Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Insights into the interactions of goat breeds and their environments

Alessandra Stella[1,2] on behalf of the NEXTGEN and ADAPTmap consortia

[1]Fondazione Parco Tecnologico Padano, Lodi, Italy
[2]IBBA-CNR, Lodi, Italy

ABSTRACT: More than 900 million goats are bred across the world, 95% of which are in developing countries, particularly Africa. Goat breeds are raised successfully in a wide variety of environments making the goat species an optimal candidate to disentangle the genetics of adaptation. The EU-funded NEXTGEN project and the goat AdaptMap have been undertaken to study this question. . The NEXTGEN strategy is based on a comparative analysis of whole genome data at the intraspecific level to optimize genetic management of livestock diversity. A bioinformatics pipeline has been developed to take advantage of whole genome sequences that are produced at a 10X coverage. Adaptation in sheep and goats is investigated by studying for each species 164 individuals sampled in places representative of the contrasted environmental conditions found all over Morocco. The ADAPTmap project (http://www.goatadaptmap.org) is an International effort launched in collaboration with the International Goat Genome Consortium (IGGC) (http://www.goatgenome.org), the African Goat Improvement Network (AGIN) - Feed the Future program of the USAID (http://www.feedthefuture.gov) and NEXTGEN to improve coordination among otherwise independent projects for genotyping and re-sequencing of goat breeds.
Keywords: Goat; Adaptation; Landscape genomics

Introduction

Goats as a model for adaptation. More than 900 million goats are bred across the world. Of the approximately 600 different breeds that have been reported, 557 have been classified as local, while 47 are transboundary (source: FAO).

The species is unique for integration into a large variety of farming systems, often in complementarity with other livestock species. Their hardiness, resistance/tolerance to disease and productivity on low quality diets are generally superior to other livestock species. Attractiveness to small-holders can be summarized in the following characteristics: 1) Relatively low production requirements in terms of land, infrastructure and capital investments, as well as the required husbandry skills; 2) a wide range of marketable products (meat, milk, fiber, hides, dung, offspring) and economic roles (wealth and resource maintenance); and 3) a short reproduction cycle and multiple offspring, allowing for faster return on investment.

The fact that goat are raised successfully in a wide variety of environments makes the species an optimal candidate to disentangle the genetics of adaptation.

The International Goat Genome Consortium (IGGC; http://www.goatgenome.org) has produced a reference genome for goats (Dong et al, 2013). Moreover, a SNP chip, featuring 52k markers, has been made available by the Illumina company (Tosser et al., 2014). The panel was initially tested with 288 goat DNA samples from 10 different breeds; currently, more than 10,000 goats have been genotyped with this chip worldwide.

Methods to detect genomic regions relevant to adaptations. SNP panels open up new perspectives to livestock genetics, in particular for the investigation of genome diversity within and among individuals and populations, population structure and inbreeding, searching for QTL controlling complex traits, performing genome-wide marker-enhanced selection of young animals, and identifying patterns of recent and past selection. This last application provides an attractive prospect for the identification of genomic regions influencing traits that are very difficult, expensive and even impossible to record experimentally, such as adaptation to extreme climates and poor fodder and resistance to disease. Such traits may become very valuable in the future, given their link to sustainability and be of key importance in a time of rapid and unpredictable change in climate. This aspect is particularly relevant in Africa, where water availability is decreasing, pasture growing seasons are shortening, and climate change is leading to the expansion of the territory of vectors and the spread tropical disease challenges outside their endemic area.

To study adaptation, phenotype-based and selective-sweep based approaches may be applied simultaneously to the same animals and reinforce each other. Selection signatures specific to adaptation have been identified in humans (e.g. Lappalainen et al., 2010) and in wild organisms (e.g. Poncet et al., 2010). but their discovery will likely be more useful in livestock species, in which they might very well become targets of selection. This highlights the need for strengthening the collaboration among scientific communities on the use of emerging technologies and new analytical approaches (Joost et al., 2011).

Traditionally, molecular genetics in conservation biology of livestock species has been confined to the analysis of neutral molecular markers such as mitochondrial DNA and microsatellites and, more recently, to dense SNP panels. However, it has always been acknowledged that

such approaches have neglected the important component of genetic variation in threatened populations that underpins their local adaptation. Recently, new methods have become available to assay adaptive variation in the genome of threatened populations, enabling the application of prioritization protocols to use unique adaptive variants as well as neutral, demographically mediated variation and even to test the association of this variation with environmental variables to identify geographic regions of priority (e.g. Bonin et al. 2007; Joost et al. 2007, 2011). Traditional conservation approaches have focused on the use of allele frequency-based differences to identify parsimonious sets of breeds whose combined diversity maximizes the genetic variation to be conserved and where additional breeds are considered for inclusion on the basis of "marginal diversity". This approach has, however, been criticized for not paying enough attention to breed viability and within-breed diversity (reviewed in Toro et al. 2009) and does not currently incorporate adaptive genetic variation in priority estimation – a limitation given the importance of local adaptation (both natural and artificially gained) in marginal livestock populations across the world.

A different approach to identify genomic regions associated to environmental variables is based on the combined use of genomics and GIScience. GIScience permits one to depict, explore and compare variables according to their geographic coordinates. This allows the detection and description of spatial synchrony, identification of data combinations associated with effects specific to a geographic area, calculation of synthetic indicators and, most importantly in this case, of hidden relationships between variables (Joost et al., 2010).The basic methodology to simultaneously assess large numbers of environmental and genetic variables using the Spatial Analysis Method (SAM) was developed by Joost et al. in 2007. The method uses high information content regression models optimized for genetic and environmental data and which is efficient enough to handle large numbers of variables.

In addition, a complementary environmental genomics approach has been taken to develop a prioritization tool for conservation biologists based on analysis of locally adapted genetic variants and their uniqueness in a set of populations (Bonin et al. 2007). The principle of complementarity is used to maximize the number of variants conserved by assessing their relative distribution in populations under conservation consideration.

The NEXTGEN and Adaptmap projects. NEXTGEN (http://nextgen.epfl.ch) is an ongoing EU FP7 project based on a comparative analysis of whole genome data at the intra-specific level to optimize genetic management of livestock diversity.

The NEXTGEN project took an explicitly spatial approach to examine geographic and environmental drivers of genomic adaptation in small ruminant populations. The goal was to detect signatures of selection without any prior information on genomic regions and traits under selection. Thus, rather than carrying out a candidate-gene approach focusing on a set of genes previously known to be involved in adaptation or selection, we scanned the genome looking for regions (and the SNPs therein) that show marked deviations from the expected distribution under neutrality, as it is expected for natural populations that have been under selective pressure in the past.

Many methods for detecting selection using molecular data, especially genomic DNA sequence and SNP data have recently been developed and applied. NEXTGEN evaluated the potential of these methods to take into account the specificity of our study and thus to be used on our dataset. A number of international projects aim to investigate diversity of goat breeds/populations by recording and relating genomic diversity, morphological traits and several geo-climatic parameters for characterizing the breeding environment. In addition, large projects have been carried out for genomic analyses of complex traits, providing extensive datasets on commercial populations (e.g. http://www.3srbreeding.eu). Genomic analyses typically comprise extensive genotyping with the Illumina 50k SNP panel. However, re-sequencing and genotyping by sequencing techniques will be applied on some of the same local and commercial populations and de-novo sequencing approaches will be applied to wild ancestors.

Integrating data from the various sources is an essential step to allow a comprehensive application of population genomics.

The ADAPTmap project (http://www.goatadaptmap.org) is an International effort started in collaboration with the International Goat Genome Consortium (IGGC) (http://www.goatgenome.org), the African Goat Improvement Network (AGIN) - Feed the Future program of the USAID (http://www.feedthefuture.gov) and the NEXTGEN project to improve coordination among otherwise independent projects for genotyping and re-sequencing of goat breeds.

Materials and Methods

Within NEXTGEN, populations of domestic sheep (*O. aries*) and goats (*C. hircus*) and their wild counterparts (mouflons: *O. orientalis*; and bezoars: *C. aegagrus*) were sequenced and genotyped (Table 1).

Table 1. A summary of the sequencing and genotyping undertaken in NEXTGEN.

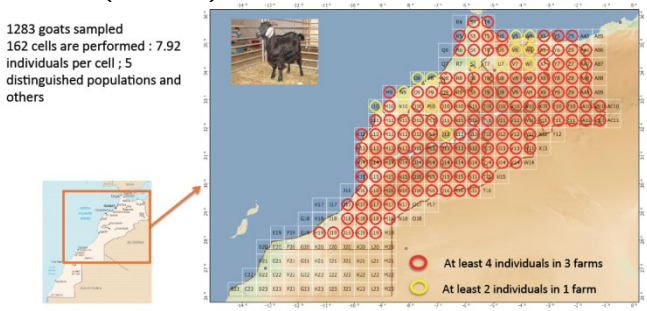
Country	Species	Animals Sequenced	Animals with SNP genotypes
Morocco	Sheep	164	30
	Goats	164	30
Iran	Mouflon	15	8
	Sheep	20	18
	Bezoar	23	9
	Goats	20	9

Morocco was chosen for the study for its extremely heterogeneous landscape, representing one of the widest possible ranges of geo-climatic conditions in such a small area. A grid system (Figure 1) was developed to guide sampling and for application of a landscape genomics

approach. A total of 1283 goats were sampled. A regular grid over the Northern part of Morocco divided the area into cells (squares of 0.50°of latitude and longitude). Partners at the Laboratory for Geographical Information Systems at EPFL (Lausanne, CH), used geo-climatic parameters for seeking the maximum spreading of environmental factors throughout the collected samples and identified 164 goats (and 164 sheep) to sequence.

Adaptation events are, in fact, studied through a spatially explicit analysis of genome diversity that is a powerful alternative to linkage or association studies for identifying genomic regions associated with key traits for sustainable breeding.

Figure 1. The grid sampling approach applied in Morocco (n = 1283).

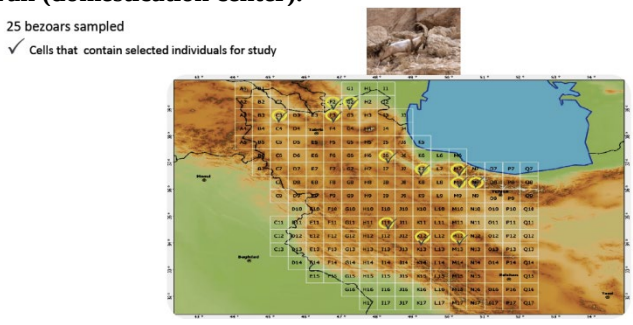


Moreover, for sheep and goats, an evaluation of the potential of wild ancestors and traditional breeds to act as reservoirs of neutral and adaptive genetic diversity was conducted, by analyzing the genomes of almost 80 individuals (Asiatic mouflons, bezoars and local Iranian domestic sheep and goats), sampled in Iran, accepted to be the domestication center for these species (Figure 2).

A bioinformatics pipeline was then developed to take advantage of whole genome sequences of the whole set of samples, produced at a 10X coverage.

The European Bioinformatics Institute (EBI), partner of NEXTGEN, developed and applied a pipeline for variant calling from the produced sequenced.

Figure 2. The sampling structure for goats (n = 65) in Iran (domestication center).



Several million (~30 million) single nucleotide polymorphisms (SNP) were derived from sequence data. Together with genotyped polymorphisms, these data were used in a coordinated effort to scan the genome for genetic signals of domestication and adaptation to the environment.

A number of statistical methods were used to scan the genome for signals of domestication and adaptation. Applied methods can be roughly grouped in single-locus methods and haplotype-based methods. Average heterozygosity at each SNP locus (e.g. Rubin et al., 2010) and the fixation index Fst (Nei, 1977) belong to the first group. As for haplotype-based methods, EHH/iHS (Sabeti et al., 2002; Voight et al., 2006), SweeD (Pavlos et al., 2013), XPCLR (Chen et al., 2010) and HapFLK (Fariello et al., 2013) were used in this study.

Signals detected with the different statistical methods were critically evaluated. Consensus signals were retained as signals of high credibility. Signals detected only by a subsets of methods were however not discarded, but interpreted in the context of the different properties of specific methods (e.g. recent vs old selection signatures; selection signatures at intermediate vs extreme frequencies).

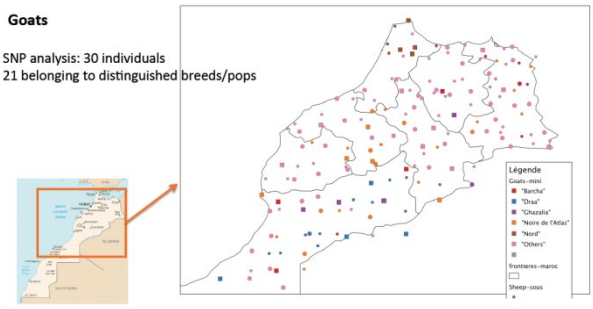
Candidate genes identified with the genome scan were later analyzed with pathway analysis (e.g. Ingenuity).

Northern Iran is the putative domestication centre for both domestic sheep and goats; animals from Morocco came from very diverse geographic and climatic regions (coast, mountains, desert). Contrasting comparisons between wild and domestic animals and among animals from different environments allowed for the detection of signals of, respectively, domestication and adaptation.

Results and Discussion

A preliminary analysis of the collected samples was based on their spatial distribution and, consequently, on the geo-climatic parameters defining their respective production environments. The analysis supported the choice of individuals for sequencing and genotyping.

Figure 3. The spatial distribution for Moroccan samples subject to genotyping with the 50k SNP chip.



De-novo sequencing of the wild ancestors and re-sequencing was performed at the sequencing production laboratory at CEA- Genoscope (France), using Illumina 2000 technology.

The SNP calling strategy that EBI developed is designed to minimize the ascertainment bias, especially when comparing wild and domestic genomes. Also, it accounts for the confounding effect of demographic history by running methods analyzing population structure (BAPS, Corander and Martinen 2006; TESS, Chen et al. 2007) in

order to detect groups of individuals suitable for further analyses.

De-novo assembly was performed for the wild ancestors. In particular, Table 2 has the summary for sequence features for the wild ancestor of goats, Capra aegagrus.

Table 2.Sequencing statitstics for the Bezoar.

Parameter	Value
Number of contigs	102,000
Total length of contigs	2.45 Gb
Contig N50	52.0 Kb
Number of scaffolds	6,676
Total length of scaffolds	2.58 Gb
Scaffold N50	1.75Mb
Ambiguities/10,000 bases	17.92

To define an optimal set of SNPs to study population genomic in small ruminants, an assessment the performance of the 50k SNP chip as a surrogate genome data source was performed. SNP chip data were compared with the whole genome sequence data for assessing genetic diversity and structure in domestic and wild small ruminants. Results showed how the sample size (10-30 individuals) and the number of SNPs (1000 - 5 Million and WGS) affected the estimation of population genomic parameters and the detection of signatures of selection in small ruminants (*O. aries*, C. hircus and their wild ancestors *O. orientalis* and *C. aegagrus*). We found that 1K SNPs was not sufficient for such estimation but random 10K SNPs allowed a good estimation of heterozygosity, inbreeding coefficient and nucleotide diversity. We showed also that a reliable estimation of LD required at least 500K SNPs. We demonstrated that the Ovis exome capture and commercial 50K-SNP Chips (Ovine and Caprine Illumina® SNP50 Beadchips designed to describe the diversity of industrial breeds) biased the estimations when studying traditional breeds and wild animals.

To implement the analysis of local adaptation to environmental conditions in Morocco, LASIG implemented methods based on the correlation of allele presence and environmental variables (SAM).

In the context of NEXTGEN, the challenge was to implement software able to process large amounts of genetic data, as sequencing identified a very large number of variants. Therefore, LASIG developed an open source toolbox dedicated to landscape genomics (Samβada). Samβada uses logistic regressions to estimate the probability that an individual carries a specific genetic marker given the habitat that characterizes its sampling site. The genetic data were recorded as binary variables and their association to the topo-climatic data was assessed with log-likelihood ratio and/or Wald tests. Models were ranked according to their scores to ease post-processing analyses.

Large SNP panels and whole-genome sequences often require sharing the computational load. When requested, Samβada splits the molecular data to distribute processing and merges the results subsequently.

While global regression models assess the overall relationships in the data, spatial patterns of associations give information about local processes at work. Sam β ada can measure the level of spatial autocorrelation in both molecular and environmental datasets using local and global Moran's I.

Results from the selection signature search identified several regions subjected to domestication and/or adaptation. For example, Figures 4 and 5 represent the iHS scores plotted along chromosome 26 for domestic goat and wild bezoars, respectively. Relatively high values indicate a longer homozygous haplotype around the ancestral or derived allele.

Figure 4. Integrated Haplotype Score (iHS) along chromosome 26 for domestic goats.

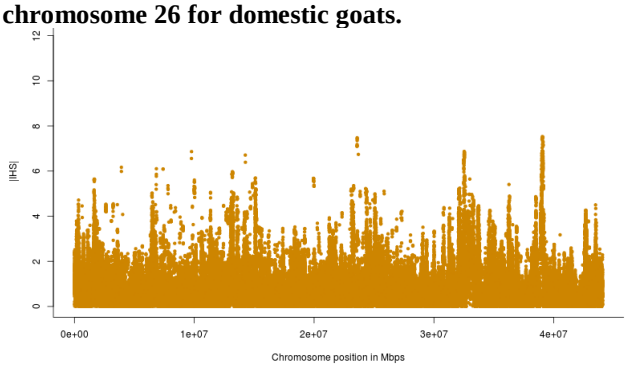
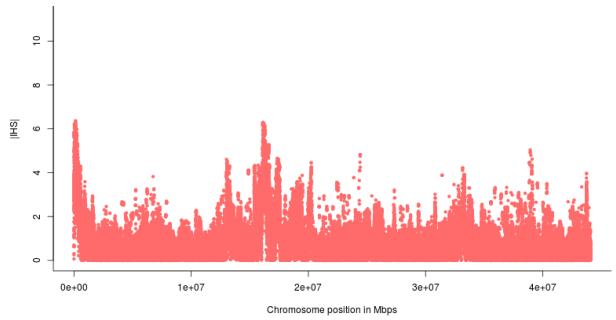


Figure 5. Integrated Haplotype Score (iHS) along chromosome 26 for wild bezoars.



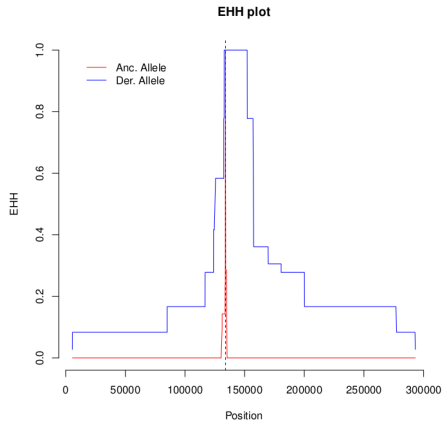
Finally, Figure 6 has the EHH plot around SNP 882. SNP 882 had the highest iHS in *O. orientalis*. The EHH plot shows the decay of homozygosity around the core SNP for the ancestral and derived alleles. A considerable difference between the area under the EHH curve substantiates the finding of a selection/adaptation signal at that locus.

ADAPTMAP database. A database structure was developed to store and interrogate genomic data (SNP genotypes), phenotypic descriptors and GIS information.

A web interface was constructed to allow uploading and query of the database. Genomic and phenotypic data can be transferred to the centralized database using a multi-file web uploader; SNPage software was developed to easily handle and safely merge SNP genomic information. SNPage is a buffer program able to transform multiple Illumina formats into a single format to enter SNPage. SNPage allows to include multiple batches using just one tool. A specific tool designed for AdaptMap

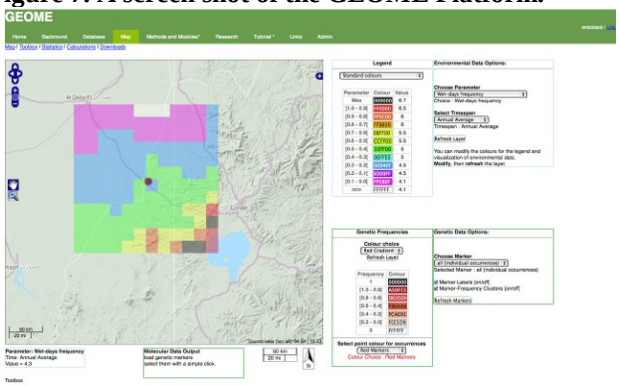
performs a few simple quality controls of the genomic data uploaded by the user. A graphical plugin was designed in order to use SNPage output file (e.g. genotypes and SNPnames) and transform it into formats suited for Samβada as well as for other most used population analysis software.

Figure 6. The Extended Haplotype Homozygosity (EHH) plot around SNP 882.



Tools for integration of geo-climatic information were built in the database. In order to simplify access to the data for AdaptMap users, colleagues at LASIG kindly agreed to integrate the AdaptMap searches within the GEOME Platform (Figure 7), a WebGIS-based platform for the integrated analysis of environmental, ecological and molecular data through the implementation of an original set of combined geocomputation, databases, spatial analysis and population genetics tools (http://lasigpc8.epfl.ch/geome/) .

Figure 7. A screen shot of the GEOME Platform.

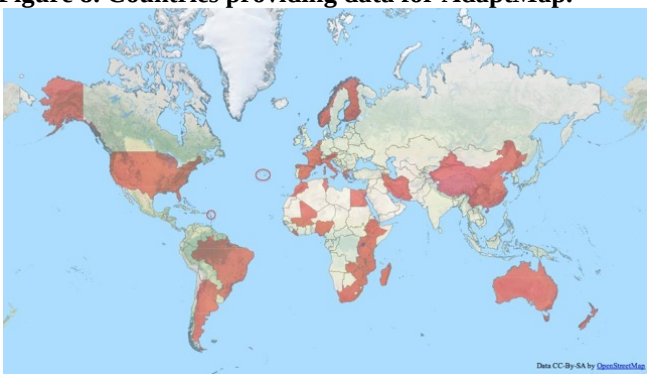


Currently, information has been so far collected from 14 projects (see Figure 8 for the distribution). The number of sampled breeds ranges from 3 to 15 per study and represent 31 countries and a wide variability of environments and production systems. Most breeds are autochthonous (80%) and mostly bred for meat production (63%).

Conclusions

Goats are an optimal model to study domestication and adaptation. Collection of descriptors for production environments together with genomic information enhances the understanding of the domestication and adaptation processes, allowing to disentangle the genotype by environment interaction at relevant loci.

Figure 8. Countries providing data for AdaptMap.



Acknowledgments

The NEXTGEN project was funded by the EU Framework Programme 7.

Literature Cited

Bonin A., Nicole F., Pompanon F., Miaud C., Taberlet P. (2007) Conserv Biol. 21, 697-708.

Chen C, Forbes F, François O (2006) Molecular Ecology Notes, 6, 980 –983.

Chen H, Patterson N, Reich D. (2010) Genome Res. 20:393–402.

Corander and Martinen (2006) Corander J, Marttinen P. Mol Ecol:15(10):2833-43

Dong Y, Xie M, Jiang Y, Xiao N, et al. (2013) Nat Biotechnol. 31(2):135-41.

Fariello M., S. Boitard, H. Naya,M. SanCristobal and B. Servin. (2013) Genetics, 193(3):929-941.

Joost, S., Bonin, A., Bruford, M.W., Després, L., Conord, C., Erhardt, G., Taberlet, P. (2007) Mol. Ecol. 16: 3955–3969.

Joost, S., Colli, L., Baret, P.V., Garcia, J.F., Boettcher, P.J., Tixier-Boichard, M., Ajmone-Marsan, P., the GLOBALDIV Consortium, (2010) Anim. Genet. 41(s1):47-63.

Joost, S., Colli, L., Bonin, A., Biebach, I., Allendorf, F., Hoffmann, I., Hanotte, O., Taberlet, P., Bruford, M. and the GLOBALDIV Consortium, (2011) Conserv Genet Resour. 3:785-788.

Lappalainen, T., Salmela, E., Andersen, P.M., Dahlman-Wright, K., Sistonen, P., Savontaus, M.L., Schreiber, S., Lahermo, P., Kere, J.. (2010) Eur. J. Hum. Genet. 18, 471-478.

McKay, S.D., Schnabel, R.D., Murdoch, B.M., Matukumalli, L.K., Aerts, J., Coppieters, W., Crews, D., Dias Neto, E., Gill, C.A., Gao, C., Mannen, H., Wang, Z., Van Tassell, C.P., Williams, J.L., Taylor, J.F., Moore, S.S. (2008) BMC Genet. 9, 37.

Nei, M. (1977)Annals of human genetics 41(2), 225–233.

Pavlos P., D. Živković, A. Stamatakis and N. Alachiotis. (2013) Molecular biology and evolution, 30(9):2224-2234.

Poncet, B.N., Herrmann, D., Gugerli, F., Taberlet, P., Holderegger, R., Gielly, L., Rioux, D., Thuiller, W., Aubert, S., Manel, S. (2010) Mol. Ecol. 19, 2896-2907.

Rubin C, Zody M, Eriksson M, Meadows JR, et al. (2010) Nature, 464(7288):587–591

Sabeti PC, Reich DE, Higgins JM, Levine, et al. (2002) Nature, 419 :832–837.

Tosser-Klopp G, Bardou P, Bouchez O, et al. International Goat Genome Consortium. (2014) PLoS One Jan 22;9(1):e86227.

Voight b., S. Kudaravalli, X. Wen, and J. K. Pritchard. (2006) PLoS biology, 4(3):e72.