

Etude de l'adaptation des petits ruminants marocains aux conditions environnementales par une approche de génomique du paysage

POMPANON F. (1, 2), BENJELLOUN B. (1, 2, 3), LEEMPOEL K. (4), ALBERTO F.J. (1, 2), BOYER F. (1, 2), JOOST S. (4), TABERLET P. (1, 2), the NEXTGEN CONSORTIUM (5)

(1) Université Grenoble Alpes, Laboratoire d'Ecologie Alpine, 3800 Grenoble, France

(2) CNRS, Laboratoire d'Ecologie Alpine, 3800 Grenoble, France

(3) INRA Maroc, CRRRA Tadla, 23000 Beni-Mellal, Maroc

(4) Laboratory of Geographic Information Systems, School of Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Suisse

(5) <http://nextgen.epfl.ch>

RESUMÉ

Pendant la dizaine de milliers d'années qui a suivi leur domestication au Moyen Orient, la chèvre et le mouton se sont répandus en Europe, Asie et Afrique. Ils ont donné de nombreuses populations gérées de façon traditionnelle, qui se sont adaptées aux conditions locales. Avec l'émergence de la notion de race il y a environ 200 ans, les pressions de sélection ont fortement augmenté pour atteindre un niveau maximal chez des races industrielles devenues très performantes mais ayant perdu une grande part de leur diversité génétique originelle. Dans ce contexte, la diversité génétique encore présente chez les races indigènes à l'échelle mondiale représente une ressource importante, sous forme d'allèles impliqués dans des adaptations à des conditions environnementales particulières (sécheresse, résistances aux pathogènes, etc.). Ainsi, la caractérisation des gènes impliqués dans les adaptations locales est un enjeu majeur, notamment dans un contexte de changement climatique, où ces ressources génétiques joueront un rôle essentiel dans la capacité d'adaptation des espèces domestiques. Dans le cadre des projets européens Nextgen et ClimGen, nous développons une étude de génomique du paysage sur l'ensemble du territoire marocain, afin de caractériser les gènes impliqués dans les adaptations locales. Nous avons séquencé les génomes complets de 161 chèvres et 160 moutons issus de localités représentatives de la diversité des conditions climatiques, écologiques et des systèmes d'élevage. Tout d'abord, nous n'avons détecté qu'une très faible structuration spatiale de la diversité génétique sur l'ensemble de la zone échantillonnée (environ 400 000 km²). Ensuite, la recherche de régions chromosomiques impliquées dans des mécanismes adaptatifs a été menée en recherchant des signatures de sélections par deux types de méthodes : des approches corrélatives recherchant à mettre en évidence les allèles dont la fréquence varie le long de gradients environnementaux (SAMBADA), des approches populationnelles recherchant les gènes qui différencient deux groupes d'individus représentatifs de conditions environnementales contrastées (e.g., XP-CLR). Nous avons ainsi mis en évidence plusieurs gènes et voies métaboliques qui apparaissent comme de bons candidats participant à l'adaptation à différentes variables (altitude, température, humidité, ...). Il apparaît que les gènes présentant la plus forte réponse sélective à une même variable environnementale ne sont pas les mêmes selon l'espèce. Ces résultats montrent la richesse des ressources génomiques adaptatives présentes dans des populations indigènes qui doivent être préservées pour la sauvegarde à long terme des espèces.

A landscape genomics approach for studying the adaptation of small ruminants to their local environment

POMPANON F. (1, 2), BENJELLOUN B. (1, 2, 3), LEEMPOEL K. (4), ALBERTO F.J. (1, 2), BOYER F. (1, 2), JOOST S. (4), TABERLET P. (1, 2), the NEXTGEN CONSORTIUM (5)

(1) Université Grenoble Alpes, Laboratoire d'Ecologie Alpine, 3800 Grenoble, France

(2) CNRS, Laboratoire d'Ecologie Alpine, 3800 Grenoble, France

SUMMARY

Since their initial domestication in the Middle East c.a. 10,000 years ago, sheep and goats have spread over Europe, Asia and Africa, resulting in local populations traditionally managed that became locally adapted to their environment. The emergence of the breed concept c.a. 200 years ago lead to the strong increase of selective pressures in order to create successful breeds but this was associated with a loss of global genetic diversity. In this context the genetic diversity still present in indigenous breeds world-widely distributed constitutes a valuable resource consisting in alleles involved in adaptations to local environmental conditions (i.e., dry, resistance to pathogens, etc.). Thus, in the context of climate change, the characterization of genes involved in local adaptations is an important issue because these genetic resources would constitute a major part of the adaptive potential of domestic breeds. Within the framework of the EU projects Nextgen and ClimGen, we developed a landscape genomics approach all over Morocco, in order to characterize genes involved in local adaptation. We produced a whole genome sequence for 161 goats and 160 sheep from localities representative of the different eco-climatic conditions and management schemes found over Morocco. First, we found a very low spatial structure for the genetic diversity over the sampled area (around 400,000 km²). Second, we looked for chromosome regions bearing signals of selection using two kinds of methods: correlated approaches that find alleles whose frequency varies along environmental gradients (SAMBADA), and population-based approaches looking for genes that differentiate groups of individuals from

contrasted environments (e.g., XP-CLR). Several genes and metabolic pathways were found to be probably involved in the adaptation to different environmental variables (e.g., altitude, humidity, temperature,...). The genes showing the strongest response to a given environmental variable appeared to be different according to the species. These results show the high potential of adaptive genomic resources present in indigenous populations which deserve to be protected for the long term conservation of domestic species.

INTRODUCTION

Les chèvres et les moutons ont été les premiers animaux d'élevage domestiqués, il y a plus de 10 000 ans, en marge du croissant fertile. Depuis cette période, ils ont accompagné les mouvements des populations humaines, et ont colonisé la planète via les migrations et les échanges commerciaux (Taberlet et al., 2008). Durant dix millénaires, les populations locales de petits ruminants ont été élevées selon des méthodes traditionnelles, les éleveurs favorisant dans chaque contexte environnemental les reproducteurs avec les meilleurs phénotypes, tout en maintenant un certain niveau de flux géniques entre phénotypes. Ce mode de gestion durable a permis l'adaptation progressive des populations aux conditions environnementales locales (climat, pathogènes et parasites, nature et disponibilité des ressources alimentaires, etc.). La situation a brutalement changé il y a environ 200 ans avec l'apparition du concept de race, impliquant la fragmentation en groupes isolés d'individus fortement sélectionnés pour présenter des phénotypes similaires. Cette situation s'est amplifiée avec l'apparition de races dites industrielles, aux phénotypes standardisés et faibles tailles efficaces. Ces races connaissent un essor croissant alors qu'un nombre de plus en plus important de races indigènes disparaissent ou deviennent en danger d'extinction (Taberlet et al., 2008). Selon la FAO, 18 % des races locales de chèvres et 30 % des races locales de moutons sont éteintes ou en danger (Source FAOSTAT 2013). Pourtant, les races industrielles qui deviennent dominantes ont perdu une grande partie de leur variabilité génétique d'origine, alors que les races locales possèdent des adaptations qui pourraient s'avérer essentielles dans un contexte de changement climatique. Elles constituent donc des ressources génomiques que notre étude a contribué à caractériser. Nous avons séquencé les génomes complets de plus de 160 chèvres et 160 moutons représentant la diversité des conditions environnementales présentes au Maroc afin de caractériser des gènes impliqués dans l'adaptation à des variables environnementales.

1. MATERIEL ET METHODES

1.1. ECHANTILLONNAGE

L'échantillonnage a été effectué sur 161 chèvres et 160 moutons sélectionnés à partir de 412 points d'échantillonnage couvrant de façon homogène l'ensemble du Maroc (400 000 km²). Le Maroc est caractérisé par une grande diversité de climats (allant de Atlantique à désertique, en passant par différentes tendances de climats Méditerranéens et montagnards), et une large gamme d'altitudes (de 0 à 4000 m). Les sites sélectionnés représentent l'ensemble de ces conditions climatiques (maximisation de la distance climatique entre points) tout en maximisant la distance géographique entre les individus (Stucki et al. 2014a, voir Figure 1). Les échantillons prélevés consistent en des fragments d'oreille.

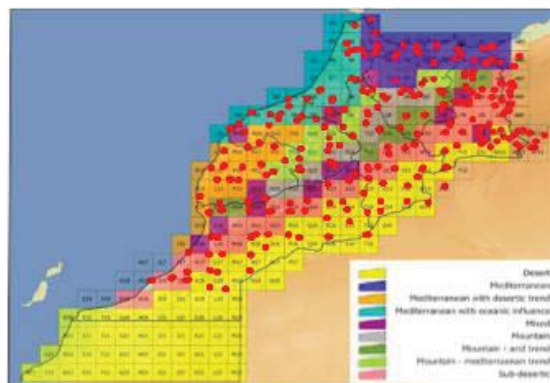


Figure 1 : Diversité des conditions climatiques au Maroc et grille d'échantillonnage. Les points rouges donnent la localisation des individus séquencés (cas de la chèvre).

1.2. PRODUCTION DES DONNEES DE GENOME COMPLET

Les extractions d'ADN ont été faites avec le kit *Puregene Tissue kit* de Qiagen® en suivant le protocole du fabricant. Des banques de fragments d'ADN de 300-600 pb ont été réalisées à partir de 500 ng d'ADN pour être séquencées sur un séquenceur *Illumina HiSeq2000* (Illumina®, USA) avec une couverture de 12 X selon le protocole décrit par Benjelloun et al. (2015). Les séquences *paired-end* obtenues ont été alignées sur les génomes de référence CHIR v1.0 et OAR v3.1, pour la chèvre et le mouton respectivement. L'identification, le filtrage et la caractérisation des sites variables (SNPs et courts indels) s'est faite selon le protocole décrit par Benjelloun et al. (2015).

1.3. ANALYSES

1.3.1. Caractérisation de la structure génétique neutre

La structure génétique du cheptel marocain a été caractérisée pour chaque espèce par une méthode de clustering (sNMF, Fricot et al., 2014).

1.3.2. Détection de signatures de sélection.

67 variables climatiques ont été extraites de la base WorldClim (Hijmans et al. 2005, <http://www.worldclim.org/current>), et 14 variables topographiques ont été extraites à partir d'un modèle DEM (Digital Elevation Model: SRTM <http://earthexplorer.usgs.gov>; U.S. Geological Survey) pour tous les points d'échantillonnage. Un grand nombre de variables climatiques étant corrélées, nous n'avons conservé que 31 et 27 des 81 variables pour les moutons et chèvres respectivement (Stucki et al. 2014a, Leempoel et al. 2015, Benjelloun 2015). Ces variables reflètent la température, les précipitations, la topographie et l'altitude.

Les signatures de sélection associées à des variations environnementales ont été recherchées par une approche corrélative (Samadina, Stucki et al., 2014b), basée sur la recherche d'une corrélation (selon une régression logistique) entre la présence/absence d'un allèle (cas de marqueurs bi-alléliques) et la variation d'un paramètre environnemental. Une approche populationnelle a également été menée sur 7 paramètres environnementaux, en contrastant les génomes des 20 individus présents à chaque extrême du gradient de variation de chaque paramètre. Les régions du génome différenciant les individus issus d'environnements contrastés ont été identifiées en fonction de (i) leur score XP-CLR (Chen et al., 2010) et (ii) des valeurs extrêmes de *Fst* (Weir et Cockerham, 1984) qu'elles présentaient (Benjelloun, 2015).

1.3.2. Analyses de Gene Ontology

L'identification des processus biologiques impliquant les gènes potentiellement sous sélection a été effectuée en utilisant GOrilla (Eden et al., 2009).

2. RESULTATS

2.1. STRUCTURATION GENETIQUE NEUTRE

Pour les chèvres comme pour les moutons, l'analyse sNMF suggère l'absence de structure génétique neutre significative sur l'ensemble du territoire marocain (K=1 groupe expliquant au mieux les données observées). Une faible structuration

géographique est cependant observée lorsque l'on considère la partition des génomes selon 3 groupes ancestraux.

2.2. GENES CANDIDATS A L'ADAPTATION

Les analyses XP-CLR ont mis en évidence de nombreuses régions chromosomiques potentiellement sous sélection (voir sur la Figure 2 les signaux trouvés chez le mouton en relation avec l'altitude). La sélection des régions ayant à la fois des scores XP-CLR et des valeurs de Fst parmi les 0,1 % les plus élevées du génome sont liées à plus d'une centaine de gènes chez les moutons et plus de 200 chez la chèvre, en fonction de la variable environnementale étudiée. En outre, plus de 60 % des SNPs identifiés sont intergéniques et moins de 1 % s'avèrent exoniques.

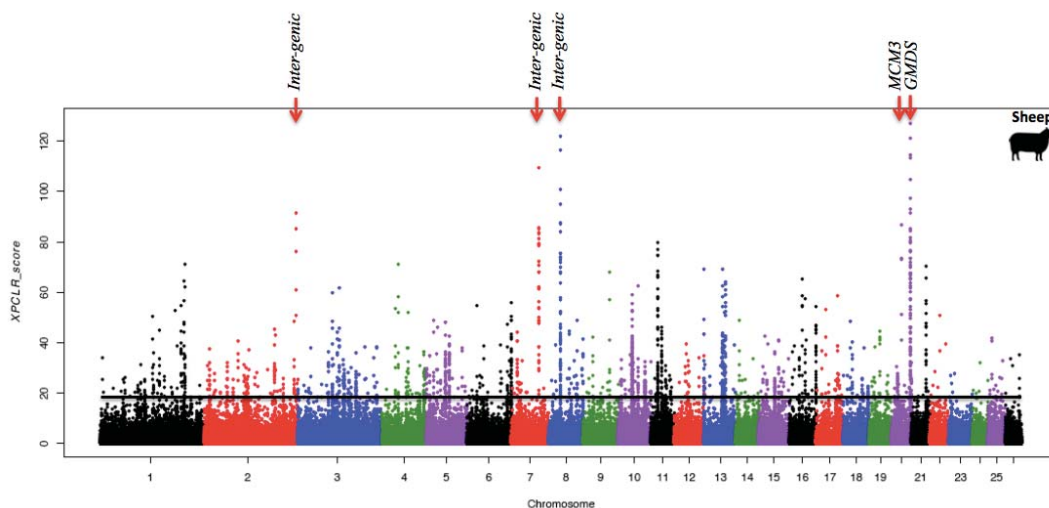


Figure 2 : Scores XP-CLR le long du génome du mouton en relation avec la recherche de SNPs sous sélection en fonction de l'altitude. Les scores sont indiqués pour les autosomes. La ligne noire indique le seuil de 0,1 % des scores les plus élevés. Les flèches rouges indiquent les régions chromosomiques associées aux 4 meilleurs scores. D'après Benjelloun 2015.

L'approche corrélative détecte beaucoup moins de SNPs potentiellement sous sélection (8 chez les moutons et 20 chez les chèvres) dont certains n'ont pas été détectés par l'approche populationnelle. Cette approche corrélative est ici très fiable dans la mesure où il n'y a pas de forte structure de population, ce qui diminue le risque de produire des faux positifs.

Sur l'ensemble de ces analyses en cours, les plus avancées concernent l'identification des gènes liés à l'altitude. Les fonctions biologiques impliquant ces gènes qui ont été identifiées par l'analyse *Gene Ontology* sont référencées dans le Tableau 1.

Tableau 1 : Voies métaboliques enrichies en gènes candidats sous sélection en relation avec l'altitude. Pour chaque espèce seuls les 3 termes les plus enrichis sont présentés. D'après Benjelloun 2015.

Terme <i>Gene Ontology</i>	Processus biologique	Gènes candidats/associés	enrichissement	P valeur
MOUTON				
GO:0010569	Régulation de la réparation des cassures double brin par recombinaison homologue	3/14	29,01	10 ⁻⁴
GO:0000018	Régulation de la recombinaison de l'ADN	4/50	10,83	5 10 ⁻⁴
GO:0070665	Régulation positive de la prolifération des leucocytes	6/106	7,66	10 ⁻⁴
CHEVRE				
GO:0060486	Différenciation des cellules Clara	2/3	46,34	6 10 ⁻⁴
GO:0060047	Contraction cardiaque	3/13	16,04	7 10 ⁻⁴
GO:0060306	Régulation de la repolarisation membranaire	4/25	11,12	4 10 ⁻⁴

3. DISCUSSION

La faible structure génétique observée sur l'ensemble du Maroc rend particulièrement efficace la détection des régions potentiellement impliquées dans l'adaptation locale, en limitant l'auto-corrélation spatiale et les effets confondants. La grande majorité des SNPs identifiés comme potentiellement associés à une adaptation sont intergéniques et introniques, comme montré dans de précédentes études (e.g. Ai et al. 2015). Ils peuvent être entraînés lors de balayages sélectifs, ou impliqués dans des mécanismes régulateurs via des

régions promotrices, des séquences régulant la traduction et la transcription, etc. De telles régions peuvent être très éloignées des gènes qu'elles régulent (Noonan and McCallion 2010; Dunham et al. 2012). Notre connaissance de ces mécanismes régulateurs est encore trop partielle pour pouvoir interpréter l'ensemble de ces données. Ces résultats posent également la question des taux de faux positifs détectés par les méthodes actuellement utilisées pour détecter des signatures de sélection. Les principales voies métaboliques qui ressortent comme potentiellement impliquées dans une réponse adaptative à l'altitude sont liées aux fonctions respiratoires et cardiaques. Chez la chèvre une des principales

voies métaboliques identifiées est associée à la différenciation de cellules de Clara. Heath et al. (1976) ont montré que ces cellules de l'épithélium pulmonaire sont plus grosses et plus nombreuses chez les lamas vivant à haute altitude. L'analyse de *Gene Ontology* montre également, parmi les gènes sous sélection, une sur-représentation de gènes impliqués dans les mécanismes de contraction du muscle cardiaque, et notamment en lien avec le potentiel d'action ventriculaire. Ce potentiel d'action est connu pour être prolongé en condition hypoxique à haute altitude (Zhou et al., 2015). Chez le mouton, les voies métaboliques enrichies en gènes sous sélection sont impliquées dans la prolifération des leucocytes. On peut penser à un lien avec la respiration, sachant que ces cellules sont impliquées dans la réaction à l'hypertension pulmonaire en condition d'hypoxie chronique (Stenmark et al. 2005). Une autre possibilité est une adaptation à une pression de pathogène plus importante aux basses altitudes. Une autre voie métabolique mise en évidence chez le mouton est liée à la contraction musculaire qu'on sait dépendre des conditions d'hypoxie (Perrey et Rupp, 2009).

A part quelques exceptions (moins de 1 % des gènes candidats pour l'altitude), les gènes potentiellement sélectionnés par un facteur environnemental sont différents chez la chèvre et le mouton.

CONCLUSION

Les approches de génomiques du paysage permettent d'identifier les voies métaboliques potentiellement impliquées dans l'adaptation locale aux conditions environnementales chez les petits ruminants. Même si l'implication effective des gènes identifiés doit être validée, ces études montrent (i) l'importance de maintenir la diversité allélique de ces gènes pour préserver le potentiel adaptatif, et (ii) le potentiel des populations locales vivants en environnements contrastés comme réserve de cette diversité.

Ces travaux sont financés par les projets européens NEXTGEN (FP7, Grant Agreement number: 244356) et ClimGen (FACCE-ERA-NET).

Ai H., Fang X., Yang B., Huang Z., Chen H., Mao L., Zhang F., Zhang L., Cui L., He W. et al., 2015. *Nature Genetics* 47 217

Benjelloun B., 2015. Thèse Univ Grenoble Alpes, 207 pp.

Benjelloun B., Alberto F.J., Streeter I., Boyer F., Coissac E., Stucki S., BenBati M., Ibnelbachyr M., Chentouf M., Bechchari A. et al., 2015. *Frontiers in Genetics* 6, 107-107.

Chen H., Patterson N., Reich D., 2010. *Genome Research* 20, 393-402.

Dunham I., Kundaje A., Aldred S.F., Collins P.J., Davis C., Doyle F., Epstein C.B., Frietze S., Harrow J., Kaul R., et al., 2012. *Nature* 489, 57-74.

Eden E., Navon R., Steinfeld I., Lipson D., Yakhini Z., 2009. *Bmc Bioinformatics* 10.

Frichot E., Mathieu F., Trouillon T., Bouchard G., Francois O., 2014. *Genetics* 196, 973.

Heath D., Smith P., Harris P., 1976. *Experimental Cell Biology* 44, 73-82.

Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G., Jarvis A., 2005. *Journal of Climatology* 25: 1965-1978.

Noonan J.P., McCallion A.S., 2010. *Annual Review of Genomics and Human Genetics* 11, 1-24.

Leempoel K., Golay F., Joost S. (Dir.), 2015. Thèse EPFL, n° 6626, doi :10.5075/epfl-thesis-6626

Perrey S., Rupp T., 2009. *High Altitude Medicine & Biology* 10, 175-182.

Stenmark K.R., Davie N.J., Reeves J.T., Frid M.G., 2005. *Journal of Applied Physiology* 98, 715-721.

Stucki S., Golay F., Joost S. (Dir.), 2014a. Thèse EPFL, n° 6014, doi:10.5075/epfl-thesis-6014

Stucki S., Orozco-terWengel P., Bruford M.W., Colli L., Masembe C., Negrini R., Taberlet P., Joost S., 2014b. arxiv 1405.7658v2.

Taberlet P., Valentini A., Rezaei H.R., Naderi S., Pompanon F., Negrini R., Ajmone-Marsan P., 2008. *Molecular Ecology* 17, 275-284.

Weir B.S., Cockerham C.C., 1984. *Evolution* 38, 1358-1370.

Zhou J.J., Ma H.-J., Liu Y., Guan Y., Maslov L.N., Li D.P., Zhang Y., 2015. *Canadian Journal of Physiology and Pharmacology* 93, 227-232.