ISJ 13: 355-373, 2016 ISSN 1824-307X

REVIEW

Tuning the host-pathogen relationship through evolution with a special focus on the echinoid Sp185/333 system

LC Smith[1], MR Coscia[2]

[1]Department of Biological Sciences, George Washington University, Washington DC, USA
[2]Institute of Protein Biochemistry, National Research Council of Italy, Naples, Italy

Accepted November 3, 2016

Abstract

Diversification of immune genes in host organisms that are in deadly arms races with pathogens has resulted in a wide range of approaches by which the host survives. Well known examples of adaptive immunity in vertebrates include somatic recombination of the immunoglobulin gene family and assembly of the variable lymphocyte receptors. The CRISPR-Cas system in bacteria and archaea is also considered adaptive. For invertebrates that survive in the absence of adaptive immunity, innate immune diversity is accomplished based on functions of clusters of immune genes such as *FREP*s, *VCBP*s, *C1q*s, *TLR*s, and *R* genes. Single copy gene diversity can be accomplished through extensive alternative splicing or increases in alleles in populations. The *Sp185/333* gene family in the purple sea urchin has multiple levels in which to generate immune diversity. These include clustered *Sp185/333* genes, genomic instability in regions harboring the genes, predicted mRNA editing, expression of broad repertoires of Sp185/333 proteins, and post translational modifications. The Sp185/333 proteins have anti-pathogen activities and an individual recombinant protein can bind to multiple foreign cells and molecular patterns. The underlying characteristics of gene clusters, many with repeats, that are present in unstable genomic regions is common to a number of these examples, and is likely of central importance for organisms that survive solely on innate immunity.

Key Words: Ig; TcR; VLR; DSCAM; FREPs; VCBP; TLR; Fu/HC; *R* genes; Sp185/333

Introduction

When Alice went through the looking glass, she joined the Red Queen in the Queen's race. During the race when Alice complained about running so hard for so long and getting nowhere, the Queen said to her, "it takes all the running you can do, to keep in the same place" (Carroll, 1871) (Fig. 1). Although this statement seems as crazy as both Wonderland and the Red Queen appeared to be, the idea has been employed as a hypothesis for evolution. The Red Queen's race is analogous to the immunological arms race or coevolution between the diversity of an immune system and its effectiveness in protecting the long-lived host vs. the continuously changing virulence of pathogens that have (most often) much shorter life spans (Dawkins and Krebs, 1979). This host-pathogen arms race is driven by the need to adapt constantly to survive for both the host and the pathogen, which in turn shapes the virulence and success of the pathogens as well as the detection and effector mechanisms of the host immune system (Haldane, 1949; Van Valen, 1973). The host-pathogen battle drives molecular evolution and affects all ranges of life including plants and animals and their pathogens, in addition to bacteria and their bacteriophages (Sironi *et al*., 2015). An example of this process in action is illustrated by an elegant study illustrating the experimental evolution of the bacterium *Pseudomonas fluorescens* SBW25 in the presence of its pathogen, bacteriophage Φ2 (Patterson *et al.,* 2010). The outcome of this arms race demonstrates that evolutionary changes occur at higher rates when the host and pathogen coevolve together, and drives greater genotype diversification in both *Pseudomonas* and the phage compared to the bacteria alone or to the phage after the host cells have been killed by the phage. As expected, the bacteria show greater genetic diversity, and remarkably the phage genes that function in infectivity are also selected to change more rapidly. Because the number of genes that are involved both in pathogen virulence and host immunity is finite,

Corresponding author:
L. Courtney Smith
Department of Biological Sciences
Science and Engineering Hall
Suite 6000, 800 22nd St NW
Washington DC 20052
E-mail: csmith@gwu.edu

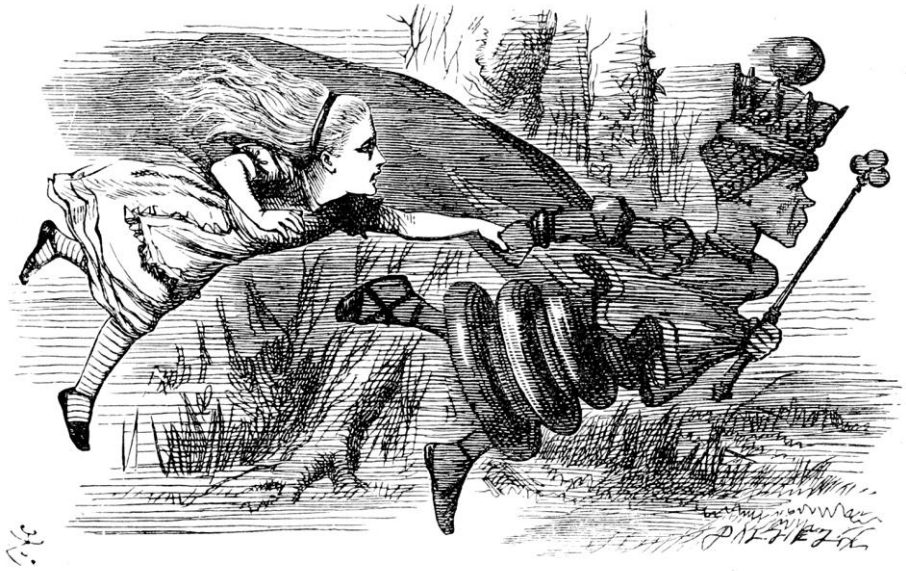

Fig. 1 Alice and the Red Queen in the Red Queen's race. They both run as fast as they can but remain in the same place. Illustration by Sir John Tenniel from "Through the Looking-Glass" (Carroll, 1871).

different approaches to diversification are incorporated into a few genes for increasing pathogen virulence or are incorporated into host immune genes to gain an advantage over the adversary in the arms race. Here, we review a range of examples of how a pathogen or a host changes or increases diversity of genes that function in the arms race. We end by focusing on the immune diversity in the *Sp185/333* system of the purple sea urchin.

Pathogen diversification for altered or increased virulence

There is a wide range of how microbes generate genomic diversity to alter their phenotype to overcome a host immune system and/or to improve avoidance tactics to hide from host attack (Deitsch *et al.*, 2009). Microbial pathogens adapt constantly to avoid mechanical clearance, circumvent barriers to invasion, avoid recognition and destruction by the host immune system, and/or alter or derail the host immune system with the aim of overwhelming the host response to result in a successful infection. The success in pathogen variation is measured by fitness or success in infection and proliferation leading to subsequent infection of other hosts. DNA recombination is employed by both bacterial and eukaryotic pathogens, and can move coding regions into an area of the genome that allows expression (Deitsch *et al.*, 2009). This process can fuse intact genes or fragments of genes to generate chimeric sequences, or can move a promoter near a gene (or vice versa) that does not have a functional promoter, leading to altered expression and/or altered protein sequences. This can change not only whether a gene is expressed but can also change the level of expression. Antigenic variation or hypervariability, which includes phase variation, is the expression of a particular antigen or set of antigens sequentially over time that alters the pathogen surface phenotype and results in temporary avoidance of the host immune response until the next surface antigen is expressed (Fig. 2). Antigenic variation can be accomplished by changes in gene expression in single or multi-copy genes that are members of expanded families. Modifications to gene sequences within families can result from gene conversion or recombination among similar genes, and slipped strand mispairing followed by excision and repair (Deitsch *et al.*, 2009; Oren *et al.*, 2016). Variation in expression among genes that are members of an expanded family that have variable sequences is a very effective means for avoiding or "out running" immune recognition by the host. Furthermore, changes in gene expression within a gene family may be related to variations in the environment, and success in infecting a host that may be based on which genes within the family are expressed and under what circumstances. An example of this type of immune avoidance is the variable surface glycoprotein (VSG) that changes expression over time on the surface of trypanosome parasites, which are the pathogens that cause African sleeping sickness (Horn, 2014) (Fig. 2). Using a large pool of mosaic *VSG* genes, the parasite is able to escape host antibody recognition and attack by continuosly changing epitopes on its protective extracellular coat. Over time, expression switches to an antigenically different *VSG* by silencing the initially expressed gene and activating the transcription of a new *VSG*, or by recombining a *VSG* from another region in the genome into the active transcription site.

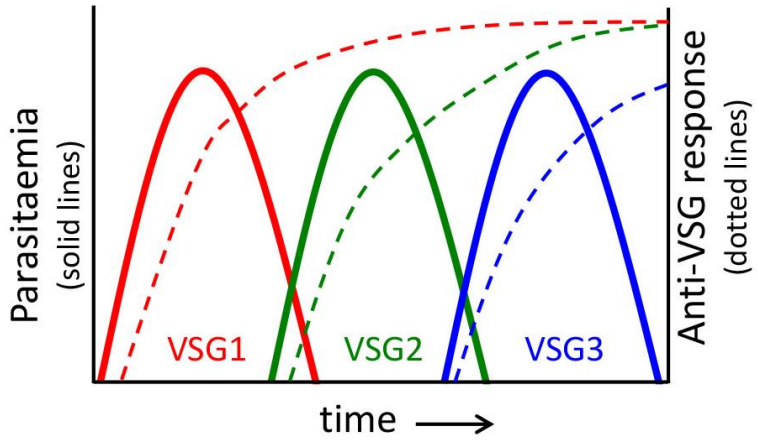

Fig. 2 *VSG* gene expression in Trypanosome parasites switches over time. This simplified scheme of antigenic phase variation over time in trypanosome infection illustrates changes in the version of VSG proteins (red, green, blue) on the parasite surface that enables it to avoid the immune response temporarily. Once the host immune response is directed to the expressed VSG (dotted lines), parasites with a variant VSG expression survive, take over the population, and are the source of the relapsing parastaemia. This figure is modified from (Horn, 2014).

Antigenic variation also occurs in bacteria and is characterized by multiple copies of genes that are slightly different, such as when variation is limited to a specific region of the gene sequence. Changes in antigens can be controlled by 1) switch mechanisms that regulate independent expression, 2) the expression of a single gene while other members of the family are silenced, and 3) continuosly changing which gene copy is expressed (Finlay and McFadden, 2006). Antigenic variation is also employed by viruses. The accumulation of mutations within viral genes that encode epitopes that are bound by antibodies is a process of antigenic drift followed by antibody-based selection, which gives rise to increased success in viral infectivity and is directly related to antigenic changes to the viruses. This is particularly frequent in the case of RNA viruses in which antigenic variation depends mostly on the higher mutational frequency of RNA replicases (Lauring *et al.,* 2013). The strategies of the human immunodeficiency virus (HIV) to escape the host immune system is a combination of a high mutation rate ($3x10^{-5}$ nucleotides per viral replication cycle) plus retroviral recombination to alter viral recognition by host cytotoxic T lymphocytes and binding by antibodies (Hamelaar, 2012). Furthermore, conformational masking and glycan shielding of the viral envelope protein and its unfolding only upon entry into a cell serve to neutralize the host antibody response (Kwon *et al.*, 2015). Antigenic variations are a good means by which to avoid, at least temporarily, the host immune response and are employed by a wide range of pathogens and parasites.

A common feature of genes that function in immune avoidance is the expansion of genes in pathogens that encode effector proteins that disrupt the host immune response. This has been documented repeatedly for bacterial pathogens of higher plants (Macho and Zipfel, 2015) and is also the means by which *Shigella* infects humans (Ashida *et al.*, 2015) and *Bacillus thuringiensis* infects *Caenorhabditis elegans* (Schulte *et al.*, 2013). For some plant pathogens and for *Shigella*, microbes inject effector peptides into the host cell cytoplasm that disrupt or derail the innate immune response of the host. Similarly, for both *Bacillus* and *Shigella*, the effector peptides block pathogen specific responses in nematodes and humans, respectively, and allow the pathogen to proliferate successfully. In addition to changes in genes encoded in the genome, prokaryotes also carry virulence genes on plasmids. A *Bacillus* pathogen of *C. elegans* has a plasmid with variable numbers of crystal toxin genes with altered sequence diversity (Schulte *et al.*, 2010) in addition to increased gene diversity in the genome (Schulte *et al.*, 2013) in response to interactions with the host. Genes encoding the effectors and other virulence genes are greatly expanded in the genomes of these pathogens when they are directly involved in the arms race and adapt to the co-adaptations of the host immune system.

There is a wide range of means by which horizontal transfer can be used by microbes to change their genotypes and phenotypes that can be beneficial for increased virulence capabilities and proliferation success. Mating, or bacterial conjugation, is a key mechanism for the horizontal transfer of DNA, including virulence genes, among microbes (de la Casa-Esperon, 2012; Cabazon *et al.,* 2015). Transfer is mediated by the formation of the Type IV secretion system by Gram negative bacteria and mating pair formation results in membrane fusion and the transfer of DNA from the donor to recipient cell. This complex is employed in "double duty" activity because it is also used to inject virulence factors into eukaryotic host cells to increase bacterial survival and proliferation. Non-mating horizontal or lateral DNA transfer is another

means for altering the genome of a pathogen (de la Casa-Esperon, 2012). Microbes can acquire DNA either from other cells of the same species, or from cells of different species that can modify the means by which the altered microbe survives in the environment, in a host, or both. The outcome of DNA acquisition can increase the pathogen diversity and alter the interactions between host and pathogen. Natural competence to take up DNA from the environment rather than directly from other cells can also transform the capabilities of microbes. Horizontal DNA transfer is frequent in marine *Vibrios* not only among the cells but also based on the uptake of environmental DNA (Sun *et al*., 2013). Acquired DNA that includes genes encoding proteins with functions that improve fitness to survive a host immune response are likely to be selected for, maintained, and passed on to daughter cells. Bacteriophage infection can also result in the acquisition of bacterial DNA fragments that have been captured within the phage genome and can be transferred to other bacteria upon subsequent infections. "Infections of infections" also occur in the form of mobile genetic elements that use phages for transfer from one bacterial cell to another and can influence the pathogenicity and evolution of microbes (Penades *et al.,* 2015). The wide variation and efficiency of microbial diversification in the arms race with their hosts, particularly given their short generation times, have driven the diversity mechanisms in the long-lived multicellular hosts.

Host immune diversification

Adaptive immune systems in vertebrates; V(D)J recombination

The long-lived (relative to microbial pathogens) host response to constant genomic diversification and virulence changes in pathogens is to diversify the sequences of single copy genes in individual immune cells and/or to expand the copies of a gene followed by diversification of the sequences in the resulting gene family. Both approaches have been well documented in animals and the second approach has been particularly well employed in higher plants. Gnathosomes diversify their immunoglobulin (Ig) family genes, which function in both pathogen detection and response, by somatic recombination of variable, diversity and joining (V(D)J) segments to generate antigen binding sites in the variable exon of the rearranged single copy Ig and T cell Receptor (TcR) genes (Tonegawa, 1983; Davis *et al*., 1984). The enzymes of central importance in this process are the recombinase activating genes (*RAG*s) that function in "cutting and pasting" segments of the Ig gene family into a functional exon (Teng and Schatz, 2015) and the activation induced deaminase (*AID*) gene that is involved in class switching and affinity maturation of Ig genes (Muramatsu *et al*., 2000). However, other variations in Ig family proteins are known that are not involved in V(D)J recombination and show sequence diversity in the constant region. For example, the hinge region in IgM from an Antarctic fish (*Trematomus bernacchii*) is unique relative to the hinge regions of all other vertebrate Ig isotypes because it is much longer, is encoded by a 3′ extension of the Cμ2 exon, and is highly polymorphic

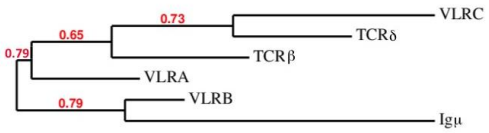

Fig. 3 The origins of the adaptive immune system in vertebrates inferred from the TM regions of immune receptors. Preliminary evidence of a common origin is derived from the phylogenetic relationships among Agnathan VLRs, Gnathostome Igμ, TcR β and δ chains based on the consensus of their transmembrane region nucleotide sequences. Sequences were aligned by ClustalX (ver. 1.8) and the phylogenetic tree was constructed by the Neighbor-joining method with 100 bootstrap iterations. Positions with gaps were excluded from the alignment used to generate the tree.

among individuals in the population (Coscia *et al*., 2012). Because a longer hinge region contributes to greater flexibility of the antibody binding sites, this may improve antigen binding capacity and effectiveness for IgM in this fish. Changes in the size of the hinge region may result from variations in the splice sites that alter the mRNA, or may be the result of non-synonymous substitutions at several specific sites that are under purifying selection compared to most other sites. Consequently, because the polymorphism mostly affects the hinge, which is long enough to be solvent exposed, it may protect this region from cleavage by pathogen proteases, such as those of the nematode parasites that cause a high incidence of infections in Antarctic fish (Coscia and Oreste, 1998; Coscia *et al*., 2000).

In addition to differences among antibodies outside of the variable regions, there are also conserved features in antibodies that are present across a wide range of vertebrates. One of these features is the motif FXXXFXXS/TXXXS (X; any amino acid) in the Ig transmembrane (TM) region that can be considered a universal attribute of Ig molecules, and likely play a significant role in their function (Varriale *et al*., 2010). Interestingly, a 72 nucleotide sequence encoding the TM region includes 13 amino acids under purifying selection and show some identity with the TM regions from Agnathan variable lymphocyte receptor (VLR)A, VLRB, VLRC (see below) (Pancer *et al*., 2005), and murine TcR β and δ chains (Coscia and Oreste, unpublished) (Fig. 3). Nearly half of the positions in the VLR TM region are conserved including the 72 nucleotides. Although VLRs are phylogenetically unrelated to Igs, both classical antibodies from the Gnathostomes and VLRs from Agnathans show regions of diversity within the assembled or copied antigen binding regions, in addition to regions outside of the antigen binding sites. Furthermore, certain regions are conserved likely because of essential functions perhaps including interactions with TM regions of other proteins on the lymphocyte surface. Thus, the presence of a shared sequence

supports the notion that a common ancestor of the jawed and jawless vertebrates likely had both humoral and cellular based adaptive immunity.

The apparent immunological "Big Bang" of adaptive immunity (Bernstein *et al*., 1996) and the sudden appearance of somatic recombination of the Ig gene family in higher vertebrates was thought to have arisen by a horizontal transfer of a transposon encoding recombinase, which was inserted into an ancestral Ig superfamily gene during the early evolution of jawed vertebrates (Bernstein *et al.,* 1996; Agrawal *et al.,* 1998; Hiom *et al.,* 1998; Schluter *et al*., 1999; Flajnik, 2004). However, *ProtoRAG* genes with similarities to vertebrate *RAG*s and Transib transposases with putative "cut and paste" activity have been found in a range of invertebrates including the purple sea urchin (*Strongylocentrotus purpuratus*), hydra (*Hydra magnipapillata*), a sea anemone (*Nematostella vectensis*), and amphioxus (*Branchiostoma floridae*) in addition to related Transib transposons identified in the fruit fly (*Drosophila melanogaster*) and the mosquito (*Anopheles gambiae*) (reviewed in (Fugmann, 2010). Although these invertebrate gene sequences show similarities to *ProtoRAG*-like sequences, most are not affiliated with terminal inverted repeats, which are typical of transposases and are associated with binding and cutting activity. However, functional analysis of a Transib transposase from the corn earworm (*Helicoverpa zea*) demonstrated DNA cleavage and ligation that is similar to vertebrate RAG1 (Hencken *et al.,* 2012). In the purple sea urchin (*S. purpuratus*), the RAG1 protein homologue, SpRag1-like, associates with RAG2 homologues including SpRag2-like from the sea urchin and RAG2 from the bull shark, and it also binds to the mouse one-turn recombination signal sequence (Fugmann et al., 2006). Furthermore, when associated with RAG2, SpRAG1-like supports low levels of V(D)J recombination *in vitro* (Carmona *et al*., 2016). The *ProtoRAG* genes in amphioxus, *Branchiostoma belcheri*, are linked and oriented "head to head" as in other animals, are associated with terminal inverted repeats (unlike the sea urchin *SpRAGL* genes), and support DNA cleavage, transposition, and the ligation of terminal inverted repeats, albeit at low effeciency, which is similar to signal joint ligation in vertebrates (Huang *et al*., 2016). The identification of *RAG* gene pairs in basal, extant deuterostomes in addition to other invertebrates with Transib transposase genes indicates that the origins of the RAG enzyme mediated adaptive immune system in higher vertebrates may have resulted from vertical inheritance of the ancestral *ProtoRAG* transposases in invertebrates rather than by lateral transfer from a microbe to a basal vertebrate (Fugmann, 2010).

Assembly of variable lymphocyte receptors

The jawless vertebrates, hagfish and lampreys, have an alternative adaptive immune system that uses a completely different mechanism to diversify the *VLR* genes, which function as antibodies and TcRs in these fish (Pancer *et* al., 2005; Boehm *et al*., 2012). There are three *VLR* genes that encode VLRA, VLRB and VLRC that have functions similar to αβTcR, Ig, and γδTcR, respectively (reviewed in (Flajnik, 2014) (Table 1). The *VLR* genes are assembled in a "copy choice" or gene conversion-like mechanism that randomly selects leucine rich repeats (LRR) from flanking cassettes of LRR segments and copies them into and replaces the non-coding region of the single copy, incomplete, germline *VLR* gene (Nagawa *et al.,* 2007; Rogozin *et al.,* 2007; Boehm *et al.,* 2012). The assembly employs short stretches of nucleotides that match between different LRRs, which can be located at any place within an individual LRR. The assembly activity takes place in the thymoid regions at the tips of the gill filaments for *VLRA* and *VLRC* (Bajoghli *et al*., 2011) and in the hematopoietic regions of the typhlosole and kidney for *VLRB* genes (reviewed in Boehm *et al*., 2012; Kasahara, 2015). Unlike the Ig system in higher vertebrates, *VLR* gene assembly does not use *RAG* encoded recombinases, but is mediated by two cytidine deaminases (CDAs) in the lamprey (Guo *et al*., 2009). Expression of the *CDA1* gene is restricted to the tips of the gill filaments in association with *VLRA* and *VLRC* expression, whereas expression of *CDA2* is associated with *VLRB* expression in the typhlosole (Bajoghli *et al*., 2011). Although the assembly results in a combinatorial set of full length LRRs in *VLR* genes that appears to maintain the reading frame (Nagawa *et al*., 2007; Rogozin *et al.,* 2007), apoptotic lymphocytes are detected and selection of *VLRA* and *VLRC* is suggested based on non-random numbers of LRRs (Sutoh and Kasahara, 2014). The astounding parallels between the Ig/TcR system and the VLR system suggest the possibility of common origins of antigen diversification and mechanisms of lymphocyte functions in the jawed and jawless vertebrates that predates RAG activities (Table 1).

Table 1 Parallels in the antigen receptors and types of lymphocytes in jawless and jawed vertebrates[1]

Jawless Vertebrates		Jawed Vertebrates	
Antigen Receptor	**Cell Type**	**Antigen Receptor**	**Cell Type**
VLRA	on the surface of T-like lymphocytes	αβTcR	on the surface of αβ T lymphocytes
VLRC	on the surface of T-like lymphocytes	γδTcR	on the surface of γδ T lymphocytes
VLRB	on the surface of and secreted from B-like lymphocytes	BCR	on the surface or and secreted from B lymphocytes

[1]modified from Kasahara and Sutoh (2014).

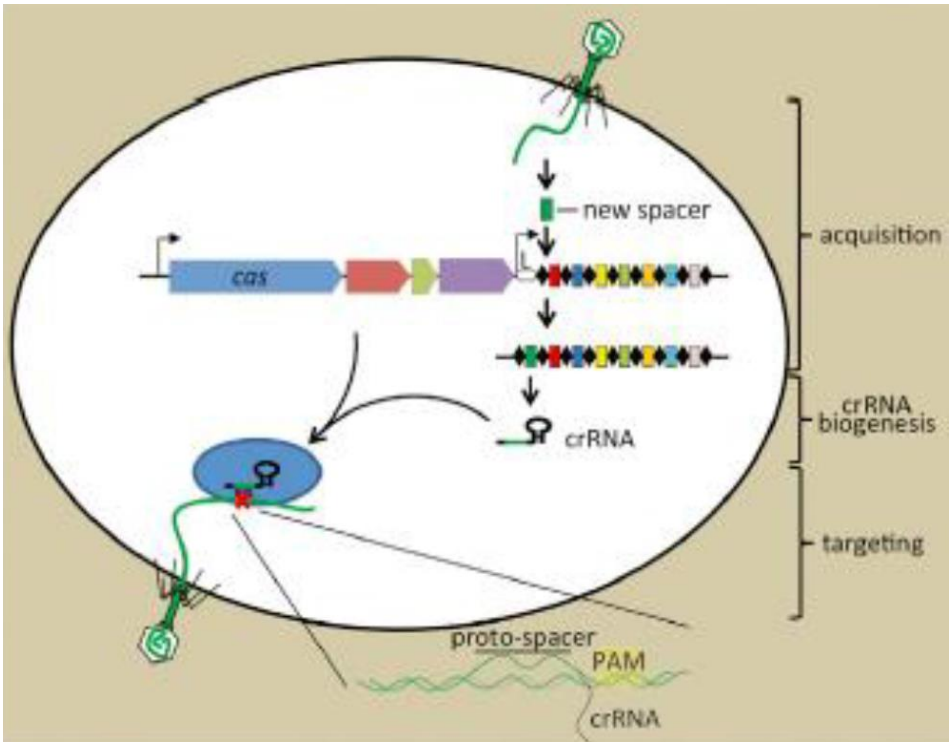

Fig. 4 CRISPR-Cas, the adaptive immune system in prokaryotes. The CRISPR locus is composed of CRISPR-associated (Cas) genes followed by a start transcription site (bent arrow), a leader (L) and a series of repeats (black triangles) separated by spacers (rectangles of various colors). Upon infection with a bacteriophage (shown at the top), a section of the phage DNA (green) is incorporated as a new spacer at the beginning of the CRISPR array. Upon subsequent infection by a phage with DNA sequence that is the same as the spacer, the expression of the CRISPR (cr)RNA that includes the "green" spacer results in a sequence match to the DNA of the second phage infection and results in cleavage and destruction of the invading DNA. This figure is from (Barrangou and Marraffini, 2014) and reprinted from Molecular Cell with permission from Elsevier.

CRISPR-Cas adaptive immunity in prokaryotes

A third version of adaptive immunity mediated by clustered regularly interspersed repeats (CRISPR) has been identified in about 50% of bacteria and most of the archaea (Sternberg *et al.,* 2016). The CRISPR system functions to protect bacteria from bacteriophage infection and lysis using activities with similarities to RNA interference in higher animals and plants. The CRISPR region of the bacterial genome acquires and incorporates short stretches of bacteriophage DNA called spacers, which act as a mechanism to recognize bacteriophage DNA in subsequent infections (Fig. 4). This sequence similarity is employed to cleave the foreign DNA and RNA using the CRISPR associated (Cas) proteins (Jiang and Doudna, 2015; Marraffini, 2015). Although there are a range of variations in CRISPR systems in different bacterial and archaeal species, in general it is the prokaryote adaptive immune system that is effective in protection against bacteriophages and other foreign DNA such as mobile genetic elements (Barrangou and Marraffini, 2014). Because the spacers are incorporated into the genome, immunity to specific phages is heritable and adaptive, although it is Lamarckian rather than Darwinian and is not anticipatory. Although it was long believed that higher vertebrates were the only group to function with adaptive immunity, the VLR and CRISPR/Cas examples demonstrate that there are multiple mechanisms for adaptive self protection against the diversity of pathogens.

Diversity of innate immunity in invertebrates

Drosophila melanogaster

Most animals and all plants survive solely on their innate immune functions. *Drosophila melanogaster* was one of the first model organisms that lent itself to analysis because of its well understood genetics, short life cycle, and easily modified genes for functional analysis to understand how the immune system responds to pathogenic threats and injury. There are four major immune signaling pathways in *Drosophila* which are Toll, IMD, JAK/STAT and JNK (Stokes *et al.,* 2015; Huang *et al.,* 2016) that enable the fruit fly to translate the detection of pathogens to the production of anti-microbial peptides (AMPs) that control pathogens. The same signaling pathways are present in the common house fly, *Musca domestica*, with similar numbers of genes encoding proteins that function in the signaling pathways

(Armitage *et al*., 2015). However, compared to *Drosophila* the house fly genome shows a great expansion in the numbers of immune genes encoding recognition and effector proteins with increased sequence diversity. Clearly, the difference in life histories between the fruit fly and the house fly - living on and eating fruit vs. the well-known preferred habitats of the house fly on which it walks, eats, and lays eggs - is reflected in the diversity of the immune system in the house fly. For any given organism, the activities and diversity of immune function is the outcome of co-evolution with the types and diversity of the pathogens with which it shares the habitat, and its ability to keep pace with rapidly evolving microbes (reviewed in (Ghosh *et al.,* 2011).

DSCAM *in arthropods*

One aspect of the *Drosophila* immune system, in addition to many other members of the arthropod phylum, is the gene encoding the Down Syndrome Cell Adhesion Molecule (*DSCAM*). *DSCAM* is present in more advanced arthropods as a single copy gene with duplications of specific exons (Fig. 5), or as a multicopy gene family in more basal arthropods that do not show exon expansions (Armitage *et al.,* 2015; Brites and du Pasquier, 2015). DSCAM proteins in vertebrates and arthropods have 10 Ig domains and six fibronectin domains in the extracellular region, a transmembrane region, and a cytoplasmic tail that includes signaling motifs (Schmucker and Chen, 2009; Ng *et al*., 2014). The single copy *DSCAM* gene in insects generates significant sequence diversity in the encoded proteins (Armitage *et al*., 2015). Diversity is based on extensive expansion of exons that encode the 5′ portions of Ig domains 2 and 3 and all of Ig 7, which show mutually exclusive alternative splicing of the duplicated exons in the primary transcripts that encode each of these Ig domains (Fig. 5). Mutually exclusive alternative splicing of duplicated exons is based on insufficient distance between two duplicated exons, incompatible splicing signals located in regions between different sets of exons, secondary stem/loop structure of the *DSCAM* mRNA, and trans-acting factors that regulate splicing (Park *et al*., 2004; Kreahling and Graveley, 2005; Schmucker and Chen, 2009; Hemani and Soller, 2012; Spoel and Dong, 2012). The transcriptional outcome for this single copy gene is a theoretical 38,016 different proteins in *Drosophila*, about half of which are expressed in neuronal tissue and function in axon guidance, while the rest are expressed in hemocytes (Schmucker *et al*., 2000; Watson *et al*., 2005; Schmucker and Chen, 2009). The resulting sequence diversity in the three Ig domains defines the shape of the extracellular region that is based on interactions among the Ig domains, and influences interactions with particular binding targets. The *DSCAM* response in mosquito and shrimp immunity in response to pathogen challenge is to mount the DSCAM proteins on the hemocyte surface and to secrete or release DSCAM proteins into the hemolymph (Deitsch *et al.,* 2009; Dong and Dimopoulos, 2009; Ng *et al.,* 2014). Remarkably, the versions of DSCAM that are present in the hemolymph and on hemocytes after challenge show elevated specificity towards the infecting pathogen and function as opsonins to induce phagocytosis, clearance, and host protection, all of which lead to survival (Dong *et al*., 2006; Hung *et al*., 2013). DSCAM appears to be a major means by which the arthropods generate immune sequence diversity in response to pathogens. Overall, there are several means for generating sequence diversity in essentially single copy or very small families of genes in animals. These include somatic recombination of duplicated gene segments for *Ig* and *TcR* genes, copy choice or gene conversion assembly of duplicated LRR cassettes in the *VLR* genes, and extensive alternative splicing of duplicated exons in *DSCAM* genes in arthropods.

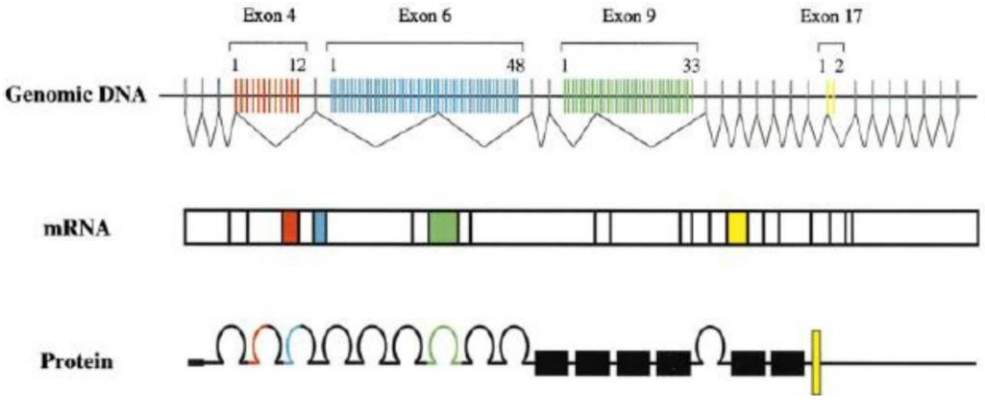

Fig. 5 The *DSCAM* gene structure and alternative splicing. The exons in the *DSCAM* gene in *Drosophila* that encode the N-terminal portions of Ig domains 2 (red) and 3 (blue), plus all of Ig domain 7 (green) are duplicated in tandem in the gene (top). In addition, the exon encoding the transmembrane region (yellow) is duplicated. The duplicated exons undergo alternative splicing in the mRNA (middle), to produce a highly diversified set of proteins (bottom) that are expressed in neurons and hemocytes. This figure is from (Schmucker *et al*., 2000) and reprinted from Cell with permission from Elsevier.

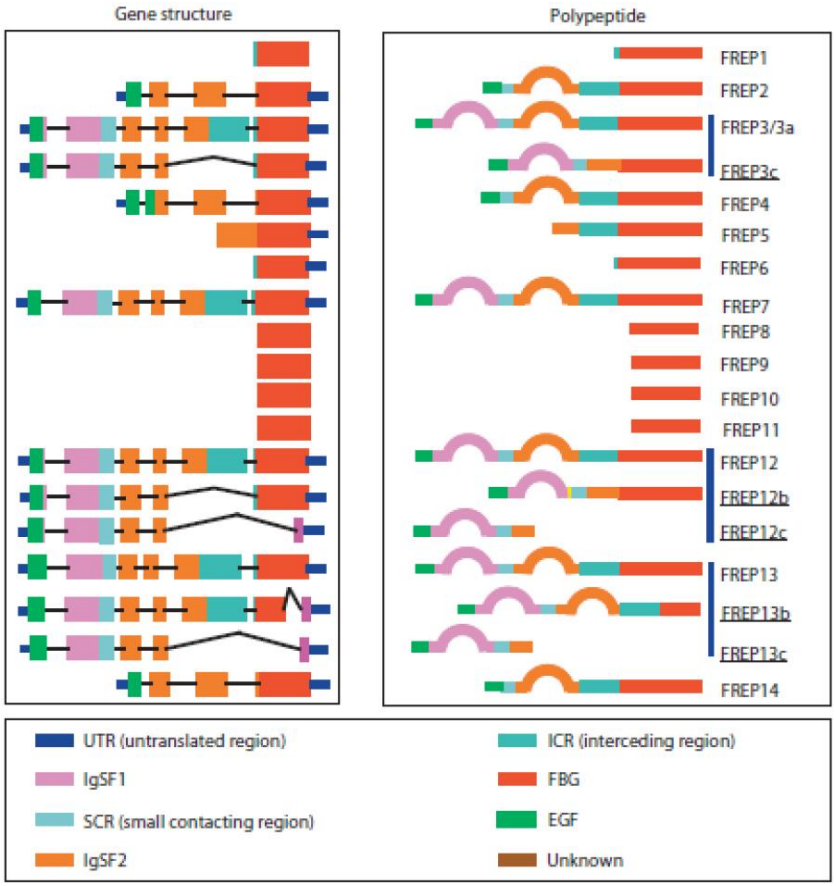

Fig. 6 Diversity in the *FREP* gene families and the encoded proteins. The *FREP* genes (on the left) have one to seven exons and encode proteins (on the right) with one or two Ig superfamily (IgSF) domains or proteins with only the fibrinogen (FBG) domain, although some include the interceding region (ICR). Protein families listed to the right and are underlined produce mRNAs that are alternatively spliced. Color coding in the genes and the proteins is defined in the legend below; epidermal growth factor domain, EGF. This figure is from (Hanington and Zhang, 2011) with minor modifications and is reprinted from the Journal of Innate Immunity with permission from Karger Publishers.

FREPs *in gastropods*

Another approach for generating sequence diversity in immune systems is through gene duplication that generates clustered families of similar genes followed by sequence variation among gene family members, which is driven by pathogen pressure. One example is the freshwater gastropod mollusc, *Biomphalaria glabrata*, that is an intermediate host for trematode parasites and serves as a vector for terminal infection in birds or mammals including humans (Gordy *et al.*, 2015; Macho and Zipfel, 2015). Snails have 14 families of fibrinogen related protein (*FREP*) genes that are expressed in response to infection and result in wide arrays of FREPs in their hemolymph (Adema *et al.*, 1997). FREPs are composed of a signal sequence, one or two Ig domains, an interceding region of variable length and sequence, and a conserved fibrinogen related domain (Gordy *et al.,* 2015) (Fig. 6). The gene structure shows great diversity including variations in the Ig exons, changes to the interceding domain exons, deletion of the fibrinogen domain exon, or the absence of all exons except for the fibrinogen domain exon (Fig. 6). In addition to variations in the structure of the genes, *FREP2*, 3, 4, and *12* gene families in the germ line DNA of *B. glabrata* appear to be expanded and diversified in hemocyte DNA (Gordy *et al.*, 2015). For example, the estimate for the number of germ line *FREP13* genes is about four, whereas the number of different sequences of the first Ig coding region in hemocyte DNA is increased by about 10 fold and may be expanded by somatic recombination (Loker *et al.*, 2004; Zhang *et al.*, 2004; Hanington *et al.*, 2010). The *FREP* gene families are interesting examples of small gene families that are expanded and diversified in the snail hemocytes (Adema, 2015), however the mechanism by which this occurs is not known. The *FREPs* highlight the concept that somatic diversification of genes encoding innate immune molecules can occur in (at least some) invertebrates, and therefore is not exclusive to vertebrates.

Immune diversity from expanded gene families and/or from expanded alleles per locus

Variable region-chitin binding proteins

A family of innate immune receptors, the variable region-containing chitin-binding proteins (VCBPs), are present in the marine Protochordates, *Branchiostoma floridae* and *Ciona intestinalis*, and encode proteins composed of a leader, two tandem Ig V-type domains, and a chitin-binding domain (Cannon *et al.,* 2002, 2004; Dishaw *et al.,* 2010, 2011); reviewed in Liberti *et al*. (2015). There are five *VCBP* genes (A-D) in *C. intestinalis*, that have little sequence diversity, however, two of the five genes in *B. floridae*, Bf*VCBP2* and Bf*VCBP5*, show significant sequence diversity relative to the other three *B. floridae* genes and all of the *VCBP* genes in *C. intestinalis*. The diversity for *BfVCBP2* and *BfVCBP5* is based on hundreds of alleles in the population in which most of the diversity is in the N-terminal Ig V-type domain. However, there are also indels in non-coding regions and inverted repeats that may influence genomic instability in the region of these loci (Dishaw *et al.,* 2010). Instability may drive sequence diversity that result from gene duplications, gene conversion, recombination and unequal crossing-over that would lead to significant diversity in allelic sequences in *B. floridae* populations. The V-type domains are likely the site of bacterial binding because both a full-length VCBP protein and a recombinant protein missing the chitin binding domain show similar binding characteristics (Dishaw *et al*., 2011). None of the four *VCBP* genes in *C. intestinalis* nor three of the *VCBP* genes in *B. floridae* (*BfVCBP1*, *3*, and *4*) show sequence diversity, suggesting that each recognizes stable non-self molecules. VCPB proteins are variably expressed in association with the stomach and intestinal epithelium, are expressed by celomocytes, and appear to mediate immune activities (Dishaw *et al*., 2011; Liberti *et al.,* 2014). Interactions may result in gastrointestinal homeostasis between host and commensal gut microbes in addition to protection from pathogenic invasion. The *VCBP* gene family in the Protochordates provides an example of immune detection genes that are both conserved without sequence diversity and others that show significant diversity.

Expanded gene families in plants and animals

In many cases, genes encoding immune detection and response proteins are simply expanded in the genome of the host organism. This has been evaluated extensively for the variety of disease resistance (*R*) genes in higher plants (Spoel and Dong, 2012), and may in part be an outcome of shared sequences among *R* genes that encode the LRR region of the proteins that may drive recombination and meiotic mispairing (McDowell and Simon, 2008; Joshi and Nayak, 2013; Oren *et al.,* 2016) in addition to ectopic duplications and gene recombination (Meyers *et al*., 2003; Kuang *et al*., 2004; Smith and Hulbert, 2005). Diversification of *R* genes also appears to be induced by contact with pathogens that activates mechanisms for non-homologous recombination among the *R* genes to diversify the sequences with potential to create new recognition capabilities for R proteins in progeny of the infected plant (Durrant *et al*., 2007). In addition to plants, significantly expanded gene copy numbers have also been identified in the genome of the Pacific oyster, *Crassostrea gigas*, particularly for complement homologues encoding C1q, fibrinogen domain containing proteins, and C-type lectin domain-containing proteins (Zhang *et al*., 2015). Gene family expansions for the *TLR* genes have been documented in sea urchins, *B. floridae*, and the annelid, *Capitella capitata* (reviewed in (Buckley and Rast, 2012) (Davidson et al., 2008)) in addition to expansions in a variety of other genes encoding pattern recognition receptors in a wide range or organisms (reviewed in (Buckley and Rast, 2015).

Sequence diversity of single copy genes

In addition to expansions in the number of members of gene families, some immune genes in invertebrates are present in only a few copies or are single copy yet have significant sequence diversity, which is based on large numbers of alleles with slightly different sequences in the population. One example is the set of linked genes in the fusion/histocompatibility (*Fu/HC*) locus in the compound tunicate, *Botryllus schlosseri*. The *Fu/HC* locus is composed of single copy genes that function in allograft recognition and rejection or fusion among individuals, which is an important life history trait for *B. schlosseri*. Natural allograft reactions occur commonly when these sessile invertebrates grow into contact with one another resulting in either fusion of their vasculature or rejection and tissue separation (Oka and Watanabe, 1957; Scofield *et al*., 1982). The responses to non-self are controlled by 1) two, very tightly linked candidate *fuhc* (*cfuhc*) genes that encode secreted and transmembrane Fu/HC proteins, 2) *fester* and *uncle fester* that encode putative receptors in non-self detection, 3) *hsp40-l* that encodes a chaperone protein, and 4) the *Botryllus* histocompatibility factor (*bhf*) that shows lower allelic polymorphism, encodes a cytoplasmic protein that is partially disordered and is associated with the cytoplasmic face of the plasma membrane (Fig. 7) (Voskoboynic

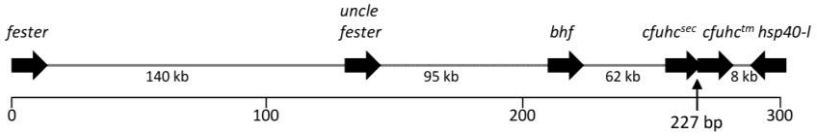

Fig. 7 The *Fu/HC* locus in *Botryllus schlosseri*. The *Fu/HC* locus spans about 305 kb and the orientations and distances between the genes are indicated including the tight linkage between *cfuhc^{sec} and cfuhc^{tm}*. Additional genes are present in this region including nine genes that match to homologues in other organisms with known function and seven with unknown function. This figure is modified from (Taketa and De Tomaso, 2015).

et al., 2013, Taketa *et al*., 2015; reviewed by (Taketa and De Tomaso, 2015)). Diversity of the polymorphic genes in the Fu/HC locus is based on up to hundreds of alleles in the population, of which the most polymorphic are the two c*fuhc* genes and *fester*, plus the *hsp40-l* region that encodes the C-terminus of the protein. Diversity for fester and uncle fester proteins can also be increased through alternative splicing.

The examples of immune genes described above for both vertebrates and invertebrates illustrate multiple ways in which to achieve sequence diversity at immune loci, either through the assembly or copying of gene segments into a non-functional single copy gene rendering it functional, through the expansion of genes into clustered families of similar genes, through the presence of many alleles for single copy genes in a population, and/or through extensive alternative splicing.

The Sp185/333 system in echinoids

Sp185/333 mRNAs

The variety of highly variable gene families has shown that different groups of organisms often use different genes with different diversification mechanisms to attain the same biological outcome of host protection from pathogens (Raftos and Raison, 2008; Ghosh *et al.,* 2011). In the purple sea urchin, *Strongylocentrotus purpuratus*, for which the genome sequence is available (Sodergren *et al*., 2006), the numbers of immune genes and immune gene families with expanded members illustrate a surprising level of complexity for immune defense in this invertebrate (Hibino *et al.,* 2006; Rast *et al.,* 2006). Significant expansions are present in the TLR family (Buckley and Rast, 2012), the NOD and NALP families (Hibino *et al*., 2006; Rast *et al.,* 2006; Buckley and Rast, 2015), the genes encoding scavenger receptors with cysteine rich domain (Pancer, 2000; Pancer, 2001), and small C-type lectin genes (Sodergren *et al*., 2006). Overall, the set of immune genes identified as homologues of those known from vertebrates indicates that the echinoid immune system is quite sophisticated (Hibino *et al.*, 2006; Rast *et al*., 2006).

Another immune response gene family are the *Sp185/333* genes that are strongly upregulated in adult coelomocytes and larval blastocelar cells in response to challenge with marine bacteria and lipopolysaccharide (LPS) (Rast *et al*., 2000; Nair *et al.,* 2005; Ho *et al.,* 2016). The *Sp185/333* sequences identified from a cDNA library screen of immune activated celomocytes using a subtracted probe (Nair *et al.,* 2005) matched to only two sequences in GenBank; EST333 [(accession number; R62081 (Smith *et al*., 1996)] and DD185 [accession number; AF228877 (Rast *et al.,* 2000)]. These two matches are the basis of a combined name for the gene family, *Sp185/333*, because the sequence provides no prediction for putative function. The paradox of the result is that many of the partial cDNA sequences do not match to each other, however, once alignments were undertaken by hand, patterns of recognizable short regions of sequence emerged called *elements* (Fig. 8A) (Terwilliger *et al*., 2006, 2007). Although this appears reminiscent of alternative splicing as in *DSCAM* from arthropods from sets of duplicated exons (Ng *et al*., 2014; Brites and du Pasquier, 2015), the *Sp185/333* genes are small with only two exons, which rules out the possibility of alternative splicing as a mechanism for sequence diversity (Terwilliger *et al*., 2006). Careful alignments of genes amplified from three sea urchins show that the element structure of the cDNAs is entirely encoded by the second exon (Fig. 8A) (Buckley and Smith, 2007). Element patterns result from the mosaic of elements that are variably present and absent in the second exon, which impart significant sequence diversity to the mRNAs and deduced proteins. Although, individual elements are recognizable based on length and sequence, the same element from different genes may have slightly different sequences that adds to the diversity of genes and mRNAs with the same element pattern. Analysis of 121 unique *Sp185/333* genes from three sea urchins shows that none share sequence among animals, although sequences of some element are shared among genes. In addition to elements, the second exon has six types of repeats that are present as two to four imperfect tandem repeats in the 5′ end of the second exon plus interspersed repeats at the 3′ end (Fig. 8A) (Buckley *et al*., 2008a). Computational predictions for the origins of the tandem repeats suggest three ancestral versions that underwent duplications, recombinations and deletions to generate the current structure of the 5′ end of the second exon. These characteristics that impart significant sequence diversity within the gene family plus increased expression in response to immunological challenge strongly suggest that these genes function in the immune response of the purple sea urchin (Smith, 2012).

It is generally assumed that transcription is a high fidelity process producing mRNAs that reflect exactly the sequence of the coding regions in the genes that has likely been optimized by selection. Consequently, the identification of early stop codons and missense sequences in the *Sp185/333* cDNAs were assumed to be the result of pseudogene expression (Terwilliger *et al.,* 2007). However, comparisons among the mRNAs and genes from each of three sea urchins show, very few identical matches (Buckley *et al*., 2008b). The gene sequences show that pseudogenes (based on coding regions) are not present in the genome (Buckley and Smith, 2007) and that the early stop codons and frame shifts are only observed in the mRNAs. Alignments of genes and mRNAs with the same element pattern from the same sea urchin show that most of the differences are in positions of cytidine in the genes that corresponded to uracil in the mRNAs (Buckley *et al*., 2008b). Speculations on the basis for these changes include mRNA editing by a cytidine deaminase for the C to U changes, and perhaps polymerase (pol)μ for other nucleotide changes. Both cytidine deaminases and polμ genes are present in the sea urchin genome. It is noteworthy that the most common cDNA sequence element pattern is *E2*, which is edited at the same nucleotide in many cDNA sequences to generate a truncated protein (Fig. 8A) (Buckley *et al*., 2008b).

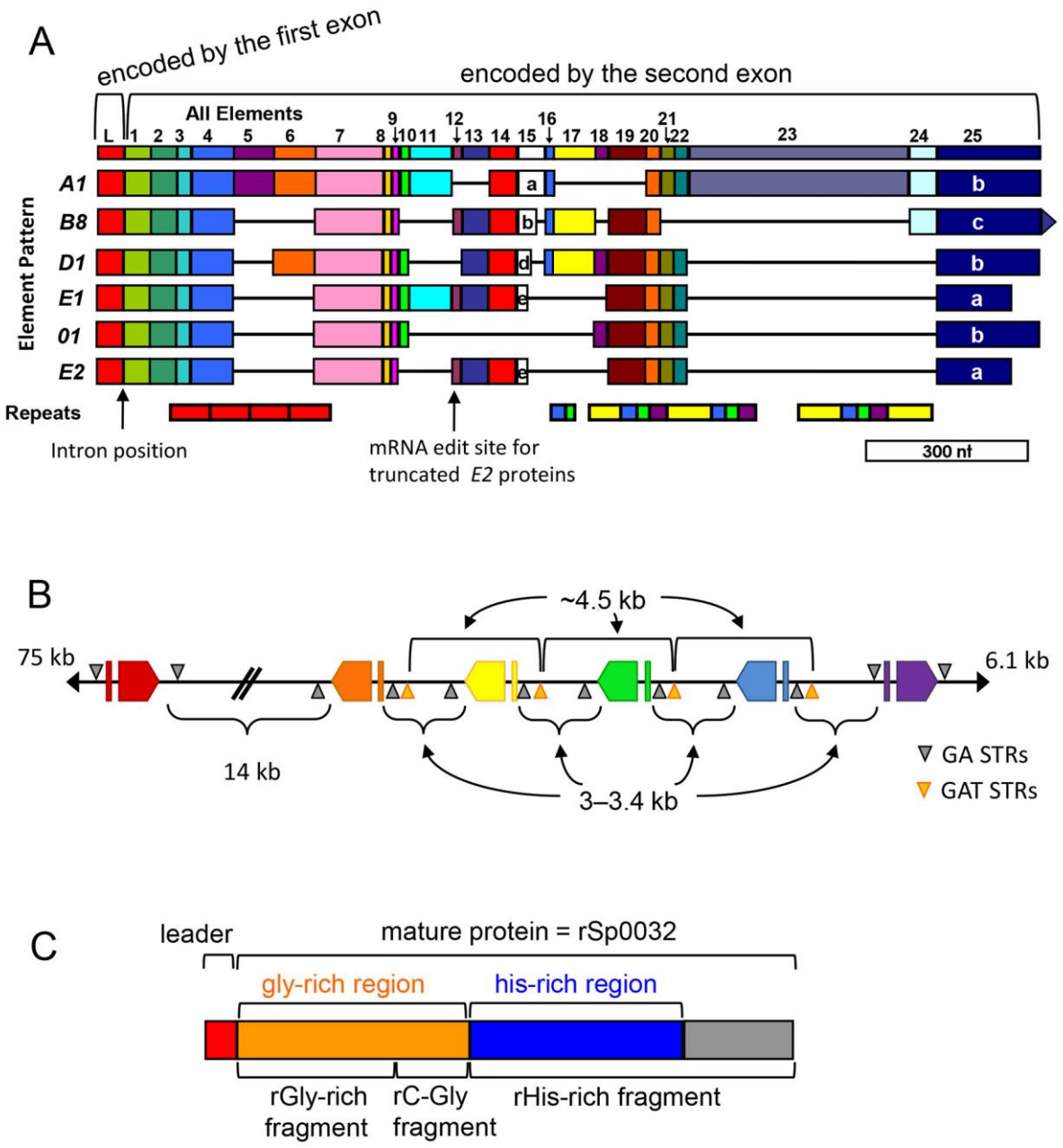

Fig. 8 Diversity of the Sp185/333 system in sea urchins. **A**) The *Sp85/333* mRNAs are encoded by the two exons in the genes that code for the leader (L) and the mature protein. The mRNAs are composed of a mosaic of blocks of sequence called elements that are shown as colored rectangles in this cartoon alignment. The combination of elements defines the element pattern that is named according to the distinctive and highly diverse sequence of element 15 (Terwilliger *et al.*, 2006). Element 25 has three versions that are defined by the position of the stop codon. A very common edit site for mRNAs is in element 12 of the *E2* element pattern (indicated) and changes a codon to a stop resulting in truncated proteins that are missing the his-rich region. Repeats are shown at the bottom and correspond to the locations of imperfect tandem repeats towards the 5' end of the second exon and the imperfect interspersed repeats in the 3' half of the second exon. This figure is modified from (Buckley and Smith, 2007). **B**) The gene family locus structure has tightly clustered genes within 34 kb that is positioned 6.1 kb from the end of the bacterial artificial chromosome (BAC) clone insert. Distances between the genes are indicated below. Each gene is surrounded by GA short tandem repeats (STRs) and segmental duplications that include three *D1* genes (yellow, green, blue) of nearly identical sequence are surrounded by GAT STRs. The size of the segmental duplications are indicated by brackets above. This figure is modified from (Miller *et al.,* 2010). **C**) The standard Sp185/333 protein structure has an N-terminal leader (red), a gly-rich region (orange), a his-rich region (blue) and a C-terminal region (gray). The recombinant fragments employed for binding analyses are indicated below. This figure is modified from (Smith and Lun, 2016).

Not only does this appear to be non-random, but edited *E2* messages encoding truncated proteins are more commonly present prior to immune challenge, whereas after challenge, the preponderance of mRNAs encoding full-length proteins increases (Terwilliger *et al*., 2007; Sherman *et al*., 2015). Given the possibility of mRNA editing, the set of mRNAs expressed from a single gene may encode a set of similar but non-identical Sp185/333 proteins. In general, the diversity of the *Sp185/333* gene family generates diverse mRNAs that increase diversity by mRNA editing to produce a broad array of Sp185/333 proteins.

The Sp185/333 *gene family*

The *Sp185/333* gene family structure shows tightly clustered genes that are spaced apart by 3 to 12 kb (Miller *et al*., 2010). One cluster of six genes is present within 34 kb, with the peripheral genes oriented opposite relative to the internal genes (Fig. 8B). All genes are surrounded by GA short tandem repeats (STRs), or microsatellites, and the regions between the STRs, which includes the genes and flanking regions of the genes, show increased sequence conservation compared to regions that are distal to the GA STRs (Miller *et al*., 2010). This has been interpreted as a result of gene duplication that is mediated by the GA STRs. Sequence conservation among regions bounded by the GA STRs is also consistent with gene conversion in which sequence exchange is initiated by short regions of shared sequences within the coding regions of the genes and is terminated at the GA STRs. This type of sequence exchange may be similar to observations reported in yeast (Gendrel *et al*., 2000). The beneficial outcome of limited gene conversion would be sequence diversification of the genes while blocking sequence homogenization of entire gene clusters (Miller *et al*., 2010). In addition to GA STRs surrounding the genes, GAT STRs are positioned at the edges of three segmental duplications that are located in tandem, are of identical size, nearly identical sequence, and include duplicated genes of the *D1* element pattern with almost identical sequence (Fig. 8B). The level of sequence identity suggests very recent GAT STR-mediated duplications within the *Sp185/333* gene cluster, which correlates with *D1* genes being the most common type based on random gene sequencing from three sea urchins (Buckley and Smith, 2007). Overall, the basis for *Sp185/333* gene sequence diversity is likely based on 1) shared blocks of sequences among genes, 2) the presence of STRs within the *Sp185/333* gene cluster, and 3) tight clustering of the genes, all of which may drive gene conversion, duplication, deletion and recombination (Miller *et al*., 2010; Oren *et al*., 2016). This complex family structure with a variety of repeats is very likely the basis for the poor assembly of the *Sp185/333* gene family in the sea urchin genome sequence in which only six genes are present in a single cluster when 50 ± 10 genes per genome have been estimated (Smith, 2012).

The Sp185/333 proteins

Sequence diversity is a characteristic of the *Sp185/333* gene family, which appears to be increased by mRNA editing, and infers great Sp185/333 protein diversity. The edited mRNAs encoding missense sequence has been suggested as another means for increasing protein diversity (Smith, 2012) as missense amino acids in Sp185/333 proteins are present in the sea urchin coelomic fluid (CF) as confirmed by mass spectrometry (Dheilly *et al.,* 2009). The Sp185/333 proteins from the CF appear much larger than expected relative to sizes predicted from the cDNA sequences suggesting the possibility of multimerization (Brockton *et al.,* 2008). This outcome is also observed for a recombinant Sp185/333 protein that appears as monomers, dimers and multimers. Sp185/333 proteins from all sea urchins show multimers, and the repertoires among individual sea urchins are quite different (Brockton *et al.,* 2008; Dheilly *et al*., 2009; Sherman *et al.,* 2015). When Sp185/333 proteins are evaluated by two dimensional (2D) Western blots, significant diversity is noted, and most of the proteins migrate to the acidic region during isoelectric focusing (Dheilly *et al.,* 2009) (Fig. 9A). However, the presence of many histidines in the C-terminal region of the deduced protein sequences (see Fig. 8C) (Terwilliger *et al*., 2006) allows isolation of full-length Sp185/333 proteins from the CF of sea urchins by nickel affinity (Sherman *et al*., 2015; Lun *et al.,* 2016). Consistent with the presence of multiple histidines, most of these nickel-isolated proteins are present in the basic region of a 2D Western blot (Sherman *et al.,* 2015) (Fig. 9B). When arrays of Sp185/333 proteins are compared among sea urchins both before and after challenges with a variety of microbes, there are no similarities among the arrays either among animals in response to the same microbe or in an individual animal responding to the same microbe over multiple challenges (Sherman *et al*., 2015). Clearly, the diversity of these proteins is very broad within sea urchins and the diversity within the population of sea urchins must be astounding.

The characteristics of the Sp185/333 immune response system in sea urchins are based on a number of attributes (Smith, 2012). 1) The differences among the genes in different individual sea urchins (Buckley and Smith, 2007), the structure of the gene family with shared sequences among genes, their tight linkage, and the association with STRs, is likely to promote swift sequence diversity among the genes (Miller *et al*., 2010; Oren *et al*., 2016). 2) mRNA editing (Buckley *et al.,* 2008b) increases the sequence diversity of the proteins beyond the diversity encoded in the genes and generates truncated and missense proteins that appear in the CF (Dheilly *et al*., 2009; Smith, 2012). These truncated proteins may be functional given their propensity to be produced prior to immune challenge (Sherman *et al*., 2015). 3) The arrays of Sp185/333 proteins are very different among individual sea urchins, and show much greater diversity than expected (Dheilly *et al.,* 2009; Sherman *et al*., 2015). The great variety of proteins may be due in part to post translational modifications that have been predicted from conserved sequence motifs for glycosylation (Terwilliger *et al.,* 2006; Sherman *et al*., 2015). In

general, these results suggest that the diversity of the Sp185/333 proteins may be highly protective for sea urchins that are in constant contact with microbes in the marine ecosystem.

The Sp185/333 proteins are expressed by the phagocyte class of coelomocytes

The Sp185/333 proteins are expressed by a subset of the phagocyte class of coelomocytes including the polygonal, discoidal and small phagocytes (Brockton *et al.,* 2008; Majeske *et al.,* 2014). Subsets of each of these cell types have Sp185/333 proteins in small cytoplasmic vesicles that are typically positioned near the nucleus. In addition, small phagocytes have Sp185/333 proteins associated with the surface of the cell. Expression of the Sp185/333 proteins in the adult tissues before vs. after challenge with LPS shows increases in coelomocytes, axial organ, esophagus, pharynx and intestine (Majeske *et al.,* 2013). Sp185/333-positive cells are present in all tissues and significant increases in these cells are noted for coelomocytes and for the axial organ after challenge with LPS (Brockton *et al.,* 2008; Majeske *et al.,* 2013). In larvae, which are free swimming, feeding members of the zooplankton, *Sp185/333* gene expression is limited to a subset of differentiated blastocoelar cells called phagocytic filopodial cells (Ho *et al*., 2016) that have similar phagocytic functions as the discoidal and polygonal phagocytes in the adult.

One of the tenets of an effective innate immune response is a quick response to the detection of pathogens before they have time to undergo significant proliferation that may overwhelm host immunity. This can be accomplished by swift up-regulation of the host immune response genes and the production of the encoded effector proteins. This has been documented for the *Sp185/333* genes in both adult and larval immune cells responding to injected bacteria or microbes in the seawater (Silva, 2000; Nair *et al*., 2005; Terwilliger *et al*., 2007; Furukawa *et al.,* 2009; Ho *et al*., 2016). It follows that the multiple *Sp185/333* genes would be expressed simultaneously in cells to optimize the numbers of effector proteins that can be produced and secreted over time. However, evaluation of *Sp185/333* gene expression in single adult phagocytes shows mRNAs of a single sequence per cell, inferring expression from a single gene (Majeske *et al*., 2014). This result suggests a complex mechanism for selecting the gene that is expressed and for silencing the others that is quite unexpected for an invertebrate innate immune response. It is expected that mRNA editing will result in a small set of proteins of the same element pattern but with slightly different sequence will be produced by single phagocytes. However, the wide range of Sp185/333 protein diversity observed for individual sea urchins (see below) will require that large numbers of phagocytes express the *Sp185/333* genes.

A recombinant Sp185/333 protein has multiple anti-pathogen activities

Hundreds of different isoforms of the Sp185/333 proteins are expressed by individual sea urchins in response to immune challenge, and they

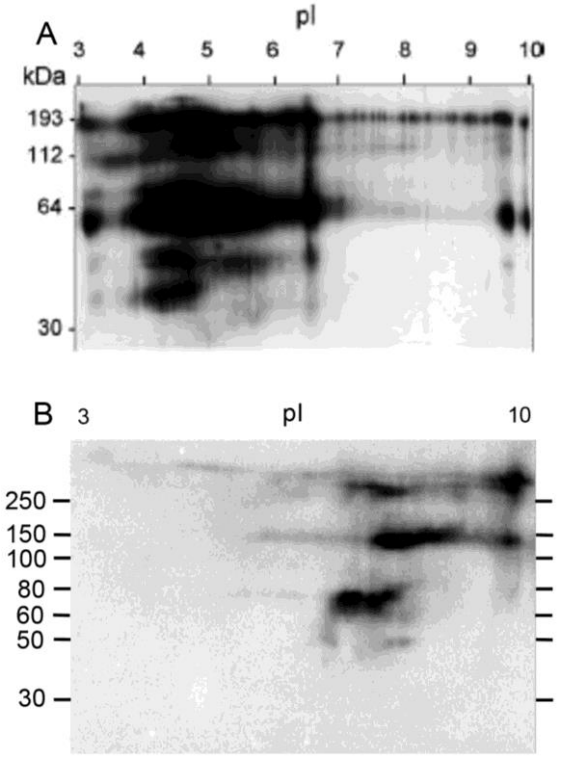

Fig. 9 The diversity of the native Sp185/333 proteins are characterized by 2D Western blots. A) Native Sp185/333 proteins from the CF of *S. purpuratus* show the diversity, but also the preponderance of proteins in the acidic pI range. The figure is from (Dheilly *et al*., 2009) and is reprinted from the Journal of Immunology with permission from the American Association of Immunologists, Inc., Copyright 2009. B) When full-length, native Sp185/333 proteins with sufficient histidines are isolated by nickel affinity, these proteins are generally present in the basic pI range. This figure is republished from (Sherman *et al.,* 2015). In both cases, the sizes of most of the proteins are larger than expected compared to sizes deduced from cDNAs and genes.

may function synergistically in response to the challenge. Understanding the activities of these proteins is essential for understanding the sea urchin immune system, however this requires that individual Sp185/333 isoforms be isolated away from other versions and from other proteins in the CF. Consequently, a recombinant Sp185/333 protein called rSp0032 with an *E1* element pattern (see Fig. 8A) was evaluated for predicted anti-microbial activities (Lun *et al.,* 2016). When rSp0032 is incubated with *Vibrio diazotrophicus*, *Bacillus subtilis*, *B. cereus*, or *Saccharomyces cerevisiae*, the protein shows saturation binding to *Vibrio* and *Saccharomyces*, but no binding is observed for the *Bacillus* species (Fig. 10). Affinity for *Vibrio* and *Saccharomyces* is high, binding sites are specific, and once bound rSp0032 cannot be dissociated from the target cells. Mass spectrometry of protein

bands from gels used to analyze rSp0032 incubation with *Vibrio* target cells shows an association between rSp0032 and flagellin indicating that binding may be mediated through pathogen associated molecular patterns (PAMPs). Expanded PAMP binding analysis shows that rSp0032 also binds tightly with flagellin from *Salmonella*, LPS from *E. coli*, and β,1-3,glucan from *Saccharomyces*, but does not bind to peptidoglycan (PGN) from *Bacillus*. This suggests that there are multiple types of foreign cells to which rSp0032 can bind, that it does not bind indiscriminately to all foreign cells, and that it uses several PAMPs as specific binding targets that have very different molecular characteristics.

Despite the diversity that has been documented for the full-length Sp185/333 proteins, they have a standard structure with an N-terminal glycine rich (gly-rich) region, a central region that is the C-terminal end of the gly-rich (C-gly) region that has an arginine-glycine-aspartic acid motif (which has not been tested for integrin binding), a more C-terminal histidine rich (his-rich) region, and a C-terminal region (Fig. 8C). Based on the significant differences in the amino acid content of the gly-rich and his-rich regions, in addition to observations that edited mRNAs (Buckley et al., 2008b) encoding truncated proteins missing the his-rich region are more prevalent prior to immune challenge (Terwilliger *et al.*, 2007; Sherman *et al.*, 2015), these two regions of the proteins are expected to have different activities or functions. Consequently, when tested, the recombinant fragments of the full-length rSp0032 (rGly-rich, rC-Gly, rHis-rich fragments; Fig. 8C) show differences in activities compared to rSp0032 (Lun *et al.,* 2016). All recombinants bind to all foreign target cells including *Bacillus* species to which rSp0032 does not bind indicating a broadening of binding characteristics (Fig. 9). Furthermore, the rC-Gly fragment multimerizes in the presence or absence of binding targets in agreement with the rGly-rich and rHis-rich fragments, which do not include the C-gly region and do not multimerize. The binding specificity of the rHis-rich fragment for yeast is identical to that of the full-length protein, however, the rGly-rich fragment shows expanded binding compared to either the full-length protein or the rHis-rich fragment. These results indicate several noteworthy aspects of the binding activities that include 1) rSp0032 binds specifically and with high affinity to quite different foreign cell targets, 2) rSp0032 binds to multiple PAMPs that have very different characteristics, 3) the central C-gly region mediates multimerization, 4) the gly-rich and his-rich regions of the protein have different activities, and 5) when all regions of the full-length protein function together, they show increased specificity for binding targets. The binding characteristics of the rGly-rich fragment predict broadened binding activities of truncated Sp185/333 proteins that are not characteristic of either the his-rich region or the full-length protein. This suggests interesting possibilities for truncated proteins, which are more likely to be present prior to challenge, and may be important for immune surveillance.

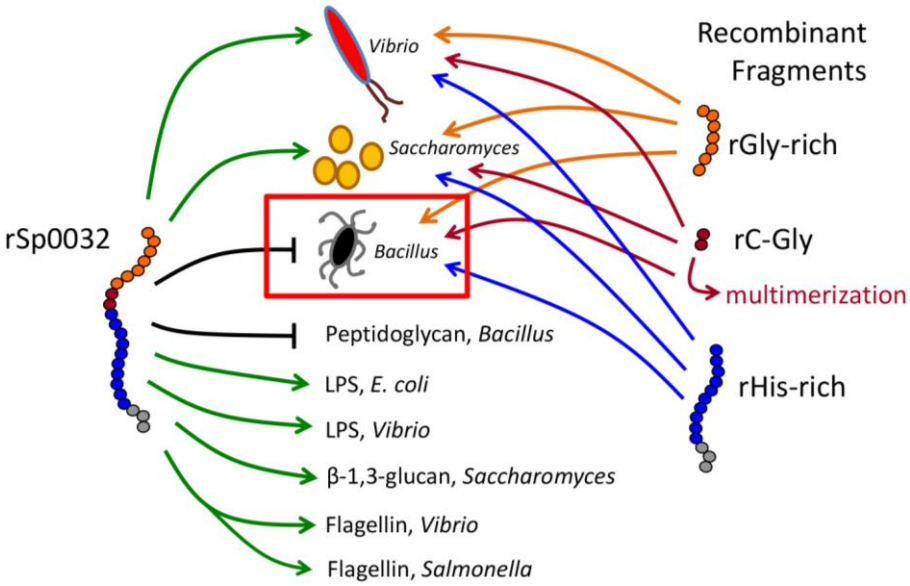

Fig. 10 Multitasking binding characteristics of rSp0032 and the recombinant fragments to microbes and PAMPs. rSp0032 (to the left) is composed of a gly-rich region (orange) including the C-terminal end of the gly-rich region (red), a his-rich region (blue), and a C-terminal region (gray). The recombinant fragments that correspond to the regions of the full-length protein are shown to the right in corresponding colors. The binding targets against which rSp0032 and the recombinant fragments were tested are shown in the middle. Colored arrows indicate binding, whereas black blocked lines indicate no binding. The major differences in binding is to *Bacillus* and to peptidoglycan (PGN) from *Bacillus*, which are highlighted with the red box. The recombinant fragments were not tested against the PAMPs. The rC-Gly fragment also mediates dimerization and multimerization of the full-length protein. The figure was modified from (Lun *et al.*, 2016; Smith and Lun, 2016).

When the multitasking activities of rSp0032 are applied to the hundreds of native Sp185/333 proteins that are expressed by coelomocytes, this has vast implications for the sea urchin immune response and its activities in host protection. If all or most of the Sp185/333 proteins have multitasking binding activities with slightly different but perhaps overlapping targets to which they bind, and many or all of the isoforms function synergistically, the outcome is a highly diverse and flexible innate immune effector response directed at controlling attack and infection from foreign microbial cells. If the expressed arrays of Sp185/333 proteins have broad binding capabilities, it is highly unlikely that microbes will be able to circumvent or defeat the Sp185/333 system because it would likely require simultaneous changes to multiple characteristics of the microbes and their PAMPs. The Sp185/333 system, including the sequence diversity noted in the members of the gene family, the (putative) directed editing of mRNAs that are translated to truncated and missense proteins, and the extraordinarily diverse arrays of proteins that are expressed among individual sea urchins strongly suggests that this system represents the central immune diversification system for echinoids.

Conclusion

The common theme reviewed here describes host immune responses to microbial pathogens and the abilities to drive diversity at one or more levels to produce broad sequence variations among immune proteins that function in detection of foreign cells and effector proteins that act to remove and/or kill the invaders. There are commonalities among animals for detecting foreign molecules including the TLR and NOD proteins among others, however, the wide variety of effector proteins and their modes of action make different immune systems appear to function quite differently. The underlying commonality is that although mechanisms to generate immune diversity are vastly different, ranging from RAG-induced somatic recombination and AID-mediated gene assembly in the vertebrate Ig and VLR systems respectively (Tonegawa, 1983; Davis *et al.,* 1984; Nagawa *et al.,* 2007; Rogozin *et al*., 2007), to extensive alternative splicing of *DSCAM* genes in arthropods (Schmucker and Chen, 2009; Ng *et al*., 2014; Brites and du Pasquier, 2015), to clusters of duplicated genes with shared sequences that drive gene recombination, deletion, duplication (Oren *et al.,* 2016), the outcome for effective immunity and survival of organisms is to stay ahead of the pathogens. The example of the *Sp185/333* system in echinoids illustrates that multiple levels of diversification may function simultaneously to generate diverse anti-pathogen protein isoforms, perhaps each with slightly different activities towards microbial invasions, that together act as effectively in host protection as the somatic rearrangement and gene assembly mechanisms that are employed in the immune systems of vertebrates.

Acknowledgements

This work was supported by funding from the Italian National Program for Antarctic Research (PNRA; PEA2009/A1.12) to MRC, and the US National Science Foundation (IOS-1550474) to LCS.

References

Adema CM. Fibrinogen-related proteins (FREPs) in mollusks. Results and problems in cell differentiation, 57: 111-129, 2015.

Adema, CM, Hertel LA, Miller RD, Loker ES. A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. Proc. Natl. Acad. Sci. USA 94: 8691-8696, 1997.

Agrawal A, Eastman QM, Schatz DG. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature 394: 744-751, 1998.

Armitage SAO, Peub R, Kurtz J. Dscam and pancrustacean immune memory - a review of the evidence. Dev. Comp. Immunol. 48: 315-323, 2015.

Ashida H, Mimuro H, Sashida C. *Shigella* manipulates host immune responses by delivering effector proteins with specific roles. Front. Immunol. 6: 219, 2015.

Bajoghli B, Guo P, Aghaallaei N, Hirano M, Strohmeier C, McCurley N, *et al.* A thymus candidate in lampreys. Nature 470: 90-94, 2011.

Barrangou R, Marraffini LA. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Mol. Cell, 54: 234-244, 2014.

Bernstein RM, Schluter SF, Bernstein H, Marchalonis JJ. Primordial emergence of the recombination activating gene 1 (RAG1): sequence of the complete shark gene indicates homology to microbial integrases. Proc. Natl. Acad. Sci, 93: 9454-9459, 1996.

Boehm T, McCurley N, Sutoh Y, Schorpp M, Kasahara M, Cooper MD. VLR-based adaptive immunity. Ann. Rev. Immunol. 30: 203-220, 2012.

Brites D, Du Pasquier L. Somatic and germline diversification of a putative immunoreceptor within one phylum: Dscam in arthropods. Results and problems in cell differentiation 57: 131-158, 2015.

Brockton V, Henson JH, Raftos DA, Majeske AJ, Kim YO, Smith LC. Localization and diversity of 185/333 proteins from the purple sea urchin - unexpected protein-size range and protein expression in a new coelomocyte type. J. Cell Sci. 121: 339-348, 2008.

Buckley KM, Munshaw S, Kepler TB, Smith LC. The 185/333 gene family is a rapidly diversifying host-defense gene cluster in the purple sea urchin, *Strongylocentrotus purpuratus*. J. Mol. Biol. 379: 912-928, 2008a.

Buckley KM, Rast JP. Dynamic evolution of Toll-like recetor multigene familes in echinoderms. Front. Immunol. 3: 136, 2012.

Buckley KM, Rast JP. Diversity of animal immune receptors and the origins of recognition commplexity in the deuterstomes. Dev. Comp. Immunol. 49: 179-189, 2015.

Buckley KM, Smith LC. Extraordinary diversity among members of the large gene family, *185/333*, from the purple sea urchin, *Strongylocentrotus purpuratus*. BMC Mol. Biol. 8: 68, 2007.

Buckley KM, Terwilliger D, Smith LC. Sequence variations in *185/333* messages from the purple sea urchin suggest posttranscriptional modifications to increase immune diversity. J. Immunol. 181: 8585-8594, 2008b.

Cabazon E, Ripoll-Rozada J, Pena A, de la Cruz F, Arechaga I. Towards an integrated model of bacterial conjugation. FEMS Microbiol. Rev. 39: 81-95, 2015.

Cannon JP, Haire RN, Litman GW. Identification of diversified genes that contain immunoglobulin-like variable regions in a protochordate. Nat. Immunol. 3: 1200-1207, 2002.

Cannon JP, Haire RN, Schnitker N, Mueller MG, Litman GW. Individual protochordates have unique immune-type receptor repertoires. Curr. Biol. 14: R465-466, 2004.

Carmona LM, Fugmann SD, Schatz DG. Collaboration of RAG2 with RAG1-like proteins during the evolution of V(D)J recombination. Genes Development 30: 909-917, 2016.

Carroll L. Through the looking-glass and what Alice found there, London, Macmillan and Co. 1871.

Coscia MR, Borriello A, Oreste U. Plasma and bile antibodies of the teleost *Trematomus* bernacchii specific for the nematode *Pseudoterranova decipiens*. Dis. Aquat. Organisms 41: 37-42, 2000.

Coscia MR, Giacomelli S, Oreste U. Allelic polymorphism of immunoglobulin heavy chain genes in the Antarctic teleost *Trematomus bernacchii*. Mar. Genomics 8: 43-48, 2012.

Coscia MR, Oreste U. Presence of antibodies specific for proteins of *Contracaecum osculatum* (Rudolphi, 1908) in plasma of several Antarctic teleosts. Fish Shellfish Immunol. 8: 295-302, 1998.

Davidson CR, Best NM, Francis JW, Cooper EL, Wood TC. Toll-like receptors genes (TLRs) from *Capitella capitata* and *Helobdella robusta* (Annelida). Dev. Comp. Immunol. 32: 608-612, 2008.

Davis MM, Chien YH, Gascoigne NR, Hedrick SM. A murine T cell receptor gene complex: isolation, structure and rearrangement. Immunol. Rev. 81: 235-258, 1984.

Dawkins R, Krebs JR. Arms races between and within species. Proc. R Soc. London Series B-Biol. Sci. 1161: 489-511, 1979.

de la Casa-Esperon E. Horizontal transfer and the evolution of host-pathogen interactions. Int. J. Evol. Biol. 2012. 2012;2012:679045.

Deitsch KM, Lukehart SA, Stringer JR. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. Nat. Rev. Microbiol. 493-503, 2009.

Dheilly NM, Nair SV, Smith LC, Raftos DA. Highly variable immune-response proteins (185/333) from the sea urchin *Strongylocentrotus purpuratus*: proteomic analysis identifies diversity within and between individuals. J. Immunol. 182: 2203-2212, 2009.

Dishaw L, Ota T, Mueller M, Cannon J, Haire R, Gwatney N, *et al.* The basis for haplotype complexity in VCBPs, an immune-type receptor in amphioxus. Immunogenetics 62: 623-631, 2010.

Dishaw LJ, Giacomelli S, Melillo D, Zucchetti I, Haire RN, Natale L, *et al.* A role for variable region-containing chitin-binding proteins (VCBPs) in host gut-bacteria interactions. Proc. Natl. Acad. Sci. USA 108: 16747-16752, 2011.

Dong Y, Dimopoulos G. Anopheles fibrinogen-related proteins provide expanded pattern recognition capacity against bacteria and malaria parasites. J. Biol. Chem. 284: 9835-9844, 2009.

Dong Y, Taylor HE, Dimopoulos G. AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. PLOS Biol. 4: e229, 2006.

Durrant WE, Wang S, Dong X. Arabidopsis SNI1 and RAD51D regulate both gene transcription and DNA recombination during the defense response. Proc. Natl. Acad. Sci. USA 104: 4223-4227, 2007.

Finlay BB, McFadden G. Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. Cell 124: 767-782, 2006.

Flajnik MF. Immunology: another manifestation of GOD. Nature 430: 157-158, 2004.

Flajnik MF. Re-evaluation of the immunological big bang. Curr. Biol. 24: R1060-R1065, 2014.

Fugmann SD. The origins of the RAG genes - from transposition to V(D)J recombination. Semin. Immunol. 22: 10-16, 2010.

Fugmann SD, Messier C, Novack LA, Cameron RA, Rast JP. An ancient evolutionary origin of the Rag1/2 gene locus. Proc. Natl. Acad. Sci. USA 103: 3728-3733, 2006.

Furukawa R, Takahashi Y, Kanajima Y, Dan-Sohkawa M, Kaneko H. Defense system by mesenchyme cells in bipinnaria larvae of the starfish, *Asterina pectinifera*. Dev. Comp. Immunol. 33: 205-215, 2009.

Gendrell C-G, Boulet A, Dutreix M. (CA/GT)n microsatellites affect homologous recombination during yeast meiosis. Genes Development 14: 1261-1268, 2000.

Ghosh J, Lun CM, Majeske AJ, Sacchi S, Schrankel CS, Smith LC. Invertebrate immune diversity. Dev. Comp. Immunol. 35: 959-974, 2011.

Gordy MA, Pila EA, Nanington PC. The role of fibrinogen-related proteins in the gastropod immune response. Fish Shellfish Immunol. 46: 39-49, 2015.

Guo P, Masayuki H, Herrin BR, Li J, Yu C, Sadlonova A, *et al*. Dual nature of the adaptive immune system in lampreys. Nature 459: 796-802, 2009.

Haldane JBS. Disease and evolution. La Ricerca Scientifica 19 (suppl.): 68-75, 1949.

Hamelaar F. The origin and diversity of the HIV-1

pandemic. Trend. Mol. Med. 18: 182-192, 2012.
Hanington PC, Forys MA, Dragoo JW, Zhang ZM, Adema CM, Loker ES. Role for a somatically diversified lectin in resistance of an invertebrate to parasitise infection. Proc. Natl. Acad. Sci. USA 107: 21087-21092, 2010.
Hanington PC, Zhang SM. The primary role of fibrinogen-related proteins in invertebrates is defense, not coagulation. J. Innate Immun. 3: 17-27, 2011.
Hemani Y, Soller M. Mechanisms of *Drosophila* Dscam mutually exclusive splicing regulation. Biochem. Soc. Transactions 40: 804-809, 2012.
Hencken CG, Li X, Craig NL. Functional characterization of an active Rag-like transposase. Nat. Struct. Mol. Biol. 19: 834-836, 2012.
Hibino T, Loza-Coll M, Messier C, Majeske AJ, Cohen A, Terwilliger DP, *et al.* The immune gene repertoire encoded in the purple sea urchin genome. Dev. Biol. 300: 349-365, 2006.
Hiom K, Melek M, Gellert M. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. Cell 94: 463-470, 1998.
Ho ECH, Buckley KM, Schrankel CS, Schuh NW, Hibino T, Solek CM, *et al.* Perturbation of gut bacteria induces a coordinated cellular immune response in the purple sea urchin larva. Immunol. Cell Biol. 94: 861-874, 2016.
Horn D. Antigenic variation in African trypanosomes. Mol. Biochem. Parasitol. 195: 123-129, 2014.
Huang S, Tao X, Yuan S, Zhang Y, Li P, Beilinson HA, *et al.* Discovery of an active RAG transposon illuminates the origins of V(D)J recombination. Cell 166: 102-114, 2016.
Hung GY, Ng TJ, Lin JH, Chiang YA, Chuang Y, Wang HC. Properties of *Litopenaeus vannamei* Dscam (LvDscam) isoforms related to specific pathogen recognition. Fish Shellfish Immunol. 35: 1272-1281, 2013.
Jiang F, Doudna JA. The structural biology of CRISPR-Cas systems. Curr. Opin. Genet. Dev. 30: 100-111, 2015.
Joshi RK, Nayak S. Perspectives of genomic diversification and molecular recombination towards *R*-gene evolution in plants. Physiol. Mol. Biol. Plants 19: 1-9, 2013.
Kasahara M. Variable lymphocyte receptors: a current overview. Results and Problems in Cell Differentiation 57: 175-192, 2015.
Kreahling JM, Graveley BR. The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the *Drosophila* Dscam pre-mRNA. Mol. Cell. Biol. 25: 10251-10260, 2005.
Kwon YD, Pancera M, Acharya P, Georgiev IS, Crooks ET, Gormann J, *et al*. Crystal structure, conformational fixation, and entry-related interactions of mature ligand-free HIV-1 Env. Nat. Struct. Mol. Biol. 22: 522-531, 2015.
Lauring AS, Frydman J, Andino R. The role of mutational robustness in RNA virus evolution. Nat. Rev. Microbiol. 11: 327-336, 2013.
Liberti A, Leigh B, De Santis R, Pinto MR, Cannon JP, Dishaw LJ, *et al*. An immune effector system in the protochordate gut sheds light on fundamental aspects of vertebrate immunity. Results and Problems in Cell Differentiation 57: 159-173, 2015.
Liberti A, Mellillo D, Zucchetti I, Natale L, Dishaw LJ, Litman GW, *et al.* Expression of *Ciona intestinalis* variable region-containing chitin-binding proteins during development of the gastrointestinal tract and their role in host-microbe interactions. PLOS ONE, 9, e94984. 2014.
Loker ES, Adema CM, Zhang S-M, Kepler TB. Invertebrate immune systems - not homogeneous, not simple, not well understood. Immunol. Rev. 198: 10-24, 2004.
Lun CM, Schrankel CS, Chou H-Y, Sacchi S, Smith LC. A recombinant Sp185/333 protein from the purple sea urchin has multitasking binding activities towards certain microbes and PAMPs. Immunobiology 221: 889-903, 2016.
Macho AP, Zippel C. Targeting of plant pattern recognition receptor-triggered immunity by bacterial type-III secretion system effectors. Curr. Opin. Microbiol. 23: 14-22, 2015.
Majeske AJ, Oleksyk T, Smith LC. The Sp185/333 immune response genes and proteins are expressed in cells dispersed within all major organs of the adult purple sea urchin. Innate Immun. 19: 569-587, 2013.
Majeske AJ, Oren M, Sacchi S, Smith LC. Single sea urchin phagocytes express messages of a single sequence from the diverse *Sp185/333* gene family in response to bacterial challenge. J. Immunol. 193: 5678-5688, 2014.
Marraffini LA. CRISPR-Cas immunity in prokaryotes. Nature 526: 55-61, 2015.
McDowell JM, Simon SA. Molecular diversity at the plant-pathogen interface. Dev. Comp. Immunol. 32: 736-744, 2008.
Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell 15: 809-834, 2003.
Miller CA, Buckley KM, Easley RL, Smith LC. An *Sp185/333* gene cluster from the purple sea urchin and putative microsatellite-mediated gene diversification. BMC Genomics 11: 575, 2010.
Muramatsu M, Kinoshita N, Fagaransan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 102: 553-563, 2000.
Nagawa F, Kishishita N, Shimuzu K, Hirose S, Miyoshi M, Nezu J, *et al.* Antigen-receptor genes of the agnathan lamprey are assembled by a process involving copy choice. Nat. Immunol. 8: 206-213, 2007.
Nair SV, Del Valle H, Gross PS, Terwilliger DP, Smith LC. Macroarray analysis of coelomocyte gene expression in response to LPS in the sea urchin. Identification of unexpected immune diversity in an invertebrate. Physiol. Genomics 22: 33-47, 2005.
Ng TH, Chiang TA, Yeh TC, Wang HG. Review of

DSCAM-mediated immunity in shrimp and other arthropods. Dev. Comp. Immunol. 46: 129-138, 2014.

Oka H, Watanabe E. Colony specificity in compound ascidians as tested by fusion experiments. Proc. Japanese Acad. Sci. 33: 657-664, 1957.

Oren M, Barella Hudgell MA, Golconda P, Lun CM, Smith LC. Genomic instability and shared mechanisms for gene diversification in two distant immune gene families: the echinoid 185/333 and the plant NBS-LRR. In: Malagoli D (ed.), The evolution of the immune system: Conservation and diversification, Elsevier-Academic Press, London, UK, pp 295-310, 2016.

Pancer Z. Dynamic expression of multiple scavenger receptor cysteine-rich genes in coelomocytes of the purple sea urchin. Proc. Natl. Acad. Sci. 97: 13156-13161, 2000.

Pancer Z. Individual-specific repertoires of immune cells SRCR receptors in the purple sea urchin (*S. purpuratus*). Adv. Exp. Med. Biol. 484: 31-40, 2001.

Pancer Z, Saha NR, Kasamatsu J, Suzuki T, Amemiya CT, Kasahara M, *et al*. Variable lymphocyte receptors in hagfish. Proc. Natl. Acad. Sci. USA 102: 9224-9229, 2005.

Park JW, Parinsky K, Celotto AM, Reenan, RA, Braveley BR. Identificatin of alternative splicing regulators by RNA interference in *Drosophila*. Proc. Natl. Acad. Sci. USA 101: 15974-15979, 2004.

Patterson S, Vogwill T, Buckling A, Benmyor R, Spiers AS, Thomson NR, *et al.* Antagonistic coevolution accelerates molecular evolution. Nature 464: 275-278, 2010.

Penades JR, Chen JC, Quiles-Puchalt N, Carpena N, Novick RP. Bacteriophage-mediated spread of bacterial virulence genes. Curr. Opin. Microbiol. 23: 171-178, 2015.

Raftos DA, Raison RL. Early vertebrates reveal diverse immune recognition strategies. Immunol. Cell Biol. 86: 479-481, 2008.

Rast JP, Pancer Z, Davidson EH. New approaches towards an understanding of deuterostome immunity. Curr. Top. Microbiol. Immunol. 248: 3-16, 2000.

Rast JP, Smith LC, Loza-Coll M, Hibino T, Litman GW. Genomic insights into the immune system of the sea urchin. Science 314: 952-956, 2006.

Rogozin IB, Iyer LM, Liang L, Glazko GV, Liston VG, Pavlov YI, *et al.* Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. Nat. Immunol. 8: 647-656, 2007.

Schluter SF, Bernstein RM, Bernstein H, Marchalonis JJ. 'Big Bang' emergence of the combinatorial immune system. Dev. Comp. Immunol. 23: 107-111, 1999.

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

Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, *et al. Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell 101: 671-684, 2000.

Schulte RD, Hasert B, Makus C, Michiels NK, Schulenberg H. Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite. Proc. Natl. Acad. Sci. USA 107: 7359-7364, 2010.

Schulte RD, Makus C, Schulenburg H. Host-parasite coevolution favours parasite genetic diversity and horizontal gene transfer. J. Evol. Biol. 26: 1836-1840, 2013.

Scofield VL, Schlumpberger JM, West LA, Weissman IL. Protochordate allorecognition is controlled by a MHC-like gene system. Nature 295: 499-502, 1982.

Sherman LS, Schrankel CS, Brown KJ, Smith LC. Extraordinary diversity of immune response proteins among sea urchins: nickel-isolated Sp185/333 proteins show broad variations in size and charge. PLOS ONE, 10, e0138892. 2015.

Silva JR. The onset of phagocytosis and identity in the embryo of *Lytechinus variegatus*. Dev. Comp. Immunol. 24: 733-739, 2000.

Sirono M, Cagliani R, Forni D, Clerici M. Evolutionary insights into host–pathogen interactions from mammalian sequence data. Nat. Rev. Genet. 16: 224-236. 2015.

Smith LC. Innate immune complexity in the purple sea urchin: diversity of the Sp185/333 system. Front. Immunol. 3: 70, 2012.

Smith LC, Chang L, Britten RJ, Davidson EH. 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. J. Immunol. 156: 593-602, 1996.

Smith LC, Lun CM. Multitasking rSp0032 has anti-pathogen binding activities predicting flexible and effective immune responses in sea urchins mediated by the Sp185/333 system. Pathogen Infect. Dis. 2: e1394, 2016.

Smith SM, Hulbert SH. Recombination events generating a novel *Rp1* race specificity. Mol. Plant Microbe In. 18: 220-228, 2005.

Sodergren E, Weinstock GM, Davidson EH, Cameron RA, Gibbs RA, Angerer RC, *et al.* The genome of the sea urchin, *Strongylocentrotus purpuratus*. Science 314: 941-952, 2006.

Spoel SH, Dong X. How do plants achieve immunity? Defence wihtout specialized immune cells. Nat. Rev. Immunol. 12: 89-100, 2012.

Stokes BA, Yadav S, Upasana S, Smith LC, Eleftherianos I. Bacterial and fungal pattern recognition receptors in homologous innate signaling pathways of insects and mammals. Front. Microbiol. 6: 19, 2015.

Sun Y, Bernardy EE, Hammer BK, Miyashiro T. Competence and natural transformation in vibrios. Mol. Microbiol. 89: 583-595, 2013.

Sutoh Y, Kasahara M. Copy number and sequence variation of leucine-rich repeat modules suggests distinct functional constraints operating on variable lymphocyte receptors

expressed by agnathan T cell-like and B cell-like lymphocytes. Immunogenetics 66: 403-409, 2014.

Taketa DA, De Tomaso AW. *Botryllus schlossseri* allorecognition: tackling the enigma. Dev. Comp. Immunol. 48: 254-265, 2015.

Teng G, Schatz DG. Regulation and evolution of the RAG recombinase. Adv. Immunol. 128: 1-39, 2015.

Terwilliger DP, Buckley KM, Brockton V, Ritter NJ, Smith LC. 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 Mol. Biol. 8: 16, 2007.

Terwilliger DP, Buckley KM, Mehta D, Moorjani PG, Smith LC. Unexpected diversity displayed in cDNAs expressed by the immune cells of the purple sea urchin, *Strongylocentrotus purpuratus*. Physiol. Genomics 26: 134-144, 2006.

Tonegawa S. Somatic generation of antibody diversity. Nature 302: 575-581, 1983.

Van Valen L. A new evolutionary law. Evol. Theory 1: 1-30, 1973.

Varriale S, Merlino A, Coscia MR, Mazzarella L, Oreste U. An evolutionary conserved motif is responsible for immunoglobulin heavy chain packing in the B cell membrane. Mol. Phylogen. Evol. 57: 1238-1244, 2010.

Watson FL, Puttmann-Holgado R, Thomas F, Lamar DL, Hughes M, Kondo M, *et al.* Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science 309: 1874-1878, 2005.

Zhang SM, Adema CM, Kepler TB, Loker ES. Diversification of Ig superfamily genes in an invertebrate. Science 305: 251-254, 2004.