

ARTICLE

<https://doi.org/10.1038/s41467-018-08197-4>

OPEN

# Penaeid shrimp genome provides insights into benthic adaptation and frequent molting

Xiaojun Zhang<sup>1,2</sup>, Jianbo Yuan<sup>1,2</sup>, Yamin Sun<sup>3</sup>, Shihao Li<sup>1,2,4</sup>, Yi Gao<sup>1,2,4</sup>, Yang Yu<sup>1,2,4</sup>, Chengzhang Liu<sup>1,2,4</sup>, Quanchao Wang<sup>1,2,4</sup>, Xinjia Lv<sup>1,5</sup>, Xiaoxi Zhang<sup>1,5</sup>, Ka Yan Ma<sup>6</sup>, Xiaobo Wang<sup>7</sup>, Wenchao Lin<sup>4</sup>, Long Wang<sup>4</sup>, Xueli Zhu<sup>4</sup>, Chengsong Zhang<sup>1,2,4</sup>, Jiquan Zhang<sup>1,2,4</sup>, Songjun Jin<sup>1,2,4</sup>, Kuijie Yu<sup>1,2,4</sup>, Jie Kong<sup>8</sup>, Peng Xu<sup>9</sup>, Jack Chen<sup>10</sup>, Hongbin Zhang<sup>11</sup>, Patrick Sorgeloos<sup>12</sup>, Amir Sagi<sup>13</sup>, Acacia Alcivar-Warren<sup>14</sup>, Zhanjiang Liu<sup>15</sup>, Lei Wang<sup>16</sup>, Jue Ruan<sup>7</sup>, Ka Hou Chu<sup>6</sup>, Bin Liu<sup>16</sup>, Fuhua Li<sup>1,2,4</sup> & Jianhai Xiang<sup>1,2,4</sup>

Crustacea, the subphylum of Arthropoda which dominates the aquatic environment, is of major importance in ecology and fisheries. Here we report the genome sequence of the Pacific white shrimp *Litopenaeus vannamei*, covering ~1.66 Gb (scaffold N50 605.56 Kb) with 25,596 protein-coding genes and a high proportion of simple sequence repeats (>23.93%). The expansion of genes related to vision and locomotion is probably central to its benthic adaptation. Frequent molting of the shrimp may be explained by an intensified ecdysone signal pathway through gene expansion and positive selection. As an important aquaculture organism, *L. vannamei* has been subjected to high selection pressure during the past 30 years of breeding, and this has had a considerable impact on its genome. Decoding the *L. vannamei* genome not only provides an insight into the genetic underpinnings of specific biological processes, but also provides valuable information for enhancing crustacean aquaculture.

<sup>1</sup>CAS Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China. <sup>2</sup>Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China. <sup>3</sup>Tianjin Biochip Corporation, Tianjin 300457, China. <sup>4</sup>Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao 266071, China. <sup>5</sup>University of Chinese Academy of Sciences, Beijing 100049, China. <sup>6</sup>School of Life Sciences, The Chinese University of Hong Kong, Shatin, N.T. 999077, Hong Kong SAR. <sup>7</sup>Agricultural Genomics Institute, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China. <sup>8</sup>Key Laboratory for Sustainable Utilization of Marine Fisheries Resources of Ministry of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China. <sup>9</sup>College of Ocean and Earth Sciences, Xiamen University, Xiamen 361102, China. <sup>10</sup>Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. <sup>11</sup>Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77843, USA. <sup>12</sup>Laboratory of Aquaculture & Artemia Reference Center, Ghent University, Coupure Links 653, Gent 9000, Belgium. <sup>13</sup>Department of Life Sciences and the National Institute for Biotechnology, Negev Ben Gurion University, Beer Sheva 84105, Israel. <sup>14</sup>Environmental Genomics Inc., P.O. Box 196 Southborough, MA 01772-1801, USA. <sup>15</sup>School of Fisheries, Aquaculture and Aquatic Sciences, Auburn University, Auburn, AL 36849, USA. <sup>16</sup>College of Life Sciences, Nankai University, Tianjin 300071, China. These authors contributed equally: Xiaojun Zhang, Jianbo Yuan, Yamin Sun, Shihao Li, Yi Gao, Yang Yu. Correspondence and requests for materials should be addressed to K.H.C. (email: [kahouchu@cuhk.edu.hk](mailto:kahouchu@cuhk.edu.hk)) or to B.L. (email: [liubin1981@nankai.edu.cn](mailto:liubin1981@nankai.edu.cn)) or to F.L. (email: [fhli@qdio.ac.cn](mailto:fhli@qdio.ac.cn)) or to J.X. (email: [jhxiang@qdio.ac.cn](mailto:jhxiang@qdio.ac.cn))

Crustacea is a dominant and diverse group of aquatic animals, among which the decapods (order Decapoda) comprise members of major ecological significance in marine and freshwater habitats and include many species with high economic values, such as shrimp, crabs, lobsters, and crayfish. In particular, the penaeid shrimp (family Penaeidae) represents the most important group in fisheries and aquaculture, and has therefore attracted considerable research attention<sup>1</sup>.

Penaeids predominantly inhabit shallow seas in tropical and subtropical areas. Their life history typically involves migration between offshore and inshore waters. After spawning, the planktonic shrimp larvae in offshore waters develop into postlarvae, which move to inshore waters and become benthic. Thus the shrimps have to adapt to different environments in their life but the underlying biological mechanisms have seldom been explored. Similar to all arthropods, penaeids exhibit discontinuous growth through intermittent molting. Different from most insects, where molting usually occurs at the larval stages for growth and metamorphosis, molting occurs throughout the lifetime of crustaceans. For instance, the penaeid shrimp experiences about 50 molts during a lifetime<sup>2</sup>, far more than in other arthropods<sup>3</sup>. Thus, shrimps could be an ideal model in studying the molting process. However, the detailed mechanisms of molting regulation are far from understood.

With an annual capture production of ~1.3 million tonnes, penaeids comprise the bulk of the global shrimp catch. Shrimp farming began in the 1970s, and has been one of the fastest growing sectors of the rapidly expanding aquaculture industry. The penaeid shrimp farming production reached >5 million tonnes, valued at over US\$32 billion in 2016<sup>4</sup>. The Pacific white shrimp *Litopenaeus vannamei* is the most important cultured shrimp worldwide, accounting for ~80% of total cultured penaeid shrimp production. While its native range is in the East Pacific Ocean, *L. vannamei* is now widely farmed in Central and South America and in Asia, particularly in China, Indonesia, Thailand, and Vietnam, with a number of breeding lines available. These lines are mostly produced through traditional selective breeding procedures, and genomic information would be extremely useful for future genetic manipulation. As early as 1997, in an international workshop on genome mapping of aquaculture animals, the penaeid shrimp was identified as one of five target organisms for genome sequencing, together with salmon, catfish, tilapia, and oyster<sup>5</sup>. The genomes of the latter four species have been published over the past decade<sup>6–9</sup>, but no complete genome of the shrimp has been reported to date, despite the efforts of a number of major research groups. For other crustaceans, high-quality genome assembly is only available for the water flea *Daphnia pulex*, the amphipod *Parhyale hawaiiensis*, and the marbled crayfish *Procambarus virginalis*<sup>10–12</sup>.

Here we present a high-quality de novo reference genome assembly for *L. vannamei*, with an extremely high proportion of simple sequence repeats (SSRs), which have been the major obstacle to genome sequencing and assembly. A prominent characteristic of the genome is the expansion of a series of genes related to visual system, nerve signal conduction, and locomotion, which apparently better equip the shrimp to adapt to its benthic habitat. The genome also sheds light on regulation of frequent molting through an intensified ecdysone signal pathway, with precise control by hormones and environmental factors. Moreover, our genomic analyses reveal that selective breeding has exerted a significant impact on the genome of *L. vannamei* broodstocks and the genetic resources acquired from this study will be useful for further genetic improvements in shrimp culture.

**Table 1 Summary of *L. vannamei* genome assembly**

**Genome assembly statistics**

Total length	1,663,559,157 bp
Number of scaffolds	4683
Longest scaffold	3,458,369 bp
Contig N50 length	57,650 bp
Scaffold N50 length	605,555 bp
Scaffold N90 length	204,841 bp

**Genome characteristics**

GC content	35.68%
Content of transposable elements	16.17%
Content of SSRs	23.93%
Predicted protein-coding gene number	25,596
Predicted noncoding RNA gene number	2777
Quantity of scaffolds anchored on linkage groups	3275
Length of scaffolds anchored on linkage groups	1,449,959,403 bp

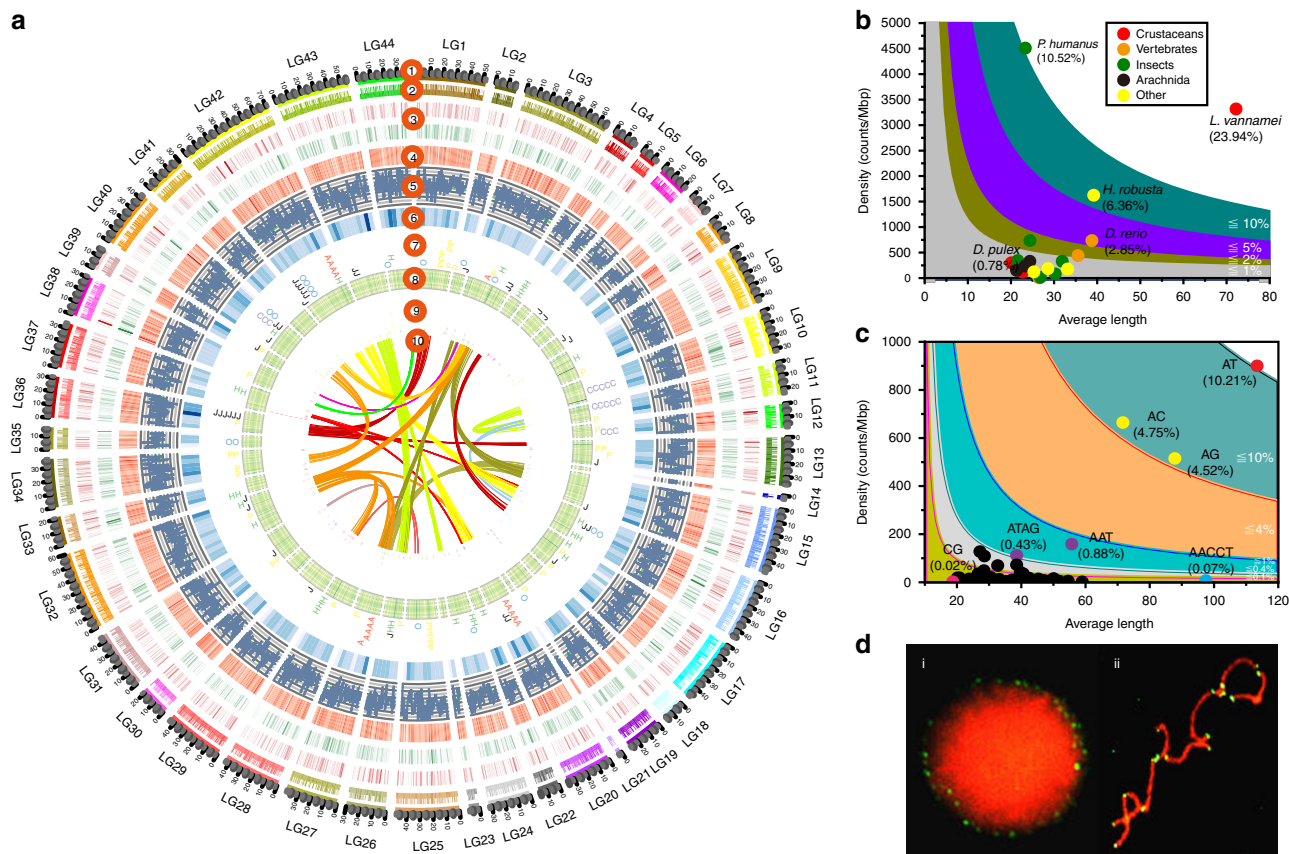
SSR simple sequence repeat, GC Guanine/Cytosine

## Results

**Genome sequencing assembly and annotation.** The *L. vannamei* genome size was measured to be 2.45 Gb by flow cytometry (Supplementary Fig. 1), similar to the size estimated by k-mer analysis (2.60 Gb, Supplementary Fig. 2). The sequencing and assembly of the *L. vannamei* genome were challenging due to the highly abundant SSRs as suggested by genome survey analysis<sup>13</sup>. We therefore applied a variety of sequencing technologies, which generated 828 Gb of Illumina clean sequences (338×), 133 Gb of PacBio long reads (54×), and 34,266 BAC end sequences (0.46×) (Supplementary Tables 1–3). We also conducted numerous conventional approaches for genome assembly, which yielded unsatisfactory results (Supplementary Table 4). Finally, we developed the WTDBG approach, which uses a fuzzy Bruijn graph method, to obtain the best assembly with the highest continuity and accuracy for the genome, that contains 1.66 Gb in 4683 scaffolds, with contig N50 of 57.65 Kb and scaffold N50 of 605.56 Kb (Table 1). The assembly was comparable to, or better than, those of other crustaceans, including the recently published genome of the marbled crayfish (scaffold N50 of 39.40 Kb)<sup>10</sup> (Supplementary Table 5).

The assembly showed high integrity and quality. Over 93% of Illumina reads could be mapped to the genome (Supplementary Table 6). The genome covered >94% of the unigenes assembled from RNA-Seq data and 94.76% of the conserved core eukaryotic genes (Supplementary Tables 7 and 8). The accuracy of our assembly is further substantiated by the 14 sequenced BAC clones, which were completely covered by corresponding scaffolds with high synteny (Supplementary Figs. 3 and 4). To assemble chromosome-level sequences, a high-density linkage map was used to anchor the scaffolds<sup>13</sup>. A total of 3275 scaffolds were anchored to 44 pseudochromosomes, representing 87.34% of the total genome (Fig. 1a, Supplementary Table 9).

The k-mer analysis indicated that repetitive sequences account for ~78% of the genome (Supplementary Fig. 2), which is more than that identified in the final assembly (49.38%) (Supplementary Table 10), indicating that some repetitive sequences are still missing from the assembly. The genome contains the highest proportion of SSRs (23.93%) among all sequenced animal genomes (Fig. 1b, Supplementary Table 11). The mean length of SSRs is 72.21 bp, which is over twice as long as those in other arthropods (20.11–31.91 bp). The SSRs density (3315.23/Mb, one SSR per 301 bp) in the *L. vannamei* genome is also, as far as we are aware, the highest among other reported genomes, except *Pediculus humanus* (4508.69/Mb) (Fig. 1b). Dinucleotide repeats are the dominant type of SSRs,



**Fig. 1** The genomic characteristics of *L. vannamei*. **a** A schematic representation of the genomic characteristics of *L. vannamei*. Track 1 (from the outer-ring): 44 linkage groups (LGs) of the shrimp genome. Track 2: Scaffolds anchored to each linkage group. Track 3: Protein-coding genes present in the scaffolds. Red represents genes on forward strand and green for genes on reverse strand. Track 4: Distribution of gene density with sliding windows of 1 Mb. Higher density is shown in darker red color. Track 5: Distribution of GC content in the genome. Track 6: Distribution of SNPs density with sliding windows of 5 Mb. Higher density is shown in darker blue color. Track 7: Distribution of six significantly expanded gene families in the genome, which are opsin (O), peritrophin-like protein (P), chitinase (H), calcified cuticle protein (A), crustacyanin (C), and JHE-like carboxylesterase 1 (J). Track 8: Distribution of SSRs in the genome. Higher density of SSRs is shown with deeper color. Track 9: Distribution of miRNA in the genome. Clusters of co-transcribed miRNAs at adjacent positions are displayed in stacked style. Track 10: Schematic presentation of major interchromosomal relationships in the shrimp genome. **b** SSR content among different animal genomes. **c** Distribution of different types of SSRs in the *L. vannamei* genome. **d** Fluorescence in situ hybridization (FISH) of  $(AACCT)_n$  type SSR to the nucleus (i) and the chromosomes (ii) of *L. vannamei*. SNP single nucleotide polymorphism, SSR simple sequence repeat

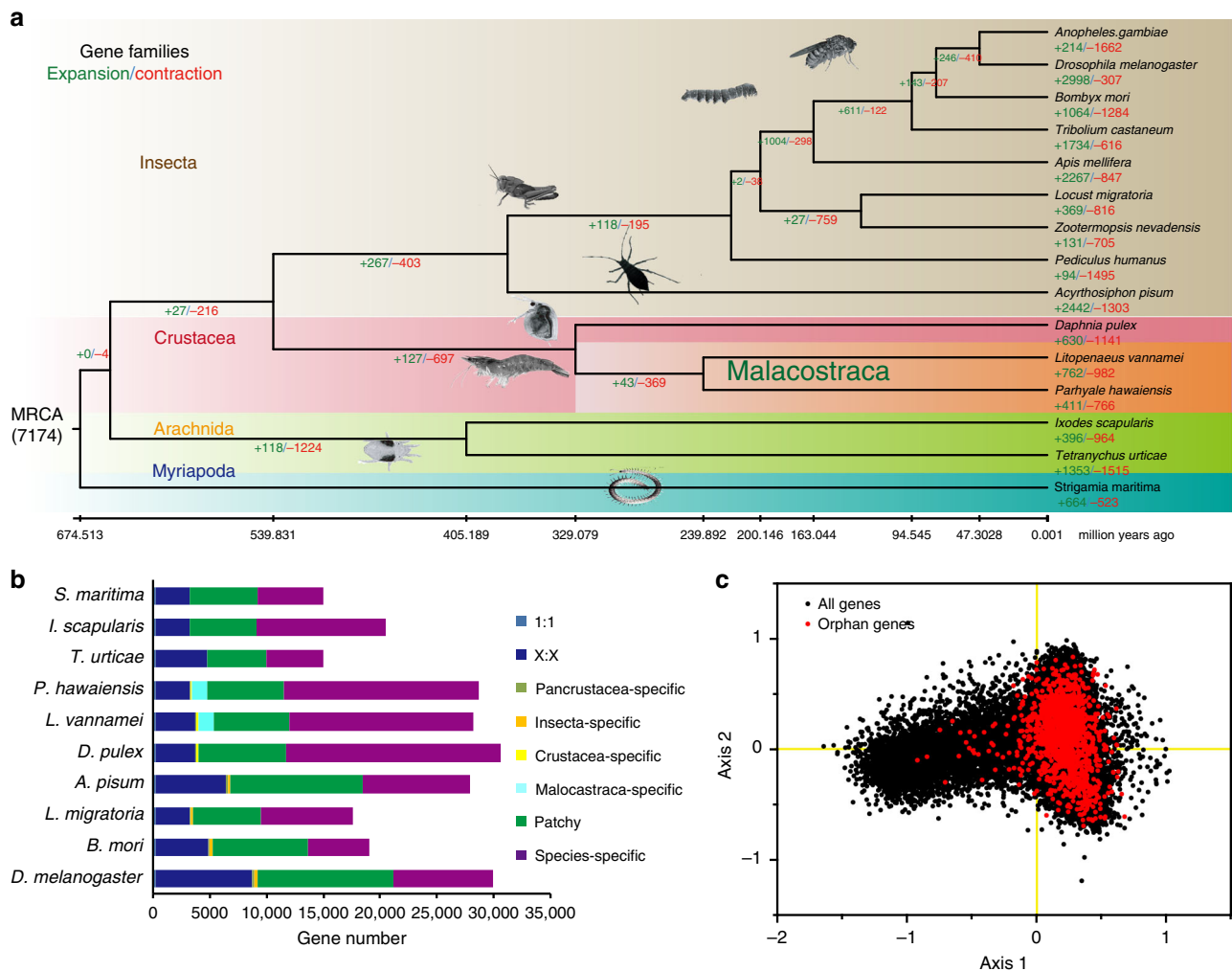
with  $(AT)_n$ ,  $(AC)_n$ , and  $(AG)_n$  accounting for 81.40% of total SSRs (Fig. 1c, Supplementary Fig. 5). Most SSRs were located in intergenic regions (24.63%) and introns of protein-coding genes (22.07%), and far fewer were found in exons (1.41%).  $(AACCT)_n$ , a telomere component identified by fluorescence in situ hybridization (Fig. 1c), is longer than many other SSRs, with the longest SSR (13,769 bp), belonging to this type (Supplementary Fig. 6). Furthermore, the length of  $(AACCT)_n$  located in introns is longer than those found in other genomic regions (Supplementary Fig. 7).

Transposable elements (TEs) account for 16.17% of the *L. vannamei* genome, with DNA transposons (9.33%) and long interspersed elements (LINES, 2.82%) comprising the two major classes. En-Spm (6.39%) was found to be the most abundant TE, with its abundance markedly higher than in *D. pulex* (0.05%), *P. virginalis* (0.01%) and *P. hawaiiensis* (0.25%)<sup>10–12</sup> (Supplementary Table 12). Most TEs in *L. vannamei* showed higher divergence (substitution rate 19–33%) than those in other crustaceans (Supplementary Fig. 8). However, LINES in *L. vannamei*, especially RTE-BovB and Penelope, showed a low divergence (Supplementary Fig. 8).

In total, 25,596 protein-coding gene models were annotated in *L. vannamei*. Compared to *D. pulex* and *P. hawaiiensis*, *L.*

*vannamei* has a longer average exon size (259 bp) and more exons per gene (5.94) (Supplementary Fig. 9, Supplementary Table 13). We also annotated noncoding small RNAs, including 1458 tRNAs, 464 rRNAs, 296 small nuclear RNAs (snRNAs), 255 small nuclear RNAs (snoRNAs), 90 ribozymes, and 214 microRNAs (miRNAs) (Supplementary Table 14).

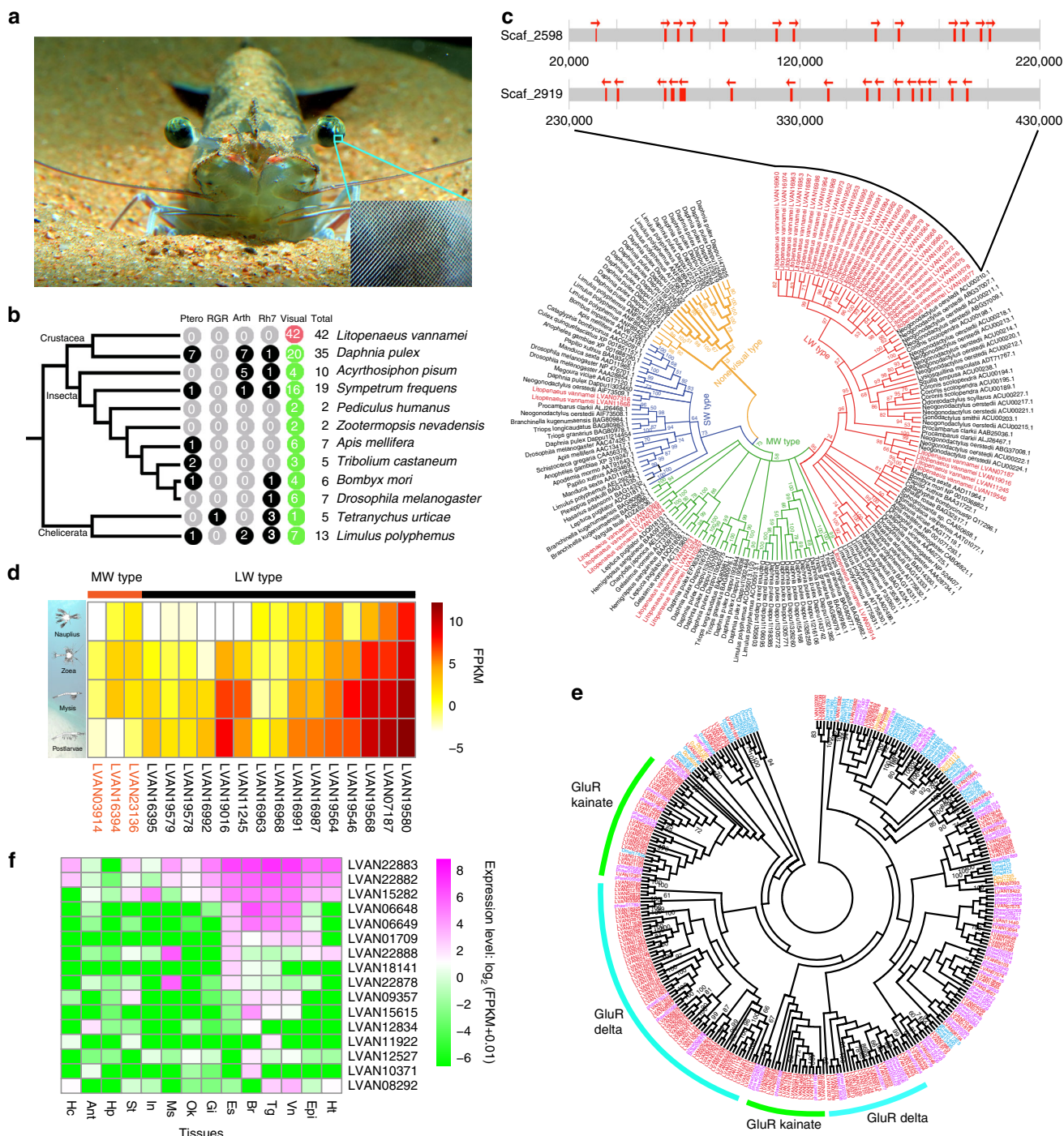
**Genome evolution.** Phylogenetic analysis indicates that crustaceans and hexapods form a monophyletic group of Pancrustacea, with the latter nested within the former, making Crustacea paraphyletic (Fig. 2a). As members of Malacostraca, *L. vannamei* (Decapoda) and *P. hawaiiensis* (Amphipoda) diverged at about 240 MYA (million years ago), i.e., in early Mesozoic. This divergence occurred after the Permian-Triassic mass extinction, when about 96% of marine species became extinct<sup>14</sup>, followed by the radiation of many new life forms. This divergence period is also consistent with the radiation of shrimp-like decapods in the early Middle Triassic<sup>15</sup>. We found three prominent genome characteristics from *L. vannamei* that might underlie the rapid evolution of penaeid shrimp, namely, abundant SSRs, a high proportion of taxon-specific genes, and extensive tandem gene duplications.



**Fig. 2** Comparative genomics analysis of *L. vannamei* and other arthropods. **a** Phylogenetic placement of *L. vannamei* in the arthropod phylogenetic tree. Numbers on branches indicate the number of gene gains (+) or losses (-). The estimated divergence times are displayed below the phylogenetic tree. Image credits: Gewin V, Martin Cooper, Bernard Dupont, Carlo Brena, Gilles San Martin, David Ludwig, S. Rae. **b** Comparison of the gene repertoire of ten arthropod genomes. “1:1” indicates single-copy genes, “X:X” indicates orthologous genes present in multiple copies in all the ten species, where X means one or more orthologs per species, “patchy” indicates the existence of other orthologs that are presented in at least one genome. **c** Principal component analysis (PCA) on relative codon usage of orphan genes

Because of their special mutational and functional qualities, SSRs play a major role in generating genetic variation underlying adaptive evolution<sup>16</sup>. Besides *L. vannamei*, two penaeid species, *Penaeus monodon* and *Marsupenaeus japonicus*, also have a high proportion of SSRs (~10%) in their genome, traceable through low coverage genome sequencing<sup>17</sup>. However, the genomes of the marbled crayfish *P. virginalis* and the amphipod *P. hawaiiensis* contain only 0.99 and 1.27% of SSRs, respectively (Supplementary Table 12). Therefore, the expansion of SSRs might stem from the common ancestor of the penaeid shrimp. SSRs play critical roles in regulating genome plasticity (including DNA recombination and replication) and gene expression<sup>18</sup>. SSRs in *L. vannamei* were widely distributed among introns (22.07%) in 16,741 genes (65.47%), whose expression may be regulated by SSR polymorphism. SSRs may also contribute to DNA recombination with TEs, as most TEs and their up- or downstream sequences contain SSRs (Supplementary Fig. 10), which make up more than 90% for ERV1, Charlie, Sola, MuDR, and En-Spm. Thus, the significant expansion of SSRs might provide a unique genetic architecture for the shrimp’s adaptive evolution.

Compared with other crustaceans, the *L. vannamei* genome has 762 expanded gene families and 16,291 species-specific genes (>57% of the entire gene repertoire, Fig. 2b), including genes related to myosin complex, chitin binding, metabolic process, and signaling transduction (Supplementary Data 1–2, Supplementary Tables 15–16, Supplementary Fig. 11). Orphan genes can contribute to lineage-specific adaptation<sup>19</sup>. A total of 3369 orphan genes were identified in the *L. vannamei* genome. They displayed a lower number of exons (4.47 exons/gene), but a greater length of exons (292 bp/gene) in contrast to the average length of all genes (5.94 exons and 260 bp per gene). They exhibited special gene structure characters, and special temporal/spatial expression patterns, suggesting their independent de novo origins (Fig. 2c, Supplementary Figs. 12 and 13). When searched against the transcriptome unigenes of other decapods, the orphan genes were found to be present in other penaeids, such as *Fenneropenaeus chinensis* (83.07%) and *P. monodon* (64.59%), but rarely present in other decapods, e.g. *Exopalamon carinicauda* (1.95%), *Eriocheir sinensis* (1.09%), and *Neocaridina denticulata* (3.82%), suggesting that these orphan genes are lineage-specific, and may contribute to penaeid-specific adaptation.



**Fig. 3** Opsin and iGluR gene family in the shrimp genome. **a** Picture of the *L. vannamei* eyes and ommatidia. **b** Opsin genes of *L. vannamei* in comparison with those in the genomes of various arthropods. The number of visual and nonvisual type opsin genes, including pteropsin (Ptero), RGR-like (RGR), arthropsin (Arth), Rh7-like (Rh7) were identified using sequence alignment with known opsins and GPCR-domain searches (Supplementary Fig. 18). **c** Phylogenetic tree of the opsin gene family in arthropods. Six clades of opsins in the *L. vannamei* genome were observed (red). The genes in largest clade are specifically expanded opsins in the *L. vannamei* genome, which are also tandemly duplicated. The arrow indicates the transcriptional orientation. **d** Expression of opsin genes (FPKM>1) during different larval stages of *L. vannamei*. **e** Expansion of ionotropic glutamate receptor (iGluR) genes in *L. vannamei*. **f** Expression of iGluR genes in different tissues: hemocyte (Hc), antennal gland (Ant), hepatopancreas (Hp), stomach (St), intestine (In), muscle (Ms), lymphoid organ (OK), gill (Gi), eyestalk (Es), brain (Br), thoracic ganglion (Tg), ventral ganglion (Vn), epidermis (Epi), heart (Ht). GPCR G protein-coupled receptor

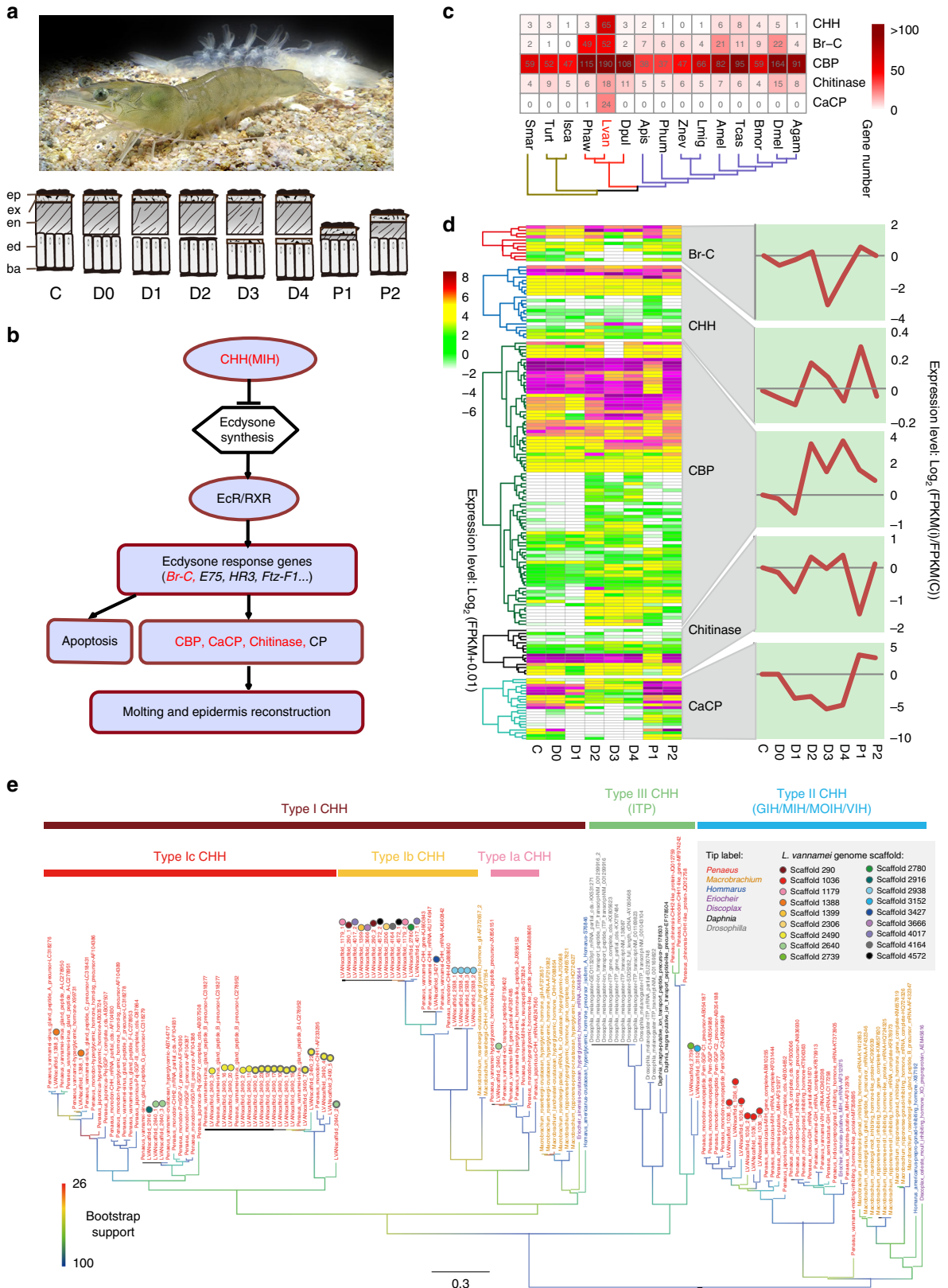
Gene tandem duplication was one of the most intriguing features of the *L. vannamei* genome. A total of 4662 genes were identified to be tandemly duplicated (Supplementary Fig. 14), and these duplicates share high sequence similarities (identity >98%), indicating recent duplications. Moreover, these genes mainly consist of opsins, crustacyanins, chitinases, cuticle proteins,

myosins, actins and heat shock proteins, which may be important for adaptive evolution of the shrimp.

**Visual system for benthic adaptation.** *L. vannamei* has a pair of large, stalked eyes with an ovoid visual surface making up of a

highly elevated number of ommatidial lenses<sup>1</sup> (Fig. 3a). Each compound eye consists of about 55,000–80,000 ommatidia<sup>20</sup>, rivaling the most acute eyes in arthropods. Adult and juvenile *L. vannamei* are benthos in turbid and shallow waters to a depth of 50 m, where no light with a wavelength <500 nm penetrates. Therefore, *L. vannamei* are expected to have developed

photoreceptors capable of capturing photons most common in this environment<sup>21</sup>. The genome sequencing of *L. vannamei* now allows us, for the first time, to infer genomic-level evolution in this delicate eye<sup>1</sup>. We identified 42 opsin genes, which are the key light-sensing factors responsible for visual signal transduction (Fig. 3b). All these opsins belong to the visual type related to color



sensation. To our knowledge, *L. vannamei* has the most visual opsin genes among animals, including creatures generally regarded as having excellent vision such as mantis shrimps (6–33 opsins) and dragonflies (20 opsins)<sup>22</sup>. The 42 visual-opsins can be divided into three subfamilies, 33 in the long-wavelength type (LW) subfamily, seven in the middle-wavelength type (MW) subfamily and two in the short-wavelength type (SW) subfamily (Fig. 3c). Intriguingly, most of the LW subfamily genes belong to two clusters in the genome (Fig. 3c), suggesting a major expansion via tandem duplication.

The expansion of LW opsins sensitive to light at the 560 nm wavelength facilitates feeding of the shrimp in benthic environment with predominantly yellow and red (long-wavelength) light<sup>21</sup>. Expansion of opsin genes was also observed in the genome of the water flea *D. pulex* (36 opsins), but this expansion is restricted to the MW subfamily, which is sensitive to light at ~530 nm wavelength associated with planktonic habitat (Supplementary Fig. 15). Considering the shift from planktonic to benthic habitat during the shrimp's life cycle, we analyzed the expression of opsin genes at different larval stages of *L. vannamei*. Interestingly, LW-type opsin genes showed significantly higher expression levels in the benthic postlarvae, whereas the MW opsin genes displayed higher expression in the planktonic larvae (Fig. 3d). These data provide strong support for the correlation of differential expression among the types of visual opsin genes with the planktonic-benthic habitats, suggesting that *L. vannamei* has a highly effective visual system adapted to its habitat transition.

**Nerve signal conduction and locomotion.** In the central nervous system of decapods, the giant nerve fibers originating from the brain and terminating at the caudal ganglion are responsible for the rapid escape reactions<sup>23</sup>. These nerve fibers of penaeid shrimp have a conduction speed of ~200 ms<sup>-1</sup>, generally regarded as the fastest conducting speed in animals<sup>24</sup>, compared with a mere 20 ms<sup>-1</sup> for squids, which have the largest nervous system among invertebrates<sup>25,26</sup>. In the *L. vannamei* genome, the gene categories for signal recognition and neural development are significantly enriched (Supplementary Data 1, Supplementary Tables 16). Marked expansion was found in several gene families, including 457 G protein-coupled receptors (GPCRs), 53 transient receptor potential channels, 21 innexins, and 47 protocadherins (Supplementary Table 17), which are considered as the core neural-development-related genes across bilaterians<sup>26</sup>. The rapid nerve conduction in shrimps is based on their heavy myelin sheath<sup>27</sup>, but the myelin substrate, cholesterol, cannot be synthesized in most invertebrates<sup>25,28</sup>. Crustacyanins, the apolipoprotein family genes which transport cholesterol, are significantly expanded and tandemly duplicated in the *L. vannamei* genome, and most of them are highly expressed in the neural and digestive system (Supplementary Data 3).

Unlike vertebrates and most protostomes that use acetylcholine as a neurotransmitter, arthropods use glutamate at the neuromuscular junction<sup>29</sup>. A total of 169 ionotropic glutamate receptors (iGluRs) and 148 glycine receptors (GlyRs) genes are present in the *L. vannamei* genome (Fig. 3e, Supplementary Fig. 16, Supplementary Table 18). iGluRs are involved in excitatory neurotransmission, which can occur quickly and last for a long time. They also play a role in regulating myelination<sup>30</sup>. GlyRs can activate Cl<sup>-</sup> conductance in neurons<sup>31</sup>. Furthermore, iGluRs and GlyRs genes are mainly expressed in the nervous system as well as in muscle (Fig. 3f, Supplementary Fig. 17). In addition, some gene families involved in muscle contraction, such as actins and myosins, are also significantly expanded in the *L. vannamei* genome (Supplementary Table 16). These features might equip the shrimp to rapidly perform neural signal transduction, enhancing its responsiveness and strengthening its locomotion, including escape reactions in adaptation to its benthic swimming life in shallow seawater.

**Intensified ecdysone signal pathway.** Molting is a basic physiological process of crustaceans (Fig. 4a), through which they realize metamorphosis, growth, and development. An ecdysone signal pathway, under hormonal regulation, controls every step in molting (Fig. 4b). Both an initial rise and a coordinated decline of circulating ecdysone concentration are necessary for successful molting<sup>32</sup>. In *L. vannamei*, the immediate function of ecdysone in promoting molting is amplified by expanded downstream genes in the ecdysone signal pathway. After entering the nucleus, ecdysone activates successive transcription of early stage ecdysone response genes (Fig. 4b). The major ecdysone response gene, *Broad-Complex (Br-C)*, was perceptibly expanded in the genome (Fig. 4c), and were highly expressed at early premolt (D1) and postmolt (P1) (Fig. 4d), consistent with their function in regulating the expression of downstream effectors, such as chitinase and cuticular proteins<sup>33</sup>. Cuticle proteins, chitin binding proteins (CBPs) and calcification-related cuticular proteins (CaCPs) are main structural proteins for constructing the cuticle<sup>34</sup>, which protects the shrimp from bacteria, fungi, viruses, and mechanical injuries. However, to allow for growth in body size, the cuticle is degraded prior to ecdysis by chitinase<sup>34</sup>. Genes encoding CBPs, CaCPs, and chitinase are significantly expanded in the *L. vannamei* genome (Fig. 4c, Supplementary Table 19). These genes were mainly expressed in the epidermis and intestinal peritrophic matrices (Supplementary Fig. 19). The structural proteins were coordinately expressed during molting, which began with the initial degradation by chitinase, and then the synthesis of new cuticle with high expression of Rebers and Riddiford 2 (R&R2)-type cuticle protein and CBPs genes from early premolt (D2) to late premolt (D4). After ecdysis (P1–P2), the CaCPs and R&R1-type cuticle proteins were produced to harden the cuticle (Fig. 4d,

**Fig. 4** Key gene families related to the ecdysone signal pathway in the *L. vannamei* genome. **a** Picture of shrimp and its shed exoskeleton. Schematic diagrams below the picture show the changes of the epidermis (ep: epicuticle, ex: exocuticle, en: endocuticle, ed: epidermis, ba: basement membrane) during the molting cycle. The molting cycle can be divided into eight stages: Intermolt (C) stage, Premolt stages (D0–D4), and Postmolt stages (P1–P2). **b** The shrimp ecdysone signal pathway. Expanded gene families are highlighted in red. **c** Distribution pattern of the expanded gene families in shrimp ecdysone signal pathway and comparison with other arthropods. **d** Expression patterns of the expanded gene family genes at the different molting stages of *L. vannamei*. The overall expression trends in different molting stages are shown in the right. **e** Phylogenetic analysis of the CHH family based on CDS sequences containing full CHH domain, reconstructed using IQ-tree 1.6.2 under TVMe + I + G4 model with ultrafast bootstrap method. Sequences from *L. vannamei* genome are annotated with circles of various colors to indicate their corresponding scaffold of origin. GenBank sequences of the CHH family from genus *Penaeus* s.l., *Macrobrachium*, *Homarus*, as well as from species *Eriocheir sinensis*, *Discoplax celeste*, *Daphnia magna* and *Drosophila melanogaster* were incorporated in this analysis. These sequences were annotated in the phylogenetic tree as follows: species name/gene description/accession number. They were classified into type I, II, and III CHH subfamilies, in which type III CHH consisted of ion transport peptides (ITPs). CHH crustacean hyperglycemic hormone

Supplementary Fig. 19). The expanded genes encoding Br-C and cuticular proteins in the genome can enhance the functioning of ecdysone, once its concentration reaches a certain threshold. The enhanced ecdysone function might be the primary mechanism to assure the frequent molting behavior of shrimps.

In *L. vannamei*, a striking expansion of the crustacean hyperglycemic hormone (CHH) family underpins the control of the enhanced ecdysone function (Fig. 4c). Ecdysone synthesis in the crustacean Y-organ is mainly under the negative control of the molt-inhibiting hormone (MIH), one of the type II peptides of the CHH family<sup>35</sup>. We identified seven tandemly duplicated MIH genes (Fig. 4e) with high sequence similarity (Supplementary Fig. 20). This might facilitate the coexpression and cofunctionality of these genes during molting. MIH transcripts were mainly located in the eyestalk and ventral nerve (Supplementary Data 4), and their expression was low at late premolt and high from postmolt to early premolt (Supplementary Data 4), the exact reverse of the hemolymph ecdysone level<sup>36</sup>. Besides, type I peptides of the CHH family are also known to suppress ecdysteroidogenesis in the Y-organ, though at a higher dose than MIH<sup>37</sup>. In the *L. vannamei* genome, we identified three major clades of type I CHH peptides: (1) type Ia CHH clade, which contains peptides more highly expressed in gut, thoracic ganglion and lymphoid organ than in eyestalk and other tissues (Supplementary Data 4), and may function in ion/osmoregulation and water uptake but lack hyperglycemic functions, (2) type Ib clade, which is highly expressed in eyestalk as well as brain and thoracic ganglion, and is less well-studied but also believed to assume an ion/osmoregulatory function<sup>38</sup>, and (3) type Ic clade, which is penaeid-specific and contains peptides involved in the regulation of molting, reproduction, energetics and ionic metabolism<sup>39</sup> (Fig. 4e). Significantly, we detected much more striking expansion in the type Ic CHH than in other type I CHH genes in the genome (Fig. 4e, Supplementary Data 5). The type Ic CHH genes showed similar expression patterns as MIH genes in different tissues and molting stages, while the other type I CHH did not (Fig. 4d, Supplementary Data 4). Most type Ic CHH genes are tandemly located in Scaffold 2490 (Fig. 4e) and share high sequence similarity (Supplementary Fig. 21). Similar gene organization and expression patterns between the type Ic CHH genes and MIH genes suggest that they might play similar roles in molting. Furthermore, MIH and CHH also increase the intracellular concentration of cyclic nucleotides (cAMP or cGMP) and  $\text{Ca}^{2+}$  in the Y-organ, which further inhibits ecdysteroidogenesis<sup>40</sup> by activating the “triggering” phase of MIH regulation<sup>50</sup>. The genes involved in the “triggering” phase of MIH regulation, including cAMP-dependent protein kinase, phosphodiesterase 1, calmodulin and  $\text{Ca}^{2+}$ /calmodulin-dependent protein kinase genes, were all under positive selection (Supplementary Fig. 22, Supplementary Data 6). The expansion and tandem duplication of CHH family genes, and positive selection of MIH “triggering” phase genes might strengthen the negative control of ecdysone synthesis. Collectively, this intensified ecdysone signal pathway, generated by gene expansion, tandem duplication and positive selection, with amplification of ecdysone function and negative control on ecdysteroidogenesis, serves to ensure the frequent and precise molting process in the shrimp.

**Key genes mediating environmental factors in molting regulation.** Molting is influenced by many environmental factors, such as nutrition and photoperiod. MIH genes of *L. vannamei* universally contain the sterol regulatory elements (SRE), i.e. the binding sites for SRE-binding protein (SREBP), in the regulatory sequence (Supplementary Fig. 23). Furthermore, all detected transcripts of MIHs and SREBP had similar expression patterns,

with high expression levels in the eyestalk at intermolt and low levels at late premolt (Fig. 5a). SREBP is a major regulator of cholesterol metabolism, and its function is inhibited by a high level of cellular cholesterol<sup>41</sup>. Ecdysone is biosynthesized with cholesterol as the substrate, and cholesterol cannot be synthesized in vivo in crustaceans<sup>28</sup>. The ecdysone level<sup>46</sup> exhibits a similar trend to the in vivo cholesterol level during molting<sup>53</sup>, and both show opposite trends to expression patterns of MIHs and SREBP. These observations suggest that a high cholesterol level may drive molting through SREBP, which may positively regulate MIH expression, thus suppressing ecdysone synthesis and molting. This hypothesis was supported by a reduction (~95%) in expression of all detected MIH genes after SREBP silencing in the eyestalk (Fig. 5b, Supplementary Fig. 24). Injection of cholesterol into shrimps had a similar inhibitory function on MIH expression, reducing MIH expression levels by >90% (Fig. 5b). This inhibition was possibly exerted through inhibiting nuclear translocation of SREBP<sup>42</sup>. MIH silencing (Supplementary Fig. 24) significantly accelerated molting of shrimps, making them develop to the late premolt (D3, Fig. 5c, i). However, most control shrimps at intermolt progressed to early premolt (D1) within the same duration (Fig. 5c, ii). By contrast, injection of Lipitor, a lipid-lowering agent, apparently retarded molting at the D3 stage (Fig. 5c, iii), as control shrimps progressed to postmolt (Fig. 5c, iv). These data indicated that SREBP silencing and a high cholesterol level had similar functions as MIH silencing, supporting the hypothesis that SREBP mediates MIH-regulated ecdysteroidogenesis.

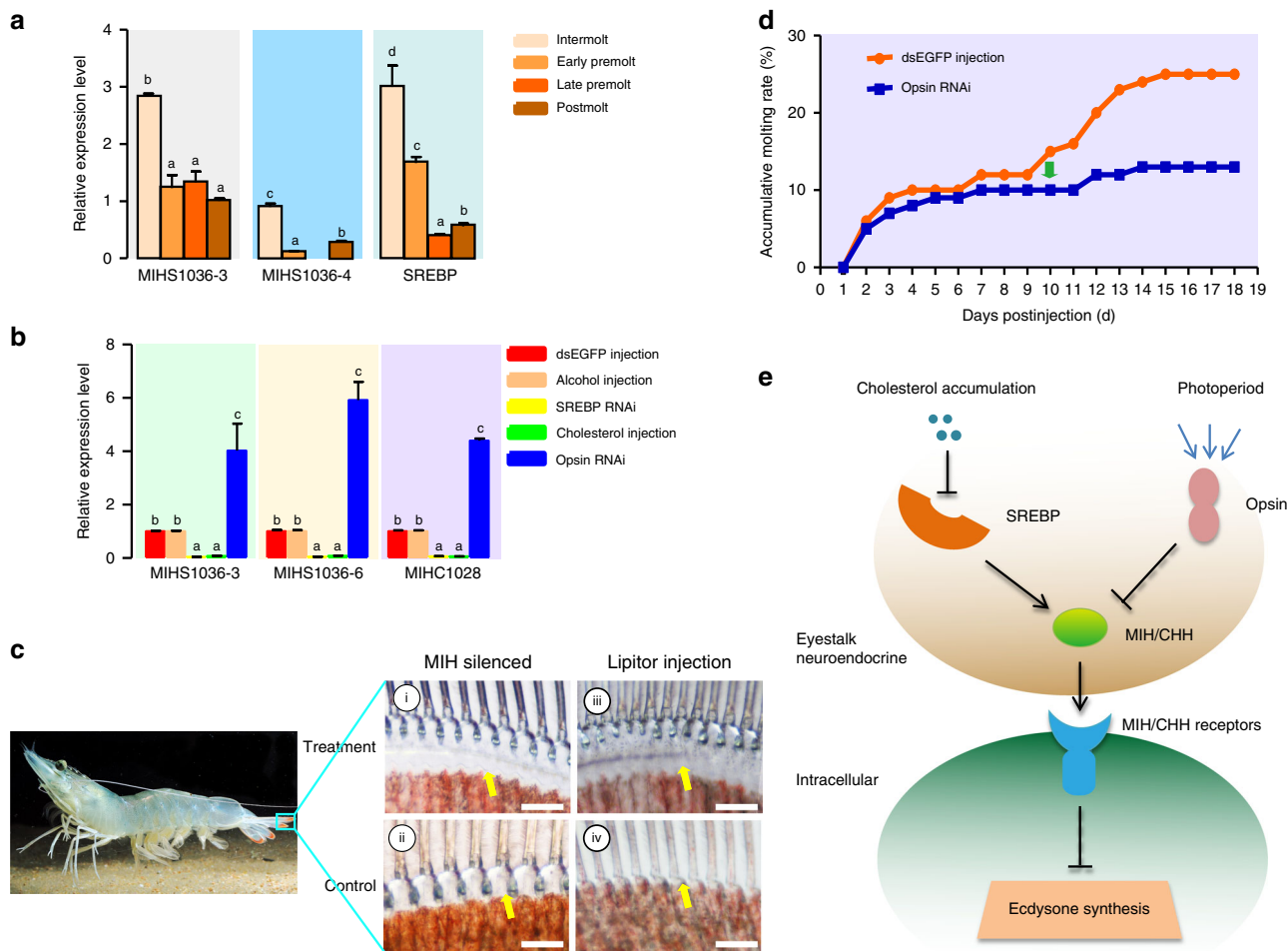
The photoperiod has evident effects on crustacean molting<sup>43</sup>. Due to the key function of opsin in visual transduction, we speculate that opsin might be the primary mediator between photoperiod and molting. This is supported by the upregulated levels of MIH genes after silencing of several LW opsin genes (Fig. 5b, Supplementary Fig. 25), indicating an inhibitory role of opsin in MIH expression. In addition, after 10 days of opsin silencing, shrimps exhibited a lower molting rate than the control (Fig. 5d).

In sum, we infer that cholesterol supply and photoperiod regulate molting via SREBP and opsin, respectively (Fig. 5e). In this model, a low cholesterol level corresponds to increased SREBP and MIH expression at postmolt. Cholesterol uptake leads to inactivation of SREBP and consequent reduction in MIH expression at early premolt. Simultaneously, ecdysone also rises to initiate the downstream signaling pathways. The model also proposes the function of opsin in mediating the effect of photoperiod through modulating the MIH-regulated molting.

**Immune protection during molting.** The newly formed cuticle of shrimps at postmolt is thin and not yet hardened, making them prone to pathogen infection. We anticipate that the shrimp may possess an intensified immune protection during ecdysis. Crustaceans mainly rely on their innate immune system, consisting of humoral immunity and cellular immunity, to defend against pathogens<sup>44</sup>. We found 111 immune-related genes with differential expression patterns at different molting stages. These genes were clustered into six groups according to their functions, including 15 antimicrobial peptides (AMPs), 59 pattern recognition receptors (PRRs), six phenoloxidase-system-related genes, 14 apoptosis-related genes and 17 homeostasis-related genes (Supplementary Data 7).

Crustins contributed to 13 of the 15 differentially expressed AMPs (Supplementary Data 7). Most of them showed the lowest expression levels during late premolt and increased expression levels immediately after ecdysis (Supplementary Fig. 26A). Crustins were mainly distributed in the epidermis,





**Fig. 5** Regulation of MIH-mediated ecdysone synthesis in shrimp. **a** The expression patterns of *SREBP* and *MIHs* in shrimp eyestalks at different molting stages. **b** The expression profiles of *MIHs* in the eyestalk with different treatments. Injections of dsEGFP and PBS serve as controls. **c** Phenotypic changes of the shrimp uropods cuticle after treatments. Shrimps at the intermolt stage were treated by RNAi: (i) 2 days after injection of dsMIH; (ii) 2 days after injection of dsEGFP. Shrimps at the early premolt stage were used for Lipitor treatment: (iii) 2 days after injection of Lipitor; (iv) 2 days after injection of PBS. Yellow arrows indicate the space between old and new cuticles at different molting stages. Bar = 50  $\mu$ m. **d** Effect of opsin silencing on accumulative molting rates, with the injection of dsEGFP serving as a control. The green arrow shows the time when the treatment led to a significantly lower accumulated ecdysis rate compared to the control group. **e** The schematic diagram of the regulation of *SREBP* and opsin on the MIH-mediated ecdysone synthesis. *SREBP* SRE-binding protein

stomach, gill, and hemocytes (Supplementary Fig. 26B). Of the 59 identified PRRs, 53 are lectins, important PRRs involved in the innate immunity of crustaceans<sup>45</sup> (Supplementary Data 7). Several lectin genes were highly expressed at P1 stage (Supplementary Fig. 26C), mainly detected in the epidermis, stomach and gill (Supplementary Fig. 26D, 27). Most other lectins were highly expressed in the hepatopancreas, which exhibited the lowest expression levels at the P1 stage and was highly expressed from P2 to early premolt (Supplementary Fig. 28). Expressions of the phenoloxidase-system-related genes apparently increased at postmolt (Supplementary Data 7) and were mainly distributed in the gill, epidermis, stomach, intestine, hemocytes, and hepatopancreas (Supplementary Fig. 26E, F, G). Recently, a phenoloxidase gene in insects was reported to play a role in skin immunity, protecting the host against fungal infection after ecdysis<sup>46</sup>. Most of the identified apoptosis-related genes, mainly expressed in the intestine and hepatopancreas, were highly expressed throughout the molting process, with two peaks at D2 and D4 (Supplementary Fig. 26E, F). At premolt, a new cuticle is generated after apolysis, when the old cuticle is separated from the epidermis. The expression

patterns of apoptosis-related genes indicated that they might be involved in apolysis rather than immunity.

In summary, AMPs (crustins, anti-lipopolysaccharide factor and penaeidin), PRRs (lectins and LGBP) and phenoloxidase system genes largely contribute to the immune protection during ecdysis (Supplementary Fig. 26H). While these immune-related genes are important in protecting the host against pathogens in the external environment, their high expression in the hepatopancreas might be related to the *in vivo* gut flora stability of the shrimps and protect the host against pathogens in internal environment as bacteria, including some pathogens are ubiquitous in the hepatopancreas.

**Impacts of breeding.** *L. vannamei* became an important aquaculture species in the 1980s, and a selective breeding program began under the “US Marine Shrimp Farming Program (USMSFP)” in 1989<sup>47</sup>. Compared to the long history of domestication in plants, terrestrial livestock and poultry<sup>48,49</sup>, the history of *L. vannamei* breeding is relatively short. However, the selection pressure in *L. vannamei* is very high, as the shrimp could be

selected one generation per year and a single spawning female can produce >300,000 offspring<sup>50</sup>. To examine the genomic consequences of such intensive selective breeding, we selected 22 individuals for genome resequencing, including 8 individuals from the wild and 14 individuals from five different broodstocks (Supplementary Table 20). We sequenced them with an average of >23× genome coverage (Supplementary Table 21), and identified 31,993,474 single nucleotide polymorphisms (SNPs) among the shrimps (Supplementary Table 22), with an average of 19 SNPs/Kb. The SNPs density in the *L. vannamei* genome is much higher than that in chickens<sup>51</sup> and pigs<sup>52</sup>, and similar to that in the oyster, which has extremely high level of heterozygosity<sup>6</sup>. Most SNPs are located at the intergenic regions (86.68%), and the coding regions display much lower genetic diversity than the introns. In the coding regions, 261,056 synonymous and 206,026 nonsynonymous SNPs were identified (Supplementary Table 23). Around 1400 SNPs are located at splice acceptor sites and 1452 SNPs at splicing donor sites. To our knowledge, this represents the largest set of high-quality SNPs obtained from *L. vannamei*, and it would constitute a valuable resource for genetic research and selection.

Significant differences were observed between the genetic diversity of wild shrimp and various broodstocks. The average nucleotide diversity ( $\pi$ ) at the whole-genome level for the two groups was  $3.69 \times 10^{-3}$  and  $2.70 \times 10^{-3}$ , respectively. Phylogenetic analyses based on the whole-genome SNPs separated the wild individuals from those in the broodstocks, which were then grouped into two clades, with the individuals in Ecuador clustered separately from the other commercial breeding lines, apparently because the former only underwent four generations of selection (Fig. 6a).

Comparison of the divergence index ( $F_{ST}$ ) and the  $\theta\pi$  ratios ( $\pi_{wild}/\pi_{breeding}$ ) between the wild and cultured populations allowed us to identify 14 regions (1.66 Mb in size) (Fig. 6b, c), containing 28 predicted genes (Supplementary Table 24), that are under strong selection. GO and KEGG enrichment analyses show that Rho protein signal transduction pathway and glycosphingolipid biosynthesis pathway are enriched. Two copies of *A-kinase anchor protein 13* (*AKAP13*), involved in the Rho protein signal transduction pathway, are strongly selected (Fig. 6c, Supplementary Fig. 29). *AKAP13* is a scaffold protein with guanine exchange factor activity and plays a role in TLR2-mediated NF- $\kappa$ B activation, which is involved in innate immunity<sup>53</sup>. The selected gene in the glycosphingolipid biosynthesis pathway is annotated as lactosylceramide 4- $\alpha$ -galactosyltransferase (*A4GALT*) (Fig. 6c). *A4GALT* is involved in protein glycosylation and protein modification process necessary for synthesis of the receptor for bacterial verotoxins in mammals<sup>54</sup>. Further, one type of anti-lipopolysaccharide factor like protein (ALF) (Supplementary Fig. 29) that is essential to immune defense<sup>55</sup> was also found to be under strong selection. These data illustrate that artificial selective breeding under culture conditions influences the disease resistance of shrimp.

## Discussion

The high-quality genome assembly of *L. vannamei* provides the opportunity to understand various biological processes of shrimp at the genome level. The most prominent characteristic of the coding region of the shrimp genome is the expansion of a series of genes related to visual system, nerve signal conduction, and molting. These gene expansions might have provided the shrimp with its excellent eyesight and rapid nerve signal conduction, better equipping it to adapt to its benthic habitat. Our genome assembly also sheds light on the molting process and provides important evidences for exploring the similarities and differences

between crustaceans and other ecdysozoans, including insects. As a vital process in shrimp, molting is tightly linked to growth, metamorphosis and reproduction. Elucidating the regulatory mechanisms of molting will provide useful clues for genetic analysis of these important processes.

Our genomic analyses have also revealed that almost 30 years of breeding has exerted a significant impact on the genome of the *L. vannamei* broodstocks. An increase in aquaculture production and efficiency can be achieved through genetic improvement of cultured stocks. The assembled shrimp genome and the large amount of SNP markers will provide a useful resource for the application of genome-wide association studies and genomic selection, and thus accelerate genetic improvements in shrimp culture.

## Methods

**Genome sequencing.** The muscle of a single male adult of *L. vannamei* (Kehai No.1 variety, Hainan, China) was used for DNA extraction and genome sequencing. Genomic DNA was prepared using TIANamp Marine Animal DNA Kits (TIANGEN, Beijing, China). The DNA was sheared using a sonication device for paired-end (PE) library construction. Libraries with insert sizes ranging from 200 bp to 20 Kb were constructed according to the instructions provided in the Illumina library preparation kit. All libraries were sequenced on HiSeq 2000 and HiSeq 2500 platform (Illumina, San Diego, USA). These raw reads were subsequently trimmed for quality using Trimmomatic v.0.35 (<http://www.usadellab.org/cms/index.php?page=trimmomatic>). Illumina sequence adaptors were removed, low-quality bases from the starts or ends of raw reads were trimmed, and reads were scanned using a 4-bp sliding-window and trimmed when the average quality per base dropped below 15. The clean reads obtained from this process were used for subsequent analysis.

For PacBio library construction, the genomic DNA of *L. vannamei* was sheared to ~20 Kb and the fragments below 7 Kb were filtered using BluePippin. Filtered DNA fragments were then converted into the proprietary SMRTbell library using the PacBio DNA Template Preparation Kit. In total, ~133 Gb of quality-filtered data were obtained from the PacBio sequencing.

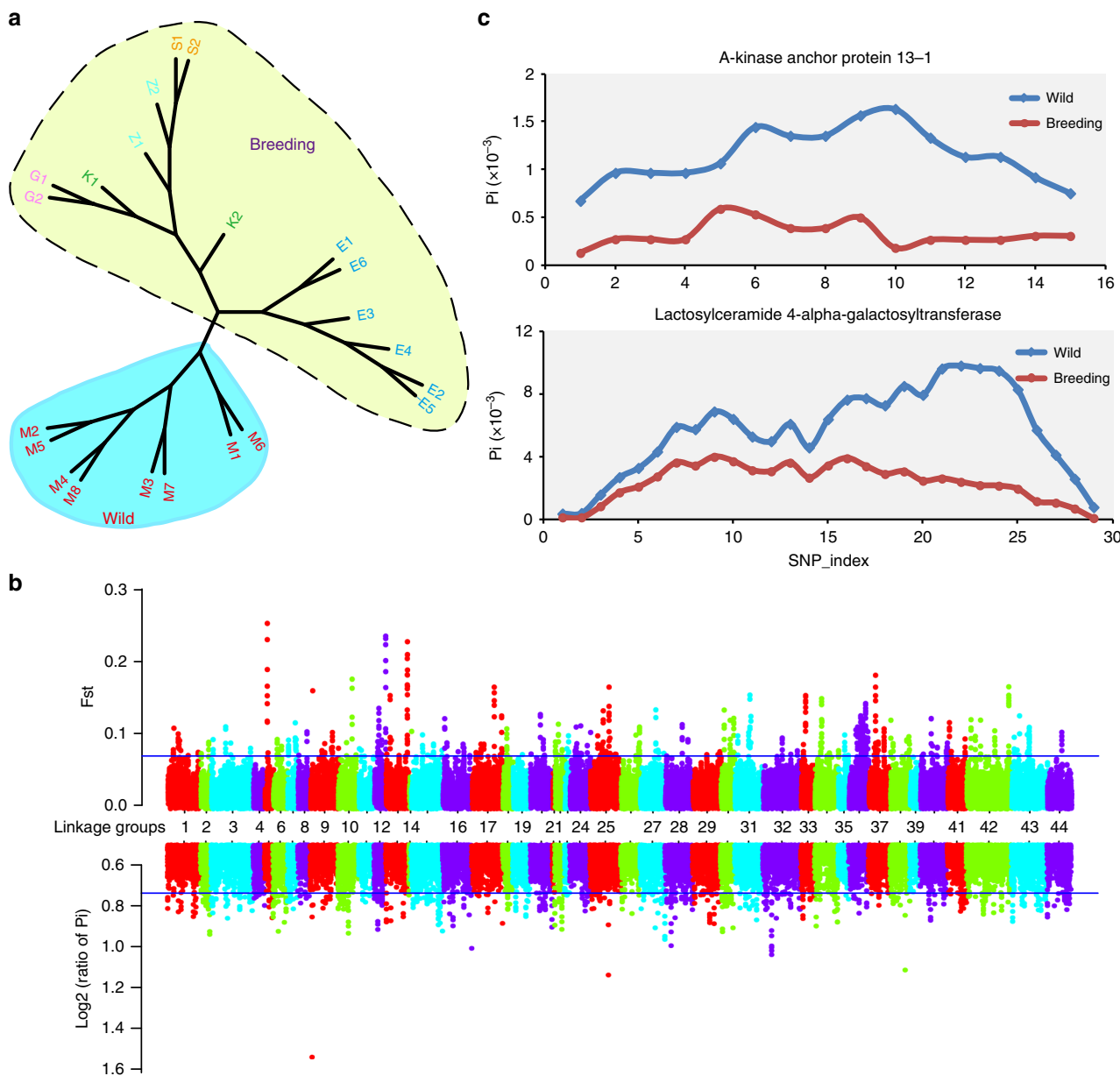
**Genome size estimation.** We adapted a method called Jellyfish<sup>56</sup>, which is based on k-mer distribution, to estimate the genome size with the high-quality reads from the short-insert size libraries (500 bp). We obtained a k-mer depth distribution from the Jellyfish analysis and could clearly observe the peak depth from the distribution data. In order to obtain an estimation of the *L. vannamei* genome size, the following formula was applied: genome size = total\_kmer\_num/kmer\_depth, where total\_kmer\_num is the total number of k-mers from all reads, and kmer\_depth is the peak depth.

**Genome assembly.** All of the subreads from the PacBio sequencing were assembled using the WTDBG software (<https://github.com/ruanjuan/wtdbg-1.2.8>). The assembled sequence was then polished using Quiver (SMRT Analysis v2.3.0) with default parameters. To ensure the high accuracy of the genome assembly, several rounds of iterative error correction were performed using the Illumina clean reads.

Finally, the de novo assembly of the PacBio read sequences was carried out using continuous long reads following WTDBG. The paired-end reads of Illumina (200, 300, and 500 bp libraries) were mapped to polish assembly sequence from Quiver by BWA (<https://github.com/PacificBiosciences/GenomicConsensus>). The SNPs and small insertions and deletions (INDELs) were called and corrected by SAMTools and an in-house script ("SNP\_correction.pl", deposited in [https://github.com/jianbone/L\\_vannamei\\_genome](https://github.com/jianbone/L_vannamei_genome)). Finally, we generated scaffolds and performed gap-filling with SSPACE 3.0 with parameter values of "-x 1 -m 50 -o 10 -z 200 -p 1" using meta-paired sequencing data (Illumina 5, 10, and 20 Kb libraries).

**Quality assessment of genome assembly.** To evaluate the quality of the genome assembly, we mapped partial reads from the short-insert size library back to the scaffolds using Bowtie2 with the following parameters: "--rdg 3,1 --rfg 3,1 --gbar 2"<sup>57</sup>. A total of 93% of the reads could be mapped back to the current assembly. To evaluate the completeness of the genome assembly in genetic regions, we used conserved core genes by running software CEGMA on the assembly of the *L. vannamei* genome<sup>58</sup>.

**Repeat annotation.** Both RepeatModeler and RepeatMasker were used to perform de novo identification and mask of repeats<sup>59</sup>. To ensure the integrity of genes in the subsequent analysis, low complexity or simple repeats were not masked in this analysis, because some of them could be found within the genes.



**Fig. 6** Genetic diversity and selective sweep analyses in *L. vannamei*. **a** The phylogenetic tree of wild (blue) and breeding shrimp (yellow) constructed using whole-genome SNPs. M1–M8 represent the wild shrimp collected from Mexico, E1–E6 represent the breeding shrimp collected from Ecuador, K1 and K2 represent the Kehai No.1 variety from China, G1 and G2 represent Guihai No.1 variety in China, S1 and S2 represent Shrimp Improvement broodstock in USA, Z1 and Z2 represent Charoen Pokphand broodstock in Thailand. **b** Genome-wide view of differentiation ( $F_{ST}$ ) and reduction in diversity ( $\pi_{wild}/\pi_{breeding}$ ) statistics associated with the wild and breeding shrimp. The blue line indicates the top 1% of rank level values for empirical percentiles. **c** The genetic diversity value  $\theta\pi$  in wild and breeding population for the genes encoding A-kinase anchor protein 13-1 (AKAP13-1) and Lactosylceramide 4-alpha-galactosyltransferase (A4GALT). A sliding-window approach (2 Kb windows with 400 bp increments) was used to calculate  $\theta\pi$  of every window for each gene, and the SNP-index indicates the sliding-window of each gene

**Transcriptome sequencing and analyses.** The *L. vannamei* samples at different developmental stages, molting stages, and different adult tissues were collected at the CAS Key Laboratory of Experimental Marine Biology in Qingdao. Total RNA was isolated and purified from all the samples using the TRIzol extraction reagent (Thermo Fisher Scientific, USA), in accordance with the manufacturer’s protocol. RNA quality and concentration were assessed using 1% agarose gel electrophoresis, and RNA concentration was measured using Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, USA). The total RNA of the same tissue was pooled to construct the Illumina sequencing libraries, and then paired-end sequences were generated using Illumina HiSeq 2000 and HiSeq 2500 platform. The clean reads were assembled using Trinity (v2.6.5, <https://github.com/trinityrnaseq/trinityrnaseq/releases>) to obtain transcripts with min\_kmer\_cov set to 2 and all other parameters set to default. The TopHat v1.2.1 package was used to map

transcriptome reads to the *L. vannamei* genome<sup>60</sup>. Cufflinks v2.2.1 (<http://cole-trapnell-lab.github.io/cufflinks/>) was used to calculate expected FPKM (Fragments Per Kilobase of transcript per Million fragments mapped) as expression values for each transcript.

**Gene prediction and annotation.** Protein-coding region identification and gene prediction were performed through a combination of homology-based prediction, de novo prediction and transcriptome-based prediction methods. Proteins from several species, including *D. pulex*, *P. hawaiiensis*, *Drosophila melanogaster*, *Anopheles gambiae*, *Locust migratoria* and *Bombyx mori*, were downloaded from NCBI. Protein sequences were mapped against *L. vannamei* with Exonerate version 2.2.0 (<http://www.ebi.ac.uk/~guy/exonerate/>). The blast hits were used in predicting the

gene structure of the corresponding genomic regions. The de novo prediction software, Augustus v2.5.5, was used to predict the coding regions in the repeat-masked genome. RNA-Seq data were mapped against the assembly using Tophat v2.1.1<sup>60</sup>. Cufflinks v2.2.1 was then used to convert the transcripts from the result of Tophat to gene models. All gene models derived from these three methods were integrated by EvidenceModeler (EVM) into a nonredundant gene set<sup>61</sup>.

Functional annotations of the obtained gene set were conducted using BLASTP with an E-value of  $1e^{-5}$  against the NCBI-NR, SwissProt databases, and KEGG database. Protein domains were annotated by mapping to the InterPro and Pfam databases using InterProScan and HMMER<sup>62,63</sup>. The pathways in which the genes might be involved were derived from gene mapping against the KEGG databases. The Gene Ontology (GO) terms were extracted from the corresponding InterProScan or Pfam results.

**Noncoding RNA prediction.** Noncoding RNA genes were predicted for the repeat-masked genome by sequence- and structure-based alignments with the Rfam noncoding RNA database (<http://xfam.org/>), with an E-value cutoff of  $1e^{-02}$  using Infernal (<http://eddylab.org/infernal/>). Specifically, expression criteria were applied for miRNA identification, and small RNA sequencing data of *L. vannamei* were downloaded from the SRA database of NCBI (SRX2648655, SRX2648646, and SRX682234). Adaptor and primer sequences were trimmed, and low-quality sequences were removed. Clean sRNA reads were compared with the Rfam database (<http://xfam.org/>) to exclude noncoding RNAs other than miRNAs. The remaining sRNA reads were subjected to miRNA identification by mapping to predicted pre-miRNA structures in the *L. vannamei* genome using miRDeep2 and miReap (<http://sourceforge.net/projects/mireap/>). Conserved miRNAs were identified by BLAST alignment to mature miRNA sequences deposited in the miRBase (<http://mirbase.org/>), requiring alignment  $\geq 15$  nt which covers the miRNA seed region (position 2–8 nt) and overall mismatch  $\leq 2$ .

**Gene family analyses.** Gene family analysis was performed using OrthoMCL<sup>64</sup>. The protein-coding genes of *L. vannamei* and nine other arthropods (downloaded from NCBI: [www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)), including *Strigamia maritima*, *Ixodes scapularis*, *Tetranychus urticae*, *P. hawaiiensis*, *D. pulex*, *Acyrtosiphon pisum*, *L. migratoria*, *B. mori*, and *D. melanogaster*, were aligned to each other using the BLASTP program. Similarity information from the pairwise sequence alignments was used as distance parameters for gene family clustering. Species-specific genes were those that had no homologs in the other species used in this analysis. Orphan genes are a group of genes identified from species-specific genes that have no homologs to any genes from NCBI Nr database. However, these orphan genes are supported by transcriptome data.

**Phylogenetic tree construction.** A phylogenetic tree was constructed using the single-copy homologous genes from *D. pulex*, *P. hawaiiensis*, *D. melanogaster*, *A. gambiae*, *L. migratoria*, *B. mori*, *Tribolium castaneum*, *Apis mellifera*, *Zootermopsis nevadensis*, *P. humanus*, *A. pisum*, *I. scapularis*, *T. urticae*, and *S. maritima*, with the RAxML software that is based on the maximum likelihood method<sup>65</sup>. Briefly, the homologous genes from the genomes of these species were clustered together using the OrthoMCL software. Each of these clusters was then filtered to obtain a total of 88 single-copy orthologs using our in-house Python script (`find_multiple_snp.py`). MUSCLE (<http://www.drive5.com/muscle>) was used for sequence alignment, at its default setting. Positions with gaps and missing data were trimmed using an in-house Python script (`allfasta2snp.py`). The final dataset contained 19,161 amino acids and was used to construct the phylogenetic tree with RAxML<sup>66</sup> under the JTT matrix-based model<sup>67</sup>. Initial trees for the heuristic search were obtained automatically by applying neighbor-joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model. The best tree had a log likelihood of  $-3308.02$ , and was used as an input tree for divergence time estimation by Bayesian relaxed molecular clock approaches implemented in MCMCTREE from the PAML package<sup>68</sup>. CODEML from the PAML package was implemented to estimate the neutral substitution rate per year among species. MCMCTREE was then used to calculate gradient vector ( $g$ ) and Hessian matrix ( $H$ ). Fossil calibrations were used as priors for the divergence time estimation<sup>69</sup>. Cauchy distributions were used with default parameters “ $p = 0.1$ ,  $c = 1$ ”. The MCMCTREE is running twice each for 1,000,000 generations with the sample frequency of 50 and a burn-in phase of 50,000 iterations. Tracer (<http://beast.bio.ed.ac.uk/>) was used to assess chain convergence and adequate effective sample sizes of all parameters. The resulting time-calibrated tree was visualized by using MEGA<sup>766</sup>.

**Positive selection.** To assess the contribution of natural selection on the single-copy orthologs in the *L. vannamei* genome, the ratios ( $\omega$ ) of nonsynonymous substitution per nonsynonymous site (dN) to synonymous substitution per synonymous site (dS) were calculated using the software package PAML v4.48a<sup>68</sup>. The homologs were aligned using MUSCLE. The branch-site model was used to detect positive selection along the foreground branch. Likelihood ratio tests were applied to test significant differences between the alternative and null models.

**Regulation of MIH-mediated ecdysone synthesis.** To examine how the cholesterol level and photoperiod regulate molting via MIH, we conducted SREBP gene RNAi, MIH genes RNAi, opsin genes RNAi, Lipitor injection, and cholesterol injection experiments, using dsEGFP RNAi and alcohol/phosphate buffer saline (PBS) injection as negative control treatments. The experiments used healthy shrimps with a body length of  $9.55 \pm 0.25$  cm and a body weight of  $11.20 \pm 1.36$  g, which were cultured and acclimated in fiberglass tanks with aerated circulating seawater for 2 days before the experiment.

To synthesize dsRNA for RNAi experiment, DNA templates were amplified from the eyestalk cDNA sample. The primer pair, dsEGFP-F and dsEGFP-R (Supplementary Table 25), was used to amplify a 289 bp DNA fragment of the EGFP gene from the pEGFP-N1 plasmid. The dsRNA of SREBP containing a fragment of 754 bp located at the 3' end of the coding region, and the dsRNA of MIH1036-6 containing a fragment of 377 bp covering the whole coding region, were amplified using the same method with primers in Supplementary Table 25. The dsRNA of LW opsin gene (*LW-1*) were amplified using two pairs of primers designed using the conserved regions of the genes (Supplementary Table 26). The PCR product (~500 bp) was generated using a T7 promoter anchor attached to the two primers used to amplify the product. The PCR products were purified using Gel Extraction Kit (OMEGA, Japan). dsRNA was synthesized with TranscriptAid T7 High Yield Transcription Kit (Thermo Fisher Scientific, USA), in accordance with the manufacturer's protocol.

We optimized the effective silencing dose of each dsRNA by injecting three different dosages, namely 1, 2, and 4  $\mu$ g, of dsMIH1036-6 or dsSREBP or dsLW-1 or dsEGFP (as control) into healthy shrimps at the intermolt stage (C). Five individuals were tested for each dosage for dsMIH1036-6 and dsSREBP, while each dosage of dsLW-1 was tested on three individuals. At 48 h after dsRNA injection, the eyestalks of the shrimps in each group were collected for RNA extraction. Expression levels of each gene were determined by qPCR. The primers used for qPCR are listed in Supplementary Tables 25 and 26. Primers 18S-qF and 18S-qR were used to detect the expression of the internal reference gene, 18S rRNA. The specificity of the qPCR product was validated by melting-curve analysis performed at the end of each qPCR reaction. The optimal silencing doses were used in subsequent experiments.

To test the effect of SREBP on MIH expression, 20 individuals at the postmolt (P) stage were divided into two groups. In the experimental group, each individual was injected with 1  $\mu$ g of dsSREBP, while individuals in the control group were injected with 1  $\mu$ g of dsEGFP each. To test the effect of high cholesterol level on MIH expression, 20 individuals at the postmolt (P) stage were divided into two groups. Each individual in the experimental group was injected with 400  $\mu$ g cholesterol dissolved in 100% alcohol. The same volume of alcohol was injected into individuals in the control group. For both experiments, the eyestalks of all shrimps in each group were collected at 48 h after injection and total RNA was extracted from different samples using RNAiso Plus reagent (TaKaRa, Japan). Expression levels of three MIH genes were estimated by qPCR, as described above.

To test the effect of opsin on MIH expression, shrimps at the same molting stage (D0–D1) were divided into six groups: three dsLW-1 groups ( $n = 20$ ) and three groups for dsEGFP injection ( $n = 20$ ) as controls. The concentration of dsRNA was 1  $\mu$ g for all groups. We examined the occurrence of molting three times a day after RNAi (at 08:00, 14:00 and 20:00). After silencing for 48, 72, and 96 h, three shrimps from each group were selected to determine the expressions of MIH and three LW opsin genes by qPCR.

During the aforementioned opsin RNAi experiment, we examined the occurrence of molting three times a day (at 08:00, 14:00, and 20:00). At 9 days after silencing, we administered a second injection to ensure the effects of silencing. The observation lasted for 18 days.

To examine the effect of MIH silencing on molting, 20 individuals at the intermolt stage (C) were divided into two groups. Each individual in the experimental group was injected with 4  $\mu$ g MIH1036-6 dsRNA, and each individual in the control group was injected with 4  $\mu$ g EGFP dsRNA. The tails of all the individuals in different groups were dissected for observation of the molting stage under a microscope at 48 h after dsRNA injection. To test the effect of low cholesterol level on molting, 20 individuals at the early premolt stage (D0–D2) were divided into two groups. Each individual in the experimental group was injected with 5  $\mu$ g atorvastatin calcium (Lipitor) predissolved in 20  $\mu$ l PBS (with 2.50% DMSO), and each shrimp in the control group was injected with 20  $\mu$ l PBS (with 2.50% DMSO). The uropods of all the individuals in the two groups were dissected for observation of the molting stage under a microscope at 48 h after injection.

**Resequencing and SNP identification.** A total of 22 shrimp individuals were selected for genome resequencing. These included eight wild individuals collected from Baja California Sur, Mexico, six cultured individuals selected for four generations in Ecuador, and eight individuals from four commercial breeding lines in China, the USA and Thailand (Supplementary Table 20), which had all been subjected to selective breeding for over 20 generations. The raw reads were mapped to the reference genome using BWA v0.7.12<sup>70</sup>. The PCR duplicates of each of the samples were removed with Picard Mark Duplicates. Reads around INDELS from the BWA alignment were realigned using the IndelRealigner option in the Genome Analysis Toolkit (GATK). The HaplotypeCaller of the Genome Analysis Toolkit

GATK was used to construct the general variant calling file (gVCF) for every individual. The individual gVCF files were combined by GenotypeGVCFs function to form a single variant calling file, which includes all the sites. A strict filtering of the SNP calls was carried out using the guidelines given by the Broad Institute using the options: “FS > 60.0 || MQ < 40.0 || QD < 2.0 || SOR > 3.0”. SNPs with minor allele frequency (MAFs) less than 5% were discarded. The missing rate of a genotype for population was 10%. Variants were annotated with the SNPEff annotation program to identify the synonymous and nonsynonymous mutations.

**Identification of candidate genes under breeding selection.** A phylogenetic tree of the 22 individuals was constructed using filtered SNP by the SNPhylo software (parameters: -l 0.6, -m 0.05). To identify the genomic regions significantly affected by artificial breeding, we calculated pairwise nucleotide diversity ( $\pi_{\text{wild}}/\pi_{\text{breeding}}$ ) and divergence index ( $F_{\text{ST}}$ ) between the wild and cultured populations using a sliding-window approach (100 Kb windows with 20 Kb increments) using VCFtools (<http://vcftools.sourceforge.net/>). The genomic regions with top 1% of rank level values for empirical percentiles detected by both methods were identified as potential selective sweeps. GO and KEGG enrichment of genes from the selective sweeps were performed using KOBAS (<http://kobas.cbi.pku.edu.cn/>).

**Code availability.** The associated Perl and Python scripts can be viewed and downloaded free at the github repository: [https://github.com/jianbone/L\\_vannamei\\_genome](https://github.com/jianbone/L_vannamei_genome). The software used for data analyses include: Trimmomatic v.0.35, BluePippin v6.31, Jellyfish-0.3, wtdbg-1.2.8, SMRT Analysis v2.3.0, BWA v0.7.12, SSPACE 3.0, Bowtie 2.3.4.3, CEGMA v3, RepeatModeler Open-1.0, RepeatMasker 4.0.5, Trinity v2.6.5, TopHat v1.2.1, Cufflinks v2.2.1, Exonerate v2.2.0, Augustus v2.5.5, Tophat v2.1.1, EvidenceModeler (EVM) r03062010, InterProScan 5, HMMER-3.0a1, miRDeep2, OrthoMCL v1.4, RAXML Workbench 1.0, MUSCLE 3.8.31, PAML v4.48a, MEGA7, Genome Analysis Toolkit (GATK) 3.6, SNPEff LGPLv3, VCFtools 1.0.0.0, KOBAS 3.0.

## Data availability

Genome sequences data that support the findings of this study have been deposited in NCBI GenBank with the accession codes of [QCYY00000000](https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP043546). RNA-Seq data were used for annotation and biological analyses: NCBI SRA SRR1460493 – SRR1460495 (<https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP043546>), SRR1460504 – SRR1460505 (<https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP043546>), SRX1098368 – SRX1098375 (<https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP061180>). The genome sequences can also be downloaded from the genome database of *L. vannamei* (<http://www.shrimpbases.net/vannamei.html>), where BLAST searching and gene structure visualization are also available. A reporting summary for this article is available as a Supplementary Information file. The source data underlying Figs. 1, 2, and 6b are provided as a Source Data file.

Received: 25 May 2018 Accepted: 7 December 2018

Published online: 21 January 2019

## References

- Dall, W., Hill, B., Rothlisberg, P. & Sharples, D. The biology of the Penaeidae. *Bailliere Tindall Cox* **27**, 1–461 (1990).
- Godin, D. M. et al. Evaluation of a fluorescent elastomer internal tag in juvenile and adult shrimp *Penaeus vannamei*. *Aquaculture* **139**, 243–248 (1996).
- Gao, Y. et al. Transcriptome analysis on the exoskeleton formation in early development stages and reconstruction scenario in growth-moulting in *Litopenaeus vannamei*. *Sci. Rep.* **7**, 1098 (2017).
- Global Aquaculture Production 1950–2016. <http://www.fao.org/fishery/statistics/global-aquaculture-production/query/en> (2018).
- Alcivar-Warren, A., Dunham, R., Gaffney, P., Kocher, T. & Thorgaard, G. First aquaculture species genome mapping workshop. *Anim. Genet.* **28**, 451–452 (1997).
- Zhang, G. F. et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49–54 (2012).
- Berthelot, C. et al. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat. Commun.* **5**, 3657 (2014).
- Xiao, J. H. et al. Signatures of selection in tilapia revealed by whole genome resequencing. *Sci. Rep.* **5**, 14168 (2015).
- Liu, Z. J. et al. The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nat. Commun.* **7**, 11757 (2016).
- Gutekunst, J. et al. Clonal genome evolution and rapid invasive spread of the marbled crayfish. *Nat. Ecol. Evol.* **2**, 567–573 (2018).
- Kao, D. et al. The genome of the crustacean *Parhyale hawaiiensis*, a model for animal development, regeneration, immunity and lignocellulose digestion. *ELife* **5**, e20062 (2016).
- Colbourne, J. K. et al. The ecoresponsive genome of *Daphnia pulex*. *Science* **331**, 555–561 (2011).
- Yu, Y. et al. Genome survey and high-density genetic map construction provide genomic and genetic resources for the Pacific White Shrimp *Litopenaeus vannamei*. *Sci. Rep.* **5**, 15612 (2015).
- Shen, S. Z. et al. Calibrating the End-Permian mass extinction. *Science* **334**, 1367–1372 (2011).
- Fransen, C. H. J. M. & Grave, S. D. *Evolution and Radiation of Shrimp-Like Decapods: An Overview* Vol. 18 (Taylor and Francis/CRC Press, Boca Raton, Crustacean Issues, 2009).
- Kashi, Y. & King, D. G. Has simple sequence repeat mutability been selected to facilitate evolution? *Isr. J. Ecol. Evol.* **52**, 331–342 (2006).
- Yuan, J. et al. Genomic resources and comparative analyses of two economical penaeid shrimp species, *Marsupenaeus japonicus* and *Penaeus monodon*. *Marine Genomics* **39**, 22–25 (2018).
- Kashi, Y. & King, D. G. Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* **22**, 253–259 (2006).
- Wissler, L., Gadau, J., Simola, D. F., Helmkampf, M. & Bornberg-Bauer, E. Mechanisms and dynamics of orphan gene emergence in insect genomes. *Genome Biol. Evol.* **5**, 439–455 (2013).
- Hu, J., Chen, K. & Bao, Z. A study of the structure of compound eyes of the shrimp, *Penaeus chinensis*. *Mar. Sci.* **2**, 30–34 (1997).
- Luchiaro, A. C., Marques, A. O. & Freire, F. A. M. Effects of substrate colour preference on growth of the shrimp *Litopenaeus Vannamei* (Boone, 1931) (Decapoda, Penaeoidea). *Crustaceana* **85**, 789–800 (2012).
- Futahashi, R. et al. Extraordinary diversity of visual opsin genes in dragonflies. *Proc. Natl. Acad. Sci. USA* **112**, E1247–E1256 (2015).
- Wine, J. J. & Krasne, J. The cellular organization of crayfish escape behavior. In *The Biology of Crustacea* (eds Sandeman, D. C. & Atwood, H. L.) 241–292 (Academic Press, New York, 1982).
- Fan, S., Hsu, K., Chen, F. & Ho, B. On the high conduction velocity of the giant nerve fiber of shrimp *Penaeus orientalis*. *Chin. Sci. Bull.* **12**, 51–52 (1961).
- Castelfranco, A. M. & Hartline, D. K. Evolution of rapid nerve conduction. *Brain Res.* **1641**, 11–33 (2016).
- Albertin, C. B. et al. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* **524**, 220–+ (2015).
- Xu, K. & Terakawa, S. Fenestration nodes and the wide submyelinic space form the basis for the unusually fast impulse conduction of shrimp myelinated axons. *J. Exp. Biol.* **202**, 1979–1989 (1999).
- Guary, J.-C. B. & Kanazawa, A. Distribution and fate of exogenous cholesterol during the molting cycle of the prawn, *Penaeus japonicus* bate. *Comp. Biochem. Physiol. Part A: Physiol.* **46**, 5–10 (1973).
- Liebkeind, B. J., Hofmann, H. A., Hillis, D. M. & Zakon, H. H. Evolution of animal neural systems. *Annu. Rev. Ecol. Evol. S* **48**, 377–398 (2017).
- Etzeberria, A. et al. Dynamic modulation of myelination in response to visual stimuli alters optic nerve conduction velocity. *J. Neurosci.* **36**, 6937–6948 (2016).
- Xu, T. L. & Gong, N. Glycine and glycine receptor signaling in hippocampal neurons: diversity, function and regulation. *Prog. Neurobiol.* **91**, 349–361 (2010).
- Chang, E. S., Bruce, M. J. & Tamone, S. L. Regulation of crustacean molting—a multi-hormonal system. *Am. Zool.* **33**, 324–329 (1993).
- Chaharbakshhi, E. & Jemc, J. C. Broad-complex, tramtrack, and bric-a-brac (BTB) proteins: critical regulators of development. *Genesis* **54**, 505–518 (2016).
- Roer, R., Abehsera, S. & Sagi, A. Exoskeletons across the Pancrustacea: comparative morphology, physiology, biochemistry and genetics. *Integr. Comp. Biol.* **55**, 771–791 (2015).
- Nakatsuji, T., Lee, C. Y. & Watson, R. D. Crustacean molt-inhibiting hormone: structure, function, and cellular mode of action. *Comp. Biochem. Phys. A* **152**, 139–148 (2009).
- Chen, H. Y., Watson, R. D., Chen, J. C., Liu, H. F. & Lee, C. Y. Molecular characterization and gene expression pattern of two putative molt-inhibiting hormones from *Litopenaeus vannamei*. *Gen. Comp. Endocr.* **151**, 72–81 (2007).
- Chung, J. & Webster, S. Molt cycle-related changes in bio-logical activity of molt-inhibiting hormone (MIH) and crustacean hy-perglycaemic hormone (CHH) in the crab, *Carcinus maenas*: from target to transcript. *Eur. J. Biochem.* **270**, 3280–3288 (2003).
- Shi, L. L., Li, B., Zhou, T. T., Wang, W. & Chan, S. M. F. Functional and evolutionary implications from the molecular characterization of five spermatophore CHH/MIH/GIH genes in the shrimp *Fenneropenaeus merguensis*. *PLoS ONE* **13**, e0193375 (2018).
- Montagne, N., Desdevises, Y., Soyez, D. & Toullec, J. Y. Molecular evolution of the crustacean hyperglycemic hormone family in ecdysozoans. *BMC Evol. Biol.* **10**, 62 (2010).

40. Covi, J. A., Chang, E. S. & Mykles, D. L. Conserved role of cyclic nucleotides in the regulation of ecdysteroidogenesis by the crustacean molting gland. *Comp. Biochem Phys. A* **152**, 470–477 (2009).
41. Adams, C. M. et al. Cholesterol and 25-hydroxycholesterol inhibit activation of SREBPs by different mechanisms, both involving SCAP and insigs. *J. Biol. Chem.* **279**, 52772–52780 (2004).
42. Brown, M. S. & Goldstein, J. L. The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* **89**, 331–340 (1997).
43. Guo, B., Wang, F., Dong, S. L. & Gao, Q. F. The effect of rhythmic light color fluctuation on the molting and growth of *Litopenaeus vannamei*. *Aquaculture* **314**, 210–214 (2011).
44. Li, F. H. & Xiang, J. H. Recent advances in researches on the innate immunity of shrimp in China. *Dev. Comp. Immunol.* **39**, 11–26 (2013).
45. Denis, M., Thayappan, K., Ramasamy, S. & Munusamy, A. Lectin in innate immunity of Crustacea. *Austin Biol.* **1**, 1001 (2016).
46. Zhang, J. et al. Prophenoloxidase-mediated ex vivo immunity to delay fungal infection after insect Ecdysis. *Front. Immunol.* **8**, 1445 (2017).
47. Wyban, J., Swingle, J., Sweeney, J. & Pruder, G. Specific pathogen-free *Penaeus vannamei*. *World. Aquaculture* **24**, 39–45 (1993).
48. Hake, S. & Ross-Ibarra, J. Genetic, evolutionary and plant breeding insights from the domestication of maize. *eLife* **4**, e05861 (2015).
49. Hickey, J. M., Chiurugwi, T., Mackay, I., Powell, W. & Cgi, I. G. S. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet.* **49**, 1297–1303 (2017).
50. Argue, B. J., Arce, S. M., Lotz, J. M. & Moss, S. M. Selective breeding of Pacific white shrimp (*Litopenaeus vannamei*) for growth and resistance to Taura Syndrome Virus. *Aquaculture* **204**, 447–460 (2002).
51. Rubin, C. J. et al. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587–U145 (2010).
52. Groenen, M. A. M. et al. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393–398 (2012).
53. Shibolet, O. et al. AKAP13, a RhoA GTPase-specific guanine exchange factor, is a novel regulator of TLR2 signaling. *J. Biol. Chem.* **282**, 35308–35317 (2007).
54. Okuda, T. et al. Targeted disruption of Gb3/CD77 synthase gene resulted in the complete deletion of globo-series glycosphingolipids and loss of sensitivity to verotoxins. *J. Biol. Chem.* **281**, 10230–10235 (2006).
55. Guo, S. Y., Li, S. H., Li, F. H., Zhang, X. J. & Xiang, J. H. Modification of a synthetic LPS-binding domain of anti-lipopolysaccharide factor from shrimp reveals strong structure-activity relationship in their antimicrobial characteristics. *Dev. Comp. Immunol.* **45**, 227–232 (2014).
56. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
57. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
58. Parra, G., Bradnam, K. & Korff, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
59. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **Chapter 4**, Unit 4 10 (2009).
60. Trapnell, C., Pachter, L. & Salzberg, S. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105 (2009).
61. Haas, B. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
62. Zdobnov, E. M. & Apweiler, R. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001).
63. Prakash, A., Jeffries, M., Bateman, A. & Finn, R. D. The HMMER Web Server for protein sequence similarity search. *Curr. Protoc. Bioinform.* **60**, 3.15.11–13.15.23 (2017).
64. Li, L., Stoeckert, C. J. Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
65. Stamatakis, A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
66. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
67. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282 (1992).
68. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
69. Oakley, T. H., Wolfe, J. M., Lindgren, A. R. & Zaharoff, A. K. Phylotranscriptomics to bring the understudied into the fold: monophyletic ostracoda, fossil placement, and Pancrustacean phylogeny. *Mol. Biol. Evol.* **30**, 215–233 (2013).
70. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

## Acknowledgements

We acknowledge financial support from the State 863 High Technology R&D Project of China (2012AA10A404 to J.X.), the National Natural Science Foundation of China (31830100 to F.L. and 41876167 to J.Y.), the grants from Qingdao National Laboratory for Marine Science and Technology (MS2017N004 to F.L.), The Senior User Project of RV KEXUE (KEXUE2018G19 to F.L. and X.Z.), the China Agriculture Research system-48 (CARS-48 to F.L.) and the Hong Kong Collaborative Research Fund (CRF) project (No. C4042-14G to K.H.C). We acknowledge the support from High Performance Computing Center, Institute of Oceanology, CAS. We would like to express our gratitude to Prof. Jun Yu, Prof. Songnian Hu, Prof. Jingfa Xiao and Dr. Tongwu Zhang (Beijing Institute of Genomics, Chinese Academy of Sciences), Prof. Ximing Guo (Rutgers University), Dr. Ruiqiang Li, Dr. Zhi Jiang and Dr. Jinbo Zhang (Novogene Bioinformatics Institute), Dr. Junyi Wang and Dr. Dongliang Zhan (Hangzhou 1Gene Ltd), for their support of the shrimp genome project, genome sequencing and assembly. We appreciate Dr. Hao Huang (Hainan Guangtai Ocean Breeding Co., Ltd) for the help with shrimp materials and Dr. Yang Zhang and Dr. Han Cao (Bionano Genomics) help for the construction of optical map. We thank Cui Zhao, Jiankai Wei, Xiaoqing Sun, Jingwen Liu, Jiangli Du, Mingzhe Sun and Yan Zhang for help with DNA, RNA extraction, and data analysis. We are grateful to Prof. Li Sun, Prof. Linsheng Song, Prof. Jinsheng Sun, Prof. Zhaoxia Cui, Prof. Baozhong Liu and Dr. Pin Huan for helpful discussions. We thank other faculty and staff at the Institute of Oceanology, Chinese Academy of Science who contributed to the shrimp genome project. We thank Dr. Jing Qin for revisions of the manuscript. We appreciate Mr. Huawei Zhang for shrimp photography.

## Author contributions

J.X., F.L., Lei Wang and Xiaojun Zhang initiated, managed, and drove the shrimp genome sequencing project. Xiaojun Zhang, C.Z., K.Y., and J.K. collected the animal material. Xiaojun Zhang, Y.S., S.L., Y.G., J.Y., and Y.Y. prepared DNA sequencing and RNA-Seq analysis. J.Y., Y.S., J.R., B.L., X.W., W.L., X. Zhu, and Long Wang performed genome assembly, gene annotation, genome structure analyses, and phylogenetic analyses. Y.G., Xiaojun Zhang, and J.Y. conducted the environment adaptation analysis. S.L., Xiaojun Zhang, Y.G., K.Y.M., and J.Y. conducted the molting regulation analysis. Y.Y., Q.W., and Y.S. conducted the population genetics analysis. S.L., Y.G., X.L., Xiaoxi Zhang, J.Z., and S.J. performed the validation experiments. J.Y., Y.S., and C.L. submitted the genome data. C.L. performed the ncRNA analysis and constructed the shrimp genome database. J.X., F.L., Xiaojun Zhang, J.Y., S.L., Y.G., Y.Y., Y.S., and C.L. wrote the manuscript and additional supplementary files. K.H.C., K.Y.M., J.C., H.Z., Z.L., P.X., A.S., P.S., J.R., A.W., B.L., and Lei Wang revised the manuscript. All authors read and approved the final manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-08197-4>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Journal peer review information:** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019