

EVOLUTIONARY BIOLOGY

"Omics" data unveil early molecular response underlying limb regeneration in the Chinese mitten crab, *Eriocheir sinensis*

Jun Wang^{1†}, Xiaowen Chen^{1†}, Xin Hou¹, Jingan Wang¹, Wucheng Yue¹, Shu Huang¹, Gangchun Xu², Jizhou Yan¹, Guoqing Lu³, Michael Hofreiter^{4*}, Chenhong Li^{1*}, Chenghui Wang^{1*}

Limb regeneration is a fascinating and medically interesting trait that has been well preserved in arthropod lineages, particularly in crustaceans. However, the molecular mechanisms underlying arthropod limb regeneration remain largely elusive. The Chinese mitten crab *Eriocheir sinensis* shows strong regenerative capacity, a trait that has likely allowed it to become a worldwide invasive species. Here, we report a chromosome-level genome of *E. sinensis* as well as large-scale transcriptome data during the limb regeneration process. Our results reveal that arthropod-specific genes involved in signal transduction, immune response, histone methylation, and cuticle development all play fundamental roles during the regeneration process. Particularly, *Innexin2*-mediated signal transduction likely facilitates the early stage of the regeneration process, while an effective crustacean-specific prophenoloxidase system (ProPo-AS) plays crucial roles in the initial immune response. Collectively, our findings uncover novel genetic pathways pertaining to arthropod limb regeneration and provide valuable resources for studies on regeneration from a comparative perspective.

INTRODUCTION

Regeneration, i.e., the process by which organisms replace amputated or damaged tissues or organs, is one of the most interesting yet largely unsolved questions in scientific research (1). Researchers have classified regeneration into different types according to the regeneration ability of the species studied, life stage involved, and the regenerative level required, including whole-body regeneration, structural/limb regeneration, organ regeneration, tissue regeneration, and cellular regeneration (2, 3). Generally, regenerative capacity varies greatly among tissues, life stages, and species involved (4). Thus, whole-body regeneration has been lost in vertebrate species and has, as a natural process, only been observed in invertebrates (3, 5, 6). In vertebrates, structural/limb, organ, and tissue regeneration were reported in fishes, amphibians, and reptiles. Although birds and mammals show limited tissue and cellular regeneration ability, structural/limb regeneration, which requires precise proximal and distal structure reconstruction, is not observed in them (2). However, their potential for it is an active field of medical research.

Thus, although the capacity to regenerate is widespread in animals, it is evident that invertebrate species show overall superior regeneration ability compared to vertebrates (3). Regeneration ability seems to have been weakened over the course of animal evolution,

and animal clades carrying more derived traits also show more restricted regeneration ability. This difference may, in part, be due to different molecular strategies in invertebrate and vertebrate species for regulating the initiation of the regeneration process (3, 7). For example, the origin of pluripotent progenitor cells was found to be different among planarians, cnidarians, arthropods, and vertebrates (7). Of the five genes that have been shown to induce pluripotency in differentiated mammalian somatic cells (*c-Myc*, *Nanog*, *Klf4*, *Oct4*, and *Sox2*), three (*Nanog*, *Klf4*, and *Oct4*) are not present in the *Hydra* genome (5), and only *Sox2* was found in the flatworm, *Macrostomum lignano* (8). Moreover, the planarian *Schmidtea mediterranea* lacks a number of highly conserved genes associated with mitosis and DNA repair, which are essential in the vertebrate regeneration process (9). Last, vertebrates have an endoskeleton, and this needs to be rebuilt during the structural/limb regeneration process. Whereas most invertebrates lack skeletons, arthropods have a prominent exoskeleton, which is, however, regularly renewed during their ontogenesis (7). Thus, it seems that large molecular differences exist between the mechanisms underlying invertebrate and vertebrate regeneration processes (7).

Substantial research has been conducted on regeneration in a number of invertebrates such as planarians and hydra (5, 6, 9). However, arthropods are the largest invertebrate phylum, and limb regeneration occurs commonly in this group, especially in the crustacean subphylum. While research has been conducted on the peculiarities of regeneration in arthropods because of their unique exoskeleton and periodic molting (10–12), the underlying molecular mechanisms of arthropod limb regeneration are not well understood, particularly not in crustaceans (7, 10, 13). Thus, we do not know whether there are any molecular pathways during the limb regeneration process that are different compared to those of or shared with vertebrate species or whether there are any similarities in the molecular control of regeneration between arthropod exoskeleton and vertebrate endoskeleton (13, 14). Research on arthropod limb regeneration should, therefore, shed light on the evolution of regeneration

¹Key Laboratory of Freshwater Aquatic Genetic Resources certified by the Ministry of Agriculture and Rural Affairs/National Demonstration Center for Experimental Fisheries Science Education/Shanghai Engineering Research Center of Aquaculture, Shanghai Ocean University, Shanghai 201306, China. ²Key Laboratory of Freshwater Fisheries and Germplasm Resources Utilization certified by the Ministry of Agriculture and Rural Affairs, Freshwater Fisheries Research Center, Chinese Academy of Fishery Sciences, Wuxi 214081, China. ³Department of Biology, University of Nebraska at Omaha, Omaha, NE 68182, USA. ⁴Evolutionary Adaptive Genomics, Institute of Biochemistry and Biology, Faculty of Science, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany.

*Corresponding author. Email hofreiter.michael@googlemail.com (M.H.); chli@shou.edu.cn (C.L.); wangch@shou.edu.cn (C.W.)

†These authors contributed equally to this work.

abilities and provide a valuable resource for studies of regeneration from an evolutionary perspective.

The Chinese mitten crab, *Eriocheir sinensis*, is native to the Liao River, the Yellow River, and the Yangtze River from the northern to the southern part of China. *E. sinensis* populations are river specific, presenting different growth and development characters regarding morphology, molting frequency, and sexual maturation time (15). *E. sinensis* is famous for its limb regeneration ability and can fully regenerate an autotomized limb as long as it can molt throughout its lifetime (10, 14). This remarkable regeneration ability allows *E. sinensis* to survive well even after substantial injury. It is likely that this trait, together with others such as its strong osmoregulatory capacity and high reproduction rate, has contributed to *E. sinensis* becoming an invasive species with worldwide distribution (16). However, the molecular mechanisms underlying limb regeneration have not been well investigated in *E. sinensis* (17), and genome information remains incomplete for the population in the Yangtze River, the largest distribution area of *E. sinensis*, although a genome was recently published for *E. sinensis* collected from the Liao River (17–19). In this study, we present a chromosome-level genome assembly for *E. sinensis* collected in the Yangtze River using single-molecule Nanopore sequencing, BioNano optical mapping, and high-throughput chromosome conformation capture sequencing technologies. We also obtained large-scale gene expression data of *E. sinensis* during its limb regeneration process. Together, these data allow us to unveil novel molecular pathways underlying the limb regeneration process of *E. sinensis* specifically and of arthropods in general.

RESULTS AND DISCUSSION

Genome sequencing, assembly, and annotation

We reconstructed a chromosome-level genome of *E. sinensis* using Nanopore long-read sequencing and Hi-C technology (fig. S1). Nearly 100 gigabases (Gb) of Nanopore sequence data along with ~90-Gb Illumina Hiseq, 480-Gb BioNano optical mapping, and 300-Gb Hi-C data were generated and assembled into a 1.67-Gb genome, accounting for 94.4% of the estimated genome size (1.77 Gb; Table 1, fig. S2, and table S1) (20, 21). The estimated genome size of *E. sinensis* by K-mer analysis is consistent with estimates from flow cytometric analysis (17). The assembled genome contained 2160 scaffolds, with scaffold and contig N50 values of 16.97 Mb and 717.34 kb, respectively (Table 1 and table S2). Hi-C scaffolding of the genome resulted in 1493 scaffolds anchored to 70 pseudochromosomes (Fig. 1A, Table 1, and table S3). We assessed the completeness and quality of the assembly using both genome and transcriptome datasets. Comparison to the BUSCO database identified 94.65% of genes with complete sequences in the assembly (fig. S3 and table S4) (22). The *E. sinensis* genome contains a high level of repetitive DNA, with 60.01% of the genome identified as repeats (Table 1, Fig. 1A, and table S5) (17, 19). The estimated heterozygosity of *E. sinensis* was 1.7% based on K-mer analysis of short-insert library reads (Table 1). Although many of the above statistics are similar to those of the recently published *E. sinensis* genome (17), our assembly consists of substantially fewer scaffolds, contains fewer gaps, and has a higher level of gene completeness (table S6).

By integrating ab initio prediction evidence, homology prediction evidence, and RNA sequencing (RNA-seq) evidence, we generated 20,286 high-quality protein-coding genes (Table 1 and table S7), a number comparable to those of *Procambarus vigilalis* (21,722),

Table 1. Assembly statistics and features of the *E. sinensis* genome.

Genome assembly statistics	
Estimated genome length	1,767,946,044
Assembled total length (scaffold)	1,767,846,446
Assembled total length (contigs, no gaps)	1,667,381,268
Number of scaffolds	2,160
Longest scaffold length	45,379,147
N50 length (contigs)	717,335
N50 length (scaffold)	16,975,517
Genomic features	
GC content	41.21%
Predicted heterozygosity	1.7%
Repeat content	60.01%
Predicted protein-coding gene number	20,286
Number of scaffolds anchored on pseudochromosomes	1,493

Parhyale hawaiiensis (28,155), *Litopenaeus vannamei* (25,596), and *Homarus americanus* (25,284) in the Crustacea subphylum (23–26). Approximately 90% of the identified genes have homologous proteins in public repositories including NR, SwissProt, RefSeq and Trembl, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Ontology (GO), indicating that most of the annotated genes are present within at least one of the different databases (fig. S4 and tables S8 and S9).

Comparative genomics

To determine the phylogenetic position of *E. sinensis* in the animal kingdom, we performed comparative genomic analyses among 18 metazoan species and, using Orthofinder software, detected 35,718 families of homologous genes. A total of 62 one-to-one single-copy orthologous genes present in all of the 18 metazoan species were identified and used for phylogenetic analysis using maximum likelihood as implemented in RAxML with *Hydra vulgaris* as root (fig. S5). Our results recovered the Ecdysozoa lineage that includes nematoda and arthropoda species, which are reported to be capable of axon, wing imaginal disc, and arthropod-limb regeneration (exoskeleton), as well as Lophotrochozoa, including annelids, and mollusks that also have strong regeneration abilities (fig. S5). Among the gene families, we identified 179 arthropod-specific gene families (with 403 genes in *E. sinensis*), 114 crustacean-specific gene families (with 183 genes in *E. sinensis*), and 340 *E. sinensis*-specific gene families (with 2426 genes in *E. sinensis*; table S10).

To conduct gene family expansion and contraction analysis in the arthropod lineage, we selected eight crustacean species with high-quality genomes and identified 1486 expanded and 1827 contracted gene families in the *E. sinensis* genome (Fig. 1B and table S11). Among them, the expanded gene families were mostly associated with retrovirus-related proteins, zinc finger proteins, transposon proteins, and insect cuticle proteins (table S11). We identified a marked expansion of zinc finger protein families in the *E. sinensis* genome compared with the other species of the crustacean lineage (table S11). Zinc finger proteins are transcription factors with finger-like domains and have been reported to play fundamental roles in

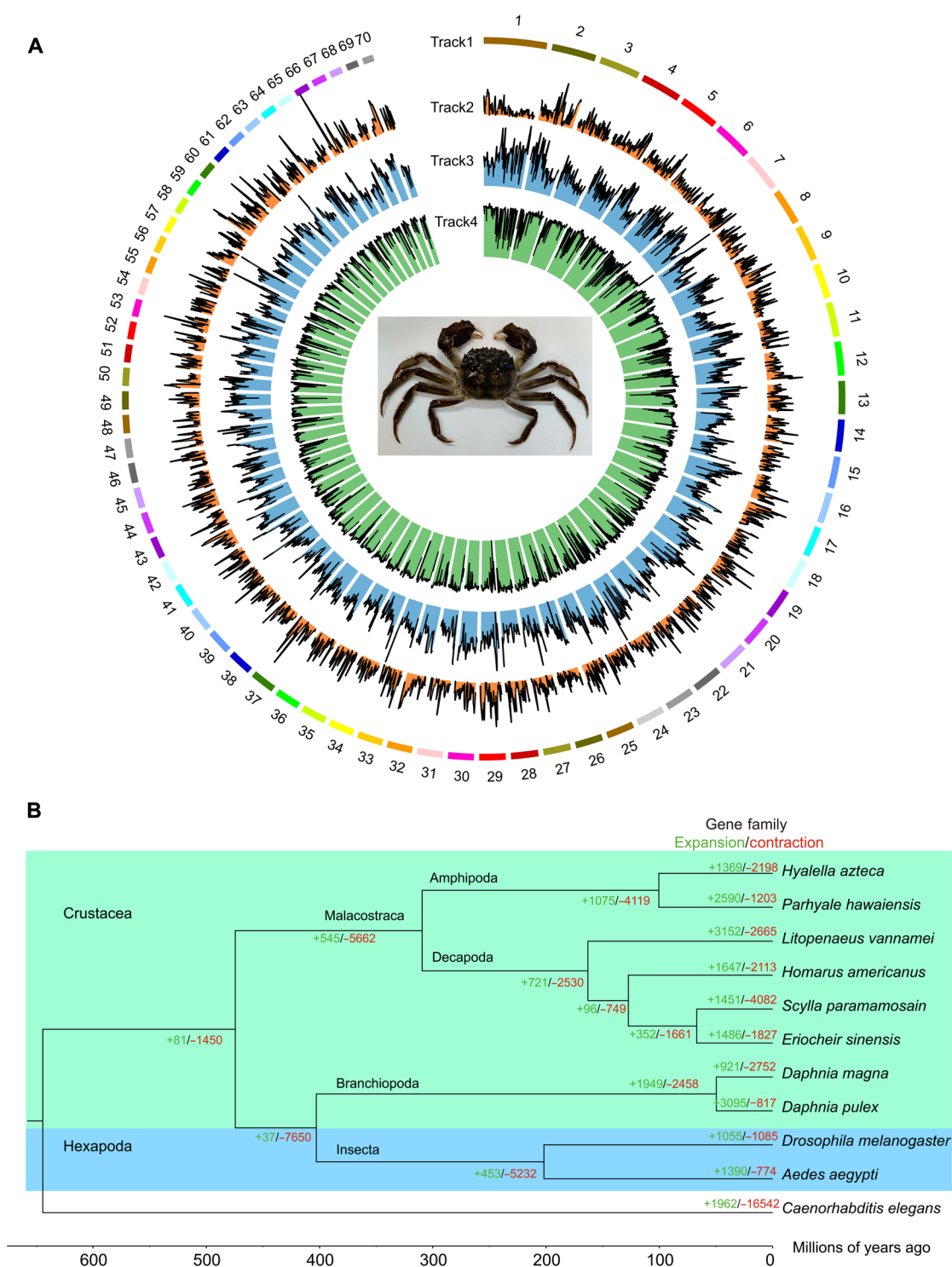


Fig. 1. Characteristics of the assembled *E. sinensis* genome and comparative genomics of *E. sinensis* and other arthropod species. (A) Circos diagram depicting the characteristics of the assembled *E. sinensis* genome. The tracks from outer to inner circles indicate the following features: Track1: chromosomes, 1 to 70 indicates chromosome 1 to chromosome 70; Track2: gene density (from 0 to 30); Track3: repeat element content (from 0% to 100%); and Track4: GC content (from 0 to 60%). All distributions were drawn using a sliding window size of 500 kb. (B) Phylogenetics tree inferred from one-to-one orthologs among the 10 arthropod species with *C. elegans* as root. The numbers represent the number of expanded (green) or contracted (red) gene families. The estimated divergence times are displayed below the phylogenetic tree.

developmental control including in the regeneration process (27). Our following RNA-seq experiment identified a total of 104 zinc finger proteins that showed significantly differential expression, indicating that the expanded zinc finger protein families may have regulatory functions during limb regeneration of *E. sinensis* (table S12). In addition, we identified expanded cuticle-related families with the Chitin_bind_4 (PF00379) and CBM_14 (PF01607) domains in the *E. sinensis* lineage. The two gene families have been suggested to be associated with cuticle formation and exoskeleton calcification (28, 29). A large number of genes from these two gene families were differentially expressed during the limb regeneration process, indicating they probably play specific roles in the limb regeneration process of *E. sinensis* (table S13). However, elucidating the biological function of expanded zinc finger proteins and other expanded gene families during limb regeneration requires further experiments. Among the contracted gene families, we identified 1548 potentially lost gene families in the *E. sinensis* lineage. To further validate these potential gene loss events, we conducted a detailed gene loss analysis and identified 60 potential gene loss events in the *E. sinensis* lineage (table S14). Among others, genes such as *Nrde2*, *Ung*, and *Mpg* associated with DNA damage response, DNA repair, and cell cycle, and genes such as *7B2*, *Syp*, and *Cacng7* associated with nervous system development may have been lost in the *E. sinensis* lineage (fig. S6 and table S14), which indicates that *E. sinensis* may have a specific mechanism of limb regeneration (9). However, the number of gene loss events identified in this study may be an overestimate and should be interpreted cautiously as further improvements of our genome assembly are still needed (there are still gaps in the assembly).

Histo-cellular characteristics of limb regeneration of *E. sinensis*

Previous studies have indicated that there are continuous developmental stages during limb regeneration of crustaceans, involving wound closure, limb bud formation and growth, and new limb growth (10, 14, 30). To gain insights into the limb regeneration process, we observed the entire process for *E. sinensis* under laboratory conditions (Fig. 2, A and B). Wound closure and scab formation finished at 1 dpa (1 day post autotomy), as seen by the development of a melanized scab (S) below the autotomy membrane (AM) at the wound surface (Fig. 2, A and B). Around 4 dpa, *E. sinensis* initiated papilla formation. The morphology of the papilla with the distal tip of the newly regenerated limb (RL) bud became apparent in less than 7 days (7 dpa). During the following days, the limb bud with the characteristic pattern of limb segments was formed, which was clearly visible after nearly half a month post autotomy (13 dpa). The limb bud continued to grow in size and, lastly, the new limb fully regenerated and appeared after molting (30 dpa; Fig. 2A) (11, 31). The regeneration process of the autotomized limb accompanied with molting can be clearly assessed (movie S1). Hematoxylin-eosin (HE) and modified Masson staining tissue sections showed that blastema cells (the blastema is a group of proliferative cells that derived from adjacent epidermal cells, and connective tissues at the wound site) are present at 1 dpa compared with 0 dpa (Fig. 2B), indicating a molecular response of *E. sinensis* rapidly after autotomy. Moreover, abundant regenerated muscle fibers below the scab could be identified already at 2 dpa. During the limb bud growth stage (13 dpa), a complete but small and folded RL was formed under the

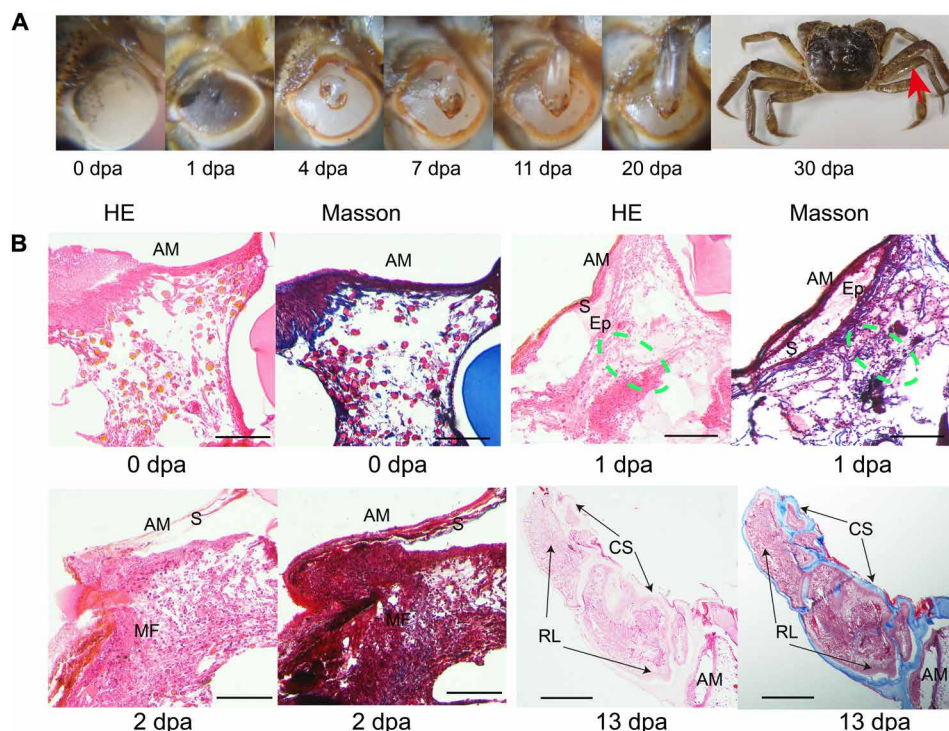


Fig. 2. Phenotypic and cellular features during limb regeneration of *E. sinensis*. (A) Observation of the whole limb regeneration process of *E. sinensis*. The red arrow indicates the regenerated limb after molting. (B) HE and modified Masson staining of cross sections at four time points post autotomy (0, 1, 2, and 13 dpa). AM, autotomy membrane; S, melanized scab; Ep, epidermal cell; MF, muscle fiber; CS, cuticular sac; RL, regenerated limb. The green dashed circle indicates the formed blastema cells. Scale bars for 0, 1, and 2 dpa, 200 μ m. Scale bar for 13 dpa, 500 μ m. In a modification of Masson trichromal staining, blue staining is collagen fiber; red staining is cytoplasm, muscle, nerve sheathe, fibronectin, and hemocyte.

encapsulated cuticular sac, indicating that the characteristic pattern of limb segments was already developed before the next molting (premolt; Fig. 2, A and B).

Expression profiles of the limb regeneration process of *E. sinensis*

During regeneration of the autotomized limb, we collected limb bud tissues at the coxa from different stages of the limb regeneration process and conducted RNA-seq analysis on scab formation stage (1 dpa), limb bud formation and growth stage (13 dpa), and new limb growth stage (30 dpa), with coxal tissues collected at 0 dpa as control (Fig. 2A).

To elucidate the chronological sequence of the regeneration process from a molecular perspective, we performed weighted gene coexpression network analysis (WGCNA) (32). We identified 22 gene coexpression modules in relation to temporal changes during the regeneration process (figs. S7 and S8). Of these modules, we identified and focused on eight modules on the basis of their highest correlation coefficients and significant *P* values (Fig. 3, fig. S8, and table S15). Three modules were dominated by coexpressed genes showing up-regulation at the early stage of the regeneration process (1 dpa; Fig. 3 and table S15). Genes in module 1 such as *Innexins*, *Slc7a5*, *Sox14*, *Ago2*, *Fli1*, *Elf2*, *Wnt8b*, and *Vgfr1* showed substantial up-regulation at 1 dpa. These coexpressed genes in module 1 were enriched in the GO categories response to stimulus, cell migration, and cell differentiation, implying an increasing level of cell communication and cell differentiation as a rapid molecular response to

autotomy (Fig. 3 and fig. S9A). Module 2 includes genes enriched in proteolysis and proteasome-related protein catabolic processes (Fig. 3 and fig. S9B). Generally, the first step of the immune response in crustaceans involves a series of proteolytic processes such as melanization encapsulation (33), and coexpressed genes in module 2 enriched in proteolytic processes may indicate a rapid immune response shortly after autotomy. Moreover, genes in module 3 were enriched in ribosome biogenesis, RNA processing, and protein folding, implying an activated protein and cellular synthesis process at 1 dpa, which may facilitate cell differentiation and cell proliferation (Fig. 3 and fig. S10A). Modules 4 and 5 contained coexpressed genes with increased expression levels later at 13 dpa (Fig. 3). Genes in module 4 were enriched in striated muscle cell differentiation and somatic muscle development, and genes in module 5 were enriched in cell cycle and cell division processes, indicating that muscle cell proliferation and growth processes were activated at this stage (figs. S10B and S11A). We also could find abundant proliferated muscle cells in the RL bud at 13 dpa through histological observation, which likely contribute to the growth of limb buds (Fig. 2B). Genes in module 7 that showed up-regulation at 30 dpa were enriched in chitin metabolic processes and cuticle development, and a large number of cuticle-related genes showed differential expression at this stage (Fig. 3, fig. S11B, and table S15). Last, genes in modules 6 and 8 showed up-regulation at 13 and 30 dpa, respectively. However, these genes were not significantly enriched in specific biological processes; they may play specific roles in the limb regeneration process but requires further research (table S15).

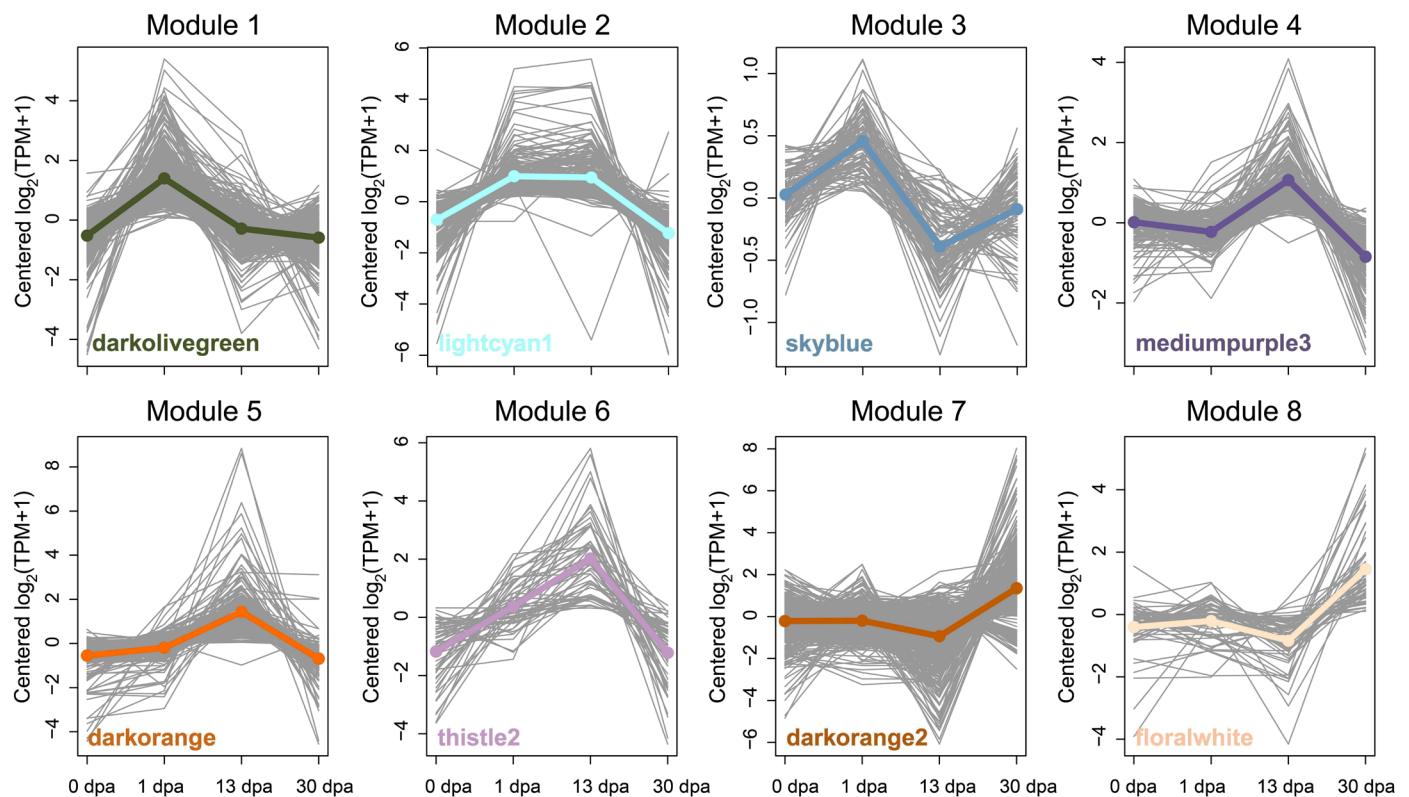


Fig. 3. Coexpression network analysis throughout the limb regeneration process of *E. sinensis*. The expression trajectory of eight modules (modules 1 to 8) showed strong correlation with different regeneration stages throughout the regeneration process by WGCNA analysis. Color labels in Fig. 3 should refer to those of all 22 gene coexpression modules shown in fig. S8.

We next investigated differentially expressed genes (DEGs) during the regeneration process. In line with the gene coexpression network analysis, GO enrichment analysis on all the DEGs ($P < 0.05$, fold change >2) indicated distinct biological processes were involved at different limb regeneration stages (fig. S12 and table S16). Specifically, genes associated with rapid molecular responses in terms of immune response, signal transduction, cell migration, and differentiation were up-regulated at 1 dpa (figs. S12B and S13). Cell cycle-related genes involved in cell proliferation and growth processes showed up-regulation at 13 dpa (figs. S12B and S14). For the 30-dpa stage, large, arthropod-specific, and expanded cuticle-related gene families participated in this process, most likely facilitating the arthropod epidermis and exoskeleton formation (figs. S12B and S15).

Overall, the DEG temporal gene expression profiles show a statistically significant enrichment for arthropod-specific genes. Thus, the proportion of arthropod-specific genes in DEGs was significantly higher than that of arthropod-specific genes in non-DEGs ($P < 0.001$, chi-square test; fig. S16). Nearly 50% (190 of 403) of the arthropod-specific genes from *E. sinensis* were differentially expressed during the limb regeneration process (table S10). GO enrichment analysis indicated that the 190 differentially expressed arthropod-specific genes are enriched in neuron axon guidance, Toll signaling pathway, cuticle development, cell morphogenesis involved in differentiation, and several other processes (fig. S17A and table S10). Last, 65 of 183 crustacean-specific genes were differentially expressed, and these genes are, among others, enriched in central nervous system formation, immune response, and response to stimulus during the limb regeneration process (fig. S17B and table S10). The increased numbers of differentially expressed arthropod- and crustacean-specific genes during limb bud development suggest that *E. sinensis* likely evolved lineage-specific mechanisms for limb regeneration relative to vertebrate species.

Novel *Innexin* gene mediated signal transduction during the early stage of limb regeneration

Innexin genes encode proteins that form gap junctions, i.e., the membrane channels that directly mediate cell-cell communication allowing ions, amino acids, or adenosine triphosphate to directly pass through a pore between cells (34). This mechanism mediates a rapid response and is assumed to play essential roles in regulating cell migration, cell differentiation, and cell fate during development in invertebrate species (35). We investigated the *Innexin* genes of *E. sinensis* in terms of gene structure, evolution, and expression patterns in the limb regeneration process (Fig. 4A, figs. S18 to S20, and table S17). A total of 16 genes have clearly identified *Innexin* domains, transmembrane domains, and conserved motifs in the gene structure (figs. S18 and S19 and table S17) (36). We also detected expression of *Innexin* genes (transcripts per million mapped reads, TPM >1) in different biological processes (molting, aerial respiration, and metamorphosis) using our previous RNA-seq data (29, 37, 38), indicating that *Innexin* genes were accurately annotated and functional (fig. S21). The ML phylogenetic tree of *Innexin* genes showed clearly distinguishable clades among selected crustacean species, and we identified a larger *Innexin* gene family in crustacean species compared with *Drosophila melanogaster* (fig. S20).

Considering the biological function of *Innexin* genes, they may show early differential expression after limb autotomy in *E. sinensis*. Further analyses of DEGs showed that six *Innexin* genes were indeed significantly up-regulated after autotomy at 1 dpa compared with

0 dpa (fold change >2 , $P < 0.05$; Fig. 4A and table S18). By conducting limb regeneration RNA-seq experiments on other crustaceans, we found that eight *Innexin* genes in *L. vannamei*, six *Innexin* genes in *Macrobrachium rosenbergii*, and four *Innexin* genes in *Macrobrachium nipponensis* were substantially up-regulated at 1 dpa compared with 0 dpa (fig. S22 and table S19). It has been reported that *Innexin* genes could be up-regulated by a Ca^{2+} wave (35). We identified slightly up-regulated Ca^{2+} concentrations and *Calmodulin* (*CaM*) gene expression at 1 dpa, indicating that the up-regulation of *Innexin* genes could indeed be caused by a Ca^{2+} wave induced by autotomy (fig. S23). Among the differentially expressed *Innexin* genes, we found that Esin.LG37.0025 (annotated as *Innexin2*, *Inx2*) was the most significantly and specifically up-regulated ($\log_2\text{FC} > 6$, $P < 0.001$) during the limb regeneration process (Fig. 4A and table S18), and quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) results from a narrow timeline indicated that *Inx2* expression was up-regulated within less than 12 hours after autotomy (fig. S23). To check whether *Inx2* (Esin.LG37.0025) is involved in other essential biological processes of *E. sinensis* or only relevant for the limb regeneration process, we used our previously published RNA-seq data (29, 37, 38) from different molting stages, metamorphosis, and aerial respiration and analyzed the expression of *Innexins* in these data. We found that *Inx2* only showed significantly differential expression during the limb regeneration process, which suggests that *Inx2* plays novel, specific, and essential roles in the limb regeneration process (Fig. 4A and fig. S21). In addition, our WGCNA coexpression network analysis indicated that the *Inx2* gene was a hub gene and presented strong coexpression with genes annotated as transmembrane transport proteins at 1 dpa (fig. S24). The coexpression network simulation showed that *Inx2* interacted with the LAT1 protein, which is encoded by the *Slc7a5* gene (fig. S24). *Slc7a5* is mainly responsible for amino acid transport involved in the mTORC1 signaling pathway, which plays essential roles in controlling cell regeneration, growth, and development by sensing major signals (e.g., amino acids) (39, 40). Using RNA fluorescence in situ hybridization (FISH) analysis, we also found that *Inx2* and *Slc7a5* were highly coexpressed in the epidermal cells and part of the blastema cells below the melanized scab at 1 dpa, indicating that the two genes may regulate cell communication and blastema cell formation during the early stage of the limb regeneration process in a coordinated way (Fig. 4B).

To further study the function of *Inx2*, we conducted RNA interference (RNAi) of the *Inx2* gene. Our results confirmed that most *E. sinensis* individuals in the *Inx2* RNAi treatment control group could regenerate papillae within 5 days (57%, 27 of 47 individuals; Fig. 4, C and D), whereas most individuals in the experimental group either died or failed to regenerate the papilla. Only 3.8% (2 of 53 individuals) in this group could regenerate their papilla within 5 days, and most individuals failed to form melanized scab, in which case the wound surface was infected by bacteria, leading to 68% (36 of 53 individuals) of individuals dying (Fig. 4, C and D). We also verified that *Slc7a5* was significantly down-regulated after *Inx2* RNAi, indicating that *Inx2* may play a role in regulating the *Slc7a5* gene as the WGCNA coexpression network results suggested (Fig. 4E).

Our RNA-seq data also showed that *Slc7a5*, *V-ATPase*, *Ragulator*, and *Rag* genes in the mTORC1 pathway were up-regulated after autotomy (fig. S25). To evaluate the relevance of the mTORC1 pathway for the regeneration process, we conducted mTORC1 pathway

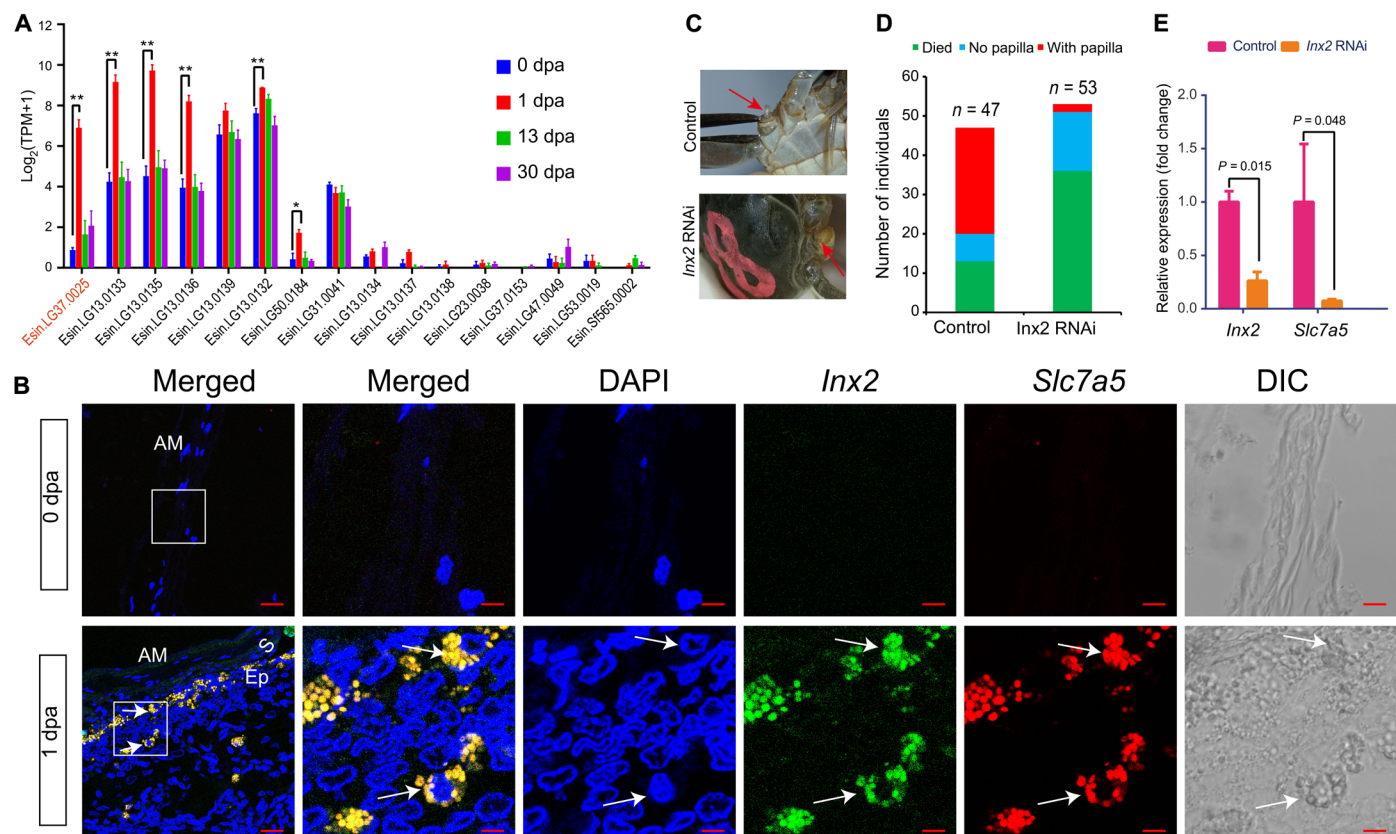


Fig. 4. *Innexin 2* regulated limb regeneration process of *E. sinensis*. (A) Expression patterns of 16 identified *Innexin* genes in the limb regeneration process. * $P < 0.05$ and fold change > 2 , ** $P < 0.01$ and fold change > 2 . (B) Fluorescence in situ hybridization of *Inx2* and *Slc7a5* at 0 and 1 dpa. Region in the white box on the left was magnified and presented on the right. The white arrow indicates that the positive signal of *Inx2* and *Slc7a5* was possibly distributed in the cell membrane of specific cells (epidermal and blastema cells). Scale bars, 20 μm for the two panels on the left and 5 μm for the other panels. DIC, differential interference contrast figure. (C) Phenotypic character of autotomized limb after RNA interference (RNAi) of the *Inx2* gene at 5 dpa. The arrow indicates the autotomy site with clear papilla formation in the control group. However, no papilla was formed in the *Inx2* RNAi group. (D) Ratio of papilla formation after autotomy with RNAi of the *Inx2* gene at 5 dpa. (E) Expression of the *Inx2* and *Slc7a5* genes after RNAi of the *Inx2* gene. Statistical significance ($P < 0.05$) was determined using the Student's *t* test.

rapamycin inhibition experiments. The length of the regenerated papilla in the experimental group (injected with rapamycin) was 1.74 ± 0.17 mm significantly smaller than that in the control group [2.84 ± 0.15 mm, injected with dimethyl sulfoxide (DMSO)] at 9 dpa ($P < 0.01$), strongly indicating that the mTORC1 pathway may be essential for the limb regeneration process of *E. sinensis* (fig. S26). Together, the results indicate that *Inx2* may regulate the limb regeneration process through the mTORC1 signaling pathway.

Enhanced and effective regenerative-immune response of *E. sinensis* during the early stage of limb regeneration

After limb autotomy/injury, quick and effective immune response and wound healing are the key biological processes to stop bleeding and bacterial or fungal infection, and thus the crucial processes for a successful limb regeneration process (41). It is well known that invertebrates lack highly evolved adaptive immunity. Therefore, they rely on a fixed and unspecific immunity to prevent pathogen entry and spread (42). Among the diverse array of the humoral immune response such as clotting cascade, antioxidant defense enzymes, and antimicrobial peptides, the most potent humoral immune response of invertebrates against pathogens is cellular melanotic encapsulation (42). By integrating comparative genomics analyses,

we found that several expanded, arthropod-specific gene families involved in the melanotic encapsulation response such as phenoloxidases (*PPO*) or phenoloxidase activating factors (*PPAF*) showed substantially differential expression during the early stage (1 dpa) of the limb regeneration process (Fig. 5 and fig. S27). *PPO* and *PPAF* have been reported to be involved in the melanization process, which is a prominent arthropod humoral response for encapsulation and killing invading pathogens as it allows a rapid response to pathogen infection and physical injury and initiates wound healing (33). This melanization process is performed by *PPO* and controlled by the prophenoloxidase (ProPO) activation cascade (42). Our results showed typical melanization was formed at the wound site of *E. sinensis* in less than 24 hours after autotomy, and large numbers of granulocytes aggregated at the wound site shortly after autotomy (2 hours post autotomy), which may activate the immune response and initiate the formation of scab rapidly (Fig. 5A) (30). *PPO* belongs to the hemocyanin family; we identified a total of 22 hemocyanin genes in the *E. sinensis* genome. Of these, 10, 8, and 4 belong to the hemocyanin, cryptocyanin, and *PPO* subfamilies, respectively. We found that the *PPO* subfamily was substantially up-regulated at 1 dpa, indicating *PPO* may play essential roles in the immune response at the early stage of limb regeneration in

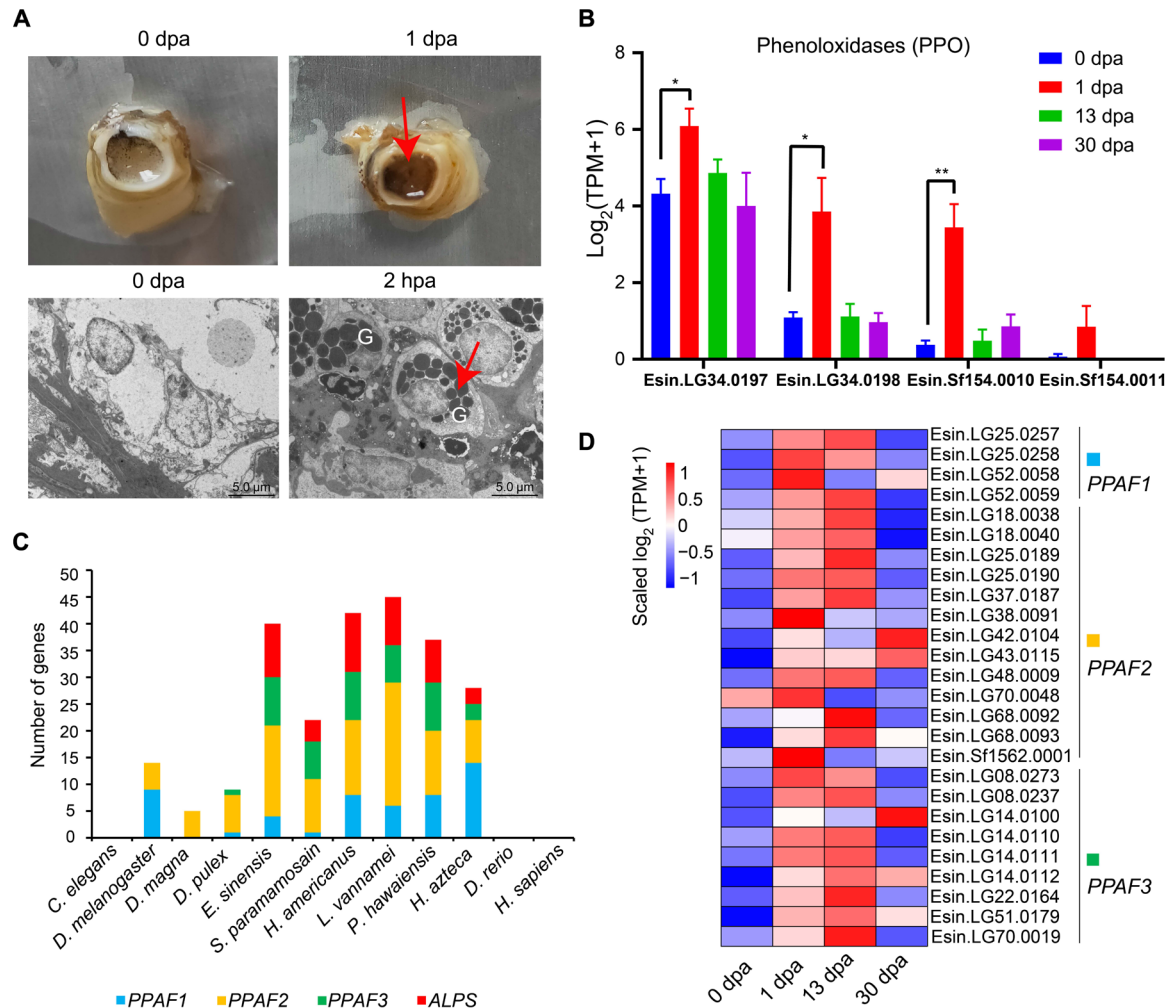


Fig. 5. Effective immune response during the limb regeneration process of *E. sinensis*. (A) Phenotype of the coxa after autotomy at 0 and 1 dpa and transmission electron microscopy (TEM) in situ observation of autotomized limb at 0 and 2 hpa (hours post autotomy). Red arrow in the top panel indicates typical melanization at the autotomy site at 1 dpa. Red arrow in the bottom panel indicates aggregated granulocytes (G). (B) Expression patterns of four PPO genes during the limb regeneration process; * $P < 0.05$ and fold change > 2 , ** $P < 0.01$ and fold change > 2 . (C) Number of arthropod-specific PPAF and ALPS genes associated with immune response in the *E. sinensis* lineage compared with other species. (D) Expression pattern of 26 differentially expressed arthropod-specific PPAF genes during the limb regeneration process.

E. sinensis (Fig. 5B and fig. S27). We also identified three PPAF families (PPAF1, PPAF2, and PPAF3) only present in the arthropod lineage, and we identified a total of 30 PPAF, with 26 of these showing differential expression during the whole limb regeneration process. Most of these genes were highly expressed at 1 and 13 dpa (Fig. 5, C and D).

The initial ProPO pathway activation is triggered upon the recognition of lipopolysaccharide (LPS) from Gram-negative bacteria, peptidoglycan from Gram-positive bacteria, and beta-1,3-glucans from fungi, which is mediated by pattern or pathogen recognition receptors (PRPs) (42). We found that lipopolysaccharide and beta-1,3-glucan binding (LGBP) genes of the PRP family were up-regulated at 1 dpa (fig. S28). In addition to LGBP genes, we also identified crustacean-specific anti-LPS factor (ALPS) genes as up-regulated at 1 dpa (Fig. 5C and fig. S28). ALPS genes have also been reported to be functional in the immune response of crustaceans, having a strong antibacterial effect and potentially participating in the ProPO system as a PRP (43, 44). We found that the ALPS gene

family is expanded in *E. sinensis*, and 3 of 10 ALPS genes identified showed differential expression (Fig. 5C and fig. S28). *C-lectin* is assumed to be a further PRP in crustaceans, and we identified a total of 83 genes annotated to have C-lectin domains with PFAM search (PF00059), 20 of which showed differential expression during the limb regeneration process (table S20). Besides the ProPO system genes, genes participating in other humoral immune response processes, such as hemolymph coagulation and clot formation, also showed differential expression during the early stage of limb regeneration. Seven clotting factor B (CFB) genes and six proclotting enzyme (PCE) genes were up-regulated at 1 dpa compared to 0 dpa, indicating that these genes play essential roles in a rapid and effective immune response (figs. S28 and S29). Thus, the effective and rapid immune response, which plays fundamental roles in wound healing, scar formation, and defense against pathogen invasion after autotomy, is driven by expanded gene families and differentially expressed arthropod-specific genes in *E. sinensis*.

Potential role of histone methylation during the early stage of limb regeneration

A key question for limb regeneration is how cells initiate gene reprogramming and redifferentiation for regeneration. The SMYD gene superfamily (both SET and MYND domain-containing proteins) plays essential roles in histone methylation, which is associated with animal development (45). As we found the arthropod-specific SMYDA gene family, which belongs to the SMYD gene superfamily, to show strong differential expression in limb bud tissue during the limb regeneration process (fig. S30 and table S10), we investigated whether histone methylation may be involved in arthropod-limb

regeneration. We discovered 39 genes, belonging to different gene families, with a conserved SET domain and annotated as histone methyltransferases in the *E. sinensis* genome (Fig. 6A). During the *E. sinensis* limb regeneration process, 10 histone methyltransferases were significantly down-regulated at 1 dpa compared with 0 dpa (fold change >2 , $P < 0.05$; Fig. 6A), with 8 of these belonging to the SMYDA gene family, while 2 belong to the SMYD4 gene family (Fig. 6A). In total, 11 SMYDA genes were identified in the *E. sinensis* genome, with 8 of them showing significant down-regulation ($P < 0.05$) at 1 dpa (Fig. 6A and fig. S30). The other three SMYDA genes also showed a down-regulation trend, although with no

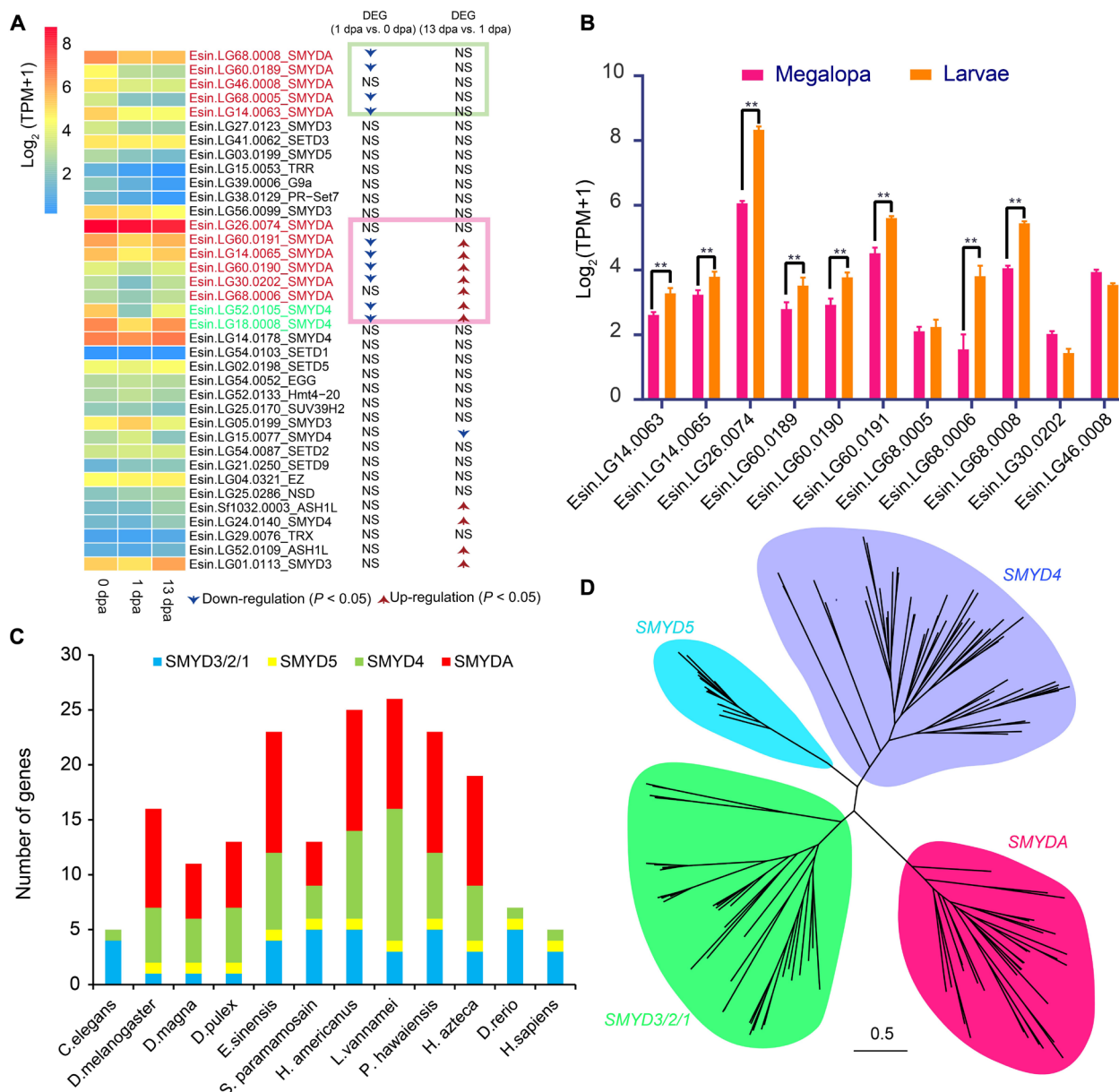


Fig. 6. Arthropod-specific SMYDA gene family during limb regeneration and metamorphosis processes. (A) Expression patterns of the identified histone methyltransferases with SET domain in the limb regeneration process. Red arrows indicate up-regulation, and blue arrows indicate down-regulation. NS, not significant. (B) Expression of identified SMYDA genes during metamorphosis with most of them showing down-regulation in the megalopa stage. $**P < 0.01$ and fold change >2 . (C) Distribution pattern of SMYD genes in representative arthropod and vertebrate species with the SMYDA gene family being present only in the arthropod lineage. (D) Maximum-likelihood phylogenetic tree of the SMYD gene family in the investigated metazoan species.

statistical significance, indicating that the whole *SMYDA* gene family was rapidly down-regulated in response to the autotomy signal (i.e., at 1 dpa; Fig. 6A and fig. S30). However, after the early response stage (1 dpa), four of the previously down-regulated *SMYDA* genes and the two previously down-regulated *SMYD4* genes were up-regulated during the limb bud growth stage (13 dpa) compared with 1 dpa, indicating that precise and coordinated histone methylation regulation may be critical for limb regeneration in *E. sinensis* (Fig. 6A). We also identified the same down-regulation trend of *SMYDA* genes in *L. vannamei* and *M. rosenbergii* but not in *M. nipponensis* after autotomy at 1 dpa compared with 0 dpa (fig. S31 and table S19). Furthermore, when we investigated our previously published RNA-seq data from whole individuals (29, 37, 38), we found that the *SMYDA* gene family also showed differential expression during metamorphosis of *E. sinensis*. This process also involves notable muscle cell differentiation, cuticle development, and morphogenesis, when megalopae turn into larvae with notable morphological changes (Fig. 6B) (29). Moreover, we investigated the expression of the *SMYDA1* (*Esin.LG14.0065*) gene through *in situ* hybridization. It was expressed in normal epidermal cells under the AM and muscle cells at 0 dpa. In contrast, reduced signals were identified in muscle cells, and positive signals were identified in epidermal and blastema cells under the scab at 1 dpa (fig. S32). Our results suggest that the *SMYDA* gene family may play specific roles during the limb regeneration process.

It has been reported that members of the *SMYD* gene family are conserved during evolution (45), which is also supported by our results that *SMYD1/2/3*, *SMYD4*, and *SMYD5* are present in almost all the species in this study (Fig. 6C). However, the *SMYDA* gene family is present only in arthropod species and absent in other Ecdysozoa lineages such as nematodes represented by *Caenorhabditis elegans*, indicating that the *SMYDA* gene family originated and diversified in arthropods (Fig. 6, C and D) (46). More intriguingly, like the *SMYDA* genes, the two *SMYD4* genes that showed dynamic expression in the limb regeneration process are present only in arthropods (fig. S33), suggesting that limb regeneration in *E. sinensis* is specifically regulated by arthropod-specific histone methyltransferases (*SMYDA* and *SMYD4*). Histone methylation has been reported to be a major regulator of gene expression in regeneration processes, and the specific *SMYDA* gene family together with the arthropod-specific *SMYD4* gene family may provide a unique epigenetic regulation mechanism for cell differentiation, proliferation, and development in arthropods and, at the same time, be responsible for the rapid and novel early molecular response for limb regeneration in *E. sinensis*. However, future work is required to validate the biological function of the *SMYDA* and *SMYD4* gene families during the limb regeneration process.

Overall, our studies revealed a series of novel genetic pathways during limb regeneration of *E. sinensis* compared with vertebrate species. Early molecular responses in terms of rapid cell signal transduction regulated by *Innexin* and other genes, effective immune responses driven by expanded and specific gene families in the ProPo pathways, and potential evidence for dynamic epigenetic regulation by the arthropod-specific *SMYDA* gene family mainly arise within less than 24 hours after autotomy (1 dpa; fig. S34). Later developmental programs such as cell proliferation and the limb bud growth process are present at 13 dpa stage with related cell cycle genes up-regulated (fig. S34). Moreover, large arthropod-specific cuticle-related genes facilitating arthropod epidermis and exoskeleton formation

are active at around 30 dpa (fig. S34). Multiple expanded and arthropod-specific gene families involved in limb regeneration indicate that the genetic basis of arthropod-limb regeneration may be unique compared to vertebrates and to other invertebrate groups such as planarians.

MATERIALS AND METHODS

Sample collection and genome sequencing

A male *E. sinensis* (316.2 g) was collected from Chongming Island breeding station off Shanghai in 2019 (Shanghai, China). The muscle tissue used for genome sequencing was collected and stored in liquid nitrogen before DNA extraction for Nanopore and Hi-C sequencing. Hemolymph, heart, testis, gill, hepatopancreas, ganglia thoracalia, and intestines were also collected for BioNano optical map construction. Our study, sampling procedures, and experimental protocols were approved by the Institutional Animal Care and Use Committee of Shanghai Ocean University (Shanghai, China) on the care and use of animals for scientific purposes. For Nanopore sequencing, high-quality DNA was extracted from muscle tissue using the Qiagen Genomic Kit. DNA concentration and purity were measured with NanoDrop and Qubit instruments, respectively. A total of 11 Nanopore sequencing libraries were constructed with insert size >20 kb and sequenced on the Oxford Nanopore Technology (ONT) sequence platform (PromethION) with 11 Flow Cells. Guppy v3.2.2 was used as ONT basecaller (parameter: -c dna_r9.4.1_450bps_fast.cfg) to process raw sequencing data, and passed reads were filtered with mean_qscore_template >7 (<https://github.com/nanoporetech/taiyaki>). After quality filtering, a total of 99.1 Gb passed sequencing reads was obtained for de novo assembly. For Illumina sequencing, sequencing libraries with 400–base pair (bp) insert size were sequenced on the Illumina HiSeq XTen platform with paired-end mode (150-bp read length).

For BioNano optical mapping, high-molecular weight DNA was extracted from heart tissue of the same *E. sinensis* male used for ONT sequencing according to the BioNano Prep Animal Tissue DNA Isolation Soft Tissue Protocol-30077. The DNA was washed with HB buffer, and DNA >250 kb was embedded in an agarose layer. It was then digested with DLE-1 single enzyme and, lastly, analyzed on the BioNano Saphyr platform to obtain optical mapping data. To obtain clean data, the raw BioNano data were filtered so that only molecular lengths >150 kb and with molecule minSites >9 were kept. After filtering, 442.0-Gb clean data with an average label density 11.00/100 kb (N50 = 319.9 kb) was obtained for the following assistant assembly. To obtain the chromosome-level assembly, another assistant assembly was conducted using Chromosome Conformation Capture-3C (Hi-C) sequencing technology (20, 21). Hi-C libraries were quantified and sequenced on an Illumina NovaSeq 6000 instrument (paired-end mode, PE150). The Hi-C library generated 321.06-Gb data (figs. S1 and S2 and table S1). Raw sequencing reads were filtered using fastp 0.12.6 software (47). After filtering, 300.47-Gb clean data were obtained for the following analysis steps.

De novo genome assembly

Before de novo genome assembly, a k-mer algorithm was applied to evaluate the *E. sinensis* genome size using Jellyfish (48). A total of 81.2-Gb Illumina clean data generated from short-insert size libraries (400 bp) were used for this analysis. The following formula was

applied to estimate the *E. sinensis* genome size: $G = K_{\text{num}}/K_{\text{depth}}$, where G indicated the estimated genome size, K_{num} indicated the number of K-mer, and K_{depth} indicated the expected depth of K-mer.

The genome was de novo assembled based on quality filtered Nanopore sequencing data (81.7 Gb) using the NextDenovo v2.0-beta.1 software (<https://github.com/Nextomics/NextDenovo>). First, the NextCorrect core module was used to correct the long raw reads with sequencing errors to get error-corrected reads, and then the NextGraph core module was used to assemble the genome with corrected reads (-a 0 -n 1832 -Q 0 -I 0.68 -S 0.61 -N 1 -r 0.43 -m 4.32 -C 81 -z 16) to obtain a preliminary assembly. Last, the Nextpolish v1.0.5 software was used to polish the preliminary assembly with both long ONT reads and Illumina reads with default parameters (49). BioNano Optical Map and Hi-C technologies were used to aid chromosome-level assembly based on the polished genome assembly. Scaffolding of the contigs with optical mapping was performed using the BioNano optical mapping technology (BioNano Genomics). First, BioNano clean data were de novo assembled, and then the polished genome assembly was mapped to the BioNano optical map to get longer super scaffolds. Last, a chromosome-level assembly was performed using Hi-C. The paired-end Hi-C reads were uniquely mapped onto the assembly contigs, chromosome agglomerative hierarchical clustering was conducted using Lachesis software, and the final assembly was obtained after Hi-C clustering (fig. S2) (50).

Quality assessment of the genome assembly

To assess the quality of our genome assemblies, we evaluated the genome assemblies using sets of benchmarking universal single-copy orthologs (BUSCO) with genome mode and lineage data from Arthropoda with BUSCO v3.1.0 (22). We mapped Illumina short-insert size library reads back to the assembly with bwa 0.7.12-r1039 (51). Mapping statistics were summarized with samtools v1.4 (52) and bcftools v1.8.0 (53). A total of 99.03% of the Illumina reads could be mapped back to the current assembly. ONT sequencing reads were mapped back to the current assembly by minimap2 v2.15 with a mapping rate of 99.86% (54).

Genome annotation

Repeat annotation

De novo repeat annotation of *E. sinensis* was carried out for tandem repeats and transposable elements (TE). GMATA v2.2 and Tandem Repeats Finder (TRF v4.07b) were used for tandem repeat identification under default parameters (55, 56). For transposable element identification, de novo repeat libraries were constructed by combining results from LTR_finder, Ltr_harvest, and RepeatModeler with default parameters (57–59). The consensus sequences in *E. sinensis*-specific de novo repeat libraries and their classification information were further combined with a library from RepeatMasker and then used to run RepeatMasker on the assembly (60).

Gene annotation

Gene structure and functional annotation

Transcriptome alignment, de novo gene prediction, and sequence homology-based predictions implemented in Evidence Modeler (EVM) were used for gene prediction (61). Briefly, for transcriptome alignment, Illumina RNA-seq reads were first mapped to the assembled genome with STAR v2.7.3a under default parameters (62), and then transcripts were assembled with Stringtie v1.3.4 (63). Assembled transcripts were aligned to the genomes to obtain gene

structure annotation information using PASA v2.3.3 (64). For de novo gene prediction, GeneMark-ET and AUGUSTUS v3.3 were used to predict genes on genome sequences hard masked for TE (65, 66), and the high-quality dataset for training these ab initio gene predictors was generated by PASA v2.3.3 (64). For sequence homology-based gene prediction, protein sequences from six species (*L. vannamei*, *Daphnia pulex*, *Lepeophtheirus salmonis*, *C. elegans*, *D. melanogaster*, and *Homo sapiens*) were downloaded and mapped against the *E. sinensis* assembled genome to generate homology gene structures using the GeMoMa v1.6.1 software (67). All predicted gene structures were integrated into consensus gene models using Evidence Modeler v1.1.1 to obtain a nonredundant gene set (61).

To determine the functional annotation of the gene models, a BLASTP search with an E-value $\leq 1e^{-5}$ was performed against several protein databases, including NR [nonredundant protein sequences in National Center for Biotechnology Information (NCBI)], SwissProt, KEGG, and KOG. InterProScan 5 was used to retrieve associated GO terms describing biological processes, molecular functions, and cellular components (default parameters). The motifs and domains of each gene model were predicted by InterProScan 5 against public protein databases, including ProDom, PRINTS, Pfam, Gene3D, CCD, SMART, PANTHER, PROSITE, and SUPERFAMILY (68).

RNA-seq and assembly

To investigate the expression patterns of *Innexins* and *SMYDA* in different essential biological processes of *E. sinensis*, Illumina raw sequencing reads from different molting stages of hepatopancreas (Intermolt and Premolt), gill tissue from the aerial respiration phase (control versus 5 days out of water), and whole individuals from megalopa and larvae I stage of *E. sinensis* during the metamorphosis process, which were generated in our previous research (29, 37, 38), were downloaded from the NCBI SRA database (PRJNA271233, PRJNA480555) and National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation (GSA: CRA003690) and used for differential expression analysis.

For limb regeneration RNA-seq experiments conducted in this study, juvenile *E. sinensis* individuals were cultured in tanks, and 3 days after molting, all crab individuals were physically stimulated on the fourth limb on the right side and autotomy took place at the coxa. One day post autotomy, 13 dpa, and after the next molting (around 30 dpa), the limb bud tissues at the coxa from the wound site (1 dpa), generated limb bud (13 dpa), and the RL (30 dpa) were sampled. The bases of coxae were collected from the fourth limb on the right side at 0 dpa for control experiments. For each stage, three crab individuals were collected as biological replicates ($n = 3$). RNA-seq libraries were constructed with the Illumina True Seq Kit and sequenced on the Illumina HiSeq 4000 platform (paired-end mode, 150 bp) for each individual. Differential gene expression was measured following the Histat2-stringtie pipeline (63). First, clean reads were mapped against our assembled *E. sinensis* genome using hisat2-2.0 software. Then, the expected TPM was calculated with the Stringtie v1.3.4 software for each individual. Last, DEGs were identified by DESeq 2 with fold change > 2 and $P < 0.05$. GO enrichment analysis was conducted through clusterProfiler 3.16.1 software (69). A WGCNA coexpression network was constructed on the basis of DEGs (32). In detail, a weighted signed network was computed on the basis of a fit to scale-free topology, with a threshold soft power of 16. A topological overlap dendrogram was used to define modules with a minimum module size of 30 genes and the

deepSplit parameter set to 2. Module and trait associations were determined by calculating the correlation coefficient between module eigengene (ME) and traits (such as days after limb autotomy). Gene coexpression networks in selected modules were visualized with Cytoscape v3.8.0 software (70). We also conducted RNA-seq experiments on limb regeneration in *L. vannamei*, *M. rosenbergii*, and *Macrobrachium nipponense*. We collected the limb bud tissues of the above crustacean species at 1 dpa and coxa tissues at 0 dpa. For each group, four to six individuals were sampled as biological replicates. For *L. vannamei*, differential gene expression was measured following the Histat2-stringtie pipeline (63) with the assembled *L. vannamei* genome as reference (NCBI accession number: GCF_003789085.1). For *M. rosenbergii* and *M. nipponense*, transcriptomes were de novo assembled using Trinity v2.8.5, the expression value (TPM) was calculated using RSEM v1.3.3 for each individual, and then the DEGs were identified by DESeq 2 with fold change >2 and $P < 0.05$ (71).

Gene family identification and phylogenetic tree construction

Protein sequences of *H. vulgaris*, *C. elegans*, *Caenorhabditis briggsae*, *L. vannamei*, *D. pulex*, *D. melanogaster*, *Capitella teleta*, *Helobdella robusta*, *Crassostrea virginica*, *Octopus sinensis*, *Apostichopus japonicus*, *Asterias rubens*, *Danio rerio*, *Xenopus laevis*, *Gallus gallus*, *Anolis carolinensis*, and *H. sapiens* were downloaded from the NCBI database (table S21). The longest transcript was selected for each gene locus. Orthologous groups were identified with Orthofinder v2.3.11 under default parameters (72). The genes that could not be clustered into any gene family and that were found in only one species were considered species specific. A total of 62 one-to-one single-copy orthologous genes were extracted for phylogenetic tree construction. Briefly, multiple protein sequence alignments were obtained with MAFFT v7.271 under default parameters, and Gblocks (0.91b) was used to extract conserved well-aligned cores with default parameters (73, 74). A maximum-likelihood (ML) phylogenetic tree was constructed with RAxML v8.2.12 with the PROTGAMMAWAG substitution model and 1000 bootstrap replicates using a final dataset containing 12,981 amino acids (75). To identify specific gene families in the arthropod lineage, gene families were first clustered by Orthofinder v2.3.11 as mentioned above. We defined arthropod-specific gene families as one or several related genes present in all arthropod species but not present in other species in this study.

To fully compare the genome characters across the arthropod lineage, we selected 10 arthropod species (*Aedes aegypti*, *D. melanogaster*, *D. pulex*, *Daphnia magna*, *E. sinensis*, *Scylla paramamosain*, *H. americanus*, *L. vannamei*, *P. hawaiiensis*, and *Hyalomma azteca*; table S21) and conducted rigorous comparative genomics in the arthropod lineage. We used MCMCTree in the PAML software package to estimate divergence times in the arthropod lineage using several calibration points as constraint, including *D. melanogaster*–*C. elegans* [~623 to 877 million years (Ma) ago], *D. melanogaster*–*A. aegypti* (~217 to 301 Ma ago), and *E. sinensis*–*S. paramamosain* (~139 to 238 Ma ago), which were obtained from the TimeTree database (76). The MCMCTree was run for 1,000,000 iterations with a burn-in of 200,000 iterations. Gene family expansion and contraction analysis was further performed using the program CAFE v3.1 (77). Then, gene loss events were identified in the *E. sinensis* lineage using the following pipeline. First, we conducted pairwise reciprocal BLAST and orthology assignment with Orthofinder v2.3.11 (72) and summarized orthology groups with respect to species presence/absence. Orthologs

present in other crustaceans but not in *E. sinensis* were considered as putatively lost in *E. sinensis*. Second, all other *E. sinensis*-deficient orthologous protein sequences were BLASTP searched against the whole proteome of *E. sinensis*. If any hits matched with the proteome of *E. sinensis*, then the affected gene was removed from the list of gene loss events. Next, we mapped all other *E. sinensis*-deficient orthologous protein sequences to the whole-genome sequence of *E. sinensis* using GenomeThreader v1.7.1 software (78) and removed genes from the list of gene loss events with positive hits. Last, to further validate the gene loss events, the *E. sinensis*-deficient orthologous protein sequences were searched against the whole genome sequence of *E. sinensis* using TBLASTN with an E-value $\leq 1e-10$, and genes were considered to be potentially lost only if no evidence of matching sequence alignments was identified in the complete *E. sinensis* genome.

For *Innexin* genes and SET domain gene family classification, first, hmmsearch software was used to search against the PFAM domain, PF00876 and PF00856, respectively, with an E-value cutoff of $1e-5$ (79). Then, we manually checked structure and annotation of the respective genes. Genes defined as *Innexin* should have a clear *Innexin* domain, and SET genes should have a clear SET domain with NCBI CDD search. Moreover, all identified genes should also have a clear annotation by BLASTP. Any resulting genes with stop codons or frameshift mutations were also manually checked. Conserved motifs and transmembrane domains of *Innexin* genes were predicted by MEME v5.4.1 software and TMHMM v2.0 software, respectively (80, 81). As for the SMYD gene family, genes should have both SET and MYND domains predicted by NCBI CDD search. The ML phylogenetic trees of the *Innexin* and SMYD gene families were constructed using RAxML v8.2.12 with 1000 bootstrap.

HE, modified Masson staining and electron microscopy

For histological analysis, limb bud tissues at the coxa from autotomized crabs at 1, 2, and 13 dpa were sampled, and the RL buds were also sampled. Samples were fixed with Bouin's fixative, and then tissue slices were prepared and stained according to the HE staining method. The tissue slices were investigated with a microscope system (Leica DM500, Germany). Modified Masson staining was also conducted using a Masson trichromatic staining kit (Solarbio, Beijing, China) following the supplier's instructions. We also conducted electron microscopy on tissue at the coxa after autotomy using transmission electron microscopy. The collected tissue was cut into 1-mm³ pieces, which were fixed in 2.5% glutaraldehyde. After being dehydrated, embedded, polymerized, cut, and stained, the samples were examined with a Hitachi transmission electron microscope (Hitachi HT7700, Japan).

RNA FISH

For RNA in situ hybridization, autotomized limb coxae at 0 and 1 dpa were fixed overnight at 4°C in 4% paraformaldehyde. Tissue sections (6 μ m) were obtained following the paraffin section method. The fluorescent probe was designed against the target gene sequences and labeled with FAM/Cy3 fluorochrome (table S22). FISH was performed according to instructions for a FISH kit purchased from Shanghai Gefan Biotech Co. Ltd. (Shanghai, China). The tissue slides were hybridized with fluorescent probes for 48 hours at 65°C and then washed with SSC and phosphate-buffered saline buffer. Images for fluorescence signals were observed using a fluorescence microscope (Leica DMi8, Germany).

Quantitative reverse transcriptase PCR

qRT-PCR was conducted on selected genes during the limb regeneration process with the ribosomal S27 fusion protein (S27) used as the reference gene for normalization (table S22). We conducted qRT-PCR using SYBR Premix Ex Taq (Takara) on a Rotor-Gene Q 2plex Platform (Qiagen, Germany). For each selected gene, four to six biological replicates were performed. Gene expression levels were measured using the $2^{-\Delta\Delta C_t}$ method.

Rapamycin injection experiment

A total of 18 crab individuals were randomly divided into two groups. The crabs in the experimental group ($n = 9$) were injected with 10 mM rapamycin (0.3 μ l/g; solubilized in DMSO) (Solarbio, Beijing, China), and those of the control group ($n = 9$) with DMSO shortly after autotomy. Three days later, rapamycin and DMSO were injected again for the experimental and control groups. The regeneration process of the limb bud was observed every day, and the length of the regenerated papillae was measured until 9 dpa.

RNAi experiment

Double-stranded RNA (dsRNA) was generated for RNAi experiments for target genes using PCR amplification of target genes (e.g., *Esin.LG37.0025-Inx2*) with a T7 promoter sequence attached to the primers (table S22). After PCR amplification, in vitro transcription was conducted using the RiboMax Promega Large Scale RNA Production Systems-T7 Kit (Promega Kit). A fragment of the green fluorescence protein (GFP) sequence from the pEGFP-N1 plasmid was used as control dsRNA. dsGFP was generated using the same method as for target genes. Crab individuals in the same molting stage were collected for the RNAi experiments, and dsRNA of *Inx2* was injected through the fourth limb on the left side at the coxa in the experimental group (3 μ g/g) after autotomy. Control dsRNA of GFP was injected in the same way (3 μ g/g). Three days later, both experimental and control groups were injected again. The regeneration situation was observed every day; the characteristics of regenerated papilla formation were recorded at 5 dpa.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <https://science.org/doi/10.1126/sciadv.abl4642>

[View/request a protocol for this paper from Bio-protocol.](#)

REFERENCES AND NOTES

- R. J. Davenport, What controls organ regeneration? *Science* **309**, 84 (2005).
- C.-H. Chen, K. D. Poss, Regeneration genetics. *Annu. Rev. Genet.* **51**, 63–82 (2017).
- J. M. W. Slack, Animal regeneration: Ancestral character or evolutionary novelty? *EMBO Rep.* **18**, 1497–1508 (2017).
- J. A. Goldman, K. D. Poss, Gene regulatory programmes of tissue regeneration. *Nat. Rev. Genet.* **21**, 511–525 (2020).
- J. A. Chapman, E. F. Kirkness, O. Simakov, S. E. Hampson, T. Mitros, T. Weinmaier, T. Rattei, P. G. Balasubramanian, J. Borman, D. Busam, K. Disbennett, C. Pfannkoch, M. Sumin, G. G. Sutton, L. D. Viswanathan, B. Walenz, D. M. Goodstein, U. Hellsten, T. Kawashima, S. E. Prochnik, N. H. Putnam, S. Shu, B. Blumberg, C. E. Dana, L. Gee, D. F. Kibler, L. Law, D. Lindgens, D. E. Martinez, J. Peng, P. A. Wigge, B. Bertulat, C. Guder, Y. Nakamura, S. Ozbek, H. Watanabe, K. Khalturin, G. Hemmrich, A. Franke, R. Augustin, S. Fraune, E. Hayakawa, S. Hayakawa, M. Hirose, J. S. Hwang, K. Ikey, C. Nishimiya-Fujisawa, A. Ogura, T. Takahashi, P. R. H. Steinmetz, X. Zhang, R. Aufschnaiter, M.-K. Eder, A.-K. Gorny, W. Salvenmoser, A. M. Heimberg, B. M. Wheeler, K. J. Peterson, A. Böttger, P. Tischler, A. Wolf, T. Gojobori, K. A. Remington, R. L. Strausberg, J. C. Venter, U. Technau, B. Hobmayer, T. C. G. Bosch, T. W. Holstein, T. Fujisawa, H. R. Bode, C. N. David, D. S. Rokhsar, R. E. Steele, The dynamic genome of *Hydra*. *Nature* **464**, 592–596 (2010).
- P. W. Reddien, The cellular and molecular basis for planarian regeneration. *Cell* **175**, 327–345 (2018).
- N. Konstantinides, M. Averof, A common cellular basis for muscle regeneration in arthropods and vertebrates. *Science* **343**, 788–791 (2014).
- K. Wasik, J. Gurtowski, X. Zhou, O. M. Ramos, M. J. Delás, G. Battistoni, O. El Demerdash, I. Falcatori, D. B. Vizoso, A. D. Smith, P. Ladurner, L. Schärer, W. R. McCombie, G. J. Hannon, M. Schatz, Genome and transcriptome of the regeneration-competent flatworm, *Macrostomum lignano*. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12462–12467 (2015).
- M. A. Grohme, S. Schloissnig, A. Rozanski, M. Pippel, G. R. Young, S. Winkler, H. Brandl, I. Henry, A. Dahl, S. Powell, M. Hiller, E. Myers, J. C. Rink, The genome of *Schmidtea mediterranea* and the evolution of core cellular mechanisms. *Nature* **554**, 56–61 (2018).
- P. M. Hopkins, Limb regeneration in the fiddler crab, *Uca pugnator*: Hormonal and growth factor control. *Am. Zool.* **41**, 389–398 (2001).
- D. L. Mykles, Interactions between limb regeneration and molting in decapod crustaceans. *Am. Zool.* **41**, 399–406 (2001).
- S. J. Khan, K. J. Schuster, R. K. Smith-Bolton, Regeneration in crustaceans and insects, in *eLS* (Wiley, 2016), pp.1–14.
- F. Alves, C. Enjolras, M. Averof, Live imaging reveals the progenitors and cell dynamics of limb regeneration. *eLife* **5**, e19766 (2016).
- E. S. Chang, M. Thiel, Regeneration in crustaceans, in *Physiology* (Oxford Univ. Press, 2015), vol. 4, pp. 545.
- J. Wang, P. Xu, G. Zhou, X. Li, Q. Lu, X. Liu, J. Zhou, C. Wang, Genetic Improvement and Breeding Practices for Chinese Mitten Crab, *Eriocheir sinensis*. *J. World Aquacult. Soc.* **49**, 292–301 (2018).
- C. Wang, S. Li, C. Fu, X. Gong, L. Huang, X. Song, Y. Zhao, Molecular genetic structure and evolution in native and colonized populations of the Chinese mitten crab, *Eriocheir sinensis*. *Biol. Invas.* **11**, 389–399 (2009).
- Z. Cui, Y. Liu, J. Yuan, X. Zhang, T. Ventura, K. Y. Ma, S. Sun, C. Song, D. Zhan, Y. Yang, H. Liu, G. Fan, Q. Cai, J. Du, J. Qin, C. Shi, S. Hao, Q. P. Fitzgibbon, G. G. Smith, J. Xiang, T.-Y. Chan, M. Hui, C. Bao, F. Li, K. H. Chu, The Chinese mitten crab genome provides insights into adaptive plasticity and developmental regulation. *Nat. Commun.* **12**, 2395 (2021).
- L. Song, C. Bian, Y. Luo, L. Wang, X. You, J. Li, Y. Qiu, X. Ma, Z. Zhu, L. Ma, Z. Wang, Y. Lei, J. Qiang, H. Li, J. Yu, A. Wong, J. Xu, Q. Shi, P. Xu, Draft genome of the Chinese mitten crab, *Eriocheir sinensis*. *Gigascience* **5**, 5 (2016).
- B. Tang, Z. Wang, Q. Liu, H. Zhang, S. Jiang, X. Li, Z. Wang, Y. Sun, Z. Sha, H. Jiang, X. Wu, Y. Ren, H. Li, F. Xuan, B. Ge, W. Jiang, S. She, H. Sun, Q. Qiu, W. Wang, Q. Wang, G. Qiu, D. Zhang, Y. Li, High-quality genome assembly of *Eriocheir japonica sinensis* reveals its unique genome evolution. *Front. Genet.* **10**, 1340 (2020).
- K. Pal, M. Forcato, F. Ferrari, Hi-C analysis: From data generation to integration. *Biophys. Rev.* **11**, 67–78 (2019).
- Y. Yuan, C. Y.-L. Chung, T.-F. Chan, Advances in optical mapping for genomic research. *Comput. Struct. Biotechnol. J.* **18**, 2051–2062 (2020).
- F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
- D. Kao, A. G. Lai, E. Stamatakis, S. Rosic, N. Konstantinides, E. Jarvis, A. Di Donfrancesco, N. Pouchkina-Stantcheva, M. Semon, M. Grillo, H. Bruce, S. Kumar, I. Siwanowicz, A. Le, A. Lemire, C. Extavour, W. Browne, C. Wolff, M. Averof, N. H. Patel, P. Sarkies, A. Pavlopoulos, A. Aboobaker, The genome of the crustacean *Parhyale hawaiiensis*: A model for animal development, regeneration, immunity and lignocellulose digestion. *eLife* **5**, e20062 (2016).
- J. Gutekunst, R. Andriantsoa, C. Falckenhayn, K. Hanna, W. Stein, J. Rasamy, F. Lyko, Clonal genome evolution and rapid invasive spread of the marbled crayfish. *Nat. Ecol. Evol.* **2**, 567–573 (2018).
- X. Zhang, J. Yuan, Y. Sun, S. Li, Y. Gao, Y. Yu, C. Liu, Q. Wang, X. Lv, X. Zhang, K. Y. Ma, X. Wang, W. Lin, L. Wang, X. Zhu, C. Zhang, J. Zhang, S. Jin, K. Yu, J. Kong, P. Xu, J. Chen, H. Zhang, P. Sorgeloos, A. Sagi, A. Alcivar-Warren, Z. Liu, L. Wang, J. Ruan, K. H. Chu, B. Liu, F. Li, J. Xiang, Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nat. Commun.* **10**, 356 (2019).
- J. M. Polinski, A. V. Zimin, K. F. Clark, A. B. Kohn, N. Sadowski, W. Timp, A. Ptitsyn, P. Khanna, D. Y. Romanova, P. Williams, S. J. Greenwood, L. L. Moroz, D. R. Walt, A. G. Bodnar, The American lobster genome reveals insights on longevity, neural, and immune adaptations. *Sci. Adv.* **7**, eabe8290 (2021).
- X.-S. Han, C. Wang, F.-h. Guo, S. Huang, Y.-W. Qin, X.-X. Zhao, Q. Jing, Neoblast-enriched zinc finger protein FIR1 triggers local proliferation during planarian regeneration. *Protein Cell* **10**, 43–59 (2019).
- A. V. Kuballa, D. J. Merritt, A. Elizur, Gene expression profiling of cuticular proteins across the moult cycle of the crab *Portunus pelagicus*. *BMC Biol.* **5**, 45 (2007).
- J. Wang, X. Chen, F. He, X. Song, S. Huang, W. Yue, Y. Chen, Z. Su, C. Wang, Global analysis of gene expression profiles provides novel insights into the development and evolution of the large crustacean *Eriocheir sinensis*. *Genomics Proteomics Bioinformatics* **18**, 443–454 (2020).

30. P. M. Hopkins, Regeneration of walking legs in the fiddler crab *Uca pugilator*. *Am. Zool.* **33**, 348–356 (1993).
31. P. M. Hopkins, D. S. Durica, Effects of all-trans retinoic acid on regenerating limbs of the fiddler crab, *Uca pugilator*. *J. Exp. Zool.* **272**, 455–463 (1995).
32. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
33. Q. Wang, M. Yin, C. Yuan, X. Liu, Z. Hu, Z. Zou, M. Wang, Identification of a conserved phenoloxidase activation pathway in cotton bollworm *Helicoverpa armigera*. *Front. Immunol.* **11**, 785 (2020).
34. A. Oshima, K. Tani, Y. Fujiyoshi, Atomic structure of the innexin-6 gap junction channel determined by cryo-EM. *Nat. Commun.* **7**, 13681 (2016).
35. G. Dahl, K. J. Muller, Innexin and pannexin channels and their signaling. *FEBS Lett.* **588**, 1396–1402 (2014).
36. V. B. Hua, A. B. Chang, J. H. Tchieu, N. M. Kumar, P. A. Nielsen, M. H. Saier Jr., Sequence and phylogenetic analyses of 4 TMS junctional proteins of animals: Connexins, innexins, claudins and occludins. *J. Membr. Biol.* **194**, 59–76 (2003).
37. S. Huang, J. Wang, W. Yue, J. Chen, S. Gaughan, W. Lu, G. Lu, C. Wang, Transcriptomic variation of hepatopancreas reveals the energy metabolism and biological processes associated with molting in Chinese mitten crab, *Eriocheir sinensis*. *Sci. Rep.* **5**, 14015 (2015).
38. X. Chen, J. Wang, X. Hou, W. Yue, Z. Li, C. Wang, Gene expression profiles of gill provide insights into the aerial respiration capacity of the Chinese mitten crab, *Eriocheir sinensis*. *Aquaculture* **506**, 148–153 (2019).
39. K. Takayama, A. Muto, Y. Kikuchi, Leucine/glutamine and v-ATPase/lysosomal acidification via mTORC1 activation are required for position-dependent regeneration. *Sci. Rep.* **8**, 8278 (2018).
40. W. Gan, X. Dai, X. Dai, J. Xie, S. Yin, J. Zhu, C. Wang, Y. Liu, J. Guo, M. Wang, J. Liu, J. Hu, R. L. Quinton, N. J. Ganem, P. Liu, J. M. Asara, P. P. Pandolfi, Y. Yang, Z. He, G. Gao, W. Wei, LATS suppresses mTORC1 activity to directly coordinate Hippo and mTORC1 pathways in growth control. *Nat. Cell Biol.* **22**, 246–256 (2020).
41. E. M. Tanaka, The molecular and cellular choreography of appendage regeneration. *Cell* **165**, 1598–1608 (2016).
42. P. Amparyup, W. Charoensapsri, A. Tassanakajon, Prophenoloxidase system and its role in shrimp immune responses against major pathogens. *Fish Shellfish Immunol.* **34**, 990–1001 (2013).
43. F. Yue, L. Pan, J. Miao, L. Zhang, J. Li, Molecular cloning, characterization and mRNA expression of two antibacterial peptides: Crustin and anti-lipopolysaccharide factor in swimming crab *Portunus trituberculatus*. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **156**, 77–85 (2010).
44. C. Sun, W.-T. Xu, H.-W. Zhang, L.-P. Dong, T. Zhang, X.-F. Zhao, J.-X. Wang, An anti-lipopolysaccharide factor from red swamp crayfish, *Procambarus clarkii*, exhibited antimicrobial activities in vitro and in vivo. *Fish Shellfish Immunol.* **30**, 295–303 (2011).
45. E. Calpena, F. Palau, C. Espinós, M. I. Galindo, Evolutionary history of the *Smyd* gene family in metazoans: A framework to identify the orthologs of human *Smyd* genes in *Drosophila* and other animal species. *PLOS ONE* **10**, e0134106 (2015).
46. F. Jiang, Q. Liu, Y. Wang, J. Zhang, H. Wang, T. Song, M. Yang, X. Wang, L. Kang, Comparative genomic analysis of SET domain family reveals the origin, expansion, and putative function of the arthropod-specific *SmydA* genes as histone modifiers in insects. *Gigascience* **6**, 1–16 (2017).
47. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
48. G. Marçais, C. Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
49. J. Hu, J. Fan, Z. Sun, S. Liu, NextPolish: A fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**, 2253–2255 (2020).
50. J. N. Burton, A. Adey, R. P. Patwardhan, R. Qiu, J. O. Kitman, J. Shendure, Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).
51. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
52. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
53. P. Danecek, S. A. McCarthy, BCFtools/csq: Haplotype-aware variant consequences. *Bioinformatics* **33**, 2037–2039 (2017).
54. H. Li, Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**, 2103–2110 (2016).
55. X. Wang, L. Wang, GMATA: An integrated software package for genome-scale SSR mining, marker development and viewing. *Front. Plant Sci.* **7**, 1350 (2016).
56. G. Benson, Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
57. Z. Xu, H. Wang, LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
58. D. Ellinghaus, S. Kurtz, U. Willhoeft, LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008).
59. J. A. Bedell, I. Korf, W. Gish, MaskerAid: A performance enhancement to RepeatMasker. *Bioinformatics* **16**, 1040–1041 (2000).
60. J. Jurka, V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, J. Walichiewicz, Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
61. B. J. Haas, S. L. Salzberg, W. Zhu, M. Pertea, J. E. Allen, J. Orvis, O. White, C. R. Buell, J. R. Wortman, Automated eukaryotic gene structure annotation using EVIDENCEModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).
62. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
63. M. Pertea, D. Kim, G. M. Pertea, J. T. Leek, S. L. Salzberg, Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).
64. B. J. Haas, A. L. Delcher, S. M. Mount, J. R. Wortman, R. K. Smith Jr., L. I. Hannick, R. Maiti, C. M. Ronning, D. B. Rusch, C. D. Town, S. L. Salzberg, O. White, Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
65. A. Lomsadze, P. D. Burns, M. Borodovsky, Integration of mapped RNA-seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* **42**, e119 (2014).
66. M. Stanke, O. Keller, I. Gunduz, A. Hayes, S. Waack, B. Morgenstern, AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
67. J. Keilwagen, F. Hartung, M. Paulini, S. O. Twardziok, J. Grau, Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics* **19**, 189 (2018).
68. E. Quevillon, V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, R. Lopez, InterProScan: Protein domains identifier. *Nucleic Acids Res.* **33**, W116–W120 (2005).
69. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
70. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
71. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
72. D. M. Emms, S. Kelly, OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
73. K. Katoh, K. Misawa, K.-i. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
74. G. Talavera, J. Castresana, Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
75. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
76. S. Kumar, G. Stecher, M. Suleski, S. Blair Hedges, TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
77. M. V. Han, G. W. C. Thomas, J. Lugo-Martinez, M. W. Hahn, Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* **30**, 1987–1997 (2013).
78. G. Gremme, V. Brendel, M. E. Sparks, S. Kurtz, Engineering a software tool for gene structure prediction in higher organisms. *Inf. Softw. Technol.* **47**, 965–978 (2005).
79. R. D. Finn, J. Clements, S. R. Eddy, HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
80. T. L. Bailey, J. Johnson, C. E. Grant, W. S. Noble, The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
81. A. Krogh, B. Larsson, G. von Heijne, E. L. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).

Acknowledgments

Funding: This work was supported by the Shanghai Agriculture Applied Technology Development Program, China (grant no. G2017-02-08-00-10-F00076) and the Agriculture Research System of Shanghai, China (grant no. 2020044). **Author contributions:** Conceptualization: Jun Wang, C.L., and C.W. Methodology: C.W., Jun Wang, C.L., J.Y., and

G.L. Investigation: C.W., J.W., X.C., X.H., W.Y., Jingan Wang, S.H., and G.X. Visualization: Jun Wang and X.C. Supervision: C.W. Funding acquisition: C.W. and Jun Wang. Writing—original draft: Jun Wang, X.C., J.Y., C.L., M.H., and C.W. Writing—review and editing: Jun Wang, J.Y., G.L., C.L., C.W., and M.H. All authors contributed to the final version of the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The *E. sinensis* genome data have been deposited at NCBI with the accession number JAHNEI000000000 under PRJNA737102 project. RNA-seq data generated for limb regeneration process (PRJNA733310), molting process (PRJNA271233), and aerial respiration process (PRJNA480555) of *E. sinensis* was deposited in NCBI's SRA database. RNA-seq data for different developmental stages of *E. sinensis* was deposited at the National Genomics Data

Center, Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation (GSA: CRA003690). RNA-seq data for limb regeneration process of *L. vannamei*, *M. rosenbergii*, and *M. nipponensis* was deposited in NCBI's SRA database (PRJNA818363, PRJNA822882, and PRJNA822548). All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 15 July 2021

Accepted 1 August 2022

Published 16 September 2022

10.1126/sciadv.abl4642

“Omics” data unveil early molecular response underlying limb regeneration in the Chinese mitten crab, *Eriocheir sinensis*

Jun WangXiaowen ChenXin HouJingan WangWucheng YueShu HuangGangchun XuJizhou YanGuoqing LuMichael HofreiterChenhong LiChenghui Wang

Sci. Adv., 8 (37), eabl4642. • DOI: 10.1126/sciadv.abl4642

View the article online

<https://www.science.org/doi/10.1126/sciadv.abl4642>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

Science Advances (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.
Copyright © 2022 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).