Molecular Ecology (2006) **15**, 2845–2855 doi: 10.1111/j.1365-294X.2006.02995.x

Detecting introgressive hybridization between free-ranging domestic dogs and wild wolves (*Canis lupus*) by admixture linkage disequilibrium analysis

A. VERARDI, V. LUCCHINI and E. RANDI
Istituto Nazionale per la Fauna Selvatica, Via Cà Fornacetta 9, 40064 Ozzano dell'Emilia (Bologna), Italy

Abstract

Occasional crossbreeding between free-ranging domestic dogs and wild wolves (*Canis lupus*) has been detected in some European countries by mitochondrial DNA sequencing and genotyping unlinked microsatellite loci. Maternal and unlinked genomic markers, however, might underestimate the extent of introgressive hybridization, and their impacts on the preservation of wild wolf gene pools. In this study, we genotyped 220 presumed Italian wolves, 85 dogs and 7 known hybrids at 16 microsatellites belonging to four different linkage groups (plus four unlinked microsatellites). Population clustering and individual assignments were performed using a Bayesian procedure implemented in STRUCTURE 2.1, which models the gametic disequilibrium arising between linked loci during admixtures, aiming to trace hybridization events further back in time and infer the population of origin of chromosomal blocks. Results indicate that (i) linkage disequilibrium was higher in wolves than in dogs; (ii) 11 out of 220 wolves (5.0%) were likely admixed, a proportion that is significantly higher than one admixed genotype in 107 wolves found previously in a study using unlinked markers; (iii) posterior maximum-likelihood estimates of the recombination parameter *r* revealed that introgression in Italian wolves is not recent, but could have continued for the last 70 (± 20) generations, corresponding to approximately 140–210 years. Bayesian clustering showed that, despite some admixture, wolf and dog gene pools remain sharply distinct (the average proportions of membership to wolf and dog clusters were $Q_w = 0.95$ and $Q_d = 0.98$, respectively), suggesting that hybridization was not frequent, and that introgression in nature is counteracted by behavioural or selective constraints.

Keywords: admixture linkage disequilibrium, *Canis lupus*, conservation genetics, introgressive hybridization, linked microsatellites, wolf and domestic dog

Received 7 January 2006; revision accepted 12 April 2006

Introduction

The fear of extensive hybridization between declining wolf populations (*Canis lupus*) and widespread free-ranging domestic dogs in Europe has been a main concern for conservation biologists over the past 30 years (Boitani 1984, 2003; Randi & Lucchini 2002). Wolves and domestic dogs are isokaryotypic, fully interfertile, and have been shown to mate successfully in captivity and in the wild when they co-occur (Wayne *et al*. 1995; Vilà & Wayne 1999). As a consequence of direct hunting and habitat depletion, almost all the European wolf populations dramatically declined in the past few centuries, with only a few relict populations surviving in the Iberian Peninsula, Italy and Scandinavia by the end of the 19th century (Boitani 2003, references therein). In Italy, legal protection, which was granted to wolves in the 1970s, the expansion of natural populations and reintroductions of wild ungulates allowed wolves to grow up to an estimated 450–500 individuals since the 1980s (Boitani 2003). Despite a substantial demographic recovery, wolves are still largely outnumbered by free-ranging dogs, which are estimated to be more than 1 million (Genovesi & Dupré 2000). There is serious concern that, as a consequence of such striking disparity in population size, the genetic integrity of wolf's gene pool might be

Correspondence: Ettore Randi, Fax: +39051 796628; E-mail: met0217@iperbole.bo.it

seriously compromised by recurrent hybridization (Boitani 2003).

However, except for anecdotical reports and occasional direct observations, there is no evidence that wolf × dog hybridization is actually widespread in Europe (Boitani 2003). Analyses of diagnostic mitochondrial DNA (mtDNA) haplotypes failed to detect introgression of dog mtDNA in wolf populations, suggesting that either hybridization is rare or strictly unidirectional, or that F_1 hybrids are not able to backcross into the wolf populations (Vilà & Wayne 1999; Randi *et al*. 2000). Hybridization in canids has been studied also using hypervariable biparental microsatellite loci (Roy *et al*. 1994; Andersone *et al*. 2002; Vilà *et al*. 2003). Randi & Lucchini (2002) analysed allelic variation at 18 unlinked canine microsatellites, aiming to assess the extent of genetic differentiation and identify possible hybrids in a sample of Italian wolves and dogs. Bayesian admixture analyses (STRUCTURE 1.0; Pritchard *et al*. 2000) led to assign the Italian wolves and dogs to two different clusters congruent with prior phenotypic identifications. Only one over 107 genotyped wolves (0.9%) was identified as a hybrid (wolf no. 334 showing an unusual black coat colour; see Table 3), suggesting that hybridization was negligible in the last few generations, and that genetic diversity in post-bottleneck Italian wolves was not sustained by introgression (Boitani 1984). However, those samples were genotyped using a few unlinked microsatellites, and the occurrence of undetected hybridization, or past introgression events was not definitely ruled out (Randi & Lucchini 2002).

Pritchard *et al*.'s model (2000) was designed to infer the number of genetic clusters, assuming Hardy–Weinberg (HWE) and linkage equilibrium (LE), and to identify the populations of origin in a sample of individuals genotyped at unlinked genetic markers, without using any prior information. However, in admixed populations Hardy–Weinberg and linkage disequilibria among unlinked markers decline rapidly with admixture time (Goodman *et al*. 1999), and the admixture signals could be lost after a few generations of backcrossing. A theoretically more sensitive model, implemented in STRUCTURE 2.1 (Falush *et al*. 2003), allows for linkage between markers and makes use of the correlation that arises between linked loci in an admixed population ('admixture linkage disequilibrium'; Stephens *et al*. 1994). Expected advantages of this model are that hybridization events can be traced further back in time, and that the population of origin of chromosomal segments can be inferred. In this study, aiming to further investigate the occurrence of introgressive hybridization, we analysed a sample of 313 Italian wolves and dogs using 16 microsatellites from four linkage groups mapping in four different canine chromosomes (Mellersh *et al*. 1997; Neff *et al*. 1999), plus four unlinked loci. Admixture analyses performed with the linkage model in STRUCTURE 2.1 were used (i) to estimate individual admixture proportions (q_i) and assign probabilistically each multilocus genotype to its population of origin (in case of no admixture), or to both parental populations (in case of admixture); (ii) to infer the ancestry of chromosomal blocks and the population of origin of alleles using a likelihood ratio test (Seldin *et al*. 2004); and (iii) to infer the age of wolf × dog admixture in Italy through posterior maximum-likelihood estimates of the recombination parameter *r* (Falush *et al*. 2003).

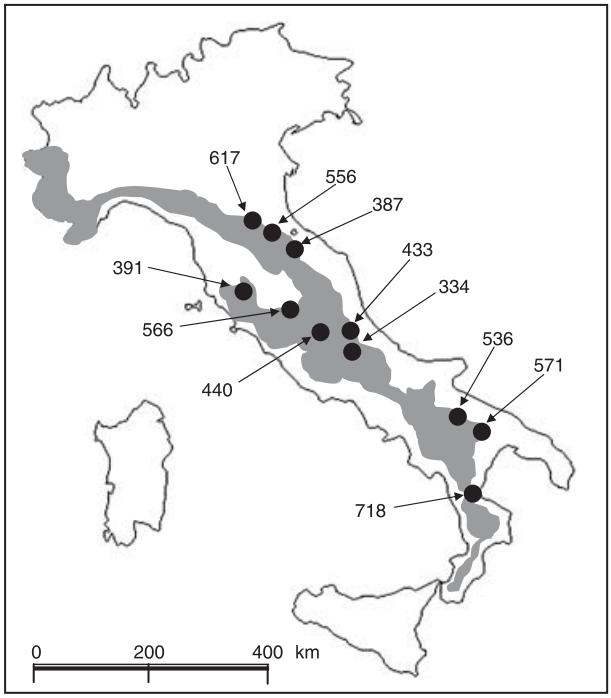


Fig. 1 Approximate wolf distribution range in Italy and locations of putative hybrid samples (see Table 3).

Materials and methods

Sampling, DNA extraction and microsatellite genotyping

Wolf samples ($n = 220$) were obtained mainly by tissue biopsies of found-dead animals collected from 1987 to 2002 across the entire range of the Italian population (Fig. 1). Dog blood samples ($n = 85$) were obtained, through veterinary practices, from feral individuals collected in areas of the Central Apennines where they are sympatric with wolves, from dog pounds and private owners. Seven animals of known hybrid origins were included in the sample as controls: three Italian wolf × dog crosses obtained in captivity, three Czech shepherd dogs, which were obtained by recurrent Eastern European wolf × German shepherd dog crosses, and one 'Lupo Italiano' belonging to a dog breed which was recently produced by crossing one founder male German shepherd dog with a female Italian

Table 1 Summary of allelic variation in wolves (*Canis lupus*) and dogs genotyped at 16 linked and four unlinked microsatellite loci (loci FH2164, C20.622, FH2593 and FH2295 were used in both the linked and unlinked data sets). Microsatellite IDs, linkage groups and map distances (expressed in megabases, Mb, roughly corresponding to Morgans) are indicated according to Mellersh *et al*. (1997), Breen *et al*. (2001), and Guyon *et al*. (2003). Deviations from Hardy–Weinberg equilibrium were estimated from F_{IS} for each locus and each population. Significant F_{IS} values (at a probability level equivalent to $P < 0.05$ after sequential Bonferroni correction for multiple comparisons) are indicated by an *; SE, standard error

	Linkage groups (Mb)	Alleles	Allele size	H_E–H_O in wolves	H_E–H_O in dogs	F_{IS} in wolves	F_{IS} in dogs
Linked loci							
CPH3	CFA6 (52.1)	14	158–188	0.72–0.61	0.78–0.77	0.150*	0.078
C06.69	CFA6 (56.6)	8	146–168	0.51–0.48	0.70–0.68	0.070*	0.031
FH2164	CFA6 (58.5)	17	302–348	0.64–0.63	0.86–0.87	0.018*	−0.005*
LEI-2A11	CFA6 (61.4)	6	175–193	0.29–0.31	0.45–0.36	−0.056*	0.192*
CPH16	CFA20 (35.7)	10	148–176	0.36–0.30	0.85–0.68	0.167*	0.200*
C20.253	CFA20 (37)	11	89–111	0.11–0.08	0.46–0.35	0.324*	0.241*
C20.622	CFA20 (41)	13	209–237	0.60–0.56	0.74–0.60	0.064	0.198*
PRKCD	CFA20 (43.1)	19	108–152	0.69–0.50	0.84–0.68	0.270*	0.197*
FH2274	CFA2 (3)	13	266–334	0.39–0.31	0.85–0.74	0.218*	0.128*
FH2087	CFA2 (11.8)	12	164–254	0.46–0.42	0.79–0.60	0.109*	0.242*
CPH7	CFA2 (19.9)	12	157–185	0.28–0.16	0.77–0.56	0.434*	0.280*
FH2593	CFA2 (31.7)	12	324–376	0.58–0.51	0.80–0.73	0.107*	0.093*
FH2017	CFA15 (47.9)	7	252–276	0.32–0.30	0.54–0.49	0.073*	0.099*
CPH4	CFA15 (52)	9	127–149	0.45–0.42	0.61–0.46	0.072*	0.242*
UOR0442	CFA15 (52.8)	8	221–239	0.55–0.45	0.64–0.54	0.181*	0.161*
FH2295	CFA15 (56.2)	17	405–507	0.57–0.48	0.87–0.72	0.165	0.175*
Unlinked loci							
FH2004	CFA11	14	102–196	0.76–0.65	0.84–0.72	0.149*	0.150*
CPH5	CFA17	11	100–122	0.68–0.68	0.67–0.54	0.002*	0.200*
FH2161	CFA21	12	129–273	0.77–0.77	0.82–0.77	−0.004	0.069*
FH2079	CFA24	8	260–292	0.67–0.65	0.61–0.63	0.032*	−0.031
Mean F_{IS} (SE)						0.127 (0.03)	0.138 (0.02)
Mean H_O (SE)				0.46 (0.18)	0.63 (0.13)		
Mean H_E (SE)				0.52 (0.18)	0.72 (0.13)		

wolf. Total DNA was extracted using guanidine thiocyanate (Gerloff *et al*. 1995). Polymerase chain reaction (PCR) amplifications of microsatellite loci were performed as described in Randi & Lucchini (2002). Genotypes were determined using an ABI 3100 automated sequencer and software GENOTYPER 2.1. Twenty microsatellite loci, originally typed in the domestic dog (Mellersh *et al*. 1997; Neff *et al*. 1999), were analysed. Sixteen of these loci belong to four different linkage groups from four different chromosomes (Table 1).

Analysis of genetic variation

Values of expected (H_E) and observed heterozygosities (H_O), F_{IS} and F_{ST}, and gametic disequilibrium among pairs of loci (estimated using Weir's *R*; 1979) in each population were computed using GENETIX (Belkhir *et al*. 2001). Deviations from HWE were tested with a simulated Fisher's exact test (Guo & Thompson 1992) as implemented in GENEPOP 3.3 (Raymond & Rousset 1995). The sequential Bonferroni test procedure was used to adjust the significance level for a 'table-wide' 5% level (Rice 1989). The hierarchical distribution of genetic diversity was analysed using AMOVA (Excoffier *et al*. 1992) as implemented in ARLEQUIN 2.0b2 (Schneider *et al*. 2000).

Bayesian admixture analyses

Admixture analyses were performed using the 'admixture' model (each individual may have ancestry in more than one parental population), and the new 'linkage' model in STRUCTURE 2.1 (Falush *et al*. 2003), which accounts for the amount of linkage disequilibrium (LD) arising by admixture. The linkage model introduces the parameter *r*, defined as the rate, per unit of map distance, at which recombination breakpoints occur in a chromosome, and that can ultimately be interpreted as an estimate of the number of generations since the admixture event. Moreover STRUCTURE 2.1 allows the use of an 'independent frequency' *I*-model, which assumes that allele frequencies in each population evolve independently, or a 'correlated allele frequency' *F*-model, which assumes that for a limited number of generations following population subdivision, or in consequence of ongoing migration, the evolution of allele frequencies in

each population is correlated to the allele frequencies in the ancestral population.

In this study, we used STRUCTURE 2.1 with the linkage model to compare the results obtained with the independent frequency (*I*) and the correlated (*F*) models. In each case, STRUCTURE was run with five repetitions of 10^5 iterations following a burn-in period of 10^4 iterations. The number of populations *K* was set at the value that maximized the increase in the posterior probability of the data Ln *P*(*D*) according to the formula [Ln $P(D)_K$ – Ln $P(D)_{K-1}$], as suggested by Garnier *et al*. (2004). For the selected *K* values, we assessed the average proportion of membership (Q_i) of the sampled populations (wolves, dogs) to the inferred clusters. Then, comparing results from the *I*- and *F*-model, we assigned each individual to the inferred clusters, using a threshold $q_i > 0.80$ for the assignment of individual genomes to one cluster, or, in the case of admixed individuals, jointly to two or more clusters, if the proportion of membership to each one was $q_i < 0.80$. In this way, we used STRUCTURE to estimate the posterior probability that each individual belongs to each population, or that it has fractions of its genome from two or more parental populations. We did not expect that the sampled dogs were hybrids. The threshold $q_i > 0.80$ was selected because it allows to correctly assign all dogs to the dog clusters.

The posterior probability of each allele to originate in each of the parental populations was estimated in wolf, dog and hybrid samples separately, with the linkage and *F*-models, with *K* = 2 and gametic phase unknown (PHASED = 0), and reported in the output by activating the SITEBYSITE option. We used the posterior probability values to estimate the following joint assignment probabilities: (i) M1P1 that, for a given locus in a given individual, both M (maternal) and P (paternal) alleles are from population 1; and (ii) M2P2 that M and P alleles come from population 2. Admixed genotypes are expressed as M1P2 and M2P1. Results are expressed as the Ln of the probability ratio of two likelihood values Ln *P*(*R*) (Seldin *et al*. 2004). Values of Ln *P*(*R*) > 0 would support the hypothesis in the numerator of the ratio, while values of Ln *P*(*R*) < 0 would support the hypothesis in the denominator.

Results

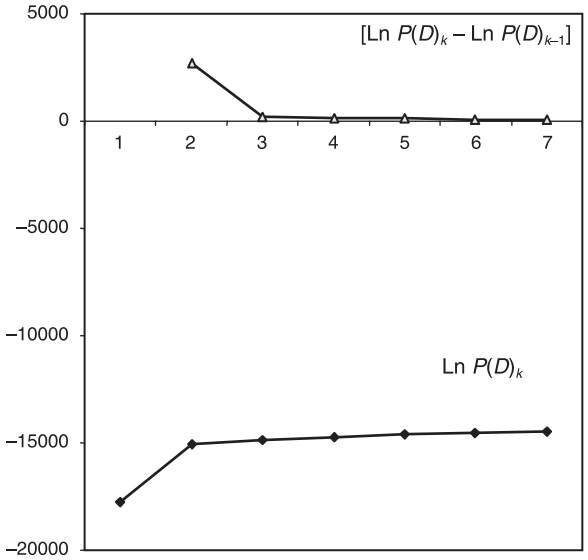
Genetic variation and linkage disequilibrium in wolves and dogs

All loci were polymorphic in wolves and dogs, showing 6–19 alleles, with values of H_E ranging from 0.11 (at locus C20.253 in wolves) to 0.87 (at locus FH2295 in dogs; Table 1). H_E and mean number of alleles per locus were significantly lower in Italian wolves (*t*-test; $P < 0.0001$ and $P = 0.03$, respectively), confirming Randi & Lucchini's (2002) findings. Both wolves and dogs showed average values of $H_E > H_O$, and significant positive values of F_{IS} ($P > 0.05$). In wolves only three loci (FH2161, FH2295 and C20.622) were in HWE ($P > 0.05$). Also, dogs showed only three loci (FH2079, CPH3 and C06.69) in HWE. A highly significant proportion of the total genetic variation ($F_{ST} = 0.24$; $P < 0.001$; AMOVA) was partitioned between wolves and dogs.

Linkage disequilibrium was estimated in a total of 190 locus combinations in each wolf and dog sample groups, of which 24 (13%) were within the linkage groups and the remaining 166 (87%) were between unlinked loci. In wolves, there were 17/24 = 71% significant comparisons ($P < 0.05$; before Bonferroni correction) within linkage groups and only 60/166 = 36% significant comparisons at unlinked loci. The overall number of significant *R* values before Bonferroni was higher than expected by chance (5% = 9.5). After sequential Bonferroni correction, the occurrence of significant *R* values did not change at linked loci (71%), while it decreased to 11/166 = 6.6% at unlinked loci. In dogs 8/24 = 33% and 3/24 = 12.5% significant tests were observed before and after Bonferroni correction at linked loci. The number of significant *R* values in the comparisons between unlinked loci was 22/166 = 13.2% and 2/166 = 1.2%, before and after Bonferroni correction. Thus, linked loci showed higher departures from LD than unlinked loci (71% vs. 6.6% in wolves; 12.5% vs. 1.2% in dogs, after Bonferroni corrections), and wolves showed stronger signals of LD than dogs at both linked and unlinked loci.

Assessing population clustering by Bayesian analyses

Genetic structuring in the total wolf and dog sample set ($n = 312$) was first assessed by running STRUCTURE 2.1 with *K* = 1–4. As expected from the AMOVA results, *K* = 1 showed the lowest posterior probability. Values of Ln *P*(*D*) increased steadily for *K* = 2 and *K* = 3, but afterwards declined with *K* = 4. Garnier *et al*.'s (2004) formula [Ln $P(D)_K$ – Ln $P(D)_{K-1}$] indicated that *K* = 3 represents the optimal clustering of the data, using either the *F*- or the *I*-model (Fig. 2). The two models produced very similar results with *K* = 2 and *K* = 4, but individual clustering was different with *K* = 3 and the *F*- or *I*-model (Fig. 3). With *K* = 2, all dogs were assigned to cluster I with average proportion of membership $Q_I = 0.99$, and wolves were mainly assigned to cluster II ($Q_{II} = 0.97$), independently from the *F*- or *I*-model (Fig. 3a). There were 11 wolves that were partially assigned to both clusters with individual $q_i < 0.80$. The known hybrids were also partially assigned to both clusters with $Q_I = 0.77$ and $Q_{II} = 0.23$ (using the *F*-model), or $Q_I = 0.72$ and $Q_{II} = 0.27$ (using the *I*-model). With *K* = 4 all dogs were assigned to cluster I, the hybrids were admixed, and individual wolf genotypes were partially split between clusters II and III (Fig. 3c) either with the *F*- or the *I*-model. With *K* = 3 and the *I*-model (Table 2), all dogs were assigned to cluster I (with



Model	*K*	Ln *P*(*D*)
F	1	–17754
	2	–15062
	3	–14923
	4	–14701
I	1	–17750
	2	–15057
	3	–14842
	4	–14641

Fig. 2 Posterior probability of the data, Ln *P*(*D*), against the number of *K* clusters (below), and increase of Ln *P*(*D*) given *K* (above), computed after Garnier *et al*. (2004). The Ln *P*(*D*) values obtained by STRUCTURE with *K* = 1–4, and using the *F*- or the *I*-model are shown in the table.

$Q_I = 0.98$), and individual wolf genotypes again were partially split between clusters II ($Q_{II} = 0.51$) and III ($Q_{III} = 0.46$). Concordantly with their known origin, the hybrids were split between the dog cluster I ($Q_I = 0.67$) and the two wolf clusters II and III ($Q_{II} + Q_{III} = 0.33$). In contrast, using the *F*-model, the wolves were assigned almost totally to a single cluster II ($Q_{II} = 0.95$), and the known hybrids were also prevalently assigned to a single cluster III ($Q_{III} = 0.86$), while dogs remained assigned totally to cluster I ($Q_I = 0.98$; Table 2 and Fig. 3b). Partial assignment of individual wolf genotypes to two clusters did not reflect their geographic origins.

Admixture analysis in the Italian wolves

Values of individual proportion of membership (q_i), and their 90% credibility intervals (*CI*), computed by STRUCTURE with *K* = 2–4 (independently from the *I*- or *F*-models) showed that all dogs had $q_d > 0.80$, and only two individuals had their lowest *CI* values lower than 0.80 (see Figs 3 and 4). All the known hybrids were assigned to their own cluster with $q_h > 0.80$, with *K* = 3 and the *F*-model, or partially to a dog and a wolf cluster with $q_h < 0.80$, with *K* = 1 or 4, and the *I*- or *F*-model. Based on these results, given that all dogs (which have no recent ancestry with wolves) and the known hybrids have been correctly assigned to their own clusters, we have selected a threshold $q_i = 0.80$ to assign wolf genotypes to a single (if $q_w > 0.80$), or to more than one cluster (if $q_w < 0.80$). For instance, using this threshold and STRUCTURE with *K* = 3 and the *F*-model, all wolves showing $q_w < 0.80$ at cluster II should be admixed. Results of STRUCTURE with *K* = 2–4, and independently from the *I*- or *F*-models, concordantly indicated that 11 wolves (5.0%) out of the 220 genotyped were likely admixed (Table 3). The lowest q_w values were observed in individual nos 617 ($q_{w617} = 0.04$), 391 ($q_{w391} = 0.11$) and 433 ($q_{w433} = 0.14$), all of them found-dead in the Tuscan-Emilian Apennines (Fig. 1). The other seven individuals that scored less than $q_w = 0.80$ were sampled from various localities in the Italian wolf distribution range, from Calabria to the northern Apennines (Fig. 1). Plottings in Fig. 3 also showed that *CI* values are negatively correlated with q_w and q_d values (in wolves: $r = –0.85$; in dogs: $r = –0.98$; $P < 0.0001$). Thus, the q_w values of putatively admixed wolves have higher credibility intervals, which make their assignment more uncertain (cf. Discussion).

The average posterior value of the recombination parameter (computed over 60 replicated runs with STRUCTURE 2.1, using only the Italian wolves, the *F*-model, *K* = 2, 10 000 iterations after 10 000 iterations of burn-in) was *r* = 70 (SD = 19.75). The map distances between linked loci were expressed in Morgans (Table 1); thus, the *r* value can be interpreted as corresponding to 70 (± 20) wolf generations after the beginning of the admixture process.

Origin of chromosomal segments in the Italian wolves

When two populations admix, recombination events generate mosaic chromosomes of mixed ancestry in hybrids and backcrosses. The linkage model in STRUCTURE 2.1 allows to infer, for each individual, the population of origin of both alleles at each locus (Falush *et al*. 2003). Using *K* = 2, we can compute the posterior probability for both maternal (M) and paternal (P) alleles to originate in population 1 (e.g. in wolves; P1M1), in population 2 (in dogs; P2M2), or in both parental populations (P1M2 and P2M1), in case of admixture. We designed a first analysis to test whether both alleles derived from either the wolf or from the dog populations, that is Ln *P*(*R*)1 = (P1M1/P2M2). Positive values for this test would support the

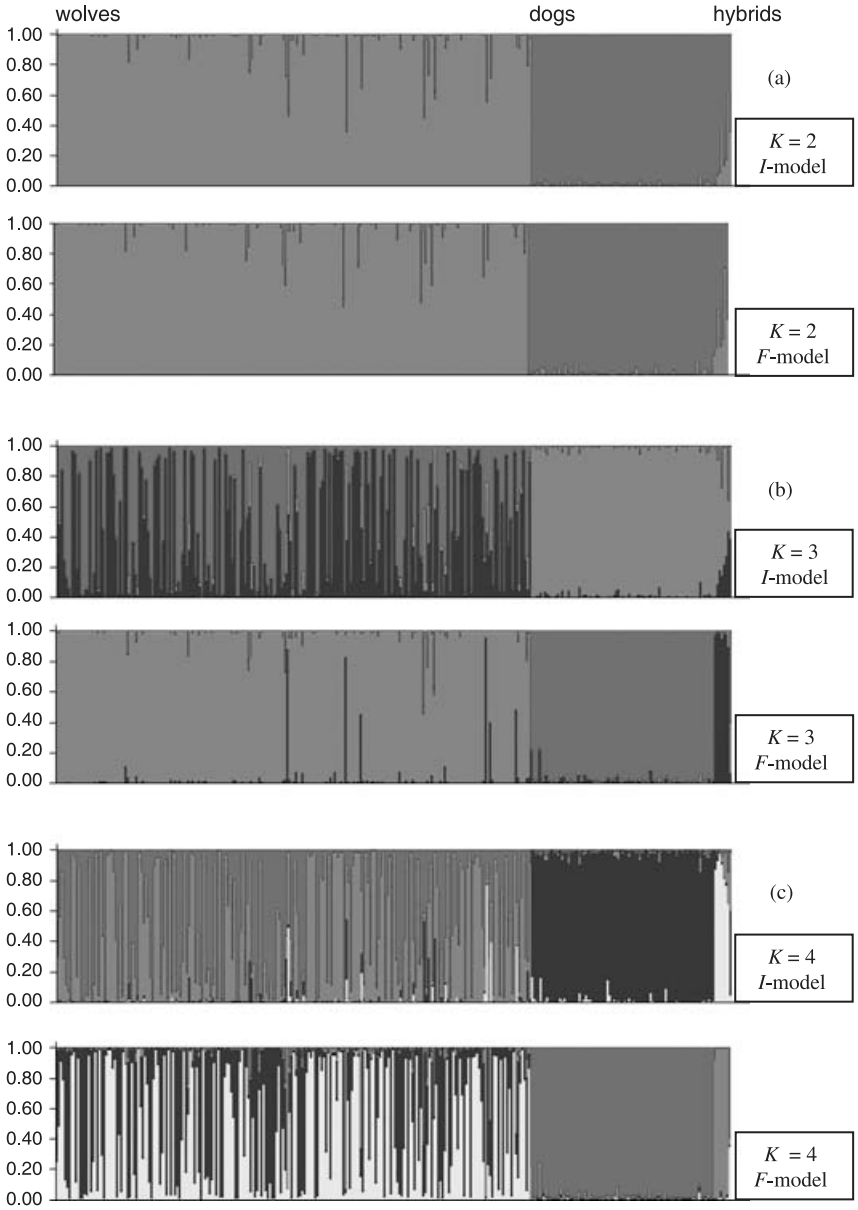


Fig. 3 Bar plotting of the results obtained from STRUCTURE using $K = 2$ (a), $K = 3$ (b), and $K = 4$ (c), with the *I*- or the *F*-model. Each individual is represented as a vertical line partitioned into K coloured segments, whose length is proportional to the individual coefficients of membership in the K clusters.

assignment of both alleles to the gene pool of the Italian wolf population. Results (Fig. 5a) showed Ln $P(R)1 > 0$ in most tests in wolves. Indeed, in 98.7% of the 4400 tests performed in wolves, both alleles were assigned to the wolf population, and only 40 (1.3%) gave negative values. These negative tests were due to 16 wolves, of which six showed negative values at only one locus, five at three loci, one at four loci, one at five loci, one at seven loci and two at 10 loci. All wolves that were identified as admixed at the $q_w > 0.80$ threshold showed also signals of recombination at one (n. 571) or more than one locus (Table 3). The admixture values observed for these 16 samples ranged from $q_w = 0.86$ to $q_w = 0.04$, with an average $q_w = 0.60 \pm 0.07$,

Table 2 Values of average proportion of membership (Q_I) of wolf, dog and hybrid samples to the inferred clusters computed using STRUCTURE with 20 loci, the *F*- or the *I*-model, and $K = 3$

$K = 3$	Cluster I	Cluster II	Cluster III
I-model	Q_I	Q_{II}	Q_{III}
Wolves	0.03	0.51	0.46
Dogs	0.98	0.01	0.01
Hybrids	0.67	0.11	0.22
F-model	Q_I	Q_{II}	Q_{III}
Wolves	0.02	0.95	0.03
Dogs	0.98	0.01	0.01
Hybrids	0.08	0.06	0.86

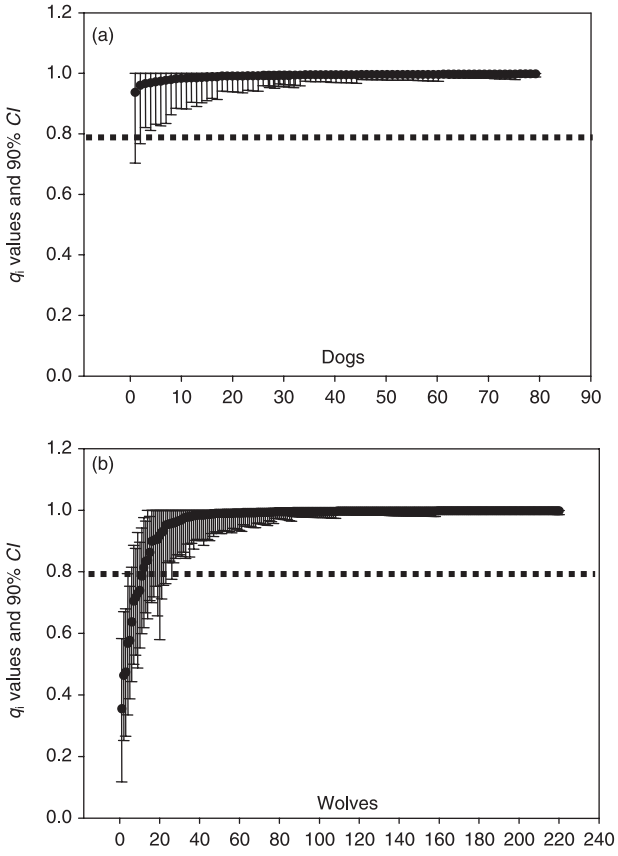


Fig. 4 Distributions of the q_i and 90% *CI* values in dogs and wolves computed with $K = 2$ and *F*-model (results obtained with the *I*-model were identical).

which is significantly lower than the average for the total wolf sample set ($q_w = 0.95 \pm 0.01$). A *Z*-test showed a highly significant negative correlation ($r = -0.85$; $P < 0.0001$) between the number of negative loci and the individual q_w values (Table 3). The known hybrids showed negative scores in 132 tests out of 160 (83%), with q_h values ranging from 0.01 to 0.60 (mean $q_h = 0.26 \pm 0.07$). As expected, none of the tests performed in dogs showed positive values. In a second analysis, we tested the likelihood that both alleles originated from either the parental populations, or that one allele originated from one population and the second allele from the other, that is: Ln $P(R)2$ = (M1P1 + P2M2)/(P2M1 + P1M2). While dogs, being a nonadmixed group, scored only positively, wolves showed 4305/4400 (98%) positive scores, and 74/4400 (2%) negative values in tests that occurred in 22 individuals, which included all the samples that had negative scores in the first analysis (Fig. 5b). The known hybrids showed positive and negative scores, with 31/160 tests (19%) showing negative values.

Discussion

In this study we genotyped, for the first time, wild-living wolves using linked microsatellite markers, aiming to detect signals of hybridization and introgression with free-ranging domestic dogs in Italy. Multilocus genotypes were analysed using the linkage model implemented in STRUCTURE 2.1 (Falush *et al*. 2003). Linked markers are frequently used in human populations to estimate LD and admixture proportions (Seldin *et al*. 2004), but they have not been applied in natural animal populations, except for wildcats (Lecis *et al*. 2006). Unlinked microsatellites are informative as far as they represent independent chromosomal markers. In this case, population structure is modelled assuming that admixture generates transient genetic disequilibria (Pritchard *et al*. 2000), which, however, are expected to decline rapidly, leading admixture signals to disappear in a few generations. In contrast, admixture LD among tightly linked markers is expected to decay more slowly, and the linkage model should improve population clustering allowing to infer more ancient admixtures.

Results of this study indicated that (i) LD, as detected by classical single locus statistics, was higher among linked vs. unlinked loci, and it was higher in wolves than in dogs; (ii) known hybrids and putative admixed wolves were identified at threshold values $q_w > 0.80$, using either the *F*- or *I*-models in STRUCTURE 2.1; (iii) likelihood ratio tests allowed assigning the population of origin of the alleles, leading to an identification of admixed wolves that was

Sample ID	Cluster I q_I	Cluster II q_{II}	Cluster III q_{III}	REC	Sampling area
334	0.260	0.735	0.005	3	L'Aquila
387	0.273	0.705	0.021	3	Arezzo
391	0.004	0.112	0.884	5	Siena
433	0.024	0.141	0.834	10	Pescara
440	0.010	0.537	0.453	4	L'Aquila
536	0.550	0.447	0.004	10	Foggia
556	0.250	0.733	0.016	3	Forli'
566	0.422	0.575	0.003	6	Grosseto
571	0.193	0.775	0.032	1	Bari
617	0.003	0.037	0.959	3	Bologna
718	0.006	0.588	0.407	3	Potenza

Table 3 List of the putatively admixed Italian wolves (Sample ID), with indication of their sampling localities (mapped in Fig. 1), and their q_i values computed using STRUCTURE with $K = 3$, 20 loci and the *F*-model. The number of inferred recombination event (REC) are also indicated

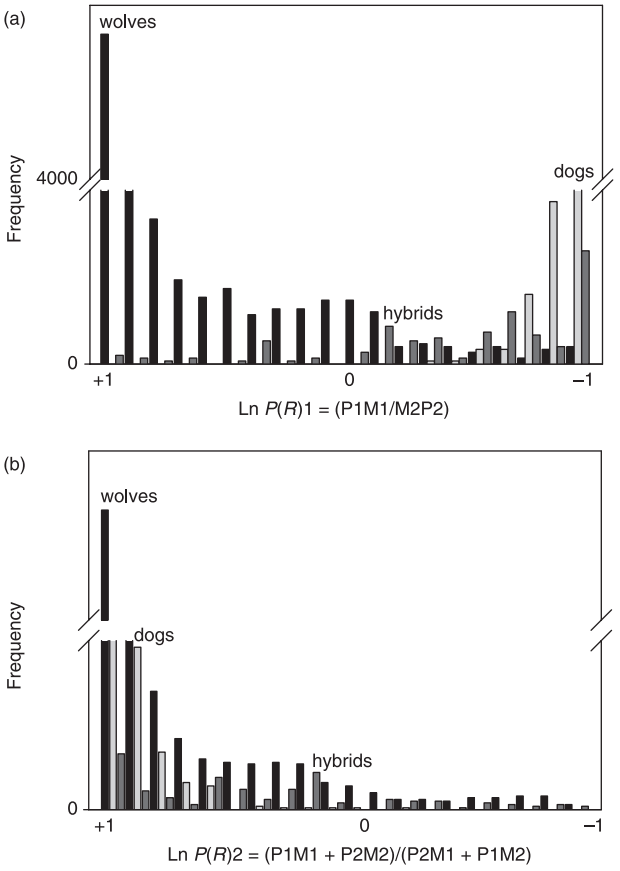


Fig. 5 Population-of-origin assignments for the four microsatellite linkage groups in dogs, Italian wolves and known hybrid obtained by plotting the frequency distributions of the likelihood ratios Ln $P(R)1$ = (P1M1/M2P2) (a), and Ln $P(R)2$ = (P1M1 + P2M2)/(P2M1 + P1M2) (b). Black bars, wolves; grey bars, dogs; dark grey bars, known hybrids.

concordant with the assignments achieved through the q_w values; (iv) threshold q_w values and likelihood ratio tests led to identify 11 out of 220 wolf genotypes (5.0%) that were likely admixed with dogs, a proportion that is significantly higher than previously found using unlinked markers (one admixed over 107 genotyped wolves; Randi & Lucchini 2002); and (v) the average posterior value of the recombination parameter $r = 70$ (± 20) suggested that dogs and wolves might have admixed during the last 70 (± 20) generations. Assuming that generation time in wolves is 2–3 years (Mech & Seal 1987), this finding reveals that wolves and dogs have admixed approximately for 140–210 years in Italy. These results, however, showed some drawbacks, which will be discussed below.

Population substructure in wolves and dogs

Hardy–Weinberg equilibrium is a basic assumption used in STRUCTURE to model population subdivisions (Pritchard *et al*. 2000). It is not known how departures from HWE may affect the reliability of individual assignments to the inferred cluster, and the detection of admixed individuals. Deviations from HWE due to lack of observed heterozygotes in European and North American wolf populations (Roy *et al*. 1994; Forbes & Boyd 1997; Randi & Lucchini 2002) have been attributed to local inbreeding and/or population substructuring (Wahlund's effect). During the past two centuries, the Italian wolves strongly declined and recently expanded again. Bottlenecks could have led to the survival of two partially isolated and genetically differentiated small subpopulations in the southern and central Italian Apennines (Zimen & Boitani 1975). However, the partial assignment of wolves to two clusters (Fig. 3b, c) does not match with different geographic origins of the samples. Additional sampling from the southern Apennines would likely help to detect eventual substructuring in the Italian wolf population. Recolonization patterns and wolf pack dynamics are poorly known (Mech & Boitani 2003). Wolf packs could be more or less stable and open to accept dispersers from other packs, but their dynamics are different in different demographic and ecological conditions (Meier *et al*. 1995, references therein). In conditions of strong human disturbance, as in Italy, possibly just one or a few breeding pairs may succeed in establishing reproductive packs in newly colonized areas. After colonization, the first pack will start to reproduce and expand, creating core areas of closely related individuals, as it was documented in Scandinavia and North America (Mech & Boitani 2003). Although new immigrants could soon arrive, active home range defence and resistance against intruding dispersers from outside the packs might led to transient local inbreeding and population substructuring. The wolf samples used in this study were obtained from carcasses collected across the entire wolf distribution range in Italy during *c*. 15 years, and they could include genetically distinct subpopulations. Due to the consequences of domestication processes and reproduction in captivity, dogs are not expected, and are not in HWE. Our samples were collected from various breeds and from free-ranging dog groups, but not from any real random breeding population. Subdivisions among dog samples were not evidenced in this study. The main subdivision was between wolves and dogs, and genetic subdivisions among dogs could have been detected using only dog samples and increasing the K values in STRUCTURE analyses (see Parker *et al*. 2004, their Fig. 3b).

Performances of the F- *and* I-*models using linked markers*

The linkage model in STRUCTURE 2.1 assumes that LD can arise from population admixture (Falush *et al*. 2003). In our data, classical single-locus statistics (Weir's R; 1979) detected significant LD, which was, as expected, higher

among linked vs. unlinked loci, as well as it was higher in wolves than in dogs. In consequence, STRUCTURE 2.1 detected a number of wolves with intermediate q_w and large *CI* values, suggesting admixture, and allowed to identify a number of putative recombination events, which were associated and significantly correlated with intermediate q_w values. Remarkably, the use of linked markers revealed an increase in the proportion of admixed Italian wolves (*c*. 5.0% vs. 0.9%). This result was expected because of the slower decay of LD among linked loci. STRUCTURE with the *I*- and *F*-model performed differently. With the *I*-model, dogs were assigned to a single cluster ($Q_I = 0.98$), and wolves were split between two clusters ($Q_{II} = 0.51$; $Q_{III} = 0.46$), perhaps reflecting departures from HWE in the Italian wolves. The known hybrids, according to their admixed origin, were partially assigned to each of the three clusters: $Q_I = 0.67$ (the dog cluster); $Q_{II} = 0.11$ and $Q_{III} = 0.22$ (the two wolf clusters). In contrast, using the *F*-model, dogs, wolves and the known hybrids were assigned to the three distinct clusters, corresponding to each a priori defined group, that is, dogs to cluster I with $Q_I = 0.98$; wolves to cluster II with $Q_{II} = 0.95$; and hybrids to cluster III with $Q_{III} = 0.86$ (see Table 2 and Fig. 3). This case study should more closely fit the assumptions embedded in the *I*-model, because (i) wolves and dogs evolved independently since 10 of thousands of generations after domestication (according to the archaeological record domestication occurred *c*. 12 000–14 000 years ago; Clutton-Brock 1999; while genetic data suggested a much older domestication time of 100 000 years; Vilà *et al*. 1997); and (ii) historical and current rates of gene flow between Italian wolves and dogs seem to be low, according to genetic findings published in this study and elsewhere (Randi & Lucchini 2002). However, the results obtained from the *F*-model are also straightforward and easy to interpret. It is not clear why the two models performed so differently. Anyway, exactly the same 11 putatively admixed wolves were identified, using either the *I*- and *F*-model. Thus, the results of individual assignment procedures were not affected by the allele frequency models. The q_w values of admixed wolves showed larger credibility intervals. Uncertainty in the assignment of admixed individuals might be due to difficulties in STRUCTURE in estimating parental allele frequencies, particularly if the true parental populations were not sampled, like in our case study, as it was pointed out by Pritchard *et al*. (2000) and Falush *et al*. (2003). Additional studies, and eventual analyses of simulated data sets, are needed to identify what factors can affect STRUCTURE performance using the *I*- and *F*-models, and how it might be possible to narrow *CI* values in admixed individuals. A recently published study on hybridization between wild and free-ranging domestic cats (*Felis silvestris*; Lecis *et al*. 2006) confirmed the results of this wolf study, suggesting that linked microsatellite markers and admixture LD analyses can significantly improve the detectability of admixed individuals in a number of hybridizing taxa.

Recombination events and times since admixture

Replicated estimates of *r* showed a unimodal distribution with an average value of 70 generations, implicating that wolves and dogs could have interbred in Italy for approximately 140–210 years. The decline of the wolf population, deforestation and the expansion of agriculture in both plain and mountain regions in Italy continued for centuries in the past. Hence, it is well possible that dog × wolf hybridization is relatively ancient. However, the admixture model implemented in STRUCTURE assumes that two parental populations hybridized only once in the past *r* generations. In reality, it is much more likely that dog × wolf hybridization, although not frequent, is recurrent. Thus, at the moment it is not clear what is the reliability of *r* as an estimator of time since admixture. For instance, the variability of *r* values around the average could have been inflated by a poor fitting of the population model to the real dynamics of hybridization.

Conclusions

The risk of introgression of dog genes in the wolf gene pool is a concern for conservation biologists. To date, however, mtDNA and microsatellite markers failed to evidence widespread hybridization and backcrossing in European wolves. Population structure analyses of multilocus genotypes clearly showed that Italian wolves and dogs belong to two distinct gene pools among which gene flow is currently sporadic (Randi & Lucchini 2002; this study). Both wolves and dogs show very high average posterior probabilities of assignment to their own clusters. Nevertheless, reproductive isolation not being complete, a few introgressed individuals were observed among wolves that showed varying levels of admixture. Interestingly but not unexpectedly, the admixed wolves were mostly confined to peripheral areas of the species' distribution range in Italy (Fig. 1). Sample nos 536 and 571, for instance, were found in two parts of Apulia at the southeast edge of the wolf distribution. Sample nos 566, 391 and 387 were collected in Tuscany close to the western border of the wolf distribution. Individual nos 556 and 617 were sampled in Emilia-Romagna along the eastern border of the wolf distribution. The admixed wolves did not show any morphological signal of hybridization, except for sample no. 334 that was unusually dark (Randi & Lucchini 2002). Wolves living at the periphery of their distribution range, sometime in areas of recent colonization, being rarer are more likely to crossbred with free-ranging dogs. Despite hybridization, wolf and free-ranging dogs remain genetically distinct in Italy, suggesting that introgression in nature might be

strongly counteracted by selection or by ethological factors (Vilà & Wayne 1999; Randi & Lucchini 2002). The admixture LD approach we applied in this study, although with some drawbacks, seems to provide a sensitive tool to detect introgressive hybridization. The performances of linked microsatellite markers and of the linkage model should be carefully assessed in the future using both empirical and simulated data sets. However, the wolf data presented in this study suggest that correlations among closely linked loci can identify more ancient admixture events more efficiently than unlinked markers. Our results suggest also that introgressive hybridization, although perhaps protracted in time, is limited and poses no serious threat on the integrity of the Italian wolf gene pool.

Acknowledgements

This project was supported by INFS and by the Italian Ministry of Environment, Direzione Conservazione della Natura. We thank everybody who helped in collecting the wolf and dog samples used for this study. Special thanks to Subject Editor R. K. Wayne and to three anonymous referees which helped to improve the quality of this paper.

References

Andersone Z, Lucchini V, Randi E, Ozolins J (2002) Hybridization between wolves and dogs in Latvia as documented using mitochondrial and microsatellite DNA markers. *Mammalian Biology*, **67**, 79–90.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2001) GENETIX *4.02, Logiciel sous Windows TM pour la gènètique des populations*. Laboratoire Gènome, Populations, Interactions, CNRS UMR 5000, Universitè de Montpellier II, France.

Boitani L (1984) Genetic considerations on wolf conservation in Italy. *Bollettino Di Zoologia*, **51**, 367–373.

Boitani L (2003) Wolf conservation and recovery. In: *Wolves. Behavior, Ecology and Conservation* (eds Mech LD, Boitani L), pp. 317–340. University of Chicago Press, Chicago, Illinois.

Breen M, Jouquand S, Renier C *et al.* (2001) Chromosome-specific single-locus fish probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. *Genome Research*, **11**, 1784–1795.

Clutton-Brock J (1999) *A Natural History of Domesticated Mammals*, 2nd edn. Cambridge University Press, Cambridge, UK.

Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Forbes SH, Boyd DK (1997) Genetic structure and migration in native and reintroduced Rocky Mountain wolf populations. *Conservation Biology*, **11**, 1226–1234.

Garnier S, Alibert P, Audiot P, Prieur B, Rasplus J-Y (2004) Isolation by distance and sharp discontinuities in gene frequencies: implications for the phylogeography of an alpine insect species, *Carabus solieri*. *Molecular Ecology*, **13**, 1883–1897.

Genovesi P, Dupré E (2000) Strategia nazionale di conservazione del lupo (*Canis lupus*): indagine sulla presenza e la gestione di cani vaganti in Italia. *Biologia e Conservazione della Fauna*, **104**, 1–36.

Gerloff U, Schlotterer C, Rassmann K *et al.* (1995) Amplification of hypervariable simple sequence repeats (microsatellites) from excremental DNA of wild living bonobos (*Pan paniscus*). *Molecular Ecology*, **4**, 515–518.

Goodman SJ, Barton NH, Swanson G, Abernethy K, Pemberton JM (1999) Introgression through rare hybridisation: a genetic study of a hybrid zone between red and sika deer (genus *Cervus*), in Argyll, Scotland. *Genetics*, **152**, 355–371.

Guo S, Thompson E (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–372.

Guyon R, Kirkness EF, Lorentzen TD *et al.* (2003) Building comparative maps using 1.5x sequence coverage: human chromosome 1p and the canine genome. *Cold Spring Harbor Symposia on Quantitative Biology 2003, Symposium 68*, pp. 171–178.

Lecis R, Pierpaoli M, Birò ZS *et al.* (2006) Bayesian analyses of admixture in wild and domestic cats (*Felis silvestris*) using linked microsatellite loci. *Molecular Ecology*, **15**, 119–131.

Mech LD, Seal US (1987) Premature reproductive activity in wild wolves. *Journal of Mammalogy*, **68**, 871–873.

Mech LD, Boitani L (2003) Wolf social ecology. In: *Wolves. Behavior, Ecology and Conservation* (eds Mech LD, Boitani L), pp. 1–34. University of Chicago Press, Chicago, Illinois.

Meier TJ, Burch JW, Mech LD, Adams LG (1995) Pack structure and genetic relatedness among wolf packs in a naturally-regulated population. In: *Ecology and Conservation of Wolves in a Changing World* (eds Carbyn LN, Fritts SH, Seip DR), pp. 293–302. Canadian Circumpolar Institute, Occasional Publication No. 35. Edmonton, Alberta, Canada.

Mellersh CS, Langston AA, Acland GM *et al.* (1997) A linkage map of the canine genome. *Genomics*, **46**, 326–336.

Neff MW, Broman KW, Mellersh CS *et al.* (1999) A second-generation genetic linkage map of the domestic dog, *Canis familiaris*. *Genetics*, **151**, 803–820.

Parker HG, Kim LV, Sutter NB *et al.* (2004) Genetic structure of the purebred domestic dog. *Science*, **304**, 1160–1164.

Pritchard JK, Stephens M, Donnelly PJ (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Randi E, Lucchini V (2002) Detecting rare introgression of domestic dog genes into wild wolf (*Canis lupus*) populations by Bayesian admixture analyses of microsatellite variation. *Conservation Genetics*, **3**, 31–45.

Randi E, Lucchini V, Christensen MF *et al.* (2000) Mitochondrial DNA variability in Italian and east European wolves: detecting the consequences of small population size and hybridization. *Conservation Biology*, **14**, 464–473.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

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

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN: *A software for population genetics data analysis*. Version 2.000. Genetics and

Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland.

Seldin MF, Morii T, Collins-Schramm HE *et al.* (2004) Putative ancestral origins of chromosomal segments in individual African Americans: implications for admixture mapping. *Genome Research*, **14**, 1076–1084.

Stephens JC, Briscoe D, O'Brien SJ (1994) Mapping admixture linkage disequilibrium in human populations: limits and guidelines. *American Journal of Human Genetics*, **55**, 809–824.

Vilà C, Wayne RK (1999) Hybridization between wolves and dogs. *Conservation Biology*, **13**, 195–198.

Vilà C, Savolainen P, Maldonado JE *et al.* (1997) Multiple and ancient origins of the domestic dog. *Science*, **276**, 1687–1689.

Vilà C, Walker C, Sundqvist A-K *et al.* (2003) Combined use of maternal, paternal and bi-parental genetic markers for the identification of wolf-dog hybrids. *Heredity*, **90**, 17–24.

Wayne RK, Lehman N, Fuller TK (1995) Conservation genetics of the gray wolf. In: *Ecology and Conservation of Wolves in a Changing World* (eds Carbyn LN, Fritts SH, Seip DR), pp. 399–407. Canadian Circumpolar Institute, Occasional Publication No. 35. Edmonton, Alberta, Canada.

Weir BS (1979) Inferences about linkage disequilibrium. *Biometrics*, **35**, 235–254.

Zimen E, Boitani L (1975) Number and distribution of wolves in Italy. *Zeitschrift für Säugetierkunde*, **40**, 102–112.

This research, which was conducted at the Conservation Genetics Laboratory, Italian Institute of Wildlife Biology (INFS), is part of a project on population and conservation genetics of carnivores in Italy and Europe. Andrea Verardi and Vittorio Lucchini worked as researcher at INFS. They are primarily interested in and population and conservation genetics of carnivores. Ettore Randi is head of Conservation Biology and Genetics at INFS.
