusium* genus the most active scallops (belonging to the ecomorph E) [15,25]. Decoding the genome from the scallop lineage with ecomorph E is crucial for accomplishing comparative genome analysis to unravel the genomic bases underlying the evolution of the swimming lifestyle in scallops.

Here, we show for the first time that the genome sizes of scallops exhibit a generally negative correlation with locomotion activity. Genome sequencing of *A. pleuronectes* and comparative genomic analyses reveal genome miniaturization and relevant gene pathways that contribute to the evolution of the swimming lifestyle, which strikingly mimics genomic changes underlying bird flight evolution and provides insights into the pivotal role of genome size fluctuation in the evolution of novel lifestyles in the animal kingdom.

Results and discussion

Negative correlation between genome size and swimming activity

Convergent and parallel evolution leads to various lifestyles in the family Pectinidae [16,25]. The order of five lifestyles, ranging from cementing, byssal-attaching, recessing, and free-living to gliding, corresponds to the increase of scallop locomotion ability [16]. To explore the relationship of genome sizes with various lifestyles, we evaluated nine scallop species with genome size information from literature or animal genome size database (http://www.genomesize.com) and lifestyle classification (Table S1) according to the studies by Stanley and colleagues [26] and Serb and colleagues [25]. We found that scallop genome size substantially co-varied with lifestyle change (**Figure 1**A). Specifically, scallops with higher swimming activity possess smaller genome sizes, showing a negative correlation ($r = -0.78$) between genome size and swimming distance during one free swimming burst (Figure 1B; Table S2). For example, the gliding scallops of the *Amusium* genus, as the most active scallops [15,17], possess the smallest genomes among the nine scallop species. Along with the decrease of motility from gliding to byssal-attaching, the genome size is gradually enlarged. Though our sampling number of scallop species is not very large but still comparable to those adopted in previous studies for investigating the relationship between genome size and biological traits [12,27]. Such observation could be further strengthened by adopting more assayed scallop species in future studies.

The energy metabolism level and shell morphology are two important influencing factors on scallop locomotion activity [15,17,18]. The energy metabolism is closely related to the environmental temperature [28,29]. We investigated the geographical distribution and environmental temperature of different scallops and found that scallops with smaller genomes reside at lower latitudes, such as the Asian moon scallops in the tropic ocean, while scallops with larger genomes, such as the Yesso scallop (*Patinopecten yessoensis*) and the Queen scallop (*Aequipecten opercularis*), reside at higher latitude oceans (Figure 1C; Tables S1 and S3). The shell morphology is predominantly associated with shell thickness and shell shape [15,17]. In terms of shell thickness, the shells of Asian moon scallop are thinner and lighter [17,25], and in terms of shell shape,

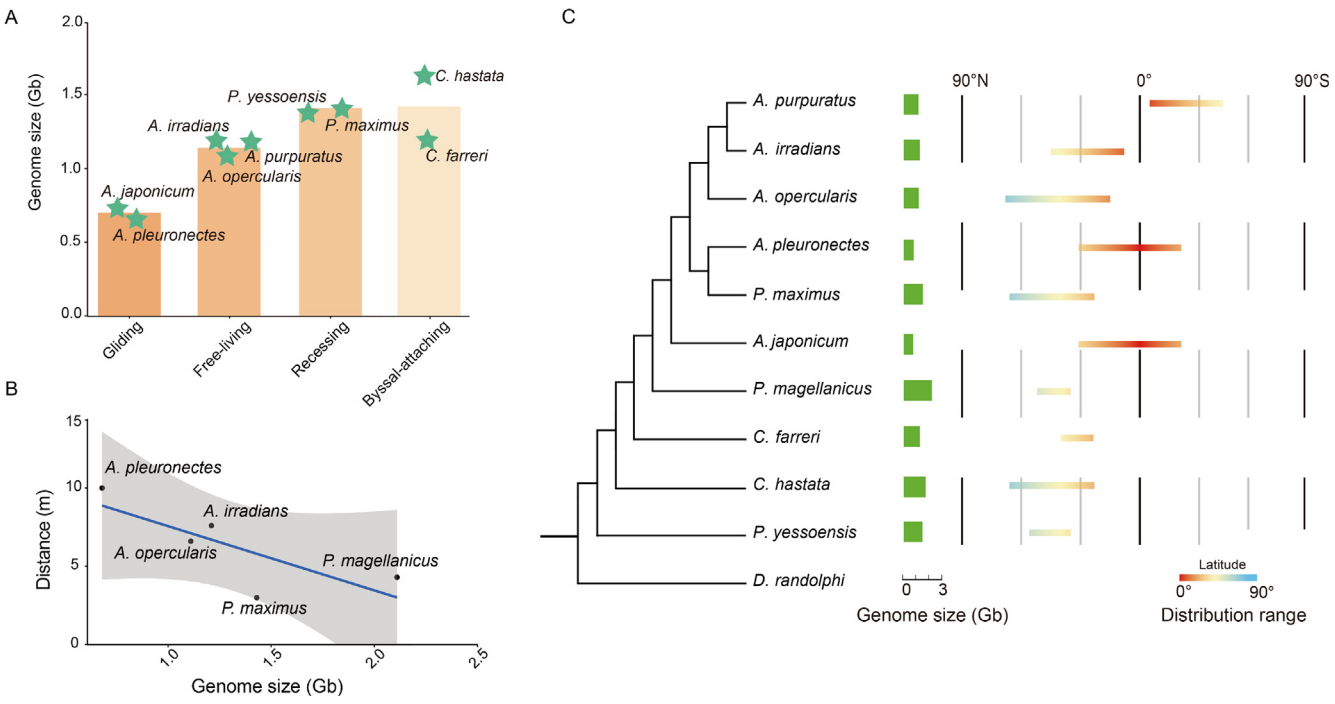

Figure 1 Relationship of genome size, locomotion activity, and geographical distribution of scallops

A. Correlation of genome sizes of nine scallop species with different lifestyles from most inactive to most active (byssal-attaching, recessing, free-living, and gliding). **B.** Correlation of genome sizes with distances during one swimming burst for five scallop species. **C.** Correlation of genome sizes and geographical distributions for ten scallop species. *A. pleuronectes*, *Amusium pleuronectes*; *A. japonicum*, *Amusium japonicum*; *A. purpuratus*, *Argopecten purpuratus*; *A. irradians*, *Argopecten irradians*; *P. yessoensis*, *Patinopecten yessoensis*; *P. maximus*, *Pecten maximus*; *A. opercularis*, *Aequipecten opercularis*; *C. farreri*, *Chlamys farreri*; *C. hastata*, *Chlamys hastata*; *P. magellanicus*, *Placopecten magellanicus*; *D. randolphi*, *Delectopecten randolphi*.

the height/width ratio (~ 1:1) of Asian moon scallop is more advantageous for swimming [17,25]. To depict the genomic bases underlying the evolution of the scallop swimming lifestyle, we put a focus on the Asian moon scallop (*A. pleuronectes*) for genome sequencing and comparative analysis, due to its exhibition of the smallest genome with strikingly high swimming activity.

The smallest but stable genome of *A. pleuronectes*

We conducted the sequencing and assembly of the smallest known scallop genome of *A. pleuronectes*, resulting in a total of 275.3 Gb of genomic data (Figure S1; Tables S4 and S5). The final genome assembly size is 626.63 Mb with contig N50 of 2.64 Mb and scaffold N50 of 34.61 Mb (Table S6), which is largely in accord with genome size estimated by *k*-mer analysis (667.07 Mb; Figure S2). A total of 609.55 Mb sequences (covering 97.27% of the total assembly) were anchored to 19 chromosomes (**Figure 2**A, Figure S3; Table S7), consistent with the dominant karyotypes of most scallops [19,30]. We performed benchmarking universal single-copy orthologs (BUSCO) analysis for the quality assessment of genome assembly, obtaining 93.9% of complete and single-copy BUSCOs (Table S8).

Totally, 24,359 protein-coding genes were annotated, of which 95.34% (23,225 genes) were supported by at least one known protein sequence (Table S9). We also annotated 1872 non-coding RNAs (ncRNAs), containing 302 microRNAs (miRNAs), 908 transfer RNAs (tRNAs), 146 ribosomal RNAs (rRNAs), and 185 small nuclear RNAs (snRNAs) (Table S10). Furthermore, we identified 246.53-Mb repeat sequences (39.34% of the genome assembly), in which tandem repeats possess the highest proportion (21.7%), followed by DNA transposons (3.53%), long interspersed nuclear elements (LINEs, 2.91%), long terminal repeats (LTRs, 2.57%), and short interspersed nuclear elements (SINEs, 1.65%) (Table S11). Genome phylogeny analysis of six scallop species suggested that *A. pleuronectes* is most closely related to the king scallop *Pecten maximus* with a divergence time of about 130 million years ago (MYA) (Figure 2B).

The genome architecture of *A. pleuronectes* was further evaluated by gene synteny analysis. For macrosynteny analysis, we found that *A. pleuronectes* showed equivalent macrosynteny conservation levels with similar karyotypes and conservation index values in comparison with other scallops, including the *P. yessoensis* genome that was recognized with ancestral bilaterian karyotype [19] (Figure 2C and D, Figure S4). For microsynteny analysis, we investigated the conservation of the *Hox* and *ParaHox* gene clusters, which play crucial roles in the anterior-posterior body plan patterning during animal development [31]. The other gene cluster (*ZMPSTE24-Taxilin-HECA-SYAP1-PSMG1*) conserved in protostomes [32] was also checked (Table S12). We found that all three gene clusters were intact with conserved gene orders in the *A. pleuronectes* genome (Figure 2E). The high conservation of chromosomal macrosynteny and microsynteny implies the

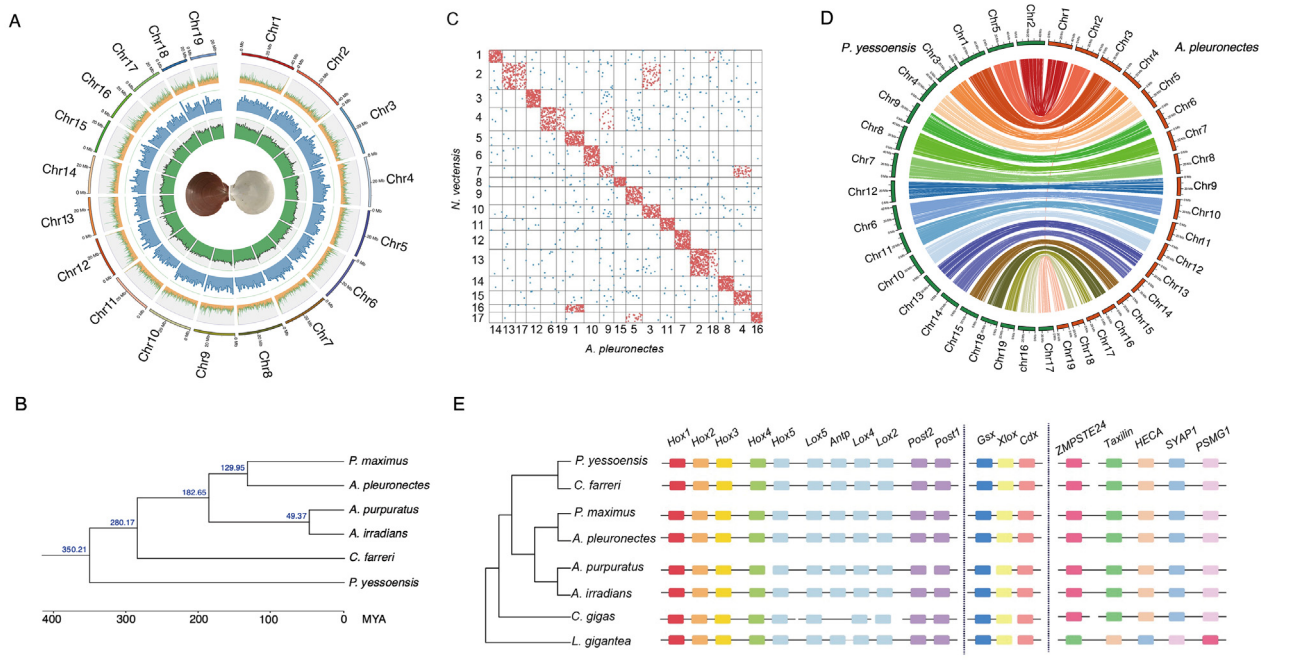

Figure 2 Genome landscape, phylogenetic analysis, and gene synteny conservation of *A. pleuronectes*

A. Circos plot showing the genomic features of *A. pleuronectes*. The tracks from the outer to the inner indicate chromosome length, repeat sequence percentage in each 100 kb genomic interval, the density of gene distribution in each 1 Mb genomic interval, and GC content of 0.5 Mb genomic interval, respectively. **B.** Phylogenetic position of *A. pleuronectes* based on 4159 highly conserved, single-copy orthologous genes derived from six scallop species. The estimated divergence time is shown in blue on each branching point. **C.** Macrosynteny analysis of *A. pleuronectes* using *N. vectensis* as an ancestral linkage group reference. **D.** Macrosynteny between *A. pleuronectes* and *P. yessoensis*. **E.** Microsynteny analysis of eight scallop species based on three conserved gene clusters. MYA, million years ago; *N. vectensis*, *Nematostella vectensis*; *C. gigas*, *Crassostrea gigas*; *L. gigantea*, *Lottia gigantea*.

smallest *A. pleuronectes* genome was contracted but not degenerated during evolution.

The smallest genome driven by transposon reduction

To depict the genome size evolution among scallop species, we conducted a comprehensive comparison of different genomic features for six scallop genomes that are currently available. Genomic architecture comparisons revealed that GC content, mean exon length, mean intron length, and coding gene number were similar among the six scallop species (Figure S5; Table S13). We found that the sizes of the genic regions among the six scallop species were all very close to the average size (average = 366 Mb, SD = 32 Mb), implying that the genic region kept stable during the diversification of scallop lineage. However, the sizes of the intergenic regions had a large variation among the six scallop species (SD = 126 Mb), and the *A. pleuronectes* genome possessed the smallest intergenic region and the smallest intergenic/genic ratio (**Figure 3**A, Figure S6). More importantly, the size of the intergenic region had a strong positive correlation with genome size (Figure S7), indicating that intergenic region size contributes greatly to the variation of scallop genome size.

We investigated the content variations of the two main repeat types [transposable elements (TEs) and tandem repeats] among scallop species, and found that TEs but not tandem repeats showed a significant reduction in *A. pleuronectes* genome (Table S13). Compared to *P. yessoensis* with the second-lowest TE contents, TEs in *A. pleuronectes* were reduced by 37.5% and 20% for the whole genome and intergenic regions, respectively. DNA transposons are the dominant TE type in scallop genomes, and we observed that *A. pleuronectes* genome lost most of the ancestral TEs that bursted about 150 MYA (Figure 3B). We found that TE expansion was remarkably inhibited in *A. pleuronectes* genome since barely any or little TE burst was observed after the divergence of *A. pleuronectes* with its closest relative *P. maximus* about 130 MYA. We further investigated the expansion/contraction of genes encoding transposition-related enzymes between *A. pleuronectes* genome and the other four scallop species by conducting a full search and comparison for genes encoding transposition-related enzymes. We found that for the several main classes (transposases, reverse transcriptases, integrases, recombinases, and endonucleases), the most dominant gene families with large gene numbers were contracted in *A. pleuronectes* genome (Figure 3C, Figure S8), including THAP domain-containing genes (PF05485) encoding proteins belonging to transposase, RVT_1 domain-containing genes (PF00078) encoding proteins belonging to reverse transcriptases, integrase_H2C2 (PF17921) and rve (PF00665) domain-containing genes encoding proteins belonging to integrases, YqaJ domain-containing genes (PF09588) encoding proteins belonging to recombinases, and HTH_Tnp_4 domain-containing genes encoding proteins belonging to endonucleases. We also investigated the expansion/contraction of genes that

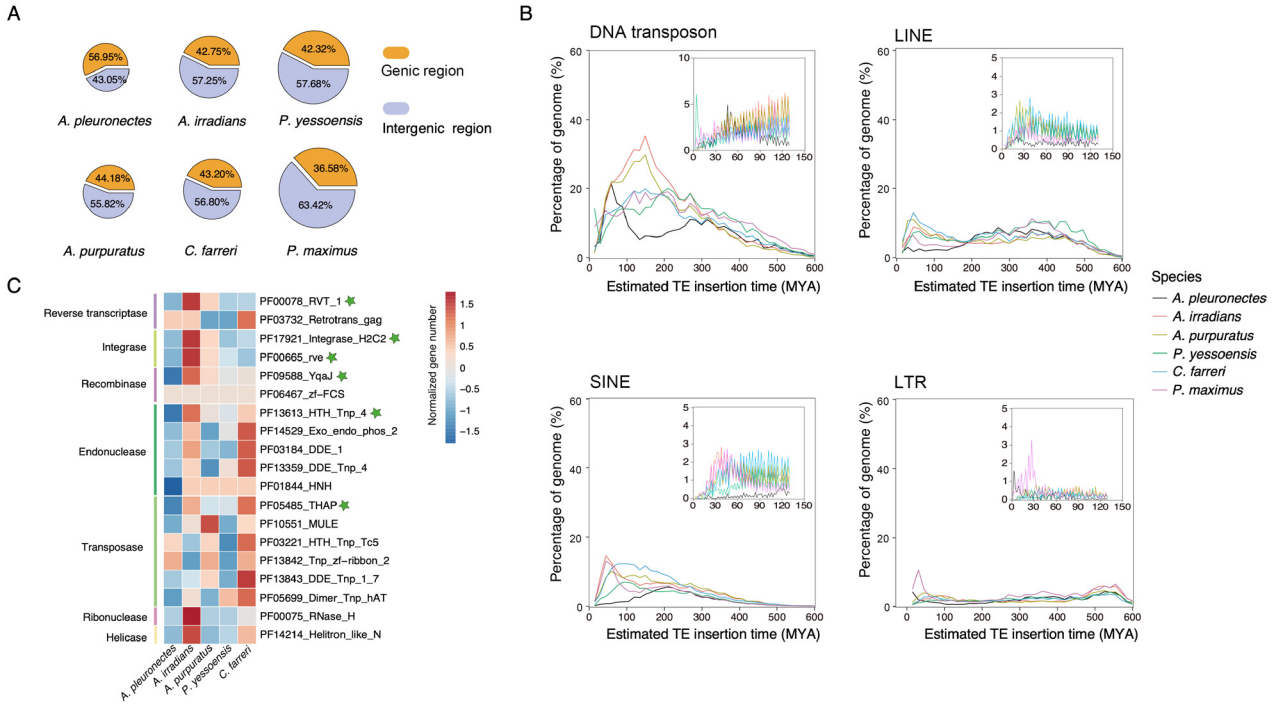

Figure 3 Genome reduction of *A. pleuronectes* in comparison with other scallop species

A. Genome size comparison among six scallop species. **B.** Estimated insertion time of four TE types (DNA transposons, LINEs, LTRs, and SINEs) in the genomes of six scallop species. **C.** Comparison of the numbers of genes encoding transposition-related enzymes among different scallop species. The green star indicates the contracted gene families. TE, transposable element; LINE, long interspersed nuclear element; LTR, long terminal repeat; SINE, short interspersed nuclear element.

inhibit transposon activity among scallops [33,34], but there were no significant trends for them (Figure S9). However, the activities of these genes represented by gene expression levels were elevated in *A. pleuronectes* (Figure S10). The TE reduction in non-coding regions of *A. pleuronectes* genome, caused by the contraction of genes encoding transposition-related enzymes and the elevated expression of TE inhibiting genes, leads to producing a "lighter" scallop genome that mimics flying birds [5,9] and therefore paves the way for the evolution of high locomotion activity.

Gene size reduction for high metabolic capacity in scallop and bird

The swimming behavior of scallops, which results from a rapid succession of adductor muscle contraction, is much related to the energy metabolism of the large striated muscle [17,18]. Our group previously revealed that the large striated muscle of the scallop was energy-dynamic and showed higher energy demand in the swimming scallop than in the cementing oyster [20]. However, it remains unclear about the molecular driving forces that determine differential locomotion activities among different scallops. We conducted comprehensive genomic and transcriptomic analyses to depict the locomotion-related molecular changes underlying genome size reduction. From a genomic perspective, we identified 3794 size-reducing genes in *A. pleuronectes* when compared with the other five scallop species (Figure S11). Functional enrichment analysis of the size-reducing genes of *A. pleuronectes* revealed that these genes were most significantly enriched in the oxidative phosphorylation pathway (**Figure 4**A, Figures S12 and S13), which is the crucial energy-producing pathway responsible for the production of 95% ATP [35]. Particularly, we found that the TCA cycle, the other important part of aerobic respiration, was enriched in both size-reducing genes and positively selected genes (Figure S14), indicating the strong selective pressure on the energy metabolism process in *A. pleuronectes*. Moreover, these size-reducing genes were also found to be significantly enriched in basal transcription and translation processes (general transcription factors and aminoacyl-tRNA biosynthesis pathways) and basal metabolic pathways (Figure 4A, Figure S12). For the analysis of size-reducing genes in birds, we selected the hummingbird (*Calypte anna*), which is one of the fastest flying birds and possesses the highest metabolic rates normalized by body mass among vertebrates [36,37]. We identified size-reducing genes in *C. anna* genome [38] (by comparing with other three selected tetrapods, including *Chelydra serpentina*, *Xenopus laevis*, and *Mus musculus*) and found that these genes were also significantly enriched in relevant pathways, including basal metabolic pathways and oxidative phosphorylation pathway (Figure 4A). These findings support the parallel evolution of gene size reduction for high metabolic capacity in both scallop swimming and bird flight.

Gene length is one of the well-known determining factors for transcriptional regulation, generally with faster and accumulated higher expression of shorter genes than longer genes [39]. To evaluate the transcriptional effect of size-reducing genes, we obtained the expression levels of these genes in three scallop

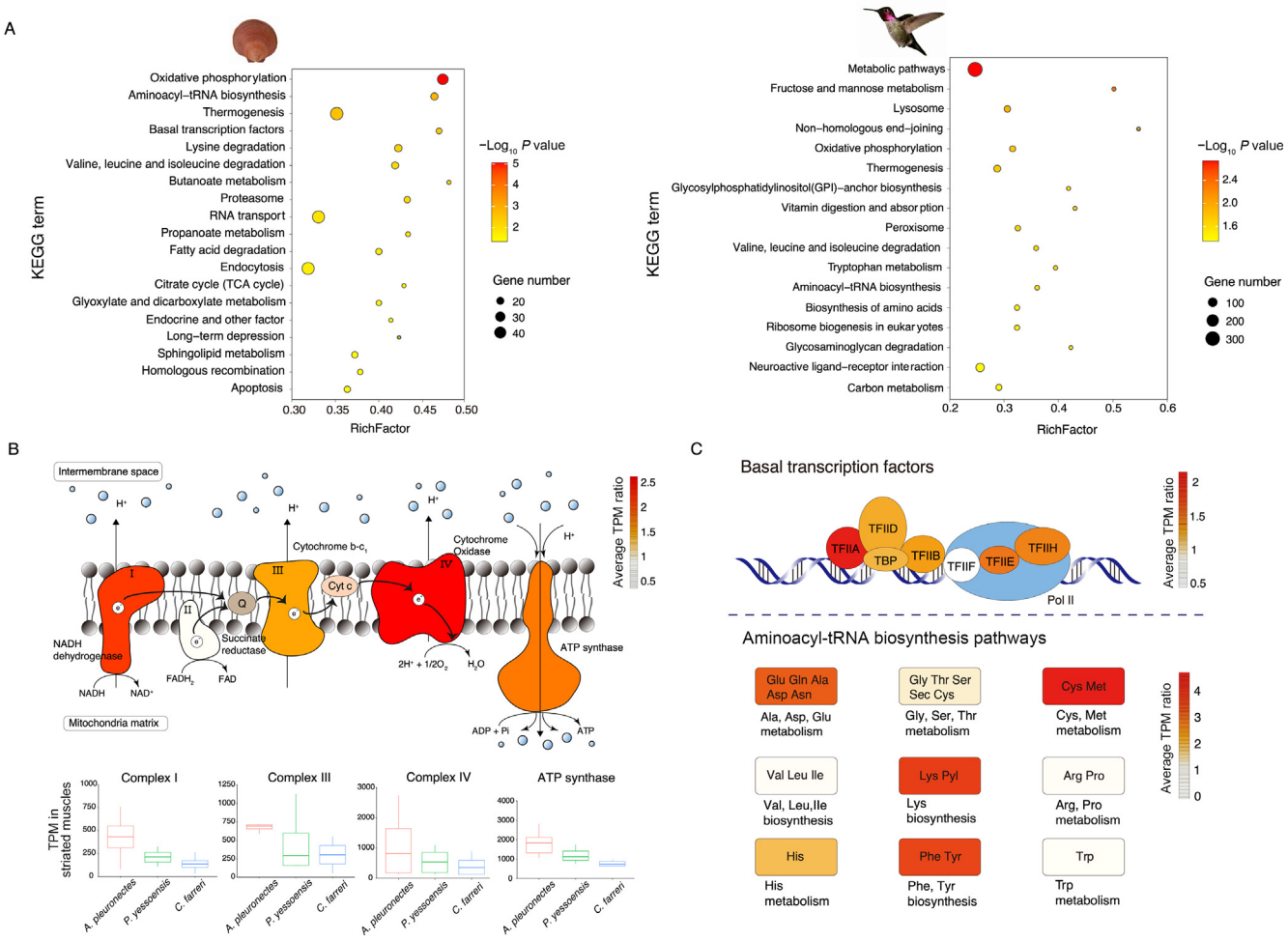

Figure 4 Functional analysis and expressional comparison of size-reducing genes in scallop (*A. pleuronectes*) and bird (*C. anna*)
A. The KEGG enrichment analysis of size-reducing genes in the *A. pleuronectes* genome compared with five less-active scallop species (*P. yessoensis*, *A. purpuratus*, *A. irradians*, *P. maximus*, and *C. farreri*) and in the *C. anna* genome compared with three non-flying tetrapods (*C. serpentina*, *X. laevis*, and *M. musculus*). **B.** Comparison of expression levels (represented by TPM values) of size-reducing genes enriched in the oxidative phosphorylation pathway in the striated muscles of three scallop species (*A. pleuronectes*, *P. yessoensis*, and *C. farreri*). Colors of complexes I-IV and ATP synthase indicate the average TPM ratios in the striated muscle of *A. pleuronectes* to the other two scallop species. **C.** Comparison of expression levels of size-reducing genes enriched in the basal transcription factors and aminoacyl-tRNA biosynthesis pathways among four scallop species (*A. pleuronectes*, *P. yessoensis*, *C. farreri*, and *A. purpuratus*). Different colors indicate the average TPM ratios in *A. pleuronectes* to the other three scallop species. TPM, transcripts per million; *C. anna*, *Calypte anna*; *C. serpentina*, *Chelydra serpentina*; *X. laevis*, *Xenopus laevis*; *M. musculus*, *Mus musculus*.

species with different motility classifications (from least active ecomorph A to most active ecomorph E) and abundant adult tissue/organ transcriptomic data enabling for a comparative analysis. The selected scallop species were classified into three ecomorph types, including the most active scallop *A. pleuronectes* (ecomorph E), the recessing scallop *P. yessoensis* (ecomorph D), and the byssal-attaching scallop *Chlamys farreri* (ecomorph B). We compared the gene expression profiles in the striated muscles of three scallop species for the three gene pathways that were most enriched with size-reducing genes in *A. pleuronectes* (Figure 4B and C). For the energy-producing oxidative phosphorylation pathway, all genes showed higher expression in the striated muscle of *A. pleuronectes* than the other scallops (Figure 4B), particularly for complexes I, III, and IV, which contribute to the production of the proton gradient, and ATP synthase (complex V), which drives ADP phosphorylation to ATP (Figure 4B). We also mixed the transcriptome data of striated and smooth muscles of *A. pleuronectes*, *P. yessoensis*, and *C. farreri* to enable the comparison with the adductor muscle transcriptome data of *Argopecten purpuratus*, which still confirmed our previous finding that *A. pleuronectes* has higher expression levels in the oxidative phosphorylation pathway than the other three species (Figure S15). For the two basal cellular processes, we also found a significant increase of gene expression in *A. pleuronectes* than in the other three scallops (Figure 4C). These transcriptional changes manifest the functional role of gene size reduction for the adaptive evolution of high metabolic capacity in *A. pleuronectes*.

Parallel repression of biomineralization genes in scallop and bird

It has been reported that shell structure and weight are closely related to scallop locomotion activity [17]. There are many known biomineralization-related genes in bivalves, such as those encoding carbonic anhydrase (CA), tyrosinase, epidermal growth factor (EGF), von Willebrand factor type A (VWA), immunoglobin (IG), and fibronectin-3 domain (FN3) [40,41]. We then investigated the copy numbers and expression profiles of these biomineralization-related genes (Table S14) in three scallop species, including *A. pleuronectes* (ecomorph E), *P. yessoensis* (ecomorph D), and *C. farreri* (ecomorph B). We found that most of the biomineralization-related genes have fewer copy numbers in *A. pleuronectes* than those in the other scallops (**Figure 5**A). The gene expression profiles of seven main gene families, including CA, chitin-binding peritrophin-A domain (CBM_14), EGF, FN3, tyrosinase, VWA, and whey acidic protein (WAP), showed lower expression levels in the mantle transcriptome of *A. pleuronectes* than in other scallop species (Figure 5B). To reveal the core biomineralization-related genes and their interaction relationships, we constructed the weighted correlation network analysis (WGCNA) gene co-expression networks (Figures S16–S18) and identified the biomineralization-related modules for *A. pleuronectes*, *P. yessoensis*, and *C. farreri*, in which the mantle-specific highly-expressed genes were significantly enriched (Figures S19–S21). As expected, several known biomineralization-related genes were found as hub genes (the top 40 genes ranked by intramodular connectivity) in the biomineralization-related modules of these three scallop species (Figure 5C, Figure S22). Notably, this biomineralization-related module was significantly enriched

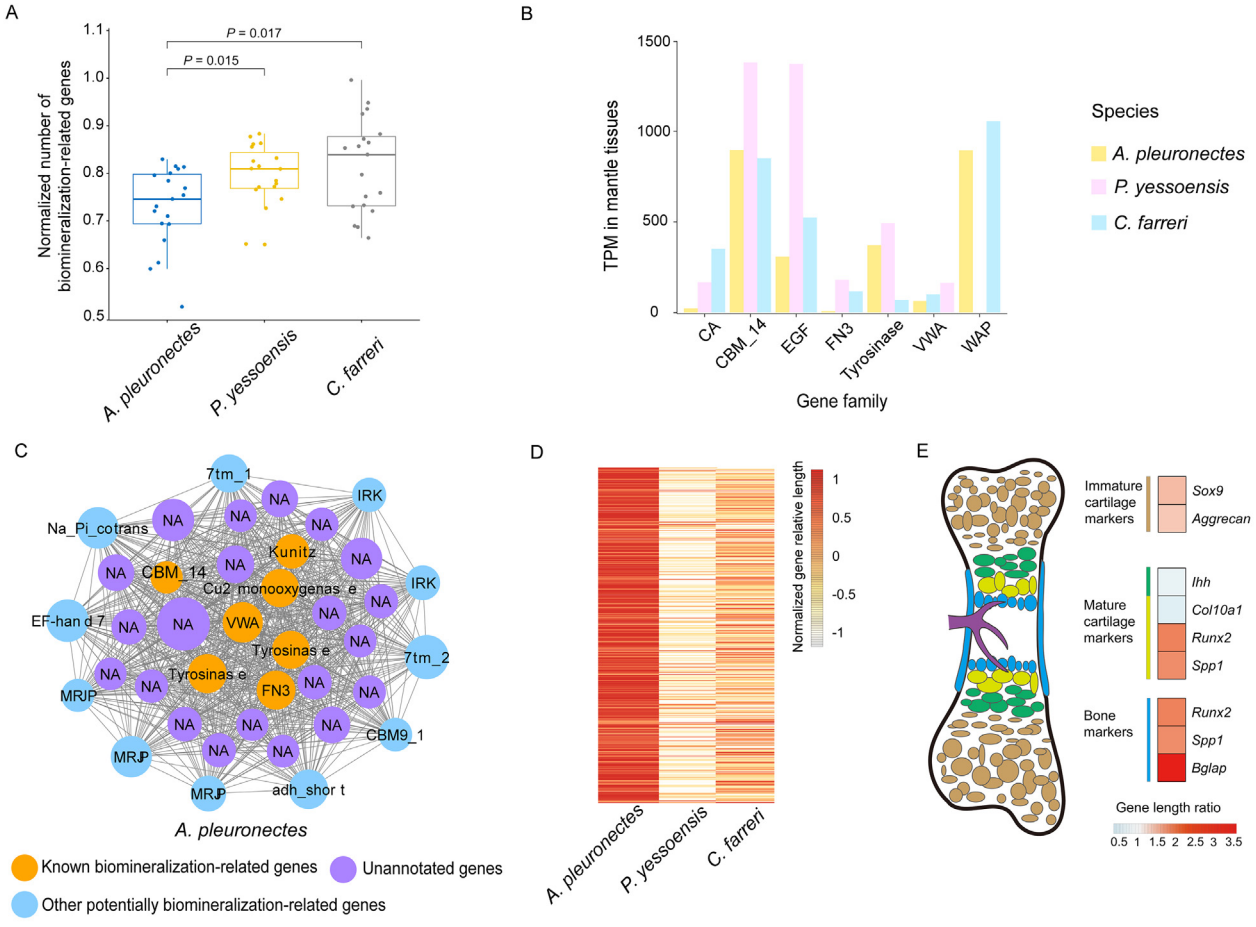

Figure 5 Analysis of biomineralization-related genes in scallop (*A. pleuronectes*) and bird (*C. anna*)

A. Comparison of biomineralization-related genes between *A. pleuronectes* and the other two scallop species. The boxplot shows the copy number variations of known biomineralization-related genes in bivalves. **B.** The histogram shows the cumulative gene expression levels (represented by TPM values) of seven main biomineralization-related gene families in the mantle tissues of three scallop species. **C.** Biomineralization-related module of *A. pleuronectes*. **D.** Relative gene length comparison of 440 biomineralization-related genes by *de novo* search among three scallop species. **E.** Comparison of bone marker genes in flying birds and non-flying tetrapods. Different colors of genes show the ratio of gene length in the flying bird (*C. anna*) to the average gene length in the other three tetrapods (*C. serpentina*, *X. laevis*, and *M. musculus*). CA, carbonic anhydrase; CBM_14, chitin-binding peritrophin-A domain; EGF, epidermal growth factor; FN3, fibronectin-3 domain; VWA, von Willebrand factor type A; WAP, whey acidic protein.

with the size-increasing genes identified in *A. pleuronectes* compared with the other two scallop species ($P < 0.05$, Figure 5D). The decrease in gene copy numbers or expression levels and increase in gene length of biomineralization-related genes in the Asian moon scallop may explain their thinner and lighter shells for more powerful motility than other scallop species. Interestingly, we also found a similar tendency in flying birds in comparison with other tetrapods, that is, the length of most bone marker genes [42] relatively increased in flying birds (Figure 5E, Figure S23), implying the convergent evolution of the depressed biomineralization process for acquisition of powerful motility in both vertebrates and invertebrates.

Conclusion

We report for the first time that genome sizes of scallops show a generally negative correlation with locomotion activity. Sequencing of the smallest genome of Asian moon scallop *A. pleuronectes* enables the decoding of co-evolution of genome size and swimming lifestyle. Transposon reduction in the *A. pleuronectes* genome, resulting from significant contraction of several transposon-related gene families, leads to the smallest scallop genome. Systematic investigation of energy metabolism and biomineralization reveals a similar tendency of molecular changes underlying genome miniaturization evolution of scallops and flying birds. Our study highlights the pivotal role of genome size fluctuation in the evolution of novel lifestyles in the animal kingdom.

Materials and methods

Sample preparation

The Asian moon scallops (*A. pleuronectes*) were gathered from the Tung Ping Chau Bay (near Tung Chung village, Guangzhou) for whole-genome sequencing. The DNA samples were stored at the key laboratory of marine genetics and breeding, Ministry of Education, Ocean University of China (specimen code: OUC-MGB-2019-Apl-08). Dissected tissues were immediately frozen in liquid nitrogen and preserved at −80 °C for DNA extraction. The conventional hexadecyltrimethylammonium bromide (CTAB) procedure was used to extract high molecular-weight genomic DNA (gDNA) from the adductor muscle [43]. The pair-end libraries with an insert size of 350 bp were built from gDNA by utilizing Illumina genomic DNA sample preparation kits (Catalog No. 20060060, Illumina, San Diego, CA), following the manufacturer's standard protocols, then sequenced on the Illumina Xten platform. A long-read DNA library for gDNA derived from the same individual was constructed using SMRTbell template prep kit 2.0 (Catalog No. 101-685-400, PacBio, Menlo Park, CA) on the Pacific Biosciences (PacBio) Sequel single-molecule real-time (SMRT) platform.

Hi-C library construction and sequencing

The Hi-C library was constructed using adductor muscle tissues obtained from the same individual of *A. pleuronectes* for high-quality DNA extraction. About 5 g of the tissue sample was cut into pieces and fixed with 2% formaldehyde solution for about 20 min at room temperature. After digesting with the *Mbo*I restriction enzyme (Catalog No. R0147M, New England Biolabs, Ipswich, MA), the sticky ends of the treated fragments were repaired with biotinylated residues to form blunt-end fragments. End repair, adaptor ligation, and polymerase chain reaction (PCR) amplification were performed before sequencing. Finally, the library was sequenced with the Illumina Xten 2500 (PE150) platform.

Genome size estimation and *de novo* assembly of the *A. pleuronectes* genome

The genome size of *A. pleuronectes* was estimated using *k*-mer frequency distribution analysis based on Illumina sequencing reads. The Trimmomatic software [44] was used to filter raw reads by trimming adaptors and discarding reads with more than 10% N bases or more than 20% low-quality bases. Then Jellyfish v2.2.5 [45] and GenomeScope v2.0 [46] were employed in measuring genome size and heterozygosity based on a 19-mer distribution. The size of the genome was estimated using the formula: genome size = *k*-mer number/*k*-mer depth.

The wtdbg2 software [47] was used to *de novo* assemble the genome using the PacBio long reads. Based on the fuzzy-Bruijn graph (FBG) algorithm, the raw reads were assembled by the wtdbg2 module, and the consensus calling of the preceding assembly was conducted with wtdbg-cns, which was also used for reducing sequencing errors. The Illumina clean reads were mapped to the assembly using Burrows-Wheeler aligner (BWA) software [48] to make sure the accuracy of the genome. Then, the Pilon software [49] was used to resolve the conflict of the assembly.

For chromosome-level scaffolding, the Hi-C sequencing technique was applied for chromosome construction for *A. pleuronectes*. The clean reads were aligned to the assembly by using BWA, and uniquely aligned reads with high mapping quality were retained for further analysis. Invalid interaction pairs were filtered by HiC-Pro software [50], then Lachesis software [51] was employed to cluster and anchor contigs to chromosomes with the agglomerative hierarchical clustering method based on the interaction matrix between sequences. The integrity and completeness of the genome assembly were evaluated by calculating the mapping rate of paired-end clean reads to the genome. For genome integrity assessment, the BUSCO v3.0.2 software [52] was applied based on the metazoan model database.

Repeat element and ncRNA annotations

Tandem repeats were identified by Tandem Repeats Finder v4.09 [53] under default parameters, and TEs were identified via both *de novo* predictions and homology-based methods. *De novo* repetitive sequence database was constructed by RepeatModeler v1.0.11 [54], and then merged with Repbase database [55] for homology-based detection by Repeatmasker v2.1 software [56]. In addition, we conducted Kimura distance-based copy divergence analysis of four types of TEs in the intergenic regions of the six scallop genomes. For the annotation of ncRNAs in *A. pleuronectes* genome, the tRNAscan-SE software [57] was employed to annotate tRNAs, and the INFERNAL software [58] was applied to search for all other RNAs, including rRNAs, miRNAs, and snRNAs against with the Rfam database [59].

Gene prediction and function annotation

The gene structure was predicted by integrating three methods, including homology-based search, *de novo* prediction, and RNA sequencing (RNA-seq) supporting alignment. The *de novo* prediction was conducted using Augustus [60], GlimmerHMM [61], SNAP [62], Geneid [63], and Genescan software [64]. For homology-based analysis, TBLASTN was used to align protein sequences from six bivalves (*P. yessoensis*, *C. farreri*, *A. purpuratus*, *Crassostrea gigas*, *Pinctada fucata*, and *Modiolus philippinarum*) to the *A. pleuronectes* genome assembly using TBLASTN. The accuracy of spliced alignments was then assessed by aligning homologous genome sequences to the matched proteins using GeneWise [65]. Then, the RNA-seq reads from adult tissues/organs were aligned to the assembly using Tophat software [66], and Cufflinks software [67] was employed to assemble and annotate transcripts. Finally, EVidenceModeler [68] was employed to integrate all gene models obtained from the aforementioned methods. NCBI non-redundant (Nr) protein database, Gene Ontology (GO) database [69], Kyoto encyclopedia of genes and genomes (KEGG) database [70], SwissProt [71], and InterPro were used to functionally annotate the predicted protein-coding genes.

Transcriptome sequencing and expression profile analysis

The mantle, striated/smooth muscle, kidney, gill, gonad, and foot tissues/organs from three adult individuals were chosen for RNA-seq. Tissues/organs were dissected and immediately frozen in liquid nitrogen, followed by preservation at −80 °C until RNA extraction. Total RNA was isolated from each tissue/organ according to previously described protocols [72]. All RNA-seq libraries were prepared using the NEB next mRNA library prep kit (Catalog No. E7770, New England Biolabs) according to the manufacturer's instructions and then sequenced on the Illumina Xten platform. The adaptors were firstly trimmed from raw reads, and reads with more than 10% N bases or more than 20% low-quality bases were filtered. STAR [73] was used to align the clean reads to the reference genome, and then FeatureCounts [74] was used to count the aligned reads to each gene. Gene expression level was represented by the count of transcripts per million (TPM), which was defined as transcripts per kilobase of exon model per million mapped reads and calculated using a custom Perl script.

Macrosynteny and microsynteny analyses

For species with chromosome-level assemblies (*A. pleuronectes*, *P. yessoensis*, *C. farreri*, and *P. maximus*), the conservation of gene macrosynteny compared to the presumed bilaterian ancestral linkage groups (ALGs) was assessed according to the method described by Wang and colleagues [19]. For species without chromosome-level assemblies (*A. purpuratus* and *Argopecten irradians*), we employed a heuristic hierarchical method to cluster the scaffolds from the draft genome and set the tree-cutting threshold as 0.25.

For detecting conserved gene clusters among the six scallop species, we conducted *Hox* and *ParaHox* cluster identification and *de novo* searching based on genome assemblies. The *Hox* and *ParaHox* genes were identified in scallop genomes using BLASTP with an *E* value threshold of 1E−10 against known *Hox* and *ParaHox* protein sequences, and the homeobox domain was further confirmed according to the conserved domain database (CDD, http://www.ncbi.nlm.nih.gov/cdd).

Phylogenetic tree construction and divergence time estimation

The phylogenetic tree of scallops with different lifestyles was constructed based on the available rRNA genes (12S, 16S, and 28S), including *A. pleuronectes* and other nine scallop species (*Amusium japonicum*, *A. purpuratus*, *A. opercularis*, *A. irradians*, *C. farreri*, *P. yessoensis*, *P. maximus*, *Chlamys hastata*, and *Placopecten magellanicus*), as well as one outgroup species (*Delectopecten randolphi*). The phylogenetic tree was built using BEAST2 v2.6 [75] based on the bayesian inference method.

Six scallop species were employed to construct the genome-based phylogeny, including *A. pleuronectes*, *P. yessoensis*, *C. farreri*, *A. purpuratus*, *A. irradians*, and *P. maximus*. The sequences of protein-coding genes of the six scallop species were retrieved from the MolluscDB database. Orthologous groups from the six genomes were identified using OrthoFinder v2.3.12 [76] with an *E* value of 1E−5 for BLASTP. Then, 4159 single-copy orthologs were used for multiple alignment analysis by using MAFFT software [77], and gaps were deleted using Gblocks [78]. All alignments were combined into a supermatrix to construct a phylogenetic tree using RAxML software [79] with 1000 rapid bootstrap analyses. Finally, the divergence time of species was estimated based on the MCMCTree module in the PAML package [80] combined with a molecular clock model. Two reference divergence time points (500−550 MYA between *Lottia* and *Crassostrea*; > 330 MYA between *Patinopecten* and *Crassostrea*) retrieved from the TimeTree database [81] were used to calibrate divergence dates of other nodes.

TE-related analysis

To investigate the dynamics of TE activities during the evolution of scallops, we employed the free-ratio model of PAML software to calculate the nucleotide substitution rates [80], and RepeatMasker was used to calculate TE sequence divergences from consensus sequences. The substitution level K of TEs was calculated according to the Jukes-Cantor formula $K = 3/4\mathrm{Ln}(1 - 4d/3)$, in which K is the genetic distance and d represents the number of observed substitutions normalized by the sequence length. We employed the formula $T = K/2r$ [82] to estimate insertion time of TEs, in which r represents the nucleotide substitution rate for each scallop.

To categorize the gene families, six scallop species were scanned using the HMM approach on known Pfam functional domains. A protein containing multiple copies of a domain was only counted once. The gene family expansion/contraction analysis for *A. pleuronectes* was conducted by comparing the average domain counts in the other five scallop species using the Fisher test in R. And the gene families with $P < 0.05$ were identified as significantly contracted or expanded ones.

Analysis of size-reducing genes in the scallop *A. pleuronectes* and the bird *C. anna*

Single-copy orthologous genes among six scallop species were identified using OrthoFinder v2.3.12 [76] with a threshold *E* value of 1E−5 for BLASTP. To identify size-reducing genes

in the scallop *A. pleuronectes*, the length of paired single-copy orthologous genes in six scallop species were standardized using gene length divided by the genic region of the genome. Genes with the relative length value in *A. pleuronectes* smaller than the average value of the other five less-active scallop species were defined as size-reducing genes. Similarly, single-copy orthologous genes from the four tetrapods were identified using OrthoFinder with an *E* value of 1E−5. To identify size-reducing genes in the bird *C. anna*, the length of paired single-copy orthologous genes in four tetrapods were standardized using gene length divided by the genic region of the genome. Genes with a relative length value in *C. anna* smaller than the average value of the other three non-flying tetrapods were defined as size-reducing genes. Genome assemblies of four tetrapods were downloaded from the NCBI genome database, including *C. anna* (NCBI genome: bCalAnn1_v1.p), *C. serpentina* (NCBI genome: ASM1885937v1), *X. laevis* (NCBI genome: Xenopus_laevis_v10.1), and *M. musculus* (NCBI genome: GRCm39).

Size-reducing genes in *A. pleuronectes* and *C. anna* were employed for KEGG enrichment analysis by EnrichPipeline [83], respectively. Remarkable enrichment of KEGG pathways ($P < 0.05$) was shown by bubble plot. We compared the expression levels (TPM) of size-reducing genes enriched in the oxidative phosphorylation pathway in striated muscles between *A. pleuronectes* and the other two scallop species (*P. yessoensis* and *C. farreri*). The complex colors of the oxidative phosphorylation pathway plots represent the up-regulation degree for genes in *A. pleuronectes* according to the ratio of the TPM of a gene in *A. pleuronectes* divided by the average TPM of the gene in the other two less-active scallop species. In addition, we downloaded the adductor muscle transcriptome data of *A. purpuratus* from NCBI (SRA: SRR6849474), and compared it with mixed transcriptome data of striated and smooth muscles of *A. pleuronectes*, *P. yessoensis*, and *C. farreri*. Transcriptome analyses of basal transcription factors and aminoacyl-tRNA biosynthesis pathways were performed among four scallop species (*A. pleuronectes*, *P. yessoensis*, *C. farreri*, and *A. purpuratus*). Transcriptome data of *C. farreri* and *P. yessoensis* were retrieved from studies of Wang et al. [19] and Li et al. [20], and nine transcriptome data of *A. purpuratus* were obtained from NCBI (SRA: SRR6849473, SRR6849474, SRR6849475, SRR7993930, SRR7993931, SRR7993936, SRR7993937, SRR7993946, and SRR7993947).

Biomineralization-related gene analysis

The identification of putative biomineralization-related genes was conducted by aligning to well-known Pfam protein domains according to the bivalve biomineralization toolbox [40]. Normalized domain number in the three scallop species (*A. pleuronectes*, *P. yessoensis*, and *C. farreri*) was compared by ggboxplot package in R, and a comparison of expression levels of biomineralization-related genes in mantle was shown in the bar plot. To construct the co-expression gene network, the R package of WGCNA [84] was employed based on the adult transcriptome data of three scallop species including *A. pleuronectes*, *P. yessoensis*, and *C. farreri*. The modules that enriched genes specifically highly expressed in the mantle tissue was considered a biomineralization-related module. The connection strength for genes in each module was represented by the intramodular connectivity value (K_{within}), which is an important factor for determining hub genes. Then, we used Cytoscape software to visualize the top 40 hub genes in the biomineralization-related modules of the three scallop species. OrthoFinder was used for detecting single-copy orthologous genes in the three scallop species, and relative gene lengths were employed to determine length-increasing genes in *A. pleuronectes* in the biomineralization-related module by comparing to the other two scallop species. And the results were shown by a heatmap plot in R.

For the analysis of bone formation-related genes, we collected marker genes of immature cartilage, mature cartilage, and bone according to the previous study [42]. Then, we verified the expression pattern of bone marker genes by the expression profile of adult organs/tissues in *G. gallus*. RNA-seq data of *G. gallus* were downloaded from NCBI (SRA: SRR594500–SRR594526, SRR16693970–SRR16693975, and SRR16693980–SRR16693982).

Code availability

The code developed for the study has been submitted to BioCode at the National Genomics Data Center (NGDC), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences (CAS) / China National Center for Bioinformation (CNCB) (BioCode: BT007309), which is publicly accessible at https://ngdc.cncb.ac.cn/biocode/tools/BT007309.

Data availability

The *A. pleuronectes* scallop genome project has been deposited in NCBI (BioProject: PRJNA797172). The genome sequencing data have been uploaded to Sequence Read Archive (SRA: SRR17752961–SRR17752967, SRR17752978, SRR17752989, and SRR17752990). The Illumina sequencing data of 20 transcriptomes for adult tissues/organs have also been uploaded to SRA (SRA: SRR17752968–SRR17752977 and SRR17752979–SRR17752988). The assembled data of *A. pleuronectes* have also been submitted to Genome Warehouse [85] at the NGDC, BIG, CAS / CNCB (GWH: GWHBJCZ00000000), and are publicly accessible at https://ngdc.cncb.ac.cn/gwh. The raw sequencing data have also been deposited in the Genome Sequence Archive [86] at the NGDC, BIG, CAS / CNCB (GSA: CRA006995), and are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

CRediT author statement

Yuli Li: Data curation, Methodology, Formal analysis, Writing - review & editing. **Yaran Liu:** Software, Formal analysis, Data curation, Methodology, Writing - original draft. **Hongwei Yu:** Formal analysis, Methodology. **Fuyun Liu:** Software, Data curation. **Wentao Han:** Formal analysis. **Qifan Zeng:** Data curation. **Yuehuan Zhang:** Data curation. **Lingling Zhang:** Writing - review & editing. **Jingjie Hu:** Writing - review & editing. **Zhenmin Bao:** Conceptualization, Project administration, Writing - review & editing. **Shi Wang:** Conceptualization, Project administration, Writing - review & editing. All authors have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

We acknowledge the grant support from the Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (Grant No. 2022QNLM050101-1), the Fundamental Research Funds for the Central Universities (Grant No. 202141011), the National Natural Science Foundation of China (Grant No. 32130107), the Key R&D Project of Shandong Province (Grant No. 2021ZLGX03), and the Taishan Scholar Project Fund of Shandong Province of China. We thank Hao Wang for downloading and preparing data from NCBI.

Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gpb.2022.07.001.

ORCID

ORCID 0000-0002-8112-1730 (Yuli Li)
ORCID 0000-0002-8147-6988 (Yaran Liu)
ORCID 0000-0002-5552-6339 (Hongwei Yu)
ORCID 0000-0001-9577-1841 (Fuyun Liu)
ORCID 0000-0003-1423-3975 (Wentao Han)
ORCID 0000-0002-1886-8444 (Qifan Zeng)
ORCID 0000-0001-9150-0163 (Yuehuan Zhang)
ORCID 0000-0002-9161-4387 (Lingling Zhang)
ORCID 0000-0002-0715-7137 (Jingjie Hu)
ORCID 0000-0001-6126-0509 (Zhenmin Bao)
ORCID 0000-0002-9571-9864 (Shi Wang)

References

[1] Canapa A, Barucca M, Biscotti MA, Forconi M, Olmo E. Transposons, genome size, and evolutionary insights in animals. Cytogenet Genome Res 2015;147:217–39.
[2] Blommaert J. Genome size evolution: towards new model systems for old questions. Proc R Soc B-Biol Sci 2020;287:20201441.
[3] Gregory TR. Genome size and developmental complexity. Genetica 2002;115:131–46.
[4] Waltari E, Edwards SV. Evolutionary dynamics of intron size, genome size, and physiological correlates in archosaurs. Am Nat 2002;160:539–52.
[5] Wright NA, Gregory TR, Witt CC. Metabolic 'engines' of flight drive genome size reduction in birds. Proc R Soc B-Biol Sci 2014;281:20132780.
[6] Decena-Segarra LP, Bizjak-Mali L, Kladnik A, Sessions SK, Rovito SM. Miniaturization, genome size, and biological size in a diverse clade of salamanders. Am Nat 2020;196:634–48.
[7] Weber JA, Park SG, Luria V, Jeon S, Kim HM, Jeon Y, et al. The whale shark genome reveals how genomic and physiological properties scale with body size. Proc Natl Acad Sci U S A 2020;117:20662–71.
[8] Liedtke HC, Gower DJ, Wilkinson M, Gomez-Mestre I. Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. Nat Ecol Evol 2018;2:1792–9.
[9] Organ CL, Shedlock AM, Meade A, Pagel M, Edwards SV. Origin of avian genome size and structure in non-avian dinosaurs. Nature 2007;446:180–4.
[10] Hughes AL, Hughes MK. Small genomes for better flyers. Nature 1995;377:391.
[11] Ritchie H, Jamieson AJ, Piertney SB. Genome size variation in deep-sea amphipods. R Soc Open Sci 2017;4:170862.
[12] Wyngaard GA, Rasch EM, Manning NM, Gasser K, Domangue R. The relationship between genome size, development rate, and body size in copepods. Hydrobiologia 2005;532:123–37.
[13] Appeltans W, Ahyong ST, Anderson G, Angel MV, Artois T, Bailly N, et al. The magnitude of global marine species diversity. Curr Biol 2012;22:2189–202.
[14] Wanninger A, Wollesen T. The evolution of molluscs. Biol Rev 2019;94:102–15.
[15] Minchin D. Introductions: some biological and ecological characteristics of scallops. Aquat Living Resour 2003;16:521–32.
[16] Alejandrino A, Puslednik L, Serb JM. Convergent and parallel evolution in life habit of the scallops (Bivalvia: Pectinidae). BMC Evol Biol 2011;11:164.
[17] Tremblay I, Guderley HE. Possible prediction of scallop swimming styles from shell and adductor muscle morphology. J Shellfish Res 2017;36:17–30.
[18] Tremblay I, Guderley HE, Himmelman JH. Swimming away or clamming up: the use of phasic and tonic adductor muscles during escape responses varies with shell morphology in scallops. J Exp Biol 2012;215:4131–43.
[19] Wang S, Zhang JB, Jiao WQ, Li J, Xun XG, Sun Y, et al. Scallop genome provides insights into evolution of bilaterian karyotype and development. Nat Ecol Evol 2017;1:120.
[20] Li YL, Sun XQ, Hu XL, Xun XG, Zhang JB, Guo XM, et al. Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins. Nat Commun 2017;8:1721.
[21] Li C, Liu X, Liu B, Ma B, Liu FQ, Liu GL, et al. Draft genome of the peruvian scallop *Argopecten purpuratus*. Gigascience 2018;7:giy031.
[22] Liu X, Li C, Chen M, Liu B, Yan XJ, Ning JH, et al. Draft genomes of two Atlantic bay scallop subspecies *Argopecten irradians irradians* and *A. i. concentricus*. Sci Data 2020;7:99.
[23] Zeng Q, Liu J, Wang C, Wang H, Zhang L, Hu J, et al. High-quality reannotation of the king scallop genome reveals no 'gene-rich' feature and evolution of toxin resistance. Comput Struct Biotechnol J 2021;19:4954–60.
[24] Kenny NJ, McCarthy SA, Dudchenko O, James K, Betteridge E, Corton C, et al. The gene-rich genome of the scallop *Pecten maximus*. Gigascience 2020;9:giaa037.
[25] Serb JM, Sherratt E, Alejandrino A, Adams DC. Phylogenetic convergence and multiple shell shape optima for gliding scallops (Bivalvia: Pectinidae). J Evol Biol 2017;30:1736–47.
[26] Stanley M. Functional morphology and evolution of byssally attached bivalve mollusks. J Paleontol 1972;46:165–212.
[27] Vinogradov AE. Nucleotypic effect in homeotherms: body-mass independent resting metabolic rate of passerine birds is related to genome size. Evolution 1997;51:220–5.
[28] Heilmayer O, Brey T, Portner HO. Growth efficiency and temperature in scallops: a comparative analysis of species adapted to different temperatures. Funct Ecol 2004;18:641–7.
[29] Peck LS. A cold limit to adaptation in the sea. Trends Ecol Evol 2016;31:13–26.
[30] Liu FY, Li YL, Yu HW, Zhang LL, Hu JJ, Bao ZM, et al. MolluscDB: an integrated functional and evolutionary genomics database for the hyper-diverse animal phylum Mollusca. Nucleic Acids Res 2021;49:D1556.
[31] Mallo M, Alonso CR. The regulation of *hox* gene expression during animal development. Development 2013;140:3951–63.
[32] Simakov O, Marletaz F, Cho SJ, Edsinger-Gonzales E, Havlak P, Hellsten U, et al. Insights into bilaterian evolution from three spiralian genomes. Nature 2013;493:526–31.
[33] Maxwell PH, Curcio MJ. Host factors that control long terminal repeat retrotransposons in *Saccharomyces cerevisiae*: Implications for regulation of mammalian retroviruses. Eukaryot Cell 2007;6:1069–80.
[34] Goodier JL. Restricting retrotransposons: a review. Mob DNA 2016;7:16.

[35] Mitchell P, Moyle J. Chemiosmotic hypothesis of oxidative phosphorylation. Nature 1967;213:137–9.
[36] Suarez RK. Hummingbird flight: sustaining the highest mass-specific metabolic rates among vertebrates. Experientia 1992;48:565–70.
[37] Suarez RK, Lighton JR, Brown GS, Mathieu-Costello O. Mitochondrial respiration in hummingbird flight muscles. Proc Natl Acad Sci U S A 1991;88:4870–3.
[38] Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 2014;346:1320–31.
[39] Heyn P, Kalinka AT, Tomancak P, Neugebauer KM. Introns and gene expression: cellular constraints, transcriptional regulation, and evolutionary consequences. Bioessays 2015;37:148–54.
[40] Yarra T, Blaxter M, Clark MS. A bivalve biomineralization toolbox. Mol Biol Evol 2021;38:4043–55.
[41] Arivalagan J, Yarra T, Marie B, Sleight VA, Duvernois-Berthet E, Clark MS, et al. Insights from the shell proteome: niomineralization to adaptation. Mol Biol Evol 2017;34:66–77.
[42] Nguyen JKB, Eames BF. Evolutionary repression of chondrogenic genes in the vertebrate osteoblast. FEBS J 2020;287:4354–61.
[43] Chong L. Molecular cloning - a laboratory manual, 3rd edition. Science 2001;292:446.
[44] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–20.
[45] Marcais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. Bioinformatics 2011;27:764–70.
[46] Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun 2020;11:1432.
[47] Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. Nat Methods 2020;17:155–8.
[48] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60.
[49] Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 2014;9:e112963.
[50] Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol 2015;16:259.
[51] Burton JN, Adey A, Patwardhan RP, Qiu RL, Kitzman JO, Shendure J. Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. Nat Biotechnol 2013;31:1119–25.
[52] Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol 2018;35:543–8.
[53] Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 1999;27:573–80.
[54] Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A 2020;117:9451–7.
[55] Bao WD, Kojima KK, Kohany O. Repbase update, a database of repetitive elements in eukaryotic genomes. Mob DNA 2015;6:11.
[56] Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics 2009;Chapter 4:4.10.1–14.
[57] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955–64.
[58] Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 2013;29:2933–5.
[59] Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. Nucleic Acids Res 2003;31:439–41.
[60] Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res 2005;33:W465–7.
[61] Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. Bioinformatics 2004;20:2878–9.
[62] Korf I. Gene finding in novel genomes. BMC Bioinformatics 2004;5:59.
[63] Blanco E, Parra G, Guigó R. Using geneid to identify genes. Curr Protoc Bioinformatics 2007;Chapter 4:Unit 4.3.
[64] Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. J Mol Biol 1997;268:78–94.
[65] Birney E, Clamp M, Durbin R. GeneWise and genomewise. Genome Res 2004;14:988–95.
[66] Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 2009;25:1105–11.
[67] Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 2010;28:511–5.
[68] Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol 2008;9:R7.
[69] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. Nat Genet 2000;25:25–9.
[70] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27–30.
[71] Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 2000;28:45–8.
[72] Hu XL, Bao ZM, Hu JJ, Shao MY, Zhang LL, Bi K, et al. Cloning and characterization of tryptophan 2,3-dioxygenase gene of Zhikong scallop *Chlamys farreri*. Aquac Res 2006;37:1187–94.
[73] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21.
[74] Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30:923–30.
[75] Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 2007;7:214.
[76] Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 2015;16:157.
[77] Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. Nucleic Acids Res 2002;30:3059–66.
[78] Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 2000;17:540–52.
[79] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312–3.
[80] Yang ZH. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 2007;24:1586–91.
[81] Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol 2017;34:1812–9.
[82] Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 1980;16:111–20.
[83] Chen SA, Yang PC, Jiang F, Wei YY, Ma ZY, Kang L. *De novo* analysis of transcriptome dynamics in the migratory locust during the development of phase traits. PLoS One 2010;5:e15633.
[84] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559.
[85] Chen M, Ma Y, Wu S, Zheng X, Kang H, Sang J, et al. Genome Warehouse: a public repository housing genome-scale sata. Genomics Proteomics Bioinformatics 2021;19:584–9.
[86] Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang A, et al. The Genome Sequence Archive Family: toward explosive data growth and diverse data types. Genomics Proteomics Bioinformatics 2021;19:578–83.