Evolution and ecology of MHC molecules: from genomics to sexual selection

Scott V. Edwards and Philip W. Hedrick

In the past few years the DNA sequence database for molecules of the MHC (major histocompatibility complex) has expanded greatly, yielding a more complete picture of the long-term rates and patterns of evolution of the MHC in vertebrates. Sharing of MHC allelic lineages between long-diverged species (trans-species evolution) has been detected virtually wherever it is sought, but new analyses of linked neutral regions and the complexities of sequence convergence and microrecombination in the peptide binding region challenge traditional phylogenetic analyses. Methods for estimating the intensity of selection on MHC genes suggest that viability is important, but recent studies in natural populations of mammals give inconsistent results concerning mate choice. The complex and interacting roles of microrecombination, parasite-mediated selection and mating preferences for maintaining the extraordinary levels of MHC polymorphism observed are still difficult to evaluate.

Scott Edwards is at the Dept of Zoology and Burke Museum, University of Washington, Box 351800, Seattle, WA 98195, USA (sedwards@u.washington.edu); Philip Hedrick is at the Dept of Biology, Arizona State University, Tempe, AZ 85287-1501, USA (hedrick@hedricklab.la.asu.edu).

As an extremely polymorphic multigene family that encodes receptors crucial to vertebrate immune defense against parasites (Box 1), the major histocompatibility complex (MHC) has relevance for a variety of fields from molecular evolution to animal behavior. When *TREE* published its first review of MHC evolution eight years ago[1], many of the molecular techniques for detecting MHC diversity, such as PCR, were still relatively new and had been applied to only a handful of rodent and great ape species. The population genetic theory that informs current analyses of MHC diversity was just being developed[2], experiments addressing the functional consequences of MHC variation for mate choice and other fitness components were still confined to inbred strains of mice, and the best examples in humans of associations between MHC and resistance to infectious disease were still forthcoming[3,4].

In the past few years, the application of new technical and theoretical approaches to studying MHC history and diversity, as well as new field experiments on non-model and outbred species, have all conspired to make the MHC an inescapable focus in ecology, animal behavior and evolutionary biology. The molecular complexity associated with this large and sometimes redundant multigene family (as well as the equally daunting nomenclature) has proved an intimidating barrier to the rigorous, detailed hypothesis-testing envisioned for many organisms (particularly nonmammalian vertebrates) by ecologists eight years ago. However, the flood of sequence data and the investigation of novel populations with new tools will continue to refine our knowledge of the causes and consequences of variation in this extraordinarily polymorphic system.

Genomics and long-term evolution

One way to appreciate the extraordinary diversity associated with the MHC is through the examination of its structure and proximate function in immune recognition. Although the molecular details of antigen-binding and processing continue to be refined, the basic function of class I and II MHC molecules in presenting degraded foreign and self-peptides to T-cells has remained stable (Box 1, Fig. 1). What will probably have more immediate impact on zoologists and human geneticists interested in MHC evolution are new discoveries emanating from the Human Genome Project. Currently, several teams in the UK, Japan and the USA aim to sequence the human and mouse MHCs in their entirety (approximately four million base pairs of DNA in each species). In addition, efforts to sequence the cosmid clusters mapped for the chicken MHC (or 'B-complex') are currently in progress (J. Kaufman, C. Auffrey, L. Hood, pers. commun.). The first reports of such projects are just appearing[5–7]. These will initiate several aspects of research on MHC ecology and evolution, including data on comparative organization, discovery of new genes, mapping of genomic regions influencing disease resistance, and the dynamics of duplications and isochores.

For example, in sequencing part of the human class I region, Mizuki *et al.*[6] discovered a shared 40-kb segment downstream of both the *HLA-B* and *-C* genes, providing detailed information on the boundaries of sequence duplication that could have given rise to these paralogous genes (i.e. genes on the same chromosome that are homologous by duplication). Guillaudeux *et al.*[7] compared 40-kb of noncoding DNA between the *HLA-B* and *-C* genes for two haplotypes and in some regions detected differences in excess of 7% of sites – a very high level of polymorphism for human noncoding DNA, perhaps because of linkage to genes under balancing selection. In addition, they discovered a novel 'class I homology unit' possessing conserved features of several class I genes, which they speculated could be involved in generating *HLA* diversity via gene conversion. Although many of the large-scale MHC sequencing efforts are still in a descriptive phase, they are expected to provide the most substantial database to date for testing molecular evolutionary hypotheses. Similar projects have already done so for several completed genome projects.

In most vertebrates investigated, MHC genes of all three classes (I–III; Box 1, Fig. 1) are linked in a cluster on the same chromosome, and in chickens the two sections of the MHC (the B-complex and the *Rfp-y* gene) that appeared to segregate independently in pedigrees turn out to be linked on the same microchromosome but separated by a region of high recombination[8]. However, in *Xenopus* frogs, a family of

  0169-5347/98/$19.00 PII: S0169-5347(98)01416-5

nonclassical class I genes (MHC genes with low levels of expression and polymorphism) exists outside and is unlinked to the main MHC genes[9]. MHC mapping in zebrafish (*Danio*) and other perciforms[10] show that class I and II genes are indeed found on different linkage groups, raising the possibility that the clustering of MHC genes in a single chromosomal region in mammals could be derived from a formerly dispersed condition. In contrast, Kasahara *et al.*[11] proposed a model in which a major chromosomal region harboring many genes duplicated as a block to another chromosome to give rise to the vertebrate MHC. This model was based on the finding that duplicated copies of several genes involved with antigen processing and other MHC-related functions (the Z subunit of the proteasome, HSP70, a heat shock protein, and ABC transporters; Box 1) map to a non-MHC bearing chromosome in mice and humans[11]. Consideration of the types and taxonomic distribution of genes participating in this duplication suggested that it may have occurred at or before the emergence of cartilaginous fish.

Recent attempts to detect MHC or proteasome subunit sequences (low molecular mass protein, LMP; Box 1) in primitive chordates have added to the controversy over the timing of the origin of MHC molecules. Several laboratories examining expressed genes in cyclostomes (jawless fish) have attempted but failed to detect MHC genes. In addition, although genes such as *LMP2* and *LMP7*, which are thought to confer class I specificity to peptides generated by the proteasome (Box 1), are known in sharks, only their constitutively expressed and proteolytically more generalized counterparts (e.g. *LMPX* and *LMPδ*) are known from jawless fish[12]. A phylogenetic and molecular clock analysis of these and other LMP genes[13] suggests that the divergence of the *LMP7* and *LMPX* genes occurred some 600 million years ago, before the divergence of jawed and jawless fish. This in turn implies that the origin of class I MHC molecules also predated the fish divergence and predicts that the MHC could have been lost in jawless fish[13]. LMP genes are among those with paralogous copies on non-MHC bearing chromosomes[11].

Box 1. The ABCs of the MHC

The MHC is the 'center of the immune universe'[50]. MHC molecules are members of the large **immunoglobulin superfamily**. The extracellular components of MHC molecules are composed of two main parts, a 'stalk' (the immunoglobulin domain) that anchors the molecule in the surface of cells, and a receptor, shaped like a groove or basket, called the **peptide binding region** (PBR). The function of class I and II MHC molecules in immune recognition is to 'display' short protein fragments ('peptides') of viruses and other pathogens on the surface of immune system cells, such as macrophages, or other cell types. **Complement** genes encode proteins that provide many effector functions critical to antibody-mediated immunity and inflammation. Class I PBRs, composed of α1 and α2 domains, are encoded by two exons in a single gene, are expressed on all cell types and bind peptides synthesized within cells, whereas class II PBRs are encoded by two genes (α and β chains), are expressed primarily on lymphocytes and bind peptides that enter such cells via phagocytosis. Most PBRs of class I and II genes are polymorphic but some, particularly of **nonclassical** MHC genes, are not. These nonpolymorphic genes can be on their way to becoming nonfunctional or may be specified to bind certain phylogenetically conserved motifs or structures of bacterial peptides, thereby ensuring a constant ability to detect a wide range of related pathogens. The PBRs of both class I and class II molecules might have originated from domains in heat shock proteins such as HSP70, which also bind peptides. In a process known as **antigen processing**, these peptides are, in the case of class I molecules, transported by polymorphic **TAP** ('transporter associated with antigen processing') molecules to a collection of protein-degrading enzymes (including low molecular mass proteins, or **LMP**) tightly-bound in a structure called the **proteasome** in the endoplasmic reticulum. In the proteasome they are then cut into appropriate sizes, usually 9–20 amino acids, depending on the peptide source and ultimate destination. During or after this process, they are loaded onto empty MHC molecules. The peptide–MHC pair are then shuttled to the surface of cells where they can be interrogated by populations of T-cells, which have on their surface **T-cell receptors**, also members of the immunoglobulin superfamily. An appropriate match between peptide, MHC molecule and T-cell receptor is necessary to produce an immune response consisting of rapid proliferation of highly specific lineages of T-cells programmed to kill cells infected with the foreign peptide, or proliferation of cells that in turn produce antibodies that can smother and prepare such cells for phagocytosis.

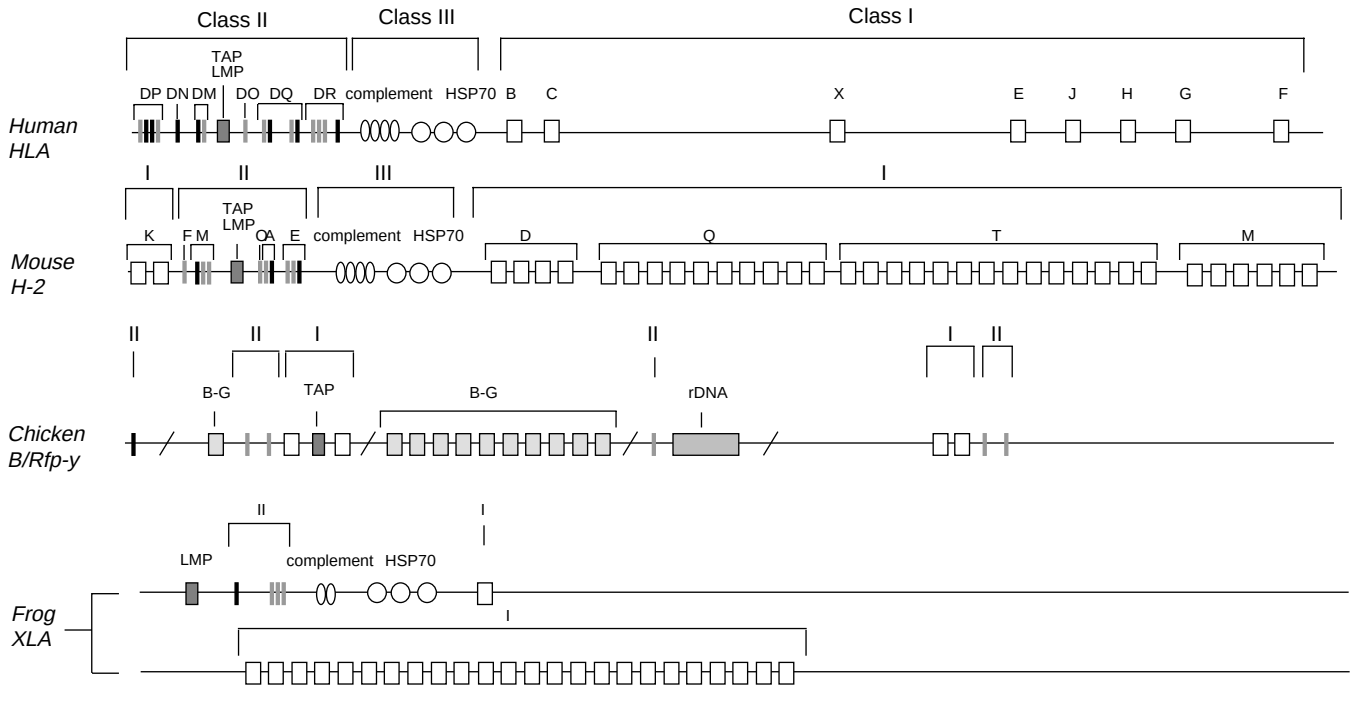


Fig. 1. Comparative genomic organization of MHC class I, II and III genes in various representative vertebrates. Species are indicated at the left. Gene and subfamily names are indicated above each chromosome, as are the class I–III regions. The smallest rectangles denote class II genes, where black and gray signify the α or β chain, respectively. The TAP, LMP, complement and *HSP70* genes (Box 1) are indicated. In the chicken, B-G signifies a large family of ubiquitously expressed and polymorphic receptor genes, the 'rDNA' signifies the nucleolar organizer region and the slashes indicate noncontiguous cosmid clusters and recombination hotspots[8,42]. The locations and numbers of genes are only approximate. *Modified, with permission, from Ref. 50.*

Therefore, the interpretation of gene maps, which favors an origin of the MHC after jawless fish branched off, and that from phylogeny and molecular dating of duplicated LMP genes are in conflict. Further investigation of the immune systems of jawless fish (which appear to lack a fully-fledged MHC region), T-cell receptors and immunoglobulins, as well as scrutiny of the constancy of evolutionary rates among proteasome subunit genes[12], will clarify the picture.

Recent phylogenetic surveys of MHC sequences[14,15] allow competing models of multigene family evolution that differ from one another in the fates of their duplicated genes and rates of homogenization to be tested (Box 2). Class I pseudogenes and nonclassical genes are often closely related to classical genes in the same species, suggesting recent duplications. Similarly, phylogenies of multiple class II MHC sequences in birds[14] and some primates[16] also imply recent postspeciation divergence. The class I and II genes in mammals conform to the 'birth-and-death' model[15], in which some duplicated genes are retained in descendent genomes for long periods, whereas others are deleted or become pseudogenes. Inevitably, when one compares class I or class II sequences between major vertebrate groups, genes within each lineage cluster together and orthologous relationships (i.e. homologies due to direct ancestry) are not retained, suggesting that ancestral MHC genes have duplicated independently in different lineages in rounds of expansion and contraction[17]. The rates and timescales over which such duplications and contractions take place need to be estimated with greater rigor in more diverse lineages. The fitness advantages of families of comparatively closely related MHC genes[14], as opposed to the long-diverged ones exemplified by several mammalian class II genes, remain obscure and appear to contradict the paradigm that MHC genes are diverse at both the level of individual genes and gene families.

Surprises continue to emerge from investigations of MHC structure in model and nonmodel organisms. Figueroa *et al.*[18] found that the MHC class II B genes of cichlid fish and other Perciformes, as well as those of representative acanthopterygian (rayfinned) fish, possess an extra intron in the immunoglobulin-like domain. This intron not only serves as a useful phylogenetic marker uniting rayfinned fish, but also provides further evidence for rounds of duplication of MHC genes after the divergence of the major groups of fish. Investigation of new MHCs inevitably reveals variation in intron length among species, and microsatellites occur in many MHC introns and flanking regions. Meager and Potts[19] found that mouse microsatellites within the MHC were more polymorphic than those flanking the MHC, suggesting increased microsatellite diversity owing to hitchhiking with selected MHC genes.

Generation of new variants

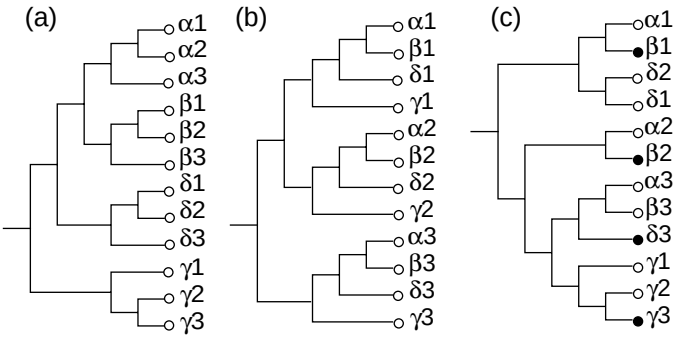
Those sequences of MHC genes that encode the peptide-binding region (PBR) are the focus of much of the population genetic research on MHC diversity (Box 1). Under the neutral theory, the rate of synonymous nucleotide substitution is predicted to be larger than the rate of nonsynonymous substitution. However, as originally shown by Hughes and Nei[20], the rate of nonsynonymous substitutions is generally higher than that for synonymous ones for PBR sites of MHC genes, suggesting the action of balancing selection.

Current evidence favors an important role for gene conversion or microrecombination in generating some MHC diversity[21], particularly in the PBR, although the number of studies attempting to detect conversion empirically (as opposed to indirect analyses of DNA sequence patterns) are few, and the relative importance of point mutation versus conversion is still controversial. Indirect evidence for microrecombination comes from alleles that differ from one another only by small segments, and the demonstration that the majority of intron sequences on both sides of the PBR have similar evolutionary histories, a pattern that makes single-crossover recombination unlikely[22]. The average size of the putative exchanges among human class I alleles is estimated to be between 15 and 20 nucleotides, and a recent global survey found evidence for only four single-point mutations, whereas at least 80 sequences suggested gene conversion[23]. An example of microrecombination with functional consequences is the *HLA-B**5301 sequence, which is thought to confer resistance to malaria and is identical to the common, nonprotective *B**3501 allele, except for exchanged codons 77 to 83 (Ref. 3). Several new alleles in South Amerindians appear to be the result of microrecombination between South Amerindian alleles[23]. Because South America has probably been populated for only the last 10 000 to 20 000 years (1000 human generations or less), the new variants, which have not yet been found in 'ancestral' Asian samples, must have arisen very recently, suggesting a high rate of generation of new variants.

There is direct evidence that the rate of microrecombination at some MHC loci is high. Zangenberg *et al.*[24] examined the rate of interallelic gene conversion in human sperm at a locus in which males were heterozygous for six highly variable regions. In 111 675 sperm, they observed nine interallelic conversions at a rate of 0.81×10^{-4} per generation, nearly 100 times greater than the prevailing recombination rate in the human genome (1% per megabase per generation)

Box 2. Multigene family evolution and the MHC

Three models that could potentially explain the long-term pattern of gene duplications in the MHC and other multigene families have recently been formalized[15]. Each of these three models has distinct predictions of the shape of the phylogenetic trees linking different genes together. In each tree in the figure, α, β, δ and γ refer to different genes, and numbers refer to species in which those genes are found. Open circles are functional genes and filled circles are pseudogenes. The divergent evolution model (a) is the simplest model, in which gene duplication is followed by maintenance of paralogues in descendent species by selection. In the case of the MHC, there would be an advantage to an increased number of MHC receptors with which to combat pathogens, but theoretically there is an optimum number of genes, above which too many immune cells potentially recognizing self would be deleted during ontogeny of the immune system, resulting in lowered fitness. The concerted evolution model (b), known for other multigene families, such as ribosomal genes, is marked by replacement of all gene family members in a genome by a single variant gene copy in the array via gene conversion and divergence of entire arrays in different species, resulting in groups of genes clustered phylogenetically by species. The concerted evolution model predicts that alleles of a given MHC locus will not cluster together relative to alleles at other loci involved in gene conversion events. Recent analyses of class I and II MHC genes in mammals[15] show that this is not the case, and that birth and death is a more appropriate model for these loci. The 'birth-and-death model' (c) posits frequent retention of ancestrally duplicated genes, as in the divergent evolution model, but also incorporates the evolution of pseudogenes and deletion of genes such that the total gene number is approximately maintained.



and several orders of magnitude higher than the estimated rate of silent substitution (Box 3). Hogstrand *et al.*[25] estimated the rate of interlocus gene conversion between two loci in mice as 0.5×10^{-6}, much lower than the rate of interallelic gene conversion but high enough to influence variation in the genes involved. Although the amount of microrecombination appears to vary over MHC loci, these observations suggest a stronger role for microrecombination in generating diversity and reveal the inadequacy of models assuming independence of nucleotide changes, an assumption fundamental to many calculations of molecular evolution. In addition, they raise the question of whether these elevated estimates of microrecombination rates are in part due to the ease of detectability of such events in the MHC, which are apparently because of the long timescale of neutral and balanced polymorphisms in this region[26].

Trans-species evolution: testing assumptions

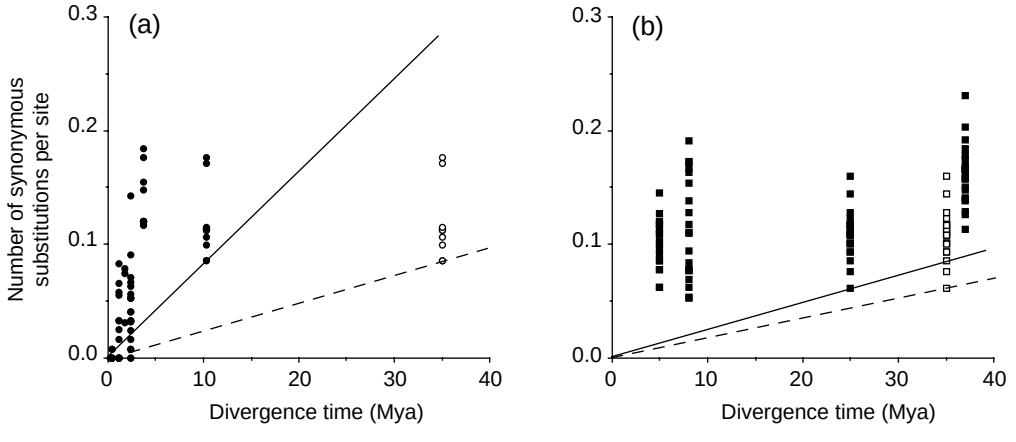
Trans-species evolution refers to the retention of alleles across multiple species divergences over a long time, and its hallmark is a discordance between gene and species trees that cannot be explained solely by the retention of neutral polymorphisms over relatively short timespans after speciation. Indeed, the population dynamics of MHC alleles closely match those for one other system only, the self-incompatibility alleles in plants[51]. It is now common for surveys of sequence variability in the PBRs of the MHC to reveal trans-species patterns when multiple species are investigated. As more data and analyses have accumulated, however, the causes of these allelic patterns are becoming less obvious. An increasing number of examples of convergence in the PBR of mammalian MHC genes at both the nucleotide and amino acid sequence levels raises doubts about some claims for extremely long persistence times, particularly from analyses based solely on PBR sites[16,27]. Phylogenies of PBRs and adjacent (and presumably neutral) introns and exons invariably conflict, with neutral sites implying a more attenuated picture of trans-species evolution[14,16,22,28].

In principle, the persistence of MHC alleles over long timespans provides unique opportunities for probing organismal histories far back in time – much farther back than would be possible with neutral genes, such as mtDNA. This property combined with novel analytic methods have been used to estimate ancestral effective population sizes of humans[29], cichlid fish (Cichlidae)[30] and most recently, Darwin's finches (*Geospiza*)[31]. Each of these studies implied less extreme founder effects in the origin of these species than previously assumed because current levels of MHC diversity in these species were too high to be explained by drastic bottlenecks in the past, assuming mutation rates of the order of 10^{-9} per site per year in the PBR (Box 3) and various selection coefficients. As these authors acknowledge, however, the large number of variables and assumptions required to implement such analyses (mutation rates, selection coefficients and absence of recombination) and difficulties in determining orthologous relationships among sequences for nonmammalian vertebrates make application to data from diverse creatures and environments challenging. For example, the much higher rates of change by nonpoint mutational mechanisms[24] and significant levels of post-speciation diversification suggested for some MHC loci might suggest alternative scenarios. Bergström *et al.*[22] sequenced the two introns flanking the polymorphic second exon of 28 class II *DRB1* alleles in 13 major allelic lineages of humans. Surprisingly, within each allelic lineage, the sequence divergence of alleles implied a coalescence time of about 250 000 years, a common ancestry that is recent for a human nuclear gene and which contrasts with the more ancient coalescence implied by variation in the PBR itself[29] or by the intron diversity across allelic lineages at that locus[22]. Their data suggest recent origins for the majority of alleles at *DRB1*, after the divergence of chimps and humans, rather than prespeciation origins as suggested by coding sequences. Combined with the ancient divergences of allelic lineages at *DRB1*, as well as analyses of neutral genes[32], the data suggest a scenario in which population size in the lineage leading to modern humans has oscillated over time, although the details of these oscillations, particularly in the past million years, will no doubt be refined as the molecular diversity for noncoding regions of MHC alleles is further investigated and compared with that of coding regions.

Box 3. Measuring evolutionary rates at MHC loci

Difficulties in measuring long-term rates of substitution arise in MHC genes because of the phylogenetic discordances with species trees created by trans-species evolution (see text). Trans-species evolution renders traditional methods of inferring relative or absolute rates of molecular evolution useless when applied to MHC variation because we cannot use species trees as a proxy for the gene trees we wish to study. Satta *et al.*[52] helped overcome these problems by noting that, even in the presence of trans-species evolution, the closest pair of alleles found in two different species are likely to have diverged close to the time of the species split. This allows traditional rate tests to be used. The authors estimated rates of synonymous (silent) substitution in primate MHC genes at about 1×10^{-9} substitutions per site per year, a surprisingly low rate given the large genetic distances between primate alleles. The long persistence time of alleles is one way to reconcile such large divergences at silent sites with a low rate of silent substitution.

Edwards *et al.*[53] used Satta *et al.*'s method and data on times of species divergence to show that silent rates in class II genes of rodents (a) were as much as four times higher than those measured in primates (b). The open symbols in each graph indicate a set of divergence times that is ambiguous. Solid and dotted lines indicate the rates estimated depending on whether the traditional or novel divergence times, respectively, are accepted for these groups. Relative rate tests, which do not depend on absolute divergence times, also suggested increased silent rates of substitution in rodents, implying that mammalian MHC genes exhibit the same types of generation-time effect as do neutrally evolving genes. Despite this consistency, estimating rates of silent substitution in this way requires assumptions of no recombination or gene conversion, the latter of which is thought to elevate synonymous substitution rates[21]. *Reproduced, with permission, from Ref. 53.*



Sources of balancing selection

The types of selection envisioned for MHC genes have traditionally included heterozygote or rare-allele advantage. Although some types of frequency-dependent selection can be ruled out[2], both mechanisms are compatible with patterns of MHC polymorphism and various selective scenarios involving disease resistance, mate choice and/or reproductive incompatibilities between individuals. (In humans, the most easily identified source of selection on MHC loci is that associated with the many autoimmune diseases, but most of these diseases are uncommon and they tend to lower fitness later in life, making them unlikely players in maintaining MHC variation.) Recent surveys still reveal a significant excess of couples with a history of spontaneous abortions sharing antigens for one or more *HLA* loci[33]. The immunological explanation for these observations, which would only apply to mammalian MHC diversity, is that the presence of an immune response occurring when the mother and fetus differ at *HLA* loci is necessary for proper implantation and fetal growth, a type of selection that could maintain many alleles at a single locus.

Following on the experiments of Yamazaki, Potts and coworkers implicating disassortative mating as a mechanism to elevate MHC diversity are further experiments in mice demonstrating an effect of early experience on imprinting and mate choice[34]. However, nonrandom mating with respect to MHC type was not detected in sheep[35], and the data on human populations are contradictory, with MHC-based mate preferences detectable in some populations[36] but not others[37]. Studies indicating that female university students preferred the odors of MHC-dissimilar males in some situations[38] (but see Refs 37 and 39) and that body odor preferences in men and women are influenced by the degree to which they share MHC alleles[40] suggest an olfactory mechanism for mate choice similar to that known for mice. Singer *et al.*[41] recently showed via gas chromatography that levels of volatile acids in mouse urine, in particular phenylacetic acid, differ between MHC types and that such quantitative differences can be discriminated by mice. Although it has still to be determined whether the evolution of chemosensory mate choice in mammals is a cause or a consequence of ancient MHC polymorphisms, the presence in urine of compounds such as phenylacetic acid, which are naturally conjugated with amino acids in mouse urine and which therefore could be bound by MHC molecules, raises intriguing possibilities as to mechanisms of chemo-discrimination.

The most traditional explanation for MHC diversity, namely heterozygote advantage (because MHC heterozygotes present a wider range of foreign peptides to T-cells), is still the most popular, although the number of examples of predicted associations between MHC genotypes and resistance to infectious disease is still small[3,4]. It is still underappreciated that some of the best examples of strong associations between MHC loci and parasite resistance or between MHC loci and expression of sexually selected traits come from birds[42,43]. For example, 95% of chickens with certain haplotypes survive infection by Marek's disease (a type of viral leukemia), whereas individuals with other haplotypes may have nearly 0% survival. MHC-disease associations in chickens might be stronger than in mammals because chicken MHC is compact, possibly intensifying the strength of selection on fewer target genes[42]. Whereas these and other[43] examples suggest selection reduces diversity because of resistance conferred by a specific haplotype, a much-needed example of selection that enhances diversity, in the form of heterozygote advantage, has recently emerged from analyses of *HLA* loci and resistance to hepatitis B virus among West Africans[4]. Consistent relationships between MHC variants and resistance could be difficult to detect even in control populations for a variety of reasons, including improper statistical design, low resolution of serotypes or RFLP markers versus sequences, geographically variable MHC adaptation owing to geographical differentiation of the parasite, adaptation to parasites no longer present, influence of other linked or epistatic loci, overlap in the peptide repertoires of individual MHC alleles, or importance to parasite resistance of loci outside the MHC[44]. Because some of the important infectious diseases of the past are epidemic, selection models for maintaining MHC variation that incorporate differential selection over time and space may be most appropriate.

Box 4. Estimating selective differences between MHC heterozygotes and homozygotes

Several approaches have been used to estimate *s*, the selective advantage of heterozygotes over homozygotes. Three of these directly estimate the amount of selection occurring in the present generation:

- From population data, *s* can be estimated from the difference between the observed homozygosity (ΣP_{ii}) and that expected under random mating, that is, Hardy–Weinberg proportions (Σp_i^2) (Ref. 46) as:

$$s = \frac{\sum p_i^2 - \sum P_{ii}}{\sum p_i^2 (1 - \sum P_{ii})}$$

where P_{ii} is the observed frequency of the homozygote for the *i*th allele, and p_i is the frequency of the *i*th allele.

- If family groups are known in which there are equal expected Mendelian proportions of heterozygous and homozygous progeny but the actual numbers are different, then selection can be estimated as:

$$s = 1 \frac{N_{ii}}{N_{i\bar{i}} + 1}$$

where N_{ii} and $N_{i\bar{i}}$ are the observed numbers of homozygous and heterozygous progeny[47]. The same approach can be used to determine selective differences in experiments with equal initial numbers and different subsequent mating preferences or mating numbers.

- The selective difference for a particular genotype or allele relative to other genotypes or alleles because of a particular parasite (or pathogen) in a population can be estimated as:

$$s = m(1 - RR)$$

where *m* is the proportion of mortality caused by the parasite and *RR* is the relative risk of the infection for the genotype[3].

Alternatively, an indirect measure of cumulative selection over the history of the allelic lineages can be estimated using nucleotide sequences as:

$$s = \frac{K_B^2 \gamma}{2\sqrt{2} N_e}$$

where γ is the ratio of the number of nonsynonymous substitutions to the number of synonymous substitutions per site, K_B is the mean number of nonsynonymous substitutions in the peptide-binding region, and N_e is the long term effective population size in the human lineage[46]. The number of interspecific coalescent events counted on trees of MHC alleles evolving in a trans-species manner seems to be influenced by the selection intensity, although no explicit methods to estimate *s* have been derived using this information[53].

Getting to '*s*'

There are various ways in which the intensity of selection has been estimated at MHC genes. In general, the selective difference (*s*) between heterozygotes and homozygotes is estimated (Box 4). Several approaches measure the combined selection intensity over all or several fitness components. For example, in some aboriginal populations (which have a limited number of alleles, are fairly isolated, and have little or no sequence variation within serotype) large deviations from Hardy–Weinberg expectations, giving *s* values of approximately 0.3 in the current generation, have been observed[44]. The level of selection derived from Hardy–Weinberg

deviations in many nonaboriginal populations can be hard to estimate either because nearly all individuals are heterozygotes for the variable *HLA* loci or because past selection pressures are no longer evident owing to better health care and sanitation. Deficiencies of MHC homozygotes have also been reported in some[43], but not all[45], natural populations investigated so far. Using another indirect method based on the unusual pattern of variation in MHC genes, Satta *et al.*[46] estimated *s* in humans to be 0.013, with the highest value of 0.042 for the *HLA-B* gene.

The extent of selection for viability, mate choice or parasite resistance can in principle be examined separately. The extent of viability selection can be determined by examining deviations from expected Mendelian segregation in families, and selection based on mate choice by determining the deviation of progeny-genotypes from random-mating proportions or by preference experiments. Recent estimates of *s* using these approaches vary from 0.254 for viability, for a large number of Amerindian progeny[47], to around 0.6 for mate choice, in laboratory mice[48] and humans in general[36]. Hill *et al.*[3] estimated *s* for particular *HLA-B* and *DRB1* alleles as 0.029 and 0.038, respectively (based on the relative risk of genotypes contracting malaria and the mortality rate from malaria), and an estimate of *s* for hepatitis B for *HLA-DR-DQ* haplotypes as 0.080[4]. The variation among estimates of *s* derived from such diverse approaches can reflect differences in *s* over time, space, species or experimental context.

The future

The amount and sophistication of theory and technology being thrown at the MHCs of diverse organisms are staggering. The information flowing from ongoing large-scale sequencing efforts will prove a goldmine for molecular evolutionary analysis. However, more extensive molecular characterizations of MHCs in nonmammalian vertebrates will be needed to infer the directions and timing of major events in MHC evolution. More realistic models incorporating mutation, microrecombination and potentially high levels of convergence in estimating population genetic parameters for MHC alleles, or further study of the conditions under which these problems are inconsequential, are needed. Finally, more experimental investigations of mate choice and infection by disease, and estimates of their selective effects in natural populations, if detectable[46], would help determine why MHC genes are so variable – for many researchers the 'holy grail' of MHC biology[49].

Acknowledgements

We thank D. Geraughty, H. Erlich, M. Flajnik, M. Kaufman, M. Kasahara, C. Ober and Y. Satta for helpful discussion and sharing of unpublished manuscripts. J. Kingsolver, M. Nei, T. Ohta, P. Parham, Y. Satta, J. Trowsdale and three anonymous reviewers provided helpful suggestions on the manuscript. S.V.E was supported, in part, by grants from the National Science Foundation.

References

1 Potts, W.K. and Wakeland, E.K. (1990) **Evolution of diversity of the major histocompatibility complex,** *Trends Ecol. Evol.* 5, 181–187
2 Takahata, N. and Nei, M. (1990) **Overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci,** *Genetics* 124, 967–978
3 Hill, A.V.S. *et al.* (1991) **Common West African HLA antigens are associated with protection from severe malaria,** *Nature* 352, 595–600
4 Thursz, M.R. *et al.* (1997) **Heterozygote advantage for HLA class-II type in hepatitis B virus infection,** *Nature Genet.* 17, 11–12
5 Beck, S. *et al.* (1996) **Evolutionary dynamics of non-coding sequences within the class II region of the human MHC,** *J. Mol. Biol.* 255, 1–13
6 Mizuki, N. *et al.* (1997) **Nucleotide sequence analysis of the HLA class I region spanning the 237-kb segment around the HLA-B and -C genes,** *Genomics* 42, 55–66
7 Guillaudeux, T. *et al.* **The complete genomic sequence of 424,015 bp at the centromeric end of the HLA class I region: gene content and polymorphism,** *Proc. Natl. Acad. Sci. U. S. A.* (in press)
8 Miller, M.M. *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,** *Proc. Natl. Acad. Sci. U. S. A.* 93, 3958–3962
9 Flajnik, M.F. *et al.* (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, 4385–4396
10 Bingulac-Popovic, J. *et al.* (1997) **Mapping of MHC class I and class II regions to different linkage groups in the zebrafish, *Danio rerio*,** *Immunogenetics* 46, 129–134
11 Kasahara, M. *et al.* (1996) **Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex,** *Proc. Natl. Acad. Sci. U. S. A.* 93, 9096–9101
12 Kasahara, M. (1997) **New insights into the genomic organization and origin of the major histocompatibility complex: role of chromosomal (genome) duplication in the emergence of the adaptive immune system,** *Hereditas* 127, 59–65
13 Hughes, A.L. (1997) **Evolution of the proteasome components,** *Immunogenetics* 46, 82–92
14 Edwards, S.V., Wakeland, E.K. and Potts, W.K. (1995) **Contrasting histories of avian and mammalian *Mhc* genes revealed by class II B sequences from songbirds,** *Proc. Natl. Acad. Sci. U. S. A.* 92, 12200–12204
15 Nei, M., Gu, X. and Sitnikova, T. (1997) **Evolution by the birth-and-death process in multigene families of the vertebrate immune system,** *Proc. Natl. Acad. Sci. U. S. A.* 94, 7799–7806
16 Trtkova, K. *et al.* (1995) ***Mhc-DRB* genes and the origin of New World Monkeys,** *Mol. Phylog. Evol.* 4, 408–419
17 Klein, J. *et al.* (1994) **The accordion model of *Mhc* evolution,** in *Progress in Immunology* (Vol. VIII) (Gergely, J. and Petraryi, G., eds), pp. 137–143, Springer-Verlag
18 Figueroa, F. *et al.* (1995) **Evidence for insertion of a new intron into an *Mhc* gene of perch-like fish,** *Proc. R. Soc. London Ser. B* 259, 325–330
19 Meager, S. and Potts, W.K. (1997) **A microsatellite-based MHC genotyping system for house mice (*Mus domesticus*),** *Hereditas* 127, 75–82
20 Hughes, A.L. and Nei, M. (1988) **Pattern of nucleotide substitution at major histocompatiblity complex class I loci reveals overdominant selection,** *Nature* 335, 167–170
21 Ohta, T. **On the pattern of polymorphism at major histocompatibility complex loci,** *J. Mol. Evol.* (in press)
22 Bergström, T.F. *et al.* (1998) **Recent origin of *HLA-DRB1* alleles and implications for human evolution,** *Nature Genet.* 18, 237–242
23 Parham, P. and Ohta, T. (1996) **Population biology of antigen presentation by MHC class I molecules,** *Science* 272, 67–74
24 Zangenberg, G. *et al.* (1995) **New *HLA-DPB1* alleles generated by interallelic gene conversion detected by analysis of sperm,** *Nature Genet.* 10, 407–414
25 Hogstrand, K. and Bohme, J. (1994) **A determination of the frequency of gene conversion in unmanipulated mouse sperm,** *Proc. Natl. Acad. Sci. U. S. A.* 91, 9921–9925
26 Satta, Y. (1997) **Effects of intra-locus recombination on *HLA* polymorphism,** *Hereditas* 127, 105–112
27 Yeager, M., Kumar, S. and Hughes, A.L. (1997) **Sequence convergence in the peptide binding region of primate and rodent MHC class Ib molecules,** *Mol. Biol. Evol.* 14, 1035–1041
28 Graser, R. *et al.* (1996) **Trans-species polymorphism of class II *Mhc* loci in danio fishes,** *Immunogenetics* 44, 36–48
29 Klein, J. *et al.* (1993) **The molecular descent of the major histocompatibility complex,** *Annu. Rev. Immunol.* 11, 269–295
30 Klein, D. *et al.* (1993) **Extensive *Mhc* variability in cichlid fishes of Lake Malawi,** *Nature* 364, 330–334
31 Vincek, V. *et al.* (1997) **How large was the founding population of Darwin's finches?** *Proc. R. Soc. London Ser. B* 264, 111–118
32 Takahata, N. and Satta, Y. (1997) **Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences,** *Proc. Natl. Acad. Sci. U. S. A.* 94, 4811–4815
33 Ober, C. *et al.* **HLA matching and fetal loss: results of a 10-year**

prospective study, *Hum. Reprod.* (in press)
34 Eklund, A. (1997) **The effect of early experience on MHC-based mate preferences in two B10.W strains of mice (*Mus domesticus*),** *Behav. Genet.* 27, 223–229
35 Paterson, S. and Pemberton, J.M. (1998) **No evidence for major histocompatibility complex-dependent mating patterns in a free-living ruminant population,** *Proc. R. Soc. London Ser. B* 264, 1813–1819
36 Ober, C. *et al.* (1997) **HLA and mate choice in humans,** *Am. J. Hum. Genet.* 61, 497–504
37 Hedrick, P.W. and Black, F.L. (1997) **HLA and mate selection: no evidence in South Amerindians,** *Am. J. Hum. Genet.* 61, 505–511
38 Wedekind, C. *et al.* (1995) **MHC-dependent mate preferences in humans,** *Proc. R. Soc. London Ser. B* 260, 245–249
39 Hedrick, P.W. and Loeschcke, V. (1996) **MHC and mate selection in humans?** *Trends Ecol. Evol.* 11, 24
40 Wedekind, C. and Füri, S. (1997) **Body odour preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity?** *Proc. R. Soc. London Ser. B* 264, 1471–1479
41 Singer, A.G., Beauchamp, G.K. and Yamazaki, K. (1997) **Volatile signals of the major histocompatibility complex in male mouse urine,** *Proc. Natl. Acad. Sci. U. S. A.* 94, 2210–2214
42 Kaufman, J. and Wallny, H.J. (1996) **Chicken MHC molecules, disease resistance and the evolutionary origin of birds,** *Curr. Top. Microbiol. Immunol.* 212, 129–141
43 von Schantz, T. *et al.* (1996) **MHC genotype and male ornamentation: genetic evidence for the Hamilton–Zuk model,** *Proc. R. Soc. London Ser. B* 263, 265–271
44 Hedrick, P.W. and Kim, T.J. **Genetics of complex polymorphisms: parasites and maintenance of MHC variation,** in *Genetics, Evolution and Society* (Singh, R.S. and Krimbas, C.K., eds), Harvard University Press (in press)
45 Boyce, W.M. *et al.* (1997) **Genetic variation of major histocompatibility complex and microsatellite loci: a comparison in bighorn sheep,** *Genetics* 145, 421–433
46 Satta, Y. *et al.* (1994) **Intensity of natural selection at the major histocompatibility complex loci,** *Proc. Natl. Acad. Sci. U. S. A.* 91, 7184–7188
47 Black, F.L. and Hedrick, P.W. (1997) **Strong balancing selection at HLA loci: evidence from segregation in South American families,** *Proc. Natl. Acad. Sci. U. S. A.* 94, 12452–12456
48 Jin, K., Speed, T.P. and Thomson, G. (1995) **Tests of random mating for a highly polymorphic locus: application to HLA data,** *Biometrics* 51, 1064–1076
49 Potts, W.P. and Slev, P.R. (1995) **Pathogen-based models favoring MHC genetic diversity,** *Immunol. Rev.* 143, 181–197
50 Trowsdale, J. (1995) **'Both bird and man and beast': comparative organization of MHC genes,** *Immunogenetics* 41, 1–17
51 Richman, A.D. and Kohn, J.R. (1996) **Learning from rejection: the evolutionary biology of single-locus incompatibility,** *Trends Ecol. Evol.* 11, 497–502
52 Satta, Y. *et al.* (1993) **The synonymous substitution rate at major histocompatibility complex loci in primates,** *Proc. Natl. Acad. Sci. U. S. A.* 90, 7480–7484
53 Edwards, S.V. *et al.* (1997) **Ancestral polymorphism of MHC class II genes in mice: implications for balancing selection and the mammalian molecular clock,** *Genetics* 146, 655–668

The future as the key to the past for palaeobotany?

D.J. Beerling

Continued increase in the concentration of atmospheric CO_2 and its possible effects on global climate has generated intense research interest on the likely responses of terrestrial plants and vegetation. Results from this new research provide quantitative information on plant function and growth in an environment with a high CO_2 concentration, but are also relevant to understanding plant growth in the distant past and to the techniques employed by palaeobotanists for reconstructing past climates from fossil plant remains. Experimental CO_2 enrichment of plants has demonstrated direct effects on leaf physiognomy, the tolerance of plants to low temperature and the relationship between tree rings, CO_2 and climate; it therefore signals the need for caution in interpreting palaeoclimates from fossils.

D.J. Beerling is at the Dept of Animal and Plant Sciences, University of Sheffield, Sheffield, UK S10 2TN (d.j.beerling@sheffield.ac.uk).

Two factors have provided a strong impetus for international experimental and modelling programmes on plant responses to global environmental change[1] (GEC): the threat of future increases in the concentration of atmospheric CO_2 and associated climatic change; and the central role of terrestrial plants and vegetation in the global carbon cycle. Revised predictions by the Intergovernmental Panel on Climate Change[2] indicate an atmospheric CO_2 concentration by the year AD 2100 of between 450 ppm and >800 ppm. These upper and lower limits are based on assumptions of economic growth, population growth and fossil fuel supply that are either low or moderate, respectively. The upper limit is not new to terrestrial vegetation and ecosystems, which have evolved and survived under such CO_2 concentrations (or higher) for at least the past 250 million years, with the exception of the short-lived fluctuations in the Quaternary period[3,4]. Consequently, results from CO_2-enrichment experiments should be of interest to palaeobotanists: they provide information relevant to understanding plant function during episodes of high CO_2 concentration in the geological past (e.g. the 'greenhouse world' of the mid-Cretaceous and mid-Eocene); and they are beginning to provide new insights into the effects of CO_2 on the relationship between plants and climate.

Past climates from plant fossils

Three aspects of the relationship between plants and their environment are used by palaeobotanists to give a signal of terrestrial climate from the fossils: leaf physiognomic characters[5]; the present climatic association of the 'nearest living relatives' (NLR) to the fossil[6,7] (a concept including the use of climatic response surfaces developed from Quaternary palynological studies[8]); and the character of tree rings in fossilized woods[9]. Each aspect has largely been developed with the specific aim of extracting information about the climate from the fossils. A neglected element, but one that is difficult to include in these studies, is an adequate consideration

0169-5347/98/$19.00 PII: S0169-5347(98)01334-2