

RESEARCH ARTICLE

Open Access



# Comparative genomic analysis of innate immunity reveals novel and conserved components in crustacean food crop species

Alvina G. Lai\* and A. Aziz Aboobaker\*

## Abstract

**Background:** Growing global demands for crustacean food crop species have driven large investments in aquaculture research worldwide. However, large-scale production is susceptible to pathogen-mediated destruction particularly in developing economies. Thus, a thorough understanding of the immune system components of food crop species is imperative for research to combat pathogens.

**Results:** Through a comparative genomics approach utilising extant data from 55 species, we describe the innate immune system of the class Malacostraca, which includes all food crop species. We identify 7407 malacostracan genes from 39 gene families implicated in different aspects of host defence and demonstrate dynamic evolution of innate immunity components within this group. Malacostracans have achieved flexibility in recognising infectious agents through divergent evolution and expansion of pathogen recognition receptors genes. Antiviral RNAi, Toll and JAK-STAT signal transduction pathways have remained conserved within Malacostraca, although the Imd pathway appears to lack several key components. Immune effectors such as the antimicrobial peptides (AMPs) have unique evolutionary profiles, with many malacostracan AMPs not found in other arthropods. Lastly, we describe four putative novel immune gene families, potentially representing important evolutionary novelties of the malacostracan immune system.

**Conclusion:** Our analyses across the broader Malacostraca have allowed us to not only draw analogies with other arthropods but also to identify evolutionary novelties in immune modulation components and form strong hypotheses as to when key pathways have evolved or diverged. This will serve as a key resource for future immunology research in crustacean food crops.

**Keywords:** Malacostraca, Innate immunity, Evolution, Crustacean, Comparative genomics, Arthropoda

## Background

The global human population is projected to escalate to 9.1 billion by 2050 [1]. With an increasing food consumption per capita and changing demands for animal proteins [2], there is a dire need for sustainable sources to avoid further degradation of the environment. It has been suggested that much of this may come from invertebrate sources including insects [3], but clearly crustaceans currently represent a source of protein that is

more culturally palatable in Europe and North America. Crustaceans already represent a significant portion of marine aquaculture produce, with the predicted annual production exceeding 10 million tonnes and sales of \$40 billion [4] that will continue to increase. The expansion of farmed crustaceans is not without major issues. It is estimated that up to 40% (\$3 billion) of just shrimp production alone can be lost each year due to disease outbreaks [4]. Pathogens and diseases impacting crustaceans have been recently extensively reviewed [4–9]. Some of the most common diseases in decapod crustaceans are the white spot disease caused by the white spot syndrome virus (WSSV) in penaeid shrimps [8],

\* Correspondence: Alvina.Lai@zoo.ox.ac.uk; alvinagrancelai@gmail.com;  
Aziz.Aboobaker@zoo.ox.ac.uk  
Department of Zoology, University of Oxford, Tinbergen Building, South  
Parks Road, Oxford OX1 3PS, UK

yellow head disease caused by the Yellow head Virus [10–12], Taura syndrome caused by the Taura Syndrome Virus [13, 14], fungal diseases in the Dungeness crab, *Cancer magister* [15–17], infections by the parasitic dinoflagellate *Hematodinium* sp. in crabs [18], the *Panulirus argus* virus 1 (PaV1) infection in lobsters [19] and bacterial diseases caused by *Vibrio* or *Aeromonas* [20]. There is broad agreement that without new interventions and better understanding of pathology and immune responses, current best practices for crustacean aquaculture cannot be improved. The use of antibiotics and chemical treatments for disease control in aquaculture is undesirable due to long-term economic and environmental ramifications [21–23]. Therefore, approaches that harness and aid the crustacean innate defence mechanism should be exploited to limit and prevent diseases, and therefore crop loss. For example, assays for the measurement of innate immune activity could provide early warnings for the presence of potential pathogens within closed aquaculture systems.

Systematic and cross-species characterisation of the crustacean immune system has not been performed, despite it being essential for the field to progress [24]. Previous comparisons amongst sequenced arthropod genomes of insects, chelicerates, the myriapod *Strigamia maritima*, the branchiopod *Daphnia pulex* and the amphipod *Parhyale hawaiiensis* have recently revealed signatures of conservation and diversity in innate immunity components across arthropod phyla [25–27]. However, not much is known about the evolutionary events that define the immune system in malacostracans, or within the order Decapoda that includes crop species. The radiation of Pancrustacea (hexapods and crustaceans) has been estimated to be between ~540 and ~666 million years ago (mya) [28, 29] while the split of Branchiopoda from Malacostraca was estimated at 614 mya [28]. Others have made estimates of similar divergence times based on crustacean hemocyanins [30]. Given the large evolutionary time scales involved, many lineage specific changes in immune system components within the Malacostraca may have occurred and using only the branchiopod *D. pulex* and a single malacostracan *P. hawaiiensis* to define immune regulation is unlikely to provide either a rich or accurate picture. Ultimately this will require both comparative and functional genomics approaches to effectively understand and exploit the immune system. Due to potential importance of crustacean food sources, such studies are of high impact and urgency. Currently, the lack of a comprehensive comparative genomics study of immunity with the Malacostraca means that a clear staging point for underpinning this work is lacking.

Here, we address this major deficit by performing an in depth comparative study amongst the broader

Malacostraca, including extant data from the order Decapoda that includes all the major food crop species (Additional file 1: Figure S1). A large number of relatively recent independent studies have started to generate publically deposited large transcriptomic data sets from food crop species and other related malacostracan species providing ample raw data for our study; complete set of references provided in Additional file 2: Table S1 [31–41]. We have annotated innate immunity genes and pathways from 69 Malacostraca transcriptome datasets from 55 species representing five Malacostraca orders: Amphipoda (7 species), Decapoda (18 species), Isopoda (27 species), Euphausiacea (2 species) and Mysida (1 species) (Additional file 1: Figure S1; Additional file 2: Table S1 and Additional file 3: Table S2) [42–44]. We used sequence, motif and domain similarity based approaches to identify 7407 genes, representing 39 immune gene families in the Malacostraca (summarised in Fig. 1). We annotate genes that encode pathogen recognition proteins, signalling components of key signal transduction pathways such as Toll, Imd and JAK-STAT, effector genes encoding proteins that perform immune protection such as antimicrobial peptides and members of the antiviral RNAi pathway. Within these key groups, we define malacostracan specific evolutionary events that suggest a previously unsuspected variation in immune gene content, and that functional genomic studies of immunity specifically with species in this group will be required for clear understanding of host defence in food crop species. A comparison across these data sets also allowed us to expand the annotation of previously discovered crustacean specific immune components, confirming their importance across the group. Finally, taking a conservative approach using orthology analyses and Pfam annotations in the sequenced amphipod genome of the crustacean *P. hawaiiensis* [27], we describe four novel gene families with immune related protein domains conserved only within the Malacostraca. We show that these novel Malacostraca genes exhibit tissue-specific expression in the amphipod *P. hawaiiensis*. Overall, our work provides a comprehensive picture of the Malacostraca innate immune system and a key staging point that will now facilitate important immunology research to underpin food crop aquaculture.

## Results and Discussion

### Pattern recognition receptors in malacostracans are dynamically evolving and exhibit family-specific expansions

While vertebrates rely on adaptive immune systems and immunological memory mediated by secreted antibodies to ward off pathogens, many invertebrates, including arthropods use a pre-encoded set of proteins known as the pattern recognition receptors (PRRs) to recognise a

	<i>P. hawalensis</i>	<i>P. meridianus</i>	<i>P. ortizi</i>	<i>L. varnanaei</i>	<i>M. nipponense</i>	<i>C. borealis</i> (nervous system)	<i>E. sinensis</i>	<i>N. awatschensis</i>	<i>E. superba</i>	<i>D. pulex</i>	<i>D. melanogaster</i>	<i>A. aegypti</i>	<i>S. maritima</i>	<i>M. merletti</i>	<i>I. scapularis</i>	
<b>Recognition</b>																
GNBP	3	5	3	5	1	3	7	3	4	11	3	7	1	0	0	
DSCAM	1	1	1	1	1	1	1	1	0	1	1	1	60	12	13	
Domeless	1	1	1	1	1	1	1	0	1	1	1	1	2	0	2	
Galectin	1	1	1	1	1	1	1	0	1	3	7	12	2	6	5	
C-type lectin	48	14	12	65	13	11	29	17	33	26	33	40	25	24	17	
TEP	4	8	5	17	11	11	12	7	8	7	6	8	4	1	3	
SCRA	4	4	4	3	3	2	5	1	0	6	5	5	4	4	3	
SCRB	8	7	10	7	4	6	5	1	6	8	13	13	7	4	3	
<b>Toll Pathway</b>																
TLR	5	3	3	5	4	7	4	0	0	5	9	12	27	14	4	
Spätzle	7	5	3	3	0	4	6	7	26	4	6	6	2	5	2	
Tube	0	0	0	1	1	1	1	0	0	0	1	1	0	1	0	
Pelle	2	1	1	2	1	1	1	0	1	2	1	1	1	0	2	
MyD88	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	
Cactus	1	1	1	1	1	1	1	1	0	2	1	1	1	2	1	
Dorsal	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	
Tollip	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	
<b>Imd pathway</b>																
Imd	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	
Relish	1	1	1	1	1	1	1	1	1	4	1	3	1	1	1	
Caspar	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
DREDD	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0	
IKKb	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	
TAK1	1	1	0	1	1	1	1	1	1	1	1	1	3	1	1	
<b>JAK-STAT pathway</b>																
JAK	1	1	1	1	1	1	1	0	0	1	1	1	1	0	1	
STAT	2	1	1	1	1	1	1	0	1	1	1	1	1	2	1	
PIAS	1	1	1	3	2	2	1	1	1	1	1	1	1	0	2	
<b>Antimicrobial peptides</b>																
ALF	8	6	7	9	5	1	6	10	8	0	0	0	0	0	0	
Crustin	9	8	7	15	11	4	10	9	50	0	0	0	0	0	0	
<b>Antiviral RNAi</b>																
Argonaute	4	3	4	3	3	3	3	1	2	1	3	4	2	6	3	
Dicer	2	2	2	2	2	2	2	0	3	3	2	2	1	0	1	
TRBP	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
Drosha	1	0	0	1	1	1	1	0	0	1	1	1	1	1	1	
Pasha	1	1	1	1	0	1	1	1	1	1	1	1	1	2	1	
<b>Novel immunity genes</b>																
Peritrophin-A	2	3	2	10	2	0	1	2	2	0	0	0	0	0	0	
Death domain	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	
von Willebrand factor A	6	1	1	3	5	1	1	1	1	0	0	0	0	0	0	
ML domain	1	1	1	1	0	0	1	1	2	0	0	0	0	0	0	
<b>Other</b>																
CLIP-serine protease	54	68	46	72	53	27	54	56	47	3	47	69	4	7	4	
Prophenoloxidase	1	1	1	2	1	1	1	1	1	1	3	10	0	0	0	

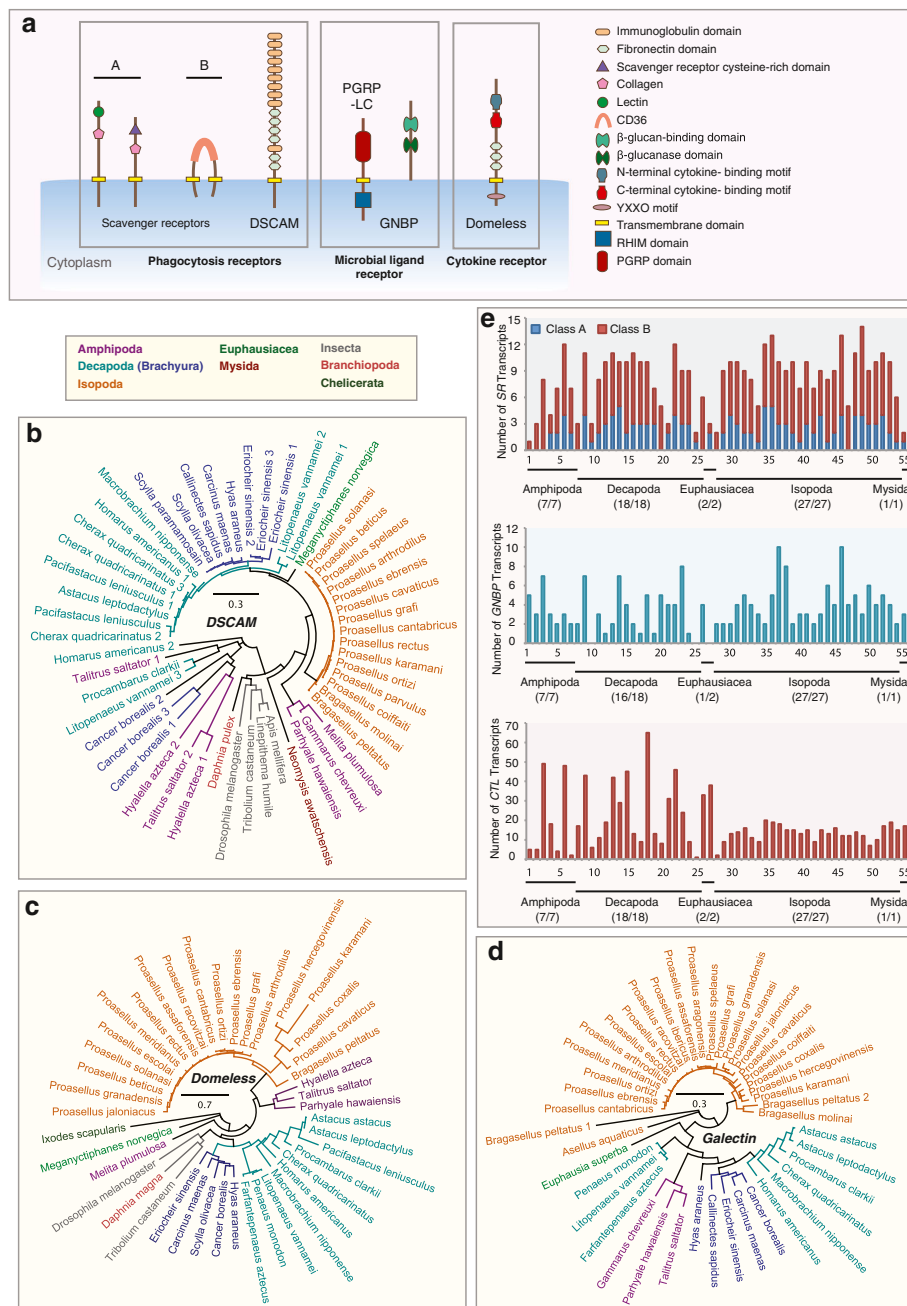
Amphipoda	Isopoda	Mysida	Branchiopoda	Chelicerata
Decapoda (Brachyura)	Euphausiacea	Insecta	Myriapoda	

**Fig. 1** Summary of gene copy number in representative malacostracans and arthropods. Gene copy number for malacostracans are determined in this study. Gene copy number for arthropods were obtained from these cited sources [25, 133]. A complete list of immunity genes identified in this study is presented in Additional file 4: Table S3, Additional file 9: Table S4, Additional file 11: Table S5, Additional file 13: Table S6, Additional file 15: Table S7, Additional file 16: Table S8, Additional file 18: Table S9

broad spectrum of microbial ligands. Arthropods PRRs facilitate microbial killing through a range of direct and indirect mechanisms [45–50] upon the detection of non-self pathogen structures known as pathogen-associated molecular patterns (PAMPs) present on the surface of microbes [51]. Some examples of PAMPs include peptidoglycans (PGN) and lipoteichoic acids (LTA) in Gram-positive bacteria, lipopolysaccharides (LPS) in Gram-negative bacteria and  $\beta$ -glucans from fungal cell walls [51, 52]. We examined seven PRR families in Malacostraca, which included the Gram-negative binding proteins (GNBPs), Down syndrome cell adhesion molecules (DSCAMs),

scavenger receptors (SRs), Domeless proteins (discussed in the signal transduction section), C-type lectins (CTLs), galectins and thioester-containing proteins (TEPs; Fig. 1; Fig. 2A). DSCAM, SRs and Domeless are transmembrane receptors (Fig. 2a). GNBPs can either be associated with the cell membrane via a glycosylphosphatidylinositol anchor or function as soluble receptors [53]. We identified 202 GNBPs, 49 DSCAMs, 443 SRs, 47 Domeless proteins, 1005 CTLs, 47 galectins and 432 TEPs in malacostracans (Additional file 4: Table S3).

GNBPs are a group of proteins that share the carbohydrate-binding  $\beta$ -glucanase domain. Multiple



**Fig. 2** Innate immunity pattern recognition receptors (PRRs) in malacostracans. **a** Domain architecture of PRRs. The neuronal transmembrane protein of the immunoglobulin (Ig) superfamily DSCAM contains tandem arrays of Ig and fibronectin domains. DSCAM is shown to participate in pathogen recognition in mosquitoes and phagocytosis flies. Scavenger receptors are a diverse group of multidomain proteins. Two classes of the membrane-associated scavenger receptors are shown. Members of the class A scavenger receptors (SCRAs) subfamily contain the characteristic scavenger receptor cysteine-rich (SRCR) domain, C-type lectin and collagenous domains. Class B scavenger receptors (SCRBs) are characterised by two transmembrane domains and a CD36 domain. PGRP-LC contains the RHIM (receptor-interacting protein homotypic interaction motif) domain. GNBP typically possess the  $\beta$ -glucan binding domain and the  $\beta$ -glucanase domain. Domeless is a cytokine receptor required for JAK-STAT signalling. Conserved protein domains of PRRs are shown in the figure inset. Phylogenetic trees of **b** DSCAM, **c** Domeless and **d** Galectin are constructed using the maximum-likelihood method from an amino acid multiple sequence alignment. Taxa labels are depicted as their respective colour codes. Bootstrap support values ( $n = 1000$ ) for all trees can be found in Additional file 26: Figure S14. Scale bar represents substitution per site. **e** The graphs represent the repertoire of putative PRR transcripts of the following gene families: SCRAs and SCRBs, GNBP and C-type lectins. The y-axis represents total number of genes identified in all 55 malacostracan species for each family. Each species is represented by a number on the X-axis and a complete list of species is available in Additional file 3: Table S2. Black horizontal bars below each graph delimit the five orders of malacostracans and the numbers in parentheses (x/y) represent the following: x = number of species in which a particular gene family is found and y = total number of species in each order



naming conventions exist for this group as they are also called lipopolysaccharide and  $\beta$ -glucan binding proteins,  $\beta$ -1,3-glucan binding proteins or  $\beta$ -1,3-glucanase-related proteins. Originally discovered in lepidopterans as proteins that can recognise  $\beta$ -1,3-glucans from fungal cell walls [54–56], others have shown that insect GNBP can also recognise Gram-negative bacteria [57–61]. Functional studies on malacostracan GNBP revealed that these proteins are expressed in hemocytes and hepatopancreas and expression is induced upon treatment with WSSV, Gram-negative and -positive bacteria [62–67]. GNBP have two active sites, identified in the  $\beta$ -glucanase domain of the silk moth *Bombyx mori* that are denoted as E188 and E193 [58]. Both glutamic acid residues are reported to be absent from most insect GNBP, which implied that insect GNBP lack catalytic activity [58, 68]. GNBP have undergone significant evolutionary changes within Arthropoda because chelicerates lack GNBP altogether [26] while *Drosophila melanogaster* and *D. pulex* have 3 and 11 GNBP proteins respectively (Fig. 1; Additional file 4: Table S3) [25]. We identified 202 GNBP genes from Malacostraca (Additional file 4: Table S3; Fig. 2e). Of these, 150 have intact  $\beta$ -glucanase domains. Alignment of these  $\beta$ -glucanase domains from Malacostraca with the *B. mori* sequence revealed that 109 out of 150 malacostracan GNBP possessed glutamic acid residues at both E188 and E193 positions (Additional file 5: Figure S2). From our analyses we show that a clear GNBP expansion has occurred within Malacostraca (Fig. 2e). Many decapods have expanded GNBP; for example *Procambarus clarkii*, *Eriocheir sinensis* and *Astacus leptodactylus* have 8, 7 and 7 GNBP respectively (Fig. 2e; Additional file 4: Table S3). GNBP expansion more broadly in Crustacea may compensate for the previously reported absence of peptidoglycan recognition proteins (PGRPs) in the crustacean lineage [25] since GNBP can also recognise Gram-positive bacteria. While we do find some PGRPs for the first time in our analysis - see Immune deficiency (Imd) signalling section, our data supports this idea. The 3 GNBP in *D. melanogaster* are catalytically inactive [26] and given that this species has 13 PGRPs, it is possible that in *D. melanogaster*, PGRPs may compensate for inactive GNBP. Thus in the malacostracan and insect lineages within Pancrustacea, these two key pathogen detection systems have undergone opposing evolutionary trajectories. This observation supports the suggestion that detailed functional genomic studies of immunity genes are required in a malacostracan species to properly make progress in food crop species immunology research.

A major component of arthropod immune systems that still requires further definition and is yet to be fully exploited is the DSCAM proteins, which undergoes

startling levels of alternative splicing (AS) in the Pancrustacean clade. The canonical DSCAM domain arrangement consists of 9 (immunoglobulin; Ig) – 4 (fibronectin; Fn) – (Ig) – 2 (Fn) [69]. Since other Ig-containing genes may confound the identification of bona fide DSCAM transcripts in malacostracans, we searched for genes/transcripts containing the Fn1-Fn2-Fn3-Fn4-Ig10-Fn5 motif set from known DSCAM protein sequences used as queries for BLAST. From this, we identified putative DSCAM transcripts in 49 out of 55 malacostracan species in our study (Fig. 2b). We observed that DSCAMs in brachyurans (except for *Cancer borealis*) are monophyletic and are likely to be orthologous (Fig. 2b). DSCAM AS in arthropods has evolved to allow versatile pathogen recognition and this is facilitated by the vast reservoir of receptor diversity resulting from alternative splicing of the hypervariable regions [27, 70–74]. Although it is likely that most malacostracan DSCAMs have multiple splice forms, accurate characterisation and annotation of splice variants from transcriptome data alone is confounded by long arrays of highly similar Ig exons. Genome and genomic DNA based approaches will be needed to assess this with accuracy. It seems likely that the DSCAM remains a key PRR in malacostracans, and will need to be further studied in the context of infection as potential diagnostic marker and effector mechanism that might be exploited in aquaculture.

Scavenger receptors (SRs) are a subclass of structurally diverse membrane-bound PRRs, first described as proteins having the ability to bind to oxidised low-density lipoproteins (LDLs) (Fig. 2a) [51, 75–79]. SRs can recognise a diverse range of cognate ligands and these include modified self-molecules (eg: oxidised LDLs) and non-self microbial structures such as LPS and LTA [80, 81]. We considered two classes of SRs in Malacostraca, namely the macrophage class A scavenger receptors (SCRAs) and the class B scavenger receptors (SCRBs; Fig. 2a). We annotated 129 SCRAs and 314 SCRBs in malacostracans (Additional file 4: Table S3; Fig. 2e). Malacostracan SCRAs are characterised by multiple domains; the cysteine-rich (SRCR) domain, C-type lectin domain, lysyl oxidase or collagen domain (Fig. 2a) [82–84]. Malacostracan SCRBs have the CD36 domain and two transmembrane domains (Fig. 2a) [79]. To date, the only SR reported in crustaceans is a homolog of Croquemort, a SCR family member in *Marsupenaeus japonicus* [85]. Humans and *Caenorhabditis elegans* only have three CD36-like proteins each [86, 87]. SCRBs in malacostracans have however, undergone multiple gene duplications; the isopod *Proasellus ortizi*, the decapod *C. borealis* and amphipod *P. hawaiiensis* have 10, 6 and 8 SCRBs respectively (Fig. 1; Additional file 4: Table S3). Major SCR gene expansion is likely to have occurred at the base of Mandibulata as *S. maritima*, *D. melanogaster* and *D. pulex*

have 7, 13 and 8 homologs respectively while the chelicerate *I. scapularis* only has three (Fig. 1). Clearly the role of SCRBS in mandibulate immunity needs further study as almost nothing is known about the significance of the SCRBS expansion in arthropods. Perhaps not all SCRBS in malacostracans are involved in host defence because CD36-like proteins have been shown to participate in other physiological roles such as facilitating cellular uptake of carotenoids required for visual chromophore formation [88], scavenging of apoptotic cells [89] and lipoprotein homeostasis [90].

Lectins have been shown to be directly relevant to the immune system of crustaceans [91, 92]. A C-type lectin (CTL) in *M. japonicus*, expressed primarily in intestinal tissues, is upregulated upon bacteria and WSSV infection and can bind LPS and PGN in a dose-dependent manner [93]. Nonetheless little is known about how many of each of the different types of lectins are present in malacostraca. CTLs are a group of diverse proteins characterised by a carbohydrate-recognition domain, some of which are  $\text{Ca}^{2+}$ -dependent and they can bind sugar and non-sugar ligands [94, 95], while galectins are another type of lectin proteins that can bind  $\beta$ -galactoside sugars and are involved in multiple cellular processes such as apoptosis, cell proliferation and immunity [96]. We identified over a thousand putative CTLs across Malacostraca with *Litopenaeus vannamei* having 65 different CTLs, in line with a general trend for decapods to have more CTLs than other malacostracan groups (Fig. 1; Fig. 2e; Additional file 4: Table S3). Although the copy number of CTLs varies greatly between malacostracan species (Fig. 2e; Additional file 4: Table S3), it is clear that divergent evolution through multiple gene duplications has occurred within this lineage, particularly in some amphipod and decapod species (Fig. 2e). More broadly we find that this appears to be a feature in many other arthropod lineages; *S. maritima*, *D. melanogaster*, *Aedes aegypti* and *D. pulex* have 25, 33, 40 and 26 genes respectively (Additional file 4: Table S3). To date, only five CTLs in *L. vannamei* have been studied in the contexts of Gram-negative bacteria agglutination and WSSV infection [97–101]. Future expression panel testing, particularly in decapods, will be required to ascertain whether CTLs may have distinct roles in recognising different pathogenic agents. Galectins in malacostracans are present as single-copy homologs except in two isopod species (*Asellus aquaticus* and *Bragasellus peltatus* that have two galectins each; Additional file 4: Table S3; Fig. 2d). Our analysis revealed that insects, chelicerates, *S. maritima* and *D. pulex* have multiple copies of galectins (Fig. 1) suggesting that with respect to galectins, malacostracans have evolved conservatively (Fig. 2d).

The thioester-containing protein (TEP) superfamily includes the vertebrate complement system, the pan-

protease inhibitor  $\alpha$ 2-macroglobulin ( $\alpha$ 2M), insect TEP-like proteins and macroglobulin complement related (MCR) proteins [102]. TEPs have the unique propensity to form covalent bonds with pathogens through their canonical thioester (GCGEQ) motifs to promote endocytotic clearance or to neutralise pathogenic proteases [103–105]. Amongst arthropods, some TEPs lack the canonical thioester motif and they presumably lack the ability to form covalent bonds with pathogenic surfaces [26]. We identified a total of 432 TEPs in all 55 malacostracan species (Additional file 4: Table S3). Decapods in general have more TEPs than other malacostracan orders (Additional file 6: Figure S3B). *P. clarkii* has at least 25 different TEPs, the highest amongst the malacostracan datasets considered here (Additional file 4: Table S3). However, only a third (147/432) of malacostracans TEPs have the GCGEQ motif. Nonetheless, as reports have indicated that a TEP protein in *D. melanogaster*, although lacking the thioester motif, could still bind to fungi and promote phagocytosis [104], this TEP diversity in malacostraca and particularly Decapoda may be immune related. We analysed phylogenetic relationships between TEP members in Malacostraca and observed that like TEPs in arthropods, they fall into three major categories:  $\alpha$ 2Ms, insect TEP-like proteins and MCRs (Additional file 6: Figure S3). The  $\alpha$ 2Ms in amphipods (except for 1 gene in *Talitrus saltator*) form a monophyletic group (Additional file 6: Figure S3A). The vertebrate C3 and C4 factors are also monophyletic, while C3 from the amphioxus *Branchiostoma belcheri* and C5 factors from mouse and human are paraphyletic (Additional file 6: Figure S3). Since we did not find any malacostracan TEPs clustering with the vertebrate complement factors and together with the observation that C3-like proteins are only found in chelicerates and myriapods [26], we predict that C3-like proteins have been lost in the Pancrustacea.

The recognition of PAMPs by PRRs is the first line of defence against invading pathogens. In this study, we have annotated known PRR families in Malacostraca and established analogies to arthropod PRRs (Figs. 1 and 2). We show that malacostracans have a large repertoire of PRR proteins to efficiently cope with a broad range of pathogens. Several PRR families, CTLs, GNBP and SCRBS, are expanded in malacostracans and this may, in part, contribute to enhanced plasticity when dealing with diverse microbial ligands. Our data will underpin comparative approaches as to how PRR activation in aquaculture affects outcomes in different conditions. Together our analyses indicate that PRRs are evolving rapidly within this lineage, reflecting the diverse selection pressure from pathogens encountered by different malacostracan groups.

### Prophenoloxidases are invented at the base of Pancrustacea

The prophenoloxidase-activating system (proPO) is another non-self pathogen recognition mechanism implicated in arthropod immunity [106–108]. Upon the recognition of LPS, PGNs or  $\beta$ -glucans by GNBP, a serine protease cascade ensues, which results in the proteolytic cleavage of proPO into active phenoloxidase (PO). PO then catalyses melanin formation [107] and this creates a physical barrier to inhibit further pathogen growth and movement [109]. Because PO plays functional roles in the melanisation pathway and wound healing [107, 110, 111], the emergence of PO is associated with the evolution of humoral immunity in arthropods. POs are thought to be members of the hemocyanin superfamily; a family that is exclusively found in arthropods [112]. Due to shared sequence similarities, it was proposed that hemocyanins could be converted to proPOs upon chemical treatments [113, 114]. Chelicerates (scorpions and spiders) and the myriapod *S. maritima* lack *sensu stricto* proPOs (Fig. 1) [25, 26, 112, 115, 116] and so perhaps they would need to rely on activated hemocyanins for melanin synthesis. To date, most crustacean proPOs were identified from decapods [117–123]. We found only two malacostracan proPOs from non-decapod species in GenBank, *Nebalia kensleyi* (Leptostraca; ACV33307.1), and *Oratosquilla oratoria* (Stomatopoda; ADR50356.1; Additional file 7: Figure S4A), indicating that proPO exists beyond decapod species. No other reports exist for proPOs in amphipods (except for *P. hawaiiensis*) [27], isopods, krills and mysid crustaceans. Some have reported that amphipods and isopods lack proPO [30, 124–127]. Failure to identify proPOs from isopods by an independent study could be due to the use of a limited EST dataset [127]. In contrast to the previous studies, we were able to identify proPOs from all five malacostracan orders (Additional file 7: Figure S4A). Since proPOs and hemocyanins have similar sequences, we confirmed that these are *bona fide* proPOs through reciprocal BLASTs and phylogenetic analysis (Additional file 7: Figure S4A). Considering that proPOs are present in insects, *D. pulex* and malacostracans but not in myriapod and chelicerate lineages (although related proteins with predicted tyrosinase activity were identified) [26], it is likely that this non-oxygen binding derivative of hemocyanin was invented at the base of Pancrustacea. ProPOs may have evolved distinct roles in immunity since we were still able to identify many other hemocyanin genes in malacostracans (Additional file 7: Figure S4B). Parallels have been drawn between the initiation of serine protease cascades and the conversion of proPOs into catalytically active POs after exposure to PAMPs [128, 129]. POs must be tightly regulated by serine proteases since PO activation

generates highly reactive toxic quinone intermediates [109, 130, 131] and CLIP-domain serine proteases are implicated in this process [132]. CLIP-domain serine proteases are expanded in Diptera insects (*D. melanogaster* has 47 genes) but not in *D. pulex*, *S. maritima* and chelicerates (Fig. 1) [26, 133]. We made similar observations on the expansion of CLIP serine proteases in Malacostraca (Additional file 7: Figure S4C). We annotated over 2163 CLIP serine proteases. The highest numbers across five malacostracan orders are: the decapod *L. vannamei* (72 genes), the amphipod *P. hawaiiensis* (54 genes), the isopod *Proasellus meridianus* (68 genes), the krill *Meganyctiphanes norvegica* (57 genes) and the mysid crustacean *Neomysis awatschensis* (56 genes; Additional file 4: Table S3). The expansion of CLIP serine proteases in malacostracans may signify a need for highly regulated PO activation and this correlates with our novel findings of proPO presence across the broader Malacostraca.

### Toll and JAK-STAT pathways are conserved in Malacostraca while several key components of the Imd pathway are lost

Signal transduction pathways link recognition of PAMPs by PRRs with transcriptional activation. Three well-studied pathways are the Toll, Imd and Janus Kinase (JAK)- signal transducer and activators of transcription (STAT) pathways. Components of the Toll pathway in malacostracans include a chain of interacting proteins: the Toll-like receptors (TLRs) [134–136], Spätzle [137–140], myeloid differentiation factor 88 (MyD88) [141], Tube, Pelle, Dorsal, Cactus and the Toll-interacting protein (TOLLIP; Fig. 3A). Our investigation of the malacostracan Toll pathway members suggests that this pathway is broadly conserved. Malacostracan TLRs appear to have undergone divergent evolution through multiple gene duplications and from our phylogenetic analysis, we saw that paralogs exhibited marked sequence divergence (Additional file 8: Figure S5). Like in dipterans (*D. melanogaster* and *A. aegypti* have 6 genes each), we discovered multiple-copies of the gene encoding Spätzle, the cytokine partner of Toll, in malacostracans. In particular we observed that *Euphausia superba* has at least Spätzle encoding genes, while we found numbers more in line with insects in *E. sinensis* (6 genes), *A. leptodactylus* (6 genes), *P. hawaiiensis* (7 genes), *M. norvegica* (5 genes), *N. awatschensis* (7 genes) and *Proasellus sonalasi* (7 genes; Fig. 3B). We conclude that the unique expansion in *E. superba* might be a unique curiosity of this species. We identified intact MyD88-Tube-Pelle complexes in only 13 out of 55 malacostracan species (Fig. 3c). As in insects [133, 142], MyD88 and Tube each exist as single-copy genes in Malacostraca (Additional file 9: Table S4). We found Tube transcripts from 17 species representing only the Decapoda and Isopoda orders. Only one report





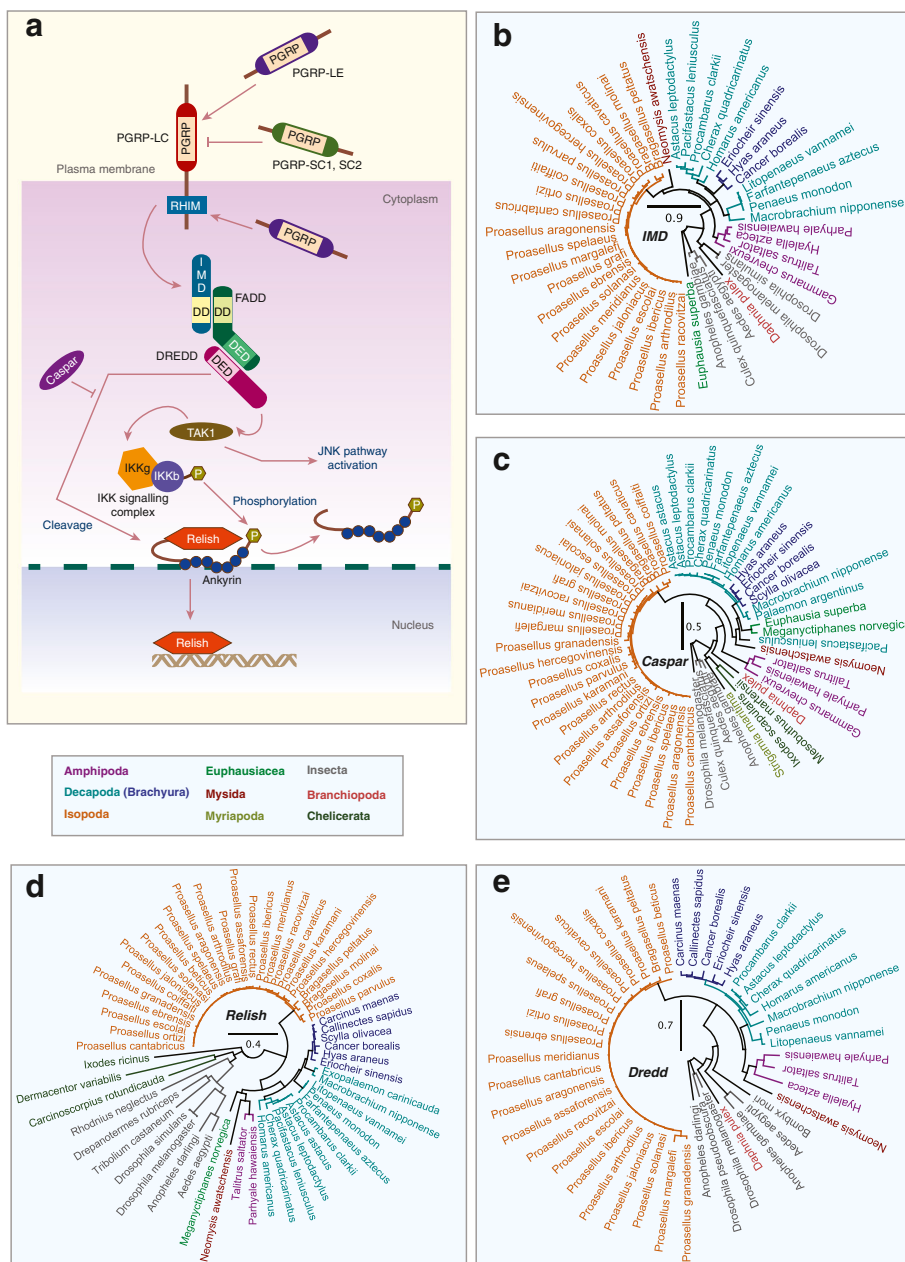
of a crustacean Tube homolog has been previously shown [143]. Although dipterans have one copy of Pelle, we observed duplications of Pelle in some malacostracan species in the Decapoda (*Homarus americanus* and *L. vannamei*), Amphipoda (*Echinogammarus veneris*, *P. hawaiiensis* and *T. saltator*) and Isopoda (*Proasellus beticus*, *P. coiffaiti*, *P. coxalis*, *P. grafi*, *P. hercegovinensis* and *P. rectus*) (fig. 3C; Additional file 9: Table S4). As two copies of Pelle were found in *D. pulex* and the deer tick *Ixodes scapularis* [26], it is possible that duplication of Pelle may have occurred at the base of arthropod lineages, with subsequent loss of one copy in insects. We identified single-copy homologs of Dorsal and Cactus in malacostracans (Fig. 3d, f and g). TOLLIP is a negative regulator of NF- $\kappa$ B in mammals [144]. Not much is known about the function of TOLLIP in invertebrates and to date, only one TOLLIP in crustaceans has been described [145]. We identified single-copy homologs of TOLLIP across five Malacostraca orders, *D. pulex*, the myriapod *S. maritima* and chelicerates (*Mesobuthus martensii* and *Ixodes scapularis*) (Additional file 9: Table S4, Fig. 3e). Amongst dipterans, we identified single-copy TOLLIP homologs in *Anopheles gambiae* but neither in *D. melanogaster* nor *A. aegypti*, although it is present in other insects like bees and ants (fig. 3E; Additional file 9: Table S4). Malacostracan TOLLIPs share similarities to mammalian TOLLIP proteins having both the protein kinase C conserved region 2 and the C-terminal coupling of ubiquitin to endoplasmic reticulum degradation domain [144, 146].

Most components of the Imd pathway are present in malacostracans except for three gene families (Fig. 4a). Imd is conserved amongst insects, myriapods and *D. pulex*, but not in chelicerates [26, 147]. Imd exists as a single gene within malacostracans across all five orders (Fig. 4b). Imd is preferentially activated by the inner PGN layer of Gram-negative bacteria through the binding of PGRP-LC to PGN [148, 149]. To our knowledge, no PGRP homologs have been previously reported in crustaceans including *D. pulex* [25] and *P. hawaiiensis* [27]. Although we failed to identify PGRPs in most malacostracans, we found four putative PGRP genes from *T. saltator* (Amphipoda), *Proasellus karamani* (Isopoda) and *H. americanus* (Decapoda; Additional file 10: Figure S6A). This could indicate a complex pattern of PGRP loss amongst crustacean taxa, that PGRPs are present but not represented in available malacostracan transcriptome and/or that the PGRP sequences we have found have evolved convergently. Sequence analysis revealed that these malacostracan PGRPs possess the amidase domain and share striking sequence similarities to *D. melanogaster* PGRP-SC1, SC2 and SB2 (Additional file 10: Figure S6D). Within this domain, five amino acid residues (H-Y-H-T-C;

marked in Additional file 10: Figure S6D) have been shown to be critical for PGRP enzymatic activity [150, 151]. These residues are present in the malacostracans PGRPs annotated here, indicating that they have the potential to be enzymatically active. These data suggest that PGRPs are present in Crustacean taxa but perhaps have greatly reduced representation. Future genome sequenced based analyses will be required to clarify this. A negative regulator of Imd signalling is the Caspar protein, a homolog of the mammalian Fas-associating factor 1 [152]. We identified single homologs of *Caspar* across all five malacostracan orders (Fig. 4C) and in other arthropods (*D. pulex*, dipterans, chelicerates and myriapod) indicating that it is conserved in Arthropoda (Additional file 11: Table S5). Concerning other Imd pathway components, we identified single-copy homologs of Relish, death-related ced-3/Nedd2-like protein (DREDD), I $\kappa$ B kinase  $\beta$  (IKK $\beta$ ) and MAPKKK transforming growth factor  $-\beta$  (TGF $\beta$ )-activated kinase 1 (TAK1) in malacostracans indicating that these components of the Imd pathway have remained intact (Fig. 4, Additional file 10: Figure S6 and Additional file 11: Table S5). However, we failed to identify clear homologs of I $\kappa$ B kinase  $\gamma$  (IKK $\gamma$ ), FAS-associated death domain (FADD) and TAK1-binding protein (TAB2) in malacostracan transcriptomes, which could be due to sequence divergence or the replacement of these components with other functionally related proteins. Independently, components of the Imd pathway have been reported in the decapod *Carcinus maenas* and the authors also did not find clear homologs for IKK $\gamma$  or FADD [37]. Not much is known about the Imd pathway in crustaceans. To date, only two homologs, Imd and Relish have been subjected to functional studies in shrimps [153–156]. Overall, the Imd pathway appears to be reduced in malacostracans. Whether this is a result of actual gene loss, sequence divergence or the utilization of alternative proteins for Imd signalling is presently unknown and warrants further investigation.

Core components of JAK-STAT include the cytokine transmembrane receptor Domeless, JAK (Hopscotch in *D. melanogaster*) and STAT proteins (Additional file 12: Figure S7A) [157–160]. Mammals have four JAK proteins and seven STATs [161] while most arthropods, except chelicerates, only have single-copy JAK and STAT homologs [26, 133]. JAK and Domeless proteins in malacostracans exist as single homologs (Fig. 2C; Additional file 12: Figure S7B, Additional file 13: Table S6). Like Domeless, STAT in most malacostracans exists as single homologs, except in two amphipod species; *Hyalella azteca* (2 genes) and *T. saltator* (3 genes; Additional file 12: Figure S7C, Additional file 13: Table S6). Negative regulators of JAK-STAT include the suppressor of cytokine signalling (SOCS) and





**Fig. 4** Immune deficiency (IMD) pathway members in malacostracans. **a** PGRPs recognise Gram-negative bacteria and activate the IMD pathway through the RHIM motifs. Although the IMD pathway is typically activated by PGRPs in *Drosophila melanogaster*, PGRPs are not necessary for IMD signalling and it was posited that an unknown protein is present upstream of the IMD signalling cascade. Like DIF from the Toll pathway, in the IMD pathway, differential activation of another NF- $\kappa$ B transcription factor, Relish, occurs. Relish is phosphorylated through the activation of I $\kappa$ B kinase (IKK) complexes and transforming growth factor- $\beta$ -activated kinase 1 (TAK1). The caspase-8 homologue death-related ced-3/Nedd2-like protein (DREDD) and FAS-associated death domain (FADD) proteins are required for IKK and TAK1 activation and Relish is cleaved through DREDD. Caspar, a homologue of mammalian Fas-associating factor 1 that is essential for antifungal immunity, negatively regulates the IMD-mediated immune response by preventing nuclear translocation of Relish. Caspar also suppresses the IMD pathway through targeting Dredd-dependent cleavage of Relish. Phylogenetic trees of **b** IMD, **c** Caspar, **d** Relish and **e** DREDD are constructed using the maximum-likelihood method from an amino acid multiple sequence alignment. Taxa labels are depicted as their respective colour codes. Bootstrap support values ( $n = 1000$ ) for all trees can be found in Additional file 26: Figure S14. Scale bar represents substitution per site

protein inhibitors of activated STAT (PIAS) [162, 163]. SOCS and PIAS are also well conserved in malacostracans (Additional file 13: Table S6; Additional file 12: Figure S7D and Additional file 14: Figure S8). Mammals have eight SOCS proteins while *D. melanogaster* only has three [160, 164]. Copy number of SOCS varies between malacostracans; some of the highest numbers are in *L. vannamei*, *P. hawaiiensis* and *M. nipponense* where they have 6, 6 and 5 genes respectively (Additional file 13: Table S6). Phylogenetic analysis of malacostracan SOCS proteins revealed that they clustered in seven major groups (Additional file 14: Figure S8). Few studies on crustacean SOCSs are available [165, 166], and whether the entire malacostracan SOCS repertoire have roles in immunity is yet unknown.

Overall we have shown that three signal transduction pathways, Toll, Imd and JAK-STAT have remained largely conserved in malacostracans. Nonetheless, several components of these pathways exhibit lineage specific diversification, for example, the loss of three Imd pathway modules (IKKY, FADD and TAB2) and the divergent evolution of core TLRs and Spätzle components of the Toll pathway.

#### Anti-lipopolysaccharide factors and crustins are malacostracan-specific antimicrobial peptides

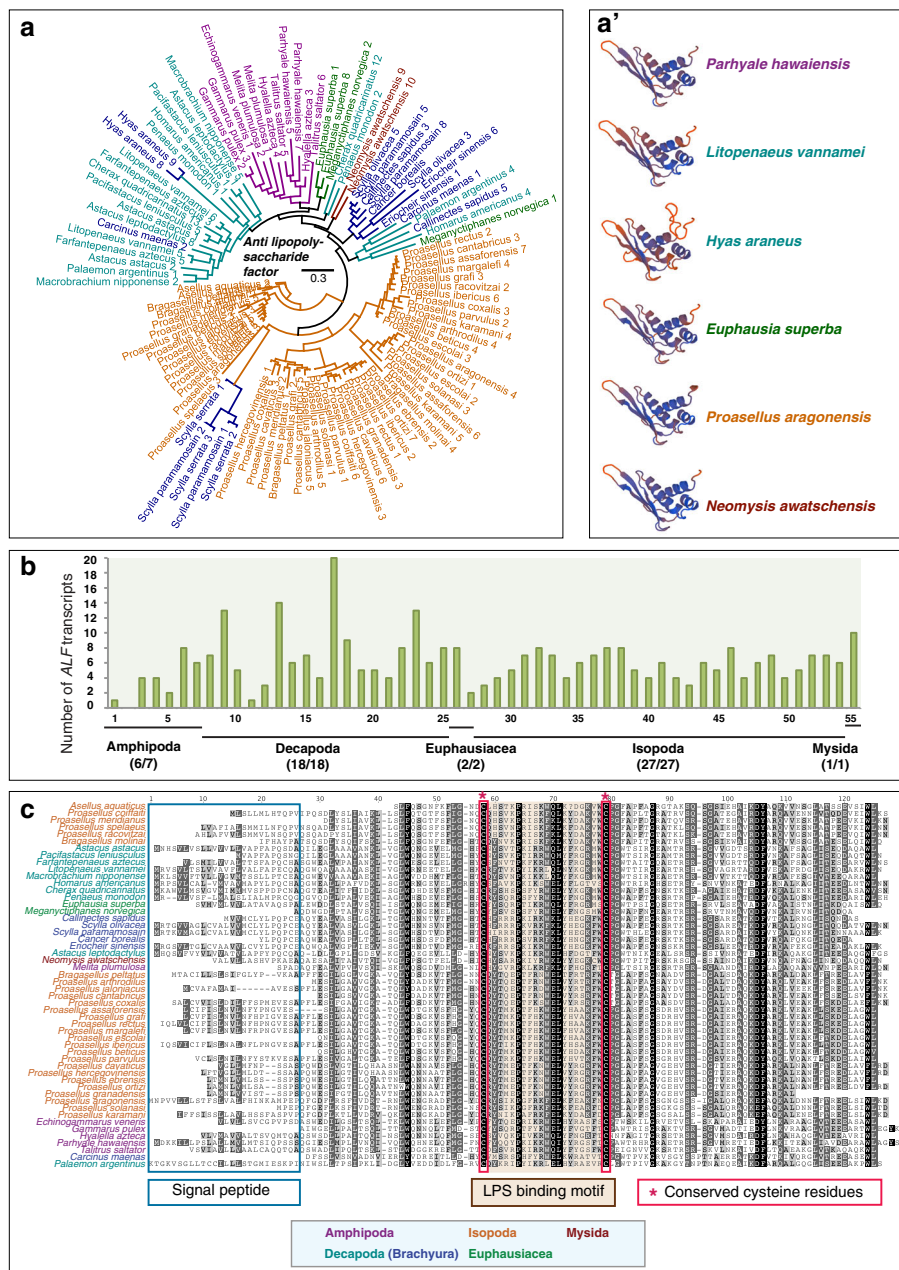
Signal transduction culminates in the activation of immune effector molecules to neutralise pathogenic agents. Antimicrobial peptides (AMPs) are rapidly evolving, highly specific effector proteins that are potent agents against a broad range of microbes [167, 168]. *D. melanogaster* has seven AMP families, but only three of these, attacins, cecropins and defensins, are shared with other dipterans [133]. To date, fifteen AMP families have been reported in crustaceans, fourteen of these are from decapods and many are lineage-specific [169, 170]. We considered two of these AMP families, anti-lipopolysaccharide factors (ALFs) and crustins, and show that they are actually well conserved in malacostracans beyond just the Decapoda (Figs. 5 and 6). While ALFs have only been reported in decapods [170, 171], crustins have been reported once in a non-decapod malacostracan species, the amphipod *Gammarus pulex* [172]. The branchiopod *D. pulex* lack ALFs and crustins or anything with sequence similarity, and we did not identify clear homologs in other arthropods, indicating that both gene families are specific to Malacostraca (Additional file 15: Table S7; Fig. 5a). We identified a total of 337 ALFs from malacostracans form a wide range of tissue samples (Additional file 15: Table s7). The decapod *Hyas araneus* has the highest number of ALFs (20 genes; Additional file 15: Table S7; Fig. 5b). Using homology modelling, we find that ALFs in malacostracans share high structural similarities, consisting of

three  $\beta$ -sheets and three  $\alpha$ -helices (Fig. 5a'). Alignment analysis of malacostracan ALFs revealed that they contained two conserved cysteine residues predicted to form a disulfide bridge (Fig. 5c) [170, 173, 174]. Between the cysteine residues, a region containing positively charged amino acids is defined as the LPS-binding domain [171]. This domain is present in all malacostracan ALFs, which suggests a conservation of LPS binding across this whole gene family (Fig. 5c).

Crustin is a cysteine-rich AMP containing a whey acidic protein (WAP) domain and was first discovered in the decapod *Carcinus maenas* to have a role in defence against Gram-positive bacteria [175]. Crustins are abundant in malacostracans and we identified 513 putative genes with *E. superba* encoding at least 50 crustins (Fig. 6a and b; Additional file 15: Table S7). In comparison with other WAP domain proteins, crustins are characterised by an additional crustin domain consisting of 12 conserved cysteine residues, in which a single WAP domain is present and we note that this the case in all malacostracans crustins (Fig. 6c). Future studies can now address the biological roles of these AMPs and questions as to whether these AMPs are differentially regulated by specific microbial ligands, whether they are broad spectrum or selective and whether they are active in specific developmental stages. Our analyses show that the immune effector phase in malacostracans has undergone substantial lineage specific evolution, expansion and sequence diversification of AMPs, reflecting their modes of action to guard against a broad range of pathogens found in their natural habitats. We also note conservation of AMPs across Malacostraca, meaning that non-food crop species that can be easily studied in the lab will be potential model systems for this aspect of malacostracan immunity.

#### Malacostracans have a canonical RNAi-based antiviral immune system

RNA interference (RNAi) is a conserved antiviral mechanism in many systems [176–181]. RNAi-mediated gene silencing is now employed as a method to prevent viral disease progression in shrimps through the targeting of viral genes in order to inhibit replication [182–186]. No direct mechanistic evidence exists regarding the involvement of the RNAi pathway components in crustacean innate immunity. Despite this, there have been increasing efforts to identify RNAi pathway members in penaeid shrimps because of the potential applicability of RNAi-derived technologies in circumventing viral diseases [180–182, 185, 187–192]. We annotated core RNAi components in malacostracans, which include Dicer, the trans-activating response (TAR) RNA-binding protein (TRBP) and Argonaute-2 (Fig. 7a). We found single-copy homologs of TRBPs across all five

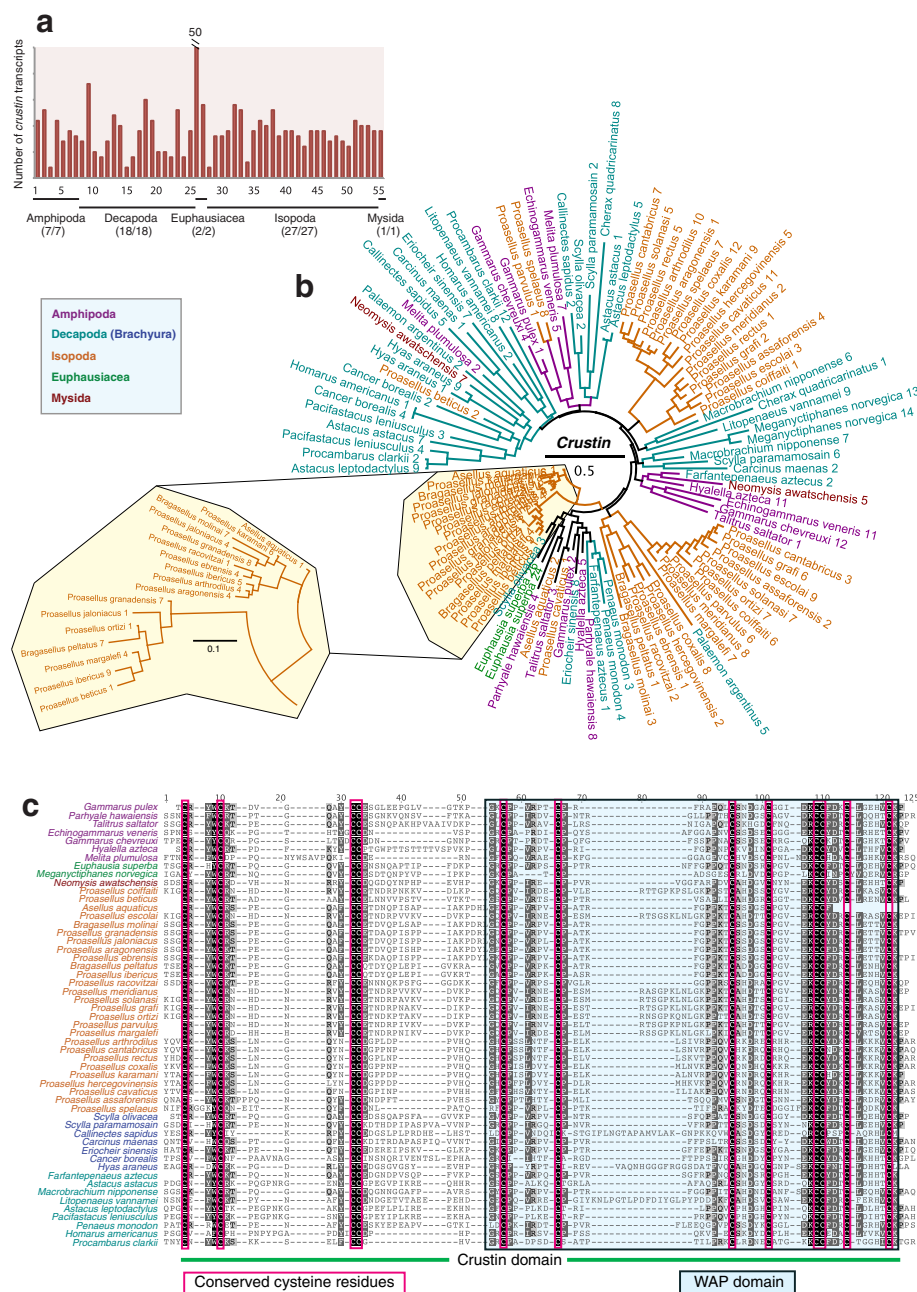


**Fig. 5** Anti-lipopolysaccharide factors (ALFs) in Malacostraca. **a** Phylogenetic tree of ALFs constructed using the maximum-likelihood method from an amino acid multiple sequence alignment. **(A)** Homology models of ALFs constructed with SWISS-MODEL revealed highly conserved predictions of the structural fold of these proteins. ALFs have conserved  $\alpha$ -helical and  $\beta$ -strand structures. **b** Graph of putative ALF transcripts. The y-axis represent total number of genes identified in all 55 malacostracan species for each family. Each species is represented by a number on the X-axis and a complete list of species is available in Additional file 3: Table S2. Black horizontal bars below each graph delimit the five orders of malacostracans and the numbers in parentheses (x/y) represent the following: x = number of species in which a particular gene family is found and y = total number of species in each order. Taxa labels are depicted as their respective colour codes. Bootstrap support values ( $n = 1000$ ) for all trees can be found in Additional file 26: Figure S14. Scale bar represents substitution per site. **c** Multiple sequence alignment of ALFs showing the putative signal peptides and LPS binding motifs characterised by two conserved cysteine residues are marked in red boxes

malacostracan orders and they share the conserved dsRNA-binding domain (Fig. 7d; Additional file 16: Table S8). We identified Dicer proteins in amphipods, isopods and krills, but not in the mysid

crustacean *N. awatschensis* (Additional file 16: Table S8; Fig. 7b). Dicer-1 and Dicer-2 have distinct roles in *D. melanogaster*, where the former is involved in microRNA (miRNA) biogenesis while the latter participates in

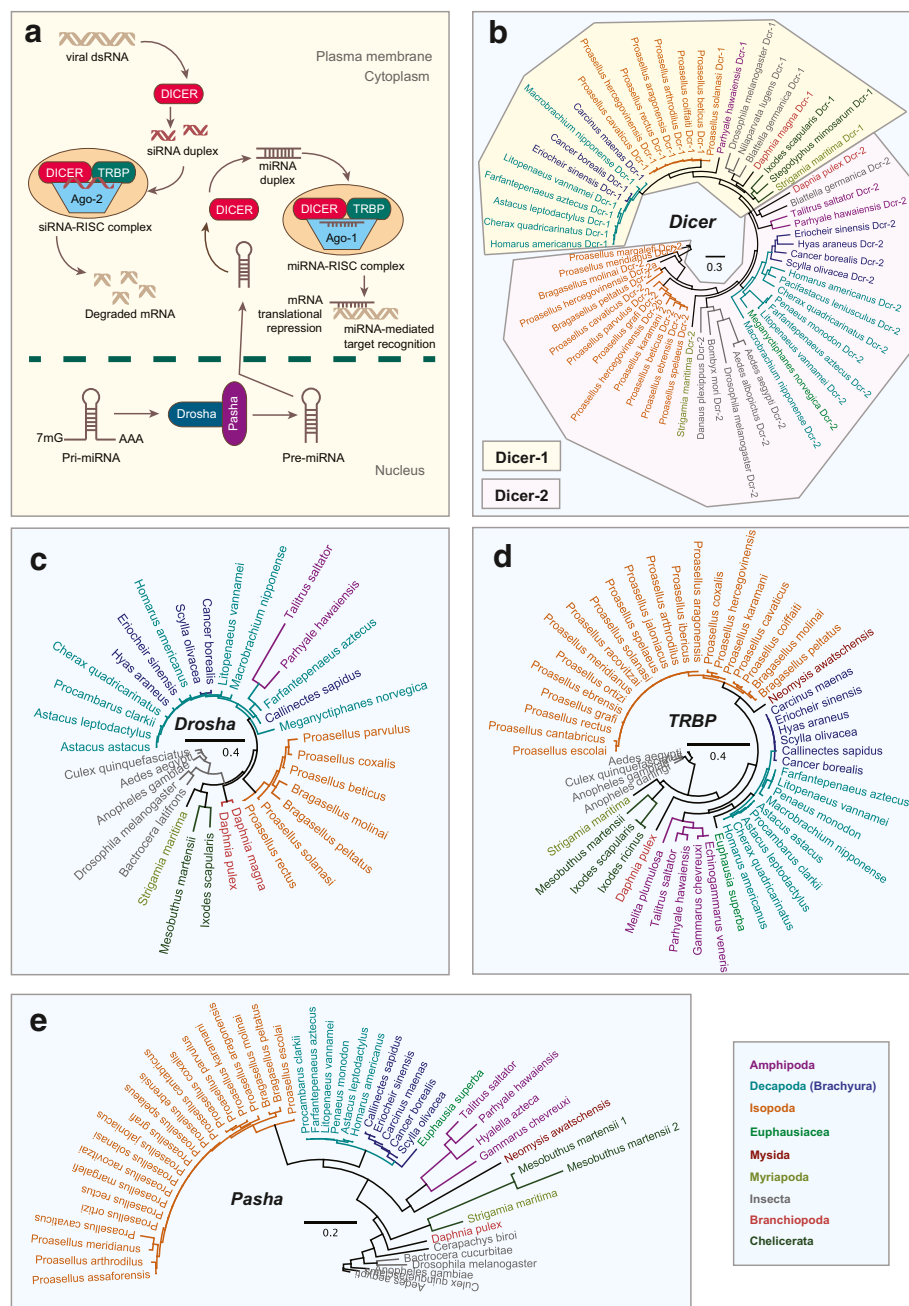




**Fig. 6** Crustin antimicrobial peptides in malacostracans. **a** Graph of putative *crustin* transcripts. The y-axes represent total number of genes identified in all 55 malacostracan species for each family. Each species is represented by a number on the X-axis and a complete list of species is available in Additional file 3: Table S2. Black horizontal bars below each graph delimit the five orders of malacostracans and the numbers in parentheses (x/y) represent the following: x = number of species in which a particular gene family is found and y = total number of species in each order. **b** Phylogenetic tree of crustins is constructed using the maximum-likelihood method from an amino acid multiple sequence alignment. Taxa labels are depicted as their respective colour codes. Bootstrap support values ( $n = 1000$ ) for all trees can be found in Additional file 26: Figure S14. Scale bar represents substitution per site. **c** Multiple sequence alignment of crustins showing the crustin domain and the WAP domain within it. The WAP domain is characterised by 8 conserved cysteine residues marked in red boxes. The crustin domain, which includes the WAP domain contains four additional cysteine residues marked in red boxes

dsRNAs processing into small-interfering RNAs (siRNAs) [193]. Phylogenetic and sequence analysis of malacostracan Dicer proteins revealed that they form two clusters

representing Dicer-1 and Dicer-2 (fig. 7B). With a few exceptions, most malacostracans have single-copy homologs of Dicer-1 and Dicer-2 proteins; the krill species *E.*



**Fig. 7** RNA interference (RNAi) pathway members in malacostracans. **a** Exogenous viral dsRNA is imported into the cytoplasm and is cleaved by the endoribonuclease from the RNase III family, Dicer. Cleaved fragments of dsRNA are known as small interfering RNAs (siRNAs). Dicer contains both Piwi/Argonaute/Zwille (PAZ) and helicase domains. Dicer activates the RNA-induced silencing complex (RISC), which is comprised of Argonaute-2 and the transactivating response RNA-binding protein (TRBP). Single-stranded siRNAs are incorporated into a RISC complex, upon which the siRNAs form complementary base-pairing to target mRNA and mRNA cleavage ensues. Other RISC-associated proteins include R2D2 and Loquacious in *D. melanogaster*. Dicer is also involved in microRNA (miRNA) biogenesis. Encoded by the genome, miRNAs are involved in the regulation of gene expression in the RNAi pathway. Transcribed by RNA polymerase II, Pri-miRNA is a long primary transcript of miRNAs and is processed into a stem-loop containing pre-miRNA by the microprocessor complex consisting of a ribonuclease III enzyme Drosha and Pasha, which is also known as DGCR8, protein. Pre-miRNA enters the cytoplasm and is cleaved by Dicer to generate a mature miRNA that is then integrated into the RISC complex. The miRNA-targeted transcript is either degraded or silenced. Phylogenetic trees of **b** Dicer, **c** Drosha, **d** TRBP and **e** Pasha are constructed using the maximum-likelihood method from an amino acid multiple sequence alignment. Taxa labels are depicted as their respective colour codes. Bootstrap support values ( $n = 1000$ ) for all trees can be found in Additional file 26: Figure S14. Scale bar represents substitution per site



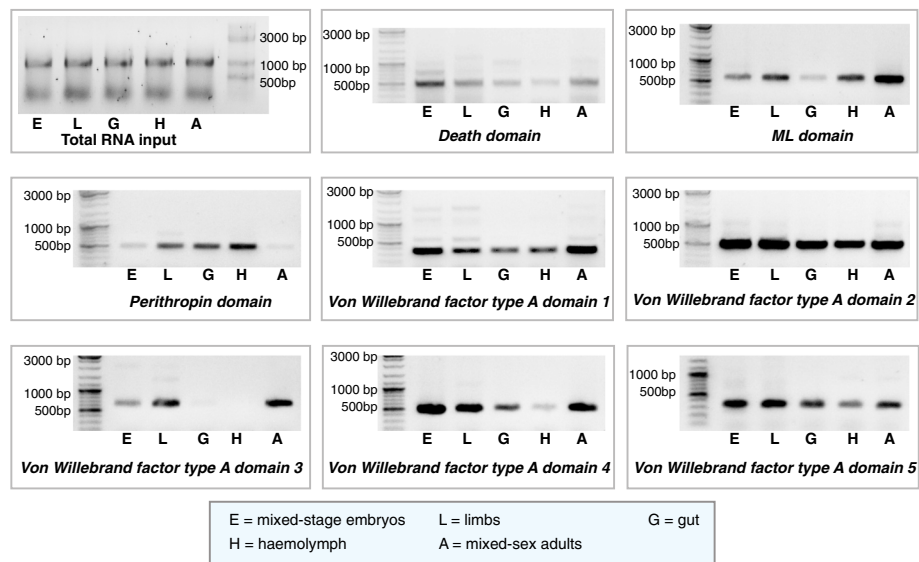
*superba* and *M. norvegica* have only Dicer-2 transcripts (Additional file 16: Table S8). With respect to Argonautes, we observed that most malacostracan species have multiple copies of this gene. We show that these putative Argonautes form a separate cluster from the closely related Piwi proteins (Additional file 17: Figure S9). We identified single-copy homologs of Argonaute-1 and multiple copies of Argonaute-2 in malacostracans (Additional file 17: Figure S9). Duplications of Argonaute-2 have occurred independently in specific lineages because variable copy number of this protein is reported in chelicerates but not in insects [26, 194]. Also, the longer branch lengths of Argonaute-2 proteins indicate that sequence evolution is higher than those of Argonaute-1 (Additional file 17: Figure S9). In *L. vannamei*, only Argonaute-2 is responsive to dsRNA [191]. Hence, it was thought that Argonaute-1 operates through the miRNA pathway in shrimps [195]. As in arthropods, the miRNA pathway is associated with crustacean antiviral defence [196, 197]. The expression of miRNAs in *M. japonicus* was differentially regulated upon viral challenge [198] and in other systems, viral infection results in the modification of host miRNA profiles [199–201]. Components of miRNA biogenesis are intact in malacostracans; we identified single-copy homologs of *Drosha* and partner of *Drosha* (*Pasha*) (Fig. 7c and e). Research in the area of RNAi-mediated antiviral immunity has remained comparatively sparse in crustaceans, despite its rich therapeutic potential. Our results provide independent evidence that malacostracans have a naturally occurring antiviral defence mechanism in place. Much more needs to be done to understand the role of RNAi in innate immunity before it can be exploited for host defence against viral infections.

#### Four novel gene families with potential involvement in malacostracan immunity

We have shown that although most canonical immunity genes and pathways in malacostracans share broad conservation with arthropods, lineage specific diversifications and gene duplications are common, which together suggests that lineage specific immune components may exist. With the advent of high-throughput sequencing, we are now able to tap into the availability of growing transcriptomic resources to find currently unknown proteins that might have potential involvement in host defence. Here, we present four novel gene families classified on the basis of shared domains implicated in immune function. We obtained a set of crustacean specific proteins from an orthology analysis using complete arthropod genomes [27]. This list contained 750 protein sequences that have no significant blast hit to any other sequences in the NCBI nr database. We filtered this list down to 82 genes based on the presence of known Pfam domains [202] and then down to a selection of 4 genes

with domains suggestive of immune function (Additional file 18: Table S9). We used *P. hawaiiensis* as a starting point for this analysis as this is the only complete Malacostraca genome available to date. While we are aware of potential limitations of this approach; for example, we may miss gene families that are not present in *P. hawaiiensis*, since we were interested in genes that are found across all five malacostracan orders, we were able to rationalise the use of *P. hawaiiensis*, an established laboratory organism, as a basis for comparison. Considering genes with known Pfam domains allowed us to investigate potential structural characteristics that have implied immune function. We observed broad conservation of all four novel gene families in Malacostraca. We have named these gene families, pending functional studies, according to their Pfam annotations: 1) chitin binding peritrophin-A family, 2) death domain family, 3) ML domain family and 4) von Willebrand factor type A family (Additional file 18: Table S9). We found that these genes are all expressed in the *P. hawaiiensis* and that they exhibited differential expression patterns between developmental stages and tissue types (Fig. 8; Additional file 19: Table S10).

Peritrophins are chitin-binding proteins originally isolated from the insect gut peritrophic membrane [203]. Peritrophic membrane is thought to constitute a barrier for midgut epithelial cells to prevent the entry of microbes [204–206]. Peritrophin-like proteins are characterised by peritrophin domains. One example is the peritrophin-A domain, which contains six conserved cysteine residues separated by other non-conserved amino acids (Additional file 20: Figure S10) [207]. It has been shown recently that crustaceans also have peritrophin-like genes. Penaeid shrimps express peritrophins during oogenesis and these proteins have roles in the protection of spawned eggs against *Vibrio* [203, 208]. A peritrophin-like gene cloned from *Fenneropenaeus chinensis* could bind chitin and Gram-negative bacteria [209] and another peritrophin-like protein from *Exopalaemon carinicauda* is involved in WSSV infection [210]. To date, only 10 peritrophin-like genes have been identified in crustaceans and no reports exist beyond decapod species [203, 208–212]. Here, we identified 80 novel peritrophin-like genes across five Malacostraca order (Additional file 18: Table S9; Additional file 20: Figure S10A). This novel family of peritrophin-like proteins have no significant similarities to other peritrophin genes previously reported in crustaceans or arthropods. They share 18 conserved cysteine residues, a chitin-binding peritrophin-A domain and multiple conserved aromatic amino acids (Additional file 20: Figure S10B). In *P. hawaiiensis*, we observed that one peritrophin-like gene is highly expressed in gut, hemocyte and limb samples but not in whole embryonic or adult tissue samples



**Fig. 8** RT-PCR results showing expression patterns for eight novel malacostracan genes in five *Parhyale hawaiiensis* tissue samples. These genes represent the four novel malacostracan gene families with potential function in innate immunity; genes containing the death domain, ML domain, peritrophin domain and the von Willebrand factor type A domain. Tissue samples were obtained from *P. hawaiiensis*: mixed stage embryos, limbs, gut, hemolymph and mixed-sex mature animals. Total RNA input (1  $\mu$ g) used in reverse transcription for all five tissue samples are shown

(Fig. 8). This corroborates the observation that other peritrophins are found in the gut peritrophic membrane. They may serve additional roles in innate immunity since we also saw increased expression in circulating hemolymph (Fig. 8).

The second novel gene family specific to malacostracans is characterised by a death domain (DD). The DD superfamily represents evolutionarily conserved proteins of four subfamilies: the DD subfamily, the caspase recruitment domain subfamily, the pyrin domain subfamily and the death effector domain subfamily [213, 214]. Many DD superfamily members are involved in the regulation of immune response. Some examples are Imd, FADD and DREDD of the Imd pathway [215–219], MyD88, Tube and Pelle of the Toll pathway [137, 220] and receptor interacting protein, tumor necrosis factor receptor-1 (TNFR1), TNFR-associated death domain and MAP kinase-activating death domain of the TNF pathway [221–225]. This new malacostracan DD gene family is a group of novel single-copy transcripts sharing a C-terminal DD (Additional file 21: Figure S11A). This family is specific to Malacostraca and has no significant blast results to the nr database or any other DD-containing proteins (Additional file 18: Table S9). The N-terminal region of this family exhibited considerable similarities within members of this group but no known domains could be determined by hidden Markov model (HMM) in this region (Additional file 21: S11B). Experimental confirmation revealed that embryonic samples show the

highest expression of a *P. hawaiiensis* DD gene (Fig. 8). This observation is intriguing because programmed cell death can be a key process during animal embryogenesis [226] and many DD proteins are shown to be involved in apoptosis [216, 227].

The third novel gene family we identified in malacostracans is characterised by the presence of a MD-2-related lipid recognition (ML) domain. The ML domain was originally identified from a group of unknown proteins that share regions of homology with the MD-2 protein [228]. We identified 39 transcripts containing the ML domain; they have no significant blast hits to any known genes so we named this the malacostracan ML family (Additional file 22: Figure S12A; Additional file 18: Table S9). The malacostracan ML family contains six conserved cysteine residues and these residues may be involved in the formation of disulfide bonds (Additional file 22: Figure S12B). Mutation of a conserved cysteine in MD-2 abolishes the response to LPS, which suggests a role of the disulfide bond in MD-2 function [229]. Members of this ML family are expressed in a wide range of tissue types; we observed expression in brain and nervous system samples (Additional file 22: Figure S12A; Additional file 18: Table S9). ML genes in *P. hawaiiensis* exhibited the highest level of expression in adult tissue samples and to lesser degrees in embryos or gut samples (Fig. 8). Since ML proteins have been implicated in lipid signalling, metabolism and immunity [228, 230–232], increased expression of *P. hawaiiensis*

ML in whole adult samples may imply a metabolic role in addition to host defence, as expression is also observed in hemocytes. Although not much is known about the direct roles of ML proteins in immunity, others have proposed that they could participate as lipid-binding cofactors in the recognition of pathogenic agents [228]. In mammals, MD-2 directly binds LPS through the MD-2-TLR4 complex of the Toll signalling pathway [233–235]. MD-2 has a leader sequence for endoplasmic reticulum targeting and secretion but lacks any transmembrane domain [236]. We predicted the presence of putative N-terminal transmembrane topologies in the malacostracans ML family, which indicates that they could be anchored to the cell membrane and have unknown PRR functions rather than being secreted (Additional file 22: Figure S12B).

The final family of novel malacostracan genes with potential involvement in immunity has the von Willebrand factor type A (VWA) domain. Von Willebrand factor (VWF) proteins were discovered in patients with blood clotting disorders, named the von Willebrand disease [237, 238]. They are ubiquitous in blood plasma and connective tissues and have functions in binding blood clotting factors and in mediating platelet adhesion at the site of vascular injury [239–241]. It was thought that enzymatic components of the blood clotting and complement system utilise similar macromolecular building blocks that existed before the protostome-deuterostome divergence [242]. Since arthropods have open circulatory systems and lack adaptive immunity, hemolymph clotting is a pivotal part of the immune response because it functions not only to prevent hemolymph loss but also to immobilise pathogens at the site of wound [243]. Arthropod hemolymph is also sensitive to small amounts of microbial polysaccharides [244–246] and clotting enzymes in arthropods emerged via convergent evolution from the assembly of domains found in vertebrate factors rather than being exact orthologs [242]. Here, we describe the novel observation of a gene family that contained a VWA domain (Additional file 18: Table S9). They are present as multiple-copy homologs; we identified 87 genes from all five malacostracan orders and they have no significant sequence homology to any other known sequences (Additional file 23: Figure S13). The malacostracans VWA family is made up of members with long transcripts up to 7 kb in length, with coding sequences translated to polypeptides of up to 1500 amino acids; this feature is commonly seen in other VWF proteins [247, 248]. We observed expression in tissue specific datasets obtained from the brain, nervous system, hepatopancreas, hemocytes, gill and eyestalk (Additional file 18: Table S9). From our RT-PCR results, the five VWA genes in *P. hawaiiensis* exhibited diverse patterns of expression; e.g. the VWA2 gene appears to

be highly expressed in embryos, limbs, gut, hemolymph and adult tissues while the VWA3 gene is not expressed in gut and hemolymph samples (Fig. 8). In summary, we have described four novel gene families specific only to malacostracans, which are good candidates for fulfilling roles in host defence. These genes offer new avenues for research and further analysis will be required to ascertain if they fit into the conceptual framework of innate immunity. Given our conservative approach to identifying gene families with domains that relate to known Pfam domains, it seems likely further comparative study, coupled with functional genomics and immunobiology approaches, will identify more malacostracan specific immune related genes.

## Conclusion

The recent availability of transcriptome sequences of distantly related malacostracan species has allowed us to describe molecular components of their innate immune systems at a new level of detail. This data is now available to the community to inform the next stages of immune research to underpin important aquaculture developments. By separating the immune response into successive phases, we observed dynamic evolutionary adaptations in the pathogen recognition phase, signal transduction and effector response systems. Malacostracans achieve flexibility in recognising infections through the divergent evolution of certain PRR families, notably the gene expansions of GNBP and CTLs. Upon recognition, several enzymatic cascades are involved in signal modulation and these too have novel evolutionary features. Malacostracans achieve diversity in modulation components through gene duplications of modulatory families involving CLIP serine proteases and Spätzle. When drawing comparisons to other arthropods, we observed novelties in these immune modulation components and are able to form strong evolutionary hypotheses as when key pathways evolved or diverged (e.g. the invention of proPO at the base of Pancrustacea). Core immune signal transduction pathways are largely conserved in malacostracans, although several components of the Imd pathways have been lost. The Imd pathway is activated through the digestion of PGNs by PGRPs in *D. melanogaster* [53].

PGRPs are previously thought to be lost in Crustacea; as *D. pulex* lack PGRPs [25] and no PGRPs are present in the *P. hawaiiensis* genome [27]. Despite this being true for most malacostracan datasets we considered here, we were able to identify four PGRP genes spread across the Amphipoda, Isopoda and Decapoda orders, which are predicted to be catalytically active based on the presence of essential residues (Additional file 10: Figure S6A and D). Although unlikely, PGRPs in these species may have appeared through convergent evolution. Future studies

will be required to determine whether the biological roles of these PGRPs have any relevance in host defence and the evolutionary events that explain their relatively patchy phylogenetic distribution.

Effector mechanisms in malacostracans, like in other arthropods, are highly divergent and lineage specific. We described two malacostracan-specific AMPs previously confined to the Decapoda, and show that members of these families are widespread in other non-decapod malacostracan species. Crustaceans are regularly exposed to viral components in their natural environments [180, 249, 250] and hence need antiviral mechanisms in place to counteract infection. We demonstrate that malacostracans have intact siRNA and miRNA components.

Finally, we present four novel gene families in Malacostraca as potential key players of the innate immunity. We only addressed the structural significance of these genes in the context of host defence based on comparisons with other immune proteins containing similar structural features. More functional studies will be required in the future to ascertain the roles of these genes and their potential function in innate immunity before they can be confirmed as crustacean immune system components. Resources presented in this study facilitate and expand the scope of both basic and applied research, in particular analyses on the mechanistic links between specific immune modules and overall host defence. Importantly our data suggest that non-decapod species, like the laboratory model *P. hawaiiensis*, may nonetheless be suitable for studying malacostracan specific immune mechanisms relevant to food crop species.

## Methods

### Innate immunity datasets and query sets

We retrieved complete transcriptome datasets for malacostracan species (available at the time of manuscript preparation) from the European Nucleotide Archive (<http://www.ebi.ac.uk/ena>) (Additional file 24). These transcriptomes included those generated from specific tissue types or developmental stages. We also included the *Parhyale hawaiiensis* transcriptome generated by a separate study [27]. All analyses were performed on a total of 69 transcriptome datasets. A full list of datasets and accessions used in this study is listed in Additional file 2: Table S1. For query sequences used in homology searches, we retrieved a set of insect immunity genes from ImmunoDB (<http://cegg.unige.ch/Insecta/immunodb>) [133] and known malacostracan entries compiled from Uniprot and GenBank. Both gene sets are consolidated to generate a core set of protein sequences used as queries.

### Identification of Malacostraca innate immunity genes

First, we use CD-HIT (<https://github.com/weizhongli/cdhit>) to generate 'non-redundant' datasets. CD-HIT

was used to collapse contigs that have at least 95% identity in order to merge potential splice variants. To generate a set of malacostracan immunity gene orthologs, we used multiple Basic Local Alignment Search Tool (BLAST)-based approaches such as PSI-BLAST, BLASTp and tBLASTn with varying Blocks Substitution matrices. This list was filtered by e-value of less than  $10^{-6}$  and best reciprocal BLAST hits against the GenBank non-redundant (nr) database. We then filtered the best reciprocal nr BLAST hits by the presence of conserved domains reported to be essential for function and tree-based approaches to compile a final non-redundant list of Malacostraca innate immunity orthologs. Fasta files for all sequences are available as Additional file 25.

### Identification of conserved domains and phylogenetic tree construction

Malacostracan transcripts were in silico translated according to the longest open-reading frames into protein sequences. Conserved domains of the malacostracan hits were annotated using the Batch CD-Search Tool by NCBI (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>). Hits without essential domains were discarded. Multiple sequence alignments of protein sequences were performed using MAFFT [251]. Phylogenetic trees were built from the MAFFT alignments using the WAG + G model in RAxML [252] to generate best-scoring maximum likelihood trees. Multiple sequence alignment images were generated using the Geneious programme [253]. Graphical representations of Newick trees were also generated using Geneious.

### Identification of novel genes with putative immune function in Malacostraca

We previously performed orthologous group analyses in a separate study using complete arthropod genomes, which included the genome of the Malacostraca *P. hawaiiensis* [27]. We retrieved a list containing 750 protein sequences that were found only in *P. hawaiiensis* and have no significant blast hits to any other sequences in the NCBI nr database [27]. We performed a scan for the presence of Pfam domains [202] using HMMER (<http://hmmer.org/>) [254] on these 750 sequences and identified 82 genes containing Pfam domains. Further examination of predicted Pfam domains revealed four genes with domains suggestive of immune function: 1) chitin binding peritrophin-A domain pfam01607, 2) death domain pfam00531, 3) von Willebrand factor type A domain pfam05762 and 4) ML domain pfam02221 (Additional file 18: Table S9). These four genes were used as query sequences for Blast against the other malacostracan transcriptomes. We scanned putative



malacostracan orthologues for transmembrane domains and signal peptides using the Phobius tool [255, 256].

### RNA extraction and reverse transcriptase-polymerase chain reaction (RT-PCR) of novel Malacostraca immunity genes in *Parhyale hawaiiensis*

Five types of *P. hawaiiensis* tissue samples were collected from: 1) 100 mixed-stage embryos, 2) amputated limb fragments from 15 adult animals, 3) dissected gut tissues from 20 adult animals, 4) hemolymph from 50 adult animals and 5) adult whole tissues from two males and two females. Embryos were dissected from gravid females and rinsed with molecular grade water. Prior to tissue collection from the adults, animals were washed with filtered artificial seawater followed by treating with a mixture of clove oil (Sigma) and milliQ water (1:5000 dilution) for anaesthetization. As soon as the animals stopped moving after several minutes, the clove oil mixture was rinsed off and the anaesthetized animals were rinsed with molecular grade water. Limb fragments were dissected using spring scissors (Fine Science Tools). Gut tissues were collected using a scalpel and fine forceps. Hemolymph samples were collected by allowing the animals to bleed out in molecular grade water. Immediately after collection, 1 mL of Trizol reagent (Thermo Fisher Scientific) was added to the tissue samples in eppendorf tubes and samples were then snap frozen on dry ice. RNA extractions were performed according to the Trizol manufacturer's instructions. Concentrations of total RNA extracts were quantified using Qubit and Nanodrop. One microgram of total RNA from each tissue type was used for cDNA synthesis using the Qiagen QuantiTect Reverse Transcription Kit according to manufacturer's instructions. PCR on each gene was performed using Phusion High-fidelity polymerase (Thermo Fisher Scientific) and the following program was used for thermal cycling: 98C 30s, followed by 25 cycles of 98C 10s, 62C 30s, 72C 45 s, and then 72C 5 m. PCR products were ran on a 1% agarose gel and stained with SYBR Safe (Thermo Fisher Scientific). Primer sequences used for PCR are listed in Additional file 19: Table S10.

### Additional files

**Additional file 1: Figure S1.** Phylogenetic relationship of Malacostraca. Malacostraca is shown within the Pancrustacea clade. Malacostraca tree is adapted from Melands and Willassen 2007. Decapod phylogeny is adapted from Scholtz and Richter 1995 and Schram [43]. Representative species are shown at each branch. Species denoted in purple are edible food crops. (PDF 153 kb)

**Additional file 2: Table S1.** A complete list of all 69 malacostracan transcriptome datasets used in this study along with information on tissue types, developmental stages, accession IDs and total number of transcripts for each transcriptome. (PDF 65 kb)

**Additional file 3: Table S2.** List of species used in graphs along with their corresponding number IDs. (PDF 46 kb)

**Additional file 4: Table S3.** Malacostracans pattern recognition receptors and proPOs. (PDF 639 kb)

**Additional file 5: Figure S2.** Multiple sequence alignment of the  $\beta$ -glucanase domains of Gram negative binding proteins of malacostracans together with the  $\beta$ -glucanase protein from *Bombyx mori* (NP\_001159614.1). Two conserved Glu active site residues are labeled as E188 and E193 based on positions in the *B. mori* protein. (PDF 924 kb)

**Additional file 6: Figure S3.** Thioester-containing protein (TEPs) family. (A) Phylogenetic tree of the TEP family is constructed using the maximum-likelihood method from an amino acid multiple sequence alignment. Amino acid sequences include the macroglobulin complement related proteins, the vertebrate C3, C4 and C5 complement factors, arthropod TEPs and the alpha-2 macroglobulin family. Taxa labels are depicted as their respective colour codes. Bootstrap support values ( $n=1000$ ) for all trees can be found in Supplementary figure 8. Scale bar represents substitution per site. (B) Graph of putative TEP family transcripts. The y-axis represent total number of genes identified in all 55 malacostracan species for each family. Each species is represented by a number on the X-axis and a complete list of species is available in Additional file 3: Table S2. Black horizontal bars below each graph delimit the five orders of malacostracans and the numbers in parentheses (x/y) represent the following: x = number of species in which a particular gene family is found and y = total number of species in each order. (PDF 1763 kb)

**Additional file 7: Figure S4.** Prophenoloxidase activation system. (A) Phylogenetic tree of prophenoloxidase (proPO) is constructed using the maximum-likelihood method from an amino acid multiple sequence alignment. Taxa labels are depicted as their respective colour codes. Bootstrap support values ( $n=1000$ ) for all trees can be found in Additional file 14: Figure S8. Scale bar represents substitution per site. The graphs represent the total number of (B) hemocyanin and (C) CLIP-domain serine protease transcripts in malacostracans. The y-axis represent total number of genes identified in all 55 malacostracan species for each family. Each species is represented by a number on the X-axis and a complete list of species is available in Additional file 3: Table S2. Black horizontal bars below each graph delimit the five orders of malacostracans and the numbers in parentheses (x/y) represent the following: x = number of species in which a particular gene family is found and y = total number of species in each order. (PDF 579 kb)

**Additional file 8: Figure S5.** Malacostracans Toll-like receptors (TLRs). (A) Phylogenetic tree of TLRs is constructed using the maximum-likelihood method from an amino acid multiple sequence alignment of toll-IL-1 receptor (TIR) domains. Taxa labels are depicted as their respective colour codes. Bootstrap support values ( $n=1000$ ) for all trees can be found in Additional file 14: Figure S8. Scale bar represents substitution per site. (B) Graph of putative TLR transcripts. The y-axis represent total number of genes identified in all 55 malacostracan species for each family. Each species is represented by a number on the X-axis and a complete list of species is available in Additional file 3: Table S2. Black horizontal bars below each graph delimit the five orders of malacostracans and the numbers in parentheses (x/y) represent the following: x = number of species in which a particular gene family is found and y = total number of species in each order. (PDF 1124 kb)

**Additional file 9: Table S4.** Malacostracans Toll pathway components. (PDF 474 kb)

**Additional file 10: Figure S6.** Additional members of the IMD pathway. Phylogenetic trees of (A) PGRP, (B) I $\kappa$ B kinase  $\beta$  (IKK $\beta$ ); also known as immune response deficient-5 (IRD-5) and (C) TAK1 are constructed using the maximum-likelihood method from an amino acid multiple sequence alignment. Taxa labels are depicted as their respective colour codes. Bootstrap support values ( $n=1000$ ) for all trees can be found in Additional file 14: Figure S8. Scale bar represents substitution per site. (D) Multiple sequence alignment of four PGRPs identified in malacostracans. Conserved amidase catalytic residues are highlighted in red boxes. (PDF 1151 kb)

**Additional file 11: Table S5.** Malacostracans Imd pathway components. (PDF 351 kb)



**Additional file 12: Figure S7.** JAK-STAT pathway members in malacostracans. (A) Activation of the JAK-STAT signalling occurs through the binding of ligands such as cytokines to the cytokine receptor Domeless. Conserved protein domains of Domeless are shown in the figure inset. This binding activates the phosphorylation of the Janus kinase (JAK) proteins, which creates docking sites for Signal Transducer and Activator of Transcription (STAT) proteins through their Src Homology 2 (SH2) domains. STATs are phosphorylated by JAKs and activated STATs dimerise and are translocated to the nucleus to induce transcription. JAK-STAT transduction is controlled by suppressors of cytokine signalling (SOCS) and protein inhibitors of activated STAT (PIAS). SOCS proteins inhibit STATs phosphorylation via two mechanisms; 1) by competing with STATs for phosphotyrosine binding sites on cytokine receptors and (2) by binding to JAKs and preventing the recruitment of STATs onto the Domeless receptor. PIAS, also known as the E3 SUMO-protein ligase PIAS, is a transcriptional co-regulator that has the ability to inhibit STAT function. Phylogenetic trees of (B) JAK, (C) STAT and (D) PIAS are constructed using the maximum-likelihood method from an amino acid multiple sequence alignment. Two groups of PIAS proteins have been identified in malacostracans. Taxa labels are depicted as their respective colour codes. Bootstrap support values ( $n=1000$ ) for all trees can be found in Additional file 14: Figure S8. Scale bar represents substitution per site. (PDF 972 kb)

**Additional file 13: Table 6.** Malacostracans JAK-STAT pathway components. (PDF 228 kb)

**Additional file 14: Figure S8.** Suppressor of cytokine signalling (SOCS) gene family in malacostracans. Phylogenetic tree of the SOCS gene family is constructed using the maximum-likelihood method from an amino acid multiple sequence alignment. Seven main groups of SOCS proteins are identified in malacostracans. Taxa labels are depicted as their respective colour codes. Bootstrap support values ( $n=1000$ ) for all trees can be found in Additional file 14: Figure S8. Scale bar represents substitution per site. (PDF 1157 kb)

**Additional file 15: Table S7.** Malacostracans antimicrobial peptides. (PDF 107 kb)

**Additional file 16: Table S8.** Malacostracans antiviral RNAi pathway genes. (PDF 378 kb)

**Additional file 17: Figure S9.** *Argonaute* and *Piwi* gene families in malacostracans. Phylogenetic tree of the *Argonaute* and *Piwi* family is constructed using the maximum-likelihood method from an amino acid multiple sequence alignment. Seven main groups of SOCS proteins are identified in malacostracans. Taxa labels are depicted as their respective colour codes. Bootstrap support values ( $n=1000$ ) for all trees can be found in Additional file 14: Figure S8. Scale bar represents substitution per site. (PDF 1315 kb)

**Additional file 18: Table S9.** Four novel malacostracan gene families. (PDF 231 kb)

**Additional file 19: Table S10.** Primer sequences used for RT-PCR of novel malacostracan genes in *Parhyale hawaiensis*. (PDF 33 kb)

**Additional file 20: Figure S10.** Novel gene family with potential innate immunity function characterised by the chitin binding peritrophin-A domain (pfam01607) found only in malacostracans. (A) Phylogenetic tree of proteins containing the chitin binding peritrophin-A domain is constructed using the maximum-likelihood method from an amino acid multiple sequence alignment. Taxa labels are depicted as their respective colour codes. Node labels represent bootstrap support values from 1000 replicates. Scale bar represents substitution per site. (B) Multiple sequence alignment of peritrophin-A proteins showing the conserved peritrophin-A domain marked with a blue box, conserved cysteine residues marked in red boxes and conserved aromatic amino acids indicated by blue asterisks. The yellow shaded box represents a group of proteins with additional isopod-specific sequences. (PDF 5816 kb)

**Additional file 21: Figure S11.** Novel gene family with potential innate immunity function characterised by the death domain (pfam00531) found only in malacostracans. (A) Phylogenetic tree of proteins containing the death domain is constructed using the maximum-likelihood method from an amino acid multiple sequence alignment.

Taxa labels are depicted as their respective colour codes. Node labels represent bootstrap support values from 1000 replicates. Scale bar represents substitution per site. (B) Multiple sequence alignment of the death domain family showing the death domain marked with a red box and a novel gene-family-specific motif marked with a blue box. (PDF 3603 kb)

**Additional file 22: Figure S12.** Novel gene family with potential innate immunity function characterised by the ML domain (pfam02221) found only in malacostracans. (A) Phylogenetic tree of proteins containing the ML domain is constructed using the maximum-likelihood method from an amino acid multiple sequence alignment. Taxa labels are depicted as their respective colour codes. Node labels represent bootstrap support values from 1000 replicates. Scale bar represents substitution per site. (B) Multiple sequence alignment of the ML family showing the ML domain characterised by six conserved cysteine residues marked with red asterisks. Transmembrane domains were predicted using the Phobius program (Käll et al., [254]; Käll et al., [255]) and are annotated with a pink box. (PDF 3188 kb)

**Additional file 23: Figure S13.** Novel gene family with potential innate immunity function characterised by the Von Willebrand factor type A domain (pfam05762) found only in malacostracans. (A) Phylogenetic tree of proteins containing the Von Willebrand factor type A domain is constructed using the maximum-likelihood method from an amino acid multiple sequence alignment. Taxa labels are depicted as their respective colour codes. Node labels represent bootstrap support values from 1000 replicates. Scale bar represents substitution per site. (B) Multiple sequence alignment of the Von Willebrand factor type A domain from this gene family. (PDF 5243 kb)

**Additional file 24:** Supplementary references (DOCX 31 kb)

**Additional file 25:** Fasta files of all sequences. (ZIP 3690 kb)

**Additional file 26: Figure S14.** Phylogenetic trees presented in this study with node labels representing bootstrap support values from 1000 replicates. Scale bars represent substitution per site. (PDF 11640 kb)

**Additional file 24:** Supplementary references (DOCX 31 kb)

**Additional file 25:** Fasta files of all sequences. (ZIP 3690 kb)

**Additional file 26: Figure S14.** Phylogenetic trees presented in this study with node labels representing bootstrap support values from 1000 replicates. Scale bars represent substitution per site. (PDF 11640 kb)

## Abbreviations

ALF: Anti-lipoplysaccharide factor; AMP: Antimicrobial peptide; AS: Alternative splicing; BLAST: Basic Local Alignment Search Tool; CTL: C-type lectin; DD: Death domain; DREDD: Death-related ced-3/Nedd2-like; DSCAM: Down syndrome cell adhesion molecule; FADD: FAS-associated death domain; Fn: Fibronectin; GNBP: Gram-negative binding protein; HMM: Hidden Markov model; Ig: Immunoglobulin; IKK $\beta$ : I $\kappa$ B kinase  $\beta$ ; IKK $\gamma$ : I $\kappa$ B kinase  $\gamma$ ; Imd: Immune deficiency; JAK-STAT: Janus Kinase - signal transducer and activators of transcription; LDL: Low-density lipoprotein; LPS: Lipopolysaccharide; LTA: Lipoteichoic acid; MCR: Macroglobulin complement related; miRNA: microRNA; ML: MD-2-related lipid recognition; MyD88: Myeloid differentiation factor 88; PAMP: Pathogen-associated molecular pattern; Pasha: Partner of Drosha; PGN: Peptidoglycan; PGRP: Peptidoglycan recognition protein; PIAS: Protein inhibitors of activated STAT; PO: Phenoloxidase; proPO: Prophenoloxidase; PRR: Pattern recognition receptor; RNAi: RNA interference; SCRA: Class A scavenger receptor; SCRB: Class B scavenger receptor; siRNA: Small-interfering RNA; SOCS: Suppressor of cytokine signalling; SR: Scavenger receptor; TAB2: TAK1-binding protein; TAK1: MAPKKK transforming growth factor  $\beta$  (TGF $\beta$ )-activated kinase 1; TEP: Thioester-containing protein; TLR: Toll-like receptor; TNFR1: Tumor necrosis factor receptor-1; TOLLIP: Toll-interacting protein; TRBP: Trans-activating response (TAR) RNA-binding protein; VWA: Von Willebrand factor type A; VWF: Von Willebrand factor; WSSV: White spot syndrome virus;  $\alpha$ 2M:  $\alpha$ 2-macroglobulin

**Acknowledgements**

We would like to thank Alessia Di Donfrancesco for her help on *P. hawaiiensis* RNA extractions.

**Funding**

This work was supported by the Biotechnology and Biological Sciences Research Council (grant number BBK0075641 to A.A.A.); Human Frontier Science Program Postdoctoral Fellowship (to A.G.L.) and the Elizabeth Hannah Jenkinson Fund (to A.G.L.). The funding bodies have no roles in design of the study, collection, analysis and interpretation of the data and in writing the manuscript.

**Availability of data and material**

The datasets supporting the conclusions of this article are included within the article (and its additional files).

**Authors' contributions**

AGL and AAA conceived and designed the study. AGL performed the experiments and data analyses. AGL and AAA finalised the data, wrote the manuscript and approved the final manuscript.

**Competing interests**

Not applicable.

**Consent for publication**

Not applicable.

**Ethics approval and consent to participate**

Not applicable. The crustacean *Parhyale hawaiiensis* was used in this study. This species is highly abundant in the wild and laboratory strains have been established. No ethics approval is required.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 31 January 2017 Accepted: 7 May 2017

Published online: 18 May 2017

**References**

- Fischer G. World food and agriculture to 2030/50. Tech. Pap. from Expert Meet. How to Feed World. 2009. p. 24–6.
- Kearney J. Food consumption trends and drivers. *Philos Trans R Soc Lond B Biol Sci.* 2010;365:2793–807.
- VanHuis A, Van Itterbeeck J, Klunder H, Mertens E, Halloran A, Muir G, et al. Edible insects: future prospects for food and feed security. *FAO*; 2013
- Stentiford GD, Neil DM, Peeler EJ, Shields JD, Small HJ, Flegel TW, et al. Disease will limit future food supply from the global crustacean fishery and aquaculture sectors. *J Invertebr Pathol.* 2012;110:141–57.
- Stentiford GD. Diseases in aquatic crustaceans: problems and solutions for global food security. *J Invertebr Pathol.* 2012;110:139.
- Bondad-Reantaso MG, Subasinghe RP, Josupeit H, Cai J, Zhou X. The role of crustacean fisheries and aquaculture in global food security: Past, present and future. *J Invertebr Pathol.* 2012;110:158–65.
- Flegel TW. Historic emergence, impact and current status of shrimp pathogens in Asia. *J Invertebr Pathol.* 2012;110:166–73.
- Lightner DV, Redman RM, Pantoja CR, Tang KFJ, Noble BL, Schofield P, et al. Historic emergence, impact and current status of shrimp pathogens in the Americas. *J Invertebr Pathol.* 2012;110:174–83.
- Shields JD. The impact of pathogens on exploited populations of decapod crustaceans. *J Invertebr Pathol.* 2012;110:211–24.
- Flegel TW. Major viral diseases of the black tiger prawn (*Penaeus monodon*) in Thailand. *World J. Microbiol. Biotechnol.* Springer; 1997;13:433–42
- Prapavorarat A, Pongsomboon S, Tassanakajon A. Identification of genes expressed in response to yellow head virus infection in the black tiger shrimp, *Penaeus monodon*, by suppression subtractive hybridization. *Dev Comp Immunol.* 2010;34:611–7.
- Junkunlo K, Prachumwat A, Tangprasittipap A, Senapin S, Borwornpinyo S, Flegel TW, et al. A novel lectin domain-containing protein (LvCTL) associated with response of the whiteleg shrimp *Penaeus (Litopenaeus) vannamei* to yellow head virus (YHV). *Dev Comp Immunol.* 2012;37:334–41.
- Lightner D, Redman R, Hasson K, Pantoja C. Taura syndrome in *Penaeus vannamei* (Crustacea:Decapoda): gross signs, histopathology and ultrastructure. *Dis Aquat Organ.* 1995;21:53–9.
- Stentiford GD, Bonami J-R, Alday-Sanz V. A critical review of susceptibility of crustaceans to Taura syndrome, Yellowhead disease and White Spot Disease and implications of inclusion of these diseases in European legislation. *Aquaculture.* 2009;291:1–17.
- Unestam T. Fungal diseases of Crustacea. *Rev Med Vet Mycol.* 1973;8:1–20.
- Fisher WS, Wickham DE. Egg mortalities in wild populations of the dungeness crab in central and northern California. *Fish. Bull. NMFS/NOAA.* 1977;75:235–7
- Armstrong DA, Buchanan DV, Caldwell RS. A mycosis caused by *Lagenidium* sp. in laboratory-reared larvae of the Dungeness crab, *Cancer magister*, and possible chemical treatments. *J Invertebr Pathol.* 1976;28:329–36.
- Stentiford GD, Evans M, Bateman K, Feist SW. Co-infection by a yeast-like organism in *Hematodinium*-infected European edible crabs *Cancer pagurus* and velvet swimming crabs *Necora puber* from the English Channel. *Dis Aquat Organ.* 2003;54:195–202.
- Behringer DC, Butler MJ, Shields JD, Moss J. Review of *Panulirus argus* virus 1—a decade after its discovery. *Dis Aquat Organ.* 2011;94:153–60.
- Wang W. Bacterial diseases of crabs: a review. *J Invertebr Pathol.* 2011;106:18–26.
- Capone DG, Weston DP, Miller V, Shoemaker C. Antibacterial residues in marine sediments and invertebrates following chemotherapy in aquaculture. *Aquaculture.* 1996;145:55–75.
- Collier LM, Pinn EH. An assessment of the acute impact of the sea lice treatment ivermectin on a benthic community. *J Exp Mar Bio Ecol.* 1998; 230:131–47.
- Grant A, Briggs AD. Use of ivermectin in marine fish farms: Some concerns. *Mar. Pollut. Bull. Marine Pollution Bulletin*; 1998;36:566–8
- Hauton C. The scope of the crustacean immune system for disease contro. *J Invertebr Pathol.* 2012;110:251–60.
- McTaggart SJ, Conlon C, Colbourne JK, Blaxter ML, Little TJ. The components of the *Daphnia pulex* immune system as revealed by complete genome sequencing. *BMC Genomics.* 2009;10:175.
- Palmer WJ, Jiggins FM. Comparative genomics reveals the origins and diversity of arthropod immune systems. *Mol Biol Evol.* 2015;32:2111–29.
- Kao D, Lai AG, Stamatakis E, Rosic S, Konstantinides N, Jarvis E, et al. The genome of the crustacean *Parhyale hawaiiensis*, a model for animal development, regeneration, immunity and lignocellulose digestion. *Elife.* 2016;5:1.
- Pisani D, Poling LL, Lyons-Weiler M, Hedges SB. The colonization of land by animals: molecular phylogeny and divergence times among arthropods. *BMC Biol.* 2004;2:1.
- Rota-Stabelli O, Daley AC, Pisani D. Molecular timetrees reveal a Cambrian colonization of land and a New scenario for ecdysozoan evolution. *Curr Biol.* 2013;23:392–8.
- Hagner-Holler S, Kusche K, Hembach A, Burmester T. Biochemical and molecular characterisation of hemocyanin from the amphipod *Gammarus roeselii*: complex pattern of hemocyanin subunit evolution in Crustacea. *J Comp Physiol B.* 2005;175:445–52.
- Christie AE. Identification of the first neuropeptides from the Amphipoda (Arthropoda, Crustacea). *Gen Comp Endocrinol.* 2014;206:96–110.
- Ghaffari N, Sanchez-Flores A, Doan R, Garcia-Orozco KD, Chen PL, Ochoa-Leyva A, et al. Novel transcriptome assembly and improved annotation of the whiteleg shrimp (*Litopenaeus vannamei*), a dominant crustacean in global seafood mariculture. *Sci Rep.* 2014;4:7081.
- Sun S, Xuan F, Fu H, Zhu J, Ge X, Gu Z. Transcriptomic and histological analysis of hepatopancreas, muscle and gill tissues of oriental river prawn (*Macrobrachium nipponense*) in response to chronic hypoxia. *BMC Genomics.* 2015;16:491.
- Li Y, Hui M, Cui Z, Liu Y, Song C, Shi G. Comparative transcriptomic analysis provides insights into the molecular basis of the metamorphosis and nutrition metabolism change from zoeae to megalopae in *Eriocheir sinensis*. *Comp. Biochem. Physiol. Part D Genomics Proteomics.* Elsevier Inc.; 2015;13:1–9.
- Meyer B, Martini P, Biscontin A, De Pittà C, Romualdi C, Teschke M, et al. Pyrosequencing and de novo assembly of Antarctic krill (*Euphausia superba*) transcriptome to study the adaptability of krill to climate-induced environmental changes. *Mol Ecol Resour.* 2015;15:1460–71.
- Stahl BA, Gross JB, Speiser DJ, Oakley TH, Patel NH, Gould DB, et al. A Transcriptomic Analysis of Cave, Surface, and Hybrid Isopod Crustaceans of the Species *Asellus aquaticus*. *Réaux S, editor. PLoS One.* 2015;10:e0140484.

37. Verbruggen B, Bickley LK, Santos EM, Tyler CR, Stentiford GD, Bateman KS, et al. De novo assembly of the *Carcinus maenas* transcriptome and characterization of innate immune system pathways. *BMC Genomics*. 2015;16:458.
38. Truebano M, Tills O, Spicer JI. Embryonic transcriptome of the brackishwater amphipod *Gammarus chevreuxi*. *Mar. Genomics*. Elsevier B.V.; 2016;28:5–6.
39. Kim H-S, Hwang D-S, Lee B-Y, Park JC, Lee YH, Lee J-S. De novo assembly and annotation of the marine mysid (*Neomysis awatschensis*) transcriptome. *Mar. Genomics*. Elsevier B.V.; 2016;28:41–3.
40. Havird JC, Mitchell RT, Henry RP, Santos SR. Salinity-induced changes in gene expression from anterior and posterior gills of *Callinectes sapidus* (Crustacea: Portunidae) with implications for crustacean ecological genomics. *Comp. Biochem. Physiol. Part D Genomics Proteomics*. Elsevier Inc.; 2016;19:34–44.
41. Tan MH, Gan HM, Gan HY, Lee YP, Croft LJ, Schultz MB, et al. First comprehensive multi-tissue transcriptome of *Cherax quadricarinatus* (Decapoda: Parastacidae) reveals unexpected diversity of endogenous cellulase. *Org Divers Evol*. 2016;16:185–200.
42. Scholtz G, Richter S. Phylogenetic systematics of the reptantian Decapoda (Crustacea, Malacostraca). *Zool J Linn Soc*. 1995;113:289–328.
43. Schram FR. Phylogeny of decapods: moving towards a consensus. *Adv. Decapod Crustac. Res.* Dordrecht: Springer Netherlands; 2001. p. 1–20.
44. Meland K, Willassen E. The disunity of "Mysidacea" (Crustacea). *Mol Phylogenet Evol*. 2007;44:1083–104.
45. De Gregorio E, Spellman PT, Tzou P, Rubin GM, Lemaître B. The Toll and Imd pathways are the major regulators of the immune response in *Drosophila*. *EMBO J*. 2002;21:2568–79.
46. Hoffmann JA. The immune response of *Drosophila*. *Nature*. 2003;426:33–8.
47. Meister M. Blood cells of *Drosophila*: cell lineages and role in host defence. *Curr Opin Immunol*. 2004;16:10–5.
48. Govind S, Nehm RH. Innate immunity in fruit flies: a textbook example of genomic recycling. *PLoS Biol*. 2004;2:E276.
49. Theopold U, Schmidt O, Söderhäll K, Dushay MS. Coagulation in arthropods: defence, wound closure and healing. *Trends Immunol*. 2004;25:289–94.
50. Brennan CA, Anderson KV. *Drosophila*: the genetics of innate immune recognition and response. *Annu. Rev. Immunol*. 2004;22:457–83.
51. Janeway CA. Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb. Symp. Quant. Biol*. 1989;54 Pt 1:1–13.
52. Beutler B. Innate immunity: an overview. *Mol Immunol*. 2004;40:845–59.
53. Ferrandon D, Imler J-L, Hetru C, Hoffmann JA. The *Drosophila* systemic immune response: sensing and signalling during bacterial and fungal infections. *Nat Rev Immunol*. 2007;7:862–74.
54. Ochiai M, Ashida M. Purification of a beta-1,3-glucan recognition protein in the prophenoloxidase activating system from hemolymph of the silkworm, *Bombyx mori*. *J Biol Chem*. 1988;263:12056–62.
55. Ochiai M, Niki T, Ashida M. Immunocytochemical localization of beta-1,3-glucan recognition protein in the silkworm, *Bombyx mori*. *Cell Tissue Res*. 1992;268:431–7.
56. Ochiai M, Ashida M. A pattern-recognition protein for beta-1,3-glucan. The binding domain and the cDNA cloning of beta-1,3-glucan recognition protein from the silkworm, *Bombyx mori*. *J Biol Chem*. 2000;275:4995–5002.
57. Lee WJ, Lee JD, Kravchenko VV, Ulevitch RJ, Brey PT. Purification and molecular cloning of an inducible gram-negative bacteria-binding protein from the silkworm, *Bombyx mori*. *Proc Natl Acad Sci U S A*. 1996;93:7888–93.
58. Zhang R, Cho HY, Kim HS, Ma YG, Osaki T, Kawabata S, et al. 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*. 2003;278:42072–9.
59. Pauchet Y, Freitak D, Heidel-Fischer HM, Heckel DG, Vogel H. Immunity or digestion: glucanase activity in a glucan-binding protein family from Lepidoptera. *J Biol Chem*. 2009;284:2214–24.
60. Genta FA, Bragatto I, Terra WR, Ferreira C. Purification, characterization and sequencing of the major beta-1,3-glucanase from the midgut of *Tenebrio molitor* larvae. *Insect Biochem Mol Biol*. 2009;39:861–74.
61. Bragatto I, Genta FA, Ribeiro AF, Terra WR, Ferreira C. Characterization of a beta-1,3-glucanase active in the alkaline midgut of *Spodoptera frugiperda* larvae and its relation to beta-glucan-binding proteins. *Insect Biochem Mol Biol*. 2010;40:861–72.
62. Roux MM, Pain A, Klimpel KR, Dhar AK. The lipopolysaccharide and beta-1,3-glucan binding protein gene is upregulated in white spot virus-infected shrimp (*Penaeus stylirostris*). *J Virol*. 2002;76:7140–9.
63. Sritunyalucksana K, Lee SY, Söderhäll K. A beta-1,3-glucan binding protein from the black tiger shrimp, *Penaeus monodon*. *Dev. Comp. Immunol*. 2002;26:237–45.
64. Cheng W, Liu C-H, Tsai C-H, Chen J-C. Molecular cloning and characterization of a pattern recognition molecule, lipopolysaccharide- and beta-1,3-glucan binding protein (LGBP) from the white shrimp *Litopenaeus vannamei*. *Fish Shellfish Immunol*. 2005;18:297–310.
65. Du X-J, Zhao X, Wang J-X. Molecular cloning and characterization of a lipopolysaccharide and beta-1,3-glucan binding protein from fleshy prawn (*Fenneropenaeus chinensis*). *Mol Immunol*. 2007;44:1085–94.
66. Lin Y-C, Vaseeharan B, Chen J-C. Identification and phylogenetic analysis on lipopolysaccharide and beta-1,3-glucan binding protein (LGBP) of kuruma shrimp *Marsupenaeus japonicus*. *Dev Comp Immunol*. 2008;32:1260–9.
67. Liu F, Li F, Dong B, Wang X, Xiang J. Molecular cloning and characterization of a pattern recognition protein, lipopolysaccharide and beta-1,3-glucan binding protein (LGBP) from Chinese shrimp *Fenneropenaeus chinensis*. *Mol Biol Rep*. 2009;36:471–7.
68. Hughes AL. Evolution of the betaGRP/GNBP/beta-1,3-glucanase family of insects. *Immunogenetics*. 2012;64:549–58.
69. Shapiro L, Love J, Colman DR. Adhesion molecules in the nervous system: structural insights into function and diversity. *Annu Rev Neurosci*. 2007;30:451–74.
70. Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, et al. *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell*. 2000;101:671–84.
71. Graveley BR, Kaur A, Gunning D, Zipursky SL, Rowen L, Clemens JC. The organization and evolution of the dipteran and hymenopteran Down syndrome cell adhesion molecule (Dscam) genes. *RNA*. 2004;10:1499–506.
72. Dong Y, Taylor HE, Dimopoulos G. AgDscam, a hypervariable immunoglobulin domain-containing receptor of the Anopheles gambiae innate immune system. *PLoS Biol*. 2006;4:e229. Public Library of Science.
73. Brites D, McTaggart S, Morris K, Anderson J, Thomas K, Colson I, et al. The Dscam homologue of the crustacean *Daphnia* is diversified by alternative splicing like in insects. *Mol Biol Evol*. 2008;25:1429–39. Oxford University Press.
74. Yue Y, Meng Y, Ma H, Hou S, Cao G, Hong W, et al. A large family of Dscam genes with tandemly arrayed 5'cassettes in Chelicerata. *Nat Commun*. 2016;7:11252.
75. Brown MS, Goldstein JL. Receptor-mediated endocytosis: insights from the lipoprotein receptor system. *Proc Natl Acad Sci U S A*. 1979;76:3330–7.
76. Krieger M. The other side of scavenger receptors: pattern recognition for host defense. *Curr Opin Lipidol*. 1997;8:275–80.
77. Medzhitov R, Janeway CA. Decoding the patterns of self and nonself by the innate immune system. *Science*. 2002;296:298–300.
78. Areschoug T, Gordon S. Scavenger receptors: role in innate immunity and microbial pathogenesis. *Cell Microbiol*. 2009;11:1160–9.
79. Canton J, Neculai D, Grinstein S. Scavenger receptors in homeostasis and immunity. *Nat Rev Immunol*. 2013;13:621–34. Nature Publishing Group.
80. Plüddemann A, Mukhopadhyay S, Gordon S. The interaction of macrophage receptors with bacterial ligands. *Expert Rev Mol Med*. 2006;8:1–25.
81. Plüddemann A, Mukhopadhyay S, Gordon S. Innate immunity to intracellular pathogens: macrophage receptors and responses to microbial entry. *Immunol Rev*. 2011;240:11–24.
82. Hampton RY, Golenbock DT, Penman M, Krieger M, Raetz CR. Recognition and plasma clearance of endotoxin by scavenger receptors. *Nature*. 1991;352:342–4.
83. Dunne DW, Resnick D, Greenberg J, Krieger M, Joiner KA. The type I macrophage scavenger receptor binds to gram-positive bacteria and recognizes lipoteichoic acid. *Proc Natl Acad Sci U S A*. 1994;91:1863–7.
84. Peiser L, De Winther MPJ, Makepeace K, Hollinshead M, Coull P, Plested J, et al. The class A macrophage scavenger receptor is a major pattern recognition receptor for *Neisseria meningitidis* which is independent of lipopolysaccharide and not required for secretory responses. *Infect Immun*. 2002;70:5346–54.
85. Mekata T, Okugawa S, Inada M, Yoshimine M, Nishi J, Kono T, et al. Class B scavenger receptor, Croquemort from kuruma shrimp *Marsupenaeus japonicus*: Molecular cloning and characterization. *Mol Cell Probes*. 2011;25:94–100.
86. Febbraio M, Hajjar DP, Silverstein RL. CD36: a class B scavenger receptor involved in angiogenesis, atherosclerosis, inflammation, and lipid metabolism. *J Clin Invest*. 2001;108:785–91.

87. Stuart LM, Deng J, Silver JM, Takahashi K, Tseng AA, Hennessy EJ, et al. Response to *Staphylococcus aureus* requires CD36-mediated phagocytosis triggered by the COOH-terminal cytoplasmic domain. *J Cell Biol.* 2005;170:477–85.
88. Kiefer C, Sumser E, Wernet MF, Von Lintig J. A class B scavenger receptor mediates the cellular uptake of carotenoids in *Drosophila*. *Proc Natl Acad Sci U S A.* 2002;99:10581–6. National Acad Sciences.
89. Franc NC, Dimarcq JL, Lagueux M, Hoffmann J, Ezekowitz RA. Croquemort, a novel *Drosophila* hemocyte/macrophage receptor that recognizes apoptotic cells. *Immunity.* 1996;4:431–43.
90. Rigotti A, Acton SL, Krieger M. The class B scavenger receptors SR-BI and CD36 are receptors for anionic phospholipids. *J Biol Chem.* 1995;270:16221–4.
91. Wang X-W, Wang J-X. Diversity and multiple functions of lectins in shrimp immunity. *Dev Comp Immunol.* 2013;39:27–38.
92. Wongpanya R, Sengprasert P, Amparyup P, Tassanakajon A. A novel C-type lectin in the black tiger shrimp *Penaeus monodon* functions as a pattern recognition receptor by binding and causing bacterial agglutination. *Fish Shellfish Immunol.* 2016;60:103–13.
93. Feng J, Huang X, Jin M, Zhang Y, Li T, Hui K, et al. A C-type lectin (MrLec) with high expression in intestine is involved in innate immune response of *Macrobrachium rosenbergii*. *Fish Shellfish Immunol.* 2016;59:345–50.
94. Drickamer K, Taylor ME. Biology of animal lectins. *Annu Rev Cell Biol.* 1993;9:237–64.
95. Drickamer K, Fadden AJ. Genomic analysis of C-type lectins. *Biochem. Soc. Symp.* 2002;59–72.
96. Cooper DN, Barondes SH. God must love galectins; he made so many of them. *Glycobiology.* 1999;9:979–84.
97. Ma THT, Tiu SHK, He J-G, Chan S-M. Molecular cloning of a C-type lectin (LvLT) from the shrimp *Litopenaeus vannamei*: early gene down-regulation after WSSV infection. *Fish Shellfish Immunol.* 2007;23:430–7.
98. Sun J, Wang L, Wang B, Guo Z, Liu M, Jiang K, et al. Purification and characterization of a natural lectin from the serum of the shrimp *Litopenaeus vannamei*. *Fish Shellfish Immunol.* 2007;23:292–9.
99. Zhang Y, Qiu L, Song L, Zhang H, Zhao J, Wang L, et al. Cloning and characterization of a novel C-type lectin gene from shrimp *Litopenaeus vannamei*. *Fish Shellfish Immunol.* 2009;26:183–92.
100. Zhao Z-Y, Yin Z-X, Xu X-P, Weng S-P, Rao X-Y, Dai Z-X, et al. A novel C-type lectin from the shrimp *Litopenaeus vannamei* possesses anti-white spot syndrome virus activity. *J Virol.* 2009;83:347–56.
101. Wei X, Liu X, Yang J, Fang J, Qiao H, Zhang Y, et al. Two C-type lectins from shrimp *Litopenaeus vannamei* that might be involved in immune response against bacteria and virus. *Fish Shellfish Immunol.* 2012;32:132–40.
102. Nonaka M, Yoshizaki F. Primitive complement system of invertebrates. *Immunol Rev.* 2004;198:203–15.
103. Blandin S, Shiao S-H, Moita LF, Janse CJ, Waters AP, Kafatos FC, et al. Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. *Cell.* 2004;116:661–70.
104. Stroschein-Stevenson SL, Foley E, O'Farrell PH, Johnson AD. Identification of *Drosophila* gene products required for phagocytosis of *Candida albicans*. *PLoS Biol.* 2006;4:e4.
105. Armstrong PB. Role of  $\alpha 2$ -macroglobulin in the immune responses of invertebrates. *Invertebr Surviv J.* 2010;165–80.
106. Söderhäll K, Cerenius L, Johansson MW. The prophenoloxidase activating system and its role in invertebrate defence. *Ann N Y Acad Sci.* 1994;712:155–61.
107. Söderhäll K, Cerenius L. Role of the prophenoloxidase-activating system in invertebrate immunity. *Curr Opin Immunol.* 1998;10:23–8.
108. Sritunyalucksana K, Söderhäll K. The proPO and clotting system in crustaceans. *Aquaculture.* 2000;191:53–69.
109. Cerenius L, Söderhäll K. The prophenoloxidase-activating system in invertebrates. *Immunol Rev.* 2004;198:116–26.
110. Aspán A, Söderhäll K. Purification of prophenoloxidase from crayfish blood cells, and its activation by an endogenous serine proteinase. *Insect Biochem.* 1991;21:363–73.
111. Ashida M, Yoshida H. Limited proteolysis of prophenoloxidase during activation by microbial products in insect plasma and effect of phenoloxidase on electrophoretic mobilities of plasma proteins. *Insect Biochem.* 1988;18:11–9.
112. Burmester T. Origin and evolution of arthropod hemocyanins and related proteins. *J Comp Physiol B.* 2002;172:95–107.
113. Decker H, Ryan M, Jaenicke E, Terwilliger N. SDS-induced phenoloxidase activity of hemocyanins from *Limulus polyphemus*, *Eurytelma californicum*, and *Cancer magister*. *J Biol Chem.* 2001;276:17796–9.
114. García-Carreño FL, Cota K, Navarrete Del Toro MA. Phenoloxidase activity of hemocyanin in whiteleg shrimp *Penaeus vannamei*: conversion, characterization of catalytic properties, and role in postmortem melanosis. *J Agric Food Chem.* 2008;56:6454–9.
115. Jaenicke E, Decker H. Functional changes in the family of type 3 copper proteins during evolution. *Chembiochem.* 2004;5:163–9.
116. Decker H, Hellmann N, Jaenicke E, Lieb B, Meissner U, Markl J. Minireview: Recent progress in hemocyanin research. *Integr Comp Biol.* 2007;47:631–44.
117. Hernández-López J, Gollas-Galván T, Vargas-Albores F. Activation of the prophenoloxidase system of the brown shrimp *Penaeus californiensis* Holmes). *Comp. Biochem. Physiol. Part C Pharmacol. Toxicol. Endocrinol.* 1996;113:61–6.
118. Lee SY, Lee BL, Söderhäll K. Processing of crayfish hemocyanin subunits into phenoloxidase. *Biochem Biophys Res Commun.* 2004;322:490–6.
119. Ko C-F, Chiou T-T, Vaseeharan B, Lu J-K, Chen J-C. Cloning and characterisation of a prophenoloxidase from the haemocytes of mud crab *Scylla serrata*. *Dev Comp Immunol.* 2007;31:12–22.
120. Cerenius L, Lee BL, Söderhäll K. The proPO-system: pros and cons for its role in invertebrate immunity. *Trends Immunol.* 2008;29:263–71.
121. Masuda T, Otomo R, Kuyama H, Momoji K, Tonomoto M, Sakai S, et al. A novel type of prophenoloxidase from the kuruma prawn *Marsupenaeus japonicus* contributes to the melanization of plasma in crustaceans. *Fish Shellfish Immunol.* 2012;32:61–8.
122. Amparyup P, Charoensapsri W, Tassanakajon A. Prophenoloxidase system and its role in shrimp immune responses against major pathogens. *Fish Shellfish Immunol.* 2013;34:990–1001.
123. Masuda T, Momoji K, Hirata T, Mikami B. The crystal structure of a crustacean prophenoloxidase provides a clue to understanding the functionality of the type 3 copper proteins. *FEBS J.* 2014;281:2659–73.
124. Pless DD, Aguilar MB, Falcón A, Lozano-Alvarez E, HeimerdelaCotera EP. Latent phenoloxidase activity and N-terminal amino acid sequence of hemocyanin from *Bathynomus giganteus*, a primitive crustacean. *Arch Biochem Biophys.* 2003;409:402–10.
125. Terwilliger NB, Ryan MC. Functional and phylogenetic analyses of phenoloxidases from brachyuran (*Cancer magister*) and branchiopod (*Artemia franciscana*, *Triops longicaudatus*) crustaceans. *Biol Bull MBL.* 2006;210:38–50.
126. Jaenicke E, Fraune S, May S, Irmak P, Augustin R, Meesters C, et al. Is activated hemocyanin instead of phenoloxidase involved in immune response in woodlice? *Dev Comp Immunol.* 2009;33:1055–63.
127. King AJ, Cragg SM, Li Y, Dymond J, Guille MJ, Bowles DJ, et al. Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes. *Proc Natl Acad Sci U S A.* 2010;107:5345–50. National Acad Sciences.
128. Kanost MR, Jiang H, Yu X-Q. Innate immune responses of a lepidopteran insect. *Manduca sexta*. *Immunol Rev.* 2004;198:97–105.
129. Tang H, Kambris Z, Lemaitre B, Hashimoto C. Two proteases defining a melanization cascade in the immune system of *Drosophila*. *J Biol Chem.* 2006;281:28097–104.
130. Smith CL, Delotto R. A common domain within the proenzyme regions of the *drosophila* snake and easter proteins and tachypleus proclotting enzyme defines a new subfamily of serine proteases. *Protein Sci.* 1992;1:1225–6.
131. Jiang H, Wang Y, Kanost MR. Pro-phenol oxidase activating proteinase from an insect, *Manduca sexta*: a bacteria-inducible protein similar to *Drosophila* easter. *Proc Natl Acad Sci U S A.* 1998;95:12220–5.
132. Jiang H, Kanost MR. The clip-domain family of serine proteinases in arthropods. *Insect Biochem Mol Biol.* 2000;30:95–105.
133. Waterhouse RM, Kriventseva E V, Meister S, Xi Z, Alvarez KS, Bartholomay LC, et al. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science. American Association for the Advancement of Science;* 2007;316:1738–43.
134. Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. *Cell.* 2006;124:783–801.
135. Pasare C, Medzhitov R. Toll-like receptors: linking innate and adaptive immunity. *Adv. Exp. Med. Biol. Boston, MA: Springer US;* 2005;560:11–8.
136. Leulier F, Lemaitre B. Toll-like receptors—taking an evolutionary approach. *Nat Rev Genet.* 2008;9:165–78.



137. Lemaitre B, Nicolas E, Michaut L, Reichhart JM, Hoffmann JA. The dorsoventral regulatory gene cassette *spätzle/Toll/cactus* controls the potent antifungal response in *Drosophila* adults. *Cell*. 1996;86:973–83.
138. Rutschmann S, Kilinc A, Ferrandon D. Cutting edge: the toll pathway is required for resistance to gram-positive bacterial infections in *Drosophila*. *J. Immunol. The Journal of {...}*; 2002;168:1542–6.
139. Shi X-Z, Zhang R-R, Jia Y-P, Zhao X, Yu X-Q, Wang J-X. Identification and molecular characterization of a *Spätzle*-like protein from Chinese shrimp (*Fenneropenaeus chinensis*). *Fish Shellfish Immunol*. 2009;27:610–7.
140. Wang P-H, Liang J-P, Gu Z-H, Wan D-H, Weng S-P, Yu X-Q, et al. Molecular cloning, characterization and expression analysis of two novel *Tolls* (*LvToll2* and *LvToll3*) and three putative *Spätzle*-like *Toll* ligands (*LvSpz1-3*) from *Litopenaeus vannamei*. *Dev Comp Immunol*. 2012;36:359–71.
141. Tauszig-Delamasure S, Bilak H, Capovilla M, Hoffmann JA, Imler J-L. *Drosophila* MyD88 is required for the response to fungal and Gram-positive bacterial infections. *Nat Immunol*. 2002;3:91–7.
142. Christophides GK, Zdobnov E, Barillas-Mury C, Birney E, Blandin S, Blass C, et al. Immunity-related genes and gene families in *Anopheles gambiae*. *Science*. 2002;298:159–65. American Association for the Advancement of Science.
143. Li C, Chen Y, Weng S, Li S, Zuo H, Yu X, et al. Presence of Tube isoforms in *Litopenaeus vannamei* suggests various regulatory patterns of signal transduction in invertebrate NF- $\kappa$ B pathway. *Dev Comp Immunol*. 2014;42:174–85.
144. Zhang G, Ghosh S. Negative regulation of toll-like receptor-mediated signaling by *Tollip*. *J Biol Chem*. 2002;277:7059–65.
145. Wang P-H, Gu Z-H, Wan D-H, Zhu W-B, Qiu W, Chen Y-G, et al. *Litopenaeus vannamei* *Toll*-interacting protein (*LvTollip*) is a potential negative regulator of the shrimp *Toll* pathway involved in the regulation of the shrimp antimicrobial peptide gene *penaeidin-4* (*PEN4*). *Dev Comp Immunol*. 2013;40:266–77.
146. Burns K, Clatworthy J, Martin L, Martinon F, Plumpton C, Maschera B, et al. *Tollip*, a new component of the IL-1RI pathway, links *IRAK* to the IL-1 receptor. *Nat Cell Biol*. 2000;2:346–51.
147. Bechsgaard J, Vanthournout B, Funch P, Vestbo S, Gibbs RA, Richards S, et al. Comparative genomic study of arachnid immune systems indicates loss of beta-1,3-glucanase-related proteins and the immune deficiency pathway. *J Evol Biol*. 2016;29:277–91.
148. Gottar M, Gobert V, Michel T, Belvin M, Duyk G, Hoffmann JA, et al. The *Drosophila* immune response against Gram-negative bacteria is mediated by a peptidoglycan recognition protein. *Nature*. 2002;416:640–4.
149. Kaneko T, Yano T, Aggarwal K, Lim J-H, Ueda K, Oshima Y, et al. *PGRP-LC* and *PGRP-LE* have essential yet distinct functions in the *drosophila* immune response to monomeric DAP-type peptidoglycan. *Nat Immunol*. 2006;7:715–23.
150. Cheng X, Zhang X, Pflugrath JW, Studier FW. The structure of bacteriophage T7 lysozyme, a zinc amidase and an inhibitor of T7 RNA polymerase. *Proc Natl Acad Sci U S A*. 1994;91:4034–8.
151. Mellroth P, Karlsson J, Steiner H. A scavenger function for a *Drosophila* peptidoglycan recognition protein. *J Biol Chem*. 2003;278:7059–64.
152. Kim M, Lee JH, Lee SY, Kim E, Chung J. Caspar, a suppressor of antibacterial immunity in *Drosophila*. *Proc Natl Acad Sci U S A*. 2006;103:16358–63.
153. Huang X-D, Yin Z-X, Liao J-X, Wang P-H, Yang L-S, Ai H-S, et al. Identification and functional study of a shrimp *Relish* homologue. *Fish Shellfish Immunol*. 2009;27:230–8.
154. Li F, Yan H, Wang D, Priya TAJ, Li S, Wang B, et al. Identification of a novel *relish* homolog in Chinese shrimp *Fenneropenaeus chinensis* and its function in regulating the transcription of antimicrobial peptides. *Dev Comp Immunol*. 2009;33:1093–101.
155. Wang P-H, Gu Z-H, Huang X-D, Liu B-D, Deng X, Ai H-S, et al. An immune deficiency homolog from the white shrimp, *Litopenaeus vannamei*, activates antimicrobial peptide genes. *Mol Immunol*. 2009;46:1897–904.
156. Li F, Xiang J. Signaling pathways regulating innate immune responses in shrimp. *Fish Shellfish Immunol*. 2013;34:973–80.
157. Shuai K, Liu B. Regulation of JAK-STAT signalling in the immune system. *Nat Rev Immunol*. 2003;3:900–11.
158. Arbouzova NI, Zeidler MP. JAK/STAT signalling in *Drosophila*: insights into conserved regulatory and cellular functions. *Development*. 2006;133:2605–16.
159. Li WX. Canonical and non-canonical JAK-STAT signaling. *Trends Cell Biol*. 2008;18:545–51.
160. Morin-Poulard I, Vincent A, Crozatier M. The *Drosophila* JAK-STAT pathway in blood cell formation and immunity. *JAK-STAT*. 2013;2:e25700.
161. Levy DE, Darnell JE. Stats: transcriptional control and biological impact. *Nat Rev Mol Cell Biol*. 2002;3:651–62.
162. Yoshimura A, Naka T, Kubo M. SOCS proteins, cytokine signalling and immune regulation. *Nat Rev Immunol*. 2007;7:454–65.
163. Grönholm J, Ungureanu D, Vanhatupa S, Rämetsä M, Silvennoinen O. Sumoylation of *Drosophila* transcription factor STAT92E. *J Innate Immun*. 2010;2:618–24.
164. Ilangumaran S, Ramanathan S, Rottapel R. Regulation of the immune system by SOCS family adaptor proteins. *Semin Immunol*. 2004;16:351–65.
165. DeZoysa M, Lee J. Suppressor of cytokine signaling 2 (SOCS-2) homologue in disk abalone: Cloning, sequence characterization and expression analysis. *Fish Shellfish Immunol*. 2009;26:500–8.
166. Zhang Y, Zhao J, Zhang H, Gai Y, Wang L, Li F, et al. The involvement of suppressors of cytokine signaling 2 (SOCS2) in immune defense responses of Chinese mitten crab *Eriocheir sinensis*. *Dev Comp Immunol*. 2010;34:42–8.
167. Yount NY, Bayer AS, Xiong YQ, Yeaman MR. Advances in antimicrobial peptide immunobiology. *Biopolymers*. 2006;84:435–58.
168. Guan-Guerra E, Santos-Mendoza T, Lugo-Reyes SO, Terán LM. Antimicrobial peptides: general overview and clinical implications in human health and disease. *Clin Immunol*. 2010;135:1–11.
169. Gueguen Y, Garnier J, Robert L, Lefranc M-P, Mougenot I, de Lorgeril J, et al. PenBase, the shrimp antimicrobial peptide *penaeidin* database: sequence-based classification and recommended nomenclature. *Dev Comp Immunol*. 2006;30:283–8.
170. Rosa RD, Barracco MA. Antimicrobial peptides in crustaceans. *Inv Surv J*. 2010;7:262–84.
171. Zhao XF, Wang JX. The antimicrobial peptides of the immune response of shrimp. *Inv Surv J*. 2008;5:162–79.
172. Smith VJ, Fernandes JMO, Kemp GD, Hauton C. Crustins: enigmatic WAP domain-containing antibacterial proteins from crustaceans. *Dev Comp Immunol*. 2008;32:758–72.
173. Gross PS, Bartlett TC, Browdy CL, Chapman RW, Warr GW. Immune gene discovery by expressed sequence tag analysis of hemocytes and hepatopancreas in the Pacific White Shrimp, *Litopenaeus vannamei*, and the Atlantic White Shrimp, *L. setiferus*. *Dev. Comp. Immunol*. 2001;25:565–77.
174. Supungul P, Klinbunga S, Pichyangkura R, Jitrapakdee S, Hirono I, Aoki T, et al. Identification of immune-related genes in hemocytes of black tiger shrimp (*Penaeus monodon*). *Mar. Biotechnol. (NY)*. Springer-Verlag; 2002;4:487–94.
175. Relf JM, Chisholm JR, Kemp GD, Smith VJ. Purification and characterization of a cysteine-rich 11.5-kDa antibacterial protein from the granular haemocytes of the shore crab, *Carcinus maenas*. *Eur J Biochem*. 1999;264:350–7.
176. Lecellier C-H, Voinnet O. RNA silencing: no mercy for viruses? *Immunol Rev*. 2004;198:285–303.
177. Cullen BR. Is RNA interference involved in intrinsic antiviral immunity in mammals? *Nat Immunol*. 2006;7:563–7.
178. Wang X-H, Aliyari R, Li W-X, Li H-W, Kim K, Carthew R, et al. RNA interference directs innate immunity against viruses in adult *Drosophila*. *Science*. 2006;312:452–4.
179. Zambon RA, Vakharia VN, Wu LP. RNAi is an antiviral immune response against a dsRNA virus in *Drosophila melanogaster*. *Cell Microbiol*. 2006;8:880–9. Blackwell Publishing Ltd.
180. Liu H, Söderhäll K, Jiravanichpaisal P. Antiviral immunity in crustaceans. *Fish Shellfish Immunol*. 2009;27:79–88.
181. Labrecque Y, Warr GW. Insights into the antiviral functions of the RNAi machinery in *penaeid* shrimp. *Fish Shellfish Immunol*. 2013;34:1002–10.
182. Robalino J, Bartlett T, Shepard E, Prior S, Jaramillo G, Scura E, et al. Double-stranded RNA induces sequence-specific antiviral silencing in addition to nonspecific immunity in a marine shrimp: convergence of RNA interference and innate immunity in the invertebrate antiviral response? *J Virol*. 2005;79:13561–71. American Society for Microbiology.
183. Yodmuang S, Tirasophon W, Roshorn Y, Chinnirunvong W, Panyim S. YHV-protease dsRNA inhibits YHV replication in *Penaeus monodon* and prevents mortality. *Biochem Biophys Res Commun*. 2006;341:351–6.
184. Kim CS, Kosuke Z, Nam YK, Kim SK, Kim KH. Protection of shrimp (*Penaeus chinensis*) against white spot syndrome virus (WSSV) challenge by double-stranded RNA. *Fish Shellfish Immunol*. 2007;23:242–6.



185. Xu J, Han F, Zhang X. Silencing shrimp white spot syndrome virus (WSSV) genes by siRNA. *Antiviral Res.* 2007;73:126–31.
186. Tirasophon W, Yodmuang S, Chinnirunvong W, Plongthongkum N, Panyim S. Therapeutic inhibition of yellow head virus multiplication in infected shrimps by YHV-protease dsRNA. *Antiviral Res.* 2007;74:150–5.
187. Westenberg M, Heinhuis B, Zuidema D, Vlaskovits JM. siRNA injection induces sequence-independent protection in *Penaeus monodon* against white spot syndrome virus. *Virus Res.* 2005;114:133–9.
188. Ding S-W, Voinnet O. Antiviral immunity directed by small RNAs. *Cell.* 2007;130:413–26.
189. Robalino J, Bartlett TC, Chapman RW, Gross PS, Browdy CL, Warr GW. Double-stranded RNA and antiviral immunity in marine shrimp: inducible host mechanisms and evidence for the evolution of viral counter-responses. *Dev Comp Immunol.* 2007;31:539–47.
190. Wu Y, Lü L, Yang L-S, Weng S-P, Chan S-M, He J-G. Inhibition of white spot syndrome virus in *Litopenaeus vannamei* shrimp by sequence-specific siRNA. *Aquaculture.* 2007;271:21–30.
191. Labreuche Y, Veloso A, de la Vega E, Gross PS, Chapman RW, Browdy CL, et al. Non-specific activation of antiviral immunity and induction of RNA interference may engage the same pathway in the Pacific white leg shrimp *Litopenaeus vannamei*. *Dev Comp Immunol.* 2010;34:1209–18.
192. Bartholomay LC, Loy DS, Dustin Loy J, Harris DL. Nucleic-acid based antivirals: Augmenting RNA interference to “vaccinate” *Litopenaeus vannamei*. *J Invertebr Pathol.* 2012;110:261–6.
193. Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, et al. Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell.* 2004;117:69–81. Elsevier.
194. Schnettler E, Tykalová H, Watson M, Sharma M, Sterken MG, Obbard DJ, et al. Induction and suppression of tick cell antiviral RNAi responses by tick-borne flaviviruses. *Nucleic Acids Res.* 2014;42:9436–46.
195. Chen Y-H, Zhao L, Jia X-T, Li X-Y, Li CZ, Yan H, et al. Isolation and characterization of cDNAs encoding Ars2 and Pasha homologues, two components of the RNA interference pathway in *Litopenaeus vannamei*. *Fish Shellfish Immunol.* 2012;32:373–80.
196. Lu L-F, Liston A. MicroRNA in the immune system, microRNA as an immune system. *Immunology.* 2009;127:291–8. Blackwell Publishing Ltd.
197. Fullaondo A, Lee SY. Identification of putative miRNA involved in *Drosophila melanogaster* immune response. *Dev Comp Immunol.* 2012;36:267–73.
198. Ruan L, Bian X, Ji Y, Li M, Li F, Yan X. Isolation and identification of novel microRNAs from *Marsupenaeus japonicus*. *Fish Shellfish Immunol.* 2011;31:334–40.
199. Li Y, Li J, Belisle S, Baskin CR, Tumpey TM, Katze MG. Differential microRNA expression and virulence of avian, 1918 reassortant, and reconstructed 1918 influenza A viruses. *Virology.* 2011;421:105–13.
200. Du P, Wu J, Zhang J, Zhao S, Zheng H, Gao G, et al. Viral infection induces expression of novel phased microRNAs from conserved cellular microRNA precursors. *PLoS Pathog.* 2011;7:e1002176. Public Library of Science.
201. Tian F, Luo J, Zhang H, Chang S, Song J. MiRNA expression signatures induced by Marek's disease virus infection in chickens. *Genomics.* 2012;99:152–9.
202. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, et al. The Pfam protein families database. *Nucleic Acids Res.* 2004;32:D138–41.
203. Loongyai W, Avarre J-C, Cerutti M, Lubzens E, Chotigeat W. Isolation and functional characterization of a new shrimp ovarian peritrophin with antimicrobial activity from *Fenneropenaeus merguensis*. *Mar Biotechnol (NY).* 2007;9:624–37.
204. Lehane MJ. Peritrophic matrix structure and function. *Annu Rev Entomol.* 1997;42:525–50.
205. Wang P, Granados RR. An intestinal mucin is the target substrate for a baculovirus enhancer. *Proc Natl Acad Sci U S A.* 1997;94:6977–82. National Acad Sciences.
206. Terra WR. The origin and functions of the insect peritrophic membrane and peritrophic gel. *Arch. Insect Biochem. Physiol.* John Wiley & Sons, Inc.; 2001;47:47–61.
207. Tellam RL, Wijffels G, Willadsen P. Peritrophic matrix proteins. *Insect Biochem Mol Biol.* 1999;29:87–101.
208. Khayat M, Babin PJ, Funkenstein B, Sammar M, Nagasawa H, Tietz A, et al. Molecular characterization and high expression during oocyte development of a shrimp ovarian cortical rod protein homologous to insect intestinal peritrophins. *Biol Reprod.* 2001;64:1090–9. Society for the Study of Reproduction.
209. Du X-J, Wang J-X, Liu N, Zhao X-F, Li F-H, Xiang J-H. Identification and molecular characterization of a peritrophin-like protein from fleshy prawn (*Fenneropenaeus chinensis*). *Mol Immunol.* 2006;43:1633–44.
210. Wang L, Li F, Wang B, Xiang J. A new shrimp peritrophin-like gene from *Exopalaemon carinicauda* involved in white spot syndrome virus (WSSV) infection. *Fish Shellfish Immunol.* 2013;35:840–6.
211. Chen K-Y, Hsu T-C, Huang P-Y, Kang S-T, Lo C-F, Huang W-P, et al. *Penaeus monodon* chitin-binding protein (PmCBP) is involved in white spot syndrome virus (WSSV) infection. *Fish Shellfish Immunol.* 2009;27:460–5.
212. Huang Y, Ma F, Wang W, Ren Q. Identification and molecular characterization of a peritrophin-like gene, involved in the antibacterial response in Chinese mitten crab, *Eriocheir sinensis*. *Dev Comp Immunol.* 2015;50:129–38.
213. Reed JC, Doctor KS, Godzik A. The Domains of Apoptosis: A Genomics Perspective. *Sci. Signal. Science Signaling.* 2004;2004:re9-re9.
214. McEntyre JR, Gibson TJ. Patterns and clusters within the PSM column in *TiBS*, 1992-2004. *Trends Biochem Sci.* 2004;29:627–33.
215. Kischkel FC, Hellbardt S, Behrmann I, Germer M, Pawlita M, Krammer PH, et al. Cytotoxicity-dependent APO-1 (Fas/CD95)-associated proteins form a death-inducing signaling complex (DISC) with the receptor. *EMBO J.* 1995;14:5759–88.
216. Chinnaiyan AM, O'Rourke K, Tewari M, Dixit VM. FADD, a novel death domain-containing protein, interacts with the death domain of Fas and initiates apoptosis. *Cell.* 1995;81:505–12.
217. Muzio M, Chinnaiyan AM, Kischkel FC, O'Rourke K, Shevchenko A, Ni J, et al. FLICE, a novel FADD-homologous ICE/CED-3-like protease, is recruited to the CD95 (Fas/APO-1) death-inducing signaling complex. *Cell.* 1996;85:817–27.
218. Boldin MP, Goncharov TM, Goltsev YV, Wallach D. Involvement of MACH, a novel MORT1/FADD-interacting protease, in Fas/APO-1- and TNF receptor-induced cell death. *Cell.* 1996;85:803–15.
219. Georgel P, Naitza S, Kappler C, Ferrandon D, Zachary D, Swimmer C, et al. *Drosophila* immune deficiency (IMD) is a death domain protein that activates antibacterial defense and can promote apoptosis. *Dev Cell.* 2001;1:503–14.
220. Sun H, Bristow BN, Qu G, Wasserman SA. A heterotrimeric death domain complex in Toll signaling. *Proc Natl Acad Sci U S A.* 2002;99:12871–6. National Acad Sciences.
221. Hsu H, Xiong J, Goeddel DV. The TNF receptor 1-associated protein TRADD signals cell death and NF-kappa B activation. *Cell.* 1995;81:495–504.
222. Stanger BZ, Leder P, Lee T-H, Kim E, Seed B. RIP: A novel protein containing a death domain that interacts with Fas/APO-1 (CD95) in yeast and causes cell death. *Cell.* 1995;81:513–23.
223. Hsu H, Shu HB, Pan MG, Goeddel DV. TRADD-TRAF2 and TRADD-FADD interactions define two distinct TNF receptor 1 signal transduction pathways. *Cell.* 1996;84:299–308.
224. Hsu H, Huang J, Shu HB, Baichwal V, Goeddel DV. TNF-dependent recruitment of the protein kinase RIP to the TNF receptor-1 signaling complex. *Immunity.* 1996;4:387–96.
225. Schievella AR, Chen JH, Graham JR, Lin LL. MADD, a novel death domain protein that interacts with the type 1 tumor necrosis factor receptor and activates mitogen-activated protein kinase. *J Biol Chem.* 1997;272:12069–75.
226. Abrams JM, White K, Fessler LI, Steller H. Programmed cell death during *Drosophila* embryogenesis. *Development.* 1993;117:29–43. The Company of Biologists Ltd.
227. Nagata S. Apoptosis by death factor. *Cell.* 1997;88:355–65.
228. Inohara N, Nuñez G. ML - a conserved domain involved in innate immunity and lipid metabolism. *Trends Biochem Sci.* 2002;27:219–21.
229. Schromm AB, Lien E, Henneke P, Chow JC, Yoshimura A, Heine H, et al. Molecular genetic analysis of an endotoxin nonresponder mutant cell line: a point mutation in a conserved region of MD-2 abolishes endotoxin-induced signaling. *J Exp Med.* 2001;194:79–88. Rockefeller University Press.
230. Kirchhoff C, Osterhoff C, Young L. Molecular cloning and characterization of HE1, a major secretory protein of the human epididymis. *Biol Reprod.* 1996;54:847–56. Society for the Study of Reproduction.
231. Okamura N, Kiuchi S, Tamba M, Kashima T, Hiramoto S, Baba T, et al. A porcine homolog of the major secretory protein of human epididymis, HE1, specifically binds cholesterol. *Biochim Biophys Acta.* 1999;1438:377–87.
232. Naureckiene S, Sleat DE, Lackland H, Fensom A, Vanier MT, Wattiaux R, et al. Identification of HE1 as the second gene of Niemann-Pick C disease. *Science.* 2000;290:2298–301. American Association for the Advancement of Science.

233. Shimazu R, Akashi S, Ogata H, Nagai Y, Fukudome K, Miyake K, et al. MD-2, a molecule that confers lipopolysaccharide responsiveness on Toll-like receptor 4. *J Exp Med*. 1999;189:1777–82.
234. Viriyakosol S, Tobias PS, Kitchens RL, Kirkland TN. MD-2 binds to bacterial lipopolysaccharide. *J Biol Chem*. 2001;276:38044–51.
235. Gangloff M, Gay NJ. MD-2: the Toll “gatekeeper” in endotoxin signalling. *Trends Biochem Sci*. 2004;29:294–300.
236. Visintin A, Mazzoni A, Spitzer JA, Segal DM. Secreted MD-2 is a large polymeric protein that efficiently confers lipopolysaccharide sensitivity to Toll-like receptor 4. *Proc Natl Acad Sci U S A*. 2001;98:12156–61.
237. Alexander B, Goldstein R. Dual Hemostatic Defect in Pseudothrombophilia. *J Clin Invest*. 1953;32:551.
238. Quick AJ, Hussey CV. Hemophilic condition in the female. *J Lab Clin Med*. 1953;42:929–30.
239. Houdijk WP, Sakariassen KS, Nievelstein PF, Sixma JJ. Role of factor VIII-von Willebrand factor and fibronectin in the interaction of platelets in flowing blood with monomeric and fibrillar human collagen types I and III. *J Clin Invest*. 1985;75:531–40. American Society for Clinical Investigation.
240. Marchese P, Murata M, Mazzucato M, Pradella P, De Marco L, Ware J, et al. Identification of three tyrosine residues of glycoprotein Ib alpha with distinct roles in von Willebrand factor and alpha-thrombin binding. *J Biol Chem*. 1995;270:9571–8.
241. Wu XX, Gordon RE, Glanville RW, Kuo HJ, Uson RR, Rand JH. Morphological relationships of von Willebrand factor, type VI collagen, and fibrillin in human vascular subendothelium. *Am J Pathol*. 1996;149:283–91. American Society for Investigative Pathology.
242. Krem MM, Di Cera E. Evolution of enzyme cascades from embryonic development to blood coagulation. *Trends Biochem Sci*. 2002;27:67–74.
243. Opal SM. Phylogenetic and functional relationships between coagulation and the innate immune response. *Crit Care Med*. 2000;28:S77–80.
244. Iwanaga S, Kawabata S, Muta T. New types of clotting factors and defense molecules found in horseshoe crab hemolymph: their structures and functions. *J Biochem*. 1998;123:1–15. Oxford University Press.
245. Armstrong PB, Armstrong MT. The decorated clot: Binding of agents of the innate immune system to the fibrils of the limulus blood clot. *Biol Bull*. 2003;205:201–3.
246. Isakova V, Armstrong PB. Imprisonment in a death-row cell: the fates of microbes entrapped in the Limulus blood clot. *Biol Bull*. 2003;205:203–4.
247. Verweij CL, Diergaarde PJ, Hart M, Pannekoek H. Full-length von Willebrand factor (vWF) cDNA encodes a highly repetitive protein considerably larger than the mature vWF subunit. *EMBO J*. 1986;5:1839–47. European Molecular Biology Organization.
248. Sadler JE. Biochemistry and genetics of von Willebrand factor. *Annu Rev Biochem*. 1998;67:395–424.
249. Loh PC, Tapay LM, Lu Y, Nadala EC. Viral pathogens of the penaeid shrimp. *Adv Virus Res*. 1997;48:263–312.
250. Fuhrman JA. Marine viruses and their biogeochemical and ecological effects. *Nature*. 1999;399:541–8.
251. Katoh K, Asimenos G, Toh H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol Biol*. 2009;537:39–64.
252. Stamatakis A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30:1312–3.
253. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28:1647–9.
254. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;39:W29–37.
255. Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol*. 2004;338:1027–36.
256. Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res*. 2007;35:W429–32.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

