J Mol Evol (1997) 45:589–598

JOURNAL OF MOLECULAR EVOLUTION

Mhc Allelic Diversity and Modern Human Origins

Robert E. Hickson,* Rebecca L. Cann

Department of Genetics and Molecular Biology, John A. Burns School of Medicine, 1960 East-West Rd., University of Hawaii at Manoa, Honolulu, HI 96822, USA

Received: 19 February 1997 / Accepted: 12 June 1997

Abstract. Thirty complete coding sequences of human major histocompatibility complex (*Mhc*) class II *DRB* alleles, spanning 237 codons, were analyzed for phylogenetic information using distance, parsimony, and likelihood approaches. Allelic genealogies derived from different parts of the coding sequence (exon 2, the 5′ and 3′ ends of exon 2, respectively, and exons 3–6) were compared. Contrary to prior assertions, a rigorous analysis of allelic genealogies in this gene family cannot be used to justify the claim that the lineage leading to modern humans contained on average at least 100,000 individuals. Phylogenetic inferences based upon the exon 2 region of the *DRB* loci are complicated by selection and recombination, so this part of the gene does not provide a complete and accurate view of allelic relationships. Attempts to reconstruct human history from genetic data must use realistic models which consider the complicating factors of nonequilibrium populations, recombination, and different patterns of selection.

Key words: Major histocompatibility complex (Mhc) — HLA — Human evolution — Phylogenetics — Coalescence — Genetic diversity — Selection

Introduction

In all the human evolutionary models currently undergoing examination, the issue of population size inferred from modern estimates of genetic diversity occupies a key position. Given the importance of effective population size for modeling adaptive evolution (Fisher 1930), the desire to extract as much information as possible from underlying patterns of DNA sequence diversity is strong. One of the chief analytical difficulties, however, is that small local populations spread over 12,000 km for 2 million years were probably not in genetic equilibrium for long periods of time (Cavalli-Sforza et al. 1994). Populations fluctuated in size, there were local extinctions and recolonizations, as well as nonrandom patterns of mating and migration (Takahata 1995; see also Goldstein et al. 1995). Complex processes of molecular evolution can present additional challenges to reconstructing history from genetic data. Careful and sophisticated analyses are, therefore, required to infer past population dynamics from current patterns of genetic diversity.

Mitochondrial DNA (mtDNA) and loci of the major histocompatibility complex (*Mhc*) have provided the most well-known examples of attempts to produce molecular estimates of long-term effective population size of modern humans. Low levels of mtDNA diversity in human populations have been used to infer relatively small effective long-term population sizes (approximately 10,000 breeding individuals; see Hartl and Clark 1989; Takahata 1995). Similar estimates have been derived from nuclear protein polymorphisms (Nei and Graur 1984), and these led to the interpretation of a ‘‘population bottleneck’’ in the evolution of modern humans (Cann et al. 1987; Vigilant et al. 1991).

Recent debate, however, has focused attention upon how accurately the mtDNA data reflects the full evolutionary history of modern humans. Levels of nuclear genetic diversity within and between human populations

**Present address:* AgResearch, Wallaceville Animal Research Centre, PO Box 40063, Upper Hutt, New Zealand
Correspondence to: R. Hickson; e-mail hicksonr@agresearch.cri.nz

Table 1. *Mhc* data sets used in the analyses: Data set relates to which figure the sequences are used in (data set A = Fig. 2; B = Fig. 3)[a]

Data set	Exons used	Species and Locus	Abbreviation	Source
A,B	Exons 2–6	Human DRB (n = 30)	DRB≠	Web site[b]
B	Exons 2–6	*Gorilla gorilla* DRB1	Gogo1*08	M77154
B	Exons 2–6	*Saguinus oedipus* DRB*01	Tamarin	M76488
B	Exons 2–6	*Canis familiaris* DRB1	Dog	U47339
B	Exons 2–6	*Bos taurus* DRB1	Cow	X92409
B	Exons 2–6	*Mus musculus E-beta-b*	Mouse	M36940
B	Exons 2–6	*Macropus rufogriseus* DBB	Wallaby	M81625

[a] Sources of *Mhc* allele sequences used for analyses are given as either their GenBank accession numbers or their World Wide Web site location. For the human alleles a hash (≠) indicates a specific numbered locus and allele, as shown in the figures. *Saguinus* is a new world monkey
[b] http://histo.cryst.bbk.ac.uk/WWWFiles/sequences/drbdna_aln.html

can be high, especially when compared to our closest primate relatives (Takahata 1995; Herbert and Easteal 1996). In particular, the very high levels of allelic diversity for some loci within the *Mhc* have been viewed as being incompatible with the relatively small population sizes inferred from the mtDNA data (Klein et al. 1993; Ayala et al. 1994). There are, for example, over 100 described alleles each for *HLA-B* and *DRB1* loci in human populations (Marsh and Bodmer 1995; Parham and Ohta 1996).

Ayala and colleagues (Klein et al. 1993; Ayala et al. 1994; Ayala 1995, 1996; Ayala and Escalante 1996) have suggested that while many alleles could pass through a bottleneck consisting of a few thousand individuals, *maintenance* of a high degree of diversity required a long-term effective population size of 100,000—an order of magnitude higher than the size inferred from mitochondrial data—and so is incompatible with the occurrence of a significant reduction in human effective population size during the last tens of millions of years.

However, we consider that some of the assumptions made by Ayala et al. (1994; Ayala 1995; Ayala and Escalante 1996) were inappropriate because of factors involved in the generation and maintenance of *Mhc* diversity. *Mhc* loci code for cell-surface receptors which are involved in the presentation of foreign antigens to T-cells of the immune system. Recognition by a T-cell of an *Mhc*-bound antigen complex activates components of the immune system to detect and destroy cells displaying the specific antigen (Klein 1986). Most of the nucleotide diversity between *Mhc* alleles is at sites directly involved in interactions with the foreign peptides (Hughes and Nei 1988, 1989; Brown et al. 1993). Maintenance of large numbers of *Mhc* alleles is thought to be due primarily to the action of balancing selection, where individuals heterozygous at a *Mhc* locus are able to bind a broader range of peptides derived from potentially pathogenic or debilitating parasites (Hughes and Nei 1992).

Recombination and/or convergence influences patterns of similarity between alleles at some *Mhc* loci (Gyllensten et al. 1991; Hughes et al. 1994; Andersson and Mikko 1995; O'hUigin 1995; Parham and Ohta 1996). Selection for, or against, specific alleles (e.g., Hill et al. 1992) may also complicate analyses. This contrasts sharply with the patterns and processes of change at other loci, such as mtDNA. Previous analyses of *Mhc* data (e.g., Ayala et al. 1994) did not adequately consider these confounding influences and, consequently, inferences about past human population sizes based upon these analyses may not be valid (see also Erlich et al. 1996).

In this paper we use more careful and rigorous analyses to examine what phylogenetic and population inferences can be made from human *DRB* sequence data and find that the conclusions of Ayala and colleagues are not well supported. We emphasize the need for analyses and population models which more accurately reflect the patterns and processes of diversification in real populations.

Methods

Many partial and some complete human *DRB* sequences are available, although the largest data set exists for sequences encoding only the second exon, which corresponds to the peptide-binding region (Andersson et al. 1987). Our analyses concentrated upon 30 sequences for which complete exon data were available (22 *DRB1* alleles, two alleles each for *DRB3* and *DRB4*, and four alleles from the *DRB5* locus). By alleles we mean nucleotide sequence variants at a locus. All of these loci are expressed and produce functional molecules in humans (Klein et al. 1992). Databases of human *Mhc* sequences are maintained by P.J. Travers on the World Wide Web at http://histo.cryst.bbk.ac.uk/WWWFiles/sequences/sequence_map.html and are also available via FTP from FTP.EMBL-Heidelberg.DE in the directory /pub/databases/hla (Marsh and Bodmer 1995).

Gyllensten et al. (1991) noted that the beta-pleated sheet region (codons 5–54) and the alpha helix region (codons 55–94) of *DRB1* exon 2 contained different phylogenetic information, so we partitioned the data set of 30 sequences into four subsets; exon 2 as a whole (codons 5–94), the 5′ end of exon 2 (codons 5–54) containing the beta-pleated sheets, the 3′ end of exon 2 (codons 55–94) including the alpha helix, and the rest of the molecule (exons 3–6, codons 95–237) corresponding to the second, transmembrane, and cytoplasmic domains of the protein. Exon 1 is not included here as it is short (four codons) and is removed during processing (Andersson et al. 1987). Comparisons were also made between primate and other mammalian *DRB*-like sequences (Table 1).

We chose neighbor joining (Saitou and Nei 1987) as an appropriate method of analysis because of variability in evolutionary rates between *DRB* alleles and loci (see Results). Previous analyses used UPGMA to reconstruct allelic genealogies (Ayala et al. 1994), but where rate varia-

tion occurs this method is likely to lead to inconsistency (Nei and Roychoudhury 1993). Bootstrap values provided an indication of the stability or robustness of phylogenies. Phylogenetic reconstructions were performed using the neighbor-joining distance method as implemented in MEGA (version 1.0, Kumar et al. 1993). A range of genetic distance measures were evaluated (Jukes-Cantor, Kimura 2-parameter, Tamura-Nei, and Gamma) and all produced similar topologies and bootstrap values. Parsimony analysis (using PAUP version 3.1, Swofford 1993) was also performed on the data set, while maximum-likelihood (Felsenstein 1991) and spectral analyses (Lento et al. 1995) were used on smaller collections of the data. All these analyses gave similar results.

Results

The distribution of amino acid variability within the molecule is shown in Fig. 1. As has been found in earlier studies (Hughes and Nei 1988, 1989), amino acid replacement changes (nonsynonymous substitutions) predominate in exon 2, especially at antigen recognition sites (ARS, see Fig. 1; Table 2). Nonsynonymous substitutions are less common in the other exons (Table 2). There were no substantial differences in base composition between sequences.

Phylogenetic Structure

Genealogies inferred from different parts of the *DRB* molecule are shown in Fig. 2. These are unrooted and based on synonymous as well as nonsynonymous changes. Consideration of just nonsynonymous substitutions as well as a range of different nucleotide correction models and optimality criteria did not affect the essential structure or robustness of the phylogenies (data not shown). Exclusion of the 24 codons which comprise the ARS (see Brown et al. 1993) greatly decreased phylogenetic resolution in exon 2 (data not shown).

There are two important features of the genealogies presented in Fig. 2. First, the alpha helix (Fig. 2C) presents a different set of relationships than the rest of the coding region (Fig. 2B,D). This is consistent with the earlier analyses of Gyllensten et al. (1991). Tree comparisons using the partition and path difference metrics (Steel and Penny 1993) indicated that Fig. 2C is substantially different from the other phylogenies. The beta sheet and exon 3–6 genealogies are similar in that alleles of the same serological types cluster together. However, the placement of *DRB1**0901 differs, and comparison of other mammalian DRB-like sequences shows that relationships based upon beta sheet and exons 3–6 data can differ (Fig. 3).

The second major point is that the low bootstrap values for the *DRB* data sets indicate that branching orders within the phylogenies are poorly supported, particularly when exon 2 as a whole (Fig. 2A) or the alpha helix region (Fig. 2C) is used to reconstruct relationships. However, for the genealogies derived from the beta sheets (Fig. 2B) and exons 3–6 (Fig. 2D) there are nine relatively well-supported clusters (bootstrap values >75%), which we refer to as allelic lineages. These correspond to distinct serological subdivisions of the alleles, although on the basis of sequence analysis, discrimination of the *DRB1*03*, **08*, **11*, **12*, and **13* alleles from each other is less clear.

Timing of Diversification

Rates of nucleotide substitutions for *Mhc* loci have been calculated by Satta et al. (1994). Their most reliable estimate for a nonsynonymous substitution rate at ARS in *DRB1* was between 6.8 and 9.3 substitution/site/10^9 years, while the synonymous substitution rate across all sites was calculated to be 1/site/10^9 years. A nonsynonymous rate estimate for sites not directly involved in peptide binding is not available but is expected to be less than the synonymous substitution rate.

However, the usefulness of such estimates for inferring divergence times is doubtful because of variation in substitution rates for both different parts of the gene (Tables 2, 3) and for different alleles (Fig. 2). Nonsynonymous substitutions appear highest at ARS in the beta sheets and lowest in exons 4–6. Comparison of synonymous substitutions within and between lineages give anomalously high values for the alpha helix region (Tables 2, 3). Applying the rate estimates of Satta et al. (1994) results in very different predicted ages of lineages depending upon which sites are used (Table 2). Given that lineages probably diverged millions of years ago, use of synonymous substitutions are unlikely to provide reliable estimates for the ages of lineages.

Within allelic lineages there are no or few nucleotide substitutions between alleles in exons 3–6 (see Fig. 2D), and alleles tend to differ from each other primarily by substitutions in the alpha helix region of exon 2 (Table 3). While synonymous substitutions per site within a lineage range up to 0.0287 for exons 3–6, one-third of the comparisons within lineages have no nucleotide differences between them.

Discussion

It is essential to place emphasis on generating a reliable tree first before favoring a particular explanation for the pattern of relationships and drawing general conclusions. Ayala et al. made several assumptions about *DRB* evolution which are not supported by examination of the data. Specifically, they did not consider whether simple analyses of *DRB* exon 2 sequence data reliably reflected the phylogenetic histories of the alleles.

Phylogenetic Structure

Our analyses identified two key features in the data. The first echoes Gyllensten et al. (1991) by showing the dif-

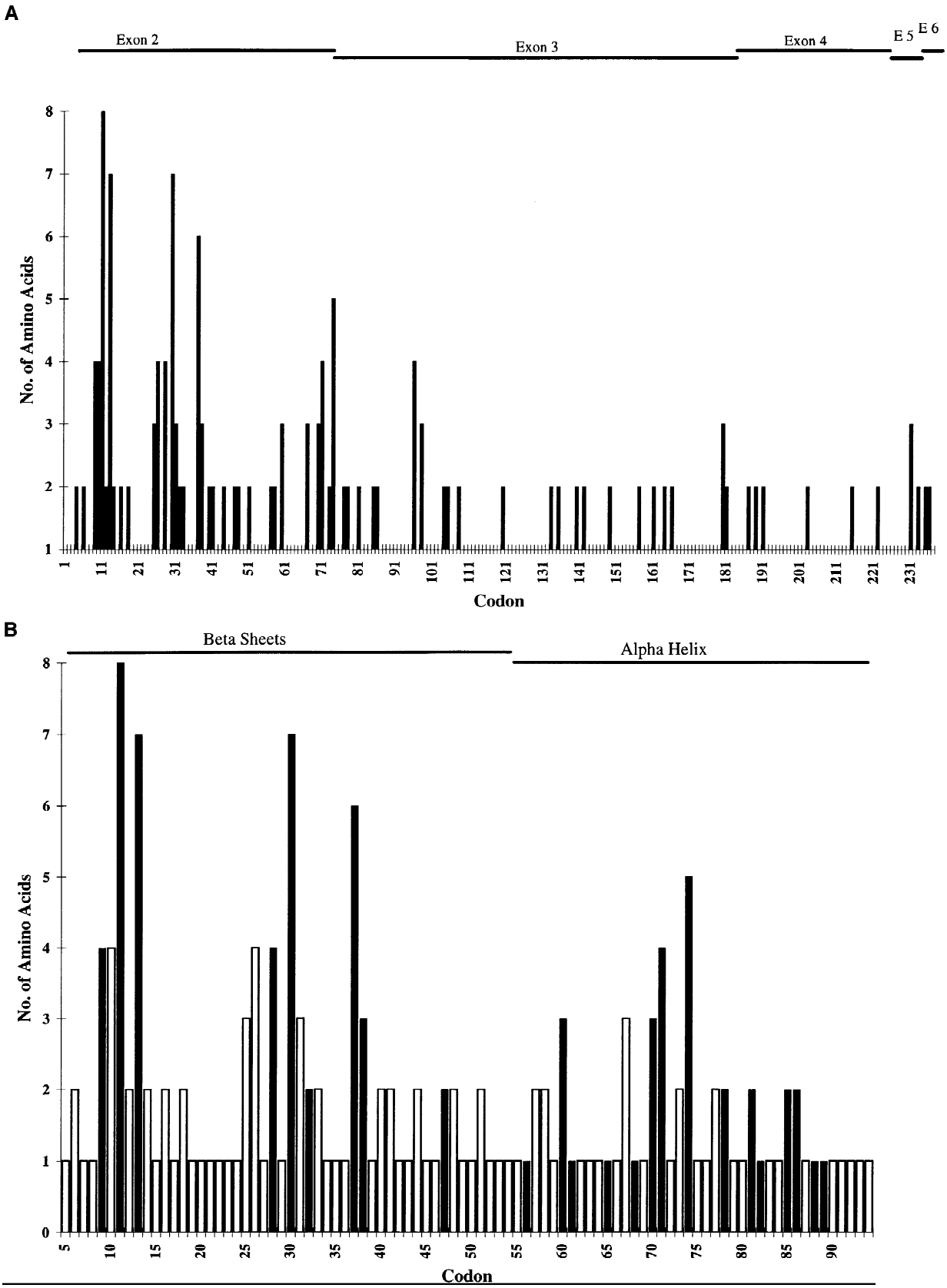

Fig. 1. Amino acid variability in the 30 *DRB* alleles. The codon position is given along the x-axis, and the number of amino acids observed at each position is plotted on the y-axis. **A** Variability across the whole coding region. Exon delineations are given above the chart. **B** variability within exon 2, the peptide-binding region. The sites involved in peptide binding (see Brown et al. 1993) are indicated by *filled bars.*

Table 2. Comparisons of substitution patterns between allelic lineages[a]

	dR	dS	dR/dS ratio	Divergence time (million years)
Exon 2, all sites	0.101 (± 0.001)	0.091 (± 0.002)	1.1	
Exon 2, ARS only	*0.27* (± 0.05)	0.057 (± 0.002)	4.7	29–40
Beta sheets; ARS only	*0.664* (± 0.014)	0.01 (± 0.002)	62.9	71–98
Alpha helix; ARS only	*0.146* (± 0.004)	0.077 (± 0.004)	1.6	16–22
Exon 2, excl. ARS	0.043 (± 0.001)	0.103 (± 0.002)	0.42	
Beta sheets, excl. ARS	0.051 (± 0.002)	0.088 (± 0.002)	0.60	
Alpha helix, excl. ARS	0.030 (± 0.001)	0.131 (± 0.004)	0.23	
Exons 3–6	0.026 (± 0.001)	*0.064* (± 0.001)	0.41	64
Exon 3	0.032 (± 0.001)	0.079 (± 0.001)	0.41	
Exons 4–6	0.023 (± 0.001)	0.039 (± 0.002)	0.59	
Exons 1–6	0.056 (± 0.001)	*0.074* (± 0.001)	0.76	74

[a] Mean number of nonsynonymous (dR) and synonymous (dS) substitutions per site (± SEM; calculated using the Jukes-Cantor correction) for different parts of the *DRB* coding region for 30 alleles. The dR/dS ratio is also shown. For the beta sheet and alpha helix regions of exon 2 the substitutions per site both including and excluding antigen recognition sites (ARS; as identified by Brown et al. 1993) are given. The italicized values are used to calculate divergence times based on Satta et al.'s (1994) rates of 6.8–9.3 substitutions/site/10^9 years for nonsynonymous substitutions at ARS and 1 substitution/site/10^9 years for synonymous substitutions across the gene. The lineages are defined as alleles within the groups *DRB1*01, 15/16, 04, 0701, 1001, 03/08/11/12/13, DRB3, DRB4,* and *DRB5* (see text)

ferent evolutionary histories contained within *DRB1* alleles. Phylogenies derived from the beta-pleated sheet region of exon 2 (Fig. 2B) and exons 3–6 (Fig. 2D) are more consistent with each other, while that obtained from the alpha helix mixes alleles from different lineages (Fig. 2C). Alleles within a lineage are often only distinct from each other by substitutions within the alpha helix, but the patterns of substitutions in this region do not always provide phylogenetically reliable information (see also Titus-Trachtenberg et al. 1994). Factors which could account for this disparity are discussed below.

The second significant feature of the analyses is that resolution of *DRB* genealogical structure is poor, irrespective of whether the alpha helix is included (Fig. 2A) or excluded (Fig. 2B) from phylogenetic analyses. Sequence analyses of the beta sheets of exon 2 and of exons 3–6 are consistent in that they identified five major *DRB1* lineages, which are all serologically distinct: DR1 (the **01* alleles), DR2 (**15/16*), DR4 (**04*), DR7 (**07*), and DR10 (**10*). The three other loci analyzed (*DRB3, DRB4,* and *DRB5*) also form their own distinct lineages. The remaining alleles (**03, *08, *11, *12,* and **13*) comprise a ninth lineage. While this last group of alleles can usually be distinguished serologically the sequence data was not adequate to resolve their relationships (Fig. 2).

A limitation of using only exon 2 sequence data is illustrated by comparison with phylogenetic inferences derived from exons 3–6. *DRB1**0901 clusters with *DRB5* alleles when the beta sheet regions are analyzed (Fig. 2B), while the exon 3–6 data places *DRB1**0901 with *DRB1**0701 (Fig. 2D). This appears to be due to interlocus recombination (Klein et al. 1992). Similarly, in the analyses of other mammalian *DRB* sequences there are differences in patterns of association between the exons (Fig. 3A, B).

Noncoding data may be expected to help clarify patterns of allele evolution. In an analysis of intron sequence data Satta et al. (1996) suggested that *DRB3* arose by gene duplication from *DRB1*03.* For the exon data this association is only apparent in the alpha helix region (Fig. 2C), which in our view does not provide the most reliable phylogenetic information. The fact that there are no uniquely shared substitutions between *DRB1*03* and *DRB3* in the exons, as well as the lack of support from other coding regions, makes the association of these lineages ambiguous. A hitchhiking effect to explain similarities between phylogenetic analyses of the alpha helix and intron data can be discounted since the introns examined by Satta et al. (1996) are not contiguous with exon 2. We regard Satta et al.'s (1996) conclusion about these relationships as provisional because they did not have data for all the *DRB1* lineages. More noncoding sequence data from *DRB1* alleles is required to help unravel the patterns of diversification.

Recombination

Inter- or intralocus recombination events have been used to explain the generation of some *Mhc* alleles (e.g., Lee et al. 1990; Belich et al. 1992; Watkins et al. 1992; McAdam et al. 1994; Andersson and Mikko 1995; Parham et al. 1995; Parham and Ohta 1996). Convergent evolution also occurs at some sites, for example at codon 86 in human *DRB1* (see Titus-Trachtenberg et al. 1994). Analysis of the *DRB* sequence data using the program Reticulate (Jakobsen and Easteal 1996) revealed a large amount of incompatibility within exon 2, especially for the alpha helix region (data not shown). This is suggestive of recombination or convergence events (Jakobsen and Easteal 1996), which would account for the discor-

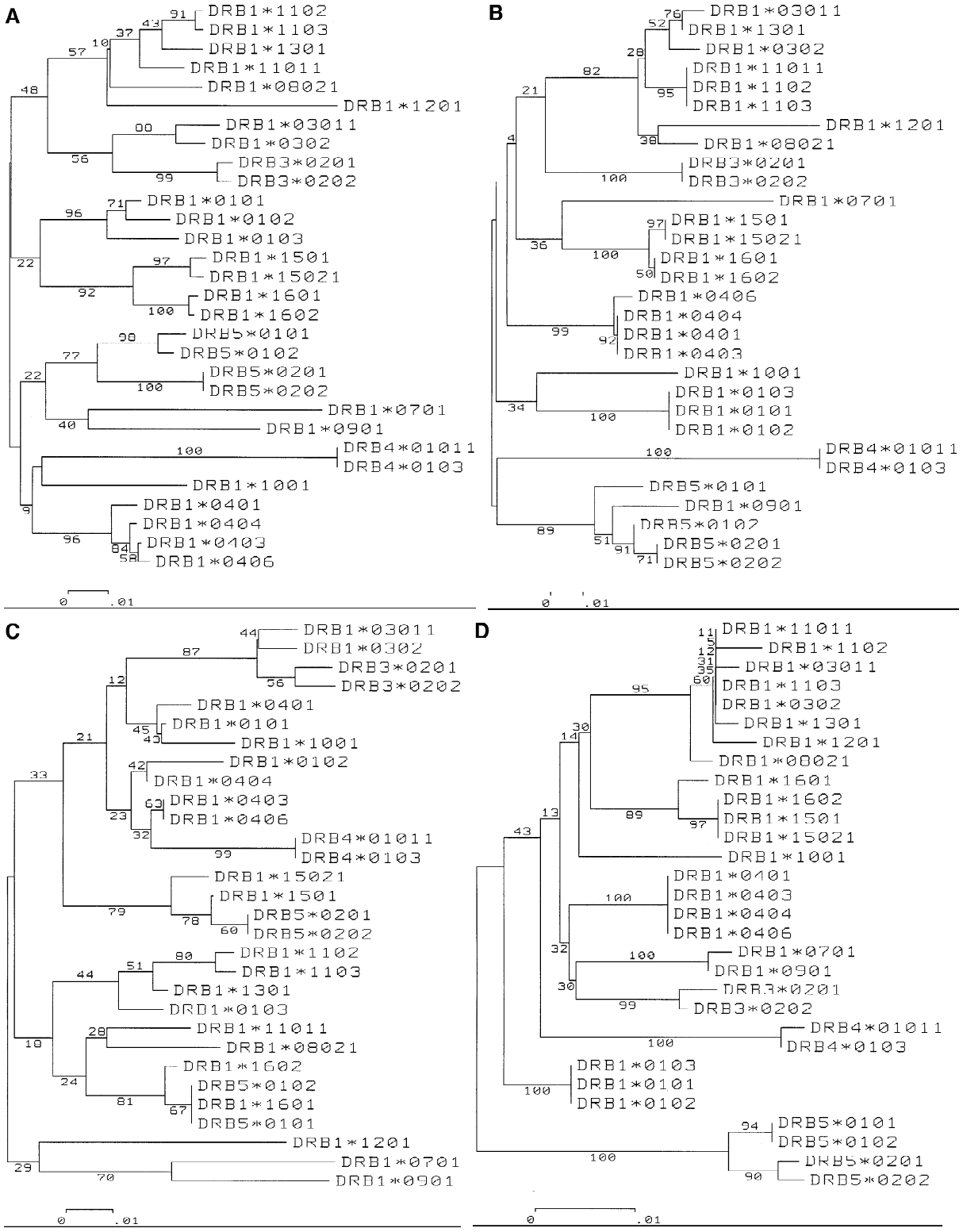

Fig. 2. Genealogies for different parts of the *DRB* gene based on the neighbor-joining distance method. Both synonymous and nonsynonymous substitutions (with Jukes-Cantor correction) were included, and the phylogenies are unrooted. The percentage of bootstrap support (based upon 100 replications) for each branch is shown. **A** Exon 2 as a whole (90 codons). **B** Only the beta-pleated sheet region of exon 2 (50 codons). **C** Only the alpha helix region of exon 2 (40 codons). **D** exons 3–6 (143 codons).

dant and poorly supported sets of relationships in the alpha helix region compared with the rest of the coding region. Interestingly, the relationships of the *DRB1*03, 11,* and **13* alleles, which make up more than half of the described human *DRB1* alleles, are hardest to resolve (see Fig. 2). Understanding how their diversity was generated is an essential prerequisite to interpreting diversity in the context of human history.

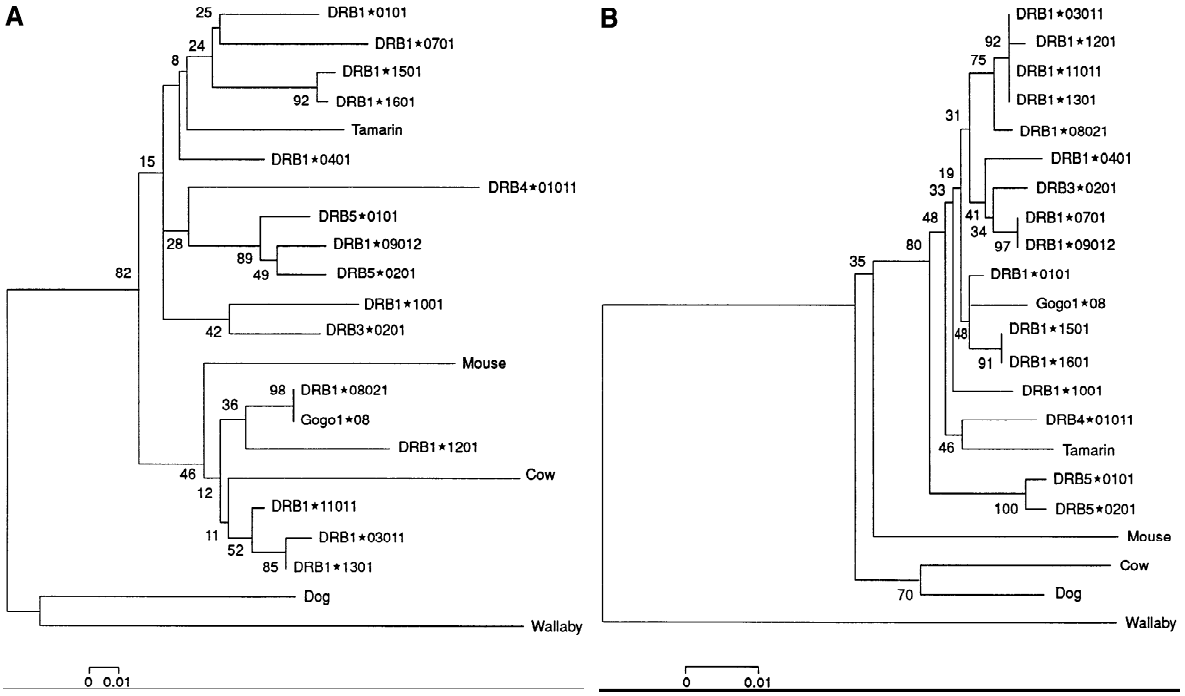

Fig. 3. Unrooted neighbor-joining phylogenies of selected mammalian *DRB* sequences based upon **(A)** the beta sheet region of exon 2 and **(B)** exons 3–6. Only nonsynonymous substitutions (with Jukes-Cantor correction) were used because of large evolutionary distances between some sequences. The percentage of bootstrap support (based upon 100 replications) for each branch is shown.

Table 3. Comparisons of substitution patterns within lineages[a]

	dR	dS	dR/dS ratio
Beta sheets	0.028 (± 0.004)	0.009 (± 0.002)	3.1
Alpha helix	0.062 (± 0.006)	0.070 (± 0.007)	0.9
Exon 3	0.002 (± 0.0005)	0.012 (± 0.002)	0.17
Exons 4–6	0.003 (± 0.0001)	0.004 (± 0.001)	0.75

[a] Mean number of nonsynonymous (dR) and synonymous (dS) substitutions per site (± SEM; calculated using the Jukes-Cantor correction) for different parts of the *DRB* coding region. The dR/dS ratio is also shown. For the beta sheet and alpha helix regions of exon 2 both ARS and non-ARS are included. Note that within a lineage alleles generally differ most from each other by substitutions in the alpha helix region. Lineages are as defined in Table 2

Klein and O'hUigin (1995) and O'hUigin (1995) have argued that intralocus recombination in the *Mhc* has often been uncritically applied and is inconsistent with known mechanisms of recombination. However, a cDNA-mediated recombination mechanism, similar to that suggested to be involved in interlocus exchange between mouse class I *Mhc* loci (Pease et al. 1993), could permit shuffling of small segments of exon 2 between *DRB1* alleles. Such a mechanism may be amenable to investigation in vitro using the polymerase chain reaction. Regardless of the extent to which recombination has played a role in generating allelic diversity, the observed complex patterns of nucleotide substitutions in exon 2 of *DRB1* (Fig. 2, Tables 2 and 3) necessitates more careful analyses than those of Ayala et al. (1994).

Timing of Diversification

Estimation of the age of alleles was central to inferring the population size in Ayala et al.'s (1994) analysis. However, the different rates of substitution for different parts of the *DRB* genes (see Table 2) and variation in rates between lineages and alleles (Fig. 2) make it very difficult to determine times of lineage or allele divergence. The occurrence of similar amino acid motifs in other primate species has been used to infer that some of the lineages arose early in primate evolution (Trtkova et al. 1993; Figueroa et al. 1994). As noted above, similarities based upon comparisons of exon 2 sequences only may be attributable to convergence and/or recombination rather than reflect historical relationships. Based on examination of intron sequences, Satta et al. (1996) also concluded that the prosimian *DRB* lineages are not orthologous to those in higher primates. Until complete *DRB1* sequences are examined from a wider range of primates the ages of lineages remain uncertain.

Even though lineages may be shared between species, the turnover rate of individual alleles can probably be quite rapid (Gyllensten et al. 1996), and this is reflected in the observation that no alleles have been found which are shared between humans and their closest primate relatives (Hughes and Hughes 1995; Parham and Ohta 1996). Relatively recent origins for individual alleles is also supported by the fact that few nucleotide differences occur between alleles when exons 3–6 are examined (see Fig. 2D).

Ayala et al. (1994) inferred the ages of alleles using a

UPGMA dendrogram and a single substitution rate estimate derived from Satta et al. (1993). As we have discussed above, rate variation within and between alleles make such applications inappropriate and unreliable.

Coalescence

The structure of the allelic genealogy and age of alleles were central to the major conclusion of Ayala and colleagues (Klein et al. 1993; Ayala et al. 1994; Ayala 1995; Ayala and Escalante 1996). On the basis of what they considered a slow evolutionary process with a clearly defined branching structure, they utilized coalescence theory to determine the population size necessary to maintain a large number of alleles. They concluded that a long-term effective population size of at least 100,000 was necessary to maintain the observed diversity of human *DRB1* alleles, an estimate 10-fold higher than that derived from mtDNA. Assuming the rate of growth of our ancestral population(s) was low, the founding population size would also have to be about 100,000 (see Ayala et al. 1994; Ayala 1996). However, the range of population sizes which could permit a large number of alleles to be maintained is broad (Takahata 1993; Ayala et al. 1994; Ayala and Escalante 1996), and the lower bound of these values is compatible with population estimates derived from mtDNA data (see Rogers and Jorde 1995; Ayala 1996; Erlich et al. 1996).

In addition, there are problems associated with the use of coalescence theory to examine *DRB* allelic history. As noted above, use of just exon 2 data to estimate the ages of lineages and of alleles is problematical and can generate unreliable estimates of population size. Rapidly expanding populations can shorten coalescence times (Griffiths and Tavare 1994), and the poorly supported internal branching structure of the *DRB* allelic genealogies (see Fig. 2) could be a result of rapid diversification (see Takahata 1993). Ayala et al.'s (1994) representation of *DRB1* allelic genealogy gives the impression of well-resolved and regular evolution and undue confidence in the pattern and timing of diversification. In contrast to the largely neutral, nonrecombining pattern of sequence polymorphism identified with human mtDNA, *Mhc* alleles are subject to positive selection and recombination (see above), features that complicate the application of coalescence theory to phylogenetic reconstruction.

Genetic Diversity and Environmental Heterogeneity

Large population sizes and long periods of time are one means of generating and maintaining high levels of genetic diversity, but other factors can also contribute. During rapid population expansion, which has occurred in human populations during the last 10,000 years, incorporation of new alleles is probable (Takahata 1993). Frequent local extinctions and recolonizations can also lead to high levels of diversity (Takahata 1994), particularly if there is environmental heterogeneity. For the *Mhc* this is likely when it is considered that during their recent history, human populations have not only moved into a large range of new environments but also domesticated and came into much closer associations with a range of animals. These changes have brought human populations into contact with a diverse range of new potentially pathogenic organisms. Different environments have their own unique parasite communities (Cameron 1956), parasites have adopted humans as new hosts following domestication (Hubbert et al. 1975), and high population densities have permitted the recent establishment of new diseases in society (McNeill 1976; Stannard 1993). These factors provided new selective pressures on human populations. The fact that *Mhc* diversity is much higher in humans (and some rodents) than other mammals can be expected on the basis of the former's wider environmental range. A correlation between environmental factors, parasitism, and *Mhc* variability in mole rats has been proposed by Nevo and Beiles (1992).

The central role of the *Mhc* is to aid in the detection of potential pathogens. To be an effective system the *Mhc* loci must be able to bind a broad range of antigenic peptides, and this is achieved by having high levels of allelic diversity. Klein and O'hUigin (1994) argued that allele evolution is slow, driven primarily by point mutation, so in their view the observed high levels of allelic diversity are not a consequence of recent changes in human ecology. Under this hypothesis, alleles of apparently recent origin should be uncommon. This does not appear to be the case. Increasingly, novel alleles are being reported for *DRB* as well as other loci, in, for example, West Africa (Hill et al. 1992), South America (Belich et al. 1992; Watkins et al. 1992; Titus-Trachtenberg et al. 1994; Garber et al. 1995), Australia (Gao and Serjeantson 1992; Lester et al. 1995), Asia (Lee 1993), and Oceania (Gao et al. 1992a,b), as well as in Europe (e.g., Anholts et al. 1995). While novel alleles may not necessarily be of recent origin, there is evidence (Belich et al. 1992; Watkins et al. 1992) that some have arisen within the recent past (see also Titus-Trachtenberg et al. 1994; Garber et al. 1995; Erlich et al. 1996; Gyllensten et al. 1996). Shuffling of amino acid motifs as a consequence of recombination offers the potential for rapid generation of alleles with new binding potentials.

Although it can be very difficult to link specific alleles to interactions with specific parasites or diseases, we interpret the observation that novel alleles appear to have evolved in different environments as support for the hypothesis that parasite-mediated selection can influence genetic diversity (see also Belich et al. 1992; Nevo and Beiles 1992; Watkins et al. 1992; Parham and Ohta 1996).

Conclusion

In their analyses, Ayala and colleagues (Klein et al 1993; Ayala et al. 1994; Ayala 1995; Ayala and Escalante 1996) considered *DRB1* allele evolution to be a slow and regular process relatively unaffected by factors such as recombination, convergence, different selective pressures within the molecule, or variation in rates of allele evolution. We have not found this to be the case and our analyses call into doubt their conclusions about diversity and population size (see also Erlich et al. 1996). Ayala et al. (1994) only analyzed exon 2 of *DRB1* and did not evaluate whether this was a reliable indicator of the evolutionary history of the alleles, despite Gyllensten et al.'s (1991) earlier indication of the phylogenetic complexity of this exon. The confidence Ayala and colleagues (Klein et al. 1993; Ayala et al. 1994; Ayala 1995; Ayala and Escalante 1996) had in their allelic genealogy is undermined by the lack of phylogenetic stability, irregular rates of substitutions, and unusual patterns of change in exon 2 (see Fig. 2).

Use of a simple UPGMA dendrogram to describe *DRB* allele evolution (Ayala et al. 1994) is inappropriate and misleading since it assumes equal rates of substitution. This in turn leads to unreliable estimates of allele ages and, consequently, biases calculation of population sizes. In our view, analyses of *DRB* allelic diversity do not present unequivocal evidence against the existence of relatively small ancestral human population sizes. Allelic diversification in the *Mhc* is a dynamic process and is likely to be influenced by factors in addition to population size.

High diversity at some nuclear loci still contrasts with the limited variation found in human mtDNA (see Takahata 1995; Herbert and Easteal 1996), so the doubts we (and others) have about some interpretations of *Mhc* allelic diversity do not resolve the issue of early human population sizes. As with earlier debates over human mtDNA population data, the patterns and processes involved in generating *DRB* diversity highlight the need to analyze genetic data by taking into account relevant features of the structure and function of the locus, as well as more realistic population models (see Takahata 1995). The analysis and interpretation of diversity at *DRB* and other *Mhc* loci is not a simple procedure and the confounding influences of selection, recombination, and convergence in nonequilibrium populations must be considered more carefully before interpreting the observed pattern of diversity. The complex history of some *Mhc*, as well as other highly variable loci, should caution investigators to beware of simplistic assertions about the evolution of alleles and populations.

Despite the complexity of their evolution *Mhc* loci have proven very useful for the reconstruction of aspects of the genetic history of human populations (see, for example, Belich et al. 1992; Gao et al. 1992b; Watkins et al. 1992; Titus-Trachtenberg et al. 1994; Takahata 1995), and have also enhanced our understanding of multifactorial inheritance and molecular interactions (see Stern et al. 1994; Parham and Ohta 1996).

Acknowledgments. We thank David Penny and two reviewers for constructive comments and discussion. John Hunt provided access to the program MEGA. R.E.H. was supported initially by an Alfred P. Sloan Foundation Fellowship in Molecular Evolution and subsequently by a New Zealand Science and Technology Fellowship.

References

Andersson G, Larhammar D, Widmark E, Servenius B, Peterson PA, Rask L (1987) Class II genes of the human major histocompatibility complex. Organization and evolutionary relationships of the DRb genes. J Biol Chem 262:8748–8758

Andersson L, Mikko S (1995) Generation of MHC class II diversity by intra- and intergenic recombination. Immunol Rev 143:5–12

Anholts JD, Verduyn W, Parlevliet A, Doxiadis II, D'Amaro J, Giphart MJ, Persijn GG, Schreuder GM (1995) Irregular polymerase chain reaction-sequence-specific oligonucleotide hybridization patterns reveal seven new HLA-DRB1 alleles related to DR2, DR3, DR6, DR8, and DR11. Implications for sequence-specific priming. Hum Immunol 42:15–22

Ayala FJ (1995) The myth of Eve: molecular biology and human origins. Science 270:1930–1936

Ayala FJ (1996) HLA sequence polymorphism and the origin of humans. Science 274:1554

Ayala FJ, Escalante AA (1996) The evolution of human populations: a molecular perspective. Mol Phylog Evol 5:188–201

Ayala FJ, Escalante A, O'hUigin C, Klein J (1994) Molecular genetics of speciation and human origins. Proc Natl Acad Sci USA 91:6787–6794

Belich MP, Madrigal JA, Hildebrand WH, Zemmour J, Williams RC, Luz R, Petzl-Erler ML, Parham P (1992) Unusual HLA-B alleles in two tribes of Brazilian Indians. Nature 357:326–329

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 364:33–39

Cameron TWM (1956) Parasites and parasitism. Methuen, London

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31–36

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, NJ

Erlich HA, Bergstrom TF, Stoneking M, Gyllensten U (1996) HLA sequence polymorphism and the origin of humans. Science 274: 1552–1554

Felsenstein J (1991) PHYLIP manual, version 3.4. University of Washington, Seattle

Figueroa F, O'hUigin C, Tichy H, Klein J (1994) The origin of the primate *Mhc-DRB* genes and allelic lineages as deduced from the study of prosimians. J Immunol 152:4455–4465

Fisher RA (1930) The genetic theory of natural selection. Clarendon Press, Oxford

Gao X, Serjeantson S (1992) Twelve HLA-DR6-related DRB1 alleles and associated DR DQ haplotypes in traditional Australians and other populations of Asia-Oceania. Eur J Immunogen 19:263–272

Gao X, Bhatia K, Trent RJ, Serjeantson SW (1992a) HLA-DR,DQ nucleotide sequence polymorphisms in five Melanesian populations. Tissue Antigens 40:31–37

Gao X, Zimmet P, Serjeantson SW (1992b) HLA-DRDQ sequence polymorphisms in Polynesians, Micronesians, and Javanese. Hum Immunol 34:153–161

Garber TL, Butler LM, Trachtenberg EA, Erlich HA, Rickards O, De Stefano G, Watkins DI (1995) HLA-B alleles of the Cayapa of Ecuador: new B39 and B15 alleles. Immunogenetics 42:19–27

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. Proc Natl Acad Sci USA 92:6723–6727

Griffiths RC, Tavare S (1994) Sampling theory for neutral alleles in a varying environment. Philos Trans R Soc Lond Biol 344:403–410

Gyllensten UB, Sundvall M, Erlich HA (1991) Allelic diversity is generated by intra-exon sequence exchange at the DRB1 locus of primates. Proc Natl Acad Sci USA 88:3686–3690

Gyllensten U, Bergström T, Josefsson A, Sundvall M, Erlich HA (1996) Rapid allelic diversification and intensified selection at antigen recognition sites of Mhc class II DPB1 locus during hominoid evolution. Tissue Antigens 47:212–221

Hartl DL, Clark AG (1989) Principles of population genetics. 2nd ed. Sinauer, Sunderland, MA

Herbert G, Easteal S (1996) Relative rates of nuclear DNA evolution in humans and old world monkey lineages. Mol Biol Evol 13:1054–1057

Hill AVS, Allsopp CEM, Kwiatkowski D, Taylor TE, Yates SNR, Anstey NM, Wirima JJ, Brewster DR, McMichael AJ, Molyneux ME, Greenwood BM (1992) Extensive genetic diversity in the HLA class II region of Africans, with a focally predominant allele, DRB*1304. Proc Natl Acad Sci USA 89:2277–2281

Hubbert WT, McCulloch WF, Schnurrenberger PR (1975) Diseases transmitted from animals to man. 6th ed. Charles C Thomas, Springfield, IL

Hughes AL, Hughes MK (1995) Natural selection on the peptide-binding regions of major histocompatibility complex molecules. Immunogenetics 42:233–243

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170

Hughes AL, Nei M (1989) Pattern of nucleotide substitution at major histocompatibility complex class II loci reveals overdominant selection. Proc Natl Acad Sci USA 86:958–962

Hughes AL, Nei M (1992) Maintenance of MHC polymorphism. Nature 355:402–403

Hughes AL, Hughes MK, Howell CY, Nei M (1994) Natural selection at the class II major histocompatibility complex loci of mammals. Philos Trans R Soc Lond Biol 346:359–367

Jakobsen IB, Easteal S (1996) A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. Comput Appl Biosci 12:291–295

Klein J (1986) Natural history of the major histocompatibility complex. John Wiley, New York

Klein J, O'hUigin C (1994) MHC polymorphism and parasites. Philos Trans R Soc Lond Biol 346:351–358

Klein J, O'hUigin C (1995) Class II *B Mhc* motifs in an evolutionary perspective. Immunol Rev 143:89–112

Klein J, Satta Y, O'hUigin C, Mayer WE, Takahata N (1992) Evolution of the primate DRB region. In: Kimiyoshi T, Aizawa M, Sasazuki T (eds) Proceedings of the 11th international histocompatibility workshop, vol 2. pp 45–56

Klein J, Takahata N, Ayala FJ (1993) MHC polymorphism and human origins. Sci Am 269:46–51

Kumar S, Tamura K, Nei M (1993) MEGA: molecular evolution genetics analysis, version 1.0. Pennsylvania State University, University Park, PA

Lee WK (1993) DR6 in Koreans. DR11 frequently acts as a recipient gene to create DR13 alleles. Hum Immunol 37:229–236

Lee WK, Johnson AH, Hurley CK (1990) Two divergent routes of evolution gave rise to the DRw13 haplotypes. J Immunol 145: 3119–3125

Lento GM, Hickson RE, Chambers GK, Penny D (1995) Use of spectral analysis to test hypotheses on the origin of pinnipeds. Mol Biol Evol 12:28–52

Lester S, Cassidy S, Humphreys I, Bennett G, Hurley CK, Boettcher B, McCluskey J (1995) Evolution in HLA-DRB1 and major histocompatibility complex class II haplotypes of Australian aborigines. Definitions of a new DRB1 allele and distribution of DRB1 gene frequencies. Hum Immunol 42:154–160

Marsh SG, Bodmer JG (1995) HLA class II region nucleotide sequences 1995. Tissue Antigens 46:258–280

McAdam SN, Boyson JE, Liu X, Garber TL, Hughes AL, Bontrop RE, Watkins DI (1994) A uniquely high level of recombination at the *HLA-B* locus. Proc Natl Acad Sci USA 91:5893–5897

McNeill WH (1976) Plagues and people. Anchor Press, New York

Nei M, Graur D (1984) Extent of protein polymorphism and the neutral mutation theory. Evol Biol 17:73–118

Nei M, Roychoudhury AK (1993) Evolutionary relationships of human populations on a global scale. Mol Biol Evol 10:927–943

Nevo E, Beiles A (1992) Selection for class II *Mhc* heterozygosity by parasites in subterranean mole rats. Experentia 48:512–515

O'hUigin C (1995) Quantifying the degree of convergence in primate MHC-DRB genes. Immunol Rev 143:123–140

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

Parham P, Adams EJ, Arnett KL (1995) The origins of HLA-A,B,C polymorphism. Immunol Rev 143:141–180

Pease LR, Horton RM, Pullen JK, Yun TJ (1993) Unusual mutation clusters provide insight into class I gene conversion mechanisms. Mol Cell Biol 13:4374–4381

Rogers AR, Jorde LB (1995) Genetic evidence on modern human origins. Human Biol 67:1–36

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Satta Y, O'hUigin C, Takahata N, Klein J (1993) The synonymous substitution rate of the major histocompatibility complex loci in primates. Proc Natl Acad Sci USA 90:7480–7484

Satta Y, O'hUigin C, Takahata N, Klein J (1994) Intensity of natural selection at the major histocompatibility complex loci. Proc Natl Acad Sci USA 91:7184–7188

Satta Y, Mayer WE, Klein J (1996) Evolutionary relationship of *HLA-DRB* genes inferred from intron sequences. J Mol Evol 42:648–657

Stannard DE (1993) Disease, human migration, and history. In: Kiple KF (ed) The Cambridge world history of human disease. Cambridge University Press, Cambridge, pp 35–42

Steel MA, Penny D (1993) Distributions of tree comparison metrics—some new results. Syst Biol 42:126–141

Stern LJ, Brown JH, Jardetsky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC (1994) Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature 368:215–221

Swofford DL (1993) PAUP: phylogenetic analysis using parsimony, version 3.1. Illinois Natural History Survey, Champaign

Takahata N (1993) Allelic genealogy and human evolution. Mol Biol Evol 10:2–22

Takahata N (1994) Repeated failures that led to the eventual success in human evolution. Mol Biol Evol 11:803–805

Takahata N (1995) A genetic perspective on the origin and history of humans. Annu Rev Ecol Syst 26:343–372

Titus-Trachtenberg EA, Rickards O, De Stefano GF, Erlich HA (1994) Analysis of HLA class II haplotypes in the Cayapa Indians of Ecuador: a novel DRB1 allele reveals evidence for convergent evolution and balancing selection at position 86. Am J Hum Genet 55:160–167

Trtkova K, Kupfermann H, Grahovac B, Mayer WE, O'hUigin C, Tichy H, Bontrop R, Klein J (1993) Mhc-DRB genes of platyrrhine primates. Immunogenetics 38:210–222

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507

Watkins DL, McAdam SN, Liu X, Strang CR, Milford EL, Levine CG, Garber TL, Dogon AL, Lord CI, Ghim SH, Troup GM, Hughes AL, Letvin NL (1992) New recombinant HLA-B alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. Nature 357:329–333