PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

MHC Polymorphism and Parasites [and Discussion]

Jan Klein, Colm O'Huigin, C. A. Mims, J. Lines, J. Deutsch and A. L. Hughes

Phil. Trans. R. Soc. Lond. B 1994 **346**, 351-358
doi: 10.1098/rstb.1994.0152

MHC polymorphism and parasites

JAN KLEIN[1,2] AND COLM O'HUIGIN[1]

[1] *Max-Planck-Institut für Biologie, Abteilung Immungenetik, Corrensstr. 42, D-72076 Tübingen, Germany*
[2] *Department of Microbiology and Immunology, University of Miami School of Medicine, Miami, Florida 33136, U.S.A.*

SUMMARY

The major histocompatibility complex (MHC) polymorphism is marked by the existence of allelic lineages that are extremely old, having been passed from one species to another in an evolutionary line of descent. Each species has several of these lineages and many of their more recent derivatives, the actual alleles. The lineages are separated by large genetic distances and are characterized by the presence of short sequence motifs which, at the protein level, have remained virtually unaltered for over 40 million years. Several explanations for the MHC polymorphism have been proposed. We argue that the only one consistent with the entire body of knowledge about the MHC is an explanation based on the immune response to parasites. Furthermore, we propose that parasites coevolving with their hosts have had a major influence on MHC polymorphism, whereas parasites that switched hosts recently and became very virulent have had little effect. The latter category includes micro- and macroparasites responsible for the major human infectious diseases. This hypothesis explains why no convincing association between human leucocyte antigen (*HLA*) alleles and resistance to infectious disease can thus far be documented, and indicates the direction in which the search for such associations should be taken.

1. SELF–NONSELF DISCRIMINATION

Vertebrates, for reasons that are not at all obvious, perhaps because of the growing size and complexity of their bodies (Klein 1989), embarked on an evolutionary pathway leading to the emergence of an entirely new defence system aimed at protecting them against both micro- and macroparasites. The essential feature of this system is that it anticipates every possible molecule the parasites might come up with. A fundamental predicament in designing it is how to distinguish non-self from self. All living creatures, parasites and their hosts included, are constructed from the same building blocks assembled in a wide variety of permutations. Limiting our considerations to proteins alone, we note that all organisms use the same 20 or so amino acids to construct unique polypeptides, which differ in sequence between parasites and their hosts. Unfortunately, it is apparently impossible to design a receptor capable of recognizing a protein molecule as a whole and thus differentiating it from all other protein molecules. Recognition is always limited to a small portion of the ligand: a peptide, a three-dimensional motif, or a combination of both. And here are the horns of the dilemma: because there is a considerable overlap between the peptides of the parasite and those of its host, the host must find a way of ignoring the overlapping peptides (otherwise the defence mechanisms would turn against the host itself) and focusing only on those unique to the parasite.

Natural selection has found a solution to this dilemma so simple and so ingenious that even the smartest speculative biologists were unable to figure it out before experimentalists discovered how it works. It requires three molecules acting in concert: the major histocompatibility complex (MHC), the T-cell receptor (TCR), and the immunoglobulin (Ig) molecules. The MHC molecule acts first. It focuses exclusively on proteins and ignores carbohydrates, lipids, nucleic acids, and other organic substances. Furthermore, it deals best with protein fragments, peptides, although it can also handle entire protein molecules. It is a receptor for peptides, but a receptor of very limited specificity. It exists in two variants, class I and class II (Klein 1986), each of which has somewhat different restrictions on peptide binding (Rammensee *et al.* 1993; Germain 1994). The peptide-binding groove of the class I molecule is closed so that only peptides of a certain length (8–10 amino acid residues) fit into it and are bound by their termini. It has pockets into which side chains of the peptides, anchor residues, must fit for optimal binding. The anchors and the fixed distances between them determine most of the specificity of the peptide binding which thus remains very broad, allowing thousands of different peptides to be accommodated by a particular MHC molecule. The groove of the class II molecule, on the other hand, is open and accommodates peptides that extend beyond its edges (length range of 12–24 residues); the peptides are not bound by their termini. The specificity of the binding is achieved

Phil. Trans. R. Soc. Lond. B (1994) **346**, 351–358
Printed in Great Britain

through the interaction between stretches of amino acid residues in the peptide and in the groove, whereby the side chains of the former must again fit into the pockets of the latter.

Peptide binding occurs when the dimeric MHC molecules are assembled from their constituent chains and the complex is then displayed on the surface of the 'antigen presenting cells' (Neefjes & Momburg 1993). The MHC molecules bind self and non-self peptides indiscriminately, most of the time the former. The discrimination is then accomplished by a neat trick during the maturation of the immune system. The peptide–MHC protein combination is recognized by the TCR and this recognition is highly specific. For each MHC-peptide complex combination, there is a T lymphocyte clone that recognizes this and virtually no other combination. Initially, there are T lymphocyte clones in the developing immune system for *any* combination of peptide and MHC protein, both self and non-self. At a later stage, when there are still very few non-self peptides in the body, all lymphocytes reactive with self peptide–MHC protein complexes are purged from the immune system, which is then left with cells capable of recognizing non-self peptide–MHC protein complexes only (Miller 1992). When such complexes occur in the body during a parasitic infection, the responding clone is stimulated and either becomes involved in the attack on the invader itself or releases mediators that stimulate B lymphocytes to secrete Ig molecules, the 'antibodies' of the classical immune response.

The amino acid residues of the groove that come into contact with the peptide, the peptide-binding region or PBR, are highly variable and the corresponding part of the specifying gene is highly polymorphic (Hughes & Nei 1989). We will illustrate the features of this polymorphism by taking one of the primate class II loci, the *DRB1* locus, as an example. The number of alleles described at the *DRB1* locus in the human species alone is edging toward 100 (Marsh & Bodmer 1993). Several hundred more have been reported for a variety of other primate species (O'hUigin *et al.* 1993), none of which, however, has been studied as thoroughly as *Homo sapiens*.

2. ALLELIC LINEAGES AND MOTIFS

A conspicuous characteristic of *DRB1* polymorphism is that the alleles fall into groups that we shall refer to as *allelic lineages*. Each lineage is distinguished by one or more diagnostic *motifs*: short stretches of nucleotide sequences with substitutions peculiar to that lineage. The motifs are localized in (non-contiguous) segments of the sequence specifying the PBR. Because the substitutions are primarily non-synonymous, the motifs can be described most succinctly when translated into amino acid sequences. Hence, when we speak of 'allelic lineages' but use amino acid notation for their motifs, it is to avoid long-winded phrases such as 'allelic lineage with a motif which in amino acid translation reads ...'. The amino acids will be given in the international one-letter code and their position will be indicated in the mature protein sequence. To give an example, the main motif shared by all the *HLA–DRB1*03* alleles (where *HLA* stands for 'human leukocyte antigen, the human MHC, and the numbers following the asterisk are allelic designations) is 9-EYST-12. At the same position, the *HLA–DRB1*04* alleles have the sequence 9-EQVK-12, the *HLA–DRB1*02* alleles 9-WQPK-12, and so on. Other motifs occur at positions 26–33, 37 and 38, and 67–74, which are also the most polymorphic segments of the *DRB1* genes. The alleles within an allelic lineage differ from one another by solitary substitutions scattered throughout the rest of the sequence.

The most remarkable observation is that the motifs, although allelic lineage-specific, occur as polymorphisms in a variety of other primate species; some even occur in non-primate eutherian mammals. For example, the lineage with the EYST motif has been found not only in humans, but in common chimpanzee (*Patr*), pygmy chimpanzee (*Papa*), gorilla (*Gogo*), orangutan (*Popy*), rhesus macaque (*Mamu*), pigtail macaque (*Mane*), Japanese macaque (*Mafu*), drill (*Male*), hamadryas baboon (*Paha*), common marmoset (*Caja*), cotton-top tamarin (*Saoe*), dusky titi (*Camo*), and northern lesser bushbaby (*Gase*): hence in apes (*Patr*, *Papa*, *Gogo*, *Popy*), Old World monkeys (*Mamu*, *Mane*, *Mafu*, *Male*, *Paha*), New World monkeys (*Caja*, *Saoe*, *Camo*), and prosimians (*Gase*) (see O'hUigin *et al.* 1993). (The four-letter italicized symbols are abbreviations of the scientific genus and species names; for example, *Patr* stands for *Pan troglodytes*.) Among the non-primates, its occurrence has been documented in sheep (*Ovar*), anoa (Asian buffalo, *Bude*), cattle (*Bota*), banteng (*Boja*), goat (*Caae*) (see Schwaiger *et al.* 1993), and mouse (*Mumu*) (see Lundberg & McDevitt 1992).

It is, however, not just the motif that is shared by the different primate species, but the entire allelic lineage. The interspecific sharing of allelic lineages is indicated by the overall sequence similarity of corresponding genes in the different species, which is reflected in genetic distances calculated from the similarities and in dendrograms based on the genetic distances (but also on parsimony considerations). In the dendrograms, genes of a given lineage but from different species cluster together and separate from clusters (clades) formed by other allelic lineages. The MHC polymorphism, therefore, has a trans-species character (figure 1) (Klein 1987). Each species possesses between five and seven major allelic lineages, all of which are shared with at least one other species. The typing has thus far not been exhaustive enough to reveal whether some species have lost any of the major lineages, although this would not be surprising, particularly in respect of those on the endangered species list, whose populations have been drastically reduced in recent times. It is clear, however, that all the major allelic lineages have passed through many speciation phases. Several of the *DRB1* allelic lineages may be more than 40 million years (Ma) old.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

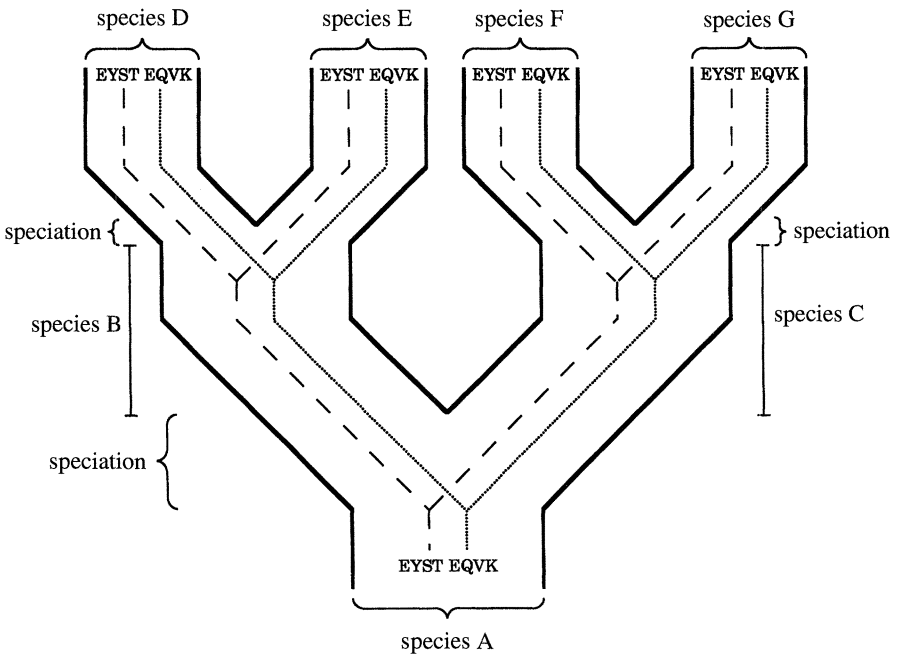

Figure 1. Passage of MHC allelic lineages through speciation phases. Three speciation events are shown. The MHC lineages (broken and dotted lines) carry characteristic motifs (amino acid residues abbreviated in the international single-letter code) that are retained through the entire evolutionary period. Not shown are the individual alleles that arise in each species from each lineage by mutations.

3. EVOLUTION BY INCORPORATION

The trans-species character of the allelic lineages is established by a special mode of the evolutionary process (Klein *et al.* 1991). Instead of the standard fixation of mutations in diverging species, the MHC allelic lineages evolve by a process that we refer to as *incorporation* (fixation within a lineage). In this process, mutations spread, one after another, through only a portion of the population and thus remain at frequencies of less than 1 over many millions of years. Hence an allele that has acquired one mutation is in a position to acquire a second before the first achieves fixation. Another group of individuals of the same species acquires a different mutation, and then a second, again without the first becoming fixed, and so on. Thus, mutation after mutation is incorporated into the lineages, which thus gradually drift apart in their sequences. It takes on average 1.3 Ma to incorporate a non-synonymous substitution into the PBR of a primate class I gene and approximately 4.5 Ma into the PBR of a class II gene. The times for the incorporation of synonymous substitutions are 3.2 Ma and 5.0 Ma, respectively (Klein *et al.* 1993). Because the average lifespan of a species is 2 Ma (Stanley 1975) and *HLA* alleles can differ by up to 55 substitutions (39 non-synonymous, 16 synonymous in *HLA–DRB1*0301* and *HLA–DRB1*0701* alleles), it is obvious that the lineages must evolve trans-specifically.

It is also obvious that the time during which mutations remain as polymorphisms in a succession of populations (species) exceeds the expectations based on the coalescence theory of neutral alleles. The mean coalescence time of two neutral alleles is $2N_e$ generations, where N_e is the effective population size (Hartl & Clark 1989). The estimated N_e in the hominoid lineage in the last 20 my is 10^5 individuals (Takahata *et al.* 1994), giving an expected mean coalescence time of neutral alleles of 2×10^5 generations. Taking a generation time of 20 years for the hominoids, we obtain a mean coalescence time of two neutral alleles of 4 Ma. In reality, the mean coalescence time of MHC alleles is at least ten times higher. The obvious conclusion is that MHC genes are not neutral and that their long persistence can be explained only by invoking a form of balancing selection (Takahata 1990). The occurrence of balancing selection in the MHC genes is also indicated by the observation that the per site frequency of non-synonymous substitutions in the PBR is several times higher than that of synonymous substitutions (Hughes & Nei 1989).

4. MHC AND PARASITES

The only known function of the classical MHC molecules is to present peptides to T lymphocytes of the immune system. It therefore seems logical to assume that the agents behind the balancing selection are the parasites, from which the immune system, with the MHC molecules at the forefront, protects the vertebrates. Involvement in the immune response is also the only postulated function of the MHC which all the vertebrates share. Neither odour-based mate (kin) selection (Potts & Wakeland 1994), nor maternal–foetal incompatibility (Gill 1994), nor any of the other selection pressures proposed for the maintenance of MHC polymorphism have the ubiquitous distribution

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

among the vertebrate taxa that the parasite-targeted immune response does.

Assuming that the generation, not only of allelic lineages with their lineage-specific motifs, but also of MHC polymorphism in general, is indeed driven and maintained by parasites, several difficult questions arise. Foremost among them is why the MHC polymorphism evolves in a trans-species manner. Undoubtedly, the answer to this question will emerge from the study of host–parasite interactions. We divide parasites in this regard into three categories, limiting our discussion to human parasites only. In the first category are parasites that have become associated with humans only in the past 10 000 years (Cameron 1956; Johnson 1986; Nelson 1988). They are often highly virulent and have been responsible for past and present epidemics and pandemics. Because they have been the cause of high mortality in human populations, they are widely believed to have exerted strong selection pressure on the evolution of MHC polymorphism. Examples include *Plasmodium falciparum*, *Trypanosoma gambiense*, *Leishmania donovani*, *Mycobacterium tuberculosis*, *Yersinia pestis*, smallpox virus, rubella virus, and influenza virus. All these organisms are believed to have become human parasites after the advent of agriculture and domestication of animals, which led to an increasing population density and aggregation of humans in large communities. For several of them, there is indeed good experimental evidence for their recent association with *Homo sapiens*. For example, sequencing of the small-subunit ribosomal RNA gene from a variety of *Plasmodium* species has revealed that *P. falciparum*, the causative agent of falciparan malaria, is so closely related to avian *Plasmodium* species that a lateral transfer of the parasite from birds to humans as recently as the start of the agricultural revolution is the most plausible explanation of its origin (Waters *et al.* 1993). In a few cases, an explanation for increased virulence of the parasite at the time of the transfer is also available. Thus, in *Yersinia pestis* (the causative agent of the bubonic plague) two mutations, one at a chromosomal locus and the other in a plasmid, may have produced hypervirulent strains of these bacteria which, when transferred from rats and fleas to humans, caused the plague epidemics (Rosqvist *et al.* 1988).

5. PARASITES AND MHC POLYMORPHISM

We argue that these recent parasites have had little effect on MHC polymorphism. We base our argument on two observations. The first is that the great majority of non-synonymous PBR substitutions differentiating *HLA* alleles and MHC lineages are much older than 10 000 years and therefore could not have been incorporated in response to this category of parasites. If a non-synonymous PBR substitution is incorporated into the human population once every 1.3–5 Ma, obviously most alleles that differ by a single substitution, and almost certainly all the alleles that differ by two or more substitutions and were not generated by complex mutations or by recombination, must have been established long before the beginning of the Neolithic revolution. Their persistence in the hominoids therefore could not have been caused by parasites that invaded the human population less than 10 000 years ago. The second observation is that the high virulence of these parasites often paralyses the host's immune system, which then provides little protection against the invaders. Often the parasite kills the host quickly, before the latter can mobilize its anticipatory immune system. There is thus very little selection on any component of the immune system, including the MHC genes.

Our exclusion of the first category of parasites from our search for the driving force of MHC polymorphism may seem to be contradicted by the studies on falciparan malaria. Several such studies have been carried out, but the largest and the most publicized of them was by Hill and his colleagues (1991) on Gambian black populations. The main conclusions of this study were that children under ten years of age and carrying the *HLA-B53* allele have a 40% lower chance of succumbing to either severe malarial anaemia or cerebral malaria than those lacking this allele. Similarly, children carrying the *DRB1*1302 – DQB1*0501* haplotype have a 50% lower chance of dying from malarial anaemia than children with other haplotypes. The *DRB1*1302 – DQB1*0501* haplotype, however, offers no protection against cerebral malaria and neither it, nor the *HLA-B53* allele, provides any protection from the initial infection by *P. falciparum*. The study, even if confirmed by extension to other populations, cannot be regarded as evidence that *P. falciparum* is one of the agents driving MHC polymorphism, nor that it is responsible for the divergence of allelic lineages. Hill *et al.* (1991) suggest that, during the past 5000 years, malaria has driven the frequencies of *HLA-B53* among African blacks from the base level of 1% up to 22% in some populations (Allsopp *et al.* 1992). If so, the increase would have been accompanied by a corresponding decrease in frequencies of other *HLA* alleles and hence by an overall loss rather than gain of MHC polymorphism, especially as *HLA-B53* heterozygotes are no more resistant to *P. falciparum* than *HLA-B53* homozygotes. In fact, it has been suggested (Carter *et al.* 1992) that the effect Hill and his colleagues observed was a manifestation of decreased rather than enhanced immune response triggered by *P. falciparum* infection in the presence of *HLA-B53* compared with response in the presence of other *HLA-B* alleles, and that the heightened reactivity resulted in autoaggression.

Into the second category we place parasites whose evolution appears to follow Fahrenholz's rule that common ancestors of present-day parasites were themselves parasites of the common ancestors of present-day hosts (Fahrenholz 1913). Because they have speciated in parallel with their hosts, the immune system, including the MHC molecules, must have coevolved with them. The parasites could therefore have been the source of selection pressure for periods that can be measured in millions of years. An example of this category are the papova viruses. DNA sequence comparisons of papova viruses BKV,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

SV40 and polyoma virus, derived from humans, monkeys and mice, respectively, yield a phylogenetic tree matching that of their hosts (Soeda *et al.* 1980; Shadan & Villarreal 1993) and suggest a host–parasite coevolution for over 30 Ma. Long-term coevolution of human ancestors with their parasites is also indicated for some trematode worms of the genus *Schistosoma* (Despres *et al.* 1992) and for non-falciparan *Plasmodium* species (Waters *et al.* 1993). Numerous cases of parasite–host coevolution have been compiled by Brooks & McLennan (1993).

Into the third category we assign organisms so well adapted that they normally not only cause no measurable damage, but their presence is often beneficial to their hosts. Only when the host becomes immunologically compromised, for example in individuals infected with the human immunodeficiency virus or individuals immunodeficient because of a genetic defect, is the true nature of these organisms revealed. Then the harmless commensals turn into *bona fide* parasites and cause an opportunistic infection resulting more often than not in death This behaviour indicates that, in a healthy individual, these opportunistic parasites are kept at bay by the immune system, which includes the MHC molecules. It has been estimated that a healthy human houses about 10^{12} bacteria on the skin, 10^{10} in the mouth, and 10^{14} in the digestive tract (Mims 1977). It must be a tremendous challenge to the immune system to survey this mass of *Gastarbeitern* and make sure that their behaviour does not violate the body's norms. The human opportunistic parasites include viruses (cytomegalovirus, herpes simplex virus, varicella–zoster virus), bacteria (*Escherichia coli*, *Pseudomonas*, *Streptococcus*), fungi (*Candida*) and protozoa (*Pneumocystis carini*). Many of these organisms are widespread in nature and are of an old age. *E. coli*, for example, is a normal constituent of the gut flora in a wide range of warm-blooded animals. It is estimated to have diverged from *Salmonella*, its closest relative, between 120 and 160 Ma ago (Ochman & Wilson 1987); during all that time, it has presumably remained a commensal and an occasional parasite of a great many different host species. The population of *E. coli* has been shown to be clonal (Selander & Levin 1980; Ochman & Selander 1984); the clones, as in other parasites, are believed to be of ancient origin. (Some of the *Trypanosoma cruzi* clones, for example, are estimated to have been isolated from each other for 40–50 Ma BP; see Tibayrenc *et al.* (1986).) It is hard to imagine that such a ubiquitous mass of parasites that has presented so great a challenge to their hosts for so long a period should not become a major driving force in the evolution of the vertebrate immune system and in particular of the MHC.

We propose, therefore, that parasites of the second and third categories are primarily responsible for the evolution and for the properties of MHC polymorphism. They are in a position to provide steadfast pressure on the MHC genes to diversify and for the variants to endure for periods exceeding those normally allotted to a species. Parasites in the first category, on the other hand, have little, if any, influence on MHC polymorphism and may actually be reducing it. If our reasoning is correct, two further deductions can be made. First, one of the explanations for the scantiness of empirical evidence regarding the effect of parasites on MHC polymorphism is that the wrong parasites (those of the first category) have been studied. Second, it will be necessary to design experiments in which the selection pressure exerted on the MHC will be measured by taking commensals or adapted parasites from a host of one MHC haplotype and transferring them to a slightly immunocompromised host of another haplotype.

6. PARASITES AND ALLELIC LINEAGES

If the reason for the trans-species evolution of MHC polymorphism is the long-lasting host–parasite association and coevolution of the two protagonists, how then does the existence of the allelic lineages fit into this concept? Could the lineages not simply be the result of the stochasticity of allelic genealogy? We believe not, for the following reason. According to the theory of gene genealogy, all the genes of a given generation coalesce into a single, most recent common ancestor $4N_e$ generations ago (Hartl & Clark 1989). When the coalescence process is dissected into separate steps, it turns out that the coalescence of all the genes to two ancestors takes $2N_e$ generations and the coalescence of these two into a single ancestor another $2N_e$ generations. In the *HLA–DRB* system, an initial single locus has duplicated repeatedly and produced nine paralogous loci (*DRB1* to *DRB9*). Most of the duplications can be dated approximately to the time of the primate emergence on the evolutionary scene; only one duplication may have occurred somewhat later (Klein *et al.* 1991). Had the emergence of allelic lineages predated the initial duplication, genes at different loci should be more closely related to some of the lineages, but this is not the case. From the genetic distances between the lineages, we estimate that all the lineages began diverging at about the same time. This observation contradicts the expectation under a purely stochastic process, according to which the divergence of the first two lineages should have taken half the time of the total divergence of all the lineages.

We believe, therefore, that there is more to the existence of MHC allelic lineages than genealogical stochasticity. The differentiation into lineages may have been assisted by one of two processes. First, the discontinuity in the spectrum of MHC variability manifested in the existence of the lineages may be a reflection of discontinuities in the parasite spectrum manifested in the existence of different species and clones within a 'species' (Klein 1991). Second, the lineages may be manifestations of a strategy to keep parasites from spreading through the entire population. A parasite that has learned to breach defences built on antigen presentation through one lineage of MHC molecules will not be able to avoid the immune response in individuals of the same species carrying other MHC lineages, if the lineages differ from one another to a great extent. We prefer the former

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

argument because the latter smacks of group selection.

The existence and perpetuation of motifs characterizing individual lineages may be explained in the same manner. Sharing of motifs between MHC genes and proteins may, however, occur in two ways: by descent and by convergence. There are good reasons to believe that shared motifs characterizing the *DRB1* lineages in different catarrhine primates owe their similarity to the common origin of the genes within the lineage. The strongest evidence for the common origin hypothesis is the sharing by these genes of the same inserts of *Alu* and retroviral elements (Y. Satta & J. Klein, manuscript in preparation). The occasional occurrence of a *DRB1* motif in non-primate mammals (such as the presence of the EYST motif in artiodactyls and rodents), on the other hand, must be the result of convergence. This conclusion is supported by two observations. First, we have examined some 500 *DRB* alleles and allelic pairs differing by single, double, and treble replacements. The location and type of replacements differentiating each member of the pair were scored and the data were used to measure (i) the relative frequency of replacement at each of the PBR sites and (ii) the types of replacement that occur. Some PBR sites were observed to undergo replacement at a much higher frequency than others; at many PBR sites, replacements are limited to a few amino acids. When we applied these observations to a coalescence model to simulate lineage divergences, we found that highly diverged alleles may resemble each other at a number of motif-bearing sites and hence that convergence has occurred (O'hUigin 1994).

Based on the high degree of shared codon usage in the DNA encoding the motifs, Lundberg & McDevitt (1992) have argued that polymorphisms in the motif-encoding DNA segments have directly descended to mouse and humans from the eutherian ancestor, whereas the lineages in each species have been generated *de novo* through recombination. However, because the third-position GC content in the class II genes examined by Lundberg & McDevitt is quite high (70–80%) and because GC content is the major determinant of mammalian codon usage (Aota & Ikemura 1986), the shared codon usage may be attributable to the high GC content.

Second, as mentioned above, the *DRB1* lineages apparently arose after the duplication of the ancestral *DRB* gene in the primates. They could not, therefore, have existed in the common ancestor of primates, artiodactyls, and rodents. The same conclusion applies to the motifs diagnostic of the individual lineages; these must have arisen independently in the three mammalian orders.

7. MHC AND SELF

We have limited our discussion of MHC polymorphism to the stimuli provided by the non-self world, the parasites. Of course, the MHC molecules interact most of the time not with non-self, but with self peptides; one may wonder, therefore, whether some minor, or even major, influence on the polymorphism may not come from the host itself. Space does not allow us to go into any details, except to state our position on this topic. A failure to bind a critical non-self peptide may endanger the life of an individual; a failure to bind a self peptide is probably to have no consequences at all because the worst that can happen is that the molecule will not be expressed on the cell surface. However, this is unlikely, because there will always be other self peptides that will bind, and even if a molecule were not expressed, the encoding element would eventually turn into a pseudogene and be replaced by another gene. In either case, there would not be much influence on the polymorphism of the functional MHC genes. It is true, however, that little is still known about the process of T-cell selection by MHC molecules in the thymus and it is therefore possible that this view will need to be modified once the actual mechanism of the selection process is elucidated.

We thank Ms Lynne Yakes for editorial assistance and Dr Robert Slade for critical reading of the manuscript.

REFERENCES

Allsopp, C.E.M., Harding, R.M., Taylor, C., Bunce, M., Kwiatkowski, D., Anstey, N., Brewster, D., McMichael, A.J., Greenwood, B.M. & Hill, A.V.S. 1992 Interethnic genetic differentiation in Africa: HLA class I antigens in the Gambia. *Am. J. hum. Genet.* **50**, 411–421.

Aota, S. & Ikemura, T. 1986 Diversity in G + C content at the third position of codons in vertebrate genes and its cause. *Nucl. Acids Res.* **14**, 6345–6355.

Brooks, D.R. & McLennan, D.A. 1993 *Parascript, Parasites and the language of evolution.* Washington: Smithsonian Institution Press.

Cameron, T.W.M. 1956 *Parasites and parasitism.* London: Methuen & Co.

Carter, R., Schofield, L. & Mendis, K. 1992 HLA effects in malaria: Increased parasite-killing immunity or reduced immunopathology? *Parasitol. Today* **8**, 41–42.

Despres, L., Imbert-Establet, D., Combes, C. & Bonhomme, F. 1992 Molecular evidence linking hominoid evolution to recent radiation of schistosomes (Platyhelminthes: Trematoda). *Molec. phylogenet. Evol.* **1**, 295–304.

Fahrenholz, H. 1913 Ectoparasiten und Abstammungslehre. *Zool. Anz., Leipzig* **41**, 371–374.

Germain, R.N. 1994 MHC-dependent antigen processing and peptide presentation: Providing ligands for T lymphocyte activation. *Cell* **76**, 287–299.

Gill, T.J. III 1994 Reproductive immunology and immunogenetics. In *The Physiology of reproduction* (ed. E. Knobil & J. D. Neil), 2nd edn, pp. 783–812. New York: Raven Press.

Hartl, D.L. & Clark, A.G. 1989 *Principles of population genetics*, 2nd edn. Sunderland, Massachusetts: Sinauer.

Hill, A.V.S., Allsopp, C.E.M., Kwiatkowski, D., Anstey, N.M., Twumasi, P., Rowe, P.A., Bennett, S., Brewster, D., McMichael, A.J. & Greenwood, B.M. 1991 Common West African HLA antigens are associated with protection from severe malaria.; *Nature, Lond.* **352**, 595–600.

Hughes, A.L. & Nei, M. 1989 Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection. *Proc. natn. Acad. Sci. U.S.A.* **86**, 958–962.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Johnson, R.B. 1986 Human disease and the evolution of pathogen virulence. *J. theor. Biol.* **122**, 19–24.
Klein, J. 1986 *Natural history of the major histocompatibility complex*. New York: John Wiley.
Klein, J. 1987 Origin of major histocompatibility complex polymorphism: The trans-species hypothesis. *Hum. Immun.* **19**, 155–162.
Klein, J. 1989 Are invertebrates capable of anticipatory immune response? *Scand. J. Immun.* **29**, 499–505.
Klein, J. 1991 Of HLA, tryps, and selection: An essay on coevolution of MHC and parasites. *Hum. Immun.* **30**, 247–258.
Klein, J., Satta, Y., O'hUigin, C., Mayer, W.E. & Takahata, N. 1991 Evolution of the primate *DRB* region. In *HLA 1991* (ed. T. Sasazuki), vol. 2, pp. 45–56. (Proceedings of the 11th International Histocompatibility Workshop, Yokohama, Japan.) Oxford University Press.
Klein, J., Satta, Y., O'hUigin, C. & Takahata, N. 1993 The molecular descent of the major histocompatibility complex. *A. Rev. Immun.* **11**, 269–295.
Lundberg, A.S. & McDevitt, H.O. 1992 Evolution of major histocompatibility complex class II allelic diversity: Direct descent in mice and humans. *Proc. natn. Acad. Sci. U.S.A.* **89**, 6545–6549.
Marsh, S.G.E. & Bodmer, J.G. 1993 *HLA* class II nucleotide sequences, 1992. *Immunogenetics* **37**, 79–94.
Martin, R.D. 1993 Primate origins: plugging the gaps. *Nature, Lond.* **363**, 223–234.
Miller, J.F.A.P. 1992 The key role of the thymus in the body's defense strategies. *Phil. Trans. R. Soc. Lond. B* **337**, 105–124.
Mims, C.A. 1977 *The pathogenesis of infectious disease*. London: Academic Press.
Neefjes, J.J. & Momburg, F. 1993 Cell biology of antigen presentation. *Curr. Opin. Immun.* **5**, 27–34.
Nelson, G.S. 1988 Parasitic zoonoses. In *The biology of parasitism*, pp. 13–41. New York: Alan R. Liss.
Ochman, H. & Wilson, A.C. 1987 Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes. *J. molec. Evol.* **26**, 74–86.
Ochman, H. & Selander, R.K. 1984 Evidence for clonal population structure in *Escherichia coli*. *Proc. natn. Acad. Sci. U.S.A.* **81**, 198–201.
O'hUigin, C. 1994 Quantifying the degree of convergence in primate Mhc-DRB genes. *Immunol Rev.* (In the press.)
O'hUigin, C., Bontrop, R. & Klein, J. 1993 Nonhuman primate MHC-*DRB* sequences: a compilation. *Immunogenetics* **38**, 165–183.
Potts, W.K. & Wakeland, E.K. 1994 Evolution of MHC genetic diversity: A tale of incest, pestilence and sexual preference. *Trends Genet.* **9**, 408–413.
Rammensee, H.G., Falk, K. & Rötzschke, O. 1993 MHC molecules as peptide receptors. *Curr. Opin. Immun.* **5**, 35–44.
Rosqvist, R., Skurnik, M. & Wolf-Watz, H. 1988 Increased virulence of *Yersinia pseudotuberculosis* by two independent mutations. *Nature, Lond.* **334**, 522–525.
Schwaiger, F.-W., Weyers, E., Epplen, C., Brün, J., Ruff, G., Crawford, A. & Epplen, J.T. 1993 The paradox of MHC-*DRB* exon/intron evolution: α-helix and β-sheet encoding regions diverge while hypervariable intronic simple repeats coevolve with β-sheet codons. *J. molec. Evol.* **37**, 260–272.
Selander, R.K. & Levin, B.R. 1980 Genetic diversity and structure in *Escherichia coli* populations. *Science, Wash.* **210**, 545–547.
Shadan, F.F. & Villarreal, L.P. 1993 Coevolution of persistently infecting small DNA viruses and their hosts linked to host-interactive regulatory domains. *Proc. natn. Acad. Sci. U.S.A.* **90**, 4117–4121.
Soeda, E., Maruyama, T., Arrand, J.R. & Griffin, B.E. 1980 Host-dependent evolution of three papova viruses. *Nature, Lond.* **285**, 165–167.
Stanley, S.M. 1975 A theory of evolution above the species level. *Proc. natn. Acad. Sci. U.S.A.* **72**, 646–650.
Takahata, N. 1990 A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2419–2423.
Takahata, N., Satta, Y. & Klein, J. 1994 Divergence time and population size in the lineage leading to modern humans. *Theor. Popul. Biol.* (In the press.)
Tibayrenc, M., Ward, P., Moya, A. & Ayala, F.J. 1986 Natural populations of *Trypanosoma cruzi*, the agent of Chagas disease, have a complex multiclonal structure. *Proc. natn. Acad. Sci. U.S.A.* **83**, 115–119.
Waters, A.P., Higgins, D.G. & McCutchan, T.F. 1993 The phylogeny of malaria: A useful study. *Parasitol. Today* **9**, 246–252.

Discussion

C. A. MIMS (*Sheriff House, Ardingly, West Sussex, U.K.*). Professor Klein suggests that in the case of infections that appeared in humans more than 10 000 years ago, the MHC has not had time to respond by developing appropriate molecules for the presentation of the relevant microbial antigens. My question is about time. We know that host genetic resistance can develop quite rapidly under pressure from a virulent infection. Host and virus changes have been thoroughly charted in myxomatosis in the Australian rabbit, an infectious disease that evolved in a highly susceptible host. Genetically resistant rabbits appeared within 10–15 years, which, after allowing for differences in generation times between humans and rabbits, is far less than the 10 000 years you suggested for human infections. Also, although here the picture is complicated by the influence of nutrition, poverty, housing, it seems probable that in Europeans during the past few hundred years there has been a significant selection for genotypes resistant to tuberculosis. Of course, it is not known to what extent any of these host genetic changes are based on MHC, but it leads me to ask why the MHC should take so long in responding to life-threatening infections.

J. KLEIN. I wanted to say that the MHC polymorphism now present in the global population is more than 10 000 years old and must have therefore been established by selection pressures other than those that appeared after the neolithic revolution. It was not my intention to deny that new variants keep emerging in local populations and that their fate may be influenced by positive selection exerted by parasites. But the time needed for their incorporation as polymorphisms into the global population is, on average, longer than 0.5 million years. The other point I was trying to make was that the MHC (or, more precisely, the anticipatory immune system) may not be able to protect us (or the rabbits) from highly virulent parasites. Genes other than the MHC may decide who is to be resistant and who susceptible, and the infection stops when the parasites do not find any more susceptible individuals to infect.

J. LINES (*London School of Hygiene and Tropical Medicine, U.K.*). Professor Klein raised the question of why we do not see more obvious associations between the MHC and infectious disease in natural populations. However, suppose the relationship between host and parasite were a matching

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

allele system, and that the allele frequencies of both partners were at selective equilibrium. At equilibrium, genotype fitnesses are by definition equal, and so at the population level there would be no visible association between a given host genotype and disease in general. Particular parasite genotypes would be associated with particular host genotypes, but this would only be seen by scoring the genotype of both parasite and host in individual infections. Have any such surveys, taking account of the genotype of both partners, been carried out with the infections that Professor Klein showed us in his table?

J. KLEIN. In all the studies that I am aware of, the parasite is regarded as being homogeneous and only the frequencies of MHC alleles of the susceptible and resistant individuals are compared.

J. DEUTSCH (*Zoology Department, University of Cambridge, U.K.*). If one were able to characterize the MHC allelic lineages in two closely related species, would it be possible to estimate the population size of initial populations at speciation based upon the proposition of allelic lineages shared by the species?

J. KLEIN. One of the main reasons why we have been sequencing MHC alleles of related primate species is to estimate the sizes of the founding populations using the population genetics theory of allelic genealogy and models of computer simulations. The results obtained thus far indicate that for *Homo sapiens*, the founding effective population size was between 10 000 and 100 000 individuals (see Klein *et al. Trends Genet.* **6**, 7–11 (1990); Klein *et al. J. med. Primatol.* **22**, 57–64 (1993); Klein *et al. Scient. Amer.* December 1993, pp. 56–62.)

A. L. HUGHES (*Department of Biology, Pennsylvania State University, U.S.A.*). I think it need not be true that *Plasmodium Falciparum* has become a human parasite only recently.

J. KLEIN. My inclusion of *Plasmodium falciparum* in the group of recent parasites is based on the data of A. P. Waters and colleagues (*Proc. natn. Acad. Sci. U.S.A.* **88**, 3140–3144 (1991)). On the basis of sequence analysis of small-subunit ribosomal RNA genes, these investigators reported that *P. falciparum* forms a monophyletic group with avian *Plasmodium* species, highly divergent from primate and rodent *Plasmodium* parasites. They argue that *P. falciparum* and avian *Plasmodium* species shared a common ancestor relatively recently and hence that their results are consistent with the hypothesis of a recent acquisition of *P. plasmodium* by humans, possibly at the time a switch-over to an agricultural-based lifestyle occurred (see also Livingstone *Am. Anthropol.* **60**, 533–560 (1958)).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES