ANNUAL REVIEWS

Annual Review of Immunology

Unfinished Business: Evolution of the MHC and the Adaptive Immune System of Jawed Vertebrates

Jim Kaufman[1,2]

[1]Department of Pathology, University of Cambridge, Cambridge CB2 1QP, United Kingdom

[2]Department of Veterinary Medicine, University of Cambridge, Cambridge CB2 0ES, United Kingdom; email: jfk31@cam.ac.uk

Annu. Rev. Immunol. 2018. 36:383–409

The *Annual Review of Immunology* is online at immunol.annualreviews.org

https://doi.org/10.1146/annurev-immunol-051116-052450

Keywords

coevolution, genome-wide duplication, synteny, natural killer, T cell, antibody

Abstract

The major histocompatibility complex (MHC) is a large genetic region with many genes, including the highly polymorphic classical class I and II genes that play crucial roles in adaptive as well as innate immune responses. The organization of the MHC varies enormously among jawed vertebrates, but class I and II genes have not been found in other animals. How did the MHC arise, and are there underlying principles that can help us to understand the evolution of the MHC? This review considers what it means to be an MHC and the potential importance of genome-wide duplication, gene linkage, and gene coevolution for the emergence and evolution of an adaptive immune system. Then it considers what the original antigen-specific receptor and MHC molecule might have looked like, how peptide binding might have evolved, and finally the importance of adaptive immunity in general.

INTRODUCTION

The major histocompatibility complex (MHC) is a genetic region that encodes a wide variety of molecules, including the classical MHC class I and II molecules that are central to the adaptive immune response of jawed vertebrates. These classical MHC molecules present peptides to thymus-derived (T) lymphocytes, and class I molecules are also recognized by natural killer (NK) cells. The importance of T cells and NK cells for survival from pathogens and tumors, as well as their role in autoimmunity and tissue graft rejection, has led to decades of investigation, discussed in numerous review articles and presented in all standard textbooks, revealing systems of enormous complexity and richness of detail. However, key genes encoding the MHC class I and II molecules, the αβ and γδ T cell receptors (TCRs), and B cell receptors (BCRs)/antibodies, have not been found in the closest relatives of the jawed vertebrates (the jawless fish, also called agnathans or cyclostomes) or in any invertebrate.

So, how did the adaptive immune system of the jawed vertebrates arise? Some important insights have come from simply considering data from humans and biomedical models like mice. Other insights have come from comparative immunologists, who initially examined animals with functional assays like graft transplantation, mixed lymphocyte reaction, and immunization for antibody production, and with immunoprecipitation to look at molecules. Really rapid progress came with the application of molecular biology techniques to a wide variety of vertebrates, with an explosion of discoveries that have led to many speculative hypotheses, for which evidence for or against has been sought. However, among nonmammalian jawed vertebrates, there has been little biochemical or functional examination, except in food animals like chicken and Atlantic salmon, owing to their economic importance.

At this moment, there is much fragmentary data but little consensus on many aspects of the evolution of the MHC and the adaptive immune system. This review focuses on the evolution of the MHC, much of it with lessons from the class I system, because class II molecules and functions in nonmammalian vertebrates have been examined far less. There are many other important issues about the origin and evolution of the adaptive immune system for which there is no room in this review to discuss. These topics include the details of transposons and recombination activating genes (RAGs), new antigen receptors (NARs), IgM and IgW/IgD, the extra NAR-like V domain on TCRs, germinal centers and affinity maturation, APOBEC and isotype switching, and innate immune cells, for each of which there are many interesting papers and some excellent reviews.

WHAT IS AN MHC?

The MHC was originally described in mice (and then humans) as the genetic locus determining the fastest rejection of tissue grafts between individuals within a species (that is, allograft rejection), with the genes responsible being the classical MHC class I and II genes (1), now having been found throughout the jawed vertebrates. These key genes encode heterodimeric glycoproteins that bind protein fragments (peptides) inside the cell and bring them to the cell surface for recognition. The heavy (α) chains of classical class I molecules, class II α chains (from A genes), and β chains (from B genes) are encoded in the MHC; but the gene for β_2-microglobulin (β_2m), which associates with the class I heavy chain, is located outside of the MHC in most jawed vertebrates. Classical class I molecules present peptides originating from proteins in the cytoplasm and nucleus to cytotoxic CD8 T cells (as well as NK cells), while classical class II molecules present peptides derived from proteins in intracellular vesicles and the extracellular space to CD4 T cells. These classical MHC molecules depend on complex pathways of biosynthesis, antigen degradation, translocation, and peptide loading, for which some of the components are encoded by the MHC; these include

the inducible proteasome components (LMPs or PSMBs), transporters for antigen presentation (TAP1 and TAP2), and peptide editor and chaperone tapasin for class I molecules as well as peptide editor and chaperone DM and inhibitor DO for class II molecules (1–4).

In addition to binding and presenting peptides, classical MHC molecules generally have high allelic polymorphism and sequence diversity (1–3). This high polymorphism reflects a molecular arms race between pathogens that evade recognition and hosts that can survive based on their MHC molecules (5, 6). There is evidence for other selective pressures for polymorphism, such as mating preference or reproductive success (7, 8).

Many other genes can be found near the classical MHC genes, including nonclassical MHC class I and II genes that may or may not have immune function and generally are not highly polymorphic (2, 3). Exceptional nonclassical class I molecules, such as MIC-A and MIC-B, are polymorphic, presumably owing to an arms race with pathogens (9, 10). The classical MHC genes appear to be ancient, with different nonclassical molecules appearing (and disappearing) at different times in the evolution of jawed vertebrates (11).

The textbook view of the MHC (2, 12, 13) is based on the HLA region on human chromosome 6 (**Figure 1**, **Supplemental Figure 1**), divided into a class I region with classical and nonclassical class I genes along with unrelated framework genes (including innate immune genes such as TRIMs), and a class II region with classical and nonclassical class II genes along with genes involved in class I antigen processing (TAP1 and TAP2, and LMPs/PSMBs) as well as a kinase (RING3/BRD2). In between is a class III region with many different kinds of genes, including some involved in immunity (such as complement components C4, C2, and factor B and cytokines of the tumor necrosis factor family, TNFs). On either side are the extended MHC regions, with the class I chaperone and peptide editor tapasin as well as many other kinds of genes (including the RXR and Notch signaling receptors) in the extended class II region and butyrophilins, olfactory receptors, and the nonclassical class I molecule HFE (involved in iron uptake) as well as other kinds of genes in the extended class I region. The cohesiveness of this region could be questioned, but many of these genes may have been part of a primordial immune region, with the descendent region still containing as many as 40% of the original genes involved in some way with immunity (2).

Supplemental Material

The organization of the MHC in other species varies enormously (14) (**Figure 1**, **Supplemental Figure 1**). Even in biomedical model species such as mice and rats (15, 16), there are differences compared to humans, including classical class I gene(s) between the class II and the extended class II regions, and several large families of nonclassical class I genes (many with functions unlike those found in humans) in the extended class I region. Other important differences are found in cattle, sheep, and pigs (17–19). There seem to be several organizations among marsupials (20–24). In the gray short-tailed opossum (*Monodelphis domestica*) and the Tamar wallaby (*Macropus eugenii*), the class III region is outside of the MHC, whereas in the Tasmanian devil (*Sarcophilus harrisii*), the different regions appear to be duplicated. The class I and II genes along with multiple copies of TAP and LMP/PMBS genes are interspersed in the opossum, giving no single class I or II region. The framework genes that in placental mammals divide up the class I region are found together in a single region in the opossum, but without class I genes. Most surprisingly, the MHC of the wallaby has no classical class I genes, which instead are located at the telomeres of several chromosomes.

In the largest group of birds, the passerines (which include songbirds among the perching birds), there are generally large families of class I and II genes, whereas in the nonpasserines there are many fewer genes in a region that is relatively small and simple (25). In most birds, the current genomic assemblies are fragmented (26–33) and clear conclusions may be hard to reach (compare the location of the class I locus in the zebra finch, *Taeniopygia guttata*; 26, 27). However, in the chicken (*Gallus gallus*), the BF-BL region within the B locus was found to be simple and

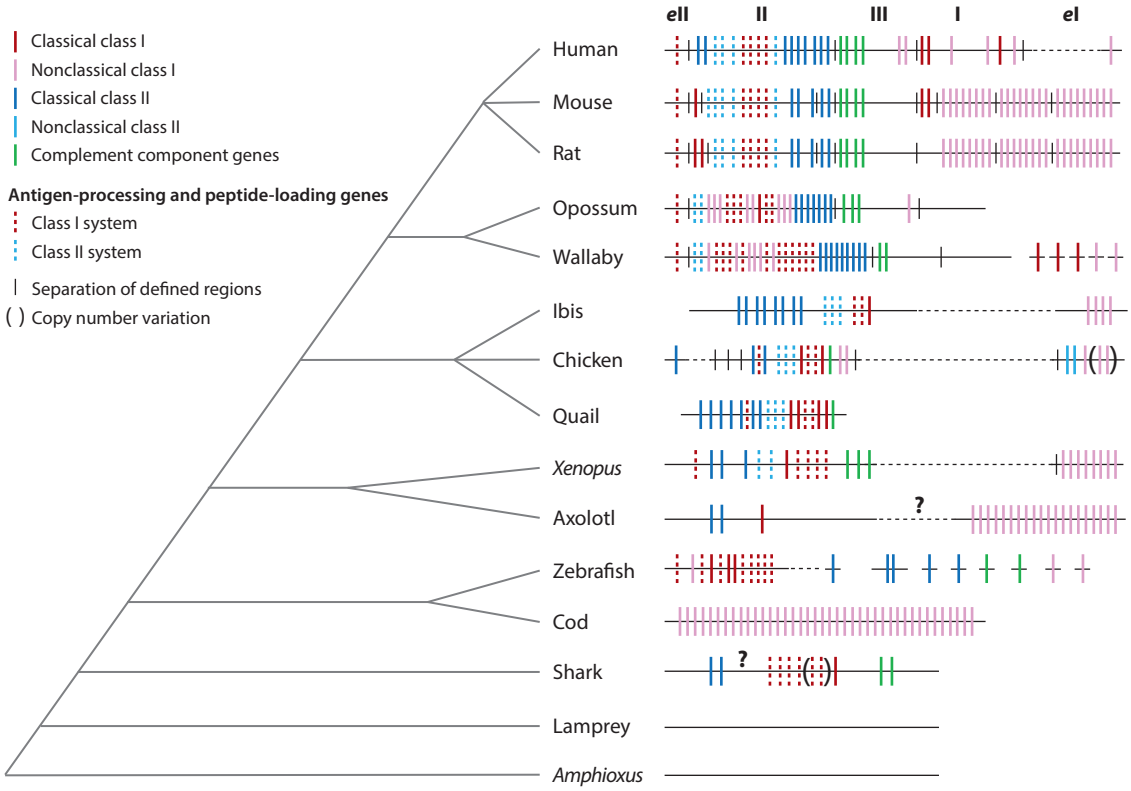

Figure 1

Only jawed vertebrates have an MHC, and the organization can vary significantly. This idealized phylogenetic tree shows the relationships of some of the organisms discussed in the text, next to idealized representations of some of the genes in the MHC syntenic region (and for some organisms, the equivalent MHC paralogous regions). Solid horizontal lines indicate contiguous sequences (proven; inferred from homology of scaffolds, as in *Xenopus*; or inferred from mapping, as in the nurse shark); dotted horizontal lines indicate considerable genomic sequences without MHC genes; question marks indicate uncertain linkage. Solid horizontal lines indicate MHC class I and II genes (*red*, classical class I; *pink*, nonclassical class I; *dark blue*, classical class II; *light blue*, nonclassical class II), as well as complement component genes (*green*). Dotted vertical lines indicate antigen-processing and peptide-loading genes (*red*, class I system; *blue*, class II system). Thin vertical lines indicate separation of defined regions, which are named only for the human MHC (*e*II, extended class II region; II, class II region; III, class III region; I, class I region; *e*I, extended class I region). Brackets indicate copy number variation. Data from References 2, 15, 16, 20, 23, 24, 32, 34, 39, 41, 45, 49, 53–60.

compact, with the class II B genes next to the DM genes, the TAP genes flanked by classical class I genes and the class III region outside of the class I and class II regions, characterized therefore as a "minimal essential MHC" (34). Since then, further work has described an extended class I region with TRIM and BG genes (related to butyrophilins) on one side of the BF-BL region, a region with the nonclassical CD1 genes located on the other, and a Y region of nonclassical class I and II genes at a distance on the same chromosome and genetically unlinked owing to high recombination (35–40). The equivalent of the BF-BL region in Japanese quail (*Coturnix japonica*) is organized in a similar way as in the chicken, but with several duplications of genes (41) (**Figure 1**).

The genomic studies in reptiles are incomplete and poorly assembled, but they have been interpreted to show various organizations of MHC genes (42–44). Like that of the chicken, the MHC of *Xenopus* frogs is organized with the antigen-processing and peptide-loading genes next to the classical class I gene, with the class III region on the outside (45), and with a region of

nonclassical (XNC) class I genes located on the same chromosome but at some distance (46, 47) (**Figure 1**). These XNC genes are recognized by T cells with semi-invariant TCRs, much like CD1 and MR1 are in mammals (48). Among salamanders, the Mexican axolotl (*Ambystoma mexicanum*) is reported to have tens of class I genes and a single predominant class II gene pair (49, 50).

Unlike tetrapods (amphibians, reptiles, birds, and mammals), which apparently have a single phylogenetic origin, fish represent several ancient lineages that have been separated for very long periods of time (**Figure 1**). The largest group, teleosts (ray-finned fish), generally have classical class I gene(s) along with TAP, tapasin, and LMP/PSMB genes in a single region, with several lineages of nonclassical class I genes located outside of the classical class I region (51–54). Class II genes are located in discrete regions on other chromosomes, and the class III region genes are scattered throughout the genome, although some stay together (54–56). As a result, some researchers use the term MH regions rather than MHC regions. The Atlantic cod (*Gadus morhua*) and related fish have no class II genes but hundreds of class I genes, some of which are proposed to take on class II function (57). Much less is known for other groups of fish, although poorly assembled genome sequences give hints of surprises to come. At the base of the jawed vertebrates, cartilaginous fish, as represented by the nurse shark (*Ginglymostoma cirratum*), have a single MHC with many genes found in the tetrapod MHC, including classical class I and II genes as well as class III genes such as the complement component C4 (58–60). However, unlike all other vertebrates, the nurse shark β_2-microglobulin (β_2m) gene is located within the MHC (61).

With the inherent plasticity of genomes and the enormous spans of time involved, it is perhaps not surprising that the MHC has undergone many changes resulting in the various organizations found among jawed vertebrates. Perhaps it is more surprising that the MHC has held together as well as it has. Some researchers perceive the possibility of a selective advantage in keeping these genes together, while others attribute the apparent cohesiveness to relatively low levels of nonhomologous recombination (2, 3). Some researchers consider all genes (including olfactory genes and the class III region) as bone fide MHC genes, either because they may be important for MHC-linked phenomena or because "MHC" is a convenient abbreviation (for instance, 38, 62). In this latter view, any region with genes related to those found in the MHC might be called an MHC region; for instance, the human MHC paralogous regions discussed below might be called MHC-1, MHC-6, MHC-9, and MHC-19. In any case, there is value in considering genes that may have been part of this region before class I and class II genes arose, allowing the evolutionary history of the region to be followed.

So, what is an MHC? Based on the concept of coevolution (63, 64) discussed below, this review considers classical class I and II genes, along with the genes located within the MHC that are part of the antigen-processing and peptide-loading pathways, as biologically coherent units that anchor the MHC based on function. Other genes are considered as part of the "MHC syntenic region" (a genomic region of colocated genes), free to come and go in evolution without deeper biological meaning. In this view, nonclassical class II genes that are involved in class II antigen presentation, such as DM and DO, are part of the MHC, whereas nonclassical class I genes may or may not be located in the MHC syntenic region.

MHC PARALOGOUS REGIONS, GENOME-WIDE DUPLICATIONS, AND THE BIG BANG

Many genes located outside of the MHC are similar to those found within the MHC, and some of these are located more or less together in parts of the genome that have become known as MHC paralogous regions. The appearance of these MHC paralogous regions (and many other such regions) is generally attributed to genome-wide duplication (65–67), first suggested by Susumu

Ohno some 50 years ago. He proposed that the appearance of duplicate genes (called paralogues, or by some ohnologs) would allow new functions to evolve without losing the old functions (68). As two rounds (2R) of these genome-wide duplication events occurred at the base of the vertebrates, it was suggested some 20 years ago that the sudden appearance of many duplicate genes without essential functions allowed a burst of evolution that led to the emergence of the adaptive immune system (widely called the Big Bang; 69, 70). Since then, the hypothesis of genome-wide duplications has become widely accepted, but many facts have emerged that modify or challenge the original notions about the evolution of the adaptive immune system, so that the whole area remains rather unsettled.

Supplemental Material

2R would yield four MHC regions, and initial observations from the human genome identified the MHC on chromosome 6 and MHC paralogous regions on chromosomes 1, 9, and 19 (**Supplemental Figure 2**). Genes from the MHC syntenic region have clear paralogues spread throughout the MHC, from the class II region (like PSMBs and RING3/BRD2), the class III region (like C4), and the extended MHC regions (like RXR and Notch). Many of these and other MHC paralogous genes have been traced back to various invertebrate genomes, including those of the protochordate *Amphioxus*, and as far back as a primitive multicellular organism, the placozoan *Trichoplax adhaerens* (71, 72). The relative proximity of these genes in particular regions has been used to support the notion of a proto-MHC syntenic region, proposed to be involved in stress responses and innate immunity (72).

The current consensus is that 2R took place at the base of the vertebrates around 500 Mya, with the lamprey genome currently interpreted to mean that both events happened before the emergence of the jawless fish (73). In nearly all vertebrates, only one of the MHC paralogous regions is responsible for graft rejection and other adaptive immune responses, encoding genes of both class I and class II systems. As described above, the MHC in teleost fish is fragmented into several regions, which fits with a third round (3R) in teleosts (and even 4R in salmonid fish around 80 Mya), followed by differential silencing or loss of genes in one or another of the paralogous regions (74). More recent genome-wide duplication events have been studied, particularly in frogs related to the laboratory model *Xenopus laevis*, from *Xenopus tropicalis* (considered basal diploids among these amphibians) to the "dodecaploid" *Xenopus ruwenzoriensis*. A single MHC is found in all these frogs except the dodecaploid (75, 76), consistent with the notion that there might be a selective advantage to keeping MHC genes together in a single diploidized region.

The emergence of the adaptive immune system from paralogous genes free to take up new functions is a seductive hypothesis, but it remains controversial. Sadly, no descendants of the many fish groups present between the appearance of the jawless fish and the cartilaginous fish have survived, leaving a gap of some 70 million years in molecular analysis. The fossil record allowed the body plans of vertebrates through this time period to be assessed, but instead of a rapid burst of evolution, only a gradual accumulation of changes is found (77). Since MHC genes or molecules have not been identified in fossils, evolutionary changes must be inferred from surviving contemporary genomes.

Various explanations for the presence of class I genes in MHC paralogous regions have been proposed (**Figure 2**). The human class I genes include the classical class I genes (HLA-A, -B, and -C) along with nonclassical class I genes (HLA-E, -F, and -G and HFE) in and around the MHC syntenic region on human chromosome 6, nonclassical class I genes (CD1A, B, C, D, and E as well as MR1) on chromosome 1, and the nonclassical class I gene FcRn on chromosome 19 (78–80). Two models suggest that class I genes were present in the primordial MHC before 2R. However, FcRn and MR1 orthologous genes have not been identified outside of mammals, nor CD1 genes outside of mammals, birds, and some reptiles. Moreover, both CD1 genes in chickens are part of the BF-BL region, which includes the MHC (37), although fragmentary reptile genome sequences

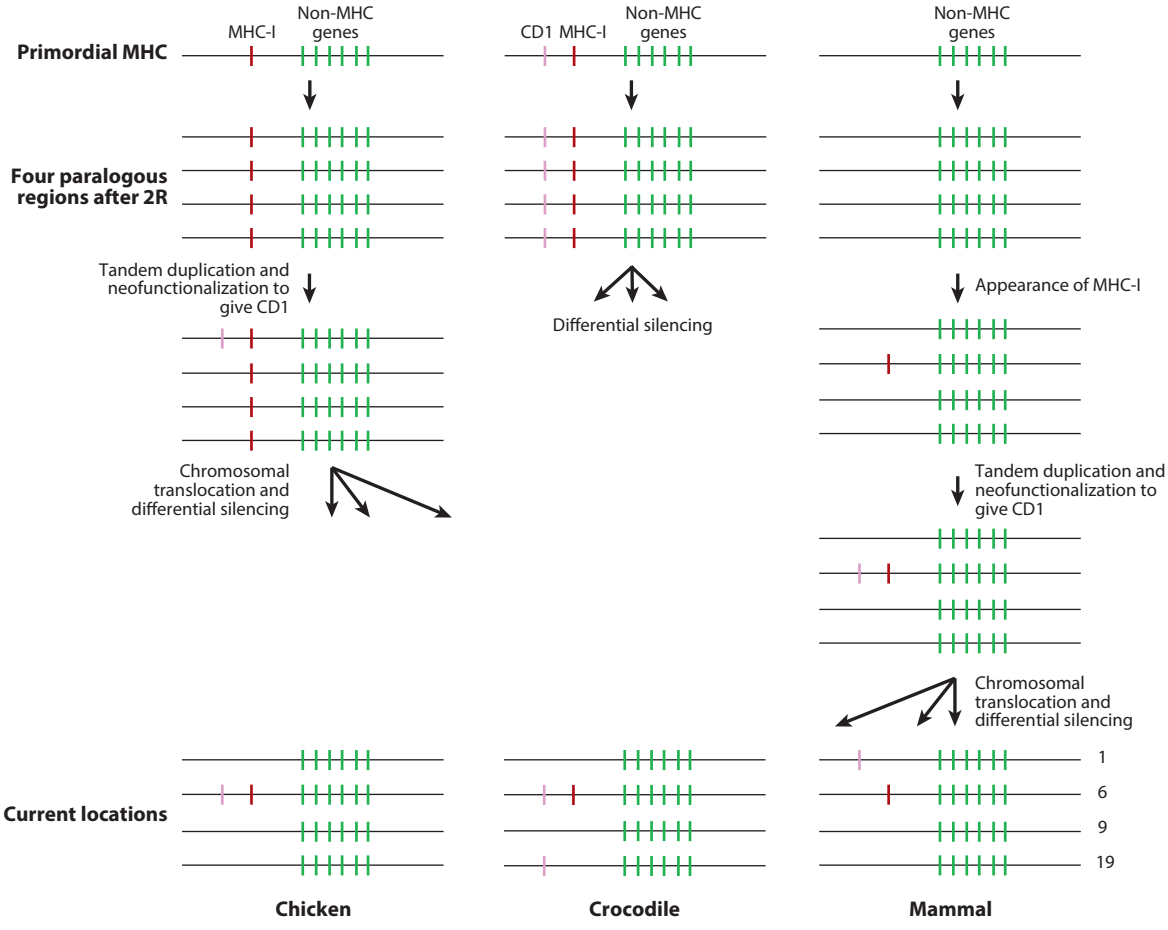

Figure 2

Three models proposed to explain class I genes in MHC paralogous regions. Classical class I genes in red, CD1 genes in pink, other genes in the MHC syntenic region (non-MHC genes) in green. The presence of CD1 genes on two chromosomes in crocodiles is inferred from the location of similar genes in mammals; however, the scaffolds for all these genes may in fact be next to each other on one chromosome. Figure modified from Reference 33 with permission.

might suggest several locations for CD1 genes (81). A third model suggests that the primordial MHC genes appeared in one MHC paralogous region after 2R, and that some nonclassical class I genes translocated near to other MHC paralogous regions in the lineage leading to mammals (82, 83).

Thus far, the available evidence does not decisively reject any of these scenarios. A careful consideration (33) shows that the MHC paralogous regions are vast (21 to 76 Mb) compared to those of the MHC (3.4 Mb), with the paralogous genes located amid many genes that are not obviously paralogous, and with the nonclassical class I genes located near the edges of the paralogous regions (**Supplemental Figure 2**). It is conceivable the class I genes were moved from the MHC to paralogous regions by homologous recombination involving flanking genes, but how frequently such events might occur is unclear. Comprehensive examination of the paralogous regions in a range of species would be helpful, and it also would be interesting to know whether

Supplemental Material

the 20- to 70-Mb region centered around the MHC contains genes from the MHC paralogous regions that at the moment are not identified as paralogues.

THE ANCESTRAL MHC AND THE APPEARANCE OF THE MAMMALIAN MHC

A major mystery about the mammalian MHC has been the presence of a class III region in between the class I and class II regions. Another mystery was the location of two LMP/PSMB genes (encoding inducible proteasome components to cut up proteins in the cytoplasm) and the two TAP genes (encoding the ABC transporter that pumps peptides into the lumen of the endoplasmic reticulum) in the class II region, as well as the location of the tapasin gene (encoding the dedicated chaperone and peptide editor) in the extended class II region, all far away from the class I genes that they serve. One suggestion was that the original location of classical class I genes was in between the extended class II and the class II region, as found for the mouse K and rat RT1 regions, and therefore near to the TAP, tapasin, and LMP/PSMB genes (15). However, the latest rat genomic sequences suggest that the RT1 region was a translocation from the other side of the class III region in the lineage that led to rodents, followed by silencing of the classical class I genes in the original class I region of rats (16).

The organization of the chicken BF-BL region with the class I region containing classical class I genes and TAP genes in between the class II and class III regions led to a scenario for the evolution from the primordial MHC in some long-lost ancestor to placental mammals. Key to this scenario is the concept of coevolution between structurally unrelated genes in the antigen-processing, peptide-loading, and peptide-presentation pathway. Evolutionary biologists have long discussed this concept using terms like coadaptation and supergenes, but for immunologists, co-evolution between genes that are not generally separated by recombination was first suggested for class II A/B gene pairs, then found for rat TAP and classical class I RT1A gene(s), and finally generalized from work with chicken TAP and class I genes (63, 84–86). In this view, a pathway with polymorphic interacting genes can only work effectively over generations of individuals if the polymorphic genes are closely linked in the genome (**Figure 3**). Conversely, unlinked polymorphic interacting genes give a variety of outcomes, as illustrated by the immunoglobulin-like NK receptors (killer inhibitory receptors, or KIRs) encoded in the leukocyte receptor complex (LRC) on human chromosome 19 that recognize classical class I molecules (HLA-A, -B and -C) encoded in the MHC on chromosome 6, with unfortunate combinations leading to deleterious outcomes for infection, autoimmunity and reproduction (8, 87).

Compared to rat genes, there is much stronger coevolution between chicken MHC genes that leads to truly profound mechanistic and functional consequences (39, 64, 88). Even though there are two classical class I genes in chickens, only the BF2 molecule is well expressed (**Figure 4*a***), with at least tenfold less RNA for the BF1 gene. Depending on the peptides presented by the dominantly expressed BF2 gene, a chicken MHC haplotype will confer resistance or susceptibility to a particular pathogen (89–91), leading to very strong genetic associations of the BF-BL region with infectious disease (39, 40). The chicken TAP genes located in between the two class I genes (and the tapasin gene nearby, located in between the two class II B genes) have high allelic polymorphism with moderate diversity between alleles. The peptide-translocation specificity of the TAP heterodimer is closely aligned with the peptide-binding specificity of the classical class I molecule encoded by the BF2 gene (92, 93), specifying several peptide positions and with a different specificity for each haplotype, with the poorly expressed BF1 molecule receiving few peptides (**Figure 4*b***). It is the rarity of recombination along the chicken MHC (the BF-BL region) that allows this strong coevolution between the TAP genes, the tapasin gene (94), and the well-expressed BF2 gene and leads to one particular peptide specificity for each MHC haplotype.

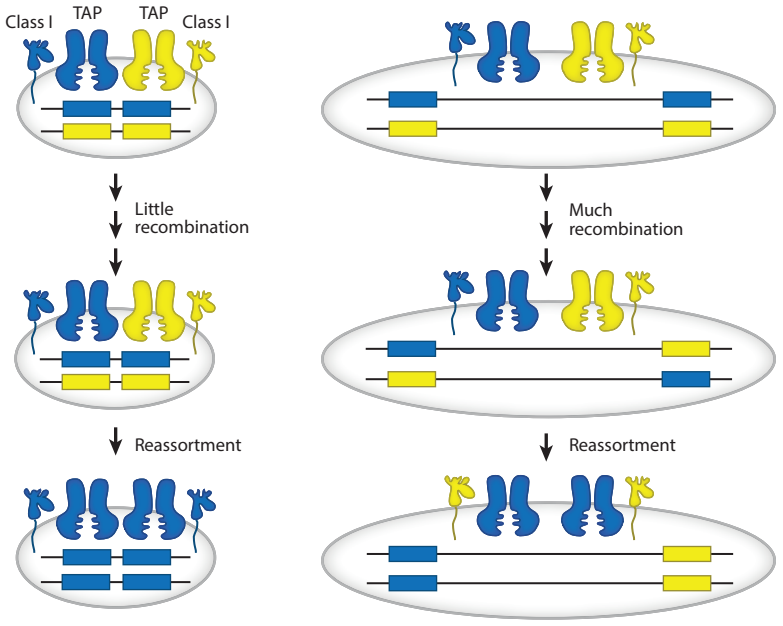

Figure 3

Close genetic linkage is required to keep advantageous combinations of genes together, as illustrated with class I and TAP alleles. Genes are represented by rectangles, with distance between rectangles representing the amount of recombination (that is, closely linked genes are close together). TAP molecules, shown in blue or yellow, pump peptides only for class I molecules drawn in the same color. Figure modified from Reference 88 with permission.

In contrast to the chicken MHC, the MHC of most placental mammals has relatively weak genetic associations with resistance and susceptibility to infectious pathogens, with monomorphic TAP, tapasin, and LMP/PSMB genes located in the class II and extended class II regions able to interact at some level with all alleles and loci of the classical class I genes located far away in the class I region (39, 64). For instance, human TAP genes pump a wide variety of peptides, of which some will be appropriate for any human classical class I molecule. As a result, each MHC haplotype will have a multigene family of class I molecules (**Figure 4*a***) that altogether confer more or less resistance to most pathogens, which can explain why MHC associations with infectious disease in humans appear relatively weak (certainly in comparison to autoimmune diseases) and have taken decades to establish convincingly (3, 95). By contrast, poultry scientists found strong genetic associations of the BF-BL region (or in many cases the larger B locus) with economically important diseases, as large numbers of chickens lived or died (40).

Although the evidence is fragmentary, it seems likely that the salient features of the chicken class I system are ancestral, being found in several key vertebrates outside of placental mammals (63, 64). As mentioned above, the class III region is on the outside of the MHC in opossum, wallaby, chicken, quail, and *Xenopus* frogs (20, 24, 34, 41, 45). At least some antigen-processing and peptide-loading genes are located near the classical class I gene(s) in opossum, chicken, quail, duck (*Anas platyrhynchos*), Oriental stork (*Ciconia boyciana*), *Xenopus* frogs, various teleost fish, and sharks, and more than one allele (or diversified families) of these genes are reported for opossum, chicken, duck, *Xenopus* frogs, and some teleost fish (20, 34, 41, 45, 51, 59, 92–94, 96–100). A single classical class I gene or haplotypes with a single class I gene are found in opossum, zebra

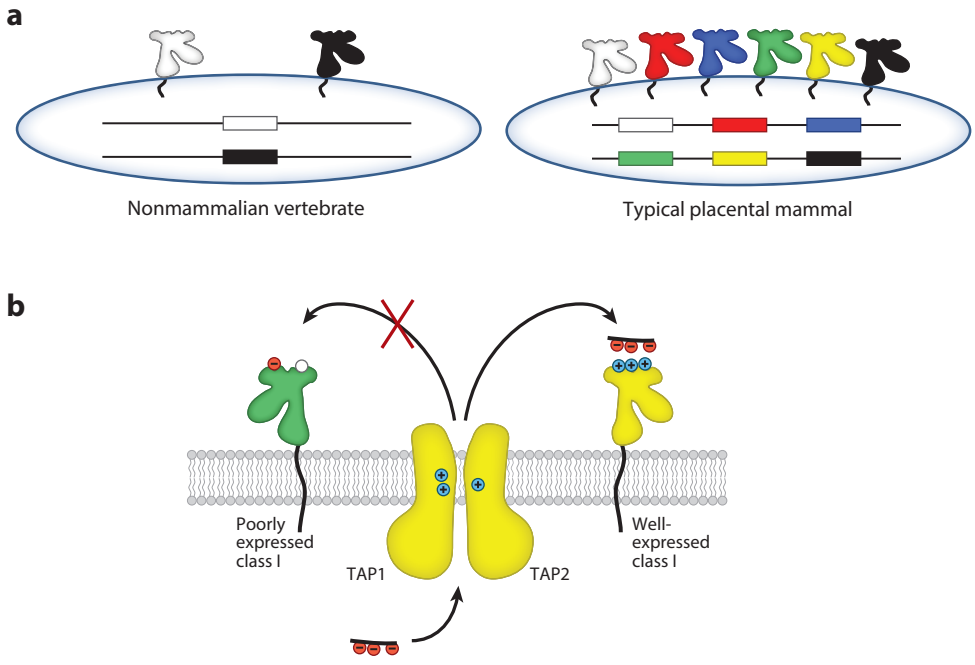

Figure 4

Coevolution to give a single class I molecule has important functional consequences. (*a*) Outside of placental mammals, many jawed vertebrates express a single classical class I molecule at a high level, so a heterozygote would express only two class I molecules. One of these two class I molecules would have to find a protective peptide for the individual to survive, and the difference between life and death determined by the MHC reads out as strong genetic associations. A typical placental mammal expresses a multigene family of class I molecules, each with a different peptide-binding specificity. Altogether, this multigene family has a good chance of finding a protective peptide, so most MHC haplotypes confer resistance to most pathogens, reading out as weak genetic associations. (*b*) Coevolution between polymorphic TAP genes and closely linked class I genes leads to a single dominantly expressed class I molecule, as illustrated by the B4 haplotype of the chicken MHC. The dominantly expressed class I molecule binds peptides with three negative anchor residues, and the TAP1-TAP2 heterodimer pumps such peptides. The poorly expressed class I molecule has a different specificity and fails to bind many peptides; this class I molecule is therefore not often recognized by T cells and eventually falls into disuse. Circles with plus signs indicate basic residues, circles with negative signs indicate acidic residues, and empty circles indicate hydrophobic residues. Panel *a* modified from Reference 178 and panel *b* modified from Reference 179, both with permission.

finch, Oriental stork, *Xenopus* frogs, Atlantic salmon (*Salmo salar*), and sharks, and a dominantly expressed class I gene in chicken, quail, duck, and house sparrow (*Passer domesticus*) (20, 26, 45, 58, 88, 89, 94, 96, 97, 101–105). There are significant genetic associations of the MHC with resistance to infectious pathogens in chicken, house sparrow, the lowland leopard frog (*Lithobates yavapaiensis*), and Atlantic salmon (40, 103, 106, 107). The obviously different organizations such as in Tamar wallaby, Atlantic cod, and Mexican axolotl may be examples of subsequent evolution from the ancestral state.

Putting this all together, the simplest scenario for the evolution of the typical mammalian MHC would involve an inversion within the ancestral MHC (92, 108). In this view (**Figure 5**), the class III region swung into the middle of the MHC and the classical class I region swung to the outside, but with the breakpoint such that polymorphic TAP, tapasin, and LMP/PSMB genes

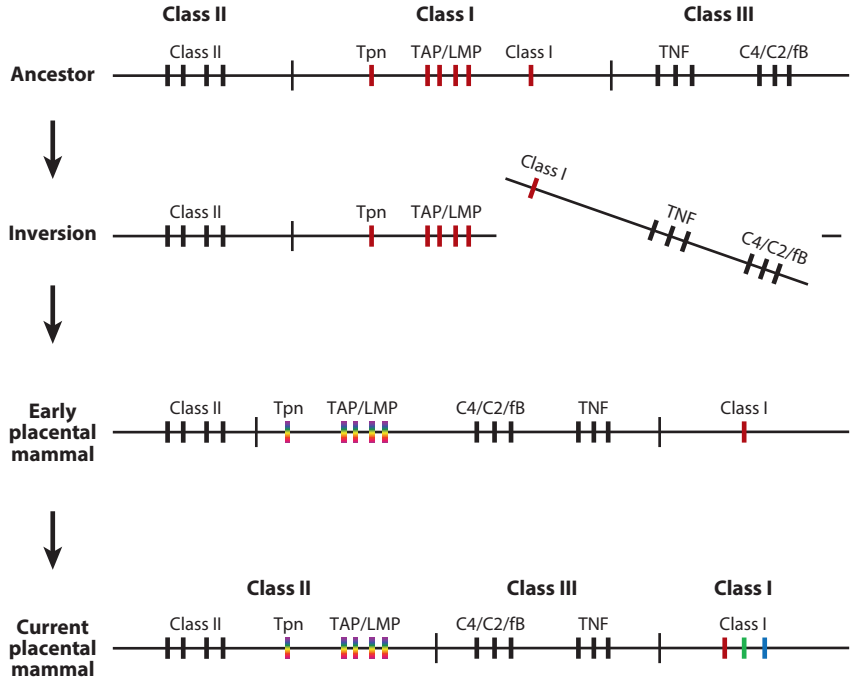

Figure 5

Textbook organization of the typical mammalian MHC can arise from the ancestral organization of nonmammalian vertebrates by an inversion. Close linkage of antigen-processing and peptide-loading genes with a single class I gene leads to coevolution with a particular specificity. After the inversion, increased recombination selects for monomorphic promiscuous antigen-processing and peptide-loading genes that can support a multigene family of class I molecules with different peptide-binding specificities. Genes are indicated by filled rectangles, all colored black except for genes that are part of the class I pathway: inducible proteasome components (low-molecular-weight proteins, LMP/PSMB), peptide transporter (transporters associated with antigen presentation, TAP), dedicated chaperone (tapasin, Tpn), and class I genes. Colors indicate specificity: red, green, and blue indicate highly specific (fastidious), while rainbow pattern indicates wide specificity (promiscuous). Other genes include tumor necrosis factor (TNF), complement component 4 (C4), C2, and factor B (fB), and class II genes.

were left behind. With the class III region in between, the strong genetic linkage between the class I gene(s) and the antigen-processing and peptide-loading genes was lost, and the coevolutionary relationships could not be sustained. At this point, there was a strong selection for the TAP, tapasin, and LMP/PSMB alleles that could function best with a wide variety of class I alleles, no matter which one might appear by recombination. In fact, alleles of such promiscuous TAP genes have been found in chickens (93). Once monomorphic genes that could work with a wide variety of class I alleles were selected, then a multigene family of well-expressed class I loci could be supported.

It appears that evolution between the two extremes of a single class I gene and a multigene family can occur. In rats, the classical class I genes apparently were translocated into the extended class II region close to the TAP genes and silenced in the class I region (16). This has allowed two lineages of TAP genes to coevolve with two lineages of class I genes, specifying amino acids in one peptide position (85, 86). Conversely, the class I genes in the Tamar wallaby have moved out of the MHC to telomeres of several chromosomes (22–24), and it would not be surprising to

find monomorphic TAPs and several well-expressed classical class I molecules. However, overall the evidence is fragmented and not very detailed outside of a few species, so careful examination in a range of animals may reveal surprises requiring additional explanation.

THE PRIMORDIAL MHC AND THE EVOLUTION OF THE ADAPTIVE IMMUNE SYSTEM

The reason for the location of the TAP, tapasin, and LMP/PSMB genes is not obvious based on the MHC of typical mammals, where they are monomorphic and located far away from the class I genes that they serve. Being monomorphic, they could have been located elsewhere in the genome (as is MECL1/LMP10/PSMB10; 66). The principle of coevolution can explain the location of polymorphic peptide-loading genes next to the classical class I gene in chickens (and presumably antigen-processing genes as well in many nonmammalian vertebrates), but it can also explain their presence in the MHC of all jawed vertebrates (88, 92, 108).

Coevolution can lead to structurally unrelated genes working together to form a pathway, in a similar way that coevolution can lead to alleles of different genes working together. The logic is the same: Close linkage in a genome means that advantageous combinations of genes stay together long enough to be selected as a group. In this way, a primordial NK receptor that did not bind peptides, an ABC transporter that did not pump peptides for MHC class I molecules, and a potential chaperone that did not hold the two together could coevolve in time to a peptide-binding MHC molecule that could interact with a tapasin molecule, which in turn interacted with a TAP heterodimer. If the genes encoding these proteins were located on different chromosomes, reassortment at every generation would mean that the chance for the interacting variants to remain in the same individual organism to allow selection would be slim. However, if these genes were located next to each other in a primordial MHC, then there would be an opportunity for the individual organism that bore them to be selected. Hence, it is likely that the pathway of antigen processing and peptide loading emerged from coevolution of closely linked genes in a primordial MHC.

A more striking speculation arose from the unexpected discovery of a lectin-like NK receptor/ligand gene pair in the chicken MHC, rather than in the NK complex (NKC) where orthologues are found in mammals (34). The presence of these genes (called B-NK and B-lec in chickens, but orthologues of NKR-P1/KLRB1 and LLT1/clr/CLEC2 in mammals) was taken to mean that the MHC and the NKC shared a common ancestral region, the primordial MHC (33, 108, 109). Other examples of unexpected genes near the MHC have been reported, notably the XMIV genes that have some similarities to TCRs being located in the *Xenopus* MHC and antibody light chain genes being located next to the elephant shark MHC (45, 110). Moreover, various other paralogous regions associated with the MHC have been proposed, including regions on human chromosomes 3, 11, 12, 14, 15, 20, and 21, which encode the tapasin-related chaperone and peptide editor TAPBPR, members of the JAM-nectin family, and the immunoglobulin-like NK receptors encoded in LRC (72, 111–115).

Although more complicated explanations might be possible (in the same sense as CD1 and MR1 potentially translocating from an MHC to an MHC paralogous region), the simplest scenario is that the primordial MHC contained both the antigen-specific receptor genes and the genes of their ligands, which could coevolve into the antigen-presentation and recognition pathway of TCRs, NK receptors, and MHC molecules (and ancillary pathways, such as antibodies) precisely because they were present in the same genomic region. A striking discovery was the presence of syntenic regions for the MHC, NKC, and LRC (based on genes not involved in adaptive immunity) found side-by-side in the urochordate *Ciona* and the protochordate *Amphioxus* (113, 116), suggesting that MHC molecules and their receptors evolved together in a primordial MHC that has been falling apart ever since.

Given the huge spans of time and the many different ways in which genomes can be broken up and reformed, it is perhaps surprising that any trace of the ancient genomic organization could be detected, but it is possible that the breakups explain some other features of contemporary adaptive immune systems. For example, it might be considered puzzling that there are separate loci for TCRβ and TCRγ gene segments, but a single intertwined locus for TCRα and TCRδ gene segments (117, 118). One speculative explanation (**Supplemental Figure 3**) is that the ancestral pair of antigen-specific receptor genes (say, TCRα and TCRβ) was present in the primordial MHC, which gave rise to three other gene pairs in paralogous regions following rounds of genome-wide duplication, and then silencing, loss, and divergence gave rise to one TCRβ locus, one TCRγ locus, and one locus that kept both TCRα and TCRδ gene segments together (with antibody H and L chain loci potentially in one or another of the paralogous loci). If 2R happened before the emergence of the adaptive immune system of jawed vertebrates (as discussed above), then gene region duplication followed by translocation, silencing, loss, and divergence would be an alternative scenario.

Supplemental Material

THE ORIGINS OF ANTIGEN-SPECIFIC RECEPTORS AND THE EVOLUTION OF PEPTIDE BINDING

The emergence of MHC molecules must be linked to the evolution of their antigen-specific receptors, the TCRs (119). All antigen-specific receptors in jawed vertebrates are heterodimers of proteins with immunoglobulin variable (V) domains in front of immunoglobulin constant (C) domains, with the sequence variation that is important for antigen binding concentrated in three loops called complementarity-determining regions (CDRs) located at one end of the V domain. One diversification mechanism that is common to all the antigen-specific receptors in jawed vertebrates is recombination (in somatic cells rather than germ line cells) that leads to rearrangement of gene segments (**Supplemental Figure 3**). It has long been postulated that an ancient transposon supplied the key features necessary for this mechanism of diversification: recombination signal sequences (RSS, typically heptamer-nonamer repeats on the end of gene segments) acted on by RAG1 and RAG2 molecules (120, 121). Such transposons have been identified, but they have also been found in the genomes of invertebrates without obvious adaptive immune systems (122, 123), so these features may be necessary but not sufficient. In any case, there seems to be little argument that the gene organization of antigen-specific receptors in jawed vertebrates is due to an insertion event by a transposon into the exon encoding a V domain in a V-C gene, followed by duplication of that gene, and a second insertion event in one of the two duplicates, eventually encoding the heterodimeric molecule (**Supplemental Figure 4**). Genes for such a heterodimeric molecule gave rise to the antigen-specific receptors in jawed vertebrates, with three obvious models based on which came first: antibody/BCR, γδ TCR, or αβ TCR.

One clearly articulated model is that αβ TCR came first, recognizing a monomorphic MHC molecule with bound peptide (124). This model is based on two facts: Contemporary jawed vertebrate receptors are diversified initially by gene rearrangements leading to variation only in CDR3 (**Supplemental Figure 4**), and many structures of TCR and MHC molecules show CDR3 interacting primarily with bound peptide, whereas CDR1 and CDR2 interact primarily with the MHC molecule (although there are exceptions; 119). To be clear, other loops in contemporary antigen-specific receptors can have high levels of diversity, but this variation arises from the presence of multiple V gene segments, or from subsequent somatic hypermutation, gene conversion, and other mechanisms of diversification. In this model, the initial selective pressure was for gene rearrangements that generated diversification to recognize a variety of peptides bound to a monomorphic MHC molecule. These αβ TCR genes then duplicated, eventually giving rise to antigen-specific

receptors that did not require recognition of an MHC molecule with bound peptide, presumably first cell-bound γδ TCRs that recognize membrane-bound molecules, and later BCRs and secreted antibodies.

An alternative view is that evolving an antigen-specific receptor to recognize a variety of molecular shapes is a simpler process, and because the CDR3 loops are located in the center of the antigen-binding site, a mechanism for diversity for that loop is the most important first step (for instance, 125). A primordial BCR gene might give rise to γδ TCR genes and then αβ TCR genes. A membrane-bound BCR with low affinity for soluble antigens but high avidity (multi-site binding affinity) for antigens on other cells (such as self-cells, bacteria, or parasites) could give rise to a soluble antibody with multiple binding sites still giving high avidity.

A primordial γδ TCR might be expected to have the same properties as suggested for a primordial BCR, with duplicates evolving to recognize MHC molecules as an αβ TCR on the one hand, and secreted as antibody on the other. However, one might instead posit that a primordial γδ TCR recognized specific cellular ligands, as do some contemporary γδ and αβ TCRs. For instance, several waves of mouse T cells with semi-invariant γδ TCRs leave the thymus to populate particular organs, and these semi-invariant T cells recognize members of the B7/butyrophilin family (126, 127). The genes for many butyrophilin family members in mammals (and BG homologues in chickens) are located in the MHC syntenic region, and the BG genes show both copy number variation and polymorphism (36, 128, 129). However, the fact that contemporary semi-invariant TCRs are selected during development after great diversification makes it difficult to understand the evolutionary selection for diversification of the first antigen-specific receptors only to recognize invariant self-ligands.

If the model for αβ TCR first is correct, how could it have come about? In particular, what could have been the ancestral molecule? One possibility is that TCR evolved from an NK cell receptor (130) that originally recognized a stress molecule (also called danger-associated molecular pattern, or DAMP) that was an ancestor of MHC molecules. In fact, NK cells are lymphocytes with many properties in common with T cells, and as mentioned above, there is evidence that originally the MHC, NKC, and LRC syntenic regions were together. Although NK cells were originally understood as recognizing the lack of self-ligands (particularly MHC class I molecules), the current view is that a balance of input from inhibiting and activating receptors on NK cells determines the outcome, with the lack of MHC molecules seen as part of a wider indication of cellular stress (or lack of homeostasis) (131).

If the MHC molecule was originally a stress ligand recognized by an NK cell receptor, how and why did it begin to bind and present peptides? One speculation derives from the fact that stressed cells change the start site of translation from acetyl-methionine to other amino acids, particularly leucine (132). A primordial MHC molecule that bound newly translated (rather than proteolyzed) proteins that began with leucine would signal stress (although the specificity also might include transport from the cytoplasm to the lumen of the endoplasmic reticulum). Presentation of such cryptic peptides by classical class I molecules to T cells, presentation of peptides with a particular first residue (formyl-methionine) by the nonclassical class I molecule encoded by the M3 gene (also called Hmt), and activation of NK cells by a peptide binding to Qa1 have all been described (10, 133–135), lending experimental credence to such a model.

THE FIRST MHC MOLECULE

Classical MHC molecules all have a peptide-binding groove with two long α helices atop an eight-stranded β sheet that is composed of two interlocking domains (encoded by separate exons), each with four β strands followed by a long α helix. Both such domains are encoded in the class I α

chain gene, with exon 2 encoding the α1 domain (with no disulfide bond) and exon 3 encoding the α2 domain (with an intradomain disulfide bond), while the equivalent domains are encoded by two class II genes, the α1 domain by exon 2 of the A gene and the β1 domain by exon 2 of the B gene. Similar intron-exon and structural organizations are found for nonclassical class I and II molecules, but not obviously in other known molecules (10, 136). So, where did MHC molecules come from?

Much of the previous discussion of how MHC molecules evolved the property of peptide binding would be solved if peptide-binding domains were acquired intact. Such a model (137), in which class I genes were derived from chaperones that bind unfolded proteins, was proposed based on a low level of sequence similarity between the α1-α2 domains of the *Xenopus* class I molecule and the C-terminal domain of HSC70 (which binds peptides in this chaperone), along with the size and the analysis of predicted secondary structure, hydrophobicity, and other properties. The model proposed that exons encoding chaperone domains were translocated in front of a β_2m-like gene to form the first MHC molecule with a class I–like structure, and that class II genes were formed by a subsequent transfer of one of the chaperone exons to another β_2m-like gene.

The notion that chaperones are the most likely origin for MHC molecules is repeated in many papers, for example, in a recent review (138) citing the protein structure of an HSC70 homologue called inducible HSP70 (139), which is encoded in the class III region of most MHC syntenic regions (2, 3). The HSC70 and HSP70 molecules contain two subdomains, one with eight β strands and the other with two long α helices (139). It is possible to imagine that such secondary structures could be rearranged during evolution to give an organization like that of MHC molecules, with domains each encoding four β strands in front of a long α helix. However, examination of the chaperone structures shows that the peptide is bound in the loops of the β strands and is covered by an α-helical lid (139), so that the mode of peptide binding as well as the interaction of the β strands and α helices are completely unlike those of contemporary MHC molecules.

An alternative model is that class II molecules came first (108, 140, 141), originally based on data from proteolysis, limited protein sequencing, and isoelectric focusing of protein domains in human class II molecules, and on considerations of symmetry. In this view (**Figure 6**), a single primordial gene encoded a class II β-like chain with two extracellular domains, which formed a homodimer. Duplication of this gene followed by evolution of one of the duplicates to encode a class II α-like chain led to a heterodimer much like contemporary class II molecules, with two extracellular domains on each chain. A similar domain organization, but rearranged to give one chain with three extracellular domains, was envisaged for class I molecules.

Subsequent protein, cDNA, and gene sequences as well as protein structures confirmed the apparent domain organization, with intron/exon organizations allowing a model for the transition from class II genes to class I genes (**Figure 6**). Class II molecules often are encoded by gene pairs in opposite transcriptional orientation, and a single inversion with asymmetric breakpoints could place the α1 exon of the class II α chain in front of the class II β chain gene. This inversion would produce a class I gene as well as a β_2m gene with a membrane-bound tail, from which soluble β_2m could be generated by subsequent evolution of a stop codon or a splice site mutation.

Other arguments have been made for a primordial class I molecule (137), based on the facts that contemporary classical class I molecules have a wide tissue distribution, are ligands for recognizing single-cell infections by both the innate immune system and the adaptive immune system (NK and T cells, respectively), and are considered more plastic in evolution (that is, able to take on a great variety of functions through the nonclassical class I molecules). By comparison, class II molecules are constitutively expressed only in certain cell types and function only for antigen presentation

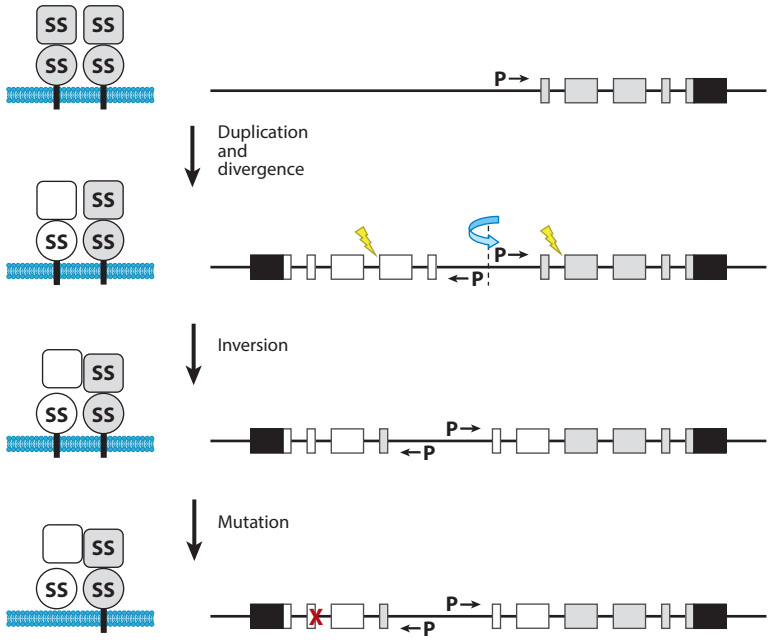

Figure 6

Class II genes can give rise to class I and β_2-microglobulin genes. (*Left*) Representations of molecules, with cell membrane in blue, peptide-binding domains as squares, immunoglobulin domains as circles, and intradomain disulfide bonds indicated as SS. (*Right*) Intron/exon organizations of genes, with rectangles indicating exons, promoters indicated as P, lightning bolts marking breakpoints for inversion, and X indicating mutation. Domains and exons originally from α chain (A gene) are in white and those from β chain (B gene) are in gray, except 3′ untranslated regions (UTRs), in black. Figure modified from References 141 and 180 with permission.

for the adaptive immune system (142). Moreover, certain fish, including the Atlantic cod, have only class I molecules (57), which could be taken to imply that class I genes came first.

These arguments for a primordial class I molecule may be countered by the arguments that a primordial class II molecule might have had a wider tissue distribution and fulfilled the same functions as contemporary class I molecules, and only became specialized after the emergence of class I molecules. Indeed, class II expression can be induced in a wide variety of cell types (143). The lack of class II molecules in Atlantic cod is clearly due to gene loss, based on the presence of gene fragments for proteins that interact with class II molecules, CD4 and invariant chain (57). Moreover, TAPL, which is the closest relative to TAP genes and is found throughout the animal kingdom, forms a homodimer that pumps peptides from the cytoplasm to the lumen of the lysosomal system (144) and thus could have supplied peptides to primordial class II molecules. The presence of the β_2m gene in the nurse shark MHC is consistent with both models, but the fact that the nurse shark β_2m gene is located on a sequencing scaffold next to a RING3/BRD2 gene (61), which is found in the class II region in tetrapods (2, 3), might be taken as evidence for emergence in the class II region. Moreover, some of the major differences in peptide binding between class I and II molecules are not as clear as they once were, including promiscuous peptide binding as well as N- and C-terminal extensions of peptides outside of the groove of class I molecules (90, 91, 145–148). The functions of contemporary class II molecules may be restricted by both chains being tied down at the membrane (with interactions between the transmembrane regions reported

to be involved in conformational changes of the extracellular domains) (149, 150). The greater evolutionary plasticity of class I molecules (10) might thus be explained by the presence of a single transmembrane region in class I molecules allowing greater flexibility in function.

One experimental approach to choose between these two models would be a direct analysis of MHC-like molecules outside of the vertebrates. Despite much examination, no such genes have been identified in the genomes of jawless fish or protochordates (although some genes potentially related to CD4 and TCR have been reported; 151). A sensitive bioinformatics method for interrogating genome sequences, which identified all known class I–like genes in mammals, chickens, and *Xenopus* as well as many class I–like genes in jawed fish, failed to identify candidates in jawless fish, other chordates, and invertebrates (152). A preliminary analysis for class II genes has given similar results (A. Papenfuss, personal communication). Another experimental approach might be to look in jawless fish, other chordates, and invertebrates for cell surface proteins that have bound (a variety of) peptides. The enormously sensitive mass spectrometry approaches (proteomics, immunopeptidomics) currently available provide the tools to examine membrane proteins for the functional properties associated with presentation by MHC molecules. However, it must be borne in mind that perhaps the primordial MHC molecule perished with the animals in the 70 million–year gap between jawless fish and jawed vertebrates, and will never be found.

THE IMPORTANCE OF ADAPTIVE IMMUNITY

Screening for genes upregulated in lymphocytes from jawless fish during an immune response identified families of variable lymphocyte receptors (VLRs) that were reported to be clonally distributed (153, 154). Instead of being based on immunoglobulin-like domains found in the antigen-specific receptors of jawed vertebrates, the extracellular regions of these VLRs are based on leucine-rich repeats (LRRs), like Toll-like receptors (TLRs) and NOD-like receptors (NLRs) of the innate immune system. Moreover, instead of diversification by the RAG genes found in jawed vertebrates, the VLR genes are diversified from nearby pseudogenes by a gene conversion mechanism (conceptually similar to diversification of chicken antibody genes; 155) using cytidine deaminases (CDA1 and CDA2) related to AID and APOBEC (153, 154).

This jawless fish system initially appeared to be analogous, rather than homologous, to the adaptive immune system of jawed fish, but closer examination showed that the two systems were functionally similar and had common cellular ancestors (156). Three kinds of VLRs have been described, each produced by a different genetic locus (157, 158). VLR-B is found on the surface of one population of lymphocytes and also secreted into the bloodstream; VLR-B molecules could bind soluble antigens with extremely high affinity, just like antibodies. Moreover, transcriptomic analysis showed that the VLR-B-bearing lymphocytes express many genes similar to those expected for jawed vertebrate B cells. In contrast, VLR-A and -C are found only on the surface of lymphocytes, which express many genes expected for jawed vertebrate T cells, with VLR-A genes in the blood (like $\alpha\beta$ T cells) and VLR-C in epithelia (like $\gamma\delta$ T cells). Moreover, the tips of the gill arches (now named thymoids) were found to express genes expected for the thymus of jawed vertebrates and to contain VLR-A- and VLR-C-bearing lymphocytes that actively expressed CDA1 (159). The repertoire of VLR-A molecules has been shown to change as a result of residing in the thymoids (160), but what kind of education might be occurring is as yet unknown, particularly since no molecules analogous to MHC molecules have been identified. However, it has been reported that VLR-B molecules from unimmunized hagfish recognize a polymorphic leukocyte alloantigen, NICIR3/ALA, which has an immunoglobulin-like V-C extracellular region (161).

If the different lymphocyte populations that gave rise to T and B cells were already in existence in the lineage that led to both jawless fish and jawed vertebrates, then was the original molecular

system based on LRRs or on immunoglobulin domains? It might be helpful to remember that the NK cell system also has receptors based on two gene superfamilies, immunoglobulin like and lectin like, used to different extents in different vertebrate taxa. Mice and rats use almost exclusively lectin-like NK receptors, humans use predominantly immunoglobulin-like NK receptors, some vertebrates use a mixture of both, and others have few if any active NK receptor genes (162, 163). Moreover, NK receptor–like genes have been described in jawed vertebrates, jawless fish, and protochordates (125, 164). It may be that immune evasion by pathogens can become so intense that a particular gene family loses effectiveness and a switch in molecules used by the host becomes strongly selected. Such a receptor switch might have occurred in the lineages leading to the jawless and jawed vertebrates. Evidence for both VLR and TCR-like genes in the lamprey, and for VLR-like genes in the protochordate *Amphioxus*, has been reported (151, 165).

The independent emergence of different molecular systems for adaptive immunity is not so far-fetched, given that there are other examples of somatically diversified molecular systems proposed as adaptive immune systems in invertebrates (166). The fibrinogen-related proteins (FREPs) found in snails, clams, and other mollusks are cell surface and secreted molecules that are highly diversified by point mutation, with some evidence that these molecules confer resistance to the schistosome parasite in snails (167). The Down syndrome cell adhesion molecule (DSCAM) of arthropods is expressed in the peripheral nerves of insects, highly diversified by alternative splicing and with much evidence for nervous system patterning in juveniles, but it is expressed in hemocytes and the fat body of adults, with some evidence that these molecules confer resistance to the malarial parasite (168). However, there is no consensus that these molecular systems actually function as antigen-specific receptors (169).

Not all invertebrates may have an adaptive immune system, and it may not be necessary. At least some invertebrates have large multigene families of what in vertebrates would be considered molecules of the innate immune system. For instance, echinoderms (like sea urchins, starfish, and sand dollars) have hundreds of expressed TLR, NLR, and scavenger receptor genes, giving rise to the notion that it is the diversity of the receptors that matters, rather than the way in which that diversity arises (170). In this view, diversity can arise from multigene families at one end of a spectrum to somatically diversified single genes at the other end (171), and with mixed strategies (as may have happened in the zebrafish, a teleost fish; 172). Another approach might be to have adaptive immunity that is not anticipatory (that is, not having produced receptors for any conceivable challenge), but dependent on recent infection history. For example, most bacteria (apparently all known archaebacterial and about half of eubacteria) employ clustered regularly interspaced short palindromic repeats (CRISPR), in which the nucleic acids of infectious pathogens such as bacteriophages are incorporated into the genome, expressed as oligonucleotides, and used to cut up the DNA of invading pathogens (173). While the CRISPR system protects single-celled clonal bacteria, the parallels with the RNAi/PIWI systems of multicellular eukaryotes are intriguing (174, 175).

Finally, pathogens must also struggle to survive. One point to consider is that hosts employ a large number of overlapping immune defense mechanisms, and a host will be protected from a particular pathogen even if only one of the many systems of immunity is successful. In contrast, a pathogen must overcome every kind of immunity erected by a host, in general with considerably less genetic material than the host (176). A second point is that pathogens reliant on their host(s) for survival would face a bleak existence if they were so virulent as to deplete the populations of their hosts, so the virulence of such pathogens may be selected to allow hosts to coexist at a reasonable level (177). Thus, the pathogens of vertebrates with adaptive immunity are not better or stronger than the pathogens of invertebrates with only innate immunity, but they may be different, focused on the particular immune barriers erected by their hosts. As a corollary, pathogens that survive

must overcome any new kinds of immunity that might evolve in their hosts, and the evolution of adaptive immunity may merely and temporarily (on the scale of evolutionary time) up the ante for the pathogens.

CONCLUSION

As set out in the introduction, there is much unfinished business to understand the origin and evolution of the MHC and the adaptive immune system of jawed vertebrates. What needs to be done? There has been enormous value in determining the genomes of a variety of animals, but many of these genomes are so fragmentary as to be almost a hindrance rather than a help in resolving the issues about the adaptive immune system. Complete and well-assembled genomes anchored on chromosomes in many animal species will be very helpful. Even more important (and much more difficult) are steps to go beyond the genomes to biochemistry and to function at the levels of cells, organisms, and populations. It is particularly important to examine natural populations of hosts and their natural pathogens in the field, in order to unite our understanding of genomic organization with the biological phenotype that is really under selection.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

With the complex mix of emotions that many years bring, I would like to thank Jack Strominger, in whose lab I first developed some ideas about the evolution of the MHC; Fritz Melchers, in whose institute I spent many years trying to figure out how to test them but mostly spent my time feeding all the animals; Don Wiley, Alan Williams, Alan Wilson, and Morten Simonsen, who were enormously helpful but never got to see where their help led; my many living friends and colleagues I have met along the way, including my special friends Karsten Skjødt and Jan Salomonsen; and especially Max Cooper, who has been an inspiration for decades. I offer apologies to my colleagues for the many papers that I have not cited, either because I missed them or because of the strict word limit. I thank Felicity Coulter, Rebecca Martin, Dan Wise, and Brian Ferguson for critical reading, and I acknowledge a Wellcome Trust Investigator Award (110106/Z/15/Z) and a BBSRC sLoLa grant (BB/K002465/1) for current support.

LITERATURE CITED

1. Klein J. 1986. *The Natural History of the Major Histocompatibility Complex*. New York: Wiley
2. Beck S, Trowsdale J. 2000. The human major histocompatibility complex: lessons from the DNA sequence. *Annu. Rev. Genom. Hum. Genet.* 1:117–37
3. Trowsdale J, Knight JC. 2013. Major histocompatibility complex genomics and human disease. *Annu. Rev. Genom. Hum. Genet.* 14:301–23
4. Blum JS, Wearsch PA, Cresswell P. 2013. Pathways of antigen processing. *Annu. Rev. Immunol.* 31:443–73
5. Bernatchez L, Landry C. 2003. MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years? *J. Evol. Biol.* 16:363–77
6. Spurgin LG, Richardson DS. 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc. Biol. Sci.* 277:979–88

7. Kamiya T, O'Dwyer K, Westerdahl H, Senior A, Nakagawa S. 2014. A quantitative review of MHC-based mating preference: the role of diversity and dissimilarity. *Mol. Ecol.* 23:5151–63
8. Parham P, Moffett A. 2013. Variable NK cell receptors and their MHC class I ligands in immunity, reproduction and human evolution. *Nat. Rev. Immunol.* 13:133–44
9. Petersdorf EW, Shuler KB, Longton GM, Spies T, Hansen JA. 1999. Population study of allelic diversity in the human MHC class I-related MIC-A gene. *Immunogenetics* 49:605–12
10. Adams EJ, Luoma AM. 2013. The adaptable major histocompatibility complex (MHC) fold: structure and function of nonclassical and MHC class I-like molecules. *Annu. Rev. Immunol.* 31:529–61
11. Dijkstra JM, Yamaguchi T, Grimholt U. 2018. Conservation of sequence motifs suggests that the non-classical MHC class I lineages CD1/PRoCR and UT were established before the emergence of tetrapod species. *Immunogenetics*. In press. **http://doi.org/10.1007/s00251-017-1050-2**
12. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, et al. 2004. Gene map of the extended human MHC. *Nat. Rev. Genet.* 5:889–99
13. Amadou C. 1999. Evolution of the *Mhc* class I region: the framework hypothesis. *Immunogenetics* 49:362–67
14. Kelley J, Walter L, Trowsdale J. 2005. Comparative genomics of major histocompatibility complexes. *Immunogenetics* 56:683–95
15. Kumánovics A, Takada T, Lindahl KF. 2003. Genomic organization of the mammalian MHC. *Annu. Rev. Immunol.* 21:629–57
16. Hurt P, Walter L, Sudbrak R, Klages S, Müller I, et al. 2004. The genomic sequence and comparative analysis of the rat major histocompatibility complex. *Genome Res.* 14:631–39
17. Ellis SA, Hammond JA. 2014. The functional significance of cattle major histocompatibility complex class I genetic diversity. *Annu. Rev. Anim. Biosci.* 2:285–306
18. Lunney JK, Ho CS, Wysocki M, Smith DM. 2009. Molecular genetics of the swine major histocompatibility complex, the SLA complex. *Dev. Comp. Immunol.* 33:362–74
19. Dukkipati VS, Blair HT, Garrick DJ, Murray A. 2006. '*Ovar-Mhc*'—ovine major histocompatibility complex: structure and gene polymorphisms. *Genet. Mol. Res.* 5:581–608
20. Belov K, Deakin JE, Papenfuss AT, Baker ML, Melman SD, et al. 2006. Reconstructing an ancestral mammalian immune supercomplex from a marsupial major histocompatibility complex. *PLOS Biol.* 4:e46
21. Cheng Y, Stuart A, Morris K, Taylor R, Siddle H, et al. 2012. Antigen-presenting genes and genomic copy number variations in the Tasmanian devil MHC. *BMC Genom.* 13:87
22. Deakin JE, Siddle HV, Cross JG, Belov K, Graves JA. 2007. Class I genes have split from the MHC in the tammar wallaby. *Cytogenet. Genome Res.* 116:205–11
23. Siddle HV, Deakin JE, Coggill P, Hart E, Cheng Y, et al. 2009. MHC-linked and un-linked class I genes in the wallaby. *BMC Genom.* 10:310
24. Siddle HV, Deakin JE, Coggill P, Whilming LG, Harrow J, et al. 2011. The tammar wallaby major histocompatibility complex shows evidence of past genomic instability. *BMC Genom.* 12:421
25. Balasubramaniam S, Bray RD, Mulder RA, Sunnucks P, Pavlova A, Melville J. 2016. New data from basal Australian songbird lineages show that complex structure of MHC class II β genes has early evolutionary origins within passerines. *BMC Evol. Biol.* 16:112
26. Balakrishnan CN, Ekblom R, Völker M, Westerdahl H, Godinez R, et al. 2010. Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biol.* 8:29
27. Ekblom R, Stapley J, Ball AD, Birkhead T, Burke T, Slate J. 2011. Genetic mapping of the major histocompatibility complex in the zebra finch (*Taeniopygia guttata*). *Immunogenetics* 63:523–30
28. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, et al. 2010. The genome of a songbird. *Nature* 464:757–62
29. Ellegren H, Smeds L, Burri R, Olason PI, Backström N, et al. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* 491:756–60
30. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–31
31. Zhang G, Li C, Li Q, Li B, Larkin DM, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346:1311–20

32. Chen LC, Lan H, Sun L, Deng YL, Tang KY, Wan QH. 2015. Genomic organization of the crested ibis MHC provides new insight into ancestral avian MHC structure. *Sci. Rep.* 5:7963
33. Rogers SL, Kaufman J. 2016. Location, location, location: the evolutionary history of CD1 genes and the NKR-P1/ligand systems. *Immunogenetics* 68:499–513
34. Kaufman J, Milne S, Göbel TW, Walker BA, Jacob JP, et al. 1999. The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401:923–25
35. Shiina T, Briles WE, Goto RM, Hosomichi K, Yanagiya K, et al. 2007. Extended gene map reveals tripartite motif, C-type lectin, and Ig superfamily type genes within a subregion of the chicken MHC-B affecting infectious disease. *J. Immunol.* 178:7162–72
36. Salomonsen J, Chattaway JA, Chan AC, Parker A, Huguet S, et al. 2014. Sequence of a complete chicken BG haplotype shows dynamic expansion and contraction of two gene lineages with particular expression patterns. *PLOS Genet.* 10:e1004417
37. Salomonsen J, Sørensen MR, Marston DA, Rogers SL, Collen T, et al. 2005. Two CD1 genes map to the chicken MHC, indicating that CD1 genes are ancient and likely to have been present in the primordial MHC. *PNAS* 102:8668–73
38. Miller MM, Goto RM, Taylor RL Jr., Zoorob R, Auffray C, et al. 1996. Assignment of Rfp-Y to the chicken major histocompatibility complex/NOR microchromosome and evidence for high-frequency recombination associated with the nucleolar organizer region. *PNAS* 93:3958–62
39. Kaufman J. 2013. The Avian MHC. In: *Avian Immunology*, ed. KA Schat, P Kaiser, B Kaspers, pp. 149–67. London/New York: Elsevier. 2nd ed.
40. Miller MM, Taylor RL Jr. 2016. Brief review of the chicken Major Histocompatibility Complex: the genes, their distribution on chromosome 16, and their contributions to disease resistance. *Poult. Sci.* 95:375–92
41. Shiina T, Shimizu S, Hosomichi K, Kohara S, Watanabe S, et al. 2004. Comparative genomic analysis of two avian (quail and chicken) MHC regions. *J. Immunol.* 172:6751–63
42. Green RE, Braun EL, Armstrong J, Earl D, Nguyen N, et al. 2014. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science* 346:1254449
43. Jaratlerdsiri W, Deakin J, Godinez RM, Shan X, Peterson DG, et al. 2014. Comparative genome analyses reveal distinct structure in the saltwater crocodile MHC. *PLOS ONE* 9:e114631
44. Miller HC, O'Meally D, Ezaz T, Amemiya C, Marshall-Graves JA, Edwards S. 2015. Major histocompatibility complex genes map to two chromosomes in an evolutionarily ancient reptile, the Tuatara *Sphenodon punctatus*. *G3* 5:1439–51
45. Ohta Y, Goetz W, Hossain MZ, Nonaka M, Flajnik MF. 2006. Ancestral organization of the MHC revealed in the amphibian *Xenopus*. *J. Immunol.* 176(6):3674–85
46. Flajnik MF, Kasahara M, Shum BP, Salter-Cid L, Taylor E, Du Pasquier L. 1993. A novel type of class I gene organization in vertebrates: A large family of non-MHC-linked class I genes is expressed at the RNA level in the amphibian *Xenopus*. *EMBO J.* 12:4385–96
47. Courtet M, Flajnik M, Du Pasquier L. 2001. Major histocompatibility complex and immunoglobulin loci visualized by in situ hybridization on *Xenopus* chromosomes. *Dev. Comp. Immunol.* 25:149–57
48. Edholm ES, Banach M, Robert J. 2016. Evolution of innate-like T cells and their selection by MHC class I-like molecules. *Immunogenetics* 68:525–36
49. Sammut B, Du Pasquier L, Ducoroy P, Laurens V, Marcuz A, Tournefier A. 1999. Axolotl MHC architecture and polymorphism. *Eur. J. Immunol.* 29:2897–907
50. Laurens V, Chapusot C, del Rosario Ordonez M, Bentrari F, Padros MR, Tournefier A. 2001. Axolotl MHC class II β chain: predominance of one allele and alternative splicing of the β1 domain. *Eur. J. Immunol.* 31:506–15
51. Takami K, Zaleska-Rutczynska Z, Figueroa F, Klein J. 1997. Linkage of LMP, TAP, and RING3 with Mhc class I rather than class II genes in the zebrafish. *J. Immunol.* 159:6052–60
52. Sato A, Figueroa F, Murray BW, Málaga-Trillo E, Zaleska-Rutczynska Z, et al. 2000. Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. *Immunogenetics* 51:108–16
53. Grimholt U, Tsukamoto K, Azuma T, Leong J, Koop BF, Dijkstra JM. 2015. A comprehensive analysis of teleost MHC class I sequences. *BMC Evol. Biol.* 15:32

54. Sambrook JG, Figueroa F, Beck S. 2005. A genome-wide survey of Major Histocompatibility Complex (MHC) genes and their paralogues in zebrafish. *BMC Genom*. 6:152
55. Dijkstra JM, Grimholt U, Leong J, Koop BF, Hashimoto K. 2013. Comprehensive analysis of MHC class II genes in teleost fish genomes reveals dispensability of the peptide-loading DM system in a large part of vertebrates. *BMC Evol. Biol.* 13:260
56. Deakin JE, Papenfuss AT, Belov K, Cross JG, Coggill P, et al. 2006. Evolution and comparative analysis of the MHC class III inflammatory region. *BMC Genom*. 7:281
57. Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrøm M, et al. 2011. The genome sequence of Atlantic cod reveals a unique immune system. *Nature* 477:207–10
58. Ohta Y, Okamura K, McKinney EC, Bartl S, Hashimoto K, Flajnik MF. 2000. Primitive synteny of vertebrate major histocompatibility complex class I and class II genes. *PNAS* 97:4712–17
59. Ohta Y, McKinney EC, Criscitiello MF, Flajnik MF. 2002. Proteasome, transporter associated with antigen processing, and class I genes in the nurse shark *Ginglymostoma cirratum*: evidence for a stable class I region and MHC haplotype lineages. *J. Immunol.* 168:771–81
60. Terado T, Okamura K, Ohta Y, Shin DH, Smith SL, et al. 2003. Molecular cloning of C4 gene and identification of the class III complement region in the shark MHC. *J. Immunol.* 171:2461–66
61. Ohta Y, Shiina T, Lohr RL, Hosomichi K, Pollin TI, et al. 2011. Primordial linkage of β2-microglobulin to the MHC. *J. Immunol.* 186:3563–71
62. Younger RM, Amadou C, Bethel G, Ehlers A, Lindahl KF, et al. 2001. Characterization of clustered MHC-linked olfactory receptor genes in human and mouse. *Genome Res*. 11:519–30
63. Kaufman J. 1999. Co-evolving genes in MHC haplotypes: the "rule" for nonmammalian vertebrates? *Immunogenetics* 50:228–36
64. Kaufman J. 2015. Co-evolution with chicken class I genes. *Immunol. Rev.* 267:56–71
65. Katsanis N, Fitzgibbon J, Fisher EM. 1996. 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* 35:101–8
66. Kasahara M, Hayashi M, Tanaka K, Inoko H, Sugaya K, et al. 1996. Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. *PNAS* 93:9096–101
67. Kasahara M, Nakaya J, Satta Y, Takahata N. 1997. Chromosomal duplication and the emergence of the adaptive immune system. *Trends Genet*.13:90–92
68. Ohno S. 1970. *Evolution by Gene Duplication*. New York: Springer-Verlag
69. Schluter SF, Bernstein RM, Bernstein H, Marchalonis JJ. 1999. 'Big Bang' emergence of the combinatorial immune system. *Dev. Comp. Immunol.* 23:107–11
70. Flajnik MF. 2014. Re-evaluation of the immunological Big Bang. *Curr. Biol.* 24:R1060–65
71. Abi-Rached L, Gilles A, Shiina T, Pontarotti P, Inoko H. 2002. Evidence of en bloc duplication in vertebrate genomes. *Nat. Genet.* 31:100–5
72. Suurväli J, Jouneau L, Thépot D, Grusea S, Pontarotti P, et al. 2014. The proto-MHC of placozoans, a region specialized in cellular stress and ubiquitination/proteasome pathways. *J. Immunol.* 193:2891–901
73. Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, et al. 2013. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat. Genet.* 45:415–21
74. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, et al. 2016. The Atlantic salmon genome provides insights into rediploidization. *Nature* 533:200–5
75. Du Pasquier L, Miggiano VC, Kobel H-R, Fischberg M. 1977. The genetic control of histocompatibility reactions in natural and laboratory-made polyploid individuals of the clawed toad, *Xenopus. Immunogenetics* 5:129–41
76. Du Pasquier L, Wilson M, Sammut B. 2009. The fate of duplicated immunity genes in the dodecaploid *Xenopus ruwenzoriensis*. *Front. Biosci*. 14:177–91
77. Donoghue PC, Purnell MA. 2005. Genome duplication, extinction and vertebrate evolution. *Trends Ecol. Evol.* 20:312–19
78. Kandil E, Egashira M, Miyoshi O, Niikawa N, Ishibashi T, Kasahara M. 1996. The human gene encoding the heavy chain of the major histocompatibility complex class I-like Fc receptor (FCGRT) maps to 19q13.3. *Cytogenet. Cell Genet.* 73:97–98

79. Kasahara M, Kandil E, Salter-Cid L, Flajnik MF. 1996. Origin and evolution of the class I gene family: Why are some of the mammalian class I genes encoded outside the major histocompatibility complex? *Res. Immunol.* 147:278–84
80. Tsukamoto K, Deakin JE, Graves JA, Hashimoto K. 2013. Exceptionally high conservation of the MHC class I-related gene, *MR1*, among mammals. *Immunogenetics* 65:115–24
81. Yang Z, Wang C, Wang T, Bai J, Zhao Y, et al. 2015. Analysis of the reptile CD1 genes: evolutionary implications. *Immunogenetics* 67:337–46
82. Hughes AL. 1991. Evolutionary origin and diversification of the mammalian CD1 antigen genes. *Mol. Biol. Evol.* 8:185–201
83. Dascher CC. 2007. Evolutionary biology of CD1. *Curr. Top. Microbiol. Immunol.* 314:3–26
84. Germain RN, Bentley DM, Quill H. 1985. Influence of allelic polymorphism on the assembly and surface expression of class II MHC (Ia) molecules. *Cell* 43:233–42
85. Deverson EV, Leong L, Seelig A, Coadwell WJ, Tredgett EM, et al. 1998. Functional analysis by site-directed mutagenesis of the complex polymorphism in rat transporter associated with antigen processing. *J. Immunol.* 160:2767–79
86. Joly E, Le Rolle AF, González AL, Mehling B, Stevens J, et al. 1998. Co-evolution of rat TAP transporters and MHC class I RT1-A molecules. *Curr. Biol.* 8:169–72
87. Moffett A, Colucci F. 2015. Co-evolution of NK receptors and HLA ligands in humans is driven by reproduction. *Immunol. Rev.* 267:283–97
88. Kaufman J. 2015. What chickens would tell you about the evolution of antigen processing and presentation. *Curr. Opin. Immunol.* 34:35–42
89. Wallny HJ, Avila D, Hunt LG, Powell TJ, Riegert P, et al. 2006. Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. *PNAS* 103:1434–39
90. Koch M, Camp S, Collen T, Avila D, Salomonsen J, et al. 2007. Structures of an MHC class I molecule from B21 chickens illustrate promiscuous peptide binding. *Immunity* 27:885–99
91. Chappell P, Meziane ELK, Harrison M, Magiera Ł, Hermann C, et al. 2015. Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. *eLife* 4:e05345
92. Walker BA, Hunt LG, Sowa AK, Skjødt K, Göbel TW, et al. 2011. The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. *PNAS* 108:8396–401
93. Tregaskes CA, Harrison M, Sowa AK, van Hateren A, Hunt LG, et al. 2016. Surface expression, peptide repertoire, and thermostability of chicken class I molecules correlate with peptide transporter specificity. *PNAS* 113:692–97
94. van Hateren A, Carter R, Bailey A, Kontouli N, Williams AP, et al. 2013. A mechanistic basis for the co-evolution of chicken tapasin and major histocompatibility complex class I (MHC I) proteins. *J. Biol. Chem.* 288:32797–808
95. McLaren PJ, Coulonges C, Bartha I, Lenz TL, Deutsch AJ, et al. 2015. Polymorphisms of large effect explain the majority of the host genetic contribution to variation of HIV-1 virus load. *PNAS* 112:14658–63
96. Mesa CM, Thulien KJ, Moon DA, Veniamin SM, Magor KE. 2004. The dominant MHC class I gene is adjacent to the polymorphic *TAP2* gene in the duck, *Anas platyrhynchos*. *Immunogenetics* 56:192–203
97. Tsuji H, Taniguchi Y, Ishizuka S, Matsuda H, Yamada T, et al. 2017. Structure and polymorphisms of the major histocompatibility complex in the Oriental stork, *Ciconia boyciana*. *Sci. Rep.* 7:42864
98. Ohta Y, Powis SJ, Lohr RL, Nonaka M, Du Pasquier L, Flajnik MF. 2003. Two highly divergent ancient allelic lineages of the transporter associated with antigen processing (*TAP*) gene in *Xenopus*: further evidence for co-evolution among MHC class I region genes. *Eur. J. Immunol.* 33:3017–27
99. Tsukamoto K, Miura F, Fujito NT, Yoshizaki G, Nonaka M. 2012. Long-lived dichotomous lineages of the proteasome subunit beta type 8 (*PSMB8*) gene surviving more than 500 million years as alleles or paralogs. *Mol. Biol. Evol.* 29:3071–79
100. McConnell SC, Hernandez KM, Wcisel DJ, Kettleborough RN, Stemple DL, et al. 2016. Alternative haplotypes of antigen processing genes in zebrafish diverged early in vertebrate evolution. *PNAS* 113:E5014–23

101. Shiina T, Hosomichi K, Hanzawa K. 2006. Comparative genomics of the poultry major histocompatibility complex. *Anim. Sci. J.* 77:151–62
102. Drews A, Strandh M, Råberg L, Westerdahl H. 2017. Expression and phylogenetic analyses reveal paralogous lineages of putatively classical and non-classical MHC-I genes in three sparrow species (*Passer*). *BMC Evol. Biol.* 17:152
103. Grimholt U, Larsen S, Nordmo R, Midtlyng P, Kjoeglum S, et al. 2003. MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*): facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogenetics* 55:210–19
104. Moon DA, Veniamin SM, Parks-Dely JA, Magor KE. 2005. The MHC of the duck (*Anas platyrhynchos*) contains five differentially expressed class I genes. *J. Immunol.* 175:6702–12
105. Okamura K, Ototake M, Nakanishi T, Kurosawa Y, Hashimoto K. 1997. The most primitive vertebrates with jaws possess highly polymorphic MHC class I genes comparable to those of humans. *Immunity* 7:777–90
106. Bonneaud C, Pérez-Tris J, Federici P, Chastel O, Sorci G. 2006. Major histocompatibility alleles associated with local resistance to malaria in a passerine. *Evolution* 60:383–89
107. Savage AE, Zamudio KR. 2011. MHC genotypes associate with resistance to a frog-killing fungus. *PNAS* 108:16705–10
108. Kaufman J. 2011. The evolutionary origins of the adaptive immune system of jawed vertebrates. In *The Immune Response to Infection*, ed. SHE Kaufmann, BT Rouse, DL Sachs, pp. 41–55. Washington, DC: Am. Soc. Microbiol.
109. Rogers SL, Göbel TW, Viertlboeck BC, Milne S, Beck S, Kaufman J. 2005. Characterization of the chicken C-type lectin-like receptors B-NK and B-lec suggests that the NK complex and the MHC share a common ancestral region. *J. Immunol.* 174:3475–83
110. Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, et al. 2014. Elephant shark genome provides unique insights into gnathostome evolution. *Nature* 505:174–79
111. Kasahara M, Watanabe Y, Sumasu M, Nagata T. 2002. A family of MHC class I-like genes located in the vicinity of the mouse leukocyte receptor complex. *PNAS* 99:13687–92
112. Du Pasquier L, Zucchetti I, De Santis R. 2004. Immunoglobulin superfamily receptors in protochordates: before RAG time. *Immunol. Rev.* 198:233–48
113. Flajnik MF, Kasahara M. 2010. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat. Rev. Genet.* 11:47–59
114. Flajnik MF, Tlapakova T, Criscitiello MF, Krylov V, Ohta Y. 2012. Evolution of the B7 family: coevolution of B7H6 and NKp30, identification of a new B7 family member, B7H7, and of B7's historical relationship with the MHC. *Immunogenetics* 64:571–90
115. Boyle LH, Hermann C, Boname JM, Porter KM, Patel PA, et al. 2013. Tapasin-related protein TAPBPR is an additional component of the MHC class I presentation pathway. *PNAS* 110:3465–70
116. Olinski RP, Lundin LG, Hallböök F. 2006. Conserved synteny between the *Ciona* genome and human paralogons identifies large duplication events in the molecular evolution of the insulin-relaxin gene family. *Mol. Biol. Evol.* 23:10–22
117. Isobe M, Russo G, Haluska FG, Croce CM. 1988. Cloning of the gene encoding the δ subunit of the human T-cell receptor reveals its physical organization within the α-subunit locus and its involvement in chromosome translocations in T-cell malignancy. *PNAS* 85:3933–37
118. Krangel MS, Carabana J, Abbarategui I, Schlimgen R, Hawwari A. 2004. Enforcing order within a complex locus: current perspectives on the control of V(D)J recombination at the murine T-cell receptor α/δ locus. *Immunol. Rev.* 200:224–32
119. Rossjohn J, Gras S, Miles JJ, Turner SJ, Godfrey DI, McCluskey J. 2015. T cell antigen receptor recognition of antigen-presenting molecules. *Annu. Rev. Immunol.* 33:169–200
120. Fugmann SD, Lee AI, Shockett PE, Villey IJ, Schatz DG. 2000. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu. Rev. Immunol.* 18:495–527
121. Koonin EV, Krupovic M. 2015. Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat. Rev. Genet.* 16:184–92
122. Fugmann SD, Messier C, Novack LA, Cameron RA, Rast JP. 2006. An ancient evolutionary origin of the *Rag1/2* gene locus. *PNAS* 103:3728–33

123. Huang S, Tao X, Yuan S, Zhang Y, Li P, et al. 2016. Discovery of an active RAG transposon illuminates the origins of V(D)J recombination. *Cell* 166:102–14
124. Davis MM, Bjorkman PJ. 1988. T-cell antigen receptor genes and T-cell recognition. *Nature* 334:395–402
125. Litman GW, Rast JP, Fugmann SD. 2010. The origins of vertebrate adaptive immunity. *Nat. Rev. Immunol.* 10:543–53
126. Abeler-Dörner L, Swamy M, Williams G, Hayday AC, Bas A. 2012. Butyrophilins: an emerging family of immune regulators. *Trends Immunol.* 33:34–41
127. Di Marco Barros R, Roberts NA, Dart RJ, Vantourout P, Jandke A, et al. 2016. Epithelia use butyrophilin-like molecules to shape organ-specific γδ T cell compartments. *Cell* 167:203–218
128. Afrache H, Gouret P, Ainouche S, Pontarotti P, Olive D. 2012. The butyrophilin (BTN) gene family: from milk fat to the regulation of the immune response. *Immunogenetics* 64:781–94
129. Rhodes DA, Reith W, Trowsdale J. 2016. Regulation of immunity by butyrophilins. *Annu. Rev. Immunol.* 34:151–72
130. Du Pasquier L. 2000. The phylogenetic origin of antigen-specific receptors. *Curr. Top. Microbiol. Immunol.* 248:160–85
131. Lanier LL. 2005. NK cell recognition. *Annu. Rev. Immunol.* 23:225–74
132. Holcik M, Sonenberg N. 2005. Translational control in stress and apoptosis. *Nat. Rev. Mol. Cell Biol.* 6:318–27
133. Starck SR, Jiang V, Pavon-Eternod M, Prasad S, McCarthy B, et al. 2012. Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science* 336:1719–23
134. Jensen PE, Sullivan BA, Reed-Loisel LM, Weber DA. 2004. Qa-1, a nonclassical class I histocompatibility molecule with roles in innate and adaptive immunity. *Immunol. Res.* 29:81–92
135. Fischer-Lindahl K, Hermel E, Loveland BE, Wang CR. 1991. Maternally transmitted antigen of mice: a model transplantation antigen. *Annu. Rev. Immunol.* 9:351–72
136. Yaneva R, Schneeweiss C, Zacharias M, Springer S. 2010. Peptide binding to MHC class I and II proteins: new avenues from new methods. *Mol. Immunol.* 47:649–57
137. Flajnik MF, Canel C, Kramer J, Kasahara M. 1991. Which came first, MHC class I or class II? *Immunogenetics* 33:295–300
138. Rock KL, Reits E, Neefjes J. 2016. Present yourself! By MHC class I and MHC class II molecules. *Trends Immunol.* 37:724–737
139. Zhang P, Leu JI, Murphy ME, George DL, Marmorstein R. 2014. Crystal structure of the stress-inducible human heat shock protein 70 substrate-binding domain in complex with peptide substrate. *PLOS ONE* 9:e103518
140. Kaufman JF, Auffray C, Korman AJ, Shackelford DA, Strominger J. 1984. The class II molecules of the human and murine major histocompatibility complex. *Cell* 36:1–13
141. Kaufman J. 1988. Vertebrates and the evolution of the major histocompatibility complex class I and class II molecules. *Verh. Dtsch. Zool. Ges.* 81:131–44
142. Benoist C, Mathis D. 1990. Regulation of major histocompatibility complex class-II genes: X, Y and other letters of the alphabet. *Annu. Rev. Immunol.* 8:681–715
143. Boehm U, Klamp T, Groot M, Howard JC. 1997. Cellular responses to interferon-γ. *Annu. Rev. Immunol.* 15:749–95
144. Zollmann T, Bock C, Graab P, Abele R. 2015. Team work at its best—TAPL and its two domains. *Biol. Chem.* 396:967–74
145. Motozono C, Pearson JA, De Leenheer E, Rizkallah PJ, Beck K, et al. 2015. Distortion of the major histocompatibility complex class I binding groove to accommodate an insulin-derived 10-mer peptide. *J. Biol. Chem.* 290:18924–33
146. McMurtrey C, Trolle T, Sansom T, Remesh SG, Kaever T, et al. 2016. *Toxoplasma gondii* peptide ligands open the gate of the HLA class I binding groove. *eLife* 5:e12556
147. Remesh SG, Andreatta M, Ying G, Kaever T, Nielsen M, et al. 2017. Unconventional peptide presentation by major histocompatibility complex (MHC) class I allele HLA-A*02:01: BREAKING CONFINEMENT. *J. Biol. Chem.* 292:5262–70

148. Li X, Lamothe PA, Walker BD, Wang JH. 2017. Crystal structure of HLA-B*5801 with a TW10 HIV Gag epitope reveals a novel mode of peptide presentation. *Cell. Mol. Immunol.* 14:1–4
149. Kaufman JF, Strominger JL. 1979. Both chains of HLA-DR bind to the membrane with a penultimate hydrophobic region and the heavy chain is phosphorylated at its hydrophilic carboxy terminus. *PNAS* 76:6304–8
150. Dixon AM, Drake L, Hughes KT, Sargent E, Hunt D, et al. 2014. Differential transmembrane domain GXXXG motif pairing impacts major histocompatibility complex (MHC) class II structure. *J. Biol. Chem.* 289:11695–703
151. Pancer Z, Mayer WE, Klein J, Cooper MD. 2004. Prototypic T cell receptor and CD4-like coreceptor are expressed by lymphocytes in the agnathan sea lamprey. *PNAS* 101:13273–78
152. Krasnec KV, Papenfuss AT, Miller RD. 2016. The UT family of MHC class I loci unique to non-eutherian mammals has limited polymorphism and tissue specific patterns of expression in the opossum. *BMC Immunol.* 17:43
153. Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL, Cooper MD. 2004. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* 430:174–80
154. Alder MN, Rogozin IB, Iyer LM, Glazko GV, Cooper MD, Pancer Z. 2005. Diversity and function of adaptive immune receptors in a jawless vertebrate. *Science* 310:1970–73
155. Weill JC, Reynaud CA. 1987. The chicken B cell compartment. *Science* 238:1094–98
156. Boehm T, McCurley N, Sutoh Y, Schorpp M, Kasahara M, Cooper MD. 2012. VLR-based adaptive immunity. *Annu. Rev. Immunol.* 30:203–20
157. Hirano M, Guo P, McCurley N, Schorpp M, Das S, et al. 2013 Evolutionary implications of a third lymphocyte lineage in lampreys. *Nature* 501:435–38
158. Li J, Das S, Herrin BR, Hirano M, Cooper MD. 2013. Definition of a third VLR gene in hagfish. *PNAS* 110:15013–18
159. Bajoghli B, Guo P, Aghaallaei N, Hirano M, Strohmeier C, et al. 2011. A thymus candidate in lampreys. *Nature* 470:90–94
160. Holland SJ, Gao M, Hirano M, Iyer LM, Luo M, et al. 2014. Selection of the lamprey VLRC antigen receptor repertoire. *PNAS* 111:14834–39
161. Takaba H, Imai T, Miki S, Morishita Y, Miyashita A, et al. 2013. A major allogenic leukocyte antigen in the agnathan hagfish. *Sci. Rep.* 3:1716
162. Yoder JA, Litman GW. 2011. The phylogenetic origins of natural killer receptors and recognition: relationships, possibilities, and realities. *Immunogenetics* 63:123–41
163. Hammond JA, Guethlein LA, Abi-Rached L, Moesta AK, Parham P. 2009. Evolution and survival of marine carnivores did not require a diversity of killer cell Ig-like receptors or Ly49 NK cell receptors. *J. Immunol.* 182:3618–27
164. van den Berg TK, Yoder JA, Litman GW. 2004. On the origins of adaptive immunity: Innate immune receptors join the tale. *Trends Immunol.* 25:11–16
165. Cao DD, Liao X, Cheng W, Jiang YL, Wang WJ, et al. 2016. Structure of a variable lymphocyte receptor-like protein from the amphioxus *Branchiostoma floridae*. *Sci. Rep.* 6:19951
166. Cerenius L, Söderhäll K. 2013. Variable immune molecules in invertebrates. *J. Exp. Biol.* 216:4313–19
167. Zhang SM, Adema CM, Kepler TB, Loker ES. 2004. Diversification of Ig superfamily genes in an invertebrate. *Science* 305:251–54
168. Watson FL, Püttmann-Holgado R, Thomas F, Lamar DL, Hughes M, et al. 2005. Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science* 309:1874–78
169. Armitage SA, Peuss R, Kurtz J. 2015. Dscam and pancrustacean immune memory—a review of the evidence. *Dev. Comp. Immunol.* 48:315–23
170. Buckley KM, Rast JP. 2015. Diversity of animal immune receptors and the origins of recognition complexity in the deuterostomes. *Dev. Comp. Immunol.* 49:179–89
171. Flajnik MF, Du Pasquier L. 2004. Evolution of innate and adaptive immunity: Can we draw a line? *Trends Immunol.* 25:640–44
172. Howe K, Schiffer PH, Zielinski J, Wiehe T, Laird GK, et al. 2016. Structure and evolutionary history of a large family of NLR proteins in the zebrafish. *Open Biol.* 6:160009

173. Koonin EV, Makarova KS, Wolf YI. 2017. Evolutionary genomics of defense systems in archaea and bacteria. *Annu. Rev. Microbiol.* 71:233–61
174. Marques JT, Carthew RW. 2007. A call to arms: coevolution of animal viruses and host innate immune responses. *Trends Genet*. 23:359–64
175. Obbard DJ, Gordon KH, Buck AH, Jiggins FM. 2009. The evolution of RNAi as a defence against viruses and transposable elements. *Philos. Trans. R. Soc. Lond. B* 364:99–115
176. Gilman RT, Nuismer SL, Jhwueng DC. 2012. Coevolution in multidimensional trait space favours escape from parasites and pathogens. *Nature* 483:328–30
177. Hedrick SM. 2004. The acquired immune system: a vantage from beneath. *Immunity* 21:607–15
178. Kaufman J, Völk H, Wallny HJ. 1995. A "minimal essential Mhc" and an "unrecognized Mhc": two extremes in selection for polymorphism. *Immunol. Rev.* 143:63–88
179. Kaufman J, Jacob J, Shaw I, Walker B, Milne S, et al. 1999. Gene organisation determines evolution of function in the chicken MHC. *Immunol. Rev.* 167:101–17
180. Kaufman J, Skjoedt K, Salomonsen J. 1990. The MHC molecules of nonmammalian vertebrates. *Immunol. Rev*. 113:83–117