Associations between genetic polymorphisms and phenotypic traits in the Lesvos dairy sheep

I. Mastranestasis [a,b], A. Kominakis [c,*], A.L. Hager-Theodorides [c], L.V. Ekateriniadou [b], Ch. Ligda [b], K. Theodorou [a]

[a] Biodiversity Conservation Laboratory, Department of Environment, University of the Aegean, 81100 Mytilene, Greece
[b] Hellenic Agriculture Organization – Veterinary Research Institute of Thessaloniki, Campus of Thermi, P.O. Box 60272, 57001 Thermi, Greece
[c] Department of Animal Breeding, Agriculture University of Athens, Athens, Greece

ARTICLE INFO

ABSTRACT

The aim of the present study was to investigate the potential associations between polymorphisms of microsatellite markers and various traits in the Lesvos dairy sheep. In total, 350 individuals using 11 microsatellite markers were genotyped and various traits including six body anatomical measurements as well as milk yield and litter size were considered. The adaptive Least Absolute Shrinkage and Selection Operator method was employed to detect significant associations between markers' alleles and the various phenotypic traits. In all but one microsatellite loci, rare alleles were detected. Markers OARJMP29, OARFCB128 and OARCP38 were found to be associated with milk yield. Marker OARJMP29 was also associated with body weight, girth circumference and horn length. Marker OARFCB128 was also significant for tail length, tail width and girth circumference while marker OARCP38 was significant for body length, as well. Current results shed more light in the associations between marker polymorphisms with various phenotypic traits in dairy sheep, including traits of economic importance that could potentially be implemented in a marker assisted selection program for increasing production efficiency in the Lesvos breed.

1. Introduction

The Lesvos dairy sheep is a fat-tailed and mixed-wool type native breed mainly located on the homonymous island of Greece. The breed represents a rich genetic resource, with high gene flow between flocks and high levels of genetic variability (Mastranestasis et al., 2015). Variation extends to morphology, coat coloration and (re)production traits (Sossidou et al., 2013).

Today, about 260,000 purebred animals are being kept on the island (Ministry of Rural Development and Food, MRDF, 2015) dispersed over 2000 farms of medium sized flocks (100–200 animals) under a semi-intensive farming system. Apart from the Lesvos island, the breed is also found in other Aegean islands and the mainland of Greece where it is used to upgrade local sheep populations by crossbreeding. The breed is well adapted to dry-heat areas and degraded soils while delivering relatively high milk production. Due to its ability to produce under unfavorable conditions, the demand on genetic material of this specific breed remains high across the country. On the Lesvos island, about 30,000 ewes are officially recorded following the A4 recording scheme (ICAR, 2014). In the recorded flocks, the average milk yield ($\pm$SD) is 224 ± 77 kg in 163 ± 33 days milking period (MRDF, 2015).

Selection of animals may extend from correction of milk yield on systematic effects to estimation of breeding values using Best Linear Unbiased Prediction (BLUP). Marker assisted selection (MAS) could improve the efficacy of the applied selection scheme, especially for traits of economic importance i.e. milk yield and litter size. The first step of MAS is the identification of molecular markers that are linked to quantitative trait loci (QTL). Various molecular markers may be used here with microsatellites (or Simple Sequence Repeats, SSRs) being excellent candidates; they are abundant, evenly distributed in the genome, highly polymorphic and co-dominantly inherited. In sheep, SSRs have been used for a variety of purposes including, parentage verification, pedigree analysis, gene tagging, genome mapping, estimation of genetic diversity, phylogenetic analyses as well as in association studies of disease resistance (e.g. Phua et al., 1999; Benavides et al., 2002) and production traits (Chatziplis et al., 2013).

Driven from the high importance of the Lesvos sheep, we have conducted the present study with the aim to detect genetic poly-

Table 1
Number of observations (n), mean (M), standard error of mean (SE) and standard deviation (SD) of variables overall and per sex.

Variable (abbreviation)	n	M ± SE	SD	Males n	Males M ± SE	Males SD	Females n	Females M ± SE	Females SD
Horn length, HL (cm)	341	13.18 ± 0.39	10.24	60	28.92 ± 0.62	6.84	281	9.82 ± 0.30	7.29
Wither height, WH (cm)	341	69.96 ± 0.15	4.00	60	76.17 ± 0.25	2.74	281	68.63 ± 0.12	2.79
Girth circumference, GC (cm)	340	93.14 ± 0.24	6.18	60	98.43 ± 0.54	5.89	281	92.00 ± 0.24	5.63
Body length, BL (cm)	341	70.42 ± 0.15	4.02	60	76.37 ± 0.32	3.52	281	69.15 ± 0.12	2.79
Tail length, TL (cm)	335	30.98 ± 0.19	5.05	60	32.53 ± 0.29	3.18	275	30.42 ± 0.16	3.74
Tail width, TW (cm)	338	13.78 ± 0.16	4.24	60	15.25 ± 0.27	2.92	278	13.26 ± 0.11	2.58
Body weight, BW (kg)	277	57.60 ± 0.47	11.03	45	70.76 ± 1.06	10.01	232	55.15 ± 0.42	9.15
Days in milk, DIM (days)	270	173.73 ± 1.18					270	173.73 ± 1.18	26.00
Milk yield, MY (kg)	270	189.25 ± 2.25					270	189.25 ± 2.25	52.35
Litter size, LS	229	1.24 ± 0.02					229	1.24 ± 0.02	0.42

morphisms that might be associated with various phenotypic traits, including traits of economic importance and that could be used to enhance the efficacy of the applied selection scheme in the breed.

2. Material and methods

2.1. Samples – DNA extraction and PCR amplification

A total of 350 random blood samples were collected from 10 officially registered flocks dispersed in five regions in the Lesvos island. Genomic DNA was extracted from fresh whole blood using the PureLink Genomic DNA Mini Kit (Invitrogen by Life Technologies). Detailed information on methodology and molecular techniques are given in Mastranestasis et al. (2015).

2.2. Description of traits

Six body anatomical traits i.e. horn length (HL), wither height (WH), girth circumference (GC), body length (BL), tail length (TL) and tail width (TW) using measuring tape, compass and staff along with body weight (BW) were recorded on the genotyped animals (rams and ewes). The number of observations available varied across traits and sex (Table 1). Ewes' records on milk yield (MY) and litter size (LS) were also made available from the official milk recording scheme (Table 1).

2.3. Marker data

A total number of 11 microsatellites (MAF65, OARFCB193, OARJMP29, OARFCB128, OARCP34, MCM140, MAF209, MAF214, MAF70, BM1824, OARCP38, ILSTS5) belonging to the sheep biodiversity panel were used from the list of the FAO-ISAG group (FAO, 2004). These markers were chosen by taking into account the level of polymorphism detected in previous studies as well as their location in different chromosomes (e.g. Al-Tarayrah, 2006; Lawson Handley et al., 2007; Ligda et al., 2009). To control for false positive error rate(s) during association analysis, an adequate representation of alleles per locus was pursued by grouping rare alleles (with frequency less than 5%) in one class using a dummy allele description (denoted by 100). Application of a non-parametric (i.e. Spearman) correlation analysis of the marker data (after grouping) revealed correlations of intermediate magnitude (0.40–0.60), among many pairs of markers' loci implying the presence of multicollinearity in our data.

2.4. Statistical analysis

Due to multiple alleles per locus, leading to limited numbers of genotypes available, analysis was focused at the allele level. This enabled the estimation of only single allelic effects and not their interaction either at single locus (dominance deviation) and/or at loci combinations (epistatic effects). Since classical statistical methods (such multi-way ANOVA, multiple regression or stepwise regression) are expecting to show degraded performance in correctly identifying significant factors and/or regressors under multicollinearity conditions (George, 2000; Mundry and Nunn, 2009), the adaptive Least Absolute Shrinkage and Selection Operator (LASSO), as implemented in procedure GLMSELECT in SAS 9.3 (2012), was employed as a variable selection method i.e. to detect associations between the markers and the various traits. For each trait, a fixed effects model was fitted by treating markers (n = 11) as the main fixed effects. During analysis of the body anatomical traits, sex (two classes: males and females), herd (ten classes) and age (in years) of rams (four classes: 2, 3, 4 and ≥5) or age (in years) of ewes (five classes: 2, 3, 4, 5 and >5) were also fitted as additional fixed effects. During analysis of MY, the previous model was extended to include days in milk as a covariate, as well. Finally, least squares (LS) means for marker alleles that were found to be statistically significant according to the adaptive LASSO were estimated, using standard ANOVA. To deal with the multiple comparison problem, because of multiple alleles per marker locus, the adjusted p-values were computed from simulated distributions of 200,000 samples. Standard ANOVA, LS means estimation and adjustment of p-values were performed by procedure GLM in SAS 9.3 (2012).

2.5. Search of QTLs and candidate genes neighboring significant markers

Chromosomal regions surrounding significant microsatellites to a distance spanning 1 Mb were searched for the presence of known QTLs and/or ovine genes in an attempt to identify the causative molecular genetic basis of the marker-trait associations. The sheepQTLdb and the NCBI genome (*Ovis aries*, assembly OAR_v3.1, NCBI annotation release 101) databases were searched for QTLs and genes, respectively. It should be noted here that due to the currently available annotation for the ovine genome, all candidate genes are assigned reference sequence status "model" (i.e. provided by the NCBI Genome Annotation pipeline and not subject to individual review); thus information about their function is still limited. Nevertheless, information about the function of respective orthologs in other species can be used to identify genes with relevant functions.

3. Results

3.1. Alleles per locus

Alleles description along with numbers and relative frequencies are shown in Table 2. In all markers, except for locus OARCP38, rare alleles were detected. Note that even after alleles grouping, the relative frequency of rare alleles was as high as 0.037, 0.035 and 0.016 for markers OARCP34, for OARFCB128 and MAF160, respec-