Molecular Ecology (2002) **11**, 1905–1913

Major histocompatibility complex variation in red wolves: evidence for common ancestry with coyotes and balancing selection

P. W. HEDRICK, R. N. LEE and D. GARRIGAN
Department of Biology, Arizona State University, Tempe, AZ 85287, USA

Abstract

We examined variation at a class II major histocompatibility complex (MHC) gene (*DRB*1) in the captive red wolf population and samples of coyotes from Texas and North Carolina. We found 4 alleles in the 48 red wolves, 8 alleles in the 10 coyotes from Texas and 15 alleles in the 29 coyotes from North Carolina. Two of the four alleles found in red wolves, *Caru*-2 and *Caru*-4, were found in both the Texas and North Carolina coyote samples. Allele *Caru*-1, previously found in gray wolves, was also found in the North Carolina sample. The most frequent red wolf allele, *Caru*-3, was not found in any of the coyote samples. However, an allele found in both the Texas and North Carolina coyote samples is only one nucleotide (one amino acid) different from this red wolf allele. Overall, it appears from examination of this MHC gene that red wolves are more closely related to coyotes than to gray wolves. There were a number of different types of evidence supporting the action of balancing selection in red wolves. Namely, there was: (i) an excess of heterozygotes compared with expectations; (ii) a higher rate of nonsynonymous than synonymous substitution for the functionally important antigen-binding site positions; (iii) an eight times higher average heterozygosity of individual amino acids at the positions identified as part of the antigen-binding site than those not associated with it; (iv) the amino acid divergence of four red wolf alleles was greater than that expected from a simulation of genetic drift; and (v) the distribution of alleles, and the distributions of amino acids at many positions were more even than expected from neutrality. Examination of the level and pattern of linkage disequilibria between pairs of sites suggest that the heterozygosity, substitution and frequencies at individual amino acids are not highly dependent upon each other.

Keywords: adaptive variation, heterozygosity, linkage disequilibrium, phylogenetic tree, substitution rate

Received 6 February 2002; revision received 6 June 2002; accepted 6 June 2002

Introduction

The red wolf, *Canis rufus*, once had a distribution throughout much of the eastern part of the USA (Nowak *et al*. 1995). However, in the early 20th century, the numbers of red wolves declined dramatically because of eradication programmes, habitat destruction, hybridization with coyotes and parasite infestation (McCarley 1962; Nowak 1979). In 1967, red wolves were listed as endangered and by 1970, they remained in only a small area of Texas and Louisiana. A captive breeding programme was initiated in 1974, which now contains contributions from 14 individuals from this remnant population; the natural population became extinct in 1975. The captive population was used to start a reintroduced wild population in eastern North Carolina in 1987 (Phillips *et al*. 1995), which has now grown to approximately 100 individuals.

Molecular genetic studies and analysis have sought to determine the origin of red wolves, i.e. whether they should be considered a separate species or are the result of hybridization between coyotes, *C. latrans*, and gray wolves, *C. lupus*. Wayne and colleagues (Wayne & Jenks 1991; Roy *et al*. 1994, 1996) suggested that red wolves were descended from hybridization between coyotes and gray wolves based on mitochondrial DNA (mtDNA) and microsatellite data. Their data demonstrated that red wolves did not

Correspondence: P. W. Hedrick. E-mail: philip.hedrick@asu.edu

appear to have molecular markers not found in either coyotes or gray wolves, suggesting that red wolves have not existed long enough to evolve unique molecular characteristics. There is evidence of extensive hybridization between red wolves and coyotes prior to the initiation of the captive population (Nowak 1979) but examination of pre-1940 specimens by Roy *et al*. (1996) did not show unique red wolf markers in these individuals. Reich *et al*. (1999) analysed these microsatellite data and suggested that red wolves may be the result of recent (with broad time limits) hybridization.

In contrast, Bertorelle & Excoffier (1998) suggested that the microsatellite data of Roy *et al*. (1994) are consistent with a model in which red wolves and coyotes are descended from the same lineage, without hybridization from gray wolves. Similarly, Wilson *et al*. (2000) suggested that red wolves and coyotes evolved in North America, separate from the old world lineage of gray wolves. Both of these analyses suggested that gray wolves would not have contributed significantly to the current red wolf population, but that the red wolves (and Wilson *et al*. claimed the eastern Canadian wolf) derived for the most part from the coyote lineage.

The major histocompatibility complex (MHC) is one of the most important genetic systems for infectious disease resistance in vertebrates (Edwards & Hedrick 1998; Hedrick & Kim 2000; Hill 2001). The association of disease resistance and MHC variation has often been difficult to document for various reasons (Hedrick & Kim 2000), but a number of human studies have documented its importance for resistance to malaria (Hill *et al*. 1991), hepatitis (Thurz *et al*. 1997) and AIDS (Carrington *et al*. 1999). In addition, O'Brien & Evermann (1988) suggested that organisms with low MHC variation, such as many endangered species, might have high susceptibility to infectious disease. In addition, it has been widely recognized that mortality from pathogens and parasites may be a significant extinction threat (Lyles & Dobson 1993; Laurenson *et al*. 1998; Murray *et al*. 1999; Lafferty & Gerber 2002).

Here we characterize the genetic variation in a highly variable class II MHC gene in red wolves, and for comparison, examine samples from coyotes. Our purposes were twofold. First, we wished to determine if variation at this important nuclear gene would provide any evidence on the relationship among red wolves, coyotes and gray wolves. Second, we wanted to document the extent of genetic variation in endangered red wolves available for an adaptive response to infectious disease and evaluate evidence for balancing selection at this MHC gene.

Materials and methods

Sample sources and DNA techniques

DNA samples from red wolves (48 animals) and coyotes from Texas (10) and North Carolina (29) were obtained from L. Waits, University of Idaho. Most the information on the molecular techniques used here is identical to that provided in Hedrick *et al*. (2000). However, here we used primers from Kennedy *et al*. (1998) to amplify the full-length 280 bp sequence of MHC class II *DRB*1 exon 2, these primers were DRB1F (5′-CCGTCCCCACAGCACATTTC-3′) and DRB1R (5′-TGTGTCACACACCTCAGCACCA-3′). These primers were used both for single strand conformation polymorphism (SSCP) analysis and subsequent cloning. Subclones with the correct size insert were screened by SSCP and clones with band profiles identical to that individual's genomic SSCP pattern were chosen for sequencing. Heterozygotes were identified using SSCP as the sum of the profiles for two allelic subclones. Subclones were sequenced on both strands on an Applied Biosystems 377 automated sequencer (division of Perkin–Elmer, Foster City, CA, USA). Sequences were identical in all replicates and from both strands.

Data analysis

To align the sequences, the sequence editor ESEE Version 3.25 was used. MEGA Version 2.1 (Kumar *et al*. 2001) was used to construct the neighbour-joining (NJ) tree using the genetic distance of Jukes & Cantor (1969) and to obtain bootstrap confidence intervals (1000 replicates). MEGA was also used to calculate the relative rate of nonsynonymous and synonymous substitutions according to Nei & Gojobori (1986) and applying the correction of Jukes & Cantor (1969) for multiple hits. Expected heterozygosity was calculated after Nei (1987) with the small sample size correction. Amino acid heterozygosity for individual sites was calculated by weighting the amino acids in given sequences by their population frequency as in Hedrick *et al*. (1991). The Ewens–Watterson test (Ewens 1972; Watterson 1978) for determining whether the distribution of allele frequencies (and distribution of amino acids at specific positions) were different from that expected by neutrality was tested using a program based on the algorithm of Stewart (1977).

Results

Amount and extent of variation

All the samples amplified successfully except for one coyote sample from North Carolina. SSCP gels resulted in band patterns consistent with four alleles in the red wolves, eight alleles in the Texas sample of coyotes and fifteen alleles in the North Carolina sample of coyotes (Table 1). (To associate alleles with the species in which they were found, we use an identifier in which the first two letters are the first two letters of the genus name and the second two letters the first two letters of the species name.) Two of the

Table 1 The observed frequencies of the 18 different alleles at the MHC *DRB*1 locus in red wolves (boldface *Caru* alleles) and three samples of coyotes (*Cala*) from California (CA) (Hedrick *et al*. 2000), Texas (TX) and North Carolina (NC) where 2*N* is the number of alleles identified. Identical alleles from red wolves, coyotes and gray wolves (*Calu*) are given on the same line. The other *Calu* alleles (six alleles) were gray wolf alleles described in Hedrick *et al*. (2000). The *Cala* alleles with an * in the Texas and North Carolina samples were identified by SSCP but sequences were not obtained

Allele	Gray wolf (2*N* = 26)	Red wolf (2*N* = 96)	Coyote CA (2*N* = 20)	TX (2*N* = 20)	NC (2*N* = 56)
Other Calu	0.731	—	—	—	—
Calu-11, ***Caru*-1**, *Cala*-10	0.269	0.208	—	—	0.054
***Caru*-2**, *Cala*-11	—	0.260	—	0.100	0.054
Caru-3	—	0.396	—	–	—
***Caru*-4**, *Cala*-12	—	0.135	—	0.050	0.036
Cala-1	—	—	0.200	0.250	0.036
Cala-2	—	—	0.100	—	—
Cala-3	—	—	0.400	—	—
Cala-4	—	—	0.050	—	—
Cala-5	—	—	0.100	—	—
Cala-6	—	—	0.050	—	—
Cala-7	—	—	0.050	—	0.054
Cala-8	—	—	0.050	—	—
Cala-13	—	—	—	0.050	—
Cala-14	—	—	—	—	0.143
Cala-15	—	—	—	—	0.036
Cala-16	—	—	—	—	0.018
Cala-17	—	—	—	—	0.018
Cala-18	—	—	—	0.350	0.411
Cala-19*	—	—	—	0.100	0.018
Cala-20*	—	—	—	0.050	—
Cala-21*	—	—	—	0.050	—
Cala-22*	—	—	—	—	0.018
Cala-23*	—	—	—	—	0.018
Cala-24*	—	—	—	—	0.071
Cala-25*	—	—	—	—	0.018

four red wolf alleles were found in the Texas sample of coyotes and three of the four red wolf alleles were found in the North Carolina sample of coyotes. Five alleles were shared in the Texas and North Carolina samples of coyotes.

Compared with expected Hardy–Weinberg proportions (corrected for small sample size), the observed heterozygosities for the red wolves and the coyote samples from Texas were higher, and from the coyote samples from North Carolina and California were lower (Table 2). However, none of the observed heterozygosities in these four samples was statistically significantly different from that expected. The average inbreeding coefficient from pedigree analysis ($\bar{f}$) for the red wolves was 0.044, which lowers the expected heterozygosity**.** We can estimate the expected heterozygosity as

$$H_{e.f} = H_e(1 - \bar{f})$$

where H_e is the small sample size corrected expected heterozygosity. The observed heterozygosity is 0.833 and using the inbreeding data, the expected heterozygosity is 0.690, giving a significant excess of observed heterozygotes over that expected ($\chi^2 = 4.59$, $P < 0.05$).

Sequences were obtained for all 4 alleles found in the sample of 48 red wolves (3 of these we found in our previous sample of 3 red wolves, Hedrick *et al*. 2000 although the sequence here is longer), for 5 of 8 alleles in the Texas sample of coyotes, and for 10 of 15 alleles in the North Carolina sample of coyotes (the amount of DNA was not adequate to obtain sequences for the remaining alleles). None of these new alleles were identical to those described in dogs. The new allele found in red wolves, *Caru*-4, was the least frequent, but the frequency of the four alleles was rather even (Table 1, also see below). Previously, we found that red wolf allele *Caru*-1 was identical to gray wolf allele *Calu*-11 (Hedrick *et al*. 2000). In this further survey, we have now also found this allele in the coyote sample from North Carolina (*Cala*-10). In addition, we have now found red wolf alleles *Caru*-2 and *Caru*-4 in both the coyote samples from Texas and North Carolina (called *Cala*-11 and

Table 2 The observed and expected (using Hardy–Weinberg proportions and small sample size correction) heterozygosities at the *DRB*1 locus for red wolves and the three samples of coyotes where *N* is the sample size. Also given is the expected heterozygosity for red wolves using the observed average inbreeding coefficient ($\bar{f}$) from pedigree data

	Heterozygosity	
Species or sample (*N*)	Observed	Expected
Red wolf (48)	0.833	0.722
Expected using $\bar{f}$	0.833*	0.690
Coyote		
Texas (10)	1.000	0.826
North Carolina (28)	0.714	0.805
California (10)	0.500	0.811

*$P < 0.05$.

Cala-12, respectively). The only red wolf allele not found in coyotes is *Caru*-3, the most common allele at a frequency of 0.396. However, coyote allele *Cala*-18, the most frequent allele found in both Texas (0.350) and North Carolina (0.411) samples, is only one nucleotide (one amino acid) different from red wolf allele *Caru*-3. In other words, only *Caru*-3 appears to be potentially unique to red wolves but it may be the result of recent evolutionary divergence from *Cala*-18. *Caru*-3 may possibly exist in coyotes from Texas and/or North Carolina because *Caru*-3 and *Cala*-18 are not differentiable by SSCP. However, the six sequences we have obtained of this mobility product from coyotes from Texas and North Carolina all have been allele *Cala*-18.

Figure 1 presents a NJ tree with the 28 *DRB* sequences (there are actually 34 sequences in the figure because five are in more than one lineage or taxa). (Sequences 228 bp in length were used here because most of the Mexican wolf, gray wolf and coyote sequences, *Cala*-1 to *Cala*-8, were of this length.) As has been found for MHC genes in other taxa, the sequences for a given taxon are dispersed throughout the phylogenetic tree. In particular, the four sequences from red wolves are widely dispersed in the tree with high bootstrap numbers separating them. As discussed above, for three red wolf alleles an identical coyote allele was found and for the fourth, *Caru*-3, a coyote allele that differed by only one nucleotide was found. However, only for red wolf allele *Caru*-1 was there an identical gray wolf allele. Further, the closest gray wolf allele to each of red wolf alleles *Caru*-2, *Caru*-3, and *Caru*-4 averaged 13.8 nucleotides (9 amino acids) difference.

Table 3 presents the nucleotide and amino acid sequence for the 23 variable codons for the 4 red wolf alleles (the complete sequences for these four alleles and the six new coyote sequences, *Cala*-13 to *Cala*-18, have been deposited in GenBank, Accession nos. AF516916 to AF51625). The average number of amino acid differences between the 4 red wolf alleles was 25 nucleotides (15.5 amino acids) and the range in difference between pairs for nucleotides (19–31) and amino acids (13–19) was not large.

Of the 93 amino acids, 23 (24.7%) were variable and of the 280 nucleotides sequenced, 62 (22.1%) were variable. For one position, 28, the four alleles each had a different amino acid, whereas for eight other positions they had three different alleles. There were two codon positions, positions 73 and 77 that had only silent variation (this same silent variation is also present in both coyotes and gray wolves, Hedrick *et al*. 2000). The amino acid positions, documented using X-ray crystallography in humans to be important in the antigen-binding site (ABS) (Brown *et al*. 1993), are indicated by an asterisk in Table 3. Note that these are different from those given in Hedrick *et al*. (2000) and the correct position numbers are 9, 11, 13, 28, 30, 32, 37, 38, 47, 56, 60, 61, 65, 68, 70, 71, 74, 78, 81, 82, 85, 86, 88 and 89. Of the 24 ABS positions, 14 (58.3%) are variable over the 4 alleles, whereas for the remaining 69 positions that are

Table 3 The nucleotide and amino acid sequence for the 25 variable positions for the four sequences found in red wolves

Position	8	9*	10	11*	13*	16	26	27	28*	30*	32*	37*	47*	57	60*	63	67	70*	71*	73	74*	77	78*	84	86*
Consensus	–tg	gag	atg	ttt	t-c	cat	ttt	ctg	g-g	agc	tat	ttc	tac	g-c	t-c	– gg	ctc	cag	aag	gcc	g-g	ac-	—	ggg	—
	–	E	M	F	F	H	F	L	—	S	Y	F	Y	D	—	—	L	Q	K	A	E	T	—	G	—
Caru-1	g–	t-c	c-	—	cc-	—	–c	g–	–aa	ta-	c–	—	—	– t-	– c-	g–	a–	—	g–	–a	ac-	–c	tac	—	att
	V	Y	Q	—	P	—	—	V	E	Y	H	—	—	V	S	G	I	—	E	—	T	—	Y	—	I
Caru-2	g–	ag-	—	—	gc-	—	—	—	– c-	—	—	—	—	—	– a-	c–	—	—	—	—	– c-	–c	tac	c–	ggc
	V	R	—	—	A	—	—	—	A	—	—	—	—	—	Y	R	—	—	—	—	A	—	Y	R	G
Caru-3	t–	—	—	–a	– t-	—	– a-	—	at-	ga-	—	c-	– t-	—	– a-	g–	—	– g-	—	—	– a-	–g	gtg	—	att
	L	—	—	L	—	—	Y	—	M	D	—	H	F	—	Y	G	—	R	—	—	—	—	V	—	I
Caru-4	t–	—	—	—	– t-	t–	c–	—	– t-	—	—	—	—	at-	– c-	c–	t–	—	– g-	—	– a-	–g	gtg	—	ggc
	L	—	—	—	—	Y	L	—	V	—	—	—	—	I	S	R	F	—	R	—	—	—	V	—	G

*Positions that are thought to be part of antigen-binding site.

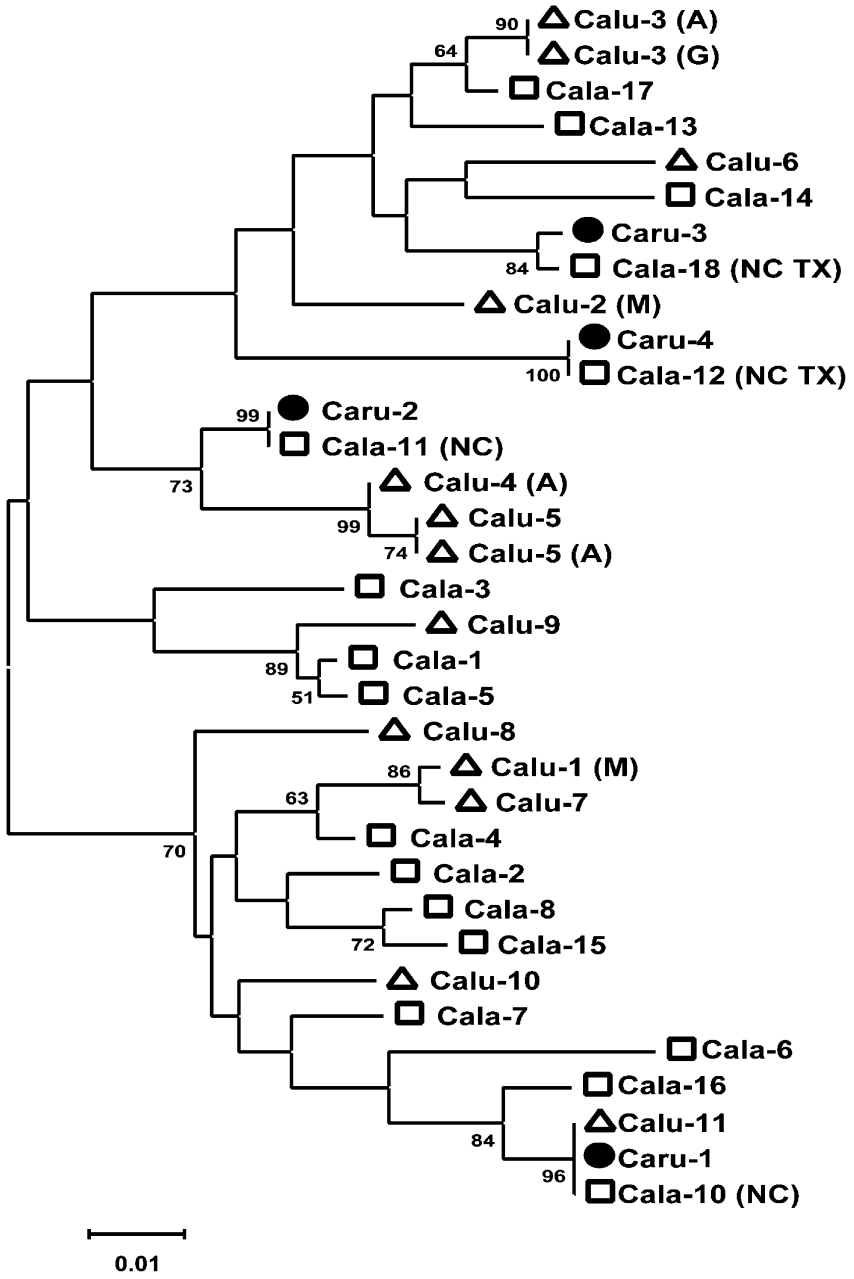

Fig. 1 A neighbour-joining tree (with bootstrap values) showing the four red wolf MHC alleles, *Caru*-1 to *Caru*-4 (●). Also given are the 14 coyote alleles (□) found in the California (CA), Texas (TX) and North Carolina (NC) samples (*Cala* alleles) and the seven gray wolf and six Mexican wolf alleles found (△). The geographical location(s) of the alleles in coyotes identical to those in red wolves and *Cala*-18, which is one nucleotide different from *Caru*-3, are also given. The lineage for the Mexican wolves, McBride (M), Aragon (A) and Ghost Ranch (G) is also given. The vertical lines indicate identical sequences found in different taxa and the scale bar indicates the number of substitutions per site.

not thought to interact with the bound peptide, only 11 (15.9%) are polymorphic and 2 of these are only variable for synonymous differences.

Using the frequencies observed, we can calculate the average heterozygosity for each of the amino acid positions (Fig. 2). The highest heterozygosity was for position 28 (0.714) and 18 amino acid positions have heterozygosities > 0.4. Most of the variation was concentrated in the ABS positions with an average heterozygosity of 0.349, whereas the nonantigen-binding sites had a significantly lower average heterozygosity of 0.043 (12% of that found for ABS positions).

Evidence for balancing selection

Although the excess of observed heterozygotes over Hardy–Weinberg proportions (taking into account inbreeding) and the concentration of heterozygosity in the functionally important ABS positions suggests the existence of balancing selection, we also carried out several specific tests. First, we estimated the rate of nonsynonymous (d_N) and synonymous (d_S) substitutions for ABS and non-ABS amino acid positions (Table 4). For the antigen-binding sites in the red wolf alleles, d_N (0.286) is significantly greater than d_S (0.075) and the ratio d_N/d_S is 3.81. For the

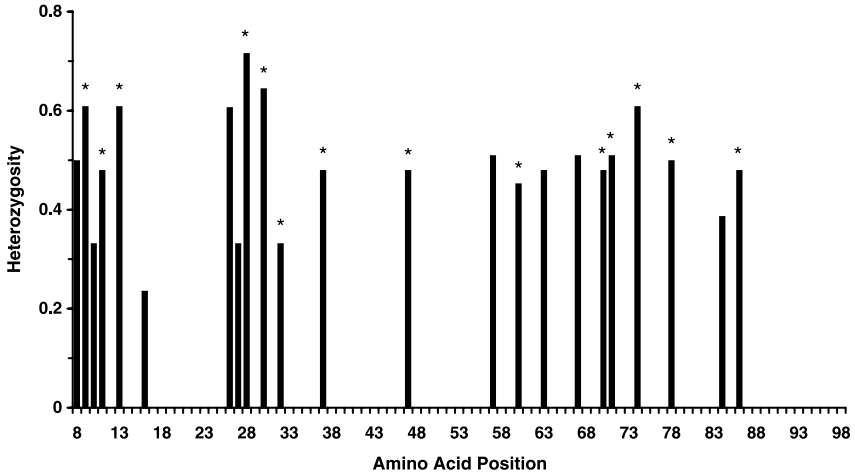

Fig. 2 The heterozygosity at individual amino acid positions in the red wolf. * indicates sites that are putative antigen-binding site positions.

Species	Positions	N	d_N	d_S	d_N/d_S	P
Red wolf	ABS	24	0.286 ± 0.058	0.075 ± 0.041	3.81	< 0.001
	Non-ABS	69	0.047 ± 0.015	0.027 ± 0.015	1.74	0.144
	All	93	0.110 ± 0.024	0.039 ± 0.015	2.82	0.001
Coyote	ABS	19	0.223 ± 0.051	0.080 ± 0.043	2.79	0.004
	Non-ABS	50	0.028 ± 0.011	0.030 ± 0.016	0.93	1.000
	All	69	0.082 ± 0.018	0.043 ± 0.018	1.91	0.044

Table 4 The estimated rates of nonsynonymous and synonymous substitutions for antigen-binding site (ABS) and nonantigen-binding site amino acid positions and their ratio for the four red wolf and 17 coyotes sequences. N is the number of codons in each category and P is the probability that d_N and d_S are different

nonantigen-binding positions, the difference is not significant (similar results were found for the coyote alleles).

Second, we used the Ewens–Watterson test to determine if the frequencies of the red wolf alleles are consistent with neutrality expectation. Using the allele frequencies in Table 1, the expected (Hardy–Weinberg) heterozygosity $(1 - F)$ is 0.714, whereas the expected (under mutation–genetic drift equilibrium) heterozygosity for 4 alleles in a sample of 96 is 0.435. The probability that this difference (a more even distribution than neutrality expectations) would occur by chance is 0.014.

Using the same test and the frequency distribution for individual amino acid sites, we calculated the probability of the distribution occurring by chance for each polymorphic amino acid position in red wolves (Fig. 3). Of the 23 polymorphic amino acid positions, 7 had probabilities of < 0.05 and 6 of these were at ABS positions. The positions with higher probabilities were scattered throughout the gene although one (position 10) was between two ABS positions.

When comparing population values, such as heterozygosity, substitution rates and frequencies, for very closely linked amino acid positions within a gene, the effects of selection on one position may influence variation at another position because they are in linkage disequilibrium (Hedrick 2000). We examined this possibility by calculating the probability of the observed disequilibrium for pairs of the 23 positions polymorphic for amino acids. There were seven pairs of positions for which the probabilities are not significant, $P > 0.05$ (Table 5). The nonsignificant pairs are spread over the whole amino acid sequence and

Table 5 The probability of significance of linkage disequilibrium using Fisher's exact test (Hedrick 2000) between pairs of amino acid positions giving the values > 0.05

	Position							
	8	10	16	27	32*	60*	63	78*
10	—							
16	—	0.064						
27	—	—	0.064					
32*	—	—	0.064	—				
60*	0.057	—	–	—	—			
63	—	—	–	—	—	1.000		
78*	—	—	–	—	—	0.057	—	
86*	—	—	–	—	—	1.000	—	—

*Positions that are thought to be part of antigen-binding site.

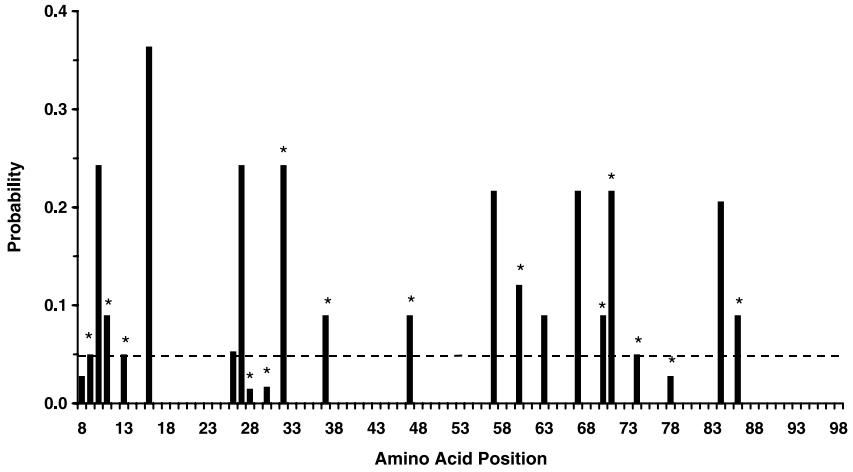

Fig. 3 The probability of significance using the Ewens–Watterson test on the distribution frequency of amino acids at different amino acid positions in the red wolf. *indicates sites that are putative antigen-binding site positions.

some of the nonsignificant pairs are very close to each other, e.g. pair 60 and 63. Further, the nonsignificant pairs contain pairs in which neither, one or both are putative ABS positions. If a simple Bonferroni correction for multiple comparisons is used then, 12 other pairs of positions are nonsignificant.

For four of the nonsignificant pairs (all the nonsignificant combinations involving position 60 in Table 5), all four gametes are present. This is somewhat surprising because with only four alleles, only if each allele has a different combination of amino acid pairs can all four gametes, given two different amino acids at each position, be present. To evaluate the extent of recombination that is necessary to explain these data, we used the four-gamete test of Hudson & Kaplan (1985). This test indicates that there has been a minimum of four recombination events in the history of the four sequences between amino acid positions 8 and 60, 60 and 63, 63 and 78, 78 and 86. However, it must be noted that these putative recombination events have created a patchwork sequential pattern of disequilibrium, in which despite the intervening recombination events, disequilibrium has been restored between positions 8 and 78 and between positions 63 and 86. Overall, this analysis suggests that there is some independence in the population values calculated on data from individual amino acid positions.

Finally, to determine the probability that four such divergent alleles in red wolves could persist by chance, assuming genetic drift, we used Monte Carlo simulation. The simulation starting point was the 28 different alleles given in Fig. 1, all with equal frequencies, and then genetic drift was allowed to reduce the allele number to four (different finite population size scenarios, e.g. a gradual decline in population size or a smaller constant size, gave very similar results). The process was replicated 1000 times and the distribution of the number of pairwise amino acid differences between the remaining four alleles was determined. The simulation results were compared with the observed mean difference (11.32 amino acids) for the four red wolf alleles, weighted by their frequency, for the 207 bp sequence (the longer sequence has not been determined for the coyote alleles found only in the California sample). The mean expected from the simulation results was 5.76 and none of the simulations resulted in a value as high as that observed (Fig. 4). In other words, it appears very unlikely that the four alleles remaining in the red wolves would be as divergent as those observed.

These results are based on several assumptions. First, the initial distribution of the sequences was equal. It is possible that earlier nonrandom events may have made the ancestral frequency of the divergent alleles higher. Second, the measure of amino acid divergence in the population uses the even frequencies observed for calculation. The effect of genetic drift from 28 to 4 alleles results in a distribution similar to that expected under neutrality with a high frequency for the most common allele and a low frequency for the most rare allele. To examine the influence of the allele frequency distribution, we assumed that all the remaining alleles were equal in frequency and calculated the pairwise amino acid difference. The observed value is still higher and only 0.039 of samples of four different alleles resulted in a pairwise difference larger. In other words, when the frequency distribution of the alleles is the same, the amino acid divergence of the alleles is still significantly different from that expected by chance.

Discussion

Ancestry of red wolves

It has been suggested that red wolves are a recent hybrid from gray wolves and coyotes (Wayne & Jenks 1991; Roy *et al*. 1994, 1996) because all the mtDNA haplotypes and microsatellite alleles found in red wolves were also found

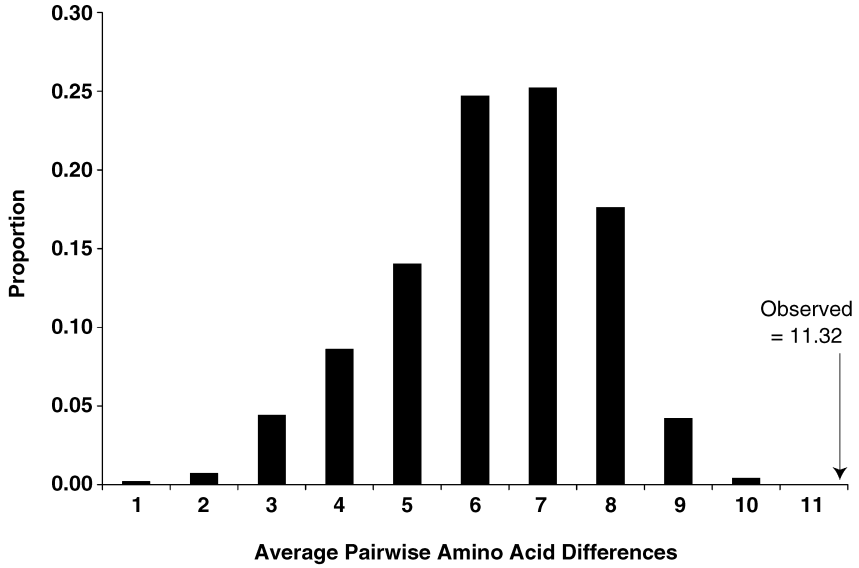

Fig. 4 The frequency distribution of the average number of pairwise amino acid differences from a simulation of genetic drift initiated with the 28 different sequences in Fig. 1 and continued until only four different alleles remain. The pairwise number of amino acid differences between the four extant alleles in red wolves is indicted to the right.

in either gray wolves or coyotes. However, in our data set of MHC sequences, only one of the four red wolf alleles was found in the gray wolves and three were found in the coyote samples. In addition, there was a coyote allele that was only one nucleotide (one amino acid) different from the fourth red wolf allele. Therefore, on average the closest coyote allele was only 0.025 nucleotides different from red wolf alleles, whereas the closest gray wolf allele was 13.8 nucleotides different. In other words, it appears that the lineage leading to the red wolf contained a larger proportion of coyote than gray wolf ancestry, supporting the suggestions of Bertorelle & Excoffier (1998) and Wilson *et al*. (2000).

Previously (Hedrick *et al*. 2000), we had only examined coyotes from California and did not find any red wolf alleles in that sample. It now appears that this population was more isolated from red wolves than the Texas and North Carolina samples that we examined in this study. It is possible that a further examination of gray wolves may reveal other alleles identical or more similar to the red wolf alleles. However, our previous Mexican wolf (Mexican wolves are the closest extant geographical wolf taxa to the last red wolf population in the USA) study did not reveal any red wolf alleles in Mexican wolves. Another possibility is to examine eastern Canadian gray wolves although the relationship of these animals to red wolves, coyotes and other gray wolves is controversial (Wilson *et al*. 2000).

Evidence for balancing selection

We found evidence for balancing selection for the red wolf alleles from several different observations and tests. First, we found an excess of heterozygotes over that expected when the observed level of inbreeding was taken into account. Second, we found an eightfold higher heterozygosity at amino acid positions in the functionally important ABS than in non-ABS positions. Third, we found a significantly higher estimate of nonsynonymous than synonymous substitution for the amino acid positions in the ABS positions, whereas there was no difference in the non-ABS positions. Fourth, we found a more even distribution of alleles than expected from neutrality and this evenness was also significant at 7 of the 23 polymorphic amino acid positions. Finally, the 4 red wolf alleles are more divergent than expected when drawn by chance from the 28 different alleles we have characterized.

Overall, there is strong evidence that balancing selection has been important at this gene and from the observed excess of heterozygotes, this selection even appears important in the current generation. However, the captive red wolf population has been managed to minimize mean kinship (Ballou *et al*. 1995) and it is possible that this management may have influenced genotypic frequencies and the distribution of allele frequencies. One check on this effect is to examine other loci that are thought to be neutral. A study is in progress to examine 19 microsatellite loci in these same red wolf individuals (J. Adams and L. Waits, personal communication). Preliminary results suggest that there is no significant excess of heterozygotes and the distribution of alleles is consistent with neutrality for these loci. In other words, it appears that there is no general effect on these observations from the breeding programme that influences neutral loci. There are detailed necropsy results on some captive red wolves (Acton *et al*. 2000) that potentially may be used in combination with MHC genetic information to examine how selection has operated recently to influence genotypic frequencies.

Acknowledgements

We appreciate the samples provided to us by J. Adams and L. Waits and to R. Fredrickson for calculating inbreeding coefficients. This research was supported by a grant from United States Fish and Wildlife Service and the Ullman Professorship to PWH.

References

Acton AE, Munson L, Waddell WT (2000) Survey of necropsy results in captive red wolves (*Canis rufus*), 1992–96. *Journal of Zoo and Wildlife Medicine*, **31**, 2–8.

Ballou JD, Gilpin M, Foose TJ, eds. (1995) *Population Management for Survival and Recovery*. Columbia University Press, New York.

Bertorelle G, Excoffier L (1998) Inferring admixture proportions from molecular data. *Molecular Biology and Evolution*, **15**, 1298–1311.

Brown JH, Jardetzky TS, Gorga JC *et al.* (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature*, **364**, 33–39.

Carrington M, Nelson GW, Martin MP *et al.* (1999) *HLA* and HIV-1: heterozygote advantage and *B*35-Cw*04* disadvantage. *Science*, **238**, 1748–1752.

Edwards S, Hedrick PW (1998) Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends in Ecology and Evolution*, **13**, 305–311.

Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3**, 87–112.

Hedrick PW (2000) *Genetics of Populations*, 2nd edn. Jones and Bartlett, Boston.

Hedrick PW, Kim TJ (2000) Genetics of complex polymorphisms: parasites and maintenance of the major histocompatibility complex variation. In: *Evolutionary Genetics: from Molecules to Morphology*, pp. 204–234 (eds Singh RS, Krimbas CB). Cambridge University Press, Cambridge.

Hedrick PW, Lee RN, Parker KM (2000) Major histocompatibility complex (MHC) variation in the endangered Mexican wolf and related canids. *Heredity*, **85**, 617–624.

Hedrick P, Whittam TS, Parham P (1991) Heterozygosity at individual amino acid sites: extremely high levels for *HLA-A* and *-B* genes. *Proceedings of the National Academy of Sciences of the USA*, **88**, 5897–5901.

Hill AVS (2001) The genomics and genetics of human infectious disease susceptibility. *Annual Reviews of Genomics and Human Genetics*, **2**, 373–400.

Hill AVS, Allsop CEM, Kwiatdowski D *et al.* (1991) Common West African HLA antigens are associated with protection from severe malaria. *Nature*, **352**, 595–600.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: *Mammalian Protein Metabolism* (ed. Munro HN), pp. 21–132. Academic Press, New York.

Kennedy LJ, Carter SD, Barnes A *et al.* (1998) Nine new dog *DLA-DRB1* alleles identified by sequence-based typing. *Immunogenetics*, **48**, 296–301.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: *Molecular Evolutionary Genetics Analysis*, *Version 2.1*. Arizona State University, Tempe, AZ, USA.

Lafferty KD, Gerber L (2002) Good medicine for conservation biology: the intersection of epidemiology and conservation theory. *Conservation Biology*, **16**, 593–604.

Laurenson K, Sillero-Zubiri C, Thompson H, Shiferaw F, Thirgood S, Malcolm J (1998) Disease as a threat to endangered species: Ethiopian wolves, domestic dogs and canine pathogens. *Animal Conservation*, **1**, 273–280.

Lyles AM, Dobson AP (1993) Infectious disease and intensive management: population dynamics, threatened hosts, and their parasites. *Journal of Wildlife Medicine*, **24**, 315–326.

McCarley H (1962) The taxonomic status of wild *Canis* (Canidae) in south central United States. *Southwestern Naturalist*, **7**, 227–235.

Murray DL, Kapke CA, Evermann JF, Fuller TK (1999) Infectious disease and the conservation of free-ranging large carnivores. *Animal Conservation*, **2**, 241–254.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

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

Nowak RM (1979) *North American Quaternary Canis*. University of Kansas Museum Natural History Monograph no. 6.

Nowak RM, Phillips MK, Henry VG, Hunter WC, Smith R (1995) The origin and fate of the red wolf. In: *Ecology and Conservation of Wolves in a Changing World* (eds Carybn LN, Fritts SH, Seip DR), pp. 409–415. Canadian Circumpolar Institute Occasional Publication no. 35.

O'Brien SJ, Evermann JF (1988) Interactive influence of infectious disease and genetic diversity in natural populations. *Trends Ecology and Evolution*, **3**, 254–259.

Phillips MK, Smith R, Henry VG, Lucash C (1995) Red wolf reintroduction program. In: *Ecology and Conservation of Wolves in a Changing World* (eds Carybn LN, Fritts SH, Seip DR), pp. 157–168. Canadian Circumpolar Institute Occasional Publication no. 35.

Reich DE, Wayne RK, Goldstein DB (1999) Genetic evidence for a recent origin by hybridization of red wolves. *Molecular Ecology*, **8**, 139–144.

Roy MS, Geffen E, Smith D, Ostrander EA, Wayne RK (1994) Patterns of differentiation and hybridization in North American wolflike canids, revealed by analysis of microsatellite loci. *Molecular Biology and Evolution*, **11**, 553–570.

Roy MS, Geffen E, Smith D, Wayne RK (1996) Molecular genetics of pre-1940 red wolves. *Conservation Biology*, **10**, 1413–1424.

Stewart FM (1977) Computer algorithm for obtaining a random set of allele frequencies for a locus in an equilibrium population. *Genetics*, **86**, 482–483.

Thurz MR, Thomas HC, Greenwood BM, Hill AVS (1997) Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nature Genetics*, **17**, 11–12.

Watterson GA (1978) An analysis of multi-allelic data. *Genetics*, **88**, 171–179.

Wayne RK, Jenks SM (1991) Mitochondrial DNA analysis implying extensive hybridization of the endangered red wolf *Canis rufus*. *Nature*, **351**, 565–568.

Wilson PJ, Grewal S, Lawford ID *et al.* (2000) DNA profiles of the eastern Canadian wolf and the red wolf provide evidence for a common evolutionary history independent of the gray wolf. *Canadian Journal of Zoologist*, **78**, 2156–2166.