

Origin and evolution of the adaptive immune system: genetic events and selective pressures

Martin F. Flajnik* and Masanori Kasahara†

Abstract | The adaptive immune system (AIS) in mammals, which is centred on lymphocytes bearing antigen receptors that are generated by somatic recombination, arose approximately 500 million years ago in jawed fish. This intricate defence system consists of many molecules, mechanisms and tissues that are not present in jawless vertebrates. Two macroevolutionary events are believed to have contributed to the genesis of the AIS: the emergence of the recombination-activating gene (RAG) transposon, and two rounds of whole-genome duplication. It has recently been discovered that a non-RAG-based AIS with similarities to the jawed vertebrate AIS — including two lymphoid cell lineages — arose in jawless fish by convergent evolution. We offer insights into the latest advances in this field and speculate on the selective pressures that led to the emergence and maintenance of the AIS.

Somatic hypermutation

Mutation of the variable gene after mature B cells are stimulated. It results in affinity maturation of the antibody response. Like the class switch, it requires activation-induced cytidine deaminase.

The adaptive immune system (AIS) is fascinating to both scientists and laymen: we have a specific yet incredibly diverse system that can fight myriad pathogens and has a ‘memory’ — the basis of vaccination — that enables a rapid response to previously encountered pathogens. The complexity of immune response regulation rivals that of the nervous system in terms of the connections forged and suppressed between cells, but immune cells must also traverse the body through blood, lymph and tissue until they encounter invading organisms. How did such a system arise, and can studies of non-mammalian vertebrates help us to understand the immunity gestalt?

Antibodies were discovered over 100 years ago, and major questions relating to the generation of diversity were solved in the 1970s with the detection of somatic hypermutation¹ and variable–diversity–joining rearrangement (VDJ rearrangement)² of antibody (or immunoglobulin (Ig) or B cell receptor (BCR)) genes. In the 1980s, T cell receptors (TCRs) were discovered and there was universal agreement that they shared a common ancestor with BCR genes, based on their similar domain organization and reliance on the same rearrangement mechanism to generate diversity³. After the discovery of enzymes that are involved in the rearrangement of BCR and TCR genes⁴ and of hypermutation of BCR genes⁵, attention shifted to asking how a system that is capable of generating such diversity evolved. Jawed fish were found to have almost all of these genes and

mechanisms, but jawless fish (Agnathans) apparently had none. This mystifying finding led to the ‘big bang’ theory of AIS emergence⁶, which is one of the main topics of this Review.

The discovery in jawless fish of a lymphoid cell-based system of adaptive immunity that is strikingly similar to the system in jawed fish was a total surprise⁷. The system in jawless fish is similar to VDJ rearrangement but has rearranging receptors that are encoded by another gene family — the variable lymphocyte receptor (VLR) family. We discuss the compelling possibility that the VLR rearrangement process is under the control of the same family of enzymes that are responsible for somatic hypermutation of BCR genes⁸, which pushes the origins of acquired immunity back to the ancestor of all vertebrates. However, despite this new finding of adaptive immunity in jawless fish, the ‘big bang’ theory concerning the origin and rapid development of the highly complex BCR- and TCR-based AIS remains intact.

In this Review, we describe the basic features of BCRs, TCRs and the major histocompatibility complex (MHC) and stress their preservation and modification over evolutionary time. We then discuss the two catastrophic, innovative events behind the ‘big bang’ origins of the BCR–TCR–MHC-based AIS — the recombination-activating gene (RAG) transposon invasion and genome-wide duplications early in vertebrate history — and we discuss the new type of AIS that has been discovered in jawless fish. Finally, we speculate on the origins of

*Department of Microbiology and Immunology, University of Maryland at Baltimore, 655 West Baltimore Street, Biosciences Research Building 3-056, Baltimore, Maryland 21201, USA.

†Department of Pathology, Hokkaido University Graduate School of Medicine, North 15 West 7, Sapporo 060-8638, Japan.

e-mails: mflajnik@som.umaryland.edu; mkasaha@med.hokudai.ac.jp

doi:10.1038/nrg2703

Published online

8 December 2009

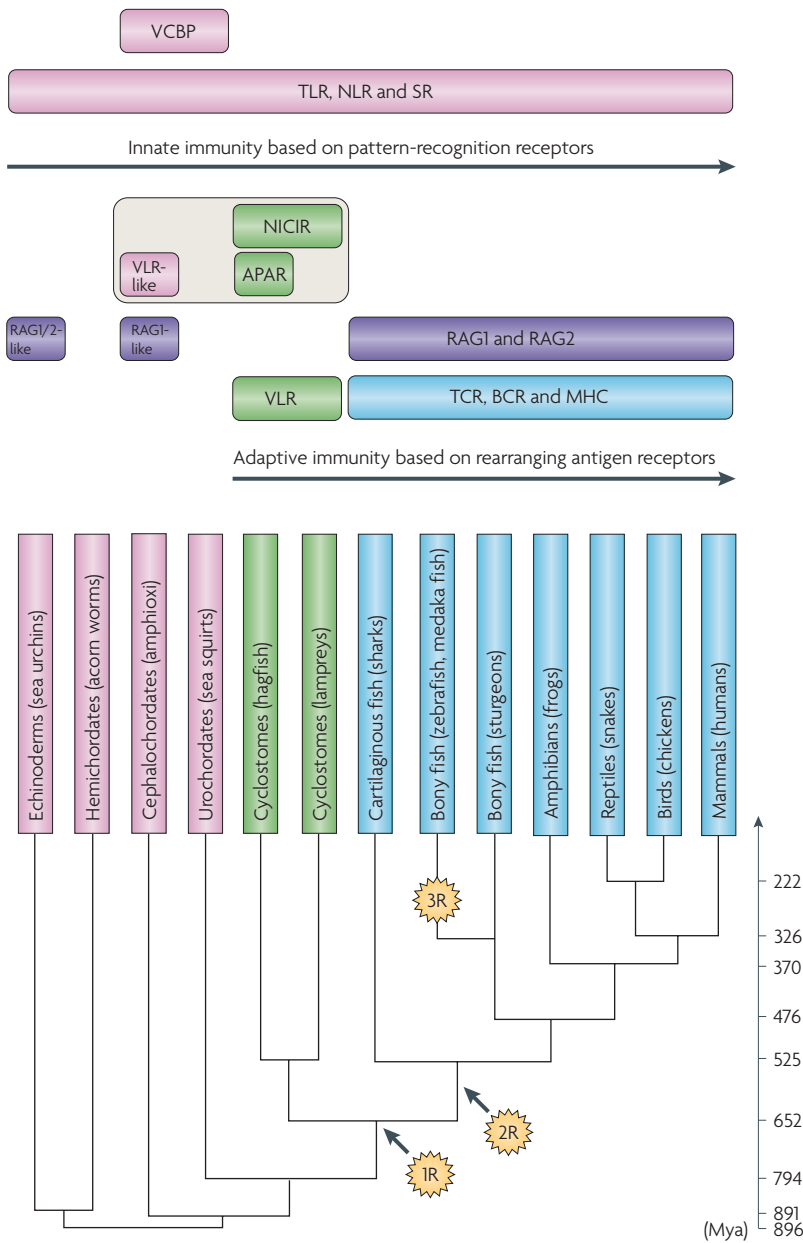


Figure 1 | Overview of the evolution of the immune system in deuterostomes. The stages in phylogeny at which the immune molecules referred to in this Review emerged. Molecules restricted to jawed and jawless vertebrates are indicated in blue and green, respectively. Molecules that emerged at the stage of invertebrates are in pink. Recombination-activating gene (RAG)-like genes (indicated in purple) are of viral or bacterial origin (from the transib transposon family) and are also present in the genomes of sea urchins and amphioxii. Agnathan paired receptors resembling antigen receptors (APAR) and novel immunoreceptor tyrosine-based activation motif-containing immunoglobulin superfamily receptor (NICIR, also known as T cell receptor (TCR)-like) are agnathan immunoglobulin superfamily (IgSF) molecules that are thought to be related to the precursors of TCRs and B cell receptors (BCRs). 1R and 2R indicate the two rounds of whole-genome duplication (WGD). Whether the 2R, the second round of WGD, occurred before or after the divergence of jawed and jawless vertebrates is controversial; the figure places it after the divergence according to the commonly held view (BOX 2). An ancestor of the majority of ray-finned fish is thought to have experienced an additional, lineage-specific WGD (designated as 3R) ~320 million years ago^{134,135}. The divergence time of animals (shown in Mya (million years ago)) is based on Blair and Hedges¹³⁶. MHC, major histocompatibility complex; NLR, Nod-like receptor; SR, scavenger receptor; TLR, Toll-like receptor; VCBP, V-region containing chitin-binding protein; VLR, variable lymphocyte receptor.

adaptive immunity in all vertebrates, emphasizing the lessons we have learned from extant immune systems.

The AIS of jawed vertebrates

The most evolutionarily ancient extant organisms in which the AIS, as defined in humans, is found are the cartilaginous fish. It is believed to have arisen in the first jawed vertebrates (gnathostomes) — the placoderms (FIG. 1). Components of the innate immune system, for example, pattern-recognition receptors (PRRs) — including the Toll-like receptors (TLRs), Nod-like receptors (NLRs) and scavenger receptors (SRs) — are found throughout the animal kingdom (FIG. 1). In some invertebrates, such as the sea urchin⁹, and in many plants¹⁰ there has been a great expansion of PRR families, which suggests that in the absence of an AIS, complex innate mechanisms might be required for defence.

Lymphoid cells — the core of the AIS — are first found in pre-vertebrate deuterostomes (that is, these are the most evolutionarily ancient living phylogenetic group in which lymphoid cells can be found). The AIS requires a large population of cells in each individual to permit clonal selection of cells with receptors specific for a particular pathogen. Large-bodied vertebrates were the first organisms to fulfil this requirement (FIG. 1). Here, we discuss important features of the AIS in jawed vertebrates and consider their conservation or plasticity over evolutionary time. The concept of plasticity was first proposed by Klein and colleagues in the context of MHC molecules¹¹, but can be extended to all features of the AIS.

B cell receptors

Immunoglobulin M. Like most molecules of the AIS as defined in humans, among living animals Igs are first found in cartilaginous fish (sharks, skates, rays and chimaeras or ratfish) and are generated by a somatic recombination mechanism. Their evolutionary features are summarized in FIG. 2a. IgM is the most ancient antibody class and has the same function in all gnathostomes¹². The transmembrane form of IgM defines the B cell lineage. After B cell stimulation — by an antigen and cognate interactions from T cells — this isotype is secreted into the plasma as a multimeric protein. In bony fish, IgM is secreted as a tetramer in different reduction or oxidation states that seem to modify the binding strength of the antibodies¹³. In cartilaginous fish, IgM is an abundant secreted monomer; this secretion correlates with a switch to a higher affinity response¹⁴. It is not known whether a cell programmed to produce multimeric IgM switches to produce the monomeric form or whether sharks have lineages of B cells that are specialized to make either the multimeric or monomeric Ig. Despite these modifications in fish, IgM is generally evolutionarily conserved.

Immunoglobulin D. The function of IgD remains poorly understood. It had generally been thought to be an evolutionary newcomer, probably arising in mammals (FIG. 2a), but the detection in bony fish¹⁵ and frogs^{16,17} of an Ig isotype that resembles IgD showed that this is not the case

(FIG. 2b). Furthermore, phylogenetic analyses suggest that another isotype in cartilaginous fish and some bony fish, IgW^{18–20}, is related to IgD. Therefore, an IgD-like isotype was present in the ancestor of all gnathostomes. In frogs and mammals, IgD/IgW occurs predominantly in the transmembrane form, but in bony fish there are high levels of secreted IgD, which is mostly bound to the surface of granulocytes. This work in fish prompted a recent re-examination of the function of IgD in humans, and it was found that the secreted form of IgD is bound to the surface of basophils and has strong proinflammatory properties²¹. Therefore, although this new study does not address the function of cell-surface IgD, the work in fish heralded an entirely new understanding of secretory IgD

function in humans. Like IgM, IgD is an ancient antibody class, but unlike IgM, it has been extraordinarily malleable in structure and (presumably) in function over evolutionary time.

Other immunoglobulins. Besides IgM and IgD, mammals have three Ig isotypes: IgG, which is involved in high-affinity memory responses; IgE, which functions in inflammatory (and allergic) responses at epithelial surfaces; and the mucosal antibody IgA (FIG. 2a). IgG and IgE originate from an isotype known as IgY, which is first found in amphibians and has the same function as IgG (at least in frogs and birds^{17,22}). Shorter forms of IgY, which are formed either by alternative splicing or

Variable–diversity–joining rearrangement
(VDJ rearrangement.) The recombination-activating gene (RAG)-mediated ligation of T cell receptor or B cell receptor variable (V), diversity (D) and joining (J) gene segments during lymphocyte ontogeny, which generates the antigen receptor repertoire.

Immunoglobulins
Also known as antibodies or B cell receptors, they are composed of two identical heavy (H) and light (L) chains that are covalently linked by disulphide bonds. Monomeric immunoglobulins are bivalent, and the binding site is made up of the amino-terminal variable domains of one H and one L chain. The class or isotype of an immunoglobulin is defined by its H chain.

Jawless fish
Primordial vertebrates without jaws, of which the only two extant forms are lampreys and hagfish.

Major histocompatibility complex
A large complex of tightly linked genes, many of which are involved in immunity. It encodes the polymorphic class I and class II molecules, which present antigens in the form of peptides to cytotoxic and helper T cells, respectively.

Jawed vertebrates
Vertebrates from cartilaginous fish to mammals. The first class of vertebrates with jaws, the placoderms, all became extinct.

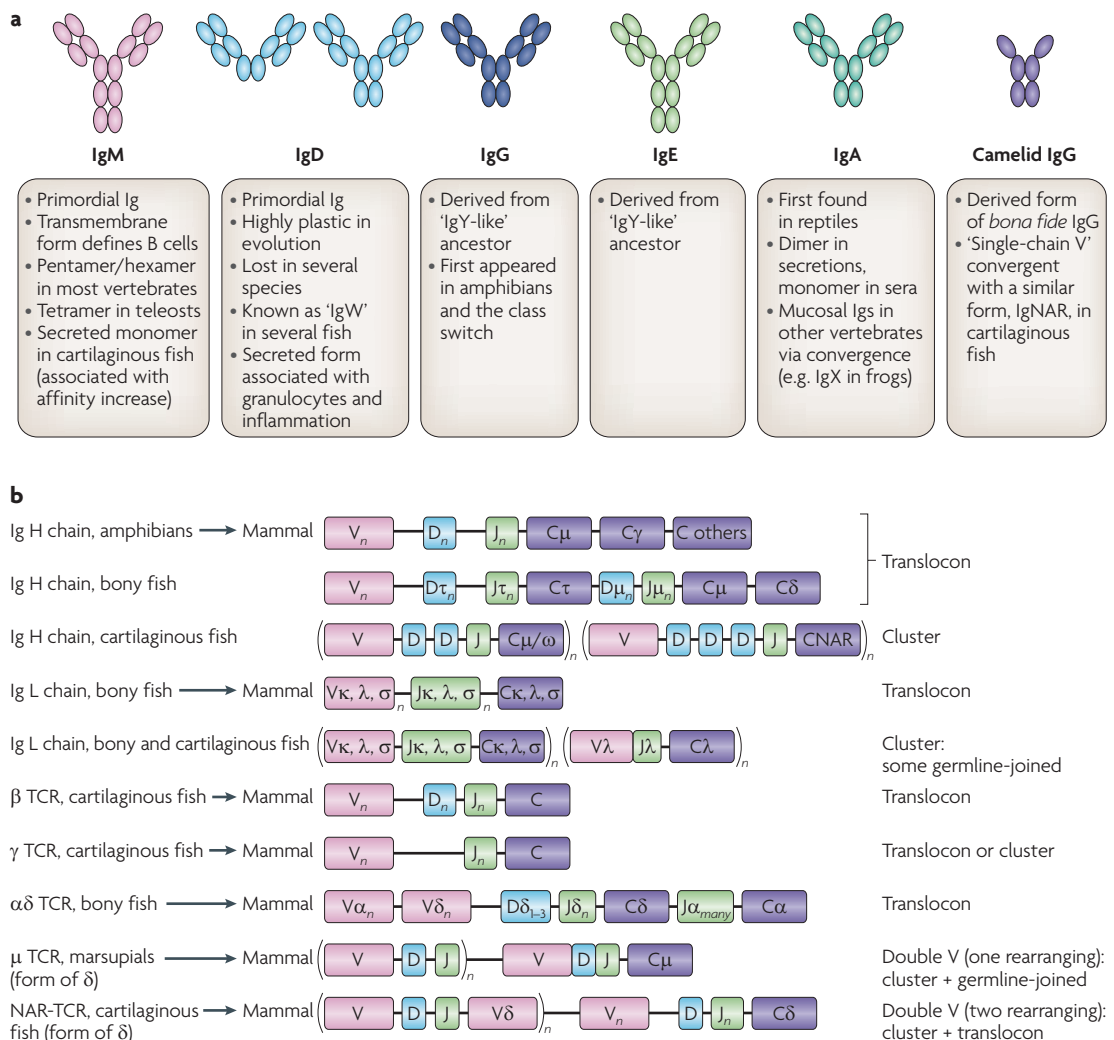


Figure 2 | Antigen receptor proteins and genes in jawed vertebrates. **a** | The different B cell receptor (BCR) isotypes in mammals and their features in other vertebrates. Each oval represents an immunoglobulin superfamily (IgSF) domain consisting of 90–100 amino acids. Immunoglobulin D (IgD) is represented in two forms, mouse (left) and human (right); note that even in mammals the IgD structure is highly plastic. Major phylogenetic features are described in the text. **b** | BCR genes (which are defined by their heavy (H) chain) and T cell receptor (TCR) genes throughout phylogeny. Recombination signal sequences are located at the 3' end of all variable (V) loci, and the 5' end of joining (J) segments and on both sides of diversity (D) segments. Note that in both the bony fish Ig H chain and vertebrate αδ TCRs, entire antigen receptor families are deleted upon rearrangement of the V segment to the downstream D and J segments (τ genes are deleted in bony fish and δ is deleted at the TCR locus). Note also that, despite its name, μ TCR in marsupials is related to δ TCR and not the IgM constant region. L chain, light chain; NAR, new antigen receptor.

Mucosal immunity

Immune responses made across epithelial surfaces, such as the gut and lung.

Immunoglobulin new antigen receptor

(IgNAR.) A specialized antibody in sharks that is composed of disulphide-linked heavy chains and no associated light chains.

Translocon organization

An organization of immunoglobulin genes in which there are multiple variable (V), diversity (D) and joining (J) segments upstream of a single constant gene.

Cluster organization

An organization of immunoglobulin genes in which there are single variable (V), diversity (D), joining (J) and constant gene segments, although sometimes two or three diversity segments occur together. Also known as minilocus organization.

Receptor editing

The re-rearrangement of antigen receptor genes to avoid self reactivity in the case of B cells or to express new receptors that can be positively selected in the case of T cells.

Class switch

The rearrangement of an existing variable–diversity–joining (VDJ) exon from the 5' end of the immunoglobulin M gene to downstream constant (C) genes. Like somatic hypermutation, it requires activation-induced cytidine deaminase.

Major histocompatibility complex restriction

T cell receptors recognize antigens in the form of small peptides ranging from 9–22 amino acids that are bound to major histocompatibility complex (MHC) class I or class II molecules. The T cell receptor recognizes both the MHC protein and the peptide antigen — this is known as 'MHC restriction.'

from separate genes, are present in multiple taxa; these short forms are believed to neutralize pathogens without activation of inflammatory responses. IgA is found first in reptiles and seems to serve the same functions in all animals. Interestingly, in lower taxa, other isotypes — for example, IgX in frogs²³ — provide mucosal immunity. Finally, in camelids, there is a modified form of IgG that does not associate with light (L) chains, and its V regions are free to interact with antigens; a similar form called immunoglobulin new antigen receptor (IgNAR) is found in cartilaginous fish; the IgNAR V region is also conserved in some TCRs^{24,25} (see below). In summary, in evolutionary terms all vertebrates have multiple isotypes that provide distinct effector functions. Amphibians are the most primitive class, with a classical memory response that is based on isotype switching to IgG and IgY after antigen stimulation.

B cell receptor light chains. It was previously believed that the two isotypes of the BCR L chains in humans — λ and κ — arose late in evolution, probably in the mammalian lineage (FIGS 2,3a). However, genome and EST projects and studies of model vertebrate species have shown that these two isotypes emerged in the common ancestor of all living vertebrates^{26–29}. Furthermore, another L chain — identified in frogs and named σ ³⁰ — also arose early in evolution and has been retained in all cold-blooded gnathostomes. The reason(s) for the emergence and maintenance of multiple L chains is (are) not known, but it might be related to the formation of specific binding sites by particular combinations of L chains and heavy (H) chains.

B cell receptor heavy chains. From bony fish to mammals, BCR H chain genes are arranged in a so-called translocon organization (FIGS 2b,3a): multiple V genes occur upstream of multiple D and J segments, followed by single genes that encode the different constant (C) regions. The primary H chain repertoire is generated by rearrangement in the primary lymphoid tissues (for example, bone marrow), which involves a single V gene (out of tens to hundreds of V genes) being rearranged to the downstream D and J segments³¹. In some species, such as chickens and rabbits, after the rearrangement of a single V gene, gene conversion events involving nearby non-functional V genes generate diversity in the primary repertoire^{32,33}. In cartilaginous fish, the Ig genes are in a cluster configuration with a single V, D, J and C gene in each cluster³⁴. RAG-mediated rearrangement occurs within a cluster, but diversity is generated by junctional diversity at the V–D, D–D and D–J rearrangement joints and by somatic hypermutation after antigenic stimulation of B cells^{35–37}. The situation is similar for the L chain genes, except they lack D segments (FIG. 2b). Also, in bony fish the L chain genes can occur in translocon organization or cluster organization; this suggests that extensive receptor editing — that is, re-rearrangement of BCR genes — can occur for L chain genes in all fish^{38–41} (FIG. 2b).

The H chain class switch first evolved in amphibians, which have a mammalian-like Ig H chain locus. In bony

fish, the Ig H chain isotypes IgM and IgD are generated by alternative splicing. The recently discovered IgT isotype is generated by alternative rearrangement: the IgT D, J and C gene segments are located between the V genes and the IgM and IgD D, J and C segments^{42,43} (FIG. 2b), and rearrangement that joins V genes to the IgM D segments deletes the IgT genes, a situation similar to that at the α D TCR locus. There is no switching apparent between clusters in cartilaginous fish, neither within the IgM class nor between the IgM and cartilaginous fish-specific IgNAR/IgW classes.

Amphibians, which have class switching and IgY, mark a major transition from the AIS in fish. A single B cell has the capacity to modify the effector function of its secreted molecule and to provide a memory response. In sharks, a 'switch' at the biochemical level from multimeric to monomeric IgM provides at least some of the functions of an IgG — for example, the capacity to enter extravascular sites — and an increase in affinity, but the memory response is clearly inferior to that of the higher vertebrates.

T cell receptors

$\alpha\beta$ T cell receptors. $\alpha\beta$ TCRs recognize both the antigen (in the form of a peptide) and the MHC class I or II molecule to which it is bound. This is known as major histocompatibility complex restriction (MHC restriction)⁴⁴. T cells develop in the thymus, and the basic evolution of the thymus and T cell development is conserved for all gnathostomes, except in some species in which the system seems to have degenerated (FIG. 1 and see below). A translocon organization of both the α and β TCR genes is found in all animals (FIG. 2b). The presence of a D segment in the β TCR locus and the absence of D segments and the presence of a large number of J segments in the α TCR locus are also evolutionarily conserved, at least from mammals to bony fish⁴⁵ (FIG. 2b). The many J segments in the α locus allow extensive receptor editing in developing T cells, resulting in greater opportunities for modification of the antigen receptor, followed by potential selection of self-MHC-restricted TCRs in the thymus⁴⁶.

$\gamma\delta$ T cell receptors. $\gamma\delta$ TCRs are quite different from $\alpha\beta$ TCRs. The basic gene organization is conserved in gnathostomes, with only J segments in the γ gene, usually two D segments in the δ gene, and a close linkage of the δ and α TCR genes. However, the receptor can be adapted in several ways. It was shown in mice that $\gamma\delta$ T cells that arose early in ontogeny bear invariant receptors and home to sites, such as the skin, where they form a first line of defence and self-renew for the life of the organism⁴⁷. Similarly, in humans, T cells with canonical $\gamma\delta$ receptors make up a large percentage of lymphocytes in the blood and become highly expanded upon infection. These cells respond to the metabolite isoprenylphosphate, which is produced by all cells but is produced in a more immunogenic form by unicellular pathogens⁴⁸. The 'innate-like' $\gamma\delta$ T cells can make up the majority of T cells in some species, but in others $\gamma\delta$ T cells do not bear canonical receptors,

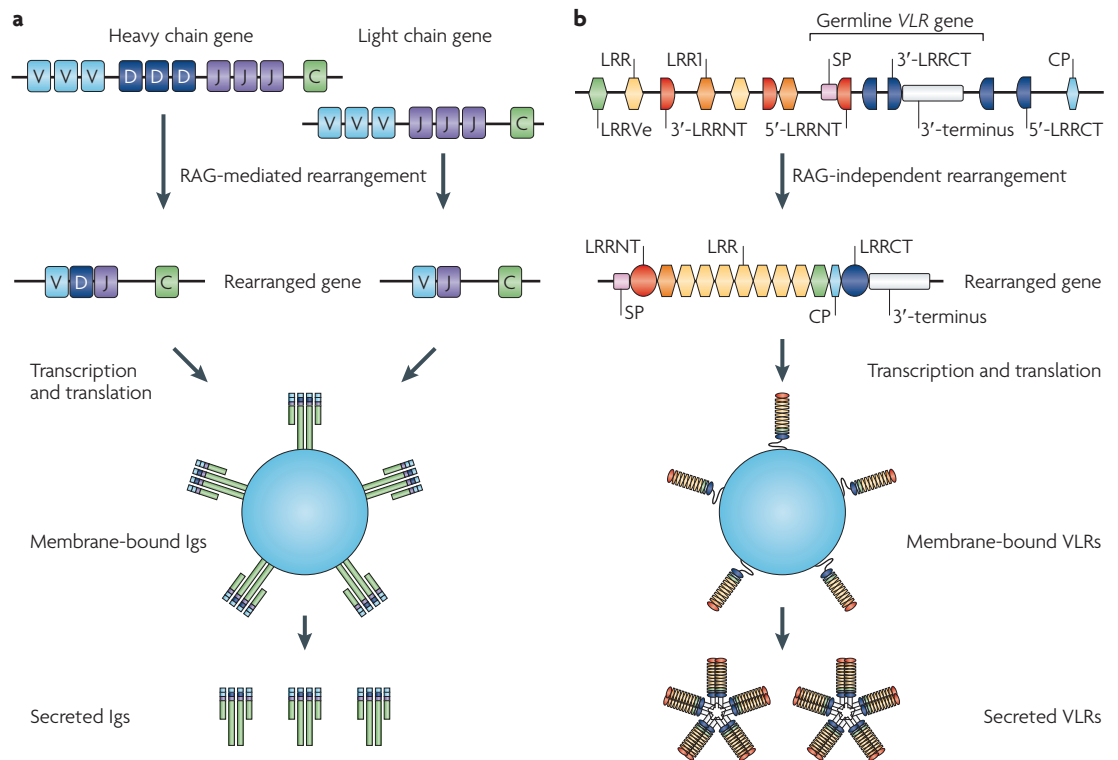


Figure 3 | Two distinct forms of adaptive immunity in vertebrates. a | Immunoglobulin (Ig) genes in jawed vertebrates. Ig genes generate repertoire diversity by recombining variable (V), diversity (D) and joining (J) gene segments with the help of the recombination-activating gene (RAG) recombinase. D gene segments occur only in the heavy chain genes. T cell receptors generate diversity in a similar manner (not shown). They occur only in a membrane-bound form. **b** | Variable lymphocyte receptors (VLRs) in jawless vertebrates. Multiple amino-terminal leucine-rich repeat (LRRNT)-, leucine-rich repeat (LRR)-, connecting peptide (CP)- and carboxy-terminal LRR (LRRCT)-encoding modules are located adjacent to the germline VLR gene. During the development of lymphoid cells, these modules are incorporated into the VLR gene. The rearranged VLR gene encodes a membrane-bound protein. VLRA has a secreted form and functions as an antibody, whereas VLRA apparently occurs only in a membrane-bound form. Whether membrane-bound VLRs occur as monomers or multimers is not known. LRR1 and LRRVe denote LRR modules located at the N- and C-termini, respectively. The organization of the VLR locus shows minor variations depending on loci and species. The figure is intended to emphasize the essential features of VLR assembly and does not accurately reproduce the organization of a specific VLR locus. SP, signal peptide.

are a large proportion of the T cells in the body, and seem to be more adaptive in nature⁴⁹. The binding sites of these receptors (even of some of the innate receptors) are more like BCR binding sites: they have one long and one short complementarity-determining region 3 (CDR3), whereas $\alpha\beta$ TCRs have similarly sized CDR3s for both chains⁵⁰.

Recently, new types of δ TCR genes have been discovered in different vertebrate taxa. In about a quarter of their δ TCR genes, sharks have a three-domain receptor chain with two amino-terminal V domains and a membrane-proximal C domain called NAR-TCR⁵¹ (FIG. 2b). Both V domains are generated by VDJ rearrangement: the N-terminal domain is encoded by an IgNAR-like cluster, and the other is generated by rearrangement of V to canonical D and J elements. It is likely that this TCR chain recognizes antigens in a similar way to the *bona fide* IgNAR, which is presumed to have a single-domain binding site. In marsupials, the δ TCR locus has been duplicated. An apparently ancient *trans* rearrangement of IgV regions to this locus

in germ cells⁵² has resulted in one set of V, D and J segments that rearrange in a cluster that is upstream of a pre-rearranged (that is, germline-joined) VDJ gene. This is predicted to result in a three-domain TCR chain that has two V domains and one C domain and is similar to the TCR chain in sharks (FIG. 2b). Although the function of these receptors is not known, it is proposed that — unlike the MHC-restricted $\alpha\beta$ TCRs — they recognize soluble antigens, which would give them a similar function to Igs⁵³. Consistent with this idea, shark γ TCR genes have recently been shown to hypermutate like BCR genes but unlike $\alpha\beta$ TCR genes⁵⁴. Therefore, a new paradigm emerging from the comparative studies is that $\gamma\delta$ TCRs probably have the capacity to interact with free antigens in a similar way to Igs. Additionally, when comparing $\alpha\beta$ TCRs and $\gamma\delta$ TCRs, the former clearly are the more evolutionarily conserved, probably because they are constrained by MHC restriction. By contrast, $\gamma\delta$ TCRs are more evolutionarily labile and have adaptive immune and innate immune functions that differ even between vertebrates in the same class.

Complementarity-determining region 3

A loop in the variable (V) region of the B cell receptor and T cell receptor chains that is encoded by the variable–diversity–joining (VDJ) intersection that is generated by recombination-activating gene (RAG)-mediated somatic rearrangement. It is the part of the antigen-binding site that is most diverse in amino acid sequence and length.

Transporter associated with antigen processing

(TAP). A transporter of the ATP-binding cassette superfamily that is involved in the transport of peptides from the cytosol into the lumen of the endoplasmic reticulum. It is a heterodimer composed of TAP1 and TAP2 subunits and has a crucial role in the transport of major histocompatibility complex class I-binding peptides.

Tapasin

A protein that facilitates the binding of peptides to major histocompatibility complex (MHC) class I molecules by forming a bridge between MHC class I molecules and transporter associated with antigen processing (TAP).

Recombination signal sequences

Conserved nucleotide sequences flanking variable (V), diversity (D) and joining (J) segments that are recognized by the recombination-activating gene (RAG) proteins to induce rearrangement.

Immunoglobulin superfamily

Domains of 90–100 amino acids composed of 7–9 β strands forming two sheets, generally stabilized by a disulphide bond. Found in B cell receptor, T cell receptor and major histocompatibility complex molecules.

Ohnologues

Paralogues, named after Susumu Ohno, that are thought to have emerged close to the origin of vertebrates by whole-genome duplication.

Paralogy group

A set of paralogons that are derived from a single ancestral region.

Paralogues

(Also known as paralogous genes.) Genes within a single species that belong to the same gene family. In contrast to 'paralogues', 'orthologues' refers to genes that diverged by speciation events.

Major histocompatibility complex

The MHC is defined by: MHC class I and class II genes; genes that encode transporter associated with antigen processing (TAP) and tapasin, which are involved in processing antigens for presentation by MHC class I molecules; and genes that are involved in general immunity (see below). The MHC, like TCRs and BCRs, is first found in cartilaginous fish⁵⁵. Genes involved in class I processing and presentation are closely linked in almost all vertebrates (except in placental mammals, in which there is close linkage of the class I processing genes and class II genes) and are believed to have co-evolved^{55,56}. As described below, genes that encode natural killer (NK) cell receptors are associated with the MHC of various living vertebrates, and these serve as connections to what seems to have been a primordial gene complex composed of receptors and ligands^{55,57,58}. Therefore, it seems likely that NK receptors (and perhaps also the forerunners of antigen receptors) were physically linked to the primordial MHC in a similar way to the links between receptors and ligands in the myriad histocompatibility systems of modern-day lower deuterostomes and protostomes⁵⁹; these links were subsequently fragmented in gnathostomes.

How did the jawed vertebrate AIS arise?

As discussed above, the BCR–TCR–MHC-based AIS is present in cartilaginous fish but absent in lower chordates. Two macroevolutionary events — the invasion of the RAG transposon and two whole-genome duplications (WGDs) — are believed to have provided the innovation and raw material, respectively, for the relatively rapid emergence of the AIS in jawed vertebrates. In this section we discuss the potential contribution of these two events.

The recombination-activating gene transposon. In the late 1970s, Sakano and Tonegawa discovered recombination signal sequences (RSSs) flanking V, D and J rearranging segments⁶⁰, and they noticed that the inverted repeats within the RSSs were reminiscent of a transposon. This suggested that a gene encoding an immunoglobulin superfamily (IgSF) exon was invaded by a transposon. Therefore, the gene could not be expressed unless it was stitched back together through the action of a recombinase. There are many IgSF genes in lower chordates that are related to BCR and TCR V regions (the so-called 'VJ' type of IgSF gene; FIG. 1), so there are many candidate genes that could have been invaded by the RAG transposon in a vertebrate ancestor⁶¹. For example, the gene family encoding the variable region containing chitin-binding proteins (VCBPs) in amphioxus is presumed to be involved in defence and could be related to the non-rearranging ancestor of the BCR and TCR genes⁶². Other genes that might be related to the ancestral VJ gene (for example, REFS 63–65) are described below.

The *RAG1* and *RAG2* genes have no homologous regions and no introns but are closely linked to each other, which suggests that they could have been acquired by vertebrates as part of the original transposon^{4,66}. Complete RAG genes have been isolated from all gnathostomes studied, but not from jawless fish. This is consistent with

the idea that acquisition of the RAG transposon helped to trigger BCR- and TCR-based immunity. Surprisingly, linked *RAG1* and *RAG2* genes were found in the genome of the sea urchin (an echinoderm)⁶⁷ (FIG. 1), which implies either that the RAG transposon invaded the genome 100 million years before the emergence of adaptive immunity and remained dormant, or that it had invaded at multiple times. The authors suggested that a *RAG1*-like gene invaded the genome and inserted near a *RAG2*-like gene that was subsequently included in the original transposon. The *RAG1*-like gene could have originated from transib — a transposon that encodes the *RAG1* core and that has been found in many organisms⁶⁸ — or, as Dreyfus has recently suggested⁶⁹, from a viral genetic element. *RAG2* then lost its original function and was co-opted to the VDJ rearrangement process. This proposition is consistent with recent work showing that *RAG2* does not contribute to rearrangement *per se*, but instead interacts with modified histones at sites of active transcription⁷⁰. We propose a working model in which *RAG2* binds to active regions of transcription and *RAG1* binds to the exposed RSS. *RAG1* interacts with, and is dependent on, *RAG2* to initiate VDJ rearrangement, but *RAG1* performs all of the major functions associated with recombination.

Regardless of the route of its emergence in chordates, the RAG transposon invasion must have been crucial for adaptive immunity, as all of the BCR and TCR genes use the same mechanism of VDJ rearrangement. Furthermore, despite arguments to the contrary⁷¹, RAG activity was likely to have been an innovative event for adaptive immunity, as jawed vertebrates have many novel mechanisms and molecules that are involved in immunity in general, and adaptive immunity in particular, compared with jawless fish (discussed further below)⁷².

Whole-genome duplication. Nearly 40 years ago, Susumu Ohno proposed that the vertebrate genome underwent one or two rounds of WGD at the stage of fish or amphibians through a tetraploidization process⁷³. Although this proposal was a matter of hot debate until recently (BOX 1), it is now widely accepted that the vertebrate genome experienced two rounds of WGD after the emergence of urochordates and before the radiation of jawed vertebrates^{74–77} (FIGS 1, 4). WGDs seem to have had a crucial role in the emergence of jawed vertebrate-type adaptive immunity^{78,79}. This is indicated by the fact that many ohnologues, which arose as a result of WGDs, are essential components of the jawed vertebrate-type AIS^{80,81}.

The MHC paralogy group (FIG. 4a) and the HOX paralogy group (Supplementary information S1 (figure)) are illustrative examples. In the case of the MHC paralogy group^{82–84}, paralogues of more than 100 gene families are distributed in four sets of paralogons (for a list of gene families, see REF. 85). As would be expected from Ohno's theory of WGD, the genomes of invertebrates, such as amphioxus^{84,86} and *Ciona intestinalis*⁸⁷, do not have multiple sets of MHC paralogons and contain only one MHC-like region, which is devoid of class I and II genes (a proto-MHC).

Recently, Olinski *et al.*⁸⁵ suggested that the neurotrophin paralogy group (FIG. 4, left) and the MHC

Box 1 | Controversies surrounding whole-genome duplication

The idea that the vertebrate genome underwent two rounds of whole-genome duplication (WGD) at a time close to the origin of vertebrates — known as the two-round (2R) hypothesis — caused hot debate for more than a decade. The major evidence in favour of this hypothesis was that paralogues that emerged close to the origin of vertebrates are not distributed at random in the human genome, but have a strong tendency to occur in clusters, known as paralogons, on multiple — typically four — chromosomes. The major arguments against this hypothesis were twofold^{127,128}. First, ohnologues (for example, A, B, C and D) generated by two rounds of WGD should display the tree topology (that is, A and B together and C and D together), which is not always the case¹²⁷. Second, paralogons occasionally contain paralogues that duplicated much earlier than the origin of vertebrates. We now know that incongruent tree topologies can occur if two waves of genome doubling occurred in close succession¹²⁹, if paralogues have evolved at different evolutionary rates¹³⁰ or if the resolving power of phylogenetic trees is not sufficient¹²⁹. Also, the divergence time of paralogues can be much older than the age of WGD if tandem duplication preceded WGD and the duplicated genes were lost differentially after WGD. Therefore, neither of these two counterarguments contradicts the 2R hypothesis.

Recently, comparison of the human and amphioxus (*Branchiostoma floridae*) genomes revealed widespread quadruple conserved synteny, in which four sets of human paralogons corresponded to one set of linked genes in *B. floridae*¹³¹. This observation provided definitive evidence for the 2R hypothesis and settled the long-standing debate. The exact timing of WGD relative to the emergence of jawless vertebrates (FIG. 1) is still controversial. A commonly held view, which we also favour, is that the first round of WGD occurred in a common ancestor of jawed and jawless vertebrates, and the second round in a common ancestor of jawed vertebrates⁷⁶. However, recent analysis of 55 gene families in jawless vertebrates suggested that both rounds of WGD might have taken place in a common ancestor of all vertebrates, although this analysis could not discount the idea that the second round of WGD occurred in a common ancestor of jawed vertebrates¹³², as depicted in FIG. 1.

paralogy group (FIG. 4, right) were derived from neighbouring regions on the same ancestral chromosome. This suggestion is based on the observation that several genes of the MHC- and neurotrophin-paralogy groups are linked in the genomes of *C. intestinalis*⁸⁵ and amphioxus⁸⁸. Interestingly, chromosomal regions 12p13.2–p12.3 and 19q13.4, which encode the human NK complex (NKC) and leukocyte receptor complex (LRC), respectively (BOX 2), are embedded in the neurotrophin paralogons⁸⁹. Therefore, Olinski *et al.*'s suggestion provides a logical basis for the earlier proposals that the ancestors of NK receptor genes encoded in the NKC and LRC were located in the proto-MHC^{55,57,90,91}.

Strikingly, the integrated MHC/neurotrophin paralogy group encodes many ohnologues that have important functions in the AIS, particularly in antigen presentation by MHC class I and II molecules (FIG. 4; [Supplementary information S2](#) (table)). In addition to TAP and tapasin, three proteasome subunits that are specialized for antigen presentation are encoded in this paralogy group, together with some of their paralogues. Similarly, the integrated paralogy group encodes 10 of the 11 cathepsin ohnologues, including cathepsins S and L, which have crucial roles in antigen presentation by class II molecules and in thymic selection. Other examples include: retinoid X receptor- β (*RXR β*) and regulatory factor X5 (*RFX5*), which encode proteins that regulate the expression of class I and II molecules, respectively; *NOTCH1*, which is involved in lymphocyte differentiation; and *VAV1*, which is required for T and B cell development and activation ([Supplementary information S2](#) (table)).

Paralogons

Chromosomal segments that contain closely linked sets of paralogues. Also known as paralogous regions.

Agglutinin

Any substance that can clump particles together.

AID-APOBEC

A family of cytidine deaminases that are involved in the hypermutation of variable-diversity-joining (VDJ) segments, in the immunoglobulin class switch and in defence against viruses. Two members of the family are implicated in generating diversity in variable lymphocyte receptors.

Allelic exclusion

The expression of a single receptor in cells with the potential to express more than two receptors.

The convergent AIS of jawless vertebrates

With some exceptions, and despite the involvement of the well-conserved and rapidly evolving molecular features described above, the AIS of jawed vertebrates has been remarkably uniform over 500 million years. Functional assays of jawless vertebrates (lampreys and hagfish) have shown signs of adaptive immunity, but neither adaptive immune molecular characteristics nor tissues, such as a thymus or spleen, were detected in these animals. For example, jawless vertebrates produce specific agglutinins when immunized with particulate and soluble antigens^{92–95}, but extensive transcriptome analysis of jawless fish leukocytes provided no evidence for the presence of MHC, BCR or TCR molecules^{96–98}. Their absence has been recently confirmed by the analysis of the *sea lamprey* (*Petromyzon marinus*) [draft genome sequence](#). How can they mount adaptive immune responses in the absence of BCR, TCR and MHC molecules? The discovery of a *VLR* gene, which generates diversity by somatically rearranging leucine-rich repeat (LRR) modules⁷, provided an answer⁹⁹. The germline *VLR* gene lacks gene segments coding for LRR modules and is not capable of encoding functional proteins (FIG. 3b). During the development of lymphoid cells, LRR modules that flank the gene are sequentially incorporated into *VLR* by a process called 'copy choice'¹⁰⁰. Incorporation is presumably mediated by cytidine deaminases of the AID-APOBEC family⁸. The mature *VLR* gene encodes a glycosyl-phosphatidylinositol-anchored polypeptide. The sequence of individual LRR modules assembled into the *VLR* gene is highly diverse, and the number of LRR modules present in a rearranged gene varies from one to nine. Therefore a single *VLR* gene can have combinatorial diversity comparable to that of Igs (~10¹⁴ combinations)¹⁰¹. The crystal structures of hagfish *VLR* monomers indicate that they adopt a horseshoe-like solenoid structure that is characteristic of LRR family proteins. The majority of variable residues in the *VLR* are located on the concave surface, which suggests that this surface is involved in antigen binding¹⁰². This was recently confirmed by *in vitro* mutagenesis experiments¹⁰³ and the crystal structure analysis of *VLR* in complex with the human blood group H-trisaccharide¹⁰⁴ or with hen egg lysozyme¹⁰⁵.

Lampreys have two *VLR* genes: *VLRA* and *VLRB*⁸. Both seem to be expressed exclusively in lymphoid cells, and they usually exhibit allelic exclusion. Stimulation with antigens causes lymphoid cells expressing specific *VLRB* molecules to undergo clonal expansion and to secrete *VLRs* in a manner analogous to the secretion of Igs by B cells^{7,101}. Secreted *VLRB* molecules occur as pentamers or tetramers of dimers and have eight to ten antigen-binding sites¹⁰³. Therefore, the subunit organization of secreted *VLRB* molecules resembles that of IgM antibodies. The multivalent structure of *VLRB* enables it to bind antigens carrying repetitive epitopes with high avidity and specificity, and it displays strong agglutinating activities¹⁰⁶. It has been suggested that, like subclasses of IgM, *VLRB* might preferentially or exclusively recognize thymus-independent antigens¹⁰⁶. By contrast, *VLRA* seems to occur only in a cell-bound form and is

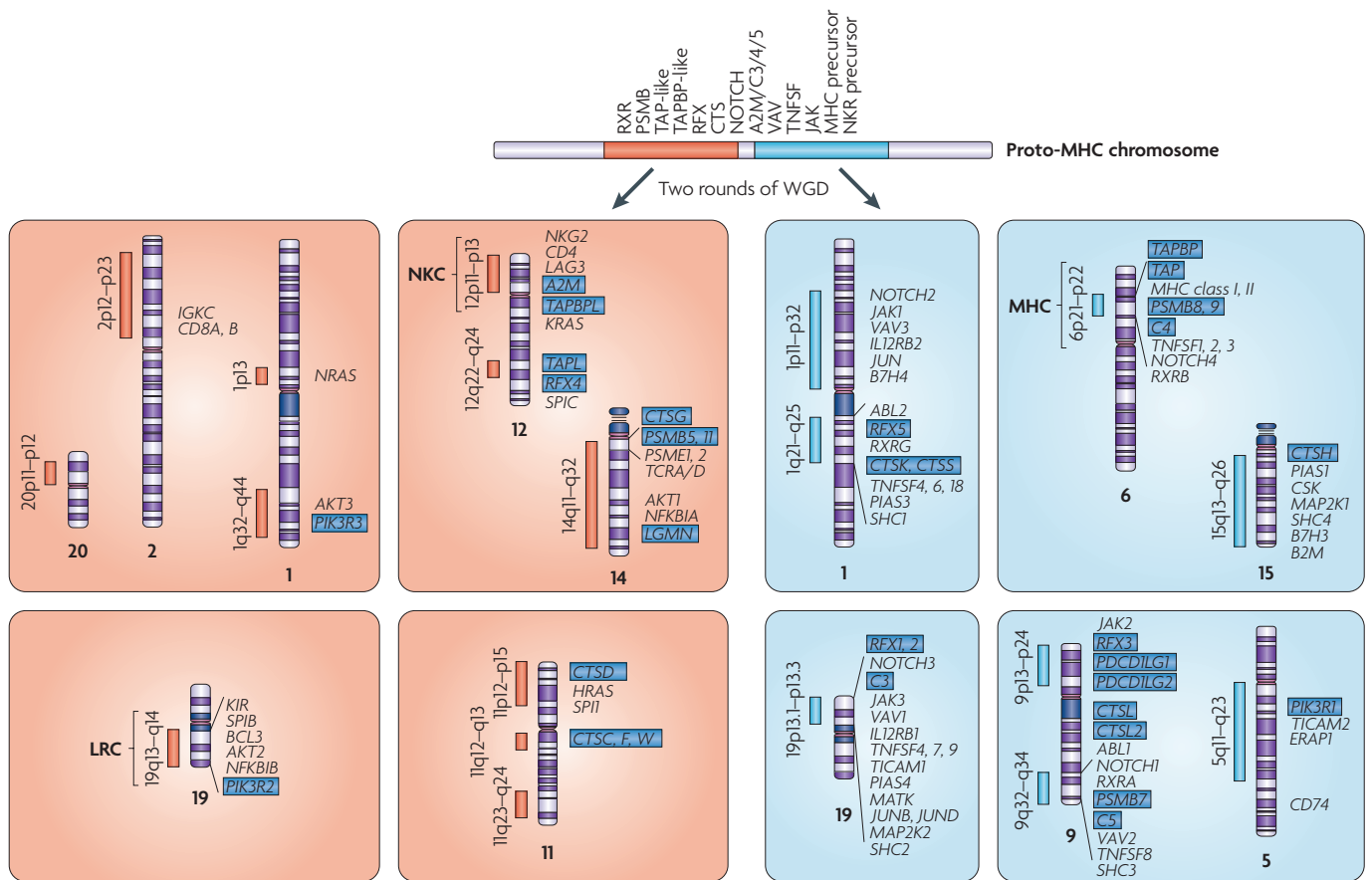


Figure 4 | **The major histocompatibility complex paralogy group and the neurotrophin paralogy group in the human genome.** Four sets of major histocompatibility complex (MHC) paralogons are located on chromosomes 1, 6, 9 and 19 (REFS 55, 137, 138). A number of smaller-sized MHC paralogons, which presumably originated from fragmentation and subsequent translocation of the major paralogons, have been identified⁵⁵. Among them, those located on 15q13–q26 and 5q11–q23 seem to have broken off from the paralogons on chromosomes 6 and 9, respectively¹³⁰. Four sets of major neurotrophin paralogons are located on chromosomes 1, 11, 12 and 19 (REF. 89). The MHC paralogy group (right) and neurotrophin paralogy group (left) are thought to have descended from neighbouring regions on a single ancestral chromosome. Paralogues that are distributed across the two paralogy groups are indicated by blue highlighted text. In addition to genes that share paralogues among the relevant paralogons, immunologically important genes, such as those encoding MHC class I, MHC class II, T cell receptors (TCRs), and immunoglobulins (Igs) are shown. AZM, $\alpha 2$ -macroglobulin; B2M, $\beta 2$ -microglobulin; BCL, B cell leukaemia/lymphoma; CSK, cytoplasmic tyrosine kinase; CTS, cathepsin; ERAP, endoplasmic reticulum aminopeptidase; IL12RB, interleukin 12 receptor, β -subunit; JAK, Janus kinase; KIR, killer Ig-like receptor; LAG3, lymphocyte activation gene 3; LGMN, legumain; LRC, leukocyte receptor complex; MAP2K, mitogen-activated protein kinase kinase; MATK, megakaryocyte-associated tyrosine kinase; NFKBIA, nuclear factor- κ B inhibitor; PIAS, protein inhibitors of activated STAT; PIK3R, phosphatidylinositol 3-kinase, regulatory subunit; PSMB, proteasome subunits, β -type; PSME, proteasome activator subunits; RFX, regulatory factor X; RXR, retinoid X receptor; SPI1, spleen focus forming virus proviral integration oncogene; TAP, transporter associated with antigen processing; TAPBP, TAP-binding protein (also known as tapasin); TAPBP-like, TAPL, TAP-like (also known as ABCB9); TICAM, TIR domain-containing adaptor molecule; TNFSF, tumour necrosis factor ligand superfamily; WGD, whole-genome duplication.

expressed on lymphoid cells that functionally resemble T cells. Therefore lampreys have two populations of lymphoid cells that seem to be functionally similar to the T and B cells of jawed vertebrates¹⁰⁷.

Hagfish also have two *VLR* genes¹⁰⁸, *VLRA* and *VLRB*, which are presumed to be orthologous to lamprey *VLRA* and *VLRB*, based on sequence comparison. They map to the same chromosome but are distant from each other, which indicates that they function as separate units¹⁰⁹. Consistent with this, the rearranged *VLRA* and *VLRB* genes do not share an LRR module with an

identical nucleotide sequence. It remains to be determined whether hagfish *VLRA* and *VLRB* are functionally similar to the corresponding lamprey genes.

Origin of the two distinct forms of AIS

Both VLRs and BCRs/TCRs rely on combinatorial diversity to generate a vast repertoire of binding specificities; however, the genes of VLRs show no structural similarity to those of BCRs or TCRs (FIG. 3). Their recombination machineries must differ because RAG1 and RAG2 are required only for BCR/TCRs. These observations suggest

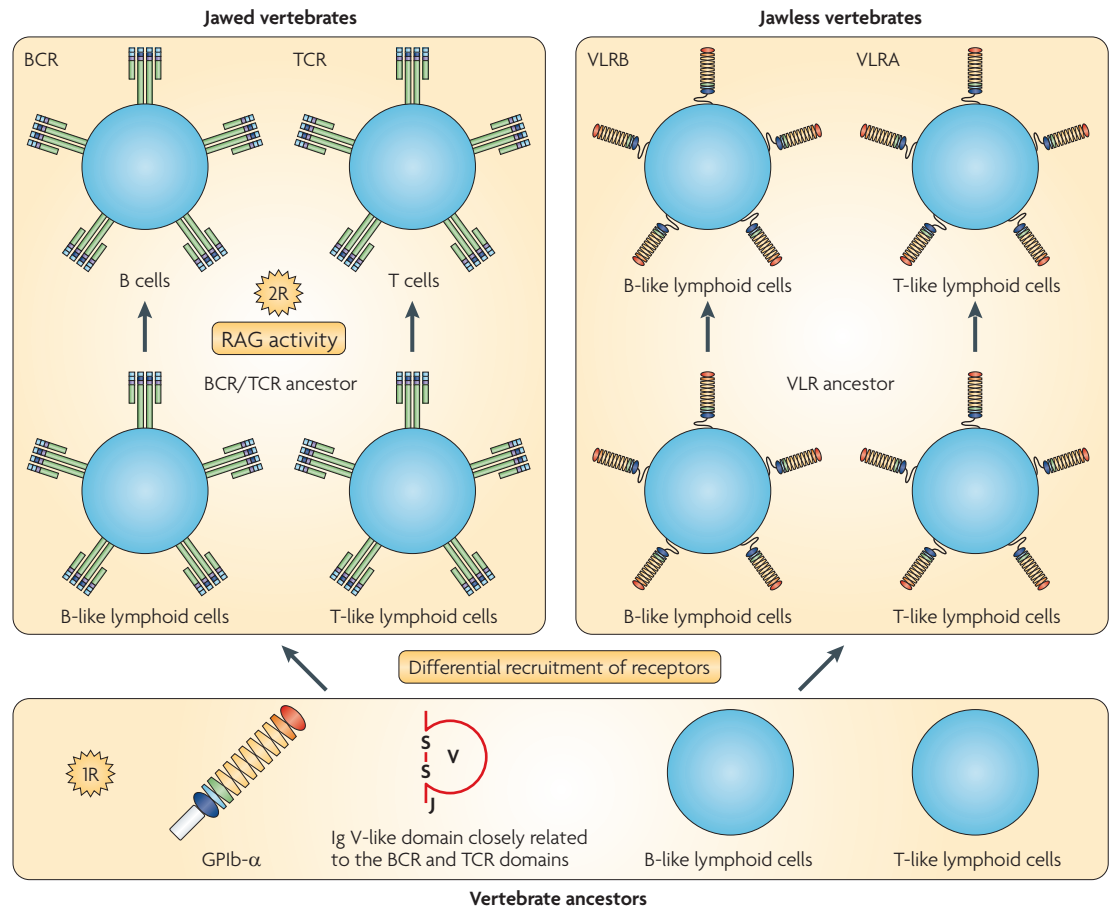


Figure 5 | A hypothetical model for the origin of the two major forms of adaptive immune system. The occurrence of B- and T-like lymphoid cells in jawless vertebrates indicates that they were likely to be present in a common ancestor of all vertebrates¹⁰⁷. Similarly, the occurrence of a V-type immunoglobulin (Ig)-like domain with a canonical joining (J) segment in jawless vertebrates indicates that a common ancestor of all vertebrates had V-type Ig-like domains that could be converted to B cell receptor (BCR) or T cell receptor (TCR) Ig-like domains. *Amphioxus*, a basal chordate, has a glycoprotein Ib- α (GPIb- α)-like protein, which is the most likely precursor to the variable lymphocyte receptor (VLR)⁸. Therefore, a common ancestor of all vertebrates presumably had all of the key structural elements of the two radically different antigen receptors. Acquisition of recombination-activating gene (RAG) recombinase activities and the second round of whole-genome duplication (WGD; indicated as 2R) are thought to have had a crucial role in the emergence of the BCR–TCR–MHC-based adaptive immune system (AIS). Little is known about the genetic events that contributed to the emergence of the VLR-based AIS in jawless vertebrates. It is possible that the first round of WGD (indicated as 1R) contributed to the emergence of the two lineages of lymphoid cells; however, we cannot rule out the possibility that their origin predates the emergence of vertebrates.

that the ancestors of jawed and jawless vertebrates developed their antigen receptors independently.

Although VLRs probably exist only in lampreys and hagfish, the origin of LRR modules is ancient; LRR module repeats are found in many animal and plant proteins, in which they seem to mediate protein interactions or microbial recognition. It has been proposed, on the basis of the similarity of gene structures and a characteristic insertion in the LRR carboxy-terminal (LRRCT) domain, that an ancestor of VLR might have emerged from a glycoprotein Ib- α (GPIb- α)-like protein (GPIb- α is a component of the platelet glycoprotein receptor complex that is conserved in all vertebrates)⁸. Furthermore, searches for IgSF proteins in jawless vertebrates have identified molecules that are thought to be related to the potential evolutionary precursors of BCRs and TCRs^{63–65}. These

are: the agnathan paired receptors resembling antigen receptors (APAR) family of innate immune receptors, which encode a group of membrane proteins carrying a single extracellular V-type Ig-like domain with a canonical J segment⁶⁴; and an immunoreceptor tyrosine-based inhibition motif-bearing membrane protein, named ‘TCR-like’⁶⁵, which has V-C2-type domain organization and seems to be closely related to the novel immunoreceptor tyrosine-based activation motif-containing IgSF receptor (NICIR) family of activating innate immune receptors¹¹⁰. The occurrence of these receptors in jawless vertebrates suggests that a common ancestor of all vertebrates had V-type Ig-like domains that evolved into the Ig-like domains of BCRs and TCRs. Hence, as mentioned above, it seems that a vertebrate ancestor had both BCR/TCR and VLR precursors (FIG. 5). In a jawed

vertebrate lineage in which RAG insertion occurred, an IgSF protein was 'chosen' as an antigen receptor, which led to the BCRs and TCRs. In a jawless vertebrate lineage with no appropriate RAG insertion, an LRR protein family that could mediate microbial recognition was 'chosen' as an antigen receptor, which led to the VLRs.

The occurrence of structurally unrelated receptors that are used for the same or similar purposes is not unprecedented (BOX 2). For example, in primates the NK receptors that interact with classical MHC class I molecules are members of the killer cell Ig-like receptor (KIR) family, but in rodents the corresponding receptors are C-type lectin-like molecules known as Ly49 receptors¹¹¹. A common ancestor of primates and rodents is thought to have had precursor genes for both types of NK receptors; the KIR gene family expanded in the primate lineage and the Ly49 family expanded in the rodent lineage^{112,113}. We suggest that the ancestors of jawed and jawless vertebrates adopted distinct molecules as their antigen receptors in an analogous manner.

The observation that jawless vertebrates have B- and T-like lymphoid cells¹⁰⁷ indicates that the two lineages of lymphoid cells emerged before the divergence of jawed and jawless vertebrates (FIG. 5). Therefore, the origin of B- and T-like lymphoid cells is likely to be more ancient than previously thought. This is compatible with recent observations that B and T cells do not share an immediate common progenitor, but differentiate from myeloid B progenitors and myeloid T progenitors, respectively^{88,113–115}. This suggests that B- and T-like lymphoid cells diverged before the divergence of myeloid and lymphoid cells. Although we must be cautious when using cell lineage information as a proxy for the order of

evolutionary emergence, it seems likely that the building blocks for both types of antigen receptors and also the two lineages of lymphoid cells were present in a common ancestor of all vertebrates (FIG. 5).

Pressures for adaptive immunity

Why did the AIS arise? As noted above, invertebrates and plants manage with innate mechanisms of defence (however, it should be mentioned that primitive adaptive systems have been described in several invertebrate taxa¹¹⁶ and that an RNAi-like mechanism provides a potent adaptive response in many organisms¹¹⁷). Indeed, it has been argued that an anticipatory system of defence is more trouble than it is worth, as it necessitates the generation of tolerance to self during the ontogeny of lymphocytes, and failure to regulate these defence mechanisms can result in autoimmunity¹¹⁸. Furthermore, the adaptive system might only be maintained when under strong selective pressure: organisms such as cod¹¹⁹, seahorses¹²⁰ and axolotl¹²¹, which might not be under strong pressure from pathogens, seem to have lost the capacity to mount effective immune responses, at least as far as they can be measured experimentally¹²¹.

However, we think that at least the BCR–TCR–MHC system arose as a potent system of defence. It is believed that this system emerged in animals with jaws, which are likely to have been predatory. Large predators generally do not have many offspring, thereby maintaining a ratio of predator-to-prey that perpetuates the *status quo*¹²². Sharks, the oldest living vertebrates with this adaptive system, have many new mechanisms of reproduction to protect young offspring — such as impenetrable egg-cases and modes of viviparity that have converged on the mechanisms in mammals. Therefore it can be proposed that a potent system of defence arose in conjunction with the need to protect offspring. In jawless fish, the AID–APOBEC family members might have enabled the diversification of a preimmune repertoire of IgSF genes, as seems to be the case for the VLR system. The addition of the RAG transposon to this adaptive system would have provided a great advantage, causing a break during rearrangement that would have created the potential for large variations in the size of one of the CDR loops that interact with antigens³⁷. This advantage, along with the raw material of a large number of newly duplicated genes that could be co-opted for immune functions, probably resulted in the highly complex AIS seen today in gnathostomes.

Conclusions

There are many characteristics of the AIS that can generally be agreed upon: jawless and jawed fish both have an AIS based on different gene families, but the BCR–TCR–MHC-based system is more complex; the RAG transposon and WGDs had major roles in the emergence of the jawed vertebrate AIS, and the major features of the system arose early in a 'big bang' and have been modified only slightly over evolutionary time; and the jawed vertebrate AIS displays both conserved and plastic features, the former presumably to allow for a general functioning of the system and the latter to permit rapid changes over evolutionary time to respond to pathogens¹¹.

Box 2 | Natural killer cell receptors

Natural killer (NK) cells are part of the innate immune system. They can produce cytokines and kill virally infected cells without a clonal proliferation phase. NK cells share many characteristics with T cells, including most cell-surface markers, elaboration of the same effector functions and origins from a similar haematopoietic precursor¹¹¹. However, the NK cell receptor (NKR) recognition strategy is different from that of T cells (although the signalling cascades are homologous). When cells become virally infected, major histocompatibility complex (MHC) class I molecules are often downregulated, making the cells 'invisible' to cytotoxic T cells. Inhibitory NKRs on NK cells recognize MHC class I molecules; when the inhibitory NKRs are not engaged, the NK cells can be stimulated by activating receptors that often recognize stress-induced molecules that are upregulated on viral infection. This is known as the 'missing self' hypothesis¹³³.

Unlike T cell receptors (TCRs), NKRs are not generated by recombination-activating gene (RAG)-mediated recombination, but rather through stochastic, clonal expression of a few receptors from a cluster of genes encoding activating and inhibitory NKRs. These receptors are encoded by the NK complex (NKC) and leukocyte receptor complex (LRC), found on different chromosomes in all mammals examined. The NKRs that recognize classical (polymorphic) class I molecules are encoded by genes in different gene families in various species, for example the C-type lectin Ly49 genes found in the NKC in mice and the killer immunoglobulin superfamily (IgSF) receptor (KIR) genes found in the LRC of humans. This is the only known case in the immune system in which receptors from different gene families perform precisely the same functions in closely related species, and it is a testament to the rapid evolution of the NK system of defence. Evidence from comparative studies of the MHC, LRC and NKC strongly suggest that the three complexes were originally linked and that their polymorphic receptors and ligands have co-evolved, as has been found in many histocompatibility systems in plants and invertebrates⁵⁹.

Many questions still need to be addressed, most of which will require incremental progress. For example, what is the nature of the RAG transposon? What are the functions of antigen receptors in different organisms, especially the $\gamma\delta$ TCR? And how is the VLR system regulated? However, there are two glaring questions: did T- and B-like lymphoid cells really emerge in the common ancestor of jawless and jawed vertebrates, and what were the pressures that selected for the emergence of the AIS? Regarding the first of these questions, besides what we have discussed above concerning the antigen receptors, it is essential to determine whether, like gnathostome $\alpha\beta$ T cells, the 'T-like lymphoid cells' of jawless vertebrates recognize processed antigens, perhaps through presentation on a convergent type of 'MHC' molecule. Instead, they might recognize native antigens in a cell-bound form, perhaps in a way similar to the recognition that occurs for some subsets of gnathostome $\gamma\delta$ T cells. Because the lamprey 'T cell' antigen receptor VLRA has been shown to bind soluble antigens and because its gene apparently mutates after antigenic stimulation, our prediction is that the $\gamma\delta$ TCR-like mechanism will prove to be correct¹²³. Regarding the second question, we will never be able to resolve this problem experimentally.

New ideas concerning selection pressures suggest that the capacity of the immune system to distinguish between commensal and pathogenic microorganisms resulted in the great complexity of the gnathostome AIS^{124,125}. The understanding of the immune system's interactions with commensal organisms is in its infancy, and perhaps we can develop defined hypotheses as this field develops. One hypothesis, which emphasizes the importance of the gut, suggests that with the advent of jaws, digested material could injure the intestine and therefore result in massive infections. The seahorse, which has seemingly lost features of adaptive immunity, is a filter-feeding animal, so it is speculated to have reverted to a pre-adaptive state¹²⁰. Our view is that the AIS arose as a system of defence and became more crucial and more sophisticated with the emergence of the large predatory jawed vertebrates that bore few offspring. We think that our position has been strengthened by recent evidence that has shown that the jawed vertebrates in which it is believed that the AIS first emerged — the placoderms — had internal fertilization¹²⁶. We hope that this Review will spark further interest in this area, especially among non-immunologists who might see the problem from a fresh perspective.

- Weigert, M. G., Cesari, I. M., Yonkovich, S. J. & Cohn, M. Variability in the λ light chain sequences of mouse antibody. *Nature* **228**, 1045–1047 (1970).
- Tonegawa, S. Reiteration frequency of immunoglobulin light chain genes: further evidence for somatic generation of antibody diversity. *Proc. Natl Acad. Sci. USA* **73**, 203–207 (1976).
- Davis, M. M., Chien, Y. H., Gascoigne, N. R. & Hedrick, S. M. A murine T cell receptor gene complex: isolation, structure and rearrangement. *Immunol. Rev.* **81**, 235–258 (1984).
- Oettinger, M. A., Schatz, D. G., Gorka, C. & Baltimore, D. *RAG-1* and *RAG-2*, adjacent genes that synergistically activate V(D)J recombination. *Science* **248**, 1517–1523 (1990).
- Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–563 (2000).
- Schluter, S. F., Bernstein, R. M., Bernstein, H. & Marchalonis J. J. 'Big Bang' emergence of the combinatorial immune system. *Dev. Comp. Immunol.* **23**, 107–111 (1999).
- Pancer, Z. *et al.* Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* **430**, 174–180 (2004).
This is a seminal paper that described a novel rearranging gene in the lamprey. This work suggested strongly that jawless vertebrates have an alternative form of adaptive immunity that is not dependent on BCRs, TCRs or the MHC.
- Rogozin, I. B., Iyer, L. M. *et al.* Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nature Immunol.* **8**, 647–656 (2007).
This paper suggests that the diversity of the VLR gene is generated by a gene conversion-like process that is presumably mediated by cytidine deaminases of the AID-APOBEC family. It also describes the identification of the lamprey VLRA gene.
- Hibino, T., Loza-Coll, M. *et al.* The immune gene repertoire encoded in the purple sea urchin genome. *Dev. Biol.* **300**, 349–365 (2006).
- Jones, J. D. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–329 (2006).
- Klein, J., Satta, Y., O'hUigin, C. & Takahata, N. The molecular descent of the major histocompatibility complex. *Annu. Rev. Immunol.* **11**, 269–295 (1993).
- Flajnik, M. F. Comparative analyses of immunoglobulin genes: surprises and portents. *Nature Rev. Immunol.* **2**, 688–698 (2002).
- Kaattari, S., Evans, D. & Klemer, J. Varied redox forms of teleost IgM: an alternative to isotopic diversity? *Immunol. Rev.* **166**, 133–142 (1998).
- Dooley, H. & Flajnik, M. F. Shark immunity bites back: affinity maturation and memory response in the nurse shark, *Ginglymostoma cirratum*. *Eur. J. Immunol.* **35**, 936–945 (2005).
- Wilson, M. *et al.* A novel chimeric Ig heavy chain from a teleost fish shares similarities to IgD. *Proc. Natl Acad. Sci. USA* **94**, 4593–4597 (1997).
- Ohta, Y. & Flajnik, M. IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. *Proc. Natl Acad. Sci. USA* **103**, 10723–10728 (2006).
- Zhao, Y. *et al.* Identification of IgF, a hinge-region-containing Ig class, and IgD in *Xenopus tropicalis*. *Proc. Natl Acad. Sci. USA* **103**, 12087–12092 (2006).
References 16 and 17 confirm that IgD is much older than previously realized and that it has evolved rapidly over evolutionary time.
- Greenberg, A. S. *et al.* A novel 'chimeric' antibody class in cartilaginous fish: IgM may not be the primordial immunoglobulin. *Eur. J. Immunol.* **26**, 1123–1129 (1996).
- Berstein, R. M., Schluter, S. F., Shen, S. & Marchalonis, J. J. A new high molecular weight immunoglobulin class from the carcharhine shark: implications for the properties of the primordial immunoglobulin. *Proc. Natl Acad. Sci. USA* **93**, 3289–3293 (1996).
- Ota, T., Rast, J. P., Litman, G. W. & Amemiya, C. T. Lineage-restricted retention of a primitive immunoglobulin heavy chain isotype within the Dipnoi reveals an evolutionary paradox. *Proc. Natl Acad. Sci. USA* **100**, 2501–2506 (2003).
- Chen, K. *et al.* Immunoglobulin D enhances immune surveillance by activating antimicrobial, proinflammatory and B cell-stimulating programs in basophils. *Nature Immunol.* **10**, 889–898 (2009).
- Warr, G. W., Magor, K. E. & Higgins, D. A. IgY: clues to the origins of modern antibodies. *Immunol. Today* **16**, 392–398 (1995).
- Mussmann, R., Du Pasquier, L. & Hsu, E. Is *Xenopus* IgX an analog of IgA? *Eur. J. Immunol.* **26**, 2823–2830 (1996).
- Desmyter, A. *et al.* Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. *Nature Struct. Biol.* **3**, 803–811 (1996).
- Greenberg, A. S. *et al.* A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature* **374**, 168–173 (1995).
- Greenberg, A. S., Steiner, L., Kasahara, M. & Flajnik, M. F. Isolation of a shark immunoglobulin light chain cDNA clone encoding a protein resembling mammalian κ light chains: implications for the evolution of light chains. *Proc. Natl Acad. Sci. USA* **90**, 10603–10607 (1993).
- Crisicciello, M. F. & Flajnik, M. F. Four primordial immunoglobulin light chain isotypes, including λ and κ , identified in the most primitive living jawed vertebrates. *Eur. J. Immunol.* **37**, 2683–2694 (2007).
- Hohman, V. S., Schuchman, D. B., Schluter, S. F. & Marchalonis, J. J. Genomic clone for sandbar shark λ light chain: generation of diversity in the absence of gene rearrangement. *Proc. Natl Acad. Sci. USA* **90**, 9882–9886 (1993).
- Rast, J. P. *et al.* Immunoglobulin light chain class multiplicity and alternative organizational forms in early vertebrate phylogeny. *Immunogenetics* **40**, 83–99 (1994).
- Schwager, J., Burckert, N., Schwager, M. & Wilson, M. Evolution of immunoglobulin light chain genes: analysis of *Xenopus* IgL isotypes and their contribution to antibody diversity. *EMBO J.* **10**, 505–511 (1991).
- Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575–581 (1983).
- Reynaud, C. A., Anquez, V., Grimal, H. & Weill, J. C. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell* **48**, 379–388 (1987).
- Knight, K. L. & Becker, R. S. Molecular basis of the allelic inheritance of rabbit immunoglobulin VH allotypes: implications for the generation of antibody diversity. *Cell* **60**, 963–970 (1990).
- Hinds, K. R. & Litman, G. W. Major reorganization of immunoglobulin VH segmental elements during vertebrate evolution. *Nature* **320**, 546–549 (1986).
- Hinds-Frey, K. R., Nishikata, H., Litman, R. T. & Litman, G. W. Somatic variation precedes extensive diversification of germline sequences and combinatorial joining in the evolution of immunoglobulin heavy chain diversity. *J. Exp. Med.* **178**, 815–824 (1993).
- Diaz, M., Greenberg, A. S. & Flajnik, M. F. Somatic hypermutation of the new antigen receptor gene (*NAR*) in the nurse shark does not generate the repertoire: possible role in antigen-driven reactions in the absence of germinal centers. *Proc. Natl Acad. Sci. USA* **95**, 14343–14348 (1998).
- Lee, S. S., Tranchina, D., Ohta, Y., Flajnik, M. F. & Hsu, E. Hypermutation in shark immunoglobulin light chain genes results in contiguous substitutions. *Immunity* **16**, 571–582 (2002).

38. Daggfeldt, A., Bengten, E. & Pilstrom, L. A cluster type organization of the loci of the immunoglobulin light chain in Atlantic cod (*Gadus morhua* L.) and rainbow trout (*Oncorhynchus mykiss* Walbaum) indicated by nucleotide sequences of cDNAs and hybridization analysis. *Immunogenetics* **38**, 199–209 (1995).
39. Hsu, E. & Criscitiello, M. F. Diverse immunoglobulin light chain organizations in fish retain potential to revise B cell receptor specificities. *J. Immunol.* **177**, 2452–2462 (2006).
40. Lange, M. D., Waldbieser, G. C. & Lobb, C. J. Patterns of receptor revision in the immunoglobulin heavy chains of a teleost fish. *J. Immunol.* **182**, 5605–5622 (2009).
41. Zimmerman, A. M., Yeo, G., Howe, K., Maddox, B. J. & Steiner, L. A. Immunoglobulin light chain (IgL) genes in zebrafish: genomic configurations and inversions rearrangements between $V_{\text{H}}-J_{\text{H}}-C_{\text{H}}$ gene clusters. *Dev. Comp. Immunol.* **32**, 421–434 (2008).
42. Danilova, N., Bussmann, J., Jekosch, K. & Steiner, L. A. The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. *Nature Immunol.* **6**, 295–302 (2005).
43. Hansen, J. D., Landis, E. D. & Phillips, R. B. Discovery of a unique Ig heavy-chain isotype (IgT) in rainbow trout: implications for a distinctive B cell developmental pathway in teleost fish. *Proc. Natl Acad. Sci. USA* **102**, 6919–6924 (2005).
44. Zinkernagel, R. M. & Doherty, P. C. Immunological surveillance against altered self components by sensitized T lymphocytes in lymphocytic choriomeningitis. *Nature* **251**, 547–548 (1974).
45. Charlemagne, J., Fellah, J. S., De Guerra, A., Kerfourn, F. & Partula, S. T-cell receptors in ectothermic vertebrates. *Immunol. Rev.* **166**, 87–102 (1998).
46. Guo, J. *et al.* Regulation of the TCR α repertoire by the survival window of CD4⁺CD8⁺ thymocytes. *Nature Immunol.* **3**, 469–476 (2002).
47. Havran, W. L. *et al.* Limited diversity of T-cell receptor γ -chain expression of murine Thy-1⁺ dendritic epidermal cells revealed by $V\gamma 3$ -specific monoclonal antibody. *Proc. Natl Acad. Sci. USA* **86**, 4185–4189 (1989).
48. Morita, C. T. *et al.* Direct presentation of nonpeptide prenyl pyrophosphate antigens to human $\gamma\delta$ T cells. *Immunity* **3**, 495–507 (1995).
49. Thedrez, A. *et al.* Self/non-self discrimination by human $\gamma\delta$ T cells: simple solutions for a complex issue? *Immunol. Rev.* **215**, 123–135 (2007).
50. Rock, E. P., Sibbald, P. R., Davis, M. M. & Chien, Y. H. CDR3 length in antigen-specific immune receptors. *J. Exp. Med.* **179**, 323–328 (1994).
51. Criscitiello, M. F., Saltis, M. & Flajnik, M. F. An evolutionarily mobile antigen receptor variable region gene: doubly rearranging NAR-TcR genes in sharks. *Proc. Natl Acad. Sci. USA* **103**, 5036–5041 (2006).
52. Parra, Z. E. *et al.* A unique T cell receptor discovered in marsupials. *Proc. Natl Acad. Sci. USA* **104**, 9776–9781 (2007).
53. Sciammas, R. & Bluestone, J. A. TCR $\gamma\delta$ cells and viruses. *Microbes Infect.* **1**, 203–212 (1999).
54. Chen, H. *et al.* Characterization of arrangement and expression of the T cell receptor γ locus in the sandbar shark. *Proc. Natl Acad. Sci. USA* **106**, 8591–8596 (2009).
55. Flajnik, M. F. & Kasahara, M. Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. *Immunity* **15**, 351–362 (2001).
56. Kaufman, J. Co-evolving genes in MHC haplotypes: the 'rule' for nonmammalian vertebrates? *Immunogenetics* **50**, 228–236 (1999).
57. Trowsdale, J. Genetic and functional relationships between MHC and NK receptor genes. *Immunity* **15**, 363–374 (2001).
58. Rogers, S. L., Viertlboeck, B. C., Gobel, T. W. & Kaufman, J. Avian NK activities, cells and receptors. *Semin. Immunol.* **20**, 353–360 (2008).
59. Nyholm, S. V. *et al.* *fester*, a candidate allorecognition receptor from a primitive chordate. *Immunity* **25**, 163–173 (2006).
60. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature* **280**, 288–294 (1979).
61. Du Pasquier, L., Zucchetti, I. & De Santis, R. Immunoglobulin superfamily receptors in protochordates: before RAG time. *Immunol. Rev.* **198**, 233–248 (2004).
62. Hernández Prada, J. A. *et al.* Ancient evolutionary origin of diversified variable regions demonstrated by crystal structures of an immune-type receptor in amphioxus. *Nature Immunol.* **7**, 875–882 (2006).
63. Yu, C. *et al.* Genes 'waiting' for recruitment by the adaptive immune system: the insights from amphioxus. *J. Immunol.* **174**, 3493–3500 (2005).
64. Suzuki, T., Shin, I., Fujiyama, A., Kohara, Y. & Kasahara, M. Hagfish leukocytes express a paired receptor family with a variable domain resembling those of antigen receptors. *J. Immunol.* **174**, 2885–2891 (2005).
65. Pancer, Z., Mayer, W. E., Klein, J. & Cooper, M. D. Prototypic T cell receptor and CD4-like coreceptor are expressed by lymphocytes in the agnathan sea lamprey. *Proc. Natl Acad. Sci. USA* **101**, 13273–13278 (2004).
66. Thompson, C. B. New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity* **3**, 531–539 (1995).
67. Fugmann, S. D., Messier, C., Novack, L. A., Cameron, R. A. & Rast, J. P. An ancient evolutionary origin of the *Rag1/2* gene locus. *Proc. Natl Acad. Sci. USA* **103**, 3728–3733 (2006).
This paper demonstrates that RAG1 and RAG2 were present in echinoderms (sea urchins) approximately 100 million years before the emergence of the AIS.
68. Kapitonov, V. V. & Jurka, J. RAG1 core and V(D)J recombination signal sequences were derived from *Transib* transposons. *PLoS Biol.* **3**, e181 (2005).
This work shows that a transposable element, transib, contains the RAG1 core element and is found in many animal species, including invertebrates.
69. Dreyfus, D. H. Paleo-immunology: evidence consistent with insertion of a primordial herpes virus-like element in the origins of acquired immunity. *PLoS ONE* **4**, e5778 (2009).
70. Liu, Y., Subrahmanyam, R., Chakraborty, T., Sen, R. & Desiderio, S. A plant homeodomain in RAG-2 that binds hypermethylated lysine 4 of histone H3 is necessary for efficient antigen–receptor–gene rearrangement. *Immunity* **27**, 561–571 (2007).
71. Klein, J. & Nikolaidis, N. The descent of the antibody-based immune system by gradual evolution. *Proc. Natl Acad. Sci. USA* **102**, 169–174 (2005).
72. Flajnik, M. F. & Du Pasquier, L. Evolution of innate and adaptive immunity: can we draw a line? *Trends Immunol.* **25**, 640–644 (2004).
73. Ohno, S. *Evolution by Gene Duplication* (Springer, 1970).
74. Furlong, R. F. & Holland, P. W. Were vertebrates octoploid? *Phil. Trans. R. Soc. Lond. B* **357**, 531–544 (2002).
75. Panopoulou, G. & Poustka, A. J. Timing and mechanism of ancient vertebrate genome duplications — the adventure of a hypothesis. *Trends Genet.* **21**, 559–567 (2005).
76. Kasahara, M. The 2R hypothesis: an update. *Curr. Opin. Immunol.* **19**, 547–552 (2007).
77. Van de Peer, Y., Maere, S. & Meyer, A. The evolutionary significance of ancient genome duplications. *Nature Rev. Genet.* **10**, 725–732 (2009).
78. Kasahara, M., Nakaya, J., Satta, Y. & Takahata, N. Chromosomal duplication and the emergence of the adaptive immune system. *Trends Genet.* **13**, 90–92 (1997).
79. Kasahara, M., Suzuki, T. & Du Pasquier, L. On the origins of the adaptive immune system: novel insights from invertebrates and cold-blooded vertebrates. *Trends Immunol.* **25**, 105–111 (2004).
80. Okada, K. & Asai, K. Expansion of signaling genes for adaptive immune system evolution in early vertebrates. *BMC Genomics* **9**, 218 (2008).
This paper provides a comprehensive analysis of ohnologues that are involved in the AIS of jawed vertebrates.
81. Kasahara, M. What do the paralogous regions in the genome tell us about the origin of the adaptive immune system? *Immunol. Rev.* **166**, 159–175 (1998).
82. Kasahara, M. *et al.* Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. *Proc. Natl Acad. Sci. USA* **93**, 9096–9101 (1996).
83. Katsanis, N., Fitzgibbon, J. & Fisher, E. M. Paralogy mapping: identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel *PBX* and *NOTCH* loci. *Genomics* **35**, 101–108 (1996).
84. Vienne, A. *et al.* Evolution of the proto-MHC ancestral region: more evidence for the plesiomorphic organisation of human chromosome 9q34 region. *Immunogenetics* **55**, 429–436 (2003).
85. Oliński, R. P., Lundin, L. G. & Hallbook, F. Conserved synteny between the *Ciona* genome and human paralogous identifies large duplication events in the molecular evolution of the insulin-relaxin gene family. *Mol. Biol. Evol.* **23**, 10–22 (2006).
Based on the comparative analysis of the insulin-relaxin family in humans and *C. intestinalis*, this paper proposes that the MHC paralogy group and the neurotrophin paralogy group were derived from a single contiguous region on an invertebrate proto-chromosome.
86. Abi-Rached, L., Gilles, A., Shiina, T., Pontarotti, P. & Inoko, H. Evidence of *en bloc* duplication in vertebrate genomes. *Nature Genet.* **31**, 100–105 (2002).
87. Azumi, K. *et al.* Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: 'waiting for Godot'. *Immunogenetics* **55**, 570–581 (2003).
88. Holland, L. Z. *et al.* The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res.* **18**, 1100–1111 (2008).
This paper provides a comprehensive analysis of the genes that encode the major biological systems in amphioxus. Among the genes analysed are those involved in host defence.
89. Hallbook, F., Wilson, K., Thorndyke, M. & Oliński, R. P. Formation and evolution of the chordate neurotrophin and *Trk* receptor genes. *Brain Behav. Evol.* **68**, 133–144 (2006).
90. Belov, K. *et al.* Reconstructing an ancestral mammalian immune supercomplex from a marsupial major histocompatibility complex. *PLoS Biol.* **4**, e46 (2006).
91. Kasahara, M. Genome dynamics of the major histocompatibility complex: insights from genome paralogy. *Immunogenetics* **50**, 134–145 (1999).
92. Marchalonis, J. J. & Edelman, G. M. Phylogenetic origins of antibody structure. III. Antibodies in the primary immune response of the sea lamprey, *Petromyzon marinus*. *J. Exp. Med.* **127**, 891–914 (1968).
93. Finstad, J. & Good, R. A. The evolution of the immune response. III. Immunologic responses in the lamprey. *J. Exp. Med.* **120**, 1151–1168 (1964).
94. Litman, G. W., Finstad, F. J., Howell, J., Pollara, B. W. & Good, R. A. The evolution of the immune response. III. Structural studies of the lamprey immunoglobulin. *J. Immunol.* **105**, 1278–1285 (1970).
95. Linthicum, D. S. & Hildemann, W. H. Immunologic responses of Pacific hagfish. III. Serum antibodies to cellular antigens. *J. Immunol.* **105**, 912–918 (1970).
96. Mayer, W. E. *et al.* Isolation and characterization of lymphocyte-like cells from a lamprey. *Proc. Natl Acad. Sci. USA* **99**, 14350–14355 (2002).
97. Uinuk-Ool, T. *et al.* Lamprey lymphocyte-like cells express homologs of genes involved in immunologically relevant activities of mammalian lymphocytes. *Proc. Natl Acad. Sci. USA* **99**, 14356–14361 (2002).
98. Suzuki, T., Shin, I., Kohara, Y. & Kasahara, M. Transcriptome analysis of hagfish leukocytes: a framework for understanding the immune system of jawless fishes. *Dev. Comp. Immunol.* **28**, 993–1003 (2004).
99. Pancer, Z. & Cooper, M. D. The evolution of adaptive immunity. *Annu. Rev. Immunol.* **24**, 497–518 (2006).
100. Nagawa, F. *et al.* Antigen-receptor genes of the agnathan lamprey are assembled by a process involving copy choice. *Nature Immunol.* **8**, 206–213 (2007).
This paper is the first to analyse the genetic mechanisms involved in VLR gene assembly. It proposes that LRR modules are tethered by a copy-choice mechanism.
101. Alder, M. N. *et al.* Diversity and function of adaptive immune receptors in a jawless vertebrate. *Science* **310**, 1970–1973 (2005).
102. Kim, H. M. *et al.* Structural diversity of the hagfish variable lymphocyte receptors. *J. Biol. Chem.* **282**, 6726–6732 (2007).
103. Herrin, B. R. *et al.* Structure and specificity of lamprey monoclonal antibodies. *Proc. Natl Acad. Sci. USA* **105**, 2040–2045 (2008).
104. Han, B. W., Herrin, B. R., Cooper, M. D. & Wilson, I. A. Antigen recognition by variable lymphocyte receptors. *Science* **321**, 1834–1837 (2008).
105. Velikovsky, C. A. *et al.* Structure of a lamprey variable lymphocyte receptor in complex with a protein antigen. *Nature Struct. Mol. Biol.* **16**, 725–730 (2009).

106. Alder, M. N. *et al.* Antibody responses of variable lymphocyte receptors in the lamprey. *Nature Immunol.* **9**, 319–327 (2008).
107. Guo, P. *et al.* Dual nature of the adaptive immune system in lampreys. *Nature* **459**, 796–801 (2009). **This paper indicates that lampreys have two lineages of lymphoid cells that resemble T and B cells of jawed vertebrates.**
108. Pancer, Z. *et al.* Variable lymphocyte receptors in hagfish. *Proc. Natl Acad. Sci. USA* **102**, 9224–9229 (2005).
109. 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. *Immunogenetics* **59**, 329–331 (2007).
110. Haruta, C., Suzuki, T. & Kasahara, M. Variable domains in hagfish: *NICIR* is a polymorphic multigene family expressed preferentially in leukocytes and is related to lamprey *TCR-like*. *Immunogenetics* **58**, 216–225 (2006).
111. Yokoyama, W. M. & Plougastel, B. F. Immune functions encoded by the natural killer gene complex. *Nature Rev. Immunol.* **3**, 304–316 (2003).
112. Takahashi, T. *et al.* Natural killer cell receptors in the horse: evidence for the existence of multiple transcribed *LY49* genes. *Eur. J. Immunol.* **34**, 773–784 (2004).
113. Bell, J. J. & Bhandoola, A. The earliest thymic progenitors for T cells possess myeloid lineage potential. *Nature* **452**, 764–767 (2008).
114. Boehm, T. One problem, two solutions. *Nature Immunol.* **10**, 811–813 (2009).
115. Wada, H. *et al.* Adult T-cell progenitors retain myeloid potential. *Nature* **452**, 768–772 (2008).
116. Litman, G. W., Cannon, J. P. & Dishaw, L. J. Reconstructing immune phylogeny: new perspectives. *Nature Rev. Immunol.* **5**, 866–879 (2005).
117. Beutler, B. *et al.* Genetic analysis of resistance to viral infection. *Nature Rev. Immunol.* **7**, 753–766 (2007).
118. Hedrick, S. M. The acquired immune system: a vantage from beneath. *Immunity* **21**, 607–615 (2004).
119. Solem, S. T. & Stenvik, J. Antibody repertoire development in teleosts — a review with emphasis on salmonids and *Gadus morhua* L. *Dev. Comp. Immunol.* **30**, 57–76 (2006).
120. Matsunaga, T. & Rahman, A. What brought the adaptive immune system to vertebrates? — The jaw hypothesis and the seahorse. *Immunol. Rev.* **166**, 177–186 (1998).
121. Kaufman, J., Volk, H. & Wallny, H. J. A 'minimal essential Mhc' and an 'unrecognized Mhc': two extremes in selection for polymorphism. *Immunol. Rev.* **143**, 63–88 (1995).
122. Flajnik, M. F. Churchill and the immune system of ectothermic vertebrates. *Immunol. Rev.* **166**, 5–14 (1998).
123. Tasumi, S. *et al.* High-affinity lamprey VLRA and VLRB monoclonal antibodies. *Proc. Natl Acad. Sci. USA* **106**, 12891–12896 (2009).
124. Travis, J. Origins. On the origin of the immune system. *Science* **324**, 580–582 (2009).
125. Fall-Ngai, M. Adaptive immunity: care for the community. *Nature* **445**, 153 (2007).
126. Long, J. A., Trinajstić, K. & Johanson, Z. Devonian arthrodire embryos and the origin of internal fertilization in vertebrates. *Nature* **457**, 1124–1127 (2009).
127. Friedman, R. & Hughes, A. L. Pattern and timing of gene duplication in animal genomes. *Genome Res.* **11**, 1842–1847 (2001).
128. Hughes, A. L. & Friedman, R. 2R or not 2R: testing hypotheses of genome duplication in early vertebrates. *J. Struct. Funct. Genomics* **3**, 85–93 (2003).
129. Gibson, T. J. & Spring, J. Evidence in favour of ancient octaploidy in the vertebrate genome. *Biochem. Soc. Trans.* **28**, 259–264 (2000).
130. Lundin, L. G., Larhammar, D. & Hallbook, F. Numerous groups of chromosomal regional paralogs strongly indicate two genome doublings at the root of the vertebrates. *J. Struct. Funct. Genomics* **3**, 53–63 (2003).
131. Putnam, N. H. *et al.* The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**, 1064–1071 (2008). **This paper, written by core members of the amphioxus genome project, describes the salient features of the amphioxus genome, with emphasis on genome evolution. It provides strong evidence to support the 2R hypothesis.**
132. Kuraku, S., Meyer, A. & Kuratani, S. Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? *Mol. Biol. Evol.* **26**, 47–59 (2009). **This paper proposes that, contrary to the commonly held view, both the first and second rounds of WGD occurred after the emergence of a common ancestor of vertebrates and before the divergence of jawed and jawless vertebrates.**
133. Karre, K. NK cells, MHC class I molecules and the missing self. *Scand. J. Immunol.* **55**, 221–228 (2002).
134. Jaillon, O. *et al.* Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **431**, 946–957 (2004).
135. Vandepoele, K., De Vos, W., Taylor, J. S., Meyer, A. & Van de Peer, Y. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. *Proc. Natl Acad. Sci. USA* **101**, 1638–1643 (2004).
136. Blair, J. E. & Hedges, S. B. Molecular phylogeny and divergence times of deuterostome animals. *Mol. Biol. Evol.* **22**, 2275–2284 (2005).
137. Kasahara, M. The chromosomal duplication model of the major histocompatibility complex. *Immunol. Rev.* **167**, 7–32 (1999).
138. Horton, R. *et al.* Gene map of the extended human MHC. *Nature Rev. Genet.* **5**, 889–899 (2004).

Acknowledgements

M.F.F. has been funded by the US National Institutes of Health grants AI027877 and RR006603. M.K. has been funded by a KAKENHI grant for the priority area 'Comparative Genomics' from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Competing interests statement

The authors declare no competing financial interests.

DATABASES

Entrez Gene: <http://www.ncbi.nlm.nih.gov/gene/RAG1> | [RAG2](#)

FURTHER INFORMATION

Masanori Kasahara's homepage: <http://www.med.hokudai.ac.jp/en/dept/outline/path/index.html>
Pre-Ensembl genome database: [http://pre.ensembl.org/Lamprey_\(Petromyzon_marinus\)_draft_genome_sequence](http://pre.ensembl.org/Lamprey_(Petromyzon_marinus)_draft_genome_sequence)

SUPPLEMENTARY INFORMATION

See online article: [S1 \(figure\)](#) | [S2 \(table\)](#)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF