

Genetic structure is influenced by landscape features: empirical evidence from a roe deer population

A. COULON,* G. GUILLOT,† J.-F. COSSON,‡ J. M. A. ANGIBAULT,* S. AULAGNIER,*
B. CARGNELUTTI,* M. GALAN*‡ and A. J. M. HEWISON*

*Comportement et Ecologie de la Faune Sauvage, INRA, B.P.52627, F31326 Castanet-Tolosan cedex, France, †Mathématique et Informatique Appliquées, INA, 16, rue Claude Bernard, F75231 Paris cedex 5, France, ‡Centre de Biologie et de Gestion des Populations, INRA, Campus International de Baillarguet, CS 30 016, F34988 Montferrier-sur-Lez, France

Abstract

The delimitation of population units is of primary importance in population management and conservation biology. Moreover, when coupled with landscape data, the description of population genetic structure can provide valuable knowledge about the permeability of landscape features, which is often difficult to assess by direct methods (e.g. telemetry). In this study, we investigated the genetic structuring of a roe deer population which recently recolonized a fragmented landscape. We sampled 1148 individuals from a 40 × 55-km area containing several putative barriers to deer movements, and hence to gene flow, namely a highway, rivers and several canals. In order to assess the effect of these landscape features on genetic structure, we implemented a spatial statistical model known as GENELAND which analyses genetic structure, explicitly taking into account the spatial nature of the problem. Two genetic units were inferred, exhibiting a very low level of differentiation ($F_{ST} = 0.008$). The location of their boundaries suggested that there are no absolute barriers in this study area, but that the combination of several landscape features with low permeability can lead to population differentiation. Our analysis hence suggests that the landscape has a significant influence on the structuring of the population under study. It also illustrates the use of GENELAND as a powerful method to infer population structure, even in situations of young populations exhibiting low genetic differentiation.

Keywords: barriers, Bayesian computations, *Capreolus capreolus*, connectivity, landscape genetics, population structure

Received 3 August 2005; revision received 14 November 2005; accepted 29 November 2005

Introduction

The habitat requirements of individuals, which may be more or less pronounced among species, determine the way they use the landscape they inhabit (e.g. Arnold *et al.* 1995 on kangaroos; Mönkkönen *et al.* 1995 on flying squirrels; Luiselli & Capizzi 1997 on snakes). The understanding of this relationship between landscape structure and species biology is the basis of landscape ecology. It can provide information about population functioning which may be valuable cues for management and conservation decisions

(Moritz 1994). An important part of this understanding concerns the study of animal movements in relation to landscape structure, i.e. landscape connectivity, defined as ‘the degree to which a landscape facilitates or impedes movement among resource patches’ by Taylor *et al.* (1993). Effective movements (i.e. movements followed by successful reproduction) determine gene flow level and direction across the landscape. Studying gene flow in relation to landscape structure can thus give valuable information about the features that influence effective movements, which is often difficult to assess by direct methods such as observation or telemetry. This type of study comes under the mantle of the recent development of landscape genetics (Manel *et al.* 2003).

One way to study gene flow within a landscape is to identify genetic units and the features that are responsible

Correspondence: Aurélie Coulon, Institut National de Recherche Agronomique, Laboratoire comportement et Ecologie de la Faune Sauvage, B.P.52627, F31326 Castanet-Tolosan cedex, France. Fax: +33 5 61 28 55 00; E-mail: coulona@toulouse.inra.fr

for the differentiation of these units, hence deducing which features restrict or promote movements of individuals (Manel *et al.* 2003). Until recently, genetic structuring was studied by defining populations on a priori basis, and gene flow was measured mainly by F_{ST} or similar parameters (e.g. Ehrich & Stenseth 2001; Schulte-Hostedde *et al.* 2001; Doums *et al.* 2002; O'Reilly *et al.* 2004). This method suffers from several limitations. First, the inferred structure is forced by the prior reasoning and depends on assumptions concerning population limits which are reliable to an unknown extent. Second, this approach is not very informative for small areas and nonclustered sampling design. Recently, the development of Bayesian clustering methods has enabled the inference of genetic units using genotypes of individuals as the sole source of information (Pritchard *et al.* 2000; Dawson & Belkhir 2001; Corander *et al.* 2003; Falush *et al.* 2003). In short, these methods attempt to partition individuals into groups at Hardy–Weinberg equilibrium. They have quickly become widely used (e.g. Caizergues *et al.* 2003; Gyllenstrand & Seppä 2003; Rueness *et al.* 2003; Hampton *et al.* 2004; Cimmaruta *et al.* 2005). A recent extension of these methods concerns the integration of spatial information (the spatial coordinates of an animal's home range) in the modelling and inference process (Guillot *et al.* 2005a). This approach has yet to be tested *in natura* on a large spatially referenced wildlife population.

In this study, we implemented two Bayesian methods to infer genetic structure of European roe deer (*Capreolus capreolus*) at a small spatial scale (55 × 40 km) for a nonclustered distribution of individuals across a fragmented landscape. Our aim was to detect which landscape features influence roe deer movements and should be considered as potential barriers between individuals, with particular respect to the differentiation of management units. Moreover, because the study area was recently recolonized by roe deer (second half of the 20th century), we expected genetic differentiation, if any, to be weak (unless the area was recolonized by several well-differentiated source populations). As a consequence, this study is an interesting case for testing the ability of available methods to detect weak population genetic structure.

The roe deer is a medium-sized ungulate that is widely spread across Europe. Populations have increased substantially since the second half of the 20th century, triggering conflicts with foresters and farmers, as well as problems of roe deer–vehicle collisions (Cederlund *et al.* 1998). Consequently, roe deer populations are managed through hunting plans, and knowledge regarding population functioning is required, particularly for defining meaningful management units.

The roe deer inhabits predominantly woodland areas, but it has also colonized fragmented and more open agricultural areas in the last decades (Hewison *et al.* 1998, 2001). In fragmented landscapes, roe deer seem to be strongly tied to wooded structures (Hewison *et al.* 2001), determining

the size and composition of their home range (Cargnelutti *et al.* 2002) and their dispersal behaviour (Coulon *et al.* 2004). Moreover, Wang & Schreiber (2001) showed that the proportion of urban areas (houses, villages and roads) between populations was significantly correlated with the level of genetic differentiation between them, indicating an effect of these features as gene flow moderators. However, information regarding the influence of other landscape features on roe deer movements remains very sparse.

Roe deer occupy spatially restricted home ranges (less than 100 ha in fragmented landscapes, Cargnelutti *et al.* 2002 and unpublished data). Movement distances are maximal during natal dispersal, which occurs at 1, or sometimes 2, years of age. Little published information is available on dispersal patterns (but see Wahlström 1994; Wahlström & Liberg 1995; Linnell *et al.* 1998), and while it seems likely that landscape fragmentation alters dispersal behaviour compared to nonfragmented landscapes, almost no information exists. Preliminary dispersal data from our study area of a fragmented landscape suggest a mean dispersal distance of around 3 km (data from 9 dispersing individuals out of 63 marked roe deer; unpublished data). Post dispersal, roe deer show high site fidelity among years (Danilkin & Hewison 1996), except for some reproductive females that may make short (a few hundred metres to a few kilometres) rut excursions, leaving their home range for a few days before returning (San José & Lovari 1998). Movements of individuals hence occur over relatively small distances which could trigger the differentiation of genetic units at a small spatial scale.

We used 11 microsatellite markers to investigate the delimitation of genetic units in our study area of fragmented woodland. We tested the hypothesis that highways, canals and rivers affect perceptibly roe deer effective movements and contribute to determining population genetic structure.

Materials and methods

Study area

The study area is a hilly region of 55 × 40 km in southwestern France, where elevation ranges from 180 to 700 m (mean = 330 m). The area is mainly rural, with cultivated fields (maize, sorghum, wheat, sunflower), meadows (for grazing), villages, farms and only a few small towns (Fig. 1). The wooded habitat is heavily fragmented, with a few main forests in the centre and numerous interspersed small patches of woods, connected to each other to a varying degree by hedgerows (10.5 wood patches per 100 ha, mean area = 2.4 ha). The area is bisected by a fenced highway, several canals and the Garonne River (Fig. 1), which we hypothesize are potential barriers to roe deer movement. The four-lane highway is at least a partial barrier as a fence of 2.5 m cannot easily be jumped by roe

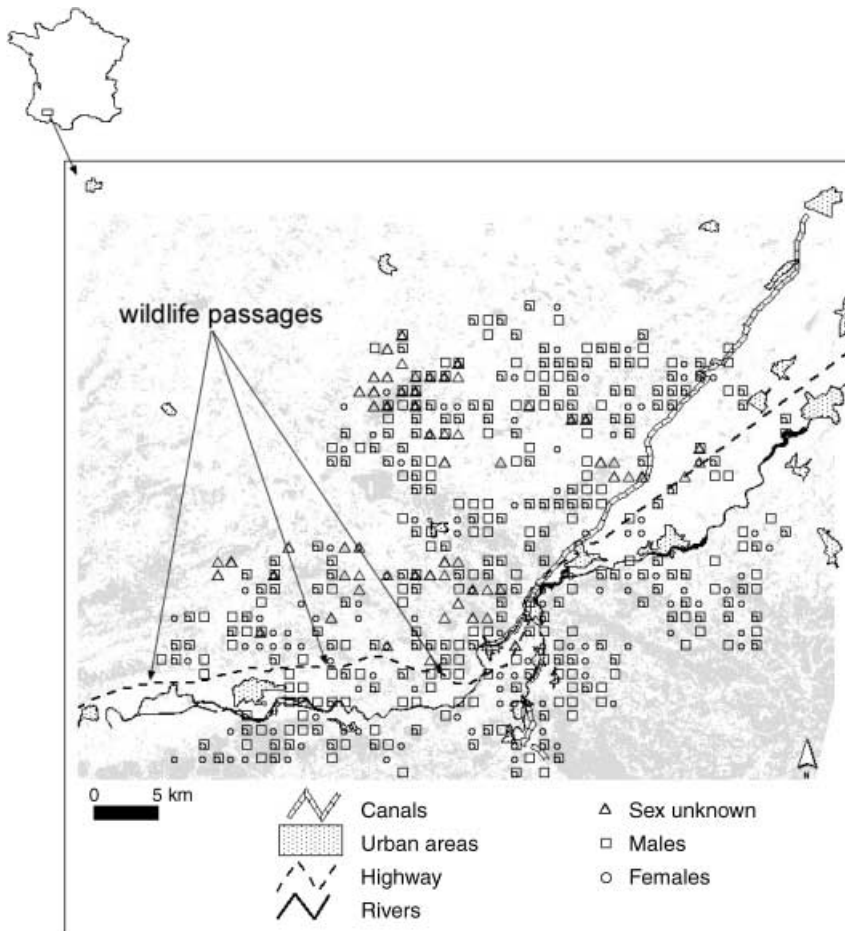


Fig. 1 Study area for the genetic structuring: sample locations and main landscape features (wooded patches are represented in grey). Arrows indicate the location of the three designated wildlife passages built along the highway.

deer. In addition, high traffic load likely leads to a high mortality risk for those animals that do manage to penetrate the fence, usually through holes due to poor maintenance. However, across the study area, there are three designated wildlife passages and numerous other bridges and drainage tunnels that provide possible routes for deer to cross the highway. The canals are also at least partial barriers as most have steep-sided concrete banks from which roe deer find it impossible to escape once they have fallen in, leading to a significant number of drownings. While the Garonne River is potentially less of a barrier, as roe deer are able to swim well (Danilkin & Hewison 1996), it is a wide river (mean width around 80 m) which is also associated with significant human activity in the form of industry and settlements.

Genetic sampling

Roe deer skin samples were collected by hunters during legal hunting (winters 2000–2001, 2001–2002, and 2003–2004) and by ourselves during live trapping carried out to fit individuals with radio-collars for other purposes (winters 2000–2001 to 2003–2004). The individuals collected before 2003 were almost exclusively from the

area to the north of the potential barriers, and analyses ran with *PARTITION* (Belkhir & Dawson 2001) showed they constituted one unique genetic unit (Coulon *et al.* 2004); most of the additional samples collected subsequently were to the south of the potential barriers, with a few taken in the north to fill spatial gaps in the previous data set. Each individual was sexed and assigned a geographical coordinate corresponding to the centre of the square kilometre of the Lambert grid where it was killed or caught (for samples provided by hunters, the site of shooting was indicated by their marking a cross on provided maps of their hunting area). These coordinates can be considered as a good estimate of the location of the individual's home range as ranges are generally less than 100 ha in our study area (Cargnelutti *et al.* 2002 and unpublished data) and, during flight, roe deer circle within or near their home range (personal observation). We obtained samples from 1148 individuals (481 females, 519 males and 148 individuals for which sex was unrecorded) distributed across the study area (Fig. 1).

DNA conservation and genotyping

Skin samples were stored in 95% ethanol prior to DNA extraction which was carried out using the DNeasy Tissue

Kit (QIAGEN). DNA was amplified using the polymerase chain reaction and genotyped with a multiplex panel of 12 microsatellites (Galan *et al.* 2003) in a genotyper ABI PRISM 310 Genetic Analyser (Applied Biosystems). Genotypes were determined using GENESCAN and GENOTYPER software (Applied Biosystems). We did not use the NVHRT48 locus of the kit for the analyses because of difficulties for interpreting the gel patterns; hence, our analysis is based on 11 microsatellite loci.

Standard population genetic analyses

We used GENEPOP 3.4 (Raymond & Rousset 1995) to perform all standard population genetic analyses. We tested for heterozygote deficiency as compared to Hardy–Weinberg equilibrium for each locus and globally using the Markov chain method [parameter values: for the test at the locus level, dememorization number = 10 000 (default value), number of batches = 400, number of iterations of batches = 7000; for the global test, the same values were used except for the number of iterations of batches which was set to 3000; default values were increased in order to obtain standard deviations for the P value estimates < 0.01 as suggested by the authors]. In order to quantify any inferred deviation from Hardy–Weinberg equilibrium, we calculated Weir & Cockerham's estimate of F_{IS} (1984), for each locus and also globally. We tested genotypic linkage disequilibrium for each pair of loci [parameter values: dememorization number = 10 000 (default value), number of batches = 800, number of iterations of batches = 8000]. To control for multiple testing, we did not choose the widely used Bonferroni correction as it has recently been criticized (e.g. Moran 2003). Instead, we used the false discovery rate (FDR) control for the P values of the Hardy–Weinberg tests on individual loci and of the linkage disequilibrium tests (Verhoeven *et al.* 2005), with the function `fdr.control` implemented in the package GENETS for R (Benjamini 1995; Storey 2002).

Delimitation of genetic units

The genetic structure of the roe deer population was investigated using two clustering methods based on Bayesian models. First, we used the STRUCTURE software version 2.1 developed by Pritchard *et al.* (2000) and enhanced by Falush *et al.* (2003), as it is the standard reference software for such analysis. Although there are some queries regarding the procedure for estimating the number of populations (Evanno *et al.* 2005), we selected this method because of the variety of modelling options available. Second, we used the GENELAND software version 0.3 recently developed by Guillot *et al.* (2005b), because it may provide a better definition of spatial genetic units by integrating spatial coordinates of samples, as explained below.

As in all Bayesian models (see, e.g. Beaumont & Rannala 2004 for a recent review of Bayesian modelling in genetics), the assumptions in STRUCTURE are of two types. There is a prior distribution for unobserved quantities (mostly clustering and allele frequencies) and a likelihood function relating these unknown parameters to the observed genotypes. In the Bayesian paradigm, the prior distribution should reflect knowledge and uncertainty about the unobserved parameters. In STRUCTURE (and actually in all other non-spatial clustering methods), the assumed prior for the clustering is uniform. This translates into the assumption that all clustering solutions are equally likely. In this classical approach, the spatial coordinates of individuals are not considered during the processing scheme. They are simply used once some structure has been inferred, to visualize inferred individual population membership on a map. This contradicts somewhat the notion of looking for differentiated populations. Indeed, except in the case of recent introgression or of behavioural barriers between populations (in which case a uniform prior for the clustering process fully makes sense), differentiation can be expected only if populations are separated either by barriers or by distance. This observation was the starting point of the recent work of Guillot *et al.* (2005a). These authors proposed to explicitly model the fact that differentiated populations tend to be structured in spatially distinct areas. Their model thus incorporates the spatial coordinates of animals at an earlier stage of the data processing. This approach generates maps of population ranges. Moreover, it is fully Bayesian in the sense that the number of populations is treated as a parameter processed by the Markov chain Monte Carlo (MCMC) scheme without any approximation. This method takes into account location errors (induced by measurement error) by introducing an additive noise to the coordinates, the true coordinates being treated as unknown and as parameters to be estimated. In both software the unknown parameters are inferred through MCMC computations.

STRUCTURE procedure and parameter values

In STRUCTURE, the number of populations K is a fixed parameter of the model; the procedure consists in running several MCMC with different values for K (several runs for each K in order to verify the consistency of the results) and inferring which K is the most likely from the approximation of their posterior probabilities. We performed several runs for each K , from $K = 1$ to 10, and calculated for each K the mean posterior probability calculated over its runs. We then used these mean values to estimate the posterior probability of each K according to the formula given by Pritchard & Wen (2003).

Roe deer colonized the study area around the 1960s and the population is thus quite young. Thus, the genetic structure, if present, was expected to be slight, and allele

frequencies to be similar in the different populations, supposing a single recent origin (which is unknown). We thus used the admixture model, where each individual is assumed to have inherited some proportion of its ancestry from each population. In this model, individuals are clustered jointly into two or more populations if their genotypes indicate that they are admixed; we allowed the degree of admixture to be different for each population. For the same reason, we used the option where the prior takes into account the fact that allele frequencies among populations may be correlated. We did not use a priori information about population affiliation. Burn-in length was fixed to 100 000 following the suggestions of the authors (Pritchard & Wen 2003). After different trials where we looked for MCMC convergence and consistency among runs with identical parameter values, an MCMC length of 1 500 000 iterations seemed the most suitable. Because such long runs are highly time-consuming, we carried out two types of run: because we suspected the presence of at least two populations, but not many more, for $K = 1$ to 4, we performed six long runs per value of K , with MCMC lengths of 1 500 000 iterations; for $K = 5$ to 10, we made five short runs per K , with 500 000 iterations. We also made two short runs per K for $K = 1$ to 4, in order to be able to compare the results for all K values. Other parameter values were set to default values.

GENELAND procedure and parameter values

In GENELAND, all the parameters (including K) are processed simultaneously by the MCMC algorithm. However, for some technical reasons discussed in Guillot *et al.* (2005a), it is better to infer K in a first run and then to run the algorithm again with K fixed at the previously inferred value in order to estimate the other parameters (mainly the assignment of individuals to the inferred populations). We ran the MCMC five times (to verify the consistency of the results), allowing K to vary, with the following parameters: 500 000 MCMC iterations, maximum rate of Poisson process fixed to 500, uncertainty attached to spatial coordinates fixed to 1 km (i.e. the precision of our sample locations, see above), minimum K fixed to 1, maximum K fixed to 30, maximum number of nuclei in the Poisson–Voronoi tessellation fixed to 200. We used the Dirichlet model as a model for allelic frequencies as it has been demonstrated to perform better than the alternative model (Guillot *et al.* 2005a). We then inferred the number of populations in our sample from the modal K of these five runs, and ran the MCMC 100 times with K fixed to this number. Other parameters remained similar to those of the runs with variable K . We calculated the mean logarithm of posterior probability for each of the 100 runs, and selected the 10 runs with the highest values. The posterior probability of population membership for each pixel of the

spatial domain was then computed for each of these 10 runs (using a burn-in of 50 000 iterations). The number of pixels was set to values notably higher than those recommended (e.g. on the GENELAND homepage www.inapg.inra.fr/ens_rech/mathinfo/personnel/guillot/Geneland.html), 500 pixels along the X axis and 380 along the Y axis (values chosen so as to avoid having two individuals in the same pixel). We then computed the posterior probability of population membership for each pixel of the spatial domain and the modal population of each individual. We finally checked the consistency of the results across these 10 runs. We also compared these results with those of three runs with 1 million MCMC each in order to verify their consistency with longer run lengths.

Standard population genetic analyses on the inferred populations

We then performed with GENEPOP the same standard population genetic analyses (as on the whole sample above) on the inferred (partitioned) population units: we tested for Hardy–Weinberg equilibrium for each locus (parameter values: dememorization number = 10 000, number of batches = 300, number of iterations per batch = 5000) and among all loci (dememorization number = 10 000, number of batches = 200, number of iterations per batch = 2000); we calculated F_{IS} values; and we looked for genotypic linkage disequilibrium [parameter values: dememorization number = 10 000 (default value), number of batches = 800, number of iterations of batches = 8000]. P values were weighted using the FDR method for multiple testing. We estimated population differentiation by calculating pairwise F_{ST} values after Weir & Cockerham (1984); bootstrapping over loci allowed us to estimate confidence intervals (tests carried out with FSTAT 2.9.3.2.; Goudet 2001). Lastly, we tested the significance of the inferred structure by performing a two-level analysis of molecular variance (AMOVA) (among and within populations) with the software ARLEQUIN 2.00 (Schneider *et al.* 2000).

Results

Standard population genetic analyses on the whole sample

The mean number of alleles per locus was 7.5, ranging from 2 to 19 (Table 1). Global F_{IS} was 0.01797 and heterozygote deficiency was significant ($P = 0.0003$) (Table 1). This deficit was due to three loci. There was no significant linkage disequilibrium.

Genetic structure analyses

Numerical results of the STRUCTURE analyses are summarized in Fig. 2. The estimated logarithm of

Table 1 Summary of basic population genetic analyses for the whole sample and for the inferred NW and SE populations from the third run of GENELAND analyses

Locus name	No. of alleles	Whole sample				NW population				SE population			
		H_O	H_E n.b.	P value	F_{IS}	H_O	H_E n.b.	P value	F_{IS}	H_O	H_E n.b.	P value	F_{IS}
IDVGA-29	2	0.3400	0.3402	0.5261	0.00054	0.3236	0.3195	0.6835	-0.013	0.4233	0.4265	0.5229	0.008
CSSM039	7	0.5528	0.5709	0.0160*	0.03164	0.5549	0.5728	0.0128*	0.031	0.5426	0.5589	0.3552	0.029
IDVGA-8	5	0.6579	0.6884	0.0744	0.04424	0.6520	0.6832	0.0622	0.046	0.6878	0.7105	0.4287	0.032
BM1706	6	0.6437	0.6488	0.3409	0.00793	0.6395	0.6387	0.4009	-0.001	0.6649	0.6836	0.3849	0.027
OarFCB304	19	0.9060	0.9172	0.0071*	0.01221	0.9034	0.9170	0.0035*	0.015	0.9144	0.9063	0.7214	-0.009
HUJ1177	8	0.6709	0.7007	0.0037*	0.04268	0.6692	0.6955	0.0053*	0.038	0.6793	0.7242	0.1679	0.062
BMC1009	7	0.6740	0.6929	0.0295	0.02732	0.6729	0.6968	0.0180*	0.034	0.6793	0.6713	0.6547	-0.012
BM848	8	0.7476	0.7875	0.0103*	0.05064	0.7586	0.7894	0.0265	0.039	0.6915	0.7583	0.0068	0.088
CSSM041	3	0.6458	0.6379	0.7879	-0.01247	0.6375	0.6400	0.5498	0.004	0.6882	0.6267	0.9538	-0.098
BM757	13	0.8485	0.8470	0.6941	-0.00178	0.8449	0.8392	0.8834	-0.007	0.8670	0.8774	0.2125	0.012
CSSM043	4	0.7046	0.6955	0.4878	-0.01301	0.7123	0.6960	0.7054	-0.024	0.6649	0.6812	0.1583	0.024
All loci	7.5	0.6720	0.6843	0.0003*	0.01797	0.6699	0.6807	0.0006*	0.016	0.6821	0.6932	0.1742	0.01601

H_O , observed heterozygosity; H_E n.b., heterozygosity expected under Hardy–Weinberg equilibrium, corrected for sampling bias. P values of the tests for heterozygosity deficiency are given in brackets, under H_E n.b.; they are followed by an asterisk when significant at the 0.05 level (after FDR control, except for the last line, where no correction was applied). F_{IS} was calculated after Weir & Cockerham (1984). The last row gives values of the mean number of alleles, the mean heterozygosities and F_{IS} calculated over all loci.

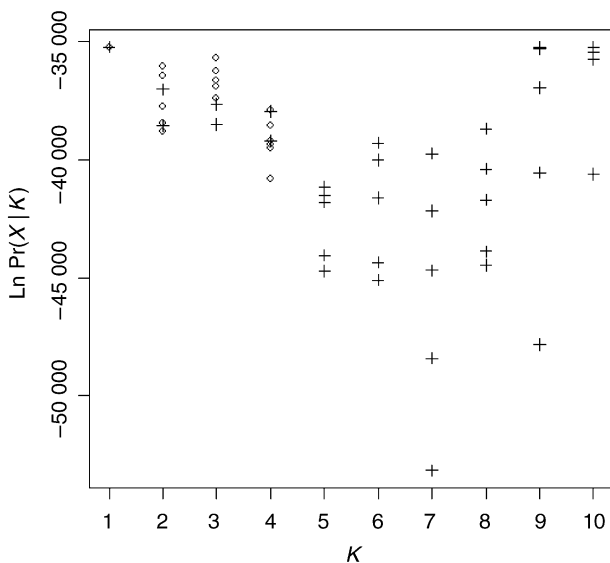


Fig. 2 Results of STRUCTURE analyses: plot of the estimated logarithms of probability of data [$\ln \Pr(X|K)$] against the number of populations tested (K), for the two types of runs carried out: long runs are symbolized by circles, short runs by crosses.

probability of the data [$\ln \Pr(X|K)$] was maximum for $K=1$ and the posterior probability of having only one population was 1. However, $\ln \Pr(X|K)$ was almost as high for $K=3, 9$ and 10 . We hence looked at the estimated membership of each individual in the clusters defined by the best runs [i.e. the runs with the highest $\ln \Pr(X|K)$] for $K=3$ and $K=10$, and assigned them to the population for

which the estimated membership was the highest. For $K=3$, 529 individuals were assigned to one population, and the remaining 619 to a second one (no individual was assigned to the third population). However, for each individual, the membership of the two populations was very close to 50% (mean membership to population 1 = 47%, 2 = 48%, 3 = 5%), and there was no obvious spatial clustering in the assignments. For $K=10$, each individual was assigned with the highest probability to a single population, with a mean membership of 18.8% (mean memberships of the other nine populations: 17.3%, 10.1%, 6.9%, 6.3%, 8.2%, 5.7%, 6.1%, 15.4%, 5.3%). Hence, there was no real support for more than one population as inferred by STRUCTURE. Moreover, we observed high fluctuations of alpha (Dirichlet parameter for the degree of admixture) during the runs, which, according to Pritchard & Wen (2003), is an indication of the absence of any real detected structure.

GENELAND runs gave different modal numbers of populations: 8, 9, 11 and 14 (twice). As GENELAND sometimes detects populations which are modal populations for none of the individuals (these populations being hereafter referred to as ghost populations), that is to say populations to which no individuals are assigned (Guillot *et al.* 2005a), we decided in the first instance to fix the number of populations to the minimum modal number (8) in the runs for estimating the other parameters. Indeed, the 10 selected runs of the 100 processed (based on their mean logarithm of posterior probability) resulted in the assignment of individuals to only two populations (except for one run, see below), so the other six of the eight populations were in fact ghost populations (Fig. 3). The areas of these ghost

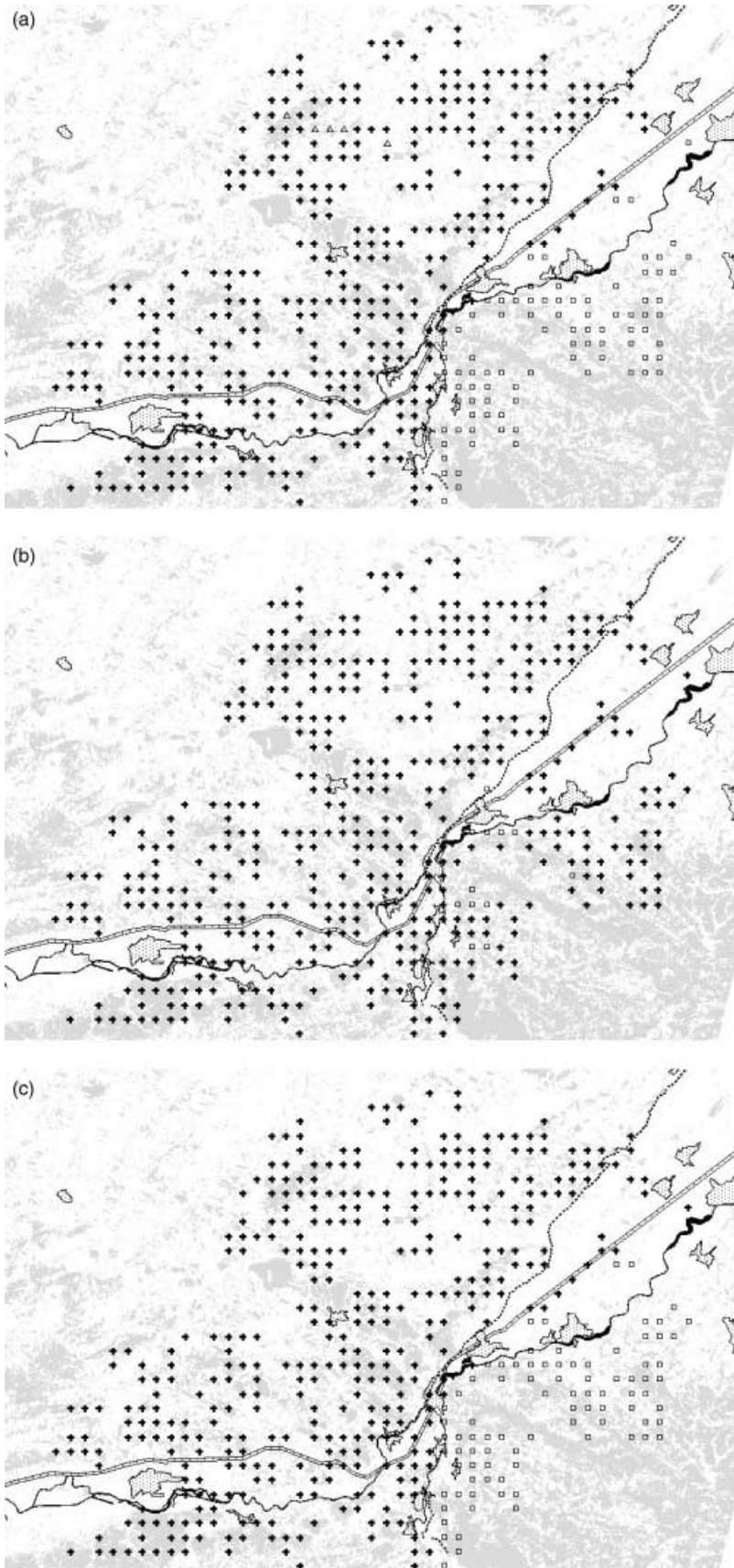


Fig. 3 GENELAND assignment of individuals in the three best runs among the 10 selected ones: individuals assigned to the northwestern population are represented by black crosses, those assigned to the southeastern population by white squares and individuals of the first run assigned to a third population are represented by white triangles. The following seven runs were similar to the third one, except for about 10 individuals of the eastern part of the area situated between the highway and the Garonne River which were alternatively assigned to one or to the other population, depending on the run. (a) GENELAND assignment in the best run, (b) GENELAND assignment in the second run, and (c) GENELAND assignment in the third run.

populations corresponded to areas of the study zone where there was no sampled individual (not shown). The results of the 10 selected runs showed globally good consistency: in eight of them, two populations were inferred, made up of 966 and 183 individuals, respectively (means across the eight runs). The southeastern population was separated from the northwestern population by a boundary zone situated between a major canal and the highway to the north, and including the Garonne River, and by a zone including a tributary of Garonne River and several minor canals to the west (Fig. 3c). Two slight deviations from this pattern were suggested by the first two runs: the first one inferred a third population, comprised of only 29 individuals, situated in the centre of the study area (Fig. 3a). The second one inferred a large main population, with only 29 individuals assigned to a second population in the southeastern part of the study area (Fig. 3b). The three runs with 1 million MCMC iterations gave results consistent with those of the eight similar runs with shorter length.

Standard population genetic analyses on the inferred populations

We ran the standard population genetic analyses on the results of the best of the 100 runs that described a pattern consistent with most of the other nine selected runs (third run). There was no significant linkage disequilibrium in either of the two inferred populations. In the northwestern population F_{IS} was 0.01595 and Hardy–Weinberg disequilibrium was significant ($P = 0.0006$). This disequilibrium was due to four loci (Table 1). In the southeastern population, F_{IS} was 0.01601 and Hardy–Weinberg disequilibrium was not significant ($P = 0.1742$). Pairwise F_{ST} between the populations was 0.008 (95% confidence interval = 0.004–0.012). The AMOVA revealed that most of the genetic variation was situated within populations (99.24%), but the variance among populations was nevertheless significant ($P < 0.0001$).

Discussion

Population structure

When looking for genetic structure of a young roe deer population across a fragmented landscape, we obtained different results using recent Bayesian methods proposed in two different software, STRUCTURE (Pritchard *et al.* 2000) and GENELAND (Guillot *et al.* 2005b). The results generated by STRUCTURE suggest one genetic unit only, while GENELAND found two. The two populations inferred in GENELAND showed very weak but significant differentiation, as indicated by values of F_{ST} and by an analysis of molecular variance. This result is interesting as it suggests that GENELAND can detect even weak spatial genetic structure,

for example in newly founded populations likely to be far from drift-gene flow equilibrium. This may be important for a variety of management and conservation issues, particularly where populations have been modified by recent human-induced changes to the landscape. No other currently available method has been able to satisfactorily deal with these cases.

A general question concerning the discrepancy between the results generated by STRUCTURE and by GENELAND is: how reliable is the structure inferred by the latter? Indeed, as is always the case with Bayesian methods, there is a possibility that the inferred clustering is related to a spurious mode in the MCMC scheme with no real statistical significance. Further, in the present context, in contrast to many scientific methods such as pattern recognition, the object of inference (namely subpopulation membership) is largely a product of our model and there is no way to validate the results using further observations. However, at least three facts support the findings of GENELAND:

- 1 The assumptions made (i.e. higher prior probability given to populations composed of spatially proximal individuals) are rather mild and comply well with field observation, as roe deer movements are restricted in space.
- 2 Many runs were carried out and although some convergence issues occurred for a few of them, ranking the runs according to the posterior density (as suggested in Corander *et al.* 2003 and Guillot *et al.* 2005a) led to a good consistency across the best runs.
- 3 The border between inferred populations coincides very well with the putative barriers to deer movements, while this was in no way included in the algorithm itself.

The main methodological difference between STRUCTURE and GENELAND is that the latter exploits the spatial information concerning the origin of the samples, while the former does not. The fact that, in contrast to STRUCTURE, GENELAND was able to detect the weak population structure provides empirical evidence to support the relevance and the usefulness of this information for inferring genetic structure. However, further studies will be necessary, especially designed for this aim, to determine for which questions and under what conditions the respective software perform well.

The whole population showed a slight but significant heterozygosity deficit, as did the northwestern population inferred by GENELAND. The deficit in the southeastern population was not significant, although the value of F_{IS} was the same as in the northwestern one, probably because of reduced statistical power due to its smaller sample size. This deficit could either be due to the existence of an underlying structure (Wahlund effect), to the presence of null alleles, or to a combination of the two. We ran GENELAND

on the individuals of the northwestern population alone with different hyper-prior parameters (corresponding to finer-scale spatial structure), to look for any further subdivision that could explain the observed deficiency. We found two populations, one of which was composed of only a few individuals (mean number of individuals over 10 runs = 70, not shown) and corresponded to the third population inferred in the best run on the whole sample (Fig. 3a). This area was sampled with a higher intensity than the rest of the study zone (intensive mark–recapture studies); we thus believe this population corresponds to a sample of highly related individuals in a small area, and not to a differentiated genetic unit per se. Furthermore, the heterozygosity deficiency is still significant in the larger northwestern unit. We cannot exclude the presence of null alleles in our data set; however, they are unlikely to be a problem for our analyses as they would generally lead to the inference of additional populations by STRUCTURE (Falush *et al.* 2003), and simulations have shown that GENE-LAND is robust to their presence (Guillot, unpublished). Alternative possible explanations for the heterozygosity deficit include subtle differentiation not detected by GENE-LAND or isolation by distance (see Coulon *et al.* 2004). There is no clustering model currently available that is capable of handling data under the latter scenario.

Another problem, encountered by both methods, was the presence of ghost populations (populations which are modal for none of the analysed individuals). The likely explanation for this phenomenon is that the data depart from the model assumptions (populations at Hardy–Weinberg equilibrium and loci without linkage disequilibrium), which creates spurious classes in the runs with varying K . It is a notoriously difficult problem that we try to address in a pending work. From a practical point of view, the empty populations should be disregarded and conclusions should be based on the nonempty populations only (Guillot *et al.* 2005a).

Absolute barriers or high resistance landscape features?

Our results showed that the southeastern unit was separated to the north from the northwestern unit by a zone including the highway, a major canal and the Garonne River, and to the west by the valley of a tributary of the Garonne River, the Salat River, including several canals and urban areas (Fig. 3). None of these landscape features alone can explain the observed population structure. Indeed, in the eastern part of the study area, the highway and the Garonne River separate well the two populations, but this is not the case in the western part. Furthermore, some (about 10) individuals situated between the highway and the Garonne River are alternatively assigned to one or to the other population, depending on the run. This lack of consistency suggests that these

individuals are first-generation migrants. The fact that the highway does not constitute an impermeable barrier could be due to the presence of three wildlife passages in the western part of the study area and of holes in the fences, which allow some individuals to traverse and thus promote a degree of homogenization of allele frequencies. In the southern part of the study area, the Salat River is situated on the boundary between the two populations. However, we believe it is unlikely to be solely responsible for the inferred structure because of its rather moderate width (around 50 m on average) and shallowness, which should allow roe deer to cross quite easily. In contrast, canals, in combination with a 2-m deer fence which runs parallel to the Salat River for 5 km along the boundary of a hunting reserve, are more likely to limit gene flow along this western boundary.

Hence it seems, from the observed population structure, that none of the landscape features that we hypothesized could act as barriers to roe deer movement (highway, rivers, canals) is in fact totally impermeable. Rather, we suggest that each of them acts as a moderator of gene flow because of a high resistance to roe deer movements, and hence their cumulative effect has led to the differentiation of two genetic units. The present study then adds evidence in support of the assertion that landscape features influence population structure. For example, Gerlach & Musolf (2000) showed that highways trigger population subdivision in bank voles (*Clethrionomys glareolus*), Funk *et al.* (2005) demonstrated that mountain ridges and elevation differences are associated with increased genetic divergence among populations of Columbia spotted frogs (*Rana luteiventris*), and the presence of roads increases the level of genetic differentiation between populations of ground beetle (*Abax parallelepipedus*) (Keller *et al.* 2004). Population genetics can indeed be considered a valuable tool for studying the permeability of landscape features and to thus infer connectivity within a landscape.

The genetic differentiation into two units that we observed may be a result of divergence at either side of the barrier zone or, more likely, the translation of two separate colonization events. Under the latter hypothesis, the northwestern and southeastern populations have separate origins and have recently met at the barrier zone where further progression and admixture is being slowed or prevented by the cumulative effect of the three barrier types. Alternatively, the area was colonized in a single event (from the northwest), slowed by the barrier zone, and subsequently, the southeastern area colonized by a few founder individuals. The founder event would then lead to a slight but significant level of genetic differentiation to either side of the barrier zone. This question remains to be explored by assignment of individuals of the two populations to samples from the potential sources of colonization.

Conclusion

Our results have shown that deer situated at the different sides of barriers such as highways and waterways may not be considered to belong to a single panmictic population. More generally, this means that landscape features that slow or impede movement have a real biological significance for the definition of management units. Up to now, most management plans for roe deer and other game species are based on administrative boundaries, not only because of the technical advantages, but also because of the absence of any empirical information for taking such decisions. We suggest that the combination of genetics, which allow the definition of population units at a rather large spatial scale, with other biological information (e.g. indicators of ecological changes, Cederlund *et al.* 1998) can provide a useful approach for defining ecologically meaningful management units (e.g. Zannèse *et al.*, unpublished data). While it may be impractical to systematically carry out large-scale studies of genetic structure for management purposes, we believe that repeating such an approach for a given species in different landscape contexts should provide valuable information on which landscape features may act as total or partial barriers to animal movement. This information can then be used as reliable and practical guide for managers to define management units over large management territories, taking into account potential barriers and hence reflecting an ecological reality.

Acknowledgements

The Fédération Régionale des Chasseurs de Midi-Pyrénées and the Fédération Départementale des Chasseurs de la Haute-Garonne gave valuable help for sample collection and participated with financial support of A. Coulon during her PhD. This work was partly financed by a grant from the Bureau des Ressources Génétiques. We thank Stéphanie Manel, Ben Sacks and an anonymous referee for giving helpful comments on a previous version of this paper.

References

- Arnold GW, Weeldenburg JR, Ng VM (1995) Factors affecting the distribution and abundance of Western grey kangaroos (*Macropus fuliginosus*) and euros (*M. robustus*) in a fragmented landscape. *Landscape Ecology*, **10** (2), 65–74.
- Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews*, **5**, 251–261.
- Belkhir K, Dawson KJ (2001) A Bayesian approach to the identification of panmictic populations and the assignment of individuals. *Genetical Research*, **78**, 59–77.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, **57**, 289–300.
- Caizergues A, Bernard-Laurent A, Brenot J-F, Ellison L, Rasplus J-Y (2003) Population genetic structure of rock ptarmigan *Lagopus*

- mutus* in Northern and Western Europe. *Molecular Ecology*, **12**, 2267–2274.
- Cargnelutti B, Reby D, Desneux L, Angibault JM, Joachim J, Hewison AJM (2002) Space use by roe deer in a fragmented landscape. Some preliminary results. *Revue d'Ecologie (Terre et Vie)*, **57**, 29–37.
- Cederlund G, Bergqvist J, Kjellander P *et al.* (1998) Managing roe deer and their impact on the environment: maximising benefits and minimising costs. In: *The European Roe Deer: The Biology of Success* (eds Andersen R, Duncan P, Linnell JDC), pp. 189–219. Oslo, Norway.
- Cimmaruta R, Bondanelli P, Nascetti G (2005) Genetic structure and environmental heterogeneity in the European hake (*Merluccius merluccius*). *Molecular Ecology*, **14**, 2577–2591.
- Corander J, Waldmann P, Sillanpää MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics*, **163**, 367–374.
- Coulon A, Cosson J-F, Angibault JM *et al.* (2004) Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual-based approach. *Molecular Ecology*, **13**, 2841–2850.
- Daniilkin A, Hewison AJM (1996) *Behavioural Ecology of Siberian and European Roe Deer*. Wildlife Ecology and Behavioural Series, Chapman & Hall, London.
- Dawson KJ, Belkhir K (2001) A Bayesian approach to the identification of panmictic populations and the assignment of individuals. *Genetical Research*, **78**, 59–77.
- Doums C, Cabrera H, Peeters C (2002) Population genetic structure and male-biased dispersal in the queenless ant *Diacamma cyaneiventris*. *Molecular Ecology*, **11**, 2251–2264.
- Ehrich D, Stenseth NC (2001) Genetic structure of Siberian lemmings (*Lemmus sibiricus*) in a continuous habitat: large patches rather than isolation by distance. *Heredity*, **86**, 716–730.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
- 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.
- Funk CW, Blouin MS, Corn PS *et al.* (2005) Population structure of Columbia spotted frogs (*Rana luteiventris*) is strongly affected by the landscape. *Molecular Ecology*, **14**, 483–496.
- Galan M, Cosson J-F, Aulagnier S *et al.* (2003) Cross-amplification tests of ungulate primers in roe deer (*Capreolus capreolus*) to develop a multiplex panel of 12 microsatellite loci. *Molecular Ecology Notes*, **3**, 142–146.
- Gerlach G, Musolf K (2000) Fragmentation of landscape as a cause for genetic subdivision in bank voles. *Conservation Biology*, **14** (4), 1066–1074.
- Goudet J (2001) *FSTAT, A program to estimate and test gene diversities and fixation indices (version 2.9.3)*. Available from www.unil.ch/izea/software/fstat.html.
- Guillot G, Estoup A, Mortier F, Cosson J-F (2005a) A spatial statistical model for landscape genetics. *Genetics*, **170** (3), 1261–1280.
- Guillot G, Mortier F, Estoup A (2005b) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes*, **5** (3), 708–711.
- Gyllenstrand N, Seppä P (2003) Conservation genetics of the wood ant, *Formica lugubris*, in a fragmented landscape. *Molecular Ecology*, **12**, 2931–2940.
- Hampton JO, Spencer PBS, Alpers DL *et al.* (2004) Molecular techniques, wildlife management and the importance of genetic

- population structure and dispersal: a case study with feral pigs. *Journal of Applied Ecology*, **41**, 735–743.
- Hewison AJM, Vincent JP, Joachim J *et al.* (2001) The effects of woodland fragmentation and human activity on roe deer distribution in agricultural landscapes. *Canadian Journal of Zoology*, **79**, 679–689.
- Hewison AJM, Vincent JP, Joachim J *et al.* (1998) Modelling the effects of woodland fragmentation on roe deer (*Capreolus capreolus*) distribution in agricultural landscapes. *Gibier et Faune Sauvage*, **15** (hors série tome, 2), 323–329.
- Keller I, Nentwig W, Lurgiader CR (2004) Recent habitat fragmentation due to roads can lead to significant genetic differentiation in an abundant flightless ground beetle. *Molecular Ecology*, **13**, 2983–2994.
- Linnell J, Wahlström LK, Gaillard J-M (1998) From birth to independence: birth, growth, neonatal mortality, hiding behaviour and dispersal. In: *The European Roe Deer: The Biology of Success* (eds Andersen R, Duncan P, Linnell JDC), pp. 257–283. Scandinavian University Press, Oslo, Norway.
- Luiselli L, Capizzi D (1997) Influences of area, isolation and habitat features on distribution of snakes in Mediterranean fragmented woodlands. *Biodiversity and Conservation*, **6**, 1339–1351.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18** (4), 189–197.
- Mönkkönen M, Reunanen P, Nikula A, Wydeven AP (1995) A regional landscape analysis and prediction of favorable gray wolf habitat in the northern Great Lakes region. *Conservation Biology*, **9**, 279–294.
- Moran MD (2003) Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos*, **100** (2), 403–404.
- Moritz C (1994) Defining 'Evolutionarily Significant Units' for conservation. *Trends in Evolution & Ecology*, **9**, 373–375.
- O'Reilly PTO, Canino MF, Bailey KM, Bentzen P (2004) Inverse relationship between F_{ST} and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chacoogramma*): implications for resolving weak population structure. *Molecular Ecology*, **13**, 1799–1814.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Pritchard JK, Wen W (2003) *Documentation for STRUCTURE software: version 2*. Available from: <http://pritch.bsd.uchicago.edu>.
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumincism. *Journal of Heredity*, **86**, 248–249.
- Rueness EK, Jorde PE, Hellborg L *et al.* (2003) Cryptic population structure in a large, mobile mammalian predator: the Scandinavian lynx. *Molecular Ecology*, **12**, 2623–2633.
- San José C, Lovari S (1998) Ranging movements of female roe deer: do home-loving does roam to mate? *Ethology*, **104**, 721–728.
- Schneider S, Roessli D, Excoffier L (2000) *ARLEQUIN version 2.000: A software for population genetic data analysis*. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland.
- Schulte-Hostedde AI, Gibbs LH, Millar JS (2001) Microgeographic genetic structure in the yellow-pine chipmunk (*Tamias amoenus*). *Molecular Ecology*, **10**, 1625–1631.
- Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society. Series B*, **64** (3), 479–498.
- Taylor PD, Fahrig L, Henein K, Merriam G (1993) Connectivity is a vital element of landscape structure. *Oikos*, **68** (3), 571–573.
- Verhoeven KJF, Simonsen KL, McIntyre LM (2005) Implementing false discovery rate control: increasing your power. *Oikos*, **108**, 643–647.
- Wahlström LK (1994) The significance of male-male aggression for yearling dispersal in roe deer (*Capreolus capreolus*). *Behavioural Ecology and Sociobiology*, **33**, 409–412.
- Wahlström LK, Liberg O (1995) Contrasting dispersal patterns in two Scandinavian roe deer *Capreolus capreolus* populations. *Wildlife Biology*, **1** (3), 159–164.
- Wang M, Schreiber A (2001) The impact of habitat fragmentation and social structure on the population genetics of roe deer (*Capreolus capreolus* L.) in Central Europe. *Heredity*, **86**, 703–715.
- Weir BS, Cockerham CC (1984) Estimating F -statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

A. Coulon is a PhD student working on dispersal behaviour in a landscape ecology (and particularly landscape genetics) framework. G. Guillot is a statistician. His interest lies in Bayesian modelling and computational methods for environmental sciences. J. F. Cosson is mainly interested in population genetics of fluctuating populations in relation with landscape ecology. J. M. Angibault works on catching, marking and monitoring techniques. S. Aulagnier's main interests are in conservation biology. B. Cargnelutti works on habitat use and data manipulation, and M. Galan on molecular biology techniques. M. Hewison's research focuses on population biology and behavioural ecology of ungulates.
