Combined use of genealogy and microsatellites in the endangered Xalda sheep

Isabel Álvarez, Juan Pablo Gutiérrez[1], Luis José Royo, Iván Fernández, Enrique Gómez and Félix Goyache

SERIDA-Somió, Camino de los Claveles 604, E33203-Gijón (Spain), e-mail: fgoyache@serida.org; [1]*Departamento de Producción Animal, Universidad Complutense de Madrid, E28040-Madrid (Spain)*

Summary

In small populations with shallow pedigree the combined use of genealogical and molecular markers information is recommendable to assess losses in genetic variability. Here we compare the 'real' molecular coancestry in the base population (Mol_0) with different estimations of this parameter in the rare Xalda sheep breed of Asturias. The Xalda individuals were classified in 4 different groups according to their pedigree knowledge using the 'equivalent generations' parameter: founder generation (0) when no parent is known, and generations 1th, 2nd and 3th. Up to 160 Xalda individuals corresponding to the 0 (46), 1th (37), 2nd (38) and 3th (39) generations were genotyped with 14 microsatellites. Molecular and genealogical coancestry coefficients have been computed for each pair of the 160 genotyped individuals. Average molecular coancestry was of 0.33, 0.36, 0.42 and 0.42 for, respectively, generations 0, 1th, 2nd and 3th. Estimations of the average molecular coancestry in the base population (M_0) were obtained as $M_0 = (M_g - f_g)/(1 - f_g)$ where f_g is the genealogical coancestry for the generation g. Difference between the estimates of M_0 and Mol_0 were 0.00, 0.05 and 0.03 for generations 1th, 2nd and 3th, respectively. When only non-founders are considered as a whole this difference was of 0.03. This methodology tends to overestimate the genetic variability in the base population, at least in the earlier generations.

Introduction

In small populations, the breeding goal is the conservation of genetic diversity. The joint effects of genetic drift and selection enhance the risk of losing alleles at selected or unselected genes and increase inbreeding in the population by changing the family structure. Various complementary criteria can be used to assess the changes in genetic variability (Rochambeau et al. 2001). Genetic variability can be described at the genealogical level, analysing demographic and pedigree information by means of parameters like inbreeding, kinship coefficients or effective population size (Caballero and Toro, 2000). At the molecular level, highly variable loci, such as microsatellites, provide a large amount of information on individual genotypes useful to clarify population structure (Balloux and Lugon-Moulin 2002; Álvarez et al., 2004). Both sources of information are based on similar assumptions: criteria based on pedigree information refer to any neutral autosomal locus while criteria based on observed genetic polymorphisms mirror phenomena referred to neutral genes or non-coding regions.

Individuals sampled for the estimation of population genetic structure should ideally belong to the same generation. However, population genetics data sets often comprise individuals from several generations. Allele frequencies vary not only over space, but also over time as populations are of finite sizes (Waples 1989). When the same site or trait is sampled over time, the absence of temporal genetic structuring can be tested by assessing the degree of differentiation between those samples. It can provide interesting information on differentiation occurred over generations (Viard et al. 1997). This can be particularly important after founder

events or bottlenecks (Hansson et al. 2000) in small populations. Preservation of rare genetic stocks makes necessary a continual monitoring of populations (Caballero and Toro 2000, 2002) especially when pedigree information is shallow and management structure leads to losses of genetic variability by drift or unobserved selection (Goyache et al. 2003). However, when the available sample size is small and genealogical information is shallow, any available tool has a limited power. In consequence, the combined use of pedigree information and molecular markers is recommendable to assess genetic diversity in small populations. Recent works have stated the mathematical relationships between coancestry coefficients computed using pedigree data and molecular tools thus encouraging the combined use of both sources of information in conservation programs (Caballero and Toro, 2000, 2002; Eding and Meuwissen, 2001; Toro et al., 2003).

The aim of this study is to check the reliability of combining genealogical and molecular information to monitor small populations. Population structure and genetic variability will be assessed over time. We will test these approaches on real data in which the pedigree is shallow and overlapping generations exists. The analysed data will come from the rare Xalda sheep breed of Asturias (Álvarez Sevilla et al., 2004; Goyache et al., 2003) which recently underwent a program of recovery and conservation of its genetic variability.

Materials and methods

We obtained the information included in the Xalda herd book since its establishment in late 80's to December 2003. A total of 1103 animals are registered in the herd book. A detailed description of the main characteristics of the Xalda's pedigree can be found in Goyache et al. (2003).

The pedigree information was analysed by using the program ENDOG v3.0 (Gutiérrez and Goyache, 2005) in order to obtain the full coancestry matrix of the pedigree. Each individual was assigned to a discrete generation by rounding the 'equivalent complete generations' parameter computed for each animal as the sum of $(1/2)^n$, where n is the number of generations separating the individual from each known ancestor (Maignel et al., 1996). Ancestors with no known parent were considered as founders (generation 0). The identified generations and total number of individuals per generation (in brackets) were: generation 0 (283), 1th (113), 2nd (204) and 3th (503).

Blood samples were obtained from a total of 160 Xalda individuals corresponding to the 0 (46), 1th (37), 2nd (38) and 3th (39) generations. 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, CSMM66, ILSTS011, McM53, RM006, ILSTS005) previously used in Álvarez et al. (2004; 2005) were analyzed in all individuals. The PCR products were labeled using a fluorescent method (Cy5 labeled primer) and genotyping was performed on an ALFexpressII automated sequencer (Amersham Biosciences, Barcelona). Molecular conacestry matrix between these 160 individuals was computed using the program MolKin v2.0 (Gutiérrez et al., 2005). The molecular coancestry between two individuals i and j is the probability that two randomly sampled alleles from the same locus in two individuals are identical by state (Caballero and Toro, 2002). Molecular coancestry between two individuals i and j at a given locus can be computed using the following scoring rules (Caballero and Toro, 2002; Eding and Meuwissen, 2001): $f_{ij,l} = \frac{1}{4}\left[I_{11} + I_{12} + I_{21} + I_{22}\right]$, 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, ¼, ½ and 1. The molecular coancestry between two individuals *i*

and *j* (f_{ij}) can be obtained by simply averaging over *L* analyzed loci $f_{ij} = \dfrac{\sum\limits_{l=1}^{L} f_{ij,l}}{L}$.

The analytical expression that relates the molecular coancestry values in the base (M_0) and current populations uses the genealogical coancestry coefficient (Toro et al., 2003). The expected value of M_0 is $E(M_g) = f_g + (1 - f_g)(M_0)$, where f_g is the genealogical coancestry in the *g* generation and M_0 the average molecular coancestry in the founder population.

Average genealogical (*f*) and molecular (*M*) coancestry for each generation (f_0, f_1, f_2, and f_3, and M_0, M_1, M_2 and M_3, respectively) was calculated. The correlation between different sets of genealogical and molecular coancestry matrices was computed using SAS/STAT[TM].

Results and Discussion

Average values and standard deviation for the genealogical and molecular conacestry values of 160 Xalda individuals classified into the 0, 1th, 2nd and 3th generations are given in Table 1. Both genealogical and molecular coancestry values increase with pedigree knowledge (from 0 to 0.09 and from 0.33 to 0.42 for, respectively, genealogical and molecular information). The Pearson and the Spearman correlations between *f* and *M* were significant and low (0.345 and 0.271, respectively) increased with pedigree knowledge to reach 0.606 and 0.428, respectively when only the individuals included into the 2nd and 3th generations are considered. Table 2 shows the Pearson and the Spearman correlations between *f* and *M* for each pair of generations. Relationships between the 1st generation and the others are not well accounted. It is needed to know at least two generations to obtain consistent correlations; at the 3th generation the Pearson and the Spearman correlation coefficients are comparable (0.420 and 0.455, respectively). Difference between the estimates of the molecular coancestry on the base population (M_0) and the 'real' molecular coancestry computed using founders were 0.00, 0.05 and 0.03 for generations 1th, 2nd and 3th, respectively. When the average molecular coancestry of non-founders is considered this difference was of 0.03.

Toro et al. (2003) encouraged the use of *M* in conservation programs based on the following reasons: a) there is a clear relationship between this measure in the current population and in the founder population through the genealogical coefficient; b) its value in the founder population coincides with the expected homozygosity, a classical parameter in population genetics (Nei, 1987); c) it is the only one that is unbiased even in presence of inbreeding.

Genealogical coancestry asses how probable an individual is identical by descent to a reference founder population in which is assumed that all the alleles are different. The estimation of *M* assumes that the founder population we consider in the genealogical study has the same allele frequencies than the sampled population and that these frequencies were both in Hardy-Weinberg and linkage equilibrium. Then, *M* basically asses the deviation of the present allele frequencies with respect those from a population showing the same allelic frequencies in Hardy-Weinberg equilibrium. These assumptions are not realistic. On one hand, genealogical information is not always the best to characterise the 'real' genetic variability existing in a population, mainly when the pedigree is shallow and we can not trace individual more than a few generations back (Goyache et al., 2003). In addition, it is not realistic to think in a virtually

infinite allelic forms when, for instance, perfect alleles in dinucleotide microsatellites rarely exceeds 30 repeats (Balloux and Lougon-Moulin, 2002). On the other hand, we can not properly expect that present populations had the same allele frequencies than founder populations because drift or selective processes they suffered during the formation of the breeding stock to be conserved (Toro et al., 2000; 2002; Goyache et al., 2003).

In this work we tried to show the reliability of using genealogical and molecular conacestry to monitor possible losses of genetic variability in small populations when it is possible to sample the founder generation. Estimates of the molecular coancestry in the base population compared with the 'thru' value of this parameter showed us that this methodology tends to overestimate the genetic variability in the base population, at least in the earlier generations. Increase of molecular coancestry seems to be faster than that for the genealogical parameter. However, when pedigree knowledge increases, correlation between these variables at individual level tend to be higher. The dataset we analysed here is not sufficient to obtain reliable conclusions. We have a good sample from the founder generation but our pedigree is very short and incomplete. Further research will be needed to obtain sound conclusions on the usefulness of this methodology.

References

Álvarez I, Gutiérrez JP, Royo LJ, Fernández I, Gómez E, Arranz JJ, Goyache F (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 LJ, Fernández I, Gutiérrez JP, Gómez E, Goyache F (2004) Genetic relationships and admixture between six Northern Spain sheep breeds assessed through microsatellites. J. Anim. Sci., 82, 2246-22520.

Álvarez Sevilla A, Gutiérrez JP, Fernández I, Royo LJ, Álvarez I, Gómez E, Goyache F (2003) 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

Caballero A, Toro MA (2000). Interrelations between effective population size and other pedigree tools for the management of conserved populations. Gen Res 75: 331-343.

Caballero A, Toro MA (2002). Analysis of genetic diversity for the management of conserved subdivided populations. Conserv Gen 3: 289-299.

Eding H., Meuwissen, T.H.E. (2001) Marker-based estimates of between and within population kinships for the conservation of genetic diversity. Journal of Animal Breeding and Genetics 118, 141-59.

Goyache F, Gutiérrez JP, Fernández I, Gómez E, Álvarez I, Díez J, Royo LJ (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.

Guo X., Elston R.C. (1999) Linkage informative content of polymorphic genetic markers. Hum Hered 49: 112-118.

Gutiérrez, J.P., Goyache, F. (2005) A note on ENDOG: a computer program for analysing pedigree information. J Anim Breed Genet, 122: in press

Gutiérrez, J.P., Royo, L.J., Álvarez, I., Goyache, F. (2005) MolKin v2.0: a computer program for genetic analysis of populations using molecular coancestry information. J. Hered., submitted

Hansson B, Bensch S, Hasselquist D, Lillandt BG, 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 fingerpriniting. Mol Ecol 9: 1529-1538.

Nei M (1987). Molecular Evolutionary Genetics. Columbia University Press, New York. University Press, New York.selected populations.

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

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

Toro, M.A., Barragán, C., Óvilo, C. (2003) Estimation of genetic variability of the founder population in a conservation scheme using microsatellites. Animal Genetics 34: 226-228.

Viard F, Justy F, Jarne P (1997). Population dynamics inferred from temporal variation at microsatellite loci in the selfing snail Bulinus truncatus. Genetics, 14: 6973-6982.

Waples RS (1989). A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics, 121: 379-391.

Table 1: Average values and standard deviation for the genealogical and molecular coancestry values of 160 Xalda individuals classified into the 0 (N = 46), 1th (N = 37), 2nd (N = 38) and 3th (N = 39) generations.

Generation	Genealogical coancestry		Molecular coancestry	
	Mean	s.d.	Mean	s.d.
0			0.33	0.12
1th	0.04	0.12	0.36	0.11
2th	0.06	0.12	0.42	0.11
3th	0.09	0.12	0.42	0.12

Table 2: Pearson and Spearman correlation coefficients computed between genealogical and molecular coancestry values by generation, excluding those pairs involving the founder generation. All the correlation coefficients were significant for p <0.001 except for that noted as non-significant (ns as superscript)

Generation	1th	2th	3th
Pearson			
1th	0.380		
2th	0.185	0.435	
3th	0.087	0.459	0.420
Spearman			
1th	0.194		
2th	0.106	0.320	
3th	0.016^{ns}	0.390	0.455