



ORIGINAL RESEARCH

Comparative Genomics Reveals Evolutionary Drivers of Sessile Life and Left-right Shell Asymmetry in Bivalves



Yang Zhang^{1,2,#}, Fan Mao^{1,2,#}, Shu Xiao^{1,2,#}, Haiyan Yu^{3,#}, Zhiming Xiang^{1,2,#}, Fei Xu⁴, Jun Li^{1,2}, Lili Wang³, Yuanyan Xiong⁵, Mengqiu Chen⁵, Yongbo Bao⁶, Yuewen Deng⁷, Quan Huo⁸, Lyping Zhang^{1,2}, Wenguang Liu^{1,2}, Xuming Li³, Haitao Ma^{1,2}, Yuehuan Zhang^{1,2}, Xiyu Mu³, Min Liu³, Hongkun Zheng^{3,*}, Nai-Kei Wong^{1,9,*}, Ziniu Yu^{1,2,*}

¹ CAS Key Laboratory of Tropical Marine Bio-resources and Ecology and Guangdong Provincial Key Laboratory of Applied Marine Biology, Innovation Academy of South China Sea Ecology and Environmental Engineering, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, China

² Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, China

³ Biomarker Technologies Co., Ltd, Beijing 101301, China

⁴ CAS Key Laboratory of Experimental Marine Biology, Center for Mega-Science, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

⁵ State Key Laboratory of Biocontrol, College of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China

⁶ Zhejiang Key Laboratory of Aquatic Germplasm Resources, College of Biological and Environmental Sciences, Zhejiang Wanli University, Ningbo 315100, China

⁷ College of Fisheries, Guangdong Ocean University, Zhanjiang 524088, China

⁸ Hebei Key Laboratory of Applied Chemistry, College of Environmental and Chemical Engineering, Yanshan University, Qinhuangdao 066044, China

⁹ Department of Pharmacology, Shantou University Medical College, Shantou 515041, China

Received 21 July 2020; revised 13 October 2021; accepted 1 November 2021

Available online 26 January 2022

Handled by Peng Cui

KEYWORDS

Comparative genomics;
Ostreoida oyster;

Abstract Bivalves are species-rich mollusks with prominent protective roles in coastal ecosystems. Across these ancient lineages, colony-founding larvae anchor themselves either by byssus production or by cemented **attachment**. The latter mode of sessile life is strongly molded by left-right **shell**

* Corresponding authors.

E-mail: carlzyu@scsio.ac.cn (Yu Z), wongnksz@163.com (Wong NK), zhenghk@biomarker.com.cn (Zheng H).

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.2021.10.005>

1672-0229 © 2022 The Authors. Published by Elsevier B.V. and Science Press on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Attachment;
Shell asymmetry;
Bivalve

asymmetry during larval development of **Ostreoida oysters** such as *Crassostrea hongkongensis*. Here, we sequenced the genome of *C. hongkongensis* in high resolution and compared it to reference bivalve genomes to unveil genomic determinants driving cemented attachment and shell asymmetry. Importantly, loss of the homeobox gene *Antennapedia (Antp)* and broad expansion of lineage-specific extracellular gene families are implicated in a shift from byssal to cemented attachment in bivalves. Comparative transcriptomic analysis shows a conspicuous divergence between left-right asymmetrical *C. hongkongensis* and symmetrical *Pinctada fucata* in their expression profiles. Especially, a couple of orthologous transcription factor genes and lineage-specific shell-related gene families including that encoding tyrosinases are elevated, and may cooperatively govern asymmetrical shell formation in *Ostreoida* oysters.

Introduction

Bivalves belong to the ancient lineages of Mollusca, comprising nearly 9600 species that thrive in aquatic environments, with notable economic and ecological importance [1]. As bilaterian organisms, they rely nutritionally on filtering phytoplankton and primarily follow a life cycle that transitions from free-swimming larvae to attached juveniles, culminating in sessile life [2]. Among filter-feeding bivalves, oysters of the superfamily *Ostreoida* serve as crucial guardians of marine ecosystems by forming oyster reefs that clean up the water and sustain biodiversity [3]. Due to climate change and coastal degradation, however, bivalves face profound challenges from warming waters and ocean acidification, which destabilize habitats, raise infection risks, and dampen the bivalve capacity of acquiring carbonate for shell formation [4].

To cope with diverse ecosystems, a variety of sessile strategies have emerged in bivalves during evolution, among which two modes of sessile life prevail. Characteristically, the majority of the bivalves, including Mytilidae (mussel), Pectinidae (scallop), and Pteriidae (pearl oyster), secrete adhesive byssal threads to stabilize themselves against marine turbulences [5,6]. In contrast, *Ostreoida* oysters have evolved highly sophisticated machinery of cemented attachment by producing organic-inorganic hybrid adhesive substances in place of byssus, which allows them to permanently fuse the left shell with rock surfaces or shells of other individuals in intertidal zones [7]. Compared with byssus, cemented attachment exhibits superiority in physical adhesion and mechanical tension, enabling oysters to efficiently create and thrive in large reef communities [2]. Developmentally, as a salient feature of their exoskeleton, shell formation processes in bivalves are strongly molded by their preferences for sessile life [8]. Quite distinctively, byssally attached bivalve species tend to possess a bilaterally symmetrical shell, whereas cement-attached oysters present a high degree of phenotypic variability and morphological asymmetry characteristic of their radically distinct left-right (L/R) shells [8]. Nevertheless, the molecular mechanisms driving these extraordinary innovations in bivalve evolution remain enigmatic, particularly in genomic contexts.

The Hong Kong oyster (*Crassostrea hongkongensis*, first described as *Crassostrea rivularis* by Gould, 1861) is an economically valuable aquacultural species endemic to the South China coastline [9]. As an ideal model for studying shell asymmetry, *C. hongkongensis* larvae follows a typical developmental cycle of cemented attachment and asymmetrical differentiation of the L/R shells. In order to elucidate the genetic basis underpinning the evolution of bivalve sessile life and asymmetry of shell formation, we sequenced and analyzed the complete

genome of *C. hongkongensis* and performed comparative genomic analysis along with several other bivalve species, including two congeneric *Ostreoida* oysters, *Crassostrea gigas* (Pacific oyster) and *Crassostrea virginica* [6,10–12]. In addition, we monitored the transcriptomic changes of *C. hongkongensis* embryos during the critical window of larval attachment, and compared the expression patterns of asymmetry-related genes in the L/R mantles of adult *C. hongkongensis* and byssus-producing pearl oyster (*Pinctada fucata*). Our comparative genomic data and associated functional assays reveal extensive molecular adaptations across the oyster genome that support the evolutionary switch from byssal to cemented attachment and divergence from the symmetrical shell in *Ostreoida* oysters.

Results and discussion

Genome assembly and evolutionary analysis of *C. hongkongensis*

Efforts on genome sequencing and assembly are inherently challenging for many marine invertebrates such as mollusks, annelids, and platyhelminths due to their remarkable genetic heterozygosity (or polymorphisms) [10,11,13]. Based on *k*-mer analysis, the genome size of a single wild-stock Hong Kong oyster (*C. hongkongensis*) individual was estimated to be 695 Mb with 1.2% of heterozygosity (Figure S1), comparable to that of the Pacific oyster (1.3%) [10]. To circumvent limitations of short-read next-generation sequencing in assembling highly polymorphic genomes, PacBio sequencing combined with Illumina sequencing was used as the dominant mode of genome sequencing in our study. We first generated 23.25 Gb of raw PacBio reads and 147.25 Gb of Illumina reads, being equivalent to 31.9× and 201.8× genome coverages, respectively (Tables S1 and S2). Following stepwise optimization of assembly algorithms, these reads were assembled into a 729.6-Mb genome with a contig N50 size of 314.1 kb, a scaffold N50 size of 500.4 kb, and the longest contig spanning 2.37 Mb (Table S3). The contig N50 size of the oyster genome is at least one order of magnitude more expansive than those of published bivalve genomes (Table S4), demonstrating the superiority of long-read sequencing technologies in coping with high polymorphism in genome assembly of marine invertebrates. However, the assembled genome size was slightly larger than that estimated by *k*-mer analysis. Such discrepancy may reflect sequence preferences of Illumina reads. The high integrity and quality of the assembly were evidenced by a productive mapping of 97.57% of sequencing reads and a low single-nucleotide error rate (Tables S5 and S6). Moreover, benchmarking universal single-copy orthologs (BUSCO)

analysis confirmed a high degree of completeness (92.84%) for the assembled genome (Table S7), which is comparable to genome completeness of other published bivalves (Table S4).

To assemble the oyster genome to the chromosomal level, we generated ~ 44.4 million valid Hi-C interaction pairs with over 50× coverage (Table S8). Then, 690.39 Mb of genome sequences were anchored into 10 pseudo-chromosomes with Hi-C data, covering 94.66 % of the assembled genome (Figure 1A, Figure S2; Table S9). Among them, 648.56 Mb of genome sequences were reoriented and anchored into chromosomes, constituting 93.94% of the total anchored sequences (Table S9). Moreover, high consistency between Hi-C-based pseudo-chromosomes with the genetic map of one congeneric species, *C. gigas*, was confirmed ($P = 0.978\text{--}0.996$, Figure S3), implicating high reliability in chromosomal genome assembly. Overall, by leveraging PacBio and Hi-C-enhanced Illumina sequencing, a very high-quality and chromosome-anchored complete genome was obtained, thus providing a robust framework for subsequent exploration of oyster biology and evolution of bivalves.

For gene annotation, we predicted 30,021 protein-coding genes in the genome by integrating results from *ab initio* prediction, homology-based searches with reference genomes, and RNA-seq-assisted prediction (Table S10), with an estimated BUSCO completeness of 91.09% (Table S11). Of these, more than 97.97% (28,329/30,021) of the predicted genes were annotated in the public databases (Table S12). The gene number here was similar to that in a close relative species, *C. gigas* (28,027 genes) [10]. In addition, transposable elements (TEs) constituted 46.2% of the *C. hongkongensis* genome, among which the predominant type of TEs was class II Helitron

(12.4%, 90.4 Mb) (Table S13). Phylogenetic analysis showed that three Ostreoida oyster species (*C. hongkongensis*, *C. gigas*, and *C. virginica*) clustered together (Figure 1B) and that Ostreoida oyster speciation occurred around 92.1 million years ago (MYA), in agreement with evidence from mitochondrial genomes [14]. Within bivalves, Ostreoida oysters are closest to the Pteriidae oyster *P. fucata*, and their divergence time was estimated to be 357.5 MYA (Figure 1B). These results further verified the hypothesis that a common ancestor of primitive Ostreoida and Pteriidae oysters existed prior to the Permian-Triassic extinction event, whereas speciation of modern Ostreoida oysters began at the end of Cretaceous-Paleogene extinction event [15]. Consistently, comparative genomic synteny analysis showed that three Ostreoida oyster genomes had high genomic collinearity except for large intra-chromosomal inversions, but substantial inter-chromosomal translocations and rearrangements occurred between chromosomes of Ostreoida oysters and *P. fucata* (Figure S4), in agreement with their phylogenetic relationships and divergence time.

Loss of the *Antennapedia* gene in Ostreoida oysters

Radical changes toward a sessile life require evolutionary innovations in the anatomical organization. In contrast to byssus-producing bivalves [6], Ostreoida oysters do not possess a byssal gland or secrete byssus during their lifetime [16], though a vestigial foot transiently appears at the veliger stage and degenerates following attachment and metamorphosis (Figure 2A). Developmentally, the homeobox (*Hox*) genes are known for their crucial roles in regulating body-plan

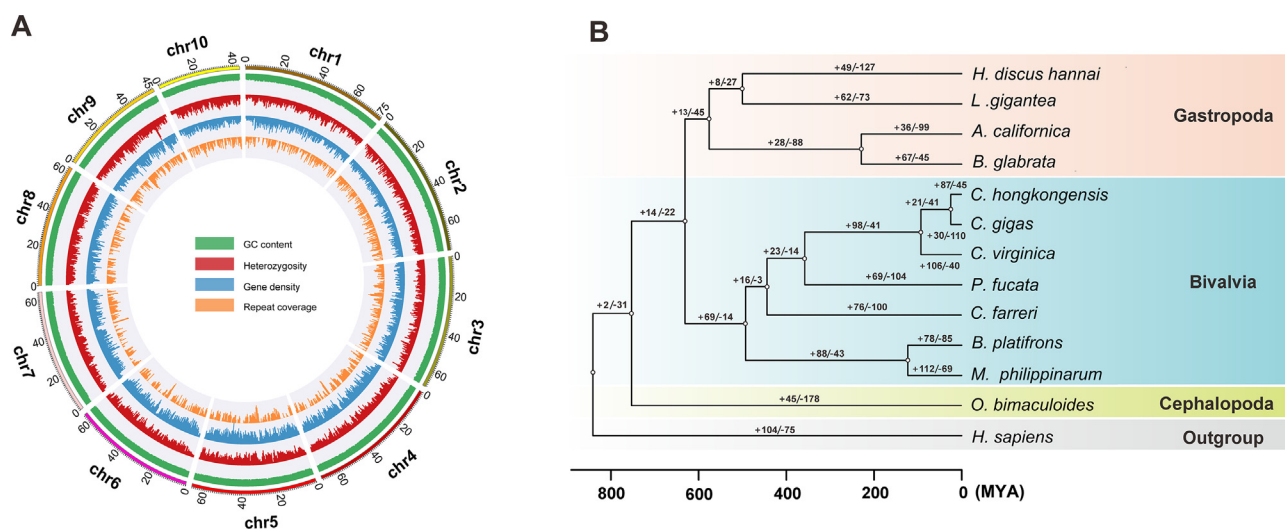


Figure 1 The genome landscape and phylogenetic analysis of the oyster *C. hongkongensis*

A. Circos plot highlights genome characteristics across 10 chromosomes in a megabase (Mb) scale. The GC content, global heterozygosity, gene density, and repeat coverage are presented from outer to inner circles in turn with non-overlapping 1 Mb sliding windows. **B.** Analysis on gene family expansion/contraction and divergence time across 12 representative mollusk species. A total of 87 gene families were expanded in *C. hongkongensis* (Hong Kong oyster). The human genome was set as outgroup. Three Ostreoida oyster species (*C. hongkongensis*, *C. gigas*, and *C. virginica*) were clustered together. Gene family expansion/contraction is indicated by a plus or minus sign. MYA, million years ago; *H. discus hannai*, *Haliotis discus hannai*; *L. gigantea*, *Lottia gigantea*; *A. californica*, *Aplysia californica*; *B. glabrata*, *Biomphalaria glabrata*; *C. hongkongensis*, *Crassostrea hongkongensis*; *C. gigas*, *Crassostrea gigas*; *C. virginica*, *Crassostrea virginica*; *P. fucata*, *Pinctada fucata*; *C. farreri*, *Chlamys farreri*; *B. platifrons*, *Bathymodiolus platifrons*; *M. philippinarum*, *Modiolus philippinarum*; *O. bimaculoides*, *Octopus bimaculoides*; *H. sapiens*, *Homo sapiens*.

development and organogenetic transitions in metazoans [17]. In view of this, we compared the clustering of *Hox* genes in byssus-producing and byssus-null bivalve species. A salient feature in byssal bivalves including *P. fucata*, *Mizuhopecten yessoensis*, *Chlamys farreri*, *Mytilus galloprovincialis*, *Bathymodiolus platifrons*, and *Modiolus philippinarum* is the intact *Hox* and *ParaHox* gene clusters (Figure 2B, Figure S5). In contrast, a disputed *Hox* gene cluster has been observed in *C. gigas* oyster genome [10], whereas a coherent *Hox* gene cluster is configured linearly in one single locus in both *C. hongkongensis* and *C. virginica*, probably in part due to the

fragmented genome assembly in *C. gigas*. Intriguingly, one of the key *Hox* members, *Antennapedia* (*Antp*), is lost in all three Ostreoida oysters (Figure 2B), implicating *Antp* as an essential driver of byssus formation. Sequence alignment reveals that *Antp* possesses a conserved homeobox domain in bivalves (Figure S6).

As evidenced in the expression profiles of three representative byssus-producing bivalve species, *Antp* and its orthologues are predominantly expressed in the byssal gland (Figure 2C). Due to the unavailability of molecular tools like clustered regularly interspaced short palindromic repeat-associated

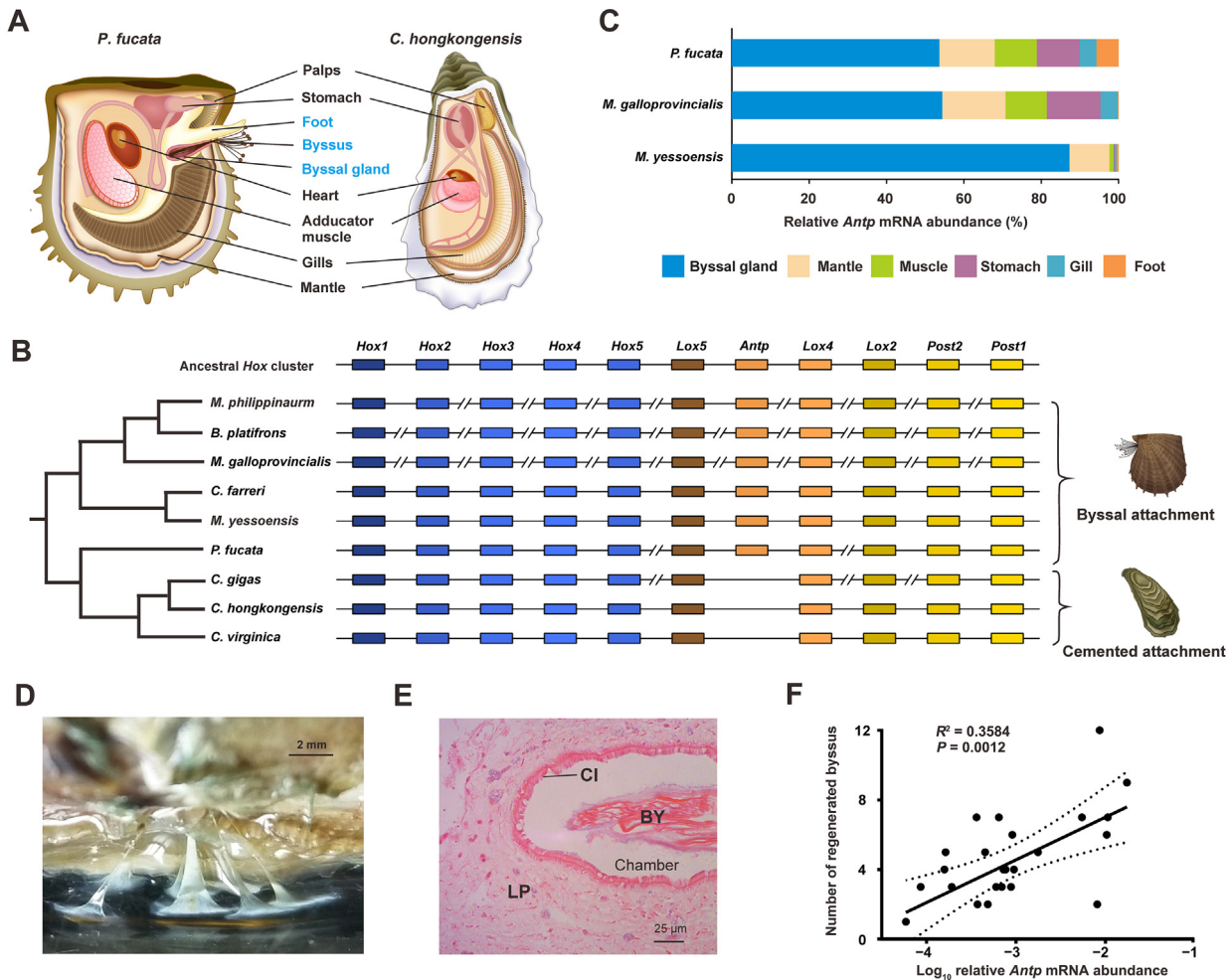


Figure 2 Loss of the *Hox* gene *Antennapedia* is implicated in an adaptive shift from byssal attachment to cemented attachment

A. Overview of the key body-plan organization in *P. fucata* and *C. hongkongensis*. *P. fucata* possesses a byssal gland and byssus, whereas adult individuals of Ostreoida oyster have lost their byssus gland and byssus. **B.** Comparison of the *Hox* cluster organization in bivalves with two distinct attachment styles, byssal attachment and cemented attachment. Unlike the disputed *Hox* gene cluster in *C. gigas* oyster genome, the *Hox* gene cluster is configured linearly in both *C. hongkongensis* and *C. virginica*. Intriguingly, *Antp* is lost in all three Ostreoida oysters. **C.** Tissue distribution of *Antp* orthologues in three byssally attached bivalves, *P. fucata*, *M. galloprovincialis*, and *M. yessoensis*. *Antp* mRNA abundance is displayed in percentage, and its expression in byssal gland accounts for more than 50%. **D.** Morphology of newly regenerated byssus 48 h after excision of original byssus. Scale bar, 2 mm. **E.** Anatomic analysis of the byssal gland of *P. fucata* with cross section. Vertical cross section of the byssal gland displaying CI, LP, and BY within a chamber. Scale bar, 25 μ m. **F.** Correlation between the relative *Antp* mRNA abundance and the number of regenerated byssus in *P. fucata*. The relative *Antp* mRNA abundance in the byssal gland was determined by real-time qPCR, while newly regenerated byssal threads were counted 48 h after excision of original byssus. Pearson's correlation coefficients and *P* values were calculated using two-tailed tests with 95% confidence. *Hox*, homobox; *Antp*, *Antennapedia*; *M. galloprovincialis*, *Mytilus galloprovincialis*; *M. yessoensis*, *Mizuhopecten yessoensis*; CI, ciliated wall; BY, byssal remnant; LP, lamina propria.

protein-9 (CRISPR/Cas9) or transcription activator-like effector nucleases (TALEN) for manipulating bivalve genomes, genetic ablation of the *Antp* gene is not yet feasible in pearl oyster for phenotypic appraisal of its function. However, histological evidence suggests that byssal gland is one of the appendage organs capable of secreting thin extended byssal threads in their mature form as observable byssus outside the organism (Figure 2D and E). Based on the fact that regenerative ability varies among individuals, we assessed *Antp* function in this phenotypic trait. Remarkably, mRNA expression levels of *Antp* are highly correlated with the number of regenerative byssus in the pearl oyster ($n = 24$, $R^2 = 0.3584$, $P = 0.0012$; Figure 2F). Taken together, our evidence strongly implicates that *Antp* is a transcriptional regulator central to byssal secretion in *P. fucata*. Further, the loss of the *Antp* gene seems to be associated with a physical loss of byssal gland in oysters. From an evolutionary perspective, *Antp* seems to play a critical role in appendage diversification in arthropods, which has previously been evidenced by its involvement in the leg formation in the crustacean *Daphnia* [18], and in the repression of the abdominal limb in the spider *Achaearanea tepidariorum* [19]. In addition, ectopic expression of *Antp* in the silkworm (*Bombyx mori*) induced the expression of the *sericin-1* gene in the posterior silk gland [20]. Collectively, these findings support a conserved function of *Antp* in secretory appendage in two distinct lineages, mollusks and arthropods.

Gene expansion and oyster attachment

In place of byssal attachment, Ostreoida oysters adopt an ingeniously cost-effective way of sessile life, namely, cemented attachment [16]. Such adhesive mechanism is characterized by extraordinary mechanical strength and superior flexibility needed to resist powerful tidal scour and absorb surge energy [21]. Cemented attachment allows oysters to efficiently anchor and thrive in marine environments and ultimately supports the genesis and health of oyster reefs. Nevertheless, the molecular mechanisms underlying oyster adhesive production remain enigmatic. Taking into account that cemented attachment is an innovation unique to Ostreoida oysters, we first ventured to investigate which gene families were expanded as a common event in three Ostreoida species. Our results showed that in *C. gigas*, *C. hongkongensis*, and *C. virginica*, there are 58, 172, and 321 species-specific expanded gene families, respectively, which can be further reduced to 32 core expanded gene families in Ostreoida oyster genomes (Figure 3A, Figure S7; Table S14).

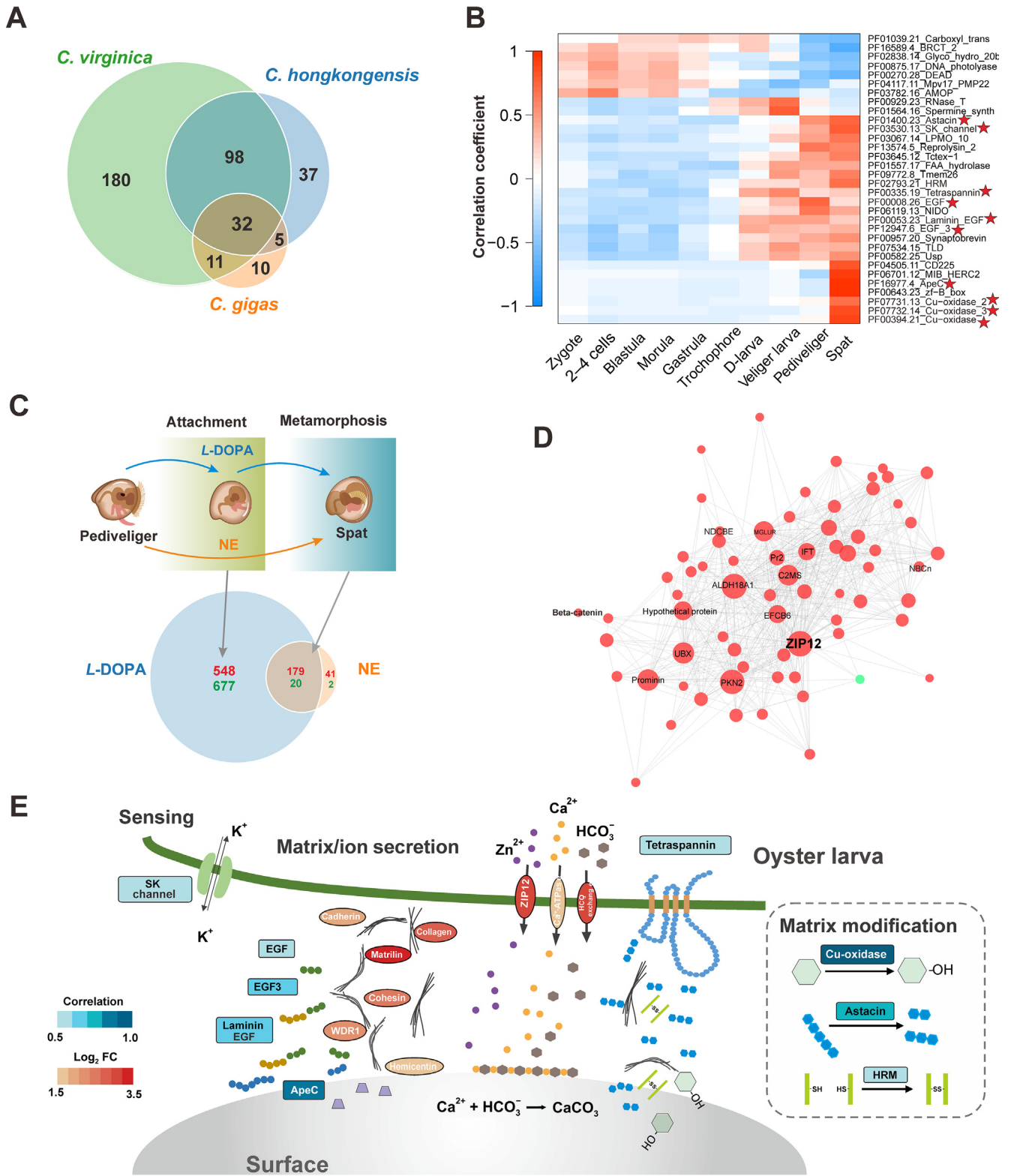
To elucidate how the expansion of these core gene families facilitates cemented attachment, we analyzed the correlations between their expression levels and specific developmental stages (Figure 3B; Table S15). Developmentally, attachment is an intricate secretion process involving a broad spectrum of chemical reactions and proteins, notably extracellular enzymes or matrices [22]. It was thus unsurprising that a small-conductance calcium-activated potassium (SK) channel gene family and nine extracellular gene families were identified to be involved in this process, which showed high correlations in the pediveliger and spat stages corresponding to the attachment initiation in larvae (Figure 3B). SK channels are widely expressed calcium-activated potassium channels in neurons, playing crucial roles in regulating dendritic excitability and synaptic plasticity [23]. Interestingly, increased expression of

expanded SK channels may aid free-swimming larvae in sensing external environments in search of an appropriate attachment site. On the other hand, the function of extracellular gene families is strictly related to key processes of shell attachment, including matrix secretion [epidermal growth factor (*Egf*), *Egf3*, lamin *Egf*, and *ApeC*], processing of matrix modification (*Cu-oxidase*, *Cu-oxidase2*, *Cu-oxidase3*, and *Astacin*), among others (Figure 3B). Indeed, many adhesive proteins contain specific protein-binding domains [24], such as EGF-like domains in the slug mucus proteins (e.g., SM40 and SM85) [25] and sea star footprint proteins (e.g., SF1) [26], raising the possibility that EGF family expansion in *C. hongkongensis* is functionally linked to the cemented attachment. Additionally, physico-chemical properties of many adhesive proteins arise in part from post-translational modifications, which ultimately support their adhesive functions [24]. Protein oxidation in marine bio-adhesives indeed contributes to enhanced crosslinking between shell disks and substrates during attachment [27]. A notable gene expansion in the copper oxidase family is likely to contribute to the stabilization of extracellular matrixes in the form of crosslinking between the oyster shell and external substrates. Copper-based enzyme lysyl oxidase is known to be essential for crosslinking and strengthening fibers in animal connective tissues via collagen oxidation [28]. Concomitantly, copper ion, as part of oxidative enzymes, is a mandatory cofactor for oxidase activity, which creates crosslinking sites from common amino acids, to enhance the cemented attachment [29,30]. Further transcriptomic analysis showed that the nine extracellular gene families were significantly up-regulated during the larvae-spat transformation of embryo development stages (Figure S8), confirming their functional importance in attachment formation.

L-3,4-dihydroxyphenylalanine-induced attachment

During larvae-spat transformation, embryonic oysters execute an intrinsic program of developmental changes, in which cemented attachment is tightly coupled to metamorphosis [31]. In this context, we set out to distinguish molecular determinants of cemented attachment from that of metamorphosis at the pediveliger stage by means of two pharmacologic agents: L-3,4-dihydroxyphenylalanine (*L*-DOPA) and norepinephrine (NE). The former simultaneously promotes normal attachment and metamorphosis, whereas the latter induces metamorphosis only but not attachment (Figure 3C, Figure S9) [31]. Based on this, a gene whose expression was up-/down-regulated by *L*-DOPA treatment rather than NE treatment was hypothesized to be a driver for attachment initiation in *C. hongkongensis*. We accordingly scrutinized 24 transcriptomes following pharmacological challenges at two time points within the temporal span of oyster attachment. Our results show that the expression of 1225 genes was specifically altered by the treatment of *L*-DOPA rather than NE (Figure 3C), confirming the essential role of *L*-DOPA as an attachment signal.

Remarkably, the expression levels of several genes encoding neurotransmitter receptors (including metabotropic glutamate receptor and neuropeptide Y receptor) were significantly increased, consistent with the assumption that neuromuscular coordination is essential for guiding embryos to settle in suitable niches and initiate attachment (Figure S10) [32]. Moreover, genes encoding metal ion channels or binding proteins



were significantly enriched, with notable examples like organic cation transporters, transient receptor potential cation channels (e.g., ZIP12), and voltage-dependent calcium channels (e.g., Ca^{2+} -ATPase), which is intuitively consistent with the well-documented stimulatory roles of selective cations in oyster larval settling. To highlight, potassium voltage-gated channel activity has been proven to be vital for oyster larval attachment, since its inhibitor tetraethyl ammonium can effectively block this developmental process [33]. Typically, attachment initiates in oyster larvae with the aid of fibrous adhesive proteins and other bioorganic substances, including mucopolysaccharides and phospholipids [2]. As a consequence, extensive extracellular matrix and adhesion proteins including collagen, cadherin, fibrocystin, and hemimentin would increase in response to *L*-DOPA stimulation, presumably paving the way for larval attachment [34].

To search out the crucial molecular determinants governing this process, we performed weighted correlation network analysis (WGCNA) to construct a potential connected gene network functionally associated with *L*-DOPA-induced attachment, wherein 15 modules were subsequently identified (Figure S11). Among them, the MEpink module was most correlated with *L*-DOPA-induced attachment ($P < 0.01$) and contained 139 genes (topological overlap > 0.3). Intriguingly, within this module, a hub forming the most connections in the network was found to be the zinc transporter ZIP12 (Figure 3D), which is a pivotal regulator of zinc flux. As a cofactor essential to a wide spectrum of proteins such as matrix metalloproteinases, zinc plays vital regulatory roles in enzymatic catalysis and macromolecular stability [35]. A high abundance of zinc is also a salient feature in aragonite- or calcite-rich shells in certain mollusks [36]. Meanwhile, among the gene families that were specifically expanded in *Ostreoida* oysters, *Astacin* encodes a cell-secreted or plasma membrane-associated protease that possesses zinc-binding activity and takes part in proteolytic processing of extracellular proteins [37]. Its expression was markedly elevated both during larvae-spat transformation and larval response to *L*-DOPA treatment (Figures S8G and S10D). Predictably, chelation of zinc potentially retarded oyster larval attachment (Figure S12), providing additional hints that the initial creation of matrix structures requires zinc and associated protein activities for cement attachment. Accordingly, based on genomic results on extracellular gene family expansion and transcriptomic

profiles at the attachment stage, we conceived a conceptual model to delineate the mechanistic determinants and processes working in the cement attachment strategy of oyster larvae (Figure 3E). We postulate that attachment formation apparently results from intricate coordination of at least three types of fundamental activities, namely, larval sensing of habitable surfaces, matrix/ion secretion, and matrix modification to mobilize adhesive processes.

Asymmetry in left-right shell formation

Symmetry is an elegant guiding principle for the implementation of body plans [38]. Across the *Bivalvia* class, the majority of bivalves display perfect or near-perfect conformity to bilaterally symmetrical shells [8,39]. In contrast, *Ostreoida* oysters may appear unorthodox in adopting morphological asymmetry in their shell formation due to functional differentiation of the L/R shells (Figure S13). The left shell is visibly much thicker and more convex than its right counterpart, which is apt for attaching to rocky surfaces or neighboring oysters within a reef community. On the other hand, the right shell is capable of physical displacement and hermetic lockdown to regulate water intake and ward off predation (Figure 4A). Moreover, structural variance in shell asymmetry is also amply reflected by a greater proportion of prismatic layers in the right shell (Figure S14), which are responsible for controlling the initiation of calcite crystal formation and growth [40]. Although asymmetry of body forms has been traditionally stereotyped as defects that may jeopardize the survival of an organism [41], the example of *Ostreoida* oysters clearly defies this rule. We reason that such an intriguing differentiation of asymmetrical shells could confer unexpected benefits such as improved population fitness in an otherwise intrinsically harsh coastal environment. With the advent of the left shell and its versatile attachment machinery, oysters can easily economize resources or secure their foothold on rocks or peers' shells within an oyster reef via cemented attachment [42]. This strategy permits oysters to lower their thresholds for founding and expanding productive colonies in demanding physical habitats, literally through the stacking of individuals at high densities, without sacrificing resistance to environmental challenges such as tidal turbulences.

To further elucidate the molecular basis of left-right asymmetry, comparative transcriptomic analysis was carried out to

Figure 3 Molecular basis of attachment initiation in *Ostreoida* oysters

A. Venn plot showing the commonly expanded gene families in three *Ostreoida* oyster species, *C. hongkongensis*, *C. virginica*, and *C. gigas*. A total of 32 core expanded gene families were identified. **B.** Heatmap illustrating the correlation between the expression levels of 32 core expanded gene families and the developmental stages of *C. hongkongensis* larvae. The high correlations between transcriptionally activated gene families and stages corresponding to attachment initiation in larvae are marked by red stars. **C.** Pharmacological responses of oyster pediveliger larvae during attachment initiation and metamorphosis. *L*-DOPA stimulates larval attachment and metamorphosis, while NE only induces metamorphosis without attachment. Venn plot shows the numbers of genes up-regulated (red) or down-regulated (green) by *L*-DOPA/NE treatment. The genes specifically altered by *L*-DOPA treatment may participate in attachment initiation. **D.** The coordinated gene network based on the zinc transporter ZIP12, which is a hub forming the highest degree of gene connections in WGCNA. Red and green dots indicate up-regulated and down-regulated genes, respectively. **E.** Schematic diagram conceptualizing the molecular basis for initiation of larval attachment in oysters. Square boxes indicate oyster-specific expanded gene families involved in larval attachment ($P < 0.001$). Filled color (blue) is scaled with correlation values at the spat stage. Ellipse boxes indicate *L*-DOPA specifically induced genes after *L*-DOPA treatment, which are filled in red scaled with values in \log_2 FC. *L*-DOPA, *L*-3,4-dihydroxyphenylalanine; NE, neopinephrine; WGCNA, weighted correlation network analysis; FC, fold change.

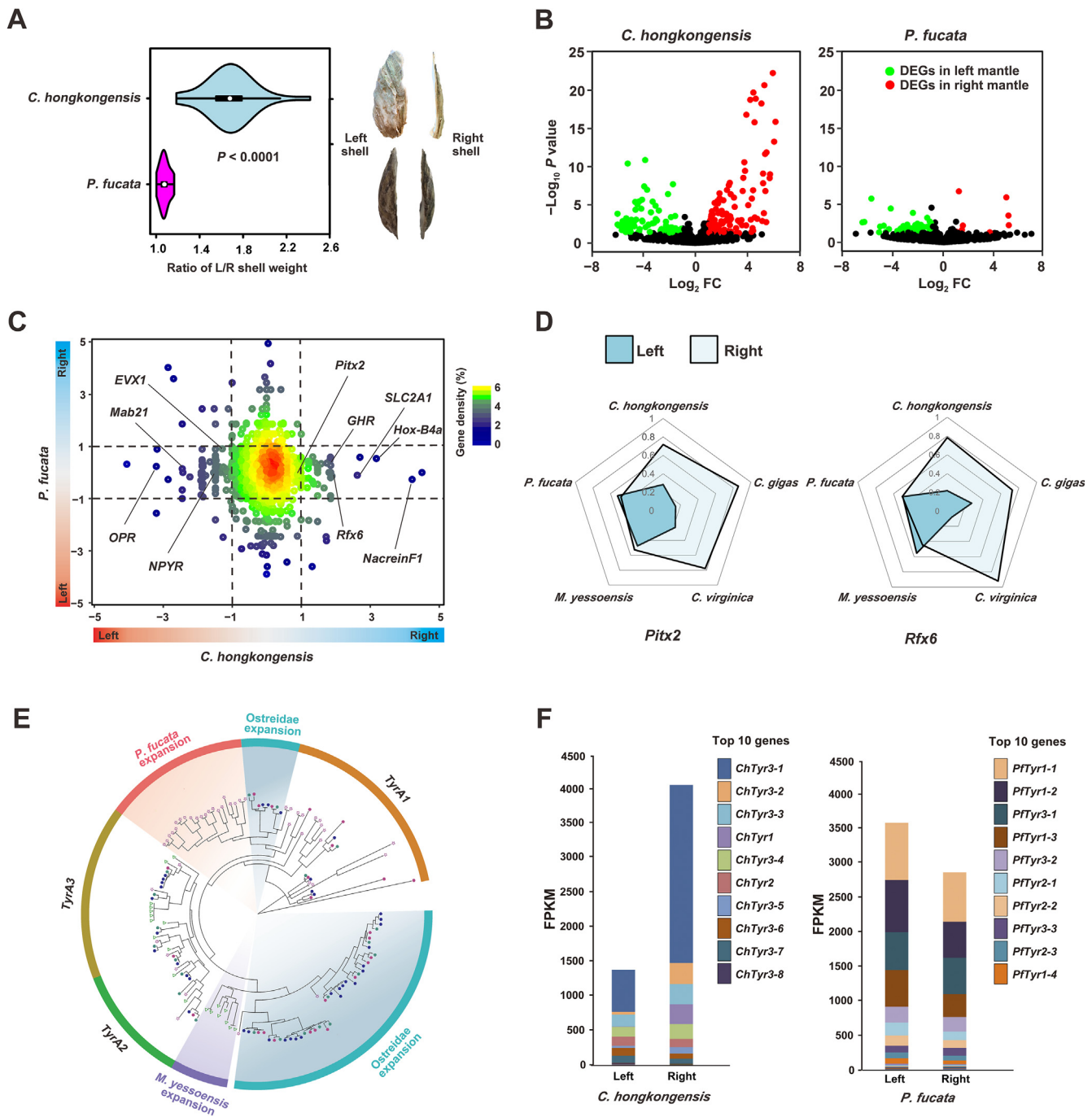


Figure 4 Left-right asymmetry of shell formation in Ostreoida oysters

A. Comparison of the ratio of L/R shell weight and morphology between *C. hongkongensis* and *P. fucata*. **B.** Volcano plot showing the left- and right-mantle DEGs, which are filtered by $|\log_2 \text{FC}| \geq 1$ with $P < 0.05$. **C.** Expression profile of paired orthologues in L/R mantles of *C. hongkongensis* and *P. fucata*. A total of 10,491 orthologues were paired and only a few asymmetry-related orthologues were specifically expressed in *C. hongkongensis*. The x- and y-axes indicate $\log_2 \text{FC}$ of expression ratio in R/L mantles of *C. hongkongensis* and *P. fucata*, respectively. **D.** Expression patterns of two pivotal transcription factor genes in left and right mantles across five bivalves. **E.** Dendrogram of known *Tyr* genes from five mollusks constructed by maximum likelihood method. Bivalve and molluscan *TyrA* orthologous groups are indicated by curvatures and annotated as A1–A3. Specific *Tyr* orthologous groups are marked with colored backgrounds and annotated with a species' name. Species are represented with different shapes: triangle, *M. yessoensis*; circle, Ostreidae; pentagon, *P. fucata*. **F.** Expression patterns of *Tyr* gene families in left and right mantles of two bivalve species, as determined by FPKM. Total FPKM of different types of orthologous genes is displayed in cumulative histograms. Different members of *Tyr* gene families are presented with different colors. L/R, left-right; DEG, differentially expressed gene; *Tyr*, tyrosinase; FPKM, fragments per kilobase per million.

quantify the gene expression profiles in the L/R mantles of *C. hongkongensis* and pearl oyster, which are the key organ controlling shell formation [43]. As expected, 188 asymmetry-related differentially expressed genes (DEGs) were identified in the L/R mantles of *C. hongkongensis*, whereas only 53 asymmetry-related DEGs were found in the L/R mantles of the pearl oyster (Figure 4B), which reflects a radical genetic divergence underpinning shell asymmetry. Next, to test the hypothesis that lineage-specific divergence of orthologues contributes to symmetry breakage, 10,491 of the orthologues were paired between the two species. Our results indicate that a few but crucial asymmetry-related orthologues are specifically expressed *C. hongkongensis* (Figure 4C), including homeobox gene paired-like homeodomain transcription factor (*Pitx2*), homeobox B4a (*Hox-B4a*), and regulatory factor *X6* (*Rfx6*). Notably, *Pitx2* is a central regulator orchestrating the Nodal cascade, which is responsible not only for directing L/R axis formation in mammals [44], but also for shell coiling and L/R asymmetry in some mollusks such as the snail [45]. Another gene of interest is *Rfx6*, recognized for its fundamental importance in guiding pancreatic islet development and insulin production in mammals [46]. While the insulin-related peptide gene is known for being a critical driver of oyster growth [47], this new evidence suggests novel roles of *Rfx6*–insulin signaling in maintaining shell asymmetry in oysters. As predicted, asymmetry-related expression of *Pitx2* and *Rfx6* in L/R mantles was confirmed by real-time qPCR in three Ostreoida lineages with asymmetrical shells, whereas such gene expression patterns were absent in three symmetrical bivalves, pearl oyster, scallop, and mussel (Figure 4D).

However, it should be noted that the majority of asymmetry-related genes in *C. hongkongensis* are not orthologous to those in the pearl oyster. For example, tyrosinase (*Tyr*) genes are one of the key gene families involved in steering shell formation and pigmentation by means of oxidation and crosslinking of o-diphenols [48]. Phylogenetic analysis revealed that more than half of *Tyr* genes (55%) clustered in several lineage-restricted clades, suggesting the rapid and independent expansion of this gene family in bivalves (Figure 4E). Remarkably, several high-abundance members of the *Tyr* family seemed to be strongly associated with L/R asymmetry and were expressed preferentially in the right mantles of *C. hongkongensis*, whereas no obvious variance was noted between left and right mantles in the pearl oyster (Figure 4F). Therefore, it seems logical to infer that rapid expansion and divergent expression of the *Tyr* gene family contribute importantly to the emergence and neofunctionalization of asymmetrical shell formation in Ostreoida lineages. Lastly, we found that 71.1% of these genes started expression at the spat stage (Figure S15), implying that a complete asymmetrical pattern becomes established in the juveniles only after metamorphosis.

Conclusion

Ostreoida oysters have evolved remarkable innovations for streamlining their body plans, which are enabled by novel cemented attachment and allied gene machinery diverging from L/R symmetry. These evolutionary breakthroughs poise oysters as highly successful reef builders and ecological guardians integral to marine ecosystems spanning the globe. To reveal the genomic changes driving these evolutionary

innovations, we sequenced the complete genome of *C. hongkongensis*, obtained active transcriptomic data developmentally critical to the attachment window, and made comparisons with other bivalve genomes. The *Anip* gene of the *Hox* cluster, found to be lost in Ostreoida oysters, is evidently a pivotal regulator of byssal secretion and expression of byssal proteins in *P. fucata*, and potentially a critical gene governing the radical switch from byssal to the cemented attachment. Furthermore, extensive extracellular gene families were expanded in the Ostreoida lineages specifically, presumably contributing to the operationalization of cemented attachment. Ion-binding genes were significantly enriched in L-DOPA-induced attachment in oysters, with zinc-binding genes being a prominent network that coordinates extracellular matrix modification and initiates adhesion. Moreover, Ostreoida divergence from shell symmetry is probably under the joint control of a suite of transcriptionally identified asymmetry-related DEGs in the L/R mantles, notably the transcription factor genes *Pitx2* and *Rfx6*, as well as the expanded lineage-specific family encoding tyrosinases. Thus, on the basis of genomic determinants and coordinated gene networks as revealed in this study, we have advanced a detailed picture of how shell asymmetry is switched on and driven in bivalves such as Ostreoida oysters. In order to provide insights into bivalve biology and disease in contexts of climate change or biological conservation, further investigation on the attachment-governing genes may be warranted.

Materials and methods

Illumina sequencing

Genomic DNA was extracted by using DNeasy Blood & Tissue Kit (Catalog No. 69582, Qiagen, Hilden, Germany) from a two-year-old single individual of *C. hongkongensis*. Two types of pair-end libraries (220 bp and 500 bp) and six types of long-insert mate-pair libraries (3 kb, 4 kb, 5 kb, 8 kb, 10 kb, and 15 kb) were constructed by using Illumina's paired-end and mate-end kits, according to the manufacturer's instructions. Libraries were sequenced on an Illumina HiSeq 2500 platform. For raw reads, sequencing adaptors were removed. Contaminated reads (such as chloroplast, mitochondrial, bacterial, and viral sequences) were screened by alignment in accordance with an NCBI-NR database by using Burrows-Wheeler alignment (BWA) v0.7.13 with default parameters. FastUniq v1.1 was used to remove duplicated read pairs. Low-quality reads were filtered out, according to the following criteria: 1) reads with $\geq 10\%$ unidentified nucleotides (N); 2) reads with > 10 nucleotides aligned to an adapter, allowing $\leq 10\%$ mismatches; 3) reads with $> 50\%$ bases with Phred quality < 5 .

PacBio sequencing

Genomic DNA was sheared by a g-TUBE device (Catalog No. 520079, Covaris, MA) with 20 kb settings. Sheared DNA was then purified and concentrated with AMPure XP beads (Catalog No. 10136224, Beckman Coulter, CA) and further used for single-molecule real-time (SMRT) bell preparation according to the manufacturer's protocol (Pacific Biosciences, CA) and 20 kb template preparation by using BluePippin size selection

(Sage Science). Size-selected and isolated SMRT bell fractions were purified with AMPure XP beads. Finally, these purified SMRT bells were used for primer and polymerase (P6) binding, according to the manufacturer's binding calculator (Pacific Biosciences). Single-molecule sequencing was performed on a PacBio RS-II platform with C4 chemistry. Only PacBio subreads ≥ 500 bp were included for performing oyster genome assembly.

Genome size estimation

About 34 Gb ($52\times$) corrected Illumina reads from the 180-bp and 500-bp libraries were selected to perform genome size estimation. The oyster genome size was estimated based on the formula: genome size = k -mer number/peak depth.

De novo genome assembly of Illumina data

Clean Illumina reads were assembled *de novo* into longer contigs by using ALLPATH-LG [49] with default parameters. Adjacent contigs were linked to scaffolds by leveraging mate-pair information with SSPACE v2.3 [50], while gaps were filled by using GapCloser v1.12 [50] implemented in a SOAPdenovo2 package [51].

De novo genome assembly of PacBio data

Canu+LoRDEC+WTDBG

We used an error correction module of Canu v1.5 [52] to select longer subreads with the settings 'genomeSize = 3,500,000,000' and 'corOutCoverage = 80', detected raw subreads overlapping through a highly sensitive overlapper MHAP v2.12 ('corMhapSensitivity = low/normal/high'), and completed an error correction through a falcon_sense method ('correctedErrorRate = 0.025'). Subsequently, output subreads of Canu were further corrected by LoRDEC v0.6 [53] with the parameters '-k 19 -s 3'. Based on these two rounds of error corrections, we generated a draft assembly by using WTDBG 1.1.006 (<https://github.com/ruanjue/wtdbg>) with the command 'wtdbg -i preads.fasta -t 64 -H -k 21 -S 1.02 -e 3 -o wtdbg'.

Hybrid genome assembly

Contigs produced by ALLPATH-LG were optimized with the aid of contigs of PacBio assembly by using quickmerge with the parameters '-hco 5.0 -c 1.5 -l 100,000 -ml 5000'. Optimized contigs were linked to scaffolds by leveraging Illumina mate-pair information by using SSPACE, and gaps were filled by using PBjelly v2.

Evaluation of oyster assembly

To estimate the genome quality, we first mapped Illumina reads to the oyster assembly by using BWA tool. Next, the completeness of genomes was verified by mapping 248 highly conserved eukaryotic genes and 908 BUSCOs in metazoa to the genomes by using CEGMA v2.5 [54] and BUSCO v3.0.2b [55], respectively.

Hi-C sequencing and assembly

Sequencing

According to the Hi-C procedure [56], nuclear DNA from muscles of oyster individuals was crosslinked, and then excised with a restriction enzyme, leaving pairs of distally located but physically intercalated DNA molecules attached to one another. The sticky ends of these digested fragments were biotinylated, which were then ligated to each other to form chimeric circles. Biotinylated circles, as chimeras of physically associated DNA molecules from the original crosslinking, were enriched, sheared, and sequenced [57]. After adaptor removal and filtering out low-quality reads, Hi-C reads were aligned to our assembled genome to evaluate ratio of mapped reads, distribution of insert fragments, sequencing coverage, and number of valid interaction pairs. Uniquely mapped reads spanning two digested fragments that are distally located but physically associated DNA molecules are defined as valid interaction pairs.

Assembly

Scaffolds of PacBio and Illumina assemblies were reduced to fragments with a length of 300 kb, which were then re-assembled by using the LACHESIS software [57] based on Hi-C data. Regions that failed to be restored to the original assembly or contained an average Hi-C data coverage of less than 0.5% were considered assembly errors and were broken into smaller scaffolds. Consistency in the assembly of Hi-C data based pseudo-chromosomes was assessed by comparisons with a genetic map for *C. gigas* [58] by using the software of ALLMAPS.

Genome annotation

Repetitive sequence prediction

Repeat composition of the assemblies was estimated by building a repeat library employing the *de novo* prediction programs LTR-FINDER, MITE-Hunter, RepeatScout, and PILER-DF. The database was classified by using PASTECClassifier [59] and then combined with the Repbase database [60] to create a final repeat library. Repeat sequences in the oyster genome were identified and classified by using the RepeatMasker program [61]. The classification criterion for long terminal repeat (LTR) family was defined as that 5'-LTR sequences of the same family would share at least 80% identity over at least 80% of their lengths.

Protein-coding gene prediction

Protein-coding genes were predicted based on *ab initio* prediction and protein homology-based approaches. The algorithms Genscan [62], Augustus [63], GlimmerHMM [64], GeneID [65], and SNAP [66] were used for *ab initio* prediction. Alignment of homologous peptides from *C. gigas*, *C. virginica*, *Lotia gigantea*, and *Danio rerio* to our assemblies was performed to identify homologous genes with the aid of GeMoMa [67]. Consensus gene models were generated by integrating the *ab initio* predictions and protein alignments using EVidence-Modeler (EVM) [68].

Functional annotation of protein-coding genes

Annotation of the predicted genes was performed by blasting their sequences against a number of nucleotide and protein sequence databases, including COG, KEGG, NCBI-NR, and Swiss-Prot, with an E-value cutoff of $1E-5$. Gene Ontology (GO) for each gene was assigned by using Blast2GO [69] based on NCBI databases.

Evolution of oysters

Protein sequences of *Haliotis discus hannai* [70], *L. gigantea* (GCF_000327385.1), *Aplysia californica* (GCF_000002075.1), *Biomphalaria glabrata* (GCF_000457365.1), *C. gigas* (GCF_000297895.1), *C. virginica* (GCF_002022765.2), *P. fucata* (<https://marinegenomics.oist.jp>), *C. farreri* (CfBase), *B. platifrons* (GCA_002080005.1), *M. philippinarum* (GCA_002080025.1), *Octopus bimaculoides* (GCF_001194135.1), and *Homo sapiens* (GCF_000001405.2) were retrieved for analysis. Proteomes of the aforementioned 12 species and that of *C. hongkongensis*, comprising a total of 295,905 protein sequences, were clustered into 38,939 orthologous groups by using OrthoMCL v3.1 [71] based on an all-to-all BLASTP strategy with an E-value of $1E-5$ and by using markov chain clustering (MCL) algorithms with default inflation parameters (1.5). Based on clustering results, *C. hongkongensis*-specific gene families were determined and annotated. To infer phylogenetic relationships, we extracted 387 single-copy gene families from all 13 species to perform multiple alignments of proteins for each family with MUSCLE v3.8.31 [72]. All of the alignments were combined into one supergene to construct a phylogenetic tree by using RAXML v8.2.12 [73] with 1000 rapid bootstrap analyses, followed by a search of the best-scoring Maximum likelihood tree in a single run. Finally, divergence time was estimated by using MCMCTree from the PAML package [74] in conjunction with a molecular clock model. Several reference-calibrated time points obtained from TimeTree database (<http://timetree.org/>) were used to date divergence time of interest. Expansion and contraction of OrthoMCL-derived homolog clusters were determined by CAFÉ v2.1 [75] on the basis of changes in gene family size with respect to phylogeny and species divergence time. In addition, we obtained domain-based expanded gene families of three *Crassostrea* species, according to previous work by Albertin and colleagues [76].

Syntenic analysis

All-to-all BLASTP analyses of protein sequences were performed between *C. hongkongensis*, *C. gigas*, *C. virginica*, and *P. fucata* with an E-value threshold set at $1E-5$. Syntenic regions within and between species were identified by using MCSan based on BLASTP results. A syntenic region was considered valid if it contained a minimum of 10 collinear genes and a maximum of 25 gaps (genes) between two adjacent collinear genes.

Hox gene analysis

Structures of *Hox* genes in oysters were determined by using the GeMoMa v1.4.2 software [77] with default parameters

based on available homeobox gene models. Predictions were handled by applying a GeMoMa annotation filter (GAF) with default parameters except for the evidence percentage filter ($e = 0.1$). These were then manually verified to achieve a single high-confidence transcript prediction per locus. Exact annotations of each homeobox gene were completed with the aid of phylogenetic relationships.

Transcriptomic analysis

Embryos at different developmental stages during oyster embryogenesis, including zygote, 2–4 cells, blastula, morula, gastrula, trochophore, D-larva, veliger, pediveliger, and spat, were collected for RNA isolation. Similarly, RNA extraction was done with various tissues, including hemocytes, muscles, gill, labial palp, hepatopancreas, gonads, and mantles. To compare asymmetry-related gene expression in the mantles of *C. hongkongensis* and *P. fucata*, their L/R mantles were collected. For both left and right mantles, unilateral tissues from five individuals were pooled as one sample, and each of the L/R mantle groups contained at least three replicates. Total RNA was isolated by using the Trizol reagent (Catalog No. 15596026, ThermoFisher Scientific, Waltham, MA), followed by treatment with RNase-free DNase I (Catalog No. M6101, Promega, WI), according to the manufacturers' instructions. RNA quality was then checked by using an Agilent 2100 Bioanalyzer. Illumina RNA-seq libraries were prepared and sequenced in a HiSeq 2500 system by a PE150 strategy following the manufacturer's instructions (Illumina, CA). After trimming raw reads based on quality scores from the quality trimming program Btrim, clean reads were aligned to the oyster assembly genome by using TopHat v2.1.1 [78] and then assembled by using Cufflinks v2.1.1 [79]. Differential expression of genes in various tissues was evaluated by using Cuffdiff [79].

WGCNA and co-expression network analysis

WGCNA [80] was applied to construct a weighted gene co-expression network for genes having a high correlation with cemented attachment. The top 10,000 differential genes exhibiting transcriptional changes in response to *L*-DOPA treatment were selected for WGCNA, wherein the modules showed a high correlation with cemented attachment. We estimated the weight for each pair of genes forming intersections within these modules and analyzed DEGs relevant to cemented attachment by using DESeq2. Cytoscape was used to delineate the co-expression network of significant gene pairs with weight > 0.3 .

Byssal regeneration

Functional relationships between *Anip* mRNA expression levels and phenotypic traits of byssal threads in adult pearl oysters (*P. fucata*) were explored. Briefly, 50–100 pearl oysters (2 years old) were collected and maintained in aerated laboratory tanks. Byssal mass comprising the byssal stem and existing old threads of pearl oysters were excised. Then, individual pearl oysters were placed in beakers (one oyster per beaker) to allow identification of subsequent regrowth of nascent thread mass. Particular care was taken in removing old threads and attachment discs from the shells. Preliminary

experiments indicate that removal of the threads did not affect subsequent thread formation. Byssal thread formation was estimated as the number of threads per oyster observed 24 h later.

Subsequently, the corresponding byssal gland of each pearl oyster was collected for RNA extraction by using TRIzol reagent, according to the manufacturer's instructions. Purified RNA samples were diluted to 1 µg/µl and pooled to perform cDNA synthesis by utilizing PrimerScript 1st Strand cDNA Synthesis Kit (Catalog No. 6110A, Takara, Japan), following the manufacturer's protocol. Real-time qPCR analysis was performed to determine *Antp* mRNA expression with gene-specific primers (Table S16).

Pharmacological treatment

Chemical compounds were obtained from Sigma-Aldrich, unless otherwise specified. Working solutions were freshly prepared in deionized (DI) water approximately 1 h before *in vivo* experiments, which were conducted in large beakers to allow observation of oyster attachment and metamorphosis. Groups of oyster larvae at the pediveliger stage were placed in three beakers containing 50 ml seawater (at a density of 20 larvae/ml). There were three groups in total: an unstimulated control, an *L*-DOPA-treated group, and a NE-treated group. Oyster larvae were challenged with different concentrations of NE (1×10^{-4} , 1×10^{-5} , 1×10^{-6} M) or *L*-DOPA (1×10^{-5} , 1×10^{-6} , 1×10^{-7} M). Previous studies have shown that this concentration range is sufficiently potent for inducing a larval response [81].

Then, oyster larvae were collected 6 h and 24 h after treatment for RNA-seq and transcriptomic analyses to determine any temporally driven differences between the *L*-DOPA-treated group (1×10^{-5} M) and the unstimulated control. By a similar design, oyster larvae were exposed to NE (1×10^{-5} M) for 6 h and 24 h, and their transcriptomic profiles were examined in relation to oyster metamorphosis.

Data availability

The raw sequence data reported in this study have been deposited in the Genome Sequence Archive [82] at the National Genomics Data Center (NGDC), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences (CAS) / China National Center for Bioinformation (CNCB) (GSA: CRA004099), and are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. The whole-genome sequence data reported in this study have been deposited in the Genome Warehouse [83] at the NGDC, BIG, CAS / CNCB (GWH: GWHBAZL00000000), and are publicly accessible at <https://ngdc.cncb.ac.cn/gwh>. The *C. hongkongensis* genome studied in this Hong Kong oyster genome project has also been deposited in the BioProject database at NCBI (BioProject: PRJNA592306), and are publicly accessible at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA592306>. Hi-C data have been deposited in the BioSample database at NCBI (BioSample: SAMN13420518), and are publicly accessible at <https://www.ncbi.nlm.nih.gov/sra/SRR10583824>. RNA-seq data of various transcriptomes have been deposited in the BioProject database at NCBI (BioProject: PRJNA588628), and are

publicly accessible at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA588628>.

CRedit author statement

Yang Zhang: Conceptualization, Methodology, Validation, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Fan Mao:** Methodology, Validation, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition. **Shu Xiao:** Methodology, Validation, Resources, Funding acquisition. **Haiyan Yu:** Methodology, Formal analysis, Investigation, Data curation. **Zhiming Xiang:** Methodology, Validation, Data curation, Funding acquisition. **Fei Xu:** Formal analysis, Validation, Data curation. **Jun Li:** Validation, Resources. **Lili Wang:** Formal analysis. **Yuanyan Xiong:** Formal analysis. **Mengqiu Chen:** Formal analysis. **Yongbo Bao:** Formal analysis. **Yuewen Deng:** Validation. **Quan Huo:** Validation. **Lyping Zhang:** Validation. **Wenguang Liu:** Validation. **Xuming Li:** Formal analysis. **Haitao Ma:** Formal analysis. **Yuehuan Zhang:** Resources. **Xiyu Mu:** Formal analysis. **Min Liu:** Formal analysis. **Hongkun Zheng:** Conceptualization, Formal analysis, Data curation, Project administration. **Nai-Kei Wong:** Writing - review & editing, Visualization. **Ziniu Yu:** Conceptualization, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. All authors have read and approved the final manuscript.

Competing interests

Haiyan Yu, Lili Wang, Xuming Li, Xiyu Mu, Min Liu, and Hongkun Zheng are current employees of Biomarker Technologies Co., Ltd. All the other authors have declared no competing interests.

Acknowledgments

We are deeply grateful to our lab members and collaborators, who have provided us with able assistance or valuable advice at all stages of this study. We acknowledge grant support from the Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), China (Grant No. GML2019ZD0407), the Key Deployment Project of Centre for Ocean Mega-Research of Science, Chinese Academy of Science (Grant No. COMS2019Q11), the National Natural Science Foundation of China (Grant Nos. 32073002 and 31902404), the China Agricultural Research System (Grant No. CARS-49), the Science and Technology Program of Guangzhou, China (Grant No. 201804020073), Natural Science Foundation of Guangdong, China (Grant No. 2020A1515011533), the Program of the Pearl River Young Talents of Science and Technology in Guangzhou of China (Grant No. 201806010003), the Institution of South China Sea Ecology and Environmental Engineering, Chinese Academy of Sciences (Grant Nos. ISEE2018PY01, ISEE2018PY03, and ISEE2018ZD01), and the Science and Technology Planning Project of Guangdong Province, China (Grant Nos. 2017B030314052 and 201707010177).

Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gpb.2021.10.005>.

ORCID

ORCID 0000-0002-0789-4938 (Yang Zhang)
 ORCID 0000-0001-6899-5591 (Fan Mao)
 ORCID 0000-0002-7276-3213 (Shu Xiao)
 ORCID 0000-0001-9709-0417 (Haiyan Yu)
 ORCID 0000-0003-1428-2910 (Zhiming Xiang)
 ORCID 0000-0002-8981-8429 (Fei Xu)
 ORCID 0000-0003-0488-0079 (Jun Li)
 ORCID 0000-0001-6357-8051 (Lili Wang)
 ORCID 0000-0002-3227-5215 (Yuanyan Xiong)
 ORCID 0000-0002-5394-8781 (Mengqiu Chen)
 ORCID 0000-0002-8268-1887 (Yongbo Bao)
 ORCID 0000-0002-4060-2220 (Yuewen Deng)
 ORCID 0000-0002-9254-4720 (Quan Huo)
 ORCID 0000-0001-9199-3623 (Lvping Zhang)
 ORCID 0000-0003-2446-0557 (Wenguang Liu)
 ORCID 0000-0002-5049-9706 (Xuming Li)
 ORCID 0000-0002-8930-614X (Haitao Ma)
 ORCID 0000-0001-9150-0163 (Yuehuan Zhang)
 ORCID 0000-0002-2631-3537 (Xiyu Mu)
 ORCID 0000-0001-9283-5942 (Min Liu)
 ORCID 0000-0002-9426-3615 (Hongkun Zheng)
 ORCID 0000-0003-1303-3170 (Nai-Kei Wong)
 ORCID 0000-0002-1049-4345 (Ziniu Yu)

References

- [1] 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.
- [2] Tibabuzo Perdomo AM, Alberts EM, Taylor SD, Sherman DM, Huang CP, Wilker JJ. Changes in cementation of reef building oysters transitioning from larvae to adults. *ACS Appl Mater Interfaces* 2018;10:14248–53.
- [3] Grabowski JH, Peterson CH. Restoring oyster reefs to recover ecosystem services. *Theor Ecol Series* 2007;4:281–98.
- [4] Kroeker KJ, Kordas RL, Crim R, Hendriks IE, Ramajo L, Singh GS, et al. Impacts of ocean acidification on marine organisms: quantifying sensitivities and interaction with warming. *Glob Chang Biol* 2013;19:1884–96.
- [5] Priemel T, Degtyar E, Dean MN, Harrington MJ. Rapid self-assembly of complex biomolecular architectures during mussel byssus biofabrication. *Nat Commun* 2017;8:14539.
- [6] Li Y, Sun X, Hu X, Xun X, Zhang J, Guo X, et al. Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins. *Nat Commun* 2017;8:1721.
- [7] Burkett JR, Hight LM, Kenny P, Wilker JJ. Oysters produce an organic-inorganic adhesive for intertidal reef construction. *J Am Chem Soc* 2010;132:12531–3.
- [8] Stanley SM. Relation of Shell Form to Life Habits of the Bivalvia (Mollusca). *Geological Society of America Bulletin* 1970;125:296.
- [9] Guo XM, Ford SE, Zhang FS. Molluscan aquaculture in China. *J Shellfish Res* 1999;18:19–31.
- [10] Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 2012;490:49–54.
- [11] Wang S, Zhang J, Jiao W, Li J, Xun X, Sun Y, et al. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat Ecol Evol* 2017;1:120.
- [12] Du XD, Fan GY, Jiao Y, Zhang H, Guo XM, Huang RL, et al. The pearl oyster *Pinctada fucata martensii* genome and multi-omic analyses provide insights into biomineralization. *Gigascience* 2017;6:1–12.
- [13] 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.
- [14] Ren J, Liu X, Jiang F, Guo X, Liu B. Unusual conservation of mitochondrial gene order in *Crassostrea* oysters: evidence for recent speciation in Asia. *BMC Evol Biol* 2010;10:394.
- [15] Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, et al. The biodiversity of species and their rates of extinction, distribution, and protection. *Science* 2014;344:1246752.
- [16] Hopkins AE. Attachment of larvae of the Olympia oyster, *Ostrea lurida*, to plane surfaces. *Ecology* 1935;16:82–7.
- [17] Frobius AC, Funch P. Rotiferan *Hox* genes give new insights into the evolution of metazoan bodyplans. *Nat Commun* 2017;8:9.
- [18] Shiga Y, Yasumoto R, Yamagata H, Hayashi S. Evolving role of Antennapedia protein in arthropod limb patterning. *Development* 2002;129:3555–61.
- [19] Khadjeh S, Turetzek N, Pechmann M, Schwager EE, Wimmer EA, Damen WGM, et al. Divergent role of the *Hox* gene *Antennapedia* in spiders is responsible for the convergent evolution of abdominal limb repression. *Proc Natl Acad Sci U S A* 2012;109:4921–6.
- [20] Kimoto M, Tsubota T, Uchino K, Sezutsu H, Takiya S. Hox transcription factor Antp regulates *sericin-1* gene expression in the terminal differentiated silk gland of *Bombyx mori*. *Dev Biol* 2014;386:64–71.
- [21] Metzler RA, Rist R, Alberts E, Kenny P, Wilker JJ. Composition and structure of oyster adhesive reveals heterogeneous materials properties in a biological composite. *Adv Funct Mater* 2016;26:6814–21.
- [22] Guerette PA, Hoon S, Seow Y, Raida M, Masic A, Wong FT, et al. Accelerating the design of biomimetic materials by integrating RNA-seq with proteomics and materials science. *Nat Biotechnol* 2013;31:908–15.
- [23] Ji H, Shepard PD. SK Ca²⁺-activated K⁺ channel ligands alter the firing pattern of dopamine-containing neurons *in vivo*. *Neuroscience* 2006;140:623–33.
- [24] Hennebert E, Maldonado B, Ladurner P, Flammang P, Santos R. Experimental strategies for the identification and characterization of adhesive proteins in animals: a review. *Interface Focus* 2015;5:20140064.
- [25] Li D, Graham LD. Epidermal secretions of terrestrial flatworms and slugs: *Lehmannia valentiana* mucus contains matrilin-like proteins. *Comp Biochem Physiol B Biochem Mol Biol* 2007;148:231–44.
- [26] Hennebert E, Wattiez R, Demeuldre M, Ladurner P, Hwang DS, Waite JH, et al. Sea star tenacity mediated by a protein that fragments, then aggregates. *Proc Natl Acad Sci U S A* 2014;111:6317–22.
- [27] Lee BP, Messersmith PB, Israelachvili JN, Waite JH. Mussel-inspired adhesives and coatings. *Annu Rev Mater Res* 2011;41:99–132.
- [28] Rucker RB, Kosonen T, Clegg MS, Mitchell AE, Rucker BR, Uriu-Hare JY, et al. Copper, lysyl oxidase, and extracellular matrix protein cross-linking. *Am J Clin Nutr* 1998;67:996S–1002S.
- [29] Walker G. A study of the cement apparatus of the cypris larva of the barnacle *Balanus balanoides*. *Mar Biol* 1971;9:205–12.
- [30] Smith AM. Multiple Metal-based Cross-Links: Protein Oxidation And Metal Coordination In A Biological Glue. In: Santos R, Aldred N, Gorb S, Flammang P, editors. *Biological And Biomimetic Adhesives: Challenges And Opportunities*. Cambridge: Royal Society of Chemistry; 2013, p.3–15.
- [31] Coon SL, Fitt WK, Bonar DB. Competence and delay of metamorphosis in the Pacific oyster *Crassostrea gigas*. *Mar Biol* 1990;106:379–87.
- [32] Bonar DB, Coon SL, Walch M, Weiner RM, Fitt W. Control of oyster settlement and metamorphosis by endogenous and exogenous chemical cues. *B Mar Sci* 1990;46:484–98.
- [33] Wang J, Wu CL, Xu CL, Yu WC, Li Z, Li YC, et al. Voltage-gated potassium ion channel may play a major role in the settlement of Pacific oyster (*Crassostrea gigas*) larvae. *Aquaculture* 2015;442:48–50.

- [34] Foulon V, Boudry P, Artigaud S, Guerard F, Hellio C. *In silico* analysis of Pacific oyster (*Crassostrea gigas*) transcriptome over developmental stages reveals candidate genes for larval settlement. *Int J Mol Sci* 2019;20:197.
- [35] Zhang T, Liu J, Fellner M, Zhang C, Sui D, Hu J. Crystal structures of a ZIP zinc transporter reveal a binuclear metal center in the transport pathway. *Sci Adv* 2017;3:e1700344.
- [36] Du Y, Lian F, Zhu L. Biosorption of divalent Pb, Cd and Zn on aragonite and calcite mollusk shells. *Environ Pollut* 2011;159:1763–8.
- [37] Bond JS, Beynon RJ. The astacin family of metalloendopeptidases. *Protein Sci* 1995;4:1247–61.
- [38] Sadeghi H, Allard P, Prince F, Labelle H. Symmetry and limb dominance in able-bodied gait: a review. *Gait Posture* 2000;12:34–45.
- [39] Weiss IM, Schonitzer V. The distribution of chitin in larval shells of the bivalve mollusk *Mytilus galloprovincialis*. *J Struct Biol* 2006;153:264–77.
- [40] Marin F, Luquet G, Marie B, Medakovic D. Molluscan shell proteins: primary structure, origin, and evolution. *Curr Top Dev Biol* 2008;80:209–76.
- [41] Splitt MP, Burn J, Goodship J. Defects in the determination of left-right asymmetry. *J Med Genet* 1996;33:498–503.
- [42] Savazzi E. Adaptational strategies of bivalves living as infaunal secondary soft bottom dwellers. *Neues Jahrb Geol P-A* 1982;164:229–44.
- [43] Wilbur KM, Saleuddin ASM. Shell Formation. In: Saleuddin ASM, Wilbur KM, editors. *The mollusca*. New York: Academic Press; 1983, p.235–87.
- [44] Yoshioka H, Meno C, Koshihara K, Sugihara M, Itoh H, Ishimaru Y, et al. *Pitx2*, a bicoid-type homeobox gene, is involved in a lefty-signaling pathway in determination of left-right asymmetry. *Cell* 1998;94:299–305.
- [45] Grande C, Patel NH. Nodal signalling is involved in left-right asymmetry in snails. *Nature* 2009;457:1007–11.
- [46] Smith SB, Qu HQ, Taleb N, Kishimoto NY, Scheel DW, Lu Y, et al. *Rfx6* directs islet formation and insulin production in mice and humans. *Nature* 2010;463:775–80.
- [47] Hamano K, Awaji M, Usuki H. cDNA structure of an insulin-related peptide in the Pacific oyster and seasonal changes in the gene expression. *J Endocrinol* 2005;187:55–67.
- [48] Nagai K, Yano M, Morimoto K, Miyamoto H. Tyrosinase localization in mollusc shells. *Comp Biochem Physiol B Biochem Mol Biol* 2007;146:207–14.
- [49] Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A* 2011;108:1513–8.
- [50] Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 2011;27:578–9.
- [51] Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. Erratum: *SOAPdenovo2*: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* 2015;4:30.
- [52] Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 2017;27:722–36.
- [53] Salmela L, Rivals E. LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 2014;30:3506–14.
- [54] Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 2007;23:1061–7.
- [55] Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;31:3210–2.
- [56] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009;326:289–93.
- [57] Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nat Biotechnol* 2013;31:1119–25.
- [58] Li C, Wang J, Song K, Meng J, Xu F, Li L, et al. Construction of a high-density genetic map and fine QTL mapping for growth and nutritional traits of *Crassostrea gigas*. *BMC Genomics* 2018;19:626.
- [59] Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 2007;8:973–82.
- [60] Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 2015;6:11.
- [61] Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* 2009;Chapter 4: Unit 4.10.
- [62] Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 1997;268:78–94.
- [63] Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 2003;19:ii215–25.
- [64] Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. *Bioinformatics* 2004;20:2878–9.
- [65] Blanco E, Parra G, Guigo R. Using geneid to identify genes. *Curr Protoc Bioinformatics* 2007;Chapter 4:Unit 4.3.
- [66] Korf I. Gene finding in novel genomes. *BMC Bioinformatics* 2004;5:59.
- [67] Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res* 2016;44:e89.
- [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] Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 2005;21:3674–6.
- [70] Nam BH, Kwak W, Kim YO, Kim DG, Kong HJ, Kim WJ, et al. Genome sequence of pacific abalone (*Haliotis discus hannai*): the first draft genome in family Haliotidae. *Gigascience* 2017;6:1–8.
- [71] Li L, Stoeckert Jr CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 2003;13:2178–89.
- [72] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–7.
- [73] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–3.
- [74] Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007;24:1586–91.
- [75] De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 2006;22:1269–71.
- [76] Albertin CB, Simakov O, Mitros T, Wang ZY, Pungor JR, Edsinger-Gonzales E, et al. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* 2015;524:220–4.
- [77] Keilwagen J, Hartung F, Grau J. GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data. *Methods Mol Biol* 2019;1962:161–77.
- [78] Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009;25:1105–11.
- [79] Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 2012;7:562–78.
- [80] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
- [81] Coon SL, Bonar DB, Weiner RM. Chemical production of cultchless oyster spat using epinephrine and norepinephrine. *Aquaculture* 1986;58:255–62.
- [82] 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.
- [83] Chen M, Ma Y, Wu S, Zheng X, Kang H, Sang J, et al. Genome Warehouse: a public repository housing genome-scale data. *Genomics Proteomics Bioinformatics* 2021;19:584–9.