Hologenome analysis reveals independent evolution to chemosymbiosis by deep-sea bivalves

Yang Guo[1†], Lingfeng Meng[2,3†], Minxiao Wang[1,4†], Zhaoshan Zhong[1], Denghui Li[2], Yaolei Zhang[2,5], Hanbo Li[2,5], Huan Zhang[1], Inge Seim[6,7], Yuli Li[8], Aijun Jiang[2], Qianyue Ji[2], Xiaoshan Su[2], Jianwei Chen[2], Guangyi Fan[2,5*], Chaolun Li[1,9,10*] and Shanshan Liu[2,11*]

Abstract

Background Bivalves have independently evolved a variety of symbiotic relationships with chemosynthetic bacteria. These relationships range from endo- to extracellular interactions, making them ideal for studies on symbiosis-related evolution. It is still unclear whether there are universal patterns to symbiosis across bivalves. Here, we investigate the hologenome of an extracellular symbiotic thyasirid clam that represents the early stages of symbiosis evolution.

Results We present a hologenome of *Conchocele bisecta* (Bivalvia: Thyasiridae) collected from deep-sea hydrothermal vents with extracellular symbionts, along with related ultrastructural evidence and expression data. Based on ultrastructural and sequencing evidence, only one dominant Thioglobaceae bacteria was densely aggregated in the large bacterial chambers of *C. bisecta*, and the bacterial genome shows nutritional complementarity and immune interactions with the host. Overall, gene family expansions may contribute to the symbiosis-related phenotypic variations in different bivalves. For instance, convergent expansions of gaseous substrate transport families in the endosymbiotic bivalves are absent in *C. bisecta*. Compared to endosymbiotic relatives, the thyasirid genome exhibits large-scale expansion in phagocytosis, which may facilitate symbiont digestion and account for extracellular symbiotic phenotypes. We also reveal that distinct immune system evolution, including expansion in lipopolysaccharide scavenging and contraction of IAP (inhibitor of apoptosis protein), may contribute to the different manners of bacterial virulence resistance in *C. bisecta*.

Conclusions Thus, bivalves employ different pathways to adapt to the long-term co-existence with their bacterial symbionts, further highlighting the contribution of stochastic evolution to the independent gain of a symbiotic lifestyle in the lineage.

Keywords Symbiosis, Hydrothermal vent, Bivalvia, *Conchocele bisecta*, Genome, Stochastic evolution, Comparative genomics, Phagocytosis, Lipopolysaccharide scavenging

---

[†]Yang Guo, Lingfeng Meng and Minxiao Wang contributed equally to this work.

*Correspondence:
Guangyi Fan
fanguangyi@genomics.cn
Chaolun Li
lcl@qdio.ac.cn
Shanshan Liu
liushanshan@mgi-tech.com
Full list of author information is available at the end of the article

Guo *et al. BMC Biology*   (2023) 21:51

Page 2 of 20

Background

Since the beginning of metazoan evolution, the development of animals has depended on the interactions with bacterial communities [1–3], and symbiotic relationships between microbial partners and animal hosts are hallmarks of these associations [4, 5]. Symbiosis has evolved independently in each lineage and is frequently associated with evolutionary adaptations [6, 7]. Thus, symbiosis-related evolution is a substantial contributor to phenotypic complexity and has long been a hot topic in biology [8–11].

Deep-sea chemosynthetic ecosystems are patchily distributed and turbulent [12]. Chemosynthetic bacteria are the primary producers of deep-sea cold seeps and vents [7, 13]. Thus, symbiosis is a common phenomenon in this ecosystem to facilitate energy utilization. As one representative of deep-sea symbiotic organisms, the widely distributed bivalves have shown variations in symbiosis-related traits, such as the spatial structure of the holobionts and intimacy with the symbionts, which make them ideal materials for studies on the correlations between symbiosis and evolution [14, 15].

According to Wein et al., the divergences of symbioses are determined by three elements: symbiotic currency, mechanism of currency exchange, and inheritance regimes [16]. Among them, symbiotic currency is constant among chemosymbiotic bivalves, where the host provides steady inorganic substrates to symbionts for carbon fixation, and in return, the bivalves obtain fixed carbons and rare metabolites, for example, the essential amino acids for bivalves [14, 17, 18]. However, the other two elements in deep-sea Bivalvia symbiotic systems have been shown to be highly varied.

The currency exchange mechanism varies and depends on the spatial structure of the holobionts. Most symbiotic bivalves belonging to Vesicomyinae and Bathymodiolinae host their symbionts intracellularly while most symbiotic thyasirids host their bacterial symbionts extracellularly [19, 20]. Correspondingly, the nutrient transfer mode may differ among taxa. Endosymbiotic bivalves, such as *Archivesica marissinica*, import gaseous substrates using carbonic anhydrase and globin so that intracellularly located symbionts can drive carbon fixation [21, 22]. In extracellular symbiotic thyasirids, the bacterial symbionts are located outside the cell membrane of bacteriocytes, which would allow the symbionts to obtain substances directly from the environment [20]. Moreover, many symbionts are alive in the bacteriocytes of endosymbiotic mussels [23]. In contrast, the extracellular symbionts of thyasirids are endocytosed in phagosomes, and the digestion rate of bacteria in phagosomes is much higher in the bacteriocytes [20, 24]. Thus, with the phenotypic divergence in spatial structure and currency exchange, how thyasirids differ from the endosymbiotic bivalves in the genome remains debated.

The inheritance regimes of bivalves can be divided into vertical inheritance from parents to offspring and horizontal transmission from the environment [25]. How the hosts recognize and police the bacterial symbionts are crucial in shaping these inheritance regimes. Different patterns in bacterial recognition and immune response have been demonstrated by comparative genomic investigations of endosymbiotic bivalves. For example, the mussel *Gigantidas platifrons* manifests horizontal transmission and expansion of gene families involved in immunological recognition [26]. The same families are contracted in the clam *A. marissinica*, which manifests vertically transmitted symbiosis [22]. The extracellularly symbiotic thyasirids might have evolved a distinct mechanism in bacterial recognition and homeostasis since they are considered less integrated with their symbionts [27]. First, they harbor bacteria either among the microvilli outside the epithelial cells of gills or, more integrally, in the apical vesicles delimited by the cell membrane and microvilli (hereafter for extracellular symbiosis to distinguish from the previous type of epi-symbiosis) [20]. In addition, thyasirids show a fluctuating dependence upon symbiont-derived nutrients [28, 29]. Thus, extracellular symbiotic thyasirids might be regarded as intermediate between endosymbiotic and asymbiotic bivalves and may exhibit distinct evolutionary adaptations. In addition, how the thyasirid hosts select and recruit their symbionts with high specificity is still unknown.

Here, we present a genome of *Conchocele bisecta* (Bivalvia: Thyasiridae) and its extracellular symbiont. By comparative genomic analysis of *C. bisecta* and endosymbiotic bivalves, genetic divergence related to the variations in currency exchange and bacterial recognition and policing are revealed. Ultrastructural observations and gene expression further corroborated our findings. We tested the hypothesis that the deep-sea bivalves evolve independently to give rise to the convergent symbiosis characteristics. Our results highlight that stochastic evolution may shape divergent symbiosis interactions by deep-sea bivalves.

Results and discussion

Distribution and constitution of symbionts in *C. bisecta*

To determine how bacterial symbionts are distributed in *C. bisecta* sampled from hydrothermal vents (Additional file 1: Supplementary note 1, Fig. S1) [30, 31], transmission electron micrography (TEM)

was performed. Densely aggregated bacterial symbionts were in apical vesicles of gill filaments bacteriocytes (Fig. 1A). In most cases, the apical vesicles were opened at the tip but covered by microvilli, while lysosomes and other organelles located near the basal membrane side (Fig. 1A, B). In addition to long, large

columnar vesicles, we noticed that there were also small ones, which were likely to be at an initial stage (Fig. 1A). In contrast to the bacterial symbionts in the apical vesicles (Fig. 1C), symbionts in intracellular vesicles had blurred cell walls and irregular shapes (Fig. 1D). It has been reported that thyasirids obtain
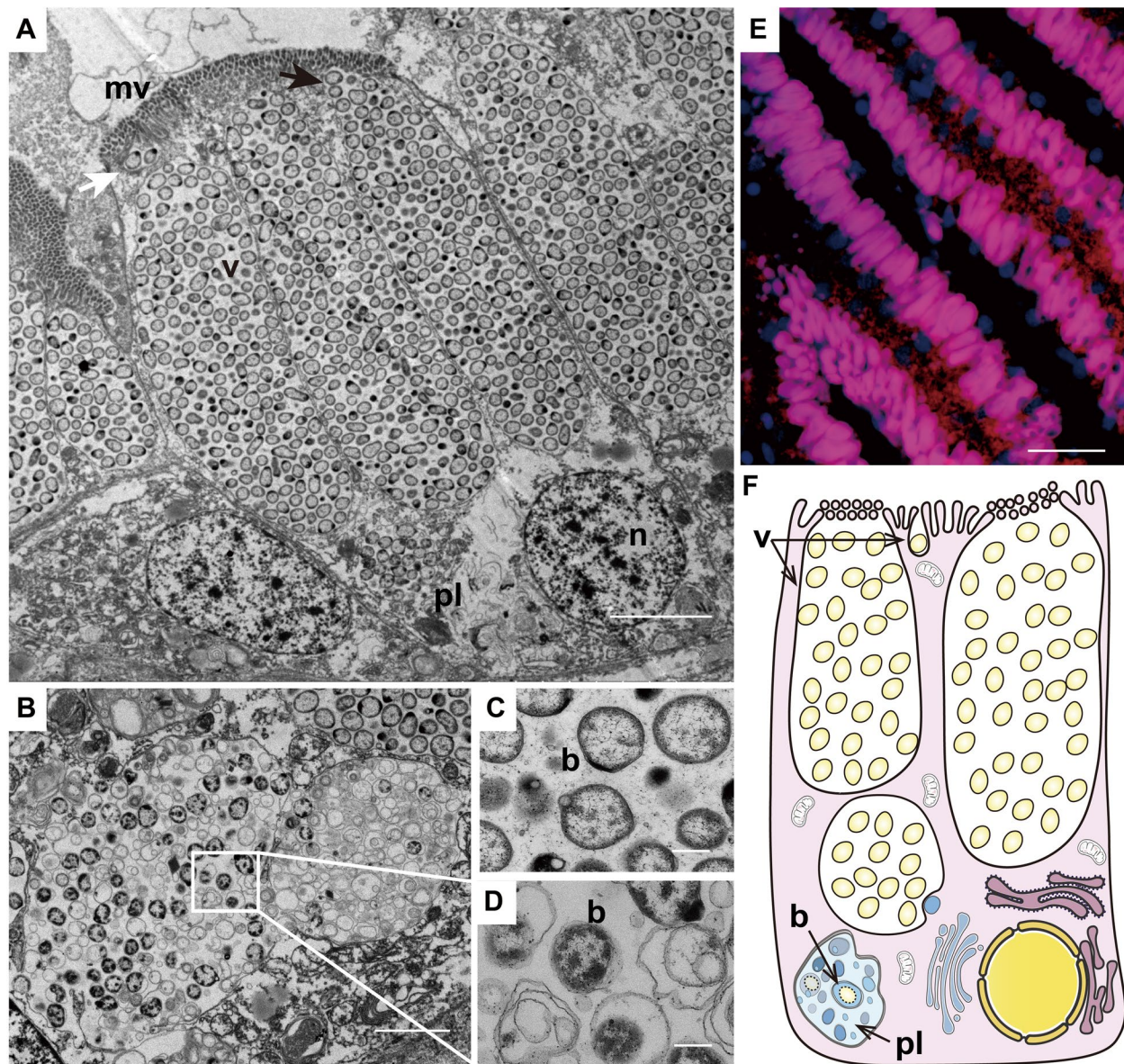
Fig. 1 Distribution of bacterial symbionts in the gill filaments of *Conchocele bisecta* collected in a deep-sea hydrothermal vent field. **A** Transmission electron micrography (TEM) of bacteriocytes in gill filaments. Extracellularly symbionts were densely aggregated in the large apical vesicles (black arrow; v: vesicle) covered by microvilli (mv) and smaller vesicles (white arrow). Phagolysosome-like organelles (pl) and nucleus (n) located in the basolateral membranes of the bacteriocytes (scale bar: 5 μm). **B** Details of phagolysosome-like organelles with whorls of lysed bacterial products (scale bar 2 μm). **C** Living bacterial symbiont (b) with a clear border located in the apical vesicles, and the symbionts were small cocci with a diameter around 500 nm (scale bar 250 nm). **D** Lysed bacterial symbionts with blurred cell walls in the phagolysosome-like organelles (scale bar: 250 nm). **E** Fluorescence in situ hybridization (FISH) of 16S rRNA of the dominate symbionts in the gill filaments. The image shows the overlay of DAPI-stained DNAs and 16S rRNA tagged with Cy3 dye (scale bar: 50 μm). **F** Schematic representation of typical bacteriocytes, including small, and large apical vesicles (v), phagolysosome-like organelles (pl), and lysed symbionts (b)

Guo *et al. BMC Biology*   (2023) 21:51

Page 4 of 20

nutrients by periodically engulfing and digesting their symbionts [24, 32]. These results suggest that most symbionts are quickly digested after being endocytosed by bacteriocytes of *C. bisecta*, and most of these intracellular vesicles should be organelles such as late phagosomes and phagolysosomes. In the endosymbiotic deep-sea mussels and vesicomyid clams, the phenotype is quite different: despite the presence of vesicles that appeared to be digesting the symbionts, there were many vesicles holding live endosymbionts, indicating that many of these vesicles might be early phagosomes or at more primitive stages [33, 34].

Based on the 16S rDNA V4 region amplicon analysis, bacteria close to the clade SUP05 dominated in the gill tissue of *C. bisecta* with a relative abundance of 94.73±3.20% (Additional file 2: Table. S1). Fluorescence in situ hybridization (FISH) was conducted to verify this observation. Probes (16S-Probe) were designed based on 16S rDNA sequences of the dominated symbionts, and a strong signal was detected in columnar vesicles arranged in the gill filaments, in accordance with the TEM examination. The negative control probe did not hybridize with both tissues, and no positive signal was observed in gonad tissue (Fig. 1E, Additional file 1: Fig. S2). The bacteria-host specificity has been widely reported in multiple lineages, including corals, nematodes, and mollusks [35–37]. Although the mechanism underlying this phenomenon is still being debated, the vertical transmission of symbionts and immune interactions were thought to be significant contributors [36–38]. However, according to our data, the dominant symbionts were absent in the gonad tissue of the *C. bisecta*, supporting the theory that thyasirids acquire symbionts horizontally [39, 40]. Thus, the genomic characterization of the holobionts would be necessary to provide insights into the interactions and specificities between host and its symbionts.

A high-quality genome of the extracellular bacterial symbiont of *C. bisecta* (hereafter SCbi) with a 92.85% completeness was recovered through a binning pipeline. The assembly consists of three single-copy rRNAs and 31 tRNAs (Additional file 1: Table S2). The 16S rDNA sequences and the GTDB (Genome Taxonomy Database) taxonomy annotation agreed to assign SCbi as *Thiodubiliella* sp. Interestingly, we found some genomic features that may contribute to symbiotic life by providing fixed carbons and rare metabolites to the thyasirid host. In brief, SCbi uses a modified Calvin cycle for carbon fixation driven by sulfur oxidation. We identified many copies of SulP-type transporters in the symbiont genome (Additional file 2: Table S3) [41, 42]. Members of the SulP family are associated with the intake of inorganic anions, such as sulfates and bicarbonates, in prokaryotes [43, 44], suggesting that these genes facilitate chemoautotrophy.

The SCbi genome exhibited complete routes for biosynthesis of 18 of 20 amino acids and five of ten cofactors (Additional file 1: Fig. S3, S4, Additional file 2: Table S4). As supported by proteomic data, missing enzymes in the biosynthesis of the amino acid threonine and folate may be compensated by a different enzyme or reaction products provided by the host (Additional file 1: Supplementary note 2, Fig. S3–7 and Table S5, Additional file 2: Table S4) [22, 45, 46]. Additionally, genes related to lipopolysaccharide (LPS) modifications and pathogenic invasion, such as homologs of *slyB* and *yeeJ*, were found in the SCbi genome and may be implicated in symbiotic interactions. For instance, according to our proteomic data, the invasin, encoded by *yeeJ* (intensity-based absolute quantification (iBAQ) per million: 512.64), and the host receptor of integrin-beta1 (iBAQ per million: 49.18) were both actively translated in the gill tissue, pointing to a potential pathway of bacteria triggered phagocytosis or invasion [47, 48].

Taken together, *C. bisecta* hosts bacterial symbionts in large apical vesicles of bacteriocytes in gill tissue, similar to the previously reported cold seep *C. bisecta*, and other thyasirids such as *Thyasirid flexuosa* [20, 49]. Moreover, *C. bisecta* harbors a dominant *Thiodubiliella* symbiont that appears adequate for providing most necessary rare metabolites to the host and interacting immunologically with the host. Moreover, our TEM results emphasize differences between the intracellularly symbiont-holding vesicles of the extracellular symbiotic *C. bisecta* and endosymbiotic mussels and clams. The absence of SCbi in gonad tissue further supports that *C. bisecta* acquired its bacterial symbiont from the surrounding environments horizontally, highlighting the importance of bacterial recognition in this symbiotic relationship.

Host genome assembly and phylogenetic analysis

To decipher the genetic machinery under these unique symbiotic features, we assembled a high-quality genome of *C. bisecta* by combining long-read PacBio sequencing reads with highly accurate short reads generated on the BGISEQ 500 platform and used Hi-C data for super-scaffolding (Additional file 1: Supplementary Note 3, Fig. S6 and Table S6–9). In detail, the 1.9 Gb assembly contained 17,791 contigs with a contig N50 length of 488 kb (Fig. 2A, Additional file 1: Table S7). We annotated 25,483 protein-coding genes and recovered 95.4% of metazoan BUSCOs (Benchmarking Universal Single Copy Orthologs) (Additional file 1: Table S7).

Our phylogenetic results showed that *C. bisecta* from Lucinida lies in a basal position of the heterodont clade, and the divergence time between *C. bisecta* and other heterodont clams is estimated as 464.9 million years ago (mya) in the mid-Ordovician (Fig. 2B). As recently
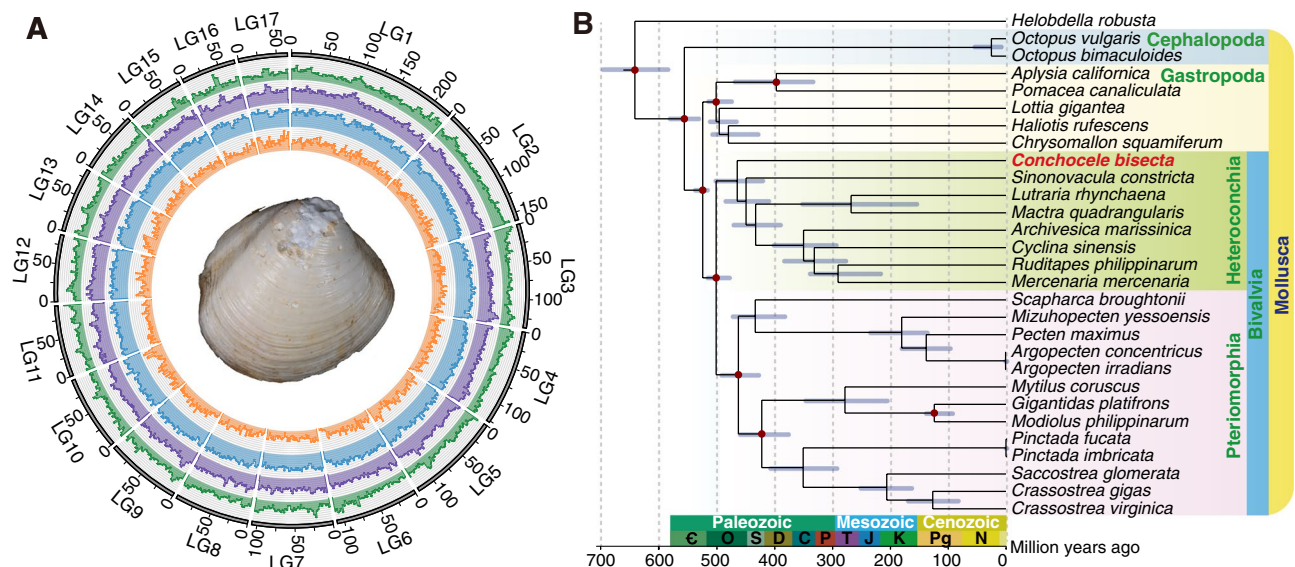
Fig. 2 Genomic characteristics of *Conchocele bisecta*. **A** Circos view of 17 linkage groups (corresponding to "pseudo-chromosomes") showing marker distributions at 2-Mb sliding windows from outer to inner circle: GC content, TE density, tandem repeat density, gene density. **B** Maximum-likelihood phylogenetic relationships were constructed using the LG+F+R6 substitution model among 28 mollusks with an annelid (*Helobdella robusta*) as an outgroup. The tree was calibrated at nine nodes (indicated by red dots) using divergence times from TimeTree (http://www.timetree.org). Blue lines indicate 95% confidence intervals for divergence times

reported, the oldest fossil record of thyasirid bivalve is *Eothyasira antiqua* from the early Jurassic [50]. However, it was previously suspected, employing 18S rRNA gene data, that Thyasiridae have a longer fossil history than its published records [51], and our phylogenetic results support this long divergence time. The conservation index (CI) of six bivalves with ancestral linkage groups (ALG) that are represented by genes of *Nematostella vectensis* were calculated by a chromosome-based macrosynteny analysis [52]. Although the CI of *C. bisecta* (0.73) is not the highest among heterodont bivalves, the CIs of heterodont bivalves fell in a narrow window (0.7–0.8), which was much higher than that of the pearl oyster (0.45) (Additional file 1: Fig. S9). The accordant high CIs of heterodont bivalves indicated the genome organization of this clade is highly conserved to bilaterian ancestral genomes, although with a long divergence time among these clams.

Integrating repeat content results from hybrid software revealed that the transposable element (TE) content of *C. bisecta* is strikingly high. About 6.5 million repetitive elements were identified, accounting for the relatively large genome of the species (Additional file 1: Fig. S10a). Comparative analysis further revealed that the genome of *C. bisecta* contains both the largest amount (1.27 Gb) and the highest percentage (66.96%) of TEs than genomes of other lineages (Additional file 1: Fig. S10a and Table S10). It has been argued that TE content contributes to larger

genome sizes [53–55]. We found that the genome size and TE content correlated in 27 lophotrochozoan genomes, except for the symbiotic clams *C. bisecta* and *A. marissinica*, which showed an elevated proportion of TEs (Additional file 1: Fig. S10b). Moreover, the two symbiotic clams had longer introns than other heterodont clams, which could be the results of high TE contents in the genomes, as seen in the African lungfish (Additional file 1: Fig. S10c) [54]. Expansions in TE-related protein domains have been shown to contribute to the high TE proportions in *A. marissinica* [22]. These protein domains, including the reverse transcriptase domain, DDE superfamily endonuclease, and the MULE transposase domain, were also significantly expanded in the *C. bisecta* genome (Additional file 1: Fig. S11). TEs can replicate independently within their host genomes and are a major source of genetic variation and novelty [56]. Thus, based on our data, the expansion in TE-related protein domains and the large amounts of TEs in these symbiotic clams may provide sufficient evolutionary material for the hosts to adapt to the symbiotic lifestyle, or at the very least, to diverse environments.

Evolution of gene families and their roles in symbiosis
Bivalves are ideal materials for studies on symbiosis and evolution due to the diverse symbiotic forms among different lineages [14, 15]. To the best of our knowledge, only two genomes of symbiotic bivalves,

A. marissinica and *G. platifrons*, have been published [22, 26]. As described, both species were endosymbiotic, with *A. marissinica* appearing to be more integrated with its symbionts than *G. platifrons* did. To investigate the evolution of symbiosis by bivalves and compare the genetic machinery under different symbiosis-related traits, we compared extracellular symbiotic *C. bisecta* and endosymbiotic *A. marissinica* and *G. platifrons* with five asymbiotic bivalves (*Crassostrea gigas*, *Lutraria rhynchaena*, *Mizuhopecten yessoensis*, *Modiolus philippinarum*, *Pinctada fucata*). The expansion and contraction events were identified both at gene and domain levels (Additional file 1: Fig. S11, Additional file 2: Table S11).

For each of the three symbiotic bivalves, there appeared to be fewer contracted KEGG orthologs (KOs) than expanded KOs, but nearly half of these contracted KOs were shared by the bivalves (Fig. 3A, B). Considering the independent origin of the taxa compared, the shared contracted KOs might be related to the convergent adaptations, such as adaptions to nutritional symbiosis in the deep sea. For instance, the cholecystokinin receptor (CCKAR) with roles in bivalve digestion was contracted in *C. bisecta* and *A. marissinica*, in agreement with their reduced digestive systems (Fig. 3C, Additional file 2: Table S11) [57, 58]. In addition, among the contracted KOs, several reduced acetylcholine receptors (CHRNAs) gene families were shared by all three symbiotic species (Fig. 3C, Additional file 2: Table S11). The adaptation to a lifestyle of coexisting with bacterial symbionts might be related to these contraction events since CHRNAs are implicated in regulating immune response and pathogen removal in both vertebrates and invertebrates (including bivalves) [59, 60].

We found many gene families expanded specifically in different lineages. Hence, we propose that the uniquely evolved KOs, especially the large amounts of expanded KOs, might be related to the variations of symbiosis-related phenotypes among the three bivalves, such as the spatial structures and the ways of symbiont inheritance. In agreement with this hypothesis, KOs related to the divergent symbiotic traits in bivalves (e.g., nutrient transport, phagocytosis, bacterial recognition, and immune response) were included in each list of the expanded KOs (Fig. 3C, Additional file 2: Table S11). This indicates that

gene family expansions may contribute to the adaptive evolution of symbiosis and in shaping the different symbiotic phenotypes in *C. bisecta* and other bivalves. Thus, we further investigated the expanded KOs.

Symbionts as the primary nutrient supply for the host

With the absence of biosynthesis capabilities of some amino acids and cofactors, it is important for mollusks to derive nutrients from external sources or their bacterial symbionts [61]. *C. bisecta* was unable to synthesize eight amino acids, including histidine, phenylalanine, isoleucine, lysine, leucine, valine, tryptophan and methionine, and genes involved in the uptake these amino acids, such as *Slc6a19*, *Slc7a9*, and *Slc16a10*, were found in the genome (Additional file 1: Fig S3, S12, Additional file 2: Table S4). SLC16A10, a transporter mediating the diffusion of amino acids across basolateral membranes of the epithelial cells [62], is both expanded and highly expressed in the gills of *C. bisecta* (Additional file 1: Fig. S12). As described, *C. bisecta* digests its symbiont intracellularly, and amino acids released by the symbionts should be accumulated in bacteriocytes, which is in correspondence to the actively transcribed *Slc16a10* (Fig. 4A). Similarly, transporters for vitamin B5 (VB5), VB7, and VB12 are highly expressed in the symbiotic tissue of *C. bisecta* (Additional file 1: Supplementary note 4, Fig. S4, S12, Additional file 2: Table S4). Therefore, the highly expressed gene family of these transporters in *C. bisecta* could emphasize the nutritional function of its gill tissues. Moreover, gene families coding for enzymes related to the filter-feeding lifestyle in mollusks, such as the glycosyl hydrolase and chymotrypsins, are found either contracted or missing in *C. bisecta*, suggesting that the clam had relied on symbiosis for nutrition for a long time (Additional file 1: Supplementary note 5, Additional file 2: Table S12) [22, 28, 63–65].

Unlike bivalves with endosymbionts, we did not find an expansion of gene families required for gaseous substance transport in *C. bisecta* (Additional file 1: Fig. S13, Fig. S14). Bacterial symbionts are located outside the cell of *C. bisecta*, and the symbionts could acquire gaseous substrates from the seawater directly (Fig. 4A). Thus, it is reasonable that there is no expansion of these transporters. As the endosymbionts obtain the gaseous substances from the cytoplasm, cytoplasmic carbonic anhydrase

(See figure on next page.)

Fig. 3 Gene family dynamics in the genomes of *Conchocele bisecta*, *Archivesica marissinica*, and *Gigantidas platifrons*. **A** Venn diagram showing the number of shared and unique expanded and contracted gene families in the three symbiotic species compared to five asymbiotic bivalves. Annotated to KEGG orthologs (KOs). **B** Total counts of KOs that expanded or contracted in each of the three bivalves, and the rates (%) of shared KOs. There are more exclusively expanded KOs than contracted ones in each bivalve. **C** Bubble plot of key expanded or contracted KOs in the three bivalves. Bubble size indicates the gene number of each KO in the eight species: *C. bisecta* (Conbis), *G. platifrons* (Gigpla), *A. marissinica* (Arcmar), *Lutraria rhynchaena* (Lutrhy), *Mizuhopecten yessoensis* (Mizyes), *Modiolus philippinarum* (Modphi), *Pinctada fucata* (Pinfuc), *Crassostrea gigas* (Cragig). Significant expansion (red), contraction (blue), or no significance (dark gray) were determined by Fisher's exact test (*P* < 0.05)

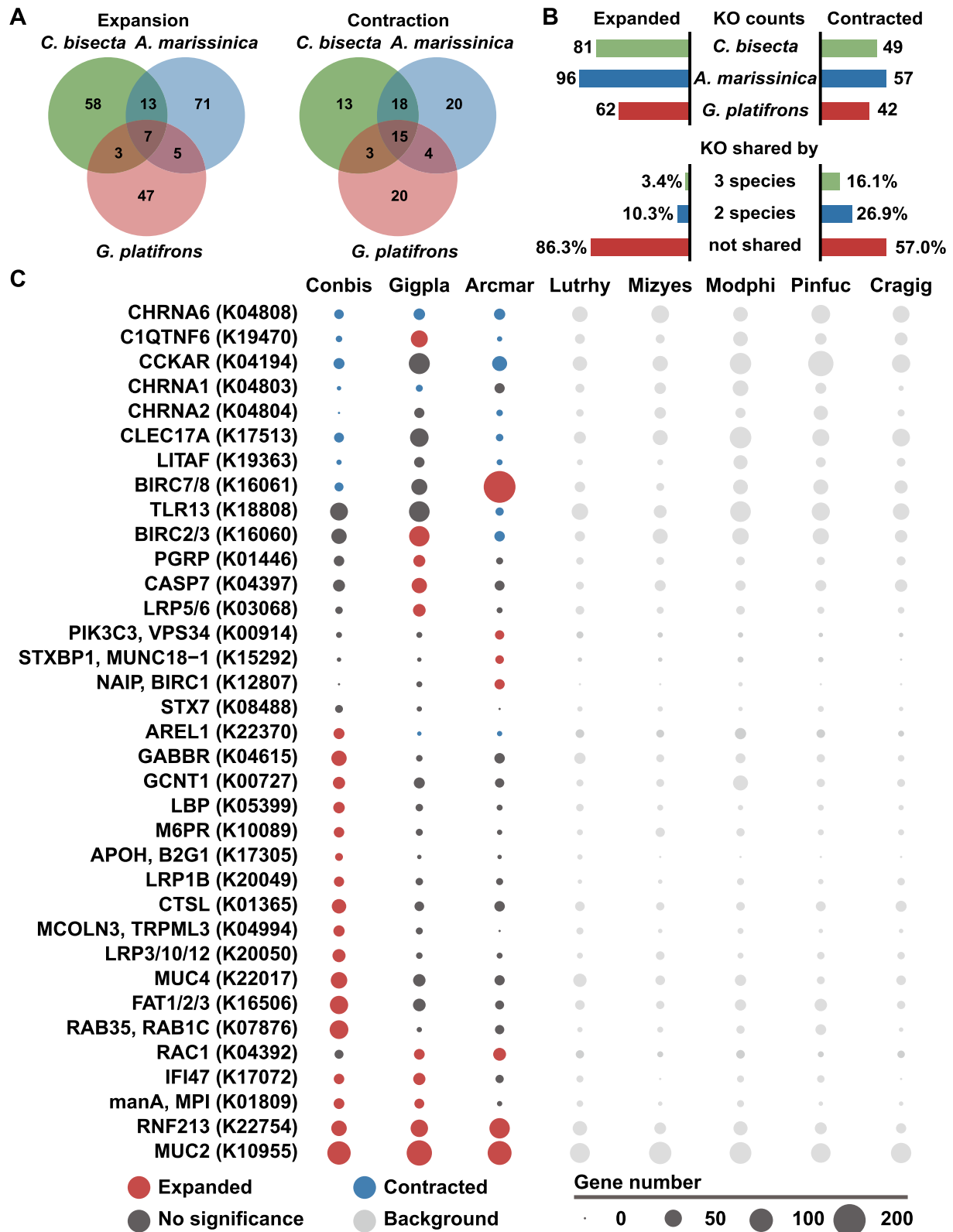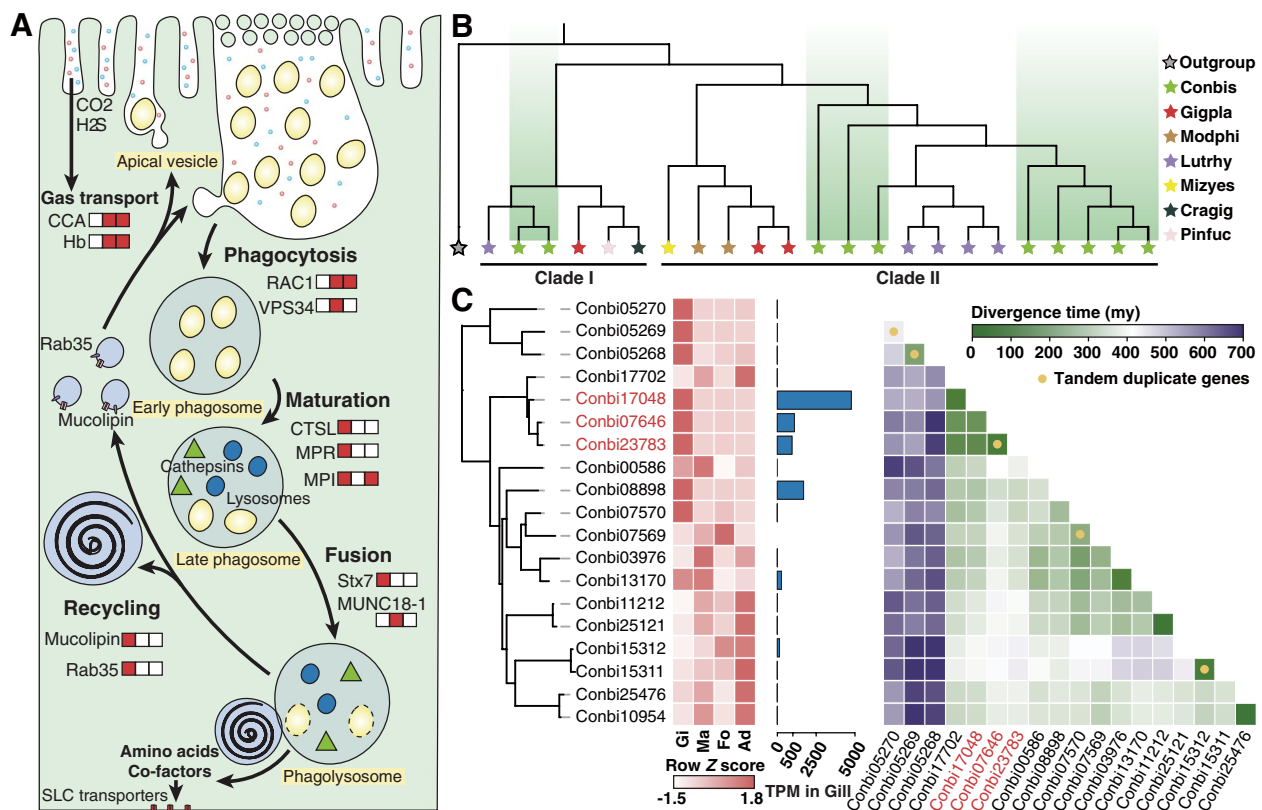Guo *et al. BMC Biology* (2023) 21:51

Page 7 of 20

Fig. 3 (See legend on previous page.)

Fig. 4 Gene families related to symbiotic currency exchange in *Conchocele bisecta*. **A** Schematic representation of currency exchange in *C. bisecta*. Expansion events of gene families responsible for substrate transports and phagocytosis are illustrated, and the boxes that followed each gene indicate if the gene family is expanded (red) or not (white) in *C. bisecta* (right), *Archivesica marissinica* (middle), and *Gigantidas platifrons* (left). **B** Phylogenetic tree based on amino acids of mucolipins (Conbis: *C. bisecta*; Gigpla: *G. platifrons*; Lutrhy: *Lutraria rhynchaena*; Mizyes: *Mizuhopecten yessoensis*; Modphi: *Modiolus philippinarum*; Pinfuc: *Pinctada fucata*; Cragig: *Crassostrea gigas*; Outgroup: *Danio rerio*). **C** Phylogenetic relationships of cathepsin L genes (*ctsL*) in *C. bisecta*. Expression levels of different tissues (Gi, gill; Ma, mantle; Fo, foot; Ad, adductor) and Row-wise *Z*-score of TPM (transcript per million) counts in gill tissue and gene divergence times (My, million years) are shown if *ctsL* copy lengths range from 600 to 1200 bp, and yellow dots indicate the tandem duplicate gene pairs. Three duplicated genes (in red) with short divergent time constitute most of the *ctsL* transcripts in gill tissue

(CCA) and hemoglobin (Hb) that may be involved in the transportation of CO_2 and sulfide and oxygen for carbon fixation are expanded in their hosts [22, 65, 66]. These findings support the hypothesis that an expansion of CCAs and Hbs in endosymbiotic bivalves is adaptive [66].

To summarize, evolutionary divergences in gene families associated with gaseous substrate transport might depend on whether symbionts are located inside bacteriocytes. In addition, the high expression of transporters of rare metabolites in symbiotic tissues and the contraction in filter-feeding-related genes revealed a high reliance on nutritional symbiosis by *C. bisecta*, and the results underlined the convergences in symbiotic currencies among the chemosymbiotic bivalves.

Enhanced phagocytic capabilities provide insights into extracellular symbiosis

As revealed by previous studies and our TEM examination (Fig. 1), phagocytosis was considered to contribute

significantly to nutritional currency exchange in extracellular symbiotic thyasirids and two endosymbiotic bivalves [20, 24, 33, 34]. Accordingly, phagocytosis-related gene families are expanded in all three symbiotic bivalves. However, the evolutionary pattern of these genes in *C. bisecta* is distinct from the endosymbiotic ones (Fig. 4A).

Phagocytosis in eukaryote cells includes bacterial internalization, digestion, and the recycling of lysosomes. Remarkably, the *C. bisecta* genome exhibit a series of expansion events that related to the processes of lysosome-mediated intracellular digestion and lysosome recycling, including the mannose-6-phosphate receptor (MPR), cathepsin L (CSTL), Syntaxin7 (Stx7), mucolipin (MCOLN), and Ras-related protein Rab-35 (Rab35) (Fig. 4A, Additional file 2: Table S11). The above processes may be critical for the regulation and digestion of symbionts, and the related genes are highly expressed in symbiotic tissues [67, 68]. However, according to our analysis, they are only significantly expanded

Guo *et al. BMC Biology*   (2023) 21:51

Page 9 of 20

in *C. bisecta*, highlighting their possible contribution on extracellular symbiosis. For instance, MPR plays a key role in the phagosome maturation by importing hydrolases to phagosome [69]. In contrast to 2 or 5 copies coded by the endosymbiotic bivalves, 15 gene copies of MPR were found in *C. bisecta* genome (Fig. 3C), and four are specifically expressed in the gill tissues (Additional file 1: Fig. S15). Moreover, an expanded set of MCOLN genes, which are crucial for lysosomes recycling [70–72], is present in *C. bisecta*. Based on phylogenetic relationships, MCOLNs split into two clades in bivalves, and thyasirid MCOLNs expanded in both clades and clustered with their heterodont relatives (Fig. 4B). However, based on our analysis, genes implicated in the internalization of the symbionts were only expanded in the endosymbiotic bivalves but the extracellular symbiotic *C. bisecta* (Fig 4A). For example, Rac family small GTPase 1 (Rac1) and phosphatidylinositol 3-kinase (VPS34) were key factors in this process [73, 74]. These two families were only found to be expanded in the endosymbiotic bivalves, but not in the extracellular symbiotic *C. bisecta* (Fig. 3C, Additional file 2: Table S11). Thus, although phagocytosis-related expansion events were observed in all symbiotic bivalves, only the thyasirid genome exhibited overall expansion of bacterial digestion and subsequent lysosome recycling genes. In contrast, the endosymbiotic bivalves likely show enhanced bacterial internalization.

Interestingly, we found some genetic evidence related to a rapid bacterial clearance rate in *C. bisecta.* Compared to the endosymbiotic bivalves, cathepsins, the hydrolases responsible for the catabolic ability of lysosomes with high expression in gills, were significantly expanded in thyasirid (Fig. 3C, Additional file 1: Fig. S15) [22, 65, 66, 75]. Based on the divergence time of *ctsL* genes, four copies (Conbi17702, Conbi17048, Conbi07646 and Conbi23783) were estimated to have expanded in a narrow window, while two of them were identified as tandem duplicate pairs (Fig. 4C). Interestingly, among the four *ctsL* copies, three genes were exclusively expressed and accounted for most of the *ctsL* transcripts in gill tissue (Fig. 4C). We speculate the expansion events enhance bacterial digestion. The processes from early phagosome through symbiont digestion may be particularly efficient in the extracellular symbiotic *C. bisecta*, which corresponded to the phenotype of the thyasirids, where most internal vesicles retain lysed bacterial symbionts [24, 49, 76]. Furthermore, expanded gene families related to lysosome recycling have been implicated in forming a large phagocytic cup via promoting membrane extensions [77–79]. The large apical vesicles harboring densely aggregated *C. bisecta* symbionts were stunning. Additional membrane is required when the apical vesicles form, which might be related to these expansion events [72]. In addition, compared to endosymbiotic bivalves, the capability of internalizing symbionts might be relatively modest in *C. bisecta*, accentuating the above phenotypic variations between the extracellular symbiotic thyasirid and endosymbiotic bivalves.

In conclusion, the massive expansion in phagocytosis is in line with the theory that thyasirids periodically engulf and digest symbionts [28, 76]. In addition, the gene families related to phagocytosis divergently evolved among different bivalves. These divergences may be related to the phenotypic variations, such as the large apical vesicles in extracellular symbiotic clams, and the large amounts of endosymbiotic vesicles with live bacteria in both *A. marissinica* and *G. platifrons*.

Immune system remodeling to the establishment and maintenance of extracellular symbiosis in *C. bisecta*

According to our comparative genomic analysis, the immune system of *C. bisecta* is remodeled differently from that of *A. marissinica* and *G. platifrons* to make room for its extracellular symbiont (Fig. 5A). The roles of pattern recognition receptors (PRRs) in bacteria selection and capture may be reduced in thyasirid as four out of 13 PRR-related Pfam domains are contracted (Fig. 5B, Additional file 2: Table S13). However, the expansion of mucosal immunity-related genes, including mucin protein families and the GCNT1 (acetylglucosaminyltransferase) family for mucin glycosylation, has been observed, and these genes were functional in the gill tissues as are supported by both transcriptomes (Fig. 3C, Additional file 1: Fig. S16). As described, TEM examination and 16S rDNA analysis found SCbi was the only dominant bacteria in the apical vesicles of bacteriocytes, indicating that inter-partner recognition of symbionts has completed before the settlement of the symbionts in the apical vesicles. Thus, we speculate that the selection of symbionts occurred extracellularly possibly at the mucin layer based on both microscopic observation and comparative genomics. The mucosal interface is the first physical encounter where the interactions of microbe and host begin through bindings of host surface glycosylated proteins and symbiont sugars [25]. In the squid-vibrio system, it has been revealed that the glycosylated surface mucus secreted by the host is the key factor in recruiting and entraining of bacterial symbionts from seawater [5]. Therefore, although both *C. bisecta* and *G. platifrons* acquire symbionts by horizontal transmission, the expanded mucins and GCNT1 enzymes in *C. bisecta* are likely to function in bacterial symbiont aggregation and specificity of the bacterial symbionts, rather than via PRRs as in deep-sea mussels.

Guo *et al. BMC Biology*   (2023) 21:51

Page 10 of 20
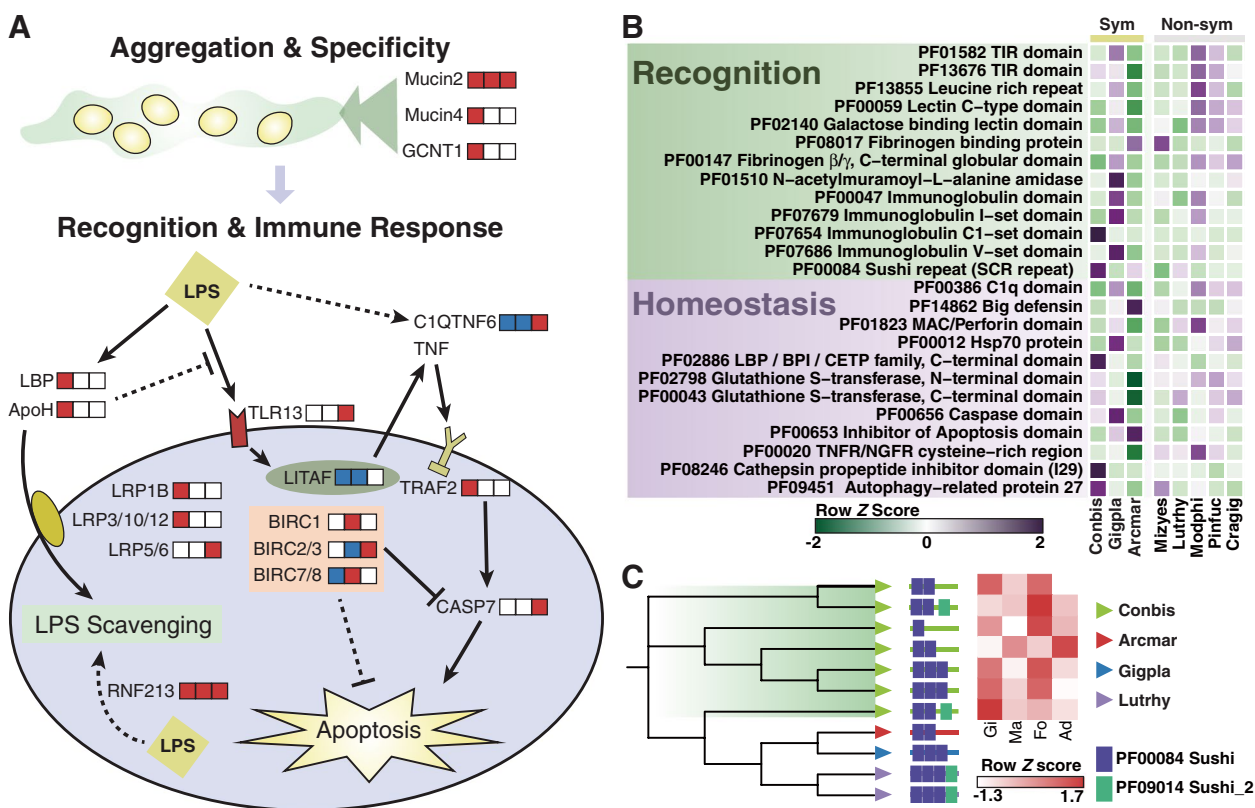
Fig. 5 Gene family related to symbiosis establishment and maintenance in *Conchocele bisecta*. **A** Schematic representation of bacterial recognition and immune response in *C. bisecta*. Both expansion and contraction events of gene families involved in these processes are illustrated, and the boxes that followed each gene indicate if a gene family has expanded (red), contracted (blue), or neither (white) in *C. bisecta* (right), *Archivesica marissinica* (middle) and *Gigantidas platifrons* (left). **B** Heat maps of significantly evolved Pfam domains implicated in recognition and homeostasis in the three symbiotic (Sym) bivalves (Conbis: *C. bisecta*; Arcmar: *A. marissinica*; Gigpla: *G. platifrons*), compared with asymbiotic (non-sym) relatives (Lutrhy: *Lutraria rhynchaena*; Mizyes: *Mizuhopecten yessoensis*; Modphi: *Modiolus philippinarum*; Pinfuc: *Pinctada fucata*; Cragig: *Crassostrea gigas*). **C** Left, phylogenetic tree based on amino acids of apolipoprotein H. Predicted Pfam domains for each gene are shown. Right, heatmap showing the tissue-specific expression (Gi, gill; Ma, mantle; Fo, foot; Ad, adductor) of *ApoH* (row-wise *Z*-score of TPM counts)

Once symbiotic relationships are established, the bacteriocyte may suffer from stressors like ROS (reactive oxygen species) due metabolic waste from symbionts. Hence, homeostasis maintenance strategies are needed to sustain the symbiosis system. Inhibitors of apoptosis protein (IAPs) are known to inhibit apoptosis and promote cell cycle progression and are expanded and highly expressed in gills of endosymbiotic bivalves to maintain cellular homeostasis (Fig. 5A) [22, 26]. However, the copy number of IAP Pfam domains in *C. bisecta* [37] is somewhat lower than the asymbiotic bivalves (56 on average) and much lower than endosymbiotic bivalves (232 for *A. marissinica*, 123 for *G. platifrons*) (Fig. 5A, B). In *C. bisecta*, most symbionts remain in apical vesicles among the microvilli of bacteriocytes and are digested once entering bacteriocytes. Thus, population control of symbionts by suppressing host cell apoptosis may not be required.

Although mainly located outside the cell, the thyasirids nevertheless need to accommodate the lifestyle of extracellular symbionts resident at their gills and foot. LPS is a common cell surface antigen of Gram-negative bacteria and one of the most potent initiators of inflammation and apoptosis, which can induce cytokine production and cause cytotoxic effects in animals, including mollusks (Fig. 5A) [80, 81]. We found that several gene families coding LPS-binding proteins expanded substantially and are highly expressed in both tissues of gill and foot (Fig. 5C). For example, apolipoprotein H (ApoH) can associate with and scavenge LPS by binding members of the low-density lipoprotein receptor-related protein (LRP) family to reduce endotoxin concentration [82, 83]. In addition to the expansion of ApoH, an expansion of three LRP families, including the LRP1B (14 vs. 3.7 gene copies in all other seven analyzed bivalves), LRP3/10/12 (26 vs. 4.9 gene copies), and LRP4 (34 vs. 21.7 gene copies), were observed in *C. bisecta* (Fig. 5A,

Guo *et al. BMC Biology*   (2023) 21:51

Page 11 of 20

Additional file 2: Table S11). LRP1B genes are the most highly expressed family among the expanded LRPs and are mainly expressed in the gills and mantle, while LRP3/10/12 genes are mainly expressed in the three muscle tissues (Additional file 1: Fig. S16). These results support a hypothesis that the host may maintain the relationship with the bacterial symbionts by neutralizing the defects of bacterial LPS. As additional evidence, other gene families involved in the immune response to LPS were significantly evolved, including the tumor necrosis factors (TNF), lipopolysaccharide induced TNF factors (LITAFs), and E3 ligases (Fig. 5A, Additional file 2: Table S11). For instance, we found that the E3 ubiquitin-protein ligase ring finger protein 213 (RNF213) has expanded in symbiotic bivalves. This family was recently reported to induce antibacterial autophagy by ubiquitylating LPS in cytosol [84].

Taken together, significant genetic differences have been observed in the two bivalves that obtain symbionts horizontally. In *C. bisecta*, glycosylated mucins, rather than the PRRs, may aid in symbiont integration, which is crucial for the horizontal transmission of symbionts in thyasirids [39, 40]. Although the convergent expansion of IAPs in endosymbiotic bivalves was not observed in *C. bisecta*, the expansion and high expression of multiple gene families may allow this thyasirid clam to resist the toxic effect of extracellular symbiosis.

Independent genome evolution drives the diversify of the deep-sea bivalve symbioses

Symbiosis is a major driven force of evolution, and this relationship shapes the phenotypic complexity of the symbiotic host [13, 85]. Among bivalves, chemosymbiosis with primary producers in the deep sea is one of the convergent adaptations to the oligotrophic environmental context [7]. However, symbiosis-related phenotypes, such as the spatial structure of bacteriocytes and the physiological adaptations between the host and the symbiont, are highly variable in bivalves [25]. The evolutionary fingerprints under these variations need further investigation.

Symbiont digestion through phagolysosome contributes significantly to the nutritional transfer from the symbiont to the host, and evidence has been obtained from both mussels and clams [33, 34]. However, as revealed by genomic comparisons, gene families related to symbiont digestion were divergently evolved among the three symbiotic bivalves. The expansion events in *C. bisecta* may enhance the phagocytic capabilities of this clam. For instance, cathepsins are representative hydrolases that participate in symbiont digestion, and the gene duplication of *ctsL* likely facilitated its specific expression in the symbiotic gill tissue. Regarding substrate transport from

the host to the symbionts, genetic divergences were also observed between the extracellular symbiotic thyasirid and its endosymbiotic relatives. Thus, these divergences in currency exchange might account for the distinct spatial structure of bacteriocytes among symbiotic bivalves, whereas these evolutionary events converged to the facilitating of currency exchange.

The immune systems of the three symbiotic bivalves were remodeled at the genome level, but the evolutionary patterns were distinct. For instance, the bacterial LPS is one of the most potent initiators of inflammation and toxic to bivalves [80, 81]. According to our analysis, the symbiotic bivalves employed different pathways to adapt to the long-term co-existence with their bacterial symbionts. For example, expanded IAP families in the endosymbiotic bivalves have been linked to the homeostatic maintenance in bacteriocytes [22, 26]. Instead, we found that the expansion events in the LPS scavenging pathway might account for the thyasirid's adaptation to the bacteria-hosting life. Likewise, although the *G. platifrons* and *C. bisecta* recruit symbiont horizontally and show a high degree of specificity with the symbiont, the bacterial recognition-related gene expansion events were divergent.

Our results suggest that distinct genome events, in particular gene duplications, underlie the independent evolution of symbiosis by deep-sea bivalves. Gene duplication results from stochastic accidents during inheritance and is considered a rich source of evolutionary substrates [86, 87]. The emerging evolutionary picture from our data is that gene duplications occurred persistently and stochastically, either before or after the establishment of symbiosis in the ancestors of symbiotic bivalves. During the many millions of years of co-evolution between the bivalves and their symbionts, selective pressure has resulted in extant species with optimal currency exchange and a high-tolerance immune strategy.

Conclusions

Biologists have long been fascinated by the evolution of symbiotic relationships. Bivalves have shown varied symbiosis forms, and phenotypic divergence-driven comparison of bivalve genomes may provide insights into this question. By characterizing the hologenome of *C. bisecta*, genetic clues for the nutritional complementarity and immune interactions between the host and the Thioglobaceae symbiont were provided. Furthermore, by comparing this extracellular symbiotic thyasirid with endosymbiotic bivalves, we show that bivalves employ divergent evolution of genes and pathways, including phagocytosis, bacterial recognition, and immune response to LPS, to adapt to the long-term co-existence with their bacterial symbionts, highlighting the contribution of stochastic evolution to the independent gain of a

symbiotic lifestyle in the lineage. Thus, our work provides a resource and valuable insights for understanding the evolution and dynamics of invertebrate symbiosis.

Methods
Sample collection
Individuals of *Conchocele bisecta* were collected during R/V *Kexue* "2018 Hydrothermal vent & Cold-seep combined expedition" at hydrothermal vents field in the East China Sea (27° 47.5′ N, 126° 53.8′ E), with the help of TV grab. Anatomy was processed on ice immediately after sampling, and tissues were stored at −80℃ unless otherwise mentioned.

FISH and TEM analysis
Both the gill tissue and gonad tissue of freshly collected clam were prepared for fluorescent in situ hybridization (FISH) analysis. In brief, the tissues were fixed in 4% paraformaldehyde at 4℃ for 16 h. The fixed tissues were washed in cold PBS for three times and dehydrated in 100% methanol, and dehydrated tissues were stored at −20℃ before use. The tissues were embedded with Paraplast plus (Sigma) first and cut into sections of 7-μm thickness with a microtome (Leica) for FISH analysis. Probe was labeled with Cy3 for FISH analysis and designed based on 16S rDNA sequence of the symbiont. The 16S-Probe was mapped to both the host genome and the assemblies of metagenomes of gill tissue to avoid unspecific hybridization. Sequential sections of gill or gonad tissue of *C. bisecta* were analyzed using both the 16S-Probe and the reverse complement one of it. The FISH experiments were performed according to the method described by Halary [88].

For transmission electron microscope (TEM) analysis, gill tissues were dissected and fixed in 2.5% glutaraldehyde and 2% paraformaldehyde at 4℃. Then the tissue was post-fixed with 1% osmium tetroxide and embedded in Ep812 resin. Sections of 70-nm thickness, which generated by a ultramicrotome (Reichert-Jung ULTRA CUT E), were double-stained with lead citrate and uranyl acetate. The sections were examined using a TEM (JEM1200, JEOL) set to 100 kV.

Amplicon sequencing of 16S rDNA V4 region
DNA templates that extracted from gill tissues of three individuals and gonad tissue of one individual were used for amplicon sequencing of 16S rDNA V4 region. The libraries were constructed and sequenced according to the instruction of BGISEQ-500 platform. Adapters and low-quality reads were removed, and operational taxonomic units (OTUs) were assigned to the SILVA ribosomal RNA database (r123) using the USEARCH pipeline with a 97% similarity cut-off. All OTUs mapped by more

than 10 reads in any of the three samples were considered to be candidates of symbionts.

Nucleic acid extraction and sequencing
Genomic DNA was extracted from the muscle tissue of *C. bisecta* using QIAamp DNA Mini Kit (Qiagen), and the DNA was examined with the Agilent 4200 Bioanalyzer (Agilent Technologies). Similarly, metagenomic DNA was exacted from tissues of gill and gonad by an identical progress, respectively.

For WGS sequencing, the DNA was fragmented to 500–800 bp in size with Covaris E220 and selected using AMPure XP beads to obtain fragments around 200 bp. Then the fragments were end-repaired and A-tailed with T4 DNA polymerase (NEB), T4 polynucleotide kinase (NEB), and *rTaq* DNA polymerase (Takara). After that, the DNA were amplified for eight PCR cycles and sequenced on the BGISEQ-500 platform with a layout of pair-end 100 bp.

For PacBio sequencing, 8 μg of extracted genomic DNA was sheared and concentrated with AMPure PB beads. The libraries were constructed using the Pacific Biosciences SMRTbell express template prep kit 2.0, and the constructed libraries were selected on a BluePippin system for molecules longer than 20 kb. Finally, the templates were primer annealed and bound to polymerases with the DNA/Polymerase Binding Kit, and sequencing was carried out on the Pacific Biosciences Sequel II platform for 15 h.

For Hi-C library construction, gonad tissue was dissociated, and cells were collected and crosslinked with 1% formaldehyde (Sigma) and 0.2M glycine (Sigma). After that, the fixed powder was resuspended in nuclei isolation buffer and then incubated in 0.5% SDS for 10 min at 62℃. Then the reaction was quenched with 10% Triton X-100 (Sigma) and the nuclei were collected by centrifugation. Then the DNA was digested with *MboI* (NEB), and the overhang was filled and biotinylated before ligated by T4 DNA ligase (NEB). Before library construction, the DNA was purified using the CTAB method. The purified DNA was sheared, and biotin-containing fragments were captured on streptavidin-coated beads using Dynabeads MyOne Streptavidin T1 (Invitrogen). Then the fragments were end-repaired and linked with adaptors before eight cycles of PCR reaction with KAPA HiFi HotStart ReadyMix (Kapa Biosystem). After that, the Hi-C library was sequenced with BGISEQ-500 platform with a layout of pair-end 100 bp.

For RNA sequencing, tissue-specific total RNA was extracted from tissues of adductor muscle, mantle, foot, and gill with TRIzol (Invitrogen), and the reverse transcription was performed with HiscriptII (Vazyme) to generate cDNA. The cDNA fragments were sequenced

Guo *et al. BMC Biology*   (2023) 21:51

Page 13 of 20

on the BGISEQ-500 platform with a layout of pair-end 100 bp.

For metagenome sequencing, selected DNA fragments were end-repaired, A-tailed, and amplified, and the libraries were sequenced on the BGISEQ-500 platform with a layout of pair-end 100 bp. In addition, DNA of gill tissues for Oxford Nanopore Technologies (ONT) long-read sequencing was extracted from gill tissues of a third individual with the Blood & Cell Culture DNA Midi Kit (Qiagen), and the library were prepared and sequenced with the Oxford Nanopore Ligation Sequencing Kits SQK-LSK109 according to the manufacturer's instructions.

Mitochondrial genome assembly and collinearity analysis

Conchocele cf. *bisecta* HPD1644 mitochondrial sequence [31], retrieved from the NCBI with sequence number "LC126312.1," was used as the "Seed Input" in the configuration file of NOVOPlasty v4.2 [89], resulting in the mitochondrial genome of *C. bisecta* in this study. The genome was annotated using online tool of MITOS (http://mitos2.bioinf.uni-leipzig.de/index.py) [90]. For collinearity analysis, the assembled mitochondrial sequences were aligned to that of *Conchocele* cf. *bisecta* HPD1644 using Sibelia v2.1.1 [91] to generate a Circos configure file and plot in Circos v0.69 [92].

Host genome assembly and chromosome anchoring

Before genome assembly, all of the generated short reads were quality filtered using SOAPnuke v1.5.2, with the parameter as "-l 20 -q 0.2 -Q 2 -d" [93]. K-mer frequency-based method was used for genome size estimation. In detail, the K-mers were counted using Jellyfish v2.2.6 with the parameter "-m 17", and the output file was employed to estimate the genome size with "histo" of Jellyfish v2.2.6 [94].

The generated PacBio reads were converted to fasta format using bam2fastq. These raw reads were assembled by Shasta-OldLinux-0.4.0 with a PacBio-CLR configure file (minReadLength = 10000, maxAlignmentCount = 50, consensusCaller = Modal) [95]. This primary assembly was then polished twice using clean short reads generated by BGI sequencer by Pilon v1.22 [96]. BUSCO v5.3 estimation of the percentage of complete metazoan (odb10) dataset reached 87.8% [97]. Contigs from this assembly were then clustered using Hi-C data, after quality control process with HiC-Pro v3.2 [98] and assembled by 3D-DNA [99], the final heat map showed these contigs were mounted into 17 pseudo-chromosomes visualized in Juicebox v1.9 [100].

Repeat annotation

Tandem Repeats Finder v 4.0.7 program was applied to detect tandem repeats in the genome [101].

Homolog-based and de novo prediction methods were integrated to identify transposable elements (TEs). For the de novo search, LTR_Finder v1.0.6 [102] and RepeatModeler v1.0.8 [103] were employed to find repetitive elements with specific consensus models. For the homology-based search, RepeatMasker v4.0.6 and RepeatProteinMask v4.0.6 against the Repbase v21.01 database were used at the nucleotide and protein levels respectively [104, 105]. Secondly, RepeatMasker was employed again to detect species-specific TEs against the database concatenated by the results of LTR_Finder and RepeatModeler together. All other species used in this work were annotated repetitive elements following the same pipeline for comparative analysis.

Gene annotation

Ab initio, homology-based and gene expression evidence were combined to predict protein-coding genes in the genome of *C. bisecta*. Augustus v3.1 was first employed on repeat-masked genome for ab initio gene prediction [106]. For the homology-based annotation, gene sets from 10 molluscan species (*Archivesica marissinica*, *Biomphalaria glabrata*, *Crassostrea gigas*, *Gigantidas platifrons*, *Lottia gigantea*, *Lutraria rhynchaena*, *Modiolus philippinarum*, *Octopus bimaculoides*, *Pinctada fucata*, and *P. canaliculate*) were used. These homologous protein sequences were first aligned onto the genome of *C. bisecta* using Blast v2.2.26 with an e-value cut-off of 1×10^{-5} [107], and then we linked the alignment hits to candidate gene *loci* by GenBlastA [108]. Secondly, genomic sequences of candidate gene regions together with their 2-kb flanking sequences were extracted and used GeneWise v2.2.0 to determine gene models [109]. Moreover, Stringtie v 1.3.4 was employed to generated gene annotation files on RNA-Seq alignments generated by HISAT v2.1.0 of different tissues (adductor muscle, mantle, foot, and gill) [110, 111]. Then these files were merged together to predict candidate coding regions open reading frames (ORFs) using Transdecoder v5.5.0 and were aligned to genomes to obtain a gene annotation file with transcript evidence. Finally, these three evidences were integrated using EVM v1.1.1 to obtain a final version of protein-coding genes [112], and their function were annotated by searching against the following public databases: Swiss-Prot v201709, KEGG v87.0, InterPro v55.0, and TrEMBL v201709. The other 7 species used in gene family analysis were functionally annotated in the same way.

Phylogenetic relationships and divergence times of mollusks and selected bivalves

Twenty-four well-assembled lophotrochozoan genomes were selected for phylogenetic analysis, include one

annelid (*Helobdella robusta*) as outgroup, 21 bivalves (*Archivesica marissinica, Argopecten concentricus, Argopecten irradians, Conchocele bisecta, Crassostrea gigas, Crassostrea virginica, Cyclina sinensis, Gigantidas platifrons, Lutraria rhynchaena, Mactra quadrangularis, Mercenaria mercenaria, Mizuhopecten yessoensis, Modiolus philippinarum, Mytilus coruscus, Pecten maximus, Pinctada fucata, Pinctada imbricata, Ruditapes philippinarum, Saccostrea glomerata, Scapharca broughtonii, Sinonovacula constricta*), 5 gastropods (*Aplysia californica, Chrysomallon squamiferum, Lottia gigantea, Haliotis rufescens, Pomacea canaliculata*), and 2 cephalopods (*Octopus bimaculoides* and *Octopus vulgaris*) [22, 26, 52, 113–132]. SonicParanoid v1.3.0 was used to define gene family clusters among different species [133]. The amino acid sequences of one-to-one single-copy orthologous genes were used to reconstruct their phylogenetic topology. The protein sequences were aligned using MAFFT v7.407 under default settings [134], and then were concatenated for phylogenetic analysis using a maximum-likelihood method implemented in IQ-TREE v 2.0.6 with the "-m MFP" parameter was applied to each protein partition [135]. To estimate divergence times, the rooted maximum-likelihood tree, along with a concatenated fourfold degenerate site sequence extracted from single-copy CDS (coding sequence), was used as the input of MCMCtree software implemented in PAML v4.8 [136]. For calibration, nine nodes were constrained by either fossil records obtained from website of TimeTree.

Host gene family analysis and domain analysis

For expansion and contraction analysis, in addition to genome of *C. bisecta*, we selected 7 representatives with good BUSCO performance out of the 15 collected bivalve genomes. The selected genomes include that of the only two published endosymbiotic bivalves (*A. marissinica, G. platifrons*) [22, 26], and 5 asymbiotic bivalves (*Crassostrea gigas, L. rhynchaena, M. yessoensis, Modiolus philippinarum, P. fucata*) which were separated in different bivalve clades and not known to host chemosynthetic bacteria. Before analysis, HMMSCAN (HMMER v3.1) was applied to identify Pfam domains in protein-coding gene sequences among the selected bivalve. The Pfam domains of the respective species were counted to construct a data frame, while multiple copies of a same domain in the same gene were counted as one.

Gene family analyses in the symbiotic bivalves were conducted using one-tailed Fisher's exact tests for either expansion or contraction. In detail, for gene expansion/contraction at the protein domain level, we first calculated the counts of each Pfam domain in each genome of the 8 analyzed species, and the Pfam domain counts in each of the symbiotic bivalves (*A. marissinica, C. bisecta, G. platifrons*) was compared against the background average domain counts of the five asymbiotic bivalve genomes (*Crassostrea gigas, L. rhynchaena, M. yessoensis, Modiolus philippinarum, P. fucata*), which method was employed by Sun et al. for comparative genomic analysis [26]. Furthermore, we conducted the same analysis with the gene counts of each KEGG ortholog on each of the three symbiotic bivalves. After that, Pfam domain or KEGG ortholog with a *P* value less than 0.05 is considered statistically expanded or contracted in the three symbiotic bivalves. Finally, the evolutionary patterns of *A. marissinica, C. bisecta*, and *G. platifrons* were compared according to the expansion/contraction results by Fisher's exact tests.

For phylogenetic analysis of each gene family, we employed Muscle v3.8.31 for multiple sequence alignment [137], and the phylogenetic trees were constructed with FastTreeMP v2.1.10 [138]. Specially, for reported expansion events of subfamilies of hemoglobin, which were not included in the KEGG database, we performed additional alignment with the sequences mentioned in Ip et al. [22] using Diamond, and phylogenetic analysis was conducted as the same.

Time of gene duplication

To evaluate the temporal dynamics of expanded gene families during the evolution of *C. bisecta*, the nucleotide substitution rates of bivalves were calculated by the branch distance divided by the estimated divergence time using MCMCtree. With default settings of MAFFT and "-automated1" option of trimAl v1.4 [139], all paralogs of the target gene family were aligned to determine the time required for gene duplication. The Nei-Gojobori pairwise codeml method was used to determine the dN values for all aligned pairs. Divergence times of gene pairs were estimated using the equation $T = K/2r$ [140], where T is the insertion time, and r is the nucleotide substitution rate. The relationships between different gene pairs are determined following the DupGen_Finder (https://github.com/qiao-xin/DupGen_finder) pipeline, using *Nematostella vectensis* as a reference genome.

Transcriptome analysis

Adaptor sequence removal and read quality filtering from the raw transcriptome datasets were performed using SOAPnuke, with the parameter as described above. HISAT2 were used to build index and align clean RNA reads to the reference genome. RNACocktail v0.2.1 were used to build genome-guided transcriptome assemblies [141]. Tissue-specific expression levels were quantified using Salmon v0.8.2 [142].

Metagenomic assembly and binning

The raw reads generated from either metagenomic libraries of the two individual gill tissues were filtered using SOAPnuke. The filtered reads which were confidently aligned to the host's genome were removed, and remained reads were assembled with MEGAHIT v1.1 (--min-count 2/3 --k-min 33 --k-max 53 --k-step 10 --no-mercy) [143]. Metagenome-assembled genomes (MAGs) were binned with Metawrap v1.1.5 [144], and the bins were combined and filtered with the "bin_refinement" module from Metawrap. To further completing the MAGs, all metagenomic short reads were mapped to each MAGs, and we employed the hybrid assembler Unicycler v0.4.8 to de novo assemble the genomes with both the mapped short reads and all of the ONT reads [145]. Finally, we calibrated the MAGs again by using the GATK pipeline to remove potential errors, and the MAGs' taxonomic assignment was completed with GTDB-tk v1.0.2 [146, 147]. CheckM v1.0.12 was used to estimate the completeness and contamination of the genomes [148].

To annotate the symbiont genome, we employed Prokka v1.14.6 to predict the CDS region [149], and tRNAscan-SE v1.3.1 to identify the tRNA gene [150], and RNAmmer v1.2 were used to identify the rRNA gene for annotation [151]. In parallel, Diamond was employed to align the protein sequences to the KEGG. Transporters were predicted by online tools of TransAPP and ABCdb [41, 42]. Metabolic potential of the bacterial symbiont in amino acids and cofactors were checked thoroughly, and missed genes were confirmed by aligning the ONT long reads to the chromosomal level genome of endosymbiont of *Bathymodiolus septemdierum* (NCBI Accession Number: NZ_AP013042.1), while the results were visualized in IGV v2.10 [152].

Metaproteomics

The gill tissue was grinded and sonicated using a high-intensity ultrasonic processor (Scientz). After centrifugation at 4°C, the supernatant was collected and the protein concentration was determined with BCA kit according to the manufacturer's instructions. Next, the protein solution was reduced with 5 mM dithiothreitol for 30 min at 56°C and alkylated with 11 mM iodoacetamide for 45 min at room temperature in darkness, and trypsin was employed to digest the protein. Tryptic peptides were dissolved and separated on a nanoElute UHPLC system (Bruker Daltonics). The peptides were subjected to mass spectrometry with the timsTOF Pro (Bruker Daltonics) in a parallel accumulation serial fragmentation (PASEF) mode. With 1.65 kV, precursors and fragments were analyzed at the TOF detector with settings including Tandem Mass Spectrometry (MS/MS) scan range from 100 to 1700 m/z, precursors with charge states 0 to 5 for fragmentation, 10 PASEF-MS/MS scans per cycle, and 30s for dynamic exclusion. The data were processed with MaxQuant search engine v.1.6.15.0. Tandem mass spectra were searched against the protein catalog of the hologenome (26928 entries) concatenated with reverse decoy database. The mass tolerance for precursor ions was 20 ppm and that for fragment ions was 0.02 Da, and FDR (false discovery rate) was adjusted to 0.01. The iBAQ values of the gill tissue were normalized as iBAQ values per million.

Abbreviations

Ad	Adductor
ALG	Ancestral linkage group
ApoH	Apolipoprotein H
Arcmar	*Archivesica marissinica*
b	Bacterial symbionts
BUSCO	Benchmarking universal single-copy orthologs
CCA	Cytoplasmic carbonic anhydrase
CCKAR	Cholecystokinin receptor
CDS	Coding sequence
CHRNA	Acetylcholine receptor
CI	Conservation index
CNGBdb	China National GeneBank DataBase
CNSA	CNGB Sequence Archive
Conbis	*Conchocele bisecta*
Cragig	*Crassostrea gigas*
CSTL	Cathepsin L
FDR	False discovery rate
FISH	Fluorescence in situ hybridization
Fo	Foot
GCNT1	Acetylglucosaminyltransferase
Gi	Gill
Gigpla	*Gigantidas platifrons*
GTDB	Genome Taxonomy Database
Hb	Hemoglobin
IAP	Inhibitor of apoptosis protein
iBAQ	Intensity-based absolute quantification
KO	KEGG ortholog
LITAF	Lipopolysaccharide induced TNF factor
LPS	Lipopolysaccharide
LRP	Low-density lipoprotein receptor-related protein
Lutrhy	*Lutraria rhynchaena*
Ma	Mantle
MAG	Metagenome-assembled genome
MCOLN	Mucolipin
Mizyes	*Mizuhopecten yessoensis*
Modphi	*Modiolus philippinarum*
MPR	Mannose-6-phosphate receptor
MS/MS	Tandem mass spectrometry
mv	Microvilli
MY	Million years
MYA	Million years ago
n	Nucleus
NCBI	National Centre for Biotechnology Information database
ONT	Oxford Nanopore Technologies
ORF	Open reading frame
OTU	Operational taxonomic unit
PASEF	Parallel accumulation serial fragmentation
Pinfuc	*Pinctada fucata*
pl	Phagolysosome-like organelles
PRIDE	Proteomics Identification database
PRR	Pattern recognition receptor
Rab35	Ras-related protein Rab-35
Rac1	Rac family small GTPase 1
RNF213	E3 ubiquitin-protein ligase ring finger protein 213

ROS	Reactive oxygen species
SCbi	The bacterial symbiont of *Conchocele bisecta*
Stx7	Syntaxin7
TE	Transposable element
TEM	Transmission electron micrography
TNF	Tumor necrosis factor
TPM	Transcript per million
v	Vesicle
VB	B vitamins
VPS34	Phosphatidylinositol 3-kinase

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-023-01551-z.

Additional file 1: Supplementary Note 1. Taxonomic identification of the bivalve samples. **Supplementary Note 2.** Metabolic potential of SCbi. **Supplementary Note 3.** The genome assembly of *Conchocele bisecta*. **Supplementary Note 4.** Transporters for B type vitamins in *C. bisecta*. **Supplementary Note 5.** Filter-feeding related gene families in *C. bisecta*. **Figure S1.** Shells and mitochondrial genome of *C. bisecta*. **Figure S2.** Fluorescence in situ hybridization (FISH) of 16S rRNA of the dominate symbionts in the gill filaments. **Figure S3.** Biosynthesis pathways of amino acids in *C. bisecta* and its symbionts. **Figure S4.** Biosynthesis pathways of cofactors in *C. bisecta* and its symbionts. **Figure S5.** Alignment results by mapping the ONT reads of SCbi to a published SUP05 genome. **Figure S6.** Expression of genes involved in biosynthesis of threonine in both the symbiont and its *C. bisecta* host. **Figure S7.** Expression of genes involved in biosynthesis and transport of folate in both the symbiont and its *C. bisecta* host. **Figure S8.** The assembly of the genome of *C. bisecta*. **Figure S9.** Chromosome-scale macro-synteny comparison between *Nematostella vectensis* and six bivalves. **Figure S10.** Relationship between genome size and transposable elements (TE) in the lophotrochozoan genomes. **Figure S11.** Expanded or contracted Pfam domains in *C. bisecta*. **Figure S12.** Expression levels of genes related in transport of rare metabolites in *C. bisecta*. **Figure S13.** Carbonic anhydrases in bivalves. **Figure S14.** Hemoglobin and hemoglobin-like proteins in bivalves. **Figure S15.** Expression levels of expanded genes that implicated in phagocytosis in *C. bisecta*. **Figure S16.** Expression levels of expanded genes that implicated in recognition and homeostasis in *C. bisecta*. **Table S2.** Genomic statistics of the SCbi (Symbionts of *C. biescta*). **Table S5.** Distribution genes involved in amino acids biosynthesis in SUP05 bacteria. **Table S6.** Statistics of sequenced genomic data. **Table S7.** Characteristics of the *C. bisecta* genome assembly. **Table S8.** The mapping rate of short reads to *C. bisecta* genome. **Table S9.** Chromosome statistics of *C. bisecta* genome. **Table S10.** TE contents (%) of lophotrochozoan genomes.

Additional file 2: Table S1. Constitution of prokaryotes in the gill of *C. bisecta* based on 16S rDNA amplicon analysis. **Table S3.** Putative transproter genes in SCbi. **Table S4.** Genes in biosynthesis pathways of amino acids and cofactors. **Table S11.** The expansion and contraction events among the three symbiotic bivalves (*Conchocele bisecta*, *Archivesica marissinica*, *Gigantidas platifrons*). **Table S12.** Counts of pfam domains related to the glycosyl hydrolase families in bivalve genomes. **Table S13.** Number of Pfam protein families related to the PRRs that involved in bacterial recognition.

Authors' contributions
Y. G., M. W., L. M., and H. L. conceived the idea. Z. Z., H. Z., and M. W. collected the sample. Y. G. and Z. Z. performed the TEM and FISH experiments. A. J., Q. J., and X. S. prepared the libraries for sequencing. L. M., Y. G., Y. Z., and Y. L. finished the genomic and transcriptomic analyses. D. L. and Y. G. contributed the metagenomic analyses. Y. G. and L. M. wrote the manuscript with the input of D. L. and Z. Z., and I. S., M. W., G. F., and C. L. revised the manuscript. M. W., G. F., C. L., and S. L. supervised the study. Y. Z., H. L., J. C., and H. Z. provided helpful comments on this study. All authors reviewed and approved the final manuscript.

Availability of data and materials
The assemblies and protein sequences of published genome used in the present study were from the NCBI repository: PRJNA629898 for *Archivesica marissinica* [22], PRJNA828432 for *Crassostrea gigas* [115], PRJNA328542 for *Gigantidas platifrons* [26], PRJNA548223 for *Lutraria rhynchaena* [121], PRJNA259405 for *Mizuhopecten yessoensis* [52], PRJNA328544 for *Modiolus philippinarum* [26], PRJNA283019 for *Pinctada fucata* [120]. All sequencing data for amplicons, genome, transcriptome, and metagenome, as well as the assemblies of both *Conchocele bisecta* and its bacterial symbiont, are available in the NCBI (National Centre for Biotechnology Information) repository under project PRJNA913320 [153]. In addition, all above datasets can also be found at CNGB Sequence Archive (CNSA) of the China National GeneBank DataBase (CNGBdb) under project CNP0003505 [153]. The metaproteomic data of *C. bisecta* was deposited to the Proteomics Identification (PRIDE) database under the accession number PXD036936 [154]. Genome assemblies and annotations of the *C. bisecta* hologenome are available in Figshare [155]. All computer commands are provided at Github [156].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Center of Deep-Sea Research, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China. [2]BGI-Qingdao, BGI-Shenzhen, Qingdao 266555, China. [3]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China. [4]Pilot National Laboratory for Marine Science and Technology, Qingdao 266237, China. [5]BGI-Shenzhen, Shenzhen 518083, China. [6]Integrative Biology Laboratory, College of Life Sciences, Nanjing Normal University, Nanjing 210046, China. [7]School of Biology and Environmental Science, Queensland University of Technology, Brisbane, QLD 4000, Australia. [8]College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China. [9]College of Marine Science, University of Chinese Academy of Sciences, Qingdao 266400, China. [10]South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, China. [11]Qingdao Key Laboratory of Marine Genomics, BGI-qingdao, Qingdao, China.

References
1. Seilacher A. Biomat-related lifestyles in the Precambrian. PALAIOS. 1999;14(1):86–93.
2. Fedonkin MA, Simonetta A, Ivantsov AY. New data on *Kimberella*, the Vendian mollusc-like organism (White Sea region, Russia): palaeoecological and evolutionary implications. J Geol Soc London Spec Publ. 2007;286(1):157–79.

Guo *et al. BMC Biology*   (2023) 21:51

Page 17 of 20

3. McFall-Ngai M, Hadfield MG, Bosch TC, Carey HV, Domazet-Loso T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci U S A. 2013;110(9):3229–36.
4. Maire J, Vincent-Monegat C, Balmand S, Vallier A, Herve M, Masson F, et al. Weevil *pgrp-lb* prevents endosymbiont TCT dissemination and chronic host systemic immune activation. Proc Natl Acad Sci U S A. 2019;116(12):5623–32.
5. Nyholm SV, McFall-Ngai MJ. A lasting symbiosis: how the Hawaiian bobtail squid finds and keeps its bioluminescent bacterial partner. Nat Rev Microbiol. 2021;19(10):666–79.
6. Bright M, Giere O. Microbial symbiosis in Annelida. Symbiosis. 2005;38:1–45.
7. Dubilier N, Bergin C, Lott C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nat Rev Microbiol. 2008;6(10):725–40.
8. Masson-Boivin C, Giraud E, Perret X, Batut J. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? Trends Microbiol. 2009;17(10):458–66.
9. McCutcheon JP, McDonald BR, Moran NA. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc Natl Acad Sci U S A. 2009;106(36):15394–9.
10. Kleiner M, Petersen JM, Dubilier N. Convergent and divergent evolution of metabolism in sulfur-oxidizing symbionts and the role of horizontal gene transfer. Curr Opin Microbiol. 2012;15(5):621–31.
11. Sørensen MES, Wood AJ, Minter EJA, Lowe CD, Cameron DD, Brockhurst MA. Comparison of independent evolutionary origins reveals both convergence and divergence in the metabolic mechanisms of symbiosis. Curr Biol. 2020;30(2):328–34 e4.
12. Baker MC, Ramirez-Llodra EZ, Tyler PA, German CR, Boetius A, Cordes EE, et al. Biogeography, ecology, and vulnerability of chemosynthetic ecosystems in the deep sea. In: McIntyre A, editor. Life in the World's Oceans: Diversity, Distribution, and Abundance: Wiley-Blackwell; 2010. p. 161–83.
13. Sogin EM, Kleiner M, Borowski C, Gruber-Vodicka HR, Dubilier N. Life in the dark: phylogenetic and physiological diversity of chemosynthetic symbioses. Annu Rev Microbiol. 2021;75:695–718.
14. Taylor JD, Glover EA. Chemosymbiotic bivalves. In: Kiel S, editor. The Vent and Seep Biota Topics in Geobiology. Dordrecht: Springer; 2010. p. 107–35.
15. Roeselers G, Newton IL. On the evolutionary ecology of symbioses between chemosynthetic bacteria and bivalves. Appl Microbiol Biotechnol. 2012;94(1):1–10.
16. Wein T, Romero Picazo D, Blow F, Woehle C, Jami E, Reusch TBH, et al. Currency, exchange, and inheritance in the evolution of symbiosis. Trends Microbiol. 2019;27(10):836–49.
17. Wentrup C, Wendeberg A, Schimak M, Borowski C, Dubilier N. Forever competent: deep-sea bivalves are colonized by their chemosynthetic symbionts throughout their lifetime. Environ Microbiol. 2014;16(12):3699–713.
18. Ponnudurai R, Kleiner M, Sayavedra L, Petersen JM, Moche M, Otto A, et al. Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. ISME J. 2017;11(2):463–77.
19. Fujiwara Y, Kato C, Masui N, Fujikura K, Kojima S. Dual symbiosis in the cold-seep thyasirid clam *Maorithyas hadalis* from the hadal zone in the Japan Trench, western Pacific. Mar Ecol Prog Ser. 2001;214:151–9.
20. Dufour SC. Gill anatomy and the evolution of symbiosis in the bivalve family Thyasiridae. Biol Bull. 2005;208(3):200–12.
21. Hongo Y, Ikuta T, Takaki Y, Shimamura S, Shigenobu S, Maruyama T, et al. Expression of genes involved in the uptake of inorganic carbon in the gill of a deep-sea vesicomyid clam harboring intracellular thioautotrophic bacteria. Gene. 2016;585(2):228–40.
22. Ip JC, Xu T, Sun J, Li R, Chen C, Lan Y, et al. Host-endosymbiont genome integration in a deep-sea chemosymbiotic clam. Mol Biol Evol. 2021;38(2):502–18.
23. Streams ME, Fisher CR, Fiala-Mdioni A. Methanotrophic symbiont location and fate of carbon incorporated from methane in a hydrocarbon seep mussel. Mar Biol. 1997;129(3):465–76.
24. Southward EC. Gill symbionts in Thyasirids and other bivalve molluscs. J Mar Biol Assoc UK. 1986;66(4):889–914.
25. Bright M, Bulgheresi S. A complex journey: transmission of microbial symbionts. Nat Rev Microbiol. 2010;8:218–30.
26. Sun J, Zhang Y, Xu T, Zhang Y, Mu H, Zhang Y, et al. Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. Nat Ecol Evol. 2017;1(5):121.
27. Laurich JR, Batstone RT, Dufour SC. Temporal variation in chemoautotrophic symbiont abundance in the thyasirid bivalve *Thyasira* cf. *gouldi*. Mar Biol. 2015;162(10):2017–28.
28. Dufour SC, Felbeck H. Symbiont abundance in thyasirids (Bivalvia) is related to particulate food and sulphide availability. Mar Ecol Prog Ser. 2006;320:185–94.
29. Batstone RT, Dufour SC. Closely related thyasirid bivalves associate with multiple symbiont phylotypes. Mar Ecol. 2016;37(5):988–97. https://doi.org/10.1111/maec.12310.
30. Oliver PG, Frey MA. *Ascetoaxinus quatsinoensis* sp. et gen. nov. (Bivalvia: Thyasiroidea) from Vancouver Island, with notes on *Conchocele* Gabb, 1866, and *Channelaxinus* Valentich-Scott & Coan, 2012. Zootaxa. 2014;3869(4):452–68.
31. Ozawa G, Shimamura S, Takaki Y, Yokobori SI, Ohara Y, Takishita K, et al. Updated mitochondrial phylogeny of Pteriomorph and Heterodont Bivalvia, including deep-sea chemosymbiotic *Bathymodiolus* mussels, vesicomyid clams and the thyasirid clam *Conchocele* cf. *bisecta*. Mar. Genomics. 2017;31:43–52.
32. Dando PR, Spiro B. Varying nutritional dependence of the thyasirid bivalves *Thyasira sarsi* and *T. equalis* on chemoautotrophic symbiotic bacteria, demonstrated by isotope ratios of tissue carbon and shell carbonate. Mar Ecol Prog Ser. 1993;92:151–8.
33. Goffredi SK, Barry JP, Buck KR. Vesicomyid symbioses from Monterey Bay (central California) cold seeps. Symbiosis. 2004;36:1–27.
34. Barry JP, Buck KR, Kochevar RK, Nelson DC, Fujiwara Y, Goffredi SK, et al. Methane-based symbiosis in a mussel, *Bathymodiolus platifrons*, from cold seeps in Sagami Bay, Japan. Invertebr Biol. 2005;121(1):47–54.
35. Durand P, Gros O. Bacterial host specificity of Lucinacea endosymbionts: Interspecific variation in 16S rRNA sequences. FEMS Microbiol Lett. 1996;140(2-3):193–8.
36. La Rivière M, Garrabou J, Bally M. Evidence for host specificity among dominant bacterial symbionts in temperate gorgonian corals. Coral Reefs. 2015;34(4):1087–98.
37. Kumar R, Kushwah J, Ganguly S, Garg V, Somvanshi VS. Proteomic Investigation of Photorhabdus Bacteria for Nematode-Host Specificity. Indian J Microbiol. 2016;56(3):361–7.
38. Lund MB, Davidson SK, Holmstrup M, James S, Kjeldsen KU, Stahl DA, et al. Diversity and host specificity of the Verminephrobacter–earthworm symbiosis. Environ Microbiol. 2010;12(8):2142–51.
39. Dufour SC, Laurich JR, Batstone RT, McCuaig B, Elliott A, Poduska KM. Magnetosome-containing bacteria living as symbionts of bivalves. ISME J. 2014;8(12):2453–62. https://doi.org/10.1038/ismej.2014.93.
40. McCuaig B, Pena-Castillo L, Dufour SC. Metagenomic analysis suggests broad metabolic potential in extracellular symbionts of the bivalve *Thyasira* cf. *gouldi*. Anim Microbiome. 2020;2(1):7.
41. Fichant G, Basse M-J, Quentin Y. ABCdb: an online resource for ABC transporter repertoires from sequenced archaeal and bacterial genomes. FEMS Microbiol Lett. 2006;256(2):333–9.
42. Elbourne LD, Tetu SG, Hassan KA, Paulsen IT. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. Nucleic Acids Res. 2017;45(D1):D320–D4.
43. Price GD, Woodger FJ, Badger MR, Howitt SM, Tucker L. Identification of a SulP-type bicarbonate transporter in marine cyanobacteria. Proc Natl Acad Sci U S A. 2004;101(52):18228–33.
44. Aguilar-Barajas E, Diaz-Perez C, Ramirez-Diaz MI, Riveros-Rosas H, Cervantes C. Bacterial transport of sulfate, molybdate, and related oxyanions. Biometals. 2011;24(4):687–707.
45. Parker AR, Moore TD, Edman JC, Schwab JM, Davisson VJ. Cloning, sequence analysis and expression of the gene encoding imidazole glycerol phosphate dehydratase in *Cryptococcus neoformans*. Gene. 1994;145(1):135–8.
46. Janmale TV, Lindsay A, Gieseg SP. Nucleoside transporters are critical to the uptake and antioxidant activity of 7,8-dihydroneopterin in monocytic cells. Free Radic Res. 2020;54(5):341–50.
47. Tran Van Nhieu G, Isberg RR. Bacterial internalization mediated by beta 1 chain integrins is determined by ligand affinity and receptor density. EMBO J. 1993;12(5):1887–95.

Guo *et al. BMC Biology*   (2023) 21:51

Page 18 of 20

48. Dupuy AG, Caron E. Integrin-dependent phagocytosis: spreading from microadhesion to new concepts. J Cell Sci. 2008;121(11):1773–83.

49. Kharlamenko VI, Kamenev GM, Kalachev AV, Kiyashko SI, Ivin VV. Thyasirid bivalves from the methane seep community off Paramushir Island (Sea of Okhotsk) and their nutrition. J Molluscan Stud. 2016;82(3):391–402.

50. Karapunar B, Werner W, Fürsich FT, Nützel A. Predatory drill holes in the oldest thyasirid bivalve, from the Lower Jurassic of South Germany. Lethaia. 2020;54(2):229–44. https://doi.org/10.1111/let.12399.

51. Taylor JD, Williams ST, Glover EA. Evolutionary relationships of the bivalve family Thyasiridae (Mollusca: Bivalvia), monophyly and superfamily status. J Mar Biol Assoc UK. 2007;87(2):565–74.

52. 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(5):120.

53. Shao F, Han M, Peng Z. Evolution and diversity of transposable elements in fish genomes. Sci Rep. 2019;9(1):15399.

54. Wang K, Wang J, Zhu C, Yang L, Ren Y, Ruan J, et al. African lungfish genome sheds light on the vertebrate water-to-land transition. Cell. 2021;184(5):1362–76 e18.

55. Zhou W, Liang G, Molloy PL, Jones PA. DNA methylation enables transposable element-driven genome expansion. Proc Natl Acad Sci U S A. 2020;117(32):19359–66.

56. Wells JN, Feschotte C. A field guide to eukaryotic transposable elements. Annu Rev Genet. 2020;54:539–61.

57. Nachman RJ, Giard W, Favrel P, Suresh T, Sreekumar S, Holman GM. Insect Myosuppressins and Sulfakinins stimulate release of the digestive enzyme ?-amylase in two invertebrates: the Scallop *Pecten maximus* and insect *Rhynchophorus ferrugineus*. Ann N Y Acad Sci. 1997;814:335–8.

58. Schwartz J, Dubos MP, Pasquier J, Zatylny-Gaudin C, Favrel P. Emergence of a cholecystokinin/sulfakinin signalling system in Lophotrochozoa. Sci Rep. 2018;8(1):16424.

59. Cao Y, Tian R, Shi S, Du X, Jiao Y. Characterization and expression analysis of tandemly duplicated nicotinic acetylcholine receptors in pearl oysters after stimulation of pathogen-related molecular patterns. Comp Biochem Physiol B Biochem Mol Biol. 2021;256:110615.

60. Du X, Tang Y, Han Y, Ri S, Kim T, Ju K, et al. Acetylcholine suppresses phagocytosis via binding to muscarinic- and nicotinic-acetylcholine receptors and subsequently interfering Ca(2+)- and NFκB-signaling pathways in blood clam. Fish Shellfish Immunol. 2020;102:152–60.

61. Lan Y, Sun J, Chen C, Sun Y, Zhou Y, Yang Y, et al. Hologenome analysis reveals dual symbiosis in the deep-sea hydrothermal vent snail *Gigantopelta aegis*. Nat Commun. 2021;12(1):1165.

62. Mariotta L, Ramadan T, Singer D, Guetg A, Herzog B, Stoeger C, et al. T-type amino acid transporter TAT1 (Slc16a10) is essential for extracellular aromatic amino acid homeostasis control. J Physiol. 2012;590(24):6413–24.

63. Zanzerl H, Salvo F, Jones SW, Dufour SC. Feeding strategies in symbiotic and asymbiotic thyasirid bivalves. J Sea Res. 2019;145:16–23.

64. Sun Y, Sun J, Yang Y, Lan Y, Ip JC, Wong WC, et al. Genomic signatures supporting the symbiosis and formation of chitinous tube in the deep-sea tubeworm *Paraescarpia echinospica*. Mol Biol Evol. 2021;38(10):4116–34.

65. Gerdol M, Fujii Y, Hasan I, Koike T, Shimojo S, Spazzali F, et al. The purplish bifurcate mussel *Mytilisepta virgata* gene expression atlas reveals a remarkable tissue functional specialization. BMC Genomics. 2017;18(1):590.

66. de Oliveira AL, Mitchell J, Girguis P, Bright M. Novel insights on obligate symbiont lifestyle and adaptation to chemosynthetic environment as revealed by the giant tubeworm genome. Mol Biol Evol. 2022;39(1):msab347. https://doi.org/10.1093/molbev/msab347.

67. Zheng P, Wang M, Li C, Sun X, Wang X, Sun Y, et al. Insights into deep-sea adaptations and host-symbiont interactions: a comparative transcriptome study on *Bathymodiolus* mussels and their coastal relatives. Mol Ecol. 2017;26(19):5133–48.

68. Lan Y, Sun J, Zhang W, Xu T, Zhang Y, Chen C, et al. Host–symbiont interactions in deep-sea chemosymbiotic vesicomyid clams: insights from transcriptome sequencing. Front Mar Sci. 2019;6:680. https://doi.org/10.3389/fmars.2019.00680.

69. Whyte JRC, Munro S. A yeast homolog of the mammalian mannose 6-phosphate receptors contributes to the sorting of vacuolar hydrolases. Curr Biol. 2001;11(13):1074–8.

70. Piper RC, Luzio JP. CUPpling calcium to lysosomal biogenesis. Trends Cell Biol. 2004;14(9):471–3.

71. Treusch S, Knuth S, Slaugenhaupt SA, Goldin E, Grant BD, Fares H. Caenorhabditis elegans functional orthologue of human protein h-mucolipin-1 is required for lysosome biogenesis. Proc Natl Acad Sci U S A. 2004;101(13):4483–8.

72. Luzio JP, Pryor PR, Bright NA. Lysosomes: fusion and function. Nat Rev Mol Cell Biol. 2007;8(8):622–32.

73. Castellano F, Montcourrier P, Chavrier P. Membrane recruitment of Rac1 triggers phagocytosis. J Cell Sci. 2000;113(17):2955–61.

74. Kinchen JM, Ravichandran KS. Phagosome maturation: going through the acid test. Nat Rev Mol Cell Biol. 2008;9(10):781–95.

75. Li Y, Tassia MG, Waits DS, Bogantes VE, David KT, Halanych KM. Genomic adaptations to chemosymbiosis in the deep-sea seep-dwelling tubeworm *Lamellibrachia luymesi*. BMC Biol. 2019;17(1):91.

76. Batstone RT, Laurich JR, Salvo F, Dufour SC. Divergent chemosymbiosis-related characters in *Thyasira* cf. *gouldi* (Bivalvia: Thyasiridae). PLoS One. 2014;9(3):e92856.

77. Wong CO, Li R, Montell C, Venkatachalam K. *Drosophila* TRPML is required for TORC1 activation. Curr Biol. 2012;22(17):1616–21.

78. Miao Y, Li G, Zhang X, Xu H, Abraham SN. A TRP channel senses lysosome neutralization by pathogens to trigger their expulsion. Cell. 2015;161(6):1306–19.

79. Klinkert K, Echard A. Rab35 GTPase: A central regulator of phosphoinositides and F-actin in endocytic recycling and beyond. Traffic. 2016;17(10):1063–77.

80. Tateda K, Matsumoto T, Miyazaki S, Yamaguchi K. Lipopolysaccharide-induced lethality and cytokine production in aged mice. Infect Immun. 1996;64(3):769–74.

81. Yu Y, Qiu L, Song L, Zhao J, Ni D, Zhang Y, et al. Molecular cloning and characterization of a putative lipopolysaccharide-induced TNF-alpha factor (LITAF) gene homologue from Zhikong scallop *Chlamys farreri*. Fish Shellfish Immunol. 2007;23(2):419–29.

82. Ağar Ç, de Groot PG, Mörgelin M, Monk SDDC, van Os G, Levels JHM, et al. beta(2)-glycoprotein I: a novel component of innate immunity. Blood. 2011;117(25):6939–47.

83. Needham BD, Trent MS. Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. Nat Rev Microbiol. 2013;11(7):467–81.

84. Otten EG, Werner E, Crespillo-Casado A, Boyle KB, Dharamdasani V, Pathe C, et al. Ubiquitylation of lipopolysaccharide by RNF213 during bacterial infection. Nature. 2021;594(7861):111–6.

85. Moran NA. Symbiosis as an adaptive process and source of phenotypic complexity. Proc Natl Acad Sci U S A. 2007;104(Suppl 1):8627–33.

86. Lynch M. Genomics. Gene duplication and evolution. Science. 2002;297(5583):945–7.

87. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290(5494):1151–5.

88. Halary S, Riou V, Gaill F, Boudier T, Duperron S. 3D FISH for the quantification of methane- and sulphur-oxidizing endosymbionts in bacteriocytes of the hydrothermal vent mussel *Bathymodiolus azoricus*. ISME J. 2008;2(3):284–92.

89. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017;45(4):e18.

90. Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, et al. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Res. 2019;47(20):10543–52.

91. Minkin I, Pham H, Starostina E, Vyahhi N, Pham S. C-Sibelia: an easy-to-use and highly accurate tool for bacterial genome comparison. F1000Res. 2013;2:258.

92. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–45.

93. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control

Guo *et al. BMC Biology*   (2023) 21:51

Page 19 of 20

and preprocessing of high-throughput sequencing data. GigaScience. 2018;7(1):1–6. https://doi.org/10.1093/gigascience/gix120.

94. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27(6):764–70.

95. Shafin K, Pesout T, Lorig-Roach R, Haukness M, Olsen HE, Bosworth C, et al. Nanopore sequencing and the Shasta toolkit enable efficient *de novo* assembly of eleven human genomes. Nat Biotechnol. 2020;38(9):1044–53.

96. 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(11):e112963.

97. Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. Methods Mol Biol. 2019;1962:227–45.

98. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 2015;16(1):1–11.

99. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016;3(1):95–8.

100. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 2016;3(1):99–101.

101. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.

102. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007;35:W265–8.

103. Smit AF, Hubley R. RepeatModeler Open-1.0; 2008.

104. Smit AF. Repeat-Masker Open-3.0. 2004. http://www.repeatmasker.org.

105. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 2005;110(1-4):462–7.

106. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUS-TUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res. 2006;34(suppl_2):W435–9.

107. Ye J, McGinnis S, Madden TL. BLAST: improvements for better sequence analysis. Nucleic Acids Res. 2006;34(suppl_2):W6–9.

108. She R, Chu JS-C, Wang K, Pei J, Chen N. GenBlastA: enabling BLAST to identify homologous gene sequences. Genome Res. 2009;19(1):143–9.

109. Birney E, Clamp M, Durbin R. GeneWise and genomewise. Genome Res. 2004;14(5):988–95.

110. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3):290–5.

111. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907–15.

112. 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(1):1–22.

113. 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(7433):526–31.

114. Liu X, Li C, Chen M, Liu B, Yan X, Ning J, et al. Draft genomes of two Atlantic bay scallop subspecies *Argopecten irradians irradians* and *A. i. concentricus*. Sci Data. 2020;7(1):99.

115. 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(7418):49–54.

116. Gómez-Chiarri M, Warren WC, Guo X, Proestou D. Developing tools for the study of molluscan immunity: the sequencing of the genome of the eastern oyster, *Crassostrea virginica*. Fish Shellfish Immunol. 2015;46(1):2–4.

117. Wei M, Ge H, Shao C, Yan X, Nie H, Duan H, et al. Chromosome-level clam genome helps elucidate the molecular basis of adaptation to a buried lifestyle. iScience. 2020;23(6):101148.

118. Sun Y, Liu X, Xie Y, Bai Y, Wang S, Teng W, et al. A high-quality chromosome-level genome assembly of the bivalve mollusk *Mactra veneriformis*. G3 (Bethesda). 2022;12(11):jkac229. https://doi.org/10.1093/g3journal/jkac229.

119. Song H, Guo X, Sun L, Wang Q, Han F, Wang H, et al. The hard clam genome reveals massive expansion and diversification of inhibitors of apoptosis in Bivalvia. BMC Biol. 2021;19(1):15.

120. Du X, Fan G, Jiao Y, Zhang H, Guo X, Huang R, et al. The pearl oyster *Pinctada fucata martensii* genome and multi-omic analyses provide insights into biomineralization. GigaScience. 2017;6(8):1–12. https://doi.org/10.1093/gigascience/gix059.

121. Thai BT, Lee YP, Gan HM, Austin CM, Croft LJ, Trieu TA, et al. Whole genome assembly of the snout otter clam, *Lutraria rhynchaena*, using Nanopore and Illumina Data, benchmarked against bivalve genome assemblies. Front Genet. 2019;10:1158.

122. Li R, Zhang W, Lu J, Zhang Z, Mu C, Song W, et al. The whole-genome sequencing and hybrid assembly of *Mytilus coruscus*. Front Genet. 2020;11:440.

123. 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(5):1–13. https://doi.org/10.1093/gigascience/giaa037.

124. Yan X, Nie H, Huo Z, Ding J, Li Z, Yan L, et al. Clam genome sequence clarifies the molecular basis of its benthic adaptation and extraordinary shell color diversity. iScience. 2019;19:1225–37.

125. Powell D, Subramanian S, Suwansa-Ard S, Zhao M, O'Connor W, Raftos D, et al. The genome of the oyster *Saccostrea* offers insight into the environmental resilience of bivalves. DNA Res. 2018;25(6):655–65.

126. Ran Z, Li Z, Yan X, Liao K, Kong F, Zhang L, et al. Chromosome-level genome assembly of the razor clam *Sinonovacula constricta* (Lamarck, 1818). Mol Ecol Resour. 2019;19(6):1647–58.

127. Sun J, Chen C, Miyamoto N, Li R, Sigwart JD, Xu T, et al. The Scaly-foot Snail genome and implications for the origins of biomineralised armour. Nat Commun. 2020;11(1):1657.

128. Masonbrink RE, Purcell CM, Boles SE, Whitehead A, Hyde JR, Seetharam AS, et al. An Annotated Genome for *Haliotis rufescens* (Red Abalone) and Resequenced Green, Pink, Pinto, Black, and White Abalone Species. Genome Biol Evol. 2019;11(2):431–8.

129. Liu C, Zhang Y, Ren Y, Wang H, Li S, Jiang F, et al. The genome of the golden apple snail *Pomacea canaliculata* provides insight into stress tolerance and invasive adaptation. Gigascience. 2018;7(9):1–13. https://doi.org/10.1093/gigascience/giy101.

130. 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(7564):220–4.

131. Zarrella I, Herten K, Maes GE, Tai S, Yang M, Seuntjens E, et al. The survey and reference assisted assembly of the *Octopus vulgaris* genome. Sci Data. 2019;6(1):13.

132. Bai CM, Xin LS, Rosani U, Wu B, Wang QC, Duan XK, et al. Chromosomal-level assembly of the blood clam, *Scapharca* (*Anadara*) *broughtonii*, using long sequence reads and Hi-C. Gigascience. 2019;8(7)1–8. https://doi.org/10.1093/gigascience/giz067.

133. Cosentino S, Iwasaki W. SonicParanoid: fast, accurate and easy orthology inference. Bioinformatics. 2019;35(1):149–51.

134. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.

135. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530–4.

136. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.

137. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.

138. Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):e9490.

139. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3.

140. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980;16(2):111–20.

141. Sahraeian SME, Mohiyuddin M, Sebra R, Tilgner H, Afshar PT, Au KF, et al. Gaining comprehensive biological insight into the

Guo *et al. BMC Biology*   (2023) 21:51

Page 20 of 20

transcriptome by performing a broad-spectrum RNA-seq analysis. Nat Commun. 2017;8(1):1–15.

142. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14(4):417–9.

143. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31(10):1674–6.

144. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome. 2018;6(1):158.

145. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595.

146. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.

147. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2019;36(6):3.

148. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043–55.

149. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9.

150. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955–64.

151. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35(9):3100–8.

152. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2012;14(2):178–92.

153. Guo Y, Meng L, Wang M, Zhong Z, Li D, Zhang Y, et al. Hologenome analysis reveals independent evolution to chemosymbiosis by deep-sea bivalves. 2023. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA 913320; https://db.cngb.org/search/project/CNP0003505.

154. Guo Y, Meng L, Wang M, Zhong Z, Li D, Zhang Y, et al. Hologenome analysis reveals independent evolution to chemosymbiosis by deep-sea bivalves. 2023. https://www.ebi.ac.uk/pride/archive/projects/PXD03 6936.

155. Guo Y, Meng L, Wang M, Zhong Z, Li D, Zhang Y, et al. Hologenome analysis reveals independent evolution to chemosymbiosis by deep-sea bivalves. 2023. https://doi.org/10.6084/m9.figshare.21820002.

156. Guo Y, Meng L, Wang M, Zhong Z, Li D, Zhang Y, et al. Hologenome analysis reveals independent evolution to chemosymbiosis by deep-sea bivalves. 2023. https://github.com/GuoYang-qd/Conbis_genome.

Publisher's Note