Major Histocompatibility Complex Variation in the Endangered Przewalski's Horse

Philip W. Hedrick, Karen M. Parker, Ellen L. Miller[1] and Philip S. Miller[2]

Department of Biology, Arizona State University, Tempe, Arizona 85287

Manuscript received December 1, 1998
Accepted for publication April 26, 1999

ABSTRACT

The major histocompatibility complex (MHC) is a fundamental part of the vertebrate immune system, and the high variability in many MHC genes is thought to play an essential role in recognition of parasites. The Przewalski's horse is extinct in the wild and all the living individuals descend from 13 founders, most of whom were captured around the turn of the century. One of the primary genetic concerns in endangered species is whether they have ample adaptive variation to respond to novel selective factors. In examining 14 Przewalski's horses that are broadly representative of the living animals, we found six different class II DRB major histocompatibility sequences. The sequences showed extensive nonsynonymous variation, concentrated in the putative antigen-binding sites, and little synonymous variation. Individuals had from two to four sequences as determined by single-stranded conformation polymorphism (SSCP) analysis. On the basis of the SSCP data, phylogenetic analysis of the nucleotide sequences, and segregation in a family group, we conclude that four of these sequences are from one gene (although one sequence codes for a nonfunctional allele because it contains a stop codon) and two other sequences are from another gene. The position of the stop codon is at the same amino-acid position as in a closely related sequence from the domestic horse. Because other organisms have extensive variation at homologous loci, the Przewalski's horse may have quite low variation in this important adaptive region.

THE major histocompatibility complex (MHC) is a fundamental part of the immune system in nearly all vertebrates (*e.g.*, Edwards and Hedrick 1998). MHC genes were first observed by the rejection of tissue and organ transplants from donors that differed genetically from the recipients and have been genetically implicated in a number of human autoimmune diseases. In addition, MHC genes have been suggested as important in mate selection and maternal-fetal interaction (*e.g.*, Hedrick 1994; Edwards and Hedrick 1998). The high variability in many MHC genes, the highest known in any vertebrate genes, is thought to be an essential aspect of the ability of an organism to recognize different parasites (*e.g.*, Apanius *et al.* 1997; Hedrick and Kim 1999; following convention, we use the term parasites to refer to viral, bacterial, protozoan, and other parasites and pathogens). As a result, documenting the amount of MHC variation in a given species should provide insight into its potential resistance or susceptibility to various parasites.

Demonstrating the connection between MHC variation and resistance to parasites is a difficult experimental challenge (Klein and O'Huigin 1994; Hedrick and Kim 1999), but there are several well-documented cases in which specific MHC haplotypes or genotypes provide resistance to parasites (Briles *et al.* 1977; Hill *et al.* 1991; Xu *et al.* 1993; Thursz *et al.* 1997; Paterson *et al.* 1998; Carrington *et al.* 1999). One of the challenges in demonstrating such associations is because the MHC is a multigene family, and it is difficult to separate the effects of specific alleles from the background genotype. In fact, because of the high variability within loci and the similarity of alleles between loci, it has often been difficult to determine which MHC sequences are allelic and which are from other genes. Partly because of this problem, there have been only a few population genetic studies using MHC alleles known by sequence in species other than humans (*e.g.*, Mikko and Andersson 1995a,b; Mikko *et al.* 1997; Fraser and Bailey 1998; Hedrick and Parker 1998; Paterson 1998). Here we document the variation in class II DRB MHC sequences from the endangered Przewalski's horse and have evidence from family, sequence, and other data that these sequences are from two genes, one of which has a nonfunctional allele. Understanding the inheritance of this variation will provide a tool both to elucidate the structure of the DRB region in horses and to examine the connection of genetic resistance to parasites in the Przewalski's horse and related equids.

Corresponding author: Phil Hedrick, Department of Biology, Arizona State University, Tempe, AZ 85287. E-mail: philip.hedrick@asu.edu

[1]*Present address:* Department of Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis, MN 55455-0392.

[2]*Present address:* Conservation Breeding Specialist Group, 12101 Johnny Cake Rd., Apple Valley, MN 55124.

PRZEWALSKI'S HORSE

The Przewalski's horse, also known as the takhi and the Asian, or Mongolian, wild horse (*Equus przewalskii*),

is the only extant species of horse other than the domestic horse, *E. caballus* (Boyd and Houpt 1994). Przewalski's and domestic horses are closely related and, although Przewalski's horse has 66 chromosomes and the domestic horse 64, they will breed and produce viable offspring. In fact, some researchers (*e.g.*, Nowak and Paradiso 1983; Ishida *et al.* 1995) suggest that Przewalski's and domestic horse are both the species *E. caballus* and that the Przewalski's chromosome number is the ancestral one (Benirschke *et al.* 1965; Ryder *et al.* 1978). The Przewalski's horse was once widespread in Eurasia but now is assumed extinct in nature with the last confirmed observation of a wild animal in the 1960s and the last wild captured animal in 1947 (Ryder and Wedemeyer 1982). The captive population of Przewalski's horses, which now numbers >1200 animals in facilities around the world, is descended from 13 animals. Twelve of these founders were Przewalski's horses, 11 of which were captured around 1900, and the other founder a domestic horse mare. In recent years, Przewalski's horses have been released to reserves in Mongolia and China in an effort to reestablish wild populations (Ryder 1993).

The pedigree of the captive population has been examined to determine the effect of over 12 generations in captivity on the expected amount of genetic variation and inbreeding (MacCluer *et al.* 1986; Thompson 1986; Geyer and Thompson 1988; Geyer *et al.* 1989; Ballou 1994). These pedigree studies suggest that the extent of genetic diversity has been greatly reduced, with 60–70% of the original alleles lost because of inbreeding and genetic drift, primarily in the early generations of captivity.

The Przewalski's horse has been examined for the level of genetic variation in blood group and allozyme loci (Bowling and Ryder 1987) and mtDNA (Ryder 1994). The level of heterozygosity in the blood group and allozyme loci was comparable to that found in domestic horses, but the average number of alleles was much lower (Hedrick 1988); part of this reduction may be a reflection of a larger sample size for the domestic horse (Bowling and Ryder 1988). Unlike the amount and pattern of variation observed for blood group loci, allozyme loci, or mtDNA, which should be determined primarily by nonselective factors (although some studies have shown that each of these genetic markers may be influenced by selection), such as inbreeding and genetic drift, the amount of variation at MHC loci should be influenced by selective as well as nonselective factors.

METHODS

Samples of genomic DNA from 14 Przewalski's horses were obtained from O. A. Ryder and L. Chemnick (Center for Reproduction of Endangered Species, Zoological Society of San Diego) and used for polymerase chain reaction. A 250-bp fragment of a MHC class II DRB3 exon 2 was amplified using the following primers: LA31, 5′-GATGGATCCTCTCTCTGCAGCACATTTCC T-3′ and LA32, 5′-CTTGAATTCGCGCTCACCTCGCC GCTG-3′, where the underlined sequences are *Bam*HI and *Eco*RI restriction sites, respectively. These are the same primers used to amplify this gene in cattle (Andersson *et al.* 1991; Sigurdardottir *et al.* 1991) and in the domestic horse (Gustafsson and Andersson 1994). Fraser and Bailey (1996) used these sequences as a basis for designing primer sequences in their study.

Thirty-five cycles of PCR were carried out in a Perkin-Elmer (Norwalk, CT) 9600 thermal cycler with denaturation at 94° for 30 sec, annealing at 50° for 30 sec, and extension at 72° for 30 sec. Each 10-μl reaction contained 40 ng of genomic DNA, 0.5 μm of oligonucleotide primers, 1.5 mm Mg, and 0.5 units of Taq polymerase (Perkin-Elmer/Cetus, Emeryville, CA) as recommended by the manufacturer. The amplified products were digested with *Bam*HI and *Eco*RI (Promega Corporation, Madison, WI) and cloned into either pUC18 or M13mp18 and M13mp19. Single-stranded conformation polymorphism (SSCP) was carried out with the same PCR conditions but included 1 μCi of [^{32}P]dATP in each reaction. Samples were electrophoresed at 4° on a 6% acylamide gel with 2.6% crosslinking at 50 W for 3.5 hr. The gel was transferred to 3MM Whatman paper, dried, and exposed overnight to X-ray film (Fuji RX). Subclones were also analyzed by SSCP to ensure that only those subclones with SSCP patterns identical to the original genomic samples were sequenced. A minimum of two subclones for each allele were sequenced on both strands by the chain termination method using a Sequenase sequencing kit (United States Biochemical Corporation, Cleveland).

The Przewalski's horse has been in captivity for a number of generations, and the total number of individuals over the history of the captive population is ~2000. As a result, analysis of the pedigree is quite complicated, and we have used Pedpack (Thomas 1986) for this purpose. From this software package, we can draw pedigrees, calculated inbreeding coefficients, and determine the expected number of copies of a given founder gene that has survived (gene survival) in a given individual or sample of individuals (see Thomas 1990 for an introduction to the theory used in the calculation of gene survival). Note that the gene survival values are not probabilities.

RESULTS

Pedigree analysis: To assist interpretation of the molecular data, we first present pedigree analysis of the sample of 14 horses. Figure 1 gives the Przewalski's horse pedigree, trimmed (Geyer and Thompson 1988) so that only the 14 sampled animals (gray symbols) and their ancestors including the 13 founders (black sym-

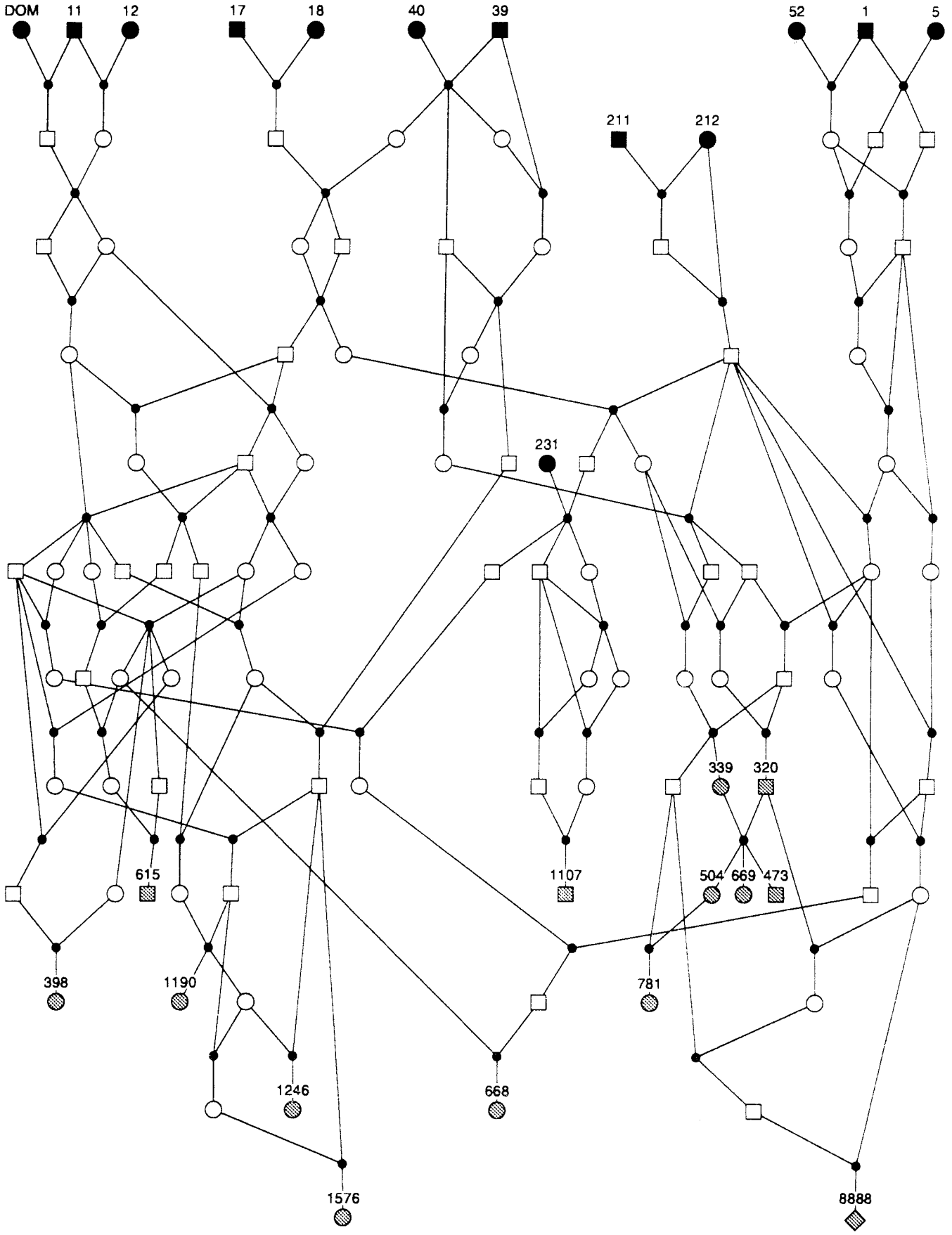

Figure 1.—Pedigree showing the 13 founders of the living Przewalski's horse (black symbols) and the 14 animals (gray symbols) that we sampled for MHC. The small black circles indicate "marriage nodes" and the numbers indicate studbook numbers.

TABLE 1

Studbook number, inbreeding coefficient, and gene survival from the 13 founders of the horses that were sequenced for class II *DRB* gene

		Founder												
No.	*f*	1	5	52	11	12	DOM	17	18	39	40	211	212	231
339[a]	0.207	0.12	0.06	0.06	—	—	—	0.06	0.06	0.31	0.26	0.22	0.62	—
320[a]	0.250	0.12	0.06	0.06	—	—	—	0.06	0.06	0.30	0.25	0.21	0.61	—
504[b]	0.334	0.11	0.05	0.05	—	—	—	0.06	0.06	0.30	0.24	0.20	0.58	—
669[b]	0.334	0.11	0.05	0.05	—	—	—	0.06	0.06	0.30	0.24	0.20	0.58	—
473[b]	0.334	0.11	0.05	0.05	—	—	—	0.06	0.06	0.30	0.24	0.20	0.58	—
1170	0.406	—	—	—	—	—	—	0.10	0.10	0.10	0.10	0.10	0.30	0.80
781	0.365	0.11	0.05	0.05	—	—	—	0.06	0.06	0.29	0.24	0.20	0.57	—
8888	0.340	0.24	0.12	0.12	—	—	—	0.02	0.02	0.13	0.11	0.23	0.67	—
398	0.464	—	—	—	0.48	0.25	0.25	0.14	0.14	0.14	0.14	—	—	—
615	0.405	—	—	—	0.48	0.25	0.25	0.15	0.15	0.15	0.15	—	—	—
1190	0.308	—	—	—	0.41	0.21	0.21	0.15	0.15	0.30	0.24	—	—	—
1246	0.317	—	—	—	0.34	0.17	0.17	0.12	0.12	0.45	0.32	—	—	—
1576	0.304	—	—	—	0.33	0.17	0.17	0.11	0.46	0.46	0.32	—	—	—
668	0.110	0.12	0.06	0.06	0.36	0.19	0.19	0.12	0.12	0.12	0.12	0.08	0.23	0.12

No., studbook number; *f*, inbreeding coefficient.
[a] Parents in the family group examined.
[b] Progeny in the family group examined.

bols) are indicated. In this pedigree representation from Pedpack (Thomas 1986), the pedigree is read from top to bottom, circles represent females, squares represent males, small circles represent "marriage" nodes with the resulting offspring below, and the numbers indicate the studbook numbers.

An attempt was made both to make our sample as representative of founder origin as possible and to have a family group to examine Mendelian segregation of the MHC gene. Table 1 gives, for each individual sampled, the overall inbreeding coefficient and the gene survival (Geyer and Thompson 1988) from each founder. The family group has parents 339 and 320 and progeny 504, 669, and 473. These individuals and 781 have quite similar founder representation, mainly from the Munich group (founders 17, 18, 39, and 40) and the Woburn group (211 and 212). Individual 1107 had high representation (0.80) from the horse (231) captured in 1947 (termed the New Askanova), 8888 had the highest representation from the Old Askanova group (1, 5, and 52), the group of 398, 615, 1190, 1246, and 1576 had representation from the Old Prague group [11, 12, and the domestic horse mare (DOM)], and 668 had representation from all 13 founders. The inbreeding coefficients ranged from 0.110 for 668 to 0.464 for 398 with an average of 0.320.

In examination of the whole captive population, Geyer and Thompson (1988) calculated that the expected number of genes surviving was 10.48 out of the 26 original genes. In our sample of 14 animals, the expected number of genes surviving is 6.19, 59% of the 10.48 total possible. The proportional distribution of these surviving genes over the five groups is not greatly different than in the world population (Geyer and Thompson 1988), *i.e.*, 11, 12, and DOM have proportions of 0.132 in our sample and 0.202 in the world population, 17, 18, 39, and 40 have 0.399 and 0.336, 211 and 212 have 0.304 and 0.166, 1, 5, and 52 have 0.242 and 0.128, and 231 has 0.141 and 0.167.

MHC sequences: Six different sequences were detected (GenBank accession nos. AF084187–AF084192), and the codons that we found variable in Przewalski's horse are given in Table 2. In addition, the amino acids observed for the same sites in the 11 published sequences from the domestic horses are given (Gustafsson and Andersson 1994; Fraser and Bailey 1996). Przewalski's sequences Eqpr-*DRB**1 and Eqpr-*DRB**4 differed by only one amino acid (position 65) and Przewalski's sequences Eqpr-*DRB**1, Eqpr-*DRB**4, Eqpr-*DRB**5, and Eqpr-*DRB**6 were all quite similar and differed from each other by one to four amino acids.

Figure 2 gives the neighbor-joining tree for the 6 Przewalski's nucleotide sequences and the 11 published domestic horse nucleotide sequences with cattle sequence BoLA-DRB3*0101 (Mikko and Andersson 1995a) as an outgroup. First, note that Przewalski's sequences Eqpr-*DRB**1, Eqpr-*DRB**4, Eqpr-*DRB**5, and Eqpr-*DRB**6 cluster together and separately from Przewalski's sequences Eqpr-*DRB**2 and Eqpr-*DRB**3. Domestic horse sequences *DRB**3 and *DRB**7 (we use the previously used designations without the species indicator for these sequences) also were quite similar to the cluster of 4 Przewalski's sequences. In all the sequences, there were only two stop codons and they were in the similar Przewalski's sequence Eqpr-*DRB**6 and domestic sequence *DRB**3 (different by seven amino acids), and

TABLE 2

Amino acid sequence of the sites variable for amino acids in the DRB gene in the 6 Przewalski's horse sequences (Eqpr) and the 11 domestic horse sequences

	10	11 +	12	13 +	16	18	22	23	25	26	27	28 +	30 +	31	32	34	35	37 +	38 +
Eqpr *DRB**1	L	V	K	H	H	S	Q	R	R	F	L	D	Y	F	Y	R	E	Y	V
Eqpr *DRB**2	Y	S	T	F	—	—	—	—	L	Y	—	H	L	—	—	G	K	—	—
Eqpr *DRB**3	Y	A	T	S	—	—	—	Q	—	—	—	—	—	—	—	G	K	T	—
Eqpr *DRB**4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Eqpr *DRB**5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Eqpr *DRB**6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
*DRB**1	Y	S	—	S	—	F	E	—	—	Y	—	—	—	—	—	G	K	—	—
*DRB**2	A	—	K	F	R	—	E	—	—	—	—	E	R	—	H	G	—	—	A
*DRB**3	—	—	K	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
*DRB**4	Y	T	—	S	—	—	—	—	—	—	—	—	—	—	—	G	K	T	—
*DRB**5	Y	S	—	S	—	—	—	—	—	—	—	—	—	—	S	G	—	T	—
*DRB**6	Y	S	—	F	—	—	—	—	L	Y	—	H	L	—	—	G	K	—	—
*DRB**7	—	—	K	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
*DRB**8	—	—	K	—	—	—	—	—	—	—	—	—	—	—	H	G	K	F	—
*DRB**9	Y	S	—	S	—	—	E	—	—	Y	—	—	—	—	—	G	K	—	—
*DRB**10	—	—	K	—	—	—	E	—	x	Y	V	F	D	V	—	—	—	H	—
*DRB**11	Y	S	—	F	—	—	—	—	L	Y	—	H	L	—	—	G	K	—	—

	44	47	50	52	56	57 +	61 +	65	67	69	70 +	71 +	74 +	78 +	84	85	86 +	87	88
Eqpr *DRB**1	V	Y	V	K	T	D	W	K	V	D	D	A	A	Y	G	V	S	D	S
Eqpr *DRB**2	L	—	L	E	P	—	—	Q	I	E	Q	K	E	V	—	—	—	E	—
Eqpr *DRB**3	—	—	—	E	P	—	L	—	—	—	—	—	—	—	—	—	F	—	G
Eqpr *DRB**4	—	—	—	—	—	—	—	Q	—	—	—	—	—	—	—	—	—	—	—
Eqpr *DRB**5	—	—	—	E	R	S	—	—	—	—	—	—	—	—	—	—	—	—	—
Eqpr *DRB**6	—	F	—	E	T	—	#	—	—	—	—	—	—	—	—	—	—	—	—
*DRB**1	—	—	L	E	P	—	—	Q	I	E	Q	K	K	—	A	—	—	E	—
*DRB**2	—	—	—	E	P	—	—	Q	I	—	E	R	—	—	—	—	I	—	G
*DRB**3	—	—	—	—	—	—	#	Q	—	—	—	—	—	—	—	—	T	—	T
*DRB**4	—	—	—	E	P	—	—	—	F	—	—	—	—	L	—	I	—	E	—
*DRB**5	—	—	—	E	P	—	—	—	—	—	—	—	—	—	A	—	—	E	—
*DRB**6	L	—	L	E	R	S	—	Q	I	E	Q	K	E	V	—	—	—	E	—
*DRB**7	—	—	—	E	P	—	—	—	—	—	—	—	—	—	—	—	—	—	—
*DRB**8	—	F	—	E	P	—	—	—	—	—	—	—	Q	—	—	I	—	—	—
*DRB**9	—	—	L	E	R	S	—	Q	I	E	Q	K	K	—	A	—	—	E	—
*DRB**10	—	F	—	E	P	—	—	—	—	E	R	K	—	—	—	—	—	—	N
*DRB**11	L	—	L	E	P	—	—	Q	I	E	Q	K	E	V	x	—	—	E	—

+, an antigen-binding site; #, a stop codon; x, an ambiguous sequence.

the stop codon was at position 61 in both. Note that the Przewalski's horse individuals (399 and 669) with Eqpr-*DRB**6 did not have ancestry from DOM. In addition, Przewalski's sequence Eqpr-*DRB**2 and domestic horse sequence *DRB**11 only differed by one amino acid at position 12, and domestic horse sequence *DRB**6 was very similar to these sequences. Both of these clusters had sequences from both species. Domestic horse sequences *DRB**1 and *DRB**9 differed only by three amino acids.

There were synonymous substitutions at only five codons in the Przewalski's sequences (Table 3), none of which were putative antigen-binding sites (ABS). Three of these were sites in which there were also nonsynonymous substitutions (positions 27, 84, and 85). There was synonymous variation at all five of these positions in the domestic horse sequences. All of the synonymous substitutions found in the Przewalski's horse sequences were also found in the domestic horse sequences.

In the total sequence examined, there were 15 positions that are ABS and 63 that are not (non-ABS). For both the Przewalski's and domestic horse, 12 (80%) of the ABS were variable while for the non-ABS, 16 (25.4%) were variable in the Przewalski's horse and 24 (38.1%) were variable in the domestic horse. There were 3.15 (0.8/0.254) and 2.10 (0.8/0.381) as many variable ABS as non-ABS per site in the Przewalski's and domestic horses, respectively.

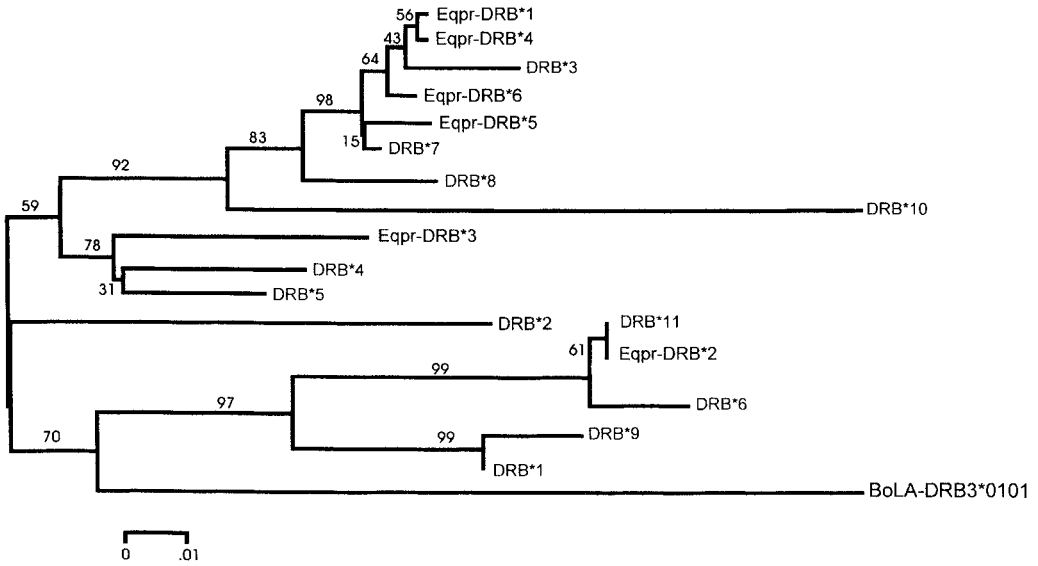

Figure 2.—A neighbor-joining tree for the 6 Przewalski's horse sequences and the 11 domestic horse sequences (the numbers indicate bootstrap values for the internal nodes).

Table 4 gives the proportions of nonsynonymous (d_n) and synonymous (d_s) differences using the Jukes-Cantor correction (Nei and Gojobori 1986; Kumar *et al.* 1993) and their ratio for ABS and non-ABS for the 17 sequences in the two species of horses. For the ABS, the nonsynonymous rate was much higher than the synonymous rate for both species with the d_n/d_s ratio much larger than unity, suggesting selection for diversity in these positions. Also in Table 4 are the two-tailed probabilities that these values are different using a *t*-test (*e.g.*, Zhang *et al.* 1997). For the antigen-binding sites in the domestic horse and for both horses, these are highly significant while for Przewalski's, the probability is only 0.06. However, as Zhang *et al.* (1997) show, these probabilities are somewhat low so the Przewalski's horse ratio is less statistically significant than the 0.06 value suggests. On the other hand, the ratios in both species for non-ABS were less than unity but not significantly different, similar to that found in other species for MHC sequences.

TABLE 3

The five positions in the 6 Przewalski's horse sequences that showed synonymous variation and the 11 sequences in the domestic horse for these positions

	27	45	73	84	85
Eqpr *DRB**1	ctg	ggg	gcc	ggc	gtc
Eqpr *DRB**2	t––	––c	––g	–––	–––
Eqpr *DRB**3	–––	–––	–––	––g	––t
Eqpr *DRB**4	–––	–––	–––	–––	–––
Eqpr *DRB**5	–––	–––	–––	–––	–––
Eqpr *DRB**6	–––	–––	–––	–––	–––
*DRB**1	t––	––c	–––	–c–[a]	–––
*DRB**2	–––	–––	––a	––g	––t
*DRB**3	–––	–––	–––	–––	–––
*DRB**4	–––	–––	–––	–––	a––[a]
*DRB**5	–––	–––	–––	–c–[a]	–––
*DRB**6	t––	––c	––g	–––	–––
*DRB**7	–––	–––	–––	–––	–––
*DRB**8	–––	–––	–––	–––	a––[a]
*DRB**9	t––	––c	–––	–c–[a]	–––
*DRB**10	g–a[a]	–––	–––	––x	–––
*DRB**11	t––	––c	––g	–x–	–––

x, indicates DNA sequence ambiguity.
[a] Codes for a different amino acid in the domestic horse.

From the SSCP analysis and cloning, two different sequences were found in 10 of the horses, three sequences in 2 (1190 and 668), and four sequences in 2 (339 and 669). Finding more than two sequences in an individual suggests amplification of sequence from more than one gene, not unexpected in a multigene family such as the MHC (Ellis *et al.* 1995; Fraser and Bailey 1996). We established a possible explanation for these patterns based on the phylogenetic clustering of the sequences (however, note that the genetic distances between sequences here are not larger than that reported in humans and cattle; *e.g.*, Mikko and Andersson 1995a) and the pattern of sequences in the family group. First, the tight clustering of sequences Eqpr-*DRB**1, Eqpr-*DRB**4, Eqpr-*DRB**5, and Eqpr-*DRB**6 suggests that these sequences are from alleles at a single gene, which we will call gene *A* with alleles A_1, A_4, A_5, and A_6. Remember, because A_6 has a stop codon, there are only three functional alleles at this putative gene.

Second, if it is assumed that the other two sequences, Eqpr-*DRB**2 and Eqpr-*DRB**3, are produced from alleles at a second locus, which we will call gene *B* with alleles B_2 and B_3, then the pattern of variation within the family group can be explained if we assume that individuals that show only one and two sequences for these two putative genes are homozygotes and heterozygotes, respectively. Table 5 shows that the pattern of sequences found in the family group can be explained by segregation of the two-locus haplotypes shown. Hap-

TABLE 4

Estimated rate of nonsynonymous (d_n) and synonymous (d_s) differences (standard error in parentheses) and their ratio in the ABS and non-ABS for the Przewalski's and domestic horses

Region	No. of codons	Differences	*E. przewalskii*	*E. caballus*	Average
ABS	15	d_n	0.200 (0.055)	0.303 (0.064)	0.267 (0.057)
		d_s	0.054 (0.055)	0.025 (0.035)	0.031 (0.034)
		d_n/d_s	3.70 ($P = 0.06$)	12.1 ($P < 0.001$)	8.61 ($P < 0.001$)
Non-ABS	63	d_n	0.078 (0.016)	0.114 (0.017)	0.103 (0.015)
		d_s	0.083 (0.027)	0.162 (0.035)	0.134 (0.031)
		d_n/d_s	0.940 ($P = 0.87$)	0.704 ($P = 0.22$)	0.769 ($P = 0.37$)
Total	78	d_n	0.098 (0.016)	0.145 (0.017)	0.130 (0.016)
		d_s	0.077 (0.024)	0.132 (0.028)	0.112 (0.025)
		d_n/d_s	1.27 ($P = 0.47$)	1.10 ($P = 0.69$)	1.16 ($P = 0.55$)

lotypes are given because it is assumed that these two genes are closely linked, as are other class II MHC genes in mammals, and the pattern of genotypes in the parents and progeny suggests the linkage association given. If these are the correct parental genotypes, then one would expect equal numbers of the two progeny genotypes (assuming no recombination). For three progeny, a ratio of two A_1A_1 B_2B_2 to one A_1A_6 B_2B_3 is as close to equality as possible.

With these assumptions, the two-locus genotypes for the 14 individuals are as given in Table 6. Using these genotypes, the observed frequencies for alleles A_1, A_4, A_5, and A_6 are 0.536, 0.107, 0.286, and 0.071, respectively, and the observed frequencies for alleles B_2 and B_3 are 0.929 and 0.071. Examining the individuals in the pedigree, it appears that alleles A_4 and A_5 are only present in individuals with the highest ancestry from the Old Prague group (11, 12, and DOM) and that allele A_6 is only present in individuals with ancestry from the Woburn group (212 and 231). Similarly, allele B_3 is only present in individuals with substantial ancestry from the Woburn group.

The observed proportions of heterozygotes for putative loci *A* and *B* are 0.286 and 0.071, respectively. Interestingly, for locus *A*, the two individuals (339 and 668) with the lowest inbreeding coefficients are both heterozygotes. To obtain an estimate of the expected heterozygosity (H_E) for the *A* locus, we calculated

$$H_E = (1 - f)(1 - \sum p_i^2),$$

where p_i is the frequency of the *i*th allele at a locus and *f* is the inbreeding coefficient. Because of the apparent separation of our sample into two major groups by both ancestry and alleles at this locus (group 1 being 339, 320, 504, 669, 473, 1107, 781, and 8888 and group 2 being 398, 615, 1190, 1246, 1576, and 668), we calculated the expected heterozygosity and observed average inbreeding coefficient separately for each group (using the individual inbreeding coefficients in Table 1 for groups 1 and 2, the average inbreeding coefficients were 0.321 and 0.318, respectively) and then weighted the expected heterozygosity by the size of the two groups. The resulting expected heterozygosity for locus *A* is 0.223, not significantly different from the 0.286 observed (based on a χ^2 test), and the expected heterozygosity for locus *B* is 0.081, not significantly different from 0.071.

DISCUSSION

Documenting the extent of MHC variation in a population can be a difficult challenge. Here the combination of SSCP to determine how many sequences there are in a given individual, phylogenetic analysis of the sequence data, and segregation of the variants in a family group allowed us to develop a reasonable hypothesis to assign sequences to two specific genes, something that was not possible for earlier work in the domestic horse (Gustafsson and Andersson 1994; Fraser and Bailey 1996). Using these assignments, heterozygosities were calculated for the two putative genes, values that were consistent with predictions from population genetics theory.

Like other studies of MHC in closely related species (*e.g.*, Klein 1987; Gyllensten *et al.* 1990; Mikko *et al.* 1997; Fraser and Bailey 1998), we found that the class

TABLE 5

The putative genotypes and haplotypes at two genes, *A* and *B*, consistent with the sequence patterns observed in the family group (studbook numbers are given in parentheses)

Parents		A_1B_2/A_6B_3 (339) × A_1B_2/A_1B_2 (320)	
Progeny	A_1B_2/A_1B_2 (504)	A_1B_2/A_6B_3 (669)	A_1B_2/A_1B_2 (473)

TABLE 6

The studbook number, the Przewalski's horse sequences for each individual, and the possible genotypes at the two putative genes, *A* and *B*

Studbook number	Sequences	Possible genotype	
		A	*B*
339	1, 2, 3, 6	A_1A_6	B_2B_3
320	1, 2	A_1A_1	B_2B_2
504	1, 2	A_1A_1	B_2B_2
669	1, 2, 3, 6	A_1A_6	B_2B_3
473	1, 2	A_1A_1	B_2B_2
1107	1, 2	A_1A_1	B_2B_2
781	1, 2	A_1A_1	B_2B_2
8888	1, 2	A_1A_1	B_2B_2
398	2, 4	A_4A_4	B_2B_2
615	2, 5	A_5A_5	B_2B_2
1190	1, 2, 5	A_1A_5	B_2B_2
1246	2, 5	A_1A_5	B_2B_2
1576	2, 5	A_5A_5	B_2B_2
668	2, 4, 5	A_4A_5	B_2B_2

II MHC sequences of the Przewalski's and domestic horses were interspersed throughout the phylogenetic tree. Although we did not find any sequences in the Przewalski's horse completely identical to sequences in domestic horses like *DRA* (Albright-Fraser *et al.* 1996), Przewalski's horse sequence Eqpr *DRB**2 and domestic horse sequence *DRB**11 only differed by one amino acid, and several other Przewalski's sequences were very similar to domestic horse sequences. These findings suggest that much of the polymorphism observed in the two species was ancestral to formation of these species and that this polymorphism has been transmitted and maintained in these descendant species, a phenomenon known as trans-species polymorphism (Klein 1987). Ishida *et al.* (1995) found that a Przewalski's mtDNA sequence was within the genetic variation of domestic horses and concluded that the two taxa were only recently diverged. Because similar MHC sequences are often found in quite divergent species, similar conclusions from our data cannot be made. Additionally, the finding that the only stop codon in the Przewalski's horse sequences was at the same exact position as the only stop codon found previously in domestic horse sequences suggests that this stop codon was also ancestral to the divergence of these two species.

As in other species (*e.g.*, Hughes and Nei 1989), we also found that the rate of nonsynonymous substitutions was higher than the rate of synonymous substitutions for putative antigen-binding sites. This finding has been interpreted to support the hypothesis that selection for diversity is very strong at these sites and results in a d_n/d_s ratio larger than unity, the opposite of that expected for purifying selection. Non-ABS, many of which are interspersed with the ABS, show no such pattern and have a ratio slightly lower than unity.

In our sample of 14 individuals, we found six different class II DRB MHC sequences. However, it appears that four of these sequences are at what we tentatively designate as gene *A*, all of which are very similar and one of which was nonfunctional because of a stop codon. The remaining two sequences at what we tentatively designate as gene *B*, Eqpr-*DRB**2 and Eqpr-*DRB**3, then represent the variation at the other gene. However, Eqpr-*DRB**3 was present in only two individuals, one of which was a daughter of the other individual with Eqpr-*DRB**3, so that these sequences were identical by descent with an overall frequency of 0.071, and the observed heterozygosity for this putative gene *B* was only 0.071.

In the domestic horse, there were eight sequences clustering with sequences Eqpr-*DRB**2 and Eqpr-*DRB**3, substantially more variation in a total sample of only six horses than we found in our Przewalski's horse sample. Also, the putatively homologous gene *DRB*3 in cattle has extensive variation with 30 different alleles in a sample of 50 European and African cattle (Mikko and Andersson 1995a). Even moose, which are thought to have low class II MHC variability, had 7 alleles for their *DRB*3 homologue in a sample of 30 Swedish animals and 4 alleles in a sample of 19 Canadian animals (Mikko and Andersson 1995b), although most of these alleles were fairly similar in sequence. Further, in a sample of 20 North American bison, which went through a bottleneck at the end of the nineteenth century, 9 different alleles were found at their *DRB*3 homologue (Mikko *et al.* 1997). In other words, it appears that the Przewalski's horse is quite depauperate in class II MHC variation compared to the domestic horse and other not too distant species.

As suggested by our pedigree analysis, the two main groups of Przewalski's horses showed substantial differentiation in the sequences found. Sequences Eqpr-*DRB**4 and Eqpr-*DRB**5 were only found in horses with a high ancestry from the Old Prague group, and sequences Eqpr-*DRB**3 and Eqpr-*DRB**6 were found in individuals with a high ancestry from the Woburn group. We do not have enough samples to determine which founders or group of founders contributed specific alleles, but we can make an educated guess from the gene survival probabilities for each individual from Table 1 (of course more than one founder may have had each sequence). For example, the most likely source of both Eqpr-*DRB**3 and Eqpr-*DRB**6 is founder 212 with founders 39, 40, and 211 having a reasonable probability of being the source. In this case, because these sequences appear to be from alleles at different loci (see Table 5), this would be the source of the Eqpr-*DRB**3-Eqpr-*DRB**6 haplotype. The most likely source of Eqpr-*DRB**5 is founder 11, with founder 39 also a likely source. Finally, the most likely source of Eqpr-*DRB**4 is also founder 11.

The overall impact of choosing mates in a managed population based on increasing the frequency of particular rare MHC alleles, although recommended by

Hughes (1991), appears to be generally detrimental to maintaining overall variation (Hedrick and Miller 1994). From the above analysis it appears that founder 11 may be the source of two of the sequences found in our study. At one time it was recommended that the contribution of DOM be systematically reduced, but detailed examination of the pedigree showed that nearly all the contribution of DOM was tightly associated with that of founders 11 and 12 (Geyer and Thompson 1988; Geyer *et al.* 1989). If a program to reduce the contribution of DOM were instituted, then the contribution of 11, the likely source of these sequences, and perhaps other genetic variation, would also be reduced.

None of the alleles we found appear to be domestic horse alleles, *i.e.*, descended from DOM. Of course, the sample of individuals from which domestic horse sequences were obtained is relatively small and it is likely that more sequence will be found as the sample size of domestic horses is increased. The six Przewalski's horses with ancestry from DOM had sequences Eqpr-*DRB**1, 2, 4, and 5. Sequences Eqpr-*DRB**1 and Eqpr-*DRB**2 were very common in the other Przewalski's horses, which had no known ancestry from domestic horses. The closest domestic horse sequence to Eqpr-*DRB**4 was *DRB**3, which differed by four amino acids and a stop codon. The closest domestic horse sequence to Eqpr-*DRB**5 was *DRB**7, which differed by three amino acids. Eqpr-*DRB**6, which had a stop codon, does not appear to be a domestic horse allele, even though the stop codon is in the same amino-acid position, because the two individuals with this allele do not have any ancestry from DOM.

Fraser and Bailey (1996) suggest that their findings support the presence of three *DRB* loci in at least some domestic horses. In our sample of 14 Przewalski's horses, 10 animals have only two sequences. Fraser and Bailey (1996) found three different sequences in two different apparently homozygous animals (*DRB**4, 5, and 6 in animal D and *DRB**7, 8, and 9 in animal F), and all six sequences appear to be present in their father. However, in another animal X, not necessarily thought to be homozygous, they found only two sequences. Interestingly, X, which had sequence *DRB**11 that was only one amino acid different from Przewalski's sequence Eqpr-*DRB**2, was an Andalusian stallion (D. G. Fraser, personal communication), an ancient breed of domestic horse. D. G. Fraser (personal communication) also found that another unrelated Andalusian horse and two Przewalski's horses also appeared to have at most two genes. These findings appear to suggest that some domestic horses have three *DRB* genes, other domestic horses have two *DRB* genes, and Przewalski's horses have two *DRB* genes. Such variation in closely related species is not unlikely because variation in the number of DRB loci has been well documented in humans (Gongora *et al.* 1997). Further data is obviously necessary to understand the patterns of *DRB* variation within and between equid species.

We greatly appreciate the samples provided by O. Ryder and L. Chemnick of the San Diego Zoo, the assistance of T. Kim and R. Sheffer with data analysis, T. Kim for producing Figure 2, and the comments of L. Andersson, S. Ellis, and D. Fraser on an earlier version of the manuscript. We thank D. Johnson and M. Carrington for their considerable efforts in attempting to determine the allelism of the different sequences.

LITERATURE CITED

Albright-Fraser, D. G., R. Reid, V. Gerber and E. Bailey, 1996 Polymorphism of *DRA* among equids. Immunogenetics **43:** 315–317.

Andersson, L., S. Sigurdardottir, C. Borsch and K. Gustafsson, 1991 Evolution of MHC polymorphism: extensive sharing of polymorphic sequence motifs between human and bovine *DRB* alleles. Immunogenetics **33:** 188–193.

Apanius, V., D. Penn, P. R. Slev, L. R. Ruff and W. K. Potts, 1997 The nature of selection on the major histocompatibility complex. Crit. Rev. **17:** 179–224.

Ballou, J. D., 1994 Population biology, pp. 93–113 in *Przewalski's Horse: The History and Biology of an Endangered Species*, edited by L. Boyd and K. A. Houpt. State University of New York Press, Albany, NY.

Benirschke, K., N. Malour, R. J. Low and H. Heck, 1965 Chromosome complement: differences between *Equus caballus* and *Equus przewalskii*, Poliakoff. Science **148:** 382–383.

Bowling, A. T., and O. A. Ryder, 1987 Genetic studies of blood markers in Przewalski's horses. J. Hered. **78:** 75–80.

Bowling, A. T., and O. A. Ryder, 1988 Letter to the editor. J. Hered. **79:** 401–402.

Boyd, L., and K. A. Houpt (Editors), 1994 *Przewalski's Horse: The History and Biology of an Endangered Species*. State University of New York Press, Albany, NY.

Briles, W. E., H. A. Stone and R. K. Cole, 1977 Marek's disease: effects of B histocompatibility alloalleles in resistant and susceptible chicken lines. Science **195:** 193–195.

Carrington, M., G. W. Nelson, M. P. Martin, T. Kissner, D. Vlahov *et al.*, 1999 *HLA* and HIV-1: heterozygote advantage and *B*35-Cw*04* disadvantage. Science **283:** 1748–1752.

Edwards, S. V., and P. W. Hedrick, 1998 Evolution and ecology of MHC molecules: from genomics to sexual selection. Trends Ecol. Evol. **13:** 305–311.

Ellis, S. A., A. J. Martin, E. C. Holmes and W. I. Morrison, 1995 At least four MHC class I genes are transcribed in the horse: phylogenetic analysis suggests an unusual evolutionary history for the MHC in this species. Eur. J. Immunol. **22:** 249–260.

Fraser, D. G., and E. Bailey, 1996 Demonstration of three DRB loci in a domestic horse family. Immunogenetics **44:** 441–445.

Fraser, D. G., and E. Bailey, 1998 Polymorphism and multiple loci for the horse DQA gene. Immunogenetics **47:** 487–490.

Geyer, C. J., and E. A. Thompson, 1988 Gene survival in the Asian wild horse (*Equus przewalskii*): I. Dependence of gene survival in the Calgary breeding group pedigree. Zoo Biol. **7:** 313–327.

Geyer, C. J., E. A. Thompson and O. A. Ryder, 1989 Gene survival in the Asian wild horse (*Equus przewalskii*): II. Gene survival in the whole population, in subgroups, and through history. Zoo Biol. **8:** 313–329.

Gongora, R., F. Figueroa and J. Klein, 1997 Complex origin of the *HLA-DR10* haplotype. J. Immunol. **159:** 6044–6051.

Gustafsson, K., and L. Andersson, 1994 Structure and polymorphism of horse MHC class II DRB genes: convergent evolution in the antigen binding site. Immunogenetics **39:** 355–358.

Gyllensten, U. B., D. Lashkari and H. A. Erlich, 1990 Allelic diversification at the class II DQB locus of the mammalian major histocompatibility complex. Proc. Natl. Acad. Sci. USA **87:** 1835–1839.

Hedrick, P. W., 1988 Letter to the editor. J. Hered. **79:** 401.

Hedrick, P. W., 1994 Evolutionary genetics of the major histocompatibility complex. Am. Nat. **143:** 945–964.

Hedrick, P. W., and T. J. Kim, 1999 Genetics of complex polymorphisms: parasites and maintenance of MHC variation, pp. 204–234 in *Genetics, Evolution, and Society*, edited by R. Singh and C. Krimbas. Harvard University Press, Cambridge, MA.

Hedrick, P. W., and P. S. Miller, 1994 Rare alleles, MHC and

captive breeding, pp. 187–204 in *Conservation Genetics*, edited by V. Loeschcke, J. Tomiuk and S. K. Jain. Birkhauser Verlag, Basel, Switzerland.

Hedrick, P. W., and K. M. Parker, 1998 MHC variation in the endangered Gila topminnow. Evolution **52:** 194–199.

Hill, A. V. S., C. E. M. Allson, D. Kwiatdowski, N. M. Anstey, P. Twunasi *et al.*, 1991 Common West African HLA antigens are associated with protection from severe malaria. Nature **352:** 595–600.

Hughes, A. L., 1991 MHC polymorphism and the design of captive breeding programs. Conserv. Biol. **5:** 249–250.

Hughes, A. L., and M. Nei, 1989 Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA **86:** 958–962.

Ishida, N., T. Oyunsuren, S. Mashima, H. Mukoyama and N. Saitou, 1995 Mitochondrial DNA sequences of various species of the genus *Equus* with special reference to the phylogenetic relationship between Przewalskii's wild horse and domestic horse. J. Mol. Evol. **41:** 180–188.

Klein, J., 1987 Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. Hum. Immunol. **19:** 155–162.

Klein, J., and C. O'Huigin, 1994 MHC polymorphism and parasites. Philos. Trans. R. Soc. Lond. B Biol. Sci. **346:** 351–358.

Kumar, S., K. Tamura and M. Nei, 1993 MEGA: molecular evolutionary genetics analysis, version 101. Pennsylvania State University, University Park, PA.

MacCluer, J. W., J. L. Vandeberg, B. Read and O. A. Ryder, 1986 Pedigree analysis by computer simulation. Zoo Biol. **5:** 147–160.

Mikko, S., and L. Andersson, 1995a Extensive MHC class II DRB3 diversity in African and European cattle. Immunogenetics **42:** 408–413.

Mikko, S., and L. Andersson, 1995b Low major histocompatibility complex class II diversity in European and North American moose. Proc. Natl. Acad. Sci. USA **92:** 4259–4263.

Mikko, S., M. Spencer, B. Morris, S. Stabile, T. Basu *et al.*, 1997 A comparative analysis of the Mhc *DRB3* polymorphism in the American Bison (*Bison bison*). J. Hered. **88:** 499–503.

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

Nowak, R. M., and J. L. Paradiso, 1983 *Walker's Mammals of the World*, Ed. 4. Johns Hopkins University Press, Baltimore.

Paterson, S., 1998 Evidence for balancing selection at the major histocompatibility complex in a free-living ruminant. J. Hered. **89:** 289–294.

Paterson, S., K. Wilson and J. M. Pemberton, 1998 Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries* L.) Proc. Natl. Acad. Sci. USA **95:** 3714–3719.

Ryder, O. A., 1993 Przewalski's horse: prospect for reintroduction into the wild. Conserv. Biol. **7:** 13–15.

Ryder, O. A., 1994 Genetic studies of Przewalski's horses and their impact on conservation, pp. 75–92 in *Przewalski's Horse: The History and Biology of an Endangered Species*, edited by L. Boyd and K. A. Houpt. State University of New York Press, Albany, NY.

Ryder, O. A., and E. A. Wedemeyer, 1982 A cooperative breeding program for the Mongolian wild horse *Equus przewalskii* in the United States. Biol. Conserv. **22:** 259–271.

Ryder, O. A., N. C. Epel and K. Benirschke, 1978 Chromosome banding studies of the equidae. Cytogenet. Cell. Genet. **20:** 323–350.

Sigurdardottir, S., C. Borcsch, K. Gustafson and L. Andersson, 1991 Cloning and sequence analysis of 14 DRB1 alleles of the bovine MHC by using the polymerase chain reaction. Anim. Genet. **22:** 199–209.

Thomas, A., 1986 Pedpack: user's manual. Technical Report 99, Department of Statistics, University of Washington, Seattle, WA.

Thomas, A., 1990 Comparison of an exact and a simulation method for calculating gene extinction probabilities in pedigrees. Zoo Biol. **9:** 259–274.

Thompson, E. A., 1986 Ancestry of alleles and extinction of genes in populations with defined pedigrees. Zoo Biol. **5:** 161–170.

Thursz, M. R., H. C. Thomas, B. M. Greenwood and A. V. S. Hill, 1997 Heterozygote advantage for HLA class-II type in hepatitis E virus infection. Nat. Genet. **17:** 11–12.

Xu, A., M. J. T. Van Eijk, C. Park and H. A. Lewin, 1993 Polymorphism on *Bo LA-DRB3* exon 2 correlates with resistance to persistent lymphocytosis caused by bovine leukemia virus. J. Immunol. **151:** 6977–6985.

Zhang, J., S. Kumar and M. Nei, 1997 Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. Mol. Biol. Evol. **14:** 1335–1338.

Communicating editor: G. B. Golding