

SHORT COMMUNICATION

Genetic evidence for a recent origin by hybridization of red wolves

D. E. REICH,* R. K. WAYNE† and D. B. GOLDSTEIN*

*Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, UK, †Department of Biology, University of California at Los Angeles, Los Angeles, CA 90095, USA

Abstract

Genetic data suggest that red wolves (*Canis rufus*) resulted from a hybridization between coyotes (*C. latrans*) and grey wolves (*C. lupus*). The date of the hybridization, however, is uncertain. According to one hypothesis, the two species came into contact as coyotes increased their geographical range in conjunction with the advance of European settlers and as grey wolves were extirpated from the American south. Alternatively, the red wolves could have originated tens of thousands of years ago as a result of climate and habitat changes that disturbed the ecology of the two parent species. To obtain an upper limit on the date of hybridization that would help to distinguish the two scenarios, we compared microsatellite allele length distributions from red wolves, coyotes and grey wolves. Subject to the assumptions of our analysis, we conclude that the red wolves originated as a result of hybridizations that occurred during the past 12 800 years, and probably during the past 2500 years.

Keywords: canids, hybridization, microsatellites, population genetics, red wolves

Received 30 April 1998; revision received 13 July 1998; accepted 13 July 1998

Introduction

Three wolf-like canids in North America have the potential to hybridize: the grey wolf (*Canis lupus*), the coyote (*C. latrans*) and the red wolf (*C. rufus*) (Kolenosky & Standfield 1975; Nowak 1979; Lehman *et al.* 1991; Wayne & Jenks 1991). Analyses of mtDNA diversity and microsatellite polymorphisms of recent and historical populations indicate that red wolves arose as hybrids between coyotes and grey wolves (Wayne & Jenks 1991; Roy *et al.* 1994, 1996). According to one hypothesis, the hybridizations occurred as recently as several hundred years ago as a result of habitat disruptions caused by European settlers (Nowak 1979; Wayne & Jenks 1991). This scenario gains plausibility in light of the analogous hybridizations between coyotes and grey wolves that have been observed during the past 90 years in the Great Lakes region (Lehman *et al.* 1991). Alternatively, the hybridizations could be ancient, as might be supposed from Pleistocene fossil records of morphologies appar-

ently intermediate between coyotes and grey wolves (Nowak 1979; Kurtén & Anderson 1980). To distinguish between theories of recent and older hybridization, we analysed polymorphisms in 10 unlinked dinucleotide microsatellite loci from 144 coyotes, 141 grey wolves, and 56 red wolves (data reported in Roy *et al.* 1994, 1996). The coyotes come from Washington State (25), Alaska (14), Alberta (20), Minnesota (22), Maine (20) and California (43). The grey wolves come from Vancouver (15), Alaska (19), Alberta (20), Minnesota (20), Quebec (37) and the Northwest Territories (30). The red wolves come from 39 captives from east Texas and 17 museum specimens. Among the 111 distinct alleles observed, 12 were unique to coyotes, 16 to grey wolves and none to red wolves. The lack of private alleles among the red wolves indicates that few mutations have occurred since hybridization and suggests an approach for calculating an upper limit on the date of hybridization.

Analysis and Results

To perform the analysis, we focused on eight alleles at four loci that are definitely 'allowed' in terms of the

Correspondence: D. B. Goldstein. Fax: +44 1865310447; E-mail: david.goldstein@zoology.ox.ac.uk

microsatellite mutation process but are absent in the parental populations. In order to restrict our study to such alleles, we consider only allele lengths that are internal to the distributions within the parental populations. This procedure identifies alleles that we can be sure are not forbidden by range constraints because we have observed alleles in the population that are both longer and shorter. By counting the number of such alleles observed among the red wolves (in fact, none was observed), and by considering the frequencies of neighbouring alleles that could have given rise to them through mutation, we set an upper limit on the proportion of lineages, i.e. lines of descent from the parental populations into the hybrids, that underwent mutation since hybridization. We thereby set an upper limit on the date of hybridization.

To estimate the probability (P_i) that any mutation among the red wolves produces an allele of interest (i), we consider a weighted average of neighbouring alleles at the same locus:

$$P_i = b_{-1} p_{i-1} + b_1 p_{i+1} + b_{-2} p_{i-2} + b_2 p_{i+2} + b_{-3} p_{i-3} + b_3 p_{i+3} + \dots \quad (1)$$

where p_{i-j} denotes the frequency of allele $i-j$, and the b_j values are the probabilities that if a mutation occurs at allele $i-j$, it would be converted into allele i .

To assign a value to P_i in eqn 1, we must choose the b_j values and estimate the p_{i-j} values from the data. To estimate the b_j values, we consider single and multistep mutations separately, beginning with the assumption that 85% of microsatellite mutations are single step (the range is 80–95% in humans, according to Di Rienzo *et al.* 1994). If these mutations are unbiased in their direction, the overall probability that the mutation will produce allele i is $b_1 = b_{-1} = (0.85)(0.5)$. As multistep mutations occur $\approx 15\%$ of the time, and assuming for simplicity that they only occur in step sizes of 2 and 3 with equal probabilities of 7.5% each, we use similar reasoning to obtain $b_2 = b_{-2} = b_3 = b_{-3} = (0.075)(0.5)$, $b_{|j| > 3} = 0$. We therefore have the equation:

$$P_i = (0.85)(0.5)(p_{i-1} + p_{i+1}) + (0.075)(0.5)(p_{i-2} + p_{i+2}) + (0.075)(0.5)(p_{i-3} + p_{i+3}) \quad (2)$$

In Table 1, we describe results for models in which 55%, 70% and 100% of mutations are single step.

To estimate the p_{i-j} values, we use the proportions of alleles observed in the red wolf population. This is justified because the red wolves are more closely related to their ancestors (also red wolves), than modern coyotes and grey wolves. Moreover, if significant genetic drift had occurred around the time of hybridization, the red wolf allele frequencies would certainly provide a better reflection of the post-hybridization frequencies than the modern coyotes and grey wolves. To assess the robustness of our date estimates to the choice of population used to estimate the p_{i-j} values, we also calculate dates based on the

Table 1 Estimates of P_i according to several mutation models. In the first four rows, p_{i-j} values are calculated from the frequencies of alleles in the red wolf sample, and we vary the proportion of mutations that are single step (100%, 85%, 70% and 55%). In the next four rows, p_{i-j} is calculated using mixtures of coyotes and grey wolves in proportions of 50%–50%, 70%–30% and 90%–10%, and we fix the percentage of mutations that are single step at 85%. In the last row of the table, we show P_i obtained from a K-allele model

Fixed condition	Variable condition	P_i values (labelled by locus ID/ allele size)										
		172/104	172/106	172/112	204/164	204/170	225/134	250/108	250/116			
Stepwise Model p_{i-j} values arecalculated from red wolves	100% of mutations are single step	0.217	0.011	0.011	0.051	0	0	0	0	0	0	0
	85% of mutations are single step	0.185	0.026	0.029	0.077	0.002	0.011	0.007	0	0	0	0
	70% of mutations are single step	0.153	0.040	0.047	0.103	0.003	0.022	0.014	0	0	0	0
	55% of mutations are single step	0.122	0.055	0.065	0.129	0.005	0.033	0.021	0	0	0	0
Stepwise Model 85% of mutations are single step	P_{i-j} from 50% coyotes, 50% grey wolves	0.125	0.012	0.047	0.148	0.020	0.019	0.006	0.009			
	P_{i-j} from 70% coyotes, 30% grey wolves	0.157	0.015	0.038	0.122	0.012	0.024	0.008	0.006			
	P_{i-j} from 90% coyotes, 10% grey wolves	0.190	0.017	0.029	0.095	0.004	0.029	0.010	0.002			
	P_{i-j} from red wolves (same as row 2)	0.185	0.026	0.029	0.077	0.002	0.011	0.007	0			
K-Allele Model	Not applicable	0.111	0.111	0.111	0.167	0.167	0.111	0.067	0.067			

coyote and grey wolf allele frequencies. It is clear from the molecular and archaeological evidence that coyotes contributed most of the genes to red wolves (Nowak 1979; Roy *et al.* 1994, 1996), and hence we use 50%–50%, 70%–30% and 90%–10% mixtures of coyote and grey wolves (Table 1). All three mixtures point to a somewhat more recent upper limit for the date of hybridization than the results obtained from red wolf allele frequencies (Table 2), and we conclude that we are not only being more accurate, but also conservative in using the red wolves to estimate the p_{i-j} values.

Thus far in our analysis, we have assumed a stepwise model to describe microsatellite mutation (Goldstein & Pollock 1997). For the purpose of comparison, however, it may also be worth considering an alternative model in which there are K allowed states at a locus, and in which a mutation has an equal probability of producing each state. To obtain a value for P_i , we assume $p_{i\pm j} = \frac{1}{K-1}$ for all $i \pm j$ within the allowed ranges. To estimate K , we count the number of states between the maximum and minimum allele lengths observed at a locus. The associated dates of hybridization for this model turn out to be substantially more recent than the dates obtained using the previously described 'stepwise mutation model' (Table 2).

We may now proceed with the analysis using the P_i values obtained by the various methods described above. For each allele of interest (i), the expected frequency in the red wolf sample (R_i), is determined by the proportion of lineages that have undergone one or more mutations (k) times the probability (S_i) that the mutations will produce the allele of interest. Given the scarcity of mutations among the red wolves, however, it seems that a lineage

will probably at most be associated with a single mutation, and P_i will be a good approximation for S_i . To obtain an upper limit on k , we then use the relation $R_i = kS_i \approx kP_i$ and observe that none of the eight alleles of interest is present in the red wolf sample. Supposing that n_i is the effective number of lineages (see below) at the locus at which allele i occurs, the probability of obtaining as few as 0 out of 8 alleles is the Cartesian product

$$Q(k) = \prod_{i=1}^8 (1 - R_i)^{n_i} \approx \prod_{i=1}^8 (1 - kP_i)^{n_i} \quad (3)$$

An upper limit for k at the 5% level is the value of k such that $Q(k) = 0.05$.

To use eqn 3, we require an estimate for the effective number of lineages n_i . We note that n_i is at least the number of parental lineages that are inherited at the locus, but is probably considerably larger because many of the ancestral lineages could have bifurcated (doubled) soon after they were introduced. For simplicity, we assume that the effective number of lineages is the same at all loci, i.e. $n_i = n$. Using the fact that 10 distinct mitochondrial DNA haplotypes were observed among 18 red wolves that were surveyed (Wayne & Jenks 1991; Roy *et al.* 1996), we infer that n is at least 10 and probably considerably larger as males will have contributed something. To obtain a more stringent restriction on n , we consider whether each of the 111 alleles that occurs at non-zero frequency r_j in the parental populations is absent or present among the red wolves. If the allele is absent, the probability of the observed result is $B_j = (1 - r_j)^n$, while if the allele is present, the probability of the observed result is $B_j = 1 - (1 - r_j)^n$. The overall likeli-

Table 2 Upper limits at the 5% level on k , $t\mu$ and years since hybridization. Different ways of estimating P_i and the mutation rate, described in the caption to Table 1, always lead to the conservative conclusion that hybridization occurred less than 15 300 years ago. These results are based on a Channel Island fox mutation rate of 3.7×10^{-5} per year, a human mutation rate of 1.9×10^{-4} per year, and an effective number of parental lineages (n) of 23. In making these calculations, we do not take into account error in our estimation of the mutation rate, or error in our extrapolation of $t\mu$ from k . Note that the quoted upper limit on the date of hybridization (12 800 years) is based on the P_i values from the second row of Table 1

Fixed condition	Variable condition	Limit on k	Limit on $t\mu$	Limit on years (mutation rate from fox)	Limit on years (mutation rate from humans)
Stepwise Model p_{i-j} values are calculated from red wolves	100% of mutations are single step	< 0.432	< 0.566	< 15 300	< 3000
	85% of mutations are single step	< 0.377	< 0.474	< 12 800	< 2500
	70% of mutations are single step	< 0.335	< 0.408	< 11 000	< 2100
	55% of mutations are single step	< 0.299	< 0.355	< 9600	< 1900
Stepwise Model 85% of mutations are single step	p_{i-j} from 50% coyotes, 50% grey wolves	< 0.331	< 0.403	< 10 900	< 2100
	p_{i-j} from 70% coyotes, 30% grey wolves	< 0.335	< 0.407	< 11 000	< 2100
	p_{i-j} from 90% coyotes, 10% grey wolves	< 0.339	< 0.414	< 11 200	< 2200
	p_{i-j} from red wolves (same as row 2)	< 0.377	< 0.474	< 12 800	< 2500
K-Allele Model	Not applicable	< 0.142	< 0.153	< 4100	< 800

hood of the observed distribution of alleles (taking the product over 111 terms) is:

$$L(n) = \prod_{j=1}^{111} B_j \quad (4)$$

The value of n for which $L(n)$ takes the largest value, that is, the maximum likelihood estimate \hat{n} , is 23 regardless of the mixture of coyotes and grey wolf populations (50%–50%, 70%–30% or 90%–10%) we use to estimate the r_j values. Note that in making this calculation, we are implicitly assuming that all red wolf alleles are inherited from the parental populations, while in principle some may have arisen via new mutations (homoplasy). In fact, it seems unlikely that a large number of red wolf alleles arose via new mutations, as no private alleles are now observed among the red wolves; however, if we did adjust for new mutations, our estimate for the effective number of lineages would be slightly lower than 23.

Finally, to assess the error in our estimate of n , we consider the log-likelihood ratio

$$R(n) = 2 \left(\frac{\ln(L(\hat{n}))}{\ln(L(n))} \right) \quad (5)$$

$R(n)$ is expected to be distributed approximately χ^2 with 1 degree of freedom (Sokal & Rohlf 1981), and we can find a 95% central confidence interval for n by identifying the smallest and largest values of n such that $R(n) < \chi^2_{1,0.05} = 3.84$. Regardless of the mixture of coyotes and grey wolf populations we use to estimate the allele frequencies (r_j values), we conclude that $15 \leq n \leq 35$.

Following the inferences about the effective number of lineages, we will now assume a model in which $n = 23$ lineages have been inherited from the parental populations, and subsequent bifurcations have occurred too recently to develop new mutations. To set an upper limit on the number of generations since hybridization (t) as a function of the mutation rate (μ), we begin by using the criterion

$$Q(k) \approx \prod_{i=1}^8 (1 - kP_i)^n = 0.05 \quad (6)$$

With P_i values taken from the second row of Table 1, and using 23 as the most probable estimate for n , we conclude that 0.377 is an upper limit on k at the 5% level. To relate k , the proportion of lineages on which one or more mutations occurred, to t , we note that $e^{-t\mu}$ is the first term in a Poisson series with parameter $t\mu$, and that it gives the probability of no mutations occurring on a lineage. Then, as $1 - e^{-t\mu}$ is the expected value for k , we can use $k \approx 1 - e^{-t\mu}$ to extrapolate an upper limit on $t\mu$ of 0.474. Of course, the expected proportion of lineages containing mutations, $1 - e^{-t\mu}$, could be larger than the proportion (k) that was actually observed. To ascertain the degree to which this sampling error could raise the upper limit on $t\mu$, we make the conservative assumption that k is equal to its largest

possible value of 0.377, and that there are 23 distinct lineages in the red wolf sample. From the binomial distribution, we expect $(n)k = (23)(0.377) = 8.7$ lineages to have one or more mutations, and a variance of $(n)k(1 - k) = (23)(0.377)(1 - 0.377) = 5.4$. In assessing the percentage deviation of k from $1 - e^{-t\mu}$, we then divide the standard deviation in the number of lineages by the expected number of lineages, i.e. $(5.4)^{1/2}/8.7$ or 27%. We conclude that even with the most conservative assumptions, k could not be expected to be different from $1 - e^{-t\mu}$ by more than approximately two standard deviations, or 54%.

To obtain a value for t in terms of years, we need to know the mutation rate. For this purpose, we rely on two California Channel Island fox populations (from Santa Cruz and Santa Rosa islands), that are known to have diverged $\approx 11\,500$ years ago because of the geological separation of the land masses (Wayne *et al.* 1991). The average $(\delta\mu)^2$ genetic distance between the populations is 1.7, where $(\delta\mu)^2$ is the square of the difference between the mean allele lengths at a locus in the two populations and the number is obtained by averaging over 19 unlinked loci (including six of the 10 used in this study). To obtain a mutation rate for the Channel Island foxes, we note that the stepwise mutation model for microsatellites predicts that $E[(\delta\mu)^2] = 2\mu\tau$, where τ is the number of years since divergence (Goldstein *et al.* 1995). The expectation for $(\delta\mu)^2$ is robust to historical fluctuations in population size (Takezaki & Nei 1996), and hence we can use the transformed equation $\mu = E[(\delta\mu)^2]/2\tau$ to obtain a mutation rate of 7.4×10^{-5} per year. Given the generation time for foxes of about 1.5 years, and the generation time for red wolves of no more than 3 years, we can conservatively extrapolate the mutation rate for red wolves to be 3.7×10^{-5} per year (Carbyn 1987; Wayne *et al.* 1991). Combining the upper limit on $t\mu$ (0.474) with the mutation rate, we then obtain an upper limit on the date of hybridization of 12 800 years (Table 2). Note that in calibrating the red wolf mutation rate using the Channel Island fox, we are being conservative in our date estimates. As the microsatellite markers used in both studies were selected because of their high polymorphism (and high mutation rate) among domestic dogs (Ostrander *et al.* 1992), and as dogs are more closely related to wolves than to foxes, the fox mutation rates at these markers will tend to be slower than in wolves, and the result will be an overestimate of the true dates of hybridization (Ellegren *et al.* 1995; Vila *et al.* 1997).

We have two reasons to believe that the mutation rate calculated from the fox divergence time can be extrapolated to red wolves. First, trees of individuals constructed using microsatellite variation cluster individuals according to their separate islands, suggesting a model of isolation among populations that is necessary in order for the

$(\delta\mu)^2$ genetic distance to be used (Goldstein *et al.* 1998). Second, mutation rates calculated using moderate-sized sets of polymorphic microsatellite markers are within a factor of about 5 for Channel Island foxes and humans (Weber & Wong 1993; Goldstein *et al.* 1998) (Table 2). As all canids are closely related (Wayne *et al.* 1997), the difference in mutation rates between Channel Island foxes and red wolves is probably considerably less. If, for the sake of comparison, we calculate the upper limit on the date of hybridization based on the mutation rate from humans, we find that it is considerably more recent than the one calculated from the Channel Island fox mutation rate. Using a human mutation rate that has been estimated fairly accurately and directly at 5.6×10^{-4} per generation (Weber & Wong 1993), and a generation time in red wolves that is at most 3 years (Carbyn 1987), the red wolf mutation rate becomes 1.9×10^{-4} per year, and the upper limit on the date of hybridization becomes 2500 years (Table 2).

Discussion

Our estimates of the mutation rate put a conservative upper limit of 12 800 years on the date when most of the hybridizations could have occurred. This is based on several assumptions and estimates; in particular, related to estimating the probability that a new mutation will generate a specific allele missing in the red wolf population. To obtain a sense of how robust the results would be if these probabilities were calculated differently, or if we used an alternative mutation model, we performed the same calculations using a variety of different parameter combinations. We found that however we performed the calculations, we never obtained a date of hybridization that was older than 15 300 years (Table 2). It is worth emphasizing that all of our calculations are dependent on estimates of the mutation rate and also subject to error in the extrapolation of $t\mu$ from k .

Another aspect of our analysis is the determination of \hat{n} , the 'effective number of lineages' in the average genealogical tree at a red wolf locus. If we know \hat{n} , which gives the total number of independent opportunities for new mutations since the time of hybridization, we have a direct way of estimating the amount of time that has passed since the hybridization. In order to set a lower boundary on \hat{n} , we estimate the number of separate lines of ancestry tracing back from the red wolves to their parental populations. However, the true opportunity for new mutations is not just provided by the ancestral lineages, but also by the lineages that bifurcate from the ancestral ones on their way to producing the modern red wolf sample size. Depending on whether most of the bifurcations in the gene genealogies occur quite recently or whether they occur close to the time of hybridization, the

inclusion of these extra branches produces an unbiased estimate for \hat{n} that is 1–5 times larger than the \hat{n} predicted by the method that we use. We conclude that the upper limit on the date of hybridization that is calculated by our own method (12 800 years) is very conservative.

There are two demographic scenarios that could complicate the interpretation of these results. First, hybridizations between coyotes and grey wolves may have occurred many times in the past. We must therefore limit our conclusions to a careful statement that a large majority of (but not necessarily all) red wolf lineages are sufficiently young to have accumulated no mutations since hybridization and to have originated less than 12 800 years ago. A second scenario is the possibility of substantial gene flow from the hybrids back to the coyote and grey wolf populations. If this occurred, it is possible that some of the alleles private to the red wolves would have ceased to be private as soon as they were copied back into the parental groups, thereby biasing our results toward more recent dates of hybridization. While this scenario cannot be ruled out, it seems unlikely that backward gene flow could have occurred on a sufficiently dramatic scale to eliminate all private red wolf alleles that arose via new mutation.

We conclude that most genes present among modern red wolves did not originate through hybridization in the Pleistocene, but instead have a more recent origin. Thus, we can rule out an ancient hybrid origin for most if not all of the red wolf gene pool, regardless of whether the hybridization occurred suddenly or over a continuous period of time. Not only do our calculations put a conservative upper limit on the dates of hybridization (12 800 years), but half the area under the likelihood curve is between 0 and 2500 years. In rejecting an old date of hybridization, our study lends support to the hypothesis of a recent hybridization between coyotes and grey wolves that may have been associated with the extensive agricultural cultivation of the southern United States by European settlers beginning around 250 years ago. This inference, if true, has relevance to the debate about whether red wolves should be given special protection under the U.S. Endangered Species Act (O'Brien & Mayr 1991; Dowling *et al.* 1992; Wayne 1992).

Acknowledgements

We are grateful to the anonymous referees for their useful comments. We also want to thank Mark Beaumont for his detailed and insightful suggestions concerning the analysis presented here.

References

- Carbyn LN (1987) Gray wolf and red wolf. In: *Wild Furbearer Management and Conservation in North America* (eds Novak M *et al.*), pp. 358–378. Ministry of Natural Resources, Ontario, Canada.

- Di Rienzo A, Peterson AC, Garza JC *et al.* (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences of the USA*, **91**, 3166–3170.
- Dowling TE, Minckley WL, Douglas ME, Marsh PC, Demarais BD (1992) Response to Wayne, Nowak, and Phillips and Henry: Use of molecular characters in conservation biology. *Conservation Biology*, **6**, 600–603.
- Ellegren H, Primmer CR, Sheldon BC (1995) Microsatellite evolution – directionality or bias. *Nature Genetics*, **11**, 360–362.
- Goldstein DB, Pollock DD (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *Journal of Heredity*, **88**, 335–342.
- Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences of the USA*, **92**, 6723–6727.
- Goldstein DB, Roemer GW, Smith DA *et al.* (1998) The use of microsatellite variability to infer patterns of migration, population structure and demographic history: An evaluation of methods in a natural model system. Submitted to *Genetics*.
- Kolenosky G, Standfield RO (1975) Morphological and ecological variation among gray wolves (*Canis lupus*) of Ontario, Canada. In: *The Wild Canids* (ed. Fox MW), pp. 62–72. Van Nostrand Reinhold, New York.
- Kurtén B, Anderson E (1980) *Pleistocene Mammals of North America*. Columbia University Press, New York.
- Lehman N, Eisenhawer A, Hansen K *et al.* (1991) Introgression of coyote mitochondrial-DNA into sympatric North-American Gray Wolf populations. *Evolution*, **45**, 104–119.
- Nowak RM (1979) North American Quaternary *Canis*. *Monograph of the Museum of Natural History, Number 6*. University of Kansas, Lawrence, Kansas.
- O'Brien SJ, Mayr E (1991) Species hybridization and protection of endangered animals. *Science*, **253**, 251–252.
- Ostrander EA, Jong PM, Rine J, Duyk G (1992) Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Proceedings of the National Academy of Sciences of the USA*, **89**, 3419–3423.
- 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.
- Sokal RR, Rohlf FJ (1981) *Biometry*, 2nd edn. W.H. Freeman, New York.
- Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**, 389–399.
- Vila C, Savolainen P, Maldonado JE *et al.* (1997) Multiple and ancient origins of the domestic dog. *Science*, **276**, 1687–1689.
- Wayne RK (1992) On the use of molecular genetic characters to investigate species status. *Conservation Biology*, **6**, 590–592.
- Wayne RK, Jenks SM (1991) Mitochondrial DNA analysis implying extensive hybridization of the endangered red wolf *Canis rufus*. *Nature*, **351**, 565–568.
- Wayne RK, George SB, Gilbert D *et al.* (1991) A morphologic and genetic study of the island fox, *Urocyon littoralis*. *Evolution*, **45**, 1849–1868.
- Wayne RK, Geffen E, Girman DJ, Koepfli KP, Marshall C (1997) Molecular systematics of the Canidae. *Systematic Biology*, **46**, 622–653.
- Weber JL, Wong C (1993) Mutation of human short tandem repeats. *Human Molecular Genetics*, **2**, 1123–1128.

D. E. Reich works on theoretical population genetics. R. K. Wayne studies the history and evolution of canids. D. B. Goldstein uses molecular markers to study human evolution and the genetics of adaptation.

Appendix

Apart from specific conclusions about the origin of the red wolves, our method may have use in estimating the dates of hybridization for other species. By using a likelihood approach, it would be possible to generalize the method to species that have a few but not too many private alleles. To calculate the likelihood function, we take the product of the probabilities of the observed results at each allele of interest i . If an allele is absent, the probability of the observed result is $C_i = R_i^n \approx (1 - kP_i)^n$, while if the allele is present, the probability of the observed result is $C_i = 1 - R_i^n \approx 1 - (1 - kP_i)^n$. The overall likelihood (taking the Cartesian product over m candidate alleles) is then

$$L(k) \approx \prod_{i=1}^m C_i \quad (7)$$

By using $L(k)$ to construct a log-likelihood ratio analogous to the one in eqn 4, we can obtain a confidence interval for k and extrapolate from there to a confidence interval on the date of hybridization.