ELSEVIER

Genetic variability and differentiation in Spanish roe deer (*Capreolus capreolus*): A phylogeographic reassessment within the European framework ☆

Luis J. Royo [a],[1], Gerardo Pajares [b],[1], Isabel Álvarez [a], Iván Fernández [a], Félix Goyache [a],*

[a] *Área de Genética y Reproducción Animal, SERIDA, E-33203 Gijón, Asturias, Spain*
[b] *Asociación del Corzo Español, Departamento de Ecología, Universidad Complutense de Madrid, C/ José Antonio Novais s/n, E-28040 Madrid, Spain*

Abstract

Here we study 109 Iberian roe deer individuals corresponding to 9 Spanish populations. Individuals were sampled at locations that are expected to have acted as refugia for the species during the 20th century. Samples were analyzed for a 436 bp mtDNA fragment from the HVI region of mitochondrial DNA and 10 microsatellites. The 109 sequences gave 31 different haplotypes that enabled identification of a new haplogroup (mainly present in Northwestern Iberian populations and representing roughly a third of our samples) that is not present in other European roe deer populations. Using microsatellites, correspondence analysis and molecular coancestry information revealed high molecular differentiation among Northwestern and Central-Southern Spain roe deer populations. Both sequence and microsatellite analysis reveal that the Spanish roe deer populations are genetically heterogeneous and have high genetic structure clearly separating the Pyrenean-reintroduced populations and two main areas for the species in Spain (Northwestern and Central-Southern) coinciding with the two main areas acting as refugia for the majority of the mammal species during glaciations. The implications of the obtained information as regards the phylogeography of the species are discussed together with suggestions as to appropriate strategies for the conservation and management of populations.
© 2006 Elsevier Inc. All rights reserved.

1. Introduction

The European roe deer (*Capreolus capreolus*) is widespread across the continent with populations distributed from Cádiz in Southern Spain, the Southern Apennines in Italy and the Southern Balkans to as far north as the Arctic Circle in Scandinavia. Several studies, mainly using mitochondrial (mt)DNA information (Wiehler and Tiedemann, 1998; Randi et al., 1998, 2004; Vernesi et al., 2002; Lorenzini et al., 2003), have shown that the roe deer population in Europe has a significant structuring consistent with the cycles of restriction and colonization induced by the last glacial oscillations. As seen for a wide variety of European flora and fauna (see Taberlet et al., 1998 and Hewitt, 1999 for reviews), these studies have consistently identified three regions: the Iberian Peninsula, Italy and the Balkans, as glacial refugia for the species. Vernesi et al. (2002) and Randi et al. (2004) identified two main mtDNA haplogroups, both characterising the roe deer populations in Western, Central and Northern Europe and in the Balkans. Within the non-Balkan haplogroups, the aforementioned studies identified a subcluster formed

by the roe deer populations in Central-Southern Italy, thus genetically supporting the existence of an endemic subspecies: *Capreolus capreolus italicus* (Festa, 1925).

The Iberian Peninsula has played a significant role in the history of the roe deer. The genetic situation of the species described above indicates that the postglacial roe deer expansion resembled Hewitt's (1999) 'bear pattern'. In this model, individuals from the Italian refugium did not spread over the Alps, and populations of Iberian origin prevail in Central-Northern Europe. Furthermore, Randi et al. (2004) recently identified a new mtDNA haplogroup mainly present in the Northwestern Iberia roe deer populations, thus suggesting the existence of different glacial refugia within the Iberian Peninsula for the roe deer. However, the genetic situation of the Iberian roe deer population is poorly known because of the limited number of Spanish samples used in the published works (Vernesi et al., 2002) or the biased distribution of the sampling locations (Randi et al., 2004). This contrasts with the extensive sampling and analysis available for other, possibly relict populations such as those from Italy (Randi et al., 1998, 2004; Vernesi et al., 2002) or Eastern European countries such as Serbia, Montenegro and Kosovo (Randi et al., 2004). In fact, better knowledge of the genetic background of the species in these Eastern countries led Randi et al. (2004) to suggest the possibility of an Eastern origin of the most relevant mtDNA haplogroup in Central-Northern European roe deer (the so-called Clade Central), thus questioning the general pattern of postglacial recolonization of Europe by the species.

In contrast with the extensive use of mtDNA for phylogenetic studies in roe deer, the use of microsatellite markers for this purpose is still underdeveloped. Microsatellite markers are usually expected to be more informative in describing the consequences of the most recent population events (isolation, expansion, translocations, current dispersal rates) than in ancestral population history (Coulon et al., 2004; Randi et al., 2004). Recently, Caballero and Toro (2002) formalized how to obtain coancestry coefficients from molecular information by applying Malécot's (1948) definition to the marker genes, although they referred to it as identity by state instead of identity by descent. The molecular coancestry between two individuals i and j is the probability that two randomly sampled alleles from the same locus in two individuals are identical by state. Molecular coancestry (f) is equivalent, in terms of allelic frequencies, to $f_{ij} = \sum_{ij} x_{ij} y_{ij}$ (Eding and Meuwissen, 2001; Caballero and Toro, 2002) where x_{ij} and y_{ij} are the frequencies of the ith allele at the jth locus within the individuals (or populations) x and y. This term was previously defined by Nei (1972) as 'the probability of identity of a gene from x and gene from y'. Nei (1972, 1987) also showed that, in the founder population, its value would coincide with that of the expected homozygosity. Eding and Meuwissen (2001), using simulated data, showed that molecular coancestry has other interesting properties, namely that average kinship between populations becomes constant very quickly after population fission, causing between-population diversity to remain constant. Molecular coancestry may be used to assess genetic diversity between populations (Eding et al., 2002; Fabuel et al., 2004; Álvarez et al., 2005). Researchers may contrast molecular coancestry information (basically consisting of unbiased estimations of the ancestral allelic frequencies of the studied populations; Eding and Meuwissen, 2001) with that offered by other parameters characterising recent genetic evolutionary changes (mainly due to drift) in order to assess recent or ancestral processes affecting population differentiation (Álvarez et al., 2005). This can be of particular interest for wild species, in which human action has had less influence on the genetic background of populations. However, the possibilities of using molecular coancestry information to study the genetic relationships between populations using actual datasets have not been widely explored (Caballero and Toro, 2002; Fabuel et al., 2004; Álvarez et al., 2005), especially in the wild.

Genetic analysis of the Spanish roe deer populations is also of interest due to the fact that up to three possible subspecies have been described in the literature on the basis of their high morphological variability (Lister et al., 1998): *C. c. canus* (Miller, 1910) in Northern Spain, *C. c. decorus* (Cabrera, 1916) in Northwestern Spain and, especially, *C. c. garganta* (Cabrera, 1916; Meunier, 1983) in Southern Spain, considered as a relictual population of endemic Iberian roe deer. Morphological differences between *C. c. decorus* and *C. c. canus* are mainly based on coat colour and size and probably do not support a taxonomical differentiation (von Lehmann and Sägesser, 1986; Fandos and Reig, 1993). *C. c. garganta* is expected to be distributed in the mountains of Cádiz and Málaga, which mark the Southwestern distribution limit of roe deer. In this area, the species has evolved in very particular ecological conditions with a low density of individuals in a Mediterranean xerophytic forest. The identity of the *C. c. garganta* subspecies is doubtful (Aragón et al., 1995b), although Lorenzini et al. (2003) suggested a genetic differentiation between Andalusian and Northern roe deer in Spain. However, based on a limited sampling in the putative area of distribution of *C. c. garganta*, Randi et al. (2004) suggest that the subspecies may have admixed origins and that its monophyly would be not supported by molecular data.

The recent population history of the species in Spain has been well characterized (Tellería and Virgós, 1997; Gortázar et al., 2000). Deforestation and over-hunting led to the decline and eradication of local roe deer populations, particularly in Central and Southern Iberia, though also in the humid, mountainous North of the Peninsula (Boutin, 1990; Aragón et al., 1995a). Subsequent to the 1970's, preservation policies and the decline in the rural human population permitted natural recolonisation of roe deer from refugia located in the main Spanish mountainous ranges (Braza et al., 1989; Aragón et al., 1995a,b; Tellería and Virgós, 1997; Gortázar et al., 2000). However, there are records of some reintroduction events carried out

using non-indigenous roe deer, such as those in Catalonia, Northeastern Spain (Rosell, 2002). The aim of the present study is to contribute to a more comprehensive molecular analysis of the roe deer populations in Spain so as to increase our knowledge of the phylogeny of the species. To do so, we shall obtain roe deer samples from most Spanish locations in which reintroduction events are not documented as having occurred, as well as obtaining molecular information from mitochondrial sequences and microsatellite markers. Mitochondrial sequences will be compared with others from the literature so as to accurately assign Spanish roe deer haplotypes to their corresponding haplogroups. Microsatellite marker polymorphisms will be used to obtain molecular coancestry information on the genetic relatedness of the sampled populations at the moment of population separation. Throughout the paper we shall attempt to address the following questions: (a) Is the Spanish roe deer population heterogeneous? and (b) What is the contribution of the Spanish roe deer in the genetic background of the species? The implications of the obtained information for establishing management and conservation strategies for roe deer in the Iberian Peninsula will be also discussed.

2. Materials and methods

2.1. Sampling

Muscle or skin samples were obtained from a total of 109 individuals corresponding to 9 roe deer populations. The sampling locations plus their abbreviations and the number of individuals sampled (in brackets) are: *Western Asturias* (As; 20; sampled in Cangas del Narcea, in the area surrounding Muniellos Nature Reserve and the county of Valdés), *Cádiz* (Ca; 15; sampled mainly in the Natural Park of Los Alcornocales), *Centre* (Ce; 14; sampled in Guadarrama and Montes de Toledo ranges), *Galicia* (Ga; 10; sampled in the Game Preserve of Los Ancares), *Lleida* (Le; 10; sampled in La Pobla de Segur), *Picos de Europa* (Pi; 10; sampled in Valdeón within the Picos de Europa National Park), *Soria* (So; 10; sampled in Urbión and Alcarama), *Sueve* (Su; 10; sample in the Sueve Range area in Northeastern Asturias) and, as an outgroup, the southern *Bordeaux* area, in France (Bo; 10).

The sampling locations include all the mountainous systems in Spain that are expected to have acted as refugia for the species during the decrease in populations size in the first half of the 20th century (Conde de Yebes, 2000), which allowed the natural recolonization of the Spanish countryside during the second half of this period: (a) Galician Massif (Ga), Cantabrian Range (As, Su and Pi), Iberian Range (So), Pyrenean Range (Le), Central Massif (Ce), including samples from the Sierra del Guadarrama and four samples from the Toledo Mountains, and Penibetica Range (Ca). The geographical distribution of the sampled populations is shown in Fig. 1. Throughout the paper, samples from As, Ga, Pi and Su will be referred as Northwestern, those
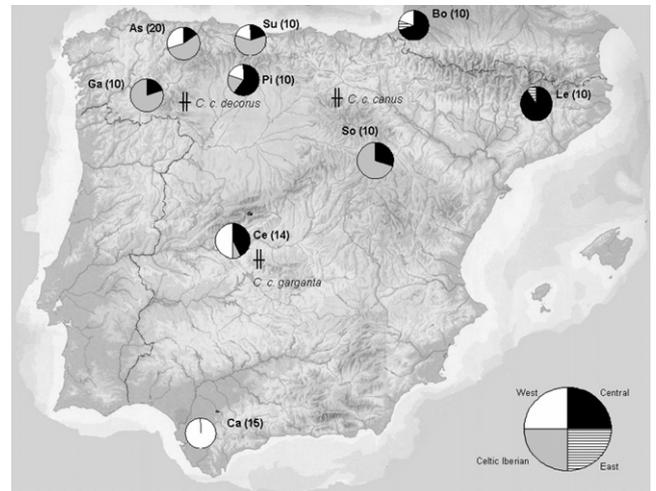

Fig. 1. Sampling locations and proportion of mtDNA haplogroups in the analyzed roe deer populations. The diagrams indicate the proportions of mtDNA control region haplotypes belonging to Clades West, Central, East and Celtic-Iberian. Abbreviations correspond to Western Asturias (As), Bordeaux (Bo), Cádiz (Ca), Centre (Ce), Galicia (Ga), Lleida (Le), Picos de Europa (PI), Soria (So) and Sueve (SU). The number of individuals sampled in each location is in brackets. The approximate location in which the putative *Capreolus capreolus* subspecies described in the literature for the Spanish roe deer is indicated.

from Ca, Ce and So as Central-Southern and those from Bo and Le as Pyrenean. The sampling performed here is expected to cover the distribution range of the nominal roe deer subspecies in Spain: *C. c. canus* (So), *C. c. decorus* (As, Ga, Pi and Su) and *C. c. garganta* (Ce and Ca).

2.2. DNA extraction and amplification

Total DNA was isolated from muscle or skin samples using the Nucleospin® tissue protocol (Macherey-Nagel). Mitochondrial DNA was amplified using the following PCR primers: 5′-CGTCAGTCTCACCATCAACCCC CAAAGC-3′ and 5′-TGAGATGGCCCTGAAGAAA GAACC-3′, which amplify a 436 bp fragment from the first hypervariable domain of the mtDNA control region (D-loop) (Douzery and Randi, 1997; Vernesi et al., 2002). The PCR protocol consisted of a 20 µl reaction mix containing approximately 50 ng of total DNA, 0.2 µM of each primer, 2.0 mM Cl_2Mg, 0.2 mM of dNTPs and 0.2 U of *Taq* polymerase (Biotools, Madrid). The PCR consisted of an initial denaturation step at 94 °C for 5 min followed by 35 cycles of 30 s at 94 °C, 30 s at 60 °C and 1 min at 72 °C, with a final elongation step of 10 min at 72 °C after the last cycle. PCR products were purified using ExoSAP-IT® (USB Corporation, Cleveland) following the manufacturer's recommendations. DNA sequencing was performed on an ALFexpressII semi-automated sequencer (Amersham Biosciences, Barcelona) using the Thermo Sequenase Cy5 Dye Terminator Cycle Sequencing Kit (Amersham Biosciences, Barcelona). The sequences were aligned using the PILEUP command (Feng and Doolittle, 1987) of the Winsconsin Package (Winsconsin Package Version 10.2,

Genetics Computer Group (GCG), Madison, Wisconsin) and were checked by eye. The numbering of nucleotide sites refers to the *Bos taurus* sequence (Genbank, Accession No. J01394), following Vernesi et al. (2002). The GenBank accession numbers and references for the sequences obtained in the present analysis are from DQ114745 to DQ114784 and from DQ384640 to DQ384708.

A set of 10 microsatellites (CSSM66, ILSTS011, INRA006, BM6506, BM1818, BM757, NVHRT48, NVHRT21, RT1, NVHRT16) were analysed in all the individuals. The first six were previously used in ovine studies (Álvarez et al., 2004, 2005), while the others have been reported to be polymorphic in roe deer by Poetsch et al. (2001). PCR products were labelled using a fluorescent method (Cy5 labelled primer) and genotyping was performed on an ALFexpressII semi-automated sequencer (Amersham Biosciences, Barcelona).

2.3. Sequence analysis

Genetic variability within populations was assessed from mitochondrial information by computing nucleotide diversity (π) using the program MEGA 3.1 (Kumar et al., 2004). Among populations *p*-distance and Reynolds' distance matrices (available on request) were computed using the program Arlequin v.2.0 (Schneider et al., 2000). The statistical significance of the values was estimated by permutation analysis using 1000 replications.

To correctly assign our sequences to the haploclusters previously identified in the literature, we integrated our data with the 161 haplotypes reported by Randi et al. (2004). Among the available sequences, further analyses were restricted to the 388 bp length forming the consensus region defined between our sequences and those reported by Randi et al. (2004). To identify biologically meaningful, empirically supported phylogenies, networks were constructed using the median-joining network method (Bandelt et al., 1999) as implemented in the program Network 4.1.1.1 (Available at http://www.fluxus-engineering.com/).

Divergence times among the identified haplogroups were estimated using the method proposed by Gaggiotti and Excoffier (2000), as implemented in the program Arlequin 2.0 (Schneider et al., 2000). This method, based on the net number of substitutions between populations, is able to remove the possible effects of bottlenecks and unequal population sizes. For the translation of mutational units into years, the sites with indels were omitted and, following Randi et al. (1998), substitution rates per site per million of years ranging between 0.04 and 0.08 were used.

The program DnaSP 4.0 (Rozas et al., 2003) was used to assess past demographic parameters, computing the following statistics: (a) the distribution of the observed pairwise nucleotide site differences (also called mismatch distribution) and the expected values for no recombination (Rogers and Harpending, 1992); the smoothness of the observed mismatch distribution was quantified by the raggedness statistic (*r*); (b) Fu's *Fs* statistics; (c) Fu and Li's (1993)

D^* and F^* tests. The D^* test statistic is based on the differences between the number of singleton sites (mutations appearing only once among the sequences) and the total number of mutations, whilst the F^* test statistic is based on the differences between the number of singleton sites and the average number of nucleotide differences between pairs of sequences (Fu and Li, 1993). Statistics (a) were computed to assess a possible growing populations scenario; statistics (b) test the presence of genetic hitchhiking, which produces a similar pattern to mismatch distributions; statistics (c) test for the assumption of neutrality of mutations thus detecting the presence of background selection. Coalescence simulations based on a neutral infinite-sites model and assuming a large constant population size (Hudson, 1990) were applied to determine the confidence intervals of the computed statistics.

2.4. Microsatellite analysis

Genetic diversity on the analyzed data set was assessed computing the expected heterozygosity (H_e) and the heterozygote deficiency within population (F_{IS}). Among-populations gene flow and genetic differentiation were assessed by computing the following between-population genetic parameters: molecular coancestry (f_{ij}, Caballero and Toro, 2002), kinship distance (D_k, Caballero and Toro, 2002) and Reynolds' distance (Reynolds et al., 1983). Following Caballero and Toro (2002), most of these parameters were computed in terms of molecular coancestry. The molecular coancestry between two individuals *i* and *j* is the probability that two randomly sampled alleles from the same locus in two individuals are identical by state (Caballero and Toro, 2002). Molecular coancestry between two individuals *i* and *j* at a given locus may be computed using the following scoring rules (Caballero and Toro, 2002; Eding and Meuwissen, 2001): $f_{ij,l} = 1/4[I_{11} + I_{12} + I_{21} + I_{22}]$, where I_{xy} is 1 when allele *x* on locus *l* in individual *i* and allele *y* on the same locus in individual *j* are identical, and zero otherwise. Notice that this value can only have four values: 0, 1/4, 1/2 and 1. The molecular coancestry between two individuals *i* and *j* (f_{ij}) may be obtained by simply averaging over *L* analyzed loci $f_{ij} = \frac{\sum_{l=1}^{L} f_{ij,l}}{L}$. The molecular coancestry of an individual *i* with itself is self-coancestry (called s_i), which is related to the coefficient of inbreeding of an individual *i* (F_i) by the formula $F_i = 2s_i - 1$. In turn, the (kinship) distance (here called D_k) between two individuals *i* and *j* is $D_k = [(s_i + s_j)/2] - f_{ij}$ (Caballero and Toro, 2002). Within- and between-populations molecular coancestry and D_k were computed by simply averaging the corresponding values for all the within- or between-population pairs of individuals. Wright's (1969) *F*-statistics, F_{IS}, F_{ST}, and F_{IT} are obtained as $F_{IS} = \frac{\tilde{F} - \tilde{f}}{1 - \tilde{f}}$, $F_{ST} = \frac{\tilde{f} - \bar{f}}{1 - \bar{f}}$ and $F_{IT} = \frac{\tilde{F} - \bar{f}}{1 - \bar{f}}$, where $\tilde{f}$, $\tilde{F}$ are, respectively, the mean coancestry and the inbreeding coefficient for the entire population, $\bar{f}$ and the average

coancestry for the subpopulation (see Formulae (3) and (6) in Caballero and Toro, 2002). Notice that $\tilde{F}$ is not the same as genealogical inbreeding, defined as the probability that an individual has two identical alleles by descent (Malécot, 1948), but rather homozygosity, referred to as the identity by state, which is defined as the probability that two alleles chosen at random from the population are the same.

The above computations were performed using the program MolKin 2.0 (Gutiérrez et al., 2005). For descriptive purposes, genetic distance matrices (available on request) are shown as plots constructed through multidimensional scaling analysis carried out using the Proc MDS of SAS/STAT™ (1999).

Individual multilocus genotypes were also investigated carrying out a canonical discriminant analysis as implemented in the program GENETIX 4.02 (Belkhir et al., 2001) to obtain an unbiased test of population structure.

2.5. Geographical analyses

Finally, geographical structuring was assessed by computing: (a) the correlation between the among-populations geographical distance matrix and the among-populations genetic distance matrices using Mantel's test; and (b) an Analysis of Molecular Variance (AMOVA). Both Mantel's test and AMOVA analyses were carried out on sequences and microsatellites information using the program Arlequin 2.0 (Schneider et al., 2000). The statistical significance of the values was estimated by permutation analysis using 1000 permutations.

3. Results

3.1. Mitochondrial DNA sequences, cluster assignment and differentiation among geographical populations

The 109 sequences (from np 15804 to 16239) gave 31 different haplotypes (Table 1) with 28 variable sites (24 transitions, 3 transversions and 1 indel). Overall haplotype (k) and nucleotide (π) diversity was, respectively, 0.28 and 0.012. In general, populations from Northwestern Spain (As, Ga, Pi and Su) have significantly higher ($p < 0.001$) values of k and π (ranging from 0.35 to 0.70 and from 0.009 to 0.013, respectively) than those of Central-Southern Spain (Ca, Ce and So) (varying between 0.20 and 0.36 and 0.002 and 0.008). Among the identified haplotypes, the one denominated H5, which is present in the Ca and Ce populations, was the most frequent (11%), followed by H29 (7.3%), which has been basically found in the So population, and H14 (6.4%), which is well distributed across Northwestern populations. Twelve haplotypes are unique. The Bo population does not share mtDNA haplotypes with any of the other sampled populations. Only two haplotypes (H9 and H16) are represented in Northwestern and Central-Southern populations. Thirteen haplotypes are only represented in the Northwestern areas, while six of them are only represented in the Central-Southern areas.

After pooling our 31 mtDNA haplotypes with those reported previously by Randi et al. (2004) ($N = 161$), we identified 95 different haplotypes of 388 pb length (see Supplementary data). Fig. 2 shows a median-joining network including most parsimonious trees supported by these 95 sequences. Four major haploclusters were identified. Because of the uneven pattern of substitution in the D-loop region (Sbisà et al., 1997) leading to homoplasies, the network was contracted for character conflicts (Forster et al., 2001) to confirm its robustness. The procedure reduced the number of plotted sequences to 43, but the four haploclusters were still clearly identifiable (not shown). Three of the four major clusters coincided with those named by Randi et al. (2004) as cluster East (including most sequences from Balkan origin), cluster Central (the most represented in the whole European territory including the sequences assigned to the subspecies *C. c. italicus*) and cluster West (sampled mainly in the Iberian Peninsula). The new haplogroup identified here is made up of 6 haplotypes representing 35 (32.1%) of our samples. Further information on the coincidences between our sampling and that of Randi et al. (2004) can be found in Supplementary data. It should be noted, however, that the new cluster identified here includes some sequences reported by Randi et al. (2004) (sampled basically in Spain and to a lesser extent in Portugal) as being linked to the cluster Central (see Fig. 2 of Randi et al., 2004). Moreover, we found no haplotypes in our sampling that coincide with those linked to the cluster West obtained in the Central Italian Alps by Randi et al. (2004).

The four haploclusters identified above are present in the samples analyzed here. A quick overview of Fig. 1 reveals that the new haplocluster identified here is mainly spread over the Iberian area usually known as 'Celtic'. Following the traditional classification of the domestic animal native strains from this geographical area (see Royo et al., 2005 for more detailed information), the new cluster will be named hereafter Celtic-Iberian. Both Pyrenean populations have a haplotype assignable to the cluster East (Balkan type). The three other clusters (Central, Celtic-Iberian and West) are represented in our samples in similar proportions (34.9%, 32.1% and 31.2%, respectively). Fig. 1 illustrates the frequencies of the different haploclusters in our roe deer populations. All the samples obtained in the Southern population (Ca) and a half of those obtained in the Ce population are assignable to the cluster West, which is in turn present in the Cantabrian roe deer population (As, Pi and Su) and in the Bo population (between 20% and 30% of the samples). The cluster Celtic-Iberian is present mainly in Northwesternmost populations (Ga, 80%; Su, 60%; As, 55%) and in the So population (70%), although it is also present at a low proportion in the Pi and the Ce populations (20% and 7.1%, respectively). Most of the sequences obtained in the Pyrenean populations and in the Pi area were assigned to the cluster Central (between 70% and 90%), which is also present in all the other populations except Ca (Fig. 1).

Table 1
Variable sites of the identified haplotypes in the available individuals of Cantabrian roe deer, number of samples (N), haplotype diversity (k) and nucleotide diversity (π) by sampled population and for the whole dataset. Abbreviations correspond to Western Asturias (As), Bordeaux (Bo), Cádiz (Ca), Central Spain (Ce), Galicia (Ga), Lleida (Le), Picos de Europa (PI), Soria (So) and Sueve (Su)

| Haplo	Cluster	Consensus	Variable sites	As	Bo	Ca	Ce	Ga	Le	Pi	So	Su	TOTAL
			1111111111111111111117111111										
			5555555555566666666666666										
			8888899999900000000001112222										
			12669235689233556689148122 3										
H01	West	C85	A-TGTTATTATTGCCGGATCGTTGTCAC			1							1
H02	West	C83	.-.........C................			6							6
H03	West	C91	.-.A...........AG...C......			7	5						12
H04	West	C94	.-.........CTG.............			1							1
H05	Central	C16	.A.......G......A..T.......		4								4
H06	East	C14	.-...C..........A..TACC...G.		1								1
H07	Central	C18	.-.......G......A.CT.C......		3								3
H08	West	C95	.-.................C......		2								2
H09	West	C63	.-...............AG...C......				2			1		2	5
H10	Celtic-Iberian	C02	.-.............T.A..TA.C.....					3		2		5	10
H11	Celtic-Iberian	C86	.A.A..........T.A..TA.C..G..					1					1
H12	Celtic-Iberian	C01	.A.A..........T.A..TA.C.....	6				2					8
H13	Celtic-Iberian	C51	.-............A..TA.C.....	2				1					3
H14	Central	C03	.-....GC...C....A..T.......	2				2		1		2	7
H15	Celtic-Iberian	C87	.A.A......C...T.A..TA.C.....					1					1
H16	Central	C19	.A.......G......A..T..C....	1					1		3		5
H17	Central	C88	.-.......G.....AA..TA..A....						1				1
H18	Central	C17	.-.....C.G......A..T..C.....						3				3
H19	East	C26	.-...C..........A..TA.C.....						1				1
H20	Central	C16	.-.....G......A..T.......						4				4
H21	West	C04	.AC.............G...C......	4									4
H22	Celtic-Iberian	C01	.-.A.........T.A..TA.C.....	3									3
H23	West	C04	.-C.............G...C......	2									2
H24	Central	C19	.-.......G......A..T..C....							1			1
H25	Central	C89	.A..C...CG......A..T.......							3			3
H26	Central	C89	GA..C...CG......A..T.......							1			1
H27	West	C90	.-...............AG...C.....T							1			1
H28	Central	C82	.-.......G.........T.CC.....				5						5
H29	Celtic-Iberian	C51	.A...........A..TA.C.....				1				7		8
H30	Central	C92	.-.......G.........T.C......				1						1
H31	Celtic-Iberian	C93	.A.A.........T.A..TA.C.C...									1	1
N				20	10	15	14	10	10	10	10	10	109
k				0.35	0.40	0.27	0.36	0.60	0.50	0.70	0.20	0.40	0.28
π				0.013	0.008	0.007	0.008	0.009	0.005	0.012	0.002	0.010	0.012
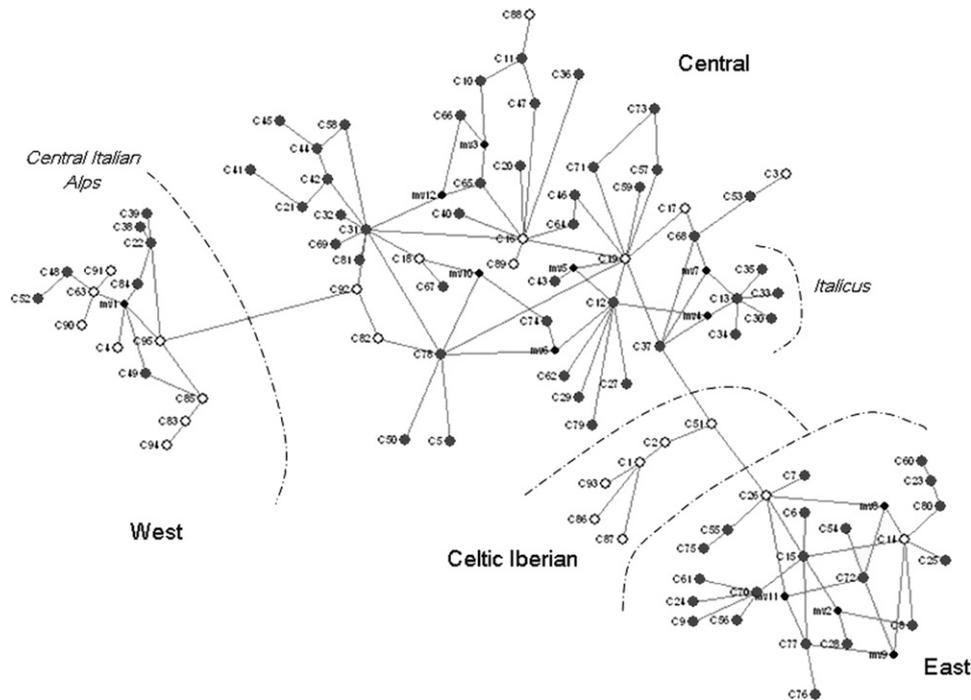
Fig. 2. Median-joining tree constructed with the 95 consensus haplotypes listed in Supplementary data. Small black circles indicate missing haplotypes. Open circles indicate haplotypes that have been identified in our sampling. Grey circles indicate additional haplotypes obtained by Randi et al. (2004). Within clusters Central and West, respectively, the haplotypes linked to the subspecies *C. c. italicus* and the haplotypes sampled in Central Italian Alps are highlighted.

3.2. Divergence times among haplogroups and demographic tests

Divergence times were estimated between the four main haplogroups identified here (Table 2). The estimations varied from 40,000 to 221,000 years depending on the pair of clusters and the mutation rate (0.04 or 0.08 substitutions per site per million years). The largest divergence times (roughly 155,000 years on average) were estimated between the cluster West and both the Celtic-Iberian and the East clusters and between the clusters Central and East. The comparisons involving the cluster Central show shorter divergence times with respect to the clusters West and Celtic-Iberian. The shortest (and the only non-significant) divergence time is between the clusters Celtic-Iberian and East.

Fig. 3 shows the observed and expected mismatch distribution for each of the four haploclusters identified here. The best fit with a pattern of population expansion was found for the cluster Central, whilst the observed mismatch distributions for clusters East and West show two different peaks, indicating two possible expansion events for each of the clusters. The cluster Celtic-Iberian shows a ragged shape. However, the raggedness for the four identified clusters was highly significant ($p < 0.000001$), thus supporting recent population growth in Spanish roe deer, though a unimodal shape was not found for all the identified clusters. Fu's (1997) *Fs* and Fu and Li's (1993) D^* and F^* tests were non-significant for all the cases (data not shown), thus leading us to respectively reject the presence of genetic hitchhiking and background selection in the analyzed dataset.

3.3. Microsatellite diversity and molecular coancestry

The parameters characterising the degree of polymorphism of the marker set used at microsatellite and population levels are given in Table 3. The number of alleles per locus varied between 4 (RT48) and 20 (BM757) and the observed heterozygosity between 0.508 (RT48) and 0.898 (BM757). Most of the markers employed showed positive F_{IS} values, ranging from -0.024 (BM656) to 0.098 (BM757). The average number of alleles per loci was rarefacted (a_r), as proposed by Hurlbert (1971), to account for the differences in sample size. The Bo, Ce and Le

Table 2
Divergence times (in thousand of years) allowing for unequal population sizes between the four clusters of mtDNA identified in this study

	Central	East	Celtic-Iberian
East	117–58		
Celtic-Iberian	91–45	80–40	
West	98–49	210–105	221–111

Intervals are estimated considering two different mutation rates (0.04 and 0.08 substitutions per site per million years). Bootstrapping (1000 replicates) indicated that the divergence times reported here were significant for $p < 0.05$ with the exception of the East-Celtic-Iberian pair ($p = 0.067$).

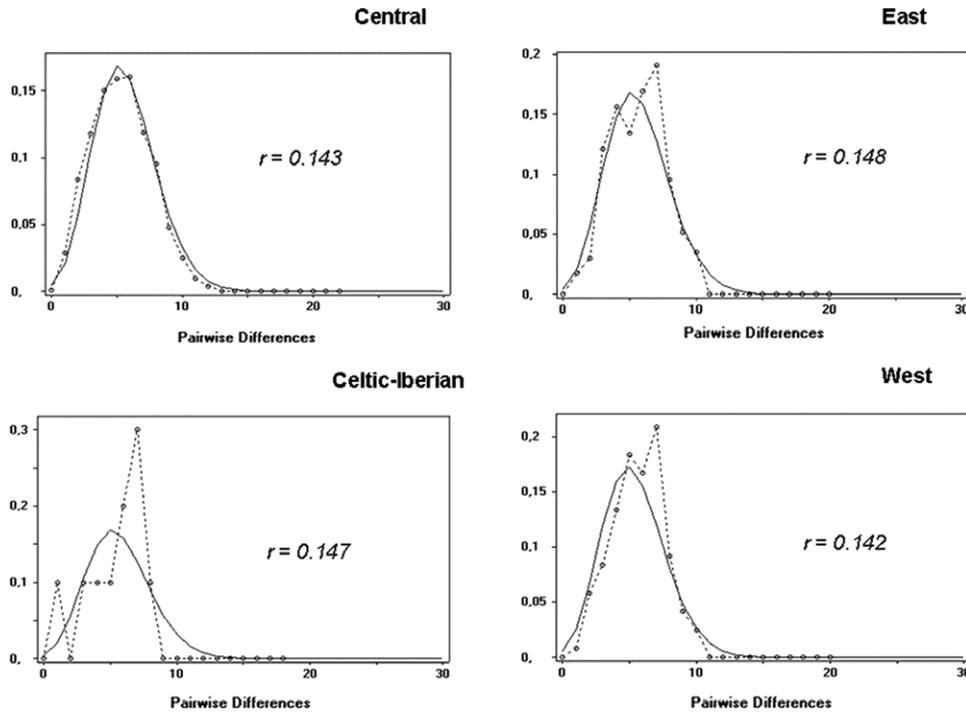Fig. 3. Expected (regular line) and observed (dotted lines) mismatch distribution and raggedness statistics (r) for Central, East, West and Celtic-Iberian mtDNA haploclusters. All the r values were statistically significant for $p < 0.000001$.

Table 3
Number of alleles per marker (n), observed heterozygosity (H_o) and heterozygote deficiency (F_{IS}) values per marker and sampled population

	n	H_o	F_{IS}	f_{ii}
Marker				
CSSM66	8	0.829	0.018	
ILSTS011	11	0.807	−0.017	
INRA006	10	0.791	−0.008	
BM6506	6	0.518	−0.024	
BM1818	10	0.708	0.070	
BM757	20	0.898	0.098	
NVHRT48	4	0.508	0.056	
NVHRT21	12	0.856	0.015	
RT1	12	0.858	0.058	
NVHRT16	11	0.712	−0.012	
Population				
As	4.5	0.680	−0.006	0.351
Bo	6.5	0.745	0.019	0.277
Ca	4.5	0.572	−0.017	0.476
Ce	6.0	0.626	0.097	0.373
Ga	5.7	0.688	−0.005	0.344
Le	6.2	0.720	0.006	0.299
Pi	4.9	0.752	−0.123	0.359
So	4.5	0.680	−0.045	0.359
Su	4.6	0.665	−0.118	0.364
All the samples	10.4	0.681	−0.015	0.251

For the statistics at population level, the parameter n was substituted by the rarefacted (Hurlbert, 1971) average number of alleles per locus (a_r), and the average molecular coancestry within population (f_{ii}) is given. Abbreviations correspond to Western Asturias (As), Bordeaux (Bo), Cádiz (Ca), Central Spain (Ce), Galicia (Ga), Lleida (Le), Picos de Europa (PI), Soria (So) and Sueve (Su).

populations have a_r values of 6.0 or higher, the So population has an intermediate value (5.7) whilst the other populations have values below 5. The highest levels of observed heterozygosity (above 0.7) were found for the Bo, Le and Pi populations, while the lowest was found for the Ca population (0.572). With the exception of Ce and Le, all the sampled populations showed heterozygote excess, characterised by negative F_{IS} values. The within-populations average molecular coancestry (f_{ii}), characterizing the degree of genetic similarity of the individuals of a given population, ranges from 0.277 (Bo) to 0.476 (Ca), whilst the average value for the whole sampled population is 0.251.

The values computed for F_{ST}, F_{IS}, F_{IT}, kinship distance (D_k) and Reynold's distance (D_R) for the whole population using microsatellites were respectively 0.147, -0.015, 0.134, 0.315 and 0.091. In general, the distance matrices (not shown) revealed that the lowest levels of differentiation coincided with the Northwestern population pairs (paired F_{ST} values below 0.07 and paired D_k values below 0.40) and with the pair formed by the Bo and the Le populations. Fig. 4 shows the multidimensional scaling plots constructed using D_R and $0.5 - f_{ij}$ paired values. The two plots give complementary information. On Dimension 1, the Ca and Ce populations are always highly differentiated from the Bo and Le populations. In the D_R plot, however, the So population tends to cluster with the Northwestern populations, whilst for the plot constructed using the $0.5 - f_{ij}$ paired values, the So population clearly clusters with the Ca and Ce populations. Notice that molecular coancestry (f_{ij}) is the second term in the formula used to compute kinship distance: $D_k = [(s_i + s_j)/2] - f_{ij}$. In an ideal
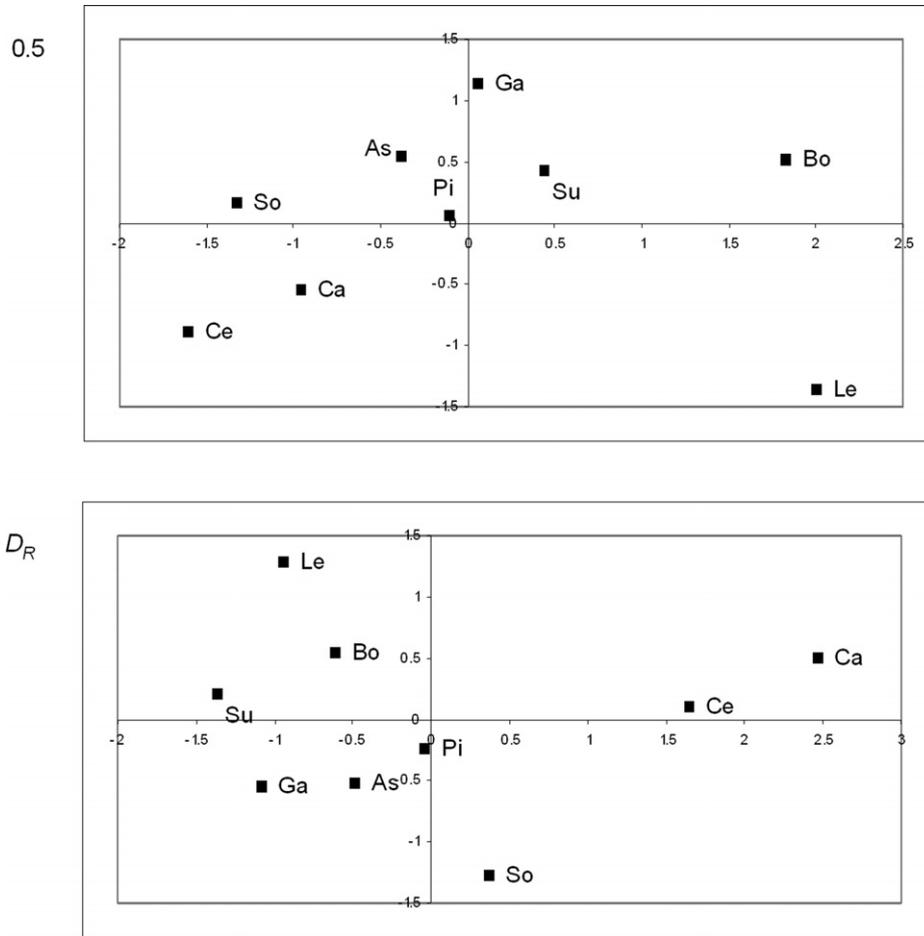
Fig. 4. Multidimensional scaling plots constructed using the between-population Reynold's distance (D_R) matrix and between-population kinships distance (D_k) assuming no differentiation after population fission ($0.5 - f_{ij}$) inferred using microsatellite information. Dimension 1 is represented on the X-axis and Dimension 2 on the Y-axis. Abbreviations correspond to Western Asturias (As), Bordeaux (Bo), Cádiz (Ca), Centre (Ce), Galicia (Ga), Lleida (Le), Picos de Europa (PI), Soria (So) and Sueve (SU).

situation of absence of differentiation after the separation of populations, the kinship distance would be $D_k = 0.5 - f_{ij}$. Thus, the plot constructed using $0.5 - f_{ij}$ illustrates the differentiation of the populations exclusively dependent on the differentiation of the populations at the time of metapopulation fission (Eding et al., 2002; Álvarez et al., 2005), whilst the D_R plot basically illustrates genetic differences mainly based on drift (Reynolds et al., 1983; Álvarez et al., 2005).

The canonical discriminant analysis identified up to 5 factors with eigenvalue $\geqslant 1$, explaining a total of 81.81% of the genetic variation contained in the microsatellite dataset. Fig. 5 illustrates the similarities among individual genotypes using the two first retained factors, which respectively explained 24.26% and 19.66% of the variation. The individuals belonging to the Ca and Ce populations are clearly differentiated from the others on Dimension 1. The individuals belonging to the populations sampled in the Pyrenean area (Bo and Le) are separated from those from the Northeastern Spain population on Dimension 2. Notably, the So individuals present intermediate coordi-

nates between the populations of Northwestern and Central-Southern Spain.

3.4. Geographical patterns of genetic variation

Correlation between different set of between-populations genetic distance matrices (D_R and p-distance for mtDNA and D_R and $0.5 - f_{ij}$ for microsatellites) and the between-populations geographical distance matrix were computed. Though they are higher for microsatellites, all the correlation coefficients were statistically significant, ranging from 0.41 ($p = 0.03$) for p-distance to 0.71 ($p = 0.001$) for D_R computed using microsatellites.

Additional analysis was conducted to assess the degree of differentiation of the geographical groups of populations for Spanish roe deer. Results of the AMOVA analysis carried out on mitochondrial and microsatellite information are given in Table 4. The analysis showed statistically significant genetic differentiation (highlighted by the Φ_{CT} statistic) among the three main groups of populations (Northwestern, Central and Pyrenean) only for microsatellite data
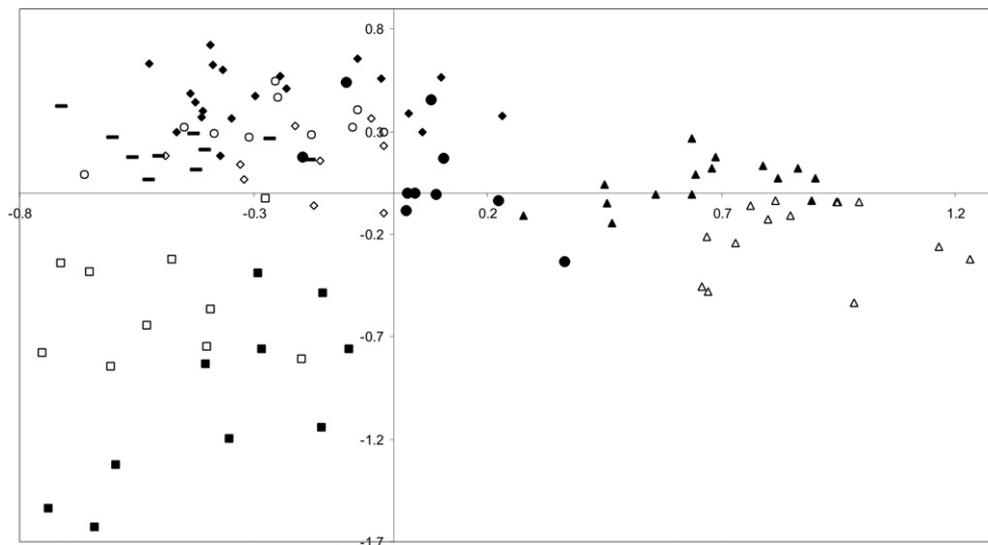
Fig. 5. Plot showing the distribution of the individual genotypes of the roe deer individuals genotyped for our microsatellite set on the bidimensional spaces formed by the two first factors computed by a canonical discriminant analysis. The first factor (on the X-axis) explained 24.26% of the genetic variance and the second factor (on the Y-axis) explained 19.66% of the variation. The Central-Southern populations are in Black (Centre) or open (Cádiz) triangles; Pyrenean populations are in open (Bordeaux) or black (Lleida) squares; Northwestern populations are in black diamonds (Western Asturias), open rectangles (Galicia), black rectangles (Sueve) and open circles (Picos de Europa). Notice that the Soria individuals (in black circles) have intermediate positions between the Northwestern and the Central-Southern populations.

Table 4
Results of the AMOVA analysis conducted on mitochondrial and microsatellite information in the analysed Spanish roe deer populations

Groups of populations	Percentage of variation components			Differentiation statistics		
	Among groups	Among populations within groups	Within populations	Φ_{CT}	Φ_{SC}	Φ_{ST}
Mitochondrial DNA						
Two-level analysis	—	29.35	70.65	—	—	**0.294**
Three-level analysis	6.80	23.93	69.27	0.068	**0.257**	**0.307**
Three-level analysis excluding the So population	19.07	11.19	69.74	**0.191**	0.138	**0.303**
Microsatellites						
Two-level analysis		12.65	87.35	—	—	**0.126**
Three-level analysis	7.54	7.00	85.46	**0.075**	0.075	**0.145**
Three-level analysis excluding the So population	9.52	5.42	85.07	**0.095**	0.060	**0.149**

The present results were first obtained in a two-level analysis without grouping the sampled populations and secondly by grouping Northwestern (As, Ga, Pi and Su), Central-Southern (Ca, Ce and So) and Pyrenean (Bo and Le) populations in a three-level hierarchical analysis. The three-level hierarchical analyses were re-run excluding the So population from the analysis. Differentiation parameters significantly different from 0 for $p < 0.01$ are in bold.

(accounting for roughly 7% of the variance). However, among-groups differentiation assessed *via* mtDNA was not statistically significant. The three-level hierarchical analysis was re-run excluding the So population, which, on account of its relationships with the other sampled population, could not be assessed without uncertainty for microsatellite information (see Fig. 4). The exclusion of the So population gave a significant Φ_{CT} value for mtDNA information, an increase in the among-groups differentiation for microsatellites (Φ_{CT}) and a decrease in the within-groups variation for both mtDNA and microsatellite data.

4. Discussion

In this study, we sampled Spanish roe deer populations in order to ascertain the genetic heterogeneity of the species in one of the possibly main refugia for the species during the last glaciation period and to contribute to the knowledge of the genetic background of the species in the European context. During the first half of the 20th century, the roe deer in Spain suffered a dramatic decrease in population size due to over-hunting and deforestation, thus causing the original population to fragment into several small local populations that became extinct in the majority of cases (Aragón et al., 1995a; Tellería and Virgós, 1997; Virgós and Tellería, 1998). Sampling was carried out here over the major mountainous systems at locations that are expected to have acted as refugia during the decrease in population size. The Spanish territory has been naturally recolonized from most of these sampled areas (Tellería and Virgós, 1997), with a small number of reports of reintroduction events with non-indigenous roe deer that have

occurred mainly in the Pyrenean area (Gortázar et al., 2000; Rosell, 2002). In this respect, the Le population (in the Catalonian Pyrenees) shows high similarities with the Bo population, which has been used as the outgroup in this study. The Bo population is an admixed population in which reintroduction events are likely to exist (Boisaubert and Boutin, 1988). Both the Bo and the Le populations show mtDNA haplotypes assignable to the cluster East (of Balkan origin), which is also present in reintroduced roe populations in Italy (Vernesi et al., 2002). Moreover, at a microsatellite level, both the Bo and the Le populations have the highest average number of alleles per locus and heterozygosity values, and the lowest within-population molecular coancestry values, which means the lowest genetic resemblance of the individuals in an area. The high variability shown by admixed-recently found roe deer populations that expanded rapidly has been previously reported by Randi et al. (2004) with respect to other reintroduced Italian populations (Liguria).

4.1. Genetic heterogeneity in the Spanish roe deer population

Our results show that the Spanish roe deer populations are genetically heterogeneous and have high genetic structure. Sequence, microsatellite and AMOVA analyses clearly point out that, not including the Pyrenean-reintroduced populations, we may differentiate two main areas for the species in Spain: (a) the Northwestern area; (b) the Central-Southern area.

Though the dispersal behaviour for the species is accepted to be male biased (Wahlström and Liberg, 1995; Coulon et al., 2004), thus leading to lesser among-population differentiation at a nuclear DNA level, we found a higher degree of among-area differentiation for microsatellites than for mtDNA sequences. Here, microsatellites informed on the short-term evolution of the populations mainly due to drift using D_R (Reynolds et al., 1983; Álvarez et al., 2005), but also on the allele frequencies in the ancestral population using molecular coancestry information (Eding and Meuwissen, 2001; Eding et al., 2002; Álvarez et al., 2005). In this respect, the genetic relationships between the So roe deer population and those of Northwestern Spain are more likely to have occurred during the recent history of the species. With respect to mtDNA information, we found that sequences linked to the cluster West (Randi et al., 2004) are highly frequent in the roe deer populations of Central-Southern Spain, except for the So population. We also identified a new haplogroup, which we have called Celtic-Iberian and which is basically spread over the Northwestern part of the Iberian Peninsula, that seems to be relict of the Iberian roe deer populations. However, the So population has a high proportion of Celtic-Iberian haplotypes, thus leading us to include it within the Northwestern populations. Further analyses will be needed to ascertain whether the So population might be an ancestrally admixed population formed in a geographical area that may have acted as a contact zone between two major and well differentiable roe deer populations in the Iberian Peninsula.

Moreover, the present study allows us to postulate that the Northwestern Iberian Peninsula, including the Cantabrian branch and the Galician Massif, and the Central-Southern part of the Iberian Peninsula including the Tajo river basin and the Mediterranean coast have acted as glacial refugia for the species in the Iberian Peninsula. This is consistent with the climatic scenario in the Iberian Peninsula, at least during the last glaciation period. Arribas (2004) summarizes the environmental conditions in the Iberian territory during this period and identifies two main areas acting as refugia for most of the mammal species: (a) the Southern half of the Peninsula and the Mediterranean coastal area; (b) the Northern side of the Cantabrian Range. The Southern part of the Iberian Peninsula has probably had free-of-ice temperate conditions, thus forming the main refugium of the Iberian Peninsula with Northward limits in the Central Massif and the Ebro river basin. The climatic conditions in the Northern part of the Peninsula would have been difficult, with important glaciers in the Pyrenees and the Cantabrian Range and extreme climatic conditions in the Northern Castilian Plateau (the present Duero river basin). Interestingly, however, the Northern side of the Cantabrian Range would have presented more appropriate habitat conditions, thus constituting a good refugium for the individuals retained by the glaciers and the Castilian Plateau during their migration to the South. A similar pattern of differentiation has been reported by Branco et al. (2000) for wild rabbit in the Iberian Peninsula using Cytochrome B sequence information. These authors identified two main clades of mtDNA, each of which spread over the Southeastern and Northeastern Iberian areas with relatively narrow overlap between contiguous geographical ranges.

4.2. Spanish roe deer mtDNA variation within a species context

It is usually accepted that the Iberian Peninsula has played a significant role in the history of the roe deer (Randi et al., 2004). If the postglacial expansion of the species exactly resembled Hewitt's (1999) 'bear pattern', the most frequent haplogroup in the species, the cluster denominated as Central by Randi et al. (2004) and previously described in earlier papers (Wiehler and Tiedemann, 1998; Vernesi et al., 2002), would be Iberian in origin. However, this hypothesis may be questioned and the demographic analyses carried out in the present study (mismatch distributions, Fu's (1997) Fs test and Fu and Li's (1993) D^* and F^* neutrality tests) may support a postglacial population expansion for the species, although the mismatch shapes for clusters East and West might reflect more than a single expansion event during successive interglacial warm periods. Moreover, the identification of the cluster West in the Iberian Peninsula (Randi et al., 2004), which

has not contributed to a large extent to the current postglacial mtDNA diversity in central Europe, and the relatively small number of haplotypes assignable to the clade Central found in the Iberian Peninsula led Randi et al. (2004) to consider a possible origin of the clade Central in Eastern Europe. This suggestion was based on the high genetic variability found for mitochondrial DNA in an extensive sampling of Serbia, Montenegro and Kosovo. In addition, we have found a new cluster (the Celtic-Iberian), exclusively spread over the Iberian territory, including sequences that were previously assigned to the cluster Central. However, the data obtained in the present analysis may shed some light on the aforementioned reports. We identified seven Iberian consensus haplotypes (six from our sampling and one from the study by Randi et al., 2004) assignable to cluster Central that are widely represented in Europe (especially those named C16 and C12; see Supplementary data) and 3 new haplotypes for this cluster (which would correspond to the consensus haplotypes C88, C89 and C92), which in conjunction confirm the existence of an important variability at the mtDNA level in the Iberian Peninsula for the so-called cluster Central of the species. On the other hand, the 104 mtDNA sequences reported in Fig. 2 in Randi et al. (2004) as sampled in Serbia, Montenegro and Kosovo are resumed in 9 of our consensus haplotypes (see Supplementary data). Moreover, we constructed a contracted median-joining network (Bandelt et al., 1999; Forster et al., 2001) using only the consensus haplotypes linked to cluster Central; contraction enables us to infer the 'real' phylogenies of the haplotypes, thus avoiding the conflicts produced by homoplasies (Fig. 6). The nodes including some of the consensus haplotypes obtained in our sampling tend to show interior locations in the clade.
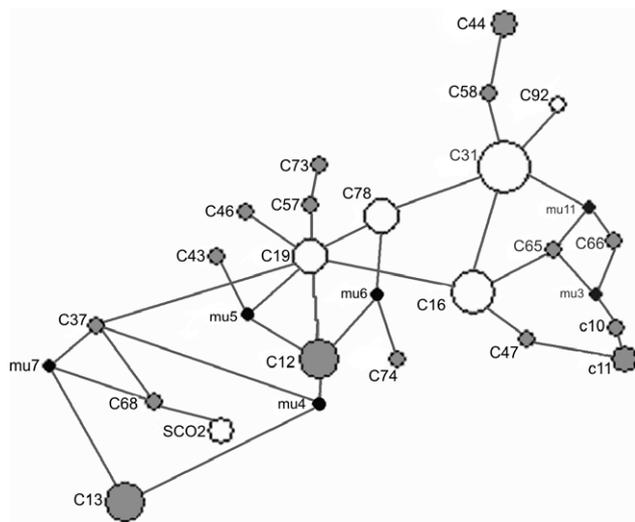

Fig. 6. Median-joining tree constructed with the consensus haplotypes linked to cluster Central. Small black circles indicate missing haplotypes. The network has been contracted for character conflicts. Open circles indicate haplotypes that have been obtained, at least once, in our sampling. Grey circles indicate haplotypes obtained by Randi et al. (2004). Within cluster Central, the haplotypes linked to the subspecies *C. c. italicus* are highlighted.

A relative temporal scale may be inferred from the hierarchy of haplotype position in a phylogenetic tree whereby most interior haplotypes are the oldest and those with outer positions are younger (Templeton et al., 1995; Branco et al., 2000). As an overall consequence, the Iberian origin of the cluster Central cannot be rejected, although it is possible that some parts of Eastern Europe, such as Serbia, may retain large genetic variability for the haplocluster.

Furthermore, the data obtained in the present analysis also suggest that the presence of haplotypes linked to cluster West in the Central Italian Alps (Randi et al., 2004) is not likely to be a consequence of an ancestral migration to this area of individuals from the Iberian Peninsula. In our sampling, we failed to find sequences coinciding with these haplotypes, suggesting that the presence of haplotypes assigned to Clade West in both Iberia and the Alps is more likely to be the result of two different glacial colonization events.

4.3. Conservation and management implications

The overall information obtained in the present analysis may provide support for differential conservation or management strategies for some Spanish roe deer populations. Our study revealed a complex genetic situation characterized by the presence of two haploclusters (West and Celtic-Iberian) that are characteristic of the Iberian Peninsula and that are mainly distributed in different Peninsular areas: the Central-Southern and the Northwestern areas, respectively. Moreover, these two geographical areas have significant genetic differentiation at microsatellite loci level. Two questions arise: (a) Can the Central-Southern and Northwestern Iberian roe deer populations be recognized as subpopulations? and (b) Should Central-Southern and Northwestern Iberian roe deer populations undergo particular conservation policies?

In recent years, scientific controversy has arisen with regard to recognizing the Andalusian roe deer population (mainly located in Cádiz at the Southwestern edge of the species distribution area) as a subspecies: *C. c. garganta* (Cabrera, 1916). This recognition would be based on evidence suggesting a genetic differentiation between Andalusian and Northern Spain roe deer at a mtDNA level (Lorenzini et al., 2003). However, Randi et al. (2004), with a limited number of samples, reported that the Cádiz population is not monophyletic for mtDNA haplotypes, most of them belonging to Clade Central, which is the most frequent clade in the whole European continent. Our study reveals that the Ca population may be monophyletic for the cluster West. However, the whole roe deer population distributed in the putative *C. c. garganta* spreading area (including the Andalusian and our Ce populations; Randi et al., 2004) are not monophyletic for the cluster West, though this cluster is predominant. The Northwestern roe deer population (which could in turn be assigned to the so-called subspecies *C. c. decorus*) would present a similar genetic scenario with a major presence of the relict Celtic-Iberian haplocluster.

When population distinctiveness is evaluated on the basis of genetic criteria, it is usually accepted that distinct genetic populations are those showing reciprocally monophyletic mtDNA alleles and significant divergence for allele frequencies at nuclear loci (Moritz, 1994). This criterion has been previously used to support the differentiation of *C. c italicus* as a subspecies (Randi et al., 2004). In our opinion, neither Northwestern nor Central-Southern populations would comply with this strict criterion and non *Capreolus capreolus* subspecies may be identified in the Iberian Peninsula. In our opinion, however, both the Northwestern and Central-Southern Iberian roe deer populations are distinguishable populations, *sensu* Crandall et al. (2000), and should undergo particular management policies.

Crandall et al. (2000) reviewed useful criteria for diagnosing population distinctiveness. Besides genetic criteria (Moritz, 1994), distinct populations may also be those that actually represent significant adaptive variation based on concordance between sets of data obtained using different techniques (Ryder, 1986) and those that are reproductively separated from other populations and have unique or different adaptations (Waples, 1991). Crandall et al. (2000) argue that the assessment of distinctiveness should appropriately incorporate adaptive differences, as evidenced by genetic and ecological data. The operational criteria for determining the genetic and ecological distinctiveness of populations proposed by Crandall et al. (2000) would reject population exchangeability (Templeton, 1989): (a) when there is evidence at the ecological level for population differentiation owing to genetic drift or natural selection or in life history traits, morphology, habitat, or selective traits, such differences being ideally demonstrably heritable; and (b) at the genetic level, there is evidence of restricted gene flow between populations or representative evidence could be obtained from microsatellites, sequences or other markers. In summary, the diagnosis of population distinctiveness should include both the factors that define the ecological niche in which the population evolves and the limits of spread of new genetic variants through genetic drift and natural selection in addition to the genetic factors defining the limits of spread of new genetic variants through gene flow (Crandall et al., 2000; Templeton, 1989).

Following Crandall et al.'s (2000) reasoning, despite their questionable identification as the *C. c. decorus* and *C. c. garganta* subspecies respectively, the Northwestern and Central-Southern Iberian roe deer populations merit further characterization efforts and the implementation of specific management policies for conservation in order to avoid their admixture with individuals reintroduced in their spreading areas. In this respect, it should be highlighted that, in Portugal, the Northern roe deer populations spread over the Duero river are managed according to specific conservation rules (Regulation no. 466/2001; A. Monzón, personal communication) and reintroductions with individuals from other areas are forbidden.

In summary, roe deer in Spain consitute an edge of variability for the species. Up to three main mtDNA haplogroups (Central, West and the herein identified Celtic-Iberian) can be found in our populations. Moreover, Spanish roe deer present a complex population structuring, which is probably the result of overlapped glacial colonisation events during the last glaciation periods. Both mtDNA and microsatellite information enable us to differentiate two main geographical areas that have probably acted as glacial refugia for the species: the Northwestern Iberian Peninsula, limited by the Duero river basin, and the Central-Southern part of the Iberian Peninsula including the Tajo river basin and the Mediterranean coast. In addition, we discuss the need for further characterization and implementation of conservation policies for the Northwestern and Central-Southern Spanish roe deer populations These roe deer populations are spread over particular ecosystems and, despite their non-monophyletic mtDNA origin, maintain mtDNA haploclusters (Celtic-Iberian and West, respectively) that may represent the remains of late glacial refugial populations as well as showing significant differentiation, as assessed through microsatellites. Consequently, these populations should be preserved and management programmes should avoid reintroduction or artificial admixture with other geographical populations. Furthermore, in this study we have demonstrated the possibilities of using molecular coancestry to characterise wild animal populations.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2006.05.020.

References

Álvarez, I., Royo, L.J., Gutiérrez, J.P., Fernández, I., Gómez, E., Goyache, F., 2004. Genetic relationships and admixture among sheep breeds from Northern Spain assessed using microsatellites. J. Anim. Sci. 82, 2246–2252.

Álvarez, I., Gutiérrez, J.P., Royo, L.J., Fernández, I., Gómez, E., Arranz, J.J., Goyache, F., 2005. Testing the usefulness of the molecular

coancestry information to assess genetic relationships on livestock using a set of Spanish sheep breeds. J. Anim. Sci. 83, 737–744.

Aragón, S., Braza, F., San José, C., 1995a. Socioeconomic, physiognomic, and climatic factors determining the distribution pattern of roe deer *Capreolus capreolus*. Spain Acta Theriol. 40, 37–43.

Aragón, S., Braza, F., San José, C., 1995b. Características morfológicas de los corzos (*Capreolus capreolus*) de las sierras de Cádiz- Málaga. Doñana Acta Vertebrata 22, 51–64.

Arribas, O., 2004. Fauna y paisaje de los Pirineos en la Era Glaciar. Fundació Territori i Paisatge, Barcelona, Spain.

Bandelt, H.-L., Forster, P., Rohl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37–48.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., Bonhomme, F., 2001. GENETIX 4.02, logiciel sous Windows TM pour la génétique des populations. Montpellier, France: Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II.

Boisaubert, B., Boutin, J.M., 1988. Le Chevreuil. Hatier, Paris, 237 pp.

Boutin, J.M., 1990. Le statu du chevreuil (*Capreolus capreolus*) au Portugal et les recherches effectuées sur cette spèce. B.M.O.N.C. 143, 35–37.

Branco, M., Ferrand, N., Monnerot, M., 2000. Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome *b* gene. Heredity 85, 307–317.

Braza, F., Varela, I., San José, C., Cases, V., 1989. Distribution actuelle du chevreuil (*Capreolus capreolus*), du daim (*Dama dama*) et du cerf (*Cervus Elaphus*) en Espagne. Zeitschrift für Säugetierkunde 54, 393–396.

Caballero, A., Toro, M.A., 2002. Analysis of genetic diversity for the management of conserved subdivided populations. Conserv. Genet. 3, 289–299.

Cabrera, A., 1916. Un nuevo cérvido español. Boletin de la Real Sociedad Española de Historia Natural. Tomo XVI. Museo Nacional de Ciencias Naturales. Madrid, pp. 175–176.

Conde de Yebes, 2000. Veinte años de caza mayor. Ediciones Al-Andalus, Sevilla, Spain.

Coulon, A., Cosson, J.F., Angibault, J.M., Cargnelutti, B., Galan, M., Morellet, N., Petit, E., Aulagnier, S., Hewison, A.J., 2004. Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual-based approach. Mol. Ecol. 13, 2841–2850.

Crandall, K.A., Bininda-Emonds, O.R.P., Mace, G.M., Wayne, R.K., 2000. Considering evolutionary processes in conservation biology. Tree 15, 290–295.

Douzery, E., Randi, E., 1997. The mitochondrial control region of Cervidae: evolutionary patterns and phylogenetic content. Mol. Biol. Evol. 14, 1154–1166.

Eding, H., Crooijmans, R.P.M.A., Groenen, M.A.M., Meuwissen, T.H.E., 2002. Assessing the contribution of breeds to genetic diversity in conservation schemes. Genet. Sel. Evol. 34, 613–633.

Eding, H., Meuwissen, T.H.E., 2001. Marker-based estimates of between and within population kinships for the conservation of genetic diversity. J. Anim. Breed. Genet. 118, 141–159.

Fabuel, E., Barragan, C., Silio, L., Rodríguez, M.C., Toro, M.A., 2004. Analysis of genetic diversity and conservation priorities in Iberian pigs based on microsatellite markers. Heredity 93, 104–113.

Fandos, P., Reig, S., 1993. Craniometric variability in two populations of roe deer (*Capreolus capreolus*) from Spain. J. Zool. 231, 39–49.

Feng, D.F., Doolittle, R.F., 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. 25, 351–360.

Festa, E., 1925. Il capriolo dell'Italia centrale (*Capreolus capreolus italicus* sbsp. *nova*). Bollettino del Museo di Zoologia e Anatomia Comparata Dell'università di Torino 7, 1–2.

Forster, P., Torroni, A., Renfrew, C., Röhl, A., 2001. Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. Mol. Biol. Evol. 18, 1864–1881.

Fu, Y.-X., 1997. Statistical tests of neutrality against population growth, hitchhiking and background selection. Genetics 147, 915–925.

Fu, Y.-X., Li, W.-H., 1993. Statistical tests of neutrality of mutations. Genetics 133, 693–709.

Gaggiotti, O.E., Excoffier, L., 2000. A simple method of removing the effect of a bottleneck and unequal population sizes on pairwise genetic distances. Proc. R. Soc. Lond., B 267, 81–87.

Gortázar, C., Herrero, J., Villafuerte, R., Marco, J., 2000. Historical examination of the status of large mammals in Aragón, Spain. Mammalia 64, 411–422.

Gutiérrez, J.P., Royo, L.J., Álvarez, I., Goyache, F., 2005. MolKin v2.0: A Computer Program for Genetic Analysis of Populations Using Molecular Coancestry Information. J. Hered. 96, 718–721.

Hewitt, G.M., 1999. Post-glacial re-colonization of European biota. Biol. J. Linn. Soc. 68, 87–112.

Hudson, R.R., 1990. Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. 7, 1–44.

Hurlbert, S.H., 1971. The non concept of species diversity: a critique and alternative parameters. Ecology 52, 577–586.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA2: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. Brief. Bioinform. 5, 150–163.

Lister, A.M., Grubb, P., Sumner, S.R.M., 1998. Taxonomy, morphology and evolution of European roe deer. In: Andersen, R., Duncan, P., Linnell, J.D.C. (Eds.), The European roe deer: the biology of success, Oslo, vol. 2, pp. 23–46.

Lorenzini, R., San José, C., Braza, F., Aragón, S., 2003. Genetic differentiation and phylogeography of roe deer in Spain, as suggested by mitochondrial DNA and microsatellite analysis. Ital. J. Zool. 70, 89–99.

Malécot, G., 1948. Les Mathématiques de l'Hérédité. Masson et Cie, Paris, France.

Meunier, K., 1983. Das Spanisches Reh. In: Hoffman, R.R. (Ed.), Wildbiologisches Informationen fur den Jager.

Miller, G.S., 1910. *Capreolus capreolus canus*. Ann. And Mag. Nat. Hist. 8th ser. VI, p. 460.

Moritz, C., 1994. Defining 'evolutionary significant units' for conservation. Trends Ecol. Evol. 9, 373–375.

Nei, M., 1972. Genetic distance between populations. Am. Nat. 106, 283–292.

Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York; University Press, New York.

Poetsch, M., Seefeldt, S., Maschke, M., Lignitz, M., 2001. Analysis of microsatellite polymorphism in red deer, roe deer, and fallow deer— possible employment in forensic applications. Forensic Sci. Int. 116, 1–8.

Randi, E., Pierpaoli, M., Danilkin, A., 1998. Mitochondrial DNA polymorphism in populations of Siberian and European roe deer (*Capreolus pygargus* and *C. capreolus*). Heredity 80, 429–437.

Randi, E., Alves, P., Carranza, J., Miloševi-Zlatanovi·, S., Sfougaris, A., 2004. Phylogeography of roe deer (*Capreolus capreolus*) populations: the effects of historical genetic subdivisions and recent nonequilibrium dynamics. Mol. Ecol. 13, 3071–3083.

Reynolds, J., Weir, B.S., Cockerham, C.C., 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 105, 767–769.

Rogers, A.R., Harpending, H., 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. 9, 552–569.

Rosell, C., 2002. Los ungulados y los ecosistemas forestales: los ejemplos del jabalí y el corzo. In: Gestión y conservación de la biodiversidad en ecosistemas forestales. Aplicación a la fauna vertebrada. Centre Tecnològic Forestal (Ed.). Universitat de Barcelona. Universitat de Lleida, pp. 368–386.

Royo, J.L., Álvarez, I., Beja-Pereira, A., Molina, A., Fernández, I., Jordana, J., Gómez, E., Gutiérrez, J.P., Goyache, F., 2005. The origins of Iberian horses assessed via mitochondrial DNA. J. Hered. 96, 663–669.

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP, DNA polymorphisms analyses by the coalescent and other methods. Bioinformatics 19, 2496–2497.

Ryder, O.A., 1986. Species conservation and systematics: the dilemma of subspecies. Trends Ecol. Evol. 1, 9–10.

SAS/STAT™, 1999. User's Guide. Release 8.2. SAS Institute Inc, Cary NC.

Sbisà, E., Tanzariello, F., Reyes, A., Pesole, G., Saccone, C., 1997. Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene 205, 125–140.

Schneider, S., Roessli, D., Excoffier, L., 2000. Arlequin v.2.0: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.G., Cosson, J.F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. Mol. Ecol. 7, 453–464.

Tellería, J.L., Virgós, E., 1997. Distribution of an increasing roe deer population in fragmented Mediterranean landscape. Ecography 20, 247–252.

Templeton, A.R., 1989. The meaning of species and speciation: a genetic perspective. In: Otte, D., Endler, J.A., (Eds.), Speciation and its Consequences, Sinauer, pp. 3–27.

Templeton, A.R., Routman, E., Phillips, C.A., 1995. Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. Genetics 140, 767–782.

Vernesi, C., Pecchioli, E., Caramelli, D., Tiedermann, R., Randi, E., Bertorelle, G., 2002. The genetic structure of natural and reintroduced roe deer (*Capreolus capreolus*) populations in the Alps and central Italy, with reference to the mitochondrial DNA phylogeography of Europe. Mol. Ecol. 11, 1285–1297.

Virgós, E., Tellería, J.L., 1998. Roe deer habitat selection in Spain: constraints on the distribution of a species. Can. J. Zool. 76, 1294–1299.

von Lehmann, E., Sägesser, H., 1986. *Capreolus capreolus Linnaeus*, 1758. Reh. In: Niethammer, J., Krapp, F., (Eds.), Handbuch der Saügetiere Europas. vol. 2, pp. 233–268.

Wahlström, L.K., Liberg, O., 1995. Patterns of dispersal and seasonal migration in roe deer (*Capreolus capreolus*). J. Zool. 235, 455–467.

Waples, R.S., 1991. Pacific salmon, Oncorhynchus spp., and the definition of 'species' under the endangered species act. Mar. Fish. Rev. 53, 11–22.

Wiehler, J., Tiedemann, R., 1998. Phylogeography of the European roe deer *Capreolus capreolus* as revealed by sequence analysis of the mitochondrial control region. Acta Theriolo. Suppl. 5, 187–197.

Wright, S., 1969Evolution and the Genetics of Populations: The Theory of Gene Frequencies, vol. 2. University of Chicago Press, Chicago, USA.