Available online at www.sciencedirect.com
ScienceDirect
Livestock Science 115 (2008) 80–88
LIVESTOCK SCIENCE
www.elsevier.com/locate/livsci

Relationship between genealogical and microsatellite information characterizing losses of genetic variability: Empirical evidence from the rare Xalda sheep breed

I. Álvarez [a], L.J. Royo [a], J.P. Gutiérrez [b], I. Fernández [a], J.J. Arranz [c], F. Goyache [a,*]

[a] SERIDA-Somió, C/ Camino de los Claveles 604, E-33203 Gijón, Asturias, Spain
[b] Departamento de Producción Animal, Facultad de Veterinaria, Avda. Puerta de Hierro s/n, E-28040-Madrid, Spain
[c] Departamento de Producción Animal, Universidad de León, E-24071 León, Spain

Received 5 December 2006; received in revised form 20 May 2007; accepted 19 June 2007

Abstract

Preservation of rare genetic stocks requires continual monitoring of populations to avoid losses of genetic variability. Genetic variability can be described using genealogical and molecular parameters characterizing variation in allelic frequencies over time and providing interesting information on differentiation that occurred after the foundation of a conservation program. Here we analyze the pedigree of the rare Xalda sheep breed (1851 individuals) and the polymorphism of 14 microsatellites in 239 Xalda individuals. Individuals were assigned to a base population (BP) or 4 different cohorts (from C1 to C4) according to their pedigree information. Genetic parameters were computed at a genealogical and molecular level, namely inbreeding (F), observed (H_o) and expected (H_e) heterozygosity, individual coancestry coefficients (f and f_m), average relatedness (AR), mean molecular kinship (Mk), average number of allele per locus (A), effective number of ancestors (f_a), effective population size (N_e and $N_{e(m)}$) and founder genome equivalents (N_g and $N_{g(m)}$). In general, the computed parameters increased with pedigree depth from BP to C4, especially for the genealogical information and molecular coancestry-based parameters (f_m, Mk and $N_{g(m)}$). However, H_o and H_e showed the highest values for C1 and the molecular heterozygote deficiency within population ($F_{IS(m)}$) showed the lowest value for C1, thus indicating that loss of genetic variability occurs very soon after the implementation of conservation strategies. Although no genealogical or molecular parameters are sufficient by themselves for monitoring populations at the beginning of a conservation program, our data suggests that coancestry-based parameters may be better criteria than those of inbreeding or homozygosity because of the rapid and strong correlation established between f and $f_{(m)}$. However, the obtaining of molecular information in well-established conservation programs could not be justified, at least in economic terms.

Keywords: Genetic variability; Inbreeding; Homozygosity; Coancestry; Conservation program; Pedigree depth

1. Introduction

The breeding goal in small populations is the conservation of genetic diversity. Genetic variability can be described at the genealogical level using parameters like inbreeding or kinship coefficients (Caballero and

* Corresponding author. Tel.: +34 985195303; fax: +34 985195310.
E-mail address: foyache@serida.org (F. Goyache).

1871-1413/$ - see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.livsci.2007.06.009

Toro, 2000). At the molecular level, highly variable loci provide a large amount of information on individual genotypes that is useful for clarifying population structure (Balloux and Lugon-Moulin, 2002; Álvarez et al., 2004). Even though genealogical analyses assume that each individual in the base population has unique alleles on each locus and molecular data are referred to finite sampling (thus being more pronouncedly affected by time forces), both sources of information are based on similar assumptions (Rochambeau et al., 2001): criteria based on pedigree information refer to any neutral autosomal locus, while criteria based on observed genetic polymorphisms mirror phenomena related to neutral genes or non-coding regions.

Allelic frequencies vary over time as populations are of finite size. The degree of temporal genetic differentiation can be assessed over time, thus providing interesting information on the differentiation that occurred after founder events or bottlenecks (Hansson et al., 2000). Preservation of rare genetic stocks requires continual monitoring of populations (Caballero and Toro, 2000), especially when pedigree information is shallow and the management structure leads to losses of genetic variability by drift or unobserved selection (Goyache et al., 2003). When the available sample size is small and genealogies are scarce, the combined use of pedigree information and molecular markers might be recommendable. Recent studies have formalized the computation of coancestry coefficients from molecular information (Caballero and Toro, 2000, 2002; Eding and Meuwissen, 2001).With these findings researchers can use theoretically equivalent parameters at the genealogical and molecular level to monitor small populations (Toro et al., 2002, 2003; Royo et al., 2007).

This study aims to test the interest of combining genealogical and molecular information to monitor genetic variability in small populations. We shall test this approach on real data from the rare Xalda sheep breed of Asturias (Álvarez Sevilla et al., 2004; Goyache et al., 2003), which has recently undergone a conservation program, providing empirical evidence of interrelationship between molecular and genealogical estimators of genetic diversity.

2. Materials and methods

2.1. Data and sampling

We obtained the information registered in the Xalda herd book, which includes a total of 1851 animals (217 males), of the breeders' association (ACOXA). Up to 1505 individuals (130 males) were alive at the moment of carry out sampling and 1152 animals (107 males) were considered reproductive individuals (older than 1 year). The Xalda flockbook includes a large number of very small sized flocks. Although some of these flocks have a short duration and do not remain within ACOXA for long the number of actives flocks per year is usually ranges between 50 and 60. A detailed description of the main characteristics of the Xalda pedigree can be found in Goyache et al. (2003). Using the program ENDOG v3.2 (Gutiérrez and Goyache, 2005), we computed the equivalent complete generations (g_e) for each individual in the pedigree. The parameter g_e is computed as the sum over all known ancestors of the term of $(1/2)^n$, where n is the number of generations separating the individual from each known ancestor (Boichard et al., 1997). Notice that, on average for a given reference population, g_e is equivalent to the 'discrete generation equivalents' (T) proposed by Woolliams and Mäntysaari (1995), thus characterizing the amount of pedigree information in data sets with overlapping generations. This parameter was used to assign the Xalda individuals to a discrete cohort by rounding the individual g_e value in order to obtain cohorts with average pedigree depth of 0 (individuals with no known parents or base population), 1, 2, 3 and 4 g_e's. The identified cohorts and total number of individuals per generation (in brackets) were: Base population (BP; 325), Cohort 1 (C1; 607), Cohort 2 (C2; 521), Cohort 3 (C3; 336) and Cohort 4 (C4; 62) (see Table 1).

Blood samples were randomly obtained from a total of 303 reproductive (or selected for reproduction) Xalda individuals, kept in a total of 58 different herds, and assigned to their corresponding cohorts. For further analyses, the following editing rules were applied: a) no full sibs were allowed; and b) within a cohort, no half sibs sampled in the same herd were allowed. After editing, the number of available samples for the BP, C1, C2, C3 and C4 were respectively 72, 49, 32, 50 and 36, altogether totalling 239 samples (38 from male individuals).

Individuals sampled for the BP are widely represented in the other cohorts, having links with 25.7%, 63.9%, 74.1% and 66.1% of the individuals in the pedigree assigned to, respectively, C1, C2, C3 and C4.

2.2. Genealogical analyses

Genealogical information was analyzed using the program ENDOG v3.2 (Gutiérrez and Goyache, 2005). Pedigree data were analyzed in order to obtain the full coancestry (f; Malécot, 1948) matrix of the Xalda pedigree, the individual coefficient of inbreeding (F)

Table 1
Description of the available data and sampling, and average genetic parameters per cohort and for the whole analyzed population

Parameter	Base population	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Whole population
Number of individuals						
In pedigree	325	607	521	336	62	1851
Genotyped	72	49	32	50	36	239
Sampled herds	25	29	24	24	16	58
Equivalent generations(g_e)						
Whole pedigree	0.0	1.1	2.0	3.0	3.8	1.6
Genotyped individuals [a]	0.0	1.0	2.0	3.1	3.7	1.7
Genealogical information						
Inbreeding (F)	–	0.011	0.023	0.061	0.141	0.026
Average relatedness (AR)	–	0.024	0.033	0.043	0.048	0.027
Coancestry (f)	–	0.013	0.022	0.040	0.066	0.013
F_{IS} [b]	–	−0.003	0.001	0.022	0.081	0.006
Effective number of ancestors (f_a)	–	53	35	20	15	31 [c]
Effective population size (N_e)	–	46.0	21.7	8.1	3.5	19.6
Founder genome equivalents (N_g)	–	39.9	23.0	12.6	7.6	37.8
Molecular information						
Average number of alleles per locus ($A_{(60)}$ [d])	8.1	7.7	7.3	7.6	7.1	7.6
Observed heterozygosity (H_o)	0.644	0.657	0.594	0.594	0.596	0.617
Expected heterozygosity (H_e)	0.728	0.736	0.716	0.711	0.696	0.696
Molecular coancestry $f_{(m)}$	0.278	0.322	0.349	0.347	0.366	0.324
Mean kinship (Mk)	0.282	0.315	0.317	0.318	0.321	0.304
$F_{IS(m)}$ [a]	0.108	0.032	0.090	0.090	0.061	0.080
Effective population size ($N_{e(m)}$)	–	−32.8	6.6	6.4	6.9	11.8
Founder genome equivalents ($N_{g(m)}$)	–	8.1	5.1	5.3	4.1	7.8
Expected $f_{(m)}$ in the base population ($E(f_{(m)0})$)	–	0.314	0.335	0.320	0.321	0.315

[a] The S.E. of the average g_e corresponding to C1, C2, C3 and C4 were, respectively, 0.07, 0.04, 0.04 and 0.04.
[b] Defined as heterozygote deficiency within a population.
[c] The effective number of ancestors (f_a) was computed using each cohort as the reference population, except for the whole population, for which the reference population included all the individuals in the pedigree with both parents known, as implemented by default in the program ENDOG (Gutiérrez and Goyache, 2005).
[d] Average number of alleles per loci rarefacted (Hurlbert, 1971) for 60 copies.

(Malécot, 1948), defined as the probability that an individual has two identical alleles by descent, and the individual average relatedness coefficient (AR) (Goyache et al., 2003; Gutiérrez et al., 2003), defined as the probability that an allele randomly chosen from the whole population in the pedigree belongs to the animal. The three parameters were averaged per cohort.

Three additional parameters were computed for each cohort and for the whole pedigree to characterize losses of genetic variability across cohorts: a) the effective population size (N_e), defined as the number of breeding animals that would lead to the actual increase in inbreeding if they contributed equally to the next generation, was computed here as $N_e = \frac{1}{2\Delta F}$, ΔF being here the average F for the tth cohort as realized inbreeding; b) the founder genome equivalents (N_g; Lacy, 1989, 1995), defined as the overall founder representation in a managed population accounting for the loss of genetic variability from unequal founder and non-founder contributions, was obtained as $N_g = \frac{1}{2\Delta f}$ (Lacy, 1995; Caballero and Toro, 2000), Δf being the average f for each cohort as realized coancestry. Parameter N_g is usually computed using complex probability calculations or Monte Carlo pedigree simulations (Lacy, 1995). The computation of N_g from the additive relationship matrix (Caballero and Toro, 2000), has the practical advantage of avoiding the need for gene dropping methodologies, thus reducing the computational effort; and c) the effective number of ancestors (f_a)

(Boichard et al., 1997), which accounts for the losses of genetic variability produced by the unbalanced use of reproductive individuals also taking into account bottlenecks in the pedigree and defined as the number of equally contributing ancestors that would be expected to produce the same genetic diversity as in the population under study. The parameter f_a was computed as $f_a = \frac{1}{\sum_{j=1}^{a} q_j^2}$, where q_j is the marginal contribution of an ancestor j; in other words, the genetic contribution made by an ancestor that is not explained by other ancestors chosen before.

From the coancestry information, Wright's (1969) F-statistics, F_{IT}, F_{ST}, and F_{IS} (defined, respectively, as heterozygote deficiency in the total population, heterozygote deficiency due to population subdivision and heterozygote deficiency within population) were obtained as $F_{IT} = \frac{\tilde{F}-\tilde{f}}{1-\tilde{f}}$, $F_{ST} = \frac{\tilde{f}-\bar{f}}{1-\bar{f}}$, and $F_{IS} = \frac{\tilde{F}-\bar{f}}{1-\bar{f}}$ (Caballero and Toro, 2000, 2002), where $\tilde{f}$, $\tilde{F}$ are respectively the mean coancestry and the inbreeding coefficient for the entire population, and $\bar{f}$ the average coancestry for each defined subpopulation (here BP, C1, C2, C3 and C4). The F_{IS} statistic computed from genealogies is equivalent to the parameter α (Caballero and Toro, 2000; Toro et al., 2000), which means the departure from random mating as a deviation from Hardy–Weinberg proportions. Positive genealogical F_{IS} values mean that the average F within a population exceeds between-individuals coancestry (Gutiérrez et al., 2005b) and occurs if matings between relatives are not avoided (Toro et al., 2000).

2.3. Molecular analyses

Total DNA was isolated from blood samples following standard procedures (Sambrook et al., 1989). A set of 14 microsatellites (BM8125, BM6526, CP34, BM757, INRA006, BM6506, BM1818, FCB128, CSSM31, CSSM66, ILSTS011, McM53, RM006, ILSTS005) previously used in Álvarez et al. (2004, 2005, 2007) were analyzed in all individuals. The PCR products were labelled using a fluorescent method (Cy5 labelled primer) and genotyping was performed on an ALFexpressII automated sequencer (Amersham Biosciences, Barcelona).

Molecular information was analyzed using the program MolKin v2.0 (Gutiérrez et al., 2005a). The following parameters were computed from microsatellite information at the cohort and the whole population level: observed (H_o) and expected (H_e) heterozygosity and number of alleles per locus (A) corrected using Hurlbert's rarefaction method (1971) as $A[g] = \sum_i \left[1 - \prod_{k=0}^{g-1} \frac{N-N_i-k}{N-k}\right]$, where g is the specified sampled size, N the number of gene copies examined in a given locus ($N>g$), and N_i the number of occurrences of the ith allele among the N sampled gene copies to account for sample size. Here, g was fitted to 60, which is twice the minimum number of individuals within a cohort with genotype known for all the microsatellites.

Additionally, the molecular coancestry matrix between the 239 genotyped individuals was computed. The molecular coancestry ($f_{(m)}$) 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 can be computed using the following scoring rules (Caballero and Toro, 2002; Eding and Meuwissen, 2001): $f_{(m)ij,1} = \frac{1}{4}[I_{11} + I_{12} + I_{21} + I_{22}]$, where I_{xy} is 1 when allele x on locus l in individual i and allele y in 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 ($f_{(m)ij}$) can be obtained by simply averaging over L analyzed loci as $f_{(m)ij} = \frac{\sum_{l=1}^{L} f_{ij,l}}{L}$. This parameter was further averaged within and between cohorts and for the whole population. The parameter $f_{(m)}$ has 'nice' theoretical properties such as that of the coincidence of its value in the founder population with the expected homozygosity ($1-H_e$; Nei, 1972, 1987). In fact, molecular coancestry can be computed from allelic frequencies as $f_{ij} = \Sigma_{ij} x_{ij} y_{ij}$ (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', further pointing out that it is equal to the Malécot's (1948) coefficient of coancestry if there is no selection and each allele is derived from a single mutation in an ancestral generation. Nei (1972, 1987) showed that its value in the founder population would coincide with that of the expected homozygosity.

The molecular effective population size ($N_{e(m)}$) and the molecular founder genome equivalents ($N_{g(m)}$) were computed from molecular information in a similar way to that detailed for N_e and N_g using genealogical information (Caballero and Toro, 2000, 2002) as $N_e = \frac{1}{2\Delta F_e}$ and $N_g = \frac{1}{2\Delta f_{(m)}}$ respectively, the increments of expected homozygosity ($F_e = 1-H_e$) and $f_{(m)}$ being computed on the corresponding values for the BP as $\Delta F_e = \frac{F_{e(t)} - F_{e(0)}}{1-F_{e(0)}}$ and $\Delta f_{(m)} = \frac{f_{(m)t} - f_{(m)0}}{1-f_{(m)0}}$, where $F_{e(t)}$ and $f_{(m)t}$ are the average molecular expected homozygosity ($1-H_e$) and molecular coancestry at the tth cohort and $F_{e(0)}$ and $f_{(m)0}$ are the average expected homozygosity and molecular coancestry in BP, respectively. Additionally, molecular mean kinship (Mk; Gutiérrez et al.,

Table 2
Pearson correlation coefficients between genealogical and molecular coancestry values computed within and between cohorts

	Cohort 1	Cohort 2	Cohort 3	Cohort 4
Cohort 1	0.713			
Cohort 2	0.226	0.730		
Cohort 3	0.136	0.239	0.670	
Cohort 4	0.181	0.316	0.411	0.710

All the correlation coefficients were significant for $p<0.0001$.

2005a) was computed as the average molecular coancestry of each individual with the rest of the population, and further averaged for each studied cohort. The lower Mk value the lower representation of a genotype in the population.

From the above parameters, Wright's (1969) *F*-statistics, $F_{IT(m)}$, $F_{ST(m)}$, and $F_{IS(m)}$ were computed from molecular information in the way detailed above for genealogical information, though using the molecular values of $\tilde{f}$, $\tilde{F}$ and $\bar{f}$ (noted as $\tilde{f}_{(m)}$, $\tilde{F}_o$ and $\bar{F}_{(m)}$).

The expected molecular coancestry value in the base population (Ef_0) was computed using each cohort as reference population and following Toro et al. (2003) as $Ef_{(m)0}=(f_{(m)t}-f_t)/(1-f_t)$, where $f_{(m)t}$ and f_t are respectively the molecular and the genealogical coancestry for the generation *t*th cohort.

2.4. Correlations

Finally, the Pearson correlation coefficients between genealogical and molecular coancestry values were computed within- and between-cohorts using the program SAS/STAT™ (1999).

3. Results

Average values of the parameters characterizing the genetic variability of the Xalda population assessed from both genealogical and molecular information using each defined cohort as reference population are given in Table 1. The S.E. of the average g_e corresponding to the sampled individuals varied from 0.04 to 0.07 thus being confident in assignment of individuals to their cohorts. The computed parameters increased, in general, with pedigree depth from the BP to C4, especially for the genealogical parameters. From the molecular information, Mk and f_m showed a similar increasing pattern with pedigree depth, though this increase is very slight after C1. However, H_o and H_e showed the highest values for C1 whilst $F_{IS(m)}$ showed the lowest value for C1.

The parameter F_{IS} computed from genealogical information showed values near 0 (0.6% for the whole population) except for C3 and C4, for which it reached values of 2.2% and 8.1%, respectively. The parameter $F_{IS(m)}$ was 8% for the whole genotyped population, presenting the highest values in the BP (10.8%) and the lowest for C1 (3.2%).

N_e and $N_{e(m)}$ do not follow a comparable pattern: N_e decreases with pedigree depth following the increase observed for *F*, whilst $N_{e(m)}$ presents a negative value for C1 due to the decrease in expected homozygosity assessed from BP to C1 (0.272 and 0.264, respectively) to be then found to be roughly constant from C2 to C4. The values obtained for f_a, N_g and $N_{g(m)}$ decreased with pedigree depth, though the decrease for $N_{g(m)}$ is less pronounced (from 8.1 for C1 to 4.1 for C4) than that observed for N_g (from 39.9 to 7.6).

After rarefaction, the average number of alleles per locus varied from 8.1 for the BP to 7.1 for C4. No clear

Table 3
Between-cohorts coancestry (f_{ij}) and F_{ST} matrices obtained from genealogical and molecular information

Cohorts	Genealogical information				Molecular information			
	Base population	Cohort 1	Cohort 2	Cohort 3	Base population	Cohort 1	Cohort 2	Cohort 3
f					$f_{(m)}$			
Cohort 1	0.001				0.280			
Cohort 2	0.001	0.013			0.288	0.316		
Cohort 3	0.001	0.016	0.027		0.289	0.313	0.332	
Cohort 4	0.001	0.018	0.029	0.043	0.289	0.313	0.336	0.340
F_{ST}					$F_{ST(m)}$			
Cohort 1	0.003				0.014			
Cohort 2	0.005	0.002			0.015	0.014		
Cohort 3	0.010	0.005	0.002		0.016	0.016	0.011	
Cohort 4	0.009	0.004	0.003	0.003	0.021	0.022	0.016	0.012

pattern of allele losses with pedigree depth was obtained. With respect to the BP, however, the loss of alleles varied from 4.50% for C1 to 11.28% for C4.

The expected value of molecular coancestry in the base population was computed using each of the analyzed cohorts as the reference population. The obtained values are always higher than the 'true' $f_{(m)}$ value of the base population (0.278), varying from 0.314 to 0.335 for C1 and C2, respectively.

The within- and between-cohorts Pearson correlations between f and $f_{(m)}$ are given in Table 2. The within-cohort Pearson correlation coefficients were similar across cohorts (roughly 0.7). Between-cohorts Pearson correlation coefficients were similar for each pair and increased with pedigree depth until reaching a value of 0.411 for the pair C3–C4.

Between-cohorts gene flow (assessed through coancestry) and differentiation (assessed *via* F_{ST}) were computed from genealogical and molecular information and are given in Table 3. The BP had similar genetic representation across cohorts. This pattern also appeared for the other cohorts and the gene flow from a cohort to the subsequent ones is substantially the same for both f and $f_{(m)}$. The F_{ST} and $F_{ST(m)}$ values obtained for the whole population were 0.007 and 0.026, thus highlighting slight overall population differentiation. Between-cohorts genealogical $F_{ST(m)}$'s were 0.5% or lower, except for the pairs formed by BP and each C3 and C4, which reached a value of roughly 1%. A similar pattern was obtained using molecular F_{ST}'s; the highest differentiations were found between each C3 and C4 and BP (higher than 2).

4. Discussion

At the beginning of *in situ* conservation programs of livestock breeds, the homogenization of type characteristics and the small number of available breeding individuals leads to losses of genetic variability. In this study, we analyzed real data from a small, rare population (the Xalda sheep breed) undergoing a program for preservation of its genetic variability in order to ascertain the importance of genealogical and molecular information on the population's monitoring during the early stages of a conservation program. In contrast with other papers devoted to this task (Toro el al., 2002), our sampling included the base or founder population, thus allowing reliable information to be obtained to achieve our goal. In this respect, two major aspects may be highlighted: a) losses of genetic variability are likely to occur very soon after the setting up of a conservation program, probably because a significant number of founders do not give progeny to the following generations; and b) coancestry can be a better criterion to monitor losses of genetic variability than other classical parameters (Lacy, 1995; Caballero and Toro, 2000) such as genealogical inbreeding and its molecular counterpart, expected homozygosity, at the beginning of conservation programs.

The number of microsatellites genotyped here is limited and results should, therefore, be interpreted with caution. Simulation studies (Eding and Meuwissen, 2001; Baumung and Sölkner, 2003) have shown that a high number of loci are needed to necessary to obtain reliable estimates for the relatedness of individuals or the detection of highly inbred (autozygous) animals. However, the informativeness of a microsatellite set depends on other conditions such as the degree of polymorphism (Baumung and Sölkner, 2003; Fernández et al., 2005) and the structure of the analysed population (Carothers et al., 2006; Fernández et al., 2005). The level of polymorphism of our loci is acceptable with an actual average number of alleles per locus of 10.5 (7.6 after rarefacted for 60 copies). Moreover, Carothers et al. (2006), in humans, using genotypes from 410 microsatellite markers and from 10,000 SNPs were not able to accurately estimate individual inbreeding coefficients from molecular information. These authors consider that likely reasons explaining this situation, were that the number of samples used (50) was not large enough to provide precise estimates of allele frequencies, that sampling was not drawn from a single population with a common history and the size of the population used as reference was too was small and came from different (sub)populations to those being studied. Our research is carried out on samples from a single population, including a wide representation of founders population, with individuals sharing a common history allowing the establishment of close relationships between genealogies and molecular variation (characterized by f and $f_{(m)}$) which are needed to obtain reliable assessments of the variation of the allelic frequencies over time (Fernández et al., 2005).

When pedigree is shallow (as in a scenario like that analyzed here) or incomplete, genealogical information is not always informative as regards the losses of genetic variability (Goyache et al., 2003) and molecular data can perform better. In this study, we found that the average number of alleles per locus decreases at least 4% (from 8.1 to 7.7; see Table 1) in the first generation after founders. However, we also found, with respect to the BP, that both H_o and H_e increase, with a large decrease in the F_{IS} value, in the individuals with one generation of ancestors in their pedigree (C1; see also Table 1). This is

related to the fact that the BP not only has high values for both H_o and H_e (consistent with its high value for A), but also the highest value for molecular $F_{IS(m)}$. The Xalda breed has suffered an intense population bottleneck during the second half of the 20th century; the number of pure ewes at the beginning of the 80's was lower than 800 (Álvarez Sevilla et al., 2004). At the beginning of the 90's the Xalda founders, analyzed here as the BP, were recovered at different genetically-isolated locations (Goyache et al., 2003). The BP individuals have indubitably suffered local genetic bottlenecks probably leading to the fixation of different alleles which, as a whole, capture the genetic variability existing in the breed before the intense bottleneck suffered by the Xalda population prior to the implementation of its conservation program (Álvarez et al., 2004, 2005, 2007). This is probably a common genetic scenario in populations for which a program for the preservation of their genetic variability is being designed and should be taken into account to carefully conserve the genetic background of the founder individuals. In the wild, Tempelton and Read (1994) reported similar behaviour than that we found in the Xalda sheep. For these authors this can be expected in finite populations with separate sexes because of random differences in allele frequency between sexes.

As shown in Table 2, the within-cohort Pearson correlation coefficients between $f_{(m)}$ and f are high and comparable to those reported by Toro et al. (2002) for real datasets formed by populations with deep pedigrees. In our study, this occurs in a population with a very shallow pedigree. This was not expected from studies using simulated datasets, including a base population, in which correlation coefficients are lower (Toro et al., 2002). In simulation studies, marker alleles were assigned at random in the base generation and thus there is no direct relationship between f and $f_{(m)}$, as the $f_{(m)}$ values tended to be lower than those of genealogical coancestry, even when a large number of microsatellites were used (Fernández et al., 2005). As the number of generations increased, 'real' relationships between f and $f_{(m)}$ are established in the simulated datasets and therefore correlation between them increases. In real conditions, however, the number of alleles per locus in the BP is probably lower than that implemented in simulated datasets (Balloux and Lugon-Moulin, 2002) and from our study it is realistic to think that the establishment of the relationships between f and $f_{(m)}$ occur very soon, when most rare alleles carried out by the BP are lost in C1. Previous non-accounted selection (at least for the adaptation to the particular environmental characteristics of the Xalda spreading area) may underlie this rapid, strong relationship established between f and $f_{(m)}$ (Bataillon et al., 1996). The within-cohort correlation coefficients tend to be moderate when there is sufficient pedigree available (in C3 and C4), thus supporting the rapid establishment of relationships between f and $f_{(m)}$.

The rapid loss of genetic variability suffered by our population, highlighted by the decrease in A and the increase in $f_{(m)}$ values, underlies the overestimation of the $E(f_{(m)0})$, computed following Toro et al. (2003) with respect to the $f_{(m)0}$ values. Genealogical coancestry assesses the probability of an individual being identical by descent to another, assuming that all the alleles in a given reference founder population are different. The parameter $E(f_{(m)0})$ assumes that the present-sampled population considered in our genealogical study has the same allele frequencies as the founder population and also that these frequencies were in both Hardy–Weinberg and linkage equilibrium (Toro et al., 2002, 2003). As highlighted by the present study, these assumptions are not realistic. We cannot actually expect present populations to have the same allele frequencies as founder populations, since drift or selective processes occurred during the formation of the breeding stock to be conserved (Toro et al., 2000, 2002; Goyache et al., 2003). Moreover, it is not realistic to consider a number of independent allelic forms in the base population (virtually infinite, depending on the founder population size), as is assumed in genealogical analyses when, for instance, perfect alleles in dinucleotide microsatellites rarely exceed 30 repeats (Balloux and Lougon-Moulin, 2002). Thus, the assumption of a straightforward relationship between the $f_{(m)}$ value in the current population and in the founder population through f (Toro et al., 2003) is unclear; at least in the practical framework presented in this paper. Notice that, in our study, $E(f_{(m)0})$ overestimates $f_{(m)0}$, thus underestimating the actual genetic variability maintained by our founders, but that $f_{(m)0}$ and the expected homozygosity ($1-H_e$) in the BP are roughly the same (0.272 and 0.278, respectively). This straightforward relationship between $f_{(m)0}$ and expected homozygosity in the founder population is consistent with the direct relationship existing at a genealogical level between the expected heterozygosity and coancestry (see formula 26 in Caballero and Toro, 2000), thus directly relating to the effective population size.

From our study, the use of both genealogical and molecular coancestry information to monitor genetic variability in small populations undergoing conservation programs can be justified more than the use of other parameters such as inbreeding and homozygosity. As shown by Caballero and Toro (2000), the parameter N_g represents a compound of contributions from the founders (the effective number of founders, N_{ef}), and from all other

individuals in the genealogy. Since the expected contribution from a particular ancestor to its descendants will become constant after a few generations in a pedigree, the use of N_g allows the genetic drift occurring every generation in the pedigree to be accounted for. Within a practical framework such as the one presented here, genealogies are usually shallow and thus less useful and H_e cannot characterize the 'real' variability in our population. Moreover, when computing $N_{e(m)}$ for C1, we would obtain a negative, non-realistic value of −32.8. If, as usual, our population had been sampled across generations for molecular studies, the genetic variability obtained would be overestimated by the excess of heterozygotes produced by the crosses between founder individuals that have experienced previous particular genetic bottlenecks, as occurs in our C1. Genealogical and especially molecular coancestry behave in a more predictable way; they both tend to increase with pedigree depth, though the increase of $f_{(m)}$ is less pronounced than that of f, and N_g and $N_{g(m)}$ tended to converge in C4. However, $f_{(m)}$ increases regardless of whether the expected homozygosity decreases, as reported from the BP to C1, thus better characterizing the loss of genetic variability resulting from the BP, and also highlighted by the parameter A.

Both f and $f_{(m)}$ can be used, in combination with other parameters such as F_{IS} (Toro et al., 2000), AR (Goyache et al., 2003; Gutiérrez et al., 2005b) or Mk (Gutiérrez et al., 2005a), to ascertain whether the genetic representation of a cohort in the subsequent ones is balanced. Table 3 shows how the gene flow from the base population to the others is similar for both f and $f_{(m)}$. However, differentiation seems to occur rapidly and C3 and C4 present F_{ST} values of roughly 1% and 2%, respectively, for genealogical and molecular information with the BP, probably indicating unaccounted losses of alleles and increases of within-cohort coancestry. The parameter AR may be used as an index to maintain the initial genetic stock as well as to predict the long-term inbreeding of a population, thus suggesting modifications to management practice for conserving the genetic makeup of a population (Gutiérrez et al., 2003; Goyache et al., 2003; Gutiérrez et al., 2005b). AR can be computed for each non-founder individual so as to know the genetic representation of this individual on the population. Moreover, the average value of AR in a random mating population for a given generation would be twice the F of the following generation. The parameter Mk is the molecular counterpart of AR and shares its properties for managing small populations. In our study, although F greatly exceeds half AR in C3 and C4, at a molecular level the observed homozygosity ($1-H_o$) only exceeds Mk for C2. An unbalanced contribution of the individuals to the genetic background of the population occurs and mating policy should be modified to ensure the maintenance of the genetic background from the founder population. The parameter F_{IS} can increase the information on the development of the population, characterizing the mating policy. We have shown that positive values appear soon ($F_{IS}=0.002$ for C3; see Table 1) after the foundation of pedigrees; this is probably due to the selection for type traits performed by Xalda breeders (Goyache et al., 2003). Positive values of the parameter also indicate that most genetic variability in our breed is between- and not within-individuals (Toro et al., 2000). Moreover, the genealogical parameter F_{IS} may be a better indicator than the former in the early stages of a conservation program for monitoring genetic variability.

5. Conclusions

Here we provide empirical evidence on genetic variation over time in real populations using data from the rare Xalda sheep breed as an example. Neither genealogical nor molecular information by themselves are sufficient for monitoring small populations when the pedigree is shallow. A depletion of genetic variability occurs very soon after the commencement of conservation program and each available parameter offers partial information for monitoring populations. In practical terms, no clear relationship between $f_{(m)}$ in the current population and in the founder population through f could be establish. The correlation between f and $f_{(m)}$ increases rapidly with pedigree depth and obtaining molecular information in well-established conservation programs may not be justified, at least in economic terms further than paternity tests.

Acknowledgments

This work was partially funded by grants from MEC-INIA, No. RZ2004-00007-C02 and from the Regional Government of the *Principado de Asturias* No. PC-REC04-27. The authors are indebted to the Xalda Breeders Association (ACOXA; http://www.xalda.com/) for its kind collaboration.

References

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

Álvarez, I., Gutiérrez, J.P., Royo, L.J., Fernández, I., Gómez, E., Arranz, J.J., Goyache, J.J., 2005. Testing the usefulness of the molecular coancestry information to assess genetic relationships in

livestock using a set of Spanish sheep breeds. J. Anim. Sci. 83, 737–744.

Álvarez, I., Royo, L.J., Gutiérrez, J.P., Fernández, I., Arranz, J.J., Goyache, F., 2007. Genetic diversity loss due to selection for scrapie resistance in the rare Spanish Xalda sheep breed. Livest. Sci. 111, 204–212.

Álvarez Sevilla, A., Gutiérrez, J.P., Fernández, I., Royo, L.J., Álvarez, I., Gómez, E., Goyache, F., 2004. Conservación de la oveja Xalda de Asturias. AGRI 34, 41–49.

Balloux, F., Lugon-Moulin, N., 2002. The estimation of population differentiation with microsatellite markers. Mol. Ecol. 11, 155–165.

Bataillon, T.M., David, J.L., Schoen, D.J., 1996. Neutral genetic markers and conservation genetics. Simulated germplasm collections. Genetics 144, 409–417.

Baumung, R., Sölkner, J., 2003. Pedigree and marker information requirements to monitor genetic variability. Genet. Sel. Evol. 35 (383), 369.

Boichard, D., Maignel, L., Verrier, E., 1997. The value of using probabilities of gene origin to measure genetic variability in a population. Genet. Sel. Evol. 29, 5–23.

Caballero, A., Toro, M.A., 2000. Interrelations between effective population size and other pedigree tools for the management of conserved populations. Genet. Res. 75, 331–343.

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

Carothers, A.D., Rudan, I., Kolcic, I., Polasek, O., Hayward, C., Wright, A.F., Campbell, H., Teague, P., Hastie, N.D., Weber, J.L., 2006. Estimating human inbreeding coefficients: comparison of genealogical and marker heterozygosity approaches. Ann. Hum. Genet. 70, 666–676.

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.

Fernández, J., Villanueva, B., Pong-Wong, R., Toro, M.A., 2005. Efficiency of the use of pedigree and molecular marker information in conservation programs. Genetics 170, 1313–1321.

Goyache, F., Gutiérrez, J.P.I., Fernández, I., Gómez, E., Álvarez, I., Díez, J., Royo, L.J., 2003. Monitoring pedigree information to conserve the genetic variability in endangered populations: the Xalda sheep breed of Asturias as an example. J. Anim. Breed. Genet. 120, 95–103.

Gutiérrez, J.P., Goyache, F., 2005. A note on ENDOG: a computer program for analysing pedigree information. J. Anim. Breed. Genet. 122, 357–360.

Gutiérrez, J.P., Altarriba, J., Díaz, C., Quintanilla, R., Cañón, J., Piedrahita, J., 2003. Pedigree analysis of eight Spanish beef cattle breeds. Genet. Sel. Evol. 35, 43–64.

Gutiérrez, J.P., Royo, L.J., Álvarez, I., Goyache, F., 2005a. MolKin v2.0: a computer program for genetic analysis of populations using molecular coancestry information. J. Heredity 96, 718–721.

Gutiérrez, J.P., Marmí, J., Goyache, F., Jordana, J., 2005b. Pedigree information reveals moderate to high levels of inbreeding and a population genetic structure in the Catalonian donkey breed. J. Anim. Breed. Genet. 122, 378–386.

Hansson, B., Bensch, S., Hasselquist, D., Lillandt, B.G., Wennerberg, L., von Schantz, T., 2000. Increase of genetic variation over time in a recently founded population of great reed warblers (*Acrocephalus arundinaceus*) revealed by microsatellites and DNA fingerprinting. Mol. Ecol. 9, 1529–1538.

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

Lacy, R.C., 1989. Analysis of founder representation in pedigrees: founder equivalents and founder genome equivalents. Zoo Biol. 8, 111–123.

Lacy, R.C., 1995. Clarification of the genetic terms and their use in the management of captive populations. Zoo Biol. 14, 565–578.

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

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.

de Rochambeau, H., Fournet-Hanocq, F., Vu Tien Khang, J., 2001. Measuring and managing genetic variability in small populations. Ann. Zootech. 49, 77–93.

Royo, L.J., Álvarez, I., Gutiérrez, J.P., Fernández, I., Goyache, F., 2007. Genetic variability in the endangered Asturcón pony assessed using genealogical and molecular information. Livest. Sci. 107, 162–169. doi:10.1016/j.livsci.2006.09.010.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor. USA.

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

Tempelton, A.R., Read, B., 1994. Inbreeding: one word, several meanings, much confusion. In: Loeschke, V., Tomiuk, J., Jain, S.K. (Eds.), Conservation Genetics. Birkhäuser Verlag Basel, Switzerland, pp. 91–105.

Toro, M.A., Rodrigáñez, J., Silió, L., Rodríguez, C., 2000. Genealogical analysis of a close herd Black Hairless Iberian pigs. Conserv. Biol. 14, 1843–1851.

Toro, M.A., Barragán, C., Óvilo, C., Rodrigáñez, J., Rodríguez, C., Silió, M., 2002. Estimation of coancestry in Iberian pigs using molecular markers. Conserv. Genet. 3, 309–320.

Toro, M.A., Barragán, C., Óvilo, C., 2003. Estimation of genetic variability of the founder population in a conservation scheme using microsatellites. Anim. Genet. 34, 226–228.

Woolliams, J.A., Mäntysaari, E.A., 1995. Genetic contributions of Finnish Ayrshire bulls over four generations. Anim. Sci. 61, 177–197.

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