Polymorphism Across an Exon-Intron Boundary in an Avian *Mhc* Class II B Gene

Daniel Garrigan[1] and Scott V. Edwards

Department of Zoology, University of Washington

Twenty-three sequence haplotypes spanning the boundary of the second exon and intron of a red-winged blackbird *Mhc* class II B gene, *Agph-DAB1,* are presented. The polymorphism of the exon segment is distributed in two divergent allelic lineages which appear to be maintained by balancing selection. The silent nucleotide diversity of the exon ($\pi = 0.101$) is more than five times that of the intron ($\pi = 0.018$) and decays rapidly across the exon-intron boundary. Additionally, genealogical reconstruction indicates that divergence from a common ancestor in the exon sample is over four times that of the intron. The intron sequences reveal a pattern of polymorphism which is characteristic of directional selection, rather than a pattern expected from linkage to a balanced polymorphism. These results suggest that the evolutionary histories of these two adjacent regions have been disassociated by recombination or gene conversion. The estimated population recombination parameter between the exon and the intron is sufficiently high ($4N_e c = 8.545$) to explain the homogenization of intron sequences. Compatibility analyses estimate that these events primarily occur from the exon-intron boundary to about 20–30 bases into the intron. Additionally, the observation that divergent exon alleles share identical intron sequence supports the conclusion of disassociation of exon and intron evolutionary histories by recombination.

Introduction

The major histocompatibility complex (*Mhc*) is a large multigene family that encodes glycoproteins which mediate the specificity of the vertebrate adaptive immune response. Class II *Mhc* proteins are expressed only in antigen-presenting cells, such as macrophages. The function of these proteins is to bind truncated foreign peptides ingested by the cell and present them to helper T cells. Interleukins secreted by the antigen-presenting cell stimulate cloning of the helper T cells that bind the antigen, which, in the humoral response, causes B cells to transform into antibody-producing plasma cells. The secondary structure of the extracellular domain of each of the two chains (α and β) comprising class II molecules consists of an α helix and a β pleated sheet. A number of key residues in these regions influence the peptide-binding capabilities of the MHC protein and are encoded by codons known as peptide-binding region (PBR) codons (see Klein et al. [1993] and Edwards and Hedrick [1998] for reviews of *Mhc* biology).

Within many vertebrate populations, the PBR codons are known to be highly polymorphic at numerous "classical" *Mhc* loci. Investigation of some class II loci, such as HLA-DRB1 in humans, reveals at least 179 alleles segregating in the population (Parham and Ohta 1996). Various forms of balancing selection have been proposed as the mechanism for maintaining these extreme levels of polymorphism (Hedrick and Thomson 1983). Two phenomena are hallmarks of *Mhc* polymorphism: first, there are a large number of alleles segregating in a population, and second, many alleles persist over an unusually large number of generations. Takahata (1990) has estimated that a sample of rodent *Mhc* alleles coalesce at about $260N_e$ generations, rather than the $4N_e$ generations expected for a neutral locus (N_e is the effective population size).

If these explanations of *Mhc* polymorphism are correct, one would predict that a neutral region linked to the PBR codons will exhibit levels of polymorphism much higher than that expected by random drift alone. Similarly, one would also predict that this elevated polymorphism will decay as recombination increasingly separates the histories of the two regions, and that regions effectively unlinked to the selected sites will exhibit neutral or nearly neutral levels of polymorphism (Takahata and Satta 1998). In this way, the extent of polymorphism in a neutral region closely linked to the PBR, such as introns, will depend on the intensity of selection in the PBR, effective population size, rate of recombination, and the rate of point mutation as recently illustrated for the human HLA-H pseudogene by Grimsley, Mather, and Ober (1998).

The majority of *Mhc* class II β-chain polymorphism studies focus on the second exon, which contains all of the PBR codons. While these data are critical for describing the forces of selection which mold PBR polymorphism, determination of noncoding sequences flanking the PBR exons is equally important in understanding the full spectrum of evolutionary dynamics acting on the exon (Graser et al. 1996). Based on data from both introns flanking the PBR exon and the PBR exon itself, Bergström et al. (1998) have suggested that many of the human DRB1 alleles are not ancient balanced polymorphisms, but a great deal of the allelic diversity has been recently generated. While there was evidence in their study that some serologically distinct groups of alleles are maintained over extended periods by selection, within each of these groups there was significantly lower variation in the flanking introns than predicted by the variation of the closely linked exons. The authors concluded that genetic mechanisms such as gene conversion

have diversified short segments only in the exons within allelic lineages.

The present study surveys polymorphism in the PBR exon and second intron of a recently characterized *Mhc* class II B gene in red-winged blackbirds (*Agelaius phoeniceus*), *Agph-DAB1* (Edwards, Gasper, and March 1998). The *Agph-DAB1* locus encodes a β-chain from a class II molecule and is believed to be highly polymorphic. However, despite characterization of the genomic sequence in and around the PBR exon, Edwards, Gasper, and March (1998) were unable to selectively amplify a single locus with primers placed in regions immediately flanking this exon, but, rather, appeared to amplify multiple loci. This result was likely due to the large number of *Mhc* class II loci in the blackbird genome (Edwards, Nusser, and Gasper 1999) and the likelihood that concerted evolution occurs at relatively high rates in the class II genes of birds (Edwards, Wakeland, and Potts 1995; Wittzell et al. 1999). Thus, another major goal of this study was to survey diversity at a single avian class II *Mhc* locus, which has only been accomplished thus far for chickens (Zoorob et al. 1993).

Materials and Methods
PCR, Cloning, and Sequencing

High-molecular-weight genomic DNA was obtained from 17 individual blackbirds from 4 populations in Florida ($n = 7$), Kentucky ($n = 6$), New York ($n = 3$), and Colorado ($n = 1$). Polymerase chain reaction (PCR) primers were designed to discriminate between the intron 2 sequence of *Agph-DAB1* and that of another blackbird class II B gene, *Agph-DAB2* (S. V. Edwards, unpublished data). The two primers Ex2F.2 (5′-ATGCC-MAGYRCWGGAACAGCGAC) and Int2R.6 (5′-ATT-GGGAATAGACCGGGAAT) amplify a 337-bp product spanning 89 bp of the 3′ end of exon 2 and 248 bp of the 5′ end of intron 2. The exon 2 region includes the majority of the α-helical region and includes 11 codons which encode residues known to contact processed peptides in a human class II protein (Brown et al. 1993). In total, these data represent about one third of the entire second exon and 65% of the second intron (Edwards et al. 1998).

Products from PCR were blunt-end-ligated into the pCR2.1 vector and transformed into INVαF′ cells from the TA cloning kit (Invitrogen). Twenty positive colonies per individual were chosen and screened via PCR for presence of the correct insert. A total of 96 clones were sequenced with the M13 forward and reverse primers on an ABI 373A automated sequencer using dye terminator chemistry. The sequences of all cloned inserts were determined in both the forward and the reverse directions. Additionally, sequences from two to eight clones found to be identical within each individual determined the consensus sequence of each haplotype. The sequences were aligned manually using the computer application SeqPup (Gilbert 1996) and have been deposited in the GenBank nucleotide database (accession numbers AF105404–AF105420 and AF133292–AF133297).

Data Analysis

To quantify polymorphism, we calculated the average number of pairwise differences between sequences (k; Tajima 1983) and the number of segregating sites (S) using the computer application DnaSP (Rozas and Rozas 1997). Nucleotide diversity (π) was determined by dividing k by the number of sites. Additionally, the parameter M of Watterson (1975) was estimated from S; this value was also divided by the number of sites for comparison between regions, as we applied these measures to the exon and the to intron separately. Plotting the distribution of nucleotide diversity across the exon-intron boundary is useful for detecting the spatial heterogeneity of polymorphism caused by selection and recombination (Hudson and Kaplan 1988). To accomplish this, π was measured within sliding windows of constant length whose midpoints were moved by 1 bp along the entire sequence, with pairwise removal of alignment gaps. The optimal window size for detecting nonrandom variation in a region of DNA sequence was determined by the method of Tajima (1991).

Under the neutral infinite-sites model, the mutation parameter $4N_e\mu$ (μ is the mutation rate) predicts the expected value of both k and S. If a locus is selectively neutral, $M = k$. If there is significant deviation from this expectation, the action of selective forces can be inferred (Tajima 1989). The D statistic of Tajima (1989) tests the null hypothesis that $k - M = 0$. If $D < 0$, mutant alleles occur at low frequency and most of the mutations occur in the external branches of the genealogy, a pattern consistent with directional selection. If $D > 0$, mutant alleles occur at high frequencies and most of the mutations occur in internal branches of the genealogy, a pattern consistent with balancing selection.

The linked-polymorphism hypothesis of *Mhc* evolution predicts that $D > 0$ in both the exon and the intron. Null distributions of D, given the number of segregating sites and the sample size, were generated by the Monte Carlo simulation method of Hudson (1990) to determine the statistical significance of the observed D value at the $\alpha = 0.05$ level with 1,000 replicates. Takahata (1990) demonstrated that the shape of the genealogy of strongly balanced lines will not necessarily deviate from that expected under neutrality; rather, it is the timescale of the balanced genealogy that will be elongated by a factor that depends on the selection coefficient in a symmetric, overdominant scenario (Takahata 1990, eq. 15). To study this possibility, genealogies of the exon and intron were estimated in order to detect possible differences in levels of divergence between all alleles. Genealogies were reconstructed using a standard phylogenetic Markov chain model of base substitution as implemented in the computer applications DNAML and DNAMLK (Felsenstein 1995). A test for constancy of rates among branches was performed with a likelihood ratio test whereby twice the difference in log likelihoods is assumed to follow a chi-square distribution with $n - 2$ degrees of freedom, where n is the number of tips in the genealogical trees.

```
                A   L   L   E   Y   K   R   T   A   V   D   R   Y   C   R   H   N   Y   R   I   D   A   P   F   S   V   E   R   R
DAB1*07   G GCC TTA CTG GAG TAC AAA CGG ACT GCG GTG GAC CGG TAC TGC CGG CAC AAC TAC CGG ATT GAT GCC CCA TTC AGC GTG GAG CGC CGA G GTG 92
DAB1*04   . ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... . ...
DAB1*01   . ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..A. ... ... ... ... ... . ...
DAB1*09   . ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... . ...
DAB1*13   . ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... . ...
DAB1*14   . ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... . ...
DAB1*16   . ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... . ...
DAB1*20   . ... ... ... ... ... ... ... .. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .A .A. ... A ...
DAB1*21   . ... ... ... ... ... ... ...A. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... . ...
DAB1*22   . ... ... ... ... ... ... ...A. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... . ...
DAB1*23   . ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...A. ... ... ... ... ... ... ... ... ... . ...
DAB1*18   . ... ..C ... ... ... GC. ... .. ... .. ... .C. ... C.. ... ... ... ACC A.. ... ... ... ... TC. ... T ...
DAB1*02   . .AA ..C ... G.. GC. ... ... ..A ..A AC. ... ... ... ... ACC A.. ... ... ... ... ... ... ... . ...
DAB1*03   . .AA ..C ... G.. GC. ... ... ..A ..A AC. ... ... ... ... ACC A.. ... ... ... ... ... ... ... . ...
DAB1*05   . .AA ..C ... G.. GC. ... ... ..A ..A AC. ... ... ... ... ACC A.. ... ... ... ... ... ... ... . ...
DAB1*06   . .AA ..C ... G.. GC. ... ... ..A ..A AC. ... ... ... ... ACC A.. ... ... ... ... ... ... ... . ...
DAB1*08   . .AA ..C ... G.. GC. ... ... ..A ..A AC. ... ... ... ... ACC A.. ... ... ... ... ... ... ... . ...
DAB1*10   . .AA ..C ... G.. GC. ... ... ..A ..A AC. ... ... ... ... ... ... ... ... ... ... ... ... ... . ...
DAB1*11   . .AA ..C ... G.. GC. ... ... ..A ..A AC. ... ... ... ... ACC A.. ... ... ... ... ... A ...
DAB1*12   . .AA ..C ... G.. GC. ... ... ..A ..A AC. ... ... ... ... ACC A.. ... ... ... ... ... ... ... . ...
DAB1*15   . .AA ..C ... G.. GC. ... ... ..A ..A AC. ... ... ... ... ACC A.. ... ... ... ... ... ... ... . ...
DAB1*17   . .AA ..C ... G.. GC. ... ... ..A ..A AC. ... ... ... ... ACC A.. ... ... ... ... ... ... ... . ...
DAB1*19   . .AA ..C ... G.. GC. ... ... ..A ..A AC. ... ... ... ... ACC A.. ... ... ... ... ... ... ... . ...

                    +           +   +               +               +       +   +           +   +       +   +
DAB1*07   AGCGCGGGGCAGAGCGTGTCCCCTCGGGCCCTGCCCTGCCAGTGACCATCCCAGTCAATCCCAGTCGCCACCAGTCCCTTCCAGTATATTCCCAGTCCCTCCCACTCCATTCCCAGTCCCCTT 215
DAB1*04   .......G.....................................................................................................................
DAB1*01   .....................................................................C.......................................................
DAB1*09   ..........A.................................................................................................................
DAB1*13   ...........................................................................................................................
DAB1*14   ......................................................T.....................................................................
DAB1*16   ..............T..T..........................................................................................................
DAB1*20   ...............AA.....T.................A...................................................................................
DAB1*21   ..G........A................................................................................................................
DAB1*22   ..G........A...T.................................A..........................................................................
DAB1*23   ............A.T.................................A..........................................................................
DAB1*18   ...............AA.........A......A........G......AT.................G...........G.................A...............
DAB1*02   ..............A.............................................................................................................
DAB1*03   ..............A.............................................................................................................
DAB1*05   ..........................................................................................................C.................
DAB1*06   ...C......G.................................................................................................................
DAB1*08   ..........................................................................................................C.................
DAB1*10   ...........................................................................................................................
DAB1*11   ...........................................................................................................................
DAB1*12   ...........................................G...............................................................................
DAB1*15   ...........................................................................................................................
DAB1*17   ...........................................................................................................................
DAB1*19   ...............A................A...A......................................................................................

DAB1*07   CGAGTCCATTCCCAGTCCCTCCCAGTCCATTCCAAATCCATTCCAGTCTATCCCACTCTCTCCCAGTCCCTCTCAGTCCATCCCAATCCCTCCAGT-CCATTCCTAGTCCCTT-CC-AGTCC 337
DAB1*04   G.....................................................................A.A.....................-..........C......-.-.--.....
DAB1*01   .....................................................................................................-............-.-.--.....
DAB1*09   .........................................C...........................................................-............-.-.--.....
DAB1*13   .........................................C...........................................................-............-.-.--.....
DAB1*14   .....................................................................................................-............-.-.--.....
DAB1*16   .....................................................................................................-............-.-.--.....
DAB1*20   ......T..................................C...........................................................-........G...-.-.--.....
DAB1*21   .....................................................................................................-............-.-.--.....
DAB1*22   .T.................C...............................................T........................T......A......-.-.--.....
DAB1*23   .T...T.........................T.....................................................................-............-.-.--.....
DAB1*18   .....................................................................................................-............-.-.--.....
DAB1*02   G....................................................................................................-............-.-.--.....
DAB1*03   .....................................................................................................-............-.-.--.....
DAB1*05   G.....................................................G..............................................-..........T.-.-.--.....
DAB1*06   .....................................................................................................-............-.-.--.....
DAB1*08   .....................................................................................................-............-.-.--.....
DAB1*10   .....................................................................................................-......A.....-.-.--.....
DAB1*11   .....................................................................................................-............-.-.--.....
DAB1*12   .....................................................................................................-............-.-.--.....
DAB1*15   ..............................................................................T......A......-..C.-.-.--.....
DAB1*17   .....................................................................................................-............-.-.--.....
DAB1*19   ..............G......................................................................................-............-.-.--.....
```
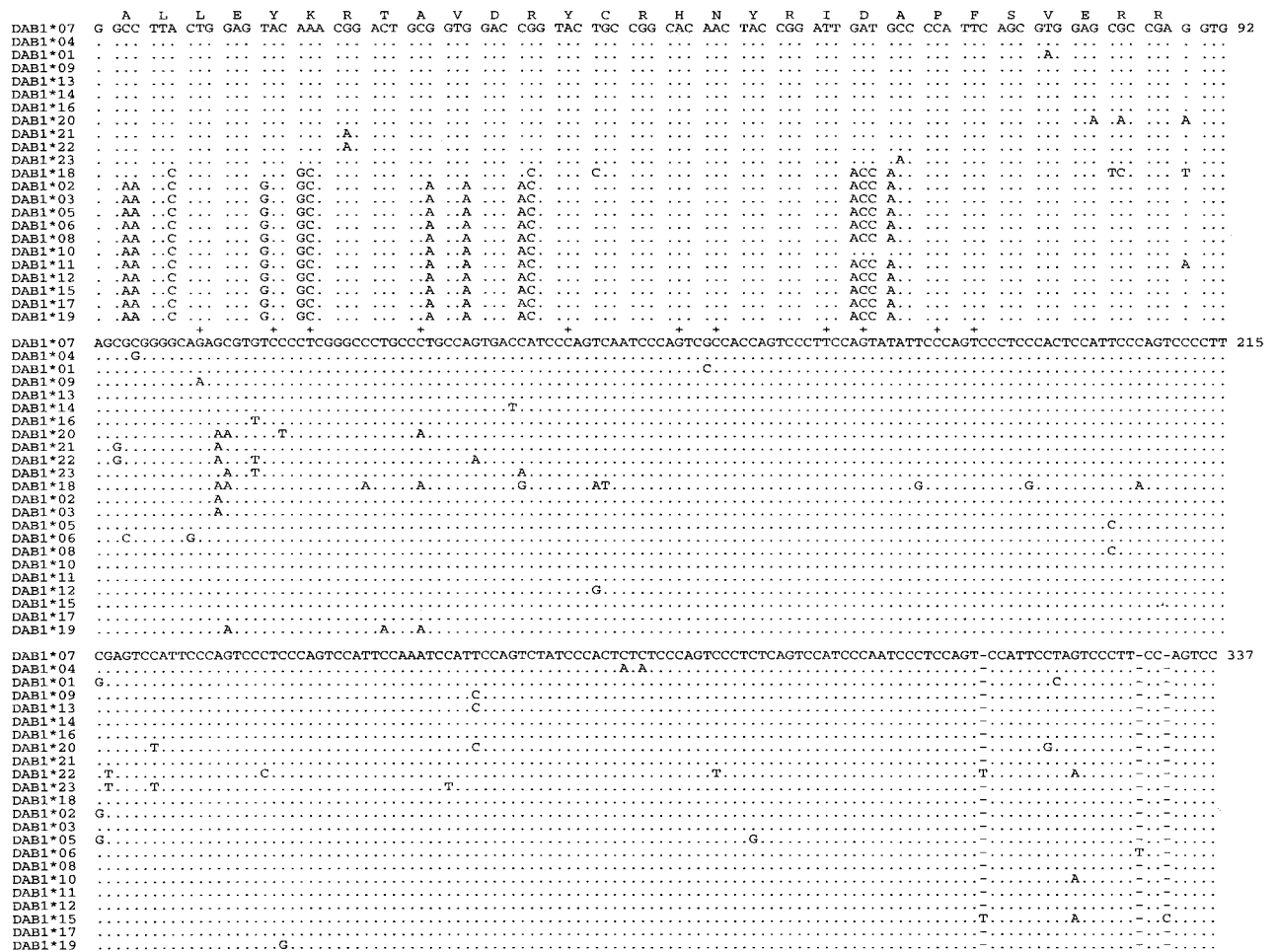
FIG. 1.—Complete nucleotide sequence from all 23 haplotypes. PBR codons are designated with plus signs. Indels occur in DAB1*22, DAB1*06, and DAB1*15. The haplotypes are abbreviated from the full name *Agph-DAB1*XX,* as suggested by Klein et al. (1990).

Results

Sequences and Polymorphism

In total, we sequenced an average of five clones (range, three to eight) per individual. No more than two sequence types were obtained from a single individual, providing strong evidence that we amplified a single locus. A total of 23 different exon-intron haplotypes were present in the sample of 34 chromosomes. By this cloning scheme, 9 of 17 blackbirds were observed to be heterozygous. For most individuals, five to eight clones were sequenced, resulting in at least a 94% chance of detecting both copies. However, only three clones were sequenced from the Colorado individual; although only a single sequence was retrieved, no inference was made as to whether this bird was either homozygous or heterozygous because of the lower probability of detecting both alleles. Thus, a total of 33 sequences were utilized in subsequent estimates of population variation. The full alignment was 337 bp, including two sequences which contained a single-base indel and one which had two single- base indels relative to the others (fig. 1). There were a total of 58 polymorphic sites in the entire 337-bp region (17%), excluding the three indels. The sequencing revealed a total of eight alleles that differed at the amino acid level, two of which occurred at high frequencies, corresponding to sequence types 1a (42.42%) and 5 (30.30%) in Edwards, Gasper, and March (1998). These two alleles exhibit observed differences at 29.6% of amino acid positions and 20.2% of nucleotide positions. Four of the nine putative heterozygotes carried both of the divergent alleles.

Because many of the rare exon alleles differ from the common types by only a single nonsynonymous substitution, the samples are hereafter referred to as belonging to "allelic lineages" 1 and 5, following the allele numbering scheme of Edwards, Gasper, and March (1998). However, *Agph-DAB1*18* and *Agph-DAB1*10* appear to be the result of gene conversion events. *Agph-DAB*10* is similar to lineage 1 sequences, except at nucleotide positions 62–65, where it is identical to lineage 5 sequences (fig. 1). *Agph-DAB1*18* resembles lineage 5 alleles, but is identical to lineage 1 alleles in a patchwork consisting of sites 7, 17–18, 36, and, again, 62–65. These recombinant alleles were excluded from the calculation of Tajima's *D* statistic (M. Slatkin, personal communication).

To further characterize patterns of polymorphism, the data were partitioned in accordance with lineages 1

Table 1
Polymorphism Statistics for the Exon and the Intron

VALUE	EXON			INTRON		
	Total	Within 1	Within 5	Total	Within 1	Within 5
n	33	15	16	33	15	16
S	22	1	7	36	14	23
k	8.057	0.133	1.292	4.373	2.991	4.717
M	0.071	0.003	0.025	0.040	0.018	0.030
π	0.101	0.002	0.015	0.018	0.012	0.020
D	1.986	−1.159	−1.366	−1.747	−1.211	−1.299
$P(D\,|\,S,\,n)$	0.016	0.000	0.100	0.017	0.102	0.086

NOTE.—Comparisons are made both for the total sample and within each exon lineage. Nucleotide diversity (π) is corrected for multiple hits by the method of Jukes and Cantor (1969). M per site is calculated by the finite-sites model of Tajima (1996). Definitions of all values can be found in the *Materials and Methods* section of the text.

and 5, which can also be equated with the "mutant allelic classes" designation of recent population genetics models (Slatkin 1996). Table 1 summarizes the measures of polymorphism in the intron and exon separately, both within and between lineages. The nucleotide diversity (π) is over five times as high between exon lineages as between the corresponding intron sequences. More importantly, the silent nucleotide diversity in the exon ($\pi = 0.101$) is also higher than the intron nucleotide diversity ($\pi = 0.018$) by a factor of 5. We employed a heteroscedastic one-tailed t-test to detect significant increases in π in silent exon sites in order to correct for unequal variances that may be caused by balancing selection in the exon ($F = 14.296$). The null hypothesis that the pairwise silent exon diversity is equal to or less than the intron diversity is rejected ($t = 16.174$, $P < 0.05$, df $= 528$).

Selection and Recombination

The neutrality test of Tajima (1989) rejects the null hypothesis that the polymorphism of exon 2 ($D = 1.986$) could be generated by random drift alone ($P = 0.016$). The observation that $D > 0$ in the exon suggests a pattern consistent with a balanced polymorphism in this region. The test also rejects the null hypothesis that random drift alone acts on the intron ($P = 0.017$). The fact that $D = -1.747$ in the intron indicates an excess of rare mutations, an unexpected result if the intron has been linked to a balanced polymorphism existing in the population for many generations. Thus, we can infer that most of the mutations observed in the intron occur on the external branches of the genealogy, rather than on deeper branches, and that the genealogy itself is starlike in shape. This pattern suggests either rapid population expansion or directional selection, the former being a more likely cause of this pattern in a noncoding region. This contrast of patterns of polymorphism indicates that the evolutionary histories of these two regions are independent of one another.

The two major allelic lineages and the two recombinant forms constitute the four distinct sequence types when only the 11 PBR codons surveyed here are considered. The nucleotide diversity of PBR replacement sites is 0.104 (0.130 per silent PBR site) based on the Nei and Gojobori (1986) method of determining silent and replacement codon sites using the computer application DnaSP. This yields a mean ratio of replacement to silent substitutions of 0.800. While this pattern is not consistent with the effects of positive selection acting on the PBR codons, it is potentially biased by the small number of silent PBR sites under examination. If we extrapolate and include data from all PBR codons from the corresponding alleles (1a and 5) from Edwards, Gasper, and March (1998), the replacement-to-silent ratio becomes 3.33. If we assume the extrapolated ratio as a measure of selection intensity, the scaling factor of Takahata (1990) is 8.855. Thus, we expect that these PBR alleles had a common ancestor approximately $35N_e$ generations ago.

Figure 2*A* and *B* illustrates the spatial distribution of observed nucleotide diversity across the exon-intron boundary and demonstrates the rapid decline in diversity across it. The comparisons of spatial nucleotide diversity within lineages show that there is a peak in the intron just downstream of the exon-intron boundary, especially within lineage 5 (fig. 2*D* and table 1). This observation might suggest that potential recombination breakpoints are not directly on the boundary of the exon and intron, but, rather, extend about 20–30 bases into the intron. This conclusion is substantiated by a maximum-likelihood analysis of spatial phylogenetic variation as measured by the computer program PLATO (Grassly and Holmes 1997). The result of this analysis implicates positions 1–157, a region that includes 19% of the 5′ end of the intron, as evolving anomolously relative to the remaining 180 bp. A compatibility matrix (fig. 3) clusters these events between base positions 65 and 110, around the exon-intron boundary at base position 90 (Jakobsen and Easteal 1996).

The genealogies (fig. 4) show a marked difference in the amount of divergence for the two regions. Constancy of rates among branches was not rejected in either tree at the $\alpha = 0.05$ level of significance. Thus, we have depicted the genealogies which assume the same substitution rate in all branches. The divergence of the exon 2 samples is approximately four times as great as that of the intron 2 samples. Perhaps even more striking, however, is the observation that identical intron sequences (i.e., *Agph-DAB1*07* and *Agph-DAB1*11*) are associated with divergent exon alleles from lineages 1 and 5. This observation also implies a recent recombination event between these two haplotypes or their ancestors.
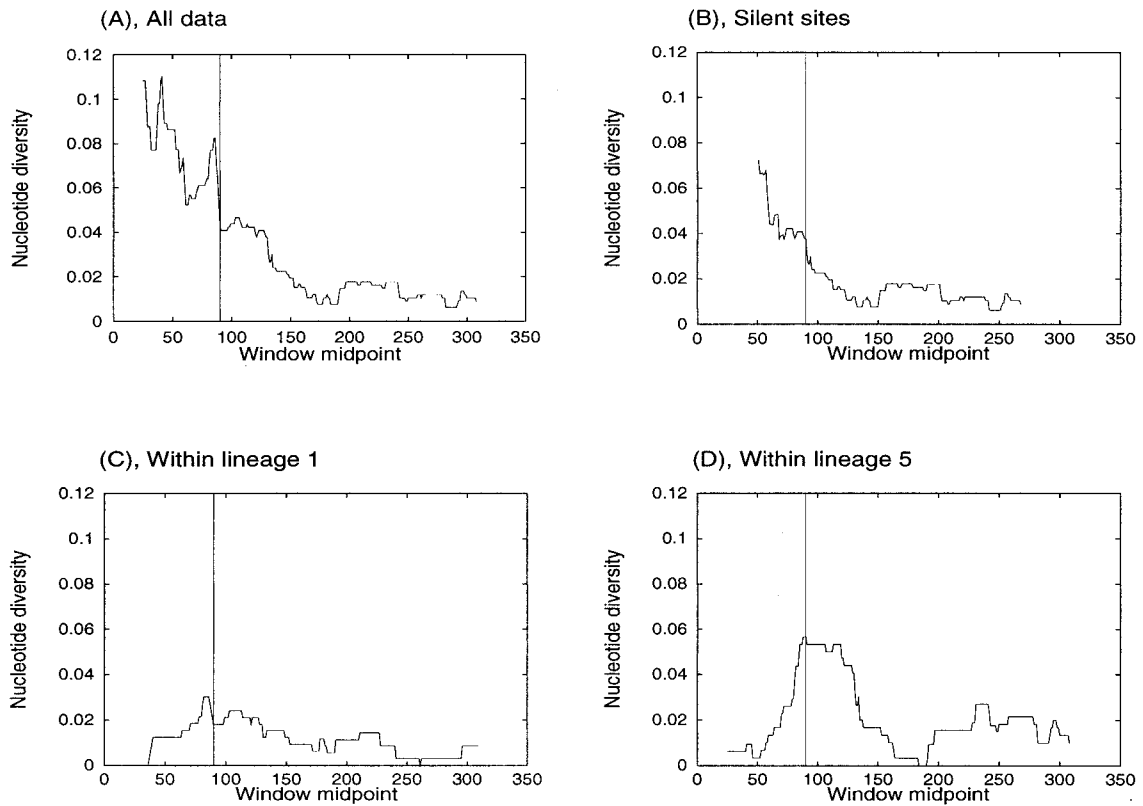
FIG. 2.—Sliding-window analysis of nucleotide diversity across the exon-intron boundary. Vertical lines indicate the exon-intron boundary at base position 90. *A,* Between all alleles sampled. *B,* Between all alleles, measuring only silent sites in the exon. *C,* Within allelic lineage 1. *D,* Within lineage 5. Window size is 50 bp, taken at each nucleotide site. Data points are offset from the X-axis because points correspond to the base positions of the midpoints of windows.
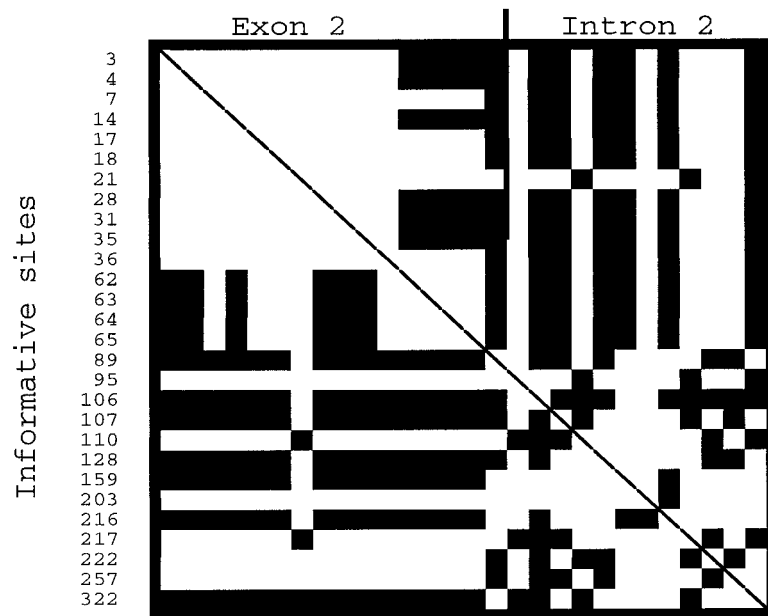

FIG. 3.—Compatibility matrix between individual informative sites generated by the RETICULATE computer application (Jackobsen and Eastel 1996). Filled squares represent pairs of sites which are phylogenetically incompatible. In the exon, the putative conversion tracts are identified, and regions around the exon-intron boundary correspond to regions of high within-lineage polymorphism.
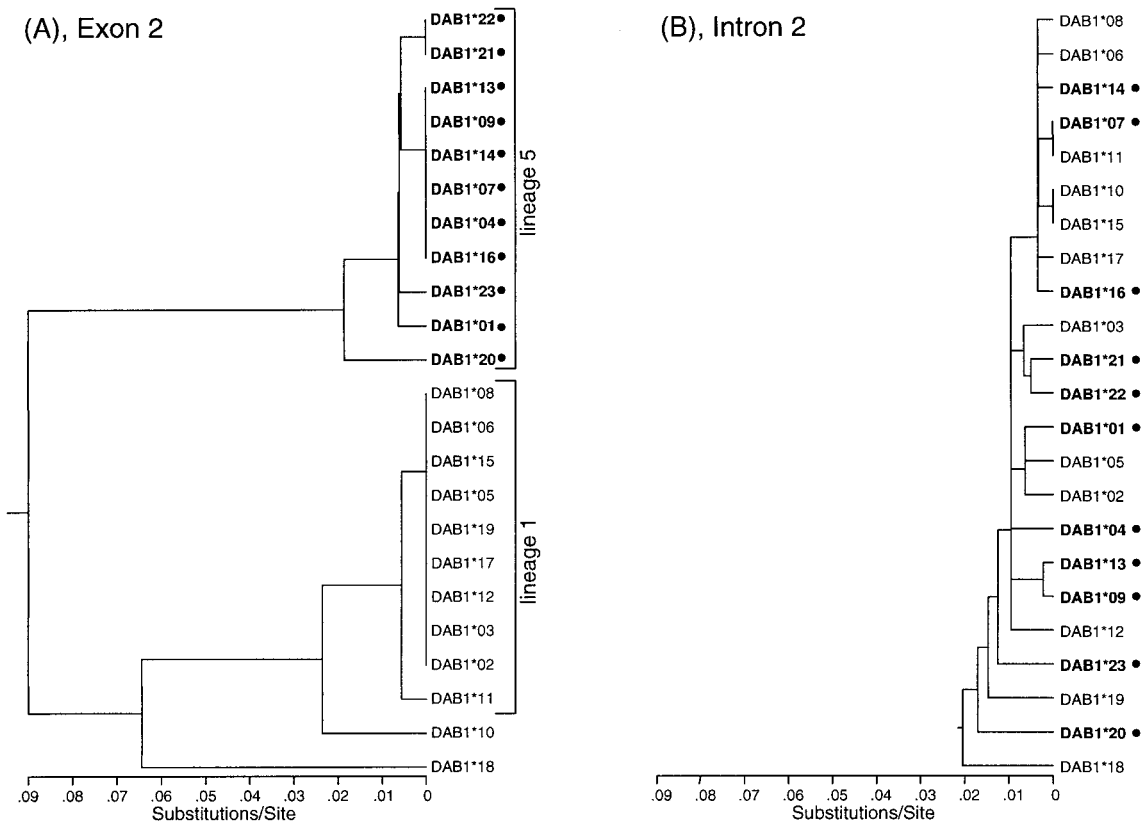
FIG. 4.—Maximum-likelihood reconstructions of allelic genealogies for (A) exon 2 (lnL = −269.24) and (B) intron 2 (lnL = −655.97); tips labeled in bold with a bullet represent samples associated with exon lineage 5. Genealogies were constructed under the assumption of constancy of rates among branches. The topological congruence test of Kishino and Hasegawa (1989) rejects the hypotheses that each tree is congruent with the other data set (P < 0.01). Genealogies are rooted at the same position as they are when reconstructed with *Agph-DAB2* as an outgroup (not shown).

An estimation of the neutral recombination parameter $4N_e c$ (c is the rate of recombination per gamete per generation) was made by the method of Hudson (1987). Also, the analogous parameter γ (Hey and Wakeley 1997) was estimated, using the computer application SITES, because of a lower bias for small segments of sequence. The time back until a recombination event in these samples will be exponentially distributed with mean $2N/(1 + 4N_e c)$ (Hudson 1990). Hudson's method yields $4N_e c = 8.545$, very similar to $\gamma = 8.509$. Hence, the average time until a pair of sequences experiences a recombination is $0.21N_e$ generations. The relative ratio of $c/\mu = 0.560$ is small and may be an artifact of μ for balanced lines being overestimated by the neutral model (Hudson 1987).

Discussion

There is strong evidence that these data do indeed represent a single class II *Mhc* locus and thus provide a first glimpse into the polymorphism of such a gene in songbirds. Although the DNA segment we have surveyed is relatively short and covers only about one third of the second exon, it nonetheless spans a critical boundary for the interpretation of *Mhc* evolution. The reduced divergence between the intron sequences which flank the PBR exon is not consistent with a pattern expected under complete linkage to a balanced polymorphism. The divergence among introns is significantly lower than the silent divergence among exons. Furthermore, the rapid decay of polymorphism across the exon-intron boundary suggests the action of recombination or gene conversion in the exon. However, recombination appears to be occurring well into the intron, rather than only at the exon-intron boundary. These patterns may be explained by invoking either of two scenarios. The first is that intraallelic recombination has disassociated the intron sequence, and a single intron allele has recently reached fixation. The second is that there are vastly different rates of mutation in the two regions.

The independent genealogical history of *Agph-DAB1* introns, as compared with the exon 2 genealogical history, suggests that our samples may be the result of historical recombination or gene conversion events. Cereb, Hughes, and Yang (1997) found a very similar pattern in an HLA class I PBR exon and flanking introns and explained it by invoking recombination and subsequent fixation of introns by random drift. This process can easily be modeled by analogy to migration between two island subpopulations. If each exon lineage is considered a subpopulation, for what rate of intraallelic recombination or conversion (analogous to migration) will panmixia occur? This amounts to recombination swap-

Table 2
Nucleotide Diversity Comparisons Within (π_w) and Between (π_b) Allelic Lineages

Region	π_b	π_w	F_{st}
Exon 2	0.0960	0.0039	0.9246
Intron 2	0.0164	0.0160	0.0123

NOTE.—Values of F_{st} correspond to the degree of subdivision by lineages. The within-lineage values are the unweighted mean from each lineage (Takahata and Satta 1998). Exon 2 values are for silent sites only.

ping alleles back and forth between lineages, which can only occur in heterozygotes. Takahata and Satta (1998) suggested the use of an F_{st} statistic to quantify the degree of interallelic differentiation present in a neutral region linked to a balanced polymorphism (Takahata and Satta 1998, eq. 8). Table 2 illustrates the use of this statistic to show how recombination can break down this differentiation. As with migration, the subpopulations or lineages in linked, neutral regions will drift as a single population when $N_e c > 1$. Thus, if selection has maintained the exon 2 alleles for approximately $35N_e$ generations and recombination occurs between the exon and the intron, we would expect that the neutral intron will at some point in its history reach fixation. This predictive model seems to be able to explain the random scattering of lineages on the intron 2 genealogy (fig. 4*B*).

An alternate explanation for the dramatic difference in polymorphism in the exon and the intron is that the mutation rate is greatly elevated in the exon or, conversely, highly conserved in the intron. If the mutation rate is indeed elevated in the exon, then divergence times of exon alleles may be overestimated. We attempted to resolve this possibility by using the primers to amplify the same region in a closely related blackbird, *Agelaius thilius,* in order to calibrate a between-species rate of nucleotide substitution. However, no PCR product was obtained; hence, further work must be done to characterize the region in that species. Data from the human *Mhc* suggest equivalence of silent rates in class II B introns and exons. Satta, Mayer, and Klein (1996) have found that while human DRB introns evolve slowly (1×10^{-9} per site per year), the rate did not significantly differ from the rate of synonymous substitution in exon 2. Regardless, the fact that divergent exon alleles in our study share identical intron types argues against the theory that discrepancies in mutation rate alone are responsible for the observed pattern.

The *Agph-DAB1*10* and *Agph-DAB1*18* haplotypes are interesting examples of the patchwork pattern found in the second exons of many *Mhc* genes (Ohta 1997). This pattern is potentially the result of a series of intragenic gene conversion events. The elevated silent divergence in the exon over that in the intron is also consistent with recurrent gene conversion acting on the exon. Simulations performed by Ohta (1999) have demonstrated that the ratio of synonymous exon divergence to that of a linked intron will increase as the rate of either intragenic or intergenic conversion increases, a pattern consistent with our data.

Is the intron polymorphism greater than what we would expect for a truly neutral locus? If we utilize female red-winged blackbird population size approximations from the mtDNA data of Avise, Ball, and Arnold (1988; $N_{ef} = 3.75 \times 10^4$), then because *A. phoeniceus* is highly polygynous, we can estimate the total $N_e \approx 5 \times 10^4$. If we also assume a conservatively high value of $\mu = 1 \times 10^{-8}$ in the intron, as in humans (Drake et al. 1998), we would expect $4N_e\mu = 0.002$. This value is an order of magnitude smaller than our estimates for the intron, suggesting the extension of the recombination breakpoint into the intron.

Evidence from the murine *Mhc* demonstrates that homologous sequence similarity increases the frequency of meiotic crossing over (Sant'Angelo, Lafuse, and Passmore 1992; Yoshino et al. 1995). It may thus be reasonable to postulate that regions of high heterozygosity serve as boundaries for crossing-over events. Sant'Angelo, Lafuse, and Passmore (1992) found that the breakpoint in recombinants occurs in the second intron of the Eb locus, about 100 bp from the boundary of the PBR exon. The distinction among this variety of forces potentially responsible for the observed pattern and magnitude of polymorphism in the intron proves difficult. However, with increasingly realistic models, additional data from the *Agph-DAB1* gene, and data from another species, perhaps the distinction among these forces will come into focus.

Acknowledgments

LITERATURE CITED

AVISE, J. C., R. M. BALL, and J. ARNOLD. 1988. Current versus historical population sizes in vertebrate species with high gene flow: a comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. Mol. Biol. Evol. **5**:331–344.

BERGSTRÖM, T. F., A. JOSEFSSON, H. A. ERLICH, and U. GYLLENSTEN. 1998. Recent origin of HLA-DRB1 alleles and implications for human evolution. Nat. Genet. **18**:237–242.

BROWN, J. H., T. S. JARDETSKY, J. C. GORGA, L. H. STERN, R. G. URBAN, J. L. STROMINGER, and D. C. WILEY. 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DRB1. Nature **364**:33–39.

CEREB, N., A. L. HUGHES, and S. Y. YANG. 1997. Locus-specific conservation of the HLA class I introns by intra-locus homogenization. Immunogenetics **47**:30–36.

DRAKE, J. W., B. CHARLESWORTH, D. CHARLESWORTH, and J. CROW. 1998. Rates of spontaneous mutation. Genetics **148**: 1667–1686.

EDWARDS, S. V., J. GASPER, and M. MARCH. 1998. Genomics and polymorphism of *Agph- DAB1* an *Mhc* class II B gene

in red-winged blackbirds (*Agelaius phoeniceus*). Mol. Biol. Evol. **15**:236–250.

EDWARDS, S. V., and P. W. HEDRICK. 1998. Evolution and ecology of MHC molecules: from genomics to sexual selection. Trends Ecol. Evol. **13**:305–310.

EDWARDS, S. V., J. NUSSER, and J. GASPER. 1999. Characterization and evolution of Mhc genes from non-model organisms, with examples from birds. *In* A. J. BAKER, ed. Molecular methods in ecology. Blackwell Scientific, Cambridge, Mass. (in press).

EDWARDS, S. V., E. K. WAKELAND, and W. K. POTTS. 1995. Contrasting histories of avian and mammalian Mhc genes revealed by class II B sequences from songbirds. Proc. Natl. Acad. Sci. USA **92**:12200–12204.

FELSENSTEIN, J. 1995. PHYLIP (phylogeny inference package). Version 3.57. Distributed by the author, Department of Genetics, University of Washington, Seattle.

GILBERT, D. G. 1996. SeqPup. Biology Department, Indiana University, Bloomington.

GRASER, R., C. O'HUIGIN, V. VINCEK, A. MEYER, and J. KLEIN. 1996. Trans-species polymorphism of class II Mhc loci in danio fishes. Immunogenetics **44**:36–48.

GRASSLY, N. C., and E. C. HOLMES. 1997. A likelihood method for the detection of selection and recombination using nucleotide sequences. Mol. Biol. Evol. **14**:239–247.

GRIMSLEY, C., K. A. MATHER, and C. OBER. 1998. HLA-H: a pseudogene with increased variation due to balancing selection at neighboring loci. Mol. Biol. Evol. **15**:1581–1588.

HEDRICK, P. W., and G. THOMSON. 1983. Evidence for balancing selection at HLA. Genetics **104**:449–456.

HEY, J., and J. WAKELEY. 1997. A coalescent estimator of the population recombination rate. Genetics **145**:833–846.

HUDSON, R. R. 1987. Estimating the recombination parameter of a finite population model without selection. Genet. Res. Camb. **50**:245–250.

———. 1990. Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. **7**:1–44.

HUDSON, R. R., and N. L. KAPLAN. 1988. The coalescent process in models with selection and recombination. Genetics **120**:831–840.

JAKOBSEN, I. B., and S. EASTEAL. 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.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. MUNRO, ed. Mammalian protein metabolism. Academic Press, New York.

KISHINO, H., and M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order of the Hominoidea. J. Mol. Evol. **29**:170–179.

KLEIN, J., R. E. BONTROP, R. L. DAWKINS, H. A. ERLICH, U. B. GYLLENSTEN, E. R. HEISE, P. P. JONES, P. PARHAM, E. K. WAKELAND, and D. I. HAWKINS. 1990. Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics **31**:217–219.

KLEIN, J., Y. SATTA, C. O'HUIGIN, and N. TAKAHATA. 1993. Molecular descent of the major histocompatibility complex. Ann. Rev. Immunol. **11**:269–295.

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

OHTA, T. 1997. Effects of intra-locus recombination on HLA polymorphism. Hereditas **127**:105–112.

———. 1999. Effect of gene conversion on polymorphic patterns at major histocompatibility complex loci. Immunol. Rev. **167**:319–325.

PARHAM, P., and T. OHTA. 1996. Population biology of antigen presentation by MHC class I molecules. Science **272**:67–74.

ROZAS, J., and R. ROZAS. 1997. DnaSP version 2.0: a novel software package for extensive molecular population genetic analysis. Comput. Appl. Biosci. **13**:307–311.

SANT'ANGELO, D. B., W. P. LAFUSE, and H. C. PASSMORE. 1992. Evidence that nucleotide sequence identity is a requirement for meiotic crossing over within the mouse Eb recombinational hotspot. Genomics **13**:1334–1336.

SATTA, Y., W. E. MAYER, and J. KLEIN. 1996. Evolutionary relationship of HLA-DRB genes inferred from intron sequences. J. Mol. Evol. **42**:648–657.

SLATKIN, M. 1996. Gene genealogies within mutant allelic classes. Genetics **143**:579–587.

TAJIMA, F. 1983. Evolutionary relationship of DNA sequences in a finite population. Genetics **105**:437–460.

———. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**:585–595.

———. 1991. Determination of window size for analyzing DNA sequences. J. Mol. Evol. **33**:470–473.

———. 1996. The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. Genetics **143**:1457–1465.

TAKAHATA, N. 1990. A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proc. Natl. Acad. Sci. USA **87**:2419–2423.

TAKAHATA, N., and Y. SATTA. 1998. Footprints of intragenic recombination at HLA loci. Immunogenetics **47**:430–441.

WATTERSON, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7**:256–271.

WITTZELL, H., A. BERNOT, C. AUFFREY, and R. ZOOROB. 1999. Concerted evolution of two Mhc class II B loci in pheasants and domestic chickens. Mol. Biol. Evol. **16**:479–490.

YOSHINO, M., T. SAGAI, K. F. LINDAHL, Y. TOYODA, K. MORIWAKI, and T. SHIROISHI. 1995. Allele-dependent recombination frequency: homology requirement in meiotic recombination at the hot spot in the mouse major histocompatibility complex. Genomics **27**:298–305.

ZOOROB, R., A. BERNOT, D. M. RENOIR, R. CHOUKRI, and C. AUFFREY. 1993. Chicken major histocompatibility complex class II B genes: analysis of interallelic and interlocus sequence variance. Eur. J. Immunol. **23**:1139–1145.