Developmental and Comparative Immunology 35 (2011) 959–974

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

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

Invertebrate immune diversity

Julie Ghosh[1], Cheng Man Lun[1], Audrey J. Majeske[1], Sandro Sacchi[1], Catherine S. Schrankel[1,2], L. Courtney Smith[*,1]

Department of Biological Sciences, George Washington University, Washington, DC United States

ARTICLE INFO

Article history:
Available online 21 December 2010

Keywords:
Penaeidins
Crustins
FREPs
Fuhc
Fester
Uncle fester
Dscam
PGRP
VCBP
Sp185/333
Innate immunity
Immune diversification
Gene family evolution

ABSTRACT

The arms race between hosts and pathogens (and other non-self) drives the molecular diversification of immune response genes in the host. Over long periods of evolutionary time, many different defense strategies have been employed by a wide variety of invertebrates. We review here penaeidins and crustins in crustaceans, the allorecognition system encoded by *fuhc*, *fester* and *Uncle fester* in a colonial tunicate, Dscam and PGRPs in arthropods, FREPs in snails, VCBPs in protochordates, and the *Sp185/333* system in the purple sea urchin. Comparisons among immune systems, including those reviewed here have not identified an immune specific regulatory "genetic toolkit", however, repeatedly identified sequences (or "building materials" on which the tools act) are present in a broad range of immune systems. These include a Toll/TLR system, a primitive complement system, an LPS binding protein, and a RAG core/Transib element. Repeatedly identified domains and motifs that function in immune proteins include NACHT, LRR, Ig, death, TIR, lectin domains, and a thioester motif. In addition, there are repeatedly identified mechanisms (or "construction methods") that generate sequence diversity in genes with immune function. These include genomic instability, duplications and/or deletions of sequences and the generation of clusters of similar genes or exons that appear as families, gene recombination, gene conversion, retrotransposition, alternative splicing, multiple alleles for single copy genes, and RNA editing. These commonly employed "materials and methods" for building and maintaining an effective immune system that might have been part of that ancestral system appear now as a fragmented and likely incomplete set, likely due to the rapid evolutionary change (or loss) of host genes that are under pressure to keep pace with pathogen diversity.

1. Introduction

Host immunity, by its very nature, is an ongoing arms race against invading pathogens, with the onus on pathogens to outwit host immunity, and on the host to successfully eliminate invaders. This long-term host-pathogen co-evolutionary history involves high rates of mutation and/or variation in microbes that generally have short generation times versus long-lived hosts with corresponding low mutation rates [1], which encounter and must defend against many different pathogen types. On the host side, animals employ pattern recognition receptors (PRRs) that generically recognize the pathogen-associated molecular patterns (PAMPs) of the important classes of microbes, including lipopolysaccharide (LPS), peptidoglycans (PGNs), double stranded RNA, and β-glucans among others. On stimulation, these PRRs activate downstream signaling pathways that result in either the release of antimicrobial molecules, such as those known for *Drosophila* [2], and/or the stimulation of a cell-mediated phagocytic response [3]. Although these host responses are broadly successful, pathogens have evolved adaptations, such as altered surface molecules [4] to enable them to evade detection or immune attack. Consequently, for survival, the host requires a repertoire of receptor and effector molecules that is capable of detecting foreign molecules more diverse than basic PAMPs, and of mounting an effective response. Because the precise pathogen adaptations that are encountered are necessarily unpredictable, the best option for the host – and the one that is characteristic of immune systems across the animal kingdom – is to find ways to generate random or near random diversification and

Abbreviations: 4DSC, four-disulfide core; aa, amino acids; AMPs, antimicrobial peptides; BAC, bacterial artificial chromosome; CRD, cysteine-rich domain; EGF, epidermal growth factor; FBG, fibrinogen; FNIII, fibronectin type III; GNBP1, Gram-negative binding protein 1; Ig, immunoglobulin; IgSF, immunoglobulin superfamily; indels, insertions/deletions; LPS, lipopolysaccharide; NACHT, NAIP/CIITA/HET-E/TP1; NITR, novel immune-type receptors; PAMPs, pathogen-associated molecular patterns; PGNs, peptidoglycans; PRD, proline-rich domain; PRR, pattern recognition receptor; *R*, resistance; RGD, arginine-glycine-aspartic acid; SCR, short consensus repeat; SEPs, secretory-excretory products; SNPs, single nucleotide polymorphisms; TCRs, T-cell receptors; TEPs, thioester-containing proteins; TIR, Toll/IL-1 receptor; TM, transmembrane; WAP, whey acidic protein.

* Corresponding author at: Tel.: +1 202 994 9211; fax: +1 202 994 6100.
E-mail address: csmith@gwu.edu (L.C. Smith).

[1] All authors contributed equally and are listed in alphabetical order.

[2] Current address: Department of Immunology, University of Toronto, Toronto Canada.

0145-305X/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.dci.2010.12.009

Penaeidin Type	Length of Signal Sequence	Length of Pro-rich domain	Length of Cys-rich domain
PEN2	21	23-25	26-28
PEN3	19	21-31	25-32
PEN4	19	22	26
PEN5	19	32	28

Fig. 1. Schematic representation of the domain organization of penaeidins. The table lists the number of amino acids in each domain by penaeidin type. SS, signal sequence. The diagram is not drawn to scale.
Sources: Penbase http://www.penbase.immunaqua.com; [6]. The table is reprinted with modifications from [6] with kind permission from Springer Science+Business Media.

expansion of immune receptors so that the largest possible range of pathogens, regardless of the type of evasive adaptation evolved, will be detectable by the immune system as foreign.

There are many different strategies among animals to expand the repertoire of immune responsiveness. Although the somatic rearrangements involved in diversification of the T cell receptor genes (TCRs) and immunoglobulins (Igs) in mammals are well known and the assembly of the variable lymphocyte receptor (VLR) genes in the agnathans [5] is an intriguing alternative solution to immune diversification, these mechanisms are not observed outside of the vertebrates. Invertebrates employ other approaches that include (i) the presence of large gene families either within individuals or within populations that encode a wide variety of protein isoforms, (ii) genomic instability within the families of similar genes that promote unequal crossovers, gene conversion, gene duplication/deletion and paralogous mispairing, all of which promote sequence diversification, (iii) a variety of modifications to mRNAs including alternative splicing, RNA editing and low fidelity RNA polymerases, and (iv) a broad array of modifications to proteins either during or after translation. Several examples of immune genes and gene families in invertebrates that show sequence diversification are described below.

2. Antimicrobial peptides (AMPs) in crustaceans

AMPs are found in a wide variety of living organisms, including bacteria, fungi, plants, and animals, and are an important aspect of the innate immune response [6]. Most AMPs are known to be immunomodulators that are active against Gram-negative and Gram-positive bacteria, yeast, fungi, parasites, enveloped viruses and tumor cells, and some AMPs kill pathogens directly *in vitro* [6–9]. AMPs are typically small cationic amphipathic molecules that range in size from 15 to 200 amino acids (aa) in length, but are rarely larger than 30 kDa [10]. They are classified into three major groups based on aa sequence, secondary structure and functional properties. There has been particular interest in identifying AMPs in shrimp because of a growing number of diseases that affect this economically important group, and the notion that understanding shrimp AMPs may lead to therapeutic applications to curb the loss of shrimp production from infections [11]. Two major shrimp AMPs are the cysteine-rich penaeidins and crustins. Each shows sequence diversity and is comprised of multiple classes (penaeidins) and types (crustins) that are synthesized mostly in hemocytes and are released into the hemolymph in response to infection [6].

2.1. Penaeidins

There are four classes of penaeidins in penaeid shrimp and each class has several isoforms (Fig. 1) [12–14]. Penaeidins are small peptides of 5–7 kDa with an N-terminal signal peptide region followed by a proline-rich domain (PRD) and C-terminal cysteine-rich domain (CRD) containing six cysteine residues ([15,16]; see PenBase http://www.penbase.immunaqua.com for all (>200) penaeidins). The N-terminal PRD is longer than the CRD and is free of disulfide bonds, thus making it less rigid, whereas the C-terminal CRD is more conserved across classes and is stabilized by three disulfide bonds [15,17,18]. The sequence diversity within the PRD among different penaeidin classes is likely the source of variation in anti-microbial responses [17,19] based on evidence that the CRD of PEN4 may not be necessary for antimicrobial function [17,20]. However, the combination of both CRD and PRD domains is essential to achieve the maximum and specific antimicrobial response and the CRD domain may function to potentiate antimicrobial activation [17]. Overall, the presence of different penaeidin classes and isoforms within and among different shrimp species (Table 1) indicates that shrimp AMPs make up a large and diverse family.

The diversity among different penaeidins can be measured at several different levels. Variation exists in gene composition and length within the same and among different classes of penaeidins within and among shrimp species. For example, there is variation in the number of exons in different penaeidin classes, and introns may be of variable length or may be absent from genes within a penaeidin class among different shrimp species [6,16,21]. The expression levels of different penaeidin classes within individual shrimp can vary as can the functional specificity of each AMP [6,16]. In general, most penaeidins show variable activity against Gram-positive bacteria and fungi and have chitin-binding properties, however, PEN5 is also active against Gram-negative bacteria [17,21–23]. In essence, variation among penaeidins is derived from variations in gene sequences, variation in gene expression, and variations in the activity of the proteins against microbes.

Table 1
Penaeidins in shrimp.

Species	Penaeidin type	No. of isoforms identified
Farfantepenaeus brasiliensis	PEN2	1
Farfantepenaeus paulensis	PEN2	2
Farfantepenaeus subtilis	PEN2	1
Fenneropenaeus chinensis	PEN3, PEN5	3, 4
Fenneropenaeus penicillatus	PEN3	2
Litopenaeus schmitt	PEN2, PEN3	2, 2
Litopenaeus setiferus	PEN2, PEN3, PEN4	2, 4, 2
Litopenaeus stylirostris	PEN2, PEN3	1, 2
Litopenaeus vannamei	PEN2, PEN3, PEN4	3, 15, 3
Penaeus monodon	PEN3, PEN5	8, 4
Penaeus semisulcatus	PEN3	1*

Modified from [6].
* Retrieved from PenBase http://www.penbase.immunaqua.com.

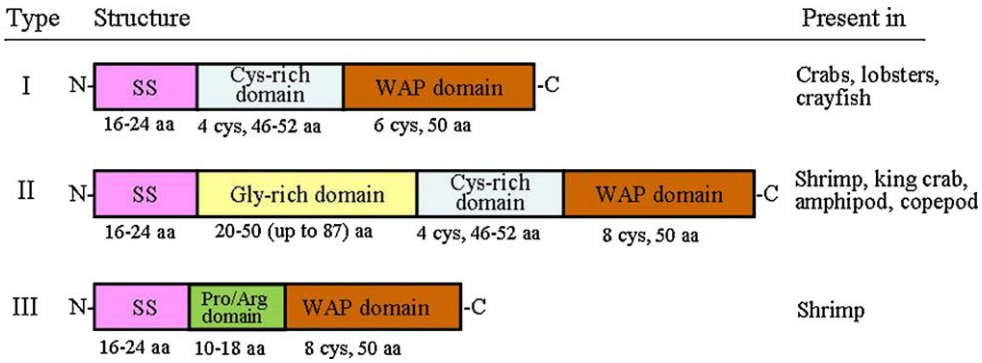

Fig. 2. Schematic representation of domain organization (not to scale) of three types of crustins. SS, signal sequence; aa, amino acids; WAP, whey acidic proteins. The diagram is not drawn to scale.

Reprinted with minor modifications from [31] with permission from Elsevier.

Recently, hyastatin, a penaeidin-like AMP, was discovered in the spider crab (*Hyas araneus*) and has a glycine-rich N-terminus, a short proline–arginine-rich region, followed by a C-terminal domain containing six cysteines [24]. Hyastatin is active against Gram-negative and Gram-positive bacteria and yeast, and a recombinant N-terminal glycine-rich fragment shows chitin-binding properties and weak activity against Gram-positive bacteria *in vitro*. These results suggest that hyastatin may have multiple domains with antimicrobial function [25] and may be a chimeric protein that is the result of a gene recombination and fusion event. The chimeric hyastatin has multiple domains that would have originally been encoded by the two parental genes, the result being an improved and more efficient variety of antimicrobial activities in the chimeric hyastatin compared to the activities of the two parental proteins. The concept of chimeric proteins may also be applied to the presence of both proline-rich and cysteine-rich domains in shrimp penaeidins, which may be under different pathogen pressures. However, this speculation will require additional work for verification.

The genetic mechanisms that generate sequence variability within the different penaeidins are not known [16]. There is no evidence to indicate that diversity among members of a class or an individual isoform is generated from post-translational modifications. Rather, all of the diversity within each class and among isoforms of penaeidins results from variation in the gene sequences, and it is therefore thought that each penaeidin is encoded by its own unique gene. Furthermore, phylogenetic analysis suggests that regardless of species, multiple copies of penaeidin genes within each class of penaeidin cluster together with strong nodal support, indicating that penaeidin genes are paralogous and that each class may have expanded by gene duplication events [26].

Positive selection from pathogen pressure acts on both the PRD and CRD, according to the average ratio of non-synonymous to synonymous substitutions [26]. However, positive selection for diversification does not appear to act on every codon, and a small number of codons in both the PRD and CRD possess a faster rate of non-synonymous substitutions over that of silent substitutions. The difference in the number of positively selected sites in the PRD vs. the CRD could be the basis for the different structural organizations of the respective domains; the PRD is less conserved compared to the CRD. This may be the result of pathogen pressure driving codon changes that leads to conformational variation among penaeidins and variable activity or specificity against different microbes [17].

2.2. Crustins

Crustins are a second type of AMP in crustaceans. The first crustin-like protein was isolated from the granular hemocytes of the shore crab, *Carcinus maenas*, was designated a carcinin, has antibacterial activity, and possesses a whey acidic protein (WAP) domain at the C-terminus [27–29]. However, similar sequences were identified in an expressed sequence tag (EST) study in the shrimp species *Litopenaeus vannamei* and *L. setiferus* [30], and a more generic term "crustin" was coined to refer to this particular type of crustacean AMP [11].

Crustins are cationic cysteine-rich AMPs, range in size from 7 to 14 kDa, with a non-conserved signal sequence at the N-terminus, and a conserved cysteine-rich WAP domain at the C-terminus (Fig. 2) [14,31]. There are three types (I–III) of crustins based on the differences in domain organization between signal sequences and the WAP domain (Fig. 2) [6]. WAP domains in Types II and III have four disulfide bonds that stabilize the domain into a tightly packed structure described as the four-disulfide core (4DSC) region. Type I crustins usually have no more than six cysteines and an incomplete 4DSC configuration [31]. Recently, a variant anti-microbial crustin was identified in the black tiger shrimp, *Penaeus monodon*, that has a glycine–proline-rich region at the N-terminus with the signature WAP domain at the C-terminus [32]. To date, more than 50 different crustins have been characterized from 50 different crustacean decapods including crabs, lobsters, shrimp, prawns and crayfish.

Antimicrobial activity of crustins from different species are variable depending on whether experiments are conducted *in vitro* or *in vivo*. *In vitro* studies suggest that all crustins are active against Gram-positive bacteria, a few are active against Gram-negative bacteria, and none show responsiveness towards fungi.. [32–35]. *In vivo* results show variable responses to both Gram-negative and Gram-positive bacteria [36–39] and no response to fungi, as measured by the mortality rate of shrimp challenged with a marine fungus [39]. Regardless of the variability in response of crustins from a wide range of crustaceans against specific microbes, it is clear that the WAP domain that is present in all crustins plays an important role in the antibacterial activity.

Crustin genes vary in length, with two to four exons, and vary in composition within and among different types and/or isoforms [31,33,40,41]. Different crustaceans express variable numbers of crustin types and isoforms [11,31,42]. This diversity is not the result of alternative splicing within an individual, but can be attributed to discrete nucleotide polymorphisms in genes and/or the transcripts either through the expression of different alleles and/or RNA editing [27,40]. The source of crustin gene diversity can be deduced from a phylogenetic analysis, which shows that the three types of crustins cluster into monophyletic clades [31]. This suggests that crustins have undergone expansions within the crustaceans resulting in families of closely related genes that encode proteins with slightly variant activities, a result similar to that for the penaeidins. It is likely that this is the outcome of positive selection for diversification in response to pathogen pressures. The variation

in microbial specificity is likely due to a combination of variation in allelic expression of the different crustin types and RNA editing resulting in the production of different crustin isoforms with variations that impart slightly different antimicrobial activities.

3. Allorecognition in protochordates: fuhc, fester and Uncle fester

The ability to discriminate self from non-self is found in almost all metazoan forms, and is a cornerstone of immunity. Among the invertebrates, the diverse histocompatibility system of the colonial tunicate *Botryllus schlosseri*, has been studied in detail. An adult *Botryllus* consists of asexually derived zooids that remain encased in a single gelatinous tunic and share a common blood supply [43]. The vasculature terminates in small protrusions called ampullae located at the periphery of the colony, and as *Botryllus* grows over various substrata, physical contact between colonies results in either fusion or rejection reactions. During fusion, ampullae penetrate the neighboring tunic, the two vasculatures reorganize together, and the two colonies eventually fuse into a chimera [44–48]. In an inflammatory rejection reaction, neighboring ampullae make contact and then disintegrate into darkened points of rejection, blood exchange is blocked and the colonies eventually separate [49–51].

3.1. The fusion/histocompatibility locus

Early studies established that this natural transplantation reaction is governed by a single "fusibility" locus [44–47]. Now known as *fuhc* (fusion/histocompatibility), this Mendelian locus is highly polymorphic with tens to hundreds of co-dominant alleles present in wild populations. The fusion/rejection rules dictate that colonies must share at least one of two alleles to fuse [52–59]. Sequence variations in the *fuhc* gene correlates 100% with predicted histocompatibility outcomes [52].

3.2. The fuhc protein

The encoded fuhc is a type I transmembrane (TM) protein of 1008 aa with an N-terminal signal sequence, an extracellular epidermal growth factor (EGF) repeat, two (possibly three) tandem C2 type Ig domains, a TM region, and a C-terminal intracellular tail (Fig. 3A) [52,reviewed in 57]. Single nucleotide polymorphisms (SNPs) and insertions or deletions (indels) in the gene are manifest in a highly polymorphic primary sequence of the fuhc proteins, and most alleles differ by 25–50 aa within the ectodomain. There are no explicitly hypervariable regions along the protein, although the polymorphisms within the extracellular domains correlate exactly with histocompatibility typing [52]. The gene has 31 exons, and alternative splicing generates different, membrane bound and secreted versions of the encoded protein (Fig. 3A). Full-length and alternatively spliced fuhc proteins are expressed in tadpole and adult tissues involved with histocompatibility [52]. In all, these characteristics suggest that fuhc is a protein that defines self in *Botryllus* histocompatibility reactions.

3.3. The fester locus

The search for other genes within the *fuhc* locus whose products could conceivably recognize and respond to *fuhc* allelic diversity identified two putative receptor genes: *fester* [60] and *Uncle fester* [57]. *Fester* has 11 exons, is located near *fuhc*, and also exhibits significant sequence polymorphism [60]. In total, 45 *fester* alleles have been classified among populations along the USA East and West coasts [57]. However, unlike *fuhc*, variations in the *fester* allele sequences do not predict histocompatibility outcomes [60]. On the other hand, the fester proteins are highly diversified through mRNA alternative splicing (see below). The second member of the fester protein family, Uncle fester, is similar in structure and topology to fester (Fig. 3B), although it shows less polymorphism [57].

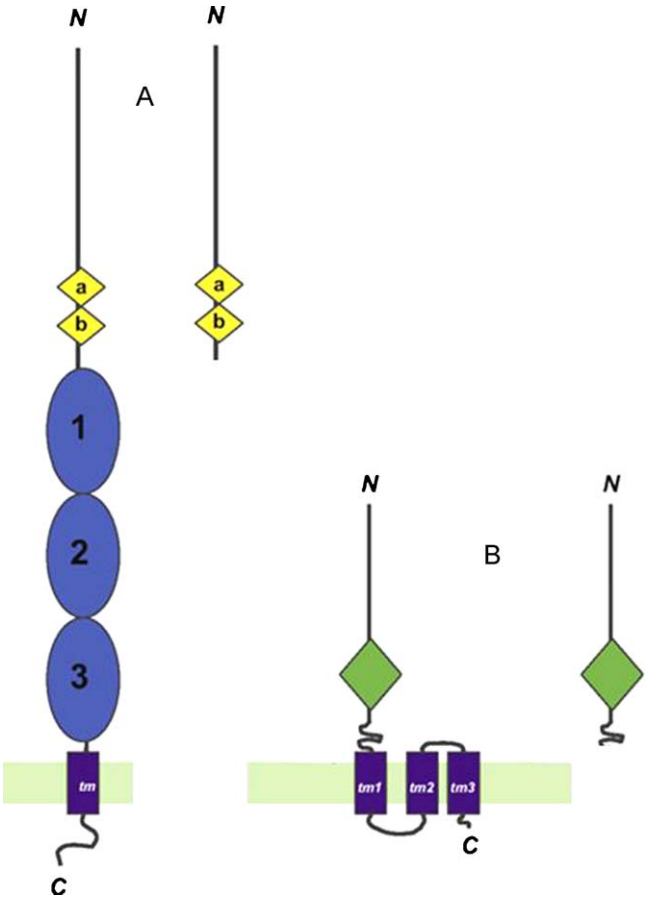

Fig. 3. The fuhc, fester, and Uncle fester proteins in *Botryllus schlosseri*. (A) The fuhc protein. The full length fuhc (on the left) has two tandem EGF domains (yellow diamonds), followed by two, possibly three, tandem Ig (C2 type) domains (blue ovals), a TM region, and a short intracellular tail. Alternative splicing generates a secreted form (on the right) that is truncated between the second EGF domain and the first Ig domain. (B) fester or Uncle fester proteins (on the left). Both proteins encode an extracellular SCR domain (green diamond), although there is no amino acid homology between the two domains. Both proteins have three predicted TM regions and an intracellular C-terminal tail. fester and Uncle fester proteins have different levels of sequence diversity among variants. A splice variant of fester deletes the TM regions and generates a secreted form (on the right) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).

Reprinted with modification from [57] with permission from Elsevier.

3.4. The fester protein

fester encodes a 368 aa, type I TM protein with a signal sequence followed by a short consensus repeat (SCR) domain, three tandem TM regions, and a short intracellular tail (Fig. 3B) [57,60]. More than 60 membrane bound isoforms of fester are generated from the variable splicing of exons encoding the ectodomain, and the deletion of sequences encoding the TM region generates another 16 secreted versions. All variants are expressed in tissues involved with histocompatibility responses, although the exact interactions between fester and fuhc proteins remains unknown. Blocking fester expression suggests that it functions as both an activating and an inhibitory receptor to fuhc [57,60,61]. Down regulation of fester using RNA interferance (RNAi) results in neither fusion nor rejection reactions between different colonies, regardless of their

respective histocompatibility. Conversely, injection of a monoclonal antibody (mAb) against fester mimics a self-recognition event and causes incompatible colonies to fuse, but leaves compatible fusion events unaffected. Thus, it appears that the fuhc/fester response is analogous to the "missing-self" recognition mechanism [62] of natural killer cells in that fester acts in an inhibitory fashion to prevent rejection during self-recognition (fuhc) events, but also has the capacity to activate both fusion and rejection reactions between colonies [57,60,61]. This dual function could be achieved through the use of different splice variants acting within independent fusion and rejection pathways, though the exact mechanisms remain to be elucidated.

3.5. Diversification of allorecognition systems in Botryllus

The *fuhc/fester* system in *Botryllus* is an intriguing example of diversification within an invertebrate immune system at both the gene and protein level. Hotspots of recombination around the *fuhc* locus likely contribute to crossovers that alter *fuhc* genotypes [54] and thus add more diversity to the system. The initial characterization of the *fuhc* locus identified 18 putative genes associated with *fuhc,* including guanine monophosphate synthetase, E-3 ubiquitin ligase, and a member of the DNA topoisomerase III family [54] and these may offer clues into how *fuhc* polymorphism is maintained. Mutations in DNA polymerase III and other replication enzymes in yeast and *E. coli* are known to increase the amount of spontaneous chromosomal breaks, which when fixed through homologous recombination, become an important source of recombination [reviewed in [63]]. In a similar vein, high levels of transcription enhances homologous recombination (transcription-associated recombination; TAR) in prokaryotes and higher eukaryotes [63]. Therefore, we speculate that the linkage of the "housekeeping" and replication genes within the *fuhc* locus may impart an increased chance of recombination events during replication stress and high rates of transcription. There may also be other, yet undiscovered genes or mechanisms related to *fuhc* and/or *fester* that maintain diversification of the putative allorecognition ligands and/or receptors.

Above the gene level, diversification via alternative splicing generates several varieties of secreted and membrane bound proteins encoded by both *fuhc* and *fester* [52,54,60], although it is unclear whether any post-translational modifications of the encoded proteins take place. Furthermore, different *fester* splice variants may be partitioned into inhibitory or activating pathways [60], thus any cross-talk of receptors and signaling pathways could be diversified through distinctive post-translational modifications of the proteins involved, or even through variable oligomerization of different fuhc and fester isoforms, if such an occurrence exists among these proteins.

The *fuhc/fester* system is likely the result of a rapid diversification in evolutionary history [61,64]. This is particularly evident in the complete lack of *fuhc/fester* homologues in the genome of the closely related solitary ascidian *Ciona intestinalis* [65]. Allorecognition systems in other animals in general do not share directly conserved components. For example, the cnidarian histocompatibility genes *alr-1* and *alr-2* offer scant similarity to *Botryllus fuhc* or the vertebrate major histocompatibility complex (MHC) [66]. Indeed, it had long been believed that the complex non-self discriminatory system in *Botryllus*, or those of any other "lower" organism, might be the basal forms of MHC in higher vertebrates [59,67,68]. However, the receptors and ligands in these invertebrate systems have proven to be largely unrelated to those of the vertebrate adaptive immune system, and the hunt for the true evolutionary precursor of MHC continues [61,64,69]. In summary, allorecognition processes exist in many invertebrates, but the mechanisms by which self/non-self discrimination occurs vary considerably, which suggests that allorecognition proteins have evolved and diversified independently across many phyla. Further work on the system present in *Botryllus* and other invertebrates will add to our growing understanding of the evolution of metazoan histocompatibility.

4. Down Syndrome cell adhesion molecule (Dscam) in insects

One of the most intriguing and well characterized hypervariable genes that functions in the innate immunity of the invertebrates encodes the Down Syndrome cell adhesion molecule (Dscam). In humans, it is a receptor with Ig domains that regulates neuronal wiring and defects in the protein are linked to Down Syndrome [70]. Orthologues of human *DSCAM* family genes are also present in arthropods: 11 species of *Drosophila*, two species of mosquitoes, *Anopheles gambiae* [71] and *Aedes aegypti*, the honey bee, *Apis mellifera* [72,73], the wax moth, *Bombyx mori* [74], the beetle, *Tribolium castaneum* [74], and two crustaceans, *Daphnia magna* and *D. pulex* [75]. Phylogenetic analysis of *Dscam* genes in different species predicts the existence of a six million year old common ancestral gene [76] that over evolutionary time has undergone a single duplication event in the human genome to generate *DSCAM* and *DSCAML1*, and has undergone extraordinary exon duplications within a single *Dscam* gene in arthropod genomes. In *Drosophila melanogaster, Dscam* has 115 exons of which 95 are duplications of exons 4, 6 and 9 that are tandemly arranged in 12, 48 or 33 clusters of different isoforms (Fig. 4A) [77,78]. Alternative splicing of the four clusters of exons is mutually exclusive in which each mRNA ends up with a single variant for exons 4, 6 and 9, which results in the formation of 38,016 putatively different mRNAs (Fig. 4B) [72].

The molecular mechanism that regulates the alternative splicing for the four sets of clustered exons is, in large part, still unclear and under investigation, however, there are several clues that the formation of secondary structure in the mRNA is involved. The mechanism of alternative splicing within the duplicated array for exon 4 suggests that an intronic sequence positioned 5′ of the exon 4 array forms a stem-loop in the mRNA (Fig. 4A) [79]. Although the deletion of this intronic sequence impairs the production of a correct Dscam protein, it is not involved with the choice of which exon 4 variant is spliced into the mRNA [80]. Alternative splicing of exon 6 is based on the presence of a highly conserved intronic region of 66 base pairs called the docking site that is located between exon 5 and the exon 6 array and a complementary selector sequence that is located 5′ of each exon 6 variant (Fig. 4A). The model proposed for alternative splicing for exon 6 suggests that the selector sequence anneals to one of the docking sequences forming a stem loop structure in the RNA. The heterogeneous nuclear ribonucleoprotein Hrp 36, which is a constitutive splicing repressor, binds the exon 6 sequences and may play a role in alternative splicing [81]. There is no current information about the mechanism for alternatively splicing exon 9 [82].

The 210 kDa Dscam protein is a PRR composed of a putative signal peptide, 10 tandemly repeated Ig domains, six fibronectin type III (FNIII) domains, one TM region and a cytoplasmic tail (Fig. 4B). The alternative splicing of the four variable exons influences the final structure of the Dscam protein at the N-terminal half of the second Ig domain (exon 4) the N-terminal half of the third Ig domain (exon 6) and the entire seventh Ig domain (exon 9) (Fig. 4B). There is also a choice of one of two possible TM regions (exon 17). *Dscam* is expressed in the larval hemocytes, the fat body (the main immune-organ of *Drosophila*), as well as in the brain [74]. Dscam has a tissue-specific expression pattern for certain isoforms of the variable exons and is present in the conditioned medium indicating that it can be secreted. When *Dscam* expression is blocked there

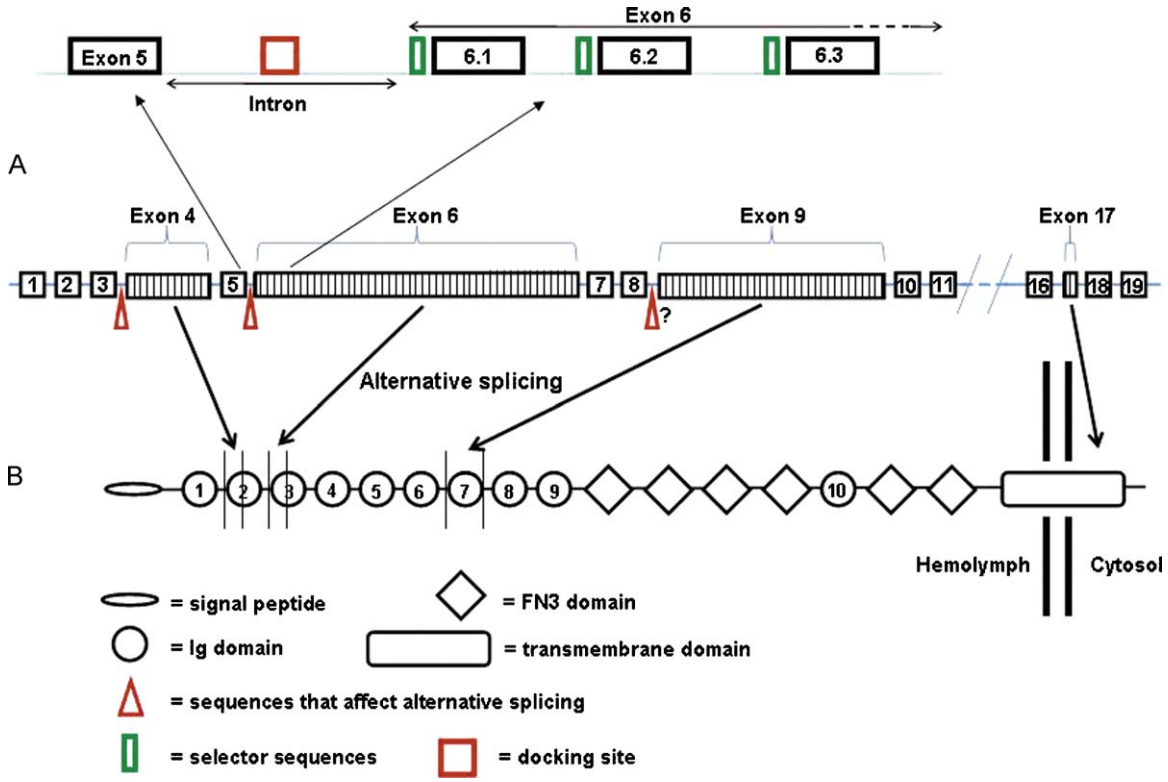

Fig. 4. The *Dscam* gene in *Drosophila melanogaster*. (A) The structure of the *Dscam* gene is shown illustrating the duplicated and non-duplicated exons. The red triangles indicate positions of sequences that may affect alternative splicing. Above is shown the locations of the docking site (red box) and selector sequences (green boxes) that control alternative splicing of exon 6 through the formation of stem-loop structures. (B) The mature Dscam protein after alternative splicing that generates hypervariable sequences in the second, third and seventh Ig domains. Two choices for the TM domain sequence also results from alternative splicing.
Reprinted with modifications from [164] with kind permission from Springer Science+Business Media.

is a 30% decrease in phagocytosis of bacteria indicating that it is involved with opsonization that augments phagocytosis [74].

The *AgDscam* gene in *A. gambiae* has similar organization to that in *Drosophila* (three clusters of duplicated exons) and the same type of alternative splicing that can generate 31,920 different mRNA molecules [83]. RNAi induced silencing of *AgDscam* leads to decreased survival of mosquitoes and decreased phagocytic capacity similar to that observed for *Drosophila*. Furthermore, upon second challenge from different types of pathogens (Gram-positive or Gram-negative bacteria or fungi) different high affinity Dscam proteins are produced. Although more work is required to clarify the precise immune role that Dscam plays and the epitopes on the pathogens to which Dscam binds, it represents a novel and very interesting solution to immune defense in arthropods.

5. Peptidoglycan recognition proteins (PGRPs) in insects

PGRPs are key molecules for detecting the presence of pathogens in a wide range of animals [84], but are best understood in arthropods. Insect PGRPs are selectively able to bind different types of PGNs that are present on the surface of pathogens [85]. PGNs are the main constituent of the cell wall of Gram-positive bacteria (lysine type) and Gram-negative bacteria, in which the lysine is replaced by diaminopimelic acid (DAP-type). The N-terminal region of the PGRP proteins is variable and specific for each type of PGRP protein, whereas the C-terminal PGRP domain (160 aa) is well conserved across different species from humans to invertebrates [86]. Some of the secreted PGRPs have type 2 amidase activity and are able to interact directly with and to degrade the PGN to a non-inflammatory product [87] (Table 2). Other secreted PGRPs cooperate with the Gram-negative binding protein 1 (GNBP1) to detect the presence of Lys-type PGN in the hemolymph that activates an enzymatic cascade to cleave Spätzle that binds to Toll, which initiates a cytoplasmic pathway resulting in the expression of *AMP* genes [88]. In contrast, the PGRP-LCs are cell surface receptors that bind to DAP-type PGNs of Gram-negative bacteria (Table 2). After binding, they dimerize and activate the Immunodeficiency (IMD) pathway [89], which also leads to the expression of *AMP* genes [85,90]. The end result is the control and removal of invading pathogens through the activities of AMPs [85,91,92].

There are 13 *PGRP* genes in the *D. melanogaster* genome [85] and some are alternatively spliced to generate a repertoire of 17 PGRP proteins [93]. The *PGRP* genes are commonly divided in two subgroups based on the length of the transcripts. Short transcripts encode PGRP-S (200 aa) with a signal peptide and one PGRP domain while long transcripts encode PGRP-L (400 aa) with one or two PGRP domains and no signal peptide (Table 2) [94]. Some of the *PGRP-L* genes are organized in clusters, possibly the result of duplications of an ancestral gene [93]. The diversity that has been characterized in this gene family likely underpins the ability of the insect immune system to recognize and discriminate among different classes of pathogens.

6. Fibrinogen related proteins (FREPs) in molluscs

Many invertebrates co-exist with their own co-evolved lineages of pathogens. Molluscs such as snails, for example, are hosts to roughly 18,000 species of digenetic trematode parasitic worms that do great damage to both the snail and vertebrate hosts [95–97]. The pulmonate freshwater snail, *Biomphalaria glabrata*, has received substantial attention as an intermediate host for the trematode *Schistosoma mansoni*, which causes human schistosomiasis, a dis-

Table 2
Peptidoglycan recognition proteins.

PGRP gene	Gene variant	Protein localization	Amidase activity	Function	PGN recognized	Pathway activated
SA		Circulating receptor	No	Recognizes Gm^+ bacteria, acts with GNPB1; no anti-fungal activity [166]	Mostly Lys type	Toll
SB-1		Circulating receptor	Yes	Unknown	DAP type	
SB-2		Circulating receptor	Yes	Unknown	DAP type[a]	
SC-1A		Circulating receptor	Yes	Induces cleavage of PGN *in vitro* [87], recognizes Gm^+ bacteria; induces phagocytosis	DAP and Lys type	Toll
SC-1B		Circulating receptor	Yes	Cleavage of PGN *in vitro* [87], recognizes Gm^+ bacteria; induces phagocytosis	DAP and Lys type	Toll
SC-2		Circulating receptor	Yes	Unknown	DAP type[a]	
SD		Circulating receptor	No	Enhances binding between PGRP-SA and GNBP1 (trimerization) [88]	DAP and Lys type	Toll
LA	*LAa*	TM	No	Unknown [167]	DAP type[a]	
	LAb	TM	No	Unknown [167]		
	LAc	TM	No	No PGRP domain (early stop codon) [93]		
LB		TM	Yes	Modulates the intensity of the IMD pathway activation [85]	DAP type	
LC	*LCa*	TM	No	Heterodimerizes with PGRP-LC [168]; induces phagocytosis [169]	DAP type	IMD
	LCx	TM	No	Recognizes polymeric DAP type LPS if alone; recognizes monomeric DAP if heterodimerized with PGRP-LCa [168]; induces phagocytosis [169]		IMD
	LCy	TM	No	Recognizes Gm^- bacteria [85]		IMD
LD		TM	No	Unknown [167]	DAP type[a]	
LE		TM	No	Multimerizes in large complexes to recognize Gm^- bacteria [170]; induces autophagy; activates PPO cascade [171]	DAP type	IMD
LF		TM	No	Decoy receptor [172]	DAP type	

[a] Predicted PGN (peptidoglycan) type recognized; TM, transmembrane protein; Gm^+, Gram positive; Gm^-, Gram negative; GNBP1, Gram-negative binding protein 1; Lys, lysine; DAP, diaminopimelic acid; PPO, prophenoloxidase; IMD, immunodeficiency

ease that afflicts 200 million people worldwide [98]. In response to trematode infections, the circulating hemocytes of *B. glabrata* secrete a diverse family of Ig superfamily (IgSF) domain containing fibrinogen-related proteins (FREPs) [99,100]. FREPs belong to the evolutionarily conserved immune related fibrinogen-related domain (FReD) lectin family, which is widely found in vertebrates and invertebrates [101] and includes fibrinogens, tenascins, microfibril-associated proteins, ficolins, tachylectins and FREPs. The last three of these proteins function in self/non-self recognition and host internal defense responses [97,99,101–104]. FREPs bind to trematode sporocysts and precipitate soluble parasite-derived secretory/excretory products (SEPs) [99,105]. FREPs also facilitate agglutination, phagocytosis and encapsulation by the hemocytes, and the release of toxic oxygen radicals upon challenge with the parasites [99,100,106,107].

FREPs have a unique molecular structure, consisting of one or two tandem N-terminal IgSF domains followed by an interceding region and a C-terminal fibrinogen (FBG) domain (Fig. 5) [97,99,100,108,109]. The FREP IgSF sequences resemble the variable (V)-type Ig domains based on predicted secondary structure [100]. Most *FREP* genes have four exons of which the first encodes the putative signal peptide, the second and third encode the IgSF domain(s), and the fourth encodes the FBG domain [100,110]. A few *FREP* genes are intronless and may be the result of retrotransposition [109].

6.1. FREP message diversity

Different subfamilies of FREP genes are expressed differentially upon trematode infection, which also confirms the general up-regulation of *FREP* gene expression in response to trematode infection [99,111]. Alternatively spliced forms of *FREP* transcripts and *FREP* retrosequences suggest that *FREP* diversification occurs at both genomic and transcriptional levels [109]. Alternative splicing generates FREP proteins with truncated FBG and IgSF domains and loss of the C-terminal half of the protein resulting in proteins with just a single IgSF domain (Fig. 5). Although the regulation of this alternative splicing is not understood, it may be associated with particular stresses or pathogen challenges.

6.2. FREP protein diversity

The FREP proteins are present in the hemolymph as multimeric proteins that are able to bind to a wide range of pathogens displaying carbohydrates on their surfaces [104]. Western blot analysis of FREP protein repertoire shows a pattern of bands of 50–100 kDa prior to immune challenge, which expands to 50–150 kDa post challenge. The smaller FREPs (65–75 kDa) predominantly bind to trematode sporocysts and their SEPs while the larger FREPs (95–125 kDa) bind to Gram-positive and Gram-negative bacteria, and to yeast. Recombinant FREPs show binding activities that are similar to native FREPs and recognize a wide range of pathogens, from prokaryotes to eukaryotes including trematode sporocysts and SEPs, and a variety of microbes including Gram-positive and Gram-negative bacteria and yeast [104]. Both the IgSF and FBG domains have potential binding ability, although it is not known which region is responsible for FREP functions. One hypothesis is that the IgSF domain may bind non-self pathogens while the FBG domain has lectin activity and is able to bind to a carbohydrate bearing ligand on a hemocyte membrane [101,104,112]. Different categories of FREPs seem to exhibit functional specialization with respect to the type of pathogen challenge.

6.3. FREP gene diversity

There are 14 different *FREP* gene subfamilies [97,99,100,108,113] that have one to eight loci per subfamily [109]. Four *FREP* genes for which the full-length sequences are known show variability within the regions encoding the IgSF and FBG domains [100,108,110]. ESTs also show a high level of

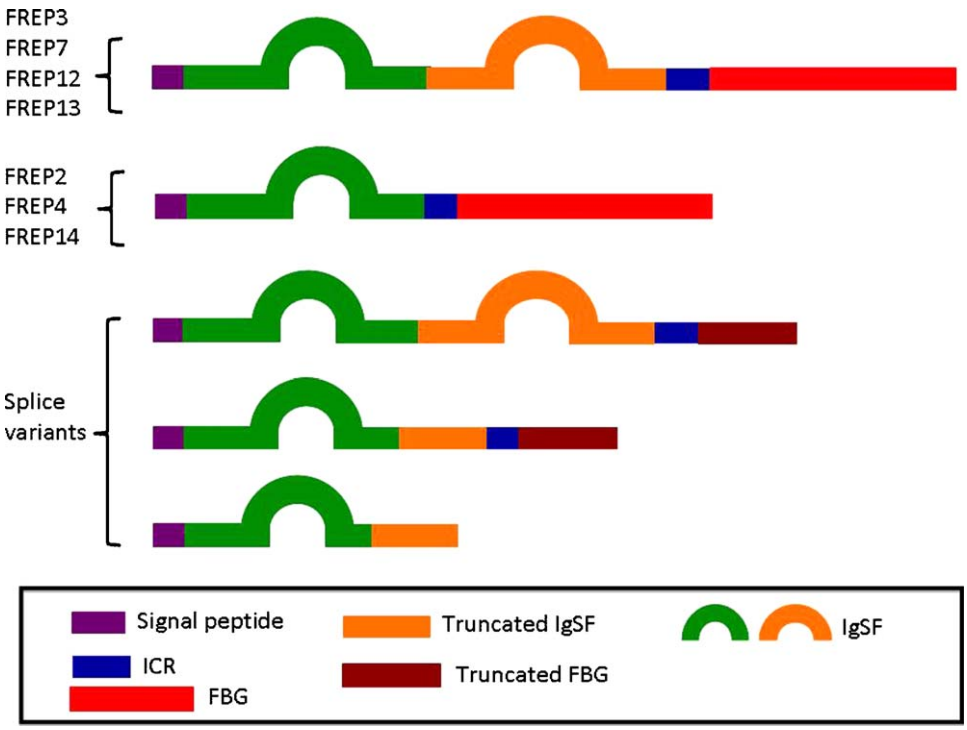

Fig. 5. Schematic representation of FREP domain structures in *Biomphalaria glabrata*, including three truncated forms. The two types of FREPs have either one or two IgSF domains plus a C-terminal FBG domain that is separated by the interceding region (ICR). *FREP* gene families that are known to encode FREPs with one or two IgSF domains are listed to the left. The structures of the IgSF regions of FREP families that are not listed (FREP1, 5, 6, 8, 9, 10, 11) are poorly known. Alternative splicing contributes to FREP diversity with tandemly arranged IgSF domains, resulting in splice variants with truncated FBG, truncated second IgSF domain, or missing C-terminal regions. FREP domains are not drawn to scale.

Reprinted with modifications from [109] with permission from Elsevier.

FREP sequence diversity [114] and targeted analysis of the IgSF1 domain from the *FREP3* cDNAs show an extraordinary level of diversity that is three to four fold higher than control genes [108]. Individual snails express various unique alleles as demonstrated by *FREP3* genes from one individual snail which yielded 45 different sequences, and 37 different sequences from a second snail, among which there was a single shared sequence [108]. Overall, a total of 314 unique *FREP3* gene sequences recovered from 22 snails encode 204 unique aa sequences. The diversification of the IgSF1 region of *FREP3* appears to be a non-random mechanism based on the absence of indels and the absence of lethal mutations that would cause protein truncation or non-synonymous changes.

The *FREP* alleles appear to be under strong selective pressure to diversify, because no single allele has dominated the population, which indicates that a diverse repertoire is maintained for immune defense against pathogens [108]. At least some of the *FREP* genes occur in tandem in the genome, which may facilitate the process of generating variant *FREP* alleles [101]. A combination of both Southern blots and computational analysis predicts that there is a small number of germline sequences that are maintained within each subfamily of *FREP* genes and that they retain sequences with the potential to account for all of the known variation observed in the cDNAs [108,111]. Some *FREP* genes may undergo diversification through somatic mutation or recombinatorial events, while other *FREPs* retain their original sequences. Mathematical modeling predicts that *FREP* diversification could potentially lead to variation in binding specificity without selection or tolerance induction [95]. The diversity suggests that the somatic recombination of *FREP* genes may be triggered by pathogen challenges [108], an observation that has been documented in families of clustered resistance (*R*) genes in higher plants (reviewed in [115]). It is noteworthy that the introns of several *FREP* genes are similar to non-long-terminal repeat retrotransposons that are present in the *S. mansoni* genome and encode a reverse transcriptase [100]. The transfer of retrotransposons from the parasite to the host may be involved with the diversification of the *FREP* genes in hemocytes. The combination of *FREP* gene diversification through point mutation and recombination [108] plus alternative splicing [109] results in a diverse set of expressed proteins. *FREP* diversification keeps pace with pathogen diversification and may be a mechanism to prevent tracking by pathogens (subsets of pathogens with specific genotypes that have increased success for infecting subsets of hosts) over multiple host generations [108].

7. Variable domain-containing chitin binding proteins (VCBPs) in amphioxus

IgSF proteins in protochordates are of particular interest in the study of the evolutionary origins of the antigen receptors that function in the adaptive immune system of jawed vertebrates – the TCRs and the Igs. These proteins and the genes that encode them, which contain both constant and variable Ig domains, are characterized by their inherent diversity, thus complicating the search for homologues by sequence alignment to infer ancestral proteins. Given these limitations, screens of cDNA libraries from fish and the amphioxus species, *Branchiostoma floridae*, followed by PCR and RACE that were designed to identify secreted and membrane bound IgSF proteins with sequence similarities to TCRs and Igs resulted in the identification of the membrane mounted novel immune-type receptors (NITRs) of fish [120,121] and the secreted variable domain-containing chitin-binding proteins (VCBP) in *B. floridae* [122]. There are five VCBP families in *B. floridae*, which have two variable-type IgSF domains at their N-terminus and a single chitin-binding domain at the C-terminus.

FREPs in other invertebrates

Other molluscs

Since the discovery of FREPs in *B. glabrata*, FREP homologues have been identified in other invertebrates, some of which may have roles in immune defense. Although *FREPs* are not present in the genome of the giant owl limpet, *Lottia gigantea* (http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html), another type of snail, the sea hare or sea slug, *Aplysia californica*, has two *FREP* genes, *AcFREP1* and *AcFREP2*, that encode proteins with two IgSF domains and a FBG domain [116]. The genomic structure of the two *AcFREP* genes are different, such that *AcFREP1* is intronless whereas *AcFREP2* has four exons. *AcFREP2* may be the parental gene whereas *AcFREP1* may be a newer intronless retrotransposed copy. The overall frequency of nucleotide substitution in the coding region of *AcFREPs* is not significantly higher than the conserved actin genes that served as controls [116] and therefore the *AcFREPs* may have functions other than defense.

Arthropod FREPs

FREP diversification in dipterans offers a tantalizing hint that diversity of immune receptors is indeed required to cope with greater or more variable range of pathogen exposures. The *D. melanogaster* genome encodes 14 *FREP* genes, while the mosquito *A. gambiae* has 57 [117]. Similar expansions of other gene families are seen in the mosquito species as that in *Drosophila*, particularly those involved in the anti-parasite response such as the complement-related thioester-containing proteins (TEPs) and leucine rich repeat (LRR) proteins [118]. This higher level of immune receptor diversity corresponds well with the higher level of pathogen exposure seen by mosquito species due to their hematophagic lifestyle, as compared with the fruit-eating *Drosophila*. Further evidence that this receptor diversity is in fact functional and positively selected, is that expression of immune receptors, in particular the FREPs, is differential between the two sexes in *A. gambiae* – sub-sets of *FREPs* are restricted to one sex or the other. This is presumed to represent differences in pathogen exposure due to different lifestyles – in *A. gambiae*, only the females are hematophagic and thus at high risk of exposure to parasites [117,119].

7.1. VCBP diversity

Initial studies of the *VCBP2* family revealed sequence variability in the form of base substitutions, giving rise to eight unique *VCBP2* sequences from cDNAs pooled from seven animals [122]. Intriguingly, the hypervariable region of VCBPs is restricted to approximately 18 aa on the N-terminal side of the first cysteine in the chitin-binding domain, and within regions of the V domains that do not correspond to the complementarity-determining regions (CDRs) of vertebrate V regions [122]. Consequently, the V domains of VCBPs can be viewed as V-like to avoid confusion with the functional properties of V regions in vertebrate antigen receptors. Further analysis of VCBP2 sequences revealed six distinct variants in a single animal, whose sequences corresponded directly with those found in the somatic and germline genome of the animal suggesting that VCBP sequence diversity is germline encoded, and not somatically derived [123]. Sequence variation among animals is significant, with pairs of animals generally sharing no more than two hypervariable regions in their VCBP2 repertoires, and 43 unique peptides predicted from the pooled sequences of 13 different animals [123]. VCBPs thus appear to share the characteristic diversity of many known invertebrate immune-related proteins at the population level.

7.2. VCBP protein structure

The VCBPs are secreted proteins, which is consistent with the screens for proteins containing a signal peptide and with their lack of a predicted TM region [122]. The solved structure of VCBP3 demonstrates strong structural similarity to the vertebrate antigen receptors, including intrachain dimerization between the two V-like regions, and three-layer packing of the presumed binding domain, in which residues of the side chains involute to form an inner layer between the two β-sheets of the V-like domain [124]. However, the hypervariable regions of VCBP3 occur in the N-terminal sequence of the V-like domains, in contrast to the those of the antigen receptors, which are found in the CDR loops that connect the β-strands [124]. While the peptide backbones of antigen receptor and VCBP variable regions are structurally highly similar and can be superimposed on a model, the hypervariable regions of each vary in position and are altered in orientation by 180° [125]. Assuming that these hypervariable regions do indeed mediate ligand binding, this structural difference suggests qualitative differences in the ligand-binding properties and mechanisms of VCBPs compared with other IgSF receptors.

A role in immunity is considered to be the likely function of the VCBPs based on their anatomical localization to the gut [122] and pharyngeal tissues in *B. floridae*, and to hemocytes in tunicates [126], as well as the diversity of the IgSF domains and chitin-binding motifs [127]. Limited evidence suggests that VCBP3 expression at least may be up-regulated on immune challenge [128], and future work is expected to elucidate the regulation, ligand-binding, and function of VCBPs.

7.3. Haplotype variability of VCBP loci

VCBP genes 1, 2, 4 and 5 are located within a single large chromosomal locus, while *VCBP3* segregates independently and is present at a separate region of the genome [129]. Within the primary locus, *VCBP* genes and gene fragments are found in either orientation and interspersed between unrelated genes. Haplotype variation is observed, including unequal crossover and copy number variation of *VCBP2* and *VCBP5* [130], which is also noted in the *VCBP3* locus. The *VCBP*-related genomic regions are relatively enriched in inverted repeats relative to the rest of the genome [129], which may contribute to instability of the region and an increased rate of recombination, resulting in gene duplication or deletion. Identification of the precise locations and orders of sites of crossover and gene conversion is, however, obscured by the multiple 'layers' of various diversification mechanisms acting on this gene family, particularly point mutations. Yet, discrete allele lineages (*VCBP2b, 5a* and *5b*) persist as recognizable entities at the population level. While functional data are not yet available, it is tempting to speculate that these conserved alleles perform distinct, selectable functions in the *B. floridae* immune system. Assuming that the VCBP proteins have an immunological role, all of the above mechanisms may be important in maintaining population-level diversity in *B. floridae* and tunicates and thereby contributing to disease resistance in the population at large.

8. The *Sp185/333* system in the purple sea urchin

The *Sp185/333* genes in the purple sea urchin, *Strongylocentrotus purpuratus*, encode a highly diversified repertoire of putative immune response proteins (reviewed in [131,132]). The messages were first identified in coelomocytes responding to immune challenge through differential display [133] and ESTs [134,135]. The *Sp185/333* genes show striking up-regulation in response to injections of bacteria, LPS, β-1,3-glucan, and

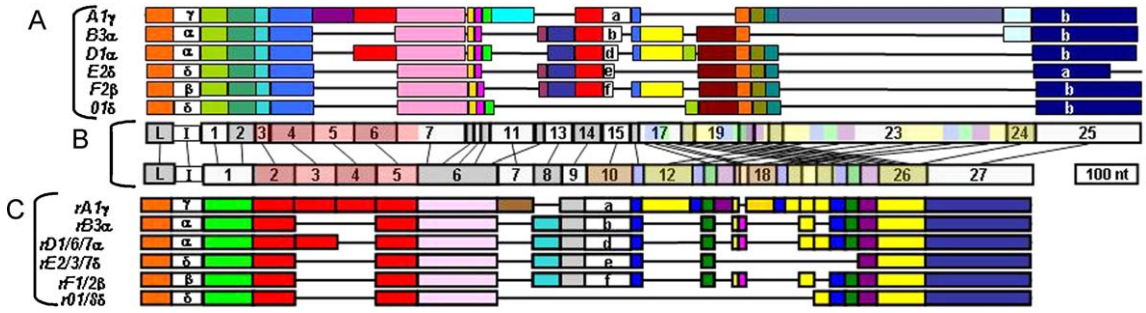

Fig. 6. Two equally feasible alignments of the *Sp185/333* gene sequences. Optimal alignments require insertion of gaps (horizontal lines) that define sequence elements (colored blocks). The intron is shown as a white block (not to scale), and the Greek letters indicate different versions of the intron [140]. A selection of representative element patterns is shown. Because of the repeats within the coding regions, two different, but equally optimal alignments are possible. (A) The cDNA-based alignment according to [135,138]. The glycine rich region is located within elements 2–7 and the histidine rich region is located within elements 10–23. (B) A comparison of elements illustrates how the two alignments correspond. Lines connecting elements for the two types of alignments indicate how the sequences were interdigitated from the cDNA alignment to produce the repeat-based alignment. This is particularly true for the repeats in element 23 from the cDNA-based alignment shown in (A). The elements are numbered; intron (I). Color shadings indicate the positions of repeats; pink, light blue, lavender, light yellow, light green. (C) The repeat-based alignment according to [140]. The type 1 repeats (red) at the 5′ end of the second exon are present in 2–4 copies. The repeats at the 3′ end of the exon are indicated in dark blue (type 2), dark green (type 3), yellow (type 4) and purple (type 5). Duplication of sets of repeats has been noted (type $[2\text{-}4\text{-}2\text{-}3\text{-}5]_2$-4; [139]). Designations of element patterns are listed to the left.
Reprinted with modification from [140] with permission from BioMed Central.

PGN [133,136,137]. Homologues have been found in other sea urchin species, *Heliocidaris erythrogramma*, *H. tuberculata*, and *Lytechinus pictus* [131], and all show sequence diversity. To date, similar sequences have not been identified outside of echinoids.

8.1. Structure of the Sp185/333 genes and messages

The *Sp185/333* messages are encoded by a gene family with an estimated size of about 50 loci ([138,139]; reviewed in [131,132]) and each gene has two exons with a small intron (Fig. 6A) [138,140]. The first exon encodes a hydrophobic leader region and the second exon encodes the mature protein. The structure of the second exon is unusual because an optimal alignment of multiple sequences requires the insertion of large artificial gaps that define blocks of sequence called *elements* (Fig. 6A) [140]. In addition to elements, the second exon has six types of repeats, of which one is present in two to four tandem copies near the 5′ end of the exon, and the rest are found as mixed interspersed repeats near the 3′ end [139]. Because of these repeats within the second exon, two equally optimal alignments of the sequences are possible (Fig. 6) [140]. To date, 171 randomly cloned genes have yielded 121 unique sequences that encode 101 unique proteins with 37 different element patterns. The genes also have many SNPs that encode non-synonymous changes in the protein sequence. Surprisingly, for a family of genes that share sequences, only a single pseudogene has been identified, which is intronless and may be a retroposon [140]. The rest of the genes have perfect, full-length open reading frames without early stop codons or frameshifts encoding missense sequence. This is an unusual finding for a family of such highly similar genes given that other multigene families typically include pseudogenes [141,142].

Much of the sequence diversity within the *Sp185/333* sequences is due to the mosaic patterns of elements that are found in recognizable patterns (Fig. 6A and C). This mosaic characteristic is an added level of diversity derived from the variant element sequences, which have an average of 11 (range of 1–28) different versions [139]. Because of the sequence diversity that is derived from the combination of the mosaic pattern of elements and the SNPs within the elements, there are no genes that share identical sequence among individual sea urchins. Paradoxically, the sequences of individual elements are commonly shared among different genes and among different individuals [140].

8.2. The Sp185/333 proteins

The deduced Sp185/333 proteins have a hydrophobic leader, a variable glycine rich region, a central region with an RGD motif (arginine–glycine–aspartic acid; integrin binding site) that is present in all but a few proteins, a variable histidine rich region with patches of poly-histidine, and a C-terminal region (see legend for Fig. 6) [135,136,138]. There are no cysteines and no obvious secondary structures can be deduced from the amino acid sequences. The proteins are expressed by two subsets of coelomocytes, the polygonal phagocytes and the small phagocytes, and are localized within perinuclear vesicles in both cell types and on the surface of the small phagocytes [143]. The expected size range of the deduced proteins is from 4 to 55 kDa depending on element pattern and truncation, however, most sea urchins show unexpectedly large Sp185/333 proteins that do not correspond to the predicted size of the monomers [137,143]. Both native and recombinant proteins appear as large multimers and/or aggregates with a wide range of isoelectric points resulting in over 260 spots on two dimensional Western blots, which appear to change in presence and intensity depending on the type of immune challenge [137]. The diversity of the protein repertoire is much greater than expected from a gene family of about 50 members, and may be the result of several levels of diversification including putative post-translational modifications.

8.3. RNA editing

In contrast to the genes, which have full length open reading frames (except for one intronless gene), many of the messages have SNPs that alter codons to early stops, and small indels that introduce frameshifts resulting in missense sequence leading to early stop codons [136]. Furthermore, the sequence diversity among the messages for an individual sea urchin is not the same as that for the genes, indicating that the variable nucleotide positions are different for the genes compared to those in the messages [144]. Searches to identify the most likely gene from which a given message is derived results in very few matches (Table 3) [144]. Furthermore, a single gene is the most likely source for 54–95% of the messages in individual animals. The sequences of the messages and the genes from which they are most likely transcribed show that nucleotide variation occurs throughout the length of the message, that 73% of the substitutions are transitions, and that the frequency of cytidine to uracil substitution is much higher than expected. These

Table 3
Very few *Sp185/333* messages and genes have identical matches.

Animal	Unique/total messages	Unique/total genes	Identical matches messages/genes	Percentage of matches vs. unique messages/genes
1	29/47	9/9	0/0	0%/0%
2	89/148	51/53	8/5	9%/9.8%
4	43/50	29/30	6/2	1.4%/6.9%

Data from [144].

results all suggest RNA editing is involved in sequence diversification, perhaps through the activities of a cytidine deaminase and/or a low fidelity polymerase such as polμ, which are both encoded in the genome [145]. This level of RNA editing is much higher than has been reported for other systems [146], and would be a powerful mechanism for increasing the diversity of the Sp185/333 proteins, producing missense and truncated Sp185/333 proteins that are present in the coelomic fluid [137]. Although it is not known whether proteins resulting from RNA editing are functional, 81% of the sequenced cDNAs from immunnoquiescent sea urchins have an *E2* element pattern with a SNP that changes a glycine codon to an early stop codon in element 13 (see Fig. 6A). After immune challenge, the frequency of messages with the truncated *E2* pattern decreases to 58% with a corresponding increase in full length E2 protein [131]. Because none of the sequenced genes with an *E2* element pattern have an early stop codon, directed transcript editing is a likely mechanism to expand the diversity of the Sp185/333 protein repertoire, and perhaps to respond to immune challenge in ways beyond the up-regulation of gene expression [144]

8.4. Gene diversity

Computational phylogenetic analysis of the *Sp185/333* gene sequences has been used to derive a possible evolutionary history of the gene family by evaluating the repeats and elements [139]. The tandem repeats in the 5′ end of the second exon (Fig. 6) are likely the result of duplications, deletions, unequal crossovers, and recombination from deduced ancestral repeat sequences to generate the extant variants of these repeats. Molecular clock analysis of the *185/333* sequences collected from several species of sea urchins were used to estimate the rate of nucleotide change over time compared to the rate of divergence of the echinoids [131]. This suggested that *185/333* diversification occurs rapidly in each species such that recombinant and parent sequences are quickly blurred, and that the genes diverge more quickly than the rate of speciation. The interspersed repeats at the 3′ end of the exon (Fig. 6) are complex, but appear to have a pattern in which a mixed set of repeats is duplicated [139]. In addition to recombination among repeats, recombination among genes is evident because there is no correlation between repeat patterns at the 5′ end of the second exon compared to the repeat patterns at the 3′ end. Similarly, shared sequence for the 5′ untranslated region and that for the leader region do not match among different messages [135]. Of the 121 genes with unique sequence, 12 appear to be recombinants such that the two ends share sequence with different genes [139]. Recombination appears to occur at any point within the gene, such that specific sequence variants of elements can be positioned next to as many as 12 variants of adjacent elements.

The structure of the genome harboring the *Sp185/333* genes provides a few clues to the diversity of the family. The genes are ≤2 kb and many are clustered as closely as ~3 kb apart [140,147,148]. Preliminary assembly of the sea urchin genome shows four clustered *Sp185/333* genes, however, finishing level sequence of a bacterial artificial chromosome (BAC) insert has six clustered genes within 34 kb, all of which are flanked by microsatellites (Fig. 7). The level of sequence diversity (or level of shared sequence) among the clustered genes is the same as that for genes randomly cloned and sequenced from three other genomes, which indicates that there is no correlation between proximity and shared sequence. However, the sequences between pairs of GA and GAT microsatellites that surround the genes, show increased similarity compared to outside of the microsatellite pairs. Together this suggests that the microsatellites may promote gene and segmental duplications. Furthermore, gene conversion may act among the members of this gene family, and that it may not be restricted to closely linked genes.

8.5. Multiple levels of diversity

In summary, there appear to be several mechanisms working at different levels to diversify the Sp185/333 protein repertoire. First, there may be a variety of sizes of DNA segments that are involved in diversification including (i) large genomic segments that encompass a gene plus its flanking sequence, (ii) the genes themselves, (iii) elements, (iv) repeats within the genes, (v) simple repeats within the larger repeats, and (vi) microsatellites that flank the genes. The *Sp185/333* sequences may be diversified by several mechanisms, including gene duplication (and likely deletion), gene conversion, recombination and paralogous mispairing occurring at meiosis [147,148]. Second, the mRNAs appear to be significantly edited, changing the encoded amino acid sequences to include truncated proteins, many with missense sequence [136,137]. Third, the proteins have conserved glycosylation sites and may be glycosylated [138], although other posttranslational modifications are possible. Both native and recombinant proteins readily dimerize and multimerize with monomers being rarely detected [137,143]. These several levels of diversification enables a gene family of about 50 members to generate a protein repertoire of over 260 isoforms [137]. Although the mechanisms for the diversification in the *Sp185/333* system are currently speculative, the outcome is a response that is likely required for survival in the marine habitat with an estimate of 10^5–10^6 microbes (bacteria and archea) per ml in surface waters [149,150] and 10^2–10^3 times more in the sedi-

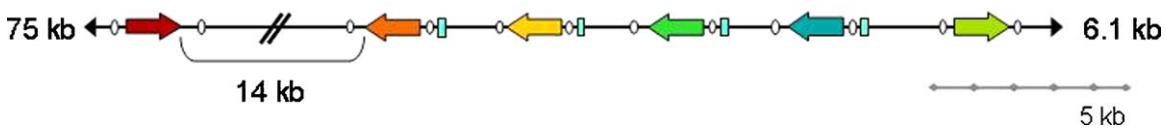

Fig. 7. A genomic cluster of *Sp185/333* genes. Six *Sp185/333* genes are clustered at one end of the BAC insert. They are positioned as close as 3.2 kb and are found in both orientations. The genes have four different element patterns (*A1*, red; *B8*, orange; *D1*, yellow, green, blue; *E2*, purple; see Fig. 6 for element pattern illustration) and are surrounded by GA microsatellites (white ovals). A larger segment, including the *D1* genes, are surrounded by GAT microsatellites (light blue squares).
Modified from [165] with permission from Chase A. Miller.

Table 4
Immune diversification in invertebrates[#].

Gene/protein family	Group or species	Level of diversity	Diversification mechanism(s)	Functional domain(s)	Activity
Crustins	Crustaceans	~50 genes characterized to date from crustaceans	SNPs, Allelic diversity, RNA editing[*]	WAP	Gm^- and Gm^+ bacteria
Dscam	Arthropods (non-immune functions in other groups)	38,016 possible protein isoforms in *D. melanogaster*, 39,120 in *A. gambiae*	Extensive alternative RNA splicing	Immunoglobulin	Bacteria
FREPs	Molluscs, arthropods	≤45 genes per individual (*Biomphalaria glabrata*), 14 subfamilies. 57 genes in *A. gambiae.*	Multiple gene families, selective pressure on genetic sequence, alternative splicing, somatic recombination	Immunoglobulin, fibrinogen	Trematodes, yeast, Gm^- and Gm^+ bacteria
Fuhc/fester	*Botryllus schlosseri*	Hundreds of fuhc alleles; 45 known alleles of *fester* in populations of *B. schlosseri*	High level of allelic diversity in population, SNPs and indels, meiotic mispairing, alternative splicing	Immunoglobulin	Allogeneic contact
Penaeidins	Shrimp and other crustaceans	4 gene classes, multiple isoforms each	Selective pressure on genetic sequences	PRD, CRD	Gm^- and Gm^+ bacteria, fungi
PGRPs	Widespread (animal kingdom)	13 genes in the *D. melanogster* genome, 17 protein isoforms	Alternative splicing, multi-gene family	PGRP	Gm^- and Gm^+ bacteria
Sp185/333	*Strongylocentrotus purpuratus* and other echinoids	~50 genes and ≤260 protein isoforms per individual	Multigene family, genomic instability, gene duplication, deletion, recombination, SNPs, RNA editing	No domains identified	Binds to Gm^- and Gm^+ bacteria (unpublished)
VCBPs	*Branchiostoma floridae*	5 gene families, ~6 unique members of VCBP2 gene family per individual	Multigene families, gene copy number variation, genomic instability, meiotic mispairing, SNPs in hypervariable domains	Immunoglobulin	Chitin (unpublished)

[*] SNPs, single nucleotide polymorphisms; WAP, whey acidic protein domain; Gm^-, Gram negative; Gm^+, Gram positive; PRD, proline rich domain; CRD, cysteine rich domain.
[#] The information in the table was collected from the publications cited in this review.

ment [151]. There may be as many as 10^{30} marine viruses in the oceans (10^6–10^7 per ml depending on region and depth), which is estimated to be ~15 fold more than marine microbes [152]. At least some of the viruses and bacteria will infect marine invertebrates [153]. The sea urchin innate immune system is both complex and sophisticated [145] and the *Sp185/333* system is a fascinating example of how this invertebrate survives this hostile environment.

9. Conclusion

Immune gene diversity is a result of a two-way, co-evolutionary interaction between populations of hosts and pathogens. Co-evolution occurs when changes in gene frequencies in one population results in selection pressure that drives changes in gene frequencies in a competing population, which in turn places slightly altered selection pressures to change gene frequencies in the first population [154]. This process continues back and forth resulting in incremental changes that accumulate in selected traits, which are the infectivity of the pathogen and/or its abilities to avoid or suppress the host immune response, and the host's ability to resist and/or clear infection. The process is an arms race with constantly changing dynamics leading to an accumulation of changes in both populations, which has been described as having Red Queen dynamics [155], a concept borrowed from a pronouncement by the Red Queen in *Through the looking glass*, "It takes all the running *you* can do, to keep in the same place" [156].

The effects of this arms race can clearly be seen in immune systems in general and the selected examples of invertebrate genes and gene families that we discuss here, which not only have a common attribute of striking sequence diversity but also exemplify the wide variety of genes that function in immunity and non-self recognition (Table 4). Groups of related organisms or individual species appear to have their own particular immune response system that is tailored to their habitat and diet for effective control of their commensals, pathogens or conspecific competitors as documented for *Botryllus*. As a result of their separate evolutionary histories, invertebrate immune responses are quite different and searches for a common "genetic toolkit" that functions in a wide variety of animals to regulate the genes involved in the immune response have not been successful. A genetic toolkit is the minimal set of conserved genes that encode transcription factors and components of signaling pathways that control development [157]. An immune specific genetic toolkit either does not exist, or perhaps more likely, the putative ancestral toolkit is likely to have been significantly altered and is currently unrecognizable in modern metazoans because of the rapid rate of evolutionary diversification imposed by pathogen pressure in different habitats. However, there are some common themes that emerge from comparisons among immune systems from a diverse set of animals (Table 4). There appears to be a set of repeatedly identified protein domains and diversification mechanisms that one might consider to be the core repertoire of "building materials" and "construction methods" (respectively) that are essential for an effective immune response, and on which the tools of the toolkit could act. The common construction methods, or widely used mechanisms that promote sequence diversification, include localized genomic instability, duplications and deletions of genes, generation of clusters of similar genes that become members of families, gene recombination, gene conversion, retrotransposition, duplications of exons within genes, alternative splicing, multiple available alleles of single copy genes, and RNA editing. Many of these construction methods are commonly employed by the invertebrate examples described here (Table 4). The gene diversification mechanisms tend to act on families of clustered genes that share similar sequences, such as *FREPs* and *Sp185/333*. The processes for generating a family of duplicated genes from a single copy gene, which then promotes continued duplications leading to sequence diversity, are not fully understood. However, recent work in yeast suggests that the re-initiation of DNA synthesis at origins of replication during S phase plus the early termination of DNA synthesis at repeat motifs may be a mechanism for duplicating genes [158]. This concept fits with the organization of the *Sp185/333* genes that are surrounded by microsatellites (Fig. 7) [147]. Perhaps clusters of genes with shared sequences is a regional genomic organization that is required to promote the diversification mechanisms listed above. Clusters of similar *R* genes in plants may actually be an essential aspect of the immune system to promote sequence diversification that will keep pace in the arms race with pathogens [115].

The putative immune system "building materials", or proteins, domains and motifs that are repeatedly found in immune systems in most animals, are generally elusive but would be expected to be present in both basal and more advanced animals [159]. A few proteins and domains have emerged that may have been part of an essential set of building materials, including a Toll/TLR system, a primitive complement effector system with C3 or TEP homologues and membrane-attack-complex/perforin-like sequences, an LPS binding protein, and a RAG core/Transib element [160–163]. Domains from metazoans that are repeatedly identified within proteins that have immune function, including the most primitive metazoans, the cnidarians and poriferans, are NACHT, LRR, Ig, death, TIR and lectin domains, and a thioester motif [160,161]. The presence of Ig domains in some of the examples described here illustrates this point (Table 4). It has been argued that this set of domains and motifs that function in immune responsiveness may actually be the result of parallel and independent evolution and not the descent from a set of common ancestral genes [162]. Conversely, many of these domains and motifs are involved in immunity in cnidarians and poriferans, which is suggestive of their role in an ancestral metazoan immune system [160] that acted effectively against all non-self contacts, including those with pathogens. Whichever of these two hypotheses may be correct, the massive diversity of molecules used by different organisms in the various versions of immune responses that are employed by metazoans illustrates that there are many possible solutions to the problem of pathogen and non-self recognition. This is becoming a central theme in the characterization of immune systems; they can be quite different yet they also appear to share one unifying feature – the diversification of the proteins on which they rely.

Acknowledgements

The authors are indebted to Drs. Curtis Suttle and Edward DeLong for confirming concentrations of viruses and microbes in the oceans, and to Lee Heiman for editorial comments on the conclusions. We are supported by an award from the National Science Foundation (MCB 0744999) to LCS.

References

[1] Haldane, J.B.S., 1949. Disease and evolution. La Ricerca Scientifica 19 (suppl), 68–75.
[2] Ferrandon, D., Imler, J.L., Hetru, C., Hoffmann, J.A., 2007. The *Drosophila* systemic immune response: sensing and signaling during bacterial and fungal infections. Nature Reviews Immunology 7 (11), 862–874.
[3] Peiser, L., Gordon, S., 2001. The function of scavenger receptors expressed by macrophages and their role in the regulation of inflammation. Microbes and Infection 3 (2), 149–159.
[4] Cigana, C., Curcuru, L., Leone, M.R., Ierano, T., Lore, N.I., Bianconi, I., et al., 2009. *Pseudomonas aeruginosa* exploits lipid A and muropeptides modification as a strategy to lower innate immunity during cystic fibrosis lung infection. PLoS One 4 (12), e8439.
[5] Herrin, B.R., Cooper, M.D., 2010. Alternative adaptive immunity in jawless vertebrates. Journal of Immunology 185, 1367–1374.
[6] Tassanakajon, A., Amparyup, P., Somboonwiwat, K., Supungul, P., 2010. Cationic antimicrobial peptides in penaeid shrimp. Marine Biotechnology (NY) 12 (5), 487–505.
[7] Brown, K.L., Hancock, R.E., 2006. Cationic host defense (antimicrobial) peptides. Current Opinion in Immunology 18 (1), 24–30.
[8] Hancock, R.E., Brown, K.L., Mookherjee, N., 2006. Host defence peptides from invertebrates–emerging antimicrobial strategies. Immunobiology 211 (4), 315–322.
[9] Yang, D., Biragyn, A., Kwak, L.W., Oppenheim, J.J., 2002. Mammalian defensins in immunity: more than just microbicidal. Trends in Immunology 23 (6), 291–296.
[10] Bulet, P., Stocklin, R., Menin, L., 2004. Anti-microbial peptides: from invertebrates to vertebrates. Immunological Reviews 198, 169–184.
[11] Bartlett, T.C., Cuthbertson, B.J., Shepard, E.F., Chapman, R.W., Gross, P.S., Warr, G.W., 2002. Crustins, homologues of an 11.5 kDa antibacterial peptide, from two species of penaeid shrimp, *Litopenaeus vannamei* and *Litopenaeus setiferus*. Marine Biotechnology (NY) 4 (3), 278–293.
[12] Cuthbertson, B.J., Shepard, E.F., Chapman, R.W., Gross, P.S., 2002. Diversity of the penaeidin antimicrobial peptides in two shrimp species. Immunogenetics 54 (6), 442–445.
[13] Destoumieux, D., Bulet, P., Loew, D., Van Dorsselaer, A., Rodriguez, J., Bachere, E., 1997. Penaeidins, a new family of antimicrobial peptides isolated from the shrimp *Penaeus vannamei* (Decapoda). Journal of Biological Chemistry 272 (45), 28398–28406.
[14] Destoumieux, D, Munoz, M., Bulet, P., Bachere, E., 2000. Penaeidins, a family of antimicrobial peptides from penaeid shrimp (Crustacea, Decapoda). Cellular and Molecular Life Sciences 57 (8–9), 1260–1271.
[15] Gueguen, Y., Garnier, J., Robert, L., Lefranc, M.P., Mougenot, I., de Lorgeril, J., et al., 2006. the shrimp antimicrobial peptide penaeidin database: sequence-based classification and recommended nomenclature. Developmental and Comparative Immunology 30 (3), 283–288.
[16] O'Leary, N.A., Gross, P.S., 2006. Genomic structure and transcriptional regulation of the penaeidin gene family from *Litopeneaus vannamei*. Gene 371 (1), 75–83.
[17] Cuthbertson, B.J., Yang, Y., Bachere, E., Bullesbach, E.E., Gross, P.S., Aumelas, A., 2005. Solution structure of synthetic penaeidin-4 with structural and functional comparisons with penaeidin-3. Journal of Biological Chemistry 280 (16), 16009–16018.
[18] Yang, Y., Poncet, J., Garnier, J., Zatylny, C., Bachere, E., Aumelas, A., 2003. Solution structure of the recombinant penaeidin-3, a shrimp antimicrobial peptide. Journal of Biological Chemistry 278 (38), 36859–36867.
[19] Yamada, K, Takabatake, T., Takeshima, K., 2000. Isolation and characterization of three novel serine protease genes from *Xenopus laevis*. Gene 252 (1–2), 209–216.
[20] Cuthbertson, B.J., Deterding, L.J., Williams, J.G., Tomer, K.B., Etienne, K., Blackshear, P.J., et al., 2008. Diversity in penaeidin antimicrobial peptide form and function. Developmental and Comparative Immunology 32 (3), 167–181.
[21] Kang, C., Xue, J., Liu, N., Zhao, X., Wang, J., 2007. Characterization and expression of a new subfamily member of penaeidin antimicrobial peptides (penaeidin 5) form *Fenneropennawus chinensis*. Molecular Immunology 44, 1535–1543.
[22] Cuthbertson, B.J., Bullesbach, E.E., Fievet, J., Bachere, E., Gross, P.S., 2004. A new class (penaeidin class 4) of antimicrobial peptides from the Atlantic white shrimp (*Litopenaeus setiferus*) exhibits target specificity and an independent proline-rich-domain function. Biochemistry Journal 381 (Pt 1), 79–86.
[23] Cuthbertson, B.J., Bullesbach, E.E., Gross, P.S., 2006. Discovery of synthetic penaeidin activity against antibiotic-resistant fungi. Chemistry and Biology of Drug Design 68 (2), 120–127.
[24] Sperstad, S.V., Haug, T., Vasskog, T., Stensvag, K., 2009. Hyastatin, a glycine-rich multi-domain antimicrobial peptide isolated from the spider crab (*Hyas araneus*) hemocytes. Molecular Immunology 46 (13), 2604–2612.
[25] Sperstad, S.V., Haug, T., Paulsen, V., Rode, T.M., Strandskog, G., Solem, S.T., et al., 2009. Characterization of crustins from the hemocytes of the spider crab, *Hyas araneus*, and the red king crab, *Paralithodes camtschaticus*. Developmental and Comparative Immunology 33 (4), 583–591.
[26] Padhi, A., Verghese, B., Otta, S.K., Varghese, B., Ramu, K., 2007. Adaptive evolution after duplication of penaeidin antimicrobial peptides. Fish and Shellfish Immunology 23 (3), 553–566.
[27] Brockton, V., Hammond, J.A., Smith, V.J., 2007. Gene characterization, isoforms and recombinant expression of carcinin, an antibacterial protein from the shore crab, *Carcinus maenas*. Molecular Immunology 44 (5), 943–949.
[28] Relf, J.M., Chisholm, J.R., Kemp, G.D., Smith, V.J., 1999. Purification and characterization of a cysteine-rich 11.5 kDa antibacterial protein from the granular haemocytes of the shore crab *Carcinus maenas*. European Journal of Biochemistry 264 (2), 350–357.
[29] Smith, V.J., Chisholm, J.R., 2001. Antimicrobial proteins in crustaceans. Advances in Experimental Medicine and Biology 484, 95–112.
[30] Gross, P.S., Bartlett, T.C., Browdy, C.L., Chapman, R.W., Warr, G.W., 2001. 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*. Developmental and Comparative Immunology 25 (7), 565–577.
[31] Smith, V.J., Fernandes, J.M., Kemp, G.D., Hauton, C., 2008. Crustins: enigmatic WAP domain-containing antibacterial proteins from crustaceans. Developmental and Comparative Immunology 32 (7), 758–772.
[32] Supungul, P., Tang, S., Maneeruttanarungroj, C., Rimphanitchayakit, V., Hirono, I., Aoki, T., et al., 2008. Cloning, expression and antimicrobial activity of crustinPm1, a major isoform of crustin, from the black tiger shrimp *Penaeus monodon*. Developmental and Comparative Immunology 32 (1), 61–70.
[33] Amparyup, P., Kondo, H., Hirono, I., Aoki, T., Tassanakajon, A., 2008. Molecular cloning, genomic organization and recombinant expression of a crustin-like antimicrobial peptide from black tiger shrimp *Penaeus monodon*. Molecular Immunology 45 (4), 1085–1093.
[34] Zhang, J., Li, F., Wang, Z., Xiang, J., 2007. Cloning and recombinant expression of a crustin-like gene from Chinese shrimp *Fenneropenaeus chinensis*. Journal of Biotechnology 127 (4), 605–614.
[35] Zhang, J., Li, F., Wang, Z., Xiang, J., 2007. Expression, purification, and characterization of recombinant Chinese shrimp crustin-like protein (CruFc) in *Pichia pastoris*. Biotechnology Letters 29 (5), 813–817.
[36] Brockton, V., Smith, V.J., 2008. Crustin expression following bacterial injection and temperature change in the shore crab *Carcinus maenas*. Developmental and Comparative Immunology 32 (9), 1027–1033.

[37] Hauton, C., Brockton, V., Smith, V.J., 2006. Cloning of a crustin-like, single whey-acidic-domain, antibacterial peptide from the haemocytes of the European lobster, *Homarus gammarus*, and its response to infection with bacteria. Molecular Immunology 43 (9), 1490–1496.

[38] Jiravanichpaisal, P., Lee, S.Y., Kim, Y.A., Andren, T., Soderhall, I., 2007. Antibacterial peptides in hemocytes and hematopoietic tissue from freshwater crayfish *Pacifastacus leniusculus*: characterization and expression pattern. Developmental and Comparative Immunology 31 (5), 441–455.

[39] Shockey, J.E., O'Leary, N.A., de la Vega, E., Browdy, C.L., Baatz, J.E., Gross, P.S., 2009. The role of crustins in *Litopenaeus vannamei* in response to infection with shrimp pathogens: an *in vivo* approach. Developmental and Comparative Immunology 33 (5), 668–673.

[40] Chen, J.Y., Chen, J.C., Lin, S.H., Pan, C.Y., Kuo, C.M., 2006. Organization and promoter analysis of a tiger shrimp *Penaeus monodon* sing WAP domain-containing protein gene. Fisheries Science 72, 1086–1095.

[41] Vatanavicharn, T., Supungul, P., Puanglarp, N., Yingvilasprasert, W., Tassanakajon, A., 2009. Genomic structure, expression pattern and functional characterization of crustinPm5, a unique isoform of crustin from *Penaeus monodon*. Comparative Biochemistry and Physiology B Biochemistry and Molecular Biology 153 (3), 244–252.

[42] Tassanakajon, A., Vatanavicharn, T., Supungul, P., Tang, S., Amparyup, P., Somboonwiwat, K., et al., 2008. Biotechnology of marine invertebrates – recent advances in shrimp and shellfish. In: Memorial book of the 5th World Fisheries Congress 2008. TERRAPUB, Tokyo.

[43] Manni, L., Zaniolo, G., Cima, F., Burighel, P., Ballarin, L., 2007. *Botryllus schlosseri*: a model ascidian for the study of asexual reproduction. Developmental Dynamics 236 (2), 335–352.

[44] Oka, H., Watanabe, H., 1957. Colony specificity in compound ascidians as tested by fusion experiments. Proceedings of the Japanese Academy of Science 33, 657–664.

[45] Oka, H., Watanabe, H., 1960. Problems of colony specificity in compound ascidians. Bulletin of the Biology Station at Asamushi 10, 153–155.

[46] Sabbadin, A., 1962. Le basi genetiche della capacita di fusione fra colonie in *Botryllus schlosseri* (Ascidiacea). Atti Accad Naz Lincei Rend 32, 1031–1035.

[47] Sabbadin, A., 1969. The compound ascidian *Botryllus schlosseri* in the field and in the laboratory. Pubbl Staz Zool Napoli 37 (Suppl), 62–72.

[48] Rinkevich, B., Weissman, I.L., De Tomaso, A.W., 1998. Transplantation of Fu/HC-incompatible zooids in *Botryllus schlosseri* results in chimerism. Biological Bulletin 195 (2), 98–106.

[49] Ballarin, L., Cima, F., Sabbadin, A., 1995. Morula cells and histocompatibility in the colonial ascidian *Botryllus schlosseri*. Zoological Science 12 (6), 757–764.

[50] Ballarin, L., Cima, F., Sabbadin, A., 1998. Phenoloxidase and cytotoxicity in the compound ascidian *Botryllus schlosseri*. Developmental and Comparative Immunology 22 (5–6), 479–492.

[51] Scofield, V.L., Nagashima, L.S., 1983. Morphology and genetics of rejection reactions between oozooids from the tunicate *Botryllus schlosseri*. Biological Bulletin 165 (3), 733–744.

[52] De Tomaso, A.W., Nyholm, S.V., Palmeri, K.J., Ishizuka, K.J., Ludington, W.B., Mitchel, K., et al., 2005. Isolation and characterization of a protochordate histocompatibility locus. Nature 438 (7067), 454–945.

[53] De Tomaso, A.W., Saito, Y., Ishizuka, K.J., Palmeri, K.J., Weissman, I.L., 1998. Mapping the genome of a model protochordate, I. A low resolution genetic map encompassing the fusion/histocompatibility (Fu/HC) locus of *Botryllus schlosseri*. Genetics 149 (1), 277–287.

[54] De Tomaso, A.W., Weissman, I.L., 2003. Construction and characterization of large-insert genomic libraries (BAC and fosmid) from the ascidian *Botryllus schlosseri* and initial physical mapping of a histocompatibility locus. Marine Biotechnology (NY) 5 (2), 103–115.

[55] De Tomaso, A.W., Weissman, I.L., 2003. Initial characterization of a protochordate histocompatibility locus. Immunogenetics 55 (7), 480–490.

[56] Grosberg, R.K., Quinn, J.F., 1986. The genetic-control and consequences of kin recognition by the larvae of a colonial marine invertebrate. Nature 322 (6078), 456–459.

[57] McKitrick, T.R., De Tomaso, A.W., 2010. Molecular mechanisms of allorecognition in a basal chordate. Seminar in Immunology 22 (1), 34–38.

[58] Rinkevich, B., Porat, R., Goren, M., 1995. Allorecognition elements on a urochordate histocompatibility locus indicate unprecedented extensive polymorphism. Proceedings of the Royal Society London B Biology 259 (1356), 319–324.

[59] Scofield, V.L., Schlumpberger, J.M., West, L.A., Weissman, I.L., 1982. Protochordate allorecognition is controlled by a MHC-like gene system. Nature 295 (5849), 499–502.

[60] Nyholm, S.V., Passegue, E., Ludington, W.B., Voskoboynik, A., Mitchel, K., Weissman, I.L., et al., 2006. fester, a candidate allorecognition receptor from a primitive chordate. Immunity 25 (1), 163–173.

[61] De Tomaso, A.W., 2009. Sea squirts and immune tolerance. Disease Models and Mechanisms 2 (9–10), 440–445.

[62] Lanier, L.L., 2005. NK cell recognition. Annual Review of Immunology 23, 225–274.

[63] Aguilera, A., Gomez-Gonzalez, B., 2008. Genome instability: a mechanistic view of its causes and consequences. Nature Reviews Genetics 9 (3), 204–217.

[64] Dishaw, L.J., Litman, G.W., 2009. Invertebrate allorecognition: the origins of histocompatibility. Current Biology 19 (7), R286–R288.

[65] Dehal, P., Satou, Y., Campbell, R.K., Chapman, J., Degnan, B., De Tomaso, A., et al., 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science 298 (5601), 2157–2167.

[66] Nicotra, M.L., Powell, A.E., Rosengarten, R.D., Moreno, M., Grimwood, J., Lakkis, F.G., et al., 2009. A hypervariable invertebrate allodeterminant. Current Biology 19 (7), 583–589.

[67] Weissman, I.L., Saito, Y., Rinkevich, B., 1990. Allorecognition histocompatibility in a protochordate species – is the relationship to MHC somatic or structural? Immunological Reviews 113, 227–241.

[68] Burnet, F.M., 1971. Self-recognition in colonial marine forms and flowering plants in relation to the evolution of immunity. Nature 232 (5308), 230–235.

[69] Flajnik, M.F., Du Pasquier, L., 2004. Evolution of innate and adaptive immunity: can we draw a line? Trends in Immunology 25 (12), 640–644.

[70] Yamakawa, K., Huot, Y.K., Haendelt, M.A., Hubert, R., Chen, X.N., Lyons, G.E., et al., 1998. DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system. Human Molecular Genetics 7 (2), 227–237.

[71] Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R., et al., 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 298 (5591), 129–149.

[72] Schmucker, D., Clemens, J.C., Shu, H., Worby, C.A., Xiao, J., Muda, M., et al., 2000. *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell 101 (6), 671–684.

[73] Graveley, B.R., Kaur, A., Gunning, D., Zipursky, S.L., Rowen, L., Clemens, J.C., 2004. The organization and evolution of the dipteran and hymenopteran Down syndrome cell adhesion molecule (*Dscam*) genes. RNA 10 (10), 1499–1506.

[74] Watson, F.L., Puttmann-Holgado, R., Thomas, F., Lamar, D.L., Hughes, M., Kondo, M., et al., 2005. Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science 309 (5742), 1874–1878.

[75] Brites, D., McTaggart, S., Morris, K., Anderson, J., Thomas, K., Colson, I., et al., 2008. The Dscam homologue of the crustacean *Daphnia* is diversified by alternative splicing like in insects. Molecular Biology and Evolution 25 (7), 1429–1439.

[76] Crayton III, M.E., Powell, B.C., Vision, T.J., Giddings, M.C., 2006. Tracking the evolution of alternatively spliced exons within the Dscam family. BMC Evolutionary Biology 6, 16.

[77] Celotto, A.M., Graveley, B.R., 2001. Alternative splicing of the *Drosophila* Dscam pre-mRNA is both temporally and spatially regulated. Genetics 159 (2), 599–608.

[78] Lee, C., Kim, N., Roy, M., Graveley, B.R., 2010. Massive expansions of Dscam splicing diversity via staggered homologous recombination during arthropod evolution. RNA 16 (1), 91–105.

[79] Graveley, B.R., 2005. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. Cell 123 (1), 65–73.

[80] Kreahling, J.M., Graveley, B.R., 2005. The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the *Drosophila* Dscam pre-mRNA. Molecular and Cellular Biology 25 (23), 10251–10260.

[81] Olson, S., Blanchette, M., Park, J., Sawa, Y., Yeo, G.W., Yakkley, J.M., et al., 2007. A regulator of Dscam mutually exclusive splicing fidelity. Nature Structural and Molecular Biology 14, 1134–1140.

[82] Schmucker, D., Chen, B., 2009. Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes. Genes and Development 23, 147–156.

[83] Dong, Y., Taylor, H.E., Dimopoulos, G., 2006. AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. PLoS Biology 4 (7), e229.

[84] Dziarski, R., Gupta, D., 2006. The peptidoglycan recognition proteins (PGRPs). Genealogy 7 (8), 232.

[85] Lemaitre, B., Hoffmann, J.A., 2007. The host defense of *Drosophila melanogaster*. Annual Review of Immunology 25, 697–743.

[86] Kurata, S., 2010. Extracellular and intracellular pathogen recognition by *Drosophila* PGRP-LE and PGRP-LC. International Immunology 22 (3), 143–148.

[87] Mellroth, P., Karlsson, J., Steiner, H., 2003. A scavenger function for a *Drosophila* peptidoglycan recognition protein. The Journal of Biological Chemistry 278 (9), 7059–7064.

[88] Bischoff, V., Vignal, C., Boneca, I.G., Michel, T., Hoffmann, J.A., Royet, J., 2004. Function of the *Drosophila* pattern-recognition receptor PGRP-SD in the detection of Gram-positive bacteria. Nature Immunology 5 (11), 1175–1180.

[89] Riddell, C., Adams, S., Schimd-Hempel, P., Mallon, E.B., 2009. Differential expression of immune defences is associated with specific host-parasite interactions in insects. PLoS One 4 (10), e7621.

[90] Kaneko, T., Yano, T., Aggarwal, K., Lim, J.H., Ueda, K., Oshima, Y., et al., 2006. PGRP-LC and PGRP-LE have essential yet distinct functions in the *Drosophila* immune response to monomeric DAP-type peptidoglycan. Nature Immunology 7 (7), 715–723.

[91] Leulier, F., Parquet, C., Pili-Floury, S., Ryu, J.H., Caroff, M., Lee, W.J., et al., 2003. The *Drosophila* immune system detects bacteria through specific peptidoglycan recognition. Nature Immunology 4 (5), 478–484.

[92] Lemaitre, B., 2004. The road to Toll. Nature Reviews Immunology 4 (7), 521–527.

[93] Werner, T., Liu, G., Kang, D., Ekengren, S., Steiner, H., Hultmark, D., 2000. A family of peptidoglycan recognition proteins in the fruit fly *Drosophila melanogaster*. Proceedings of National Academy of Science United States of America 97 (25), 13772–13777.

[94] Charroux, B., Rival, T., Narbonne-Reveau, K., Royet, J., 2009. Bacterial detection by *Drosophila* peptidoglycan recognition proteins. Microbes and Infection 11 (6–7), 631–636.

[95] Loker, E.S., Adema, C.M., Zhang, S.-M., Kepler, T.B., 2004. Invertebrate immune systems – not homogeneous, not simple, not well understood. Immunological Reviews 198 (1), 10–24.

[96] Olson, P.D., Cribb, T.H., Tkach, V.V., Bray, R.A., Littlewood, D.T., 2003. Phylogeny and classification of the Digenea (Platyhelminthes: Trematoda). International Journal of Parasitology 33 (7), 733–755.

[97] Loker, E.S., 2010. Gastropod immunobiology. In: Soderhall, K. (Ed.), Invertebrate immunity. Landes Bioscience and Spring Science Business Media.

[98] Pearce, E.J., MacDonald, A.S., 2002. The immunobiology of schistosomiasis. Nature Reviews Immunology 2 (7), 499–511.

[99] Adema, C.M., Hertel, L.A., Miller, R.D., Loker, E.S., 1997. A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. Proceedings of National Academy of Science United States of America 94 (16), 8691–8696.

[100] Leonard, P.M., Adema, C.M., Zhang, S.M., Loker, E.S., 2001. Structure of two *FREP* genes that combine IgSF and fibrinogen domains, with comments on diversity of the *FREP* gene family in the snail *Biomphalaria glabrata*. Gene 269 (1–2), 155–165.

[101] Stout, B.A., Adema, C.M., Zhang, S.M., Loker, E.S., 2009. The biology of FREPs: diversified lectins with fibrinogen-related domains from the freshwater snail *Biomphalaria glabrata*. In: Ahmen, V.A. (Ed.), Animal lectins: A functional view. CRC Press, Taylor & Francis, pp. 475–491.

[102] Adema, C.M., Hanington, P.C., Lun, C.M., Rosenberg, G.H., Aragon, A.D., Stout, B.A., et al., 2010. Differential transcriptomic responses of *Biomphalaria glabrata* (Gastropoda, Mollusca) to bacteria and metazoan parasites, *Schistosoma mansoni* and *Echinostoma paraensei* (Digenea. Platyhelminthes). Molecular Immunology 47 (4), 849–860.

[103] Hanington, P.C., Lun, C.M., Adema, C.M., Loker, E.S., 2010. Time series analysis of the transcriptional responses of *Biomphalaria glabrata* throughout the course of intramolluscan development of *Schistosoma mansoni* and *Echinostoma paraensei*. International Journal of Parasitology 40 (7), 819–831.

[104] Zhang, S.M., Zeng, Y., Loker, E.S., 2008. Expression profiling and binding properties of fibrinogen-related proteins (FREPs), plasma proteins from the schistosome snail host *Biomphalaria glabrata*. Journal of Innate Immunity 14 (3), 175–189.

[105] Uttenweiler-Joseph, S, Moniatte, M., Lagueux, M., Van Dorsselaer, A., Hoffmann, J.A., Bulet, P., 1998. Differential display of peptides induced during the immune response of *Drosophila*: a matrix-assisted laser desorption ionization time-of-flight mass spectrometry study. Proceedings of National Academy of Science United States of America 95 (19), 11342–11347.

[106] Hertel, L.A., Stricker, S.A., Monroy, F.P., Wilson, W.D., Loker, E.S., 1994. *Biomphalaria glabrata* hemolymph lectins: binding to bacteria, mammalian erythrocytes, and to sporocysts and rediae of *Echinostoma paraensei*. Journal of Invertebrate Pathology 64 (1), 52–61.

[107] Humphries, J.E, Yoshino, T.P., 2008. Regulation of hydrogen peroxide release in circulating hemocytes of the planorbid snail *Biomphalaria glabrata*. Developmental and Comparative Immunology 32 (5), 554–562.

[108] Zhang, S.M., Adema, C.M., Kepler, T.B., Loker, E.S., 2004. Diversification of Ig superfamily genes in an invertebrate. Science 305 (5681), 251–254.

[109] Zhang, S.M., Loker, E.S., 2003. The FREP gene family in the snail *Biomphalaria glabrata*: additional members, and evidence consistent with alternative splicing and FREP retrosequences. Developmental and Comparative Immunology 27 (3), 175–187.

[110] Zhang, S.M., Leonard, P.M., Adema, C.M., Loker, E.S., 2001. Parasite-responsive IgSF members in the snail *Biomphalaria glabrata*: characterization of novel genes with tandemly arranged IgSF domains and a fibrinogen domain. Immunogenetics 53 (8), 684–694.

[111] Zhang, S.M., Loker, E.S., 2004. Representation of an immune responsive gene family encoding fibrinogen-related proteins in the freshwater mollusc *Biomphalaria glabrata*, an intermediate host for *Schistosoma mansoni*. Gene 341, 255–566.

[112] Renwrantz, L., Schancke, W., Harm, H., Erl, H., Liebsch, H., Gercken, J., 1981. Discrimative ability and function of the immunobiological recognition system of the snail *Helix pomatic*. Journal of Comparative Physiology B 141, 477–488.

[113] Zhang, S.M., Nian, H., Zeng, Y., Dejong, R.J., 2008. Fibrinogen-bearing protein genes in the snail *Biomphalaria glabrata*: characterization of two novel genes and expression studies during ontogenesis and trematode infection. Developmental and Comparative Immunology 32 (10), 1119–1130.

[114] Mitta, G., Galinier, R., Tisseyre, P., Allienne, J.F., Girerd-Chambaz, Y., Guillou, F., et al., 2005. Gene discovery and expression analysis of immune-relevant genes from *Biomphalaria glabrata* hemocytes. Developmental and Comparative Immunology 29 (5), 393–407.

[115] Friedman, A.R., Baker, B.J., 2007. The evolution of resistance genes in multi-protein plant resistance systems. Current Opinion in Genetics and Development 17, 493–499.

[116] Gorbushin, A.M., Panchin, Y.V., Iakovleva, N.V., 2009. In search of the origin of FREPs: characterization of *Aplysia californica* bifrinogen-related proteins. Developmental and Comparative Immunology 34 (4), 465–473.

[117] Dong, Y., Dimopoulos, G., 2009. *Anopheles* fibrinogen-related proteins provide expanded pattern recognition capacity against bacteria and malaria parasites. Journal of Biological Chemistry 284 (15), 9835–9844.

[118] Waterhouse, R.M., Kriventseva, E.V., Meister, S., Xi, Z., Alvarez, K.S., Bartholomay, L.C., et al., 2007. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. Science 316 (5832), 1738–1743.

[119] Wang, X., Zhao, Q., Christensen, B.M., 2005. Identification and characterization of the fibrinogen-like domain of fibrinogen-related proteins in the mosquito, *Anopheles gambiae*, and the fruitfly, *Drosophila melanogaster*, genomes. BMC Genomics 6 (1), 114.

[120] Strong, S.J., Mueller, M.G., Litman, R.T., Hawke, N.A., Haire, R.N., Miracle, A.L., et al., 1999. A novel multigene family encodes diversified variable regions. Proceedings of National Academy of Science United States of America 96 (26), 15080–15085.

[121] Yoder, J.A., Mueller, M.G., Wei, S., Corliss, B.C., Prather, D.M., Willis, T., et al., 2001. Immune-type receptor genes in zebrafish share genetic and functional properties with genes encoded by the mammalian leukocyte receptor cluster. Proceedings of National Academy of Science United States of America 98 (12), 6771–6776.

[122] Cannon, J.P., Haire, R.N., Litman, G.W., 2002. Identification of diversified genes that contain immunoglobulin-like variable regions in a protochordate. Nature Immunology 3 (12), 1200–1207.

[123] Cannon, J.P., Haire, R.N., Schnitker, N., Mueller, M.G., Litman, G.W., 2004. Individual protochordates possess unique immune-type receptor repertoires. Current Biology 14, R465–R466.

[124] Hernandez Prada, J.A., Haire, R.N., Allaire, M., Jakoncic, J., Stojanoff, V., Cannon, J.P., et al., 2006. Ancient evolutionary origin of diversified variable regions demonstrated by crystal structures of an immune-type receptor in amphioxus. Nature Immunology 7 (8), 875–882.

[125] Cannon, J.P., Dishaw, L.J., Haire, R.N., Litman, R.T., Ostrov, D.A., Litman, G.W., 2010. Recognition of additional roles for immunoglobulin domains in immune function. Seminars in Immunology 22 (1), 17–24.

[126] Litman, G.W., Rast, J.P., Fugmann, S.D., 2010. The origins of vertebrate adaptive immunity. Nature Reviews Immunology 10, 543–553.

[127] Litman, G.W., Dishaw, L.J., Cannon, J.P., Haire, R.N., Rast, J.P., 2007. Alternative mechanisms of immune receptor diversity. Current Opinion in Immunology 19, 526–534.

[128] Huang, G., Liu, H., Han, Y., Fan, L., Zhang, Q., Liu, J., et al., 2007. Profile of acute immune response in Chinese amphioxus upon *Staphylococcus aureus* and *Vibrio parahaemolyticus* infection. Developmental and Comparative Immunology 31 (10), 1013–1023.

[129] Dishaw, L.J., Mueller, M.G., Gwatney, N., Cannon, J.P., Haire, R.N., Litman, R.T., et al., 2008. Genomic complexity of the variable region-containing chitin-binding proteins in amphioxus. BMC Genetics 9 (1), 78.

[130] Dishaw, L., Ota, T., Mueller, M., Cannon, J., Haire, R., Gwatney, N., et al., 2010. The basis for haplotype complexity in VCBPs, an immune-type receptor in amphioxus. Immunogenetics, 1–9.

[131] Ghosh, J.G., Buckley, K.M., Nair, S.V., Raftos, D.A., Miller, C.A., Majeske, A.J., et al., 2010. *Sp185/333*: A novel family of genes and proteins involved in the purple sea urchin immune response. Developmental and Comparative Immunology 34, 235–245.

[132] Smith, L.C., Ghosh, J., Buckley, K.M., Clow, L.A., Dheilly, N.M., Haug, T., et al., 2010. Echinoderm immunity. In: Soderhall, K. (Ed.), Invertebrate immunity. Landes Bioscience and Spring Science+Business Media.

[133] Rast, J.P, Pancer, Z., Davidson, E.H., 2000. New approaches towards an understanding of deuterostome immunity. Current Topics in Microbiology and Immunology 248, 3–16.

[134] Smith, L.C., Chang, L., Britten, R.J., Davidson, E.H., 1996. Sea urchin genes expressed in activated coelomocytes are identified by expressed sequence tags. Complement homologues and other putative immune response genes suggest immune system homology within the deuterostomes. Journal of Immunology 156 (2), 593–602.

[135] Nair, S.V., Del Valle, H., Gross, P.S., Terwilliger, D.P., Smith, L.C., 2005. Macroarray analysis of coelomocyte gene expression in response to LPS in the sea urchin, Identification of unexpected immune diversity in an invertebrate. Physiological Genomics 22 (1), 33–47.

[136] Terwilliger, D.P., Buckley, K.M., Brockton, V., Ritter, N.J., Smith, L.C., 2007. Distinctive expression patterns of *185/333* genes in the purple sea urchin, *Strongylocentrotus purpuratus*: an unexpectedly diverse family of transcripts in response to LPS, beta-1,3-glucan, and dsRNA. BMC Molecular Biology 8, 16.

[137] Dheilly, N.M., Nair, S.V., Smith, L.C., Raftos, D.A., 2009. Highly variable immune-response proteins (185/333) from the sea urchin *Strongylocentrotus purpuratus*: proteomic analysis identifies diversity within and between individuals. Journal of Immunology 182, 2203–2212.

[138] Terwilliger, D.P., Buckley, K.M., Mehta, D., Moorjani, P.G., Smith, L.C., 2006. Unexpected diversity displayed in cDNAs expressed by the immune cells of the purple sea urchin, *Strongylocentrotus purpuratus*. Physiological Genomics 26 (2), 134–144.

[139] Buckley, K.M., Munshaw, S., Kepler, T.B., Smith, L.C., 2008. The *185/333* gene family is a rapidly diversifying host-defense gene cluster in the purple sea urchin, *Strongylocentrotus purpuratus*. Journal of Molecular Biology 379, 912–928.

[140] Buckley, K.M., Smith, L.C., 2007. Extraordinary diversity among members of the large gene family, 185/333, from the purple sea urchin, *Strongylocentrotus purpuratus*. BMC Molecular Biology 8, 68.

[141] Gilad, Y., Man, O., Paabo, S., Lancet, D., 2003. Human specific loss of olfactory receptor genes. Proceedings of National Academy of Science United States of America 100 (6), 3324–3327.

[142] Rast, J.P., Messier-Solek, C., 2008. Marine invertebrate genome sequences and our evolving understanding of animal immunity. Biological Bulletin 214, 274–283.

[143] Brockton, V., Henson, J.H., Raftos, D.A., Majeske, A.J., Kim, Y.O., Smith, L.C., 2008. Localization and diversity of 185/333 proteins from the purple sea

urchin – unexpected protein-size range and protein expression in a new coelomocyte type. Journal of Cell Science 121 (3), 339–348.
[144] Buckley, K.M., Terwilliger, D.P., Smith, L.C., 2008. Sequence variations between the *185/333* genes and messages from the purple sea urchin suggest post-transcriptional modifications. Journal of Immunology 181 (12), 2203–2212.
[145] Hibino, T., Loza-Coll, M., Messier, C., Majeske, A.J., Cohen, A., Terwilliger, D.P., et al., 2006. The immune gene repertoire encoded in the purple sea urchin genome. Developmental Biology 300, 349–365.
[146] Chester, A., Scott, J., Anant, S., Navaratnam, N., 2000. RNA editing: cytidine to uridine conversion in apolipoprotein B mRNA. Biochimica et Biophysica Acta 1494 (1–2), 1–13.
[147] Miller, C.A., Buckley, K.M., Easley, R.L., Smith, L.C., 2010. An *Sp185/333* gene cluster from the purple sea urchin and putative microsatellite-mediated gene diversification. BMC Genomics 11, 575.
[148] Smith, L.C., 2010. Diversification of innate immune genes: lessons from the purple sea urchin. Disease Models and Mechanisms 3, 274–279.
[149] Massana, R., Murray, A.E., Preston, C.M., DeLong, E.F., 1997. Vertical distribution and phylogenetic characterization of marine palnktonic *Archaea* in the Santa Barbara Chanel. Applied and Environmental Microbiology 63 (1), 50–56.
[150] Whitman, W.B., Coleman, D.C., Wiebe, W.J., 1998. Prokaryotes: the unseen majority. Proceedings of National Academy of Science United States of America 95 (12), 6578–6583.
[151] Llobet-Brossa, E., Rossello-Mora, R., Amann, R., 1998. Microbial community composition of Wadden sea sediments as revealed by fluorescence in situ hybridization. Applied and Environmental Microbiology 64 (7), 2691–2696.
[152] Suttle, C.A., 2007. Marine viruses – major players in the global ecosystem. Nature reviews Microbiology 5, 801–812.
[153] Munn, C.B., 2006. Viruses as pathogens of marine organisms – from bacteria to whales. Journal of the Marine Biological Association of the UK 86, 453–467.
[154] Woolhouse, M.E.J., Webster, J.P., Domingo, E., Charlesworth, B., Levin, B.R., 2002. Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nature Genetics 32, 569–577.
[155] Van Valen, L., 1973. A new evolutionary law. Evolutionary Theory 1, 1–30.
[156] Carroll, L., 1871. Through the looking-glass and what Alice found there. Macmillan and Co, London.
[157] Carroll, S.B., Grenier, J.K., Weatherbee, S.D., 2005. From DNA to diversity, Second ed. Blackwell Publishing, Malden MA, 258.
[158] Green, B.M., Finn, K.J., Li, J.J., 2010. Loss of DNA replication control is a potent inducer of gene amplification. Science 329 (5994), 943–946.
[159] Khalturin, K., Anton-Erxleben, F., Sassmann, S., Wittlieb, J., Hemmrich, G., Bosch, T.C., 2008. A novel gene family controls species-specific morphological traits in Hydra. PLoS Biology 6 (11), e278.
[160] Hemmrich, G., Miller, D.J., Bosch, T.C.G., 2007. The evolution of immunity: a low-life perspective. Trends in Immunology 28 (10), 449–454.
[161] Litman, G.W., Cannon, J.P., Dishaw, L.J., 2005. Reconstructing immune phylogeny: new perspectives. Nature Reviews Immunology 5 (11), 866–879.
[162] Zhang, Q., Zmasek, C.M., Godzik, A., 2010. Domain architecture evolution of pattern-recognition receptors. Immunogenetics 62, 263–272.
[163] Kapitonov, V.V., Jurka, J., 2004. Harbinger transposons and an ancient HARBI1 gene derived from a transposase. DNA Cell Biology 23 (5), 311–324.
[164] Kurata, S., 2006. Recognition and elimination of diversified pathogens in insect defense systems. Molecular Diversity 10 (4), 599–605.
[165] Miller, CA. Assembly and analysis of a *Sp185/333* gene cluster from the sea urchin. Evidence for microsatellite-mediated gene and segmental duplication of this highly variable immune response gene family. MS thesis: Program in Genomics and Bioinformatics, George Washington University 2009.
[166] Michel, T., Reichhart, J.M., Hoffmann, J.A., Royet, J., 2001. *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. Nature 414 (6865), 756–759.
[167] Royet, J., Dziarski, R., 2007. Peptidoglycan recognition proteins: pleiotropic sensors and effectors of antimicrobial defences. Nature Reviews Microbiology 5 (4), 264–277.
[168] Mellroth, P., Karlsson, J., Hakansson, J., Schultz, N., Goldman, W.E., Steiner, H., 2005. Ligand-induced dimerization of *Drosophila* peptidoglycan recognition proteins *in vitro*. Proceedings of National Academy of Science United States of America 102 (18), 6455–6460.
[169] Ramet, M., Manfruelli, P., Pearson, A., Mathey-Prevot, B., Ezekowitz, R.A., 2002. Functional genomic analysis of phagocytosis and identification of a *Drosophila* receptor for *E. coli*. Nature 416 (6881), 644–648.
[170] Lim, J.H., Kim, M.S., Kim, H.E., Yano, T., Oshima, Y., Aggarwal, K., et al., 2006. Structural basis for preferential recognition of diaminopimelic acid-type peptidoglycan by a subset of peptidoglycan recognition proteins. Journal of Biological Chemistry 281 (12), 8286–8295.
[171] Takehana, A., Katsuyama, T., Yano, T., Oshima, Y., Takada, H., Aigaki, T., et al., 2002. Overexpression of a pattern-recognition receptor, peptidoglycan-recognition protein-LE, activates imd/relish-mediated antibacterial defense and the prophenoloxidase cascade in *Drosophila* larvae. Proceedings of National Academy of Science United States of America 99 (21), 13705–13710.
[172] Maillet, F., Bischoff, V., Vignal, C., Hoffmann, J., Royet, J., 2008. The *Drosophila* peptidoglycan recognition protein PGRP-LF blocks PGRP-LC and IMD/JNK pathway activation. Cell Host and Microbe 3 (5), 293–303.