



Genomic bases of local adaptation in sheep and goats

Benjelloun B., Pompanon F., Ibnelbachyr M.

in

López-Francos A. (ed.), Jouven M. (ed.), Porqueddu C. (ed.), Ben Salem H. (ed.), Keli A. (ed.), Araba A. (ed.), Chentouf M. (ed.). Efficiency and resilience of forage resources and small ruminant production to cope with global challenges in Mediterranean areas

Zaragoza : CIHEAM Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 125

2021 pages 613-618

Article available on line / Article disponible en ligne à l'adresse :

http://om.ciheam.org/article.php?IDPDF=00008074

To cite this article / Pour citer cet article

Benjelloun B., Pompanon F., Ibnelbachyr M. **Genomic bases of local adaptation in sheep and goats.** In : López-Francos A. (ed.), Jouven M. (ed.), Porqueddu C. (ed.), Ben Salem H. (ed.), Keli A. (ed.), Araba A. (ed.), Chentouf M. (ed.). *Efficiency and resilience of forage resources and small ruminant production to cope with global challenges in Mediterranean areas.* Zaragoza : CIHEAM, 2021. p. 613-618 (Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 125)



http://www.ciheam.org/ http://om.ciheam.org/



Genomic bases of local adaptation in sheep and goats

B. Benjelloun^{1,2}, F. Pompanon² and M. Ibnelbachyr³

¹Institut National de la Recherche Agronomique Centre de Tadla, BP 567, Béni Mellal (Morocco) ²Laboratoire d'Ecologie Alpine, UMR 5533 CNRS-UGA Grenoble (France) ³Institut National de la Recherche Agronomique, CRRA de Errachidia (Morocco)

Abstract. Since their domestication 10.000 years ago followed by a worldwide spread, sheep and goats have accumulated highly valuable adaptive traits allowing them to be raised across highly diverse environments. Besides the current rapid development and wide spread of just a few productive cosmopolitan breeds marked by limited genetic diversity, indigenous populations may keep adaptive traits that would constitute key genomic resources in the context of environmental changes. Within the EU-FP7 funded NextGen project, we applied a landscape genomics framework to identify genes involved in local adaptation. We sequenced the genomes of 160 indigenous sheep and 161 goats representative of the Moroccan-wide diversity in ecology, climate and geographic origin. We detected 39 million variants in sheep and 32 million in goats showing a very weak geographic structure over the country in both species. Using different approaches we identified several sets of loci and genes that likely have a role in local adaptation globally to altitude, slope, sunshine, rainfall, temperature and their variation. The main adaptive pathways were associated with respiration and circulation for the adaptation to altitude. The major genes identified showed different patterns of variation of allele frequencies along environmental gradients. Candidate genes for adaptation to the same environmental variable were generally different between the two species, suggesting different adaptive mechanisms in sheep and goats. However, similar or functionally linked genes responding to the same environmental variable were also found. Results from this study would help in setting up efficient plans of genomic selection based on adaptive traits and be then of major interest for a sustainable management of genetic resources in sheep and goats.

Keywords. Sheep - Goats - Local breeds - Adaptation - Genomics.

Bases génomiques de l'adaptation locale chez les ovi-caprinés

Résumé. Depuis leur domestication il y a 10 000 years, suivis d'une propagation dans le monde entier, les ovins et les caprins ont accumulé des traits adaptatifs très précieux leur permettant d'être élevés dans des environnements très diversifiés. En dépit du développement rapide actuel et la large diffusion de quelques races cosmopolites productives caractérisées par une diversité génétique limitée, les populations autochtones peuvent conserver des caractéristiques d'adaptation qui constitueraient des ressources génomiques essentielles dans le contexte des changements environnementaux. Dans le cadre du projet de l'UE FP7 NextGen, nous avons appliqué un cadre de génomique du paysage pour identifier les gènes impliqués dans l'adaptation locale. Nous avons séquencé les génomes de 160 moutons et 161 chèvres indigènes, représentatifs de la diversité écologique, climatique et géographique de l'ensemble du Maroc. Nous avons détecté 39 millions de variantes chez le mouton et 32 millions chez la chèvre, présentant une structure géographique très faible sur le pays chez les deux espèces. En utilisant différentes approches, nous avons identifié plusieurs ensembles de locus et de gènes qui jouent probablement un rôle dans l'adaptation locale à l'altitude, à la pente, à l'ensoleillement, aux précipitations, à la température et à leurs variations. Les principales voies d'adaptation étaient associées à la respiration et à la circulation pour l'adaptation à l'altitude. Les gènes candidats à l'adaptation à la même variable environnementale étaient généralement différents entre les deux espèces. Cependant, des gènes similaires répondant à la même variable environnementale ont également été trouvés. Les résultats de cette étude présenteraient un intérêt majeur pour une gestion durable des ressources génétiques chez les ovins et les caprins.

Mots-clés. Ovin – Caprins – Races locales – Adaptation – Génomique.

I – Introduction

Threats to biodiversity in farm animals are increasing, both in terms of extinction rates and destruction of agro-ecosystems, or the loss of genetic diversity. Indeed, the formulation of the modern breed concept during the mid-1800s (Porter 2002) and its application to breeding and its practices led to the formation of well-defined breeds and their exposure to strong artificial selection. The development of farming practices such as artificial insemination and embryo transfer, feeding techniques and the use of vaccines and treatments against endemic diseases have allowed the spread of industrial livestock farming. This has led breeders to gradually replace indigenous breeds, well adapted to local conditions, with highly productive cosmopolitan breeds (Taberlet et al., 2008). As a result, a significant number of sheep and goat breeds have become extinct and several other breeds are endangered (Altshuler et al., 2012). In a context of climate change, indigenous breeds that are well adapted to their farming conditions constitute a capital asset of adaptive alleles. Similarly, understanding the mechanisms involved in the processes of their adaptation could have major repercussions and practical applications for the management and preservation of livestock genetic resources. This paper aims to understand the process of adaptation of goats and sheep to different environmental conditions with innovative approaches of genomics. More specifically, it aims at identifying sets of loci putatively related to local adaptation in both species.

II – Materials and methods

1. Data production, processing and analysis

Following the wide sampling of sheep and goats in Morocco and the establishment of the representative sample bank of these species in Morocco (1412 sheep and 1283 goats), 160 sheep and 161 goats were chosen so as to have the widest distribution of animals in terms of geo-climatic variables, as explained below. These individuals were subjected to DNA extraction and sequencing of their entire genomes on an Illumina® Hi-Seq 2000 platform in order to have a minimum coverage rate of 10x. As described by Benjelloun *et al.* (2015), Illumina paired-end reads for sheep were mapped to the sheep reference genome (OAR v3.1, GenBank assembly GCA_000317765.1 (Jiang *et al.*, 2014)) and those for goats were mapped to the goat reference genome (CHIR v1.0, GenBank assembly GCA_000317765.1 (Dong *et al.*)) using BWA mem (Li and Durbin). The BAM files produced were then sorted using Picard SortSam and improved using Picard Markduplicates (http://picard.sourceforge.net), GATK RealignerTargetCreator, GATK IndelRealigner (DePristo *et al.*) and Samtools calmd (Li *et al.*, 2009). Variant calling was done using three different algorithms: Samtools mpileup, GATK UnifiedGenotyper (McKenna *et al.*, 2010) and Freebayes. Genotypes were improved and phased by Beagle 4 (Browning and Browning, 2013).

2. Neutral genetic diversity

Neutral genomic variation was calculated to evaluate the level of genetic diversity present in Moroccan sheep and goats. The total number of variants and the number of variants within each population were calculated. The level of nucleotide diversity (π) was calculated in each species and averaged over all of the biallelic and fully diploid variants for which all individuals had a called genotype using Vcftools (Danecek *et al.*). Pairwise linkage disequilibrium (LD) was assessed through the correlation coefficient (r2). It was estimated in 5 segments of 2Mb on different chromosomes (physical positions between 5 Mb and 7 Mb on chromosomes 6, 11, 16, 21 and 26). LD was estimated either by using the whole set of reliable variants or after discarding rare variants with a minor allele frequency (MAF) less than 0.05. For both estimations, r2 values between all pairs of biallelic variants (SNPs and indels) on the same segment were calculated using Vcftools. Inter-SNP

distances (kb) were binned into the following 7 classes: 0-0.2, 0.2-1, 1-2, 2-10, 10-30, 30-60 and 60-120 kb and observed pairwise LD was averaged for each inter-SNP distance class and used to draw LD decay.

3. Adaptive genomic variation

A genome scan method based on population genetics models was applied on our datasets. We worked on 7 variables representing various environmental categories, i.e. climatic variables temperature, precipitations and DEM-derived altitude and slope with their respective original resolutions. For each variable, two pools of 20 individuals were constituted, each representing one extreme of the gradient of variation of the variable. The XP-CLR method (Chen *et al.*, 2010) was then run to identify potential regions differentially selected in each extreme pool. It is a likelihood method for detecting selective sweeps that involves jointly modelling the multi-locus allele frequency differentiation between two populations. It is based on a reference population and an object one. This test was combined to a single-locus *Fst* test (Weir and Cockerham 1984) to detect selective sweeps identified by both approaches.

For each species we aimed at depicting the pattern of differentiation of the top candidate genes under selection across the environmental gradients. For that, for each environmental variable we ranked the 160 sheep (and 161 goats, respectively) according to the ranking of their geographic position on the environmental gradient considered. A sliding limit moving by steps of 10 individuals was applied to define 2 groups among which the *Fst* value (Weir and Cockerham 1984) was estimated based on the candidate variants associated to those genes. The minimum number of individuals per group was 20 and the maximum 140. Then, this allowed plotting the variation of the *Fst* value along the environmental gradient.

III – Results and discussion

1. Neutral genomic diversity

We mapped unambiguously 99.4% (\pm 0.1%) of sheep reads on the OAR v3.1 assembly and 98.9% (\pm 0.1%) of goat reads on the CHIR v1.0 assembly. 38,599,873 variants were successfully called in sheep, among which 38,278,356 were polymorphic. For goats, 31,743,850 variants were discovered in the total dataset among which 31,650,083 were polymorphic. The whole genome nucleotide diversity was 0.174 in sheep and 0.126 in goats. Linkage disequilibrium was assessed by the pairwise r^2 value between polymorphic sites in the studied genomic regions. Using the whole set of reliable variants, the genomic distance at which it decayed to less than 0.15 was 655 bp in sheep and 166 bp in goats. Moreover, r^2 decayed to less than 0.1 in 3.12 kb and 2.1 kb in sheep and goats respectively. When withdrawing rare variants (MAF<0.05), the average r^2 decayed to less than 0.2 in 3.6 kb in sheep and 5.8 kb in goats. It decayed to less than 0.15 in 4.4 kb in sheep and 8.1 kb in goats.

Sheep and goats displayed very large counts of genomic variants (38.6 M and 31.7 M respectively) enlarging substantially the worldwide catalogue of ovine and caprine variants. Sheep showed 6.9 million more variants than goats, with a higher nucleotide diversity that could be linked to a higher percentage of rare variants in goats. Linkage disequilibrium was slightly lower in goats. However, *LD* value is highly influenced by the percentage of rare variants and, when we removed them, sheep displayed even lower *LD*. This fact could also partly explain differences in heterozygosity and inbreeding coefficients between the two species. Generally, *LD* extents found here complete on one hand the findings of Benjelloun *et al.* (2015) who found a longer *LD* extent ($r_{0.15}^2$ =1.33kb using the whole set of variants and $r_{0.15}^2$ =12 kb when excluding rare variants) using a subset of our goat dataset. The difference in reported *LD* results from the fact that we used here many more animals

for that estimate. On the other hand, *LD* values reported here are shorter than all those reported on other domestic animals (i.e. horses, cattle, pigs) where it largely exceeds 10 kb for r^2 =0.20 (Villa-Angulo *et al.*, 2009; Wade *et al.*, 2009). Our results would illustrate a large effective population size and the effect of the very common extensive breeding systems favouring high gene flows among Moroccan sheep and goats and the absence until now of very strong selection pressure. This was particularly suitable for identifying selective sweeps likely associated to local adaptation.

2. Genomic bases of local adaptation

Combining the XP-CLR and *Fst* methods, we highlighted 5981 (\pm 746) different candidate variants and 141 (\pm 20) different candidate genes on average in each one of the 5 studied variables in sheep (Table 1) and 4930 (\pm 564) candidate variants and 214 (\pm 25) candidate genes in each extreme group for the 5 studied variables in goats (Table 1). In sheep, we identified 136 candidate genes related to altitude and 112 candidate genes were identified for rainfall in April. Similarly, 165 genes were identified for temperature annual range (bio7), 150 genes for mean temperature of July (temp7) and 144 genes were identified for slope (Table 1). Candidate genes in goats were 252 for altitude, 209 for rainfall in March (prec3), 201 for rainfall seasonality (bio15), 221 for mean temperature of July (temp7) and 185 genes for slope (Table 1).

	Mean ± SD	Altitude	Rainfall in March	Rainfall in April	Temp. annual range	Rainfall seasonality	Mean temp. July	Slope
Sheep								
No. of genes	141±20	136	-	112	165	_	150	144
No. of variants	5981±746	5436	-	6250	5811	_	7133	5275
Goats								
No. of genes	214±25	252	209	_	_	201	221	185
No. of variants	4930±564	5408	4963	-	_	5470	4719	4090

Table 1. Number of candidate genes and variants under positive selection detected in Moroccan she	eep
and goats in relation with 7 environmental variables	

The differentiation of the identified variants and genes along environmental gradients showed different patterns, generally with a highest differentiation close to one or both extremes of the gradient forming "U" or "S" shapes (Fig. 1).

Most of the identified genes were directly associated to biological processes related to the resistance to extreme levels of the environmental parameter to which they were linked. The case of genes *NFIB* and *GATA6* that were associated to altitude in goats are involved in Clara cell differentiation. This is consistent with the nature of these cells, which are epithelial on the luminal surface of airways of the mammalian lung. In addition to their secretory and xenobiotic roles, they are the progenitor cells in small pulmonary airways (Giangreco *et al.*, 2002). They were shown to be numerous and prominent with big apical caps in llama living at high altitude (Heath *et al.*, 1976). They presented also signs of pathological alteration and marks of their compensatory proliferation after exposure to hypoxia in rabbits (Uhlik *et al.*, 2005). Other processes identified in goats for altitude comprise circulatory mechanisms that are useful in hypoxia conditions. From another hand, genes associated with ATP biosynthetic processes that have been identified for adaptation of goats to slope is consistent with a higher need for synthesised energy in animals raised in steep slopes (mountainous areas) in comparison with moderate-slope goats.

In sheep, genes associated to altitude are mainly related to the regulation of leucocyte, lymphocyte and mononuclear proliferation. Indeed, leukocyte invasion into hypoxic tissues is well-known and circulating monocytes and/or mononuclear fibrocytes are recruited to the pulmonary circulation of chronically hypoxic animals. These cells play an important role to face the pulmonary hypertensive process in response to low-input oxygen conditions (Stenmark *et al.*, 2005). This suggests that regulation of leukocyte, lymphocyte and mononuclear proliferation would be implied in sheep adaptation to high-altitude and genes enriched in these categories (*CLCF1*, *TMIGD2*, *ZP4*, *TLR4*, *KITLG* and *EBI3*) may play a certain role in this adaptation through the mechanism cited above. Several other cases of genes are under study.



Fig. 1. Evolution of differentiation index (*Fst*) for a sliding limit along an altitudinal gradient in the eight top-score candidate genes identified in sheep.

IV – Conclusion

Our study used a landscape genomic framework to depict the genetic bases of local adaptation in farm animals. The 321 sheep and goat whole genome sequences, collected from a wide range of biotic and abiotic conditions, represent a unique resource for studying evolutionary processes. We identified several sets of candidate variants, genes and biological processes that are likely involved in local adaptation to various eco-climatic conditions. Several other genes and biological processes are under study. Therefore, this study showed the effect of local adaptation on genomes in two livestock species. This contributes to our understanding on how local adaptation could act and opened new horizons for a sustainable management of the worldwide sheep and goats.

References

- Altshuler DM Durbin RM Abecasis GR Bentley DR Chakravarti A Clark AG Donnelly P Eichler EE Flicek P Gabriel SB et al., 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422): 56-65.
- Chen H, Patterson N, Reich D, 2010. Population differentiation as a test for selective sweeps. *Genome Research* 20(3): 393-402.

- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST et al., 2011. The variant call format and VCFtools. *Bioinformatics* 27(15): 2156-2158.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M et al., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43(5): 491-+.
- Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, Tosser-Klopp G, Wang J, Yang S, Liang J et al., 2013. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nature Biotechnology 31(2): 135-141.
- Giangreco A, Reynolds SD, Stripp BR, 2002. Terminal bronchioles harbor a unique airway stem cell population that localizes to the bronchoalveolar duct junction. *American Journal of Pathology* 161(1): 173-182.
 Heath D, Smith P, Harris P, 1976. CLARA CELLS IN LLAMA. *Experimental Cell Biology* 44(2): 73-82.
- Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, Wu C, Muzny DM, Li Y, Zhang W et al., 2014. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344(6188): 1168-1173.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14): 1754-1760.
- Porter V, 2002. Mason's world dictionary of livestock breeds, types and varieties, Wallingford, UK.
- Stenmark KR, Davie NJ, Reeves JT, Frid MG, 2005. Hypoxia, leukocytes, and the pulmonary circulation. Journal of Applied Physiology 98(2): 715-721.
- Taberlet P, Valentini A, Rezaei HR, Naderi S, Pompanon F, Negrini R, Ajmone-Marsan P, 2008. Are cattle, sheep, and goats endangered species? *Molecular Ecology* 17(1): 275-284.
- Uhlik J, Konradova V, Vajner L, Adaskova J, 2005. Normobaric hypoxia induces mild damage to epithelium of terminal bronchioles in rabbits (ultrastructural study). *Veterinarni Medicina* 50(10): 432-438.
- Weir BS, Cockerham CC, 1984. Estimating f-statistics for the analysis of population-structure. *Evolution* 38(6): 1358-1370.