Genet. Res., Camb. (2003), **82**, *pp.* 89–99. *With 3 figures.* © 2003 Cambridge University Press
DOI: 10.1017/S0016672303006347 Printed in the United Kingdom

Relative roles of mutation and recombination in generating allelic polymorphism at an MHC class II locus in *Peromyscus maniculatus*

ADAM D. RICHMAN[1]*, L. GERARDO HERRERA[2], DEANNA NASH[1] AND MIKKEL H. SCHIERUP[3]
[1] *Plant Sciences Department, MSU Bozeman, Bozeman, MT 59717, USA*
[2] *Department Zoology, Universidad Nacional Autónoma de México, Instituto de Biología, A.P. 70-153, 04510 México D.F., México*
[3] *Bioinformatics Research Center (BiRC), Department of Ecology and Genetics, University of Aarhus, Ny Munkegade, Building 540, DK-8000 Aarhus C, Denmark*

(*Received 19 December 2002 and in revised form 3 April 2003*)

Summary

The MHC class II loci encoding cell surface antigens exhibit extremely high allelic polymorphism. There is considerable uncertainty in the literature over the relative roles of recombination and *de novo* mutation in generating this diversity. We studied class II sequence diversity and allelic polymorphism in two populations of *Peromyscus maniculatus*, which are among the most widespread and abundant mammals of North America. We find that intragenic recombination (or gene conversion) has been the predominant mode for the generation of allelic polymorphism in this species, with the amount of population recombination per base pair exceeding mutation by at least an order of magnitude during the history of the sample. Despite this, patchwork motifs of sites with high linkage disequilibrium are observed. This does not appear to be consistent with the much larger amount of recombination versus mutation in the history of the sample, unless the recombination rate is highly non-uniform over the sequence or selection maintains certain sites in linkage disequilibrium. We conclude that selection is most likely to be responsible for preserving sequence motifs in the presence of abundant recombination.

1. Introduction

Extreme polymorphism at class II loci responsible for antigen presentation is believed to be maintained by multi-allelic overdominant selection. Overdominance is probably mediated by resistance to pathogens (Carrington *et al*., 1999; Paterson *et al*., 1998; Thursz *et al*., 1997), presumably owing to an enhanced capacity for antigen presentation in heterozygotes relative to homozygotes. Consistent with this model, most of the allelic sequence variation at class II genes is found within the second exon, which encodes the entirety of the peptide binding region (PBR) responsible for antigen presentation (Brown *et al*., 1993). Moreover, the sites of the PBR show a significant relative excess of non-silent (amino acid) substitutions, whereas other sites in exon 2 do not, indicating diversifying selection is restricted to sites responsible for antigen binding (Hughes & Yeager, 1998).

There is considerable uncertainty about the relative importance of intragenic recombination (or homologous gene conversion) and *de novo* mutation in contributing to observed allelic diversity. In what follows, we do not differentiate between gene conversion and recombination; although they have different mechanisms, their effects on sequence evolution of small fragments (<1000 base pairs) are expected to be similar. Although only mutation can create new site variation, recombination can generate novel sequences, some of which might present a different spectrum of antigens and would therefore be favoured by diversifying selection. If intragenic recombination is sufficiently frequent, it might be a major source of new allelic specificities once sufficient variation has accumulated. However, although many authors have argued for an important role for intragenic

* Corresponding author.

https://doi.org/10.1017/S0016672303006347 Published online by Cambridge University Press

recombination at antigen presenting genes of the MHC (Bergstrom *et al*., 1998; Gyllensten *et al*., 1991; She *et al*., 1990), based on observation of a patchwork of shared sequence polymorphism (often called sequence 'motifs') within and among species, others have argued that mutation and selective convergence cannot be ignored as the cause of shared sequence variation (Takahata, 1995; Takahata *et al*., 1992).

The debate about the relative importance of recombination and mutation in generating MHC diversity has continued largely because of the difficulty of assessing the relative contributions of recurrent mutation and recombination. Recent developments in the modelling of recombination using coalescent theory and a full maximum likelihood model (Fearnhead & Donnelly, 2001) now permit a quantitative treatment of this problem. Although a full likelihood model is currently too slow for analysis of large data sets, Hudson (2001) has shown that a composite likelihood model based on pairwise sample configurations is a very efficient and precise approximation. Recently, Hudson's method has been extended to the case of recurrent mutation (McVean *et al*., 2002). This is particularly relevant to the problem of estimating the recombination rate for genes under diversifying selection, in which recurrent substitution is expected to be frequent.

A contemporaneous development has been technical advances in sequence acquisition for Class I and II genes of the MHC, which have permitted a dramatic expansion in the number and scale of studies of these genes in diverse taxa. We suggest that these theoretical and technical developments create a new opportunity for revisiting old questions and the framing of new ones regarding the evolution of sequence and allelic diversity at the MHC. For example, because the amount of population recombination and mutation are functions of effective population size, the relative importance of these processes in generating sequence diversity and allelic polymorphism might be particularly clear in taxa with large effective population sizes, such as might be expected for many rodent species.

Here, we present an analysis of population recombination and mutation in two population samples of variation at the Class II *EB* locus in the Deer Mouse (*Peromyscus maniculatus*), one of the most abundant and widespread mammals of North America. In a previous study of a single population sample, we found unparalleled levels of pairwise sequence diversity in this species compared to samples of Class II MHC diversity in other taxa (Richman *et al*., 2001). In the present study, we find extensive polymorphism both within and between population samples. We show that intragenic recombination has been the predominant mode for the generation of extreme allelic polymorphism at this locus.

2. Methods

(i) *Sampling*

P. maniculatus individuals were collected at single mainland sites in northern Baja California (San Quintin, Baja California Norte, Mexico) and in southern California (Oceanside County, California, USA). Animals were live trapped, anaesthetized and killed. Spleen and liver samples were collected and immediately frozen in liquid nitrogen for transport back to the laboratory.

(ii) *Sequence acquisition*

MHC Class II β sequences were obtained by RT-PCR from spleen cDNAs using locus specific primers as described elsewhere (Richman *et al*., 2002). The sequence heterogeneity of amplification products was assessed using denaturing gradient gel electrophoresis (DGGE). Double stranded PCR products exhibit a sharp slow down in their rate of migration as they undergo partial denaturation at a position along the gradient determined by their primary sequence, forming distinct bands. The different bands were excised from the gel, reamplified and directly sequenced using the original amplification primers (Richman *et al*., 2001). Sequences reported here are based on results of both forward and reverse sequencing reactions. For instances in which a sequence was recovered from a single individual in the sample, multiple (and therefore independent) isolations of each sequencing template were analysed.

(iii) *Analyses of allelic polymorphism*

Actual and expected heterozygosities, and the standard measure of population structure based on allele frequencies (F_{ST}), were obtained using Arlequin (Schneider *et al*., 2001). Exact tests for deviation from random mating (Rousset & Raymond, 1995) and population differentiation of allele frequencies (Goudet *et al*., 1996) were also performed using Arlequin.

(iv) *Analyses of sequence diversity*

The relative amounts of substitution at silent (d_s) and non-silent (d_n) sites and their standard errors were estimated using the evolutionary pathways method (Nei & Gojobori, 1986), implemented in MEGA2 (Kumar *et al*., 2001).

In counting the number of silent and non-silent substitutions between a pair of aligned coding sequences, the evolutionary pathways method considers all possible evolutionary pathways (excluding those that include termination codons) leading from one codon to another. For some comparisons, there are

https://doi.org/10.1017/S0016672303006347 Published online by Cambridge University Press

multiple possible evolutionary pathways, which were weighted equally. The evolutionary pathway method makes fewer assumptions than other methods and is therefore expected to provide conservative (minimum) estimates of numbers of substitutions, which again provides a conservative estimate of d_n/d_s against the hypothesis $d_n/d_s > 1$ (Nei & Kumar, 2000).

The partitioning of molecular sequence variation within and among populations (F_{ST}) was estimated using Arlequin (Schneider *et al*., 2001). The hypothesis of no differentiation among samples in site polymorphism was assessed using bootstrap resampling.

(v) *Patterns of linkage disequilibrium*

Linkage disequilibrium between each pair of informative sites was estimated and tested by Fisher's exact test using DnaSP 3.51 (Rozas & Rozas, 1999). These analyses were summarized as a matrix of significant linkage disequilibrium values with significance levels $P < 0{\cdot}01$ and $P < 0{\cdot}001$.

(vi) *Analyses of recombination*

In the presence of recombination, pairwise site compatibility and linkage disequilibrium are expected to decrease with the distance between sites. This expectation forms the basis for a statistical test for the presence of recombination (Awadalla & Charlesworth, 1999). We used the program r2 (http://www.brics.dk/~compbio/r2/) to estimate the correlation of pairwise site compatibility with distance between nucleotide sites. Two segregating sites are compatible if they both fit a single genealogical scenario, free of homoplasies. Incompatibility of two sites therefore indicates recombination and/or recurrent mutations (homoplasy). Recombination is expected to exchange sets of contiguous sites, generating blocks of compatible sites that are incompatible with adjacent blocks. By contrast, site incompatibilities generated by repeated mutations are not expected to be clustered together.

The amount of population recombination ρ in the history of a set of aligned sequences was estimated using a composite likelihood estimation method (Hudson, 2001), implemented in the program LDhat (McVean *et al*., 2002). The value of the composite likelihood estimate of recombination (ρ_{CLE}) is estimated conditioned on the estimate of θ, obtained using a finite-series version of the Watterson estimator, also estimated using LDhat. The method assumes a constant mutation rate and, for this reason, analysis is restricted to polymorphic sites with two different segregating sites. A limitation of the method is that confidence levels of the estimate cannot be obtained, but ρ_{CLE} can be compared statistically to the hypothesis of no recombination ($\rho = 0$) using a permutation test. Despite simplifying assumptions of the likelihood model, the method appears to be less prone to false positives than other tests for the presence of recombination because it explicitly considers the probability of recurrent substitutions, which can contribute to pairwise incompatibility.

In order to assess the generality of our results, we estimated ρ_{CLE} and θ for several homologous Class II data sets, specifically, a DR (E-type) β locus in humans (*HLA DRB1* (Bergstrom *et al*., 1998)), White Tailed Deer (*Odocoileus virginianus*) (*Odvi DRB1* (Van Den Bussche *et al*., 1999)) and mouse (*Mus domesticus*) (*Mudo EB* (Edwards *et al*., 1997)). Analyses using LDhat were performed as described above.

As a check of the robustness of our estimates of recombination, we examined the performance of LDhat with regard to sequences evolving under symmetric balancing selection, using computer simulation. Sequences were simulated under the model of balancing selection and recombination described by Schierup *et al*. (2001). This model assumes that sequences are assigned a specificity determined by the left endpoint of the sequence. A single representative of each of a number of specificities is assumed to have been sampled and the coalescent with recombination process is simulated, where only sequences with the same specificity can coalesce. If recombination occurs, the ancestor to the left part of the sequence retains its specificity whereas the ancestor to the right part of the sequence is assigned a random specificity among the possible specificities (i.e. it is assumed that selection maintains all specificities in the population at equal frequencies). Sequences may change their associated specificity at a specificity turnover rate T, eventually allowing the whole sequence sample to coalesce. The turnover rate to new allelic specificities was fixed at 0·1, with the number of different specificities held constant at 20. These parameters imply that the persistence of allelic specificities exceeds that of (selectively neutral) sequence changes within a specificity approximately 1000-fold. Rates of recombination and mutation in the simulations were chosen to mimic the amounts of recombination and mutation estimated from the empirical data sets. This procedure yields an allelic genealogy with expected amounts of mutation and recombination along each branch, which is used to simulate a sample of sequences according to the Jukes–Cantor model. Simulated data sets were analysed using LDhat using the same settings used for empirical data sets.

Lastly, we examined the fit of ρ_{CLE} to the data by comparing the difference between ρ_{CLE} and estimates of ρ obtained from each pair of (polymorphic) sites included in the analysis, in order to identify pairs of sites that show a relative excess or deficit of linkage disequilibrium. We used this measure of the fit of the likelihood model, which assumes that the rate of

https://doi.org/10.1017/S0016672303006347 Published online by Cambridge University Press

```
                    1         11        21        31        41        51        61        71        81
                    |* *       |       * *    **   |       *   |       *   | **   *   * **  *   *  **  |**
OC14a          1    QVKSECHFYN GTQRVQFLDR YFYNREEYVR FDSD--VGEF IALTELGRPD AEYWNSQKDF LEDRRAVVDT ACRHNYRVSE -SFIVQRR
SQPM15e        2    .A.Y..... ...H..Y.V. ...I..... .........Y R.V..... .K...G..EI ..QK..EIE. V.....G..D ........
OC4a           1    .A.Y..... ...H..Y.V. .......... .........Y R.V..... .K...G..EI ..QK..EIE. V.....G..D ........
OC17a          3    .A.C..... ...S.RY.E. ........A. .......... R.VN..... .K...G.E.L ..QK..AIE. W..Y...GFD .......K
OC10a          1    .A.Y..... ......Y.V. .I........ .......... R.VN.....S .K...G.... M.QK..EIE. V.....G.FD ........
OC22a          3    .A.Y..... .....RY.V. .IH....N.. .......... .VV.....L. ..N......L ..QK..EIE. V.....GIFD .......T
SQPM20a        1    .A.Y..... ...S.RY.V. LI.....F.. .......... ..V.....GI ..HL.....L ..RL..EIE. V......GFD ........
OC19b          1    .A.C..... .....R..E. .......... .........Y REV.....GM ..NL.....L ..N...Q... V.....G.LD ........
SQPM26a        2    .G.C..... ...S.RY.E. HI.....FI. .........Y R.V.....GI ...Y...... ..QT..L..I Y.....G.LD ........
OC3a           2    .G.C..... ...S.RY.E. HI.....FM. .........Y R.V.....GI ...Y.....Y ..QT..S..I Y.....G.FD ........
SQPM23a        4    .......... .....RL.N. ........A. .........Y R.V......S ........EI ..N...A..N Y..Y..G.LD ........
SQPM18f        2    ...H...... .....RY... .I........ ....SD...Y .......... .......E.. ..QK.SR..N Y.....G.LD ........
SQPM2g         2    .A.C..... .....RY.E. HI.....FMH .......... ..V......I ..NY..R.EI ..QK..A..Y Y.....G.FD ........
OC20a          1    .A.C..... .....RY.E. HI.....FMH .......... .........I ..NY..R.EI ..QK..A..Y Y.....G.A. E.......
SQPM17c        4    .G........ .....R..E. .I......A. .........Y R.V..... .....G.... ..QK..Q..N Y.K..HG.FD ........
SQPM19a,OC1a   1,2  .L........ .....R..E. HIH.....A. .........Y R.V..... .....G...L ..QK..Q..N Y.....G... ........
OC3b           2    .L........ .....R..E. .IH....NL. .........Y R.V..... ..D..G...L ..QK..Q..N Y.....E..D ...L....
SQPM10b        1    .L........ .....R..E. .IH....NL. .........Y R.V..... ..D..G...L ..QK..Q..N Y.....G.FD ........
OC5a           3    .A.Y..... .....R..Y. .I.....FL. .........H ..V.....RQ ...A.....L ..RA..A..N Y.....E..D ...L....
SQPM3b         5    .A.Y..... .....R..Y. .I.....FL. .........H ..V.....RQ ...A.....L ..RA..A..N Y......GFD ........
OC5b           3    .A........ .....R.... FI......A. Y........KY ........A. .....G.E.L ..RA..A..N Y..Y..G.A. K.......
SQPM11f        5    .A........ .....R..N. .I......A. Y........KY .......... .....G.E.L ..RA..A..N Y..Y..G.A. K.......
OC16a          1    .A.Y..... .....R..E. HIH.....A. .........Y R.V..... .K...G...L ......Q..N Y.....G..D ...L...K
SQPM7g,OC2a    3,2  .A.Y..... .....R..E. RIH.....A. .......... R.V..... .K...G...L ......Q... V.....G.LD .......K
SQPM15b        5    ...Y..... .....LY.E. RIH.....A. .......... R.V..... .K...G...L ......S..N Y.....G.FD .......K
OC11a          2    ...H..... .......... .......F.. .......... R.V.....GI ..N..G...I .......... ......G.FD ........
SQPM25a,OC4b   1,3  ...H..... .....RY.E. HIH.....A. .......... .......... .....G..EI ......Q..N Y..Y..G.FD ........
SQPM26b,OC15a  2,3  .A.H..... .....RY.E. HIH.....A. ......L... .......... .....G...L ......Q..N Y..Y..G... ........
SQPM24b,OC8a   1,3  ...H..... .....RY.E. HIH.....A. .......... .......... .....G...L ......Q..N Y..Y..G... ........
```

Fig. 1. Inferred amino acid sequences of expressed Class II *EB* alleles obtained by RT-PCR from spleen cDNA for *Peromyscus maniculatus*. Sequence identifiers: OC, Oceanside, CA, USA; SQPM, San Quintin, BC Norte, Mexico. Starred positions indicate sites, which are expected to participate in antigen binding, based on crystallographic data for the homologous HLA DRB1*01 molecule (Brown *et al*., 1993).

recombination is the same for all sites, to investigate whether the amount of recombination varied across the sequence in a systematic fashion.

3. Results

(i) *Analyses of allelic polymorphism*

16 and 18 different sequences were obtained from samples of 22 and 20 individuals from the San Quintin (Genbank accessions AF312748-AF312764) and Oceanside (Genbank accessions AF516929-AF516946) samples, respectively (Fig. 1). A total of 29 different haplotypes were found, of which five were recovered in both samples. All the sequences obtained differed by at least one nonsynonymous substitution in the PBR, suggesting that they might also differ in their antigen binding properties.

There is no evidence for deviation from random mating with respect to allelic polymorphism at the *EB* locus in either population sample (San Quintin, $H_E = 0{\cdot}94$ vs $H_O = 0{\cdot}90$, N.S.; Oceanside, $H_E = 0{\cdot}95$ vs $H_O = 0{\cdot}89$, N.S.), an unsurprising result in view of the extensive polymorphism detected in these samples. Although a relative excess of heterozygotes is expected under a hypothesis of overdominant selection, extensive polymorphism means that most individuals are expected to be heterozygous even in the absence of such selection, resulting in limited statistical power.

A test of differentiation in haplotype diversity among samples finds significant evidence for population structure ($F_{ST} = 0{\cdot}061$, $P < 0{\cdot}001$). This significant result is due both to limited sharing of (five) haplotypes between samples and to the identity of the most frequent alleles differing between samples (Fig. 1). Although the small magnitude of F_{ST} suggests only very limited differentiation, this interpretation is not justified in the present case. Extensive polymorphism within samples necessarily implies that F_{ST} values can vary only within a narrow range.

(ii) *Analyses of sequence diversity*

Estimates of d_s and d_n indicate the maintenance of extensive sequence variation in *P. maniculatus* consistent with expectation for balancing selection (Table 1). Although a relative excess of nonsilent substitutions is expected under balancing selection (Hughes & Nei, 1988; Hughes & Nei, 1989), there is only moderate evidence for an increased rate of substitution at nonsynonymous versus synonymous sites in the PBR in the *P. maniculatus* sequences. This result appears to reflect saturation of the limited number of nonsynonymous sites at the PBR; the relative substitution rate using only recently diverged pairs of alleles yields an elevated rate of nonsynonymous substitution in *P. maniculatus* similar to those obtained for taxa with lower amounts of sequence variation, including mouse and human (Richman *et al*., 2001).

The molecular analysis of variance (AMOVA) finds little evidence for structuring of sequence variation among populations ($F_{ST} = 0{\cdot}015$, N.S.). In contrast to the relatively limited sharing of haplotypes among the samples, virtually every site polymorphism is recovered in both samples. Extensive sharing of sequence variation among samples is probably due to the age of allelic variation maintained at Class II loci (Takahata, 1990; Takahata & Nei, 1990), implying that most allelic sequence variation pre-dates the formation of population structure.

https://doi.org/10.1017/S0016672303006347 Published online by Cambridge University Press

Table 1. *Estimates of synonymous* (d_s) *and nonsynonymous substitutions* (d_n) *and their associated standard errors* (*in parentheses*) *for exon 2 E-type* β *sequences recovered from two* Peromyscus maniculatus *population samples for* (*1*) *codons of the peptide binding region* (*PBR*) *and* (*2*) *non-PBR codons. Estimates of* d_s *and* d_n *were obtained using the evolutionary pathways method* (*Nei & Gojobori, 1986*) *implemented in MEGA2* (*Kumar* et al.*, 2001*)

	PBR sites		Non-PBR sites	
	d_s	d_n	d_s	d_n
San Quintin	0·271 (0·091)	0·356 (0·048)	0·100 (0·022)	0·130 (0·019)
Oceanside	0·285 (0·099)	0·300 (0·066)	0·101 (0·045)	0·129 (0·029)

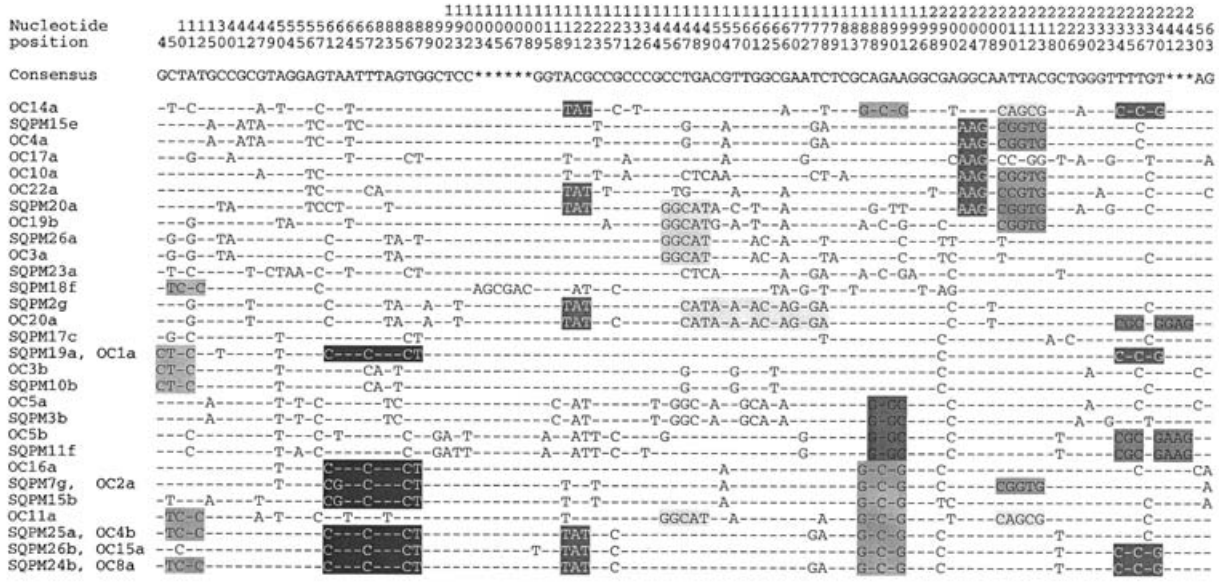

Fig. 2. Polymorphic nucleotide sites of expressed Class II *EB* alleles obtained by RT-PCR from spleen cDNA for *Peromyscus maniculatus*. Sequence identifiers: OC, Oceanside, CA, USA; SQPM, San Quintin, BC Norte, Mexico. Selected blocks of sites in linkage disequilibrium are differently coloured. The complex pattern of incompatibility among the different blocks suggests frequent recombination between blocks.

(iii) *Analyses of linkage disequilibrium and recombination*

Extensive sharing of site polymorphism among population samples, despite limited sharing of haplotypes, suggests extensive homoplasy and/or recombination has been important in the generation of allelic polymorphism. The inference that intragenic recombination rather than homoplasy is responsible for these results is supported by several analyses. First, there are multiple blocks of segregating sites that are mutually incompatible with at least one other block (Fig. 2); that is, these blocks cannot have arisen just once by mutation under a single genealogical scenario, indicating either recombination or homoplasy. Because these blocks involve three or more segregating sites, it is very unlikely these blocks arose independently by homoplasious (recurrent) substitutions.

Second, the inference that recombination has reshuffled these blocks is supported by a statistically significant negative correlation between pairwise site compatibility and the distance between sites, consistent with recombination (Table 2). The same result is found using the conventional *r2* measure of linkage disequilibrium (Table 2). Third, a plot of significant linkage disequilibrium values for informative sites shows a preponderance of significant values among adjacent sites (Fig. 3). This rapid decay of linkage disequilibrium with distance across the sequence suggests extensive recombination.

Although the preceding analyses provide ample evidence for extensive intragenic recombination, they do not provide an estimate of the amount of recombination. The likelihood method implemented in LDhat estimates the amount of recombination ρ_{CLE} for the data as $\geqslant 100$, the maximal amount evaluated. This estimate is significantly different from $\rho = 0$

https://doi.org/10.1017/S0016672303006347 Published online by Cambridge University Press

Table 2. *Statistical tests for recombination for four MHC Class II DR* (*E-type*) β *exon 2 data sets, obtained using the program r2* (*http://www.brics.dk/~compbio/r2/*). n *is the number of segregating sites used in the analysis, which is restricted to sites segregating for two different alleles, where the frequency of the rare allele equals or exceeds 0·20.* CM *and* r2 *are the Pearson correlations of pairwise compatibility and linkage disequilibrium with nucleotide distance, respectively. Significance levels:* *, P < *0·05;* **, P < *0·01*

Locus	*n*	*CM*	*r2*
Pema EB	24	−0·25**	−0·28**
Odvi DRB	24	−0·35**	−0·37**
Mudo EB	23	−0·11*	−0·16**
HLA DRB1	19	−0·45**	−0·31**

Table 3. *Estimates of* θ *and* ρ *for four MHC Class II DR* (*E-type*) β *exon 2 data sets, obtained using the program LDhat* (*McVean* et al., *2002*). n *is the number of segregating sites used in the analysis, which considers only those sites segregating for two different alleles, where the frequency of the rare allele equals or exceeds 0·10.* θ *is the Watterson estimate of the amount of population mutation* (*4*Nu)*;* ρ *is the estimated amount of population recombination* (*4*Nr) *for the* n *segregating sites.* *, P < *0·001. References for data sets analysed: Pema EB, this paper; Odvi DRB, Van Den Bussche* et al. (*1999*)*; Mudo EB, Edwards* et al. (*1997*)*, HLA DRB1, Bergstrom* et al. (*1998*)

Locus	*n*	*θ*	*ρ*
Pema EB	35	8·5	>100*
Odvi DRB	33	9·5	>100*
Mudo EB	33	10·9	>100*
HLA DRB1	30	8·7	>100*

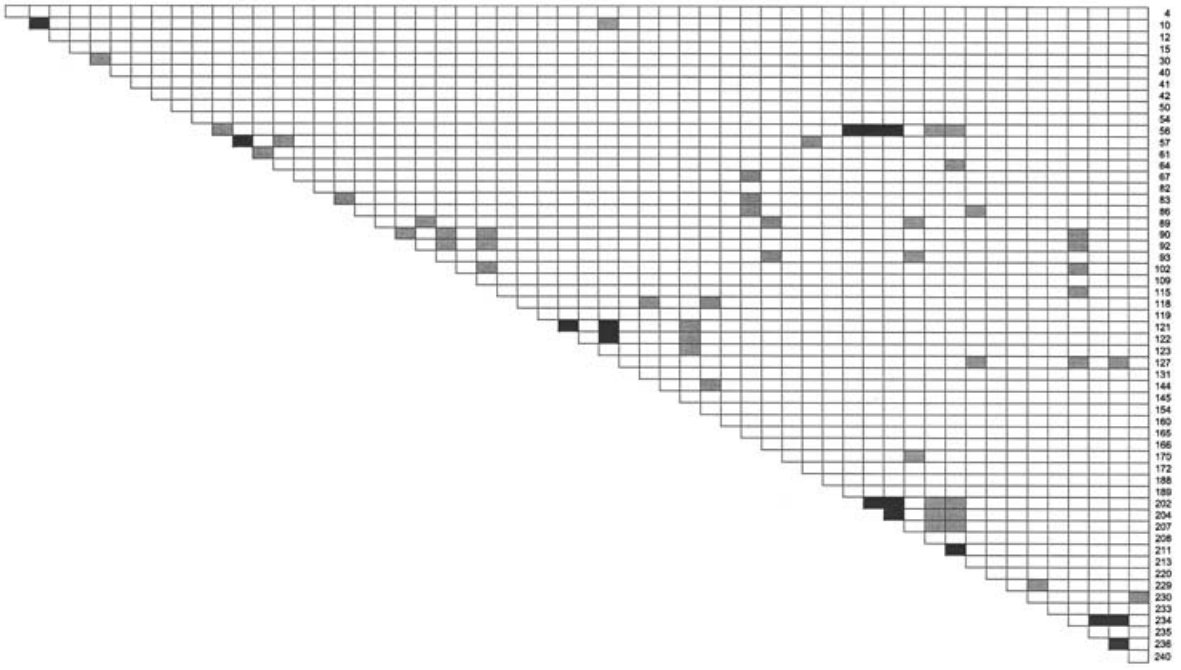

Fig. 3. Plot of significant linkage disequilibrium values for informative sites. Significance was assessed using Fisher's exact test, implemented in DnaSP 3.51 (Rozas & Rozas, 1999). Significance values: white, not significant; grey, $P < 0{\cdot}01$; black, $P < 0{\cdot}001$.

according to the likelihood permutation test implemented in LDhat, again supporting the inference of intragenic recombination (Table 3). The estimate of ρ greatly exceeds the estimate of the amount of mutation (θ) estimated from the same set of polymorphic sites, suggesting that the accumulation of new recombinants exceeds that of new mutations by at least an order of magnitude. Similarly large estimates of ρ were obtained using an alternative (and somewhat larger) estimate of θ obtained using all sites (not shown). Finally, similar results were obtained in analysis of homologous data sets from other taxa (Table 3).

The simulation results indicate that LDhat recovers reasonably accurate estimates of recombination, even given the very large amounts of recombination and mutation expected to accumulate under balancing selection (Table 4). When the amount of recombination in the simulations is zero, LDhat does not

https://doi.org/10.1017/S0016672303006347 Published online by Cambridge University Press

Table 4. *LDhat analysis of simulated data sets incorporating both recombination and balancing selection* (*see Methods*). *20 sequences of 500 base pairs were simulated and results were averaged over 20 replicates for each set of recombination and mutation parameters. The table lists the simulated mutation and recombination rate parameters and their ratio, and estimates of* ρ *and* θ *and their ratio obtained from analysis of the simulated data using LDhat*

Simulation parameters			Estimated parameters		
Recombination rate	Mutation rate	Ratio	ρ_{CLE}	θ	Ratio
0	0·005	0	0	67	0
0·01	0·007	1·4	37	42	0·9
0·1	0·05	2	107	46	2·3

falsely detect its presence, even when the amount of mutation greatly exceeds that estimated for the empirical data (Table 4, line 1; $\rho_{CLE}=0$ for all 20 replicates, where ρ was evaluated for 20 sampling points between 0 and 200). When recombination is present, LDhat does not yield biased estimates of ρ even when the amounts of both recombination and mutation are large, as indicated by the agreement between the expected and observed ratios of these parameters obtained for the simulations and LDhat respectively (Table 4, lines 2 and 3).

The composite likelihood estimate of ρ is the average of all pairwise estimates of ρ for each of the sites included in analysis (Table 3). The difference between the pairwise estimates and ρ_{CLE} may be used to assess the overall fit of the model (of a constant rate of recombination for all sites) to the data. Deviations of pairwise estimates of ρ from ρ_{CLE} indicate that neighbouring pairs of sites tend to show a large deficit of linkage disequilibrium (Table 5). This trend might be attributed in part to the minimum estimate of ρ_{CLE} ($\geqslant 100$), suggesting that an even higher estimate would provide a better fit to the data. However, this can explain only the absolute number of deviating sites, not the trend for deviating pairs of sites to be near one another. Pairs of sites showing a deficit of linkage disequilibrium define the edges of blocks of sequence that experienced comparatively little recombination, sometimes called sequence motifs (Fig. 2). The maintenance of sequence motifs in the presence of abundant recombination is thus the major factor contributing to deviation of the data from the model assumption of a constant rate of recombination across all sites. Finally, the observation that neighbouring sites across the sequence show large deficits in linkage disequilibrium indicates extensive recombination that is not restricted to any particular region within the sequence (Fig. 3, and see Discussion).

4. Discussion

(i) *Estimates of recombination under balancing selection*

The analyses presented here suggest a predominant role for intragenic recombination in generating allelic polymorphism in *P. maniculatus*. The observation of virtually no site variation among the population samples (indicated by the very low site F_{ST} in the AMOVA analysis) despite limited sharing of haplotypes is consistent with extensive polymorphism generated by intragenic recombination rather than point mutations. This inference is further strengthened by multiple lines of evidence indicating extensive intragenic recombination in the history of these sequences; there are a large number of incompatible sequence blocks (Fig. 2) and pairwise measures of site compatibility and linkage disequilibrium are significantly negatively correlated with distance between sites. The value of ρ_{CLE} ($\geqslant 100$) indicates a very large amount of recombination in the history of these sequences, consistent with the rapid decay of significant linkage disequilibrium with distance across the sequence (Fig. 3). Interestingly, the estimate of ρ_{CLE} greatly exceeds the estimate of θ for the same data, suggesting that new recombinants are accumulating much faster than new point mutants.

The likelihood model in LDhat does not include the effects of selection, raising the possibility that balancing selection has effected our estimates. However, simulations of a simple model of symmetric balancing selection with recombination indicate that in fact LDhat performs well even given the very large amounts of recombination and mutation expected to accumulate among sequences maintained by balancing selection (Table 4). Although selection on MHC alleles is undoubtedly more complex than the simple model of balancing selection examined here, the simulation results do suggest that our chief inference (that the amount of recombination exceeds that of mutation by at least an order of magnitude in our data) is not the result of an estimation bias resulting either from the presence of balancing selection, or from the accumulation of large amounts of recombination and/or mutation *per se*.

Are the findings reported here of general significance, or particular to the species studied? We examined several recently published Class II data sets, and found evidence for extensive recombination in all of them (Table 3). We conclude that intragenic recombination is an important and conspicuous feature of MHC Class II polymorphism.

https://doi.org/10.1017/S0016672303006347 Published online by Cambridge University Press

Table 5. *Differences between pairwise estimates of* ρ (ρ_{pair}) *and the CLE* (ρ_{CLE}) *for* Peromyscus maniculatus *MHC Class II* EB *data, for the 35 sites used to estimate the recombination parameter* ρ_{CLE} (*Table 3*). *Positive deviations indicate a deficit of linkage disequilibrium for the particular pairwise comparison compared with the model* ($\rho_{pair} > \rho_{CLE}$), *which assumes a constant rate of recombination for all sites; negative deviations indicate an*

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	−0·02	−0·01	−0·02	1·66	−0·02	−0·08	−0·64	−0·11	−0·49	0·70	−0·21	−0·97	−0·26	0·29	−1·03	−0·88	−0·89
2		−0·07	−0·04	−0·13	−0·10	−0·23	0·85	−0·32	1·01	0·59	0·71	0·26	0·22	−0·02	−0·27	−0·34	−0·35
3			−0·01	−0·05	−0·04	−0·11	1·23	−0·16	1·38	1·11	0·92	0·73	0·30	0·32	−0·02	−0·03	−0·03
4				−0·05	−0·05	1·57	1·76	0·97	0·74	0·71	0·77	0·51	0·37	0·00	0·00	0·15	0·15
5					1·44	−0·04	**2·63**	**2·34**	1·84	−0·56	−0·21	1·78	0·44	0·76	0·06	0·07	0·07
6						−0·00	−0·10	2·79	2·46	2·00	2·39	−0·29	0·70	1·00	0·33	0·21	0·21
7							**2·33**	**2·12**	**2·41**	**2·52**	1·53	−0·49	1·57	0·79	0·19	0·25	0·24
8								−0·13	**3·78**	**2·66**	**3·18**	**3·21**	**2·02**	1·91	0·03	0·00	0·00
9									−0·03	**2·34**	**2·69**	**2·30**	**2·15**	−0·67	0·83	0·52	0·47
10										−0·02	**2·30**	0·46	**2·41**	1·34	0·52	0·39	0·34
11											−0·10	1·46	**2·53**	1·61	0·49	0·85	0·79
12												−0·09	**2·39**	**2·00**	1·07	0·46	0·42
13													−0·46	1·13	0·90	1·30	1·24
14														−0·10	**2·31**	1·81	1·76
15															**2·26**	**2·44**	**2·38**
16																**2·09**	**2·84**
17																	−0·20
18																	
19																	
20																	
21																	
22																	
23																	
24																	
25																	
26																	
27																	
28																	
29																	
30																	
31																	
32																	
33																	
34																	

(ii) *Allelic polymorphism and the role of population subdivision*

Our findings of extensive MHC polymorphism within population samples and only limited sharing of haplotypes among population samples at the MHC *EB* locus in *P. maniculatus* suggest that population subdivision and its effect on effective population size might play an important role in the generation of MHC polymorphism in this species. Theoretical studies indicate that high effective migration of genetic polymorphism maintained by balancing selection militates against the formation of geographic structure (Muirhead, 2001; Schierup *et al.*, 2000). To date, this expectation has been evaluated in few taxa. We find significant evidence for structuring of haplotype diversity among population samples of MHC Class II *EB* diversity in *P. maniculatus*. Only five of the 29 alleles identified (Fig. 1) were found in both samples. Although this might be due in part to incomplete sampling, large differences in the frequencies of particular alleles also suggest subdivision of MHC diversity. Although the estimate of F_{ST} is small, suggesting only limited structure, this interpretation is unjustified with respect to extreme polymorphism maintained by selection. In particular, the maintenance of very high haplotype diversity within populations necessarily constrains F_{ST} within a narrow range bounded by zero. In addition, because a very large fraction of allelic sequence diversity undoubtedly predates current population structure, the observation of widespread shared sequence differences among population samples presents an obvious problem in interpreting the results of sequence based analyses of structure such as AMOVA. We conclude therefore that, although structure exists among the *P. maniculatus* samples, it is unclear how extensive local structuring of variation is or whether this structure is due to limited gene flow,

https://doi.org/10.1017/S0016672303006347 Published online by Cambridge University Press

excess of linkage disequilibrium ($\rho_{pair} < \rho_{CLE}$). *Large deviations* ($\rho_{pair} - \rho_{CLE} > 2$) *are indicated in bold. The shaded region indicates comparisons between sites on either side of the recombination signal proposed by Gyllensten* et al. (*1991*)

19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
−0·41	−0·57	−0·26	−0·46	−0·73	−0·18	−0·22	−0·96	−0·36	−0·36	−0·36	−0·68	−0·79	−0·19	0·01	−0·01	0·01
−1·17	−0·78	0·10	−1·36	−0·06	0·16	−0·60	−1·05	−1·06	−1·06	−1·07	−0·30	−0·09	−0·47	−0·02	−0·00	−0·02
−0·58	−0·18	0·17	−0·65	0·08	0·17	−0·29	−0·23	−0·55	−0·56	−0·56	0·03	0·02	−0·23	0·04	−0·03	0·04
−0·10	−1·03	−0·45	−0·17	0·05	−0·27	0·03	0·01	−0·08	−0·09	−0·09	0·06	0·01	0·01	−1·23	−0·00	−1·25
−0·54	−0·71	−0·09	−0·62	0·09	−0·22	0·14	−1·31	0·14	0·13	0·13	0·05	0·04	0·11	0·02	0·03	0·02
−0·33	−0·43	−0·10	0·22	−0·03	0·22	0·19	−0·87	0·06	0·06	0·05	0·11	0·08	−0·05	−0·62	−1·17	−0·63
0·11	0·15	−0·37	0·04	0·11	−0·25	−0·38	0·02	−0·04	−0·04	−0·05	0·10	0·03	−0·28	0·02	0·00	0·02
0·36	−0·01	0·27	0·15	0·15	0·19	−0·05	0·16	**−2·53**	**−2·55**	**−2·50**	0·11	0·06	0·05	0·06	0·03	0·06
0·10	−0·82	0·31	−0·07	0·17	−0·23	0·18	−1·59	**−5·24**	**−5·52**	**−5·47**	**−5·70**	−3·57	0·09	−1·13	0·09	−1·15
0·18	0·36	0·24	0·28	0·00	0·01	−2·18	−0·11	−0·57	−0·58	−0·59	−0·61	−0·36	−0·01	0·04	0·08	0·03
0·57	0·26	0·39	0·28	0·24	−1·01	0·01	−0·24	−0·27	−0·27	−0·28	−0·16	−0·24	0·05	−0·11	0·04	−0·12
0·83	0·33	0·34	−0·02	0·27	0·30	−0·52	0·18	−0·15	−0·16	−0·16	−0·60	−1·07	0·06	0·00	0·05	0·00
1·00	0·82	−1·30	0·26	−0·35	−0·84	−1·24	−0·45	−0·04	−0·04	−0·05	0·05	−0·01	−1·05	0·04	0·05	0·03
−0·44	1·25	0·28	−0·83	−0·20	−0·27	−0·01	−0·05	0·11	0·10	0·09	0·15	0·07	0·09	0·00	0·02	0·00
2·30	1·33	0·43	0·42	−0·17	0·13	0·27	−0·34	−0·18	0·19	−0·20	−0·36	−0·19	0·07	−0·37	0·09	−0·39
3·75	**2·58**	**2·16**	1·97	1·34	0·69	0·08	−0·06	0·24	0·23	0·22	0·03	−0·01	0·10	0·04	0·07	0·02
3·97	−0·40	**2·33**	−1·26	0·93	0·70	−1·29	0·39	0·13	0·13	0·12	−0·02	−0·04	0·10	−0·31	0·07	−0·35
3·02	−0·33	**2·38**	−1·22	0·98	0·74	−1·28	0·40	0·13	0·13	0·12	−0·02	−0·04	0·10	−0·30	0·07	−0·34
	3·49	**2·35**	0·38	−1·09	0·95	0·45	0·35	−0·02	−0·02	−0·03	−0·13	−0·12	−0·39	0·15	0·12	0·14
		2·36	−0·60	1·68	0·70	0·48	0·08	−1·04	−1·05	−1·07	−0·10	−0·10	0·26	−0·75	−0·16	−0·86
			−0·08	1·92	**3·36**	1·20	1·39	0·70	0·59	0·44	0·28	0·34	0·38	0·07	−0·03	0·07
				−0·30	**2·60**	−0·26	**2·01**	0·49	0·43	0·34	0·26	0·13	0·31	−1·37	0·30	−1·41
					3·13	**2·13**	**2·26**	1·88	1·77	1·61	1·46	1·34	0·03	0·07	−0·01	0·06
						2·07	−0·16	**2·06**	1·97	1·83	−0·22	1·44	0·81	0·43	0·42	0·33
							−0·01	**2·09**	**2·79**	**2·65**	**2·26**	**2·10**	1·05	1·28	1·21	0·97
								−0·28	−0·33	−0·39	**2·70**	**2·75**	1·39	1·32	1·35	0·98
									−0·22	−0·55	1·00	−0·49	1·85	−0·41	**2·02**	−0·49
										−0·33	1·04	−0·40	1·94	−0·39	**2·13**	−0·46
											1·10	−0·27	**2·07**	−0·35	**2·29**	−0·42
												−0·20	**2·53**	1·64	**2·51**	1·40
													1·88	1·68	**2·47**	1·44
														3·20	**2·02**	**2·98**
															−0·11	−0·88
																−0·23

local selection or both. Progress requires more extensive samples of local variation and the use of more sophisticated approaches to analysis of population structure than is possible using a single measure such as F_{ST} (see Muirhead (2001) for one such approach).

Studies of population structure at the MHC remain relatively few, making it difficult to generalize about patterns of population differentiation. Although there is abundant data on human geographic variation at MHC loci, most population differences can be explained by stochastic sorting of pre-existing variation, owing to the recency of human geographic populations. Humans appear to be a particularly poor choice for studies of the effects of population subdivision, given that most MHC variation appears to be the legacy of more extensive population subdivision in the distant past (Takahata & Satta, 1997). A study of four populations of Atlantic salmon found large differences among populations in the frequency of alleles, which again were attributed to relatively recent demographic changes (Langefors *et al*., 2001). A study of MHC variation in the cotton rat (*Sigmodon hispidus*) in North America (Pfau *et al*., 2001) found limited evidence for population differentiation at a *DQ* (A-type) α locus among population samples across the continental USA, largely owing to differences in the frequencies of widespread alleles. Perhaps the most similar result to that reported here is for a comparison of population samples of an Class I MHC locus in Chinook salmon, which found only limited sharing of haplotypes among samples (Miller & Withler, 1997). Interestingly, the estimate of F_{ST} was somewhat larger than that obtained for a set of microsatellite markers, suggesting that effective gene flow was actually lower for the MHC locus than for ostensibly neutral genes. One possible explanation is that local selection limited gene flow among demes at the MHC locus (Muirhead, 2001).

https://doi.org/10.1017/S0016672303006347 Published online by Cambridge University Press

(iii) *Maintenance of linkage disequilibrium and the role of balancing selection*

Ironically, although the observation of patchwork motifs has been routinely cited as evidence for the importance of recombination at the MHC, such motifs appear to be inconsistent with the high rate of intragenic recombination inferred here. The (minimally) order-of-magnitude difference in the amounts of recombination versus mutation suggests that any linkage disequilibrium built up by mutation would be broken down as fast it was created. We suggest that the patchwork motifs are maintained by selection despite frequent recombination (She *et al*., 1990). There is no difficulty with this explanation in principle, given that recombination is expected to be a weak force in comparison to selection in MHC Class I and II genes (Satta *et al*., 1994). This inference is supported by the observation that sequence motifs are often found in the same positions within the second exon in different taxa. We note that this observation is also consistent with a highly variable recombination rate across the exon and has been cited as evidence for a recombination signal embedded within the second exon (Gyllensten *et al*., 1991). However, in view of evidence for extensive recombination distributed throughout the exon (Table 5), such a mechanism appears to be both insufficient and unnecessary (see also Satta, 1997); the evidence for a site-specific recombination signal is in this view an artefact of selection in the presence of extensive recombination.

We thank the following for research and/or collecting permission: the Secretary of Foreign Relations of Mexico (Permit DAN 02578), the Secretary of Environment, Natural Resources and Fisheries of Mexico (Permit DOO 02.5075) and the US Forest Service (Permit CHIS-98-011). We thank A. Narvaez (Office of the Environment, Science and Technology Affairs, US Embassy, Mexico City) for logistical assistance. Grant support from the US National Science Foundation and National Geographic Society is gratefully acknowledged.

References

Awadalla, P. & Charlesworth, D. (1999). Recombination and selection at *Brassica* self-incompatibility loci. *Genetics* **152**, 413–425.

Bergstrom, T. F., Josefsson, A., Erlich, H. A. & Gyllensten, U. (1998). Recent origin of HLA-DRB1 alleles and implications for human evolution. *Nature Genetics* **18**, 237–242.

Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L. & Wiley, D. C. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364**, 33–39.

Carrington, M., Nelson, G. W., Martin, M. P., Kissner, T., Vlahov, D., Goedert, J. J., Kaslow, R., Buchbinder, S., Hoots, K. & O'Brien, S. J. (1999). HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* **283**, 1748–1752.

Edwards, S. V., Chesnut, K., Satta, Y. & Wakeland, E. K. (1997). Ancestral polymorphism of MHC class II genes in mice: implications for balancing selection and the mammalian molecular clock. *Genetics* **146**, 655–668.

Fearnhead, P. & Donnelly, P. (2001). Estimating recombination rates from population genetic data. *Genetics* **159**, 1299–1318.

Goudet, J., Raymond, M., de Meeus, T. & Rousset, F. (1996). Testing differentiation in diploid populations. *Genetics* **144**, 1933–1940.

Gyllensten, U. B., Sundvall, M. & Erlich, H. A. (1991). Allelic diversity is generated by intraexon sequence exchange at the *DRB1* locus of primates. *Proceedings of the National Academy of Sciences of the USA* **88**, 3686–3690.

Hudson, R. R. (2001). Two-locus sampling distributions and their application. *Genetics* **160**, 1231–1241.

Hughes, A. L. & Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**, 167–169.

Hughes, A. L. & Nei, M. (1989). Nucleotide substitution at major histocompatibility complex class II: evidence for overdominant selection. *Proceedings of the National Academy of Sciences of the USA* **86**, 958–962.

Hughes, A. L. & Yeager, M. (1998). Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics* **32**, 415–435.

Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–1245.

Langefors, A., Lohm, J. & von Schantz, T. (2001). Allelic polymorphism in MHC class II B in four populations of Atlantic salmon (*Salmo salar*). *Immunogenetics* **53**, 329–336.

McVean, G. A. T., Awadalla, P. & Fearnhead, P. (2002). A coalescent-based method for detecting recombination from gene sequences. *Genetics* **160**, 1231–1241.

Miller, K. M. & Withler, R. E. (1997). MHC diversity in Pacific salmon: population structure and trans-species allelism. *Hereditas* **127**, 83–95.

Muirhead, C. A. (2001). Consequences of population structure on genes under balancing selection. *Evolution; International Journal of Organic Evolution* **55**, 1532–1541.

Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous substitutions. *Molecular Biology and Evolution* **3**, 418–426.

Nei, M. & Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. New York: Oxford University Press.

Paterson, S., Wilson, K. & Pemberton, J. M. (1998). Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries* L.). *Proceedings of the National Academy of Sciences of the USA* **95**, 3714–3719.

Pfau, R. S., Van Den Bussche, R. A. & McBee, K. (2001). Population genetics of the hispid cotton rat (*Sigmodon hispidus*): patterns of genetic diversity at the major histocompatibility complex. *Molecular Ecology* **10**, 1939–1945.

Richman, A. D., Herrera, G. & Nash, D. (2001). MHC Class II beta sequence diversity in the deermouse (*Peromyscus maniculatus*): implications for models of balancing selection. *Molecular Ecology* **10**, 2765–2773.

Richman, A. D., Herrera, G. & Nash, D. (2002). Characterization of *Peromyscus* MHC Class II beta sequences by LA RT-PCR and DGGE. *European Journal of Immunogenetics* **29**, 213–217.

Rousset, F. & Raymond, M. (1995). Testing heterozygote excess and deficiency. *Genetics* **140**, 1413–1419.

https://doi.org/10.1017/S0016672303006347 Published online by Cambridge University Press

Rozas, J. & Rozas, R. (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175.

Satta, Y. (1997). Effects of intra-locus recombination of HLA polymorphism. *Hereditas* **127**, 105–112.

Satta, Y., O'Huigin, C., Takahata, N. & Klein, J. (1994). Intensity of natural selection at the major histocompatibility complex loci. *Proceedings of the National Academy of Sciences of the USA* **91**, 7184–7188.

Schierup, M. H., Vekemans, X. & Charlesworth, D. (2000). The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genetical Research* **76**, 51–62.

Schierup, M. H., Mikkelsen, A. M. & Hein, J. (2001). Recombination, balancing selection and phylogenies in MHC and self-incompatibility genes. *Genetics* **159**, 1833–1844.

Schneider, S., Roessli, D. & Excoffier, L. (2001). *Arlequin: a Software for Population Genetics Analysis*. University of Geneva, Switzerland: Genetics and Biometry Laboratory.

She, J. X., Wakeland, E. K. & Boehme, S. (1990). The generation and maintenance of MHC class II gene polymorphism in rodents. *Immunological Reviews* **113**, 207–226.

Takahata, N. (1990). A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proceedings of the National Academy of Sciences of the USA* **87**, 2419–2423.

Takahata, N. (1995). MHC diversity and selection. *Immunological Reviews* **143**, 225–247.

Takahata, N. & Nei, M. (1990). Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124**, 967–978.

Takahata, N. & Satta, Y. (1997). Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. *Proceedings of the National Academy of Sciences of the USA* **94**, 4811–4815.

Takahata, N., Satta, Y. & Klein, J. (1992). Polymorphism and balancing selection at major histocompatibility complex loci. *Genetics* **130**, 925–938.

Thursz, M. R., Thomas, H. C., Greenwood, B. M. & Hill, A. V. (1997). Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nature Genetics* **17**, 11–12.

Van Den Bussche, R. A., Hoofer, S. R. & Lochmiller, R. L. (1999). Characterization of MHC-DRB allelic diversity in white-tailed deer (*Odocoileus virginianus*) provides insight into MHC-DRB allelic evolution within Cervidae. *Immunogenetics* **49**, 429–437.

https://doi.org/10.1017/S0016672303006347 Published online by Cambridge University Press