MICROSATELLITE VARIATION AND POPULATION GENETIC STRUCTURE OF SELECTED MERINO SHEEP FLOCKS

R. Al-Atiyat[1], W. Flood[1], I. Franklin[2], B.P. Kinghorn[1], A. Ruvinsky[1]

[1] Genetics, SRSA, University of New England, NSW Australia
[2] CSIRO Livestock Industries, Prospect, NSW Australia

INTRODUCTION

Mapping microsatellite (MS) loci in sheep (Maddox, 2001) has opened an opportunity for detailed investigation of genetic structure of populations. These studies may contribute to better understanding of population dynamics including estimation of linkage disequilibrium (LD). The latest is useful for estimation selective forces, random factors and the history of populations. This paper presents an initial step in investigation of three sheep populations. We used 22 markers with the following aims : 1) to assess the genetic variation and population parameters in three Australian Merino flocks, which experienced specific selection pressures over a number of generations, 2) to study LD, using 14 markers located on 6 different chromosomes with at least 2 markers on each chromosome.

MATERIAL AND METHODS

Animals. Three different populations were studied. The first was a control flock. The other 2 populations were drawn from pedigreed flocks maintained by the CSIRO, Australia. One of them was selected for low parasite resistance (LR) and the other for high (HR). Both flocks originated from the same initial population and have been closed since 1976 (Woolaston *et al*., 1990). Mate allocation was designed to ensure no mating between relatives closer than four generations apart. The effective population size was equal to 20 for each of CSIRO's flocks and to 72 for the control flock.

Microsatellite markers and Genotyping. All studied animals were genotyped for 22 MS markers. The PCR primers were fluorescently labeled and scored automatically by ABI^{TM} microsatellite typing system. The amplified fluorescent products were visualized using GENESCAN^{TM} Software; genotypic data were scored, interpreted and transferred into Excel sheet files by using GENOTYPER^{TM} Software (Applied Biosystems, 1994).

DNA-Based Parenting and basic population genetic analysis. Cervus software (Marshall *et al.,* 1998) was used for parentage analysis. Allele frequencies and Hardy-Weinberg proportion (HWP) were calculated from genotypic data using GDA software (Lewis and Zaykin, 2001).

Estimation of LD. To avoid bias caused by sires' contributions, only maternal gametes transmitted to offspring were utilised to estimate LD using Arlequin software (Schneider *et al.,* 2000). The LD was studied at two levels. Firstly we estimated the disequilibrium for each pair of alleles (haplotype), so called interallelic LD. Then an overall LD analysis, which aggregated data on disequlibrium between all alleles at the two loci was conducted (Zapata *et al.,* 2001).

RESULTS AND DISCUSSION

Allele Frequency and Polymorphism. A total of 465 individuals were genotyped. The average number of alleles per locus varied as follows : 11.95, 10.09 and 8.68 in control, LR and HR populations, respectively. The number of alleles per MS locus varied from 14 to 5. Average expected heterozygosity was close to 0.75 in the studied populations.

DNA-Based Parenting. The level of allele mismatches between parents and offspring was low – about 0.1%. Mismatches between results of DNA-parentage test and breeding records were significant : 22% for dams and 8% for sires in the LR and HR populations. We have used results of parentage testing rather than breeding records for further analysis.

Inbreeding coefficients and Hardy Weinberg proportions. The inbreeding coefficients were low. The explanation for the control population is a regular gene flow from outside. In the experimental populations, strict breeding policy excludes mating between all but weakly linked relatives. Significant deviations from Hardy Weinberg proportions (HWP) were found only in 11 of the 66 cases studied and these deviations were distributed randomly.

Linkage Disequilibrium. *Interallelic Linkage Disequilibrium (LD).* The intensity of LD depends upon the recombination between two loci located on the same chromosome. In large outbred populations, which do not experience strong pressures from outside, LD is expected only for very tightly linked markers (Kruglyak, 1999). However the majority of populations can be far away from such conditions (Zapata et al., 2001). Our data presented in the Table 1 show that LD spread over long linkage distances in the populations studied. Altogether 11 syntenic pairs of loci were examined. In the control population (data not shown) these pairs generated 503 haplotypes, in LR 464 haplotypes and in HR 485 haplotypes. Only a small fraction of these haplotypes is shown in Table 1. In the majority of cases values of significant LD coefficients were positive, indicating that particular haplotypes were over represented. In the control and HR all D_{ij} coefficients except one did not reach highly significant χ^2 values. In contrast, for LR half of the values were highly significant. All five haplotypes representing pair of loci CSRD2105-CSRD254 located on Chromosome 2 demonstrate extreme values of LD. These were over represented at about 48.5% comparing with HWP, thus squeezing off the majority of other haplotypes from the LR. The same haplotypes have not shown significant deviations from HWP neither in the control nor in HR populations, except one case in the HR. The pair of loci McM147-McM512 also demonstrates the same pattern. Similar observations but with the opposite sign can be found when comparing syntenic pairs of loci CSRD2108-McM058, McM147-CSRD2105 and McM147-CSRD254. In this case all deviations are found in the HR, but not in the LR population. There are many significant LD values between remote alleles (186 cM, 145 cM, etc.). Such long map distances correspond to several crossing-over events during each meiosis and should lead to rapid disappearance of over represented haplotypes regardless of the cause promoting the haplotype. A survival of some haplotypes could be explained by frequent use of some sires. However LD was calculated using maternally derived gametes only. One could explain these observations assuming that a selective force either block recombination or promote recombinant chromosomes with even number ofrecombination events between two remote markers.

Table 1. Interallelic linkage disequilibrium in three studied populations[A]

	LR Population				HR Population			
Locus Pair	Haplotype	Dij	χ^2	D'	Haplotype	Dij	χ^2	D'
CSRD2108		NS			135/181	0.038	4.34	0.242
-McM058					137/195	0.041	4.61	0.30
McM147CS		NS			278/181	0.043	5.27	0.45
-RD2105					286/183	0.065	9.29*	0.782
					276/203	0.042	10.2*	1
McM147		NS			260/104	0.033	4.82	0.328
-CSRD254					284/104	0.033	4.82	0.328
McM147	280/119	0.057	13.7*	0.71		NS		
-McM512								
CSRD2105	175/100	0.075	**33.55**	1	181/104	0.042	6.55	0.455
-CSRD254	175/104	0.050	**21.66**	1				
	181/114	0.210	**54.75**	0.93				
	183/116	0.078	**22.65**	0.72				
	183/118	0.072	**27.37**	1				
CSRD2105	121/175	0.045	3.87	0.48	117/203	0.044	4.96	0.337
-McM512	117/183	0.046	4.69	0.30				
CSRD254	121/104	0.039	7.11*	1	101/104	0.029	5.22	0.29
-McM512					117/118	0.043	5.85	0.284
McM053	140/193	0.041	5.05	0.42	126/195	0.023	5.22	1
-MCMA14								
McMA10	219/140	-0.059	8.38*	-0.4	217/144	-0.051	5.10	-0.59
-CSRD240	219/150	0.051	10.0*	0.46				
CSRD247-	NS					NS		
McM10								
CSRD2148-	NS				286/150	0.054	5.62	0.47
McM136								

[A] Sample Size (No. of maternal haplotypes) in the control, LR and HR is 94, 71 and 87, respectively. Numbers in bold represent χ^2 probability of P≤0.001. * - χ^2 probability is P≤0.01 ; The rest of χ^2 probabilities are at P≤0.05. Degree of freedom is 1.
$D_{ij}=p_{ij}-p_ip_j$ and $D'_{ij}= D_{ij}/D_{max}$ (Lewontin, 1964)

Taking into consideration extraordinary concentration of highly significant LD on the *CSRD2105-CSRD254* interval on the Chromosome 2, one may suspect that this region was affected during selection for low resistance to nematodes.

Overall LD disequilibrium. We used the exact test (Schneider *et al.*, 2000) in order to check whether accumulated effects of interallelic LD were statistically significant (Table 2).

Table 2. Significant overall linkage disequilibrium cases between some studied loci

Locus	CSRD2105	CSRD2148	McMA10	CSRD2108	McM512	McM104
CSRD254	•				■	
McM136		+				
CSRD240			+			
McM058				■		
McM147					■	
CSRD24						■

■ Significant overall LD (P≤0.05) in control population ; • Significant overall LD (P≤0.05) in LR population ; + Significant overall LD (P≤0.05) in HR population.

Highly significant overall LD in the LR population was found between *CSRD254* and *CSRD2105* loci (Chr. 2), where several strong cases of interallelic LD (P ≈ 0) were observed. In HR the overall LD were found between loci : *McM136* and *CSRD2148* (Chr. 23) ; *CSRD240* and *McMA10* (P = 0.001; Chr. 9). In the control population the following cases of overall LD were observed: *McM058* and *CSRD2108* (Chr. 1) ; *CSRD254* and *McM512* (Chr. 2) ; *McM512* and *McM147* (Chr. 2) ; *CSRD247* and *McM104* (Chr. 14). Interestingly overall LD in LR and HR populations were different, while HR and control populations show presence of overall LD in the neighboring regions of Chr. 2. This data set needs to be expanded before any final conclusion is made.

ACKNOWLEDGEMENTS
We are thankful to S. Eady, L. Piper, I. Purvis for the support of this project and permission to use animals.

REFERENCES

Applied Biosystems, A Division of the Perkin-Elmer Corporation. (1994) GENESCAN™ and GENOTYPER™
Lewis, P. and Zaykin, D. (2001) GDA, ver1d16c. http://lewis.eeb.uconn.edu/lewishome/
Lewontin, R.C. (1964) *Genetics* **49** : 49-67.
Maddox, J.F. *et al.* (2001) *Genome Res.***11** : 1275-1289.
Marshall, T.C., Slate, J. and Pemberton, J.M. (1998) *Mol Ecol.* **7** : 639-655.
Schneider, S. Roessli, D. and Excoffier, L. (2000) Arlequin, ver2000. http://anthropologie.unige.ch/arlequin/
Woolaston, R.R, Barger, I.A, Piper, L.R. (1990) *Int. J. Parasitol.* **20** : 1015-1018.
Zapata, C., Rodriguez, S., Visedo, G. and Sacristan, F. (2001) *Genetics* **158** : 235-251.