Clam Genome Sequence Clarifies the Molecular Basis of Its Benthic Adaptation and Extraordinary Shell Color Diversity

HIGHLIGHTS

- We present a new genome assembly of the Manila clam Ruditapes philippinarum.
- Analysis of gene family expansions and transcriptome characterization were conducted.
- Tyr and mitf genes were potentially involved in shell color patterns of Manila clam.
- Expansion of GPCRs and immune-related genes were found in R. philippinarum.

Yan et al., iScience 19, 1225–1237 September 27, 2019 © 2019 The Author(s).
https://doi.org/10.1016/j.isci.2019.08.049
Clam Genome Sequence Clarifies the Molecular Basis of Its Benthic Adaptation and Extraordinary Shell Color Diversity

Xiwu Yan,1,3,4,* Hongtao Nie,1,3,* Zhongming Huo,1,3 Jianfeng Ding,1,3 Zhenzhen Li,2,3 Lulu Yan,1 Liwen Jiang,1 Zhengqiang Mu,1 Huamin Wang,1 Xiangyu Meng,1 Peng Chen,1 Mengyan Zhou,2 Md. Golam Rbbani,1 Guangjian Liu,2,* and Dongdong Li1

SUMMARY

*Ruditapes philippinarum* is an economically important bivalve with remarkable diversity in its shell coloration patterns. In this study, we sequenced the whole genome of the Manila clam and investigated the molecular basis of its adaptation to hypoxia, acidification, and parasite stress with transcriptome sequencing and an RNA sequence analysis of different tissues and developmental stages to clarify these major issues. A number of immune-related gene families are expanded in the *R. philippinarum* genome, such as TEP, C3, C1qDC, Hsp70, SABL, and lysozyme, which are potentially important for its stress resistance and adaptation to a coastal benthic life. The transcriptome analyses demonstrated the dynamic and orchestrated specific expression of numerous innate immune-related genes in response to experimental challenge with pathogens. These findings suggest that the expansion of immune- and stress-related genes may play vital roles in resistance to adverse environments and has a profound effect on the clam’s adaptation to benthic life.

INTRODUCTION

Bivalves are represented by approximately 20,000 species, distributed throughout aquatic habitats (Daniel, 2013). The Manila clam, *Ruditapes philippinarum*, is a bivalve mollusk with a worldwide distribution and important commercial, culinary, and ecological value. The production of *R. philippinarum* reached over 4.0 million tons, equivalent to 3.7 million USD, in 2015 (FAO, 2017). As a filter-feeding animal living buried in sediment, *R. philippinarum* is an important “sentinel” species used to assess the quality of the marine environment (Milan et al., 2011). Benthic bivalves play important roles in food chains, including as food for humans, and most benthic bivalves are critical in bioturbation, bioirrigation, and the breakdown of organic matter (Norkko and Shumway, 2011). However, massive mortality in the Manila clam has been reported arising from parasitic infections, which has severely affected the Manila clam industry (Choi et al., 2002; Pretto et al., 2014). Moreover, ocean acidification, increased water temperatures, and hypoxic events are increasing worldwide problems (Steckbauer et al., 2015). The molecular mechanisms underlying the tolerance of benthic organisms for diverse biotic and abiotic factors, such as parasites, ocean acidification, and hypoxia, have received much attention in recent years (Lim et al., 2006; Shimokawa et al., 2010; Zhao et al., 2018; Kodama et al., 2018).

Bivalve mollusks have a biphasic life cycle, in which planktonic larvae (trochophores and veligers) settle (as pediveliger larvae) and undergo metamorphosis (to the juvenile stage). The transition of benthic bivalves from a pelagic to a benthic lifestyle during their early development, their burrowing to avoid predation, and their tolerance of hypoxic environments have been attractive characteristics for research (Tamai, 1993; Kodama et al., 2018). However, knowledge of the underlying molecular mechanisms that regulate these processes is still very limited (Findlay and Battin, 2016). The Manila clam is an excellent marine model system, with a robust and multifaceted immune system (Dyachuk, 2016). The immune system of bivalves relies on the recognition of conserved pathogen-associated molecular patterns shared by broad classes of microorganisms, which allows the bivalves to successfully defend themselves against infection (Gestal et al., 2008; Ki et al., 2006; Song et al., 2010). Sediment is a specific habitat of bivalves, with characteristic microbial diversity, low input of sunlight, and a low concentration of oxygen (Wang et al., 2012). Benthic bivalves live in sediment and play important roles in natural biochemical cycles and in the material exchanges between water bodies and sediments (Vaughn and Hakenkamp, 2001). How benthic bivalves adapt to and survive in such hypoxic and pathogen-rich environments is of fundamental interest.
Despite its burrowing lifestyle, *R. philippinarum* displays a large range of shell colors and patterns (Protas and Patel, 2008; Zhang and Yan, 2010). The Manila clam has several different shell colors, including white, zebra, and orange in its natural habitat (Figure S1), and it has been selected over several generations for high stress resistance and distinctive shell color with selective breeding programs (Zhang and Yan, 2010). Previous experimental crosses have shown that shell color has a genetic basis in the Manila clam (Peignon et al., 1995). Color is one of the most conspicuous phenotypic traits in *R. philippinarum*, and the diverse colors of molluscan shells are generally believed to be determined by the presence of biological pigments (Mann and Jackson, 2014; Lemer et al., 2015; Feng et al., 2018). Several studies have been conducted of the physiological, biochemical, and molecular responses to different types of biotic and abiotic stress in different shell color strains of the Manila clam (Zhang and Yan, 2010; Nie et al., 2017a, 2017b). In recent years, RNA sequencing (RNA-seq) has been used to identify the genes differentially expressed in the differently colored shells of several bivalves (Sun et al., 2015; Feng et al., 2015; Yue et al., 2015). However, the genes potentially responsible for shell color formation and variation in strains of *R. philippinarum* are still largely unknown. The expression of genes for tyrosinase (*Tyr*) and melanogenesis-associated transcription factor (*Mitf*) is responsible for the production of melanin from tyrosine, promoting melanin production and stimulating pigmentation (Krumholz et al., 2011). In mollusks, tyrosinase is secreted and appears to contribute to shell pigmentation and melanin synthesis, which can also physically encapsulate pathogens, and is therefore, an important component of the immune system (Cerenius et al., 2008).

The main goals of this study were to provide a robust reference genome for the Manila clam and to analyze its transcriptomic responses to various stimuli and different ontogenetic stages. Here, we report the whole genome sequence of the Manila clam and its *de novo* assembly and a comparative analysis with a Korean strain of the Manila clam (Mun et al., 2017). We also examined the expression of pigmentation-related genes (*Tyr* and *Mitf*), immune pathways, and defense mechanisms in the Manila clam. To better understand these biological issues, RNA-seq analyses of hypoxia-challenged, acidification-exposed, and parasite-infected clams were performed, as were RNA-seq analyses of different shell color strains and developmental stages of the Manila clam. This study provides clues to the molecular mechanisms underlying the shell pigmentation, immune defenses, and resistance to different stresses in this clam. Comparative genomic analyses of gene expansion, contraction, and positive selection in *R. philippinarum* and *Lottia gigantea* (Simakov et al., 2013), *Crassostrea gigas* (Zhang et al., 2012), *Pinctada fucata* (Takeuchi et al., 2012), and *Patinoppecten yessoensis* (Wang et al., 2017) also identify important genetic differences among these species and provide insights into the molecular basis of adaptation to a benthic lifestyle in clams.

**RESULTS**

**Genome Characterization**

The genome of *R. philippinarum* is shown in Figure 1A. The Manila clam genome has a high level of heterozygosity and an estimated size of 1.32 Gb (Table S1 and Figure S2). To reduce the heterozygosity, genomic DNA isolated from an inbred individual (six generations, 2.0% estimated heterozygosity) (Figure S1D) was used for whole-genome shotgun sequencing (286.1 Gb of high-quality paired-end data and 135.1 Gb of mate-pair data; Table S2 and Figure S3). The *R. philippinarum* assembly comprised 30,811 scaffolds with an N50 length of 345 kb and a total length of 1.12 Gb (Table S3). The heterozygosity of the assembly was reduced to 1.038%, compared with that of a wild-type individual (1.69%) (Table S4). Evaluation of the genome showed the completeness of the assembly (Tables S5–S7). A total of 233 CEGMA (Core Eukaryotic Genes Mapping Approach) core genes with a ratio of completeness of 93.95% (Table S8), together with 91.0% complete and 3.9% fragmented Metazoa BUSCO (Benchmarking Universal Single-Copy Orthologs) orthologs (Table S9), were identified in the assembled genome, indicating the high degree of completeness of the gene regions. Repetitive regions comprised ~38.30% of the *R. philippinarum* assembly (Table S10, and Figure S4), and combined transposable elements were predicted to constitute 35.8% of the *R. philippinarum* genome (Table S11). Transcriptome-based, *ab initio*, and homology-based predictions identified 27,652 protein-coding genes (Table S12), a similar number to other Lophotrochozoa genomes (Table S13 and Figure S5). Of these, 24,570 (88.85%) were annotated based on various public databases (Table S14), with the best BLASTp hits in the SwissProt, TrEMBL, and NCBI nonredundant (NR) protein databases. Gene domain annotations were made by searching the InterPro database. All genes were functionally annotated by their alignment against Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) terms. Noncoding RNA was also annotated (Table S15).
Comparative Analysis of R. philippinarum (Rph) and Korean Clam (Hrph) Genomes

We compared the genomic statistics for Rph and Hrph (Table 1). The contig N50 for Rph was 28,111 bp, whereas that for Hrph was 6,520 bp. The complete BUSCO (C) ratio of the Rph genome was 92.2%, which was much higher than that of the Hrph genome (69.5%). The number of protein-coding genes in Rph was 27,652, whereas that in Hrph was 108,034, and the complete BUSCO (C) ratios of the genes were 91.0% and 81.5%, respectively (Table 1). There was about 67% collinearity between the Rph and Hrph genomes.

To analyze the gene families, the genes of Hrph were filtered from 108,034 to 19,776. A gene family cluster analysis was performed between Rph, Hrph, and six other species, including C. gigas, L. gigantea, P. fucata, Patinopecten yessoensis, Octopus bimaculoides, and Lingula anatina, and 436 orthologous gene families were shared among these species. Rph shared 1,266 orthologous gene families with the other species, excluding Hrph. Of these 1,266 genes, 1,240 were not annotated and 26 genes were annotated in more than one copy in the Hrph genome, and 1,246 were annotated in the collinear regions of the Rph and Hrph genomes. Hrph shared 56 orthologous gene families with other species, excluding Rph. Of the 56 genes, 16 genes were not annotated and 40 genes were annotated in more than one copy in the Rph genome, and 48 genes were annotated in the collinear regions of the Rph and Hrph genomes. Compared with Hrph, 7 expanded and 78 contracted gene families were detected in Rph with Fisher’s exact test (Table S16). A GO enrichment analysis of the expanded gene families in Rph, relative to Hrph, revealed that active transmembrane transporter activity, neurotransmitter transport, hydrolase activity, and scavenger receptor activity are involved in their immune functions and ecological adaptation (Table S16).

Comparative Genome Analysis of the Gene Families

The results of a gene family cluster analysis of R. philippinarum and other 13 species are shown in Figure S6. Compared with four other Mollusca species (C. gigas, L. gigantea, P. fucata, and Patinopecten yessoensis), R. philippinarum has 1,582 unique gene families and shares 6,975 gene families with other species (Figure S7, Tables S17 and S18). A phylogenetic tree was constructed using 288 shared single-copy orthologues of 14 species (Figure 1B). As the closest bivalve clade to the class Gastropoda (L. gigantea), R. philippinarum was estimated to have split from it about 476.9 million years ago (Figure 1B). Compared with the four other Mollusca species, 2,197 expanded and 3,943 contracted gene families were identified in R. philippinarum with Fisher’s exact test (Tables S19 and S20–S22), and 29 positively selected genes were detected in the R. philippinarum genome as the foreground branches and genes of four other Mollusca species as background branches (Tables S23 and S24). GO and KEGG enrichment analyses of the expanded genes revealed that they are involved a number of immune-related pathways in
R. philippinarum, including in the defense response to scavenger receptor activity, cytokine activity, apoptotic process, G-protein-coupled receptor activity, sulfotransferase activity, tumor necrosis factor (TNF) receptor binding, neuropeptide signaling pathway (Table S25), complement and coagulation cascades, phagosome, Toll-like receptor signaling pathway, NF-κB signaling pathway, TNF signaling pathway, and PI3K–AKT signaling pathway (Table S26).

Differential Expression of Tyr and Mitf Genes in Manila Clam

Manila clams are capable of producing an extraordinary diversity of shell color patterns. We examined the molecular processes of the melanogenesis and tyrosine metabolism pathways, which are involved in pigment synthesis. Tyrosinase (encoded by Tyr) is a crucial enzyme in the process of melanin synthesis, and the Tyr genes are regulated by a melanogenesis-associated transcription factor (encoded by Mitf). In this study, transcriptomic and gene expression analyses of the Tyr and Mitf genes in four shell color strains of R. philippinarum were conducted with RNA-seq and quantitative PCR (qPCR) (Figure 2). In total, 21 Tyr genes and 12 homologues of the Mitf gene were identified in the R. philippinarum genome. Most of the Mitf mRNAs were upregulated in the orange strain (O), and three of them were uniquely and strongly expressed in O compared with the other three color strains (Figure 2A). To confirm the differential expression of Tyr and Mitf in the different shell color strains, six Tyr and seven Mitf genes were further evaluated with RT-qPCR in four different shell color strains of R. philippinarum, and all of them were differentially expressed in the four shell color strains (Figure 2). Seven Mitf genes were upregulated in the mantle of the O strain, compared with the zebra strain, and three Mitf genes (Mitf1, Mitf2, and Mitf5) were upregulated in the mantle of the white zebra (WZ) strain (Figure 2A). An RNA-seq analysis was conducted in the different developmental stages of R. philippinarum, including the trochophore, D-shaped larval, umbo-veliger larval, spat, early juvenile before shell pigmentation, and late juvenile after shell pigmentation.

Table 1. Comparative Analysis between the Current Genome of Manila Clam and the Genome of Korean Manila Clam

| Type                                      | Rph          | Hrph          |
|-------------------------------------------|--------------|---------------|
| Estimated genome size (Gb)                | 1.32         | 1.37          |
| Assembly total length (bp)                | 1,122,973,377| 1,078,771,101|
| Sequencing depth                          | 320.63       | 74.2          |
| Contig N50 (bp)                           | 28,111       | 6,520         |
| Scaffold N50 (bp)                         | 56,467,786   | 119,518       |
| Average contig length (bp)                | 12,041       | 3,186         |
| Max contig length (bp)                    | 249,659      | 80,451        |
| Average scaffold length (bp)              | 56,456,578   | 80,439        |
| Max scaffold length (bp)                  | 204,629,219  | 1,050,406     |
| Number of contig > 2 Kb                   | 61,395       | 121,896       |
| Number of scaffolds > 2 Kb                | 19           | 13,318        |
| Genome BUSCO assessment                   | C:92.2%(S:90.3%,D:1.9%),F:1.6%,M:6.2%,n:978 | C:69.5%(S:66.6%,D:2.9%),F:11.2%,M:19.3%,n:978 |
| Repeat percentage                         | 38.29        | 26.38         |
| Gene models number                        | 27,652       | 108,034       |
| Average gene length (bp)                  | 12,875       | 5,117         |
| Average number of exons per gene          | 7.30         | 4.17          |
| Average exon length (bp)                  | 200.30       | 232           |
| Average number of introns per gene        | 6.29         | 3.17          |
| Average intron length (bp)                | 1,697.30     | 1,230         |
| Gene BUSCO assessment                     | C:91.0%(S:89.3%,D:1.7%),F:3.9%,M:5.1%,n:978 | C:81.5%(S:58.9%,D:22.6%),F:13.2%,M:5.3%,n:978 |

R. philippinarum, including in the defense response to scavenger receptor activity, cytokine activity, apoptotic process, G-protein-coupled receptor activity, sulfotransferase activity, tumor necrosis factor (TNF) receptor binding, neuropeptide signaling pathway (Table S25), complement and coagulation cascades, phagosome, Toll-like receptor signaling pathway, NF-κB signaling pathway, TNF signaling pathway, and PI3K–AKT signaling pathway (Table S26).
pigmentation stages. The expression of four Mitf genes was upregulated between the early juvenile stage before shell pigmentation and the juvenile stage after shell pigmentation (Figure 2A). Interestingly, the Tyr gene family showed a distinctive expression pattern and stage-specific expression in pelagic larvae (trochophore, D-shaped larva, and umbo-veliger larva), benthic juveniles (early juvenile and late juvenile), and adult clams (Figure 2B).

**Figure 2. The Expression of Key Genes Impacting Shell Color Patterns in *R. philippinarum* and Evolutionary Analysis of Tyr Genes**

(A) RNA-seq hierarchical clustering of Mitf genes at four different developmental stages, at noncolored and colored developmental stages in two different shell color strains, and in four different shell color strains, as well as RT-qPCR expression of seven Mitf genes in four different shell color strains. 

(B) RNA-seq hierarchical clustering of Tyr genes at four different developmental stages, at noncolored and colored developmental stages in two different shell color strains, and in four different shell color strains, as well as RT-qPCR expression of six Tyr genes in four different shell color strains. The abbreviations include trochophora (A11), D-shaped larvae (A12), umbo larvae (A23), pediveliger larvae (z3), uncolored zebra clam (Za), colored zebra clam (Zb), uncolored orange clam (Oa), colored orange clam (Ob), Orange strain (O), White strain (W), White Zebra strain (WZ), and Zebra strain (Z).

**Expansion of Immune-Related Genes in *R. philippinarum***

Compared with four other Mollusca species (*C. gigas, P. fucata, L. gigantea*, and *Patarinopecten yessoensis*), a number of immune-related genes are expanded in *R. philippinarum*, including the thioester-containing...
proteins (TEP), complement component C3 (C3), C1q domain-containing proteins (C1qDC), lysozymes, sialic acid-binding lectin (SABL), and heat shock protein 70 (Hsp70) (Figure 3). These gene families expanded in *R. philippinarum* are enriched in complement and coagulation cascades, phagosome, and lytic pathways (Figure 3). Pathogen recognition in *R. philippinarum* can be attributed to pathogen recognition receptors C1qDCs, Bf, and carbohydrate-binding lectins, such as SABLs and mannann-binding lectin, through the complement pathways (Figure 3A).

*Ruditapes philippinarum* has the largest number of genes in the C3 and CD109/iTEP subgroups (7 and 17 members, respectively), compared with *C. gigas*, *L. gigantea*, *P. fucata*, and *P. yessoensis* (Table S27), suggesting species-specific gene family expansion in *R. philippinarum*. Phylogenetic analyses also indicated extensive expansion of the Hsp70 genes from the Hspa12 subfamily in *R. philippinarum* and also in the bivalve lineages (Figure 3C). The expansion of immune-related genes (SABL, C3, C1qDC, lysozyme, and Hsp70) provides evidence of an association between gene duplication and the organism’s adaptation to pathogen-rich environments (Figure 4). The characteristics and expression of five complement C3 genes (xfsc0000076.27, xfsc0000154.2, xfsc0000773.3, xfsc0000773.2, and xfsc0001170.6) were investigated in the blood and digestive gland of *R. philippinarum* at 3, 6, 12, 24, 48, 72, and 96 h after treatment with lipopolysaccharide (LPS), peptidoglycan (PGN), or polyinosinic–polycytidylic acid (poly[I:C]). All these complement C3 genes were significantly upregulated in the immune responses of *R. philippinarum* to different pathogenic stimuli (LPS, PGN, and poly[I:C]) (Figure S8), indicating that the complement C3 genes play critical roles in the immunity of the Manila clam.

**RNA-Seq Analyses at Different Ontogenic Stages**

To identify the key genes involved in the development and metamorphosis processes, RNA-seq analyses were conducted in the different ontogenic stages of *R. philippinarum*. The Hox gene clusters and their expression at different ontogenic stages are shown in Figure S5. The expression of 11 Hox genes was detected in 17 different stages, and the Post1 and Lox5 genes were expressed in the gastrula stage (Figure 5B). Most Hox genes were strongly expressed in the early trochophore stage, when the larvae begin to secrete the shell to coat and protect the embryo. The Hox1, Post2, Hox3, and Lox4 genes were mainly expressed in trochophore and veliger larvae (Figure S9). In contrast, the Hox2 and Hox4 genes were strongly expressed in the umbo larvae (Figure 5B). The Post2 and Lox4 genes were strongly expressed in the trochophore, veliger, and umbo larvae. The Antp and Lox2 genes were strongly expressed in the umbo larvae, spats, and juveniles, whereas Lox5 was strongly expressed in the gastrula stage (Figure S5B). The expression of immune-related genes TEP and C3 was negligible throughout the different planktonic
tissues tested, including the gill, mantle, and gonad (Figure S14). Overall, the RNA-seq analysis revealed that the expanded qPCR (Figure S14). In the infected clams, the expression of complement and 44 downregulated genes (Table S29). Eleven candidate genes associated with immune, antioxidant, seq analysis of clams infected with digenetic trematodes identified 460 DEGs, including 416 upregulated apoptosis (regulated) were selected from the DEGs in several canonical pathways of interest and analyzed with To confirm the DEGs identified in clams under acidic conditions, 11 genes (six upregulated and five downregulated) were selected from the DEGs in several canonical pathways of interest and analyzed with RT-qPCR (Figure S13). A number of vital complement- and immune-responsive genes were significantly upregulated in the mantle and gonad. The expression of inhibitor of Ikappa, melatonin receptor type 1A (IxSc0000104.14) was upregulated (Figure S12), whereas the expression of genes encoding TNF, NFX1, GTPase IMAP (IAP), and glutathione S-transferase (GST) were downregulated when analyzed with RT-qPCR after exposure to hypoxia for 0, 2, 5, or 8 days (Figure S12). These genes are implicated in different physiological pathways, including respiration and energy metabolism (carbonic anhydrase, SLC2A1) and antioxidant and immune defense (CFB, C-type lectins, Hsp70, lysozyme, Tyr, TNF, ANKK1, IxB, GPx).

Transcriptomic Responses to Hypoxia, Acidification, and Parasites
We hypothesized that the expansion of the complement and immune-related genes plays a critical role in the resistance of the Manila clam to different stresses and in its benthic adaptation. The transcriptomic responses of the Manila clam to hypoxia, acidification, and parasites were analyzed to identify the complement- and immune-responsive genes. An RNA-seq analysis identified 102 differentially expressed genes (DEGs), including 75 upregulated genes and 27 downregulated genes, corresponding to 12 major physiological functions in the gills of R. philippinarum after exposure to hypoxia for 2 days (Table S28). The expression of hypoxia-responsive genes encoding Hsp70, C-type lectin, complement factor B-1 (CFB), Ikappa, melatonin receptor type 1A (MR1Ta), serine/threonine-protein phosphatase (STP), growth arrest, DNA damage-inducible protein (GADD), phosphoenolpyruvate carboxykinase (PECK), and Tyr (xfs0000104.14) was upregulated (Figure S12), whereas the expression of genes encoding TNF, NFX1, GT-Pase IMAP (IMAP), E3, and hemicentin was downregulated when analyzed with RT-qPCR after exposure to hypoxia for 0, 2, 5, or 8 days (Figure S12). These genes are implicated in different physiological pathways, including respiration and energy metabolism (carbonic anhydrase, SLC2A1) and antioxidant and immune defense (CFB, C-type lectins, Hsp70, lysozyme, Tyr, TNF, ANKK1, IxB, GPx).

To confirm the DEGs identified in clams under acidic conditions, 11 genes (six upregulated and five downregulated) were selected from the DEGs in several canonical pathways of interest and analyzed with RT-qPCR (Figure S13). A number of vital complement- and immune-responsive genes were significantly differentially expressed, including complement C1q, SABL, C-type lectin, Hsp70, and CYP450. An RNA-seq analysis of clams infected with digenetic trematodes identified 460 DEGs, including 416 upregulated and 44 downregulated genes (Table S29). Eleven candidate genes associated with immune, antioxidant, and molecular chaperone functions were chosen from the DEGs to confirm their expression levels with qPCR (Figure S14). In the infected clams, the expression of complement C1q and C-type lectin was significantly upregulated in the mantle and gonad. Expression of the CYP450 gene was significantly downregulated in the gills but significantly upregulated in the mantle and gonad. The expression of inhibitor of apoptosis (IAP) and glutathione S-transferase (GST) was significantly elevated in the three tissues (Figure S14). The expression of Hsp70 was significantly elevated in the gills of R. philippinarum after 2 and 5 days of hypoxia and after 12 h of acidification stress (both p < 0.05). A transcriptomic analysis of clams exposed to the parasite revealed that the expression of Hsp70 was significantly increased (p < 0.05) in three tissues tested, including the gill, mantle, and gonad (Figure S14). Overall, the RNA-seq analysis revealed that the expanded Hsp70 genes are involved in multistress resistance in R. philippinarum.

Expansion of the G-Protein-Coupled Receptor-Related Genes
G-protein-coupled receptors (GPCRs) were enriched in the unique and expanded genes of R. philippinarum. The numbers of genes containing the seven-transmembrane-associated domain in R. philippinarum, Patinopecten yessoensis, P. fucata, C. gigas, and L. gigantea were 858, 463, 533, 606, and 385, respectively (Table S30). Compared with the four other Mollusca species (C. gigas, P. fucata, L. gigantea, and Patinopecten yessoensis), the number of GPCR-related genes was expanded in
**R. philippinarum**, including the chemokine receptor, D dopamine receptor, FMRFamide receptor, G-protein-coupled receptor GRL101, neuropeptide receptor, probable G-protein-coupled receptor 139, probable G-protein-coupled receptor Mth, thyrotropin-releasing hormone receptor, and cysteinyl leukotriene receptor (Table S30, Figure S15). These expanded GPCR gene families in **R. philippinarum** may participate in the regulation of various physiological functions, especially immune functions and the inflammatory response, which are potentially important for stress resistance and the clam’s adaptation to benthic life.

**DISCUSSION**

Notably, the contig N50 of **R. philippinarum** was 28.1 kb, which is similar to those of other species in the bivalve lineage. Moreover, the complete BUSCOs in the genome of **R. philippinarum** was 92.2%, which was higher than that in **Modiolus philippinarum**, Limnoperna fortunei, Argopecten purpuratus, C. gigas, O. bimaculoides, and Saccostrea glomerata (Table S31 and Figure S16). In summary, the genome assembly index of **R. philippinarum** was similar to that of other species of Mollusca.

**Ruditapes philippinarum** has a variety of colors and a distinct set of genes that are either expanded or highly expressed and facilitate its adaptation to extreme environments. Many bivalves express multiple tyrosinases, and the expansion of the tyrosinase genes appears to be a common feature of bivalves, with more than 20 gene family members present in **R. philippinarum**, C. gigas, P. fucata, and Patinopecten yessoensis (Table S32). In the present study, qPCR analyses of the *Mitf* and *Tyr* genes showed that the *Tyr* genes have a
distinctive expression pattern in different developmental stages and that Mitf gene expression increases significantly from the early development stages that lack pigmentation to the juvenile stages with colored shells. This indicates the potential roles of these genes in shell biomineralization and shell color pigmentation. Three Mitf genes are uniquely expressed in the O strain, suggesting that these Mitf proteins are involved in the pigmentation responsible for its orange color. The distinctive expression patterns of the Mitf and Tyr genes that were demonstrated with RNA-seq and qPCR analyses shed light on the molecular basis of shell pigmentation and the color diversity in the Manila clam.

Here, we have provided an overview of the immune-related genes of the Manila clam that were identified with genomic, transcriptomic, and comparative genomic analyses. The ability of the immune system to detect pathogen-derived molecules is mediated by pattern recognition receptors (PRRs) (Janeway and Medzhitov, 2002). Several groups of invertebrate PRRs have been characterized in the Manila clam, such as TEPs, Toll-like receptors, peptidoglycan recognition proteins, fibrinogen-related proteins, galectins, C-type lectins, and complement-related proteins. In this study, the TEP superfamily genes were shown to be expanded in the R. philippinarum genome. TEPs bind to the surfaces of exogenous microorganisms with conserved thioester bonds, triggering a series of immune responses, including the activation of the complement system and phagocytosis (Fujita, 2002).

Complement component 3 (C3) is central to the complement system, and plays an important role in immune defense, immune regulation, and immune pathology. C3 is the core member of the complement system because it is the intersection of several activation pathways, and depends on a positive feedback loop from C3b. Upon its proteolytic activation, C3b binds covalently to the surface molecules of microbes via intrachain thioester bonds and initiates the formation of the membrane-attack complex (MAC), which eliminates invading pathogens via the lytic pathway (Anderluh and Gilbert, 2014). A C3a-like peptide has been shown to mediate the chemotaxis of hemocytes in Ciona intestinalis (Pinto et al., 2003) and Pyura stolonifera (Raftos et al., 2003). The activated fragments of C3 (C3a and C3b) are also involved in the activation of macrophages and mastocytes (Gao et al., 2013).

C3 is a gene conserved throughout evolution, from cnidarians to humans (Sahu and Lambris, 2001). The number of C3 gene isoforms in highly evolved vertebrates is small, but there is usually a large number of C3 genes in primitive vertebrates and invertebrates, although the number varies between species (Huan et al., 2012). C3 is an essential molecule in the complement system, acting as an opsonin in the phagocytosis of microorganisms. Therefore, as an important part of the immune system, C3 is particularly crucial to shellfish health because most shellfish lack an adaptive immune system. The number of genes in the TEP superfamily (C3, CD109, and A2M) is higher in the Manila clam than in other bivalve species (Table S27), suggesting that TEP expansion is related to the immune defense system and benthic adaptation of the Manila clam. The complement pathway genes (SABL, C1qDCs, C3, C3a, C3b) are also expanded in the Manila clam, which is evidence of an association between gene duplication and pathogen-rich ecological conditions. The pathogen recognition mechanisms of R. philippinarum include the pathogen-recognition receptors C1qDCs, Bf, and carbohydrate-binding lectins. The expansion of the C1q genes in the Manila clam has also been reported by Mun et al. (2017). Zhang et al. (2015) demonstrated that the C1qDC genes are expanded in the Pacific oyster and that C1qDC expression is upregulated by both biotic and abiotic challenges.

Transcriptomic analyses and qPCR confirmation showed that Hsp70 expression increased significantly in R. philippinarum after hypoxia stress for 2 or 5 days. The expression of Hsp70 also increased significantly after exposure to a parasite or acidification. Therefore, the Hsp70 genes are considered to be involved in the multistress resistance of R. philippinarum. Hierarchical clustering showed that the Hsp70 genes have different expression patterns at different developmental stages (Figure S11). The first cluster of eight Hsp70 mRNAs is only expressed in the pelagic larval stages, whereas other Hsp70 genes are mainly expressed in the pediveliger larval stages, implying that Hsp70 has potential roles in the immune responses in the different larval stages of the Manila clam. Overall, the expansion of the immune-related gene families (HSP70, C3, SABL, and lysozyme) is vital to the resistance to extreme conditions and the benthic adaptation of R. philippinarum.

The Hox gene family directs morphological development and tissue differentiation along all the principal axes of an embryo. Many Hox genes are expressed during early embryogenesis, suggesting their roles in
development (Krumlauf et al., 1987; Utset et al., 1987; Dolecki et al., 1988). The overall form assumed during embryonic development is controlled by the actions of the Hox proteins, their cofactors, and ultimately by the DNA regulatory elements that control their expression (Lufkin, 2005; Hejnol and Martindale, 2009; Hinman et al., 2003). In Gibbula varia, the strong expression of Gva-Hox2 was observed in trophophore and D-shaped larvae, whereas Gva-Post2 was expressed in cells around the shell field in the early trophophore, which form the future mantle edge (Samadi and Steiner, 2009). In the present study, Post2 was strongly expressed in the trophophore, veliger, and umbo larvae, demonstrating its significant role in larval development. Lox5 was mainly expressed in the umbo larva, suggesting that Lox5 is involved in regulating morphogenesis. In G. varia, the expression of Gva-Lox5 is lost in the velum of the posttorsional veliger, although the gene is still expressed in the cerebral ganglia and their commissures (Samadi and Steiner, 2009). High levels of Hox1 and Hox4 expression were detected in the D-shaped larvae of both the oyster and scallop (Zhang et al., 2012; Wang et al., 2017). The Hox proteins play pivotal roles in molluscan larval development and shell formation (Lufkin, 2005).

In conclusion, the expansion of the immune-related genes and complement pathways plays an important role in the benthic adaptation of the Manila clam. The ecology of a benthic lifestyle implies that these animals interact with high concentrations of microorganisms at the low oxygen levels in sediments. However, the Manila clam has well-adapted defense strategies to protect itself from pathogens and is recognized as a robust species that is extremely resistant to adverse environmental conditions.

Limitations of the Study
Here, we have provided a comprehensive framework for understanding the genetic adaptations of R. philippinarum to the benthic environment. Although the quality and contiguity of the R. philippinarum genome assembly was carefully validated and was generally reliable for the current study, the contig/scaffold N50 did not reach the optimal level for molluscan genome assemblies. With the development of the high-throughput sequencing technology, the quality of the assembly will be improved in future studies. The asymmetry in the quality of the two clam genomes compared means that the findings from this comparison must be considered with caution. Functional experimental assays are also required to confirm the expansion of these gene families and to identify the targets involved in the adaptation of R. philippinarum to its hypoxic, pathogen-rich environment.

METHODS
All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION
Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2019.08.049.

ACKNOWLEDGMENTS
We thank Professor Guofan Zhang (Institute of Oceanology, Chinese Academy of Sciences) for his insightful comments and constructive suggestions on this work. Many thanks are also given to Prof. Ximing Guo from Rutgers University, Prof. Li Li from Institute of Oceanology, Chinese Academy of Sciences, Prof. Kenneth M. Halanych and Dr. Yuanning Li from Auburn University for giving helpful suggestions, Prof. Kyudong Han and Seyoung Mun from Dankook University for sharing the data with us. This study was supported by the National Key R&D Program of China (2018YFD0901400), the Modern Agro-industry Technology Research System (CARS-49), Liaoning Revitalization Talents Program (XLYC1807271), Liaoning BaiQianWan Talents Program, the Scientific Research project of Liaoning Education Department (QL201703), and Dalian high-level talent innovation support program (Dalian Youth Science and Technology Star Project Support Program) (2016RQ065).

AUTHOR CONTRIBUTIONS
X.Y. and H.N. conceived the study and designed major scientific objectives. X.Y., H.N., Z.H., J.D., and M.Z. coordinated the whole project. M.Z., L.Y., L.J., Z.M., H.W., and X.M. prepared clam materials for genome and transcriptome sequencing. L.J. and H.W. conducted DNA extraction, and M.Z. prepared large-insert genomic libraries. H.N., L.Y., Z.M., and X.M. participated initial genome analysis. M.Z. and H.N. conducted genome sequencing, assembly, annotation, and gene family analysis. M.Z. conducted library preparation for genome resequencing, and G.R., L.Y., L.J., Z.M., H.W., and X.M. conducted mRNA extraction and library
preparation for transcriptome sequencing. H.N., Z.M., J.D., and L.Y. participated in genome polymorphism analysis. L.Y. and L.J. participated in shell color-related genes and pigment pathway analysis, and H.W. and C.C. participated in hypoxia transcriptome analysis. L.Y. and Z.M. participated in developmental transcriptome analysis, and H.Y. and T.H. performed ocean acidification experiments. X.M. participated in parasites transcriptome analysis. D.L. and P.C. conducted the immune response of R. philippinarum under different pathogenic stimuli (LPS, PGN, and Poly I:C). Z.L. provided computational services and technical support. H.N., Z.H., J.D., and Z.L. participated in final data analysis and interpretation. H.N., Z.L., and G.L. did most of the writing with input from other authors.

DECLARATION OF INTERESTS
The authors declare no competing interests.

Received: February 3, 2019
Revised: June 5, 2019
Accepted: August 27, 2019
Published: September 27, 2019

REFERENCES
G. Anderluh, and R. Gilbert, eds. (2014). MACPF/ CDC Proteins - Agents of Defence, Attack and Invasion (Springer). https://doi.org/10.1007/978-94-017-8881-6.

Cerenius, L., Lee, B.L., and Österhäll, K. (2008). The proPO-system: pros and cons for its role in invertebrate immunity. Trends Immunol. 29, 263–271.

Choi, K.S., Park, K.I., Lee, K.W., and Matsuoka, K. (2002). Infection intensity, prevalence and histopathology of Perkinsus sp. in the manila clam, Ruditapes philippinarum, in Ishiaya Bay, Japan. J. Shellfish Res. 21, 119–125.

Daniel, B. (2013). Marine Environmental Biology and Conservation (Jones & Bartlett Learning), p. 46.

Dolecki, G.J., Wang, G., and Humphreys, T. (1988). Stage- and tissue-specific expression of two homeo box genes in sea urchin embryos and adults. Nucleic Acids Res. 16, 11543–11558.

Dyachuk, V.A. (2016). Hemato poiesis in bivalvia larvae: cellular origin, differentiation of hemocytes, and neoplasia. Dev. Comp. Immunol. 65, 253–257.

FAO (2017). Fisheries department publications in FAO fisheries and aquaculture department. http://www.fao.org/3/a-i7989t.pdf.

Feng, D., Li, Q., Yu, H., Zhao, X., and Kong, L. (2015). Comparative Transcriptome analysis of the pacific oyster Crassostrea gigas characterized by shell colors: identification of genetic bases potentially involved in pigmentation. PLoS One 10, e0145257.

Feng, D., Li, Q., Yu, H., Kong, L., and Du, S. (2018). Transcriptional profiling of long non-coding RNAs in mantle of Crassostrea gigas and their association with shell pigmentation. Sci. Rep. 8, 1436.

Findlay, R., and Battin, T. (2016). The microbial ecology of benthic environments, p 4.2.1-1-4.2.1-20. In Manual of Environmental Microbiology, Fourth Edition, M. Yates, C. Nakatsu, R. Miller, and S. Pillai, eds. (ASM Press). https://doi.org/10.1128/9781555818821.ch4.2.1.

Fujita, T. (2002). Evolution of the lectin–complement pathway and its role in innate immunity. Nat. Rev. Immunol. 2, 346–353.

Gao, Z., Li, M., Wu, J., and Zhang, S. (2013). Interplay between invertebrate C3a with vertebrate complement arms: functional characterization of immune activities of amphioxus C3a. Fish Shellfish Immunol. 35, 1249–1259.

Gestal, C., Roch, F., Renault, T., Pallavicini, A., Paillard, C., Novoa, R., Oubella, R., Venier, P., and Figueras, A. (2008). Study of diseases and the immune system of bivalves using molecular biology and genomics. Rev. Fish Sci. 16 (51), 133–156.

Hejno, A., and Martindale, M.Q. (2009). Coordinated spatial and temporal expression of Hox genes during embryogenesis in the acel Convolutuloida longissima. BMC Biol. 7, 65.

Hinman, V.F., O’Brien, E.K., Richards, G.S., and Degnan, B.M. (2003). Expression of anterior Hox genes during larval development of the gastropod Haliotis asinine. Evol. Dev. 5, 508–521.

Huan, P., Wang, H., and Liu, B. (2012). Transcriptomic analysis of the clam Meretrix meretrix on different larval stages. Mar. Biotechnol. 14, 69.

Janeway, C.A., and Medzhitov, R. (2002). Innate immune recognition. Annu. Rev. Immunol. 20, 197–216.

Ko, Y.M., Park, K.I., Choi, K.S., Alvarez, R.A., Cummings, R.D., and Cho, M. (2006). Lectin from the Manila clam Ruditapes philippinarum is induced upon infection with the protozoan parasite Perkinsus olsenii. J. Biol. Chem. 281, 26854–26864.

Kodama, K., Waku, M., Sone, R., Miyawaki, D., Ishida, T., Akatsuka, T., and Horiguchi, T. (2018). Ontogenetic and temperature-dependent changes in tolerance to hypoxia and hydrogen sulfide during the early life stages of the Manila clam Ruditapes philippinarum. Mar. Environ. Res. 137, 177–187.

Krumholz, A., Yao, J., Wang, L.V., vanVickle-Chavez, S.J., Fleming, T., and Gilllanders, W.E. (2011). Photoacoustic microscopy of tyrosinase reporter gene in vivo. J. Biomed. Opt. 16, 080503.

Krumlauf, R., Holland, P.W., McVey, J.H., and Hogan, B.L. (1997). Developmental and spatial patterns of expression of the mouse homeobox gene, Hox-2.1. Development 99, 603–617.

Lemer, S., Saulnier, D., Gueguen, Y., and Planes, S. (2015). Identification of genes associated with shell color in the black-lipped pearl oyster, Pinctada margaritifera. BMC Genomics 16, 568.

Lim, H.S., Diaz, R.J., Hong, J.S., and Schaffner, L.C. (2006). Hypoxia and benthic community recovery in Korean coastal waters. Mar. Pollut. Bull. 52, 1517–1526.

Luftin, T. (2005). Hox Genes: Embryonic Development (eLS).

Mann, K., and Jackson, D.J. (2014). Characterization of the pigmented shell-forming proteome of the common grove snail Cepaea nemoralis. BMC Genomics 15, 249.

Milan, M., Coppe, A., Reinhardt, R., Cancela, L.M., Leite, R.B., Saavedra, C., Ciofi, C., Chelazzi, G., Patanello, T., Bortoluzzi, S., and Bargelloni, L. (2011). Transcriptome sequencing and microarray development for the Manila clam, Ruditapes philippinarum: genomic tools for environmental monitoring. BMC Genomics 12, 234.

Mun, S., Kim, Y.-J., Markkandian, K., Shin, W., Oh, S., Woo, J., Yoo, J., An, H., and Han, K. (2017). The whole-genome and transcriptome of the Manila clam (Ruditapes philippinarum). Genome Biol. Evol. 9, 1487–1498.

Nie, H., Liu, L., Huon, Z., Chen, P., Ding, J., Yang, F., and Yan, X. (2017a). The Hsp70 gene expression responses to thermal and salinity stress in wild and cultivated Manila clam Ruditapes philippinarum. Aquaculture 470, 149–156.
Nie, H., Chen, P., Huo, Z., Chen, Y., Hou, X., Yang, F., and Yan, X. (2017b). Effects of temperature and salinity on oxygen consumption and ammonia excretion in different color strains of the Manila clam, Ruditapes philippinarum. Aquaculture Res. 48, 2778–2786.

Norkko, J., and Shumway, S.E. (2011). Bivalves as bioturbators and bioirrigators. In Shellfish Aquaculture and the Environment Chapter 10, S.E. Shumway, ed. (Wiley Science Publishers), pp. 297–317.

Peignén, J.M., Gérard, A., Naciri, Y., Ledu, C., and Phelipot, P. (1995). Analysis of shell color determinism in the Manila clam Ruditapes philippinarum. Aquat. Living Resour. 8, 181–189.

Pinto, M.R., Chinneci, C.M., Kimura, Y., Melillo, D., Marino, R., Spruce, L.A., De Santis, R., Parrinello, N., and Lambis, J.D. (2003). CiC3-1a-mediated peptide stimulates chemotaxis by hemocytes from an invertebrate chordate-the tunicate, Ciona intestinalis. J. Immunol. 171, 5521–5528.

Pretto, T., Zambon, M., Civettini, M., Caburlotto, G., Boffo, L., Rossetti, E., and Arcangeli, G. (2014). Excretion in different color strains of the Manila clam, Ruditapes philippinarum. Aquaculture Res. 45, 2778–2786.

Pretto, T., Zambon, M., Civettini, M., Caburlotto, G., Boffo, L., Rossetti, E., and Arcangeli, G. (2014). Massive mortality in Manila clams (Ruditapes philippinarum) farmed in the lagoon of Venice, caused by Perkinsus olseni. Bull. Eur. Assoc. Fish Pathol. 34, 43–53.

Protas, M.E., and Patel, N.H. (2008). Evolution of coloration patterns. Annu. Rev. Cell Dev. Biol. 24, 425–446.

Rafatos, D.A., Robbins, J., Newton, R.A., and Nair, S.V. (2003). A complement component C3a-like peptide stimulates chemotaxis by hemocytes from an invertebrate chordate-the tunicate, Pyura stolonifera. Comp. Biochem. Physiol. A 134, 377–386.

Sahu, A., and Lambris, J.D. (2001). Structure and biology of complement protein C3, a connecting link between innate and acquired immunity. Immunol. Rev. 180, 35.

Samadi, L., and Steiner, G. (2009). Involvement of Hox genes in shell morphogenesis in the encapsulated development of a top shell gastropod (Gibbula varia L.). Dev. Genes Evol. 219, 523–530.

Shimokawa, J., Yoshinaga, T., and Ogawa, K. (2010). Experimental evaluation of the pathogenicity of Perkinsus olseni in juvenile Manila clams Ruditapes philippinarum. J. Invertebr. Pathol. 105, 347–351.

Simakov, O., Mareta, F., Cho, S.J., Edsinger-Gonzales, E., Hlavak, P., Helisten, U., Kuo, D.H., Larsson, T., Lu, J., Arendt, D., Savage, R., et al. (2013). Insights into bilateral evolution from three spiralian genomes. Nature 493, 526–531.

Song, L., Wang, L., Qiu, L., and Zhan, H. (2010). Bivalve immunity. In Invertebrate Immunity. Advances in Experimental Medicine and Biology, 708, K. Söderhäll, ed. (Springer), pp. 44–65.

Steckbauer, A., Ramajo, L., Ramajo, L., Hendriks, I.E., Fernandez, M., Lagos, N.A., Prado, L., and Duarte, C.M. (2015). Synergistic effects of hypoxia and increasing CO2 on benthic invertebrates of the central Chilean coast. Front. Mar. Sci. 2, 1–12.

Sun, X., Yang, A., Wu, B., Zhou, L., and Liu, Z. (2015). Characterization of the mantle transcriptome of Yesso scallop (Pinctopecten yessoensis): identification of genes potentially involved in biomeralization and pigmentation. PLoS One 10, e0122967.

Takeuchi, T., Kawashima, T., Koyanagi, R., Gyoja, F., Tanaka, M., Ikuta, T., Shoguchi, E., Fujiwara, M., Shinzato, C., Hisata, K., et al. (2012). Draft genome of the pearl oyster Pinctada fucata: a platform for understanding bivalve biology. DNA Res. 19, 117–130.

Tamura, K. (1993). Tolerance of Theora fragilis (Bivalvia: Semelidae) to low concentrations of dissolved oxygen. Nippon Suisan Gakkaishi 59, 615–620.

Utset, M.F., Awgulewitsch, A., Ruddle, F.H., and McGinnis, W. (1987). Region-specific expression of two mouse homeobox genes. Science 235, 1379–1382.

Vaughn, C.C., and Hakenkamp, C.C. (2001). The functional role of burrowing bivalves in freshwater ecosystems. Freshw. Biol. 46, 1431–1446.

Wang, Y., Sheng, H., He, Y., Wu, J., Jiang, Y., Tam, N.F., and Zhou, H. (2012). Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of illumina tags. Appl. Environ. Microbiol. 78, 8264–8271.

Wang, S., Zhang, J., Jiao, W., Li, J., Sun, X., Sun, Y., Guo, X., Huan, P., Dong, B., Zhang, L., et al. (2017). Scallop genome provides insights into evolution of bilaterian karyotype and development. Nat. Ecol. Evol. 1, 120.

Yue, X., Nie, Q., Xiao, G., and Liu, B. (2015). Transcriptome analysis of shell color-related genes in the clam Meretrix meretrix Mar. Biotechnol. 17, 1–11.

Zhang, G.F., and Yan, X.W. (2010). In Clam Aquaculture (Science Press), in chinese.
Supplemental Information

Clam Genome Sequence Clarifies
the Molecular Basis of Its Benthic Adaptation
and Extraordinary Shell Color Diversity

Xiwu Yan, Hongtao Nie, Zhongming Huo, Jianfeng Ding, Zhenzhen Li, Lulu Yan, Liwen Jiang, Zhengqiang Mu, Huamin Wang, Xiangyu Meng, Peng Chen, Mengyan Zhou, Md. Golam Rbbani, Guangjian Liu, and Dongdong Li
**Fig. S1** The shell color and appearance of different strains of *R. philippinarum*. A Zebra strain. B White strain. C White Zebra strain. D Orange strain, Related to Figure 2.
**Fig. S2** Distribution of 17-mer frequency in Manila clam genome. The horizontal axis represents the K-mer depth, namely the number of times occurred. The volume of K-mer is plotted against the frequency at which they occur. The left-hand peak at low frequency and high volume represented K-mer containing essentially random sequencing errors. The main volume peak of K-mer was 38, and there was a hybrid peak at about half of the main volume peak of K-mer depth (~20 X). The genome size was estimated as 1.32 Gb by the formula 'Genome size=total_kmer_num/kmer_depth', Related to Figure 1.
Fig. S3 The sequencing depth distribution. All high-quality reads generated from short insert size libraries were mapped onto the assembled genome, the sequencing depth of all contigs in the genome (green), contig from heterozygous regions (red) and contig from homozygous regions (blue) were plotted. The peak of contigs from homozygous contigs was 123 X, and from the heterozygous contigs was 61 X, Related to Figure 1.
**Fig. S4** Divergence distribution of classified families of TEs. Red indicated DNA transposons (DNA), dark yellow indicated long interspersed nuclear element (LINE), green indicated long terminal repeat (LTR), light blue indicated short interspersed nuclear element (SINE), and purple indicated unknown TEs, Related to Figure 1.
Fig. S5 Comparison of gene parameters among *R. philippinarum* genome and other Lophotrochozoa genomes. The similar gene parameters among *R. philippinarum*, *C. gigas*, *C. teleta*, *H. robusta*, *L. anatina*, *L. gigantea* and *O. bimaculoides* indicate the high quality gene structure annotation in *R. philippinarum* genome, Related to Figure 1.
Fig. S6 The distribution of genes in different species. Pink indicated single-copy orthologs, yellow indicated multiple-copy orthologs, dark yellow indicated unique genes, and green indicated other orthologs, Related to Figure 1.
Fig. S7 The venn diagram showed common and unique gene families among five Mollusca species (*R. philippinarum, L. gigantea, C. gigas, P. fucata* and *P. yessoensis*). Compared with other four Mollusca species, *R. philippinarum* owned 1,582 unique gene families and shared 6,975 gene families with others, Related to Figure 1.
Fig. S8 The expression of key genes under different stimuli (LPS, PGN and Poly I:C), hypoxia, acidification and parasites. RT-qPCR expression of genes at 3, 6, 12, 24, 48, 72 and 96 h after the treatment of lipopolysaccharide (LPS), peptidoglycan (PGN), and polyinosinic polycytidylic acid (Poly I:C), Related to Figure 3.
Fig. S9 RT-qPCR expression of eleven Hox genes at seventeen different developmental stages (Egg: eggs; Zyg: zygotes; Pol1: first polar body; Pol2: second polar body; C2: 2 cell; C4: 4 cell; C8: 8 cell; C16: 16 cell; C32: 32 cell; Gas: gastrula; ET: early trochophora; LT: late trochophora; Vel: veligar larvae; UL: umbo larvae; Ped: pediveliger larvae; Juv: juvenile; HC: have color juvenile), Related to Figure 5.
Fig. S10 RT-qPCR expression of TEP and C3 genes, Related to Figure 3.
**Fig. S11** Heat map and hierarchical clustering showing expression of Hsp70 genes at different developmental stages, Related to Figure 3.
**Fig. S12** RT-qPCR expression of genes at 0, 2, 5 and 8 d of hypoxia exposure, Related to Figure 3.
**Fig. S13** RT-qPCR and DEGs expression of genes in acidification stress, Related to Figure 3.
**Fig. S14** RT-qPCR expression of genes in gills, mantle and gonad exposed to digenetic trematode, Related to Figure 3.
**Fig. S15** The expanded function statistics of GPCR in *R. philippinarum* compared with *P. yessoensis*, *C. gigas*, *P. fucata* and *L. gigantea*, Related to Figure 4.
Fig. S16 The statistics of genomes of *R. philippinarum* and other homologous species. C, complete BUSCOs; F, fragmented (partial) BUSCOs; M, missing BUSCO. Related to Figure 1.
**Transparent Methods**

**Sampling**

The fifth generation of orange shell color strains of *R. philippinarum* was collected from the Zhangzidao aquaculture farm in Dalian, Liaoning Province, China. The clams with a similar shell length averaging at 20.90±1.34 mm, respectively, were used in this study. After being transported from the field to the laboratory, the clams were acclimated in aerated 70 L plastic tanks, containing water at 20 °C with salinity of 30. All the samples were acclimated in the laboratory for one week before the sampling. The Manila clam is not an endangered or protected species, so no specific permits were required for the study. One of the clams was sacrificed for DNA extraction.

The present study was approved by College of Fisheries and Life Science, Dalian Ocean University. Two adult individuals of orange strain of Manila clam *Ruditapes philippinarum* were selected for whole genome sequencing and genomic DNA was extracted from the whole individuals for short and long libraries of the whole-genome shotgun sequencing. One of the F6 Orange strains of Manila clam with distinctive orange shell color and highly inbred individual was selected for whole genome sequence (Fig. S1D).

**Genome sequencing**

Genomic DNA was broken into random fragments according to the whole genome shotgun sequencing strategy. Illumina paired-end (PE) libraries were constructed with insert size ranging from 230 bp to 500 bp. Illumina mate-pair (MP) libraries were constructed targeting insert size of 2 Kbp, 5 Kbp, 10 Kbp and 20 Kbp (35 lanes, Table S2). All the libraries were sequenced by Illumina Hiseq 2500 platform. Totally 286.1 Gb raw data was produced by PE libraries, and 135.1 Gb raw data produced by MP libraries.

To reduce the effect of sequencing error to the assembly, the following type of reads would be filtered: reads with ≥10% unidentified nucleotides for libraries, reads from short-insert libraries having more than 50% bases with quality score less than 5, reads having the sequencing adapters, read1 and read2 of two paired-end reads that were completely identical (considered to be products of PCR duplication). After filtering out the adapter sequences, low quality and duplicated reads, a total of 361.4 Gb (~274-fold coverage) data were remained as clean data for the whole-genome assembly.
To characterize the distribution of single-nucleotide polymorphism (SNP) and estimate the level of heterozygosity in the *R. philippinarum* genome, short pair-ended reads from the inbred and wild samples were aligned to the assembled genome. We identified 33,830,021 and 11,758,346 heterozygous SNPs in the assembled (inbred) and resequenced (wild) genome with the polymorphism rate of 1.038% and 1.690%, respectively (Table S4). As expected, inbreeding has significantly reduced its polymorphisms. The level of polymorphism in *R. philippinarum* is similar to that in Lingula genome (1.6%), but higher than that found in the scallop *Patinopecten yessoensis* (1.04%) and the Pacific oyster *Crassostrea gigas* (1.30%), indicating that the *R. philippinarum* genome exhibits comparatively high heterozygosity.

**Estimation of genome size**

The distribution of K-mer frequency is widely used for the estimation of genome size. Total of 58 Gb high-quality short-insert reads (230 bp) were used to estimate the Manila clam genome size by analyzing the K-mer frequency (defined as 17 bp here), the coverage sequencing depth of which follows Poisson distribution (Li et al., 2010). The formula ‘Genome size = total_kmer_num/kmer_depth’ (Liu et al., 2013), where total_kmer_num is the total number of K-mer and kmer_depth is the peak sequencing depth on the K-mer frequency distribution map, was used to estimate the genome size (Fig. S2). And the genome size of *R. philippinarum* were determined to be 1.32 Gb with high ratio of heterozygosity (Table S1).

**Genome assembly**

It is a challenging task to obtain a high-quality genome of *R. philippinarum* due to the high levels of heterozygosity (~2.0%). To tackle this challenge, an innovation assembly software NOVOheter2.0, developed by Novogene Bioinformatics Institute (Beijing, China) on basis of SOAPdenovo package (Li et al., 2010), was applied for the Manila clam genome assembly. Firstly, the heterozygous K-mers were determined from coverage information obtained from de Bruijn graph in *R. philippinarum* genome. All high-quality reads from short-insert size libraries were used to split and counted to generate 19-mer frequency distribution, and 19-mers with sequencing depth ranged from 23 to 124 were regarded as the initial heterozygous 19-mer set. Next, a de Bruijn graph was constructed by splitting the high-quality reads from short-insert size libraries into 45-mers with SOAPdenovo setting parameters as ‘-e 1 -M 1 -R’, and contigs with sequencing
depth ranged from 20 to 90 were regarded as heterozygous contigs (Fig. S3).

Secondly, PE and MP reads were mapped to the heterozygous K-mer set to obtain the coverage relationship, and an effective link between two heterozygous contigs was defined if there were at least three read pairs mapped. Two contigs in a bubble represents two potential haploids in a diploid genome. The heterozygous contigs were clustered together as bubble contigs considering the order and distance between two heterozygous contigs. After clustering and classifying heterozygous contigs, there were two big groups of bubble contigs waiting to be removed in the next step.

The heterozygous contigs were matched back to the corresponding edges in de Bruijn graph constructed by 45-mers, one of the heterozygous contig groups on the edges were removed to solve the bubble structures or intricate structures in the graph caused by heterozygosity of diploid R. philippinarum genome. After merging bubble contigs, the contig sequences were obtained finally.

After obtaining contig sequences, all high-quality short-insert PE reads and large-insert MP reads were realigned onto these contigs by Platanus assembler (v1.2.1) (Kajitani et al., 2014) with scaffold protocol (default parameters). And the shared paired-end relationships between each pair of adjacent contigs were used to construct scaffolds.

Gaps in the constructed scaffolds were filled by GapCloser (English et al., 2012) (http://sourceforge.net/projects/soapdenovo2/files/GapCloser) with parameters setting as ‘-l 150 -p 30’ based on the paired-end information of the read pairs that had one end mapped to the unique contig and the others located in the gap region.

After removing the heterozygosity in R. philippinarum genome by HaploMerger (Huang et al., 2012) with default parameters, a total of 30,811 scaffolds were assembled covering total length over 1,123 Mb. And a scaffold N50 length of 345,005 bp, contig N50 length of 28,111 bp and total length of 1.12 Gb were gained in the final Manila clam genome assembly (Table S3).

**Genome evaluation**

The high-quality short insert PE reads were aligned onto the assembled genome to evaluate the accuracy by using BWA (Li and Durbin, 2009) (http://bio-bwa.sourceforge.net/) with parameters setting as ‘-o 1 -i 15’. The distribution of the sequencing depth at each position was calculated to
measure the completeness of the genome assembly. The assembly with GC content equal to 31.89% (Table S5) was evaluated through mapping reads with approximately 93.32% of mapping rate and 99.83% of coverage (Table S6), explaining the completeness of the genome assembly.

Raw data of transcriptomes from the adductor muscle (5.84 Gb) were sequenced by Illumina sequencing platform. Transcripts were assembled by Trinity (Grabherr et al., 2011) with parameters setting as ‘-ss 0.5 -jc 0 -minkmercov 2 -minglue 2’ to be as expressed sequence tags (ESTs). The 78,505 random ESTs were aligned to the scaffolds by Blat (Kent, 2002) (http://genome.ucsc.edu/goldenpath/help/blatSpec.html) with default parameters, 81.52% of which (> 200 bp) were more than 90% covered by a scaffold (Table S7). The final assembly version covered almost of the genome and the region of coding genes.

CEGMA (Parra et al., 2007) (Core Eukaryotic Genes Mapping Approach, http://korflab.ucdavis.edu/dataseda/cegma/) defined a set of conserved protein families that occur in a wide range of eukaryotes, and present a mapping procedure that accurately identifies their exon-intron structures in a novel genomic sequence. Through mapping to the 248 ultra conserved eukaryotic genes, a total of 233 genes with a ratio of 93.95% were found (Table S8), supporting a high ratio of core conserved gene families, which indicated that our assembly reached a better integrity of the core genes.

BUSCO sets (Simao et al, 2015) (Benchmarking Universal Single-Copy Orthologs: http://busco.ezlab.org/) were used to assess completeness of genome assembly combining with TBLASTN (Altschul et al, 1997), Augustus (Stanke and Waack, 2003; Stanke et al., 2006) (version 3.0.2) and HMMER (Mistry et al., 2013) (version 3.1b2). Analysis of BUSCO showed that the assembly contains 90.39% complete and 5.93% fragmented metazoa BUSCO orthologues, which suggested that the assembly is of high quality. (Table S9)

Annotation of repeats

After the genome assembly, repeat annotation was carried out for the R. philippinarum genome. Repetitive sequences include transposable elements (TEs) and tandem repeats. Two approaches were used to discover TEs. RepeatMasker (Smit, Chen, 2004) (version 3.3.0) found TEs in a integrate repeat library which was derived from known repeat library (Repbase 15.02) and the de novo repeat library, built by RepeatModeler (Smit, 2017) (Vision 1.0.5), RepeatScout (Price et al,
2005), and LTR_FINDER (Xu and Wang, 2007). RepeatProteinMask (Chen, 2004) was performed to detect TEs in *R. philippinarum* genome by comparing the TE protein database. Tandem repeats were ascertained in the genome using Tandem Repeats Finder (Benson, 1999). Repetitive sequences comprised ~38.30% of the *R. philippinarum* genome, and the results were shown in detail in Table S10, Table S11 and Fig. S4.

**Annotation of protein coding genes**

De novo predictions, homolog-based predictions and RNA-seq based predictions were employed to annotate the protein coding genes in *R. philippinarum* genome. Five ab initio gene prediction programs were used to predict genes, including Augustus (Stanke and Waack, 2003; Stanke et al., 2006) (version 3.0.2), Genescan (Salamov and Solovyev, 2000) (version 1.0), Geneid (Parra et al., 2000), GlimmerHMM (Majoros et al., 2004) (version 3.0.2) and SNAP (Korf, 2004). Protein sequences of ten homologous species (Anopheles gambiae, Caenorhabditis elegans, Crassostrea gigas, Capitella teleta, Drosophila melanogaster, Helobdella robusta, Lingula anatina, Lottia gigantea, Octopus bimaculoides, Strongylocentrotus purpuratus) were downloaded from Ensembl or NCBI. Homologous sequences were aligned against the repeat-masked *R. philippinarum* genome using TBLASTN (Altschul et al., 1997) (E-value ≤ 1E-05). Genewise (Birney et al., 2004) (version 2.2.0) was employed to predict gene models based on the alignment sequences. There were two ways to assemble the RNA-seq data into the unique sequences of transcripts. One was mapping the RNA-seq data to the *R. philippinarum* genome using Tophat (Trapnell et al., 2009) (version 2.0.8) and using cufflinks (Trapnell et al., 2010) (version 2.1.1) (http://cufflinks.cbcb.umd.edu/) for transcript assembly. The other was applying Trinity (Grabherr et al, 2011) to assemble the RNA-seq data, and then PASA (Haas et al., 2003) software (http://pasapipeline.github.io/) improved the gene structures. A weighted and non-redundant gene set was generated by EVidenceModeler (EVM) (Haas et al., 2008) which merged all genes models predicted by the above three approaches. Combining with transcript assembly, PASA adjusted the gene models generated by EVM. At last, the gene set were filtered according to the following standards: coding region lengths of ≤ 50 amino acids, supported only by de novo methods and with FPKM < 5. The final reference gene set contained 27,652 protein coding genes. (Table S12, Table S13 and Fig. S5)
**Functional annotation**

Functional annotation of protein coding genes was obtained according to best BLAST hit by BLASTP (E-value $\leq 1E^{-05}$) against SwissProt (Bairoch and Apweiler, 2000), TrEMBL (Bairoch and Apweiler, 2000) and NCBI non-redundant (NR) protein databases. Motifs and domains were annotated by using InterProScan (Mulder and Apweiler, 2007) (version 4.7) to search against InterPro (Mulder and Apweiler, 2007) (v29.0) databases, including Pfam, PRINTS, PROSITE, ProDom and SMART. A Gene Ontology (Ashburner et al., 2000) (GO) term for each gene were achieved from the corresponding InterPro descriptions. Additionally, gene set was mapped to a KEGG (Kanehisa and Goto, 2000) pathway to identify the best match classification for each gene. Finally, 24,570 protein coding genes (accounting for 88.85%) owned functional annotation. (Table S14)

**Non-coding RNA annotation**

The tRNA genes were predicted by tRNAscan-SE software (Lowe and Eddy, 1997). The miRNA and snRNA fragments were identified by INFERNAL (Nawrocki et al, 2009) software against the Rfam (Griffiths-Jones et al., 2005) database. The rRNA were found by using BLASTN (E-value $\leq 1E^{-10}$) against invertebrate rRNA database. (Table S15)

**Gene family cluster**

Gene families were generated by OrthoMCL (Li et al., 2003) (http://orthomcl.org/orthomcl/). First, nucleotide and protein data of 13 species (Amphimedon queenslandica, Branchiostoma floridae, Capitella teleta, Crassostrea gigas, Daphnia pulex, Drosophila melanogaster, Helobdella robusta, Homo sapiens, Lingula anatina, Lottia gigantea, Octopus bimaculoides, Patinopecten yessoensis, Pinctada fucata) were downloaded from Ensembl (Release 70), NCBI and http://marinegenomics.oist.jp/gallery/. Before an “all against all” BLASTP (E-value $\leq 1E^{-07}$) program, the longest transcript was selected from alternative splicing transcripts belonging to one gene and genes with $\leq 50$ amino acids were removed. The alignments with high-scoring segment pairs (Hsps) were conjoined for each gene pair by solar (Yu et al., 2006). To identify homologous gene-pairs, more than 30% coverage of the aligned regions in both homologous genes was required. Finally, the alignments were clustered into gene families using OrthoMCL with 1.5
inflation index. After clustering, 33,076 gene families and 288 single-copy orthologs were detected across *R. philippinarum* and 13 other species (Fig. S6 and Fig. S7). GO enrichment and KEGG enrichment of unique gene families in *R. philippinarum* were presented in Table S16 and Table S17.

**Phylogenetic tree construction and divergence time estimation**

The above mentioned 288 shared single-copy orthologs were utilized to construct the phylogenetic tree. Protein sequences of these orthologs were aligned by MUSCLE (Edgar, 2004), then protein alignments were transformed to CDS alignments. Using this CDS alignments, the phylogenetic tree was constructed by the ML (maximum likelihood) TREE algorithm in RAxML software (Stamatakis, 2006; Stamatakis et al., 2008) (version 7.2.3). Then mcmctree program of PAML (Yang, 2007) (http://abacus.gene.ucl.ac.uk/software/paml.html) was applied to estimate divergence time among 14 species with main parameters burn-in=100,000, sample-number=100,000, and sample-frequency=2. Four calibration points were selected from Simakov et al., (2015) and TimeTree (Hedges et al., 2015) website (http://www.timetree.org) as normal priors to restrain the age of the nodes, such as 518~581 Mya for TMRCA of *H.sapiens*-B.*floridae*; 500~550Mya for TMRCA of *C.gigas*-L.*gigantea*; 305~581Mya for TMRCA of *C.teleta*-H.*robusta*; 510~543Mya for TMRCA of *D.pulex*-D.*melanogaster*. The phylogenetic tree confirmed the grouping of Lophotrochozoa. The split of *R. philippinarum* was estimated at 476.9 Mya, closing to that reported by many others.

**Expansion and contraction of gene families**

To avoid extreme gene families, families with gene number ≥ 200 in one species and ≤ 2 in all other species were filtered at first. Cluster size of each gene family in the *R. philippinarum* genome was compared with that of *C.gigas*, *L.gigantea*, *P.fucata* and *P.yessoensis* by using Fisher exact test to identify the comparatively expansion and contraction of gene families in *R. philippinarum* to the other three species. P-value of Fisher exact test was adjusted by the false discovery rate (FDR). Judged by adjusted p-value of 0.05, 2,197 gene families significantly expanded and 3,943 gene families contracted in the *R. philippinarum* genome. (Table S18) GO enrichment and KEGG enrichment of expansion and contraction gene families in *R. philippinarum* were presented in Table S19-S22.
Positively selected genes

The protein alignments of 4,494 single-copy gene families in R.philippinarum, C.gigas, L.gigantea, P.fucata and P.yessoensis were generated using MUSCLE. Gblocks (Castresana, 2000) was applied to filter poorly aligned positions and divergent regions of the protein alignments before transformed to CDS alignments. With the R. philippinarum as foreground branch, positive selection sites were detected based on branch-site models (Zhang et al., 2005) of PAML using the CDS alignments. P-values were computed using the $\chi^2$ statistic and adjusted by FDR method. In addition, in order to ensure the accuracy of the positive selection results, the genes with low alignment quality were filtered with the following criteria (PMID:25362486): gene length more than 300 bp, and no gap within 3 amino acids around the positive selected site in the five species. Finally, 29 positively selected genes were found in R. philippinarum genome. GO enrichment and KEGG enrichment of positively selected genes in R. philippinarum were presented in Table S23 and Table S24.

Comparative genome analysis between R. philippinarum and Korean Manila clam genome

The Korean Manila clam genome sequencing data (Mun et al., 2017) previously published was used to conduct the comparative analysis between the two draft Manila clam genome. Since the gene prediction of the genome of Korean clams is the genomic data (V1) without deleting heterozygosity effectively, there is a large redundancy in the resulting of 108,034 gene models. The software of HaploMerger removes the hybrid segment without interrupting the original genomic version (Huang et al., 2012). Based on which, we mapped the genome data of the last version (V2) of Korean clams to the genome of V1 version using blast. The sequences would be removed except the sequences were matched perfectly without “N”. The number of genes finally obtained was 19,776 according to the gene models in the matched sequences.

Linkage map-based chromosome anchoring

A total of 20,581 marker sequences with physical distance from 19 linkage group were obtained from a linkage map of R. philippinarum reported by Nie et al. (2017c). Only markers with a unique location were used for anchoring and orienting scaffolds to different linkage groups. Scaffolds in conflict with the genetic map (such as markers from a different linkage on the same scaffold) were checked manually with 10 kb mate-paired reads. At last, 3,888 markers were used
to construct the chromosome.

Collinearity analysis between *R. philippinarum* (Rph) and Korean clams (Hrph)

The sequences of Hrph were aligned to the genome assembly of Rph using LASTZ (version 1.02.00) with parameters “M=254 K=4500 L=3000 Y=15000 --seed=match12 --step=20 --identity=80”.

RNA-seq of four color strains of *R. philippinarum*

Four selective breeding strains of *R. philippinarum* exhibiting steady and hereditary with different kinds of shell color morphs including zebra strain with brown zigzag shells (Z); white strain with white shells with a black band in the edge (W); orange strain with orange shell color (O); white zebra strain with both zigzag and white shells with a black band in the edge (WZ) (Fig.S1). The original population of different shell color including O, W, WZ, Z and partially pigmented *R. philippinarum* were selected from cultured populations in Dalian, Liaoning, China. These four selected strains were selected by 4-7 years successive generations selection and the shell color traits were steadily hereditary. One of the main functions of the mantle is to secrete biomineralization proteins to form the shell. The mantle pallial and mantle edge are the main secretory tissues. In this experiment, the mantle of four color strains was sampled for RNA extraction. Three biological replicates per each color were collected. A total amount of 3 μg RNA from mantles of each shell color strains (O, W, Z, and WZ) were used for library construction and transcriptome sequencing.

RNA-seq of different developmental stages of *R. philippinarum*

Healthy *R. philippinarum* (shell length: 3.2±0.3 cm) were collected from Zhangzidao (Dalian, Liaoning province, China) and maintained in 70 L tanks containing aerated sand-filtered seawater (salinity: 31±1 ppt, pH 8.1±0.1) at 22±1°C for one week prior to experimentation. Artificial breeding larvae were cultivated and collected at different stages including trochophore (A11), D-shape larva (A12), umbo larva (A23) and pediveliger (Z3). The early phase of juvenile without shell color and late phase of juvenile with shell color were also sampled for RNA-seq in the orange strain and zebra strain, respectively. The collected samples were stored at -80°C before RNA extraction.

Transcriptome response to acidification, parasite and hypoxia stress
After preliminary acclimation, one hundred clams dividing into two groups were cultured on two 70 L tank for 3 months at pH 8.0 (control) and 7.4 (treated). For water acidification, CO$_2$ gas was diffused directly into seawater contained in each tank to achieve testing pH conditions. Seawater pH was continuously monitored and controlled using a pH meter. Total alkalinity (TA), pH, temperature and salinity, were measured daily to determine CO$_2$ partial pressure (pCO$_2$), bicarbonate (HCO$_3^-$) and carbonate (CO$_3^{2-}$) ions concentrations. Three biological replicates per each condition were collected.

The Manila clams were collected from the parasitic infected zone in Dalu Island. Preliminary experiments were first performed to identify the infected tissue and gonads seriously injured in the infected clams. The normal and infected gonads were selected as the samples for RNA-seq analysis. For RNA extraction, the clams were dissected in situ and the gonads were isolated from the infected and healthy individuals. Three biological replicates per each sample were collected. Then the gonads were immediately frozen in liquid nitrogen and transported in dry ice for subsequent experiments.

The experiment was performed in bottle tanks with an effective water volume of 30 L. Tanks were supplied with a continuous flow of water at 12±0.3 °C and 32 ppt salinity. Adult *R. philippinarum*, collected from Zhuanghe, Dalian, China, were divided into two groups of 50 animals. After a 7-day acclimatization period in tanks supplied with aerated filtered seawater, clams were exposed to hypoxia (Dissolved oxygen, DO: 2 mg/L) and normoxia (DO: 8 mg/L) as control groups. At day 0, the start of the experiment, the DO in the water was decreased to 2 mg/L using an oxygen depletion system through a fine bubble aerator where nitrogen was injected. Oxygen removal was controlled by nitrogen flow. Surface gas exchange in the rearing tank was limited by setting the water inflow under the water surface. The O2 concentration in the tank was monitored using YSI oxymeter and adjusted when necessary to keep hypoxia level constant all along the experiment. Normoxia was obtained by equilibrating seawater with air. The vitality and motility of the clams suffered no adverse affects when DO level changed and no mortality were observed either in the control or in the hypoxia-exposed clams.

Clams under hypoxia and normoxia conditions were sampled at 0, 2, 5, 8 day, respectively. Three biological replicates per each time and condition were collected. Total RNA was extracted
from the gills of 12 controls and 12 hypoxia exposed clams at 0, 2, 5, 8 days of exposure using RNAprep pure Tissue Kit (TianGene, Beijing, China), according to the manufacturer’s protocol. RNA purity was determined with a NanoPhotometer® spectrophotometer (Implen, CA, USA). The RNA concentration was measured with the Qubit® RNA Assay Kit in a Qubit® 2.0 Fluorometer (Life Technologies, CA, USA). RNA integrity was assessed with the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). Six libraries (Hypoxia and Normoxia) constructed using gills after 2 days hypoxia treatments were used for transcriptome sequencing.

**Sequencing libraries and RNA-seq pipeline**

In short, the extraction workflow included isolation of poly-adenylated RNA molecules using poly-T oligo-attached magnetic beads, enzymatic RNA fragmentation, cDNA synthesis, ligation of bar-coded adapters, and PCR amplification. Then, sequencing libraries were generated using NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA) following manufacturer’s recommendations and index codes were added to attribute sequences to each sample. All the seven libraries were sequenced on Illumina HiSeq platform with PE150.

Sequencing reads were aligned to the *R. philippinarum* genome using hisat2 with default parameters. Gene expression levels in terms of FPKM were estimated by custom Perl scripts. Differentially expressed gene (DEG) analysis was carried out using edgeR with three biological replicates, and genes with a fold change value ≥ 2 and adjusted p-value < 0.05 were defined as significant DEGs.

**Identification of the immune-related genes**

The immune-related protein sets from NCBI and Uniprot (reference proteins) were searched against the five protein sets (*R.philippinarum, C. gigas, L. gigantea, P. fucata, P. yessoensis*) using BLASTP with the parameter "-e 1e-5". The genes were identified as candidates following criterion: 1) the genes have a similarity of more than 40% with the reference proteins; 2) the genes have the same domain structures with the reference proteins; 3) the annotation of the genes are same reference proteins. The motifs and domains identification of the genes were mainly performed with the Pfam and SMRT database. The genes were functionally annotated based on homologous searches against databases of NR database (NCBI).
Identification of the G protein-coupled receptor (GPCR) related genes

The protein-coding genes of *R. philippinarum*, *P. yessoensis*, *P. fucata*, *C. gigas* and *L. gigantea* were annotated based on homologous searches against the database of PFAM, NR and SwissProt. The genes with the 7 transmembrane (TM) domain were considered as the GPCR-associated family genes, that were classified into different function according the NR and SwissProt annotation. The P-values were obtained from Chi-square tests for the significant expansion gene function in the *R. philippinarum* and the total number of each function was used as background.

*R. philippinarum* and immune challenge

Manila clams were collected from a local commercial farm and transported to the laboratory. Animals were acclimated in aerated seawater (salinity 32, temperature 20 ± 1 °C) with 60 animals per tank (70 L, flat bottom) for 1 week prior to experimentation. After acclimation, the individuals in the experimental groups were challenged with 50 μl of immune stimulant [LPS 100 μg ml⁻¹, Poly (I:C) 100 μg ml⁻¹, and PGN 20 μg ml⁻¹, InvivoGen, USA] and the control group was injected with 50 μl PBS. Injected clams were returned to seawater tanks and four individuals from each group were randomly sampled at 0, 3, 6, 12, 24, 48, 72 and 96 h. Gill samples were collected by dissection and stored in RNAstore (Tiangen, China) until analysis by quantitative real-time RT-PCR.

Real-time RT-PCR confirmation of Illumina sequencing data

To validate our Illumina sequencing data, differentially expressed genes were selected for quantitative RT–PCR (RT–qPCR) analysis, using the same treated groups and control groups RNA samples as were used for transcriptome profiling. The primers were designed with the Primer5 software (Premier Biosoft International). The β-actin gene was used as the internal control for the qPCR analysis. RNA (500 ng) from each sample was measured and treated with RQ1 RNase-Free DNase (Promega) to remove any genomic DNA. The cDNA was synthesized from the treated RNA using a reverse transcriptase reagent kit (PrimeScript™ RT reagent Kit, Takara). The qPCR was performed with SYBR Premix Ex Taq II (Takara). The reactions were carried out in a total volume of 25 μl containing 2.5 μl of diluted cDNA, 2.5 μl of each primer, and 12.5 μl of SYBR Green PCR Master Mix, with the following cycling profile: 95 °C for 15 min for polymerase activation, followed by 40 cycles at 95 °C for 15s, at 55 °C for 30s, and at 70 °C for 30s. Each
sample was processed in triplicate in the Roche LightCycler 480 Real-Time PCR System (Roche).
All data were analyzed using the $2^{-\Delta\Delta C_T}$ method.

**Data availability**

This genome project has been registered in NCBI under the BioProject accession PRJNA479743. The *R. philippinarum* assembly has been deposited with the accession number QUSP00000000 under PRJNA479743. All sequence data of *R. philippinarum* genome have been deposited in NCBI Sequence Read Archive under the accession numbers of SRR7716263-SRR7716297.
Supplemental References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-402

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al., (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 25, 25-29

Bairoch, A., Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 28, 45-48

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573-580

Birney, E., Clamp, M., and Durbin, R. (2004). Genewise and genomewise. Genome Res. 14, 988-995.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17, 540-552.

Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. Chapter 4, Unit 4

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792-1797

English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C., et al., (2012). Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS One. 7, e47748

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury R, Zeng Q, et al., (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol. 29, 644-652

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., Bateman, A. (2005). Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 33, D121-124.
Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, CR., Wortman, J.R., et al., (2008). Automated eukaryotic gene structure annotation using evidence modeler and the program to assemble spliced alignments. Genome Biol. 9, R7

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith Jr, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D. Salzberg, S.L., White, O. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res, 31, 5654-5666

Hedges, S.B., Marin, J., Suleski, M., Paymer, M., Kumar, S. (2015). Tree of life reveals clock-like speciation and diversification. Mol Biol Evol. 32(4), 835-845

Huang, S., Chen, Z., Huang, G., Yu, T., Yang, P., Li, J., Fu, Y., Yuan, S., Chen, S., Xu, A. (2012). HaploMerger: Reconstructing allelic relationships for polymorphic diploid genome assemblies. Genome Res. 22, 1581-1588

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Yoshitoshi, H., Ogura, M., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H., et al., (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. 24, 1384-1395

Kanehisa, M., Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27-30

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. Genome Res. 12, 656-664

Korf, I. (2004). Gene finding in novel genomes. BMC Bioinformatics 5, 59

Li, L, Stoeckert CJ Jr., Roos DS. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Research 13, 2178-2189.

Li, H., Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-60

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J., Wang, J. (2010). De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 20, 265–272.

Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., Huang, Q., Cai, Q., Li, B., Bai, Y., Zhang, Z., et al., (2010). The sequence and de novo assembly of the giant panda genome. Nature. 463, 311-317.
Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., Li, Z., Chen, Y., Mu, D., Fan, W. (2013). Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. arXiv preprint arXiv, 1308.2012.

Lowe, T.M. & Eddy, S.R. (1997). tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucleic Acids Res. 25, 955-964

Majoros, W. H., Pertea, M., Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open-source ab initio eukaryotic gene-finders. Bioinformatics 20, 2878-2879

Mistry J., Finn R. D., Eddy S. R., Bateman A., Punta M. (2013). Challenges in Homology Search: HMMER3 and Convergent Evolution of Coiled-Coil Regions. Nucleic Acids Research, 41, e121

Mulder, N., Apweiker, R. (2007). InterPro and InterProScan: tools for protein sequence classification and comparison. Methods Mol Biol. 396, 59-70

Mun, S., Kim, Y.-J., Markkandan, K., Shin, W., Oh, S., Woo, J., et al. 2017. The whole-genome and transcriptome of the manila clam (Ruditapes philippinarum). Genome Biol Evol 9(6):1487-1498.

Nawrocki, E.P., Kolbe, D.L., Eddy, S.R. (2009). Infernal 1.0: inference of RNA alignments. Bioinformatics 25, 1335-7

Parra, G., Blanco, E., Guigo, R. (2000). GeneID in Drosophila. Genome Res. 10, 511-515

Parra, G., Bradnam, K., Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 23, 1061-7

Price, A.L., Jones, N.C., Pevzner, P.A. (2005). De novo identification of repeat families in large genomes. Bioinformatics. 21 Suppl 1, 351-358

Salamov, A.A., Solovyev, V.V. (2000). Ab initio gene finding in Drosophila genomic DNA. Genome Res. 10, 516-522.

Simakov, O., Kawashima, T., Marlétaz, F., Jenkins, J., Koyanagi, R., Mitros, T., Hisata, K., Bredeson, J., Shoguchi, E., Gyoja, F., Yue, J.X., Chen, Y.C., Freeman, R.M. Jr, Sasaki A. (2015). Hemichordate genomes and deuterostome origins. Nature. 527(7579), 459-465.

Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31, 3210–3212.
Smit, A. F. A., Hubley, R., Green, P. RepeatMasker at <http://www.repeatmasker.org/RepeatModeler.html>.

Stamatakis, A. (2006). RAxML VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690

Stamatakis, A., Hoover, P., Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57, 758–771

Stanke, M., Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19, i215–225

Stanke, M., Schoffmann, O., Morgenstern, B., Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics 7, 62

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511

Trapnell, C., Pachter, L., Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105-1111

Xu, Z., Wang, H. (2007). Ltr_finder: an efficient tool for the prediction of full-length ltr retrotransposons. Nucl. Acids Res. 35, W265-268

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24, 1586–1591

Yu, X.J., Zheng, H.K., Wang, J., Wang, W., Su, B. (2006). Detecting lineage-specific adaptive evolution of brain-expressed genes in human using rhesus macaque as outgroup. Genomics 88, 745-751

Zhang, J., Nielsen, R., Yang, Z. (2005). Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level. Mol Biol Evol. 22, 2472-2479
Table S1 Statistics for K-mer analysis of *R. philippinarum* genome. Related to Figure 1.

| K-mer | K-mer Number | K-mer Depth | Genome Size (Mb) | Heterozygous Ratio (%) |
|-------|---------------|-------------|------------------|------------------------|
| 17    | 50,054,797,752 | 38          | 1,317.23         | 2.00                   |

Table S2 Genome sequencing of *R. philippinarum*. Related to Figure 1.

| Platform | Insert size | Read length (bp) | Raw data (G) | Sequence coverage |
|----------|-------------|------------------|--------------|------------------|
| Illumina | 230 bp      | 125              | 91.5         | 69.48            |
|          | 300 bp      | 125              | 95.7         | 72.67            |
|          | 500 bp      | 125              | 98.9         | 75.90            |
|          | 2 Kbp       | 125              | 57.9         | 43.96            |
|          | 5 Kbp       | 125              | 39.6         | 30.07            |
|          | 10 Kbp      | 125              | 13.1         | 9.95             |
|          | 20 Kbp      | 125              | 24.5         | 18.60            |
|          | Total       |                  | 421.2        | 320.63           |

Note: The estimated genome size is ~1.32G, and sequence coverage = Total clean bases/1.32G.

Table S3 Statistics of the assembly of the *R. philippinarum* genome. Related to Figure 1.

| Sample ID | Length | Number |
|-----------|--------|--------|
|           | Contig (bp) | Scaffold (>100 bp) | Contig (bp) | Scaffold (>100 bp) |
| Total     | 1,062,832,364 | 1,122,973,377 | 88,274 | 30,811 |
| Max       | 249,659 | 2,001,120 | - | - |
### Table S4 The number of SNPs and genome annotation characterization of wild and inbred *R. philippinarum*. Related to Figure 1.

| Sample  | Upstream/Downstream | Exonic | Intergenic |
|---------|---------------------|--------|------------|
|         |                     | Stop gain | Stop loss | Synonymous | Non-synonymous | Intronic | Splicing | Downstream | Upstream/Downstream | Intergenic | Het rate (%) | Total        |
| Wild    | 737,992             | 6,915   | 1,070     | 752,89     | 483,38       | 9,463,9   | 2,290     | 882,366    | 21,459,38          | 1.690      | 33,830       |
| Inbred  | 266,777             | 1,268   | 444       | 365,18     | 175,82       | 3,488,4   | 870       | 303,942    | 7,061,957          | 1.038      | 11,758       |

Note: Only the length of scaffold greater than 100 bp was counted. N50 refers to the length of sequence equal to or greater than the half of total sequences length.

### Table S5 Statistics of genome content. The table shows GC content of *R. philippinarum* genome. Related to Figure 1.

| Number (bp) | % of genome |
|-------------|-------------|
| A           | 361,902,804 | 32.23       |
| T           | 361,997,214 | 32.24       |
C 169,455,540 15.09
G 169,476,806 15.09
N 60,141,013 5.36
GC 367,211,996 31.89
Total (bp) 1,122,973,377 100

Note: GC content of the genome without N.

Table S6 Assessment the genome coverage rate using raw data. Related to Figure 1.

| Sample ID      | Percentage |
|---------------|------------|
| Reads Mapping rate (%) | 93.32      |
| Average sequencing depth | 137.3      |
| Coverage (%)   | 99.83      |
| Genome Coverage at least 4X (%) | 99.55      |
|               | Coverage at least 10X (%) | 99.08      |
|               | Coverage at least 20X (%) | 98.27      |

Table S7 Assessment the gene coverage rate using EST data. Related to Figure 1.

| Dataset | Number | Total length (bp) | Sequences covered by assembly | with >90% sequence in one scaffold | with >50% sequence in one scaffold |
|---------|--------|-------------------|-------------------------------|-----------------------------------|-----------------------------------|
|         | Number | Percent (%)       | Number                        | Percent (%)                       | Number                           | Percent (%)                       |
| >200 bp | 44,321 | 34,136,266        | 99.75                         | 36,129                            | 81.52                            |
| >500 bp | 18,197 | 26,149,430        | 99.89                         | 14,560                            | 80.01                            |
| >1 Kbp  | 9,056  | 19,790,574        | 99.98                         | 7,066                             | 78.03                            |
| >2 Kbp  | 3,626  | 12,154,398        | 99.97                         | 2,807                             | 77.41                            |
Table S8 CEGMA results of *R. philippinarum* genome. Related to Figure 1.

| Species       | Complete                  | Complete + partial          |
|---------------|---------------------------|-----------------------------|
|               | Prots | % completeness | Prots     | % completeness |
| *R. philippinarum* | 214   | 86.29%         | 233       | 93.95%         |

Note: CEGMA (Core Eukaryotic Genes Mapping Approach) defined the number of 248 ultra-conserved CEGs that occur in a wide range of eukaryotes. A protein is classified as complete if the alignment of the predicted protein to the HMM profile represents at least 70% of the original KOG domain, otherwise is classified as partial.

Table S9 BUSCO (Benchmarking Universal Single-Copy Orthologs) results of *R. philippinarum* genome. Related to Figure 1.

| Species       | Size     | BUSCO notation assessment results |
|---------------|----------|-----------------------------------|
| *R. philippinarum* | 1123Mbp  | C:92.2%[S:90.3%,D:1.9%],F:1.6%,M:6.2%,n:978 |

Note: The recovered matches were classified as ‘complete’ (C) if their lengths were within the expectation of the BUSCO profile match lengths. If these were found more than once they were classified as ‘duplicated’ (D). The matches that were only partially recovered were classified as ‘fragmented’ (F), and BUSCO groups for which there were no matches that pass the tests of orthology were classified as ‘missing’ (M). The number of total BUSCO groups (n) was 843.

Table S10 The prediction of repeats elements in *R. philippinarum* genome. Related to Figure 1.

| Type                      | Repeat Size (bp) | Percent (%) |
|---------------------------|------------------|-------------|
| TRF                       | 69,164,254       | 6.16        |
| RepeatMasker              | 389,644,949      | 34.70       |
| RepeatProteinMask         | 26,094,608       | 2.32        |
| Total                     | 430,065,124      | 38.30       |

Note: Repeats in *R. philippinarum* genome were consisted with tendam repeats and interpersed repeats. The TRF (Tandem repeats finder) was used to predict tandem repeats. RepeatMasker and RepeatProteinMask were used to predict interpersed repeats.
Table S11 Categories of TEs predicted in *R. philippinarum* genome. Related to Figure 1.

| Category          | Repeatmasker Length (bp) | % in Genome | TE Proteins Length (bp) | % in Genome | Combined TEs Length (bp) | % in Genome |
|-------------------|--------------------------|-------------|-------------------------|-------------|-------------------------|-------------|
| DNA transposon    |                          |             |                         |             |                         |             |
| DNA              | 212,422,142              | 18.916      | 6,059,255               | 0.54        | 216,075,767             | 19.241      |
| LINE             | 53,578,378               | 4.771       | 15,019,016              | 1.337       | 61,586,774              | 5.484       |
| Retrotransposon   |                          |             |                         |             |                         |             |
| LTR              | 104,648,756              | 9.319       | 5,050,079               | 0.45        | 107,571,925             | 9.579       |
| SINE             | 3,507,191                | 0.312       | 0                       | 0           | 3,507,191               | 0.312       |
| Other            |                          |             |                         |             |                         |             |
| Other*           | 268                      | 0.000024    | 0                       | 0           | 268                     | 0.000024    |
| Unknown*         | 60,134,172               | 5.355       | 0                       | 0           | 60,134,172             | 5.355       |
| Total            | 389,644,949              | 34.698      | 26,094,608              | 2.324       | 402,086,301             | 35.806      |

Note: The genomic abundance of TEs were predicted by Repeatmasker through ab-initio method and homology based method. The TE proteins were predicted based on TE protein database by RepeatProteinMask. The final results of TEs in *R. philippinarum* genome were integrated of both methods without redundancy. "Other" refer to the repeats that can be classified by RepeatMasker, but not included by the classes above; "Unknown" refer to the repeats that can’t be classified by RepeatMasker.

Table S12 General statistics of predicted protein-coding genes. Related to Figure 1.

| Gene set          | Number | Average transcript length (bp) | Average CDS length (bp) | Average exons per gene | Average exon length (bp) | Average intron length (bp) |
|-------------------|--------|-------------------------------|-------------------------|------------------------|--------------------------|---------------------------|
| De novo*          | Augustus | 37,212 | 8970.26                  | 1318.06               | 6.09                     | 216.40                   | 1503.11                  |
|                   | Genscan | 33,272 | 17857.67                | 1311.80               | 5.45                     | 240.77                   | 3719.46                  |
|                   | GlimmerHMM | 101,946 | 9113.13               | 590.15                | 3.68                     | 160.29                   | 3178.12                  |
|                   | SNAP    | 93,644 | 10462.10               | 760.63                | 5.74                     | 132.41                   | 2044.87                  |
|                   | Geneid  | 51,889 | 9438.01                | 791.52                | 3.19                     | 248.49                   | 3956.73                  |
### Table S13 General statistics of protein-coding genes of homolog species. Related to Figure 1.

| Species                        | Number | Average transcript length (bp) | Average CDS length (bp) | Average exons per gene | Average exon length (bp) | Average intron length (bp) |
|--------------------------------|--------|-------------------------------|-------------------------|------------------------|--------------------------|---------------------------|
| *Ruditapes philippinarum*      | 27,652 | 12875.81                      | 1462.72                 | 7.30                   | 200.30                   | 1697.30                   |
| *Crassostrea gigas*            | 26,089 | 8072.40                       | 1344.93                 | 7.16                   | 187.75                   | 1091.49                   |
| *Capitella teleta*             | 31,440 | 2652.33                       | 1074.05                 | 5.37                   | 200.03                   | 361.20                    |

Note: # Statistics calculated from the gene set predicted form each method. $ Statistics calculated from the gene set predicted by homolog proteins from each species.
* Statistics calculated from the *R.philippinarum* genome.
| Species                | Number | Average Length (bp) | Total Length (bp) | % of Genome |
|------------------------|--------|---------------------|-------------------|-------------|
| *Helobdella robusta*   | 23,350 | 3762.63             | 1127.79           | 5.99        |
| *Lingula anatina*      | 32,828 | 6526.03             | 1414.25           | 7.51        |
| *Lottia gigantea*      | 23,205 | 5045.43             | 1150.31           | 6.01        |
| *Octopus bimaculoides* | 30,845 | 8553.10             | 779.28            | 3.59        |

Table S14 General statistics of mapping rate to functional database of protein-coding genes. Related to Figure 1.

| Type          | Number | Average Length (bp) | Total Length (bp) | % of Genome |
|---------------|--------|---------------------|-------------------|-------------|
| miRNA         | 600    | 109.16              | 65,497            | 0.0058      |
| tRNA          | 3,788  | 78.15               | 296,030           | 0.0264      |

Table S15 General statistics of non-coding RNA of the genome. Related to Figure 1.
| GO ID   | GO Term                              | GO Class | P-value  | Adjusted P-value | Gene Number |
|--------|--------------------------------------|----------|----------|------------------|-------------|
| GO:0005215 | transporter activity                  | MF       | 1.53E-67 | 3.45E-65         | 58          |
| GO:0051179 | localization                          | BP       | 8.56E-57 | 6.45E-55         | 69          |
| GO:0044765 | single-organism transport              | BP       | 3.75E-54 | 2.12E-52         | 58          |
| GO:0022804 | active transmembrane transporter activity | MF     | 1.08E-48 | 4.90E-47         | 41          |
| GO:0016887 | ATPase activity                       | MF       | 1.99E-38 | 6.43E-37         | 36          |
| GO:0030286 | dynein complex                        | CC       | 2.40E-37 | 5.43E-36         | 19          |
| GO:0055524 | ATP binding                           | MF       | 1.09E-32 | 2.24E-31         | 39          |
| GO:0055085 | transmembrane transport               | BP       | 6.12E-31 | 9.88E-30         | 32          |
| GO:0037777 | microtubule motor activity            | MF       | 4.91E-30 | 7.39E-29         | 22          |
| GO:007018  | microtubule-based movement            | BP       | 4.64E-29 | 6.27E-28         | 22          |
| GO:0042626 | ATPase activity, coupled to transmembrane movement of substances | MF | 5.27E-29 | 6.27E-28         | 20          |
| GO:0017111 | nucleoside-triphasphatase activity    | MF       | 1.55E-28 | 1.67E-27         | 42          |
| GO:0005328 | neurotransmitter:sodium symporter activity | MF | 8.48E-28 | 7.10E-27         | 21          |

Table S16 GO enrichment of expanded gene families in Rph compared with Hrph. Related to Table 1.
| GO          | Term                                      | Type | FDR_1 | FDR_2 | Rank |
|-------------|-------------------------------------------|------|-------|-------|------|
| GO:0006836  | neurotransmitter transport                | BP   | 1.55E-27 | 1.25E-26 | 21   |
| GO:0000166  | nucleotide binding                        | MF   | 9.87E-25 | 7.19E-24 | 45   |
| GO:0044699  | single-organism process                   | BP   | 8.68E-24 | 5.45E-23 | 90   |
| GO:0001539  | cilia or flagellum-dependent cell motility| BP   | 2.46E-23 | 1.46E-22 | 11   |
| GO:0005858  | axonemal dynein complex                   | CC   | 3.37E-21 | 1.65E-20 | 10   |
| GO:0016020  | membrane                                  | CC   | 4.96E-18 | 1.65E-17 | 68   |
| GO:0007160  | cell-matrix adhesion                      | BP   | 1.55E-16 | 4.92E-15 | 11   |
| GO:0022892  | substrate-specific transporter activity    | MF   | 8.96E-16 | 2.81E-15 | 26   |
| GO:0016787  | hydrolase activity                        | MF   | 8.53E-15 | 2.54E-14 | 43   |
| GO:0044763  | single-organism cellular process           | BP   | 1.49E-14 | 4.31E-14 | 64   |
| GO:0016021  | integral component of membrane            | CC   | 2.01E-13 | 5.67E-13 | 46   |
| GO:0005509  | calcium ion binding                       | MF   | 5.73E-11 | 1.52E-10 | 21   |
| GO:0009190  | cyclic nucleotide biosynthetic process    | BP   | 4.21E-10 | 1.06E-09 | 9    |
| GO:0016849  | phosphorus-oxygen lyase activity          | MF   | 4.21E-10 | 1.06E-09 | 9    |
| GO:0043167  | ion binding                               | MF   | 2.36E-09 | 5.62E-09 | 60   |
| GO:0005044  | scavenger receptor activity               | MF   | 2.38E-08 | 5.43E-08 | 9    |
| GO:0009987  | cellular process                           | BP   | 4.78E-06 | 1.01E-05 | 65   |
| GO:0005319  | lipid transporter activity                 | MF   | 6.20E-06 | 1.30E-05 | 5    |
| GO:0006869  | lipid transport                            | BP   | 7.12E-06 | 1.46E-05 | 5    |
| GO:0004713  | protein tyrosine kinase activity           | MF   | 8.90E-04 | 1.69E-03 | 9    |
| GO:0003824  | catalytic activity                         | MF   | 2.54E-03 | 4.62E-03 | 53   |
| GO:0035556  | intracellular signal transduction          | BP   | 5.24E-03 | 9.33E-03 | 9    |
| GO:0006468  | protein phosphorylation                    | BP   | 5.88E-03 | 1.04E-02 | 9    |
| GO:0008281  | sulfonylurea receptor activity             | MF   | 7.36E-03 | 1.29E-02 | 1    |
Table S17 GO enrichment of unique gene families in *R. philippinarum* compared with four other Mollusca species (*C. gigas, L. gigantea, P. fucata* and *P. yessoensis*). Related to Figure 1.

Table S18 KEGG enrichment of unique gene families in *R. philippinarum* compared with four other Mollusca species (*C. gigas, L. gigantea, P. fucata* and *P. yessoensis*). Related to Figure 1.

| Map ID    | MapTitle                                  | P-value      | Adjusted P-value | Gene Number |
|-----------|-------------------------------------------|--------------|------------------|-------------|
| map04080  | Neuroactive ligand-receptor interaction    | 3.21E-55     | 8.98E-53         | 321         |
| map04512  | ECM-receptor interaction                   | 2.24E-16     | 3.14E-14         | 144         |
| map00513  | Various types of N-glycan biosynthesis    | 6.84E-12     | 6.39E-10         | 74          |
| map05206  | MicroRNAs in cancer                       | 1.11E-11     | 7.54E-10         | 115         |
| map04510  | Focal adhesion                            | 1.35E-11     | 7.54E-10         | 191         |
| map04151  | PI3K-Akt signaling pathway                | 6.12E-09     | 2.63E-07         | 170         |
| map04540  | Gap junction                              | 6.57E-09     | 2.63E-07         | 64          |
| map05033  | Nicotine addiction                        | 1.01E-08     | 3.54E-07         | 41          |
| map00601  | Glycosphingolipid biosynthesis - lacto and neolacto series | 1.68E-08 | 5.24E-07 | 35 |
| map04725  | Cholinergic synapse                        | 5.62E-08     | 1.57E-06         | 66          |
| map04514  | Cell adhesion molecules (CAMs)             | 7.89E-08     | 2.01E-06         | 53          |
| map05130  | Pathogenic                                | 1.42E-07     | 3.32E-06         | 53          |
| Map ID   | Pathway Description                                           | FDR       | Log2FoldChange | Rank |
|---------|--------------------------------------------------------------|-----------|---------------|------|
| map05204| Escherichia coli infection                                  | 1.77E-06  | 3.31E-05      | 45   |
| map00906| Chemical carcinogenesis                                       | 3.05E-06  | 5.03E-05      | 11   |
| map04672| Intestinal immune network for IgA production                | 2.74E-05  | 3.48E-04      | 8    |
| map04145| Phagosome                                                    | 6.36E-05  | 6.59E-04      | 102  |
| map04974| Protein digestion and absorption                            | 3.69E-04  | 2.92E-03      | 85   |
| map05145| Toxoplasmosis                                                | 1.19E-03  | 7.89E-03      | 72   |
| map05020| Prion diseases                                               | 1.21E-03  | 7.89E-03      | 29   |
| map04913| Ovarian steroidogenesis                                      | 1.27E-03  | 7.92E-03      | 28   |
| map00100| Steroid biosynthesis                                         | 1.47E-03  | 8.78E-03      | 18   |
| map04726| Serotonergic synapse                                         | 1.72E-03  | 9.83E-03      | 46   |
| map04610| Complement and coagulation cascades                          | 2.07E-03  | 1.09E-02      | 52   |
| map04020| Calcium signaling pathway                                   | 3.29E-03  | 1.56E-02      | 104  |
| map04115| p53 signaling pathway                                       | 5.73E-03  | 2.40E-02      | 39   |
| map00533| Glycosaminoglycan biosynthesis - keratan                    | 5.93E-03  | 2.44E-02      | 20   |
| Map ID   | MapTitle                                | P-value   | Adjusted P-value | Gene Number |
|---------|-----------------------------------------|-----------|------------------|-------------|
| map00534| sulfate Glycosaminoglycan biosynthesis - heparan sulfate / heparin | 8.90E-03  | 3.32E-02         | 15          |
| map00140| Steroid hormone biosynthesis            | 9.19E-03  | 3.37E-02         | 26          |
| map05222| Small cell lung cancer                  | 9.90E-03  | 3.51E-02         | 58          |
| map04210| Apoptosis                               | 1.18E-02  | 4.04E-02         | 49          |

Table S19 GO enrichment of expanded gene families in *R. philippinarum* compared with four other Mollusca species (*C. gigas, L. gigantea, P. fucata* and *P. yessoensis*). Related to Figure 1.

Table S20 KEGG enrichment of expanded gene families in *R. philippinarum* compared with four other Mollusca species (*C. gigas, L. gigantea, P. fucata* and *P. yessoensis*). Related to Figure 1.
| Map ID     | Pathway Description                                                                 | E1    | E2    | P   |
|------------|-------------------------------------------------------------------------------------|-------|-------|-----|
| map05146   | Amoebiasis                                                                          | 7.03E-19 | 3.49E-17 | 136 |
| map04610   | Complement and coagulation cascades                                                  | 1.41E-14 | 6.02E-13 | 95  |
| map05322   | Systemic lupus erythematosus                                                        | 4.23E-12 | 1.58E-10 | 74  |
| map05150   | Staphylococcus aureus infection                                                     | 6.21E-12 | 2.05E-10 | 73  |
| map04514   | Cell adhesion molecules (CAMs)                                                      | 4.71E-10 | 1.40E-08 | 71  |
| map04725   | Cholinergic synapse                                                                 | 5.70E-10 | 1.50E-08 | 88  |
| map05020   | Prion diseases                                                                      | 6.04E-10 | 1.50E-08 | 49  |
| map05204   | Chemical carcinogenesis                                                             | 1.31E-08 | 2.68E-07 | 61  |
| map04611   | Platelet activation                                                                | 2.44E-08 | 4.27E-07 | 130 |
| map04145   | Phagosome                                                                           | 9.81E-08 | 1.33E-06 | 144 |
| map00513   | Various types of N-glycan biosynthesis                                              | 2.45E-07 | 2.92E-06 | 82  |
| map04640   | Hematopoietic cell lineage                                                          | 9.99E-07 | 1.03E-05 | 38  |
| map03450   | Non-homologous end-joining                                                          | 1.64E-06 | 1.57E-05 | 30  |
| map05206   | MicroRNAs in cancer                                                                 | 2.38E-06 | 2.22E-05 | 129 |
| Map       | Description                                                                 | FDR 1 | FDR 2     | Count |
|-----------|------------------------------------------------------------------------------|-------|-----------|-------|
| map05033  | Nicotine addiction                                                          | 3.32E-06 | 3.00E-05 | 46    |
| map00906  | Carotenoid biosynthesis                                                      | 7.85E-06 | 6.16E-05 | 12    |
| map05222  | Small cell lung cancer                                                       | 1.00E-05 | 7.65E-05 | 88    |
| map04020  | Calcium signaling pathway                                                   | 3.34E-05 | 2.17E-04 | 148   |
| map00532  | Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate      | 6.93E-05 | 4.17E-04 | 45    |
| map05144  | Malaria                                                                       | 1.18E-04 | 6.77E-04 | 38    |
| map05130  | Pathogenic Escherichia coli infection                                        | 5.13E-04 | 2.22E-03 | 57    |
| map00140  | Steroid hormone biosynthesis Cytosolic                                       | 1.04E-03 | 3.73E-03 | 36    |
| map04623  | DNA-sensing pathway                                                         | 2.51E-03 | 8.39E-03 | 38    |
| map05145  | Toxoplasmosis Intestinal immune network for IgA production                   | 2.72E-03 | 8.77E-03 | 92    |
| map04672  | Cell cycle                                                                   | 2.95E-03 | 9.26E-03 | 74    |
| map05142  | Chagas disease                                                               | 3.60E-03 | 1.12E-02 | 54    |
| GO ID    | GO Term                          | GO Class | P-value | Adjusted P-value | Gene Number |
|----------|----------------------------------|----------|---------|-----------------|-------------|
| GO:0007155 | cell adhesion                    | BP       | 1.71E-06 | 1.89E-04        | 7           |
| GO:0000786 | nucleosome                       | CC       | 1.30E-05 | 4.66E-04        | 3           |
| GO:006334  | nucleosome assembly              | BP       | 2.18E-05 | 4.66E-04        | 3           |
| GO:0004527 | exonuclease activity             | MF       | 4.96E-05 | 6.86E-04        | 3           |

Table S21 GO enrichment of contracted gene families in *R. philippinarum* compared with four other Mollusca species (*C. gigas, L. gigantea, P. fucata* and *P. yessoensis*). Related to Figure 1.
| Map ID    | MapTitle                                | P-value    | Adjusted P-value | Gene Number |
|----------|-----------------------------------------|------------|------------------|-------------|
| map04512 | ECM-receptor interaction                | 4.29E-08   | 2.56E-06         | 11          |
| map04510 | Focal adhesion                          | 5.56E-08   | 2.56E-06         | 13          |
| map04151 | PI3K-Akt signaling pathway              | 1.89E-06   | 5.79E-05         | 11          |
| map05146 | Amoebiasis                              | 2.27E-05   | 5.22E-04         | 7           |
| map04974 | Protein digestion and absorption        | 4.85E-04   | 6.70E-03         | 6           |
| map05200 | Pathways in cancer                      | 5.22E-04   | 6.70E-03         | 7           |
| map00860 | Chlorophyll metabolism                  | 5.55E-04   | 6.70E-03         | 3           |
| map04611 | Platelet activation                     | 5.82E-04   | 6.70E-03         | 6           |
| map00053 | Ascorbate and aldolase metabolism      | 7.93E-04   | 8.02E-03         | 3           |

Table S22 KEGG enrichment of contracted gene families in *R. philippinarum* compared with four other Mollusca species (*C. gigas, L. gigantea, P. fucata* and *P. yessoensis*). Related to Figure 1.
| Map ID      | Description                                                      | p-value (FDR) | p-value (BH) | Rank |
|------------|------------------------------------------------------------------|---------------|--------------|------|
| map05221   | Acute myeloid leukemia                                           | 8.72E-04      | 8.02E-03     | 3    |
| map05322   | Systemic lupus erythematosus                                     | 1.11E-03      | 9.32E-03     | 4    |
| map04120   | Ubiquitin mediated proteolysis                                   | 1.54E-03      | 1.10E-02     | 5    |
| map00983   | Drug metabolism - other enzymes                                  | 1.62E-03      | 1.10E-02     | 3    |
| map00040   | Pentose and glucurionate interconversions                        | 1.68E-03      | 1.10E-02     | 3    |
| map05202   | Transcriptional misregulation in cancer                          | 1.79E-03      | 1.10E-02     | 4    |
| map05164   | Influenza A                                                      | 2.10E-03      | 1.21E-02     | 5    |
| map00140   | Steroid hormone biosynthesis                                    | 2.60E-03      | 1.41E-02     | 3    |
| map00980   | Metabolism of xenobiotics by cytochrome P450                     | 2.76E-03      | 1.41E-02     | 3    |
| map04610   | Complement and coagulation cascades                              | 2.96E-03      | 1.43E-02     | 4    |
| map05034   | Alcoholism                                                       | 3.12E-03      | 1.43E-02     | 4    |
| map00830   | Retinol metabolism                                              | 3.58E-03      | 1.57E-02     | 3    |
| map05168   | Herpes simplex                                                  | 4.02E-03      | 1.68E-02     | 4    |
| GO ID     | GO Term                                                                 | GO Class | P-value     | Adjusted P-value | Gene Number |
|-----------|--------------------------------------------------------------------------|----------|-------------|------------------|-------------|
| GO:0046034| ATP metabolic process hydrolase activity, acting on acid anhydrides, catalyzing transmembrane | BP       | 7.98E-09    | 2.43E-06         | 4           |
| GO:0016820| movement of substances                                                   | MF       | 2.48E-08    | 2.43E-06         | 5           |
| GO:1902600| hydrogen ion transmembrane transport proton-transporting two-sector ATPase | BP       | 2.95E-08    | 2.43E-06         | 4           |
| GO:0033178| complex, catalytic domain                                                | CC       | 2.25E-07    | 7.72E-06         | 3           |

Table S23 GO enrichment of positively selected genes in *R. philippinarum* compared with four other Mollusca species (*C. gigas, L. gigantea, P. fucata* and *P. yessoensis*). Related to Figure 1.
| GO Term             | Description                                           | BP   | MF   | Count |
|---------------------|-------------------------------------------------------|------|------|-------|
| GO:0009205          | purine ribonucleoside triphosphate metabolic process  | 4.57E-07 | 1.15E-05 | 6     |
| GO:0046128          | purine ribonucleoside metabolic process               | 5.82E-07 | 1.25E-05 | 6     |
| GO:0009150          | purine ribonucleotide metabolic process               | 1.14E-06 | 1.68E-05 | 6     |
| GO:0015991          | ATP hydrolysis coupled proton hydrolase activity, acting on acid | 1.21E-06 | 1.68E-05 | 3     |
| GO:0016817          | anhydrides                                            | 4.12E-06 | 4.55E-05 | 8     |
| GO:0006812          | cation transport                                      | 4.70E-05 | 4.36E-04 | 5     |
| GO:0006810          | transport                                             | 1.10E-04 | 9.84E-04 | 10    |
| GO:0008060          | ARF GTPase activator activity                         | 1.33E-04 | 1.04E-03 | 2     |
| GO:0032312          | regulation of ARF GTPase activity                     | 1.33E-04 | 1.04E-03 | 2     |
| GO:1901363          | heterocyclic compound binding                         | 1.48E-04 | 1.10E-03 | 12    |
| GO:0050790          | regulation of catalytic activity                      | 1.50E-04 | 1.10E-03 | 3     |
| GO:0097159          | organic cyclic compound binding                        | 1.51E-04 | 1.10E-03 | 12    |
| GO:0019220          | process                                               | 1.98E-04 | 1.32E-03 | 3     |
| GO:0000166          | nucleotide binding                                    | 2.00E-04 | 1.32E-03 | 9     |
| GO:0006139          | metabolic process                                     | 3.16E-04 | 1.93E-03 | 8     |
| GO:0006796          | metabolic process                                     | 3.80E-04 | 2.24E-03 | 7     |
| GO:0005524          | ATP binding                                           | 3.94E-04 | 2.26E-03 | 7     |
| GO:0019829          | cation-transporting ATPase activity                   | 4.61E-04 | 2.31E-03 | 2     |
| GO ID          | Term                           | Domain | Start | End   | Width |
|---------------|--------------------------------|--------|-------|-------|-------|
| GO:0035639    | purine ribonucleoside triphosphate binding | MF     | 4.66E-04 | 2.31E-03 | 8     |
| GO:0032550    | purine ribonucleoside binding | MF     | 4.77E-04 | 2.31E-03 | 8     |
| GO:0032555    | purine ribonucleotide binding | MF     | 4.77E-04 | 2.31E-03 | 8     |
| GO:0030117    | membrane coat | CC     | 7.50E-04 | 3.26E-03 | 2     |
| GO:0015031    | protein transport | BP     | 7.64E-04 | 3.28E-03 | 4     |
| GO:0016787    | hydrolase activity | MF     | 1.27E-03 | 5.25E-03 | 9     |
| GO:0051017    | actin filament bundle assembly | BP     | 1.46E-03 | 5.69E-03 | 1     |
| GO:0051764    | actin crosslink formation | BP     | 1.46E-03 | 5.69E-03 | 1     |
| GO:0043234    | protein complex | CC     | 1.51E-03 | 5.83E-03 | 5     |
| GO:0006886    | intracellular protein transport | BP     | 2.04E-03 | 7.63E-03 | 3     |
| GO:0017111    | nucleoside-triphosphatase activity | MF     | 2.34E-03 | 7.86E-03 | 5     |
| GO:0004127    | cytidylate kinase activity | MF     | 2.92E-03 | 9.17E-03 | 1     |
| GO:00010389   | regulation of G2/M transition of mitotic cell cycle | BP     | 4.37E-03 | 1.22E-02 | 1     |
| GO:0044237    | cellular metabolic process | BP     | 4.46E-03 | 1.22E-02 | 10    |
| GO:0004017    | adenylate kinase activity | MF     | 5.82E-03 | 1.55E-02 | 1     |
| GO:0016192    | vesicle-mediated transport proton-transporting ATP synthase activity, rotational mechanism | BP     | 5.99E-03 | 1.58E-02 | 2     |
| GO:0046933    | activity, rotational mechanism | MF     | 7.27E-03 | 1.88E-02 | 1     |
| GO:0005488    | binding | MF     | 1.05E-02 | 2.60E-02 | 19    |
| GO:0015986    | ATP synthesis coupled proton transport regulation of cyclin-dependent protein serine/threonine kinase activity | BP     | 1.31E-02 | 3.11E-02 | 1     |
| GO:0000079    | ATP synthesis coupled proton transport regulation of cyclin-dependent protein | BP     | 1.31E-02 | 3.11E-02 | 1     |
| GO:0019901    | protein kinase binding | MF     | 1.31E-02 | 3.11E-02 | 1     |
| GO:00030131   | clathrin adaptor complex | CC     | 1.74E-02 | 3.87E-02 | 1     |
Table S24 KEGG enrichment of positively selected genes in *R. philippinarum* compared with four other Mollusca species (*C. gigas, L. gigantea, P. fucata* and *P. yessoensis*). Related to Figure 1.

| Map ID   | MapTitle                          | P-value  | Adjusted P-value | Gene Number |
|----------|----------------------------------|----------|------------------|-------------|
| map00190 | Oxidative phosphorylation        | 9.72E-06 | 3.60E-04         | 4           |
| map04966 | Collecting duct acid secretion   | 2.68E-04 | 4.96E-03         | 2           |
| map04144 | Endocytosis                      | 1.37E-03 | 1.69E-02         | 4           |
| map04721 | Synaptic vesicle cycle           | 5.42E-03 | 3.98E-02         | 2           |
| map05110 | Vibrio cholerae infection        | 6.03E-03 | 3.98E-02         | 2           |
| map05323 | Rheumatoid arthritis             | 7.32E-03 | 3.98E-02         | 2           |
| map05010 | Alzheimer's disease              | 7.53E-03 | 3.98E-02         | 3           |
| map05120 | Epithelial cell signaling in     |          |                  |             |
|          | *Helicobacter pylori* infection  | 1.06E-02 | 4.92E-02         | 2           |

Table S25 Immune related GO enrichment in *R. philippinarum*. Related to Figure 3.

Table S26 Immune related KEGG enrichment in *R. philippinarum*. Related to Figure 3.

| Map ID   | MapTitle                          | P-value  | Adjusted P-value | Gene Number |
|----------|----------------------------------|----------|------------------|-------------|
| map04512 | ECM-receptor interaction          | 5.40E-58 | 1.25E-55         | 210         |
| map04151 | PI3K-Akt signaling pathway        | 6.11E-43 | 7.08E-41         | 250         |
| map04974 | Protein digestion and absorption | 7.74E-41 | 5.98E-39         | 156         |
| map04080 | Neuroactive ligand-receptor interaction | 4.35E-38 | 2.52E-36         | 298         |
| MapID     | Description                                                      | p-value1 | p-value2 | Count |
|-----------|------------------------------------------------------------------|----------|----------|-------|
| map05146  | Amoebiasis                                                       | 6.16E-34 | 2.86E-32 | 133   |
| map04510  | Focal adhesion                                                  | 8.31E-33 | 3.21E-31 | 247   |
| map05150  | Staphylococcus aureus infection                                  | 6.62E-21 | 2.19E-19 | 72    |
| map05020  | Prion diseases                                                  | 3.74E-18 | 1.08E-16 | 50    |
| map05142  | Chagas disease (American trypanosomiasis)                       | 6.74E-17 | 1.74E-15 | 69    |
| map04610  | Complement and coagulation cascades                             | 4.61E-15 | 1.07E-13 | 79    |
| map04145  | Phagosome                                                       | 1.19E-14 | 2.50E-13 | 133   |
| map04611  | Platelet activation                                             | 1.83E-14 | 3.53E-13 | 119   |
| map00513  | Various types of N-glycan biosynthesis                          | 2.99E-13 | 5.33E-12 | 78    |
| map04514  | Cell adhesion molecules (CAMs)                                   | 5.18E-13 | 8.58E-12 | 63    |
| map00601  | Glycosphingolipid biosynthesis - lacto and neolacto series     | 9.41E-13 | 1.43E-11 | 41    |
| map04640  | Hematopoietic cell lineage                                      | 9.83E-13 | 1.43E-11 | 39    |
| map00532  | Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate | 1.46E-12 | 1.99E-11 | 49    |
| map05322  | Systemic lupus erythematosus                                    | 4.51E-12 | 5.51E-11 | 61    |
| map00603  | Glycosphingolipid biosynthesis - globo series                   | 1.46E-11 | 1.69E-10 | 31    |
| Map       | Pathology                                                      | q-value (FDR) | FDR (p-value) | Rank |
|-----------|----------------------------------------------------------------|---------------|---------------|------|
| map05140  | Leishmaniasis                                                  | 2.24E-10      | 1.92E-09      | 36   |
| map05145  | Toxoplasmosis                                                  | 3.85E-09      | 3.08E-08      | 91   |
| map05222  | Small cell lung cancer                                         | 4.49E-09      | 3.36E-08      | 79   |
| map05133  | Pertussis                                                      | 9.39E-08      | 5.89E-07      | 86   |
| map04725  | Cholinergic synapse                                            | 6.66E-07      | 3.51E-06      | 65   |
| map05144  | Malaria                                                        | 6.99E-07      | 3.60E-06      | 35   |
| map05152  | Tuberculosis                                                   | 3.40E-06      | 1.61E-05      | 102  |
| map00533  | Glycosaminoglycan biosynthesis - keratan sulfate                | 5.13E-06      | 2.34E-05      | 26   |
| map05162  | Measles                                                        | 1.29E-05      | 5.46E-05      | 60   |
| map04210  | Apoptosis                                                      | 1.58E-05      | 6.45E-05      | 60   |
| map05321  | Inflammatory bowel disease (IBD)                               | 1.89E-05      | 7.32E-05      | 19   |
| map00960  | Tropane, piperidine and pyridine alkaloid biosynthesis         | 3.38E-05      | 0.000124404   | 12   |
| map03450  | Non-homologous end-joining                                    | 3.62E-05      | 0.000131329   | 23   |
| map04620  | Toll-like receptor signaling pathway                           | 4.50E-05      | 0.000160671   | 49   |
| map04064  | NF-kappa B signaling pathway                                   | 8.42E-05      | 0.000279211   | 49   |
| map00906  | Carotenoid biosynthesis                                        | 0.000343364   | 0.000995755   | 9    |
| map00950  | Isoquinoline alkaloid biosynthesis                             | 0.000587249   | 0.001621925   | 18   |
Sphingolipid metabolism  
Betalain biosynthesis  
Nicotine addiction  
MicroRNAs in cancer  
Asthma  
Fatty acid biosynthesis  
TNF signaling pathway  
Parkinson's disease  
Rheumatoid arthritis

| Gene          | R. philippinarum | C. gigas | P. fucata | L. gigantea | P. yesoensis |
|---------------|------------------|----------|-----------|-------------|--------------|
| C3            | 7                | 1        | 1         | 1           | 1            |
| CD109         | 17               | 4        | 9         | 6           | 5            |
| A2M           | 2                | 2        | 1         | 2           | 1            |

Table S28 RNA-seq differentially expressed genes (DEGs) under hypoxia exposure. Related to Figure 2.

Table S29 RNA-seq differentially expressed genes (DEGs) of R. philippinarum gonads exposed to digenetic trematode in nature. Related to Figure 2.

Table S30 The classified statistics of the GPCR-associated genes in the R. philippinarum, P. yesoensis, C. gigas, P. fucata and L. gigantea. Related to Figure 4.

Note: red indicates the expansion of the GPCR genes in Manila clam, and blue indicates genes uniquely observed in Manila clam.


| species          | length | contig N50 | scaffold N50 | BUSCO (genome)                                   |
|------------------|--------|------------|--------------|--------------------------------------------------|
| *R. philippinarum* | 1.12 Gb | 28.1 Kb    | 345.0 Kb     | C:92.2% [S:90.3%, D:1.9%], F:1.6%, M:6.2%, n:978 |
| *B. platifrons*   | 1.66 Gb | 13.2 Kb    | 343.4 Kb     | C:93.2% [S:91.9%, D:1.3%], F:2.6%, M:4.2%, n:978 |
| *M. philippinarum*| 2.63 Gb | 19.7 Kb    | 100.2 Kb     | C:88.9% [S:86.0%, D:2.9%], F:4.9%, M:6.2%, n:978 |
| *L. fortune*      | 1.67 Gb | 32 Kb      | 312 Kb       | C:81.9% [S:78.6%, D:3.27%], F:7.36%, M:10.73%, n:978 |
| *A. purpuratus*    | 724.8 Mb | 80 Kb      | 1.02 Mb      | C:90.4% [S:89.5%, D:0.9%], F:3.9%, M:5.7%, n:978  |
| *P. yessoensis*    | 988 Mb  | 38 Kb      | 804 Kb       | C:94.0% [S:92.0%, D:2.0%], F:1.3%, M:4.7%, n:978  |
| *C. farreri*       | 780 Mb  | 21.5 Kb    | 602 Kb       | C:93.0% [S:90.9%, D:2.1%], F:2.4%, M:4.6%, n:978  |
| *P. fucata*        | 991.02 Mb | 21.4 Kb   | 59.02 Mb     | C:87.3% [S:77.4%, D:9.9%], F:3.1%, M:9.6%, n:978  |
| *C. virginica*     | 685 Mb  | 1.9 Mb     | 75.9 Mb      | C:94.6% [S:91.9%, D:2.7%], F:1.1%, M:4.3%, n:978  |
| *C. gigas*         | 558 Mb  | 31.2 Kb    | 401.7 Kb     | C:86.9% [S:82.0%, D:4.9%], F:1.1%, M:12.0%, n:978 |
| *S. glomerate*     | 796.5 Mb | 39.4 Kb   | 804.2 Kb     | C:91.6% [S:88.2%, D:3.4%], F:3.8%, M:4.6%, n:978  |
| *O. bimaculoides*  | 2.34 Gb | 5.5 Kb     | 475.1 Kb     | C:89.2% [S:88.7%, D:0.5%], F:3.6%, M:7.2%, n:978  |
| Species                  | Gene number |
|-------------------------|-------------|
| Homo sapiens            | 1           |
| Mus musculus            | 1           |
| Gallus gallus           | 1           |
| Xenopus tropicalis      | 1           |
| Danio rerio             | 1           |
| Octopus bimaculoides    | 7           |
| Lottia gigantea         | 2           |
| Crassostrea gigas       | 24          |
| Patinopecten yessoensis | 26          |
| Pinctada fucata         | 31          |
| Ruditapes philippinarum | 21          |