BMC Genomics

RESEARCH

Open Access

Transcriptome analysis of interna rootlets of the rhizocephalan *Parasacculina sinensis* reveals potential mechanisms of parasite host control

Wenjie Li[1,2], Yiheng Feng[2], Haocheng Liu[2], Niklas Dreyer[3,4,5] and Yue Him Wong[2,6*]

Abstract

Background The endoparasitic rhizocephalan *Parasacculina sinensis* has a radically simplified morphology and primarily infests decapods crustaceans. Rhizocephalan barnacles usually absorb nutrients from the host through a complex rootlet system (the interna), and also change the morphology, physiology and behavior of their hosts. However, little is known about the transcriptomic landscape, ultrastructural details and gene expression of the interna rootlets in the group. In this study, we investigated the structural and molecular signatures of the interna of *P. sinensis* by using detailed histological staining and transcriptomic analyses.

Results The interconnected F-actin nodal network, lipid droplets, and nucleus of interna rootlets were visualized using fluorescence straining. We successfully obtained a clean transcriptome of *P. sinensis* and conducted functional analyses of interna embedded within host hepatopancreas, claw muscle, and eyestalk. The gene ontology (GO) terms related to translation, metabolic process, biosynthetic process, cellular process were highly expressed in the top 10% transcripts from the interna. The GO category of shared differential expression of genes (DEGs) among internae was related to embryonic development. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway neuroactive ligand-receptor interaction and the GO term neurotransmitter transporter activity were identified in the shared DEGs among internae. The interna entangled within host hepatopancreas, eyestalk and claw muscle fibers had similarities and differences in the functional biology. Additionally, the interna specific candidate genes probably involved in host immune, lipid metabolism, molting and growth were identified.

Conclusions Our study demonstrates an in-depth function of interna rootlets of *P. sinensis* and reveals potential mechanisms of parasite host control. This study provides novel information to further investigate the evolutionary drivers of parasitism in barnacles.

Keywords Parasitic barnacles, *Parasacculina sinensis*, Transcriptomics, Function of interna, Parasite host control

*Correspondence:
Yue Him Wong
timwong@szu.edu.cn
Full list of author information is available at the end of the article

Background

Parasitism has evolved independently multiple times across the Tree of Life, playing a critical role in the diversification and macro-evolutionary success of diverse lineages [1]. The evolutionary transition from a free-living to a parasitic lifestyle comes with some challenges, such as developing an effective infection mechanism and evading host defense systems [2]. The barnacles (Cirripedia) mostly sport a free-living lifestyle, although a number of its members are entirely parasitic. One such member, Rhizocephala, exhibits the hallmark of a parasitic taxon. Primarily infesting decapods crustaceans, these parasites may exert morphological, physiological, and behavioral control over their hosts [3, 4]. Given their potential impacts on coastal aquaculture and biodiversity [5], it is therefore surprising that little is known about the genetic mechanisms underlying the adaptations to this fascinating extreme parasitic lifestyle.

Unlike other crustaceans, rhizocephalans exhibit an extremely reduced morphology and lack even "classical" arthropodal features, such as segmented limbs, in their adult stages [3, 4]. After infection, the female parasite develops a rootlet system called the interna, which spreads throughout the haemolymph of the host to absorb nutrients for the externa [4, 6]. The externa is an external reproductive structure and often visible under the abdomen of the host [4, 7]. Notwithstanding extreme morphological changes in their hosts, such as the "feminisation" of male hosts whereby the latter exhibit wider abdomens and shortened gonopods post infection [8, 9], rhizocephalans have evolved the ability to manipulate host physiology. For instance, gonadal atrophy occurs in both male and female hosts, with reproductive function inhibited following infestation [10, 11]. Infestation by rhizocephalans may also increase the concentrations of certain proteins, haemocyanin and glucose in the host haemolymph [12], potentially indicating increased energy expenditures and metabolism of infected hosts. Furthermore, rhizocephalans induce host sterilization, inhibit molting (anecydsis), and reduce growth, thereby potentially impacting coastal crustacean aquaculture [5].

Few studies have mechanistically explored the functional biology of the interna and its role in inducing physiological changes in infested hosts (e.g., [13, 14]). The rootlets simply consist of a cuticle, an epidermis and a subjacent layer of axial cells [15]. Interestingly, the interna rootlets of *Sacculina carcini* sampled around host midgut and thoracic ganglia shows differential patterns of gene expression [13]. However, one knowledge gap concerns whether functional differences exist in the penetration of the interna into other host organs, such as the hepatopancreas and claw muscle fibers. Additionally, Martin et al. (2022) identified candidate genes from the interna tissues of *S. carcini*, including immunity-related genes, juvenile hormone binding protein and two crustacean neurohormones (known to suppress ecdysis and gonad development in other species) [13]. Zatylny-Gaudin et al. [14] conducted subsequent haemolymph proteome sequencing of healthy and infected green crabs (*Carcinus maenas*) parasitized by *S. carcini* in, it is plausible that these parasites may inhibit host melanization for self-protection, while promoting the presence of immune factors to cope with possible bacterial super-infections [14]. In crustaceans, the hepatopancreas and X-organ-sinus gland complex (XO-SG) of the eyestalk ganglion are two major endocrine system centers. The hepatopancreas plays a critical role in the storage of nutrients and energy, regulation of nutritional status, physiological metabolism, molting, and immunity [16, 17]. The endocrine regulation in molting, metabolism, and osmotic balance is mainly controlled by an XO-SG of the eyestalk ganglion, which is a crucial neuroendocrine regulatory system [18]. Despite their importance, the molecular information of the interna penetration in hepatopancreas and eyestalk remains largely unknown. It is unclear whether the interna rootlets are present within the host crab eyestalk. Thus, whether functional differences exist in the penetration of interna into the hepatopancreas, eyestalk, and other host organs requires further investigation. In addition, how parasitic barnacle infestation affects key physiological processes in the hosts, such as molting inhibition, metabolic change, and immune response requires more insights.

Here, we examine the transcriptomic signature of the internae of *Parasacculina sinensis* (Boschma, 1933), a parasitic rhizocephalan barnacle that infects the intertidal crab, *Leptodius exaratus* (H. Milne Edwards, 1834). While previous studies have focused on the morphology of the externa in *P. sinensis*, no research to date has addressed the histology or gene expression of the interna in this species. In this study, we first removed host transcripts in the transcriptome generated from the host-rhizocephalan mixture sample using an *in-silico* subtraction method. To test the hypothesis that the functional biology of interna penetration differs across host organs, we performed fine-grained transcriptomic profiling of interna entangled within the host hepatopancreas, eyestalk and claw muscle fibers. To determine whether the parasites inhibit or promote the host immunity, we explored immune factors within internae. By bioinformatically isolating interna specific secretory protein transcripts, we identified candidate genes or proteins potentially involved in host metabolism and molt cycle control. The present data provides novel insights into the evolutionary drivers of parasitism in barnacles.

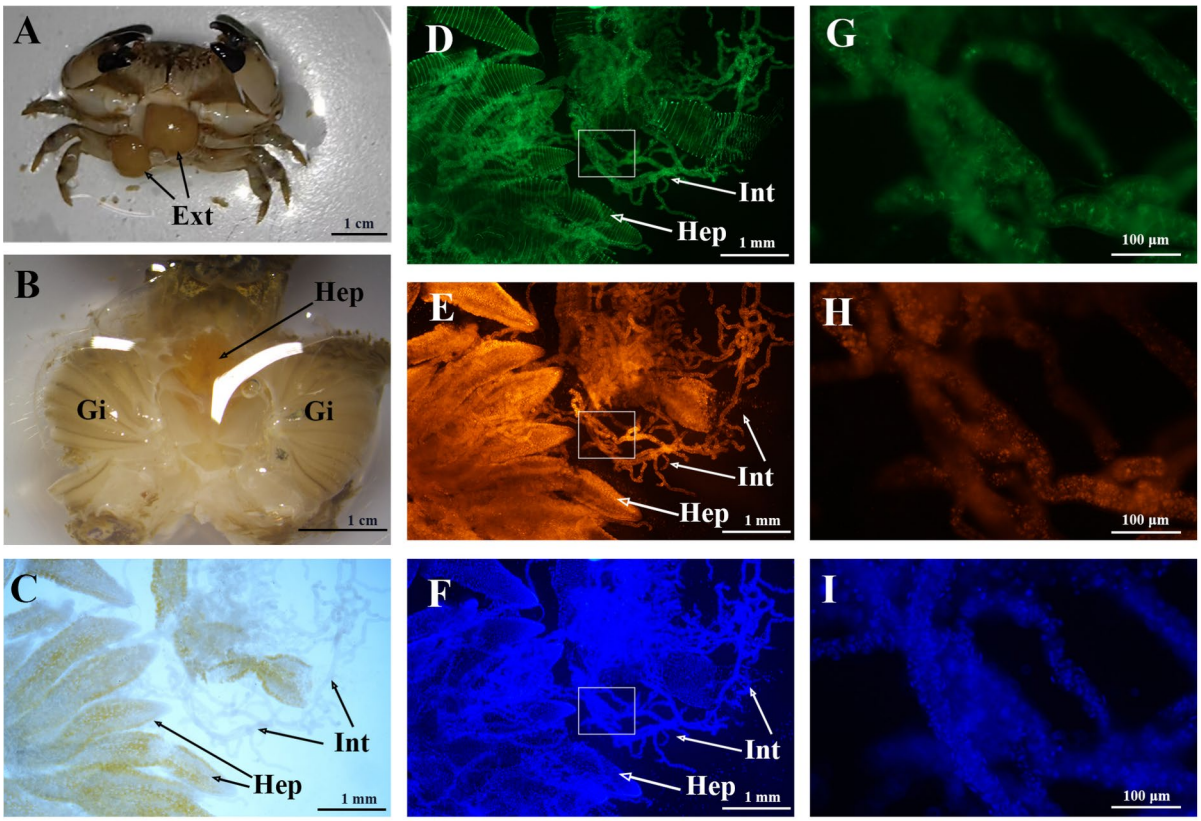

Fig. 1 *Parasacculina sinensis* and the host crab. **A** *L. exaratus* specimen parasitized by *P. sinensis*, as characterized by the emergence of two externae from the host abdomen plate. Ext: externa. **B** The interna body cavity of the infected host. Gi: gill. Hep: hepatopancreas. **C** Dissection of hepatopancreas, with interna rootlets entangling the host organ. Int: interna rootlets. **D**, **E**, **F** Phalloidin, Nile Red and DAPI staining of the hepatopancreas, in 10 × objective. **G**,**H**, **I** Magnified view of the white box highlighted area in D-F, in 40 × objective

Results

Structural overview of the interna

Based on the sequence data of *18S* rRNA and *COXI* extracted from the transcriptome assembly (details provided below), the parasitic barnacle was identified as *P. sinensis*. This species comprises two external reproductive organs (externa) and a nutrient-absorbing rootlet system that infiltrates the host (interna) (Fig. 1). The interna of *P. sinensis* is intricately associated with the host hepatopancreas (Fig. 1B, C). The structure of the parasite interna was further investigated using fluorescence staining of the hepatopancreas in infected crabs (Fig. 1D-I). At the time of sampling, the interna had developed a ramified root system infiltrating the host crab (Fig. 1D-F) and served as a nutrient-uptake device [15]. The F-actin, lipid droplets, and nuclei of the interna root system were also visualized by staining for fluorescence (Figs. 1G-I and 2). Phalloidin staining revealed that F-actin forms an interconnected nodal network throughout the interna, with individual nodes visible in Figs. 1G and 2. The Nile Red staining revealed that lipid droplets were a major cellular component within the interna (Figs. 1H and 2). DAPI staining showed that nuclei were surrounded by the interconnected F-actin nodal network (Figs. 1I and 2).

In-silico subtraction using host transcriptome reads to generate a clean transcriptome of parasitic barnacle

Using the RNA-seq read data generated from 35 *P. sinensis* samples, a transcriptome comprising 221,310 transcript sequences was assembled. The *18S* rRNA and *COXI* sequences from the host crab *L. exaratus* and *P. sinensis* were all present in the transcriptome assembly, indicating that the interna samples comprised a host-interna mixture, as predicted. To ensure the integrity of the *P. sinensis* transcriptome, an *in-silico* subtraction approach was adopted, in which the transcriptome read data generated from non-infected crab samples was used to map against the initial transcriptome assembly. Transcripts with one or more mapped reads from the clean host RNA-seq read data were identified as host contaminating transcripts and removed from the original transcriptome assembly. The *in-silico* subtraction method removed 78,334 host transcripts, retaining 142,976 transcripts as the clean *P. sinensis* transcriptome. All raw read data, including both

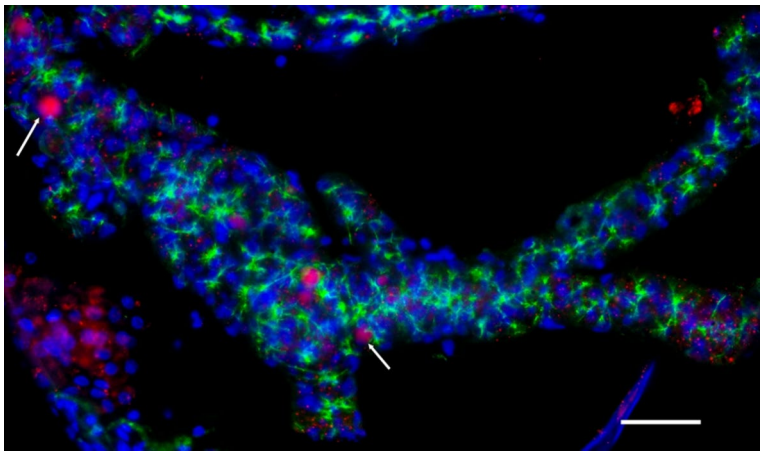

Fig. 2 *Parasacculina sinensis* interna sampled from the host crab hepatopancreas. The interna-hepatopancreas sample was co-stained with DAPI (blue), Phalloidin (green) and Nile Red (red). The images of the three fluorescent staining were captured using a 40× objective lens and merged using ImageJ. The white arrows indicate two Nile Red stained lipid droplets within the interna rootlet. Scale bar: 50 μm

35 *P. sinensis* and 10 uninfected *L. exaratus* samples, have been deposited into the NCBI Sequence Read Archive (SRA) website under accession number PRJNA1185833. The annotated genes showed a completeness of 91.1%, according to assessment by BUSCO (Additional file 1: Table S1). The annotation of *P. sinensis* transcripts were used for the subsequent functional analyses. A total of 131,864 transcripts in externa (ext), 125,627 transcripts in interna-hepatopancreas (int-HP), 42,097 transcripts in interna-claw (int-CL), and 34,497 transcripts in interna-eyestalk (int-ES) were mapped to the assembled clean *P. sinensis* transcriptome. Additionally, the BUSCO completeness scores for the transcriptome were 76.5% for the ext, 74.0% for the int-HP, 83.4% for the int-CL and 72.4% for the int-ES. Using a transcript database based on gene predictions from the *Drosophila melanogaster* genome as a reference, 12,167 transcripts in the filtered *P. sinensis* transcriptome were successfully annotated.

PCA analysis of the transcriptome of *P. sinensis* samples

The gene expression profiles of 9 externae, 14 int-HP, 7 int-CL, 5 int-ES samples were computed by mapping the clean reads from these 35 samples to the clean *P. sinensis* transcriptome. To assess the similarity and dissimilarity of the transcriptomes across samples, principal component analysis (PCA) was conducted to visualize gene expression patterns across the entire dataset (Fig. 3). The first principal component accounted for 16.4% of the variance, while the second component accounted for 6.79% of the variance. Together, these components successfully distinguished the int-HP and externa samples.

The externa and the majority of int-HP samples formed two distinct clusters in the PCA analysis. While the int-CL and int-ES samples were distinct from both ext and int-HP samples, they were not distinguishable from each other. Notably, host sex did not seem to have a major impact on the transcriptome profiles of *P. sinensis* interna and externa samples.

Functional categories of interna top 10% transcripts

The gene expression profiles of 14 int-HP, 7 int-CL, and 5 int-ES samples were examined. Transcripts expressed in at least one sample were selected for functional analyses. To gain insights into the biological functions of the interna, the top 10% transcripts (based on expression levels in each interna sample) were selected for functional enrichment analyses. A total of 12,563 transcripts, 4,210 transcripts and 3,450 transcripts were abundantly expressed in the int-HP, int-CL and int-ES, respectively. Among these top 10% transcripts, 5,275 from int-HP, 841 from int-CL, and 592 from int-ES were successfully annotated. Based on the *D. melanogaster* annotation, these transcripts were firstly mapped to the KEGG "Metabolic pathway". The mapped KEGG "Metabolic pathway" for int-HP, int-CL and int-ES samples were highly overlapping and included pathways involved in carbohydrate metabolism, energy metabolism, lipid metabolism, nucleotide metabolism, amino acid metabolism, glycan metabolism, metabolism of cofactors and vitamins, biosynthesis of terpenoids and polyketides (Fig. 4).

To further understand the biological functions of interna, we conducted gene ontology (GO) term enrichment analysis ($p < 0.05$). For all interna samples, the enriched GOBP terms were largely associated with translation, metabolic process, biosynthetic process, and certain cellular processes (Fig. 5). The specific enriched GOBP terms were protein transport (int-HP), response to stress (int-CL, int-ES) and reproduction

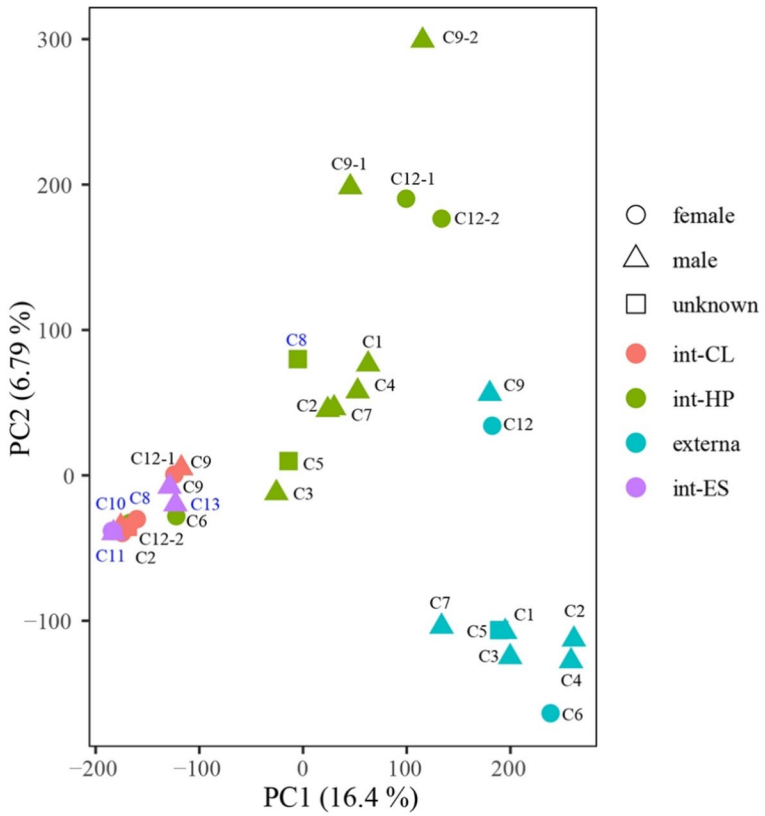

Fig. 3 The principal component analysis (PCA) of the transcriptome data. The different colors represent the int-CL, int-HP, int-ES and externa samples. The different shapes represented the female *L. exaratus*, male *L. exaratus* and the samples with unknown gender (no document). Hosts with externa parasites were indicated in black, and hosts with no bearing externa parasites were indicated in purple

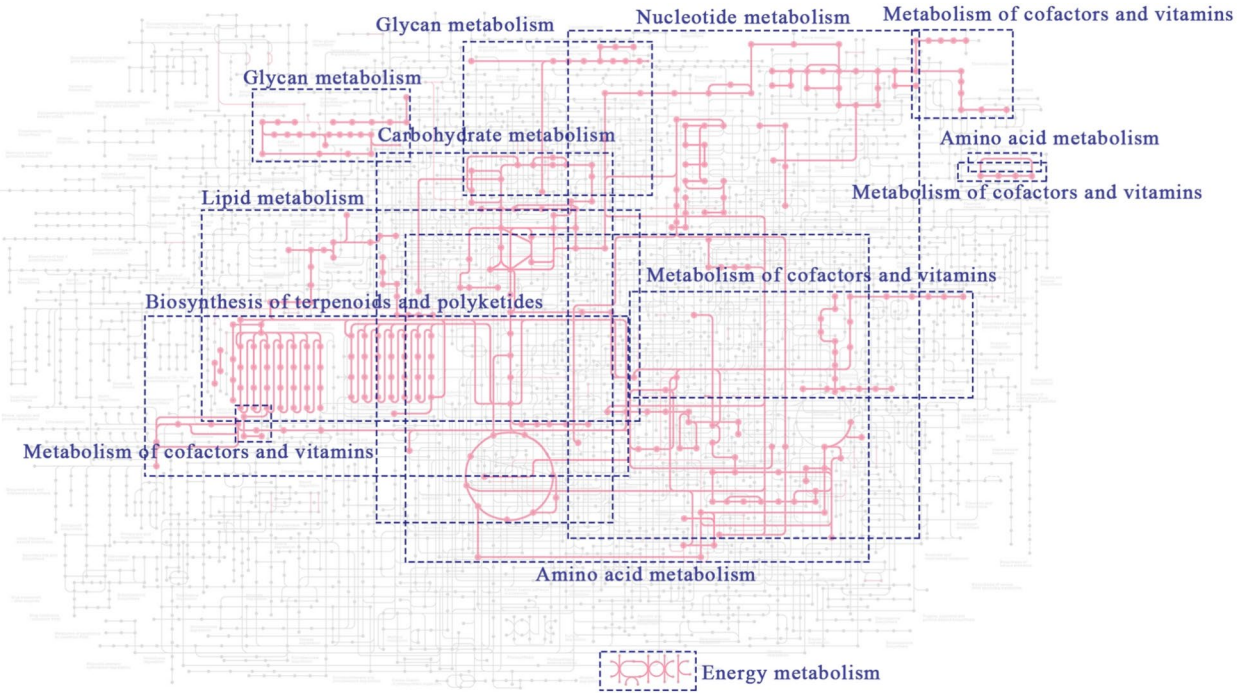

Fig. 4 KEGG metabolic pathways of the interna top 10% transcripts

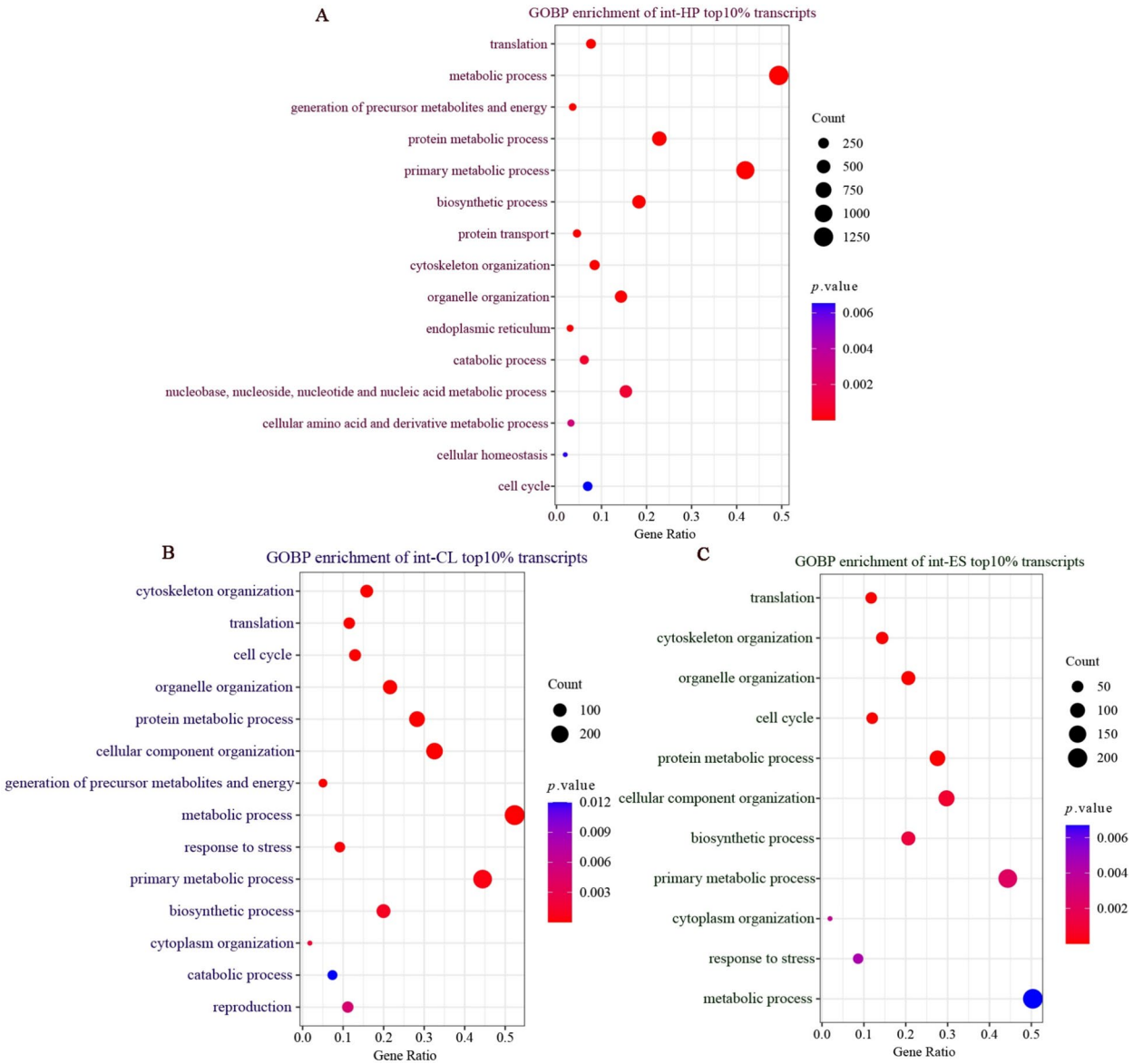

Fig. 5 Gene ontology (GO) enrichment analyses of interna top 10% transcripts. **A** GOBP enrichment of int-HP. **B** GOBP enrichment of int-CL. **C** GOBP enrichment of int-ES

(int-CL) (Fig. 5). The GO terms of molecular function (GOMF) in the int-HP and int-CL were enriched in RNA binding, structural molecule activity, nucleotide binding and peptidase activity. The GOMF terms in the int-HP were also associated with translation factor activity, nucleic acid binding and catalytic activity. The GOMF term in the int-ES was only enriched in structural molecule activity. The enriched GO terms of cellular components (GOCC) in all samples were enriched in organelle, intracellular, cell and cytoplasm. The specific enrichment GOCC terms in the int-HP were enriched in mitochondrion and nucleolus.

The significant KEGG pathways ($p < 0.05$) of int-HP, int-CL and int-ES were 20, 6 and 3, respectively (Fig. 6). The KEGG pathways of the int-HP were involved in metabolism (e.g., lipid metabolism, carbohydrate metabolism, amino acid metabolism, energy metabolism), genetic information processing (e.g., spliceosome, protein export, proteasome, nucleotide excision repair) and cellular processes (endocytosis) (Fig. 6A). The KEGG pathways of the int-CL were associated with energy metabolism (oxidative phosphorylation), genetic information processing (e.g., ribosome, proteasome, spliceosome) and cellular processes (phagosome) (Fig. 6B). The KEGG pathway of

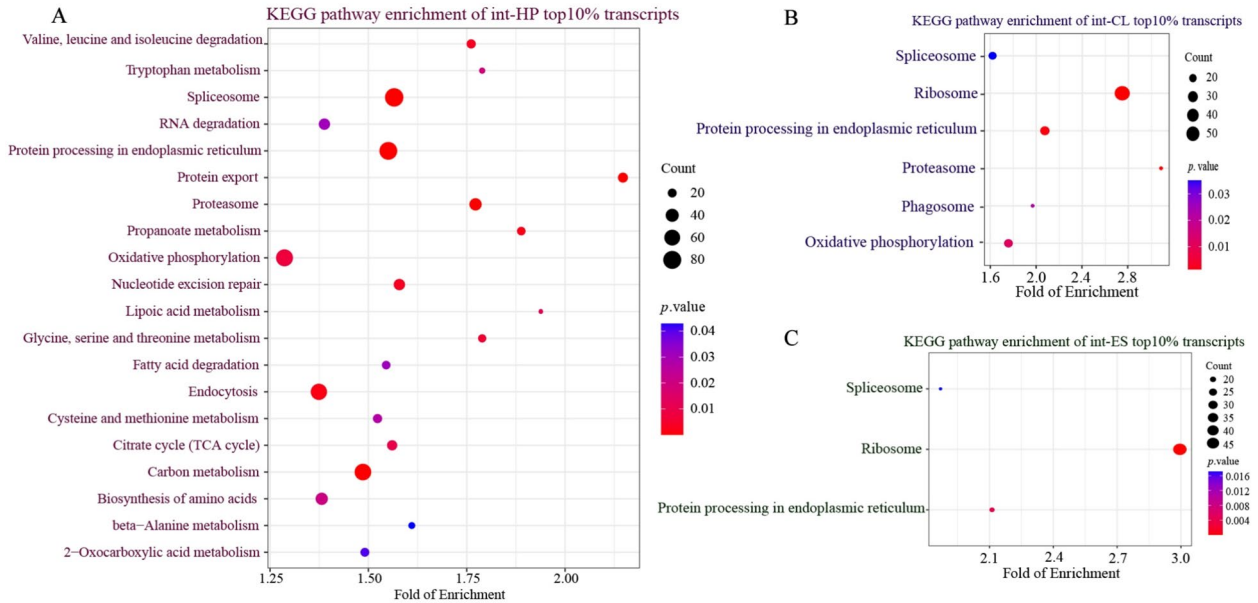

Fig. 6 KEGG pathway enrichment analyses of the interna top 10% transcripts. **A** KEGG pathway of int-HP. **B** KEGG pathway of int-CL. **C** KEGG pathway of int-ES

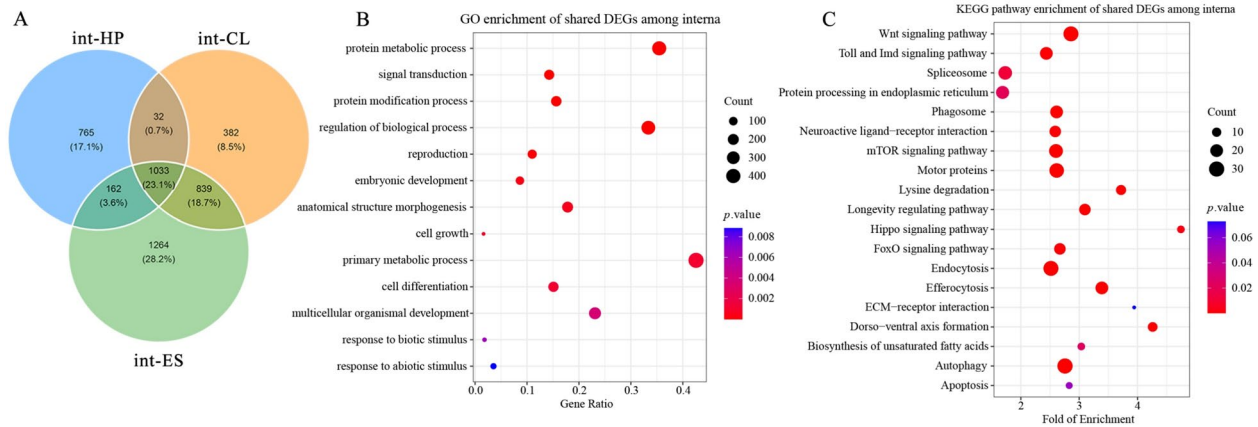

Fig. 7 The analyses of differential expression of genes among internae. **A** The Venn diagram of DEGs among internae. **B** GO terms of biological process of shared DEGs among internae. **C** KEGG pathway of shared DEGs among internae

int-ES was only enriched in genetic information processing (Fig. 6C).

Identification and function enrichment analysis of differential expressions of genes

The interna samples (int-HP, int-CL and int-ES) were compared to the externa samples to screen for interna differentially expressed genes. A total of 1992 DEGs from int-HP/ext comparison, 2286 DEGs from int-CL/ext comparison and 3298 DEGs from int-ES/ext comparison were found to be significantly up-regulated in the interna. Among these DEGs, 699 DEGs of int-HP, 636 DEGs of int-CL, 767 DEGs of int-ES were successfully translated to protein sequences. The Venn diagram revealed that 1033 interna up-regulated DEGs were shared among internae, which accounted for 51.8%, 45.2%, 31.3% of total DEGs in int-HP, int-CL and int-ES, respectively (Fig. 7A). These shared DEGs among internae represented a subset of genes highlighting the core functions of interna. The enriched GOBP terms of shared DEGs were associated with metabolic process, signal transduction, reproduction, response to stimulus, embryonic development, morphogenesis, etc. (Fig. 7B). The enriched GOMF terms of shared DEGs were related to enzyme activity (protein

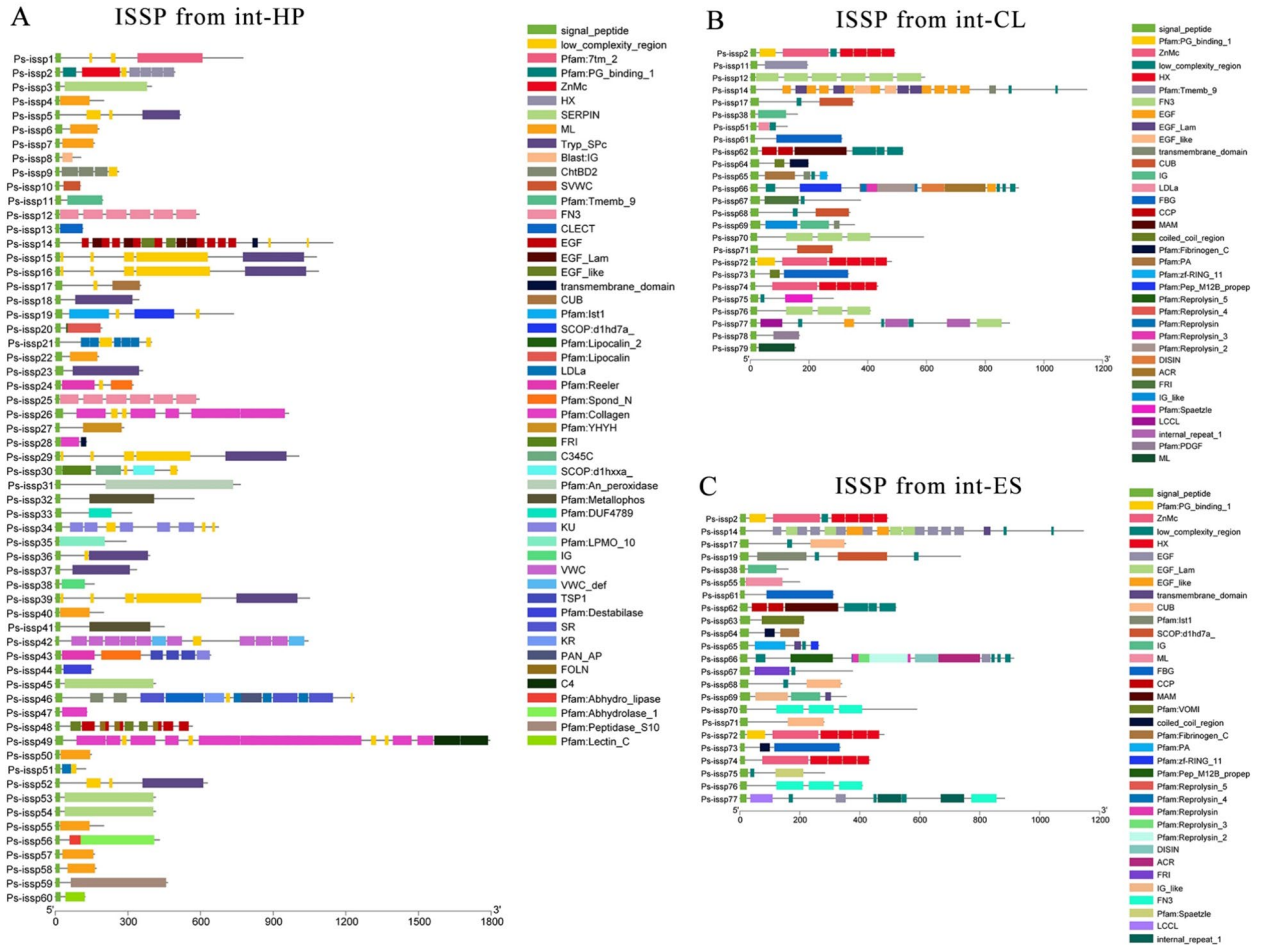

Fig. 8 The specific secretory protein domains of int-HP, int-CL and int-ES

kinase, peptidase, hydrolase, transferase), binding activity, receptor activity, motor activity, neurotransmitter transporter activity, signal transducer activity, and catalytic activity. The KEGG pathways of shared DEGs were involved in metabolism, cellular process (e.g., autophagy, efferocytosis, endocytosis), genetic information processing (spliceosome, protein processing in endoplasmic reticulum), organismal systems (e.g., longevity regulating pathway, Toll and Imd signaling pathway), and environmental information processing (e.g., Wnt signaling pathway, neuroactive ligand-receptor interaction) (Fig. 7C).

We found compelling evidence for a large proportion of specific DEGs of interna samples, indicating that the function of the interna might differ according to the host region they penetrate. The enriched GOBP terms of int-HP and int-CL DEGs were related to response to stress and cell recognition. The enriched GOBP terms of int-ES DEGs were related to protein transport and mitochondrion organization. The GOMF term transcription regulator activity was found to be enriched in the int-HP and int-CL DEGs. The GOMF terms electron carrier activity and structural molecule activity were found to be enriched in int-ES DEGs. For KEGG pathway enrichment analysis, ATP-binding cassette (ABC) transporters were found to be significantly enriched in int-ES DEGs.

Interna-specific secretory proteins

A total of 154, 69 and 83 protein sequences from int-HP, int-CL and int-ES DEGs, respectively, were predicted to contain an N-terminal signal peptide, suggesting that these proteins were secretory proteins. We name these proteins as Interna-specific secretory proteins (ISSP). All ISSPs were subjected to protein domains prediction by SMART domain analysis with Pfam protein domain database as the reference. A total of 45 protein domains from int-HP ISSPs, 31 domains from int-CL ISSPs and 30 domains from int-ES ISSPs were detected (Fig. 8; Additional file 2: Table S2). Among these ISSPs, 11 SMART/Pfam domains were shared in all interna samples, including IG, PG_binding_1, ZnMc, HX, FN3, EGF, EGF_Lam,

EGF_like, CUB, ML and FRI (Table 1). These domains were involved in immunity and lipid metabolism (IG, PG_binding_1, ML), enzymatic digestion (ZnMc), growth and development (EGF, EGF_Lam, EGF_like, CUB). Serine protease related protein domains, including serpin, KU, peptidase_S10, serine proteases trypsin, Tryp_SPc domain containing ISSPs were detected among int-HP and int-ES samples (Additional file 2: Table S2; Additional file 3: Table S3). Additionally, 31 protein domains were only detected in int-HP ISSPs; PDGF domain was only detected in int-CL ISSPs; VOMI domain was only detected in int-ES ISSPs (Fig. 8; Additional file 2: Table S2).

Discussion

The transcriptome data of *P. sinensis* reported in this study represents the first genomic resource for this species, providing comprehensive functional analyses of its interna rootlets. The findings support the hypothesis that the functional biology of interna entangled within the host hepatopancreas, eyestalk and claw muscle fibers exhibit partial differences, while fundamental biological functions such as metabolism, growth and development are consistently active across these tissues. Additionally, several protein domains which participated in the prophenoloxidase (proPO) system signaling pathway and antimicrobial immunity in crustaceans were identified. Furthermore, we list some candidate interna specific secretory proteins (ISSPs) potentially involved in the lipid metabolism and molting processes of host crabs.

Structural specialization and gene expression profile of the interna

The interna tissue network represents the endoparasitic, adult body of rhizocephalans. In *P. sinensis*, the randomly branching and highly ramified interna is primarily concentrated around the host hepatopancreas. A similar distribution pattern has been reported in *P. pilosella* and *S. pugettiae* [19]. DAPI staining showed that the cell nuclei of varying sizes and shapes are arbitrarily distributed throughout the interna (Fig. 2).

Functional enrichment analyses of the interna top 10% transcripts showed that the GOCC terms were involved in organelle, intracellular, cell, cytoplasm, mitochondrion, and nucleolus. Additionally, GO terms related to cell biological processes, cytoskeleton organization, organelle organization, cell cycle, endoplasmic reticulum, cytoplasm organization were significantly enriched (Fig. 5). While Bresciani and Høeg suggested that rhizocephalan internae are generally composed of two cell layers, namely the inner axial cell layer and the epithelial cell layer [15], our results indicated a complex network of interna cells. We observed a unique interconnected F-actin nodal network, an unprecedented structure in crustacean species. Despite the presence of highly expressed cytoskeletal genes, this structure bears no resemblance to skeletal muscle, smooth muscle, or adhering junction. Further investigations are needed to elucidate the details and function of this unique network. Additionally, the high abundance of lipid droplets distributed throughout the interna, which was associated with KEGG pathway "lipid metabolism" in the interna (Fig. 4).

As the interna of rhizocephalan barnacles may intricately invade all parts of the host crab [3, 15] and obtaining sequence reads free of host contaminants is difficult, the biological functions and metabolic capabilities of interna rootlets remain largely unexplored. In order to acquire a clean transcriptome of interna rootlets, the present study applies an *in-silico* subtraction method to remove host contaminating transcripts generated from the host-interna sample mixture. This approach allowed us to identify abundantly expressed transcripts in interna

Table 1 Details of eleven interna specific secretory protein domains shared in int-HP, int-CL and int-ES

Protein domains	SMART/Pfam Accessions	Descriptions
IG	SM00409	Immunoglobulin
PG_binding_1	PF01471	Peptidoglycan binding domain (PGBD)
ZnMc	SM00235	Neutral zinc metallopeptidases
HX	SM00120	Hemopexin-like repeats
FN3	SM00060	Fibronectin type 3 domain
EGF	SM00181	Epidermal growth factor-like domain
EGF_Lam	SM00180	Laminin-type epidermal growth factor-like domain
EGF_like	SM00001	EGF domain, unclassified subfamily
CUB	SM00042	Mostly among developmentally-regulated proteins
ML	SM00737	Domain involved in innate immunity and lipid metabolism
FRI	SM00063	Frizzled

embedded within the host hepatopancreas, claw muscle, and eyestalk. As GO and KEGG enrichment analyses were applied to these abundantly expressed transcripts, we found that GOBP terms related to translation, metabolic process, biosynthetic process, cellular process and GOMF terms involved in RNA binding, structural molecule activity, nucleotide binding, peptidase activity were enriched across all interna samples. These findings strongly implicate that growth, metabolism and development take place in the interna rootlets. The protein transport and several metabolic pathways were significantly enriched in int-HP, suggesting that this might be the major location for absorbing nutrients from the host. This aligns with the hepatopancreas's role in crustaceans as a critical organ for energy storage and breakdown, nutrient accumulation, carbohydrate and lipid metabolism [16]. In addition, the GO terms related to response to stress and reproduction were observed in int-ES and int-CL, indicating these regions might be responsible the adaptation to stress change, the protrusion of externa and infestation of parasites. These results suggest that the interna entangled within the host hepatopancreas, eyestalk and claw muscle fibers harbor similarities and differences in biological function.

Functions of differentially expressed interna genes

The GO terms analyses of shared DEGs among internae were significantly enriched in molecular function and biological process. The GOBP categories of the interna were largely related to embryonic development. The virgin externa has a brood chamber, containing the paired receptacles that eventually host one or two successful males [7]. Thus, our results demonstrate that a subset of molecular functions of the interna is involved with embryonic nutrition and development, possibly through secretion to the externa, where embryos are brooded. This role in embryonic development is particularly important as all rhizocephalans hatch as lecithotrophic (yolk-feeding or non-plankton-feeding) nauplius or cypris larvae [3].

Our GO enrichment analyses also demonstrate that the interna are involved in a variety of enzyme activity and most of them were associated with serine proteases. Serine protease is known to be involved with the prophenoloxidase (proPO) system of crustaceans [20, 21]. The proPO system or melanization is one of the major immune responses in invertebrates and the activation of this system needs the participation of serine proteases (SP) cascades [20, 21]. Similarly, we detected two serine protease inhibitor protein domains from int-HP ISSPs (details provided below). The KEGG pathways neuroactive ligand-receptor interaction and the GOMF terms neurotransmitter transporter activity were found in the interna, suggesting the presence of a functional, or possibly active, neuronal system within the interna. This is surprising and highly interesting as the larval nervous system is entirely disintegrated during the metamorphosis to first the kentrogon, the stage responsible for host penetration and subsequent injection of primordial parasite cells, and is absent in the vermigon, the invasive stage that travels through the host haemolymph and gives rise to the interna [22].

In addition, the shared DEGs among internae only accounted for 31.1–51.8% of the total DEGs in int-HP, int-CL and int-ES (Fig. 7A). This is similar to the findings in a recent study [13] where it was demonstrated that the up-regulated genes between interna samples from the host midgut and thoracic ganglia exhibited partial differences. The KEGG pathway ABC transporters are specifically enriched in int-ES DEGs. ABC transporters are one of the largest transporter families and have crucial roles in various biological processes such as growth, development and immune defense in eukaryotes [23, 24].

Protein domains involved in immunity

Several protein domains were observed in the ISSPs of the *P. sinensis* transcriptome, directly or indirectly associated with immune mechanisms. Among the proteins involved in the proPO system signaling pathway of invertebrates, two serine protease inhibitors domains, serpin and KU, were detected in the int-HP ISSPs (Additional file 2: Table S2; Additional file 3: Table S3). Notably, the inhibition of the melanization by serpins has been reported previously in the Chinese mitten crab *Eriocheir sinensis* infested by *Polyascus gregaria* [25] and the green crab (*Carcinus maenas)* parasitized by *S. carcini* [14]. Interestingly, three protein domains associated with serine proteases were also observed in the int-HP and int-ES ISSPs, including the serine proteases trypsin domain, Peptidase S10 domain (serine carboxypeptidase), and Tryp_SPc domain (Trypsin-like serine protease) (Additional file 2: Table S2; Additional file 3: Table S3). The functions of these domains-containing proteins might be similar to the serine protease, suggesting that these domains play a role in the catalytic activity of melanization processes. These observations suggest that *P. sinensis* could cause the hosts to be more susceptible to accept parasites by inhibiting the proPO system.

Some antimicrobial immune factors involved with immune recognition and elimination in crustaceans were also observed in *P. sinensis* internae. C-type lectin-like, Lectin_C, and CLECT domains which are involved in lectins were detected in the int-HP and int-ES ISSPs (Additional file 2: Table S2; Additional file 3: Table S3). Lectins are proteins or glycoproteins that specifically recognize and bind to glycoproteins on the surface of

microbes to initiate an immune response [26, 27]. C-type lectin contains at least one carbohydrate recognition domain and belongs to the group of proteins known as pattern-recognition receptors in invertebrates [28, 29]. The Myeloid differentiation protein 2 (MD2)-related lipid recognition (ML) domain, PG_binding_1 (peptidoglycan binding domain), and IG (Immunoglobulin) domain were detected in all ISSPs (Table 1). The ML domain-containing protein (*Es*ML3) could bind and agglutinate bacteria by interaction with peptidoglycan and/or lipopolysaccharide (LPS), promoting bacterial clearance and phagocytosis against bacterial infection in crabs [30, 31]. Similarly, the LCCL domain, which may play a role in LPS binding, was identified in the int-CL and int-ES ISSPs. The LPS binding domain-containing proteins were also found in the green crab *C. maenas* parasitized by *S. carcini* [14]. Additionally, other domains that might contribute to the antimicrobial and antiviral immunity in invertebrates were identified in the int-HP ISSPs, such as single domain von Willebrand factor type C domain (SVWC) [32], phospholipase A2 domain [33], destabilase and SR (Scavenger receptor Cys-rich).

Protein domains involved in lipid metabolism, molting and growth

In this study, the ML domain was identified in the int-HP, int-ES and int-CL ISSPs of the *P. sinensis* transcriptome, which was associated with lipid metabolism and molting (ecdysis). Molting is a cyclic process that takes place in all crustaceans. The shedding of old cuticles allows the underlying epidermis to expand and thus grow [34]. Crustaceans must halt feeding and activate lipid metabolism during molting [35]. Thus, lipid metabolism in crustaceans is closely linked to the molting process [36]. Many members of ML family have been shown to regulate lipid metabolism and facilitate other cellular functions involved in lipid recognition [37]. The low-density lipoprotein receptor domain class A (LDLa) was observed in the int-HP and int-CL ISSPs. The low-density lipoprotein receptor-related protein (LRP) plays vital roles in cholesterol uptake and transport [38]. Cholesterol is an indispensable nutrient that regulates molting and growth in crustaceans [39]. As a component of lipoprotein, cholesterol impacts the absorption and transportation of lipids in animals [40]. Thus, we speculate that ML domain and LDLa domain containing proteins may be potential candidates involved with the regulation and intervention of the lipid metabolism and molting process of host crabs.

Finally, the epidermal growth factor (EGF) and EGF like domains were found in all internae. EGF is a critical cell regulatory factor and the mitogen of a variety of cells, involving in regulating cell proliferation, migration and differentiation through the interaction with its receptor (EDFR) [41]. EGFR has been characterized and contributes to the growth and development in *Macrobrachium rosenbergii* [42] and *Scylla paramamosain* [43]. Therefore, the EGF domain containing proteins from *P. sinensis* might be potential candidates to intervene the cell process, growth and development of host crabs.

Conclusion

Our detailed transcriptomic profiling of the *P. sinensis* interna strongly imply that growth, metabolism and development all take place in interna rootlets. There are similarities and differences in the functional biology of interna entangled within host hepatopancreas, eyestalk and claw muscle fibers. Our results suggest *P. sinensis* could manipulate hosts susceptibility to future parasite invasion by inhibiting the proPO system. This study also finds antimicrobial immune factors that are involved in immune recognition and elimination in interna specific secretory proteins. Overall, the present research reveals functional differentiation in the rhizocephalan interna and provides novel insights into host-parasite interactions in the group, laying the groundwork for future studies on the evolution of parasitism in Thecostraca (barnacles in the broadest sense).

Materials and methods

Specimen collection

During May to September 2023, 13 infected and 10 non-infected *L. exaratus* were collected from Wong Chuk Hang (114°29'N, 22°40'E), Hong Kong. Crab species identification was based on the morphological descriptions provided by Amer et al. [44]. To ensure the relatively balanced samplings, four female and seven male hosts (the gender of two hosts was no documented) were collected (Additional file 4: Table S4). As some infected individuals may not have carried externa, it was necessary to assess the health of host individuals by transcriptome sequencing.

All crabs were anesthetized on ice for 5–10 min in laboratory. The hepatopancreas, claw, eyestalk with dense interna, externa from infected crabs (Additional file 4: Table S4) and these tissues from non-infected crabs were immediately dissected and preserved in −80 °C freezer until RNA extraction. Due to the radically simplified morphology of rhizocephalan barnacles, species identification of them was typically determined by their host species. Because the crab *L. exaratus* is the host for three parasitic barnacles [45] and the externa morphology of them is similar, rhizocephalan species identification with morphological tools is best done in concert with molecular data (either Sanger sequencing of *18S* rRNA and *COX1* genes or extracting of these from the RNA-seq data).

Fixation and staining for fluorescence studies

One infected crab was selected for fluorescence studies and was anesthetized upon collection by placing it in crushed ice. This infected crab specimen was dissected and the parasite interna and host hepatopancreas isolated immediately thereafter. Samples were fixed in 4% paraformaldehyde (Macklin, China) dissolved in autoclave filtered seawater overnight at 4 °C. Subsequently, the sample was rinsed three times with 1 × PBS (Solarbio, China). Then the sample was permeabilized in 1 × PBS containing 0.1% Triton X-100 (Sigma, Germany) for 30 min, during which it was incubated at 4 °C on a laboratory shaker. After the permeabilization was completed, the sample was again washed three times with 1 × PBS. The permeabilized sample was then treated with 1 μg/mL of Nile Red (Macklin, China) and 1 μg/mL of iFlour™ 488 Phalloidin (Yeasen Biotechnology, China) for 30 min with orbital shaking. After the staining was completed, the sample was rinsed three times with 1 × PBS. After treatment with DAPI (Beyotime, China) containing 1 μg/mL for 5 min, the sample was mounted onto a glass slide with ProLong™ Diamond Antifade Mountant (Thermofisher, USA). The slides were observed on upright fluorescence microscope (Motic, China). Photographs were taken at 4 ×, 10 × and 40 × magnification. The 40 × images of DAPI, Phalloidin and Nile Red staining of interna were merged using ImageJ [46].

RNA extraction, library construction and Illumina Hiseq sequencing

Total RNA of each tissue sample was extracted using TGuide S96 Magnetic Universal RNA Kit (TIANGEN, China) following manufacturer's protocol. The quantity and quality of the extracted RNAs were checked using Nanodrop (Thermo Fisher Scientific Inc., USA) and Agilent 2100 Bioanalyzer (Agilent Technologies, USA), respectively. Total RNA of 35 parasitic samples (14 interna-hepatopancreas (int-HP), 7 interna-claw (int-CL), 5 interna-eyestalk (int-ES) and 9 externa) from 13 infected *L. exaratus* and 10 non-infected *L. exaratus* samples were used to prepare Illumina cDNA library following the manufacturer's protocol of TruSeq v3 kit . The cDNA library was sequenced using an Illumina Novaseq 6000 machine (Yuda Biotechnology; Guangzhou, China).

Species identifications

To identify the species of parasitic barnacles, we extracted the sequences of two phylogenetic marker genes (*18S* rRNA and *COXI*) from transcriptomes of all parasite individuals externae or internae. The *18S* rRNA sequence of *Balanus eburneus* (GenBank: L26510.1) and *COXI* protein sequence of *P. yatsui* (GenBank: BAF49625.1) were used as queries to extract sequences from the transcriptomes using BLASTN and TBLASTN, respectively. The extracted sequences were then used for preliminary identification of parasitic barnacle using the Basic Local Alignment Search Tool (BLAST) from the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast). The results indicated that all parasitic barnacle samples were identified as *P. sinensis* (Additional file 5: Table S5).

To further confirm the species identification of parasitic barnacles, Maximum Likelihood (ML) phylogenetic tree was reconstructed using *18S* rRNA sequences from all parasite individuals (Additional file 5: Table S5) and other reference species (Additional file 6: Table S6). *Amphibalanus amphitrite* (Cirripedia: Thoracica: Balanomorpha) was selected as the outgroup. Sequences were manually inspected and edited using SeqMan (DNAStar Inc., Madison, WI, USA). All sequences were aligned using MAFFT with default parameters [47]. The best-fit model, GTR + F + R2, was selected based on the Akaike Information Criterion (AIC) by using ModelFinder [48]. ML analysis was performed using IQ-TREE v 2.1.1 [49] with 1000 bootstrap replicates (BP). The phylogenetic tree was visualized using FigTree v1.4.2 [50]. All parasite individuals and the *18S* rRNA sequences of *P. sinensis* obtained from GenBank clustered into a single clade, further confirming that the parasitic barnacles were *P. sinensis* (Additional file 7: Fig. S1).

De novo assembly, removal of host contamination and annotation

The raw paired-end reads were trimmed and quality-controlled with Trimmomatic v0.39 [51] with default parameters. Clean reads from parasitic barnacle samples and non-infected *L. exaratus* samples were obtained and used to conduct de novo assembly of transcriptome using Trinity v2.3.2 [52] with default parameters.

Clean reads from 35 *P. sinensis* samples were mapped against the *P. sinensis* assembly and transcript abundance estimation was performed using Salmon v1.10.1 [53]. As interna rootlets of parasitic barnacles may invade all parts of the host crab [15], the interna-host tissue samples may comprise *P. sinensis* internae and host *L. exaratus* tissues or cells (considered contamination) and therefore the initial transcriptome assembly could contain both *P. sinensis* and host' transcripts. To remove these contaminant transcript sequences, clean reads from non-infected *L. exaratus* samples were mapped against the transcriptome assembly using Salmon v1.10.1 [53]. Sequences with one or more mapped reads from any of the non-infected *L. exaratus* samples were considered as host contamination and removed. The filtered *P. sinensis* assembly was collapsed using CD-HIT-EST v4.8.1 [54] with the overlapping cut-off set at 95%.

BUSCO v5.5.5 [55] was used to evaluate the completeness of the filtered *P. sinensis* transcriptome assembly with the reference database "metazoa_odb10". To facilitate a better functional analysis of the transcriptome, the filtered *P. sinensis* transcriptome was annotated with the genome of *D. melanogaster* using BLASTX (*E*-value cut-off was set as 1^{e-5}). Principal component analysis (PCA) was conducted to assess the similarity and dissimilarity of transcriptomes across samples. The analysis was processed with FactoMineR [56] and the R package ggplot2 [57] was used for visualization.

Functional enrichment analyses of interna top10% transcripts and differential expression of genes

The top10% transcripts abundantly expressed in the interna were selected and annotated with the *D. melanogaster* genome using BLASTX (*E*-value cut-off was set as 1^{e-5}). Gene ontology (GO) enrichment analysis was performed using BiNGO v3.0.5 [58], a plugin of Cytoscape v3.9.1 [59]. The top10% transcripts from each interna sample were selected as the test set and all transcripts expressed in any of interna samples were selected as the reference set. Enrichment analysis results were assessed using Fisher's Exact test with a significant threshold of $p < 0.05$. The enriched terms were visualized using the R package ggplot2 [57]. Additionally, these top10% transcripts were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to annotate their potential metabolic pathways [60]. Transcript-level abundance estimates for the parasitic barnacles were quantified and aggregated into gene abundance metrics (transcripts per million, TPM) and read counts using Salmon v1.10.1 [53]. Transcript expression levels were calculated by aligning each sample to the reference transcriptome (filtered *P. sinensis* transcriptome). Genes with a zero TPM value across all samples were excluded, and only the expressed genes were subjected to the subsequent analyses.

Differential expression analysis of genes between internae samples (int-HP, int-CL and int-ES) and externa samples were conducted using edgeR [61], with thresholds at False discovery rate (FDR) < 0.001 and $|\log_2(\text{fold change})| > 3$. To visualize the distribution of differential expression of genes (DEGs) across samples, a Venn diagram was constructed using Venn diagram R package [62]. All DEGs were used for functional enrichment analysis, including GO and KEGG pathways. GO enrichment analysis was performed using the Cytoscape v3.9.1 [59] plugin BiNGO v3.0.5 [58] tool. The DEGs were selected as the test set and all expressed transcripts as the reference. KEGG pathways analysis was implemented using the Database for Annotation, Visualization and Integrated Discovery (DAVID) [63] (available at https://david.ncifcrf.gov/). A p-value < 0.05 was considered the threshold for significantly enriched GO terms and KEGG pathways.

Identification of signal peptide proteins and protein domains

The TransDecoder v.5.7.1 tool (http://transdecoder.github.io) was used to translate the DEGs of interna samples into protein sequences. These protein sequences were then aligned against the NCBI non-redundant (nr) protein database to evaluate sequence similarity to *D. melanogaster* proteins using BLASTP searches, with an *E*-value cut-off of 1^{e-5}. Protein sequences with more than 98% similarity were filtered to remove redundancy using CD-HIT-EST v4.8.1 [54]. Then, signal peptide prediction of proteins was performed in SignalP 6.0 server [64]. Sequence comparison of these protein sequences and other sequences was performed using BLASTP searches (*E*-value cut-off was set as 1^{e-10}) on the NCBI website (https://blast.ncbi.nlm.nih.gov/). Additionally, conserved protein domains were predicted using the SMART (Simple Modular Architecture Research Tool) database (http://smart.embl-heidelberg.de/) [65] and the Pfam HMM database [66].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11315-4.

Additional file 1.
Additional file 2.
Additional file 3.
Additional file 4.
Additional file 5.
Additional file 6.
Additional file 7.

Acknowledgements
We would like to express our sincere gratitude to Mr. Daniel K. H. Yu, an active high school student (from The Independent School Foundation (ISF) Academy, Hong Kong) and fish-lover, for the information on the geographic location of the prevalence hotspot of rhizocephalan in Hong Kong. Without his findings, this work would not have been possible.

Authors' contributions
YHW conceived and designed the research. HCL and YHF conducted the sample collection. WJL, YHF and HCL performed the experiments and acquired the data. WJL analyzed the data and wrote the manuscript. ND revised the manuscript. YHW revised the manuscript and funded this research program. All authors have read and approved the final version of the manuscript.

Funding
This research was supported by the Guangdong Major Project of Basic and Applied Basic Research (grant number: 2023B0303000017), the National Science Foundation of China General Project (award no. 42276104) awarded to YHW and the Innovation Team Project of Universities in Guangdong Province (No. 2023KCXTD028).

Data availability
The raw Illumina data generated during the current study were uploaded to the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) with accession number: PRJNA1185833.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]College of Civil and Transportation Engineering, Shenzhen University, Shenzhen, Guangdong 518000, China. [2]Institute for Advanced Study, Shenzhen University, Shenzhen 518000, Guangdong, China. [3]Nanopore Sequencing Platform, Centre de Recherche du CHU Sainte-Justine, Montréal, QC, Canada. [4]Centre de Recherche du CHU Sainte-Justine, Montréal, QC, Canada. [5]Département de Pédiatrie, Université de Montréal, Montréal, QC, Canada. [6]Present address: Institute for Advanced Study, Shenzhen University, Shenzhen 518000, Guangdong, China.

Received: 28 November 2024 Accepted: 3 February 2025
Published online: 22 February 2025

References

1. Price PW. General concepts on the evolutionary biology of parasites. Evolution. 1977;31(2):405–20. https://doi.org/10.2307/2407761.
2. Poulin R, Randhawa HS. Evolution of parasitism along convergent lines: from ecology to genomics. Parasitology. 2015;142(S1):S6–15. https://doi.org/10.1017/S0031182013001674.
3. Høeg JT. The biology and life cycle of the Rhizocephala (Cirripedia). J Mar Biol Assoc UK. 1995;75(3):517–50. https://doi.org/10.1017/S0025315400038996.
4. Høeg JT, Lützen J. Life cycle and reproduction in the Cirripedia Rhizocephala. Oceanogr Mar Biol. 1995;33:427–85.
5. Waiho K, Glenner H, Miroliubov A, Noever C, Hassan M, Ikhwanuddin M, Fazhan H. Rhizocephalans and their potential impact on crustacean aquaculture. Aquaculture. 2021;531:735876. https://doi.org/10.1016/j.aquaculture.2020.735876.
6. Walker G. Introduction to the Rhizocephala (Crustacea: Cirripedia). J Morphol. 2001;249(1):1–8. https://doi.org/10.1002/jmor.1038.
7. Høeg JT, Glenner H, Shields J. Cirripedia Thoracica and Rhizocephala (barnacles). In: Rohde K, editor. Marine parasites. Wallingford and Collingwood: CSIRO; 2005. p. 154–65.
8. Waiho K, Fazhan H, Glenner H, Ikhwanuddin M. Infestation of parasitic rhizocephalan barnacles *Sacculina beauforti* (Cirripedia, Rhizocephala) in edible mud crab. Scylla olivacea PeerJ. 2017;5:e3419. https://doi.org/10.7717/peerj.3419.
9. Fazhan H, Waiho K, Wee HB, Surzanne MA, Ma H, Ikhwanuddin M. Predicting the sacculinid *Sacculina beauforti* infection status of the orange mud crab *Scylla olivacea* by discriminant analysis. Aquaculture. 2018;491:128–34. https://doi.org/10.1016/j.aquaculture.2018.03.009.
10. Rubiliani C. Action of a rhizocephalan on the genital activity of host male crabs: characterization of a parasitic secretion inhibiting spermatogenesis. International journal of invertebrate reproduction. 1983;6(3):137–47. https://doi.org/10.1080/01651269.1983.10510036.
11. Isaeva VV, Shukalyuk AI, Trofimova AV, Korn OM, Rybakov AV. The structure of colonial interna in *Sacculina polygenea* (Crustacea: Cirripedia: Rhizocephala). Crustacean Research. 2001;30:133–46. https://doi.org/10.18353/crustacea.30.0_133.
12. Shirley SM, Shirley TC, Meyers TR. Hemolymph responses of Alaskan king crabs to rhizocephalan parasitism. Can J Zool. 1986;64(8):1774–81. https://doi.org/10.1139/z86-267.
13. Martin S, Lesny P, Glenner H, Hecht J, Vilcinskas A, Bartolomaeus T, et al. Genomic adaptations to an endoparasitic lifestyle in the morphologically atypical crustacean *Sacculina carcini* (Cirripedia: Rhizocephala). Genome Biol Evol. 2022;14(10):evac149. https://doi.org/10.1093/gbe/evac149.
14. Zatylny-Gaudin C, Hervé O, Dubos MP, Rabet N, Henry J, Liittschwager K, Fabienne A. Differential analysis of the haemolymph proteome of *Carcinus maenas* parasitized by *Sacculina carcini* (Cirripeda, Rhizocephala) reveals potential mechanisms of parasite control. Fish Shellfish Immunol. 2023;141:109064. https://doi.org/10.1016/j.fsi.2023.109064.
15. Bresciani J, Høeg JT. Comparative ultrastructure of the root system in rhizocephalan barnacles (Crustacea: Cirripedia: Rhizocephala). J Morphol. 2001;249(1):9–42. https://doi.org/10.1002/jmor.1039.
16. Wang L, Yan B, Liu N, Li Y, Wang Q. Effects of cadmium on glutathione synthesis in hepatopancreas of freshwater crab. *Sinopotamon yangtsekiense* Chemosphere. 2008;74(1):51–6. https://doi.org/10.1016/j.chemosphere.2008.09.025.
17. Xu ZN, Liu A, Li SK, Wang GZ, Ye HH. Hepatopancreas immune response during molt cycle in the mud crab, *Scylla paramamosain*. Sci Rep. 2020;10(1):13102. https://doi.org/10.1038/s41598-020-70139-2.
18. Chen HY, Toullec JY, Lee CY. The crustacean hyperglycemic hormone superfamily: progress made in the past decade. Front Endocrinol. 2020;11:578958. https://doi.org/10.3389/fendo.2020.578958.
19. Golubinskaya DD, Korn OM, Sharina SN, Selin NI. Sympatric two-species infestation by rhizocephalan barnacle parasites in the spider crab Pugettia aff. ferox Ohtsuchi & Kawamura, from Peter the Great Bay (northwestern Sea of Japan). Zool Stud. 2019;2021(60):54. https://doi.org/10.6620/ZS.2021.60-54.
20. Sritunyalucksana K, Söderhäll K. The proPO and clotting system in crustaceans. Aquaculture. 2000;191(1–3):53–69. https://doi.org/10.1016/S0044-8486(00)00411-7.
21. Cerenius L, Söderhäll K. The prophenoloxidase-activating system in invertebrates. Immunol Rev. 2004;198(1):116–26. https://doi.org/10.1111/j.0105-2896.2004.00116.x.
22. Glenner H. Cypris metamorphosis, injection and earliest internal development of the Rhizocephalan *Loxothylacus panopaei* (gissler). Crustacea: Cirripedia: Rhizocephala: Sacculinidae. J Morphol. 2001;249(1):43–75. https://doi.org/10.1002/jmor.1040.
23. Dermauw W, Van Leeuwen T. The ABC gene family in arthropods: Comparative genomics and role in insecticide transport and resistance. Insect Biochem Mol Biol. 2014;45:89–110. https://doi.org/10.1016/j.ibmb.2013.11.001.
24. Luo SS, Chen XL, Wang AJ, Liu QY, Peng M, Yang CL, et al. Genome-wide analysis of ATP-binding cassette (ABC) transporter in *Penaeus vannamei* and identification of two ABC genes involved in immune defense against *Vibrio parahaemolyticus* by affecting NF-κB signaling pathway. Int J Biol Macromol. 2024;262:129984. https://doi.org/10.1016/j.ijbiomac.2024.129984.
25. Feng CC, Zhang JB, Bao J, Luan DW, Jiang N, Chen QJ. Transcriptome analysis of germ cell changes in male Chinese mitten crabs (*Eriocheir sinensis*) induced by rhizocephalan parasite. Polyascus gregaria Front Mar Sci. 2023;10:1144448. https://doi.org/10.3389/fmars.2023.1144448.
26. Dam TK, Brewer CF. Lectins as pattern recognition molecules: the effects of epitope density in innate immunity. Glycobiolog. 2010;20(3):270–9. https://doi.org/10.1093/glycob/cwp186.
27. Lin Z, Wang JL, Cheng Y, Wang JX, Zou Z. Pattern recognition receptors from lepidopteran insects and their biological functions. Dev Comp Immunol. 2020;108:103688. https://doi.org/10.1016/j.dci.2020.103688.
28. Zelensky AN, Gready JE. Comparative analysis of structural properties of the C-type-lectin-like domain (CTLD). Proteins. 2003;52(3):466–77. https://doi.org/10.1002/prot.10626.
29. Christophides GK, Vlachou D, Kafatos FC. Comparative and functional genomics of the innate immune system in the malaria vector *Anopheles gambiae*. Immunol Rev. 2004;198(1):127–48. https://doi.org/10.1111/j.0105-2896.2004.0127.x.
30. Wang Y, Zhao S, Zhang B, Ma HY, Fang WH, Sheng WQ, et al. A novel ML domain-containing protein (*Sp*MD2) functions as a potential LPS receptor involved in anti- *Vibrio* immune response. Dev Comp Immunol. 2020;103:103529. https://doi.org/10.1016/j.dci.2019.103529.
31. Song Y, Zhou KM, Nan XY, Qin YK, Zhao K, Li WW, et al. A novel ML protein functions as a pattern recognition protein in antibacterial responses in

Eriocheir sinensis. Dev Comp Immunol. 2022;127:104310. https://doi.org/10.1016/j.dci.2021.104310.
32. Qin YK, Luo Z, Zhao K, Nan XY, Guo YN, Li WW, et al. A new SVWC protein functions as a pattern recognition protein in antibacterial responses in Chinese mitten crab (*Eriocheir sinensis*). Fish Shellfish Immunol. 2022;131:1125–35. https://doi.org/10.1016/j.fsi.2022.11.004.
33. Lin CY, Xu WB, Li BZ, Shu MA, Zhang YM. Identification and functional analysis of cytosolic phospholipase A2 (cPLA2) from the red swamp crayfish *Procambarus clarkii*: The first evidence of cPLA2 involved in immunity in invertebrates. Fish Shellfish Immunol. 2023;140:108944. https://doi.org/10.1016/j.fsi.2023.108944.
34. Dittel AI, Epifanio CE. Invasion biology of the Chinese mitten crab *Eriochier sinensis*: A brief review. J Exp Mar Biol Ecol. 2009;374(2):79–92. https://doi.org/10.1016/j.jembe.2009.04.012.
35. Huang X, Feng Y, Duan J, Xiong G, Fan W, Liu S, et al. Antistarvation strategies of *E. sinensis*: regulatory networks under hepatopancreas consumption. Oxidative Med Cell Longev. 2020;2020:6085343. https://doi.org/10.1155/2020/6085343.
36. Li Z, Zhang GB, Pan KW, Niu XJ, Shu-Chien AC, Chen T, et al. Functional transcriptome reveals hepatopancreatic lipid metabolism during the molting cycle of the Chinese mitten crab *Eriocheir sinensis*. Comp Biochem Physiol A-Mol Integr Physiol. 2023;284:111474. https://doi.org/10.1016/j.cbpa.2023.111474.
37. Inohara N, Nuñez G. ML - a conserved domain involved in innate immunity and lipid metabolism. Trends Biochem Sci. 2002;27(5):219–21. https://doi.org/10.1016/S0968-0004(02)02084-4.
38. Cortes VA, Busso D, Maiz A, Arteaga A, Nervi F, Rigotti A. Physiological and pathological implications of cholesterol. Front Biosci. 2014;19(3):416–28. https://doi.org/10.2741/4216.
39. Zhu TT, Zhou QC, Yang Z, Zhang YY, Luo JX, Zhang XS, et al. Dietary cholesterol promotes growth and ecdysone signalling pathway by modulating cholesterol transport in swimming crabs (*Portunus trituberculatus*). Anim Nutr. 2022;10:249–60. https://doi.org/10.1016/j.aninu.2022.05.001.
40. Yepiz-Plascencia G, Vargas-Albores F, Higuera-Ciapara I. Penaeid shrimp hemolymph lipoproteins. Aquaculture. 2000;191(1–3):177–89. https://doi.org/10.1016/S0044-8486(00)00427-0.
41. Barker S, Vinson GP. Epidermal growth factor in breast cancer. Int J Biochem. 1990;22(9):939–45. https://doi.org/10.1016/0020-711X(90)90199-D.
42. Sharabi O, Ventura T, Manor R, Aflalo ED, Sagi A. Epidermal growth factor receptor in the prawn *Macrobrachium rosenbergii* : Function and putative signaling cascade. Endocrinology. 2013;154(9):3188–96. https://doi.org/10.1210/en.2013-1259.
43. Lu B, Jiang QL, Liu A, Huang HY, Ye HH. Stimulatory roles of epidermal growth factor receptor (EGFR) in ovarian development of mud crab *Scylla paramamosain*. Gen Comp Endocrinol. 2020;299:113616. https://doi.org/10.1016/j.ygcen.2020.113616.
44. Amer MA, Naruse T, Reimer JD. Morphological and molecular investigation of some xanthid crabs from the Egyptian coast of the Red Sea. Thalassas. 2023;39(1):273–86. https://doi.org/10.1007/s41208-022-00510-9.
45. Chan BKK, Poon DYN, Walker G. Distribution, adult morphology, and larval development of *Sacculina sinensis* (Cirripedia: Rhizocephala: Kentrogonida) in Hong Kong coastal waters. J Crustac Biol. 2005;25(1):1–10. https://doi.org/10.1651/C-2495.
46. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671–5. https://doi.org/10.1038/nmeth.2089.
47. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059–66. https://doi.org/10.1093/nar/gkf436.
48. Kalyaanamoorthy S, Minh B, Wong T, von Haeseler A, Jermiin L. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9. https://doi.org/10.1038/NMETH.4285.
49. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530–4. https://doi.org/10.1093/molbev/msaa131.
50. Rambaut A. FigTree: tree figure drawing tool. 2014. http://tree.bio.ed.ac.uk/software/figtree. Accessed 9 Jul 2014.
51. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170.
52. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8(8):1494–512. https://doi.org/10.1038/nprot.2013.084.
53. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14(4):417–9. https://doi.org/10.1038/nmeth.4197.
54. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150–2. https://doi.org/10.1093/bioinformatics/bts565.
55. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2. https://doi.org/10.1093/bioinformatics/btv351.
56. Lê S, Josse J, Husson F. FactoMineR: An R package for multivariate analysis. J Stat Softw. 2008;25:1–18. https://doi.org/10.18637/jss.v025.i01.
57. Villanueva RAM, Chen ZJ. ggplot2: Elegant Graphics for Data Analysis, 2nd edition. Meas-Interdiscip Res Perspect. 2019;17(3):160–7. https://doi.org/10.1080/15366367.2019.1565254.
58. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics. 2005;21(16):3448–9. https://doi.org/10.1093/bioinformatics/bti551.
59. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504. https://doi.org/10.1101/gr.1239303.
60. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2007;36(suppl 1):D480–4. https://doi.org/10.1093/nar/gkm882.
61. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40. https://doi.org/10.1093/bioinformatics/btp616.
62. Chen H, Boutros PC. VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 2011;12:1–7. https://doi.org/10.1186/1471-2105-12-35.
63. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022;50(W1):W216–21. https://doi.org/10.1093/nar/gkac194.
64. Teufel F, Almagro JJA, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40(7):1023–5. https://doi.org/10.1038/s41587-021-01156-3.
65. Letunic I, Khedkar S, Bork P. SMART: recent updates, new developments and status in 2020. Nucleic Acids Res. 2021;49(D1):D458–60. https://doi.org/10.1093/nar/gkaa937.
66. Finn RD, Clements J, Eddy SR. HMMER web server: Interactive sequence similarity searching. Nucleic Acids Res. 2011;39(suppl 2):W29–37. https://doi.org/10.1093/nar/gkr367.

Publisher's Note