CLASSIFICATION OF CROSSBRED ANIMALS USING MICROSATELLITES

A. Farid[1], A. Almudevar[2], C. Dollard[1] and E. O'Reilly[3]

1-Dept. Animal Sci., Nova Scotia Agricultural College, Truro, N.S., Canada, B2N 5E3.
2-Dept. Math., Stat. & Computer Sci., Dalhousie Univ., Halifax, N.S. Canada, B3H 3J5.
3-Dept. Forest Resources & Agriculture, St. John's, Newfoundland, Canada, A1B 4J6.

SUMMARY

Unrelated purebred sheep from 10 breeds (25/breed), including North Country Cheviot, Suffolk, Romanov, Texel, Cheviot, Icelandic, Finnsheep, Dorset, Scottish Blackface and Red Masai were genotyped at 10 microsatellite loci. Samples of 1000 purebred and F_1 individuals were generated by Monte Carlo simulation using observed allele frequency distributions. Individuals were assigned to one of the 10 breeds where likelihood for the genotypes was maximum. A modified maximum likelihood (ML) procedure was used to assign samples of 500 simulated F_1 individuals to all possible purebred and crossbred groups, using combined allele frequency distributions of the breeds involved in each cross. Large differences among the 10 breeds for allele frequency distributions at the microsatellite loci resulted in 97.9% to 99.9% correct assignment of purebred individuals. Small numbers of F_1 individuals were correctly assigned to the parental breeds, and in most cases a larger number was assigned to unrelated breeds, indicating that allele frequency distributions of purebred breeds are not appropriate for assignment of crossbred animals. The modified maximum likelihood procedure resulted in correct assignment of 53% to 87% of crossbred animals.

Keywords: microsatellites, sheep breeds, assignment test, maximum likelihood

INTRODUCTION

One of the applications of microsatellites is assignment of individuals of unknown origin to one of several subpopulations. When allele frequency distributions of purebred breeds are known, and random mating and linkage equilibrium within breeds are assumed, individuals can be assigned to the correct breed using maximum likelihood estimates (Buchanan *et al.* 1994; Paetkau *et al.* 1995). We show here that this procedure is not satisfactory for determining parental breeds of crossbred animals when using microsatellite loci, and we propose the proper maximum likelihood procedure.

MATERIALS AND METHODS

Source of animals. Blood or semen samples were collected from 25 purebred sheep from each of 10 breeds, including North Country Cheviot (NC), Suffolk (SU), Romanov (RO), Texel (TX), Cheviot (CH), Icelandic (IC), Finnsheep (FN), Dorset (DO), Scottish Blackface (SB) and Red Masai (RM). Samples from each breed were taken from several farms to ensure obtaining a representative samples of alleles from each breed. DNA samples from most of the IC were kindly provided by Dr. A. Palsdottir, Institute of Experimental Pathology, Keldur, Iceland. DNA of RM sheep was supplied by Drs. A.J. Teal and O. Hanotte, International Livestock

Research Institute, Kenya. NC, DO and SB samples were collected from Canadian flocks, FN, SU and RO originated from flocks in Canada and the US, and samples of CH and TX came from Canada, US and Europe. Most Canadian sheep were registered and unrelated for at least two generations.

Laboratory procedures. Ten microsatellite loci were selected for this study: MAF64, MAF35, OarFCB128, OarFCB11, OarFCB5, OarCP26, OarCP21, OarAE129, OarAE101 and TGLA53 (Crawford *et al.* 1995). PCR reactions were carried out in 10-15 μL volumes containing 100-200 ng sample DNA, 0.2 mM dNTPs, 0.3-0.4 U Taq polymerase, 1X PCR buffer, 0.1% Tween 20, 34 ρM unlabelled primer and 6 ρM labelled primer. Primers were end labelled using γ-[^{32}P]ATP and T4 polynucleotide kinase. Two step thermal cycling conditions were used for all microsatellites except for TGLA53, which used touchdown PCR (Crawford *et al.* 1995). PCR products were subjected to PAGE on 8% polyacrylamide gels. Alleles were visualized using autoradigraphy and sized using M13mp 18 sequencing ladder as a standard.

Statistical analyses. For each population at each locus, the number of alleles and allele frequency distributions were calculated. Three Monte Carlo simulations were performed. First, genotypes of 1000 individuals at every locus were simulated from each purebred breed using the observed allele frequency distributions. For each individual, the likelihood of the genotypes was calculated for each breed, and the individual was assigned to that breed giving the maximum likelihood. Second, genotypes of 1000 crosses between every two breeds (F_1) were generated by taking gametes from frequency distributions of pairs of breeds. These individuals were then assigned to all purebred breeds as before. Third, samples of 500 individuals were simulated from every possible purebred and F1 group. The likelihood for the genotypes was computed for each configuration, and the individual assigned to the configuration giving the maximum likelihood. The likelihood for a crossbred incorporates the allele frequencies from both populations, to give the correct genotype frequency for a crossbred population.

RESULTS AND DISCUSSION

Primers to loci MAF64, MAF35, OarFCB128, OarFCB11, OarFCB5, OarCP26, OarCP21, OarAE129, OarAE101 and TGLA53 generated 16, 7, 8, 10, 3, 14, 6, 8, 9, and 13 alleles, respectively. Except for IC at locus OarFCB5, all other loci were polymorphic in every breed. The large differences among the breeds for allele frequency distributions at the 10 microsatellite loci resulted in correct classification of a high proportion of animals from each sample, ranging from 97.9% in Cheviot to 99.9% in Romanov (Table 1). The numbers of animals from each breed that were misassigned to other breeds, although very small, were the largest among DO, NC, CH and SB breeds. No individual from any of the breeds was assigned to Red Masai, indicating that the number of animals from a breed that was misassigned into other breeds was inversely related to the time of divergence and genetic distance among these breeds. The proportion of correct assignment was high even for the breeds such as CH and NC that have separated from each other rather recently. The results suggest that this panel of 10 microsatellite loci is highly characteristics of these breeds, and is a valuable tool for assigning animals to a correct breed, given that they belong to one of these breeds.

Table 1. Results of assigning 1000 individuals from each breed (row headings) to every purebred breed (column headings)

Purebred Breeds	DO	NC	SB	BC	FN	IC	RO	SU	TX	RM
Dorset (DO)	989	1	3	2	1	0	0	0	4	0
N.C. Cheviot (NC)	1	986	3	10	0	0	0	0	0	0
Scottish Blackface (SB)	1	3	988	4	1	0	0	1	2	0
Cheviot (CH)	3	9	3	979	1	1	0	2	2	0
Finnsheep (FN)	0	0	0	2	996	1	1	0	0	0
Icelandic (IC)	0	0	2	0	0	998	0	0	0	0
Romanov (RO)	0	0	0	1	0	0	999	0	0	0
Suffolk (SU)	0	1	0	5	1	0	0	993	0	0
Texel (TX)	4	0	2	2	1	4	0	1	986	0
Red Masai (RM)	0	0	0	0	0	1	1	0	0	998

Table 2 shows the results of assigning 1000 simulated individuals from five crossbred groups to each of the 10 purebreds. In these and all other cases (not shown) a small proportion of F_1 individuals were assigned to the purebred breeds that made up the cross. This proportion was larger for crosses among recently-diverged breeds, such DO x NC and NC x CH, than the crosses between more distinctly related breeds. In the latter case, large proportions of animals were not assigned to any of the 10 breeds as a result of alleles present in F_1 individuals but absent in one of the parental breeds. Interestingly, larger proportions of crosses between more distinct breeds, such as RM x DO, DO x RO and IC x TX, were assigned to breeds other than the parental breeds. These results show that allele frequency distributions of purebred breeds are not appropriate for determining breed composition of crossbred animals, and the chance of misassignment into entirely unrelated breeds increases as the differences between breeds for allele frequency distributions increases.

Percentages of F_1 individuals that were correctly assigned to each breed type using the modified ML procedure are shown in Table 3. Diagonal elements are proportions of correct assignment of purebreds to their own breed. These values are smaller than those in Table 1 because some purebred individuals were assigned to crossbred groups, which were not examined in the previous case. The proportions of correct assignment of crossbreds ranged from 0.534 in SU x CH to 0.872 in RM x RO, and were generally higher in unrelated breeds and their crosses than in related breeds. In conclusion, using highly polymorphic microsatellite markers when assigning crossbred individual to purebred breeds, may result in incorrect classification. To eliminate this problem, the ML estimates using combined allele frequency distributions of purebred breeds should be used. This procedure is useful in protecting the integrity of purebred breeds.

Table 2. Results of assigning 1000 F_1 individuals (column headings) to every purebred breed

	DO*NC	NC*CH	RM*DO	DO*RO	IC*TX
Dorset (DO)	299	11	2	27	64
N.C. Cheviot (NC)	173	342	57	26	19
Scottish Blackface (SB)	91	46	125	178	48
Cheviot (CH)	31	367	37	36	21
Finnsheep (FN)	4	0	23	54	31
Icelandic (IC)	0	0	4	3	52
Romanov (RO)	0	0	1	24	0
Suffolk (SU)	0	2	27	48	2
Texel (TX)	24	2	53	76	193
Red Masai (RM)	0	0	83	10	0
Not assigned	370	230	588	518	570

Table 3. Percentage of correctly assigned F_1 individuals to its own breed type

	DO	NC	SB	CH	FN	IC	RO	SU	TX	RM
Dorset (DO)	0.812									
N.C. Cheviot (NC)	0.604	0.766								
Scottish Blackface (SB)	0.568	0.560	0.694							
Cheviot (CH)	0.546	0.556	0.558	0.750						
Finnsheep (FN)	0.626	0.658	0.576	0.584	0.850					
Icelandic (IC)	0.710	0.664	0.604	0.598	0.718	0.876				
Romanov (RO)	0.742	0.762	0.696	0.674	0.806	0.824	0.974			
Suffolk (SU)	0.690	0.648	0.634	0.534	0.660	0.696	0.816	0.880		
Texel (TX)	0.614	0.678	0.664	0.620	0.686	0.668	0.766	0.704	0.850	
Red Masai (RM)	0.762	0.770	0.744	0.704	0.770	0.762	0.872	0.808	0.824	0.948

REFERENCES

Buchanan, F.C., Adams, L.J. Littlejohn, R.P. Madox J.F. and Crawford, A.M. (1994). *Genomics* **22**:394-403.

Crawford, A.M., Dodds, K.G., Ede, A.J., Pierson, C.A., Montgomery, G.W., Garmonsway, H.G., Beattie, A.E., Davies, K., Maddox, J.F., Kappes, S.W., Stone, R.T., Nguyen, T.C., Penty, J.M., Lord, E.A., Broom, J.E., Buitkamp, J., Schwaiger, W., Epplen, J.T., Matthew, P., Matthews, M.E., Hulme, D.J., Beh, K.J., McGraw, R.A. and Beattie, C.W (1995). *Genetics*. **140**:703-724.

Paetkau, D., Calvert, W., Stirling, I. and Strobeck, C. (1995). *Mol. Ecol.* **4**:347-354.