Review

Journal of Innate Immunity

J Innate Immun 2016;8:129–142
DOI: 10.1159/000441475

Received: May 29, 2015
Accepted after revision: October 5, 2015
Published online: November 19, 2015

High Innate Immune Specificity through Diversified C-Type Lectin-Like Domain Proteins in Invertebrates

Barbara Pees Wentao Yang Alejandra Zárate-Potes Hinrich Schulenburg Katja Dierking

Department of Evolutionary Ecology and Genetics, Zoological Institute, Christian Albrechts University of Kiel, Kiel, Germany

Key Words
C-type lectin-like domain proteins · Immune specificity · Innate immunity · Insects · Crustaceans · *Caenorhabditis elegans*

Abstract
A key question in current immunity research is how the innate immune system can generate high levels of specificity. Evidence is accumulating that invertebrates, which exclusively rely on innate defense mechanisms, can differentiate between pathogens on the species and even strain level. In this review, we identify and discuss the particular potential of C-type lectin-like domain (CTLD) proteins to generate high immune specificity. Whilst several CTLD proteins are known to act as pattern recognition receptors in the vertebrate innate immune system, the exact role of CTLD proteins in invertebrate immunity is much less understood. We show that CTLD genes are highly abundant in most metazoan genomes and summarize the current state of knowledge on CTLD protein function in insect, crustacean and nematode immune systems. We then demonstrate extreme CTLD gene diversification in the genomes of *Caenorhabditis* nematodes and provide an update of data from CTLD gene function studies in *C. elegans*, which indicate that the diversity of CTLD genes could contribute to immune specificity. In spite of recent achievements, the exact functions of the diversified invertebrate CTLD genes are still largely unknown. Our review therefore specifically discusses promising research approaches to rectify this knowledge gap.

© 2015 S. Karger AG, Basel

Introduction

Animals depend on the specificity of their immune reaction to efficiently and economically protect themselves against pathogens. For a long time, the generation of a highly specific immune response was considered a hallmark of adaptive immunity, based on its immense diversity of T and B cell receptors generated by somatic recombination and hypermutation. There is, however, accumulating evidence that invertebrates, which solely rely on innate defense mechanisms and lack T and B cells, can differentiate between pathogens on the species and even strain level [1–3]. Whether and how such high levels of specificity can be generated by the innate immune system is one of the main questions in current immunity research.

KARGER
E-Mail karger@karger.com
www.karger.com/jin

© 2015 S. Karger AG, Basel
1662–811X/15/0082–0129$39.50/0

Dr. Katja Dierking
Zoological Institute, Christian-Albrechts-Universität zu Kiel
Am Botanischen Garten 1–9
DE–24118 Kiel (Germany)
E-Mail kdierking@zoologie.uni-kiel.de

A prerequisite for the generation of a pathogen-specific immune reaction is a repertoire of diversified receptors, which enables the organism to specifically recognize and respond to different immune challenges. In vertebrates the well-characterized pattern recognition receptors (PRRs) of the innate system, such as the Toll-like receptors (TLRs), retinoic acid-inducible gene-I-like receptors (RLRs), and NOD-like receptors (NLRs), recognize structures that are conserved among microbial species and mediate a broader pattern of specificity (i.e. on the level of a particular pathogen group, such as Gram-negative or Gram-positive bacteria). The diversity of ligand binding can be expanded by the formation of receptor homo- or heterodimers and/or the interaction with other PRR coreceptors, as is well described for mammalian TLRs. However, in invertebrates, high levels of specificity (on the species and strain level) are unlikely mediated by TLRs, NLRs or RLRs, as these are too few in number in most, although not all (see below), animal species.

There are few examples of the diversification of alternative putative immune receptors in invertebrates. The hypervariable protein Down syndrome cell adhesion molecule (Dscam), which is generated by alternative somatic splicing, was shown to be implicated in the immune response of insects and crustaceans [4, 5], and suggested to act as a pathogen-specific recognition molecule. Further experimental evidence to support this hypothesis is, however, still scarce [6]. A high level of genetic polymorphism was observed in the immunoglobulin (Ig)-type variable (V) region-containing chitin-binding proteins (VCBPs) of the cephalochordate *Branchiostoma floridae* [7], and in the Ig domain of fibrinogen-related proteins (FREPs) of the snail *Biomphalaria glabrata* [8]. However, the molecular mechanisms underlying the diversification of VCBPs and FREPs and their functional significance for immune specificity are not well understood. Yet another example of diversification of putative innate immune system proteins is the Sp185/333 system in echinoderms, such as the purple sea urchin *Strongylocentrotus purpuratus*. The members of this highly diverse, novel gene family are not homologous to any other known genes, they are highly upregulated after immune challenge and expressed in the phagocyte class of sea urchin coelomocytes. The function of Sp185/333 proteins is unknown. The high diversification of Sp185/333 genes and proteins is achieved through different mechanisms, including maintenance of high polymorphism, gene recombination and modifications at the RNA and protein level [9]. Similar multigene expansions, usually based on repeated gene duplications that potentially produce high-sequence diversification, were observed for TLR and NLR gene families in the genomes of the sea urchin *S. purpuratus* [10] and the lancelet *B. floridae* [11], but the functional consequence of this diversified repertoire of potential receptors remains elusive. Here, we argue that another group of known, yet neglected, PRRs – the C-type lectins – show surprisingly high diversity in most metazoan genomes and have the potential to generate immune specificity.

C-type lectins are a family of both soluble and membrane-bound proteins with a characteristic carbohydrate-recognition domain (CRD) that mediates ligand binding. The CRD has a distinct protein fold determined by highly conserved residues that form a characteristic sequence motif [12]. It was first described in Ca^{2+}-dependent ('C-type') carbohydrate binding animal proteins ('lectins'), but has subsequently been found in many other proteins that do not bind sugar and/or calcium. Thus, we here use the term C-type lectin-like domain (CTLD) in its broadest sense, as proposed previously [13], and use sequence similarity to identify genes containing CTLDs in different metazoan genomes. CTLD proteins have a wide range of functions and bind a wide variety of ligands, such as sugars, proteins, lipids and inorganic compounds [13].

In vertebrates, CTLD proteins are known to fulfill important tasks in the immune system by acting as PRRs [for a review, see 14], as well as effector proteins with bactericidal activity [15]. As we will demonstrate further below, CTLD genes are highly abundant in many metazoan genomes. Yet, our knowledge on CTLD protein function in immunity mainly derives from the few well-studied vertebrate CTLD proteins, while the function of the majority of CTLD proteins is unknown. In invertebrates, it has repeatedly been suggested that the diversified CTLD proteins contribute to immunity, yet exact functional genetic information is largely absent and the few exceptions mainly focus on the assessment of single CTLD proteins. Moreover, the particular role of CTLD proteins in generating immune specificity is completely unexplored. Therefore, we require a systematic research program that aims at elucidating the role of diversified CTLD proteins in immune specificity.

Here, we highlight that CTLD genes constitute expanded gene families in many metazoan taxa and summarize the available evidence for a role of CTLD proteins in invertebrate immunity, mainly inferred from studies on crustaceans and insects. We then turn to the nematode genus *Caenorhabditis*, in which several complete genomes are available, to demonstrate extreme CTLD gene diversification, followed by a summary of data from *C. elegans* that strongly indicate a role of the diversified CTLD proteins in immune specificity. Finally, we empha-

Table 1. Number of CTLD genes (proteins) in different metazoan genomes

Species	Taxon	Genome size, Mb	Total coding genes	Total InterPro[1] genes (proteins), n	Total Pfam[1] genes (proteins), n	Immune function[2]	Ref.	Genome assembly
H. sapiens	Mammalia	3,209.29	20,805	100 (277)	94 (268)	e.g. DC-SIGN, DECTIN-1, DECTIN-2, MINCLE, Langerin (CD207), BDCA-2, MBL2, COLEC11, MRC-1, MRC-2, SELP, SELE, SELL, CLEC5A, CLECSF8, DCIR, CLEC9A, DEC-205, and others	reviewed in 14, 61, 62, 73	GRCh37
M. musculus	Mammalia	2,798.79	23,148	132 (355)	129 (339)	e.g. SIGNR1, SIGNR3, SIGNR5, Dectin-1, Dectin-2, Mincle, Langerin (Cd207), MBL-A, MBL-C, Colec11, Mrc-1, Mrc-2, Selp, Sele, Sell, Clec5a, Clecsf8, Dcir, Dcir2, DNGR-1 (Clec9a), DEC-205, and others	reviewed in 14, 61, 62, 73	GRCm38
D. rerio	Teleostei	1,412.46	26,459	176 (381)	169 (360)			Zv9
B. floridae	Cephalochordata	521.90	30,601[3]	692 (694)	646 (647)	AmphiCTL1	74	v1.0
S. purpuratus	Echinodermata	936.58	28,987	293 (303)	241 (244)	SpEchinoidin	75, 76	Spur_3.1
A. gambiae	Insecta	265.03	12,810	37 (99)	33 (93)	CTL4, CTLMA2	34	AgamP3
B. mori	Insecta	481.82	14,623	21 (32)	17 (25)	BmLBP, BmMBP	21, 26, 77	ASM15162v1
D. melanogaster	Insecta	139.49	13,937	56 (96)	48 (84)	DL1, DL2, DL3	18–20	BDGP 5
C. brenneri	Nematoda	190.37	30,667	319 (320)	159 (159)			WS227
C. briggsae	Nematoda	108.48	21,936	183 (183)	111 (111)			WS224
C. elegans	Nematoda	100.29	20,541	283 (288)	166 (185)	CLEC-39, CLEC-49, CLEC-79 + 9 *clec* genes (see table 2)	see table 2	WS235
C. japonica	Nematoda	166.26	29,964	111 (139)	58 (72)			WS227
C. remanei	Nematoda	145.44	31,444	292 (292)	158 (158)			WS185
N. vectensis	Cnidaria	356.61	24,773	67 (67)	62 (62)			ASM20922v1
A. queenslandica	Porifera	166.70	30,060	2 (2)	1 (1)			Aqu1
T. adhaerens	Placozoa	105.63	11,520	45 (45)	40 (40)			ASM15027v1

The total number of CTLD genes are listed for each species.
[1] The number of genes (proteins) with at least one CTLD were obtained by automatic annotation from the InterPro (v53.0) [16] and Pfam (v28.0) databases [17]. The number of CTLD proteins is higher than the CTLD gene number as it includes different isoforms encoded by the same gene. Proteins with multiple CTLDs are counted only once.
[2] CTLD proteins for which a function in immune defense is supported by experimental evidence. The list is nonexhaustive for human and mouse CTLD genes, for which gene name synonyms are shown that are commonly used in the literature.
[3] Noncurated annotation.

size the current lack of exact understanding of the immune functions of the highly diverse CTLD proteins across metazoan taxa and define promising avenues for future research that integrate functional genetic with protein analysis approaches to rectify this knowledge gap.

CTLD Genes Are Abundant in Metazoan Genomes and Mediate Immune Responses in a Variety of Invertebrate Taxa

While CTLD proteins have been shown to fulfill diverse and fundamental tasks in the vertebrate immune system, their function in invertebrate immunity is much less understood. This lack of data on CTLD function in invertebrates is surprising given the fact that genes encoding CTLD proteins are present in all invertebrate genomes sequenced to date and in some cases in very large numbers. In table 1 we list the available genome sequence data for some selected species as representatives of the major invertebrate groups and show the number of CTLD genes and proteins as provided by the InterPro v53.0 [16] and Pfam v28.0 databases [17]. CTLD genes are highly abundant in many metazoan genomes, including, for example, 132 members in *Mus musculus* (the 30th most abundant gene family in *M. musculus*) or 283 gene family members in the nematode *C. elegans* (the 7th most numerous gene family in *C. elegans*). The CTLD encod-

ing gene repertoire in the analyzed invertebrate species varies from two in the sponge *Amphimedon queenslandica* to more than 600 in the cephalochordate *B. floridae* [11]. We also looked for the total number of CTLD genes in the genomes of nine plant species *(Arabidopsis thaliana, Glycine max, Zea mays, Oryza sativa japonica* subspp., *Populus trichocarpa, Medicago truncatula, Triticum aestivum, Hordeum vulgare* and *Sorghum bicolor)*. Interestingly, the plant genomes encode only very few (1–4) CTLD genes, which further emphasizes that the expansion of CTLD genes in metazoans is remarkable. Experimental evidence of an immune function of most invertebrate CTLD-encoding genes is missing, although their role in immunity is usually assumed. Within the existing information on invertebrate CTLD proteins, the functional data is biased towards certain insect, crustacean and nematode taxa. In the following paragraphs, we will summarize the evidence of CTLD proteins functioning in the immune defenses of insects and crustaceans, before focusing on the nematode genus *Caenorhabditis*.

Insects

Although genes encoding CTLD proteins are numerous in many insect genomes (table 1), research on insect CTLD proteins has as yet been focused on assessment of single CTLD proteins from different species. For example, only 3 of the more than 50 CTLD-encoding genes in *Drosophila melanogaster* were characterized in more detail [18–20]. Several insect CTLD proteins were suggested to function as PRRs, including those found to mediate important insect immune defense mechanisms, such as hemocyte nodule formation, encapsulation, melanization and activation of phagocytosis [18, 21, 22]. Most insect CTLD proteins were first isolated from hemolymph either due to their ability to agglutinate red blood cells [19, 22], to bind to bacterial cells [21], or were selectively purified from larval plasma by carbohydrate affinity chromatography after bacterial challenge [23]. The purified CTLD proteins were subsequently characterized biochemically in vitro, for example by testing their binding properties and hemagglutination activities. These protein-level studies demonstrated CTLD proteins to possess both broad and more specific binding abilities.

While most studied insect proteins only have one CTLD, some characterized proteins from lepidopteran species, such as *Bombyx mori* [24], *Manduca sexta* [23], and the cotton bollworm *Helicoverpa armigera* [25], have two CTLDs in tandem. The *B. mori* multibinding protein (BmMBP) is an example of an insect CTLD protein with a broad binding capacity. Each of its two divergent tandem CTLDs has distinct binding specificities, enabling the molecule to bind to a large variety of microorganisms ranging from Gram-positive to Gram-negative bacteria, as well as yeasts. After binding to its target, BmMBP is able to trigger cellular immune responses, such as aggregation of hemocytes and nodule formation [26]. The expression of the four characterized *M. sexta* CTLD proteins immulectin-1 to -4 is induced upon bacterial exposure in the fat body. Immulectin-1 and -2 function in activation of the prophenoloxidase [27] and immulectin-3 and -4 act in enhancing encapsulation and melanization [28, 29]. They bind to microbe-associated molecular patterns (MAMPs) [30] with different affinities and specificities in vitro. While immulectin-1, -3 and -4 bind to a broad range of MAMPs, immulectin-2 specifically binds to lipopolysaccharide [28, 29, 31].

These and other studies provide important information on the potential immune functions of insect CTLD proteins. However, there is hardly any experimental evidence from functional genetic studies on the role of CTLD proteins in insect immune defenses in vivo. The three exceptions tested the effect of CTLD gene expression knockdown or knockout in a lepidopteran, the fruit fly and a mosquito. RNAi knockdown of the *M. sexta* CTLD gene immulectin-2 renders animals more susceptible to infection with the insect pathogen *Photorhabdus luminescens* [32]. The galactose-specific *D. melanogaster* CTLD protein DL1, whose expression is induced by injury [19] and which specifically binds to *Escherichia coli* and *Erwinia chrysanthemi*, but not to the Gram-positive *Staphylococcus aureus* or the yeast *Saccharomyces cerevisiae*, agglutinates *E. coli* and promotes the association of a *Drosophila* hemocytes-derived cell line with *E. coli* [20]. A *DL1* deletion mutant did not show any defects in antimicrobial peptide expression after pricking or infection with *E. coli*, indicating that DL1 does not participate in the humoral immune response in vivo, but might still be involved in cellular immune responses such as phagocytosis [20]. In a functional genetic screen of about 100 immune candidate genes using RNAi in the mosquito *Anopheles gambiae* (Diptera) two CTLD proteins, CTL4 and CTLMA2, were discovered to affect *Plasmodium berghei* development by acting as negative regulators of the vector's melanization response to *P. berghei* ookinetes [33]. In a subsequent genetic analysis [34], CTL4 or CTLMA2 silencing by RNAi significantly reduced the survival of adult female mosquitoes infected with *E. coli*, *Enterobacter cloacae* and Pseudomonadaceae H2.26, but not with *S. aureus*, *Micrococcus luteus* and *Enterococcus faecalis*, suggesting a specific role in defense against Gram-negative

bacteria. CTL4 and CTLMA2 expression is induced by *E. coli* and *S. aureus* infection and they both produce a heterodimer that is secreted into the mosquito hemolymph [34].

In summary, while some of the characterized insect CTLD proteins have the ability to bind to a broad range of microorganisms or MAMPs, others exhibit more specific binding properties. Insect CTLD proteins are involved in the activation of cellular immune responses rather than the humoral immune response, but also act as opsonization molecules. A direct effector function such as antimicrobial activity has not been described and their target signaling pathways are unknown. Overall, the potential immune functions of the high variety of insect CTLD proteins (table 1) are far from being understood.

Crustaceans

The second most comprehensive data set on invertebrate CTLD protein functions comes from research on crustacean species that are of economic importance, such as the Chinese white shrimp *Fenneropenaeus chinensis* or the Chinese mitten crab *Eriocheir sinensis*. CTLD protein research in these species mainly aims at the development of antiparasitic drugs, as parasitic threats are a particular concern in crustacean aquaculture. For example, the white spot syndrome virus infects most of the commercially important shrimp species and can cause major production losses [35]. Since CTLD proteins are known to participate in the immune response of other invertebrates and since many expressed sequence tags encoding lectins were identified in cDNA libraries from both healthy and immune-challenged crustaceans, characterization of immune-relevant CTLD proteins in these economically important species has been a particular focus of the current research activities. Comprehensive data sets are thus available on the molecular cloning, isolation and in vitro biochemical characterization of purified crustacean CTLD proteins (either the recombinant or the native protein), especially from the economically important shrimp species, as reviewed in depth previously [36]. Like insect CTLD proteins, crustacean CTLD proteins have repeatedly been suggested to act as PRRs in various species. They can enhance encapsulation, promote phagocytosis and induce the production of reactive oxygen intermediates in hemocytes in vitro [36]. In addition, several crustacean CTLD proteins contribute to pathogen clearance in vivo [37–39].

Interestingly, some crustacean CTLD proteins exhibit in vitro antibacterial or antiviral activity and are thus assumed to act as antimicrobials. For example, Fc-hsL from *F. chinensis* inhibits the growth of Gram-positive as well as Gram-negative bacteria and fungi, with the highest activity against Gram-positive bacteria [40]. Similarly, the two *E. sinensis* CTLD proteins EsLecA and EsLecG produce antimicrobial activity against both Gram-positive and Gram-negative bacteria in in vitro growth-inhibitory activity assays [41]. Most characterized crustacean CTLD proteins are expressed in the gills, intestine, hemocytes and hepatopancreas (the counterpart of the fat body in insects), and are likely to be secreted into the hemolymph upon immune challenge. Indeed, almost all crustacean CTLD proteins that have been studied so far show an inducible expression by either bacterial and/or viral pathogens. The majority of characterized crustacean CTLD proteins have one CTLD, the remaining have two [36]. Surprisingly, all studied crustacean CTLD proteins possess broad binding properties active towards all the bacterial strains tested. One exception might be the *Penaeus monodon* PmLec, which binds lipopolysaccharide and mainly agglutinates Gram-negative bacteria in vitro [42].

Since the genomes of these economically important crustacean species have not been fully sequenced yet, and shrimps and crabs have (as yet) not been amenable to genetic manipulation, genomic analyses and functional genetic studies on the crustacean CTLD genes are almost completely missing. The exceptions are three recent studies in which expression of crustacean CTLD proteins was genetically manipulated in vivo. Overexpression of the CTLD protein PcLec4 from the red swamp crayfish *Procambarus clarkii*, which is induced by infection with *Vibrio anguillarum*, facilitated bacterial clearance and increased survival of crayfish infected with *V. anguillarum* [39]. Knockdown of the expression of the CTLD protein MjHeCL from the kuruma shrimp *Marsupenaeus japonicus* by RNAi caused increased bacterial proliferation in the hemolymph and led to a change in expression of several antimicrobial peptides in hemocytes both in vivo and in vitro. It was thus suggested that MjHeCL plays a role in restricting the growth of the hemolymph microbiota by regulating antimicrobial peptide expression [43]. Similarly, the role of another *M. japonicus* CTLD protein, hFcLec4, and of FcLec4 from *F. chinensis* in promoting phagocytosis by acting as opsonins in hemocytes was supported by the finding that clearance of the pathogenic *V. anguillarum* from the hemolymph was delayed by RNAi knockdown of hFcLec4 and FcLec4 [44].

Interestingly, FcLec4 and hFcLec4 interact with the transmembrane protein β-integrin via their N-terminal domains and this interaction is required for their opsonic activity [44]. β-Integrins are known to participate in

phagocytosis in other invertebrates, such as flies [45]. Shrimp FcLec4 and hFcLec4 may thus act as PRRs recognizing bacteria via their CTLD and their binding to β-integrin in the membrane of hemocytes leads to cytoskeletal reorganization, which induces phagosome formation to ingest the bacteria [44]. This study is one of only two studies providing the first insight into interacting and signaling partners of invertebrate CTLD proteins. In the other study, a conserved binding motif for the transcription factor NF-κB was identified in the flanking promoter sequence of the CTLD gene LvCTL3 from *Litopenaeus vannamei*. LvCTL3 contributed to opsonization and thus marking bacteria and cells for phagocytosis. Deletion of the NF-κB binding site led to the abolishment of LvCTL3 expression, suggesting its regulation by NF-κB. This finding represents a first step towards the incorporation of CTLD proteins into the poorly understood immune signaling network in crustaceans [46].

Taken together, the functional role of several crustacean CTLD proteins was inferred at the protein level by isolating the respective proteins and biochemically analyzing their physical interaction with microorganisms or MAMPs in vitro. Almost all crustacean CTLD proteins are able to bind to a broad range of microorganism and/or MAMPs. They seem to be involved in the activation of cellular immune responses, such as encapsulation and phagocytosis, as well as in the direct elimination of pathogens by exhibiting bactericidal activity. However, the actual number of genes encoding CTLD proteins, their genomic and structural organization, and the underlying impact on the specificity of the interactions with microorganisms remains unknown due to the lack of genome sequences and further experimental analysis. Particularly, studies of the protein function in the context of the whole organism inferred from altered gene products, such as mutant or overexpression phenotypes, are rare.

CTLD Genes Are Extremely Diversified in *Caenorhabditis* Genomes

The nematode genus *Caenorhabditis* provides an impressive example of extreme CTLD gene diversification. This was previously noted for the model nematode *C. elegans* [47], which possesses 283 CTLD genes (*clec* genes) in its genome (online suppl. table S1; for all online suppl. material, see www.karger.com/doi/10.1159/000441475; note that this is an updated number and thus different from the previous report due to further improvements of the whole genome sequence). Of the 283 *clec* genes, 81% are predicted to have a signal peptide [48] and are thus likely secreted proteins, highlighting a particular role of the CTLD proteins in the pseudocoel, in the intestinal lumen and/or as components of the secreted cuticle. In contrast, 21% of the 283 *clec* genes possess membrane-spanning properties, suggesting that these function as components of cellular membranes. *clec* gene evolution is likely highly dynamic and subject to repeated duplication events, as indicated by their genomic localization in clusters (125 out of 283, equivalent to 44%; online suppl. table S1) and the high phylogenetic similarity of the genes within a cluster [47]. It is possible that the repeated duplications are favored by pathogen-mediated selection if *clec* gene duplicates allow the worm to recognize and/or eliminate a larger variety of pathogen strains [47].

In recent years, complete genome sequences became available for additional *Caenorhabditis* species, allowing cross-species comparison of *clec* gene diversification. Next to *C. elegans*, four additional species have comparatively well-annotated genomes, namely *C. briggsae*, *C. remanei*, *C. brenneri* and *C. japonica*. These five *Caenorhabditis* genomes contain between 111 *(C. japonica)* and 319 *(C. brenneri) clec* genes (online suppl. table S2). Overall, the largest group of *Caenorhabditis clec* genes are those with a single CTLD, making up half of the CTLD genes in three of the five analyzed *Caenorhabditis* species and approximately one third in the remaining two species. With one exception *(C. japonica)*, at least one third of the *Caenorhabditis clec* genes also have additional domains. These include the complement C1r/C1s Uegf Bmp1 (CUB) domain, the conserved cysteine and tryptophan residues (CW, also called PAN-3) domain, or the von Willebrand factor type A (VWA) domain. The joint presence of one or multiple CTLDs with other domains in one protein could indicate possible functions in immune signaling. Moreover, several *clec* genes (27% of all *clec* genes for *C. remanei*, 24% for *C. elegans*, 23% for *C. briggsae*, 19% for *C. brenneri* and 10% for *C. japonica*) encode proteins with more than one CTLD within the same protein. Such internal domain duplications might increase protein versatility; BmMBP from *B. mori* is one example for a tandem repetition of CTLDs that leads to extended binding abilities [26].

The majority of the *Caenorhabditis* CTLD proteins are predicted to be secreted. Besides having a putative cell-autonomous recognition function, these secreted CTLD proteins might act as highly specific immune effectors eliminating pathogens in analogy to the vertebrate lectins that are part of the complement system [49].

In summary, *clec* genes show a remarkable diversity in *Caenorhabditis* species, encoding secreted and membrane-bound proteins with one or multiple CTLDs, as well as multidomain proteins. The binding capabilities of CTLD proteins are potentially enlarged by repetitions of CTLDs within one protein. An additional functional diversification is conceivable at the protein level through dimerization or oligomerization of the CTLD proteins, as previously reported for CTLD proteins from other organisms such as CTLD protein homodimers, heterodimers and oligomers in humans and snakes [13], or the heterodimer CTL4-CTLMA2 in *A. gambiae* [34]. CTLD protein-protein interactions might have synergistic effects on CTLD binding specificity (i.e. two or multiple CTLD proteins are able to generate a higher degree of specificity if they interact with each other and with other proteins than if they act alone) and can potentially further increase the diversity available at the genetic level. Such synergistic interactions between CTLD proteins are, however, currently unexplored for the *Caenorhabditis* taxon and for invertebrates in general. Taken together, the *Caenorhabditis* CTLD proteins are particularly diverse and may thus be able to mediate immune specificity at both the level of pathogen recognition and elimination.

CTLD Proteins in *C. elegans* Show Potential for Generating High Immune Specificity

The most comprehensive data on *clec* gene function is available from the model nematode *C. elegans*. Some of the data strongly suggests that *clec* gene diversity could indeed contribute to immune specificity.

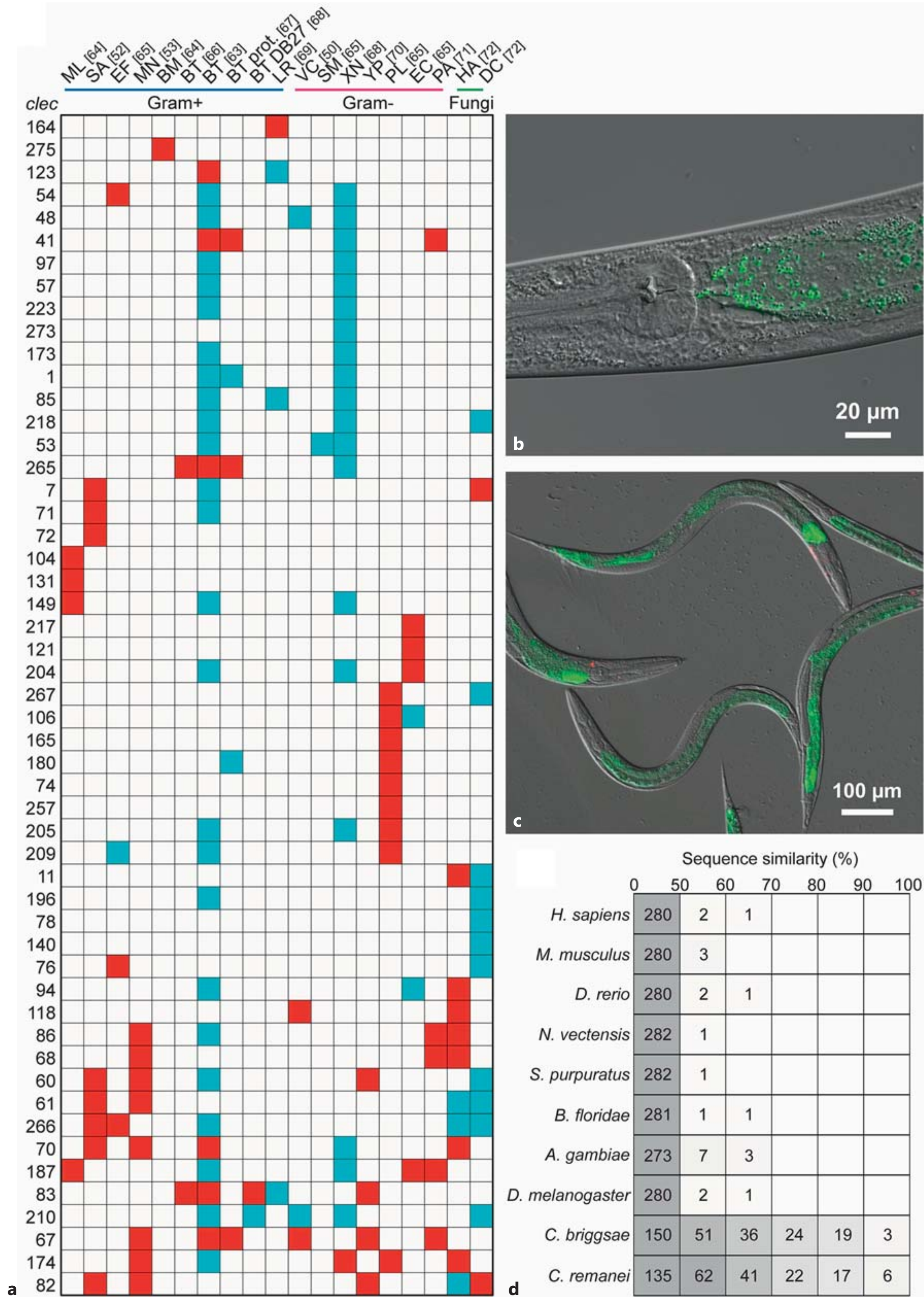
C. elegans clec Genes Are Differentially Regulated following Pathogen Infection

A large number of previous transcriptome studies in *C. elegans* demonstrate that putative immune recognition and effector gene families, such as the *clec*, lysozyme *(lys* and *ilys)* and caenopore (spp.) families, show pathogen-dependent activation upon infection [47]. Because of their extreme diversification, the *clec* genes have the highest potential to generate immune specificity. Of all *C. elegans clec* genes, 84% (237 out of 283) are differentially expressed following an infection with pathogens; 104 *clec* genes are only upregulated, 103 *clec* genes are both, up- and downregulated (online suppl. table S1), and 30 *clec* genes are only downregulated upon pathogen exposure. Importantly, some *C. elegans clec* genes are only differentially expressed in response to a specific pathogen, suggesting their highly specific regulation. In contrast, other *clec* genes are differentially expressed after infection with several pathogens. For these genes, gene induction does not appear to be influenced by general pathogen characteristics, such as the route or site of infection (e.g. the epidermis or intestine) or general pathogen surface properties (e.g. no obvious difference between Gram-negative or Gram-positive bacteria). Instead of these broader patterns of specificity expected to be common for innate immunity, a highly specific pattern of regulation is displayed (fig. 1a). For instance, *clec-67* shows upregulation upon a relatively broad range of pathogens, including both Gram-positive and Gram-negative bacteria, whereas *clec-165* is induced only by the Gram-negative *P. luminescens*. Other *clec* genes are specifically downregulated, such as *clec-273* by infection with the Gram-negative *Xenorhabdus nematophila* (fig. 1a). Interestingly, *clec* genes within a genomic cluster are not coexpressed, but seem to underlie regulation at the individual level. For example, *clec-58*, *clec-59*, *clec-60*, *clec-61* and *clec-143* lie within a cluster on chromosome II and have the same domain architecture. However, *clec-59*, *clec-60* and *clec-61* are differentially expressed after infection with several pathogens, while *clec-58* has so far not been shown to be differentially regulated upon pathogen exposure, and expression of *clec-143* is only induced following infection with *P. aeruginosa*. Different expression of *clec* genes within a genomic cluster might further indicate functional diversification.

C. elegans clec Genes Function in Pathogen Resistance

Our knowledge of the actual function of CLEC proteins in *C. elegans* is still in its infancy and it is as yet unknown why so many *clec* genes are down- or upregulated after exposure to pathogens. Few studies have tried to further elucidate the exact involvement of *clec* genes in *C. elegans* immunity by using functional genetic approaches and as yet only one study could confirm a role for CLEC proteins in pathogen binding (table 2).

The most common approach, which is taken to further analyze the function of *clec* genes, is survival analysis of either the respective *clec* mutant or *clec* RNAi-treated worms [49, 50]. In some studies, survival analysis was combined with the generation of transgenic *C. elegans* strains carrying a transcriptional *clec* gene reporter and the subsequent analysis of spatial as well as infection-induced expression patterns in vivo [51–53]. These studies demonstrated that most of the investigated *clec* genes are expressed in the intestine (fig. 1c), where most bacterial infections take place (fig. 1b), and that some *clec* genes are indeed required for resistance against infection (table 2).



1

(For legend see next page.)

Table 2. Functional analyses of CTLD genes/proteins in *C. elegans*

Gene/ protein	Approach	Phenotypic measurement	Phenotype/result	Pathogen	Ref.
clec-17	RNAi knockdown	infection	hyperdeformed anal region (Dar) response, severe constipation	MN	53
clec-39 *clec-49*	knockout mutant	survival egg laying	higher susceptibility lower egg laying		
CLEC-39 CLEC-49	ELISA-based assay with recombinant proteins	binding assays	binding to live/dead pathogen	SM	55
clec-52	promoter fusion reporter promoter fusion reporter	fluorescence induction fluorescence induction	expression in intestine expression in intestine	SA	52
clec-60	RNAi knockdown promoter fusion reporter	infection fluorescence induction	hyperdeformed anal region (Dar) response, severe constipation expression in intestine	MN SA	53 52
clec-60/61 cluster	overexpression	survival	higher susceptibility	PA	52
clec-65	RNAi knockdown	survival	higher susceptibility	EC	51
clec-67	promoter fusion reporter	fluorescence induction	expression in intestine	SE	56
clec-70	RNAi knockdown promoter fusion reporter	survival fluorescence induction	higher susceptibility expression in intestine	SA	52
clec-70/71 cluster	overexpression	survival	higher resistance higher susceptibility	SA PA	52
CLEC-79	glycoconjugate microarray with recombinant protein	binding assay	binding to Galb1-3GalNAc		54
clec-85	promoter fusion reporter	fluorescence intensity	dependence on immune components *tir-1*, *nsy-1*, *dbl-1, daf-16*	SA PA	78
clec-86	RNAi knockdown	infection	hyperdeformed anal region (Dar) response, severe constipation	MN	53
clec-174	RNAi knockdown	survival	higher susceptibility	VC	50

EC = *E. coli* strain LF82; SE = *S. enterica.* For other abbreviations, see the legend to figure 1.

Fig. 1. The highly specific pattern of *clec* gene regulation through pathogen exposure, intestinal expression of *C. elegans clec-66*, and analysis of protein sequence similarity between *C. elegans* CTLD proteins and CTLD proteins from 10 other taxa. **a** Hierarchical clustering of differential expressions of randomly selected *C. elegans clec* genes (rows) after exposure to different microbes (columns), grouped in Gram-positive bacteria (blue), Gram-negative bacteria (pink) and fungi (green). Each column represents *clec* up- (red) or downregulation (blue) according to published transcriptome studies. DB27 and I-3690 are strains of *B. thuringiensis* and *Lactobacillus rhamnosus*, respectively. *clec* expression after infection with *B. thuringiensis* [63] represents merged data from two *B. thuringiensis* concentrations and two time points. BM = *B. megaterium*; BT = *B. thuringiensis*; DC = *D. coniospora*; EC = *E. carotovora*; EF = *E. faecalis*; HA = *Harposporium* spp.; LR = *L. rhamnosus*; ML = *M. luteus*; MN = *M. nematophilum*; PA = *P. aeruginosa*; PL = *P. luminescens*; SA = *S. aureus*; SM = *S. marcescens*; VC = *V. cholerae*; XN = *X. nematophilus*; YP = *Y. pestis*. **b** GFP-labeled spores of the pathogenic *B. thuringiensis* strain MYBT18247 accumulate in the *C. elegans* anterior intestine at an early time point after infection. **c** Animals carrying a p*clec-66*::GFP reporter that is expressed throughout the intestine with highest expression in the anterior and posterior intestine. The coinjection marker p*ttx-3*::RFP (red) is expressed in head neurons. **d** The number of *C. elegans* CTLD proteins showing different levels of sequence similarity to the CTLD protein repertoire of other vertebrates and invertebrates. Sequence similarity was assessed with BLAST and a comparison of the 283 *C. elegans* CTLD proteins to those of 10 selected vertebrate and invertebrate taxa. The matrix shows the number of proteins falling into different percentage similarity categories. The observed variation is further highlighted by different shades of gray (dark gray for large numbers, light gray for small numbers).

There have been only two studies investigating the function of CLEC proteins on the protein level. One of these provided experimental evidence for the sugar-binding qualities of CLEC-79 [54], though not in the context of an immune response to infection, while the other demonstrated that CLEC-39 and CLEC-49 were able to bind live and dead *Serratia marcescens* in a Ca^{2+}-independent manner [55]. Both *clec-39* and *clec-49* seem to be required for nematode fitness in the course of an encounter with *S. marcescens* since knockout mutants produced lower reproductive and survival rates. The authors reasoned that due to a lack of bactericidal activity these CLEC proteins could act as PRRs, inducing downstream immune pathways after recognition of the pathogenic bacteria.

There is thus accumulating evidence that *clec* genes function in *C. elegans* resistance to pathogen infection. However, the exact role of the vast majority of CTLD proteins in *C. elegans* immunity remains unclear.

Expression of C. elegans clec Genes Is Regulated by Stress and Immunity Pathways

In *C. elegans* several highly conserved signaling pathways (e.g. p38 MAPK, insulin signaling and TGFβ) are required for the induction of immune effector molecules and, ultimately, host survival. The expression of numerous *clec* genes was shown to be under the control of these *C. elegans* immunity pathways and other known regulators of immunity, such as the intestine-specific GATA transcription factor ELT-2 [56] or the E-box transcription factor HLH-30 [57].

Taken together, CTLD proteins are highly diversified in the *C. elegans* genome with a potential further extension of binding capabilities through oligomerization and internal domain duplications. Experimental evidence demonstrates a *clec* gene function in *C. elegans* innate immunity, their regulation by known immunity pathways, and their highly specific response to different pathogens. Especially the latter emphasizes the particular potential of *C. elegans* CTLD proteins to generate immune specificity.

Similarity of *C. elegans* CTLD Proteins to Those from Other Taxa

Although CTLD proteins from the various taxa are related through the presence of the conserved CTL domain, it is not clear how similar the complete protein sequences are to each other. To address this point, we used the *C. elegans* CTLD protein repertoire as an example and determined their sequence similarity to the CTLD proteins from a selected range of 10 invertebrate and vertebrate taxa, including *Homo sapiens*, *M. musculus*, *D. melanogaster* and *Nematostella vectensis*. This analysis revealed that similarities above 70% are only found for comparisons within the same genus (fig. 1d). In comparisons to distinct taxonomic groups, similarity scores for the large majority of proteins are below 50% and none of the comparisons revealed a value of more than 60%. Interestingly, even in the comparisons to the other *Caenorhabditis* taxa, approximately half of the proteins produce similarity scores of less than 50%.

These results highlight that: (i) CTLD proteins are primarily related through the presence of the CTL domain and a likely function in immunity (at least indicated for numerous CTLD proteins), but otherwise seem to show high diversification in the various phylogenetic lineages, (ii) CTLD proteins thus appear to comprise a highly dynamic, fast-evolving gene family, possibly in response to selective pressure from pathogens, and (iii) information on the function of a certain CTLD protein from one taxon, for example *C. elegans*, cannot be directly transferred onto a specific CTLD protein from a different taxon, for example humans, even though general functional similarities can be deduced through a comparative approach. These findings reinforce our notion that CTLD proteins represent a gene family that is of particular interest to our understanding of the evolution of innate and invertebrate immune systems, and especially their dynamic nature within different phylogenetic lineages.

Current Limitations in Our Understanding of Invertebrate CTLD Proteins

The recently described, surprisingly high degree of specificity of innate immune systems must be based on as yet unknown underlying molecular mechanisms. One of the possibilities may rely on genomically diversified gene families, such as those of the CTLD genes. These genes are numerous in many invertebrate species (table 1) and possess several additional characteristics that support a role in high immune specificity. In the model nematode *C. elegans*, the expression of many *clec* genes is regulated through pathogen exposure in a highly specific form (fig. 1a) [47]. CTLD proteins play an important role in the innate immune system of various invertebrate taxa by activating cellular, and in one case humoral [43], immune responses. Several insect CTLD proteins were found to specifically bind to Gram-negative [18, 20, 31, 34] or Gram-positive bacteria [25]. On the protein level, further

functional diversity of CTLD proteins might be achieved through oligomerization and internal domain duplications.

To date, however, the exact role of invertebrate CTLD proteins in generating immune specificity is completely unexplored. In most cases, only single CTLD proteins from a particular invertebrate species have been studied, but not a larger set of the gene family. On the protein level, some invertebrate CTLD proteins were found to bind to either a broad range of microorganisms (both Gram-positive and Gram-negative bacteria) or alternatively only to a particular pathogen group (Gram-negative or Gram-positive bacteria), already indicating a certain level of immune specificity. At the moment, none of the studied CTLD proteins specifically bind to a single pathogen species. The reasons for this might be that most, if not all, binding analyses on CTLD proteins are performed in vitro and might not reflect what is happening during pathogen infection in vivo. Moreover, they usually do not consider the pathogen variants, which coexist and coevolve with the host in nature and against which high immune specificity may thus have evolved across time. In the future, a larger variety of CTLD proteins should be studied for a particular species. For these studies, it is important that CTLD gene function is analyzed both in vitro and in vivo. The latter is particularly important (yet also challenging) because it can reveal to what extent interactions with other proteins, dimerization and oligomerization might influence the binding properties of CTLD proteins. In addition, the identification of highly specific host-pathogen interactions requires a study system in which the host is infected in a natural way with a natural pathogen and obviously with different strains of the same pathogen species, a setup that is so far not available for studies in *C. elegans*, crustacean and insect models.

Moreover, at the moment, the general functions of the majority of invertebrate CTLD proteins have also not been characterized. While the studies on single CTLD proteins in invertebrates provide important evidence on their role in innate immunity, an understanding of their exact properties and overall functions is far from complete. For example, we lack detailed information regarding how the expression of CTLD proteins is regulated and controlled, or their exact binding properties and targets. In *C. elegans*, CTLD proteins have repeatedly been suggested to be involved in immune defense by acting as antimicrobial effectors, and signaling pathways regulating CTLD gene expression have been identified, whereas in insects and crustaceans CTLD proteins are mainly assumed to act as PRRs activating cellular immune responses. This discrepancy might be due to the fact that in *C. elegans* comprehensive genomic and transcriptomic data sets are available, which revealed that the majority of *C. elegans* CTLD proteins are induced upon pathogen infection and also contain a signal peptide and are thus likely to be secreted. In addition, several *C. elegans* CTLD gene knockout or knockdown animals have been phenotypically characterized, supporting the view that CTLD proteins are important for host resistance to infection. Nevertheless, for this nematode, we currently lack data on CTLD protein interactions with other proteins, the possible downstream molecular signaling pathways, and also data from biochemical assays to infer CTLD binding properties on the protein level.

In contrast, such analyses at the protein level are available for different crustaceans and insects. In particular, in vitro studies using recombinant or native proteins revealed the binding characteristics of several CTLD proteins and provided information on the involvement of CTLD proteins in activating cellular immune defenses. However, in these taxa information from functional genetic studies to confirm a role of CTLD proteins in immune defense in vivo are extremely scarce. Such functional genetic analyses may represent a particular challenge for such large gene families because of functional redundancy among paralogs. Thus, a single gene knockout or knockdown may not necessarily result in a visible phenotype even if the gene is of functional importance. Two approaches have been successfully used to target functionally redundant genes and may prove useful in the future: the simultaneous introduction of mutations in several or all members of a gene family using the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein 9 (Cas9) system [58, 59] and the simultaneous knockdown of several genes using RNAi-based approaches [60].

To conclude, it remains unclear for most invertebrate CTLD proteins if they either act as PRRs, directly recognizing pathogens by binding to MAMPs, as signaling mediators or as antimicrobial effectors. To identify mechanisms of CTLD protein action biochemical analyses on the purified native or recombinant proteins to identify binding capabilities and interacting proteins need to be combined with assays on bactericidal effects and also in vivo functional genetic analyses in the respective host taxa. The study by Wang et al. [44] represents a first example of such an integrated approach.

Even though our review focused on invertebrates, we would like to note that similar CTLD protein diversifica-

tion within and across species is also found in vertebrates. The potential role of CTLD proteins in contributing to immune specificity may thus also be of relevance in vertebrate taxa. Consistent with this idea, Drickamer and Taylor [61] proposed in a recent review that mammalian CTLD proteins, which bind to endogenous or self-glycans, are generally conserved, while those which bind to MAMPs are generally divergent and vary in number among species. Examples of such species-specific differences are: (i) two SIGN molecules expressed in humans (DC-SIGN and DC-SIGNR) compared to eight SIGN molecules expressed in mice, (ii) two loci coding for functional mannose-binding lectin in mice compared to only one in humans [61], (iii) one BDCA-2 gene in humans with no reported homologue in mice and (iv) one DCIR-coding locus in humans compared to four (Dcir1–4) in mice [62]. The CTLD proteins in mice and humans do not only differ in number but also in their binding specificities, suggesting that they could indeed contribute to specific responses to pathogens and that their diversification may be a consequence of an arms race with specific coevolving pathogens enabling the host to meet the particular challenges imposed by rapidly evolving pathogens.

Acknowledgements

We thank Andrei Papkou for providing the picture of *C. elegans* infected with GFP-labeled *B. thuringiensis* spores. B.P., W.Y., H.S. and K.D. are supported by grants from the German Science foundation (Grant DI 1687/1-1 to K.D. and SCHU 1415/9-2 to H.S.). A.Z.-P. and W.Y. are members of the IMPRS for Evolutionary Biology. K.D. is additionally supported by institutional funding from Kiel University.

References

1 Little TJ, O'Connor B, Colegrave N, Watt K, Read AF: Maternal transfer of strain-specific immunity in an invertebrate. Curr Biol 2003; 13:489–492.
2 Roth O, Sadd BM, Schmid-Hempel P, Kurtz J: Strain-specific priming of resistance in the red flour beetle, *Tribolium castaneum*. Proc R Soc B Biol Sci 2009;276:145–151.
3 Sadd BM, Schmid-Hempel P: Insect immunity shows specificity in protection upon secondary pathogen exposure. Curr Biol 2006; 16:1206–1210.
4 Watson FL: Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science 2005;309:1874–1878.
5 Watthanasurorot A, Jiravanichpaisal P, Liu H, Söderhäll I, Söderhäll K: Bacteria-induced Dscam isoforms of the Crustacean, *Pacifastacus leniusculus*. PLoS Pathog 2011;7:e1002062.
6 Armitage SAO, Peuss R, Kurtz J: Dscam and pancrustacean immune memory – a review of the evidence. Dev Comp Immunol 2014;48: 315–323.
7 Cannon JP, Haire RN, Schnitker N, Mueller MG, Litman GW: Individual protochordates have unique immune-type receptor repertoires. Curr Biol 2004;14:R465–R466.
8 Zhang S-M, Adema CM, Kepler TB, Loker ES: Diversification of Ig superfamily genes in an invertebrate. Science 2004;305:251–254.
9 Ghosh J, Buckley KM, Nair SV, Raftos DA, Miller C, Majeske AJ, et al: Sp185/333: a novel family of genes and proteins involved in the purple sea urchin immune response. Dev Comp Immunol 2010;34:235–245.
10 Hibino T, Loza-Coll M, Messier C, Majeske AJ, Cohen AH, Terwilliger DP, et al: The immune gene repertoire encoded in the purple sea urchin genome. Dev Biol 2006;300:349–365.
11 Huang S, Yuan S, Guo L, Yu Y, Li J, Wu T, et al: Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. Genome Res 2008;18:1112–1126.
12 Weis WI, Taylor ME, Drickamer K: The C-type lectin superfamily in the immune system. Immunol Rev 1998;163:19–34.
13 Zelensky AN, Gready JE: The C-type lectin-like domain superfamily. FEBS J 2005;272: 6179–6217.
14 Hoving JC, Wilson GJ, Brown GD: Signalling C-type lectin receptors, microbial recognition and immunity: C-type lectins in immunity. Cell Microbiol 2014;16:185–194.
15 Vaishnava S, Yamamoto M, Severson KM, Ruhn KA, Yu X, Koren O, et al: The antibacterial lectin RegIIIγ promotes the spatial segregation of microbiota and host in the intestine. Science 2011;334:255–258.
16 Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al: InterPro: the integrative protein signature database. Nucleic Acids Res 2009;37:D211–D215.
17 Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al: Pfam: the protein families database. Nucleic Acids Res 2014;42:D222–D230.
18 Ao J, Ling E, Yu X-Q: *Drosophila* C-type lectins enhance cellular encapsulation. Mol Immunol 2007;44:2541–2548.
19 Haq S, Kubo T, Kurata S, Kobayashi A, Natori S: Purification, characterization, and cDNA cloning of a galactose-specific C-type lectin from *Drosophila melanogaster*. J Biol Chem 1996;271:20213–20218.
20 Tanji T, Ohashi-Kobayashi A, Natori S: Participation of a galactose-specific C-type lectin in *Drosophila* immunity. Biochem J 2006;396: 127–138.
21 Koizumi N, Morozumi A, Imamura M, Tanaka E, Iwahana H, Sato R: Lipopolysaccharide-binding proteins and their involvement in the bacterial clearance from the hemolymph of the silkworm *Bombyx mori*. Eur J Biochem 1997;248:217–224.
22 Wilson R, Chen C, Ratcliffe NA: Innate immunity in insects: the role of multiple, endogenous serum lectins in the recognition of foreign invaders in the cockroach, *Blaberus discoidalis*. J Immunol 1999;162:1590–1596.
23 Yu X-Q, Zhu Y-F, Ma C, Fabrick JA, Kanost MR: Pattern recognition proteins in *Manduca sexta* plasma. Insect Biochem Mol Biol 2002; 32:1287–1293.
24 Koizumi N, Imai Y, Morozumi A, Imamura M, Kadotani T, Yaoi K, et al: Lipopolysaccharide-binding protein of *Bombyx mori* participates in a hemocyte-mediated defense reaction against Gram-negative bacteria. J Insect Physiol 1999;45:853–859.
25 Chai L-Q, Tian Y-Y, Yang D-T, Wang J-X, Zhao X-F: Molecular cloning and characterization of a C-type lectin from the cotton bollworm, *Helicoverpa armigera*. Dev Comp Immunol 2008;32:71–83.
26 Watanabe A, Miyazawa S, Kitami M, Tabunoki H, Ueda K, Sato R: Characterization of a novel C-type lectin, *Bombyx mori* multibinding protein, from the *B. mori* hemolymph: mechanism of wide-range microorganism recognition and role in immunity. J Immunol 2006;177:4594–4604.

27 Kanost MR, Jiang H, Yu X-Q: Innate immune responses of a lepidopteran insect, *Manduca sexta*. Immunol Rev 2004;198:97–105.
28 Yu X-Q, Tracy ME, Ling E, Scholz FR, Trenczek T: A novel C-type immulectin-3 from *Manduca sexta* is translocated from hemolymph into the cytoplasm of hemocytes. Insect Biochem Mol Biol 2005;35:285–295.
29 Yu X-Q, Ling E, Tracy ME, Zhu Y: Immulectin-4 from the tobacco hornworm *Manduca sexta* binds to lipopolysaccharide and lipoteichoic acid. Insect Mol Biol 2006;15:119–128.
30 Ausubel FM: Are innate immune signaling pathways in plants and animals conserved? Nat Immunol 2005;6:973–979.
31 Yu X-Q, Kanost MR: Immulectin-2, a lipopolysaccharide-specific lectin from an insect, *Manduca sexta*, is induced in response to Gram-negative bacteria. J Biol Chem 2000; 275:37373–37381.
32 Eleftherianos I, Millichap PJ, Ffrench-Constant RH, Reynolds SE: RNAi suppression of recognition protein mediated immune responses in the tobacco hornworm *Manduca sexta* causes increased susceptibility to the insect pathogen *Photorhabdus*. Dev Comp Immunol 2006;30:1099–1107.
33 Osta MA, Christophides GK, Kafatos FC: Effects of mosquito genes on plasmodium development. Science 2004;303:2030–2032.
34 Schnitger AKD, Yassine H, Kafatos FC, Osta MA: Two C-type lectins cooperate to defend *Anopheles gambiae* against Gram-negative bacteria. J Biol Chem 2009;284:17616–17624.
35 Sánchez-Paz A: White spot syndrome virus: an overview on an emergent concern. Vet Res 2010;41:43.
36 Wang X-W, Wang J-X: Diversity and multiple functions of lectins in shrimp immunity. Dev Comp Immunol 2013;39:27–38.
37 Huang Y, Huang X, Wang Z, Tan J-M, Hui K-M, Wang W, et al: Function of two novel single-CRD containing C-type lectins in innate immunity from *Eriocheir sinensis*. Fish Shellfish Immunol 2014;37:313–321.
38 Wang X-W, Zhang X-W, Xu W-T, Zhao X-F, Wang J-X: A novel C-type lectin (FcLec4) facilitates the clearance of *Vibrio anguillarum* in vivo in Chinese white shrimp. Dev Comp Immunol 2009;33:1039–1047.
39 Zhang X-W, Liu Y-Y, Mu Y, Ren Q, Zhao X-F, Wang J-X: Overexpression of a C-type lectin enhances bacterial resistance in red swamp crayfish *Procambarus clarkii*. Fish Shellfish Immunol 2013;34:1112–1118.
40 Sun Y-D, Fu L-D, Jia Y-P, Du X-J, Wang Q, Wang Y-H, et al: A hepatopancreas-specific C-type lectin from the Chinese shrimp *Fenneropenaeus chinensis* exhibits antimicrobial activity. Mol Immunol 2008;45:348–361.
41 Jin X-K, Li S, Guo X-N, Cheng L, Wu M-H, Tan S-J, et al: Two antibacterial C-type lectins from crustacean, *Eriocheir sinensis*, stimulated cellular encapsulation in vitro. Dev Comp Immunol 2013;41:544–552.
42 Luo T, Yang H, Li F, Zhang X, Xu X: Purification, characterization and cDNA cloning of a novel lipopolysaccharide-binding lectin from the shrimp *Penaeus monodon*. Dev Comp Immunol 2006;30:607–617.
43 Wang X-W, Xu J-D, Zhao X-F, Vasta GR, Wang J-X: A shrimp C-type lectin inhibits proliferation of the hemolymph microbiota by maintaining the expression of antimicrobial peptides. J Biol Chem 2014;289:11779–11790.
44 Wang X-W, Zhao X-F, Wang J-X: C-type lectin binds to β-integrin to promote hemocytic phagocytosis in an invertebrate. J Biol Chem 2014;289:2405–2414.
45 Shiratsuchi A, Mori T, Sakurai K, Nagaosa K, Sekimizu K, Lee BL, et al: Independent recognition of *Staphylococcus aureus* by two receptors for phagocytosis in *Drosophila*. J Biol Chem 2012;287:21663–21672.
46 Li M, Li C, Ma C, Li H, Zuo H, Weng S, et al: Identification of a C-type lectin with antiviral and antibacterial activity from pacific white shrimp *Litopenaeus vannamei*. Dev Comp Immunol 2014;46:231–240.
47 Schulenburg H, Hoeppner MP, Weiner J, Bornberg-Bauer E: Specificity of the innate immune system and diversity of C-type lectin domain (CTLD) proteins in the nematode *Caenorhabditis elegans*. Immunobiology 2008;213:237–250.
48 Petersen TN, Brunak S, von Heijne G, Nielsen H: SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 2011;8:785–786.
49 Schulenburg H, Kurz CL, Ewbank JJ: Evolution of the innate immune system: the worm perspective. Immunol Rev 2004;198:36–58.
50 Sahu SN, Lewis J, Patel I, Bozdag S, Lee JH, LeClerc JE, et al: Genomic analysis of immune response against *Vibrio cholerae* hemolysin in *Caenorhabditis elegans*. PLoS One 2012; 7:e38200.
51 Simonsen KT, Møller-Jensen J, Kristensen AR, Andersen JS, Riddle DL, Kallipolitis BH: Quantitative proteomics identifies ferritin in the innate immune response of *C. elegans*. Virulence 2011;2:120–130.
52 Irazoqui JE, Troemel ER, Feinbaum RL, Luhachack LG, Cezairliyan BO, Ausubel FM: Distinct pathogenesis and host responses during infection of *C. elegans* by *P. aeruginosa* and *S. aureus*. PLoS Pathog 2010;6:e1000982.
53 O'Rourke D, Baban D, Demidova M, Mott R, Hodgkin J: Genomic clusters, putative pathogen recognition molecules, and antimicrobial genes are induced by infection of *C. elegans* with *M. nematophilum*. Genome Res 2006;16: 1005–1016.
54 Takeuchi T, Sennari R, Sugiura K, Tateno H, Hirabayashi J, Kasai K: A C-type lectin of *Caenorhabditis elegans*: its sugar-binding property revealed by glycoconjugate microarray analysis. Biochem Biophys Res Commun 2008;377:303–306.
55 Miltsch SM, Seeberger PH, Lepenies B: The C-type lectin-like domain containing proteins Clec-39 and Clec-49 are crucial for *Caenorhabditis elegans* immunity against *Serratia marcescens* infection. Dev Comp Immunol 2014;45:67–73.
56 Kerry S, TeKippe M, Gaddis NC, Aballay A: GATA transcription factor required for immunity to bacterial and fungal pathogens. PLoS One 2006;1:e77.
57 Visvikis O, Ihuegbu N, Labed SA, Luhachack LG, Alves A-MF, Wollenberg AC, et al: Innate host defense requires TFEB-mediated transcription of cytoprotective and antimicrobial genes. Immunity 2014;40:896–909.
58 Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, et al: One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 2013;153:910–918.
59 Peng D, Kurup SP, Yao PY, Minning TA, Tarleton RL: CRISPR-Cas9-mediated single-gene and gene family disruption in *Trypanosoma cruzi*. mBio 2014;6:e02097-14.
60 Kim J, Badaloni A, Willert T, Zimber-Strobl U, Kühn R, Wurst W, et al: An RNAi-based approach to down-regulate a gene family in vivo. PLoS One 2013;8:e80312.
61 Drickamer K, Taylor ME: Recent insights into structures and functions of C-type lectins in the immune system. Curr Opin Struct Biol 2015;34:26–34.
62 Dambuza IM, Brown GD: C-type lectins in immunity: recent developments. Curr Opin Immunol 2015;32:21–27.
63 Yang W, Dierking K, Esser D, Tholey A, Leippe M, Rosenstiel P, et al: Overlapping and unique signatures in the proteomic and transcriptomic responses of the nematode *Caenorhabditis elegans* toward pathogenic *Bacillus thuringiensis*. Dev Comp Immunol 2015; 51:1–9.
64 Coolon JD, Jones KL, Todd TC, Carr BC, Herman MA: *Caenorhabditis elegans* genomic response to soil bacteria predicts environment-specific genetic effects on life history traits. PLoS Genet 2009;5:e1000503.
65 Wong D, Bazopoulou D, Pujol N, Tavernarakis N, Ewbank JJ: Genome-wide investigation reveals pathogen-specific and shared signatures in the response of *Caenorhabditis elegans* to infection. Genome Biol 2007;8:R194.
66 Boehnisch C, Wong D, Habig M, Isermann K, Michiels NK, Roeder T, et al: Protist-type lysozymes of the nematode *Caenorhabditis elegans* contribute to resistance against pathogenic *Bacillus thuringiensis*. PLoS One 2011; 6:e24619.
67 Treitz C, Cassidy L, Höckendorf A, Leippe M, Tholey A: Quantitative proteome analysis of *Caenorhabditis elegans* upon exposure to nematicidal *Bacillus thuringiensis*. J Proteomics 2015;113:337–350.

68 Sinha A, Rae R, Iatsenko I, Sommer RJ: System wide analysis of the evolution of innate immunity in the nematode model species *Caenorhabditis elegans* and *Pristionchus pacificus*. PLoS One 2012;7:e44255.
69 Grompone G, Martorell P, Llopis S, González N, Genovés S, Mulet AP, et al: Anti-inflammatory *Lactobacillus rhamnosus* CNCM I-3690 strain protects against oxidative stress and increases lifespan in *Caenorhabditis elegans*. PLoS One 2012;7:e52493.
70 Bolz DD, Tenor JL, Aballay A: A conserved PMK-1/p38 MAPK is required in *Caenorhabditis elegans* tissue-specific immune response to *Yersinia pestis* infection. J Biol Chem 2010; 285:10832–10840.
71 Shapira M, Hamlin BJ, Rong J, Chen K, Ronen M, Tan M-W: A conserved role for a GATA transcription factor in regulating epithelial innate immune responses. Proc Natl Acad Sci USA 2006;103:14086–14091.
72 Engelmann I, Griffon A, Tichit L, Montañana-Sanchis F, Wang G, Reinke V, et al: A comprehensive analysis of gene expression changes provoked by bacterial and fungal infection in *C. elegans*. PLoS One 2011;6:e19055.
73 Drummond RA, Brown GD: Signalling C-type lectins in antimicrobial immunity. PLoS Pathog 2013;9:e1003417.
74 Yu Y, Yu Y, Huang H, Feng K, Pan M, Yuan S, et al: A short-form C-type lectin from *Amphioxus* acts as a direct microbial killing protein via interaction with peptidoglycan and glucan. J Immunol 2007;179:8425–8434.
75 Multerer KA, Smith LC: Two cDNAs from the purple sea urchin, *Strongylocentrotus purpuratus*, encoding mosaic proteins with domains found in factor H, factor I, and complement components C6 and C7. Immunogenetics 2004;56:89–106.
76 Terwilliger DP, Clow LA, Gross PS, Smith LC: Constitutive expression and alternative splicing of the exons encoding SCRs in Sp152, the sea urchin homologue of complement factor B: implications on the evolution of the Bf/C2 gene family. Immunogenetics 2004;56:531–543.
77 Koizumi N, Imamura M, Kadotani T, Yaoi K, Iwahana H, Sato R: The lipopolysaccharide-binding protein participating in hemocyte nodule formation in the silkworm *Bombyx mori* is a novel member of the C-type lectin superfamily with two different tandem carbohydrate-recognition domains. FEBS Lett 1999;443:139–143.
78 Alper S, McBride SJ, Lackford B, Freedman JH, Schwartz DA: Specificity and complexity of the *Caenorhabditis elegans* innate immune response. Mol Cell Biol 2007;27:5544–5553.
79 Troemel ER, Chu SW, Reinke V, Lee SS, Ausubel FM, Kim DH: p38 MAPK regulates expression of immune response genes and contributes to longevity in *C. elegans*. PLoS Genet 2006;2:e183.
80 Mallo GV, Kurz CL, Couillault C, Pujol N, Granjeaud S, Kohara Y, et al: Inducible antibacterial defense system in *C. elegans*. Curr Biol 2002;12:1209–1214.
81 Huffman DL, Abrami L, Sasik R, Corbeil J, van der Goot FG, Aroian RV: Mitogen-activated protein kinase pathways defend against bacterial pore-forming toxins. Proc Natl Acad Sci USA 2004;101:10995–11000.
82 Pujol N, Zugasti O, Wong D, Couillault C, Kurz CL, Schulenburg H, et al: Anti-fungal innate immunity in *C. elegans* is enhanced by evolutionary diversification of antimicrobial peptides. PLoS Pathog 2008;4:e1000105.