Journal of Invertebrate Pathology 207 (2024) 108198

Contents lists available at ScienceDirect

Journal of Invertebrate Pathology

journal homepage: www.elsevier.com/locate/jip

Comparative genomic analysis of copepod humoral immunity genes with sex-biased expression in *Labidocera rotunda*

Jimoon Jun [a,1], Eun-Jeong Kim [a,1], Donggu Jeon [a], Jihye Yang [a], Hyeon Gyeong Jeong [b], Hyungtaek Jung [d], Taeho Kim [c,*], Seong-il Eyun [a,*]

[a] Department of Life Science, Chung-Ang University, Seoul 06974, South Korea
[b] Department of Taxonomy and Systematics, National Marine Biodiversity Institute of Korea, Seocheon 33662, South Korea
[c] Department of Marine Production Management, Chonnam National University, Yeosu 59626, South Korea
[d] National Centre for Indigenous Genomics, Australian National University, Acton, Australia

ARTICLE INFO

Keywords:
Crustacea
Fish aquaculture
Pattern recognition receptors
Innate immune

ABSTRACT

Studies of innate immune system function in invertebrates have contributed significantly to our understanding of the mammalian innate immune system. However, in-depth research on innate immunity in marine invertebrates remains sparse. We generated the first *de novo* genome and transcriptome sequences of copepod *Labidocera rotunda* using Illumina paired-end data and conducted a comparative genome analysis including five crustaceans (four copepods and one branchiopod species). We cataloged the presence of Toll, Imd, JAK/STAT, and JNK pathway components among them and compared them with 17 previously reported diverse arthropod species representative of insects, myriapods, chelicerates, and malacostracans. Our results indicated that copepod Gram-negative binding proteins may function in direct digestion or pathogen killing. The phylogenetic analysis of arthropod TEP and copepod-specific GCGEQ motif patterns suggested that the evolutionary history of copepod TEPs may have diverged from that of other arthropods. We classified the copepod Toll-like receptors identified in our analysis as either vertebrate or protostome types based on their cysteine motifs and the tree built with their Toll/interleukin-1 receptor domains. LrotCrustin, the first copepod AMP, was identified based on the structure of its WAP domain and deep-learning AMP predictors. Gene expression level analysis of *L. rotunda* innate immunity-related transcripts in each sex showed higher Toll pathway-related expression in male *L. rotunda* than in females, which may reflect an inverse correlation between allocation of reproductive investment and elevated immune response in males. Taken together, the results of our study provide insight into copepod innate immunity-related gene families and illuminate the evolutionary potential of copepods relative to other crustaceans.

1. Introduction

All organisms are exposed to outside pathogens throughout their lifetimes. The strategies whereby they handle invaders are generally categorized as their innate immune system and adaptive immune system based on somatic receptor diversity and immunological memory (Hoffmann et al., 1999). Lacking the adaptive immune system of vertebrates (Myllymäki et al., 2014), invertebrate organisms defend themselves using only the innate immune system, which distinguishes self from non-self and produces effectors that target and kill invaders such as viruses, bacteria, and fungi (Palmer and Jiggins, 2015).

Many invertebrates rely on a pre-encoded set of proteins known as pattern recognition receptors (PRRs) to recognize various microbial ligands (Lai and Aboobaker, 2017). Arthropods, such as insects and crustaceans, possess a diverse array of PRRs, including Toll-like receptors (TLRs), peptidoglycan recognition proteins (PGRPs), and C-type

Abbreviations: A2M, alpha-2 macroglobulin; ALF, anti-lipopolysaccharide factor; AMP, antimicrobial peptide; CTL, C-type lectin; CTLD, C-type lectin-like domain; DEG, differentially expressed gene; Dome, Domeless; Dif, Dorsal-related immunity factor; FC, fold change; GNBP, Gram-negative binding protein; Hop, Hopscotch; LRR, leucine-rich repeats; MCR, macroglobulin complement-related; ORF, open reading frame; PRR, pattern recognition receptor; PAMP, pathogen-associated molecular pattern; PGRP, peptidoglycan-recognition protein; SPZ, Späetzle; TEP, thioester-containing protein; TIR, Toll/interleukin-1 receptor; TLR, Toll-like receptor.

* Corresponding authors.
E-mail addresses: kimth@chonnam.ac.kr (T. Kim), eyun@cau.ac.kr (S.-i. Eyun).

[1] These authors contributed equally to this study.

https://doi.org/10.1016/j.jip.2024.108198
Received 4 January 2024; Received in revised form 31 August 2024; Accepted 11 September 2024
Available online 21 September 2024
0022-2011/© 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

lectins. In insects like *Drosophila melanogaster*, TLRs are less numerous and structurally simpler than those in vertebrates, reflecting a specialized immune response. In contrast, mollusks, such as snails and oysters, have an extensive array of PRRs, including TLRs, NOD-like receptors, and scavenger receptors (Lin et al., 2020). The expansion of TLRs in mollusks, like the Pacific oyster (*Crassostrea gigas*), suggests a complex immune system capable of responding to diverse pathogens (Zhang et al., 2015). Cnidarians, including corals and jellyfish, possess a simpler set of PRRs, with fewer TLRs and a greater reliance on antimicrobial peptides, reflecting their ancient evolutionary origins (Miller et al., 2007). Annelids, such as earthworms and leeches, also exhibit diverse PRRs, like TLRs and lectin-like receptors, likely due to the varied functions of these PRRs in their diverse habitat (Prochazkova et al., 2020).

The PRRs do not detect every possible invader but rather focus on pathogen-associated molecular patterns (PAMPs) that are essential for the survival of the microorganism and are therefore difficult for the microorganism to alter (Akira et al., 2006; Medzhitov and Janeway, 2000). Examples of PAMPs include peptidoglycans and lipoteichoic acids in Gram-positive bacteria, lipopolysaccharides in Gram-negative bacteria, and beta-glucans from fungal cell walls (Lai and Aboobaker, 2017). Here, we dealt with four distinct PRR families: PGRPs, Gram-negative binding proteins (GNBPs), C-type lectins (CTLs), and thioester-containing proteins (TEPs).

The PGRPs are responsible for recognizing the peptidoglycan component of the cell wall of bacteria (Dziarski and Gupta, 2006). The catalytic PGRPs such as PGRP-SC1/2, PGRP-SB1/2, and PGRP-LB have amidase activity that can enzymatically break down the peptidoglycan of the invaders. They can negatively regulate the activated immune response or serve as effectors that can kill bacteria by degrading the peptidoglycan layer that comprises the bacterial cell wall (Palmer and Jiggins, 2015). On the other hand, noncatalytic PGRPs, which do not possess hydrolyzing peptidoglycan ability such as PGRP-LC, PGRP-LE, and PGRP-SA, are the PRRs that are necessary to activate the TLR and Imd pathways in *Drosophila*.

In *Drosophila*, GNBP3 is responsible for Toll pathway-independent yeast and fungal recognition (Gottar et al., 2006; Matskevich et al., 2010). Conversely, GNBP1 functions as a linker between PGRP-SA and modular serine proteases during Toll pathway engagement and helps facilitate recognition of Gram-positive bacteria (Buchon et al., 2009).

The CTLs are a superfamily of more than 1,000 proteins that are defined by having one or more CTLDs (Brown et al., 2018). They can bind and recognize various substances, from carbohydrates, such as mannose-type and galactose-type carbohydrates, to a broader repertoire of ligands including proteins or lipids (Zelensky and Gready, 2005). From an antimicrobial perspective, their prodigious ability to recognize and bind to a wide variety of substrates, such as the carbohydrate-rich cell walls of fungi (Brown et al., 2018), cell wall glycolipid of bacteria (Yonekawa et al., 2014), and glycoconjugate structures possessed by parasites including protozoa, nematodes, and helminths, have been reported consistently (Vázquez-Mendoza et al., 2013).

TEPs are named for their conservative thioester (TE) GCGEQ motif. When they encounter various stressful conditions, the activated TE motif binds to nearby hydroxyl and amine groups, which also include the molecules of pathogen surfaces (Bou Aoun et al., 2011). The activated GCGEQ motif also promotes endocytotic clearance and neutralizes pathogenic proteases. These highly conserved protein families include the vertebrate complement system, the pan-protease inhibitor A2M, insect TEP-like proteins, and MCR proteins (Nonaka and Yoshizaki, 2004).

After an organism recognizes invaders, the innate immune system starts to eliminate non-self cells. In *Drosophila*, the innate immune system can be further subdivided into two different classes: cellular immunity and humoral immunity. Cellular immunity relies on the function of hemocytes that circulate throughout the whole body of *Drosophila* via the hemolymph, which serves a similar physiological function as blood in the mammalian circulatory system. The hemocytes contribute to the innate immune response by phagocyting small molecules, as a vertebrate macrophage or monocyte does, participating in myelinization or encapsulating relatively big molecules such as parasitic wasp eggs that plasmacytes cannot phagocytize (Yu et al., 2022). Unlike cellular immunity, humoral immunity, also known as the systemic immune response, relies on antimicrobial peptides (AMPs). The recognition of pathogens activates signaling pathways such as the Toll and the Imd pathways.

The Toll pathway is mainly activated by Gram-positive bacteria and fungi (Michel et al., 2001). It contributes to both cellular and humoral immunity, as well as development. In the innate immune response, once the invader PAMPs are recognized by host PRR (β-glucan of fungi/GNBP3 in *Drosophila*, for example), the signals stemming from recognition processes are integrated by several serine proteases, such as modular serine protease, Gram-positive-specific serine protease, and Spirit (Buchon et al., 2009). This cascade activates the Späetzle-processing enzyme, which finally cleaves Späetzle (SPZ), a ligand of the Toll receptor. Under basal conditions, the NFκB family transcription factor Dorsal and Dorsal-related immunity factor (Dif) are bound to Cactus and primed for nuclear translocation. When the Toll receptor is activated by binding of SPZ, the downstream adaptor protein MyD88 forms a complex with the kinase Pelle and the adaptor protein Tube. The MyD88-Pelle-Tube complex phosphorylates and degrades Cactus. Dorsal and Dif are then translocated to the nucleus, which promotes immune-related gene expression (Yu et al., 2022).

The Imd pathway is mainly activated by Gram-negative bacteria (Myllymäki et al., 2014). When PGRP-LC recognizes peptidoglycans (PGNs) of Gram-negative bacteria, a complex made up of Imd, Fadd, and Dredd is formed. The activated Dredd cleaves Imd, which induces the recruitment and activation of a complex composed of Tab2 and TAK1. The Imd-Tab2-TAK1 complex then activates the IKKβ/IKKγ complex. Finally, the complex phosphorylates and cleaves Rel to Rel-49 and Rel-68, and Rel-68 is translocated to the nucleus to promote immune-related gene expression (Yu et al., 2022). The activated cascades lead to the production of AMPs that are subsequently released into the hemolymph, which destabilizes the cell walls of invading microorganisms, ultimately leading to lysis and death.

The JAK/STAT signaling pathway is triggered by a local immune response and promotes epithelial repair (Agaisse and Perrimon, 2004). In *Drosophila*, the ligands Upd1, Upd2, and Upd3 bind to Domeless (Dome) (Brown et al., 2001). When Dome activates Hopscotch (Hop; a JAK unique to *Drosophila*), Hop phosphorylates STAT92E, which facilitates nuclear translocation. In *Drosophila*, the activated JAK/STAT pathway promotes the expression of Turandot A, which enhances resistance to multiple forms of stress, such as bacterial infection, heat shock, and ultraviolet light exposure. It has also been reported that the JAK/STAT pathway plays a role in cellular immunity processes, such as plasmatocyte and lamellocyte differentiation (Yu et al., 2022).

The JNK pathway is composed of Msn, TAK1, Hep, and Bsk. Activated Bsk sequentially phosphorylates JRA and KAY, which translocate into the nucleus and form a heterodimer transcription factor, AP-1. With target gene expression enhanced by AP-1, the JNK pathway contributes to multiple wound-healing processes, and also plays an antitumorigenic role. In humoral immunity, it has been reported that the Imd pathway and JNK pathway generate AMPs in a synergistic fashion (Yu et al., 2022).

Copepods, our primary focus in this study, are a group of small crustaceans that have undergone numerous adaptations that have allowed them to thrive in diverse environments, from the open ocean to freshwater lakes and rivers (Eyun, 2017; Eyun et al., 2017; Jeon et al., 2024).

Representing the largest biomass of all animals on earth (Song et al., 2021; Song et al., 2024), copepods play a key role as intermediate hosts of parasites. While the study of innate immune defenses in terrestrial organisms has advanced explosively in recent years, innate immunity in aquatic organisms has been largely ignored in the field until now. The

relationships between their widely varied evolutionary adaptations and their immune systems may provide key insight into the history of evolutionary change across all arthropods. Moreover, since copepods serve as intermediate hosts of parasites, our work here will also be invaluable to the cultivation of aquatic food crop species that are prone to parasitic damage.

Here, we conducted an extensive analysis of gene families that contribute to pathogen recognition and the signaling pathways engaged specifically in the humoral immune response (Fig. 1). The study was performed mainly by comparing a copepod *Labidocera rotunda* that we sequenced and assembled from this study to other copepods and comparing copepods including *L. rotunda* to *Drosophila* for the insufficient reports of copepod's innate immunity property. We also identified the first known *Labidocera* AMPs, compounds known to be produced by humoral immunity-associated signaling cascades in other species. We further evaluated gene expression of putative AMPs, segmenting our analysis by sex in order to evaluate potential sexual dimorphism in copepods under basal conditions (free of invading pathogens). To sum up, this study aimed to thoroughly characterize the innate immune system of the non-insect invertebrate copepod *L. rotunda* (Jun et al., 2022) by comparing 24 major components of four key innate immunity signaling pathways (Toll, Imd, JAK/STAT, and JNK) and by interrogating sex-specific differences in relative gene expression levels.

2. Materials and methods

2.1. Sample collection for Labidocera rotunda and next generation sequencing

We collected *Labidocera rotunda* from Ihotaewoo Beach in Jeju Island, Korea (33°31′51.3″N, 126°25′15.4″E and 33°31′54.5″N, 126°25′33.6″E) on July 21st, 2021 using a conical net (mesh size 200μm; mouth diameter 60cm). The total genomic DNA was extracted from five individuals. With the total genomic DNA, a paired-end 151 bp Illumina sequencing library was generated for sequencing using the Illumina Novaseq 6000 platform (Illumina, USA). All libraries were prepared using the TruSeq DNA Nano 550bp kit (Illumina, USA) according to the manufacturer's suggested protocols. mRNA was isolated from two male and two female *L. rotunda* specimens collected on the same day and at the same location as described above. We generated a paired-end 101 bp Illumina sequencing library for *L. rotunda* RNA-seq analysis using the Illumina Novaseq 6000 platform.

2.2. De novo genome and transcriptome assembly

The sequenced genomic DNA reads were trimmed and assembled using the MaSuRCA assembler (ver. 4.0.5) (Zimin et al., 2013) with

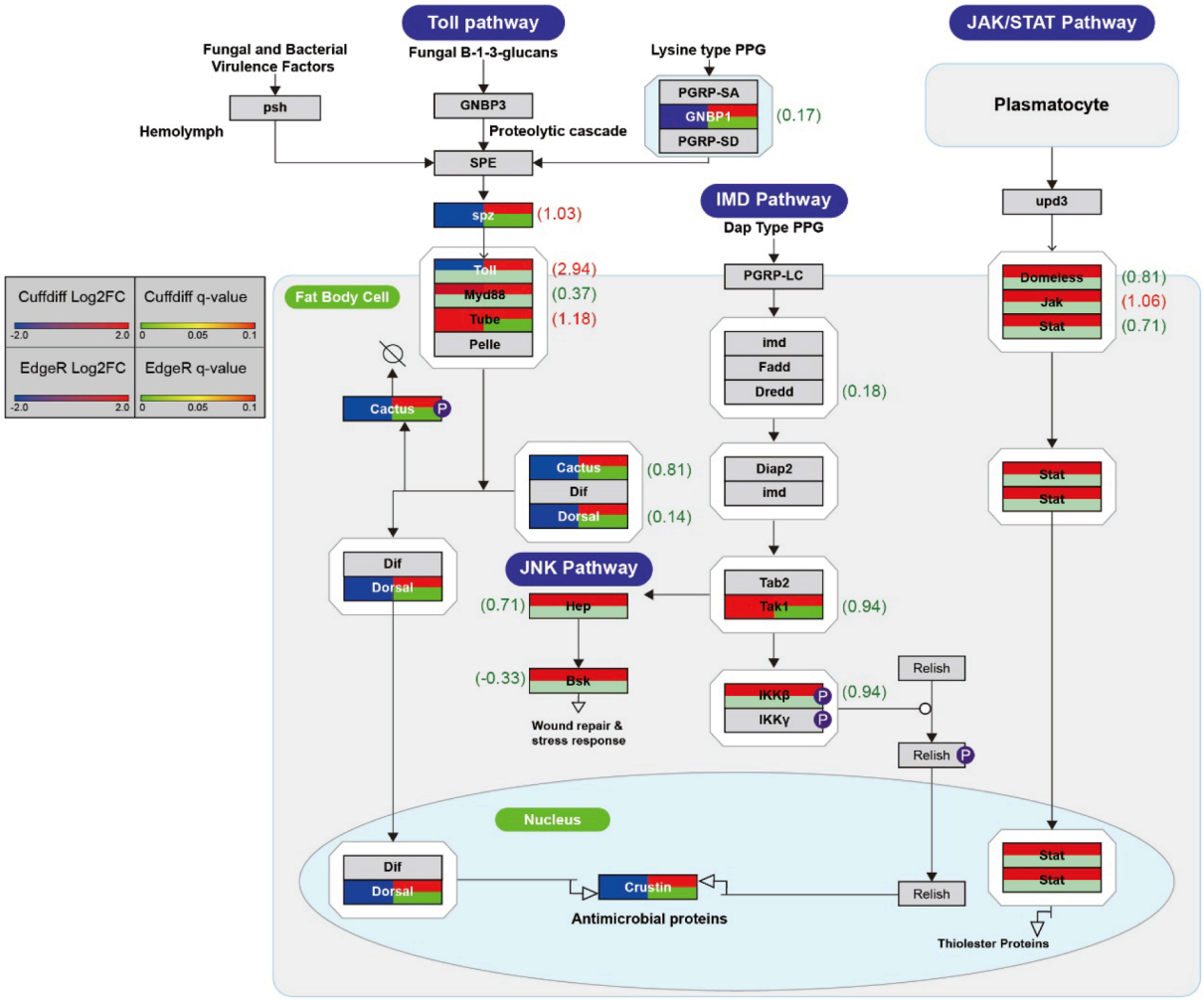

Fig. 1. Humoral immunity pathway-associated DEG analysis between sexes. Schematic depicts the predicted humoral innate immunity-related pathway of *L. rotunda* with DEG analysis results. The colored squares indicate differential gene expression levels between sexes in the $\log_2$FC scale and q-value. The expression levels that were omitted in EdgeR are depicted in light green. The pathway was based on the *D. melanogaster* humoral innate immunity-related pathway (Igboin et al., 2012). Parentheses indicate *D. melanogaster* differential expression levels between males and females ($\log_2$FC scale) (Gnad and Parsch, 2006).

default parameters except for jellyfish hash size (2.0×10^{10}). mRNA sequence trimming and *de novo* assembly of the *L. rotunda* transcriptome were carried out using four different strategies, described below as A, B, C, and D in order to select the most optional assembly: (A) Adapter removal, quality trimming, and length trimming were performed with Trim Galore! (ver. 0.6.10) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore) using the following parameters: `-quality 30 -length 80 -max_n 0` (Jeon et al., 2023; Jung et al., 2020). The trimmed reads were *de novo* assembled in Trinity (ver. 2.12.0) (Grabherr et al., 2011). (B) Trimming was performed using identical parameters as in (A) and *de novo* assembled using CLC Genomics Workbench (ver. 22.1) with default parameters. (C) Trimmed using a 0.05 thinning rate and *de novo* assembled in CLC Genomics Workbench. (D) Trimmed with 0.005 thinning rates and *de novo* assembled by CLC Genomics Workbench.

For trimming, the total number of contigs that were trimmed by CLC remained high. Average read lengths after trimming differed from each other, since CLC only partially removes low-quality sequences throughout the read, whereas Trim Galore! scans for low-quality sequences that fall below a given parameter and deletes the entire read containing such a sequence. Average read lengths after trimming were also shorter when a lower thinning rate was applied. While the CLC assembly contained larger numbers of contigs, the number of contigs >1,000 bp was 7 times larger in the Trinity assembly. Detailed analysis of the contig distribution of each assembly also showed that CLC contigs largely consisted of short-length contigs (Supplementary Fig. 1 and Supplementary Table 1). The additional open reading frame (ORF) investigation by TransDecoder (ver. 5.5.0) also indicated that each assembly had (A) 52.11%, (B) 4.50%, (C) 4.29%, and (D) 4.67% ORF-containing contigs (Supplementary Fig. 2). Therefore, we concluded that strategy A was the optimal method with which to prepare our dataset.

We clustered the transcripts containing the longest ORFs and possessing 95% sequence identity using the CD-HIT EST software package (ver. 4.8.1) (Fu et al., 2012; Moreno-Santillán et al., 2019) to generate non-redundant representative transcripts. A total of 36,959 transcripts were used for identification and differentially expressed gene (DEG) analysis of *L. rotunda* innate immunity-related gene families.

2.3. Functional annotation of L. rotunda de novo genome assembly

We used the MAKER2 genome annotation pipeline (Holt and Yandell, 2011) to annotate putative protein-coding regions of the *L. rotunda* genome assembly. Within the first round of MAKER (ver. 3.01.03), the *de novo* repeat library was built by RepeatMasker (Flynn et al., 2020). We used the *de novo* transcriptome of *L. rotunda* and the UniProt-derived representative proteomes of *Caenorhabditis elegans* (UP000001940), *Drosophila melanogaster* (UP000000803), *Apis mellifera* (UP000005203), *Tribolium castaneum* (UP000007266), and *Daphnia pulex* (UP000000305) as evidence (Supplementary Table 2). Two trained *ab initio* gene predictors (AUGUSTUS (ver. 3.4.0) (Stanke and Morgenstern, 2005) and SNAP (ver. 2006–07-28) (Korf, 2004)) were used in the first round, after which the second iterative run of MAKER was performed. Since sequences obtained from the second MAKER iteration were longer and contained significantly more complete ORF regions than sequences obtained after the first MAKER run, we mainly used second-run MAKER sequences to identify homologs of innate immunity-related gene families.

2.4. Proteome data and query sets

Having obtained *de novo* transcriptome assembly of *L. rotunda*, we retrieved four additional proteome sets for calanoid copepod *Eurytemora affinis* (GCF_000591075.1, NCBI RefSeq), siphonostomatoid copepod *Lepeophtheirus salmonis* (GCF_016086655.3, NCBI RefSeq), and harpacticoid copepod *Tigriopus californicus* (GHUE01000001-GHUE01092866; (Graham and Barreto, 2019)). The non-copepod crustacea *D. pulex* (GCF_021134715.1, NCBI RefSeq) was also retrieved to compare with five copepods. The set of key immunity genes for arthropods was selected from previously studied papers (Lai and Aboobaker, 2017; Palmer and Jiggins, 2015) and compiled from UniProt and Flybase to be used as query sequences for homology searches (Supplementary File 1).

2.5. Identification of homologs

To identify sequence homologs of copepod innate immunity-related gene families, we performed `blastp` on each copepod proteome using query sets with default parameters. The matched proteins were then filtered by the presence/absence of conserved domains known to be essential to function by NCBI Batch CD-Search (Marchler-Bauer and Bryant, 2004) (Supplementary Table 3). Additionally, a reciprocal BLAST search with the *D. melanogaster* translation database (r6.45, Flybase) was conducted for each identified gene. The candidates were assigned to gene families only if the top BLAST hit matched identified innate immunity homologs in *D. melanogaster*. The identified copepod innate immunity-related gene families were added to the query for homolog identification taken from the *L. rotunda de novo* genome and transcriptome (Supplementary Table 4).

For the *L. rotunda* innate immunity gene families, we performed `blastp` using the MAKER-retrieved translated sequences and the query set above. For *L. rotunda* innate immunity transcript identification, the non-redundant representative transcripts were submitted to `blastp` searches using the same query set. The subsequent steps were also identical to those described for the identification of copepod innate immunity-related gene families.

To count gene numbers across arthropods, the multiple proteins that were translated from a single gene locus were counted as single genes based on NCBI gene information. We drew a distinction between the gene copy number and transcript numbers of *L. rotunda* innate immunity-related gene families for clarity.

2.6. Structure analysis of toll-like receptors

Domain structure searches for *D. pulex*, *E. affinis*, *L. rotunda*, *T. californicus,* and *L. salmonis* were performed using the SMART web resource (E-value ≤ 10) (Letunic et al., 2020). All TLRs that contained one cysteine motif (LRR_CT; SM000082 or LRR_NT; SM000013) belonged to the vertebrate type (V-type) TLRs. All TLRs that contained at least two cysteine motifs belonged to the protostome type (P-type) TLRs. Leucine-rich repeat (LRR) counts included LRR_TYP (SM000369), LRR (SM000370), and Pfam LRR_8 domains. In the case of *D. melanogaster*, we referred to Imler and Hoffmann (2001).

2.7. Copepod AMP identification

For crustin-type AMP identification, we performed `blastp` using 125 crustin-like proteins as a query set from the non-redundant representative *L. rotunda* transcripts (Uniprot; Supplementary File 1). All matched transcripts were filtered by their ORF-containing and domain presence (WAP; pfam00095). The remaining candidate transcripts were then evaluated to determine the likelihood that they represented *bona fide* AMPs with AMPscanner (Veltri et al., 2018) and AMPdiscover on ProtDCal-AMP_RF model manner (Pinacho-Castellanos et al., 2021). The method for identification of anti-lipopolysaccharide factors (ALF)-type AMPs was the same as that employed for crustin-type AMP identification with the exception that the domain presence criterion (DUF3254; pfam11630) differed for ALF-type AMP identification.

2.8. Measuring gene expression levels between sexes

The previously trimmed reads obtained with Trim Galore! were mapped to the non-redundant representative *L. rotunda* transcripts by

bowtie2 (ver. 2.3.5) (Langmead and Salzberg, 2012) with high sensitivity for each replicate (two male replicates and two female replicates). The samtools flagstat results showed that each replicate was mapped and properly paired to the *L. rotunda* transcriptome with rates of 88.52%, 91.82%, and 91.79%, respectively. Since all mapping rates exceeded 85%, we determined that the mapping was appropriate (Conesa et al., 2016). All the reads were then sorted by SAMtools (ver. 1.10) (Danecek et al., 2021). This yielded three sorted bam files that were then used to quantify differential gene expression between sexes in two different programs.

We first implemented the Cuffdiff pipeline, wherein Cufflinks (ver. 2.2.1) measured gene expression under two conditions, male and female. Information on each genomic feature derived from Cufflink was merged by Cuffmerge and compared in Cuffdiff (ver. 2.2.1) (Trapnell et al., 2013). These analyses yielded a gene expression value of 32,019 transcripts, and all FPKM were subsequently adjusted to add 0.01 to both male and female expressions. The $\log_2$(fold change, FC) was also recalculated according to the adjusted FPKM values. We defined significance threshold criteria for candidate transcripts with $|\log_2 FC| > 2$ and q-value < 0.05 for differential expression between sexes.

Independently, we conducted DEG analysis using the EdgeR software package (ver. 3.32.1) (Robinson et al., 2010). Sorted bam files described previously were converted to count values by featureCounts (ver. 2.0.2) (Liao et al., 2014) with the Cuffmerge-assembled GTF. Any transcripts that did not exhibit overlapping expressions from all three replicates were omitted. The biological coefficient of variation (BCV) was set to 0.2, for no replicate of the female library. We used the TMM method to normalize library size and the Benjamini-Hochberg (BH) method to adjust P-values. The same criteria for significance as described above were applied to the resulting 11,243 transcripts.

2.9. Multiple sequence alignment and phylogenetic analysis

All multiple sequence alignments from the present study were compared using local alignment in MAFFT (ver. 7.505) (Katoh et al., 2002) with a maximum iterate value of 1,000. The best model for the phylogenetic tree was selected using IQ-TREE (ver. 2.2.0) (Kalyaanamoorthy et al., 2017). For each model, the tree was constructed by RAxML-NG (ver. 1.1.0) (Kozlov et al., 2019). The bootstrapped confidence interval was based on 1,000 replications.

3. Results and discussion

3.1. Establishment of copepod dataset from three copepods and a crustacean proteome

To date, a comprehensive study of the innate immune system across copepods has not been conducted. To ensure the determination of *Labidorera rotunda* innate immunity-related gene families, we first established a copepod dataset (Supplementary Table 4) that included known PRR genes and four gene families representing innate immunity-associated signaling pathways (Toll, Imd, JAK/STAT, and JNK). Gene families were defined according to their domain structures and reciprocal BLAST search results. Although the gene families in our final dataset have been studied in other arthropods, such as chelicerates, myriapods, and crustaceans (Lai and Aboobaker, 2017; Palmer and Jiggins, 2015), this is the first time they have been explored across various copepod species.

For the copepod dataset, we retrieved the protein sequences of *Eurytemora affinis*, *Tigriopus californicus*, and *Lepeophtheirus salmonis*, since they were openly available and have been relatively well-studied among the copepods. The protein sequences of the non-copepod crustacean *Daphnia pulex* were also included to be compared with copepod sequences (see Materials and Methods). The resulting innate immunity-related gene family dataset from copepod is described in Supplementary Table 4.

3.2. Identification of innate immunity-related gene families from *L. rotunda* de novo genome and transcriptome

From the *L. rotunda* predicted protein dataset retrieved using MAKER, we identified innate immunity-related gene families according to the same criteria we applied to the prior copepod dataset. For this round of analysis, we also added the copepod dataset organized as described above to the initial blastp query. After examining and comparing domain composition and reciprocal blastp against the *Drosophila melanogaster* protein database, we uncovered several gene families related to the innate immunity of *L. rotunda* (Table 1).

Having obtained a validated transcript dataset for *L. rotunda*, we identified their coding regions on the longest ORF and used TransDecoder to examine their domain composition. From 94,341 candidate ORFs, the sequences that contained 95% identical amino acids were clustered using the CD-HIT EST package (Fu et al., 2012; Moreno-Santillán et al., 2019). The resulting final transcript dataset was made up of 36,959 amino acid-translated transcripts. The remaining steps for the identification of innate immunity-related gene families from the transcript dataset were the same as above. The reciprocal blastp based on the *D. melanogaster* protein database revealed that most copepod transcripts were ideally matched with known innate immune system-related gene families of *D. melanogaster* except Tube and Pelle, the evolution of which are thought to have originated from the same common genes (Towb et al., 2009).

To clearly define the properties of the copepod Tube and Pelle sequences, we aligned Tube and Pelle homologs previously identified from among several species (Supplementary Fig. 3). Multiple sequence alignments revealed that the Death domain of Tube contained conserved residues that are required to bind MyD88 (Arginine, Lysine, Arginine; R, K, R) and shared one core Pelle-binding residue (Alanine; A). The kinase domain of Tube also contained a signature RD residue, the definitive marker of kinases regulated by activation loop phosphorylation (Supplementary Fig. 4). Since only Tube-like kinases, but not Pelle or Pelle-like proteins, are RD kinases (Towb et al., 2009), Tubes and Pelles could be accurately distinguished and classified.

Gene families identified from the transcript dataset are summarized in Supplementary Table 5. Comprehensive gene family classifications assembled from other reported arthropod genomes and transcriptome data were detailed in Table 2 and in the following paragraphs.

3.3. Copepod PRR family properties

We identified four PRR families (PGRPs, GNBPs, CTLs, and TEPs) in copepods. We found that PRR families were highly reduced especially in Siphonostomatoida (*L. salmonis*) compared to other copepods. It also appeared that the presence of PGRP-like proteins was ambiguous across Crustacea.

While they serve diverse functions and act in multiple pathways, PGRPs were not found in *D. pulex*, *L. rotunda*, and *L. salmonis* in the present study. As far as we know, PGRPs have been identified in mollusks, echinoderms, and vertebrates, but not in plants or some metazoans including nematodes such as *C. elegans* (Dziarski and Gupta, 2006). McTaggart et al. (2009) reasoned that the GNBPs of *D. pulex* may be expanded to compensate for their lack of PGRPs, permitting an alternative mechanism for the recognition of Gram-positive bacteria. However, we did not observe a notable expansion of GNBPs in copepods. The gene copy numbers of CTLs and TEPs were also similar to those of other arthropods, which did not support the GNBP compensation theory as an explanation for our findings in copepods.

With respect to GNBPs, the catalytic activity of the GH16 domain persisted in Crustacea, while those of *Drosophila* were inactivated, based on E188 and E193 criteria (Buchon et al., 2009; Zhang et al., 2003). In the two GH16-bearing copepods we identified, the catalytic domains in *E. affinis* and *L. rotunda* were still present. For its potential persistence of catalytic activity, we speculated that it may function in digestion or

Table 1
Maker2-retrieved innate immunity-related gene family of *Labidocera rotunda*.

Gene family	Protein id	Top hit against *Drosophila* (gene name and FlyBase ID)
GNBP	LrotGNBP	GNBP1-PA, CG6895-PA
CTLs	LrotCTL1	tfc-PC, CG9134-PC
	LrotCTL2	tfc-PB, CG9134-PB
	LrotCTL3	CG15818-PA, CG15818-PA
	LrotCTL4	lectin-21Cb-PB, CG13686-PB
	LrotCTL5	lectin-24Db-PA, CG2958-PA
	LrotCTL6	CG42339-PB, CG15204-PA
	LrotCTL7	lectin-24A-PA, CG3410-PA
	LrotCTL8	lectin-24Db-PA, CG2958-PA
	LrotCTL9	tfc-PC, CG9134-PC
	LrotCTL10	tfc-PC, CG9134-PC
	LrotCTL11	lectin-22C-PB, CG42295-PB
	LrotCTL12	Lectin-galC1-PB, CG9976-PB
	LrotCTL13	tfc-PC, CG9134-PC
	LrotCTL14	capu-PF, CG3399-PF
	LrotCTL15	Sema5c-PA, CG5661-PA
	LrotCTL16	lectin-24Db-PA, CG2958-PA
	LrotCTL17	lectin-24Db-PA, CG2958-PA
	LrotCTL18	lectin-21Ca-PA, CG2826-PA
	LrotCTL19	CG43797-PA, CG43797-PA
	LrotCTL20	lectin-24Db-PA, CG2958-PA
	LrotCTL21	bark-PA, CG3921-PA
	LrotCTL22	CG9095-PB, CG9095-PB
	LrotCTL23	lectin-24Db-PA, CG2958-PA
	LrotCTL24	lectin-21Ca-PA, CG2826-PA
	LrotCTL25	tfc-PC, CG9134-PC
	LrotCTL26	CG9095-PD, CG9095-PD
	LrotCTL27	CG7518-PG, CG7518-PG
	LrotCTL28	trol-PBE, CG33950-PBE
	LrotCTL29	tfc-PC, CG9134-PC
	LrotCTL30	fat-spondin-PA, CG6953-PA
	LrotCTL31	CG15818-PA, CG15818-PA
	LrotCTL32	Gal-PA, CG9092-PA
	LrotCTL33	lectin-33A-PB, CG16834-PB
	LrotCTL34	CG7763-PD, CG7763-PD
	LrotCTL35	tfc-PC, CG9134-PC
	LrotCTL36	tfc-PC, CG9134-PC
	LrotCTL37	lectin-24Db-PA, CG2958-PA
	LrotCTL38	lectin-24Db-PA, CG2958-PA
	LrotCTL39	lectin-24Db-PA, CG2958-PA
	LrotCTL40	CG9095-PC, CG9095-PC
	LrotCTL41	lectin-24Db-PA, CG2958-PA
	LrotCTL42	Cont-PA, CG1084-PA
	LrotCTL43	lectin-24Db-PA, CG2958-PA
	LrotCTL44	lectin-46Cb-PB, CG1652-PB
	LrotCTL45	tfc-PB, CG9134-PB
	LrotCTL46	uif-PC, CG9138-PC
	LrotCTL47	tfc-PC, CG9134-PC
	LrotCTL48	CG6055-PB, CG6055-PB
	LrotCTL49	CG11211-PA, CG11211-PA
	LrotCTL50	CG7763-PD, CG7763-PD
	LrotCTL51	slf-PB, CG3244-PB
	LrotCTL52	CG7763-PD, CG7763-PD
	LrotCTL53	CG2926-PA, CG2926-PA
	LrotCTL54	tfc-PB, CG9134-PB
	LrotCTL55	lectin-24Db-PA, CG2958-PA
	LrotCTL56	CG42339-PB, CG15204-PA
	LrotCTL57	lectin-24A-PA, CG3410-PA
	LrotCTL58	lectin-24Db-PA, CG2958-PA
	LrotCTL59	CG11211-PA, CG11211-PA
TLRs	LrotTLR1	Tollo-PA, CG6890-PA
	LrotTLR2	Toll-6-PC, CG7250-PC
SPZs	LrotSPZ1	spz-PI, CG6134-PI
	LrotSPZ2	NT1-PH, CG42576-PH
	LrotSPZ3	NT1-PH, CG42576-PH
	LrotSPZ4	NT1-PH, CG42576-PH
	LrotSPZ5	spz-PB, CG6134-PB
	LrotSPZ6	spz4-PB, CG14928-PB
	LrotSPZ7	NT1-PD, CG42576-PD
	LrotSPZ8	NT1-PH, CG42576-PH
	LrotSPZ9	NT1-PH, CG42576-PH
Dorsal	LrotDorsal	dl-PC, CG6667-PC

pathogen killing (Fig. 2).

GNBPs were sparsely represented across copepods. Only two copepods contained the GNBP gene family (*E. affinis* and *L. rotunda*) and the copy numbers were lower than in other arthropods (except the chelicerates, in which it is absent entirely). Their catalytic activity was preserved, similar to the activity observed across 10 *D. pulex* GNBPs and two *S. martina* GNBPs (Palmer and Jiggins, 2015). Still, the distinct difference between copy numbers in *D. pulex* (11 GNBPs) and copepods (1 to 3 GNBPs) may indicate the dynamic loss and gain of gene expression during their speciation. Meanwhile, the reciprocal BLASTp matched copepod GNBPs to GNBP1 and GNBP3. As *E. affinis* and *L. rotunda* lack the noncatalytic PGRPs or a whole PGRP family but possess Toll pathway-related genes, their GNBPs may induce the Toll pathway on their own or with another as yet unknown partner.

The CTLs were moderately conserved across arthropods. Retrieving 252C-type lectin-like domains (CTLDs) from a total 142 copepod CTLs (nine genes that do not contain CTLDs based on SMART web resource were excluded), we identified their EPN motifs to learn more about the binding affinity of copepod CTLDs (Runsaeng et al., 2017) and four conserved cysteine residues to compare them with reported characteristics of vertebrate CTLDs (Zhu et al., 2020). We found 40 CTLDs that contained EPN motifs and possessed possible specific binding affinity to mannose, similar to the known crustacean immune-related CTL FmLC3 (Runsaeng et al., 2017). This analysis also uncovered 104 CTLDs that lack their cysteine residues partially or absolutely, which may affect their ability to form disulfide bonds and consequently their stability in general. The arranged multiple sequence alignment of copepod EPN motifs and cysteine residue status is attached in Supplementary File 2 and also presented in Supplementary Table 6.

The CTL gene family was well conserved across copepods and identifiable by its diverse domain composition. As they possess an EPN motif resembling reported crustacean immune-related CTLs, the 40 copepod CTLDs identified here may provide deep insight into the principle of invader recognition in the arthropod innate immune system.

The TEP gene families were highly conserved across Arthropoda. Fig. 3 shows the phylogenetic analysis using TEP genes (19 TEP genes identified in this study), which include two *L. rotunda* transcripts, and 139 previously reported arthropod TEP gene sequences (Lai and Aboobaker, 2017; Palmer and Jiggins, 2015). The copepod TEP genes scattered to the alpha-2 macroglobulin (A2M) subgroup, macroglobulin complement-related (MCR) subgroup, and *Drosophila* TEP 1–4-like subgroup. TEP genes lacking a conserved GCGEQ thioester motif were assigned to the MCR subgroup, consistent with patterns observed among other arthropod gene families (Supplementary File 3). Some copepod TEP sequences have a unique pattern at the GCGEQ motif site, which rarely appears within other arthropod TEPs (Supplementary Fig. 5).

In our study, the copepod TEP phylogeny revealed the lost-progressive pattern of copepod TEPs with their minor dismiss status on each subfamily and four noncanonical copepod-specific TEPs. According to Palmer and Jiggins (2015), the TEP gene families of arthropods can be classified into four subfamilies: the vertebrate A2M subfamily, the C3-C5-like complement factors subfamily, the MCR subfamily, and the *Drosophila* TEP 1-4-like subfamily. The phylogenetic tree including Palmer and Jiggins's sequences suggested that all arthropods except *L. rotunda* and *L. salmonis* contained at least one copy of an MCR subfamily-type TEP, which indicated that the arthropod-common type TEP was lost from those two species (Palmer and Jiggins, 2015). The distinct indels of *T. californicus* and *E. affinis* TEPs also supported their progressive evolutionary change. Of 158 TEPs selected from diverse arthropods and mammalians, only copepod TEPs had a 'PxG' residue insertion on the upstrand of the GCGEQ thioester motif (Supplementary File 3). On the other hand, we found vertebrate A2M subfamily-claded and *Drosophila* TEP 1-4 like subfamily-claded TEPs in all copepods. No copepod TEPs belonged to the C3-C5-like complement factors subfamily, which implies their distant evolutionary relationship with vertebrate C3-C5 type TEPs.

Table 2
Comprehensive gene families from genomes and transcriptomes determined in this study and 17 other previously sequenced arthropods.

Pathway	Genefamilies	*Labidocerarotunda*	*Eurytemoraaffinis*	*Tigriopuscalifornicus*	*Lepeoptheirussalmonis*	*Daphniapulex*	*Parhyalehawaiensis*[a]	*Procellusmeridianus*[a]	*Litopenaeusvannamei*[a]	*Eriocheirsinensis*[a]	*Neomysisawatschensis*[a]	*Euphausiasuperba*[a]	*Metaseiulusoccidentalis*[b]	*Tetranychusurticae*[b]	*Ixodesscapularis*[b]	*Mesobuthusmartensii*[b]	*Parasteatodatepidariorum*[b]	*Drosophilamelanogaster*[b]	*Anophelesgambiae*[c]	*Aedesaegypti*[c]	*Triboliumcastaneum*[c]	*Apismellifera*[c]	*Strigamiamaritima*[b]
Recognition	PGRP	0	1	4	0	0	NA	NA	NA	NA	NA	NA	1	1	4	0	11	13	7	8	6	4	20
	GNBP	1 (3)	3	0	0	11	3	5	5	7	3	4	0	0	0	0	0	3	7	7	3	2	1
	CTL	59 (121)	43	32	17	34	48	14	65	29	17	33	NA	NA	NA	NA	NA	40	25	39	16	10	NA
	TEP	0 (2)	8	4	2	3	4	8	17	12	7	8	4	3	3	1	8	6	13	8	4	4	4
Toll	TLR	2 (2)	10	11	7	7	5	3	5	4	0	0	5	2	4	14	18	9	10	12	9	5	27
	SPZ	9 (13)	18	19	11	49	7	5	3	6	7	26	1	5	2	5	7	6	6	9	7	2	2
	Myd88	0 (1)	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
	Tube	0 (1)	1	1	1	1	0	0	1	1	0	0	0	0	0	1	0	1	1	1	1	1	0
	Pelle	0	1	1	1	1	2	1	2	1	0	1	1	2	2	0	2	1	1	1	1	1	1
	Dorsal	1 (1)	2	2	3	1	1	1	1	1	0	1	1	1	1	1	3	1	1	2	0	2	1
	Cactus	0 (1)	1	1	1	2	1	1	1	1	1	0	1	1	1	2	1	1	1	1	1	3	1
	Dif	0	0	0	0	0	NA	NA	NA	NA	NA	NA	0	0	0	0	0	1	NA	NA	NA	NA	0
Imd	Imd	0	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
	Tak1	0 (1)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3
	Ikkg	0	0	0	1	1	NA	NA	NA	NA	NA	NA	0	0	1	0	1	1	NA	NA	1	1	1
	Ikkb	0 (1)	1	0	1	1	1	1	1	1	1	0	2	0	1	1	1	1	NA	NA	2	1	1
	Fadd	0	0	0	0	0	NA	NA	NA	NA	NA	NA	0	0	0	3	2	1	1	1	1	1	1
	Dredd	0	1	0	0	5	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	0
	Relish	0	1	0	0	2	1	1	1	1	1	1	0	1	1	1	3	1	2	3	2	2	1
JAK/STAT	Dome	0 (1)	2	1	2	1	1	1	1	1	1	0	1	1	2	0	2	1	NA	NA	1	1	2
	Hop	0 (1)	1	1	1	1	1	1	1	1	0	0	0	1	1	0	0	1	NA	NA	1	1	1
	Stat92	0 (1)	2	2	2	1	2	1	1	1	0	1	5	2	1	2	2	1	2	1	1	1	1
JNK	Hep	1 (1)	1	1	1	1	NA	NA	NA	NA	NA	NA	1	1	1	1	1	1	NA	NA	1	NA	1
	Bsk	0 (2)	1	1	1	2	NA	NA	NA	NA	NA	NA	3	1	1	2	2	1	NA	NA	3	NA	1

Gene copy numbers were analyzed across 22 arthropods. This study determined the gene copy numbers for copepods and *D. pulex*. The transcript numbers of *L. rotunda* are shown in parenthesized. Proteins encoded by a single locus were counted as a single gene based on NCBI gene and protein information. Species are color-coded as follows: red for copepods and a branchiopod (*D. pulex*), blue for malacostracans, brown for chelicerates, green for insects, and purple for myriapods. Gene copy numbers for other arthropods were obtained from the following references; [a]Lai and Aboobaker (2017), [b]Palmer and Jiggins (2015), [c]McTaggart *et al.* (2009). The colored table can be found online in Appendix A Supplementary Material.

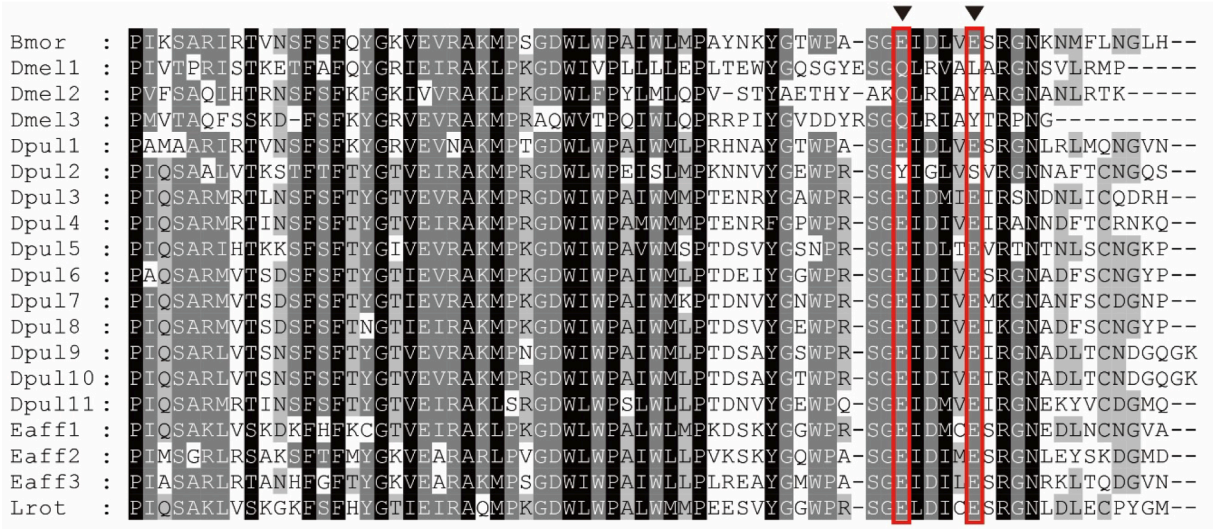

Fig. 2. Comparative analysis of the catalytic activity of copepod and *Drosophila* GNBPs. Multiple sequence alignment was performed using the L-INS-I method on 19 GH16 domains from GNBPs of *Bombyx mori* (Bmor), *D. melanogaster* (Dmel1-Dmel3), *D. pulex* (Dpul1-Dpul11), *E. affinis* (Eaff1-Eaff3), and *L. rotunda* (Lrot). The two conserved E residues indicated the GNBPs of copepods and most of *D. pulex* possessed catalytic activity, unlike *D. melanogaster*. The sequences comprising the multiple sequence alignment are noted in Supplementary Table 4 except for *B. mori* (NP_001159614.1) and *D. melanogaster* (FBpp0074817, FBpp0074861, and FBpp0076237).

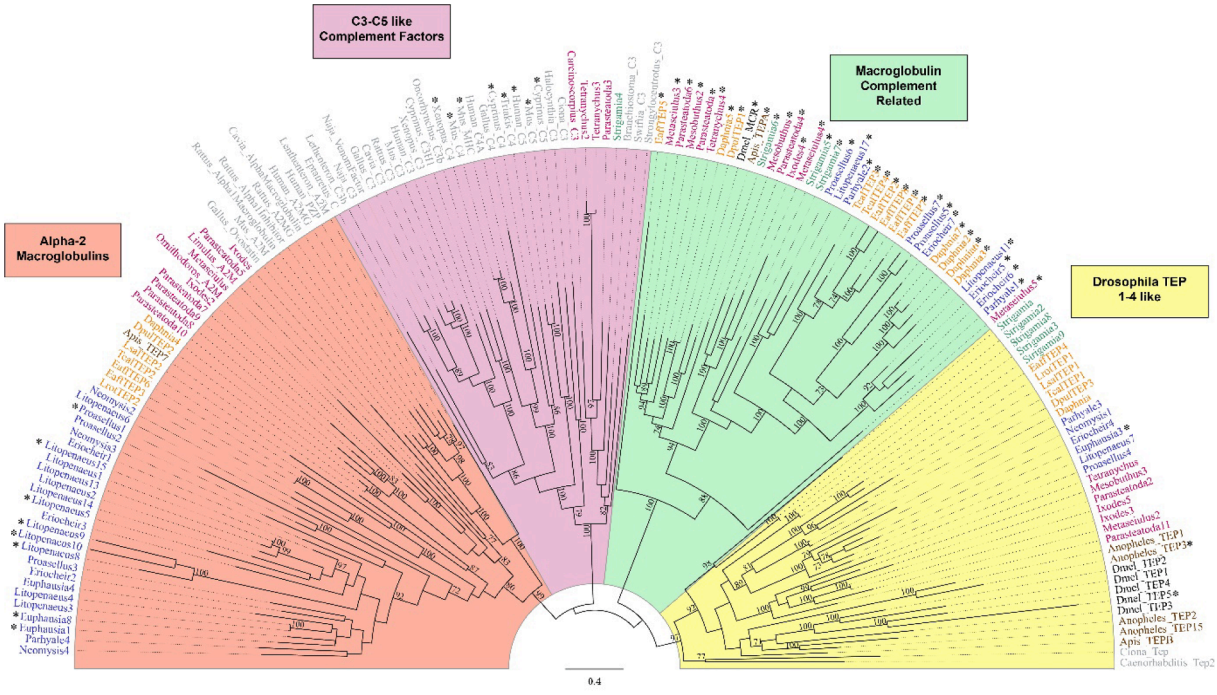

Fig. 3. The arthropod TEP gene trees diverged into 4 subfamilies. A total of 158 TEPs were used to construct a midpoint tree (WAG + F + I + G4 model). The genes without a thioester motif were denoted with an "*". Sequences included non-arthropod TEPs (mainly vertebrates; gray), chelicerates (purple), malacostraca (blue), insects (brown; except for *D. melanogaster*), and copepods (including *D. pulex*; yellow). All arthropods except for *L. rotunda* and *L. salmonis* contained at least one copy of an MCR subfamily-type TEP. The non-arthropod, chelicerate sequences were referenced from previous studies (Palmer and Jiggins, 2015). The malacostraca sequences were retrieved from a previous study (Lai and Aboobaker, 2017) and filtered by domain composition (only A2M_comp-containing (pfam07678) or A2M_2 (cd02897) domains remained). The sequences of *D. pulex* were derived from both the present study and a previous study (Palmer and Jiggins, 2015). Bootstrap values ≥ 70 were noted. (Palmer and Jiggins, 2015).

3.4. Structural variation of copepod toll-like receptors

TLRs share several conserved structures: LRRs that mediate PAMP recognition, a transmembrane domain, and a cytoplasmic Toll/interleukin-1 receptor (TIR) domain that initiates downstream signaling (Kawasaki and Kawai, 2014). Multiple studies have established that TLRs can be classified into different types based on the number of cysteine motifs around their LRRs. The V-type TLRs contain a single cysteine cluster at the end of the LRRs adjacent to the cell membrane, whereas the P-type TLRs have multiple cysteine motifs (Imler and Zheng, 2004; Leulier and Lemaitre, 2008). Table 3 outlines the types of TLRs and numbers of LRRs uncovered in our analyses (SMART E-value <

Table 3
TLR classification and LRR counts in *D. melanogaster*, *D. pulex*, and four copepods.

Toll-like receptor	V-type	P-type	NA	LRRs
Drosophila melanogaster	1	8	0	13–28
Daphnia pulex	2	5	0	9–26
Eurytemora affinis	0	8	2	6–21
Labidocera rotunda	0	2	0	15, 22
Tigriopus californicus	0	9	2	6–27
Lepeophtheirus salmonis	0	6	1	2–22

The domain composition and LRR numbers of the *D. melanogaster* genome were derived from a previous study (Imler and Hoffmann, 2001). TLRs that lack any cysteine motif were counted as NA (not applicable).

10), classified according to these criteria. The numbers of LRRs varied widely across species, from two to 28. The domain search revealed that all cysteine cluster-containing copepod TLRs were comprised of multiple cysteine clusters, consistent with P-type TLRs (Table 3).

As a complementary analysis to confirm the type of these receptors, we performed an additional tree search to characterize copepod TLRs based on their TIR domain sequences. A total of 149 TIR domains were sampled from diverse arthropods including copepods, a branchiopod (*D. pulex* from our study and Palmer and Jiggins (2015) were both retrieved), malacostracans (Lai and Aboobaker, 2017), a myriapod, chelicerates, and mammalians (Palmer and Jiggins, 2015) were claded into the V-type TLR-like subgroup and DmelToll 1-8 like subgroups as was demonstrated by the previous study, albeit with low support (Fig. 4) (Palmer and Jiggins, 2015). Our tree search also revealed that no clear homologs of *D. melanogaster* Toll-1 existed among copepod TLRs, which was similar to results obtained from some myriapod and chelicerate gene analyses (Palmer and Jiggins, 2015). Among the copepod TIR domains, four sequences (XP_023340581.1, TRY71807.1, TRY71808.1, and XP_040567498.1) were assigned to the V-type TLR-like subgroups. Surprisingly, those four sequences also belonged to the NA group in Table 3, as they lacked cysteine clusters. Subsequently, multiple sequence alignment with whole TLR proteins was performed, revealing minor variation (F/W/I > W) between the V-type TLRs and the P-type TLRS on the cysteine motif starting locus of the V-type TLRs (Supplementary Fig. 6 and Supplementary File 4). However, the detailed mechanisms whereby this sequential change or other variation occurred to affect the loss of the cysteine cluster and differentiation from V-type TLRs (or *vice versa*) remain to be elucidated.

3.5. Copepod AMP identification

AMPs are the end products and effectors of the innate immunity signaling pathway, leading to the death of the invading microorganisms. While 12 diverse AMPs have occasionally been reported in crustaceans, only crustins and ALFs are present in the majority of crustacean species (Matos and Rosa, 2022). Here, we have identified the first crustin-type *Labidocera* AMP sequences and, to the best of our knowledge, have carried out the first detailed investigation into their functional properties.

To identify candidate sequences, we first retrieved all `blastp` matched with UniProt-derived AMP-like proteins (125 crustin-like proteins for crustin identification and 18 ALF-like proteins for ALF identification) from the transcript dataset of *L. rotunda*, which was the same dataset with which we identified immunity-related gene families (Supplementary File 5). From among the 670 crustin-type and 44 ALF-type matched transcripts, we selected 282 crustin-type and 13 ALF-type transcripts that contained complete ORFs using TransDecoder and analyzed their domain composition by NCBI CD-Search (Marchler-Bauer and Bryant, 2004). For crustin identification, only the WAP domain (pfam00095)-containing transcripts remained, and DUF3254 (pfam11630)-containing transcripts remained for ALF identification. Only one sequence emerged as a putative copepod crustin. We analyzed its AMP properties with two distinct programs: AMPscanner (Veltri et al., 2018) and AMPdiscover (Pinacho-Castellanos et al., 2021). Between passes carried out by both programs, we identified this transcript as the first example of a crustin-type AMP in *Labidocera*, LrotCrustin. On

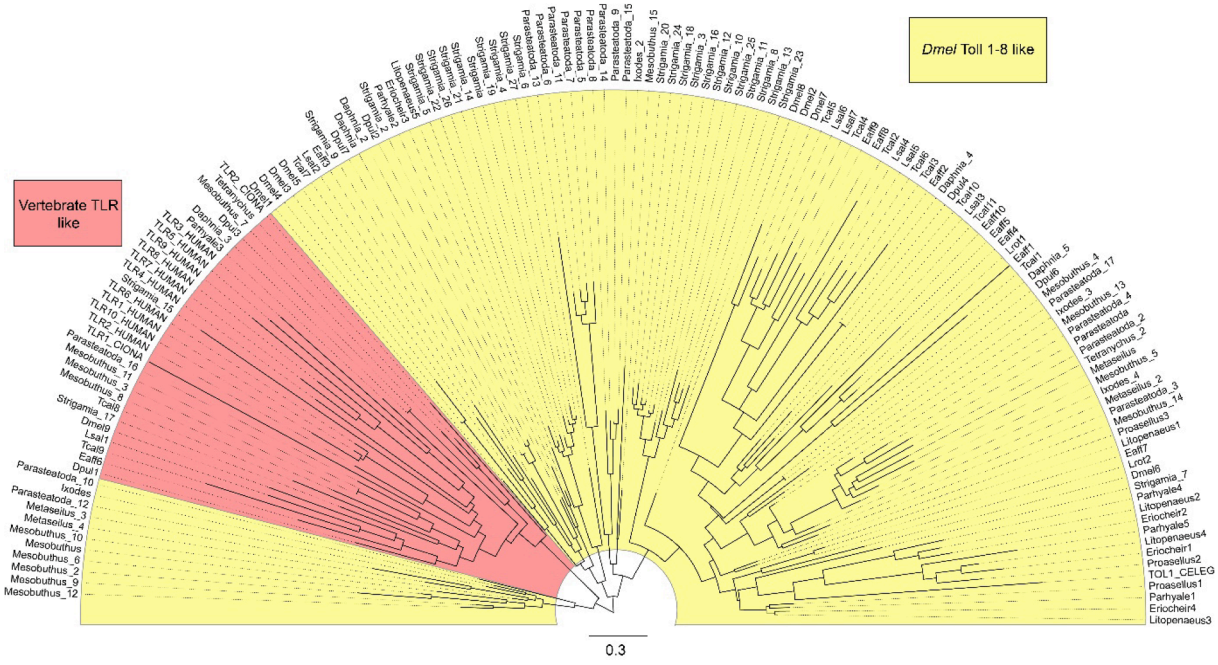

Fig. 4. Phylogenic tree built from Arthropod TIR domains. The maximum-likelihood tree (LG + G4) was built from a total of 149 TIR domains of Arthropod and Mammalian TLR sequences. There were no clear homologs of *D. melanogaster* Toll 1 among copepod TLRs. In addition, the tree revealed that four copepod TIR domain sequences belonged to the vertebrate TLR-like subfamily (Tcal8, Lsal1, Tcal9, and Eaff6). The TIR domains from copepods, malacostracans, and *D. pulex* were retrieved by SMART. Other TIR domain sequences were referenced from a previous study (Palmer and Jiggins, 2015). The *D. pulex* TIR domain sequences from (Palmer and Jiggins, 2015) were also used to construct the tree (indicated as *Daphnia* in Fig. 3.).

the other hand, for ALF identification, since there were no DUF3254 domain-containing sequences from candidate sequences, we could not isolate an ALF-type AMP from *L. rotunda*. Further research is needed to explore the similarities and differences between LrotCrustin and AMPs from other crustaceans.

3.6. Copepod Toll, Imd, JAK/STAT, and JNK pathway components

Our study revealed that the Toll pathway components across arthropods were highly conserved. Most species studied, including copepods, contained core proteins of the Toll pathway, such as TLR, SPZ, the Myd88-Tube-Pelle complex, and the Dorsal/Dif-Cactus complex, which was consistent with previous studies (Lai and Aboobaker, 2017; Palmer and Jiggins, 2015). We also observed notable expansion of copepod SPZ, compared to other classes. Since the functional role of TLR-SPZ includes modulating responses to the outside environment, the diversity of TLR-SPZ copy number variants across arthropods may reflect different selective pressures from different life environments (Lima et al., 2021).

The components of the Myd88-Tube-Pelle complex across arthropods were moderately diversified, except for insects. The loss event for Tube or Pelle was common for Chelicerates and Malacostraca; some species had lost both. It also appeared that *L. rotunda* from copepods lost its Pelle, which was the only instance of such an event among copepods. We previously discussed the divergence of Tube and Pelle from a common origin (Towb et al., 2009). Our results here indicated that the time point at which the two genes diverged was earlier than the divergence of the Arthropod common ancestor.

Compared to the Toll pathway-related gene families, the Imd pathway components across arthropods examined in our study were reduced in copepoda. All gene families except TAK1 showed dynamic presence in the clade, even for Fadd, a ubiquitous protein in the animal kingdom. These tendencies toward reduction were previously shown in Chelicerate, which was explained by secondary loss occurring in the subphylum (Palmer and Jiggins, 2015). With relatively conserved Imd pathway components in a myriapod (*S. maritina*) and insects, the loss event in copepods seemed to arise from a common ancestor of Crustacea, although further work is needed to establish this.

For the copepod JAK/STAT and JNK pathways, all copepods showed that they have at least one gene copy (transcript in *L. rotunda*) for each component, which suggests that the two pathways are highly conserved across copepods. The high conservation of JNK pathway components in copepods indicates that the JNK pathway originated from the common ancestor of arthropods.

3.7. DEG analysis of sex-biased L. rotunda innate immunity gene families

We performed DEG analyses with two different programs between two sexes to understand the sexual dimorphism of the *L. rotunda* innate immune system. According to Cuffdiff DEG analysis, 8.5% of the 32,019 transcripts analyzed here showed sex-biased expression patterns ($|\text{adjusted } \log_2\text{FC}| \geq 2$, q-value < 0.05). Among them, 69.0% of the transcripts were male-biased and 31.0% of them were female-biased. EdgeR DEG analysis suggested that 11.0% of total transcripts showed

Table 4
Innate immunity-related transcripts that were differentially expressed between two sexes.

Transcript name	Transcript ID	Adjusted $\log_2$FC (Cuffdiff)	q-value (Cuffdiff)	$\log_2$FC (EdgeR)	q-value (EdgeR)
LrotCTL4	TRINITY_DN6301_c0_g1_i12.p1	−2.131776003	0.361731	−2.080659674	5.63E-07
LrotCTL9	TRINITY_DN17713_c0_g1_i1.p1	−2.677221067	0.383464	−2.709124164	1.52E-09
LrotCTL14	TRINITY_DN701_c0_g1_i5.p1	−2.445538108	0.36139	−2.348091606	3.68E-08
LrotCTL16	TRINITY_DN701_c0_g1_i9.p1	−2.594529149	0.383464	−2.55597647	1.8E-09
LrotCTL19	TRINITY_DN5272_c0_g1_i16.p1	−3.852541429	0.420454	−3.391499918	4.53E-14
LrotCTL23	TRINITY_DN1501_c0_g1_i2.p1	−3.594376954	0.298651	−3.609408637	1.52E-15
LrotCTL25	TRINITY_DN9_c0_g1_i3.p1	−2.570929948	0.449285	−2.481164843	4.59E-09
LrotCTL26	TRINITY_DN9_c0_g1_i8.p1	−2.913537914	0.430653	−2.81919119	8.3E-11
LrotCTL27	TRINITY_DN11030_c0_g1_i1.p1	−8.240094855	0.039063	NA	NA
LrotCTL32	TRINITY_DN5321_c0_g2_i12.p1	−4.693514938	0.111815	−4.568649512	5.36E-21
LrotCTL33	TRINITY_DN5321_c0_g2_i14.p1	−4.683938791	0.097943	−4.504512173	1.84E-20
LrotCTL43	TRINITY_DN1046_c0_g1_i1.p1	−3.304532927	0.276942	−3.349204115	9.35E-14
LrotCTL44	TRINITY_DN1046_c0_g1_i3.p1	−3.189560844	0.248159	−3.227125361	4.27E-13
LrotCTL57	TRINITY_DN158_c0_g1_i9.p1	−4.86249057	0.405608	−4.689830603	3.63E-22
LrotCTL61	TRINITY_DN122_c0_g1_i10.p1	−2.68406659	0.340334	−2.730557194	3.79E-10
LrotCTL62	TRINITY_DN122_c0_g1_i8.p1	−3.385316645	0.306905	−3.334365581	4.24E-13
LrotCTL63	TRINITY_DN122_c0_g2_i1.p1	−2.470044584	0.323035	−2.448140017	7.57E-09
LrotCTL67	TRINITY_DN7708_c0_g1_i2.p1	−11.85018684	0.005272	NA	NA
LrotCTL71	TRINITY_DN2279_c0_g1_i12.p1	−2.457048578	0.341636	−2.404951985	2.21E-08
LrotCTL72	TRINITY_DN16821_c0_g1_i1.p1	−9.192684878	0.031434	NA	NA
LrotCTL76	TRINITY_DN48609_c1_g1_i1.p1	−10.24027855	0.018856	NA	NA
LrotCTL81	TRINITY_DN11126_c0_g1_i1.p1	−9.058546499	0.036762	NA	NA
LrotCTL93	TRINITY_DN253_c0_g1_i1.p1	−3.361336929	0.411893	−3.313384794	9.19E-14
LrotCTL94	TRINITY_DN253_c0_g1_i18.p1	−3.157849774	0.418025	−3.072733115	2.38E-12
LrotCTL96	TRINITY_DN309_c0_g1_i18.p1	−3.165996015	0.287761	−3.230542553	6.06E-13
LrotCTL97	TRINITY_DN309_c1_g1_i1.p1	−3.269953025	0.383464	−3.232881466	3.24E-13
LrotCTL99	TRINITY_DN309_c1_g1_i2.p1	−2.419126996	0.354555	−2.471327383	5.97E-09
LrotCTL100	TRINITY_DN309_c1_g1_i31.p1	−3.12334785	0.257382	−3.087485071	2.65E-12
LrotTEP1	TRINITY_DN11287_c0_g1_i3.p1	−3.404394275	0.266392	−3.484516713	3.37E-14
LrotSPZ5	TRINITY_DN2138_c0_g1_i17.p1	−3.275081578	0.31459	−3.327710618	3.28E-13
LrotSPZ8	TRINITY_DN3341_c0_g1_i11.p1	−2.049009075	0.384675	−2.203448453	2.62E-07
LrotSPZ9	TRINITY_DN3341_c0_g1_i12.p1	−1.909273169	0.36115	−2.086590999	9.16E-07
LrotSPZ12	TRINITY_DN500_c0_g1_i1.p1	−2.927933336	0.275752	−2.921277025	2.17E-11
LrotTube	TRINITY_DN17978_c0_g1_i1.p1	2.911360453	0.158431	2.941143581	3.62E-15
LrotCactus	TRINITY_DN10898_c0_g1_i1.p1	−2.800342017	0.316667	−2.832789907	7.69E-11
LrotDredd	TRINITY_DN6846_c0_g1_i6.p1	−2.509060049	0.334921	−2.547754088	2.83E-09
LrotCrustin	TRINITY_DN13938_c0_g1_i1.p1	−2.683772616	0.383464	−2.645961845	1.01E-09

All FPKM were adjusted by adding 0.01 to both male and female expression values to prevent division by zero. The $\log_2$FC was also recalculated according to adjusted FPKM values. We set the thresholds for significance for differential expression between sexes at $|\log_2\text{FC}| \geq 2$ and q-value < 0.05.

a sex-biased expression pattern ($|\log_2 FC| \geq 2$, q-value < 0.05). Of these transcripts, 62% of DEGs exhibited male bias, while 38% exhibited female bias.

Among 147 innate immunity-related transcripts that were identified using the same criteria as those applied to the genome dataset, we found 37 transcripts that were significantly differentially expressed between the two sexes (Table 4). The analysis by Cuffdiff revealed 5 male-biased transcripts, which all belonged to CTL families ($|\text{adjusted } \log_2 FC| \geq 2$, q-value < 0.05; Table 4). On the other hand, a separate DEG analysis conducted with EdgeR showed 24 male-biased recognition-related DEGs, six Toll pathway-related DEGs (five male-biased and one female-biased), one male-biased Imd pathway-related DEG and one male-biased DEG for LrotCrustin ($|\log_2 FC| \geq 2$, q-value < 0.05; Table 4). Overall expression levels are depicted based on the *Drosophila* model of the humoral innate immune system pathway (Fig. 1).

The differentially expressed gene analyses between male and female *L. rotunda* revealed that the Toll pathway is more highly activated in males than in females. When we investigated gene expression levels of innate immunity-related transcripts, *D. melanogaster* innate immunity-related gene families were largely comprised of weak non-biased transcripts ($|\log_2 FC| < 1$) and weak female-biased transcripts ($1 \leq \log_2 FC < 2$) except for the female-biased DEG Toll-1 ($\log_2 FC \geq 2$) (Gnad and Parsch, 2006). In contrast to *D. melanogaster*, for *L. rotunda*, there was a relative increase in the expression level of male-biased DEGs ($\log_2 FC \leq -2$) among the same components. Notably, Toll-1 and SPZ, which exhibited female bias in *D. melanogaster*, were male-biased in *L. rotunda*, although it is unclear whether the gene expression in *L. rotunda* TLRs and SPZs contributes to innate immunity as *D. melanogaster* Toll-1 and SPZ-1 did. Together with the elevated expression levels of TLRs and SPZs, the significantly increased expression of Dorsal/Cactus and LrotCrustin in male *L. rotunda* could be powerful evidence that the Toll pathway is highly activated in male *L. rotunda*.

Meanwhile, the genes corresponding to the JAK/STAT and JNK pathways showed female-biased expression, albeit at low q-values. It has been reported that the JAK/STAT pathway induces Tep1 expression and increases the phagocytic activity of hemocytes in *D. melanogaster* (Agaisse and Perrimon, 2004). However, it was difficult to correlate *L. rotunda* TEP expression with the JAK/STAT pathway since their reciprocal BLAST result with *D. melanogaster* matched with Tep2 and Tep3. However, with respect to the point that both the JAK/STAT and the JNK pathways were found to be highly conserved systems across arthropods, the differences between the slightly male-biased Bsk in *D. melanogaster* ($\log_2 FC$: −0.33) and significantly female-biased expression in *L. rotunda* ($\log_2 FC$: 3.27, q-value 0.40) were noteworthy for their potential evolutionary implications.

3.8. Susceptible male hypothesis for differential expression of immunity-related genes

One hypothesis that has been advanced in order to address differences in immune responses between males and females is that of the susceptible male (Belmonte et al., 2020). The hypothesis suggests that sexual dimorphism occurs as a result of one sex evolving a mating ornament (primarily males) and the other selecting mating partners based on this ornamentation (primarily females). From this perspective, the males allocate the majority of the reproductive investment between species, which comes at the evolutionary cost of investment of resources for building immunity. In *Drosophila,* the primary species referenced in the present study for gene expression comparisons, cuticle color, a sexually dimorphic feature in *Drosophila*, has in fact been reported to directly correlate with immune responses, which supports the susceptible male hypothesis. However, the mating behavior of the copepod *Labidocera* follows a radically different scheme than that suggested by the susceptible male hypothesis. Once the male *Labidocera* recognizes the mating partner female, it captures and holds the female by its characteristic large fifth leg, which initiates a mating ritual. It has been reported that the male strokes the female genital segment before the male copulates its spermatophore, which seems to serve as a check as to whether the male has captured the ideal female (not the cryptic female) and whether the female already contained another mating partner's spermatophore (Blades and Youngbluth, 1979). These examples suggest that when it comes to the trade-off between reproductive potential and immunity for males in either species, female choice is critical in *Drosophila,* while in *Labidocera* male choice may also be a factor, albeit to a different or lesser extent. The magnitude of the importance of male choice may therefore result in relatively higher activation of the Toll pathway in male *L. rotunda.*

If the relative amount of reproductive investment affects sexual selection, the female *Labidocera* must also be compared with the *Drosophila* female. In *Labidocera*, eggs spawned in specific seasons (mainly in cold winter) become diapause eggs. The diapause eggs sink to the bottom of the sea and hatch in the following spring or fall, after at least four weeks (Marcus and Fuller, 1986). In the other diapause egg spawning copepod, *Centropages tenuiremis*, it has been reported that the biochemical profile (dry weight, lipid, protein, and carbohydrate contents, etc.) of diapause eggs was elevated overall compared to the profile of subitaneous eggs (Hansen, 2019). Considering that the largest part of females' reproductive investment is egg reproduction (Titelman et al., 2007), the reproduction of diapause eggs may represent a very large resource investment for the female. Compared to the *Drosophila*, which does not spawn eggs in dormancy (Lirakis et al., 2018), the female *Labidocera* may allocate more reproductive investment than the female *Drosophila* does, which has compromised their immune response to an even greater degree than the male *Labidocera*.

Sex-biased expression patterns in copepod immunity genes have significant implications for fitness and reproductive success. Males with a stronger immune response are better equipped to resist diseases and pathogens, which enhances their health, longevity, and competitive edge in securing mates. However, this increased immune function may come at the cost of reduced resources for growth and other functions, potentially affecting their overall fitness. Conversely, females often allocate more resources towards reproduction, particularly in producing energy-intensive diapause eggs. While this high reproductive investment can lead to a greater number of offspring and ensure survival through adverse conditions, it also results in a reduced immune response, making females more susceptible to infections. This trade-off highlights a strategic divergence: males focus on maintaining health and mating success, whereas females prioritize reproductive output, even at the risk of decreased immune defense. These differences in immune strategy reflect adaptations to their respective reproductive roles and environmental pressures, shaping the dynamics of fitness and survival within copepod populations.

The recently reported differential immune response between sexes in *Drosophila* depends on the type of pathogen and the level of infection (infection rate, survival rate), and it therefore cannot be concluded that one sex has categorically stronger immunity than the other. However, the difference between the terrestrial ecosystem and the marine ecosystem is one of the factors that make it difficult to directly compare the pathogens exposed to *Labidocera* and *Drosophila*. Furthermore, studies on the immune response as it relates to sex are rare in copepods due to the difficulty of cultivating them in the laboratory. However, studying copepods with diverse life cycles would be a good opportunity to test various hypotheses that link immune response and sexual selection.

4. Conclusion

Our study is the first to conduct a comprehensive comparative analysis of innate immune systems across various copepod species. The findings reveal both similarities and differences when compared to other invertebrates. In *Labidocera rotunda*, the innate immune pathways (Toll, Imd, JAK/STAT, and JNK) are well-conserved, similar to those found in

insects and other crustaceans. However, unlike insects, copepods lack certain PRRs like PGRPs, reflecting a more simplified immune system. Additionally, copepods possess a more limited set of AMPs compared to other invertebrates. These differences suggest that copepods have evolved to retain essential immune functions while adapting to their specific aquatic environments. Additionally, we identified 24 key gene families that contribute to humoral immunity mainly found male-biased Toll pathway expression, offering new insights into the relationship between immune response and sexual selection.

Funding

This work was supported by the National Research Foundation of Korea (2022R1A2C4002058) and Korea Institute of Marine Science & Technology Promotion (RS-2022-KS221676) funded by the Ministry of Oceans and Fisheries, and Marine Life Classification and Diversity Research (MABIK 2024M00200) funded by the National Marine Biodiversity Institute of Korea.

Credit authorship contribution statement

Jimoon Jun: Formal analysis, Data curation. **Eun-Jeong Kim:** Formal analysis. **Donggu Jeon:** Formal analysis. **Jihye Yang:** Formal analysis. **Hyeon Gyeong Jeong:** Formal analysis. **Hyungtaek Jung:** Data curation. **Taeho Kim:** Funding acquisition, Formal analysis. **Seong-il Eyun:** Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jip.2024.108198.

References

Agaisse, H., Perrimon, N., 2004. The roles of JAK/STAT signaling in *Drosophila* immune responses. Immunol Rev 198, 72–82.
Akira, S., Uematsu, S., Takeuchi, O., 2006. Pathogen recognition and innate immunity. Cell 124, 783–801.
Belmonte, R.L., Corbally, M.-K., Duneau, D.F., Regan, J.C., 2020. Sexual dimorphisms in innate immunity and responses to infection in *Drosophila melanogaster*. Front Immunol 10.
Blades, F., Youngbluth, M., 1979. Mating behavior of *Labidocera aestiva* (Copepoda: Calanoida). Mar Biol 51, 339–355.
Bou Aoun, R., Hetru, C., Troxler, L., Doucet, D., Ferrandon, D., Matt, N., 2011. Analysis of thioester-containing proteins during the innate immune response of *Drosophila melanogaster*. J Innate Immun 3, 52–64.
Brown, S., Hu, N., Hombría, J.C.-G., 2001. Identification of the first invertebrate interleukin JAK/STAT receptor, the Drosophila gene domeless. Curr Biol 11, 1700–1705.
Brown, G.D., Willment, J.A., Whitehead, L., 2018. C-type lectins in immunity and homeostasis. Nat Rev Immunol 18, 374–389.
Buchon, N., Poidevin, M., Kwon, H.-M., Guillou, A., Sottas, V., Lee, B.-L., Lemaitre, B., 2009. A single modular serine protease integrates signals from pattern-recognition receptors upstream of the *Drosophila* Toll pathway. PNAS 106, 12442–12447.
Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., Mortazavi, A., 2016. A survey of best practices for RNA-seq data analysis. Genome Biol 17, 1–19.
Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. Gigascience 10, giab008.
Dziarski, R., Gupta, D., 2006. The peptidoglycan recognition proteins (PGRPs). Genome Biol 7, 232.
Eyun, S., 2017. Phylogenomic analysis of Copepoda (Arthropoda, Crustacea) reveals unexpected similarities with earlier proposed morphological phylogenies. BMC Evol Biol 17, 23.
Eyun, S., Soh, H.Y., Posavi, M., Munro, J.B., Hughes, D.S.T., Murali, S.C., Qu, J., Dugan, S., Lee, S.L., Chao, H., Dinh, H., Han, Y., Doddapaneni, H., Worley, K.C., Muzny, D.M., Park, E.-O., Silva, J.C., Gibbs, R.A., Richards, S., Lee, C.E., 2017. Evolutionary History of Chemosensory-Related Gene Families across the Arthropoda. Mol Biol Evol 34, 1838–1862.
Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., Smit, A.F., 2020. RepeatModeler2 for automated genomic discovery of transposable element families. PNAS 117, 9451–9457.
Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152.
Gnad, F., Parsch, J., 2006. Sebida: a database for the functional and evolutionary analysis of genes with sex-biased expression. Bioinformatics 22, 2577–2579.
Gottar, M., Gobert, V., Matskevich, A.A., Reichhart, J.-M., Wang, C., Butt, T.M., Belvin, M., Hoffmann, J.A., Ferrandon, D., 2006. Dual detection of fungal infections in *Drosophila* via recognition of glucans and sensing of virulence factors. Cell 127, 1425–1437.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29, 644–652.
Graham, A.M., Barreto, F.S., 2019. Loss of the HIF pathway in a widely distributed intertidal crustacean, the copepod *Tigriopus californicus*. PNAS 116, 12913–12918.
Hansen, B.W., 2019. Copepod embryonic dormancy: "An egg Is not just an egg". Biol Bull 237, 145–169.
Hoffmann, J.A., Kafatos, F.C., Janeway, C.A., Ezekowitz, R.A.B., 1999. Phylogenetic perspectives in innate immunity. Science 284, 1313–1318.
Holt, C., Yandell, M., 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinf 12, 491.
Igboin, C.O., Griffen, A.L., Leys, E.J., 2012. The Drosophila melanogaster host model. J Oral Microbiol 4.
Imler, J.-L., Hoffmann, J.A., 2001. Toll receptors in innate immunity. Trends Cell Biol 11, 304–311.
Imler, J.-L., Zheng, L., 2004. Biology of Toll receptors: lessons from insects and mammals. J Leukoc Biol 75, 18–26.
Jeon, M.S., Jeong, D.M., Doh, H., Kang, H.A., Jung, H., Eyun, S., 2023. A practical comparison of the next-generation sequencing platform and assemblers using yeast genome. Life Sci Alliance 6.
Jeon, D., Song, C.U., Jeong, H., Ohtsuka, S., Lee, W., Soh, H., Eyun, S., 2024. An integrated phylogenomic approach for potential host-associated evolution of monstrilloid copepods. Oceanogr Mar Biol Annu Rev 62, 350–375.
Jun, J., Jeong, H.G., Choi, H., Woo, H., Jeon, D., Seo, Y.J., Eyun, S., 2022. The complete mitochondrial genome of *Labidocera rotunda* Mori, 1929 (Copepoda: Calanoida) from Jeju Island, Korea. Mitochondrial DNA B Resour 7, 1702–1703.
Jung, H., Ventura, T., Chung, J.S., Kim, W.-J., Nam, B.-H., Kong, H.J., Kim, Y.O., Jeon, M.S., Eyun, S., 2020. Twelve quick steps for genome assembly and annotation in the classroom. PLoS Comput Biol 16, e1008325.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14, 587–589.
Katoh, K., Misawa, K., Kuma, K.i., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30, 3059–3066.
Kawasaki, T., Kawai, T., 2014. Toll-like Receptor Signaling Pathways. Front Immunol 5, 461.
Korf, I., 2004. Gene finding in novel genomes. BMC Bioinf 5, 59.
Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A., 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35, 4453–4455.
Lai, A.G., Aboobaker, A.A., 2017. Comparative genomic analysis of innate immunity reveals novel and conserved components in crustacean food crop species. BMC Genomics 18, 389.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359.
Letunic, I., Khedkar, S., Bork, P., 2020. SMART: recent updates, new developments and status in 2020. Nucleic Acids Res 49, D458–D460.
Leulier, F., Lemaitre, B., 2008. Toll-like receptors–taking an evolutionary approach. Nat Rev Genet 9, 165–178.
Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930.
Lima, L.F., Torres, A.Q., Jardim, R., Mesquita, R.D., Schama, R., 2021. Evolution of Toll, Spatzle and MyD88 in insects: the problem of the Diptera bias. BMC Genomics 22, 562.
Lin, Z., Wang, J.-L., Cheng, Y., Wang, J.-X., Zou, Z., 2020. Pattern recognition receptors from lepidopteran insects and their biological functions. Dev & Comp Immunol 108, 103688.
Lirakis, M., Dolezal, M., Schlötterer, C., 2018. Redefining reproductive dormancy in *Drosophila* as a general stress response to cold temperatures. J Insect Physiol 107, 175–185.
Marchler-Bauer, A., Bryant, S.H., 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res 32, W327–W331.
Marcus, N.H., Fuller, C.M., 1986. Subitaneous and diapause eggs of *Labidocera aestiva* Wheeler (Copepoda : Calanoida): Differences in fall velocity and density. J Exp Mar Biol Ecol 99, 247–256.
Matos, G.M., Rosa, R.D., 2022. On the silver jubilee of crustacean antimicrobial peptides. Rev Aquac 14, 594–612.
Matskevich, A.A., Quintin, J., Ferrandon, D., 2010. The *Drosophila* PRR GNBP3 assembles effector complexes involved in antifungal defenses independently of its Toll pathway activation function. Eur J Immunol 40, 1244–1254.

McTaggart, S.J., Conlon, C., Colbourne, J.K., Blaxter, M.L., Little, T.J., 2009. The components of the *Daphnia pulex* immune system as revealed by complete genome sequencing. BMC Genomics 10, 175.
Medzhitov, R., Janeway, C., 2000. Innate immunity. N Engl J Med 343, 338–344.
Michel, T., Reichhart, J.-M., Hoffmann, J.A., Royet, J., 2001. *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. Nature 414, 756–759.
Miller, D.J., Hemmrich, G., Ball, E.E., Hayward, D.C., Khalturin, K., Funayama, N., Agata, K., Bosch, T.C., 2007. The innate immune repertoire in cnidaria–ancestral complexity and stochastic gene loss. Genome Biol 8, 1–13.
Moreno-Santillán, D.D., Machain-Williams, C., Hernández-Montes, G., Ortega, J., 2019. *De Novo* Transcriptome Assembly and Functional Annotation in Five Species of Bats. Sci Rep 9, 6222.
Myllymäki, H., Valanne, S., Rämet, M., 2014. The *Drosophila* imd signaling pathway. J Immunol 192, 3455–3462.
Nonaka, M., Yoshizaki, F., 2004. Primitive complement system of invertebrates. Immunol Rev 198, 203–215.
Palmer, W.J., Jiggins, F.M., 2015. Comparative genomics reveals the origins and diversity of Arthropod immune systems. Mol Biol Evol 32, 2111–2129.
Pinacho-Castellanos, S.A., García-Jacas, C.R., Gilson, M.K., Brizuela, C.A., 2021. Alignment-free antimicrobial peptide predictors: improving performance by a thorough analysis of the largest available data set. J Chem Inf Model 61, 3141–3157.
Prochazkova, P., Roubalova, R., Dvorak, J., Navarro Pacheco, N.I., Bilej, M., 2020. Pattern recognition receptors in annelids. Dev & Comp Immunol 102, 103493.
Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.
Runsaeng, P., Puengyam, P., Utarabhand, P., 2017. A mannose-specific C-type lectin from Fenneropenaeus merguiensis exhibited antimicrobial activity to mediate shrimp innate immunity. Mol Immunol 92, 87–98.
Song, C.U., Choi, H., Jeon, M.S., Kim, E.J., Jeong, H.G., Kim, S., Kim, C.G., Hwang, H., Purnaningtyas, D.W., Lee, S., Eyun, S., Lee, Y.H., 2021. Zooplankton diversity monitoring strategy for the urban coastal region using metabarcoding analysis. Sci Rep 11, 24339.
Song, C.U., Purnaningtyas, D.W., Choi, H., Jeon, D., Kim, S., Hwang, H., Kim, C., Lee, Y. H., Eyun, S., 2024. Do red tide events promote an increase in zooplankton biodiversity? Environ Pollut 361, 124880.
Stanke, M., Morgenstern, B., 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res 33, W465–W467.
Titelman, J., Varpe, Ø., Eliassen, S., Fiksen, Ø., 2007. Copepod mating: chance or choice? J Plankton Res 29, 1023–1030.
Towb, P., Sun, H., Wasserman, S.A., 2009. Tube is an IRAK-4 homolog in a Toll pathway adapted for development and immunity. J Innate Immun 1, 309–321.
Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., Pachter, L., 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol 31, 46–53.
Vázquez-Mendoza, A., Carrero, J.C., Rodriguez-Sosa, M., 2013. Parasitic infections: a role for C-type lectins receptors. Biomed Res Int 2013, 456352.
Veltri, D., Kamath, U., Shehu, A., 2018. Deep learning improves antimicrobial peptide recognition. Bioinformatics 34, 2740–2747.
Yonekawa, A., Saijo, S., Hoshino, Y., Miyake, Y., Ishikawa, E., Suzukawa, M., Inoue, H., Tanaka, M., Yoneyama, M., Oh-Hora, M., Akashi, K., Yamasaki, S., 2014. Dectin-2 is a direct receptor for mannose-capped lipoarabinomannan of mycobacteria. Immunity 41, 402–413.
Yu, S., Luo, F., Xu, Y., Zhang, Y., Jin, L.H., 2022. *Drosophila* innate immunity involves multiple signaling pathways and coordinated communication between different tissues. Front Immunol 13, 905370.
Zelensky, A.N., Gready, J.E., 2005. The C-type lectin-like domain superfamily. FEBS J 272, 6179–6217.
Zhang, R., Cho, H.Y., Kim, H.S., Ma, Y.G., Osaki, T., Kawabata, S., Söderhäll, K., Lee, B.L., 2003. Characterization and properties of a 1,3-beta-D-glucan pattern recognition protein of Tenebrio molitor larvae that is specifically degraded by serine protease during prophenoloxidase activation. J Biol Chem 278, 42072–42079.
Zhang, L., Li, L., Guo, X., Litman, G.W., Dishaw, L.J., Zhang, G., 2015. Massive expansion and functional divergence of innate immune genes in a protostome. Sci Rep 5, 8693.
Zhu, Y., Yu, X., Cheng, G., 2020. Insect C-type lectins in microbial infections. In: Hsieh, S.-.-L. (Ed.), Lectin in Host Defense against Microbial Infections. Springer Singapore, Singapore, pp. 129–140.
Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L., Yorke, J.A., 2013. The MaSuRCA genome assembler. Bioinformatics 29, 2669–2677.