Published in final edited form as:

Curr Opin Immunol. 2007 October ; 19(5): 526–534.

Alternative Mechanisms of Immune Receptor Diversity

Gary W. Litman^{1,2,3}, Larry J. Dishaw^{1,2}, John P. Cannon^{2,3}, Robert N. Haire³, and Jonathan P. Rast⁴

¹Department of Molecular Genetics, All Children's Hospital, 801 Sixth Street South, St. Petersburg, FL 33701

²Immunology Program, H. Lee Moffitt Cancer Center and Research Institute, 12902 Magnolia Boulevard, Tampa, FL 33612

³Department of Pediatrics, University of South Florida College of Medicine, USF/ACH Children's Research Institute, 830 First Street South, St. Petersburg, FL 33701

⁴Sunnybrook Research Institute and Department of Medical Biophysics, University of Toronto, 2075 Bayview Avenue, Toronto, Ontario M4N 3M5, Canada

Abstract

Our views of both innate and adaptive immunity have been significantly modified by recent studies of immune receptors and immunity in protostomes, invertebrate deuterostomes and jawless vertebrates. Extraordinary variation in the means whereby organisms recognize pathogens has been revealed by a series of recent findings, including: novel forms of familiar immune receptors, high genetic polymorphism for new receptor types, germline rearrangement for non-Ig domain receptors, somatic variation of germline-encoded receptors and unusually complex alternative splicing of genes with both immune and non-immune roles. Collectively, these observations underscore pathways in the evolution of immune recognition and suggest universal processes by which immune systems co-opt and integrate existing cellular mechanisms to effect diverse recognition functions.

Introduction

Our general concepts of innate and adaptive immunity are based historically on extensive investigations in two groups of eutherian mammals (primates and rodents) and in an avian species (chicken). Over the last several decades, two prevailing themes regarding immune receptor diversification have emerged from this work: 1) innate receptors are encoded as single copy genes or in multigene families, neither of which undergo genetic recombination during cellular differentiation and 2) adaptive immunity is mediated by immunoglobulins (Igs) and T cell antigen receptors (TCRs) that are encoded in relatively complex multigene systems. Rearranging Ig and TCR systems serve as our prototype of dynamic adaptive immunity but neither system has been identified in jawless vertebrates, protochordates or non-chordate invertebrates [1;2]. The question thus arose: are there analogous systems in other animal groups and if so, what rules govern their function?

Like mammals, jawless vertebrates, protochordates and other invertebrates inhabit an environment filled with complex, shifting populations of microorganisms, some of which are

Corresponding author: Litman, Gary W. (litmang@allkids.org).

e-mail addresses: ldishaw@health.usf.edu, jcannon@health.usf.edu, rhaire@health.usf.edu, jrast@sri.utoronto.ca

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

potential pathogens and symbionts. The demands placed on the immune systems of these animals are comparable to those confronting jawed vertebrates [3]. Therefore, it should be expected that the immune systems in these organisms have their own sophisticated recognition systems to deal with these challenges. Recent studies have indeed uncovered a number of alternative solutions to both innate and adaptive immunity throughout the metazoan phyla. In this review, we will focus on select recent observations that in some cases illuminate the rich evolutionary history of our own immune system and in others reveal completely independent solutions to confronting immune challenges. These findings provide long-needed independent perspectives on the intense evolutionary forces that shape animal immunity.

Variable lymphocyte receptors and gene rearrangement

Early studies in comparative immunology revealed a striking distinction between the immune response of the sea lamprey, a jawless vertebrate, and those characterized in the jawed vertebrates. In particular, an inducible, specific immune response could be elicited by immunization with bacterial or red blood cells; however, the structure of this mediator appeared to be markedly different from that of jawed vertebrate antibodies [4]. Mononuclear cells with morphological features of lymphocytes were identified in various tissues and in the peripheral blood and were characterized definitively in recent studies; however, expressed sequence tag (EST) surveys performed in parallel on these cells have failed to reveal unequivocal lamprey homologs of many crucial lymphocyte-specific cell-surface determinants [5].

Antigen/mitogen challenges of sea-lamprey result in the significant expansion of the lymphocyte-like population. A subtractive EST survey of these stimulated cells led to the identification of sequences encoding a particularly complex set of leucine-rich repeat (LRR)-containing extracellular molecules, termed variable lymphocyte receptors (VLRs) [6]. Two homologs of VLRs are found on a single chromosome in hagfish [7;8], representing the only other extant lineage of jawless vertebrate. Sea lamprey VLR proteins are comprised of a short N-terminal region, an N-terminal LRR, followed by up to nine variable LRRs, a small connecting peptide, a C-terminal LRR, a threonine/proline stalk and a C-terminal hydrophobic tail, including a glycosyl-phosphatidyl-inositol (GPI) anchor. The number and sequence of the internal LRR motifs are highly diverse and have been predicted to form the antigen binding region [9], which has been supported by structural findings with hagfish VLRs [10]. Portions of the N- and C-terminal domains are invariant. Notably, expression of VLRs in lymphocyte-like cells appears to exhibit allelic exclusion [6;11], a hallmark of the rearranging immune gene loci of the jawed vertebrates. Furthermore, VLRs exhibit specificity for a bacterial antigen used in a direct immune challenge [9].

VLRs employ an unusual mechanism for generating a mature transcript. A single copy gene, possessing exons that correspond to the 5' and 3' parts of the mRNA produced in activated lymphocytes, exists in the germline as an "empty" VLR cassette. Sequences encoding the variable, internal LRRs are present in germline segments residing outside the empty VLR cassette (Fig. 1a). During development of the lymphocyte, a complex DNA-level process assembles the outlying segments [9;11], which can be in either orientation relative to the transcription of the VLR gene itself, and "pastes" them into the empty germline VLR gene (Fig. 1a-c); the integration is independent of RNA splicing [9;11].

The assembly of the VLR gene involves the use of short regions of nucleotide identity that initiate the transfer of LRR segments into the maturing VLR locus [11] (Fig. 1b). In somatically expanding cell lineages, a single VLR begins to assemble via replication-coupled template switching (e.g., copy-choice recombination) [11]. This can occur when the DNA replicating polymerase pauses, disassociates, and reassociates with a homologous region elsewhere [12-14], most likely due to intrinsic features of the VLR locus, as well as the repetitive nature

of multiple LRR cassettes [11;13]. Typically, secondary structure facilitates copy-choice rearrangements. The potential role in VLR diversification of two recently identified putative deaminases of the AID-APOBEC family, which exhibit *in vitro* mutagenesis and recombination activity, is of considerable interest [15].

The postulated role of VLRs as mediators of adaptive immunity may have an ancient history within the broader functional roles of other metazoan LRR-containing immune-related proteins, which include the Toll-like receptor (TLR) [16] and the intracellular NACHT-LRR containing proteins (NLRs) [17;18] (see below). Growing evidence discerned from LRR-containing genes in the sea urchin and amphioxus genomes suggests that the specific genome-wide architecture that facilitates VLR rearrangement may have been present before the evolutionary divergence of the vertebrate lineage.

Down Syndrome Cell Adhesion Molecule

Despite the marked structural differences between VLRs and Ig/TCR as adaptive immune mediators, both systems utilize site-directed forms of genetic rearrangement and gene conversion. Basic questions arise as to whether or not all processes that diversify immune receptors rely on genetic recombination, what mechanisms of recombination are involved and whether such recombination is somatic or meiotic.

Recent investigations have uncovered a previously unrecognized form of immune receptor diversification in *Drosophila melanogaster* (fruit fly), based on complex alternative splicing of the *Drosophila* homolog of the Down syndrome cell adhesion molecule (Dscam), a single-copy gene that encodes a relatively large Type I transmembrane molecule containing extracellular Ig and fibronectin type 3 domains, as well as a cytoplasmic domain with potential signaling motifs; Dscam protein is essential for neuronal wiring in the *Drosophila* embryo [19]. Diversity in the Dscam protein is produced by mutually exclusive alternative splicing of four duplicated exons (4,6,9 and 17) (Fig. 2). Three of these four exons (4,6 and 9) encode segments of IgSF domains and are dramatically expanded in the genome; it is predicted that the *Dscam* gene can produce on the order of 40,000 distinct isoforms through alternative splicing alone [19-21].

Expression of Dscam in fat body tissue and hemocytes, as well as the distinctive use of multiple exons to encode diversified Dscam proteins in these tissues, suggested a function for Dscam in immunity. This hypothesis was supported using mutant animals and RNAi knockdown studies, as well as by indications of direct microbial binding by certain Dscam isoforms [20]. The *Dscam* gene of the mosquito *Anopheles gambiae*, another Dipteran insect, shares many structural features with its *Drosophila* homolog, including expansion of 3 exons encoding Ig domains and the ability to produce over 30,000 possible isoforms by alternative splicing. Studies in *A. gambiae* have shown that hemocytes of this insect appear to respond to pathogenic challenge by alteration of their repertoires of Dscam isoforms in patterns specific to different microbial challenges [21]. RNAi silencing of the *A. gambiae* *Dscam* gene compromises resistance to infection by at least two different bacterial species and by the malaria parasite *Plasmodium berghei*. Dscam silencing also renders the organism susceptible to opportunistic infections. The induced splice form repertoires further demonstrated increased affinity in *in vitro* microbial binding assays to the challenge pathogen, but not necessarily to other pathogens. These data, together with the observations in *Drosophila*, demonstrate a function of *Dscam* in Dipteran immunity and strongly implicate alternative splicing as central to mediating that function (Fig. 2). Taken as a whole, these results indicate that the same molecule has been recruited for two different fundamental processes (neuronal development and immunity) and not only provide an important precedent for hypotheses of the evolutionary origins of receptor recruitment to immune responses but also raise the possibility of a more global role for

differential RNA processing, including the targeting of alternative splice sites, in the generation of receptor diversity.

Other diversified Ig domains in invertebrates

Two other gene families encoding Ig domain-containing proteins have emerged as strong candidates for mediators of alternative immune recognition: the fibrinogen-related proteins (FREPs) of the snail *Biomphalaria glabrata* and the variable domain-containing chitin-binding proteins (VCBPs) of the amphioxus *Branchiostoma floridae*, a cephalochordate.

Snail FREPs were identified initially as proteins that are induced after trematode infection and are capable of precipitating pathogen-derived proteins and binding to trematode sporocysts [22]. FREPs consist of one or two N-terminal Ig domains plus a C-terminal fibrinogen-like domain [23]. Cloning of many separate FREP cDNAs showed that they encode a diversified set of proteins, present in multiple subfamilies. Subsequent studies of sequences encoding the N-terminal Ig domain of FREP3 indicate potential somatic diversification of this region in individual snails [24]. Although molecular mechanism(s) underlying this diversification and its functional significance are not known, this finding suggests the presence of somatic diversifying mechanisms in another major branch of bilaterian animal phylogeny.

Amphioxus VCBPs were identified in a generic search for genes encoding secreted or transmembrane IgSF molecules [25]. Like FREPs, VCBPs are structural chimeras, consisting of two tandem N-terminal Ig V domains plus a C-terminal chitin-binding domain. VCBPs are encoded by five small, diversified families of genes; in adults their expression is confined to discrete cells in the intestine. The N-terminal V domains of VCBP2 and VCBP5 are related closely and contain a region of extensive, interindividual amino acid variation across approximately 10-13 residues at the N-terminus of the domain; a second region of hypervariation has been observed at the same position in the C-terminal V domain of VCBP2 [25;26]. This unusually high level of variation appears to be maintained in the population rather than being generated through somatic mechanisms. The role of a meiotic recombination hotspot in the hyperdiversification of VCBP2/5 genes currently is under investigation (L. Dishaw, unpublished observations). Hyperdiversity at the population level also is a characteristic of the *Fu/Hc* and *fester* gene systems that mediate self-nonself recognition in the protochordate *Botryllus schlosseri*, a colonial tunicate [27], although it is not thought to function in a classical immune-protective sense (i.e., against pathogenic microbes). Recently, it has been proposed that polymorphic components of complement receptor-like proteins may serve as allorecognition markers in another urochordate [28].

The structure of the tandem V domains of a VCBP molecule (VCBP3) has been solved at 1.85 Å resolution, revealing a unique orientation of the V domains, juxtaposed from the traditional configurations seen typically in antibodies or TCRs [29]. Hypervariation in VCBP2/5 lies in a portion of the molecule that may serve as a surface for ligand interactions, although the ligands recognized by VCBPs presently are unknown.

Toll-like receptors and NACHT-leucine-rich repeat (NLR) proteins

TLR proteins are characterized by an extracellular leucine rich repeat (LRR) domain, a single transmembrane (TM) region and an intracellular Toll/IL1R (TIR) signaling interaction domain. The genomic complement of TLR genes in vertebrates and protostome invertebrates ranges from one in the nematode *Caenorhabditis elegans* to about 15 genes in the bony fish representatives. Two variants of TLR structure can be recognized that differ principally in the arrangement of their extracellular LRR motifs: one that accounts for the majority of protostome TLRs and a second that accounts for all vertebrate receptors and the *Drosophila* TLR Toll9

[30]. Analysis of jawed vertebrate TLRs suggests that they have evolved in a conservative fashion at least since they were present in a bony fish common ancestor [31].

The recent annotation of the sea urchin genome revealed more than 220 TLR genes [32-34]. Most of these belong to families with similar sequences and are likely the products of relatively recent expansions, although they show clear signs of significant functional diversification [32] (Fig. 3a). Specifically, the members of these diversified subgroups are phylogenetically more closely allied to each other than to the TLRs known from chordates or protostomes. Comparison of the coding regions among members of these families reveals a highly variable LRR ectodomain and relatively conserved TM and TIR domains. Diversity in the LRRs takes the form of localized amino acid substitutions, small insertions within and between the LRRs, and for one subfamily, insertion-deletions of complete LRR elements [32]. Eleven of the remaining TLRs can be grouped in separate categories, some of which are protostome-like and others with unusually short extracellular regions. Many of the sea urchin TLR genes are encoded in genomic clusters of single exon genes (Fig. 3a). Nearly 30% of the genes in the TLR families are pseudogenes, suggesting a high evolutionary turnover. Collectively, these characteristics paint a dynamic picture of sea urchin TLR evolution that contrasts sharply with the conservative mode of vertebrate TLR evolution.

In the sea urchin the protostome and vertebrate-type TLR genes are expressed in larval cells as they differentiate after gastrulation and in the adult, but are not expressed during embryogenesis. Feeding larvae and adult coelomocytes (blood-like cells) express complex TLR repertoires that are distinct but overlapping. An expanded set of 26 putative TLR adaptors could serve as TIR domain signaling intermediates and may functionally partition the expanded sea urchin TLR recognition capacity.

The vertebrate NLR proteins (including the nucleotide-binding oligomerization domain [Nod] and NALP genes) are encoded by about 20 genes, and are implicated in cytoplasmic innate recognition and gut associated immunity [35]. In the sea urchin, homologs of the NLR proteins and another family of innate receptors (scavenger receptor cysteine-rich proteins) are encoded by more than 200 genes [34]. Similar expansions of both the TLRs and NLRs are seen in the protochordate amphioxus (unpublished observation). As in vertebrates, the sea urchin NLR genes are expressed at high levels in the adult gut and it may be that gut pathogens or symbiots are an important factor in the sea urchin NLR and possibly TLR expansions.

The multiplicity and patterns of sequence divergence among sea urchin TLRs suggest they are utilized differently from their vertebrate counterparts. This may apply also to the NLR genes. While it is unclear at present how this TLR diversity is employed, the possibility of restricted expression (either combinatorial or exclusive) suggests a realm of testable mechanisms by which immune specificity could be generated (Fig. 3b). Restricted expression of unidirectional tandem encoded genes has been seen previously in the olfactory receptors, in which exclusive expression is achieved among clusters even located in separate genomic loci [36].

Conclusions

Consideration of alternative forms of immunity in only four of the ~30 extant animal phyla has provided many new insights into the evolutionary innovations that shape animal immune system. The genomes of early deuterostomes encoded homologs of the principal components of DNA damage repair systems that have become integral to the rearrangement, junctional diversification and other genetic variation in the Ig/TCR systems in the modern jawed vertebrates. Furthermore, a divergent Rag1/2-like cluster homolog has been characterized in the sea urchin (an echinoderm) [37] and new evidence is emerging regarding a transposon origin for Rag1 and recombination signal sequences [38]. Connecting receptor structures to

various types of recombination signal sequences can occur in a limited number of steps and these new findings may help to more clearly define the evolutionary pathway to rearranging genes in jawed vertebrates. The emergence of cellular control mechanisms needed to make use of this diversity are likely a much more complicated matter. Here a comparison of the pathways in which VLR and Ig/TCR diversity are integrated into potentially common lymphocyte mechanisms may underscore a shared ancestral system.

To the best of our knowledge, Ig and TCR only have roles in immune protection. In contrast, the Dscam system illustrates the separate adoption of a single genetic diversification process for two distinct functions: neuronal guidance and immunity. This particular path of receptor variation is likely to become more commonly observed as we come to recognize the role of differential RNA processing in expanding genome potential. Like Dscam, the VLRLs also may have roles in other developmental processes (C. Amemiya, personal communication).

Predicting which other forms of molecules are targets for diversified immune function is difficult. The variety of molecular forms that are known to function in immune processes in other invertebrates and protochordates is quickly expanding. Some of these, like the novel sea urchin 185/333 immune response genes [39], are possibly the targets of further somatic diversification mechanisms. The vast and evolutionarily diverse array of antimicrobial peptides that are found throughout the animals and plants also emerge as potential candidates for anticipatory immune variation [40]. In addition, lectins, which function throughout metazoan immunity and are often found in extensive families, could be adaptively hyperdiversified in other phyla whose immune systems have not yet been explored in detail [41-43]. The C-type lectin fold already has been shown to provide a structural matrix for extraordinarily high levels of genetic variation in a bacteriophage system [44].

Our views of immunity are drawn from our current working knowledge of relatively few organisms. These organisms have been chosen based on qualities such as short life span, small size and small genome that enhance their usefulness as genetic models but also may significantly bias our view of animal immune complexity. Genetic diversification, clonal commitment/allelic exclusion, clonal expansion and memory are integral features of adaptive immunity in jawed vertebrates and likely are characteristics of immunity in jawless vertebrates. Some of these features may be characteristic of analogous systems in other animal groups, although they could be integrated and/or regulated in manners that differ markedly from our current concepts of mammalian adaptive responses. For example, diversification could be reduced to only a few isoforms of a gene, multiple cell types could express the same receptor, the same cell type could co-express several receptors, memory may not be an obligate requirement in shorter life cycles, etc. However, polymorphism, specificity and some form of clonal expansion likely are shared features of all such molecules/recognition systems.

We know from work in sea urchin and amphioxus that massive expansion and diversification of multigene families of innate receptors can occur. Some of these receptor families appear to have functions that differ in essential ways from what is seen in vertebrates. What we perceive as rather static innate immune recognition systems in some animal groups may be associated with more dynamic recognition mechanisms in others. Notably, the phylogenetic distribution of this type of highly expanded innate system likely indicates the ancestral state from which the early deuterostome and possibly vertebrate systems arose. This suggests that some of the cellular mechanisms employed by vertebrates for regulating adaptive diversity in Ig/TCR or VLRLs may have emerged from systems that were already in place to control the highly complex innate systems in deuterostome ancestors.

It is essential to keep an open mind as we seek a wider understanding of immune diversification and receptor evolution. Investigations of alternative mechanisms of immune recognition have

the potential to reveal how intense evolutionary forces have reassembled existing systems to exploit new mechanisms of receptor diversification. Observations made in the course of work described here suggest strategies to identify and characterize new immune mediators and regulatory pathways that may revolutionize our explanation of how immune systems evolve and increase our knowledge of the many layers of regulation upon which our own immune system is based.

Acknowledgements

We thank B. Pryor for editorial assistance and M. Sexton for help with graphic design. This work is supported by the National Institutes of Health grants R01AI23338 and AI57559 to GWL, a postdoctoral fellowship from the H. Lee Moffitt Cancer Center and Research Institute to LJD, the Canadian Institutes for Health Research grant MOP74667 to JPR and the National Science and Engineering Research Council of Canada grant 458115/211598 to JPR.

Reference List

1. Flajnik MF. Comparative analyses of immunoglobulin genes: surprises and portents. *Nat Rev Immunol* 2002;2:688–698. [PubMed: 12209137]
2. Litman GW, Cannon JP, Dishaw LJ. Reconstructing immune phylogeny: new perspectives. *Nat Rev Immunol* 2005;5:866–879. [PubMed: 16261174]
3. Loker ES, Adema CM, Zhang S-M, Kepler TB. Invertebrate immune systems - not homogeneous, not simple, not well understood. *Immunol Rev* 2004;198:10–24. [PubMed: 15199951]
4. Litman GW, Frommel D, Finstad J, Howell J, Pollara BW, Good RA. The evolution of the immune response. VIII. Structural studies of the lamprey immunoglobulin. *J Immunol* 1970;105:1278–1285. [PubMed: 4099053]
5. Uinuk-Ool T, Mayer WE, Sato A, Dongak R, Cooper MD, Klein J. Lamprey lymphocyte-like cells express homologs of genes involved in immunologically relevant activities of mammalian lymphocytes. *Proc Natl Acad Sci USA* 2002;99:14356–14361. [PubMed: 12391333]
6. Pancer Z, Amemiya CT, Ehrhardt RA, Ceitlin J, Gartland GL, Cooper MD. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* 2004;430:174–180. [PubMed: 15241406] First description of a complex somatic rearranging non-Ig-based immune recognition system outside of jawed vertebrates.
7. Pancer Z, Saha NR, Kasamatsu J, Suzuki T, Amemiya CT, Kasahara M, Cooper MD. Variable lymphocyte receptors in hagfish. *Proc Natl Acad Sci USA* 2005;102:9224–9229. [PubMed: 15964979]
8. Kasamatsu J, Suzuki T, Ishijima J, Matsuda Y, Kasahara M. Two variable lymphocyte receptor genes of the inshore hagfish are located far apart on the same chromosome. *Immunogenet* 2007;59:329–331.
9. Alder MN, Rogozin IB, Iyer LM, Glazko GV, Cooper MD, Pancer Z. Diversity and function of adaptive immune receptors in a jawless vertebrate. *Science* 2005;310:1970–1973. [PubMed: 16373579] In this follow-up description, the authors provide a detailed annotation of VLR diversity and estimate the mode of selection along the predicted antigen binding site, and characterize rearrangement intermediates, and provide evidence for specific VLR up-regulation in response to a bacterial toxin. Data from this work would later strengthen the copy-choice model by Nagawa et al 2007. The original lamprey work, unfortunately, could not relate single animal cDNA studies to a reference genomic library from the same animal.
10. Kim HM, Oh SC, Lim KJ, Kasamatsu J, Heo JY, Park BS, Lee H, Yoo OJ, Kasahara M. Structural diversity of the hagfish variable lymphocyte receptors. *J Biol Chem* 2007;282:6726–6732. [PubMed: 17192264] Crystal structure characterization of Hagfish VLR-A and -B. The authors demonstrate that VLRs contain a typical horseshoe-like solenoid structure characteristic of LRR-containing proteins and support the previous prediction by Alder et al [9] that the integral variable LRR modules form a putative antigen binding site lined with hypervariable amino acids. Conservation of critical structural residues among all VLRs in hagfish and lamprey suggest that this model represents the consensus VLR structure.
11. Nagawa F, Kishihita N, Shimizu K, Hirose S, Miyoshi M, Nezu J, Nishimura T, Nishizumi H, Takahashi Y, Hashimoto S, et al. Antigen-receptor genes of the agnathan lamprey are assembled by a process involving copy choice. *Nature Immunology* 2007;8:206–213. [PubMed: 17187071] First detailed description of the mechanism of VLR rearrangement. The structure of the VLR locus

was confirmed in a second lamprey species, and the authors provide a single animal reference genomic library where diversity of cDNA products, as well as recombination intermediates, could be related to specific genomic clones. Mechanism of rearrangement is revealed to involve a replication-coupled template switching event, and is not dependent on break-repair mechanisms of homologous or site-specific recombination events.

12. Moumen A, Polomack L, Roques B, Buc H, Negroni M. The HIV-1 repeated sequence R as a robust hot-spot for copy-choice recombination. *Nucleic Acids Res* 2001;29:3814–3821. [PubMed: 11557813]
13. Viguera E, Canceill D, Ehrlich SD. Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J* 2001;20:2587–2595. [PubMed: 11350948]
14. d'Alençon E, Petranovic M, Michel B, Noirot P, Aucouturier A, Uze M, Ehrlich SD. Copy-choice illegitimate DNA recombination revisited. *EMBO J* 1994;13:2725–2734. [PubMed: 8013470]
15. Rogozin IB, Iyer LM, Liang L, Glazko GV, Liston VG, Pavlov YI, Aravind L, Pancer Z. Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat Immunol*. 2007doi:10.1038/ni1463
16. Bell JK, Mullen GE, Leifer CA, Mazzoni A, Davies DR, Segal DM. Leucine-rich repeats and pathogen recognition in Toll-like receptors. *Trends Immunol* 2003;24:528–533. [PubMed: 14552836]
17. Meylan E, Tschopp J, Karin M. Intracellular pattern recognition receptors in the host response. *Nature* 2006;442:39–44. [PubMed: 16823444]
18. Martinon F, Tschopp J. NLRs join TLRs as innate sensors of pathogens. *Trends Immunol* 2005;26:447–454. [PubMed: 15967716]
19. Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL. *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 2000;101:671–684. [PubMed: 10892653]
- 20••. Watson FL, Püttmann-Holgado R, Thomas F, Lamar DL, Hughes M, Kondo M, Rebel VI, Schmucker D. Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science* 2005;309:1874–1878. [PubMed: 16109846] Usage of alternatively spliced *Dscam* exons was examined in the brain, fat body and hemocytes of the fruitfly *Drosophila melanogaster*. Fat body cells and hemocytes appear to express overlapping but distinct subsets of the exons that are expressed in the brain; secreted forms of Dscam also were detected in hemolymph serum. Functional experiments showed that certain Dscam isoforms are capable of binding to *Escherichia coli* bacteria and that hemocyte-specific loss impairs the phagocytic uptake of bacteria.
- 21••. Dong Y, Taylor HE, Dimopoulos G. AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. *PLoS Biol* 2006;4:1137–1146. Variation in usage of *Dscam* exons in the mosquito *Anopheles gambiae* was examined in a hemocyte cell line and in whole organisms in response to infectious challenge with different species of bacteria or malaria parasites. The study found that *A. gambiae* responds to challenges by alteration of its hemocyte *Dscam* splice form repertoires, in patterns specific to each different challenge; Dscam proteins in responding organisms also demonstrated increased, challenge pathogen-specific binding. Transient RNAi silencing of *Dscam* compromised resistance to experimental infections and also increased vulnerability to opportunistic infection.
22. 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* 1997;94:8691–8696. [PubMed: 9238039]
23. Zhang S-M, Loker ES. The *FREP* gene family in the snail *Biomphalaria glabrata*: additional members, and evidence consistent with alternative splicing and *FREP* retrosequences. *Dev Comp Immunol* 2003;27:175–187. [PubMed: 12590969]
- 24••. Zhang S-M, Adema CM, Kepler TB, Loker ES. Diversification of Ig superfamily genes in an invertebrate. *Science* 2004;305:251–254. [PubMed: 15247481] Genomic sequences encoding the IgSF1 domain of FREP3 in the snail *Biomphalaria glabrata* were examined in individual organisms. Results indicate that this region of *FREP3* is diversified at the genomic level in individuals, at higher rates than those observed for control genes. Mathematical projections suggest that, in a given individual, all observed *FREP3* sequences (45 in one snail, 37 in a second) could arise from a single set of nine *FREP3* source sequences either by point mutations or by a recombination mechanism that is possibly somatic.

25. Cannon JP, Haire RN, Litman GW. Identification of diversified genes that contain immunoglobulin-like variable regions in a protochordate. *Nat Immunol* 2002;3:1200–1207. [PubMed: 12415263]
- 26•. Cannon JP, Haire RN, Schnitker N, Mueller MG, Litman GW. Individual protochordates possess unique immune-type receptor repertoires. *Curr Biol* 2004;14:R465–R466. [PubMed: 15203016]
Genomic sequences encoding amino acids at the N-terminus of the VCBP2/5 family of proteins were examined in multiple individual specimens of the amphioxus *Branchiostoma floridae*. Sequence comparisons revealed that while certain predicted amino acid sequences across this region were found in nearly all animals examined, no two animals exhibited the same repertoire of predicted VCBP2/5 proteins.
27. De Tomaso AW, Nyholm SV, Palmeri KJ, Ishizuka KJ, Ludington WB, Mitchel K, Weissman IL. Isolation and characterization of a protochordate histocompatibility locus. *Nature* 2005;438:454–459. [PubMed: 16306984]
28. Kürn U, Sommer F, Hemmrich G, Bosch TCG, Khalturin K. Allorecognition in urochordates: identification of a highly variable complement receptor-like protein expressed in follicle cells of *Ciona*. *Dev Comp Immunol* 2007;31:360–371. [PubMed: 17074389]
- 29••. Hernandez Prada JA, Haire RN, Allaire M, Jakoncic J, Stojanovic N, Cannon JP, Litman GW, Ostrov DA. Ancient evolutionary origin of diversified variable regions revealed by crystal structures of an amphioxus VCBP. *Nat Immunol* 2006;7:875–882. [PubMed: 16799561] The structure of the tandem variable domains of *B. floridae* VCBP3 was solved to 1.85 Å resolution; a second structure, of the N-terminal V region alone, also was solved to 1.15 Å. The two VCBP V domains are oriented in a head-to-tail manner, in contrast to the head-to-head orientation typically found in dimerized V domains of Ig or TCR. This orientation displaces the regions of VCBP3 that would correspond to the complementarity determining regions of Ig/TCR and suggests an alternative binding surface for potential VCBP ligands.
30. Rock FL, Hardiman G, Timans JC, Kastelein RA, Bazan JF. A family of human receptors structurally related to *Drosophila* Toll. *Proc Natl Acad Sci USA* 1998;95:588–593. [PubMed: 9435236]
31. Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Davidson B, Smith KD, Hood L, Aderem A. The evolution of vertebrate toll-like receptors. *Proc Natl Acad Sci USA* 2005;102:9577–9582. [PubMed: 15976025]
32. Hibino T, Loza-Coll M, Messier C, Majeske AJ, Cohen AH, Terwilliger DP, Buckley KM, Brockton V, Nair SV, Berney K, et al. The immune gene repertoire encoded in the purple sea urchin genome. *Dev Biol* 2006;300:349–365. [PubMed: 17027739]
33. Pancer Z, Cooper MD. The evolution of adaptive immunity. *Ann Rev Immunol* 2006;24:497–518. [PubMed: 16551257]
- 34••. Rast JP, Smith LC, Loza-Coll M, Hibino T, Litman GW. Genomic insights into the immune system of the sea urchin. *Science* 2006;314:952–956. [PubMed: 17095692] This review lends perspective to comprehensive immune findings from the purple sea urchin genome (described in detail in [32]) including an innate immune gene repertoire of unprecedented multiplicity and gene homologs of important relevance to the vertebrate adaptive system.
35. Fritz JH, Ferrero RL, Philpott DJ, Girardin SE. Nod-like proteins in immunity, inflammation and disease. *Nat Immunol* 2006;7:1250–1257. [PubMed: 17110941]
36. Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, Axel R. Interchromosomal interactions and olfactory receptor choice. *Cell* 2006;126:403–413. [PubMed: 16873069]
- 37••. 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* 2006;103:3728–3733. [PubMed: 16505374] A *Rag1/2*-like gene cluster is identified from the purple sea urchin genome. Genomic organization, transcript sequence, expression and protein-level data is presented, providing new experimental opportunity for inference into the evolutionary origins of V(D)J recombination.
- 38•. Kapitonov VV, Jurka J. RAG1 core and V(D)J recombination signal sequences were derived from transib transposons. *PLoS Biol* 2005;3:e181. [PubMed: 15898832] The *transib* class of transposable elements exhibits significant similarity to Rag1-core sequence and are associated with RSS-like terminal inverted repeats, providing the first direct evidence of a homologous relation between transposable elements and Rag mediated recombination.
39. 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* 2007;8(16):16. [PubMed: 17331248]
40. Bulet P, Stocklin R, Menin L. Anti-microbial peptides: from invertebrates to vertebrates. *Immunol Rev* 2004;198(16984):169–184. [PubMed: 15199962]
41. Schwarz RS, Hodes-Villamar L, Fitzpatrick KA, Fain MG, Hughes AL. A gene family of putative immune recognition molecules in the hydroid *Hydractinia*. *Immunogenet* 2007;59:233–246.
42. Quesenberry MS, Ahmed H, Elola MT, O'Leary N, Vasta GR. Diverse lectin repertoires in tunicates mediate broad recognition and effector innate immune responses. *Integr Comp Biol* 2003;43:323–330.
43. Zhu Y, Ng PM, Wang L, Ho B, Ding JL. Diversity in lectins enables immune recognition and differentiation of wide spectrum of pathogens. *Int Immunol* 2006;18:1671–1680. [PubMed: 17035349]
44. McMahan SA, Miller JL, Lawton JA, Kerkow DE, Hodes A, Marti-Renom MA, Doulatov S, Narayanan E, Sali A, Miller JF, Ghosh P. The C-type lectin fold as an evolutionary solution for massive sequence variation. *Nature Struct Mol Biol* 2005;12:886–892. [PubMed: 16170324]
45. Canceill D, Ehrlich SD. Copy-choice recombination mediated by DNA polymerase III holoenzyme from *Escherichia coli*. *Proc Natl Acad Sci U S A* 1996;93:6647–6652. [PubMed: 8692872]
46. Mondragon-Palomino M, Gaut BS. Gene conversion and the evolution of three leucine-rich repeat gene families in *Arabidopsis thaliana*. *Mol Biol Evol* 2005;22:2444–2456. [PubMed: 16120808]
47. Parniske M, Hammond-Kosack KE, Golstein C, Thomas CM, Jones DA, Harrison K, Wulff BB, Jones JD. Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. *Cell* 1997;91:821–832. [PubMed: 9413991]

Figure 1.

Schematic representation of the assembly of a mature VLR gene [11]. Most VLR genes exist as empty cassettes with 5' and 3'-encoding LRR genetic segments (blue) and several kilobases of non-coding intervening sequence (a). In somatically expanding cell lineages a single VLR gene begins to assemble (possibly coupled to another unknown event) via a replication-dependent copy-choice mechanism (also known as template switching), which replaces the intervening non-coding sequence with LRR-encoding segments from elsewhere in the locus. This process may occur on either (or both) strand(s) during replication. Copy-choice does not involve break-repair mechanisms associated with other forms of recombination, but instead relies on several structural features of the locus [12;13] that can influence the replicating ends

of DNA. In these cases, the advancing replication fork is susceptible to secondary structures [14;45], which bring units of DNA into close enough proximity that if short segments of similarity exist, template-switching may occur. At this point, the advancing DNA polymerase pauses, disassociates and reanneals to another segment sharing compatible stretches of similarity (in this case, encoding a LRR). Although regions of sequence similarity are required for template switching, this mechanism is distinct from homologous recombination. In the VLRs, the stretches of similarity exist most often at or near the ends of the LRRs (a-b, colored diagonal ends). Template switching across the VLR gene (b) replaces the intervening sequence with numerous LRR segments, producing a final gene product (c) that is capped with conserved 5' and 3' LRR segments (blue) and a variable number of LRR segments that vary in sequence and size. Diversity and polymorphism of the mature VLRs can be achieved by the random incorporation of LRRs (random choice of homology segments), which also leads to a variation in the cross-over (template-switching) sites and may produce variation in LRR size. Mismatches in the switching sites can also be incorporated as polymorphisms in the final product. Because the VLR locus is rich with scrambled LRRs, it remains possible that the individual LRR segments throughout the locus can undergo inter- and intra-allelic gene conversion to further diversify some segments [46;47]. Although not shown to scale in the illustration, the lengths of the LRR segments can vary, as well as the sizes of the short similarity segments (depicted as diagonal colored ends). Variable LRR segments used in the assembly are colored in shades of green; template-switching can occur within the short similarity segments depicted as diagonal ends; diagonal colored ends and the corresponding colored bars represent overlap and do not imply major sequence changes; scrambled LRR segments not used in the assembly of this gene are depicted as black squares in (a). NT-LRR, invariant N-terminal LRR (blue); CT-LRR, invariant C-terminal LRR (blue); CP, connecting peptide found associated with the final LRR (red bar); SP, signal peptide; GPI, glycosyl-phosphatidylinositol membrane anchor.

Figure 2.

Diversification of Dipteran *Dscam* genes by differential mRNA processing, extrapolated from studies in *Drosophila melanogaster* and *Anopheles gambiae*. A representative *Dscam* gene (*D. melanogaster Dscam*, which contains 12 copies of exon 4 [cyan], 48 copies of exon 6 [red], 33 copies of exon 9 [green] and 2 copies of exon 17 [magenta]) is depicted at top. Mutually exclusive alternative splicing (indicated by asterisk, *) of Ig domain-encoding exons 4, 6 and 9, as well as transmembrane segment-encoding exon 17, results in the generation of an enormous array of diversified *Dscam* transcripts (approaching 40,000 possible forms in *D. melanogaster* and over 30,000 forms in *A. gambiae*). At least three different cell types (neurons, hemocytes and fat body cells) express distinctive repertoires of *Dscam* ectodomain isoforms in *Drosophila*. Observations in *A. gambiae* suggest that the diversification of *Dscam* proteins in hemocytes is influenced by the particular microbial organism used in challenge experiments, indicating that the hemocyte *Dscam* repertoire may be tailored *in vivo* to different potential pathogens. The precise identities of the alternatively spliced exons depicted in the processed *Dscam* transcripts at bottom are not meant to imply functional or representational significance of any particular exon. Exons are not drawn to scale. The complete repertoires of *Dscam* transcripts in the three cell types shown, as well as variation of these repertoires among organisms in a population, are not yet known.

Figure 3.

Expansion of innate immune receptors. TLRs in most animals are present as small gene families; however in the sea urchin and at least one other invertebrate deuterostome, they have undergone enormous expansions in gene number. Diversity among these genes suggests a dynamic and rapidly evolving system that is more similar to some immune receptors of vertebrate adaptive immunity than to vertebrate or insect TLR gene families. (a) An evolutionary innovation confers a modified immune function (*) and is expanded into a multigene family of increasing recognition potential; many of the single exon genes are encoded in tandem genomic clusters. (b) The mechanisms by which this diversity is exploited by the sea urchin are the subject of ongoing investigation; however a range of possibilities include expression of many gene types on each cell, expression of overlapping but distinct gene subsets on each cell, or completely exclusive expression of a single TLR allele. These possibilities vary in the level of specificity that can be generated in an immune response.