Diversification of innate immune genes: lessons from the purple sea urchin

L. Courtney Smith[1]

Disease Models & Mechanisms • DMM

Pathogen diversification can alter infection virulence, which in turn drives the evolution of host immune diversification, resulting in countermeasures for survival in this arms race. Somatic recombination of the immunoglobulin gene family members is a very effective mechanism to diversify antibodies and T-cell receptors that function in the adaptive immune system. Although mechanisms to diversify innate immune genes are not clearly understood, a seemingly unlikely source for insight into innate immune diversification may be derived from the purple sea urchin, which has recently had its genome sequenced and annotated. Although there are many differences, some characteristics of the sea urchin make for a useful tool to understand the human immune system. The sea urchin is phylogenetically related to humans although, as a group, sea urchins are evolutionarily much older than mammals. Humans require both adaptive and innate immune responses to survive immune challenges, whereas sea urchins only require innate immune functions. Genes that function in immunity tend to be members of families, and the sea urchin has several innate immune gene families. One of these is the *Sp185/333* gene family with about 50 clustered members that encode a diverse array of putative immune response proteins. Understanding gene diversification in the *Sp185/333* family in the sea urchin may illuminate new mechanisms of diversification that could apply to gene families that function in innate immunity in humans, such as the killer immunoglobulin-like receptor genes.

The arms race and immune diversification

Host-pathogen interactions constitute a constant, long-term evolutionary arms race. This arms race is described as a competition between high rates of mutation and/or variation in microbes with short generation times versus long-lived hosts with corresponding low mutation rates (Haldane, 1949). This conflict is based on the relatively frequent appearance of new pathogen variants that may be more virulent, and thereby more successful and so could become established in the population. On the other side of the arms race, the host immune system must respond to pathogen variation within time scales that may be significantly shorter than host generation times. To survive, hosts employ a variety of mechanisms to diversify their immune response. Higher vertebrates use somatic recombination to generate large numbers of slightly variant immunoglobulin (Ig) family proteins that interact with pathogens (Neuberger, 2008). The alternative adaptive immune response that has recently been characterized in lampreys and hagfish has a unique mechanism that assembles variable lymphocyte receptor (*VLR*) genes from cassettes of gene segments encoding leucine-rich repeats (LRRs) (Nagawa et al., 2007; Rogozin et al., 2007). The *VLR* genes are expressed by two types of lamprey lymphocytes that resemble B and T cells (Guo et al., 2009). The level of diversity of the *VLR* genes generated by gene assembly has been estimated to be at least as great as that for the Ig family resulting from somatic recombination.

The *KIR* genes – at the intersection of adaptive and innate immunity

The major histocompatibility complex (MHC) gene family in higher vertebrates is composed of a number of closely clustered genes, each with multiple alleles. The diversity that is generated by multiple alleles at multiple loci is central to pathogen recognition and to the generation of specific immune responses. The regions of the MHC genes that have the highest levels of polymorphism are those that encode the peptide-binding groove and that are under selection for successful presentation of pathogen peptides to T-cell receptors (for reviews, see Vogel et al., 1999; Woelfing et al., 2009). The killer immunoglobulin-like receptors (KIR) are encoded by a highly diverse gene cluster, and function at the intersection of adaptive and innate immunity in vertebrates (Martinsohn et al., 1999; Biassoni, 2009). KIR proteins are expressed on natural killer (NK) cells and interact with MHC class I molecules encoded by a variety of the class I genes and alleles. Two types of KIR proteins are displayed on human NK cells: inhibitory KIR proteins that block the cytotoxic activity of NK cells, and activating KIR proteins that promote NK cell killing through association with adaptor signaling proteins. Inhibitory KIR proteins survey the level of MHC expression on self cells, which is a self-monitoring system where high levels indicate that the cell is normal and that cytotoxic activity of the NK cells is inhibited. Alternatively, virally infected cells have reduced levels of MHC expression to which NK cells respond cytotoxically. NK cells deploy an array of inhibitory and activating KIR proteins to regulate cytotoxic responses to self, altered self (virus infection) and pathogens (Biassoni, 2009).

Diversity of the MHC genes is driven by pathogen pressure, and the diversity of the *KIR* genes is driven by the MHC diversity (Marinez-Borra and Khakoo, 2008). In humans, 15 to 17 *KIR* genes cluster in a head-to-tail orientation about 2 kb apart

[1]Department of Biological Sciences, George Washington University, 340 Lisner Hall, 2023 G Street NW, Washington, DC 20052, USA
e-mail: csmith@gwu.edu

from each other. They are positioned in two clusters separated by a region of 14 kb with many repeat elements (Wilson et al., 2000) that may function either as an 'anchor' to stabilize proper alignment of the two clusters during meiosis, or it may promote recombination between the two clusters (Uhrberg, 2005). Within the clusters, different haplotypes can be composed of different numbers and types of *KIR* genes. This diversity is generated by intra- and intergenic recombination, gene conversion, domain shuffling, gene duplication/deletion, and single nucleotide polymorphisms (SNPs) that alter the coding sequence (Martin et al., 2004), but the detailed molecular mechanisms that promote (or block) the DNA variations are not known, only the results of the genomic instability are typically observed.

Innate immunity

Diversification of the vertebrate innate immune system is likely to occur at least for some gene families, and an understanding of the underlying mechanisms is emerging from studies of innate immunity in non-vertebrates. Invertebrates constitute most of the animals on the planet and they lack an adaptive immune response, yet they must also respond effectively to a wide range of pathogens. Both *Caenorhabditis elegans* and *Drosophila melanogaster* are excellent animals for investigating molecular aspects of innate immunity, in part, because they are relatively easy to manipulate genetically. Initial studies of the pathogen detection molecules in Drosophila were instrumental in initiating significant interest in the Toll-like receptors (TLRs) and the innate immune system in humans (Janeway, 1989; Medzhitov et al., 1997). For innate immune genes in animals that do not employ rearrangement or assembly mechanisms, as observed for the *Ig* and *VLR* families, investigations of immune diversification may be applied to understanding innate immune diversity mechanisms in humans.

Advances in genome sequencing now allow us to enumerate and study immune genes and their functions in animals with closer phylogenetic relationships to vertebrates and humans than those that are commonly used in comparative immunobiology, such as flies and round worms. There are a number of animals with attributes that provide more relevant information for comparisons to the human innate system

because they are more closely related and consequently may facilitate new directions in research. The annotation of the immune genes in the purple sea urchin (Hibino et al., 2006) enables comparisons to human innate immunity that can be placed within a relevant evolutionary context and may modify our thinking about innate immune diversification in humans.

The purple sea urchin

The purple sea urchin, *Strongylocentrotus purpuratus*, is a marine invertebrate that lives in the near-shore habitat of the Pacific Ocean along the west coast of the USA and Canada (see supplementary material in Sodergren et al., 2006). Compared with invertebrates such as Drosophila and *C. elegans* that are employed in studies of development and immunology, the purple sea urchin is very large with an adult size of ~2.5 inches or more in diameter (Fig. 1). Sea urchins have radial symmetry, generally have spherical bodies, and are covered with spines of variable numbers, shapes and lengths. The purple sea urchin and its sister species, the red sea urchin, have life spans similar to humans, ranging from 50 to 100 years (Ebert, 1967). Sea urchins are members of the Echinoderm phylum that is grouped within the deuterostome assemblage of animals that also includes the Chordate phylum, in which humans are classified (Fig. 2). Consequently, humans are much more closely related to sea urchins

Fig. 1. Purple sea urchins, *Strongylocentrotus purpuratus*. Four adult sea urchins are shown. The image was taken by Dr Susan Fuhs during a dive in the near-shore habitat of the Pacific Ocean of Southern California, USA. Figure reproduced with permission from Cambridge University Press (Smith, 2005).

than they are to fruit flies and round worms. Several sea urchin species have been used extensively for investigations of early development, which have direct applications to understanding the regulation of development in humans. Sea urchins also have the potential to provide relevant information about innate immune function in humans compared with studies from animals that are members of more distant phyla, such as arthropods (flies). The importance of purple sea urchins in biomedical research has been underscored with the sequencing and assembly of the genome (Sodergren et al., 2006) [for detailed genome annotation analysis, see *Developmental Biology* (2006) 300 (1)], which has revealed an innate immune system that is both complex and sophisticated.

Diversity of the immune response; the *Sp185/333* cDNA sequences and the encoded proteins

Many immune genes are members of large families that are composed of closely linked, duplicated genes with a similar sequence. Gene families may be a basic requirement for (and/or a result of) the diversification of non-rearranging genes that is selected for in response to pathogen pressure. In the genome of the purple sea urchin, a number of large gene families with putative or known immune function have been identified (Hibino et al., 2006; Rast et al., 2006; Sodergren et al., 2006). Early investigations of the immune functions of bacterially activated immune cells from the purple sea urchin identified a set of expressed sequence tags (EST), of which about 70% matched to two sequences, DD185 (Rast et al., 2000) and EST333 (Smith et al., 1996). These transcripts, originally called *185/333* (Nair et al., 2005) and now called *Sp185/333* to differentiate between the different species of sea urchins that express these genes (Ghosh et al., 2010), are readily induced in the immune cells of the sea urchin in response to immune challenge from bacteria and pathogen-associated molecular patterns, such as lipopolysaccharide from Gram-negative bacteria, β-1,3-glucan which is typical of fungi, and double-stranded RNA which is a signature of viruses (Rast et al., 2000; Nair et al., 2005; Terwilliger et al., 2007). Based on the gene expression patterns and sequence diversity, the *Sp185/333* transcripts are considered to be an important component of the sea urchin immune
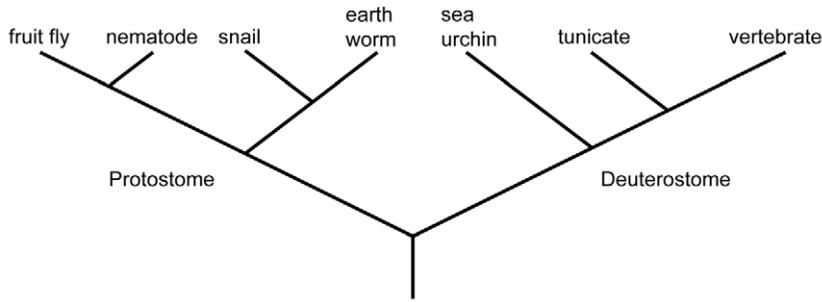
Fig. 2. Simplified phylogenetic tree of the animal kingdom. The tree illustrates the evolutionary relationship between sea urchins and vertebrates as members of the deuterostome branch of the animal kingdom. *Caenorhabditis elegans* and *Drosophila melanogaster* are members of the protostome branch.

Disease Models & Mechanisms ● DMM

response, providing the host with a diverse array of innate immune proteins.

Message diversity

The surprising level of sequence diversity of the *Sp185/333* cDNAs became evident when optimal alignments required the insertion of large artificial gaps. These artificial gaps defined blocks of sequence called *elements* that were variably present or absent within individual cDNAs, resulting in repeatedly identifiable *element patterns* in individual messages (Fig. 3) (Nair et al., 2005; Terwilliger et al., 2006). Element patterns are mosaic compositions of six to 22 elements, of a possible 27, such that most of the diversity within the *Sp185/333* transcripts is imparted by the element pattern (Buckley and Smith, 2007). Additional diversity results from numerous SNPs and small insertions/deletions (indels) that are present throughout the sequences. The

Sp185/333 transcripts are an example of an extreme level of sequence diversity, and are expressed by the immune cells of the purple sea urchin upon immune challenge.

Unexpectedly, half of the *Sp185/333* cDNAs have SNPs and small indels (one to a few nucleotides) that encode truncated proteins, some with a missense sequence (Terwilliger et al., 2007). This was particularly surprising because all but one of the sequenced genes (a total of 171) have perfect open reading frames encoding full-length proteins (Buckley et al., 2008a). When cDNAs and genes with matching element patterns are compared, nucleotide differences range from 5.8-16.7%. Furthermore, very few of the messages match exactly to the genes from an individual sea urchin. The SNPs in the messages and the corresponding positions in the genes from which they are most likely transcribed, indicate that the most common change from

gene to message is cytidine to uridine. This is consistent with message editing by cytidine deaminase (Chester et al., 2000), which is a member of a protein family that includes one protein involved with affinity maturation of antibodies by editing the Ig gene sequences in B cells (Liu and Schatz, 2009). The range of sequence variations between genes and messages also suggests that a low-fidelity polymerase, such as polymerase μ (Ruiz et al., 2001), may transcribe the *Sp185/333* gene family. Several gene models for cytidine deaminases and one for polymerase μ are present in the sea urchin genome (Hibino et al., 2006). Consequently, message editing and/or low-level transcription fidelity may be employed by sea urchins to increase the diversity of the *Sp185/333* messages and to expand the repertoire of the encoded Sp185/333 proteins responding to microbial challenge.

Protein diversity

The Sp185/333 proteins are present in or on subsets of immune cells (Fig. 4) (Brockton et al., 2008), and may associate with the cell surface through interactions between the arginine-glycine-aspartic acid (RGD) motif on the Sp185/333 proteins and integrins, which are integral membrane proteins with cell surface expression on coelomocytes. Integrin gene models are present in the sea urchin genome (Sodergren et al., 2006; Whittaker et al., 2006), and some integrins are expressed by the immune cells (Smith et al., 2006). The diversity of the Sp185/333 protein repertoire in coelomocytes results in patterns of bands or spots on one- and two-dimensional western blots that are not shared among individual sea urchins (Brockton et al., 2008; Dheilly et al., 2009). The number of Sp185/333-positive spots on two-dimensional western blots suggests that individual sea urchins may produce significantly more protein variants, which includes truncated and missense forms, than the estimated number of genes (Dheilly et al., 2009). The combination of *Sp185/333* gene diversification (Buckley et al., 2008b) plus message editing (Buckley et al., 2008a) results in an extraordinarily diverse set of expressed proteins.

Sp185/333 gene diversity

The *Sp185/333* gene family is composed of about 50 members that are highly polymorphic among individual sea urchins (Terwilliger et al., 2006; Buckley et al., 2008b).
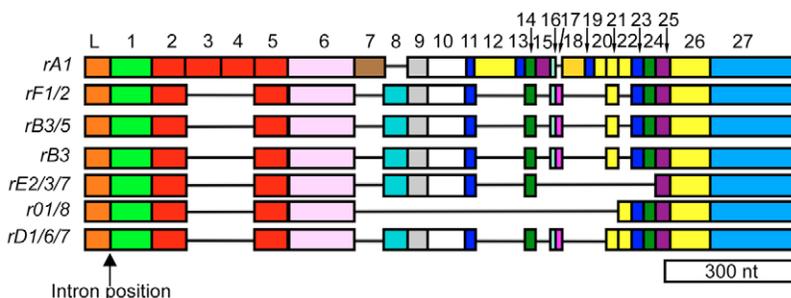

Fig. 3. Alignment of the *Sp185/333* sequences. An optimal alignment of the *Sp185/333* cDNA sequences requires the insertion of artificial gaps (black horizontal lines) that define blocks of sequence called *elements* (colored rectangles). Element numbers are listed at the top; L, leader. Different cDNAs have different subsets of elements that create a mosaic of *element patterns*. A selection of cDNAs with representative element patterns is shown. Designations of the element patterns are listed to the left. Some of the elements are also imperfect tandem or interspersed sequence repeats, and are indicated as red, dark blue, dark green, yellow and purple blocks. The other elements do not have repeated sequences. Figure modified with permission from BioMed Central Ltd (Buckley and Smith, 2007).
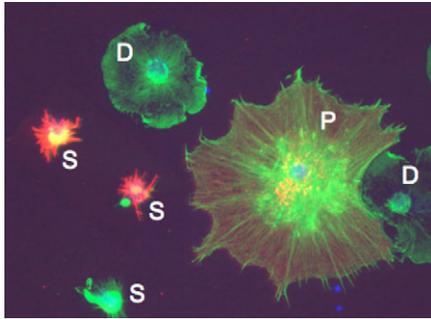
Fig. 4. Sp185/333-positive coelomocytes. The phagocyte class of coelomocytes (immune cells) include polygonal (P) cells, discoidal (D) cells, and small (S) phagocytes. A subset of the small phagocytes are Sp185/333-positive. Some of the polygonal cells have Sp185/333-positive perinuclear vesicles. The discoidal phagocytes are negative for Sp185/333. The Sp185/333 proteins are recognized by anti-185/333 antibodies and detected with a secondary antibody labeled with a red fluorochrome. Actin is recognized with an anti-actin antibody and detected with a secondary antibody labeled with a green fluorochrome. The nuclear DNA is detected by DAPI staining, which is blue (for details on immunofluorescent staining, see Brockton et al., 2008).

The genes are ≤2 kb with two exons, of which the second exon encodes the protein (except for the leader) including all of the elements (Buckley and Smith, 2007). The variable composition of the element patterns appear as a mosaic of sequences in the second exon and impart the greatest sequence diversity to the genes. An average of 71% of all genes sequenced from three animals are unique, and none of the genes are shared among all three animals employed in the analysis (Buckley et al., 2008a). However, when individual element sequences are compared between the same three sea urchins, 28% are found to be shared among two or more animals (i.e. element sequences are shared but full-length gene sequences are not). This illustrates the diversity and complexity of these short genes and infers the high level of polymorphism within the gene family and within the population of sea urchins as a whole. Only a single pseudogene with a deletion and frameshift mutation has been identified from 171 genes. Genes lacking introns are present and may be the result of retrotransposition, but it is not known whether they are expressed (Buckley and Smith, 2007).

The 5′-end of the genes encode type 1 repeats (red blocks in Fig. 3, see figure leg-

end) that are present in two to four copies and show evidence of single and multiple repeat duplications, deletions and recombination events (Buckley et al., 2008b). There is no correlation between the element patterns at the 5′-end of the second exon with the patterns at the 3′-end, indicating a rapid rate of recombination within the exon. In fact, there is evidence that recombination may occur at any point along the sequence of a gene with no particular hotspot (Buckley et al., 2008b). Analysis of the elements within the *Sp185/333* genes also illustrates diversity that is based on variations in element patterns. Within the 171 genes that have been analyzed, a given specific sequence version of an element may be adjacent to as many as 12 different variants of neighboring elements, in addition to being adjacent to different types of neighboring elements. In the face of extensive and extraordinarily rapid recombination, there may also be mechanisms to block the formation of pseudogenes because gene fragments or isolated elements, and the homogenization of gene sequences have not been found. Furthermore, there may also be mechanisms to block homogenization of gene sequences.

The spatial relationships between six clustered *Sp185/333* genes in a bacterial artificial chromosome (BAC) clone shows that many of the linked genes are positioned as close together as 3.2 kb (C. Miller, K. Buckley, R. Easley and L.C.S., unpublished; L.C.M., unpublished). The outer flanking

genes are oriented in the same direction and are opposite to those within the cluster (Fig. 5A). A screen of sea urchin BAC libraries shows that *Sp185/333*-positive clones have at least two genes clustered tightly enough to allow PCR amplification of intergenic regions. Each gene on the BAC is flanked by GA microsatellites, and GAT microsatellites are present surrounding segmental duplications that include a gene (Fig. 5A) (Buckley and Smith, 2007) (C. Miller, K. Buckley, R. Easley and L.C.S., unpublished). Significant genome instability in the region where the *Sp185/333* genes are located is probably because of the high level of sequence similarity among the genes, which is ≥88% (Buckley and Smith, 2007), the repeats within the genes, and the presence of microsatellites that surround the genes. This would promote sequence diversification from gene duplication, deletion, segmental duplication, gene recombination and gene conversion, in addition to the possibility of significant variations in *Sp185/333* gene copy number in different individuals resulting from unequal crossovers and meiotic mispairing.

Conclusions

There is a preponderance of gene families that function in immunity with members that are clustered and that share sequences. Examples include the Toll-like receptors, nucleotide-binding and oligomerization domain (NOD)-like receptors, the *KIR* genes and the *Sp185/333* genes, in addition
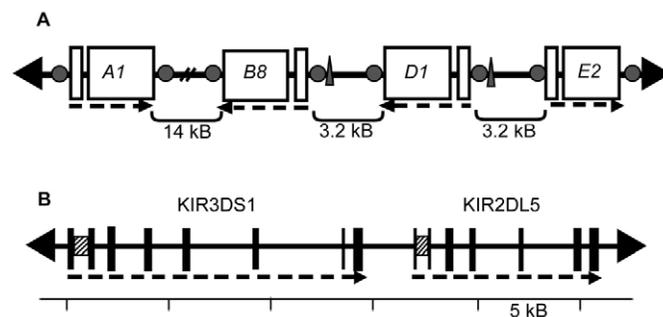

Fig. 5. Microsatellites surround the *Sp185/333* genes and minisatellites fill the first intron of the *KIR* genes. (A) Four *Sp185/333* genes (~1.8-2.0 kb each) are shown as assembled in the *S. purpuratus* genome (ver 2.0), and the two exons for each gene are shown as boxes. Gene names correspond to the element patterns in the second exon (see Fig. 3). Two types of microsatellites, GA repeats (circles) and GAT repeats (triangles) are associated with the *Sp185/333* genes. The region shown is about 28 kb and is not to scale. Figure modified with permission from BioMed Central Ltd (Buckley and Smith, 2007). (B) Two linked *KIR* genes are shown with seven or eight exons (vertical bars). The first intron of the *KIR* genes is composed entirely of a minisatellite (diagonal striped boxes). The scale is shown below the *KIR* genes. The dotted line below the genes indicates the direction of transcription. Figure modified with permission from the National Academy of Sciences, U.S.A. (Wilson et al., 2000).

to many others in both plants and animals. It is commonly stated in the literature that immune gene families are generated by a variety of mechanisms including gene and exon duplication, segmental duplication, meiotic mispairing, and unequal crossovers in response to pathogen pressure, but that the exact DNA sequences mediating the molecular scrambling that occurs in tightly clustered gene families are not known. The sequence similarities among family members may be the basis for expansions and contractions of family size through meiotic mispairing (etc., as listed above), but other, non-coding sequences may be involved. The *KIR* genes show significant sequence similarity in both the coding regions and the intergenic regions, a characteristic that is consistent with segmental duplications rather than just gene duplications (Wilson et al., 2000). The first intron of the *KIR* genes is composed entirely of a minisatellite, which is 30 to 60 repeats of a GC-rich sequence of 19 to 20 nucleotides (Fig. 5B). It is noteworthy that the Ig-like transcript (*ILT*) gene cluster that is adjacent to and much older than the *KIR* cluster does not have associated minisatellites and shows significantly less sequence similarities among the genes, with no similarities in the intergenic regions (Wilson et al., 2000). The presence of minisatellites in the *KIR* family and the microsatellites in the *Sp185/333* gene cluster tempts speculation on the possible importance of these types of repeats in the diversification of tightly clustered genes within families. Perhaps the *KIR* minisatellites and the *Sp185/333* microsatellites function similarly to promote gene duplication or deletion, gene conversion, and particularly segmental duplication (duplication of a region that might include one or more genes) that provides the raw material for sequence diversification. Newly generated genes or genes with altered sequences would be acted upon by selection, either directly (*Sp185/333*) or indirectly (*KIRs* through the MHC), for responses to pathogens. The notion that mini- and microsatellites may be involved in driving the generation of immune gene families has emerged from investigations of immunity in the sea urchin. The conservation of mechanisms employed in sea urchin immune diversification can be tested in yeast to determine whether the presence of mini- and microsatellites destabilize a region of DNA over a number of genera-

Innate immune diversity in the purple sea urchin as a system to direct investigations of gene and protein diversification in higher vertebrates

- The purple sea urchin has a sophisticated innate immune system that functions in the absence of adaptive immune capabilities, yet responds to pathogens with a host of diverse proteins to fight infection

- The purple sea urchin is an echinoderm, a sister group to the chordates that includes humans. Sea urchins are therefore evolutionarily related more closely to humans than other organisms that are used to evaluate immune function, such as fruit flies and round worms

- A family of immune genes in the purple sea urchin, *Sp185/333*, produces a highly diverse array of proteins in response to pathogens. The mechanisms that appear to generate this diversity may provide insight into how diversity is achieved for some innate immune gene families in humans

tions. Yeast have been used to investigate the expansion and contraction of repeat numbers in micro- and minisatellites, and their general instability that results from recombination and other genetic variations in linked genes (Richard and Paques, 2000; Bagshaw et al., 2008). Only rarely do studies address the opposite question of whether the presence of simple repeats affect genetic changes in linked genes (Shen et al., 1981; Treco and Arnheim, 1986). Future work to investigate whether the presence of micro- and minisatellites drive the recombination, conversion, crossovers, meiotic mispairing, etc., of closely linked genes may expand and perhaps redirect our thinking about how the human innate immune system copes with multitudes of diverse pathogens.

REFERENCES
Bagshaw, A. T., Pitt, J. P. and Gemmell, N. J. (2008). High frequency of microsatellites in *S. cerevisiae* meiotic recombination hotspots. *BMC Genomics* **9**, 49.

Biassoni, R. (2009). Human natural killer receptors, co-receptors, and their ligands. *Curr. Protoc. Immunol.* **Chapter 14**, Unit 14.10.

Brockton, V., Henson, J. H., Raftos, D. A., Majeske, A. J., Kim, Y. O. and 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. *J. Cell Sci.* **121**, 339-348.

Buckley, K. M. and Smith, L. C. (2007). Extraordinary diversity among members of the large gene family, *185/333*, from the purple sea urchin, *Strongylocentrotus purpuratus*. *BMC Mol. Biol.* **8**, 68.

Buckley, K. M., Terwilliger, D. P. and Smith, L. C. (2008a). Sequence variations between the *185/333* genes and messages from the purple sea urchin suggest post-transcriptional modifications. *J. Immunol.* **181**, 2203-2212.

Buckley, K. M., Munshaw, S., Kepler, T. B. and Smith, L. C. (2008b). 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.

Chester, A., Scott, J., Anant, S. and Navaratnam, N. (2000). RNA editing: cytidine to uridine conversion in apolipoprotein B mRNA. *Biochim. Biophys. Acta* **1494**, 1-13.

Dheilly, N. M., Nair, S. V., Smith, L. C. and Raftos, D. A. (2009). Highly variable immune response proteins from the sea urchin, *Strongylocentrotus purpuratus*: proteomic analysis of diversity within and between individuals. *J. Immunol.* **182**, 2203-2212.

Ebert, T. A. (1967). Negative growth and longevity in the purple sea urchin *Strongylocentrotus purpuratus* (Stimpson). *Science* **157**, 557-558.

Ghosh, J. G., Buckley, K. M., Nair, S. V., Raftos, D. A., Miller, C., Majeske, A. J., Hibino, T., Rast, J. P., Roth, M. and Smith, L. C. (2010). *Sp185/333*: a novel family of genes and proteins involved in the purple sea urchin immune response. *Dev. Comp. Immunol.* **34**, 235-245.

Guo, P., Masayuki, H., Herrin, B. R., Li, J., Yu, C., Sadlonova, A. and Cooper, M. D. (2009). Dual nature of the adaptive immune system in lampreys. *Nature* **459**, 796-802.

Haldane, J. B. S. (1949). Disease and evolution. *La Ricerca Scientifica* **19**, 68-75.

Hibino, T., Loza-Coll, M., Messier, C., Majeske, A. J., Cohen, A. H., Terwilliger, D. P., Buckley, K. M., Brockton, V., Nair, S., Berney, K. et al. (2006). The immune gene repertoire encoded in the purple sea urchin genome. *Dev. Biol.* **300**, 349-365.

Janeway, C. A., Jr (1989). Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harbor Symp. Quant. Biol.* **54**, 1-13.

Liu, M. and Schatz, D. G. (2009). Balancing AID and DNA repair during somatic hypermutation. *Trends Immunol.* **30**, 173-181.

Marinez-Borra, J. and Khakoo, S. I. (2008). Speed and selection in the evolution of killer-cell immunoglobulin-like receptors. *Int. J. Immunogenet.* **35**, 89-96.

Martin, A. M., Kulski, J. K., Gaudieri, S., Witt, C. S., Freitas, E. M., Trowsdale, J. and Christiansen, F. T. (2004). Comparative genomic analysis, diversity and evolution of two KIR haplotypes A and B. *Gene* **335**, 121-131.

Martinsohn, J. T., Sousa, A. B., Guethlein, L. A. and Howard, J. C. (1999). The gene conversion hypothesis of MHC evolution: a review. *Immunogenetics* **50**, 168-200.

Medzhitov, R., Preston-Hurlburt, P. and Janeway, C. A., Jr (1997). A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. *Nature* **388**, 394-397.

Nagawa, F., Kishishita, N., Shimizu, K., Hirose, S., Miyoshi, M., Nezu, J., Nishimura, T., Nishizumi, H., Takahashi, Y., Hashimoto, S. et al. (2007). Antigen-receptor genes of the agnathan lamprey are assembled by a process involving copy choice. *Nat. Immunol.* **8**, 206-213.

Nair, S. V., Del Valle, H., Gross, P. S., Terwilliger, D. P. and 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. *Physiol. Genomics* **22**, 33-47.

Neuberger, M. S. (2008). Antibody diversification by somatic mutation: from Burnet onwards. *Immunol. Cell Biol.* **86**, 124-132.

Rast, J. P., Pancer, Z. and Davidson, E. H. (2000). New approaches towards an understanding of deuterostome immunity. *Curr. Top. Microbiol. Immunol.* **248**, 3-16.

Rast, J. P., Smith, L. C., Loza-Coll, M., Hibino, T. and Litman, G. W. (2006). Genomic insights into the immune system of the sea urchin. *Science* **314**, 952-956.

Richard, G.-F. and Paques, F. (2000). Mini- and microsatellite expansions: the recombination connection. *EMBO Rep.* **1**, 122-126.

Rogozin, I. B., Iyer, L. M., Liang, L., Glazko, G. V., Liston, V. G., Pavlov, Y. I., Aravind, L. and Pancer, Z. (2007). Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat. Immunol.* **8**, 647-656.

Ruiz, J. F., Dominguez, O., Lain de Lera, T., Garcia-Diaz, M., Bernad, A. and Blanco, L. (2001). DNA polymerase mu, a candidate hypermutase? *Philos. Trans. R. Soc. London Ser. B, Biol. Sci.* **356**, 99-109.

Shen, S., Slightom, J. L. and Smithies, O. (1981). A history of the human fetal globin gene duplication. *Cell* **26**, 191-203.

Smith, L. C. (2005). Host responses to bacteria: innate immunity in invertebrates. In *The Influence of Cooperative Bacteria on Animal Host Biology* (*Advances in Molecular and Cellular Microbiology 10*) (eds M. J. McFall-Ngai, B. Henderson and E. G. Ruby), pp. 293-320. New York, NY: Cambridge University Press.

Smith, L. C., Chang, L., Britten, R. J. and 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. *J. Immunol.* **156**, 593-602.

Smith, L. C., Rast, J. P., Brockton, V., Terwilliger, D. P., Nair, S. V., Buckley, K. M. and Majeske, A. J. (2006). The sea urchin immune system. *Invertebrate Survival Journal* **3**, 25-39.

Sodergren, E., Weinstock, G. M., Davidson, E. H., Cameron, R. A., Gibbs, R. A., Angerer, R. C., Angerer, L. M., Arnone, M. I., Burgess, D. R., Burke, R. D. et al. (2006). The genome of the sea urchin, *Strongylocentrotus purpuratus*. *Science* **314**, 941-952.

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

Terwilliger, D. P., Buckley, K. M., Brockton, V., Ritter, N. J. and 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 Mol. Biol.* **8**, 16.

Treco, D. and Arnheim, N. (1986). The evolutionarily conserved repetitive sequence d(TA. AC)$_n$ promotes reciprocal exchange and generates unusual recombinant tetrads during yeast meiosis. *Mol. Cell. Biol.* **6**, 3934-3947.

Uhrberg, M. (2005). The KIR gene family: life in the fast lane of evolution. *Eur. J. Immunol.* **35**, 10-15.

Vogel, T. U., Evans, D. T., Urvater, J. A., O'Connor, D. H., Hughes, A. L. and Watkins, D. I. (1999). Major histocompatibility complex class I genes in primates: co-evolution with pathogens. *Immunol. Rev.* **167**, 327-337.

Whittaker, C. A., Bergeron, K. F., Whittle, J., Brandhorst, B. P., Burke, R. D. and Hynes, R. O. (2006). The echinoderm adhesome. *Dev. Biol.* **300**, 252-266.

Wilson, M. J., Torkar, M., Haude, A., Milne, S., Jones, T., Sheer, D., Beck, S. and Trowsdale, J. (2000). Plasticity in the organization and sequences of human KIR/ILT gene families. *Proc. Natl. Acad. Sci. USA* **97**, 4778-4783.

Woelfing, B., Traulsen, A., Milinski, M. and Boehm, T. (2009). Does intra-individual major histocompatibility complex diversity keep a golden mean? *Philos. Trans. R. Soc. London B, Biol. Sci.* **364**, 117-128.