

Detecting footprints of selection in *Ovis aries* by a spatial analysis approach

S. Joost¹, R. Negrini¹, E. Milanese², M. Pellicchia¹,
P. Ajmone Marsan¹, the Econogene Consortium

¹ Istituto di Zootecnica. Università Cattolica del S. Cuore, Piacenza, Italy

² Istituto di Zootecnica Generale. Università di Milano, Italy

Corresponding author: Stéphane Joost. Istituto di Zootecnica. Università Cattolica del S. Cuore. Via Emilia Parmense 84, 29100 Piacenza - Tel. +39 0523 599205 - Fax: +39 0523 599276 - Email: stephane.joost@econogene.eu

ABSTRACT: Detecting adaptive loci in the genome is essential as it gives the possibility to understand what proportion of a genome or which genes are under the pressure of natural selection. In this paper, we used a Spatial Analysis Method (SAM) recently developed to detect signatures of natural selection in sheep breeds. With the contribution of Geographical Information Systems, environmental variables, and AFLP data, multiple univariate logistic regressions are run to test for association between allelic frequencies at marker loci and environmental variables. The results of the application of this method to sheep breeds are compared with those obtained with a standard population genetics approach.

Key words: Spatial analysis, Natural selection, Landscape genomics, AFLP.

INTRODUCTION – The detection of signatures of natural selection within the genome of organisms allows to understand what proportion of a genome or which genes are under the influence of natural selection. Regions of the genome under selection are generally of functional importance, hence inferences regarding selection may provide useful information for identifying important genes (Nielsen, 2005).

Different methods have been developed to reveal genomic regions target of natural selection (Vasemägi and Primmer, 2005). Some of them belong to the “candidate-gene” category, and others aim at identifying Quantitative Trait Loci (QTL) involved in the expression of adaptive traits. Nevertheless, their application is often limited to relatively few well-studied species (Phillips, 2005) and they require *a priori* genomic information that may not be easily accessible for non-model organisms (Erickson *et al.*, 2004). Population genomics is another strategy. It relies on the principle that all loci across the genome are influenced by genome-wide evolutionary forces (genetic drift, gene flow), whereas locus-specific forces such as selection imprint a particular variability pattern on a few loci only (Luikart *et al.*, 2003). By comparing the genetic diversity of many loci across the genome, it is then possible to reveal loci displaying an atypical variation pattern (outlier loci), which are likely to be affected by selection (Black *et al.*, 2001). Strategies making use of population genomics are free from any prior knowledge about the selectively advantageous gene or phenotype and do not focus on a few loci only but rather depict the effect of selection over the whole genome (Storz, 2005). However their main drawback is the impossibility to link outlier loci with specific any selection pressure (environmental for example).

In this communication, we propose to broach the issue from the environmental angle, in order to complement results obtained by population genetic models. We applied a recently developed Spatial Analysis Method (SAM) (Joost *et al.*, submitted to Molecular Ecology) to detect signatures of natural selection in sheep genotyped with AFLP markers. The analysed breeds were sampled in the context of the Econogene project, whose goal was to address the conservation of sheep and goat genetic resources in marginal agrosystems in Europe (<http://www.econogene.eu>).

MATERIAL AND METHODS – *Sampling:* a total of 624 sheep representing 40 breeds were sampled in farms located in 10 European countries (http://www.econogene.eu/list_of_breeds.html). As several animals (up to 3) could be sampled in the same farm, 577 distinct locations were characterized by environmental parameters.

Molecular analysis: DNA extraction was performed using GenElute Mammalian Genomic DNA kit (Sigma). EcoRI/TaqI AFLP molecular markers were produced according to the protocol described in Ajmone Marsan *et al.* (1997). Three highly polymorphic primer pairs carrying ACA/AAC, ACA/ACT and ATG/AAC as selective nucleotides were assayed and the AFLP polymorphisms were visually scored as dominant markers, coding with 1 the presence and with 0 the absence of the band.

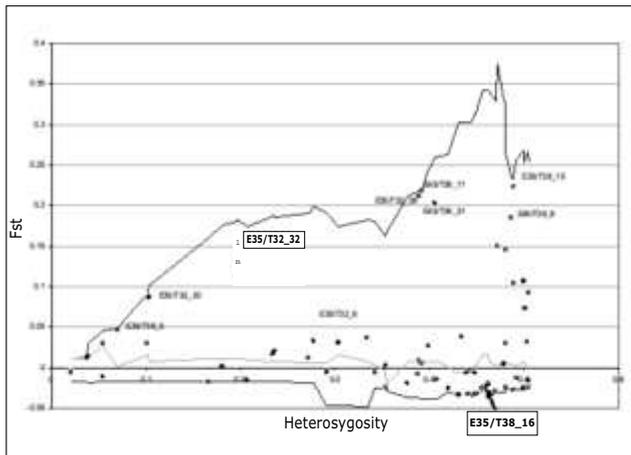
Environmental data: the environmental parameters considered were i) elevation, calculated with the 30 arc second digital elevation model of the Shuttle Radar Topography Mission (SRTM30, NASA, <http://www2.jpl.nasa.gov/srtm/>), and ii) climatology, produced by the Climatic Research Unit of Norwich (CRU) (see description in New *et al.*, 2002).

Spatial Analysis Method (SAM): to connect genetic information with geo-environmental data, we resorted to the SAM (Joost *et al.*, submitted to Molecular Ecology) which refers to the spatial coincidence concept. This consists in associating information levels and to compare them thanks to their common geographical coordinates. The SAM requires a geo-referenced data set made up of one or more environmental variables describing the sampling location (for instance mean monthly precipitations in millimetres), and a geo-referenced data set containing the genotyping of the animals investigated. Logistic regression is used to provide a measure of the association level between the frequency of molecular markers and the environmental parameters. The significance of the models constituted by all possible [marker x environmental variable] pairs is assessed, and the markers involved in the most significant models are deemed to be possibly implied in adaptation processes. The significance of the models constituted by all possible “marker x environmental variable“ pairs is assessed by the Wald and the Likelihood ratio (G) statistical tests, and the markers involved in the most significant models are deemed to be possibly implied in adaptation processes.

To assess the results provided by the SAM, the AFLP data set was also analyzed with the Dfdist software, modified from Fdist (Beaumont and Nichols, 1996). The Fdist method is a Fst-outlier based test that resort to computer simulations to model the behaviour of neutral loci under a symmetrical island model of population structure (Wright, 1951). It is based on the principle that genetic differentiation between populations is expected to be higher for loci under selection (that behave as outliers) than for the rest of the genome, considered as neutral.

RESULTS AND CONCLUSIONS – The 3 AFLP primer pairs produced 92 polymorphisms. Markers with overall frequency higher than 95% or lower than 5% were considered uninformative and removed from the analysis. Thus calculations were made for 62 AFLP markers. SAM identified two AFLP candidate-markers possibly involved in adaptation processes. The best candidate, E35/T32_32, is associated with precipitations and with the coefficient of variation of precipitations (significance level of 1.37E-13 for both Wald and G tests, equivalent to a confidence level of 99.9999999). On the other hand, the marker E35/T38_16 is associated with the number of wet days (significance level of 1.37E-11).

Figure 1. Outlier loci detection with Dfdist: Spanish Manchega vs. Welsh Mountain breed. Lines delimit the 95% confidence interval. Black dots and squared marker names identify the position of AFLPs resulted significant by the SAM method.



To compare these results with the theoretical approach in population genetics, and on the basis of the results obtained with the SAM, we contrasted pairs of breeds living in opposite extreme environments (dry vs. humid). Among the contrasted pairs (Spanish Manchega vs. Welsh Mountain, Spanish Manchega vs. Albanian Shkodrane), it was not possible to detect any outlier locus with a 95% confidence level, although both markers appear close to the significant thresholds (Figure 1). In conclusion, SAM revealed two highly significant markers deserving further investigation, to identify their map position and seek for nearby markers for confirming the effect detected and nearby candidate genes.

The authors would like to thank Aurélie Bonin for her support about the population genomics approach.

S. Joost was funded by the European Commission (Quality of Life, Contract QLK5-CT-2001 02461, "ECONO-GENE")

REFERENCES – **Ajmone-Marsan**, P., Valentini, A., Cassandro, M., Vecchiotti-Antaldi, G., Bertoni, G., Kuiper, M., 1997. AFLP (TM) markers for DNA fingerprinting in cattle, *Anim Genet.*, 28 (6):418-426. **Beaumont**, M.A., Nichols, R.A., 1996. Evaluating loci for use in the genetic analysis of population structure. *Proc Biol Sci.*, 263:1619-1626. **Black**, W.C., Baer, C.F., Antolin, M.F., DuTeau, N.M., 2001. Population genomics: genome-wide sampling of insect populations. *Ann Rev Entomol.*, 46:441-469. **Erickson**, D.L., Fenster, C.B., Stenoien, H.K., Price, D., 2004. Quantitative trait locus analyses and the study of evolutionary process. *Mol Ecol.*, 13:2505-2522. **Luikart**, G., England, P.R., Tallmon, D., Jordan, S., Taberlet, P., 2003. The power and promise of population genomics: From genotyping to genome typing. *Nature Rev Genet.*, 4:981-994. **New**, M., Lister, D., Hulme, M., Makin, I., 2002. A high-resolution data set of surface climate over global land areas. *Climate Res.*, 21:1-25. **Nielsen**, R., 2005. Molecular Signatures of Natural Selection. *Ann Rev Genet.*, 39:197-218. **Phillips**, P.C., 2005. Testing hypotheses regarding the genetics of adaptation. *Genetica*, 123:15-24. **Storz**, J.F., 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol.*, 14:671-688. **Vasemägi**, A., Primmer, C.R., 2005. Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Mol Ecol.*, 14:3623-3642. **Wright**, S., 1951. The Genetical Structure of Populations. *Ann Eugenics*, 15: 323-354.