The Evolution of Major Histocompatibility Complex in Birds

The Evolution of the Major Histocompatibility Complex in Birds

CHRISTOPHER M. HESS AND SCOTT V. EDWARDS

The genetic region that scientists today call the major histocompatibility complex (MHC) was discovered in the 1930s by Peter Gorer in his pioneering studies of antigenic responses to transplanted sera by inbred mouse strains (Gorer 1936). The MHC was genetically defined more precisely by George Snell (1948), who first used the term. Contrary to what many believe, the discovery of cellular antigens in chickens *(Gallus gallus)*, which have functions similar to those in mice, occurred before comparable discoveries in humans (Briles and McGibbon 1948). Thus, studies of the MHC in birds have more than a 50-year history. The immunological and molecular biological revolutions of the 1970s and 1980s, culminating most recently in the complete sequencing of (a) the human leukocyte antigen (HLA; Beck et al. 1999) and (b) the B complex in chickens (Kaufman et al. 1999), revealed that the vertebrate MHC is a complex, multigene family comprising loci encoding receptors on the surfaces of a variety of immune and nonimmune cells. These receptors bind amino acid fragments (or peptides) from foreign pathogens, upon which a cascade of immunological events known as the adaptive immune response is initiated (figure 1). The fundamental position of MHC molecules for initiation and maintenance of both the T cell–mediated and humoral (antibody) arms of the adaptive immune responses has led some immunologists to conclude that the MHC is "the center of the immune universe" (Trowsdale 1995).

Those aspects of MHC genes that intrigue immunologists—their function in disease resistance, their unusually high polymorphism and tight linkage into a single "supergene complex"—also intrigue evolutionary biologists and may provide keys to understanding adaptive polymorphism in general and the genetic basis of pathogen resistance in particular. The hypothesis that genetic diversity of MHC genes underlies resistance to the diversity of infectious pathogens arose out of Zuckerkandl and Pauling's pioneering experiments in mice and population genetics theory, both of which revealed the immunological advantages of heterozygosity, sexual reproduction, and outbreeding (Clarke and O'Donald 1964, Zuckerkandl and Pauling 1965). However, despite the elegant theory that parasitism should drive genetic diversity, empirical examples of the advantages of MHC heterozygosity for rapid parasite clearance have been few, with only a handful of recent but convincing case studies in mammals, for example, HIV (Carrington et al. 1999) and hepatitis (Thursz et al. 1997). Other examples in humans implicate specific MHC haplotypes (multilocus alleles) in resistance to infectious disease, but not heterozygosity per se (e.g., malaria; Hill 1991).

Happily for ornithologists, the most striking associations between specific MHC haplotypes and disease resistance are known from chickens—specifically, resistance to a virus causing Marek's disease. In this case, having the beneficial B-21 haplotype makes an individual 95% resistant, whereas individuals with the B-19 haplotype suffer 100% mortality (Cole 1968). It has been suggested that the tighter linkage of MHC

SCALING UP AND TAKING A GENOMIC APPROACH TO THE MAJOR HISTOCOMPATIBILTY COMPLEX (MHC) OF BIRDS REVEALS SURPRISING DEPARTURES FROM GENERALITIES FOUND IN MAMMALS IN BOTH LARGE-SCALE STRUCTURE AND THE MECHANISMS SHAPING THE EVOLUTION OF THE MHC

Christopher Hess (e-mail: cmhess@u.washington.edu) is a doctoral candidate in the Department of Zoology at the University of Washington, Box 351800, Seattle, Washington, 98195. Scott Edwards (e-mail: sedwards@u.washington.edu) is an associate professor in the Department of Zoology and the curator of genetic resources at the Burke Museum of Natural History and Culture, University of Washington.

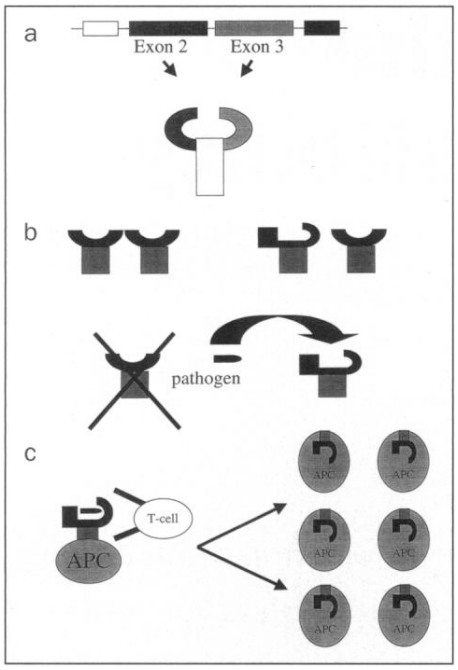


Figure 1. MHC molecules and genes. (a) In the case of class I MHC genes, the second and third exons encode the highly polymorphic peptide-binding region. (b) One of the predominant theories for the maintenance of variation at the MHC is a heterozygote advantage. The homozygote in this case can only bind a limited repertoire of pathogens whereas the heterozygote can bind a larger repertoire. (c) The adaptive immune response is initiated when a peptide bound to a MHC receptor on the antigen-presenting cell (APC) surface interacts with a cytotoxic T cell. After binding, numerous memory cells with the same T-cell receptors are produced (located on the inside of the cell surface for class I MHC genes—class I receptors are rarely found on the outside of a cell before antigen binding). Upon a new infection, the immune response will be stronger and faster than the original response.

haplotype variation and disease resistance in chickens, compared with mammals, may ultimately derive from the compact architecture of the chicken MHC, its reduced redundancy, and its lack of superfluous genes unrelated to immune function (Kaufman 1999). Do birds other than chickens possess a similarly compact MHC? Are MHC–disease associations in other birds as strong as in chickens? The tight relationship between MHC haplotype and disease in chickens, in conjunction with the rise of the growing field of ecological immunology—the study of the fitness effects of parasitism and the costs of adaptive immune responses in the wild (Moller et al. 1999)—has driven the characterization of MHC structure and allelic variation in a diversity of bird species that are models not of immunobiology but of ecology and evolutionary biology.

Understanding the molecular evolution of the MHC in birds will be essential to informed integration of MHC variation into ecological studies in the wild. The advent of techniques for obtaining large contiguous stretches of DNA sequence makes it increasingly feasible to investigate MHC structure in nonmodel species and, in concert with immunological studies, gives the potential to understand the molecular mechanisms shaping the avian MHC. With its staggering diversity at several scales, the MHC offers an extreme view of processes occurring at the genomic level. It is a dynamic, multigene family with high levels of duplication, pseudogene formation, gene conversion, point mutation, and natural selection (Klein et al. 1993). Understanding these processes in detail will not only facilitate inferences about genome-wide processes but will also forge a link between immune system–gene variation and the evolution and ecology of the organisms themselves.

Defining MHC genes

MHC genes can be defined evolutionarily by clustering with respect to other known MHC genes (genes that have been shown to be involved in antigen presentation) in a phylogenetic context. If the gene falls within a cluster of known MHC genes, it shares a common ancestor with genes that were involved in antigen presentation as well as other aspects of the adaptive immune response. Evolutionary relationships can be used to determine whether a gene should be designated an MHC gene, meaning MHC genes that are not highly transcribed or have become pseudogenes can still be defined as MHC genes, albeit genes without function. Alternatively, a strict definition may be used, one that requires that genes that participate in graft rejection be reclassified as MHC genes, but such demonstration will be beyond the purview of most ecological immunologists, who mainly study nonmodel species. Therefore, we prefer the use of evolutionary relationships in defining MHC genes.

The term "classical" is often used to indicate that an MHC gene is polymorphic and highly expressed. "Nonclassical" genes are MHC genes that cluster with other MHC genes phylogenetically, but they are usually not expressed at high levels and are not polymorphic. Another important designation of MHC genes is the difference between class I, class II, and class III genes. Class I genes encode receptors that are presented on the surfaces of nearly all nucleated cells; they are mainly responsible for facilitating immune responses to intracellular pathogens such as viruses. They interact with cytotoxic CD8+ T cells after antigens are bound. Class II genes are found only on a subset of cells, such as B cells and macrophages, and are primarily involved in the immune response against extracellular pathogens such as bacteria. Class II receptors bound with foreign antigen bind with CD4 helper T cells. Many other types of genes are found within the region defined as the MHC. These include class III genes that make up the complement component of an immune response rather than the adaptive immune response, as well as natural killer receptor (NKR) genes and Tapasin, a gene involved in antigen presentation and peptide binding. As in humans, another group of genes, TAP1 and TAP2, are also found in the

MHC of chickens and facilitate loading of peptides onto class I MHC molecules (Kaufman 1999).

Features of mammalian MHC evolution

Studies of the MHCs of mammals are the most rigorous of any vertebrate group. The functional axes on which MHC genes are categorized were first defined in mammals and are where the major features of MHC evolution were first revealed. Below we list briefly some of these key features that provide a necessary starting point for avian MHC studies.

Large-scale genomic structure. The MHC was an early focus of the human genome project. The human MHC, or HLA, is located in a guanine-cytosine (GC)-rich region and spans approximately 4 megabases (MB), about 1/100th of the haploid genome and about the size of an *Escherichia coli* genome (Beck et al. 1999). The MHC is the most gene-rich region sequenced in the human genome, with over 200 loci identified. All HLA genes are found in a single cluster on chromosome 6 (save for the MR1 class 1–like genes, found on chromosome 1), and it is estimated that 40% of the expressed genes in the region have immune system function (Beck et al. 1999). The centromeric class II region and the telomeric class I genes are separated by the class III region. The MHC class I and class II regions of humans, mice, and other mammals all have higher incidences of pseudogenes that are intermingled with functional genes than is typical for the human genome (Beck et al. 1999).

Origin and maintenance of genetic variation. MHC genes are the most diverse in the human genome. Some loci possess over 200 alleles in the human population (Parham and Ohta 1996; Hickson and Cann 1997). These levels of polymorphism are much higher than the variation found at neutral coding and noncoding loci in the human genome. Answers to how this tremendous variation is generated and maintained are still controversial. Are the levels of diversity solely the result of balancing selection (heterozygote advantage or frequency-dependent selection) or does some form of hypermutation elevate diversity beyond that maintained by selection? Phylogenetic approaches to this question invariably show that some MHC allelic lineages are tens of millions of years old, often predating divergences between species currently carrying those alleles. Klein and colleagues (1993) estimated that it takes one to six million years for a new mutation to be fixed and that the mutation rate at MHC loci is in fact at the lower end for genes in primates. They concluded that selection was the driving force behind MHC diversity, not high levels of mutation. However, Bergstrom and colleagues (1997) concluded on the basis of comparisons of MHC coding and noncoding data that, in addition to balancing selection, hypermutation produced by gene conversion (box 1a) was prevalent at MHC loci and could explain the observed levels of variation.

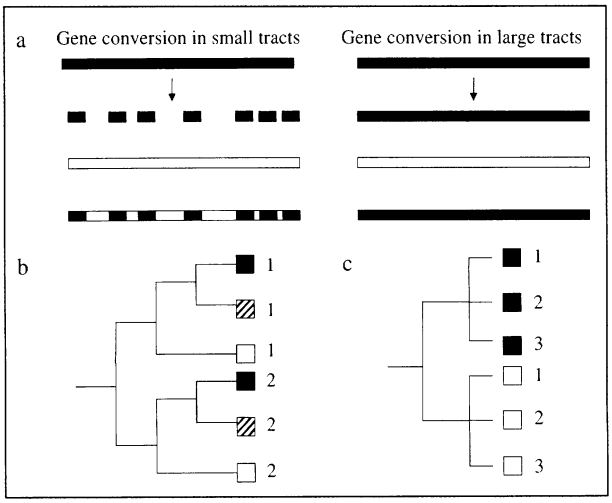
Multigene family evolution. An understanding of the mechanisms for multigene family evolution is a crucial prerequisite to explaining the diversity of MHC genes in different clades. The preponderance of evidence from the mammalian MHC is that different loci are maintained independently of each other after duplication events—the divergent evolution model (Ota and Nei 1994)—rather than being constantly homogenized by gene conversion, as in the concerted evolution model (box 1b). The presence of pseudogenes also implicated birth-and-death processes in the mammalian MHC (Hughes 1995). Some mammalian MHC loci are thought to have duplicated prior to the divergence of marsupial and placental mammals and have been evolving largely independently ever since.

Evolution of the MHC in birds

The chicken MHC. Until recently, the avian MHC was known only from work on the chicken MHC, or B complex (Guillemot et al. 1989, Kaufman et al. 1999). In sharp contrast to the mammalian MHC, chickens possess a genomic region very densely packed with genes (figure 2). The B complex is only 92 kb in length and contains only 19 genes making it approximately 1/20th the size of the human MHC (figure 3; Kaufman et al. 1999). B-complex genes themselves are streamlined, having introns that are an order of magnitude smaller than their mammalian counterparts. Kaufman has coined the phrase *minimal essential* to characterize the streamlined nature of the chicken MHC both in structure and function. Unlike the human MHC, which varies in the percentage of GC from 38% to 55%, the percentage of GC of the chicken B complex hovers consistently around 60% throughout its length, consistent with the idea that vertebrate coding regions usually are associated with higher GC isochores, between which we find high percentage regions of adenine–thymine (AT). Also in contrast to the mammalian MHC, chickens have two genetically distinct regions that both contain MHC genes. Aside from the B complex, the locus responsible for graft rejection, chickens have a second region, called Rfp–Y, that also contains polymorphic class I genes. Variation in Rfp–Y genes has been implicated in resistance to Marek's disease (Wakenell et al. 1996), although other results have contradicted this finding (Vallejo et al. 1998). Rfp–Y genotype is also suspected to affect resistance to another tumor-forming virus called Rous sarcoma (LePage et al. 2000). The Rfp–Y has not yet been sequenced, but it can be mapped to the same chromosome as the B complex and is physically separated from the B complex by the nuclear organizing region, a region of frequent recombination (Afanassieff et al. 2001).

The class III region of chickens (represented in this case by only a single gene, C4) is located outside the class I and II genes and not between them as is the case in humans. In addition, the TAP genes (which are involved in class I antigen presentation) are located in the vicinity of class I genes in chickens but in the class II region in mammals (Kaufman 1999).

Box 1. Gene conversion and multigene family evolution. Gene conversion is an important process shaping diversity at MHC genes both within and between loci.



Gene conversion is the unidirectional transfer of genetic material between loci. The process could occur through a number of different mechanisms including unequal crossing over or through a transposable element intermediate. Hogstrand and Bohme (1999) examined the frequency of gene conversion events in mice by sperm typing and estimated rates on the order of 2×10^4 conversions per generation. Sometimes gene conversion can decrease variation between loci, but it can also generate new variants by converting smaller DNA tracts, thereby creating chimaeric sequences (a). Loci evolving independently without gene conversion after duplication create a pattern in which all of the orthologous genes (homologous genes in different species) should cluster together in a phylogenetic tree (b). In this figure, the different shaded boxes represent different species, and the numbers represent different loci. Incidentally, this is the same pattern expected if gene conversion occurs in small tracts. A contrasting model (concerted evolution) predicts that gene conversion between loci within a species will homogenize the different loci (c). This process makes all of the genes within a species more similar to each other than they are to the orthologous loci in other species. When the concerted evolution model dominates, paralogues (genes in the same genome that share descent through duplication) should cluster together in a phylogenetic tree, because gene conversion events have occurred more recently than the duplication events that created the genes originally. This pattern is expected when gene conversion tracts are large.

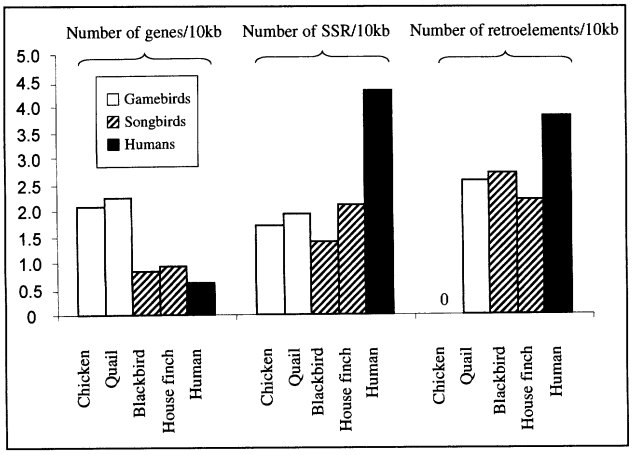


Figure 2. Differences among genetic elements of humans and birds. Preliminary sequencing of large fragments of DNA containing MHC genes suggests that the density of genes in chickens and quail is higher than in humans. Songbird MHC gene densities appear intermediate, although orthology to the chicken B complex is uncertain. The number of simple sequence repeats (SSRs [microsatellites]) is similar between gamebirds and songbirds but smaller in all birds than in mammals. The number of retroelements found in the quail and songbirds is smaller than that found in humans. There have not been any retroelements found in the chicken MHC.

The small size of the chicken MHC appears to restrict recombination between different loci. Recombinant individuals between the B–L and B–F loci (class I and class II) are extremely rare (Hala et al. 1988). Some hypothesize that the tight linkage between loci such as class I and TAP could help confer resistance to diseases, because polymorphisms at two loci covary synergistically with each other (Kaufman 1999).

The Japanese quail (Coturnix japonica) MHC.

Shiina and colleagues have sequenced a large region of the quail class I MHC (Shiina et al. 1999). They have found no evidence for class II genes being intermingled with class I genes as in chickens. Characterization of two cosmids yielded 12 MHC genes as well as TAP1. This pattern is consistent with a structurally minimal essential MHC, which has genes tightly packed together (figure 2).

Shiina and colleagues (1995) found at least four transcribed and presumably functional class I genes by examining a cDNA library. These four class I loci differ most in the amino acids that comprise the peptide binding region, a pattern consistent with functional class I genes in mammals. It is unknown whether only one of these loci is highly expressed as in chickens, but the general pattern appears more complex than the one observed in chickens. Phylogenetic analyses of the quail loci show that all four loci cluster together rather than with potential orthologues from chickens. This pattern suggests the concerted evolution or recent duplication models

(box 1) rather than the divergent model that is seen in mammals.

More recently, Shiina and colleagues have finished the sequencing of the entire region homologous to the B complex in the quail (Takashi Shiina [Tokai University School of Medicine, Kanagawa, Japan], personal communication, 2002). In addition to the class I genes described previously, they found seven class IIB genes, TAP1, TAP2, Tapasin, RING3, and other genes associated with the B complex. Interestingly, they found that there are multiple independent duplications of B–G genes (seven times), B-lectin (six times), and NKR genes (four times; figure 3). All of these genes are found as single copy genes in the chicken B complex. None of these MHC-associated genes have been found in any avian species other than chicken and quail.

The ring-necked pheasant (Phasianus colchichus) MHC. Wittzell and colleagues have found two class II MHC genes (Phco–DAB1 and DAB2) in ring-necked pheasants, a species also closely related to chickens (Wittzell et al. 1999a). Both of these genes are expressed in the spleen but at unknown levels. They differ mainly in the 3′ untranslated region (3′-UTR), with the transcription length of this region differing by 145 base pairs. Comparison of the 3′-UTRs of pheasant and chicken genes showed that Phco–DAB1 and DAB2 of the ring-necked pheasant are orthologous to the chicken B–LBI and B–LBII, respectively, as alleles from Phco–DAB1 and DAB2 clustered with their counterparts in chickens, making this the first demonstration of orthologous genes between two bird species. The split between pheasants and chickens occurred some 20 million years ago, and the duplication event that produced these two genes must predate that organismal split (Wittzell et al. 1999a). Despite the orthologous relationship between the two loci, the pheasant paralogues share a 10-base pair stretch in the third exon that is different from either of the orthologous genes in the chicken. This observation, along with the highly diverged 3′-UTR region, suggest that there is concerted evolution occurring between the exons of the two pheasant genes. The second exons of the Phco class II genes encode the extracellular peptide binding domain, they are highly variable, and they do not cluster by locus in a phylogenetic analysis. Wittzell and colleagues (1999a) suggest that genetic exchange between loci may be aided by the short intron between the second and third exons, an intriguing idea given the much smaller introns in gamebirds in contrast to those in mammals.

Like chickens, pheasants also possess a second, genetically independent region that also contains MHC genes (Wittzell et al. 1999a). It is unclear whether this second cluster is homologous to the chicken Rfp–Y because no large-scale

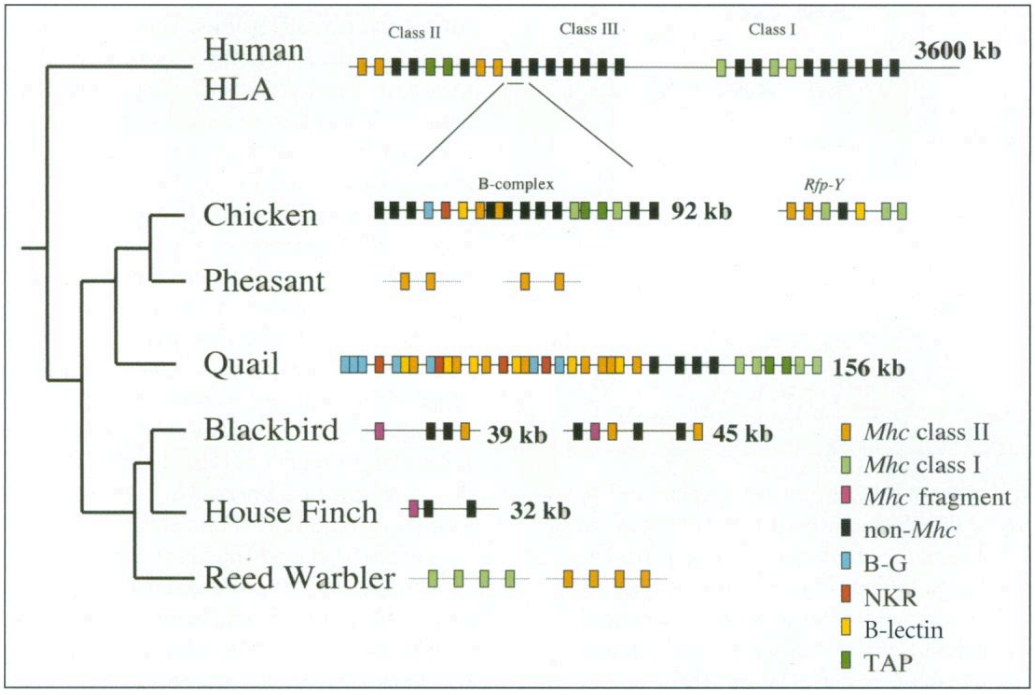
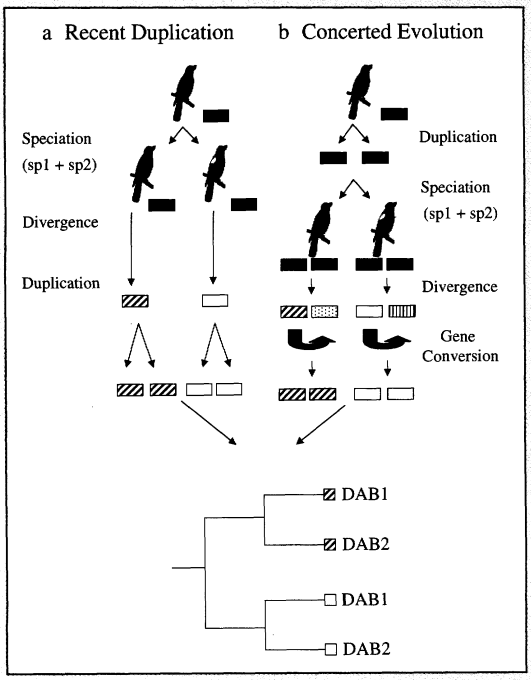


Figure 3. Comparison of genomic MHC structures in birds and human. The HLA (human leukocyte antigen) is much larger (3600 kb) than the corresponding region in the chicken (92 kb). The quail's pattern is similar to that of the chicken. The orange boxes represent class II genes, the light green—class I MHC genes, and the purple boxes represent MHC fragments or pseudogenes. Black squares represent non-MHC genes, except for B–G genes (light blue), natural killer receptor (NKR) genes (red), B-lectin genes (yellow), and TAP genes (green). (In the HLA schematic, each square represents multiple genes except for the TAP genes (Beck et al. 1999). The sizes of the pheasant and the reed warbler MHCs and orientation of these genes are unknown at the genomic level at this time (dashed line).

Box 2. Models of MHC evolution in birds.

The two main hypotheses for the mechanisms behind the clustering by species seen in birds are (a) the recent duplication model and (b) the concerted evolution model. Both models yield loci within a species more similar to each other than to the loci from closely related species. In recent duplication (a), speciation is followed by divergence in a single gene, which then duplicates to make two very similar copies of the gene. In concerted evolution (b), duplication occurs prior to speciation, but the duplicated genes are homogenized by gene conversion, once again making all of the loci within species similar. The different classes of shaded boxes represent different species, and the different numbers signify different loci.



sequences have been obtained from this species, and the flanking regions of the chicken Rfp–Y region are still unknown. Aside from some songbird species, MHC genes have also been examined by RFLP (restriction fragment length polymorphism) patterns in sandhill cranes (*Grus canadensis pratensis;* Jarvi et al. 1999) and by PCR (polymerase chain reaction) methods in seabirds (alcids) (Walsh 1999).

Songbird class I genes. Songbirds are the most diverse group of birds, comprising over half of all recognized species. They also have been the focus of many research programs on behavior, ecology, and evolution, making them an important group for linking organisms, diversity, and the MHC. In a groundbreaking study, Westerdahl and colleagues (1999) amplified the polymorphic α2 domain of great reed warbler, *Acrocephalus arundinaceus,* MHC class I genes, encoded by the third exon. Cloning these products produced as many as seven different sequences per bird, suggesting that there were at minimum four loci being amplified. Without the ability to characterize variation at a single locus, it was difficult to measure the intensity of selection acting upon the warbler class I genes. Still, the only class I genes yet characterized from a passerine bird revealed extensive variation despite estimated small population sizes from field studies (Westerdahl et al. 1999). Frameshift mutations found in some of these expressed genes suggest they are nonfunctional pseudogenes, indicating a genomic "noisiness" similar to that found in mammals.

Southern blot analyses in great reed warblers and starlings *(Sturnus vulgaris)* produce numerous bands reminiscent of DNA fingerprints when probed with a conspecific class I probe, indicating a large number of class I MHC loci in the genomes of these birds and MHC structures that are not minimal essential (Wittzell et al. 1999b). Although the class I MHCs of birds other than chickens clearly contain more loci than chickens, the critical experiments examining relative levels of expression of these loci have not been done. A large number of genomic or expressed MHC genes is not necessarily inconsistent with a functionally minimal essential MHC, provided that there is a single highly expressed locus.

Songbird class II genes. The class II MHC loci of songbirds have been characterized in a greater diversity of species than have class I genes. Like the great reed warbler studies, work has focused on ecologically and behaviorally well-characterized species such as red-winged blackbirds *(Agelaius phoenieceus),* house finches (*Carpodacus mexicanus),* and scrub jays *(Aphelocoma* spp.). Initial amplifications of class II genes with degenerate PCR primers from these species as well as from alligators showed that there was substantial diversity in these groups in the exon encoding the peptide-binding region (PBR) (Edwards et al. 1995). Most recently, Westerdahl and colleagues (2000) examined the class IIB genes of the great reed warbler and a closely related species, the willow warbler, *Phylloscopus trochilus.* There appeared to be a large number of class II loci in the warbler and starling, as was the case for class I loci. The researchers found substantial variation in RFLP genotype for both species and showed that the class I and class IIB restriction fragments were linked. Analysis of a great reed warbler cDNA library found seven considerably diverged sequences at PBR sites, which suggests that balancing selection is acting on these loci. One of the alleles had two deletions, making it a putative pseudogene.

Although degenerate PCR primers provide a critical foothold on class II MHC sequences, single loci are difficult to amplify by this method because of the sequence similarity of priming sites caused by conversion during concerted evolution or recency of duplication. Introns appear to escape these homogenizing events or diverge quickly postduplication; by placing PCR primers in the introns flanking the target exon, single loci have usually been effectively amplified (Hess et al.

2000). Surprisingly, obtaining intron sequences has proven resistant to most efforts using PCR and has required the implementation of genomic cloning and sequencing methods.

Cosmid cloning methods similar to those used to characterize the chicken MHC are being applied to songbirds. Using shotgun sequencing techniques, Hess and colleagues (2000) sequenced a 32 kb clone that contained an MHC gene, Came–DAB1, which was identified by southern blot. Paradoxically, it is now suspected that polymorphic and likely functional MHC genes frequently produce weakly hybridizing bands rather than strong ones on genomic blots (Dennis Hasselquist [Lund University, Sweden], personal communication, 2002). Although the cosmid sequenced by Hess and colleagues (2000) hybridized strongly to a house finch cDNA probe, there are numerous frameshift mutations in both the second and third exons of Came–DAB1, and the sequence is unalignable after the beginning of the third intron. Despite its apparent pseudogene status, Came–DAB1 has a higher level of genetic diversity than levels of variation typically detected in mammalian genomes. This increased diversity could be the result of hitchhiking with a locus under balancing selection nearby, a pattern observed in human MHC pseudogenes (Grimsley et al. 1998). Hess and colleagues (2000) found only a single MHC gene in over 32 kb of sequence. Although it is not known whether this region is homologous to the B complex in chickens, it is less gene-dense than the chicken or quail.

Red-winged blackbirds have received the most extensive treatment for class II genes. So far three genes (including one pseudogene) and flanking regions and two MHC fragments have been characterized on two sequenced cosmids (Edwards et al. 1998, 2000; Gasper et al. 2001). The pseudogene Agph–DAB2, which showed little polymorphism, and a highly polymorphic and presumably functional gene, Agph–DAB1, were estimated to have diverged from one another around 40 million years ago—relatively recently compared to most mammalian MHC counterparts. This calculation was made using the equation $d = 2\mu T$, where d is the percentage of sequence divergence between the two regions, μ is the silent (nonamino acid-changing) site divergence calculated in humans (1×10^{-9}), and T is the time since divergence. Additionally, the ratio of nonsynonymous to synonymous divergence between these genes was higher than expected, as was the number of reconstructed amino acid substitutions leading to Agph–DAB2 from the common ancestral gene, which suggests that the pseudogene was at one time experiencing the adaptive divergence characteristic of a functional MHC locus. A comparison of evolutionary histories of the polymorphic exon 2 of Agph–DAB1 and the adjacent intron 2 suggested that recombination is high between the two regions: The exon has a greater diversity than the intron, and the divergence time of exon alleles from a common ancestor was estimated to be more than four times that for the intron (Garrigan and Edwards 1999). Nonetheless, intron 2 is still suspected to be hitchhiking with the exon, as its diversity is much greater than that of neutral introns in blackbirds.

The sequence of a 45 kb fragment containing two MHC genes exhibits some surprising similarities to the HLA. Gasper and colleagues (2001) found that two large duplications spanning more than 20 kb were shared with another cosmid and contained 13 retroelements (figure 2), a surprisingly high density (6% of total sequence) for birds. Unlike the pheasant class II alleles, alleles at the three separate blackbird genes tended to cluster by locus, suggesting recent interlocus gene conversion or duplication.

Paleodemography

The long coalescence times of MHC alleles provide unexpected opportunities for probing the paleodemography of species—the long-term history of population size prior to and during speciation. Vincek and colleagues (1997) examined class II MHC sequences of Darwin's finches *(Geospiza* spp.), a group of species all of which have recently radiated from a common ancestor. The finches diverged within the last 5 million years, since the formation of the Galapagos Islands, although the evidence suggests they diverged from the outgroups very recently. The researchers found a presumably functional sequence (no frameshift mutations) in Bengalese finches *(Lonchura striata)* and used this information to design PCR primers for use on Darwin's finches. They were interested in employing MHC allelic variation to estimate the size of the founding population of the finch common ancestor. Phylogenetic analysis showed that there were four clusters of alleles, suggesting four separate loci. The MHC diversity suggested a minimum population size of 30 individuals for the founding group for this radiation. The substitution rate for chickens and pheasants was applied to the finch phylogenetic tree, a difficult calibration given the nature of selection acting on MHC genes, the possibility of concerted evolution, and the likely violation of assumptions of a molecular clock. Regardless, this study is an intriguing use of the variation at MHC genes to probe population history.

Link between MHC and phenotypic variation

In an early effort to link molecular variation at the MHC and phenotypic traits whose expression is likely influenced by disease, von Schantz and colleagues (1996) found that spur length in male pheasants correlates with both viability and MHC restriction fragment pattern. This study is still the only attempt at a link between variation at immune genes, a condition-dependent trait (spur length), and fitness in a natural population of birds. However, the proximate mechanisms facilitating this relationship remain to be described, although there are logical connections between them. Von Schantz and colleagues (1996) did not claim to show that birds with shorter spurs are more susceptible to disease or that birds with different MHC haplotypes vary in their disease-fighting ability. This level of detail awaits a new generation of ecological immunology studies involving the MHC, several of which are now under way.

The future

Much progress has been made in characterizing the MHCs from nonmodel species, particularly in birds. What sorts of questions will characterization of this region in multiple species—a comparative genomics approach—allow us to answer?

Coadapted gene complexes. An intriguing idea is that the compactness of the chicken MHC might promote long-term associations between different loci leading to enhanced ability to prevent disease. Specific variants at nearby loci are inherited in tandem (as single haplotypes), making them a single coevolving unit. It will be important to try to compare variation at MHC genes and nearby loci such as TAP1 or TAP2 to test whether these genes evolve in tandem (Kaufman 1999). Probing avian cosmid libraries for these linked but single copy, non-MHC genes may also prove fruitful in finding genomic regions homologous to the B complex, which are likely to be directly involved in the adaptive immune response and might play a greater role in determining susceptibility and resistance. The number of MHC genes in some species appears to be too great for the MHC genes themselves to be used efficiently as probes to target regions syntenic with the B complex, and this alternative approach might be more effective.

Long-term MHC evolution. The patterns of molecular evolution in avian MHC seem to differ from the divergent evolution model typical of mammals. Instead of most loci being old and maintained independently over time, the avian MHC is characterized by recent duplication and pseudogene formation. Periodic bouts of gene conversion appear to have made different loci within species more similar to each other than to their orthologues in closely related species. What is the empirical relationship between number and physical proximity of genes and the frequency of interlocus gene conversion? Given that this homogenizing force appears common and tends to remove variation, how can the genetic diversity at those genes be explained? Walsh (1986) has shown that gene families that undergo additive selection and gene conversion can respond to weak selection more readily than single loci, where genetic drift might dominate. In another simulation study, Ohta (1991) found that gene conversion, especially from nonexpressed loci, could contribute greatly to the high levels of polymorphism. Identity excess (or linkage disequilibrium) among loci increased with stronger selection and facilitated diversification even when selection was weak. Systematic sampling of closely related species will determine at what time scale the diversification and homogenization of avian MHC genes occur, will provide estimates of the size of the conversion units, and should clarify the nature of interactions between selection, linkage, and gene conversion.

Ecological immunology. Parasites have long been thought to have a dramatic effect on host evolution. Is there a relationship between genetic variation in the immune system and the ability to mount an effective response to a pathogen in bird populations? We expect that over time, there will be strong selection for specific variants or heterozygosity. Brown (1997) has proposed that mate choice by females may function to increase heterozygosity in offspring. In this scenario, females prefer to mate with individuals who differ at the MHC, a pattern seen in seminatural populations of mice (Potts et al. 1991). The ability to discern differences in MHC type in the mice is olfactory in nature, a prospect that at first seems unlikely in most birds (save for some seabirds and vultures that have demonstrated powers of olfaction) but becomes more plausible when individual species are examined in detail (Zelano and Edwards forthcoming). Other mechanisms of sexual selection include linkage of MHC type to strong sexually selected characters that are costly to produce. Studies comparable to those done by von Schantz and colleagues (1996) on chickens are under way in a diversity of bird species. New techniques for estimating various aspects of the immune response provide a multipronged approach to these questions (Moller et al. 1999, Zelano and Edwards forthcoming).

The MHC and disease resistance in birds appears to be more tightly linked than in mammals. Branching out into other avian species promises to greatly improve our understanding of the interplay between forces operating at the genetic level as well as those affecting organisms. Evolutionary biologists from the organismal to the genetic level have long sought to understand the forces leading to biodiversity and genetic diversity. The immune system, particularly the MHC, is one of the most spectacular examples of this kind of diversity. The combination of recent advances in assaying aspects of the immune system (Moller et al. 1999) and large-scale techniques to uncover molecular variation underlying these traits promises to make birds both a focus of the growing field of ecological immunology and as important models in this emerging area of evolutionary biology.

Acknowledgments

We thank Takashi Shiina and Marcia Miller for giving us access to results that at the time of writing were not published, as well as Ming Lee and Marc Rieder for technical assistance. We also thank the members of the Rohwer lab "green table" discussion group, Jim Kaufman and three anonymous reviewers for helpful comments on the manuscript, as well as Rebecca Chasan for initially requesting this review. This work was supported by NSF grants DEB 9797548 and 9815800 to S. V. E. and a graduate recruitment award to C. M. H.

References cited

Afanassieff M, Goto RM, Ha R, Sherman MA, Zhong L, Auffray C, Coudert F, Zoorob R, Miller MM. 2001. At least one class I gene in restriction fragment-Y (Rfp-Y), the second MHC gene cluster in the chicken, is transcribed, polymorphic, and shows divergent specialization in antigen binding region. Journal of Immunology 166: 3324–3333.

Beck S, et al. 1999. Complete sequence and gene map of a human major histocompatibility complex. Nature 401: 921–923.

Bergstrom TF, Josefsson A, Erlich HA, Gyllensten UB. 1997. Analysis of intron sequences at the class II HLA-DRB1 locus: Implications for the age of allelic diversity. Hereditas 127: 1–5.

Briles WE, McGibbon WH. 1948. Heterozygosity of inbred lines of chickens at two loci affecting cellular antigens. Genetics 33: 605.

Brown JL. 1997. A theory of mate choice based on heterozygosity. Behavioral Ecology 8: 60–65.

Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, O'Brien SJ. 1999. HLA and HIV-1: Heterozygote advantage and B*35-Cw*04 disadvantage. Science 283: 1748–52.

Clarke B, O'Donald P. 1964. Frequency-dependent selection. Heredity 19: 201–206.

Cole RK. 1968. Studies on the genetic resistance to Marek's disease. Avian Diseases 12: 9–28.

Edwards SV, Grahn M, Potts WK. 1995. Dynamics of Mhc evolution in birds and crocodilians: Amplification of class II genes with degenerate primers. Molecular Ecology 4: 719–729.

Edwards SV, Gasper J, March M. 1998. Genomics and polymorphism of AgphDAB1, an Mhc class II B gene in red-winged blackbirds *(Agelaius phoeniceus)*. Molecular Biology and Evolution 15: 236–250.

Edwards SV, Gasper J, Garrigan D, Martindale D, Koop BF. 2000. A 39-kb sequence around a blackbird Mhc class II gene: Ghost of selection past and songbird genome architecture. Molecular Biology and Evolution 17: 1384–1395.

Garrigan D, Edwards SV. 1999. Polymorphism across an exon-intron boundary in an avian Mhc class IIB gene. Molecular Biology and Evolution 16: 1599–1606.

Gasper JG, Shiina T, Inoko H, Edwards SV. 2001. Songbird genomics: Analysis of 45 kb upstream of a polymorphic Mhc class II gene in a red-winged blackbird *(Agelaius phoeniceus)*. Genomics 75: 26–34.

Gorer PA. 1936. The detection of antigenic differences in mouse erythrocytes by the employment of immune sera. British Journal of Experimental Pathology 17: 42–50.

Grimsley C, Mather KA, Ober C. 1998. HLA-H: A pseudogene with increased variation due to balancing selection at neighboring loci. Molecular Biology and Evolution 15: 1581–8.

Guillemot F, Kaufman JF, Skjoedt K, Auffray C. 1989. The major histocompatibility complex in the chicken. Trends in Genetics 5(9): 300–304.

Hala K, Chaussé A-M, Bourlet Y, Lassila O, Hasler V, Auffray C. 1988. Attempt to detect recombination between B-F and B-L genes within the chicken B complex by serological typing, in vitro MLR and RFLP analyses. Immunogenetics 28: 433.

Hess CM, Gasper J, Hoekstra HE, Hill CE, Edwards SV. 2000. MHC class II pseudogene and genomic signature of a 32-kb cosmid in the house finch *(Carpodacus mexicanus)*. Genome Research 10: 613–623.

Hickson RE, Cann RL. 1997. Mhc allelic diversity and modern human origins. Journal of Molecular Evolution 45: 589–98.

Hill AVS, et al. 1991. Common West African HLA antigens are associated with protection from severe malaria. Nature 352: 595–600.

Hogstrand K, Bohme J. 1999. Gene conversion can create new MHC alleles. Immunological Reviews 167: 305–17.

Hughes AL. 1995. Origin and evolution of HLA class I pseudogenes. Molecular Biology and Evolution 12: 247–258.

Jarvi SI, Goto RM, Gee GF, Briles WE, Miller MM. 1999. Identification, inheritance, and linkage of B-G-like and MHC class I genes in cranes. Journal of Heredity 90(1): 152–159.

Kaufman J. 1999. Co-evolving genes in MHC haplotypes: The "rule" for nonmammalian vertebrates? Immunogenetics 50: 228–36.

Kaufman J, Milne S, Göbel TWF, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S. 1999. The chicken B locus is a minimal-essential major histocompatibility complex. Nature 401: 923–925.

Klein J, Satta Y, O'hUigin C, Takahata N. 1993. The molecular descent of the major histocompatibility complex. Annual Review of Immunology 11: 269–95.

LePage KT, Miller MM, Briles WE, Taylor RL Jr. 2000. Rfp-Y genotype affects the fate of Rous sarcomas in B^2 B^5 chickens. Immunogenetics 51: 751–754.

Medawar P. 1958. Part played by deoxyribonucleic acid in transplantation immunity. Nature 182: 62.

Moller AP, Christe P, Lux E. 1999. Parasitism, host immune function, and sexual selection. Quarterly Review of Biology 74: 3–20.

Ohta T. 1991. Role of diversifying selection and gene conversion in the evolution of major histocompatibility complex loci. Proceedings of the National Academy of Sciences 88: 6716–6720.

Ota T, Nei M. 1994. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin V_H gene family. Molecular Biology and Evolution 11: 469–482.

Parham P, Ohta T. 1996. Population biology of antigen presentation by MHC class I molecules. Science 272: 67–74.

Potts WK, Manning CJ, Wakeland EK. 1991. Mating patterns in seminatural populations of mice influenced by MHC genotype. Nature 352: 619–21.

Shiina T, Ando A, Imanishi T, Kawata H, Hanzawa K, Gojobori T, Inoko H, Watanabe S. 1995. Isolation and characterization of cDNA clones for Japanese quail *(Coturnix japonica)* major histocompatibility complex (MhcCoja) class I molecules. Immunogenetics. 42: 213–216.

Shiina T, Shimizu C, Oka A, Teraoka Y, Imanishi T, Gojobori T, Hanzawa K, Watanabe S, Inoko H. 1999. Gene organization of the quail major histocompatibility complex (MhcCoja) class I gene region. Immunogenetics 49: 384–394.

Snell GD. 1948. Methods for the study of histocompatibility genes. Journal of Genetics 49: 87–108.

Thursz MR, Thomas HC, Greenwood BM, Hill AVS. 1997. Heterozygote advantage for HLA class-II type in hepatitis B virus infection. Nature Genetics 17: 11–12.

Trowsdale J. 1995. 'Both bird and man and beast': Comparative organization of MHC genes. Immunogenetics 41: 1–17.

Vallejo RL, Pharr GT, Liu HC, Cheng HH, Witter RL, Bacon LD. 1997. Nonassociation between Rfp-Y major histocompatibility complex-like genes and susceptibility to Marek's disease virus-induced tumours in 6(3) x 7(2) F–2 intercross chickens. Animal Genetics 28(5): 331–337.

Vincek V, O'hUigin C, Satta Y, Takahata N, Boag PT, Grant PR, Grant BR, Klein J. 1997. How large was the founding population of Darwin's finches? Proceedings of the Royal Society of London Series B 264: 111–118.

Von Schantz T, Wittzell H, Goransson G, Grahn M, Persson K. 1996. MHC genotype and male ornamentation: Genetic evidence for the Hamilton-Zuk model. Proceedings of the Royal Society of London Series B 263: 265–71.

Wakenell PS, Miller MM, Goto RM, Gauderman WJ, Briles WE. 1996. Association between the Rfp-Y haplotype and the incidence of Marek's disease in chickens. Immunogenetics 44: 242–5.

Walsh JB. 1986. Selection and biased gene conversion in a multigene family: Consequences of interallelic bias and threshold selection. Genetics 112: 699–716.

Walsh HW. 1999. Polymorphism at neutral and non-neutral loci in least and crested auklets *(Aethia* spp.) and implications for speciation. Master's thesis, Queen's University, Kingston, Ontario, Canada.

Westerdahl H, Wittzell H, von Schantz T. 1999. Polymorphism and transcription of Mhc class I genes in a passerine bird, the great reed warbler. Immunogenetics 49: 158–170.

———. 2000. Mhc diversity in two passerine birds: No evidence for a minimal essential Mhc. Immunogenetics 52: 92–100.

Wittzell H, Bernot A, Auffray C, Zoorob R. 1999a. Concerted evolution of two Mhc class II B loci in pheasants and domestic chickens. Molecular Biology and Evolution 16: 479–490.

Wittzell H, Madsen T, Westerdahl H, Shine R, von Schantz T. 1999b. MHC variation in birds and reptiles. Genetica 104: 301–9.

Zelano B, Edwards SV. An Mhc component to kin recognition and matechoice in birds: Predictions, progress, and prospects. American Naturalist. Forthcoming.

Zuckerkandl E, Pauling L. 1965. Evolutionary divergence and convergence in proteins. Pages 97–166 in Bryson V, Vogel HJ, eds. Evolving Genes and Proteins. New York: Academic Press.