



Mécanismes génétiques et épigénétiques sous-jacents aux relations phénotype - environnement chez les petits ruminants

Laure Denoyelle

► To cite this version:

Laure Denoyelle. Mécanismes génétiques et épigénétiques sous-jacents aux relations phénotype - environnement chez les petits ruminants. Biodiversité. Université Grenoble Alpes [2020-..], 2020. Français. NNT : 2020GRALV045 . tel-03205678

HAL Id: tel-03205678

<https://theses.hal.science/tel-03205678>

Submitted on 22 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Biodiversité – Écologie – Environnement**

Arrêté ministériel : 25 mai 2016

Présentée par

Laure DENOYELLE

Thèse dirigée par **François POMPANON**, Professeur, Université
Grenoble Alpes, et

Co-dirigée par **Gwenola TOSSER-KLOPP**, Directrice de
recherche, **INRAE Toulouse**

préparée au sein du **Laboratoire d'Écologie Alpine (LECA)**
dans l'**École Doctorale Chimie et Sciences du Vivant (EDCSV)**

Mécanismes génétiques et épigénétiques sous-jacents aux relations phénotype – environnement chez les petits ruminants

Thèse soutenue publiquement le **17/12/2020**,
devant le jury composé de :

Mme Anne BLONDEAU – DA SILVA

Maître de Conférences, Faculté des Sciences de Limoges, Examinatrice

Mme Laurence FLORI

Chargée de recherche, INRAE Montpellier, Rapporteuse

Mr Olivier FRANCOIS

Professeur, Université Grenoble Alpes, Président du jury

Mr Christophe PLOMION

Directeur de recherche, INRAE Bordeaux, Rapporteur

Mr François POMPANON

Professeur, Université Grenoble Alpes, Co-directeur de thèse

Mme Gwenola TOSSER-KLOPP

Directrice de recherche, INRAE Toulouse, Co-directrice de thèse



Remerciements

Je voudrais tout d'abord remercier les membres de mon jury qui ont accepté d'évaluer mon travail de thèse : Laurence Flori et Christophe Plomion pour leurs rapports, Olivier François pour avoir accepté la présidence du jury et d'avoir braver le virus pour être présent à ma soutenance. Un très grand merci à Anne Blondeau – Da Silva qui a suivi l'avancement de ma thèse jusqu'à son dénouement ! Ainsi qu'à Michaël Blum qui a accepté de participer à mes comités de suivi de thèse.

Tout ce travail n'aurait pas été possible sans mes directeurs de thèse. Tout a commencé par l'acceptation de ma candidature par François. Ensuite Gwenola a accepté de co-encadrer mon travail et nous voilà parti pour 4 ans de recherche. Merci à tous les deux de m'avoir fait confiance pour mener à bien ce projet !

J'ai eu la chance d'avoir deux laboratoires d'accueil amenant des rencontres enrichissantes aussi bien d'un point de vue scientifique, qu'humain.

J'aimerais remercier tous les membres de GenPhySE qui m'ont toujours accueilli avec bienveillance dans leurs locaux (et ont su me guider dans les couloirs biscornus...). Un immense merci à l'équipe GenROC qui a réussi à m'intégrer malgré la distance Toulouse – Grenoble. Merci pour l'accueil dans vos bureaux, les délires pendant les réunions, mais aussi les discussions constructives concernant nos travaux de recherche. Surtout ne changez rien ! Merci également à Philippe Bardou qui a toujours répondu présent pour débbugger mes scripts ou m'expliquer certaines subtilités bio-informatiques ! Bertrand Servin et Simon Boitard ont également pris le temps, régulièrement, de faire un point avec moi sur l'avancée de mes analyses et pour cela, je tenais à les remercier. J'ai souvent sollicité Isabelle Palhière et Coralie Danchin pour collecter des informations plus précises sur les races caprines françaises. Je voulais les remercier pour leur patience et pour toutes les informations complémentaires qu'elles ont pu m'apporter !

Merci à tous les autres que j'ai pu côtoyer lors des pauses café ou aux repas à la cafétéria.

Même si j'avais effectué mes stages de master au LECA je connaissais assez peu le laboratoire de Grenoble. Merci à tous pour votre accueil. Je tiens particulièrement à remercier l'équipe « petits ruminants » pour notre entraide, nos échanges et les 'râles' de début de réunion. Merci aussi à Bruno pour les débogages CIMENT ainsi que Christian et Delphine pour leur aide sur mes manip. Un grand merci également à Glenn et l'équipe du LECA Chambéry sans qui, il aurait été très peu probable que je me lance dans l'aventure de la recherche...

Le bureau 305 a été mon repaire pendant 4 ans. Avec Stéphanie, ma co-bureau de toujours, nous y avons connu des déménagements, des départs et des arrivées ! Comme celui de Charlotte, petite stagiaire, puis doctorante avec qui j'ai pu échanger sur des astuces bioinfos, de jolies figures R, les animaux de compagnie (on va inclure les pigeons), certains livres et autres délires (non je n'ai pas toujours une chanson débile en tête !).

À l'étage d'en dessous, j'ai aussi fait la rencontre de Marie avec qui nous avons découvert de nombreux centres d'intérêts communs, ce qui a entraîné des pauses méridiennes un peu longues (en particulier l'été !). Ce sont ensuite Joaquim et Adeline qui sont arrivés et avec qui nous avons pu échanger des délires en particulier aux moments des repas, mais aussi en dehors du labo.

Je remercie ma famille qui s'est toujours intéressée à mes travaux. Ils ont entendu parler de renards, de chamois puis de chèvres et de moutons sans toujours très bien comprendre ce que je leur faisais (ou pas d'ailleurs, car je ne fais que traiter des données !). Ils ont suivi ma rédaction en m'aidant des fois à choisir les termes les plus appropriés, et m'ont soutenu lors de ma soutenance !

Un dernier merci à mes ami.e.s qui ont suivi de près ou de loin mon aventure doctorale ! Ainsi qu'à toutes les autres personnes que j'ai pu oublier précédemment...

Table des matières

INTRODUCTION	5
I. L'EVOLUTION DES ESPECES	7
II. L'INFLUENCE DES MECANISMES GENETIQUES SUR LE PHENOTYPE.....	11
<i>Les mutations ponctuelles</i>	11
<i>Les variants structuraux</i>	12
<i>Les signatures de sélection</i>	14
III. L'INFLUENCE DES MECANISMES EPIGENETIQUES SUR LE PHENOTYPE	17
IV. LE CAS DES ESPECES DOMESTIQUES	21
V. LE CONTEXTE DE LA THESE	23
<i>Les petits ruminants</i>	23
<i>Problématiques</i>	24
<i>Les jeux de données</i>	24
 MECANISMES GENETIQUES	27
<u>PRESENTATION DU JEU DE DONNEES VARGOATS</u>	29
<i>Résumé de l'article</i>	29
<i>Abstract</i>	30
<i>Data description</i>	31
<i>Methods</i>	38
<i>Genetic diversity</i>	43
<i>Re-use potential</i>	45
<i>Supplementary tables</i>	45
 <u>ANALYSE DES CHEVRES FRANÇAISES</u>	46
I. PRESENTATION DES RACES FRANÇAISES	46
<i>Les races locales</i>	48
<i>Les races en sélection</i>	56
<i>Problématique</i>	58
II. MATERIELS ET METHODES	58
<i>Échantillonnage</i>	58
<i>Variants génétiques</i>	59
<i>Analyse de la structure</i>	60
<i>Relation entre les races</i>	61

<i>Histoire démographique des races</i>	62
<i>Signature de sélection</i>	63
Analyse globale	64
Analyse ciblée	65
III. RESULTATS	66
<i>Analyse de la structure</i>	66
<i>Relation entre les races</i>	68
<i>Histoire des races</i>	71
<i>Signature de sélection</i>	78
Analyse globale	78
Analyse ciblée	83
IV. DISCUSSION	87
<i>Diversité neutre</i>	87
<i>Sélection</i>	91

MECANISMES EPIGENETIQUES 97

<i>Résumé de l'article</i>	99
<i>Introduction</i>	100
<i>Material & Methods</i>	102
<i>Results</i>	107
<i>Discussion</i>	113
<i>Conclusion</i>	117

DISCUSSION GENERALE119

I. MARQUES GENETIQUES : TEMOINS DE L'HISTOIRE DES ESPECES	121
<i>Début de la domestication</i>	121
<i>Sélection agronomique</i>	122
<i>Sélection induite par le climat</i>	129
<i>Influence des données de génomes complets vs puce ADN</i>	130
<i>Différentes cibles pour la sélection</i>	132
II. MARQUES EPIGENETIQUES : TEMOIN DE L'ACCLIMATATION	134
III. RELATION ENTRE LES DEUX MECANISMES	134
<i>Localisation des variants</i>	135
<i>Effet des variants</i>	136
<i>Différents processus évolutifs pour répondre à l'environnement</i>	137
IV. CONCLUSION	139

BIBLIOGRAPHIE	141
----------------------------	------------

ANNEXES	167
----------------------	------------

I. MECANISMES GENETIQUES ET SELECTION	169
<i>Diversité neutre</i>	169
<i>Signatures de sélection</i>	173
Analyse globale	173
Analyse ciblée	203
II. MECANISMES EPIGENETIQUES ET ACCLIMATATION	211
<i>Pathway approach</i>	214
<i>Gene approach</i>	217

INTRODUCTION

Les organismes vivants peuvent être confrontés à des variations plus ou moins importantes de leur environnement (qu'il soit climatique, sanitaire, etc...). En réponse à ces variations, trois types de processus peuvent être mis en place pour permettre l'adéquation entre phénotype (c'est-à-dire les caractères visibles composant l'anatomie, la morphologie, la physiologie, le comportement etc.) et environnement : le choix d'un habitat favorable, l'adaptation et l'acclimatation.

Le choix de l'habitat ne fait pas intervenir de processus modifiant le phénotype. L'organisme se déplace vers un environnement qui est optimal pour le phénotype qu'il porte. La méta-analyse de Chen et al. (2011) portant sur différentes espèces (animaux ou plantes) présentes dans l'hémisphère Nord, a montré un déplacement de leurs aires de répartition vers le Nord en réponse aux changements climatiques. Cependant cette migration vers un environnement favorable n'est pas toujours possible, comme par exemple, pour les espèces domestiques qui ont dû s'adapter à différentes conditions d'élevage.

Les processus d'adaptation et d'acclimatation entraînent des modifications du phénotype des organismes. Ces processus peuvent être mis en évidence en étudiant respectivement la présence de marques génétiques et épigénétiques chez les individus d'une population.

Dans cette partie introductive nous allons tout d'abord retracer les courants de pensées qui ont mis en lumière l'évolution des espèces et donc pour la première fois la transformation des phénotypes en fonction de l'environnement. Ensuite nous détaillerons les processus d'adaptation et d'acclimatation en définissant les mécanismes les composant, comment ils peuvent être détectés et prendrons quelques exemples d'études mettant en évidence la présence de ces mécanismes dans la relation phénotype – environnement. Nous introduirons ensuite le processus de domestication qui a conduit à la création des espèces domestiques que nous allons étudier. Pour finir, les problématiques traitées dans ce manuscrit seront présentées ainsi que les jeux de données nous ayant permis de les traiter.

I. L'évolution des espèces

L'existence d'une transformation du phénotype des êtres vivants en fonction de leur environnement n'a pas toujours été admise. La première théorie la mettant en avant est celle

établie par le naturaliste français Jean-Baptiste Lamarck (1744-1829) en 1809 dans son ouvrage *Philosophie zoologique*. Celle-ci se nomme « hérédité des caractères acquis » et repose sur deux lois :

1. L'utilité d'un caractère fera que celui-ci va se développer pour être plus performant alors qu'à l'inverse un caractère peu utilisé finira par disparaître.
2. Ce développement ou cette perte de caractère se transmet dans la descendance.

L'exemple le plus communément utilisé pour illustrer cette théorie est le cou de la girafe (Figure 1) : des circonstances extérieures (ici la hauteur des feuilles sur l'arbre) va entraîner le besoin de la girafe de tirer sur son cou pour pouvoir se nourrir donc le cou de cette dernière va s'allonger graduellement au cours du temps.

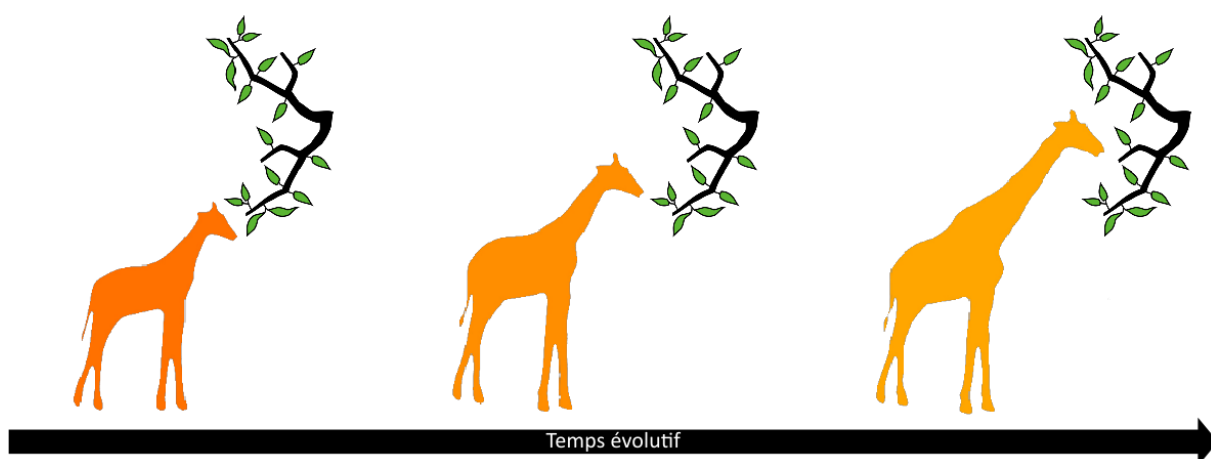


Figure 1 : Transformation évolutive du cou de la girafe selon la théorie de l'hérédité des caractères acquis de Jean-Baptiste Lamarck

Cinquante ans plus tard, Charles Darwin (1809-1882) publie son ouvrage *L'origine des espèces* (1859) où il développe sa théorie de l'évolution du vivant par « descendance avec modification ». Cette théorie de l'évolution s'appuie sur deux notions :

1. Des variations héréditaires.
2. Les organismes portant une variation conférant un avantage seront eux-mêmes favorisés par rapport aux autres (sélection naturelle).

Si nous reprenons l'exemple de la girafe (Figure 2) : des variations de la taille du cou apparaissent dans la population. Les girafes ayant un long cou sont avantagées pour se nourrir car elles ont accès à plus de feuilles. Celles-ci sont plus aptes à survivre et à se reproduire par rapport aux autres girafes donc, au cours du temps, il ne restera que des girafes à long cou.

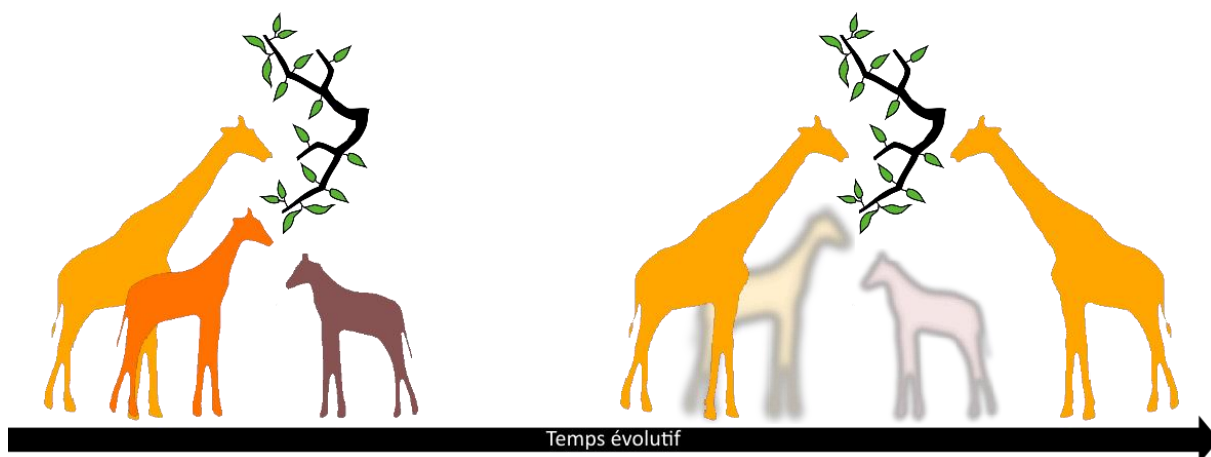


Figure 2 : Transformation évolutive du cou de la girafe selon la théorie de l'évolution de Charles Darwin.

Cet exemple représente un cas d'adaptation à l'environnement. Celle-ci se définit comme un processus évolutif qui améliore l'aptitude d'un organisme à vivre dans un environnement donné. Cette notion prend donc en compte, en plus de l'apparition d'une nouvelle valeur pour un caractère, sa valeur sélective (ou « fitness ») qui est la résultante de la survie et la fécondité de l'organisme. Si cette valeur sélective s'accroît avec l'apparition d'un caractère ou de sa variation, il est alors possible de dire que ce dernier permet à l'organisme qui le porte de s'adapter à son environnement.

Les observations des naturalistes reposaient uniquement sur la morphologie des organismes étudiés. Il faut attendre la naissance de la génétique, à la fin du XIX^{ème} siècle, pour comprendre comment ces variations phénotypiques apparaissent et se transmettent. Cela a permis de repenser les deux théories présentées ci-dessus avec un éclairage génétique et l'élaboration, dans les années 1930-1940, de la théorie synthétique de l'évolution (ou néo-darwinisme) qui regroupe les théories de la génétique et de la sélection naturelle.

En plus de la sélection naturelle, il existe 3 autres forces évolutives :

1. La mutation qui correspond à des changements héréditaires spontanés des allèles dus aux erreurs dans la répllication de l'ADN. Cela peut donner lieu à l'apparition d'un nouveau caractère ou une nouvelle valeur phénotypique pour un caractère déjà existant. Si celle-ci apporte une augmentation de la valeur sélective d'un individu alors la mutation pourra être sélectionnée. Cependant il semblerait que l'adaptation provienne principalement de la variabilité résidente dans le génome que de nouvelles mutations (Savolainen et al., 2013).

2. La migration qui induit un flux d'allèles entre différentes populations. Celle-ci peut entraver l'adaptation d'une population car ce flux ne permet pas aux allèles apportant un avantage sélectif dans un environnement donné, de se fixer dans la population (sélection diversifiante entre deux populations vivant dans des environnements différents).
3. La dérive qui est un changement aléatoire de la fréquence des allèles qui se produisent par hasard dans les populations de taille finie. Celle-ci peut peut-être responsable de la fixation de mutations délétères ou de la perte de mutations bénéfiques quand elles sont encore rares (Elena & Lenski, 2003). Cela suggère une action défavorable de la dérive sur l'adaptation, en particulier chez les populations à faibles effectifs car la probabilité de fixer ou perdre des allèles est la plus forte.

Ces forces ont pour effet des modifications de fréquences alléliques dans les populations, mais ne permettent pas d'expliquer l'observation de changements phénotypiques sans modification du génotype. C'est par exemple le cas chez les abeilles où une différence d'alimentation des larves femelles va avoir un impact sur la méthylation de l'ADN et l'expression de certains gènes induisant l'apparition d'un phénotype « reine » ou « ouvrière » (Barchuk et al., 2007; Kucharski et al., 2008). Ces changements sont appelés épigénétiques, c'est-à-dire que différents environnements vont avoir des impacts différents sur un même génotype qui vont induire un changement de phénotype. Ce mécanisme ainsi que les changements phénotypiques uniquement dus à l'environnement sont regroupés sous l'appellation "plasticité phénotypique".

Pour résumer, la valeur d'un phénotype dépend de plusieurs composantes : le génotype (ensemble des caractéristiques génétiques) de l'organisme, l'environnement dans lequel il vit et l'interaction de ces deux paramètres ($V_P = V_E + V_G + V_{G \times E}$). Nous allons maintenant détailler comment chacun de ces paramètres peut influencer le phénotype d'un organisme.

Tout d'abord l'environnement seul peut impacter un phénotype. Par exemple, la quantité de ressources disponibles dans le milieu pourra avoir un impact sur la croissance et la taille à maturité de l'organisme.

II. L'influence des mécanismes génétiques sur le phénotype

La génétique est une discipline scientifique qui s'intéresse aux gènes c'est-à-dire une séquence d'ADN qui code pour l'apparition et ou expression d'un caractère chez un organisme. Les différentes versions d'un même gène, appelées allèles, et leur combinaison, peuvent être à l'origine des différents phénotypes observés au sein d'une population.

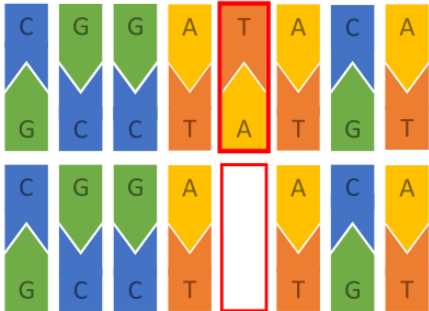
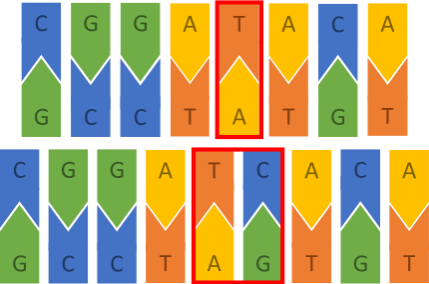
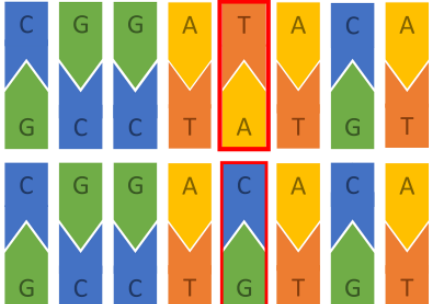
Cependant l'ADN n'est pas uniquement composé de gènes mais aussi de séquences non-codantes qui peuvent par exemple contenir des séquences régulatrices de gènes. L'étude de l'ensemble du matériel génétique (région codante et non-codante) au niveau du génome entier est appelée génomique.

Il existe deux catégories de mécanismes entraînant des modifications d'une séquence d'ADN : les mutations ponctuelles et les variants structuraux.

Les mutations ponctuelles

Les mutations ponctuelles correspondent à un changement spontané et aléatoire d'un nucléotide dans une séquence d'ADN. Il en existe 3 types : la délétion, l'insertion et la substitution qui sont définies et illustrées dans le Tableau I.

Tableau I : Présentation des différents types de mutations ponctuelles. Chaque double brin correspond à une séquence d'ADN chez des individus différents.

Mutation ponctuelle	Définition
	Suppression d'un nucléotide :
Délétion	
	Ajout d'un nucléotide :
Insertion	
	Changement d'un nucléotide par un autre :
Substitution	

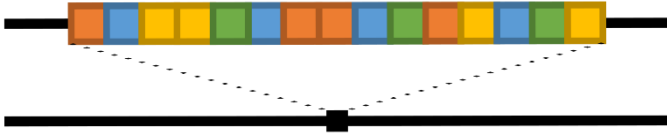
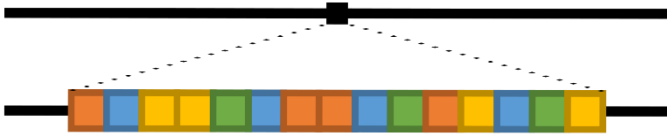
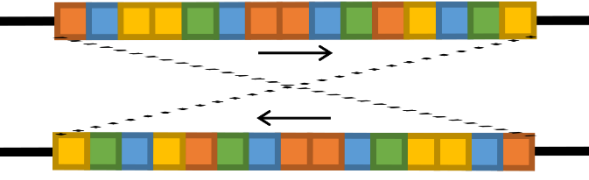
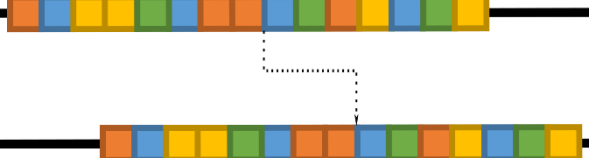
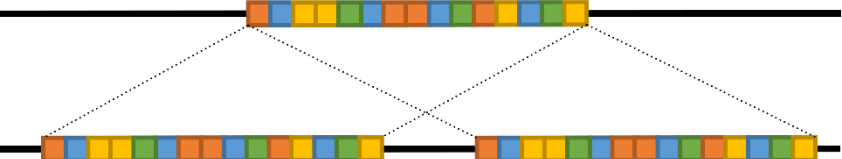
Une autre terminologie peut être trouvée dans la littérature scientifique où les substitutions sont nommées SNPs (pour « Single Nucleotide Polymorphism »), et les insertions délétions INDELs. Si les fragments insérés ou supprimés représentent plus de 50 paires de bases (pb) alors ils sont considérés comme des variants structuraux.

Les variants structuraux

Les variants structuraux modifient la structure du chromosome ce qui correspond à des modifications de longues séquences d'ADN (supérieures à 50 pb, Tattini et al., 2015). Ils regroupent les insertions, délétions, inversions de nombreuses nucléotides ; la variabilité du nombre de copies d'une séquence d'ADN (« copy number variation », CNV) et les

réarrangements chromosomiques. Ces différents variants structuraux sont définis et illustrés dans le Tableau II.

Tableau II : Présentation des différents variants structuraux. Chaque ligne correspond à une séquence d'ADN chez des individus différents.

Variants structuraux	Définition
Délétion	Une séquence d'ADN est perdue : 
Insertion	Plusieurs nucléotides sont ajoutés dans une séquence d'ADN : 
Inversion	Un segment du chromosome est inversé : 
Réarrangement chromosomique	Une partie du chromosome est localisée à un autre endroit sur un chromosome homologue ou non : 
Variabilité du nombre de copies	Une partie du génome est répétée : 

Si ces différentes modifications sont localisées dans une région régulatrice d'un gène ou dans celui-ci elles pourront être à l'origine d'un nouvel allèle et potentiellement de l'apparition d'un nouveau phénotype. C'est pourquoi toutes ces modifications du génome sont intéressantes à étudier afin de comprendre quels mécanismes sont à l'origine de l'adaptation des organismes.

Les signatures de sélection

Une grande partie du génome est considérée comme neutre, c'est-à-dire n'ayant pas d'effet avantageux ou désavantageux sur un individu. Nous retrouvons dans ces régions des variants neutres créés par mutation et dont la fréquence peut être modifiée par différentes forces évolutives comme la dérive génétique (succession de phénomènes aléatoires pouvant fixer ou éliminer des variants) ou la migration (transfert d'allèles d'une population à une autre).

Quand une mutation avantageuse apparaît dans une population, elle peut se transmettre au cours des générations avec la séquence d'ADN l'entourant ce qui a pour effet de réduire toute la variabilité génétique autour du variant et d'obtenir un haplotype (groupe d'allèles de différents loci transmis ensemble) sélectionné. Ce phénomène est appelé balayage sélectif (« selective sweep ») et on en distingue 2 types principaux et un intermédiaire (Figure 3) :

1. Le balayage complet (« hard sweep »), c'est-à-dire que l'haplotype sélectionné est fixé au sein de la population.
2. Le balayage doux (« soft sweep ») où plusieurs haplotypes peuvent être présents dans une population. Dans ce cas-là la force de sélection ne permet pas la fixation d'un seul haplotype dans la population.
3. Le cas intermédiaire appelé balayage incomplet (« incomplet sweep ») où plusieurs haplotypes peuvent être présents mais ce cas peut continuer d'évoluer au cours du temps jusqu'à fixation de l'haplotype sélectionné et donc devenir un balayage complet.

Il est impossible de faire la différence entre un balayage doux et un balayage incomplet en réalisant une analyse des balayages sélectifs présents chez des populations à un instant donné. Si l'analyse se réitère après plusieurs générations alors il sera possible de voir s'il y a eu une modification du nombre d'haplotype dans la région et si c'est le cas alors nous sommes en présence d'un balayage incomplet.

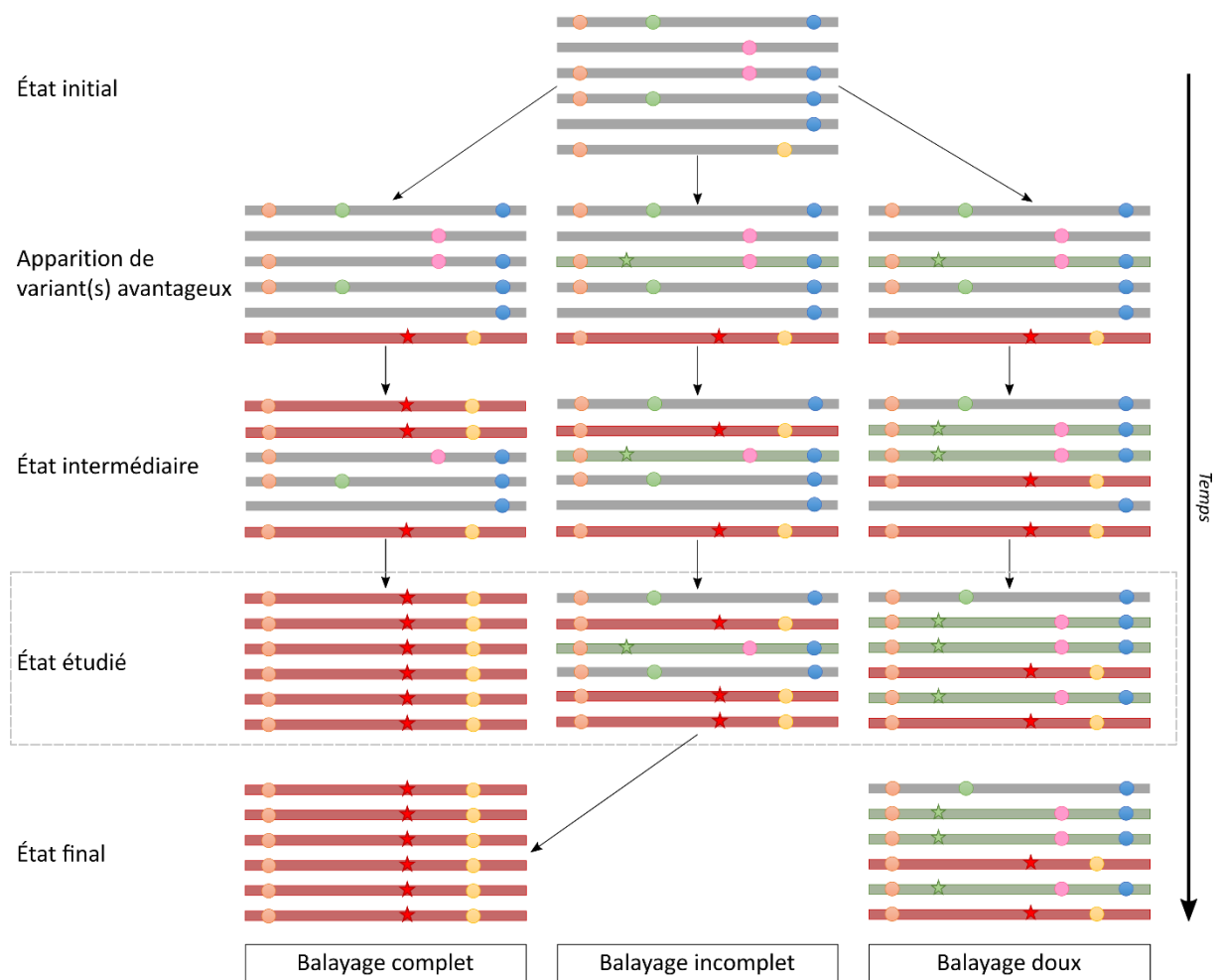


Figure 3 : Exemples des différents types de balayages sélectifs. Chaque ligne correspond à l'haplotype d'un individu de la population, les ronds sont des variants neutres et les étoiles les variants sélectionnés.

Pour détecter ces balayages sélectifs au sein de populations il existe plusieurs méthodes :

1. Chercher les balayages sélectifs au sein d'une population en utilisant des modèles statistiques pour déterminer les régions neutres et sélectionnées dans le génome d'individus appartenant à la même population. Les balayages sélectifs ont pour effet d'augmenter le déséquilibre de liaison (LD) et réduire le spectre de fréquence allélique (AFS) localement (Nielsen, 2005). Les programmes permettant de réaliser ce type d'analyse sont par exemple freqHMM (Boitard et al., 2009) qui utilise l'AFS et un modèle de Markov caché afin d'inférer l'état neutre, intermédiaire ou sélectionné des SNPs dans le génome ; REHH (Gautier & Vitalis, 2012) et SWEED (Pavlidis et al., 2013) qui détectent respectivement les LD et AFS modifiés par la présence d'un balayage sélectif.
2. Contraster des populations entre elles pour trouver les régions sélectionnées chez certaines d'entre elles par rapport aux autres. Cette méthode se base donc sur les différences génétiques entre individus ou populations avec des programmes comme

hapFLK (Bonhomme et al., 2010; Fariello et al., 2013) qui utilise la différence génétique (F_{ST}) entre populations et le LD afin de détecter des haplotypes sous sélection entre différentes races ; ou encore PCA_{dapt} (Luu et al., 2017) qui se base sur le résultat de l'Analyse en Composante Principale (ACP) des individus pour prendre en compte la structure du jeu de données, puis teste les corrélations entre la variation génétique et les premières composantes principales pour détecter les SNPs sous sélection.

3. Détection des variants corrélés à un phénotype d'intérêt ou des variables environnementales (étude d'association pangénomique, « genome-wide association study », GWAS). Le programme GEMMA (Zhou & Stephens, 2012) utilise un modèle linéaire mixte pour tester la combinaison des effets des variants, d'effets aléatoires et fixes et des résidus du modèle pouvant expliquer le phénotype observé, et LFMM (Frichot et al., 2013) utilise un modèle mixte de facteurs latents afin de tester la combinaison des effets du phénotype, d'autres variables latentes et des résidus du modèle pouvant expliquer le génotype observé.

La liste de programmes présentée ci-dessus n'est pas exhaustive et de nouvelles méthodes sont constamment en développement. Les programmes cités ci-dessus utilisent des SNPs pour détecter les signatures de sélection. C'est la modification génomique la plus couramment utilisée pour détecter des signatures de sélection car ils sont plus faciles à détecter dans le génome (au vu des méthodes standardisées que nous connaissons) que les variants structuraux car toutes ne peuvent pas être étudiées avec un seul outil (Mills et al., 2011). Pourtant ces derniers peuvent avoir un impact fort sur le phénotype et la fitness des individus (Feuk et al., 2006).

Exemples d'adaptation à l'environnement

Les différentes catégories de méthodes présentées ci-dessus permettent de détecter des SNPs ou SV sous sélection chez toutes sortes d'espèces. Concernant les SNPs, nous pouvons par exemple citer le gène GRMZM2G465226 associé à la résistance du charbon de l'épi dans le maïs (Wang et al., 2012) ; les gènes NHE et NBC qui permettent aux poissons d'eau de mer de réguler l'entrée d'ion sodium dans leur organisme (Brennan et al., 2018) ; et les gènes codant pour l'hémoglobine (HbA et HbD) chez les oiseaux volant à de hautes altitudes (McCracken et al., 2010). Pour les variants structuraux, l'étude de Christmas et al. (2019) a mis en évidence une adaptation à l'environnement montagnard due à la présence d'inversions chromosomiques dans

deux régions du génomes chez les abeilles d'Afrique de l'Est qui vivent dans les montagnes par rapport à celles vivant en plaine. Cumer, (2017) a mis en évidence des variants structuraux en relation avec l'adaptation à l'environnement chez les petits ruminants marocains, avec par exemple une délétion intronique dans le gène *MAGI2* qui est impliqué dans l'adaptation à l'altitude chez les cochons tibétains (Ai et al., 2014) ou un CNV intronique dans le gène *RNLS* qui est lié à l'efficacité alimentaire (Serão et al., 2013) et qui concorde avec l'adaptation de la race caprine Draa qui a la capacité de maintenir son apport alimentaire même lorsque l'eau manque (Hossaini-Hilali & Mouslih, 2002).

III. L'influence des mécanismes épigénétiques sur le phénotype

Le terme épigénétique a été utilisé pour la première fois par Conrad Waddington en 1939 dans son ouvrage intitulé « *An introduction to modern genetics* ». Il correspond à l'ensemble des modifications d'activité des gènes qui ne sont pas dues à des variations génomiques. À partir de ce concept général de nombreux champs de recherche avec leurs problématiques spécifiques se sont appropriés ce terme et en ont donné des définitions différentes (Figure 4).

Conception	Définition	Champs de recherche	Problème
W-épi (l'épigénétique selon Waddington) Années 1930-40	Les mécanismes causaux impliqués dans le développement par lesquels les gènes produisent des effets phénotypiques	• Génétique classique et embryologie expérimentales • Biologie du développement	Développement (au niveau de l'organisme)
N-épi (l'épigénétique selon Nanney) Années 1950-60	Systèmes intégratifs auxiliaires régulant l'expression des potentialités génétiques	• Génétique chimique (moléculaire) et biologie du développement	Développement (au niveau de la cellule)
RH-épi (l'épigénétique selon Riggs et Holliday) Années 1970 à 1990-2000 & M-épi (l'épigénétique moléculaire) Années 2000-2010	Les changements héréditaires, par mitose et/ou méiose, de la fonction des gènes qui ne peuvent pas être expliqués par des changements de la séquence d'ADN Toute modification de la chromatine ayant un impact sur l'expression des gènes, que cette modification soit héréditaire ou pas	• Génétique et épigénétique moléculaire	Développement (au niveau moléculaire)
ED-épi (l'épigénétique selon l'évo-dévo) Années 1990-2010	Les mécanismes développementaux (au dessus du niveau de la séquence d'ADN) qui sont à l'origine du phénotype et de ses modifications au cours de l'évolution	• Génétique du développement • Biologie évolutive du développement (évo-dévo) • Biologie des systèmes	L'origine de la variation phénotypique et l'interaction entre développement et évolution
ES-épi (l'épigénétique selon la Synthèse Étendue) Années 2000-2010	Mélange de N-épi & ED-épi Focalisation sur l'hérédité épigénétique transgénérationnelle	• Biologie évolutive du développement (évo-dévo) • Biologie de l'évolution • Biologie des systèmes	L'origine de la variation phénotypique et l'évolution vers une synthèse évolutionnaire étendue

Figure 4 : Résumé des différents conceptions de l'épigénétique, leurs définitions, les champs de recherche concernés et les problématiques auxquels ils essayent de répondre. Ce tableau provient de la Prospective épigénétique, écologie et évolution du CNRS (2018, <https://inee.cnrs.fr/fr/prospective-epigenetique-ecologie-et-evolution>) et inspiré par Nicoglou & Merlin (2017).

La conception de l'épigénétique qui nous intéresse est l'épigénétique moléculaire (M-épi). Comme il est inscrit dans la Figure 4, ce concept s'intéresse aux modifications de la chromatine qui est une structure nucléoprotéique composée des histones ainsi que de la séquence d'ADN qui les entoure (Figure 5).

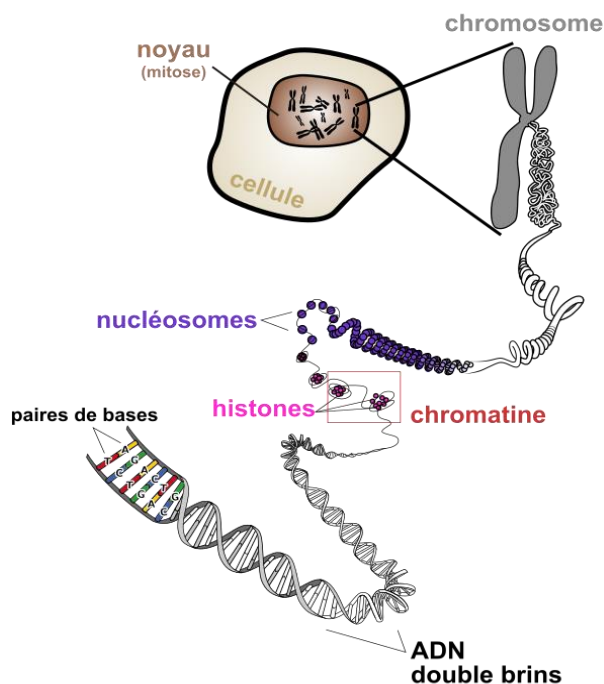


Figure 5 : Compaction de l'ADN jusqu'au chromosome conservé dans le noyau d'une cellule eucaryote

Les 2 états de compaction de cette structure portent des noms différents et ont des impacts différents sur l'expression des gènes :

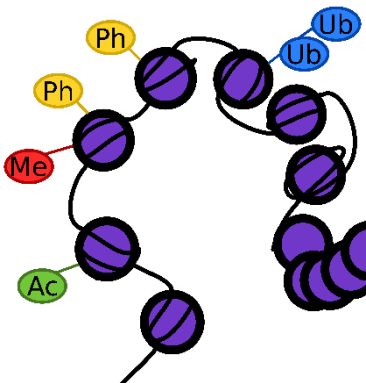
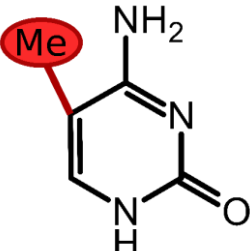
1. L'hétérochromatine correspond à une compaction très dense de la chromatine. Cela ne permet pas l'accès à l'ADN donc les gènes présents dans ces régions sont généralement peu exprimés.

2. L'euchromatine, au contraire, correspond à une compaction plus lâche de la chromatine donc l'accès aux gènes est facilité et ceux-ci peuvent être exprimés.

Les différents mécanismes pouvant avoir un effet sur la compaction de la chromatine sont la modification des histones et la méthylation de l'ADN (Tableau III).

Les cytosines méthylées sont particulièrement présentes sur les dinucléotides CG (Bollati & Baccarelli, 2010) ou au niveau des éléments transposables (Suzuki et al., 2007; Hollister & Gaut, 2009).

Tableau III : Présentation des différents mécanismes épigénétiques ayant un impact sur la compaction de la chromatine.

Mécanismes épigénétiques	Définition
Modification des histones	Méthylation (Me), phosphorylation (Ph), acétylation (Ac) ou ubiquitination (Ub) des histones : 
Méthylation de l'ADN	Ajout d'un groupement méthyle sur le 5^{ème} carbone du cycle pyrimidique d'une cytosine : 

Relation épigénétique / environnement

Les études recherchant des signatures génomiques en relation avec l'environnement mettent en évidence des mécanismes d'adaptation, alors que pour les études cherchant les relations entre l'environnement et des marques épigénétiques il s'agit de mécanismes d'acclimatation. Ces études se basent principalement sur la méthylation de l'ADN en mettant en évidence des Régions Différentiellement Méthylées (« differentially methylated regions », DMRs) entre des organismes vivant dans des conditions différentes.

Beaucoup d'entre elles sont réalisées en environnement contrôlé afin que seul l'effet du changement du paramètre souhaité puisse être observé. C'est par exemple le cas de l'étude de Kim et al. (2017) qui ont cultivé des *Arabidopsis thaliana* dans les mêmes conditions expérimentales excepté pour la dose de glyphosate à laquelle elles ont été soumises (0%, 5%, 10% ou 15%) et ont trouvé des DMRs en relation avec l'action du glyphosate et où les niveaux de méthylation étaient positivement corrélés à la gravité de la lésion causée par l'herbicide. Si l'on s'intéresse à un modèle animal, l'étude de Dolinoy et al. (2007) utilise des souris femelles élevées dans les mêmes conditions expérimentales à l'exception d'une injection ou non de bisphénol A

et de l'ingestion de compléments alimentaires. Cela a permis de montrer que l'exposition au bisphénol impliquait une hypométhylation chez la descendance, mais qu'une supplémentation alimentaire de la mère avec une nourriture riche en donneurs de méthyle (acide folique par exemple) a annulé cet effet.

Depuis quelques années, les études dans un environnement naturel et sur des organismes non modèles augmentent car l'acclimatation permet de comprendre certaines réponses des organismes à un nouvel environnement. C'est par exemple le cas des espèces invasives qui ont une variabilité génétique faible à cause de l'effet fondateur, mais qui pourtant peuvent coloniser rapidement de nouveaux espaces. Hu et al. (2019) ont introduit 211 lézards provenant d'une même île sur 8 autres, avant de les recapter 4 jours après pour comparer les différences de méthylation de l'ADN. Des cytosines différenciellement méthylées ont été mises en évidence chez les populations introduites et sont en relation avec des gènes impliqués dans le système nerveux, la réponse immunitaire, le rythme circadien et la croissance. Toutes ces voies biologiques peuvent aider les individus à coloniser des environnements présentant de nouvelles caractéristiques biotiques et abiotiques.

De plus, les études des marqueurs épigénétiques sur des populations naturelles, permet d'étudier des espèces sauvages, de comprendre quelle peut être la cause de certaines variations phénotypiques non expliquées par la génétique et comment les organismes s'acclimatent à leur environnement (Bossdorf et al., 2008).

Maintenant que nous avons détaillé les différents mécanismes permettant l'adaptation et l'acclimatation des organismes à leur environnement nous allons nous intéresser plus particulièrement aux espèces domestiques car chez elles, des intérêts agronomiques peuvent être à l'origine des changements de leurs phénotypes.

IV. Le cas des espèces domestiques

L'élevage et l'agriculture sont associés à la révolution Néolithique qui a commencé il y a environ 12 000 ans au Moyen-Orient avant de s'étendre dans le reste du monde. Durant cette période

les humains ont domestiqué différentes espèces sauvages (Figure 6) afin de s'assurer une ressource ou un service prévisible.

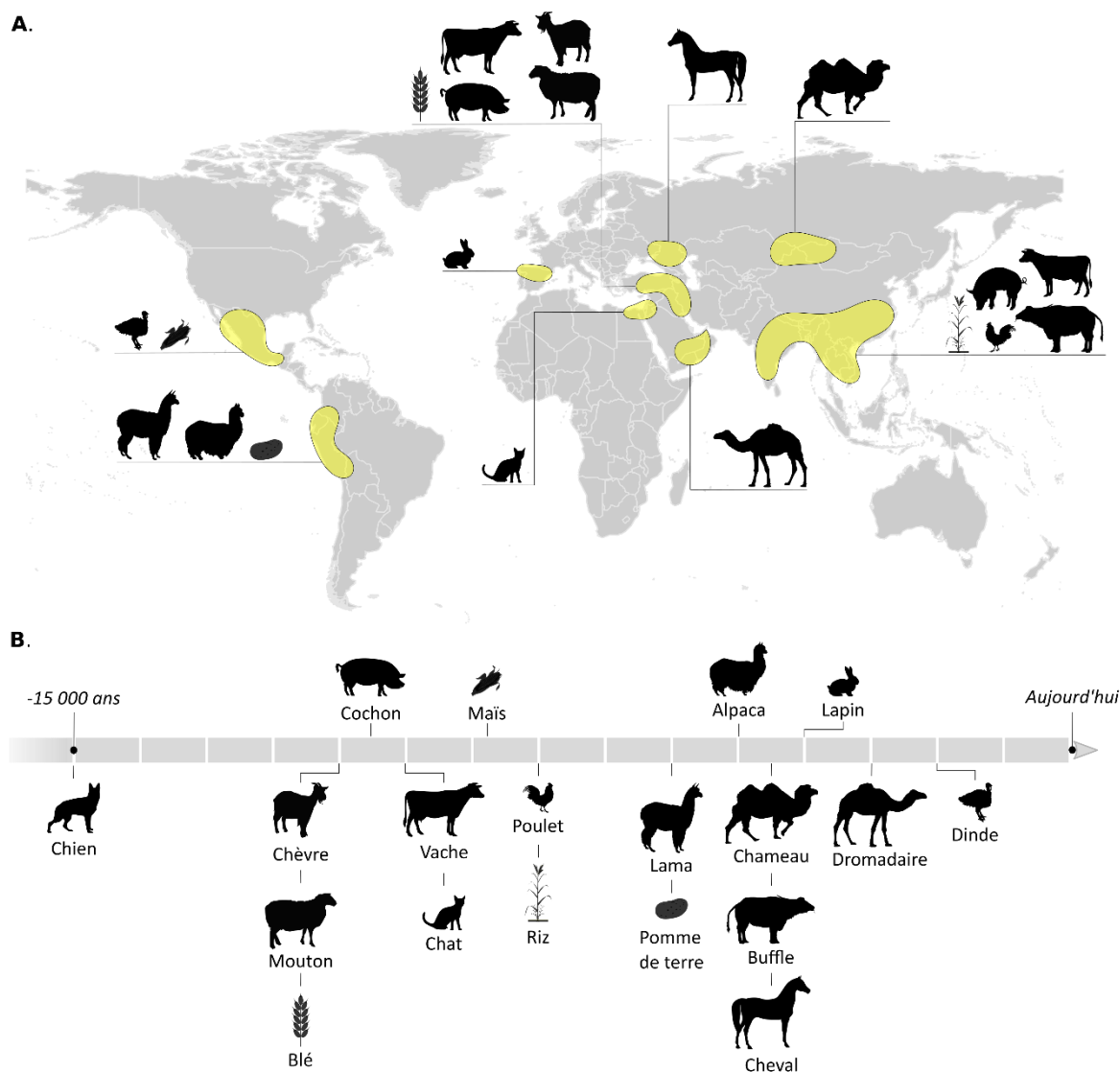


Figure 6 : Domestication de différentes espèces : A. Centres et B. dates de domestications d'après Brush et al., 1995; Piperno et al., 2009; Charmet, 2011; Gross & Zhao, 2014; Larson & Fuller, 2014.

Différents caractères ont ainsi été sélectionnés pour faciliter l'élevage de ces espèces comme la docilité chez les animaux, les caractères liés à la reproduction et la production. Chez les animaux domestiques, l'amélioration des caractères de production a mené dans la fin des 1980 à la création de différentes races par les éleveurs. D'après le rapport sur la gestion des ressources génétiques des animaux d'élevage (Food and Agriculture Organization of the United Nations, 1999) une race se définit comme « soit un sous-groupe spécifique d'animaux domestiques présentant des caractéristiques externes définissables et identifiables qui lui permettent d'être séparé par évaluation visuelle d'autres groupes définis de manière similaire au sein de la même

espèce, soit un groupe pour lequel la séparation géographique et/ou culturelle de groupes phénotypiquement similaires a conduit à l'acceptation de son identité distincte ».

On se rend bien compte que dans ce contexte ce n'est plus à un environnement naturel que les espèces doivent s'adapter ou s'acclimater mais aux conditions d'élevage. Le processus de sélection est anthropique car les humains vont favoriser les croisements entre les animaux portant des caractères d'intérêt afin que la fréquence de ceux-ci augmente dans l'élevage.

Comme pour la sélection naturelle, il est possible de détecter les régions sélectionnées pour des caractères en relation avec la domestication. Des études ont comparé les SNPs (Alberto et al., 2018), variants structuraux (Cumer, 2017) ou régions méthylées (Janowitz Koch et al., 2016) différenciellement présents entre les animaux domestiques et leurs proches sauvages. L'étude de Bélteky et al. (2018) a quant à elle consistait à sélectionner des coqs sauvages (*Gallus gallus*) en fonction de leur peur face aux humains afin de mettre en évidence des régions différenciellement méthylées dans l'hypothalamus des animaux. Les régions détectées dans ces études sont principalement en lien avec le système nerveux ou des caractères de production.

De nombreuses études se sont intéressées à la détection de variants associés à des phénotypes de production. C'est par exemple le cas de la qualité de la viande chez le porc (Amaral et al., 2019; Dall'Olio et al., 2020), la production de lait chez les vaches (Sanchez et al., 2017) ou les chèvres (Scholtens et al., 2020) ou la production d'œuf chez les poules (Wolc et al., 2014).

V. Le contexte de la thèse

Les petits ruminants

Durant ma thèse je me suis focalisée sur les petits ruminants c'est-à-dire les moutons (*Ovis aries*) et les chèvres (*Capra hircus*). Ces espèces ont été les premiers ongulés à être domestiqués il y a environ 11 000 ans dans le croissant fertile (Zeder & Hesse, 2000) à partir de mouflons (*Ovis orientalis*) et de chèvres égagres (*Capra aegagrus*). Ces animaux ont ensuite été dispersés hors de leur berceau de domestication au gré des migrations humaines et d'échanges commerciaux (Vigne, 1988; Zeder, 2017).

Problématiques

Dans le cadre de ma thèse j'ai voulu mettre en évidence des mécanismes génétiques et épigénétiques présents chez les petits ruminants, et discuter des potentiels effets qu'ils pouvaient avoir sur l'adaptation ou l'acclimatation des animaux à leur environnement.

Ma première étude se focalise sur la détection de signatures de sélection chez différentes races de chèvres. L'objectif est de voir ce qui différencie les races caprines françaises. Pour cela nous avons exploré l'histoire de ces races et détecté les régions génomiques sélectionnées chez celles-ci.

Dans le cadre de ma seconde étude, j'ai cherché des marques génétiques mais aussi épigénétiques chez les deux espèces citées. L'objectif est de mettre en évidence un mécanisme d'acclimatation chez des petits ruminants élevés en plein air et de voir si ce sont les mêmes gènes ou voies biologiques qui sont impactés par l'adaptation chez ces mêmes animaux.

Les jeux de données

Signature de sélection chez les chèvres

Afin de répondre à la première question concernant la mise en évidence de signatures de sélection chez les chèvres, nous avons extrait les SNPs présents dans les génomes de chèvres françaises échantillonnées dans le cadre du projet VarGoats (www.goatgenome.org/vargoats.html). Ce projet collaboratif constitue la première étape d'un projet 1 000 génomes caprins permettant de caractériser la diversité génétique des chèvres à une échelle mondiale et ainsi permettre d'étudier l'histoire évolutive des chèvres domestiques.

Acclimation vs adaptation chez les petits ruminants

La recherche d'acclimatation et d'adaptation chez les petits ruminants a été effectuée dans le cadre du projet ClimGen (climgen.bios.cf.ac.uk). Ce projet cherche à identifier les outils génomiques et les biomarqueurs qui peuvent être utilisés pour prévoir l'adaptation des animaux d'élevage aux changements globaux. Dans ce contexte, nous avons utilisé l'échantillonnage des chèvres et moutons marocains issus du projet NextGen (<https://www.epfl.ch/labs/nextgen/>) pour extraire 2 groupes d'animaux vivant dans des climats différents. Durant ce projet un séquençage du génome complet de ces animaux a été effectué et un fichier de variants a été généré (Benjelloun et al., *Submitted to MBE*). Nous avons extrait de ce dernier les SNPs présents chez les animaux que nous avons gardés pour notre étude. Nous avons utilisé la technique

MEDIP-Seq afin d'isoler et séquencer les fragments d'ADN portant des marques de méthylations (Jacinto et al., 2008).

Ce manuscrit va se diviser en deux parties, chacune d'elles traitant une des problématiques présentées ci-dessus. Elles seront suivies d'une discussion générale permettant de mettre en perspective nos principaux résultats face à d'autres recherches concernant l'effet des mécanismes génétiques et épigénétiques sur la relation phénotype – environnement.

MECANISMES GENETIQUES

PRESENTATION DU JEU DE DONNEES VARGOATS

Résumé de l'article

Cet article est la description d'une nouvelle ressource comprenant 1 160 séquences de génome entier de chèvres domestiques (*Capra hircus*) et de leurs apparentés sauvages (35 animaux issus de 8 espèces de chèvres sauvages) produites par le projet VarGoats. Plus précisément, les génomes de 652 chèvres ayant une distribution mondiale ont été séquencés et 508 séquences de chèvres ont été récupérées dans des bases de données publiques. L'alignement sur le génome ARS1, et la recherche de variants dans ces séquences nous ont permis de détecter 74 274 427 SNPs et 13 607 850 INDELS. L'analyse de ces données permettra de caractériser la diversité génétique des chèvres à une échelle mondiale et d'étudier l'histoire évolutive des chèvres domestiques.

État d'avancement :

Cet article doit être mis au format « Short Communication » de la revue *Genetics Selection Evolution* avant soumission.

VarGoats project: a 1,160 whole-genome sequence dataset to dissect *Capra hircus* global diversity

Laure Denoyelle, Estelle Talouarn, Philippe Bardou, Licia Colli, Adriana Alberti, Coralie Danchin, Marcello Del Corvo, Stéfan Engelen, Céline Orvain, Isabelle Palhière, Rachel Rupp, Julien Sarry, Mazdak Salavati, Marcel Amills, Emily Clark, Paola Crepaldi, Thomas Faraut, Clet Wandui Masiga, François Pompanon, Benjamin D Rosen, Alessandra Stella, Curtis P Van Tassell, Gwenola Tosser-Klopp and the VarGoats Consortium

Abstract

Background

Since their domestication 10,500 years ago, goat populations with distinctive genetic backgrounds have adapted to a broad variety of environments and breeding conditions. The VarGoats project is an international 1,000 genomes resequencing program designed to understand the consequences of domestication and breeding on the genetic diversity of domestic goats, as well as to elucidate how speciation and hybridization have modeled the genomes of a set of species representative of the genus *Capra*.

Findings

A dataset comprising 652 sequenced goats and 508 public goat sequences, including 35 animals from 8 wild species, has been collected worldwide. We identified 74,274,427 SNPs and 13,607,850 INDELs when aligning sequences to the latest version of the goat reference genome (ARS1). By performing a genetic differences analysis, we have found that goats from Africa, Asia and Europe tend to group in independent clusters. Since goat breeds from Oceania and Caribbean (Creole) all derive from importations, they are distributed along the tree according to their ancestral geographic provenance.

Conclusions

Here we report an unprecedented international effort to characterize the genome-wide diversity of domestic goats. This wide panel of sequenced individuals represents an opportunity to ascertain how the demographic and selection processes associated with post-domestication history have shaped the diversity of this species. Data generated in the project might be also extraordinarily useful to identify deleterious mutations as well as

polymorphisms with causal effects on complex traits, thus generating new knowledge that could be used in genomic prediction and genome-wide association studies.

Keywords

Whole-genome sequencing, Goat, *Capra*, Genomic variation, Genetic diversity

Data description

Context

Goats (*Capra hircus*) were domesticated around 10,500 years ago in the Fertile Crescent (Zeder & Hesse, 2000). Since then, goat populations have gradually specialized and now provide milk, meat, fibers (Skapetas & Bampidis, 2016) and fuel from manure, to become an important pillar of the livestock sector in many countries around the world. According to FAOSTAT, (www.fao.org/faostat/en) the world goat population increased by 38% since 1994, reaching over 1.03 billion heads in 2017. This makes goats the 3rd most important ruminant production species. Goats have undergone an intense process of adaptation and occupy several diverse agroecological regions of the world.

In 2010, the first goat reference whole-genome sequence was assembled (Dong et al., 2013) and the International Goat Genome Consortium (IGGC) was created to further support the development of genomic tools for studying the genetic variation of domestic goats. Later on, a 50k-SNP (Single Nucleotide Polymorphism) panel, the GoatSNP50 BeadChip was developed (Tosser-Klopp et al., 2014) to facilitate QTL discovery through genome-wide association studies (Martin et al., 2016, 2017; Mucha et al., 2018) and to enable genomic selection (Carillier et al., 2013; Carillier-Jacquin et al., 2016). This resource became a tool that facilitated collaboration, because data from diverse sources were interchangeable. The ADAPTmap project (Liu et al., 2018; Bertolini, Cardoso, et al., 2018; Bertolini, Servin, et al., 2018; Cardoso et al., 2018; Colli et al., 2018; Stella et al., 2018; Talenti et al., 2018), was one such collaboration, compiling goat genotypes from across the globe and exploring genetic diversity. However, the ADAPTmap dataset was limited to a subset of countries, and did not fully represent the variability of the *Capra* species worldwide. Indeed, wild goat species other than the bezoar (*Capra aegagrus*) were not investigated. Moreover, data generated from SNP chips are known to be affected by ascertainment bias (Albrechtsen et al., 2010), a limitation that can be overcome by carefully filtering whole genome sequencing data of sufficient depth (Benjelloun et al., 2019). Here we report an international resequencing effort: the VarGoats

project, which has generated a data set of 1,160 goat genomes through the generation of new sequencing data as well as by retrieving existing genome sequences from public databases. The VarGoats dataset will pave the way towards obtaining an unprecedented perspective about the natural and human-mediated evolutionary forces which have shaped genomes and the genetic diversity of domestic goats.

Ethics Statement

Blood collection or ear-tags samples were carried out in accordance with the national regulations from the countries where such samples were collected. In the case of samples sequenced at Génoscope (Evry, France), DNA was imported into France either with authorization 31 555 50, delivered on May, the 24th 2016 for European countries or covered by an import permit from DDPP (Direction Départementale de la Protection des Populations) for non-European countries. The only exception was the African animals for which sequencing was performed at Edinburgh Genomics. The DNA for these libraries was imported into Scotland by permission of the Scottish Government Animal Health and Welfare Division and the UK under generic license IMP-GEN-2008-03.

Individual selection

Animals sequenced in the VarGoats project were selected to represent the international genetic diversity in goats. We selected 468 animals already included in ADAPTmap dataset (with 457 public genotypes) and added 184 animals from different breeds and locations in order to improve the representation of the current levels of genetic diversity in goats. After sequencing all samples (652), 217 additional samples were retrieved from NextGen Consortium projects (PRJEB3134, PRJEB3135, PRJEB4371, PRJEB5166, PRJEB3136 and PRJEB5900 studies) and 291 additional sequences from public sequence data repositories (extraction from NCBI database on the 2019/02/07) were added to the overall VarGoats data set.

However, some individuals were selected for specific research purposes. This is the reason why we can observe a distortion in the number of sequenced Alpine and Saanen for instance. Besides some individuals can be closely related, an analysis is therefore advised to anyone using the data for any analysis.

Among the 1,160 animals, 35 represented wild *Capra* species (3% of the dataset, *C. aegagrus*, *Capra caucasica*, *Capra cylindricornis*, *Capra falconeri*, *Capra ibex*, *Capra nubiana*, *Capra*

pyrenaica and *Capra sibirica*) distributed with domestic goats native to Africa (450 goats, 39%), Europe (443 goats, 38%), Asia (226 goats, 20%), Oceania (25 goats, 2%) and Caribbean (16 goats, 1%). This geographical distribution is shown on Figure 7.

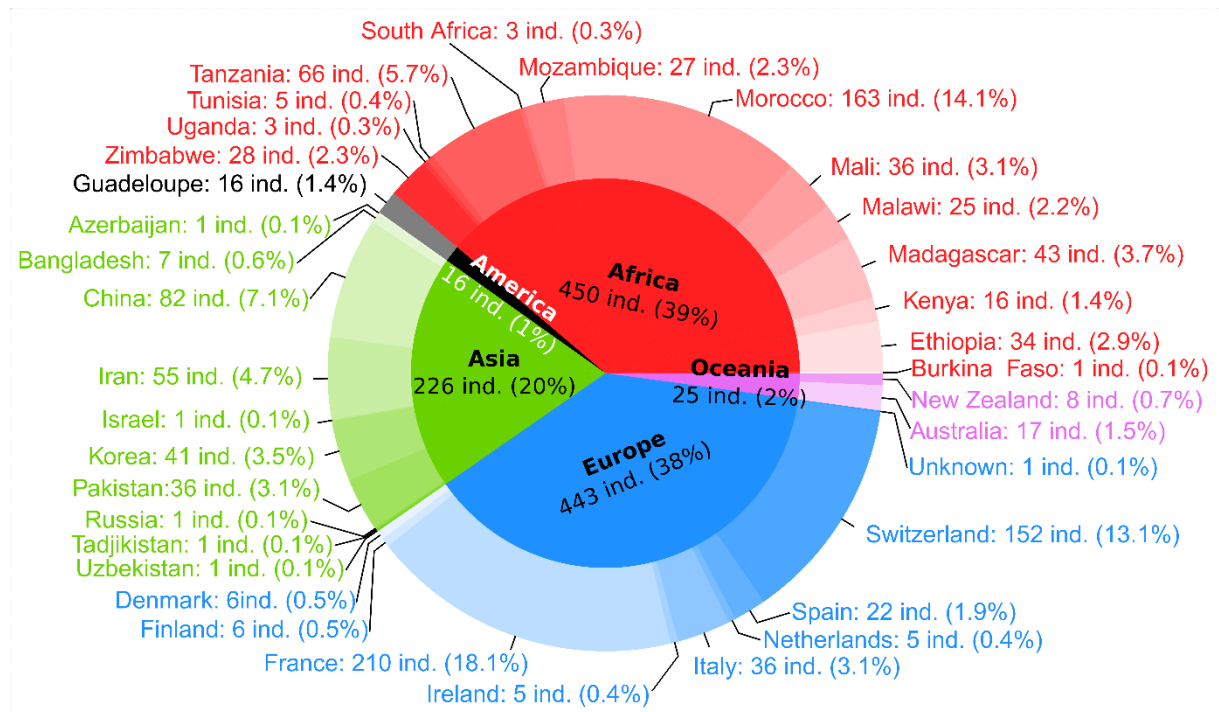


Figure 7: Geographical distribution of the 1,160 sampled *Capra* individuals included in the dataset.

A unique identification (original ID) was defined for each individual. For resequenced animals and 139 animals published by University of Bern, it starts with 4 letters: the first two correspond to the country of origin (based on ISO 3166-1 alpha-2 codes), the following two define species (CH for *C. hircus*, CA for *C. aegagrus* etc...). Letters 5 to 7 indicate the breed (UNK for unknown) followed by a number that identifies each individual specifically (the meaning of the breed and country codes are explained in Supplementary Table S1). For NextGen data, the sample alias was used as the original ID whereas we used the run accession number for other public data to avoid any duplicate. Concerning the public data, the original ID corresponds to the run accession number.

We also proposed a working name for each individual to facilitate the analysis interpretation for unknown-breed animals. This identifier is composed in the same way as the original IDs presented above, except that the breed code is followed by the 2 letters of the country before the 4-digit number that identifies each individual.

For public data, information was extracted from BioSample (NCBI, <https://www.ncbi.nlm.nih.gov/>) in order to determine the sex, breed and geographic origin of

each animal. For the VarGoats samples, information was collected by VarGoats collaborators. All individual details are provided in Supplementary Table S2.

As 50k genotypes were available for AdaptMap individuals, we wanted to make sure that no sample had been mixed up. Among available 50k genotypes, 457 were retrieved in the framework of the AdaptMap project which collected samples of 4,653 animals across 148 populations and 35 countries (Stella et al., 2018). Concordance rates (CR) between sequence variants and 50k genotypes of these 457 individuals were checked. Among the 46,654 SNPs of AdaptMap genotypes, 44,691 were found in the vcf files. For each individual the genotype concordance rate was calculated for the SNPs with available genotypes. As shown on Figure 8, the CR is lower for low coverage sequences. For 14 individuals, albeit decent sequencing depth (between 7.2 and 23.6), the CR was lower than 70%, thus indicating a technical problem resulting in a lack of correspondence between the genotyped and sequenced samples (Figure 8, A.).

To check the correctness of the breed indicated in the ID of these animals we performed a Principal Component Analysis (PCA) with PLINK (version 1.9, Purcell et al., 2007) taking into account all the animals belonging to the breeds with problematic samples. The genetic data come from the SNPs contained on autosomes in the VCF file. First of all we removed markers with more than 5% of missing data. This filtering step yielded 1,890,194 SNPs which were pruned, using the indep-pairwise function on PLINK. Each SNP that had a squared correlation value (R^2) greater than 0.1 with another SNP of a 50-SNPs sliding window moving by 10 SNPs each time (--indep-pairwise 50 10 0.1) was removed. This step reduced their number to 667,949 which were used for the analysis. Twelve of the fourteen animals clustered with their breed counterparts (Figure 8, B.), while the breed assignment of 2 individuals (AUCH-BOE-0038 and ITCH-CCG-0014) could not be confirmed.

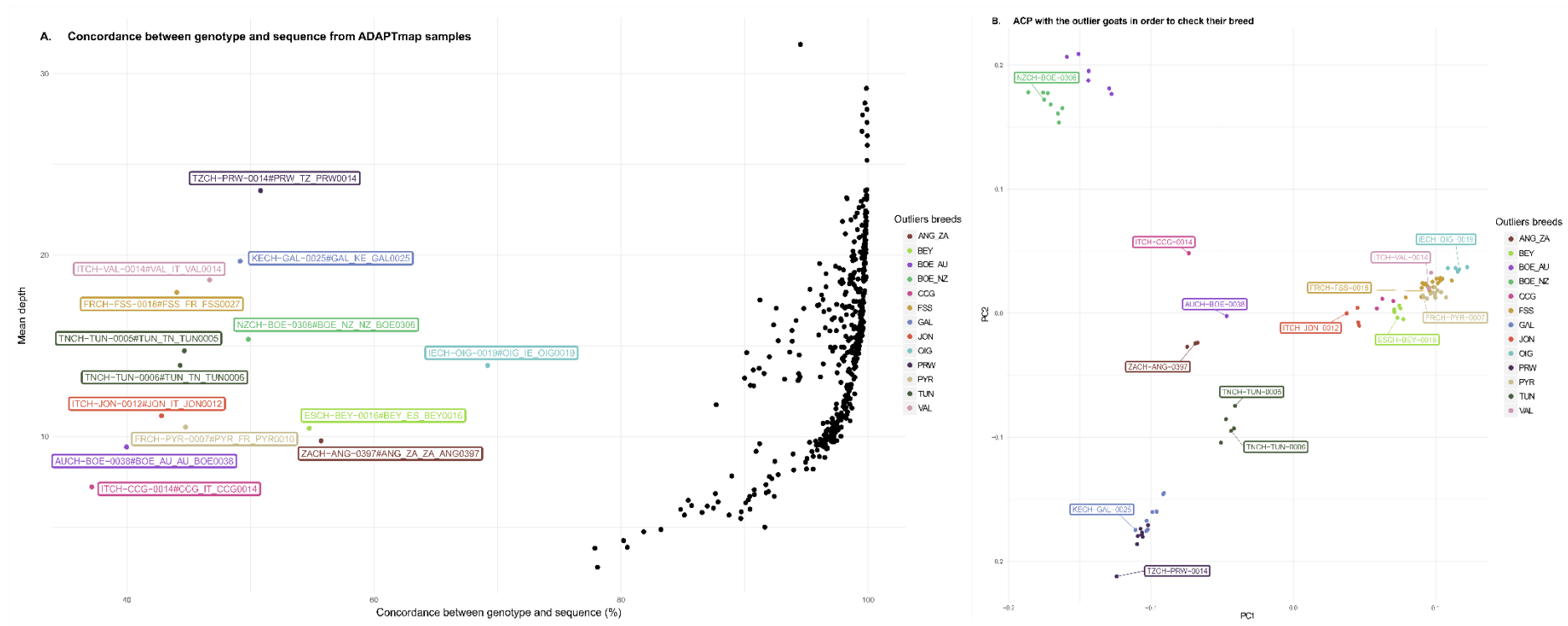


Figure 8: AdaptMap samples analysis: A.concordance rates of the 457 AdaptMap individuals calculated on the 46,654 common SNPs between sequence data and 50k genotypes; B. PCA to verify the breed of outlier animals.

AdaptMap IDs of the 11 animals with missing genotypes are italicized. The 14 low concordance rate individuals are written in bold. The 2 outliers with unconfirmed breed status were labeled as UNK. These 25 problematic individuals have been reported using an asterisk in Supplementary Table S2.

As recommended for any dataset, we advise to perform a global analysis, like a PCA or a structural analysis on the dataset or subset before deeper and specific studies to identify any outlier.

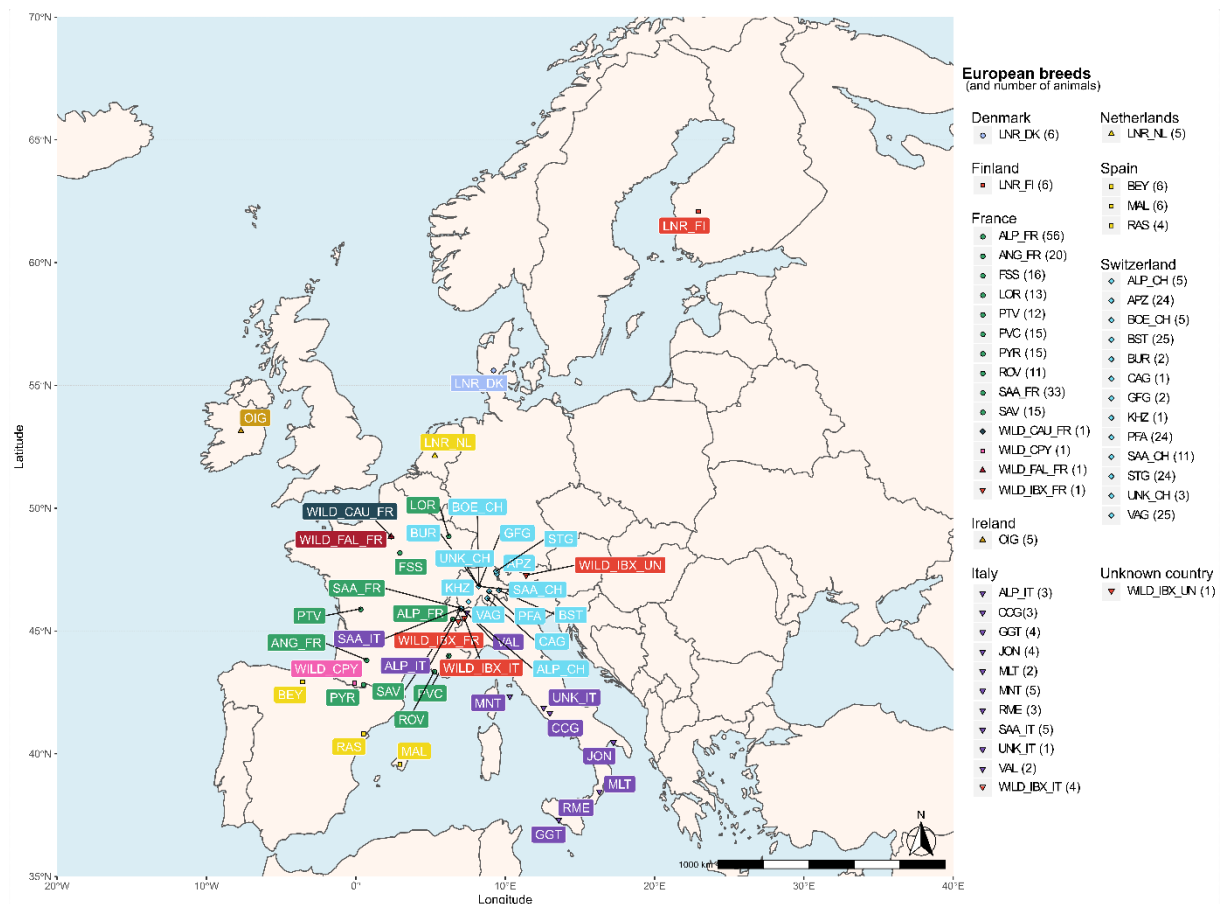


Figure 9: Geographic distribution of European breeds represented by 3 letters corresponding to the breed code (Supplementary Table S1). If a breed is present in several countries, the breed code is followed by the country code (2 letters). Each combination of color and symbol corresponds to domestic goats in a single country; the wild goats are identified with a specific color and symbol.

Breed information

The VarGoats dataset encompasses a total of 8 wild species (BEZ, CAU, CPY, CYL, FAL, IBX, NUB, SIB) and 125 breeds belonging to the *C. hircus* species (Supplementary Table S1). They are distributed as follows: 46 in Africa (35%), 40 in Europe (30%), 34 in Asia (25%), 4 in Oceania (3%) and 1 in Caribbean (1%).

The geographical distribution of the breeds was determined based on bibliographical research (www.fao.org; www.racesdefrance.fr; eng.agraria.org/goat.htm; etc.) which allowed us to define more precisely the provenance of each one of them (barycentre of province or geographic region). In the absence of precise information, GPS coordinates of the barycentre of the country of origin were assigned to the sample. Their locations are represented for each continent in Figures 9, 10, 11, except for Creole which is the only one from Caribbean (West Indies, to be more precise). The dataset includes cosmopolitan breeds such as Alpine, Boer or Saanen as well as local breeds that are only present in specific regions of the world.

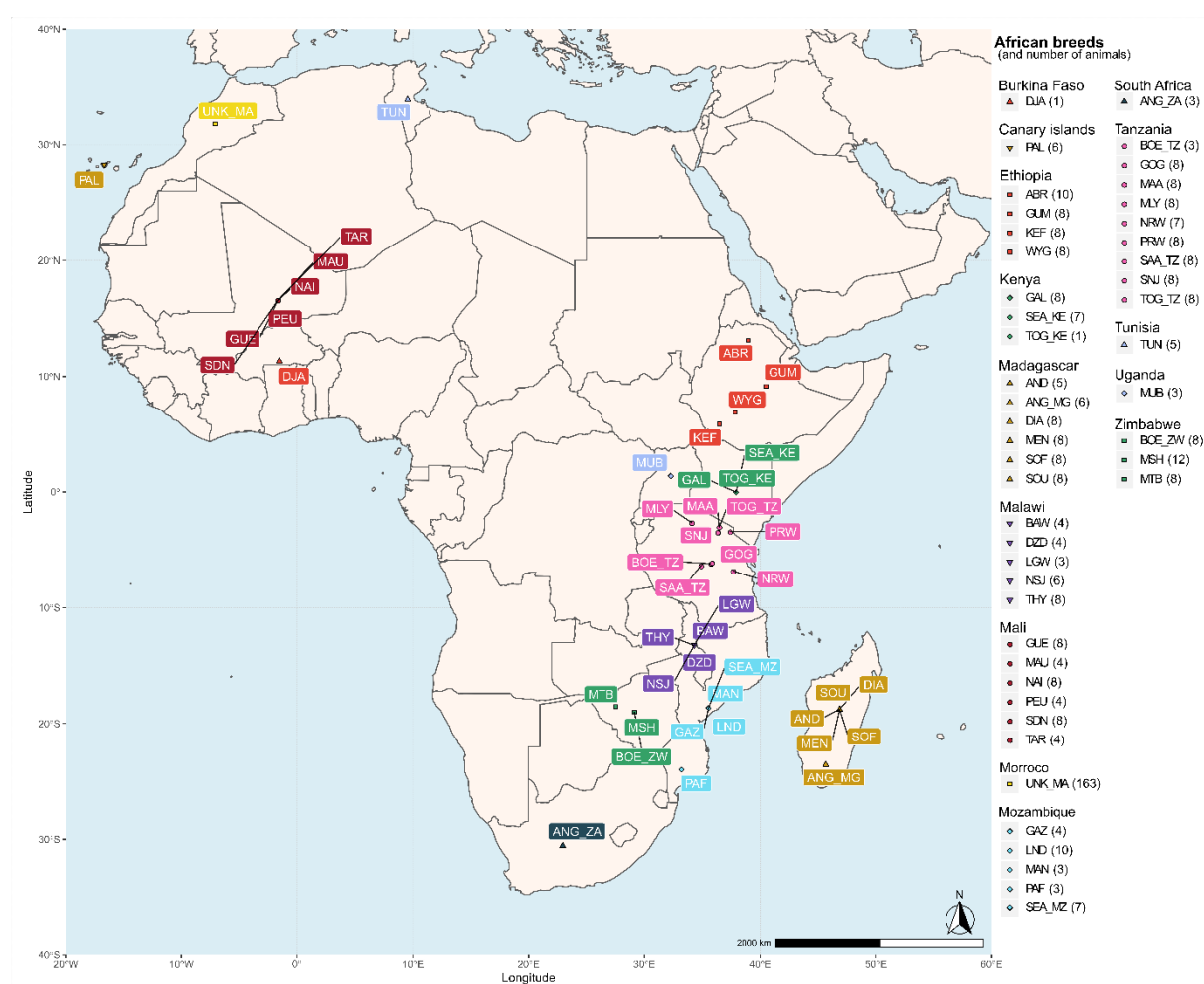


Figure 10: Geographic distribution of African breeds represented by 3 letters corresponding to the breed code (Supplementary Table S1). If a breed is present in several countries, the breed code is followed by the country code (2 letters). Each combination of color and symbol corresponds to domestic goats in a single country.

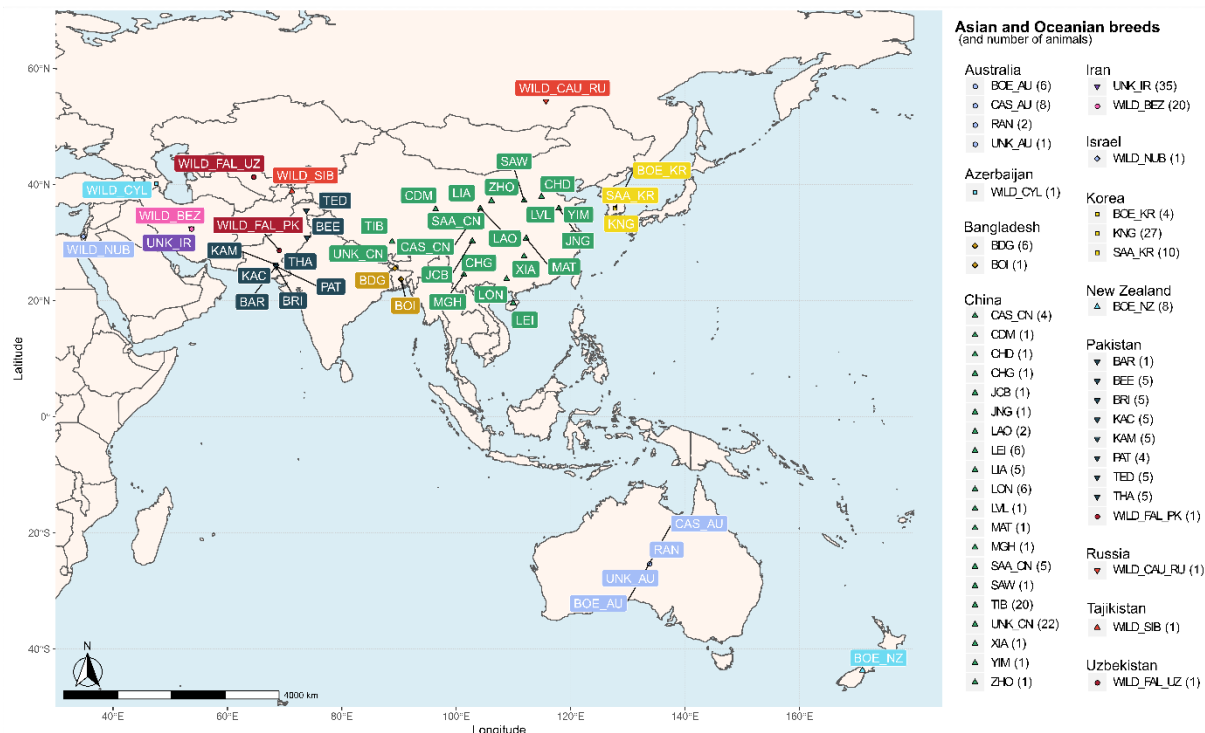


Figure 11: Geographic distribution of Asian and Oceanian breeds represented by 3 letters corresponding to the breed code (Supplementary Table S1). If a breed is present in several countries, the breed code is followed by the country code (2 letters). Each combination of color and symbol corresponds to domestic goats in a single country; the wild goats are identified with a specific color and symbol.

Methods

Library construction and sequencing

The library preparation protocol was chosen on the basis of the DNA extraction yield. When available, 250 ng of genomic DNA were sonicated using the E210 Covaris instrument (Covaris, Inc., USA) and the NEBNext DNA Modules Products (New England Biolabs, MA, USA) were used for end-repair, 3'-adenylation and ligation of NextFlex DNA barcodes (Bioo Scientific Corporation). After two consecutive 1x AMPure XP clean-ups, the ligated fragments were amplified by 12 PCR cycles by using the Kapa Hifi Hotstart NGS library Amplification kit (Kapa Biosystems, Wilmington, MA, USA), followed by 0.6x AMPure XP purification. When the nucleic acids extraction yielded low DNA quantities, 10-50 ng of genomic DNA were sonicated. Fragments were end-repaired, 3'-adenylated and NEXTflex DNA barcoded adapters were added by using NEBNext Ultra II DNA Library prep kit for Illumina (New England Biolabs). After two consecutive 1x AMPure clean-ups, the ligated products were PCR-amplified with NEBNext® Ultra II Q5 Master Mix included in the kit, followed by 0.8x AMPure XP purification.

All libraries were subjected to size profile analysis, with an Agilent 2100 Bioanalyzer (Agilent Technologies, USA) as well as to qPCR quantification (MxPro, Agilent Technologies, USA). Libraries were sequenced using 151 base-length read chemistry in a paired-end flow cell on an Illumina HiSeq 4000 sequencer (Illumina, USA). On average, 120 million paired-end reads were obtained for each sample after clean up (12X expected genomic coverage).

Libraries for the African animals were prepared by Edinburgh Genomics using the TruSeq Nano DNA High Throughput library preparation kit (Illumina, USA). The following preparation protocol was applied: 1 µg of genomic DNA was sheared to fragments of 450bp mean size using a Covaris LE220 focused-ultrasonicator (Covaris, Inc., USA). DNA fragments were then blunt ended, A-tailed, size selected and adapters were ligated into fragment ends according to the Illumina TruSeq PCR-free library preparation kit protocol. Insert size of the libraries was evaluated using a PerkinElmer LapChip GX Touch with an HT DNA 1k/12K/Hi SENS LabChip and HT DNA Hi SENS Reagent Kit (PerkinElmer, Inc., MA, USA). Final library concentration was calculated by qPCR using a Roche LightCycler 480 (Roche Molecular Systems, Inc., Switzerland) and a Kapa Illumina Library Quantification kit and Standards. Then libraries were normalized to a loading concentration of 150 nM. All the library processing steps were carried out on Hamilton MicroLab STAR liquid handling robots coupled to BaseSpace Clarity LIMS X Edition. Libraries were loaded into a HiSeq X Flow cell v2.5 and clustered using an Illumina cBot2. Clustered flow cells were sequenced at 15X coverage using a HiSeq X Ten Reagent kit v2.5.

An Illumina filter was applied to remove the unreliable data from the analysis. Raw data were filtered to remove clusters with excessive intensity in bases other than the called one. Adapters and primers were removed from the whole read and low quality nucleotides were trimmed from both ends (quality value lower than 20). Sequences between the second unknown nucleotide (N) and the end of the read were also removed. Reads shorter than 30 nucleotides after trimming were discarded. Finally, the reads and their mates that mapped onto run quality control sequences (PhiX genome) were removed.

These trimming steps were achieved using an in-house software based on the FastX package (<http://www.genoscope.cns.fr/externe/fastxtend/>).

Data generation and preparation

- Read alignment and variant calling

Each sample was processed using a pipeline based on the domain reference tools and the Genome Analysis Toolkit (GATK) best practices: BWA version 0.7.15 for alignment (Li & Durbin, 2009), SAMtools version 1.6 for handling SAM/BAM formats and calling variants (Li et al., 2009), Picard tools version 2.1.1 for labelling duplicated reads (<http://broadinstitute.github.io/picard>), as well as GATK version 3.6 for Insertion/Deletion (INDEL) realignment, base recalibration and calling variants (McKenna et al., 2010), BCFtools version 1.6 for handling VCF/BCF formats, Freebayes version 1.1.0 for calling variants (Garrison & Marth, 2012), and snpEff version 4.3t for VCF annotation (Cingolani et al., 2012). Reads were mapped to the latest ARS1 genome version (Genbank accession GCA_001704415.1) of the *C. hircus* species (Bickhart et al., 2017) using BWA-MEM software with default parameters except for “-t 14 -M” and “-R” to add read groups. The SAM output files were converted to sorted BAM using SAMtools.

Pre-processing steps (marking duplicates, INDEL-based realignment and base quality score recalibration - BQSR) were done using Picard-MarkDuplicates and GATK. The “known variants” file needed for the BQSR step was computed on a subset of 13 samples (see details in the next paragraph). Variants fulfilling the following conditions were included in the “known variants” file: (1) presenting at least 6 genotypes harbouring an alternative allele ("snpSift filter 'countVariant(>6'") (2) being called by both Freebayes and GATK-HaplotypeCaller.

Variant calling per sample was done with GATK-HaplotypeCaller in ERC mode with a minimum read mapping quality of 30 (this is required to consider a read as valid) and a minimum phred-scaled confidence threshold of 30 (“-stand_call_conf 30.0 -mmq 30 -ERC GVCF -variant_index_type LINEAR -variant_index_parameter 128000”).

Due to the large number of samples, GVCF files were combined (CombineGVCFs) before the joint genotyping step (GenotypeGVCFs) to produce the raw VCF files by chromosome/scaffold.

- Filtering process

A Variant Quality Score Recalibration (VQSR) step was performed on the raw VCF files. In order to set up training resource sets for VQSR calibration, 13 goats (AUCH-CAS-0038, BFCH-DJA-0012, CHCH-BOE-0229, ESCH-PAL-0008, ESCH-RAS-0011, ETCH-ABR-0036, FICH-LNR-0122, FRCH-ALP-0030, FRCH-CRE-0014, FRCH-SAA-0032, ITCH-GGT-0026, MZCH-PAF-0003 and

ZACH-ANG-0374), representing 13 breeds were chosen out of the 248 animals sequenced in the first batch. They belong to 11 out of the 15 gene pools determined by Colli et al. (2018) with the software Admixture (Alexander et al., 2009). We added one individual from an inbred breed (Palmera) and an animal of the Creole breed (different from the Creole individuals genotyped in the ADAPTmap dataset), as a representative of the American gene pool. Two true sites training resources were built. The first one, corresponding to the highest quality calls, used the variants consistently identified with GATK, Mpileup and Freebayes ("known=false,training=true,truth=true,prior=15.0"). The second one used the 60,000 SNPs selected by Tosser-Klopp et al. (2014) to generate the set of polymorphisms included in the Goat SNP50 BeadChip (Illumina) ("known=false,training=true,truth=true,prior=12.0"). The non-true sites training resource was built using the variants exclusively called by GATK ("known=false,training=true,truth=false,prior=10.0").

We only retained biallelic SNPs with a GATK quality score (QUAL) over 100 and with at least two individuals carrying the alternative allele.

Filtering resulted in a high confidence set of 74,274,427 SNPs and 13,607,850 INDELS.

- Main features of the dataset

The GCF_001704415.1_ARSL1_genomic.gff annotation file was used to annotate the variants using SnpEff. Table IV summarizes the variant types found on each chromosome.

Table IV: INDELs distribution and annotation of SNPs identified when sequences were aligned to the ARS1 reference genome and using the GCF_001704415.1_ARS1_genomic.gff annotation file.

CHI	Indels		SNPs														
	Number	Number	Number of effects	Downstream	Exon	Intergenic	Intron	Splice acceptor	Splice donor	Splice region	Transcript	Upstream	UTR 3'	UTR 5'	Missense	Nonsense	Silent
1	818757	4445377	12032012	458743	66648	2869811	4038272	79	90	4866	4076351	453685	51524	11943	21817	172	34232
2	681619	3704710	11025471	500555	86105	2218485	3767113	74	80	7110	3886499	493018	53529	12903	27607	155	47266
3	582990	3219852	9723073	683259	99623	1904576	3067833	97	147	7253	3210217	676123	57845	16100	34728	321	49798
4	612055	3356824	10217296	402190	63972	1869682	3685704	72	113	4137	3747652	394011	40808	8955	21883	229	32329
5	589455	3254170	10416604	608369	90816	1884801	3521587	87	117	6961	3628991	604367	56456	14052	28723	298	47717
6	628940	3423894	9712976	356081	49936	2216046	3311482	49	89	3940	3376966	352520	36787	9080	16818	109	24491
7	534662	2892084	9828378	636813	159888	1819601	3156821	133	111	10127	3347312	631783	51708	14081	63405	447	83894
8	558356	3063425	9133856	382194	59762	1965117	3107183	63	69	4452	3189032	374406	42124	9454	20793	209	30688
9	461502	2494211	7674249	270908	48106	1572204	2714425	138	46	3556	2769209	265609	24246	5802	15956	125	23479
10	493652	2698185	9192746	476511	71317	1511953	3242782	52	96	5079	3352743	472321	47467	12425	24247	221	37332
11	515256	2831023	9890955	553577	92208	1724793	3395147	92	136	7371	3508281	543462	52415	13473	27137	241	51761
12	491274	2667175	5825552	193632	34358	1765985	1782658	33	48	2654	1831608	190567	19095	4914	12736	158	17266
13	389692	2171122	7859635	394282	60365	1203702	2832529	77	91	4557	2919170	390791	43851	10220	19308	201	32517
14	486379	2667881	7004865	265510	38581	1771173	2298648	41	52	3041	2326532	265584	27828	7875	11194	96	19368
15	419873	2348100	6935036	479502	71589	1493439	2136669	53	84	4060	2229600	472382	38348	9310	25431	347	33595
16	394016	2171945	6453431	342598	55559	1256466	2169250	56	60	4402	2234817	339736	40836	9651	18793	155	29062
17	365130	1965116	5032075	269775	45016	1219875	1562271	38	91	3551	1631168	265631	26028	8631	13371	108	24014
18	328462	1847282	5375176	640087	97496	1052854	1380659	112	152	6698	1501048	636762	40352	18956	35046	523	45254
19	303457	1657902	6819329	622933	100340	808006	2215917	76	65	7868	2374994	615022	58429	15679	33492	286	57841
20	375188	2088922	4276376	152982	22276	1434583	1232290	25	25	1684	1260516	152063	16729	3203	7060	39	11033
21	346701	1955870	5590244	336130	53444	1238872	1791260	100	118	3630	1805530	327239	26317	7604	16845	199	24557
22	291852	1579747	7603338	316589	53838	756258	3024571	51	52	4586	3093398	309797	34524	9674	15493	103	29721
23	261537	1461938	5452237	421685	68240	852094	1785250	47	91	4651	1878809	402412	28019	10939	24936	313	34341
24	321051	1769342	5080572	172939	27021	1192978	1727081	12	27	2061	1766194	166665	21637	3957	8023	62	14178
25	205425	1182353	4215848	386841	64350	533065	1350422	47	88	4715	1456023	375692	33162	11443	19585	144	32896
26	262405	1481440	5195116	225962	34955	839548	1896442	41	74	2605	1943238	222634	23591	6026	11600	139	16083
27	232467	1278104	3190769	130257	20512	844522	1003995	32	31	1512	1037321	132486	15547	4554	6764	43	10924
28	228740	1306693	5179451	191839	29690	693876	2002455	47	27	2308	2040071	191412	22278	5448	9736	66	14952
29	257510	1508568	6275068	344027	54609	844994	2303306	65	78	3832	2360477	333712	22868	7100	17940	133	28259
MT-Others	1169447	5781172	9317433	534885	76267	4875855	1611366	133	149	4586	1648637	520101	33765	11689	22371	370	31779
Total	13607850	74274427	221529167	11751655	1896887	46235214	73115388	2022	2497	137853	75432404	11571993	1088113	295141	632838	6012	970627

Sex assignment

As the sex was not determined a priori for each sequenced animal, 100 SNPs mapping on to Y-chromosome contigs were selected (NW_017189563.1, NW_017189610.1, NW_017189618.1, NW_017189628.1, NW_017189685.1, NW_017189696.1, NW_017189885.1, NW_017189985.1, NW_017190040.1, NW_017190154.1, NW_017195709.1). An unknown genotype for these 100 Y-chromosome variants was used as evidence that the sequenced individual was a female. Therefore, for each variant, the number of unknown genotypes was counted for 69 Alpine and Saanen goats from the VarGoats dataset for which the sex was known (i.e., 65 males and 4 females). Among the 100 variants, 17 were retained as they successfully predicted the sex of all these individuals. For assigning the sex of other sampled goats, an individual was considered to be a male when it showed more than genotypes for more than half of 17 aforementioned variants (i.e. 8 genotypes).

Genetic diversity

A subset of 100,000 SNPs was randomly extracted, using the thin-count function from PLINK, from the 667,949 SNPs described for the PCA analysis (individual selection section). Starting from this reduced dataset, a matrix of between-population Reynolds distances was calculated with hapFLK v.1.3.0 (Bonhomme et al., 2010; Fariello et al., 2013) and used to construct a Neighbor-Joining tree (Figure 12). The wild species closest to domestic goats (*C. aegagrus*) was used as an outgroup to root the tree.

In the Neighbor-joining tree displayed in Figure 12, goats grouped according to their ancestral geographic provenance (Africa, Europe, Asia or Middle East). This outcome was consistent also for the Oceanian breeds, since Boer goats from Australia (BOE_AU) and New Zealand (BOE_NZ) clustered with other populations of the Boer breed sampled in Africa. Similarly, Australian Cashmere (CAS_AU) grouped with Angora goats due to their common Middle Eastern origin. A similar clustering pattern based on the ancestral geographic origin could be also observed in transboundary breeds sampled in different countries (i.e., Alpine, Boer and Saanen).

The first branching among domestic goats separated the Asian breeds from the remaining populations. This branch carried the most basal cluster which was mainly composed of breeds originating in South-Western Asia (Iranian and Pakistani goats) and thus geographically close

to the center of domestication (Zeder & Hesse, 2000). The second branching included a cluster composed by long-haired breeds, then, two additional African and European clusters emerged in less basal positions. In each continental group, geographically coherent sub-clusters were clearly discernible (e.g. Northern and Southern Europe, North-Western Africa, Eastern Africa and Madagascar).

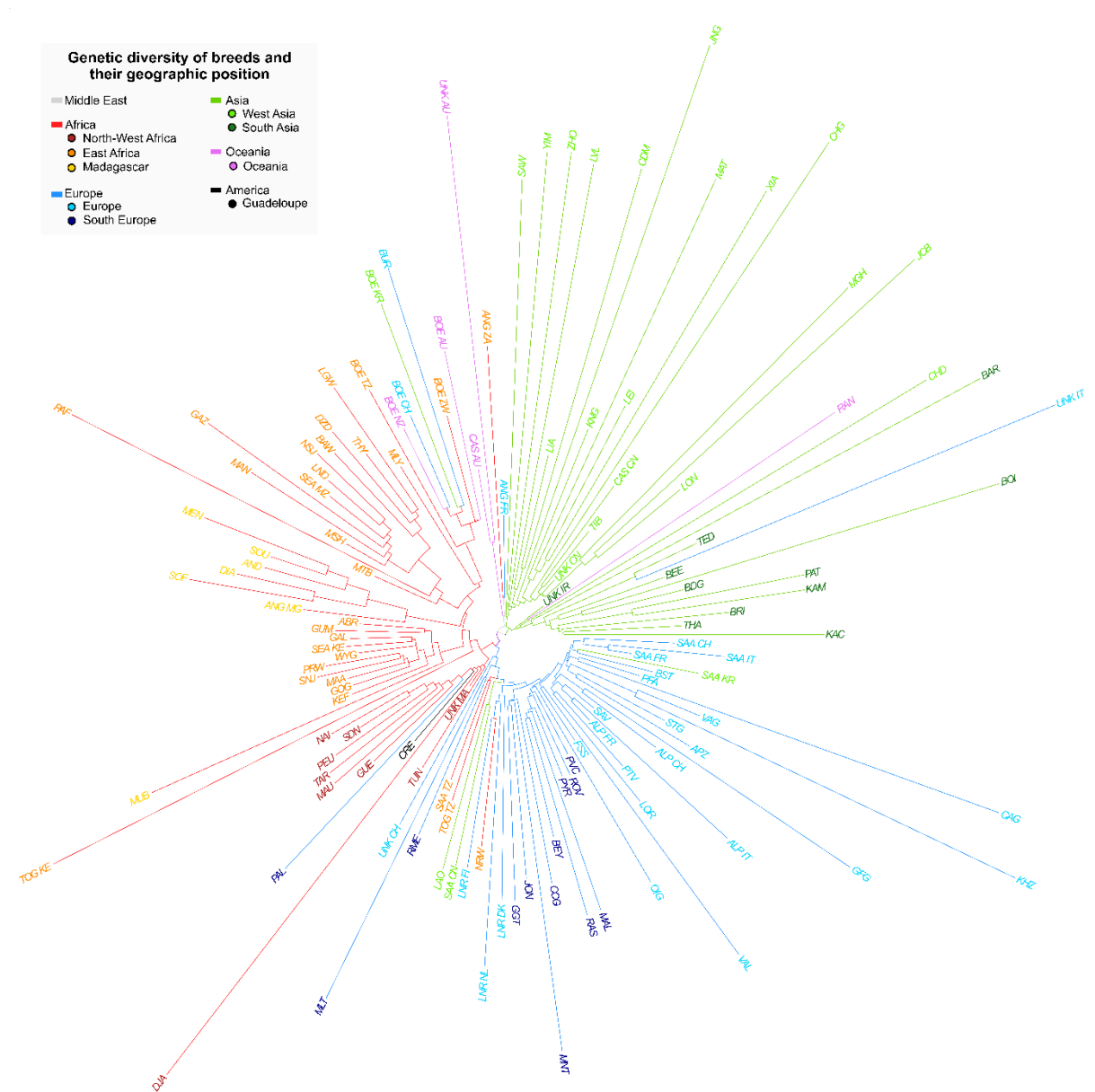


Figure 12: Neighbor-joining tree representing the genetic diversity of domestic goat populations analyzed in the context of the VarGoats project.

Concerning the two animals with an ambiguous breed status: one of them (ex AUCH-BOE-0038) grouped with long-haired goats while the other one (ex ITCH-CCG-0014) did so with Pakistani goats.

Re-use potential

The analyses planned in the framework of the VarGoat project are expected to provide a deeper understanding of the evolutionary history of domestic goats and their wild relatives. The VarGoats data set will be updated with publicly available sequences plus a last batch (60 sequenced at the Genoscope + 38 public sequences extracted on the 2020/02/14), by the end of 2020, containing new sequences generated by our Consortium.

Supplementary tables

Supplementary Table S1: Number of individuals per breed and country of origin

Description: Distribution of sequenced individuals per breed and abbreviations explanation

http://www.goatgenome.org/vargoats_data_access.html

Supplementary Table S2: Detailed information for each sequenced individual

Description: Description of each individual (species, breed, country of origin, localization, sex, sample provider and details about its sequence)

http://www.goatgenome.org/vargoats_data_access.html

ANALYSE DES CHEVRES FRANÇAISES

I. Présentation des races françaises

Le Ministère de l'Agriculture français reconnaît officiellement 14 races caprines françaises répertoriées dans l'Arrêté du 29 avril 2015 (Tableau V, Ministre de l'agriculture, de l'agroalimentaire et de la forêt, 2015).

Tableau V : races reconnues par le Ministère de l'Agriculture et leur appartenance aux catégories « races locales » ou « races menacées d'être perdues pour l'agriculture » d'après l'annexe de l'Arrêté du 29 avril 2015.

Races reconnues	Locales	Menacées	Races reconnues	Locales	Menacées
Alpine	Non	Non	Massif Central	Oui	Oui
Angora	Non	Non	Péï	Oui	Oui
Boer	Non	Non	Poitevine	Oui	Oui
Corse	Oui	Non	Provençale	Oui	Oui
Créole	Oui	Oui	Pyrénées	Oui	Oui
Des Fossés	Oui	Oui	Rove	Oui	Non
Lorraine	Oui	Oui	Saanen	Non	Non

La race Savoie a été officiellement reconnue le 31 juillet 2020 (Ministre de l'agriculture et de l'alimentation, 2020) mais n'apparaît pas encore dans l'annexe du premier Arrêté.

Conformément à l'article 4, l'Institut National de la Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE) est l'organisme habilité à comptabiliser le nombre de femelles reproductrices par race. Ce nombre permet de définir différentes catégories de races. L'article 3 définit une race locale comme étant une race ayant un fort lien avec un territoire donné (30 % des femelles reproductrices se situent dans un seul département ou 70 % dans trois départements limitrophes deux à deux). Pour savoir si une race locale est menacée d'être perdue pour l'agriculture, la moyenne des notes (entre 0, pas de menace, et 5, menace maximale) de 6 indicateurs (le nombre total de femelles reproductrices ; l'évolution de ce nombre durant les cinq dernières générations ; la proportion de croisement ; la taille efficace de la population ; l'organisation des éleveurs et l'appui technique ; le contexte socio-économique) est calculée. La méthode de conversion utilisée pour chaque indicateur est décrite en détail dans l'article de (Verrier et al., 2015). Si cette note atteste d'une menace que

la race soit perdue pour l'agriculture, des mesures de conservation des ressources génétiques animales peuvent être entamées comme le spécifie l'article 5.

Les races reconnues par le ministère de l'agriculture n'appartenant pas à des races locales sont soit des races avec de forts effectifs répartis sur une grande partie du territoire et soumises à des schémas de sélection afin d'améliorer la qualité de leur production ; soit des races importées mais dont un schéma de sélection français a été mis en place. Cette sélection est effectuée au niveau national par Capgènes¹.

Un suivi démographique de toutes ces races est effectué régulièrement par l'institut de l'élevage (IDELE²) et les structures associatives pour les races locales caprines.

Parmi les 226 animaux français échantillonnés dans VarGoats, 222 sont des chèvres domestiques réparties dans 11 races (Tableau VI) dont 8 sont locales (Créole, des Fossés, Lorraine, Poitevine, Provençale, Pyrénées, Rove, Savoie), et 3 en sélection (Alpine, Saanen, et Angora).

Tableau VI : récapitulatif des races de chèvres présentes dans le projet VarGoats ainsi que leurs effectifs. Les effectifs nationaux proviennent des estimations effectuées par Capgènes (ALP, ANG, SAA), les inventaires de l'INRAE (CRE), de l'IDELE et des associations d'éleveurs pour les autres races.

Races	Codes	Code de la race	Effectifs VarGoats	Effectifs Nationaux
Alpine	ALP	13	56	660 000
Angora	ANG	71	20	5 000
Créole	CRE	52	16	3 147 (2019)
Des Fossés	FSS	44	16	1 840 (2018)
De Lorraine	LOR	46	13	1 480 (2018)
Poitevine	PTV	76	12	4 147 (2019)
Provençale	PVC	41	15	1 926 (2018)
Pyrénées	PYR	42	15	4 400 (2019)
Rove	ROV	43	11	9 398 (2017)
Saanen	SAA	11	33	396 000
Savoie	SAV	49	15	926 (2018)

¹ Capgènes est un organisme et une entreprise de sélection (OES) créée en 2008 qui se charge de la sélection des races, de la production et collection de semence caprine en France. (www.capgenes.com)

² L'institut de l'élevage est une association (loi de 1901), qui a pour objectif d'améliorer la compétitivité des élevages de ruminants français ainsi que les filières qui leur sont associés. Pour cela elle apporte un appui technique aux éleveurs par le biais de rapports d'activités et de recherches. (idele.fr)

Pour chacune de ces races nous allons détailler son histoire pouvant expliquer, en partie, son appartenance aux races en sélection, locales menacées ou non. Nous répertorierons aussi les caractéristiques morphologiques et d'élevage de chacune d'elles afin de commencer à percevoir la diversité des phénotypes présents en France.

Les races locales

Chèvre Créole



Figure 13 : Chèvres Créoles (© Capgènes)

Morphologie : La couleur de la robe n'est pas un caractère fixé dans la population qui compte plusieurs combinaisons de couleurs (noir, fauve et gris) avec une prédominance de la couleur noire. Nous n'avons pas de données chiffrées concernant la taille des animaux, mais les poids moyens sont de 38 kg pour les boucs et 28 kg pour les chèvres.

Histoire : Cette race, élevée en Guadeloupe, est issue d'importations de chèvres venant d'Afrique, d'Europe et d'Inde au XIX^{ème} siècle.

Élevage : Ces animaux sont élevés principalement pour la viande de cabris qui représente un intérêt pour la consommation locale (Alexandre et al., 2008). Dans les années 2000, la coopérative Cabricoop collabore avec la Chambre d'Agriculture de la Guadeloupe et l'INRAE pour créer un programme de sélection de la race. Cette sélection se base sur 300 mères réparties dans 10 élevages et a pour objectif l'amélioration des caractères de production (croissance et rendement de la carcasse), un maintien de la bonne reproduction de la race ainsi que de son adaptation au milieu tropical.

Chèvre des Fossés



Figure 14 : Chèvres des Fossés (© Aurélie Clémence)

Morphologie : La Chèvre des Fossés est de taille petite à moyenne (taille au garrot moyenne de 75 cm et 65 cm pour les boucs et les chèvres avec un poids moyen respectif de 55 kg et 35 kg). Les patrons de robe sont très variés (excepté pour la robe chamoisée qui pourrait être signe de croisement). Le poil doit être long sur les

cuisse ou sur tout le corps pour les chèvres, et long sur tout le corps pour les boucs. Les animaux mottes (sans cornes) sont acceptés mais généralement les chèvres ont les cornes fines, parallèles et arquées vers l'arrière. Les boucs ont des cornes très développées et torsadées.

Histoire : Cette race tient son nom du fait que les chèvres glanaient leur nourriture sur le bord des routes dans le Grand-Ouest (Bretagne, Normandie et Pays de la Loire). Cette race très rustique et docile fournissait du lait, de la viande et pouvait servir de bête de somme aux ruraux sans terre et aux déshérités.

Suite aux guerres mondiales ainsi qu'à la révolution agricole qui a suivi (industrialisation et urbanisation des modes de vie), ses effectifs chutent drastiquement pour ne plus compter qu'une centaine d'individus dont certains sont retournés à l'état sauvage. Vers le milieu des années 1990, l'Institut de l'Élevage, l'Écomusée du Pays de Rennes et la Chambre d'Agriculture de la Manche recherchent des animaux permettant de reconstituer un troupeau de chèvres des Fossés. Cela donna naissance à la création d'un comité pilote rassemblant de nombreuses organisations locales (Conservatoire du Littoral, Parc Naturel Régional d'Armorique etc...) et à des actions de particuliers passionnés visant le maintien de l'élevage des chèvres afin d'aider à la sauvegarde de la race. En 1999, un répertoire des animaux et de leurs généalogies est créé par l'Institut de l'Élevage et mis à jour avec l'aide des différents partenaires. En 2007, l'Association de Sauvegarde et de Promotion (ASP) de la Chèvre des Fossés est créée avec pour base les membres du groupe de pilotage créé précédemment (www.chevredesfosses.fr). Elle a pour objectif de faciliter la communication entre les éleveurs mais aussi d'aider l'Institut de l'Élevage à tenir à jour une base de données de tous les animaux. On y repère les caractères morphologiques ainsi que les animaux intéressants pour de la collecte de semence.

Élevage : La race est adaptée au climat océanique doux et humide où elle est élevée principalement dans un double objectif de conservation de la race et d'entretien des espaces naturels. Certains éleveurs valorisent aussi le lait de ces chèvres pour en faire du fromage.



Figure 15 : Logo de l'Association de Sauvegarde et de Promotion de la Chèvre des Fossés

Chèvre de Lorraine



Figure 16 : Chèvre de Lorraine
(© Capgènes)

Morphologie : La race se caractérise par des chèvres de taille moyenne (taille au garrot moyenne de 73 cm et 68 cm de pour les boucs et les chèvres avec un poids moyen respectif de 80 kg et 60 kg) avec une robe grise herminée (couleur brune mouchetée tolérée) et des poils mi-longs. La barbiche est obligatoire chez les boucs, recherchée chez les femelles et les animaux mottes ne sont pas acceptés.

Histoire & Élevage : La Lorraine a toujours été le carrefour de migrations humaines. Son changement de patrie au grès des guerres et ses cités minières attirant de nombreuses populations immigrées en a fait un territoire avec de nombreuses influences et une diversité d'origines retrouvée jusqu'au sein des « chèvres communes » de la région (introduction volontaire ou non d'animaux étrangers). Celles-ci ont servi de base à la formation de la race qui prend le nom de chèvre de Lorraine dans les années 90. Au départ cette race est élevée de deux manières différentes. Soit un élevage familial où les chèvres fournissent du lait, de la viande et aident parfois au travail. Soit un élevage conjoint avec des ovins ou bovins en transhumance où les chèvres peuvent être utilisées pour conduire le troupeau et se nourrir des plantes délaissées par les autres espèces. Avec la révolution agricole et la recherche de plus de productivité dans les années 70, les éleveurs délaissent cette race au profit de chèvres Alpines et Poitevines qui sont introduites dans la région. Des cas d'hybridations entre ces races sont connus et en particulier entre Lorraine et Alpine entraînant une « alpinisation » de la race. Dans les années 1970, un éleveur passionné s'attache à récupérer les derniers animaux. Le troupeau sera dispersé dans les années 1990 et un travail de conservation, appuyé par l'École Nationale Supérieure d'Agronomie & des Industries alimentaires de Nancy, démarre en 2006 alors que la race ne compte plus que 78 individus. Cela a donné naissance à l'Association des Amis de la Chèvre de Lorraine, créée en 2007 (chevredelorraine.fr), qui a pour but de sauvegarder la race, maintenir la diversité de ce qui a pu échapper à l'alpinisation et développer les qualités productives afin que les éleveurs puissent en vivre. L'association s'occupe également de tenir à jour le livre généalogique de la race, de faciliter la communication entre les éleveurs et de promouvoir la race dans des foires ou divers manifestations. L'Institut de l'Élevage intervient en appui pour aider à la gestion de la variabilité génétique de la race.



Figure 17 : Logo de l'Association des Amis de la Chèvre de Lorraine

Chèvre Poitevine



Figure 18 : Chèvres Poitevine
(© Association pour la Défense et le Développement de la Chèvre Poitevine)

Morphologie : les chèvres Poitevines se caractérisent par leur grande taille (taille au garrot moyenne de 80 cm et 70 cm de pour les boucs et les chèvres avec un poids moyen respectif de 65 kg et 55 kg), leurs poils mi-longs et leur robe « cape de maure » : majoritairement noir, avec le dessous et la face intérieure des membres blancs, ainsi qu'une bande blanche de chaque côté des naseaux.

Histoire : Le berceau de cette race est situé dans le Centre-Ouest de la France. On comptabilisait plus de 55 000 animaux au début du XX^{ème} siècle, avant qu'en 1920 une épidémie de fièvre aphteuse ne vienne décimer les populations. Afin de reconstituer les troupeaux, des chèvres provenant des Alpes sont introduites. Cela finit par entraîner une concurrence entre les chèvres Alpines et Saanen introduites qui sont plus productives, et les Poitevines dont les effectifs continuent de diminuer. Six cents chèvres sont répertoriées en 1980, ce qui entraîne la création de l'Association pour la Défense et le Développement de la Chèvre Poitevine en 1986 (www.chevre-poitevine.org). L'association entame la sauvegarde de la race en répertoriant les caprins poitevins, et reprend le travail de sélection génétique entamé dès 1947 pour améliorer les performances laitières et faisant la promotion de la race et de ses produits.



Figure 19 : Logo de l'Association pour la Défense et le Développement de la Chèvre Poitevine

Élevage : Cette race est élevée pour produire un lait dont la qualité fromagère est reconnue.

Chèvre Provençale



Figure 20 : Chèvres Provençale
(© Association pour la Sauvegarde de la Chèvre Provençale)

Morphologie : La race n'a pas été standardisée sur ses caractères morphologiques. La "pureté" d'une chèvre Provençale repose uniquement sur son histoire et ses origines. On retrouve une grande diversité de phénotypes dans cette population avec des caractères recherchés tels que la taille moyenne (taille moyenne au garrot inférieure à 80 cm pour les boucs et 70 cm pour les chèvres), l'absence de corne, la longueur du poil.

Certains autres sont proscrits comme la robe chamoisée des Alpines ou les cornes des Rove.

Histoire : Cette race est originaire de la « Provence des collines ». Comme pour les autres races locales que nous avons citées précédemment. L'érosion des effectifs de cette race débute dans les années 1970. Dans les années 1990, on ne dénombre plus que 200 animaux correspondant aux standards de la race. Pour contribuer à la sauvegarde de la race, l'Association pour la Sauvegarde de la Chèvre Provençale est créée en 1993 (lachevreprovencale.wordpress.com). Son objectif est de développer la race dans son berceau d'origine et de maintenir sa diversité génétique caractéristique. Le programme de conservation a démarré en 1999 avec la mise en place d'un répertoire des animaux par l'Institut de l'Élevage.



Figure 21 : Logo de l'Association pour la Sauvegarde de la Chèvre Provençale

Élevage : Cette chèvre est avant tout élevée pour son lait, en général transformé en fromages à la ferme. Elle fait partie des races dont le lait sert à la fabrication du fromage AOC Banon.

Chèvre du Rove



Figure 22 : Chèvre du Rove
(© Association de Défense
des caprins du Rove)

Morphologie : Cette race se caractérise par de grandes cornes torsadées, des poils courts avec une robe principalement rouge même si des mouchetures blanches ou des panachures peuvent exister. Les chèvres ont un corps massif et musclé avec un gigot plus développé que les autres races caprines françaises (taille moyenne au garrot de 95 cm et 75 cm pour les boucs et les chèvres avec un poids moyen respectif de 80 kg et 50 kg).

Histoire : Le berceau de cette race est la Basse Provence (dans l'arrière-pays méditerranéen). Elle accompagnait les troupeaux ovins et transhumait dans les Alpes avec eux. Ces chèvres avaient différents rôles : elles conduisaient le troupeau, pouvaient allaiter les chevreaux seuls et fournissaient du lait et de la viande (cabris) pour les bergers. Plusieurs rapports dans les années 1970 insistaient sur les mesures de sauvegarde à mettre en place pour éviter la disparition de la race après l'arrêt de la transhumance à pied, et c'est ainsi qu'en 1979, l'Association de Défense des caprins du Rove voit le jour (chevredurove.com). Un répertoire des éleveurs est mis à jour tous les trois ans avec l'appui de l'Institut de l'Élevage et de la Chambre d'Agriculture des Bouches du Rhône.



Figure 23 : Logo de
l'Association de Défense
des caprins du Rove

Élevage : Cette race est aussi bien élevée pour ses aptitudes laitières que bouchères. En effet, 59 % des élevages sont allaitants (production de cabris pour la consommation) et les 41% restant fromagers (d'après l'inventaire de la race fait en 2017, Figure S4 A. en Annexe). En revanche, si on s'intéresse aux effectifs des animaux par élevage on trouve 39% des femelles allaitantes contre 61% qui sont traites (d'après l'inventaire de la race fait en 2017, Figure S4 B. Annexe). Son lait permet la confection d'un fromage nommé la « Brousse du Rove » qui vient d'obtenir l'AOP. Elle est aussi utilisée pour défricher des terrains difficiles d'accès et participe alors à la valorisation et la sauvegarde des espaces méditerranéens.

Chèvre des Pyrénées



Figure 24 : Bouc de la race Pyrénées
(© Gilles Cazaban)

Morphologie : Il existe une grande variabilité phénotypique chez cette race mais quelques caractères sont spécifiques. Les animaux sont généralement grands (taille moyenne au garrot de 85 cm et 75 cm pour les boucs et les chèvres avec un poids moyen inférieur à 90 kg pour les boucs), une barbe et une frange frontale chez les deux sexes, des cornes assez développées, un poil raide mi-long à long et un patron de robe traditionnellement noir / brun foncé avec des taches plus claires localisées.

Histoire : La chèvre des Pyrénées est autochtone du massif dont elle porte le nom. Elle était présente dans les troupeaux d'ovins pour fournir du lait aux bergers en estive. Ce lait était aussi consommé en ville. Entre 1870 et 1930 certains chevriers amenaient leurs chèvres jusqu'à Paris et dans le Nord de la France pour y vendre directement leur lait aux consommateurs (Passal, 2013). Les guerres du début du XX^{ème} siècle, l'exode rural et la concurrence des races Alpine et Saanen ont fait que la race a quasiment disparu au début des années 90. À ce moment-là, des éleveurs passionnés commencent à répertorier les chèvres de race Pyrénées pour reconstituer des troupeaux. En 2004 l'Association de la Chèvre de race Pyrénéenne (www.chevredespirenees.org) est créée pour sauvegarder cette race et continuer les actions de développement entamées dès 1993 (inventaire des troupeaux, contrôle laitier etc.).



Figure 25 : Logo de l'Association de la Chèvre de race Pyrénéenne

Élevage : Cette race habituée aux territoires de montagnes participe à la sauvegarde et la valorisation des espaces naturels tout en produisant des produits pouvant être commercialisés (lait, fromage et viande de chevreaux).

Chèvre des Savoie

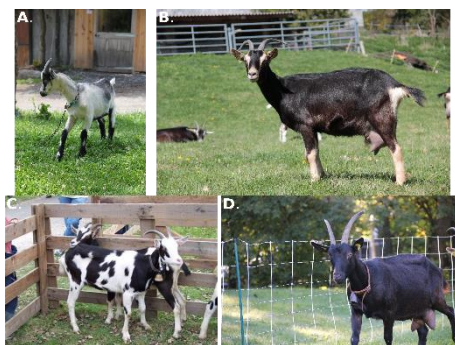


Figure 26 : Différentes robes des chèvres de Savoie : A. « cou clair », B. « à barrettes », C. pie, D. unicolore (© l'Association de Sauvegarde de la Chèvre des Savoie)

Morphologie : Cette population a des caractéristiques morphologiques marquées (poil ras, des oreilles fines et droites, cornes parallèles) mais comme beaucoup des autres races locales, la couleur est variable avec 4 types de coloration recherchés : « cou clair » (la partie antérieure du corps est plus claire que l'arrière qui reste noir, Figure 26 A.), le type « à barrettes » (robe noire brune ou grise avec des bandes blanches allant des cornes au nez, les oreilles et les pattes sont blanches elles aussi, Figure 26 B.), pie (panachures blanches,

Figure 26 C.), et enfin unicolore (noire, rousse, chocolat, Figure 26 D.). Ces chèvres sont de grande taille (taille moyenne au garrot de 85 cm et 75 cm de pour les boucs et les chèvres avec un poids moyen respectif inférieur à 90 kg et aux alentours de 60 kg).

Histoire : Cette population a été reconnue tardivement (31 juillet 2020) par le ministère de l'agriculture, de l'agroalimentaire et de la forêt en raison de son histoire mêlée à celle de l'Alpine. Autrefois les chèvres savoyardes formaient une grande population avec une multitude de patrons de couleurs allant du noir au blanc et incluant le type chamoisé. Ce type-ci s'est exporté dans d'autres régions à la suite de la 2nde guerre mondiale donnant la race Alpine. Dans les années 90, le journal de FERME (www.association-ferme.org) constate qu'il n'y a presque plus de chèvres « Alpine polychromes » dans les montagnes savoyardes. C'est ainsi qu'en 2001 l'Association de Sauvegarde de la Chèvre des Savoie (association-chevre-savoie.fr) est créée et que ces chèvres prennent le nom de chèvres des Savoie. Comme les autres associations de sauvegarde, celle-ci recense les animaux typés Savoie, aide l'Institut de l'Élevage dans la collecte des données généalogiques (répertoire des animaux), met en relation les éleveurs et promeut la population dans différentes manifestations.

Élevage : Les éleveurs sont tous producteurs de fromages (Chevrotin AOP, tomme, fromage à raclette, fromages lactiques) ou de yaourts.



Figure 27 : Logo de l'Association de Sauvegarde de la Chèvre des Savoie

Les races en sélection

Chèvre Alpine



Figure 28 : Chèvre Alpine (© Capgènes)

Morphologie : Animal de grande taille (taille moyenne au garrot de 95 cm et 75 cm pour les boucs et les chèvres avec un poids moyen respectif de 90 kg et 60 kg), très longiligne, au poil ras avec une couleur chamoisée. Ce type de robe est caractérisé par à une couleur brune dominante sur le corps et du noir sur la face, le dos jusqu'à la queue, le ventre et les pattes.

Histoire : Cette races est originaire des Alpes. Elle prend le nom d'Alpine chamoisée quand la race est officialisée en 1930 avec la création de son livre généalogique. Elle s'exporte sur tout le territoire français dans les années 1950 où elle remplace les races du Centre et de l'Ouest.

Élevage : C'est aujourd'hui la race la plus répandue en France. Depuis les années 1970, cette race est sélectionnée afin d'améliorer ses qualités laitières (qualité et quantité de lait) et depuis les années 2000 il y a également une sélection sur la morphologie mammaire et le comptage des cellules somatiques (Palhière et al., 2015). Ce schéma de sélection est coordonné par Capgènes (www.capgenes.com) en collaboration avec les professionnels de la filière caprine, l'IDELE, et l'INRAE (création des index de sélection et aide à la sélection génomique). Les Alpines représentent 55% des chèvres suivies en contrôle laitier. Son lait est utilisé pour la fabrication de fromages fermiers mais aussi pour la transformation industrielle.



Figure 29 : Logo de Capgènes

Chèvre Saanen



Figure 30 : Chèvre Saanen (© Irmgard)

Morphologie : Les chèvres ont un grand gabarit (taille au garrot moyenne de 95 cm et 75 cm de pour les boucs et les chèvres avec un poids moyen respectif de 100 kg et 65 kg), le poil court et uniformément blanc et sont souvent mottes.

Histoire : Le berceau de cette race est la vallée de la Saanen en Suisse. En 1950, l'élevage de cette race

s'étend sur le territoire français où elle va remplacer les races du Centre et de l'Ouest. Cependant son développement est moins rapide que celui des Alpines citées ci-dessus.

Élevage : Cette race est la plus répandue mondialement. Le même schéma de sélection que pour les chèvres Alpines a été mis en place afin d'améliorer sa production laitière (qualité et quantité de lait, morphologie mammaire et nombre de cellules somatiques) et sa morphologie (supprimer les taches rosées sur la robe). Les Saanen représentent 45% des chèvres suivies en contrôle laitier par Capgènes et sont particulièrement bien adaptées au système d'élevage intensif. Comme pour les Alpines, son lait est utilisé pour la fabrication de fromages fermiers mais aussi pour la transformation industrielle.

Chèvre Angora



Figure 31 : Chèvre Angora (© Association La Mohair des Fermes de France)

Morphologie : Ces chèvres sont de petite taille (taille au garrot moyenne de 65 cm de pour les chèvres avec un poids moyen de 50 kg et 35 kg pour les boucs et les chèvres) avec de longs poils blancs, les femelles peuvent être mottes alors que les mâles ont de longues cornes spiralées.

Histoire : Originaires d'Asie Centrale, cette race s'est exportée dans le monde entier au XV^{ème} siècle, avant d'être importée en France dans les années 1980 à partir de différents pays (Canada, États-Unis, Australie et Afrique du Sud). Dès 1982 l'Association Nationale des Éleveurs de Chèvres Angoras et Utilisateurs de Poil Mohair (ASECAUM) est créée pour développer la race en France, mettre en relation les éleveurs et défendre leurs intérêts. Cette association devient finalement l'Association Mohair des Fermes de France (www.mohair-france.com/laine-mohair/mohair-des-fermes-de-france) qui promeut la marque « le mohair des fermes de France » et continue de faire progresser la filière pour obtenir une cohérence collective d'élevage et de vente.

Un schéma de sélection français a été mis en place en 2011 par Capènes en collaboration avec l'INRAE et l'Institut de l'Élevage pour améliorer la qualité, la finesse (fibre inférieure à 30 microns) et l'homogénéité de la fibre mohair. Tous les ans, plus de 500 chèvres âgées de 18 mois sont soumises au contrôle de performance (filiation, poids de la toison, longueur et forme des mèches). Ce contrôle est complété par une analyse du



Figure 32 : Logo de l'Association Mohair des Fermes de France

rendement, une fois la fibre lavée, et de la finesse de la fibre. Ces résultats permettent aux éleveurs d'affiner la sélection au sein de leur élevage mais aussi pour l'ensemble de la filière dans la cadre du schéma collectif de sélection.

Élevage : Cette race est élevée sur l'ensemble du territoire français. La principale production des élevages est la fibre mohair qui peut être utilisée seule ou mélangée avec d'autres fibres pour confectionner différents textiles (vêtements, lainages, draperies, etc.).

Problématique

Avec cette présentation des différentes races caprines présentes en France, nous pouvons nous rendre compte que chacune d'elle a des spécificités liées à son histoire et à son élevage. Dans cette étude nous avons voulu voir si celles-ci pouvaient être retrouvées au niveau de leurs génomes et donc comprendre ce qui différencie les races caprines françaises entre elles. Pour cela nous avons analysé la diversité neutre, et cherché les régions sélectionnées chez les différentes races.

II. Matériels et Méthodes

Échantillonnage

Tout d'abord nous avons extrait du fichier de variants obtenu pour 1 160 animaux (présenté précédemment), les 222 chèvres domestiques françaises. Nous avons vérifié qu'aucun des animaux n'avait un taux de données génétiques manquantes supérieur à 5% (fonction « missing-indv » de VCFtools, version v0.1.15, Danecek et al., 2011). Afin de représenter au mieux la diversité de chaque race, les animaux apparentés, issus d'hybridations récentes ou étant sélectionnés sur des caractères divergents par rapport au standard de leur race ont été retirés du jeu de données.

Pour retirer les animaux apparentés nous avons utilisé le coefficient d'apparentement A_{jk} calculé entre dyades d'animaux de la même race par la fonction « relatedness » (Yang et al., 2010) de VCFtools. Nous avons ensuite supprimé au hasard un des deux animaux des dyades ayant un coefficient d'apparentement supérieur à 0,125 (relation entre les petits-enfants et

les grands-parents). Les animaux supprimés à cette étape sont principalement des Alpines (17 chèvres sur les 56 de départ), des Saanen (9 sur 33) et 1 Angora (sur les 20 de départ).

Une analyse de structure réalisée avec le programme ADMIXTURE (version v1.3.0, Alexander et al., 2009) a été réalisée afin de vérifier l'homogénéité des races. Pour réaliser cette analyse avons utilisés les 195 chèvres restantes et un nombre de populations ancestrales (K) allant de 1 à 14 (le détail de la méthode est expliqué dans la partie « analyse de structure »). Nous utilisons le résultat de l'analyse pour un nombre de populations ancestrales égal au nombre de races présentes dans notre jeu de données (K=11) pour supprimer les animaux qui avaient plus de 20% de leur génome d'origine exogène. Cela est le cas pour 7 animaux : 3 Savoie, 2 Provençales, 1 Fossés et 1 Angora (Figure S1 en Annexe).

Pour finir nous savons que 16 chèvres Alpines sont issues d'une lignée sélectionnée sur des caractères divergents (comme la concentration de cellules somatiques du lait, Caillat et al., 2011) dans l'unité expérimentale de Bourges. Nous supprimons donc les 10 animaux appartenant à cette population, restants dans notre jeu de données.

Notre échantillonnage final comprend donc 178 chèvres domestiques (29 Alpines, 18 Angoras, 16 Créoles, 15 Fossés, 13 Lorraines, 12 Poitevines, 13 Provençales, 15 Pyrénées, 11 Rove, 24 Saanen et 12 Savoie) dont 43 femelles et 135 mâles, représentatifs des standards de leur race. Celles-ci se répartissent sur une grande partie du territoire comme le montre la carte d'échantillonnage métropolitain (Figure 33).

Variants génétiques

Les SNPs correspondant à ces 178 animaux ont été extraits à partir du fichier de variants VarGoats. Les SNPs monomorphes ont été éliminés avec la fonction « filter » du programme SnpEff (version v4.3T, Cingolani et al., 2012). Nous avons ensuite utilisé la fonction « missing-site » du programme VCFtools pour supprimer les SNPs pour lesquels le pourcentage de données manquantes était supérieur à 5%. Enfin un seuil de fréquence allélique minimale de 1% a été ajouté. Nous obtenons finalement un ensemble de 25 333 920 SNPs sur lesquels différents filtres pourront être utilisés selon les analyses.

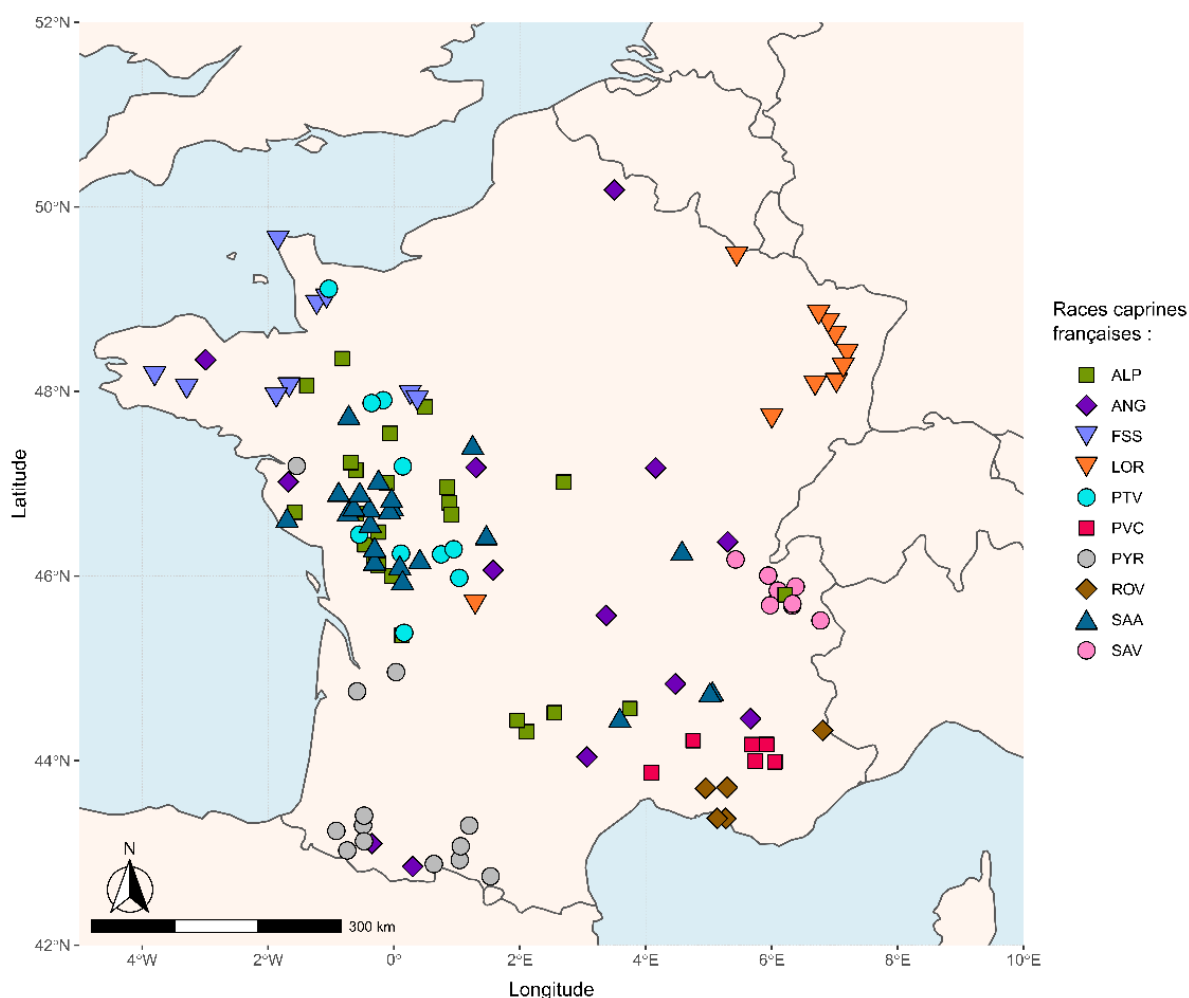


Figure 33 : Échantillonnage des races caprines françaises présentes dans le projet VarGoats. La race Créole n'est pas représentée ici car elle est endémique à la Guadeloupe. Chaque race est représentée par une couleur et un symbole différents.

Analyse de la structure

Nous réalisons une nouvelle analyse de structure avec le programme ADMIXTURE sur les 178 animaux gardés. Pour cette analyse et la précédente (permettant de filtrer les animaux), le nombre de variants a été réduit en supprimant tous les SNPs compris dans une fenêtre glissante de 50 SNPs, se déplaçant de 10 SNPs, qui avaient une corrélation entre le compte de chaque allèles supérieur à 0.1 (fonction « indep-pairwise » 50 10 0.1 du programme PLINK, version v1.90b5.3, Purcell et al., 2007).

Nous effectuons également une Analyse en Composantes Principales (ACP) avec PLINK (fonction « pca ») sur le même jeu de données. L'analyse est effectuée deux fois : une fois avec l'ensemble des chèvres de notre jeu de données puis seulement avec les races originaires de France métropolitaine (les Angoras et les Créoles sont retirées).

Relation entre les races

Nous utilisons deux méthodes différentes pour calculer la distance génétique entre races et inférer les croisements possibles entre elles. Le principe est similaire pour les deux : une première étape consistant à calculer les distances génétiques entre les races est effectuée. Un arbre de distance est construit à partir de celles-ci en fonction de différents modèles de reconstruction d'arbre. Si les résidus sont importants cela veut dire que l'arbre explique mal la relation entre les races concernées. Dans ce cas, l'ajout de croisement(s) entre les races peut permettre de réduire ces résidus et donc de trouver un modèle expliquant mieux les relations entre races. Ces analyses ont été effectuées à partir du jeu réduit de variants utilisé pour l'analyse de structure. Étant donné que les Angoras et les Créole ont des origines différentes par rapport aux autres races françaises, elles sont retirées du jeu données pour ces analyses afin que les différences entre races métropolitaines puissent être visibles.

Le premier programme utilisé est la deuxième version du programme hapFLK développé par Bertrand Servin (non publié). Comme pour la première version du programme (version v1.3.0, Fariello et al., 2013), la première étape consistant à calculer les distances génétiques entre les races est effectuée à partir de 100 000 SNPs tirés au hasard dans les données génomiques. La race Angora nous sert de groupe externe ici mais elle n'est pas prise en compte dans l'analyse. L'arbre de distance est construit suivant la technique du plus proche voisin (« neighbour joining »). Nous utilisons la nouvelle option « covkin » afin d'estimer les croisements entre populations. Cette analyse contraint la topographie de l'arbre et cherche tous les croisements possibles dans le jeu de données sans qu'on puisse lui spécifier un nombre à tester.

Le programme TreeMIX (version v1.13, Pickrell & Pritchard, 2012) a été utilisé pour tester des événements de migration (m) allant de 0 à 6. Ces tests ont été effectués 5 fois afin de pouvoir réaliser l'étape de validation croisée (« cross-validation », CV) avec la fonction Evanno du « package » OptM (Fitak, submitted) dans R (version 3.5.3, R Core Team, 2019). Cela nous a permis de déterminer quel était le nombre de migration expliquant le mieux nos données. Avec ce programme les arbres de distance sont construits avec la méthode du maximum de vraisemblance et la topographie n'est pas contrainte donc nous avons choisis l'arbre ayant le meilleur score de vraisemblance pour représenter le résultat de l'analyse.

L'analyse des pedigrees (Danchin-Burge et al., 2017) dans 8 races (il manque les Créoles, les Rove et les Savoie), permet à l'IDELE de fournir des indicateurs de la diversité génétique de ces races (IDELE, 2019). Cette analyse prend en compte environ 10 générations, ce qui apporte

des informations récentes sur les races comme par exemple des croisements entre races que nous pourrions comparer à nos résultats.

Histoire démographique des races

L'inférence démographique a été effectuée avec le programme PopSizeABC (Boitard et al., 2016) qui se base sur une méthode bayésienne (« Approximate Bayesian Computation », ABC) pour inférer la variation de la taille efficace des races au cours du temps. Pour réaliser cette analyse nous sommes repartis de l'ensemble des variants présents chez les chèvres françaises et nous les avons phasés programme Shapelt2 (version v2-r904, Delaneau et al., 2013). Le phasage consiste à inférer les haplotypes. La qualité de celui-ci peut être améliorée avec l'utilisation d'une carte de recombinaison qui estime la liaison génétique entre les SNPs. Concernant les caprins, cette carte a été réalisée récemment sur le dernier génome de référence (ARS1) par Rachel Rupp et Bertrand Servin (non publiée) donc nous avons donc pu inclure cette information dans l'analyse. Les fichiers de résultats de Shapeit2 ont été convertis en format vcf pour l'analyse PopSizeABC sur les 11 races caprines. Les statistiques sommaires utilisées pour chaque race correspondent au spectre de fréquence (suppression des SNPs avec une MAF inférieure à 5% pour ce calcul) et au déséquilibre de liaison (suppression des SNPs avec une MAF inférieure à 20% pour ce calcul) pour chacune des 21 fenêtres temporelles explorées par le programme. Nous avons créé 100 000 simulations de changement de taille efficace au cours du temps pour chaque jeu d'haplotypes (nombre d'haplotypes présents dans chaque race) en utilisant la fonction « simul_data » de PopSizeABC. Nous avons ensuite calculé pour chaque simulation une distance entre les statistiques sommaires simulées et celles observées qui nous a permis de sélectionner les histoires les plus probables et d'estimer les changements de taille de population à partir de la distribution postérieure commune des simulations gardées.

Dans les rapports des indicateurs de variabilité génétique (IDELE, 2019) nous trouvons une estimation de la taille efficace actuelle des races calculée avec 2 méthodes différentes : Wright-Fisher (Wright, 1931) ou Cervantès (Cervantes et al., 2011). Concernant les races Créole, Rove et Savoie nous avons calculé leurs tailles efficaces à partir des effectifs connus (inventaires des races faits par l'IDELE et les associations d'éleveurs pour les race Rove -Figure S4 A. en Annexe- et Savoie -Figure S5 en Annexe-, et communication personnelle de Nathalie Mandonnet -INRA Guadeloupe- pour les Créoles). Pour ce faire nous utilisons également la méthode de Wright-Fisher pour calculer la taille efficace à partir du nombre de mâles et de

femelles dans une population : $N_e = \frac{4N_m N_f}{N_m + N_f}$. Nous avons réalisé un test du χ^2 afin de comparer les estimations des tailles efficaces actuelles fournies par PopSizeABC avec celles présentées dans les rapports ou que nous avons calculées à partir des effectifs des populations.

Nous nous intéressons ensuite à calculer différents indices de consanguinité. Cela nous permettra de savoir quel est l'impact des variations démographiques et du système d'élevage (choix des croisements par les éleveurs) sur la diversité génétique des races.

Pour cela nous calculons le coefficient de consanguinité (F_{IS}) de chaque race avec la fonction « het » de PLINK. Le nombre et la longueur des fragments homozygotes (R_{OH}) présents chez nos races caprines sont cherchés en utilisant les paramètres par défaut de la fonction « homozyg » de PLINK sur le fichier réduit de variants.

Pour ces deux analyses, nous avons réalisé un test de Kruskal-Wallis afin de vérifier si les coefficients de consanguinité ou la proportion de R_{OH} dans le génome étaient différents entre races. En cas de test significatif, des tests unilatéraux de Wilcoxon par paire, corrigés pour les tests multiples par la méthode du taux de fausses découvertes (« false discovery rate », FDR), ont été effectués afin de chercher les races qui avaient des valeurs significativement plus petites ou grandes que les autres.

Afin de vérifier si les valeurs que nous trouvons sont proches des données issues des pedigrees, nous récupérons les consanguinités moyennes (probabilité d'identité des gènes en pourcentage) et leur évolution sur 10 ans dans les indicateurs de variabilité génétique de l'IDELE (2019).

Signature de sélection

La seconde partie de notre analyse consiste à chercher les signatures de sélection. Pour cela nous avons utilisé 2 approches :

1. Une première, globale, où nous avons comparé toutes les races entre elles pour trouver les signaux présents dans l'ensemble de notre jeu de données ;
2. Une seconde, ciblée, où nous avons comparé une race par rapport à toutes les autres afin de trouver les signatures spécifiques à celle-ci.

Pour ces analyses les chèvres Créoles et Angoras ont été retirées car elles ne sont pas originaires de France donc cette différence ne permettrait pas de différencier les signatures de sélection au sein des races françaises. Le fichier de variants utilisé est le fichier phasé obtenu après exécution du programme Shapelt (voir la partie « Histoire des races »).

Analyse globale

Pour effectuer cette recherche des signatures de sélection, nous avons utilisé la version 2 du programme hapFLK. La première étape de cette analyse correspond au calcul des distances génétiques et à l'ajout de potentiel(s) croisement(s) entre les races, qui sont décrits dans la partie « Relations entre les races ». Ce fichier de croisement a été utilisé dans le reste de l'analyse afin de donner l'information d'apparentement entre nos races (paramètre « kinship ») au programme.

Pour pouvoir exécuter l'analyse hapFLK nous avons besoin de déterminer un nombre de groupes d'haplotypes dans notre jeu de données (paramètre « K » du modèle). Pour cela nous utilisons le programme fastPHASE (version v1.4.0, Scheet & Stephens, 2006) sur l'ensemble d'un chromosome. Nous avons arbitrairement choisi le chromosome 20 car il est assez petit pour que l'analyse ne soit pas trop longue sans être le plus petit de l'espèce. Nous testons des groupes d'haplotypes entre 5 et 30 par pas de 5 (préconisé par les auteurs pour les gros jeux de données), et la procédure de validation croisée nous permet déterminer que le K optimal est 15 (Figure S2 en Annexe).

L'analyse est effectuée sur des fenêtres de 2 méga-bases (Mb) avec un chevauchement de 1 kilo-base (Kb). Dans chacune de ces fenêtres le programme va prendre en compte les différences de fréquence des haplotypes entre les populations pour détecter les régions sous sélection.

Nous regroupons les SNPs significatifs (q-valeurs inférieures ou égales à 0,05) s'ils sont distants de moins de 100 Kb afin de définir des régions candidates. Nous exécutons une seconde fois le test hapFLK uniquement sur ces régions en ajoutant les options « annot » et « eigen » à l'analyse afin de pouvoir représenter les haplotypes présents dans chaque race. Cette représentation, associée aux pics significatifs (regroupement des SNPs significatifs distants de moins de 5 Kb) nous a permis de délimiter manuellement les haplotypes sélectionnés dans chaque région et de pouvoir réaliser une analyse locale sur ceux-ci. Cette analyse consiste à calculer les distances génétiques entre les races uniquement sur la partie du génome

sélectionnée et à réaliser à partir de cette distance un arbre local nous permettant de déterminer quelle(s) race(s) porte(nt) l'haplotype sélectionné. Nous avons gardé les races dont la différence de distances génétiques (longueur des branches) entre l'arbre local et l'arbre sur l'ensemble du génome était significative (p -valeur $< 0,05$). Pour finir nous cherchons si ces haplotypes sont distants de moins de 10 Kb de gènes annotés sur le génome ARS1.

Afin de savoir quels effets pourraient avoir les variants détectés sous sélection par hapFLK, nous avons utilisé l'interface web de l'outil « Variant Effect Predictor » d'Ensembl (www.ensembl.org/Tools/VEP) sur l'annotation du génome ARS1 de la chèvre. Les paramètres par défaut de l'application ont été utilisés pour réaliser l'analyse sur les variants sous sélection mais également l'ensemble des variants de notre jeu de données. Les résultats de prédiction des effets ont été regroupés en 6 catégories (inter-génique, en amont du gène, en aval du gène, dans un intron, dans un exon ou autre) en fonction de leur position par rapport au gène. Les UTR (« Untranslated Transcribed Region ») ont été classés dans les positions en amont et en aval des gènes, les régions d'épissage avec les introns, et toutes les pertes ou gain de codons avec les exons. Une différence de distribution des annotations entre le jeu de données global et les variants sous sélection a été testé avec un test du χ^2 .

Analyse ciblée

La fonction « hapflkadapt » du programme hapFLK permet d'ajouter des covariables à l'analyse initiale. Cela peut permettre de tester l'association d'haplotypes sélectionnés avec des phénotypes, mais aussi l'appartenance d'un haplotype à une race spécifique si l'on contraste une race par rapport aux autres. C'est ce que nous avons fait afin de détecter les signatures spécifiques à chacun des races étudiées.

Pour cela nous avons créé 9 fichiers de covariables (1 par race) dans lesquels la race d'intérêt a une valeur qualitative correspondant au code de sa race (voir Tableau VI) et les autres animaux ont un 0. Ce fichier est utilisé dans l'analyse hapFLK_{adapt} qui est effectuée sur les mêmes fenêtres glissantes de 2 Mb avec un chevauchement de 1 Kb que celles utilisées pour le test hapFLK. L'analyse est répétée 9 fois avec à chaque fois un fichier de covariables permettant d'analyser une race par rapport aux 8 autres.

Comme pour l'analyse hapFLK, nous avons regroupé les SNPs significatifs (q-valeurs inférieures ou égales à 1.10^{-4}) s'ils étaient distants de moins de 100 Kb pour définir des régions candidates. Si leur nombre excédait 20 alors nous n'avons gardé que les 20 plus significatives (basé sur la q-valeur minimale de chaque région) pour la suite de l'analyse. Nous avons extrait les pics significatifs (groupement des SNPs significatifs distants de moins de 5 Kb) dans ces régions et cherché s'ils étaient distants de moins de 10 Kb de gènes provenant du même fichier d'annotation que celui utilisé précédemment.

III. Résultats

Analyse de la structure

L'analyse avec le programme STRUCTURE nous donne un nombre optimal de populations ancestrales de 2. Dans ce cas, un groupe est composé des races Angora et Créole, et l'autre des races originaires de France métropolitaine. En augmentant le nombre de populations ancestrales, les races se séparent au fur et à mesure. Quand ce nombre correspond au nombre de races présentes dans notre jeu de données ($K=11$, Figure 34 B.) on peut noter que le programme différencie 2 origines pour la race Angora et que les races Pyrénées et Rove forment un seul groupe (ce qui n'est pas le cas dans l'analyse précédente où la race Pyrénées s'agrégeait avec les Fossés, Figure 34 A.). En augmentant le nombre de populations ancestrales à 12 (Figure 34 C.) alors nous retrouvons toutes les races de notre jeu de données bien séparées et des chèvres Savoie et Alpines ayant des portions de génome de même origine. Au-delà de 12 populations ancestrales, les populations ajoutées différencient des animaux au sein des 11 races étudiées, sans que des animaux des races différentes ne se regroupent entièrement au sein d'une même population ancestrale.

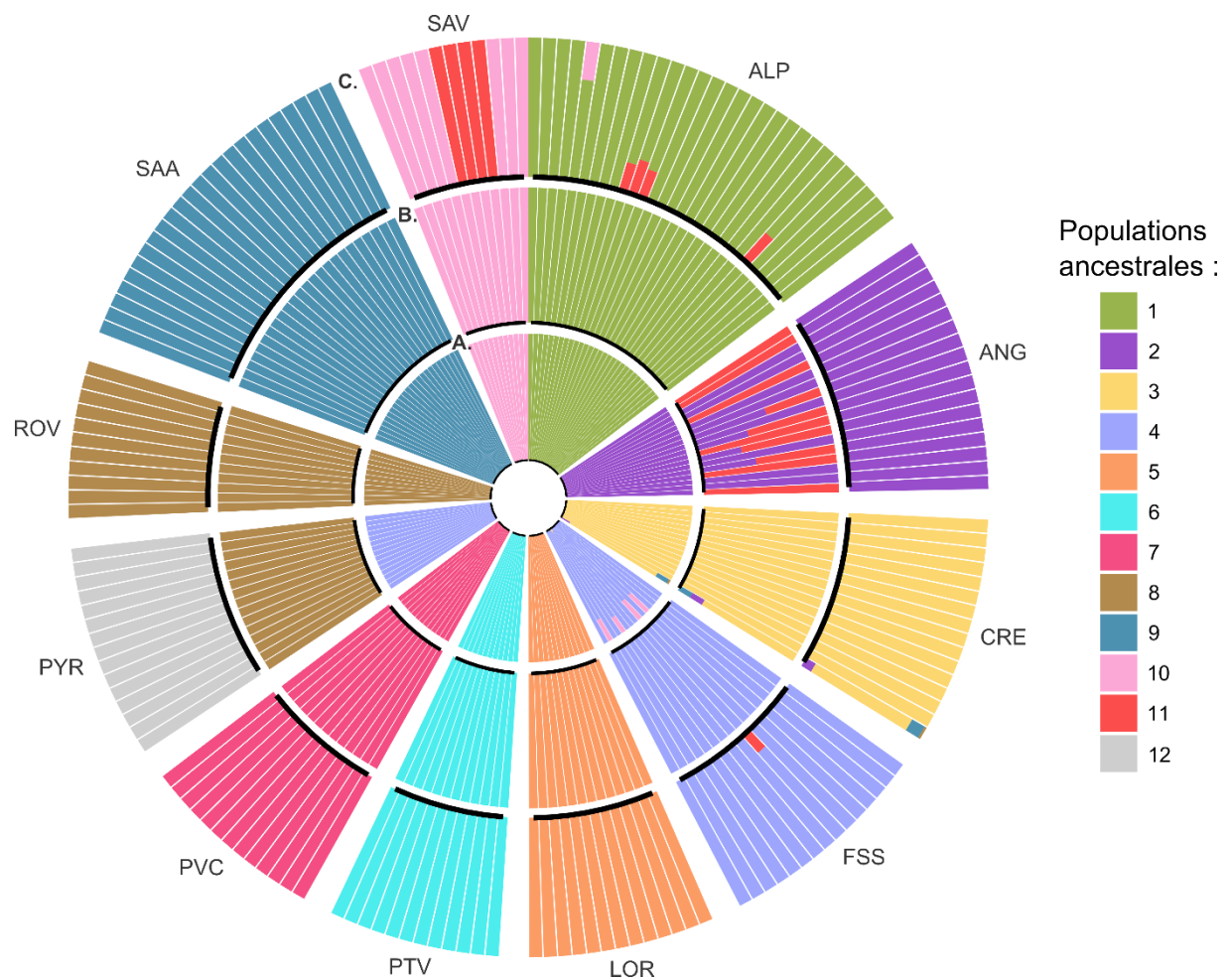


Figure 34 : Analyse de structure avec les 178 chèvres françaises, chacune représentée par une barre. Chaque cercle correspond à une analyse avec un nouveau nombre de population ancestrale (K) : A. $K = 10$, B. $K = 11$ et C. $K = 12$. Chaque couleur correspond à une population ancestrale.

Le résultat de l'analyse ACP sur l'ensemble des races caprines montre la formation de 3 groupes distincts : un correspondant aux Angoras, un autre aux Créoles et enfin les races d'origine métropolitaines (Figure 35 A.). Si nous ne conservons que les races métropolitaines nous voyons qu'elles se différencient et que chacune forme un groupe homogène (Figure 35 B.).

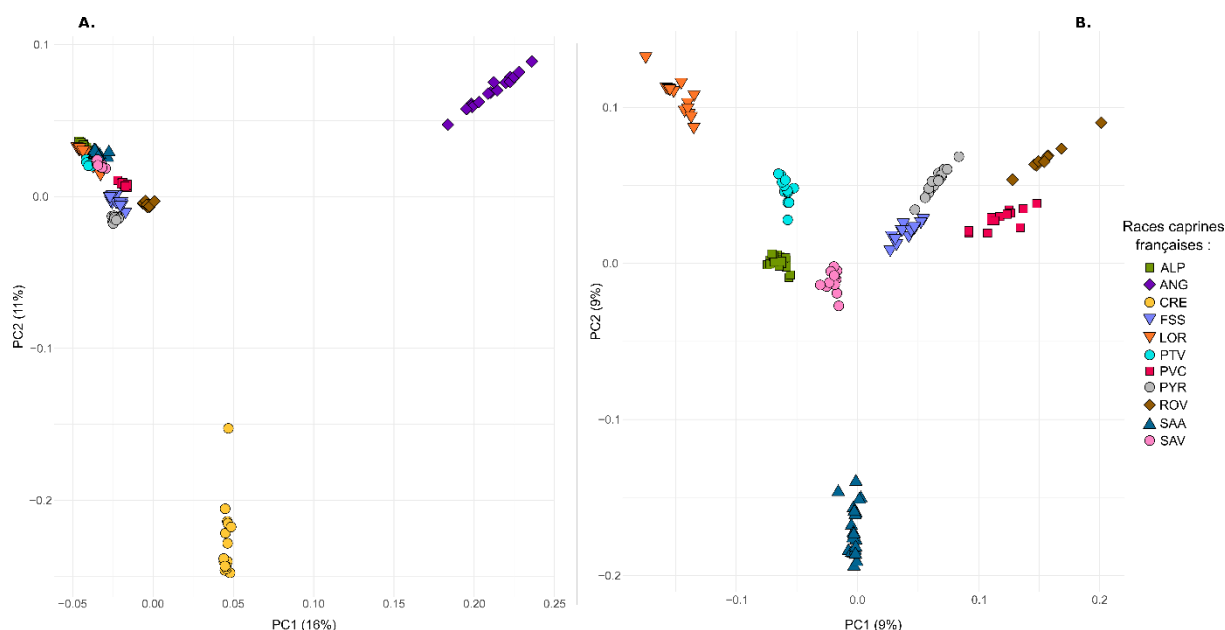


Figure 35 : Axes 1 et 2 des ACP avec A. toutes les races caprines de notre jeu de données (178 chèvres) ou B. les races caprines originaires de France métropolitaine (144 chèvres). Chaque race est représentée par une couleur et un symbole différents.

Relation entre les races

Les analyses des distances génétiques effectuées avec les programmes hapFLK ou TreeMix ont permis de créer les arbres de distances génétiques respectifs aux 2 méthodes. Les topologies de ceux-ci sont similaires (Figure 36 A. et C.). Nous pouvons remarquer qu'avec les deux méthodes les races caprines métropolitaines se regroupent par origine géographique (cas des races Provençale et Rove originaires de Provence ; et des Alpines, Savoie et Saanen originaires des Alpes), mais aussi en fonction des croisements récents (nous retrouvons les Lorraines et les Poitevines proches des Alpines dont nous savons qu'elles ont été utilisées pour renforcer les effectifs de ces 2 races). Nous observons avec le résultat hapFLK (Figure 36 A.) que la race Lorraine est celle qui possède la plus grande distance génétique avec les autres races.

Les matrices des résidus associés aux arbres (Figure 36 B. et D.) indiquent que certaines relations entre races ne sont pas bien expliquées par les modèles. Avec les deux méthodes nous observons un résidu important entre les races Alpine et Lorraine ; puis d'autres moins forts entre la race Saanen avec la Savoie ou la Provençale, et aussi entre la race Pyrénées et la Poitevine avec l'analyse TreeMIX.

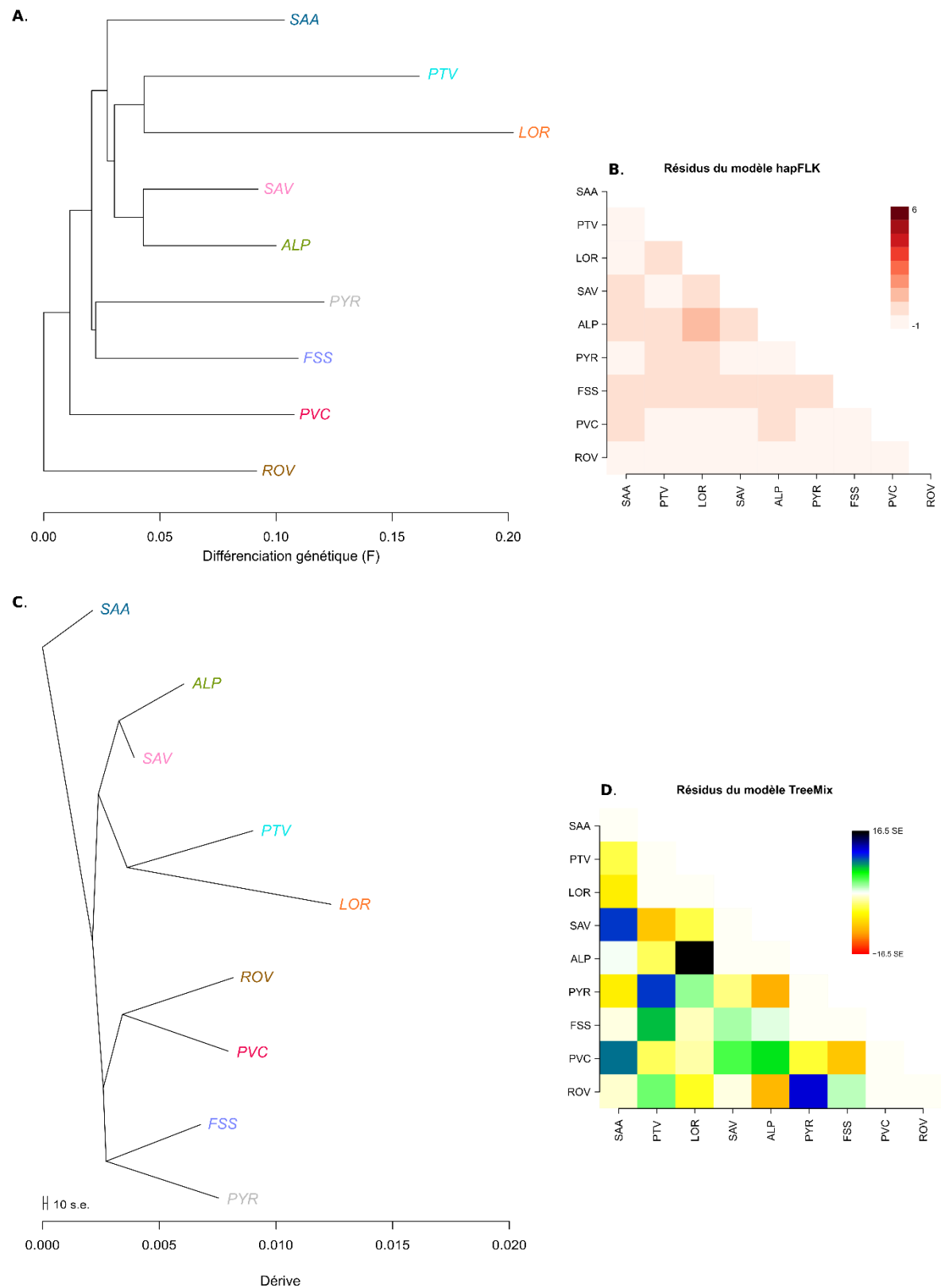


Figure 36 : Arbres phylogénétiques des races métropolitaines et leurs résidus. A. Arbre réalisé avec le programme hapFLK avec B. les résidus observés associés. C. Arbre réalisé avec TreeMix pour 0 migration et D. les résidus observés associés.

Sur la base des valeurs des résidus de l'arbre de distance génétique sans croisement, le programme hapFLK introduit un croisement entre les chèvres Alpines et Lorraines (Figure 37 A.). Celui-ci permet d'améliorer l'explication de la distance génétique entre ces races car le résidu du modèle est plus faible qu'à l'origine (Figure 37 B.).

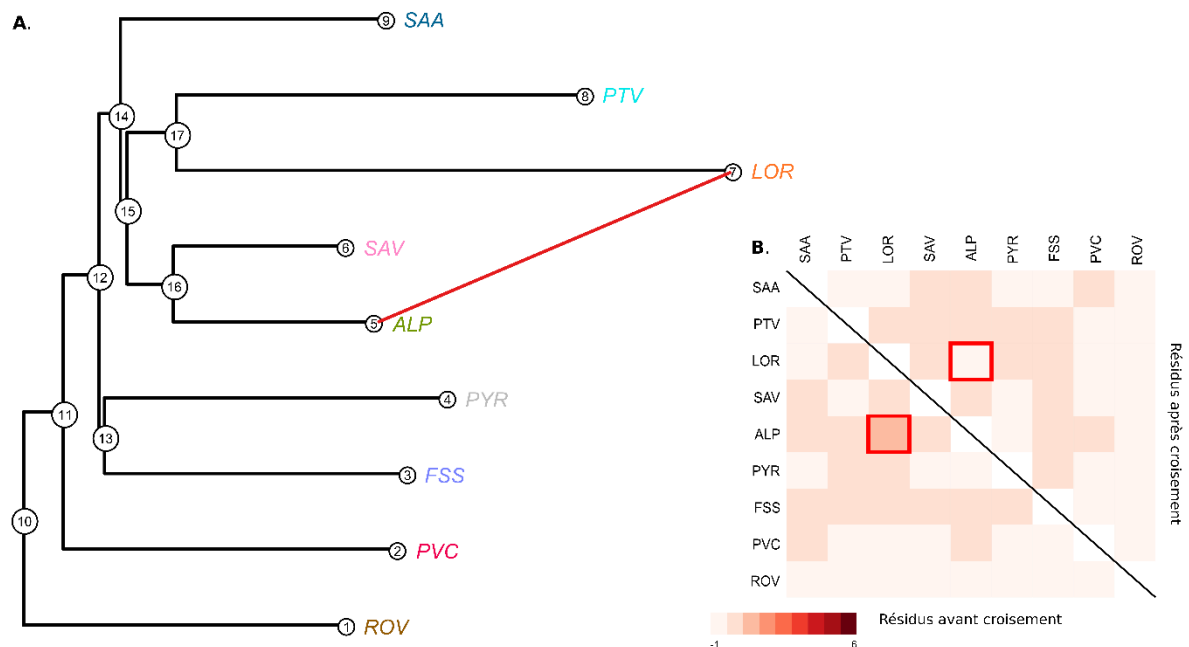


Figure 37 : Résultats de l'analyse des croisements entre les races caprines métropolitaines effectuée avec hapFLK. A. Arbre de croisement entre les races B. Matrice des résidus avant (matrice triangulaire inférieure) et après (matrice triangulaire supérieure) le croisement. La case rouge correspond au résidu impliqué dans le croisement.

Concernant l'analyse TreeMix, l'étape de validation croisée (Figure 38 A.) indique que nous obtenons le meilleur résultat avec 3 migrations. Concernant les résidus obtenus pour ce modèle (Figure 38 B.), nous pouvons voir que ceux en relation avec les 3 migrations sont devenus nuls et que de manière générale l'ensemble des résidus est plus faible sur le modèle avec les migrations que sans.

Nous observons un croisement d'un poids (fraction d'allèles dans la population descendante qui provient de chaque population parentale si le mélange s'effectue en une seule génération) de 0.33 qui relie les Provençales aux Rove ; un autre de 0.14 entre les Poitevines et les races Fossés et Pyrénées ; enfin un dernier croisement qui relie les Alpines et les Lorraines avec un poids de 0.13 (Figure 38 C.).

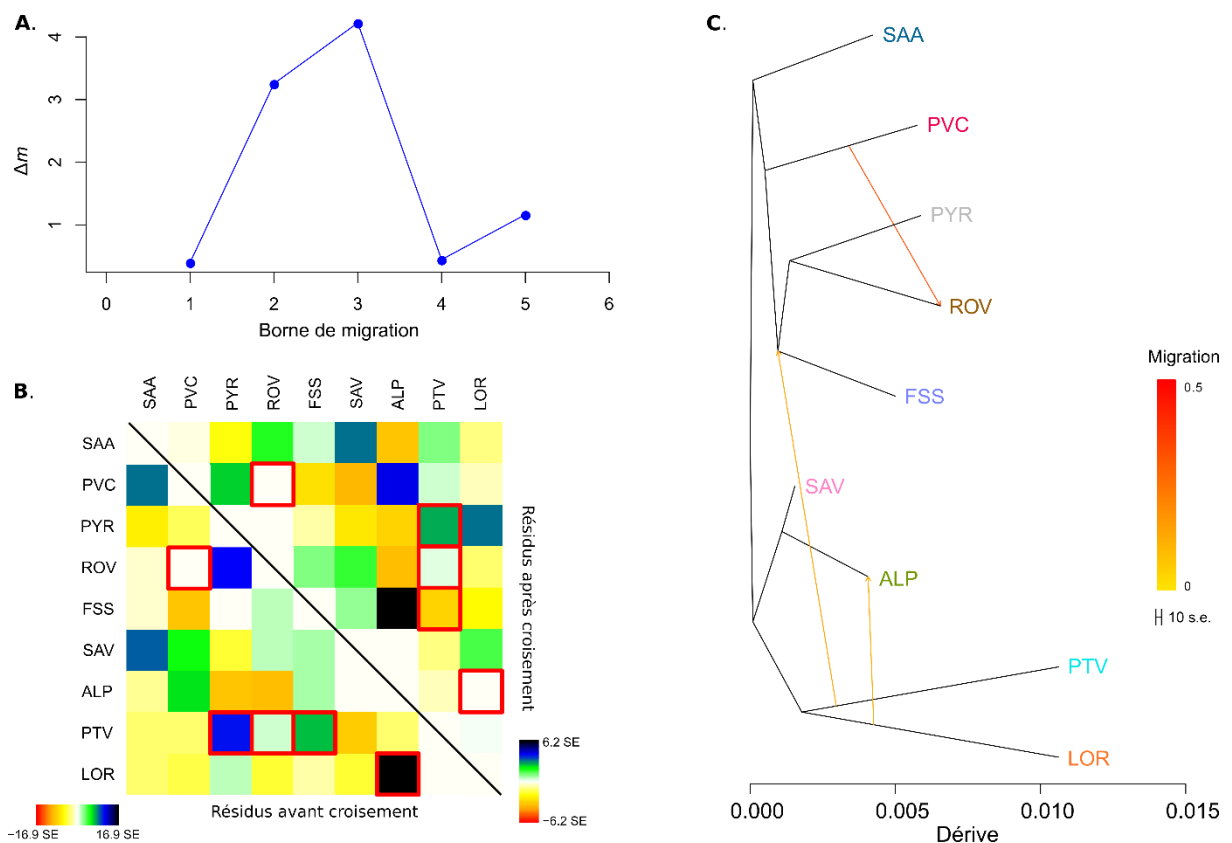


Figure 38 : Résultats de l'analyse des croisements entre les races caprines métropolitaines effectuée avec TreeMix. A. Étape de validation croisée (« cross-validation ») permettant de déterminer le nombre optimal de migrations (m) dans notre jeu de données. B. Matrice des résidus avant (matrice triangulaire inférieure) et après (matrice triangulaire supérieure) les 3 croisements. Les cases rouges correspondent aux résidus impliqués dans les croisements. C. Arbre de croisement des races obtenu pour $m = 3$, les croisements sont indiqués par les flèches entre les branches de l'arbre. Plus la couleur de cette flèche est proche du rouge plus le croisement est fort.

Histoire des races

Inférence démographique

Nous pouvons découper les variations démographiques des races caprines françaises (Figure 39) en 4 périodes de temps.

Une première période avant la domestication, se composant des temps les plus anciens jusqu'à la domestication. Durant cette période nous observons que les courbes des différentes races augmentent de la même façon, traduisant une histoire commune entre elles. Ceci est cohérent car la divergence entre les chèvres domestiques (*Capra hircus*) et sauvages (*Capra aegagrus*) n'avait pas encore eu lieu.

La seconde période correspond à la domestication qui a eu lieu aux alentours de 10 000 ans dans le croissant fertile (Zeder & Hesse, 2000). Nous trouvons ensuite une période post-domestication où les effectifs restent stables jusqu'à environ 500 générations.

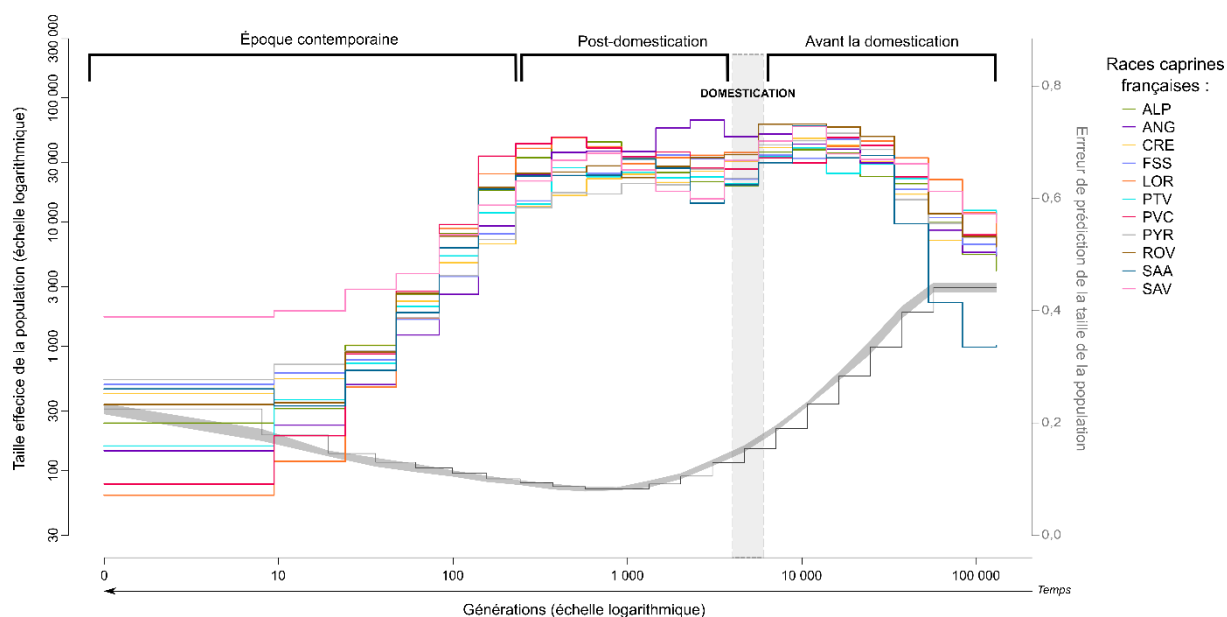


Figure 39 : Démographie des races caprines françaises. Chaque couleur représente une race et la courbe grise représente l'erreur d'estimation de la taille efficace des populations. La période de domestication est indiquée par la région en pointillée aux alentours de 10 000 ans (si nous prenons un intervalle de génération de 2 ans).

La dernière période correspondant à l'époque contemporaine (les 1 000 dernières années environ) durant laquelle tous les effectifs chutent, de manière plus ou moins rapide suivant les races, jusqu'aux tailles efficaces actuelles.

L'erreur de prédiction de la taille efficace du modèle a tendance à augmenter dans les temps récents (les cents dernières générations environ) donc nous avons voulu comparer les tailles efficaces que nous avons obtenues grâce aux simulations de notre jeu de données avec d'autres obtenues à partir de données réelles sur les races (rapports et inventaires de l'IDELE, Tableau VII). De plus ces rapports sont plus exhaustifs de la diversité des races et prennent en compte l'histoire récente des races (10 dernières années) qui peuvent nous donner des indications sur les pratiques d'élevage plus anciennes.

Tableau VII : Comparaison des estimations des tailles efficaces actuelles des races caprines françaises entre le programme PopSizeABC, les estimations fournies dans les rapports d'indicateurs de variabilité génétique (IDELE, 2019) ou les calculs à partir des sex-ratio dans les races Créoles, Rove et Savoie. Rappel des effectifs nationaux.

Races	Taille efficace			Effectifs nationaux
	PopSizeABC	Méthode de Wright	Méthode de Cervantès	
ALP	233	21 552	189	660 000
ANG	104	543	75	5 000
CRE	383	1 505	-	3 147 (2019)
FSS	550	572	94	1 840 (2018)
LOR	71	162	32	1 480 (2018)
PTV	173	857	157	4 147 (2019)
PVC	69	336	62	1 926 (2018)
PYR	640	617	249	4 400 (2019)
ROV	354	1 873	-	9 398 (2017)
SAA	430	13 645	146	396 000
SAV	1 816	207	-	926 (2018)

Nous pouvons voir que les estimations du programme PopSizeABC se situent presque toujours entre les valeurs calculées par la méthode Wright ou de Cervantès. Aucun des tests du χ^2 comparant les effectifs obtenus entre les différentes méthodes ne permet de trouver de différence entre elles (p-valeurs = 0,23 à chaque fois).

La race Provençale est celle qui a la plus faible taille efficace ($N_e = 69$) dans notre jeu de données. Nous trouvons ensuite la Lorraine ($N_e = 71$), l'Angora ($N_e = 104$), la Poitevine ($N_e = 173$), l'Alpine ($N_e = 233$), la Rove ($N_e = 354$), la Créole ($N_e = 383$), la Saanen ($N_e = 430$), la Fossés ($N_e = 550$), la Pyrénées ($N_e = 640$) et pour finir la Savoie ($N_e = 1 816$).

Consanguinité

Les coefficients de consanguinité (F_{IS}) de chaque race sont présentés dans la Figure 40.

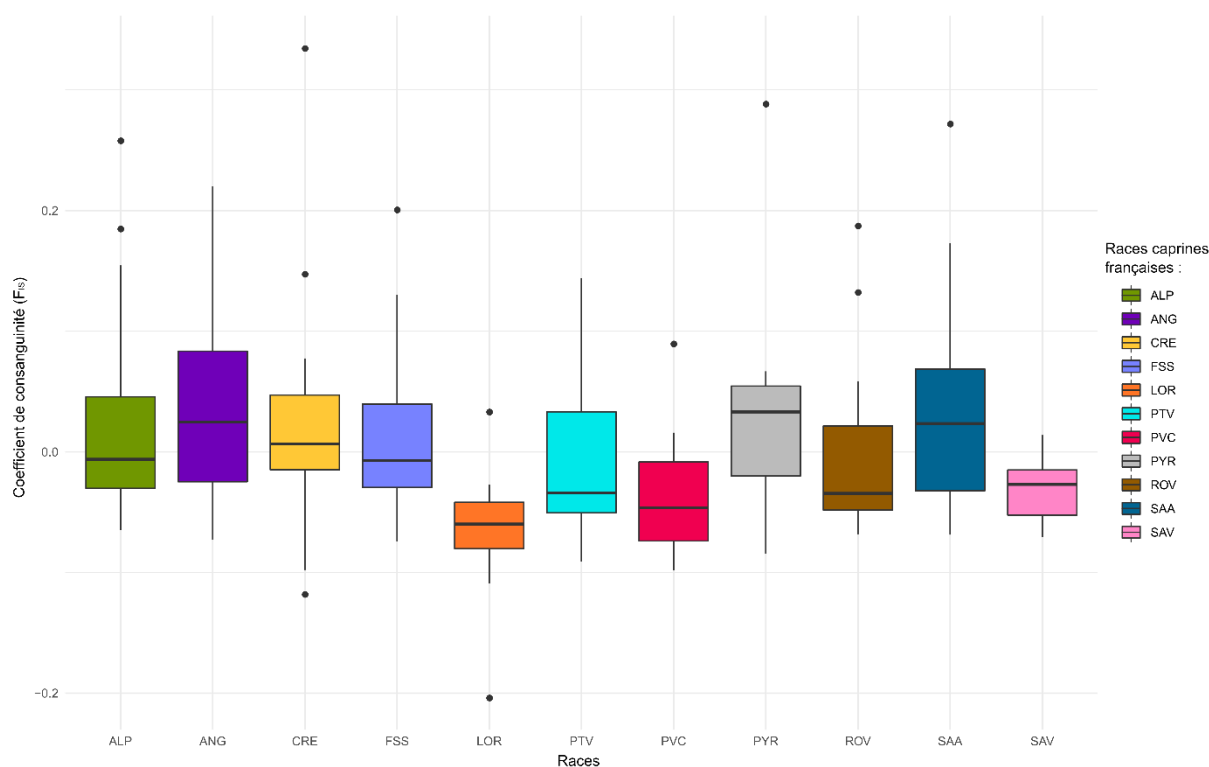


Figure 40 : Coefficients de consanguinité pour chaque race. Les boîtes à moustaches représentent la distribution des valeurs de F_{IS} pour chaque race (la partie colorée est délimitée par la valeur de F_{IS} correspondant à 25% et 75% du jeu de données. La ligne noire au milieu correspond à la valeur médiane).

Tableau VIII : p-valeurs du test de Wilcoxon par paire pour trouver les coefficients de consanguinité plus élevés (matrice triangulaire supérieure, dégradé d'orange) ou plus faibles (matrice triangulaire inférieure, dégradé de bleu) que les autres. Les cases claires sont significatives (p -valeurs $\leq 0,05$), et les intermédiaires hautement significatives (p -valeurs $\leq 0,01$) et les foncées extrêmement significatives (p -valeurs $\leq 0,001$).

	ALP	ANG	CRE	FSS	LOR	PTV	PVC	PYR	ROV	SAA	SAV
ALP		0,758	0,758	1,000	1,000	1,000	1,000	0,690	1,000	0,844	1,000
ANG	1,000		1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
CRE	1,000	0,755		1,000	1,000	1,000	1,000	0,947	1,000	1,000	1,000
FSS	0,822	0,541	0,561		1,000	1,000	1,000	0,758	1,000	0,812	1,000
LOR	0,001	0,003	0,007	0,010		0,161	0,657	0,003	0,161	0,003	0,206
PTV	0,243	0,190	0,243	0,409	1,000		1,000	0,229	1,000	0,507	1,000
PVC	0,040	0,039	0,065	0,087	1,000	0,338		0,579	0,657	0,044	0,947
PYR	1,000	0,822	1,000	1,000	1,000	1,000	1,000		1,000	1,000	1,000
ROV	0,243	0,243	0,243	0,409	1,000	0,828	1,000	0,243		0,579	1,000
SAA	1,000	0,744	0,879	1,000	1,000	1,000	1,000	0,702	1,000		1,000
SAV	0,065	0,065	0,039	0,147	1,000	0,561	1,000	0,029	0,702	0,039	

Le test de Kruskal-Wallis que nous avons effectué, afin de comparer les coefficients obtenus pour chaque race, est significatif ($p\text{-valeur} = 1,975.10^{-4}$) donc il y a au moins une race différente des autres. D'après les résultats du test de Wilcoxon par paire (Tableau VIII), les races Lorraine, Provençale et Savoie ont des coefficients de consanguinité plus faibles par rapport à certaines autres races. La race Lorraine a des coefficients plus faibles que 6 autres races : les Alpines, Angoras, Créoles, Fossés, Pyrénées et Saanen. Pour les Provençales il s'agit des races Alpine, Angora et Saanen. Pour finir les Savoie ont des coefficient plus faibles que les Créoles, Pyrénées et Saanen. Le reste des races ne se différencient pas.

■ R_{OH}

Les résultats des proportions de R_{OH} présents dans les génomes de chaque race sont présentés dans la Figure 41.

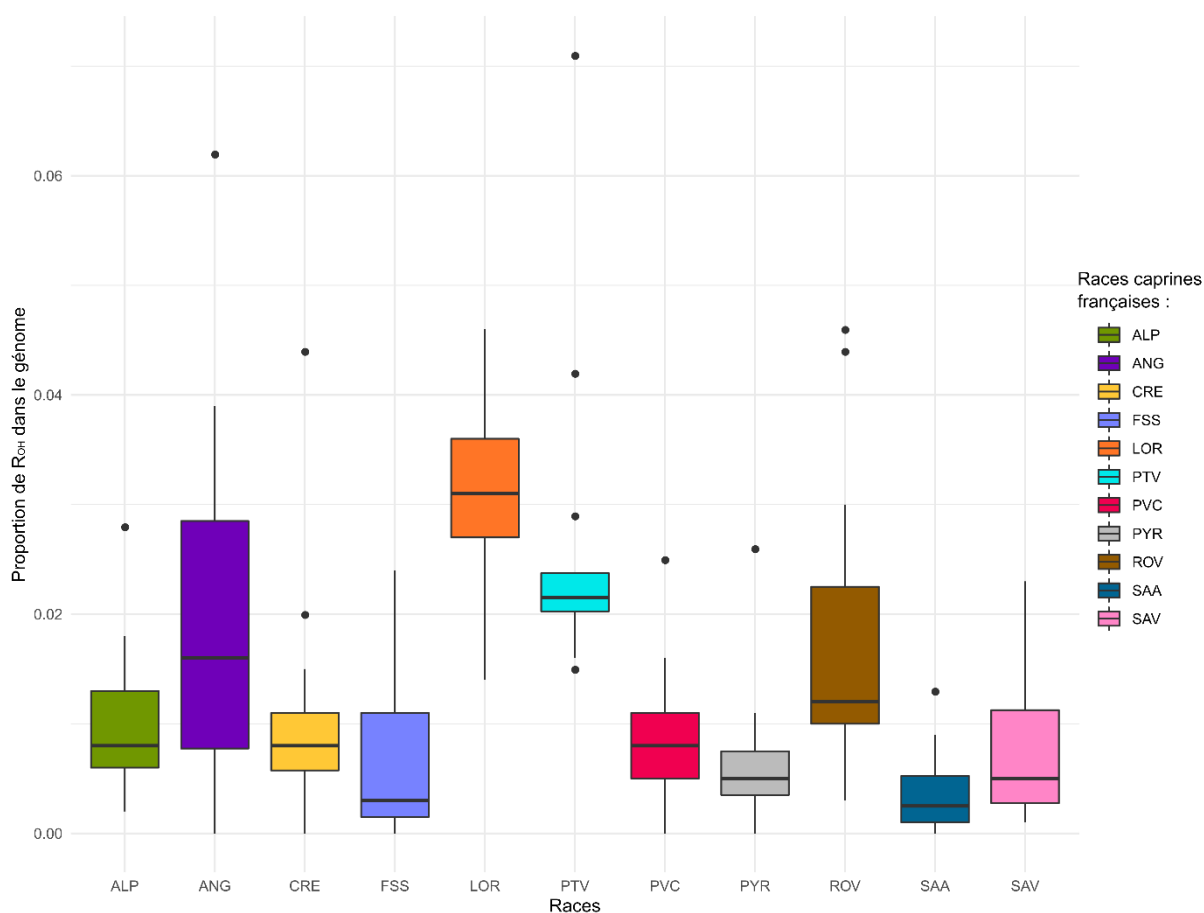


Figure 41 : Proportion de ROH dans le génome pour chaque race. Les boîtes à moustaches représentent la distribution des valeurs de FIS pour chaque race (la partie colorée est délimitée par la valeur de FIS correspondant à 25% et 75% du jeu de données. La ligne noire au milieu correspond à la valeur médiane).

Tableau IX : p-valeurs du test de Wilcoxon par paire pour trouver les proportions de R_{OH} au sein du génome plus élevées (matrice triangulaire supérieure, dégradé d'orange) ou plus faibles (matrice triangulaire inférieure, dégradé de bleu) que les autres. Les cases claires sont significatives (p -valeurs $\leq 0,05$), les intermédiaires hautement significatives (p -valeurs $\leq 0,01$) et les foncées extrêmement significatives (p -valeurs $\leq 0,001$).

	ALP	ANG	CRE	FSS	LOR	PTV	PVC	PYR	ROV	SAA	SAV
ALP		0,087	1,000	1,000	3.10^{-4}	7.10^{-5}	1,000	1,000	0,141	1,000	1,000
ANG	1,000		1,000	1,000	0,030	0,165	1,000	1,000	1,000	1,000	1,000
CRE	0,487	0,076		1,000	6.10^{-4}	8.10^{-4}	1,000	1,000	0,120	1,000	1,000
FSS	0,030	0,011	0,167		2.10^{-4}	7.10^{-4}	0,393	0,921	0,030	1,000	0,847
LOR	1,000	1,000	1,000	1,000		1,000	1,000	1,000	1,000	1,000	1,000
PTV	1,000	1,000	1,000	1,000	0,050		1,000	1,000	1,000	1,000	1,000
PVC	0,512	0,075	0,830	1,000	2.10^{-4}	7.10^{-4}		1,000	0,141	1,000	1,000
PYR	0,024	0,011	0,129	1,000	1.10^{-4}	3.10^{-4}	0,148		0,011	1,000	1,000
ROV	1,000	0,715	1,000	1,000	0,033	0,052	1,000	1,000		1,000	1,000
SAA	4.10^{-5}	2.10^{-4}	0,006	0,351	2.10^{-5}	2.10^{-4}	0,011	0,073	2.10^{-4}		0,141
SAV	0,087	0,030	0,416	1,000	2.10^{-4}	7.10^{-4}	0,416	0,915	0,030	1,000	

Le test de Kruskal-Wallis que nous avons effectué, afin de comparer les proportions de R_{OH} au sein du génome de chaque population, est significatif (p -valeur = $2,446.10^{-13}$) donc il y a au moins une population différente des autres. Les résultats du test de Wilcoxon par paire (Tableau IX) nous permettent de dire que les races Lorraine et Poitevine ont une proportion de R_{OH} plus grande que la quasi-totalité des autres races. À l'inverse la race Saanen possède la plus faible proportion de R_{OH} dans son génome.

Nous avons extrait de nos données les consanguinités et proportion de R_{OH} moyens par race qui ont été mis en relation avec les données de consanguinité fournies par l'IDELE (2019). Celle-ci est calculée suivant la méthode de Meuwissen & Luo (1992). Le tout est présenté dans le Tableau X.

Tableau X : moyenne des F_{IS} et des proportions de R_{OH} trouvés dans les races caprines françaises ainsi que la probabilité d'identité des gènes issue des indicateurs de variabilité génétique de l'IDELE (2019). Si la race a une valeur moyenne supérieure ou inférieure à plus de la moitié des autres races (6/10) alors la case est colorée en rouge ou bleue respectivement.

Races	F_{IS}	R_{OH}	Consanguinité moyenne (%)	Évolution sur 10 ans (%)
ALP	$0,017 \pm 0,074$	$0,010 \pm 0,005$	2,8	+ 0,23
ANG	$0,038 \pm 0,086$	$0,019 \pm 0,016$	5,5	+ 4,13
CRE	$0,029 \pm 0,103$	$0,010 \pm 0,010$	-	-
FSS	$0,016 \pm 0,075$	$0,006 \pm 0,007$	4,8	- 1,27
LOR	$-0,066 \pm 0,055$	$0,031 \pm 0,009$	5,8	- 1,21
PTV	$-0,009 \pm 0,065$	$0,027 \pm 0,016$	3,3	+ 0,14
PVC	$-0,034 \pm 0,051$	$0,009 \pm 0,007$	4,7	+ 0,33
PYR	$0,033 \pm 0,084$	$0,006 \pm 0,006$	1,4	- 0,7
ROV	$0,003 \pm 0,086$	$0,018 \pm 0,015$	-	-
SAA	$0,033 \pm 0,085$	$0,004 \pm 0,003$	3,3	+ 0,25
SAV	$-0,031 \pm 0,028$	$0,007 \pm 0,007$	-	-

Nous pouvons observer que les résultats que nous avons obtenus sont très peu concordants avec ceux fournis par l'étude des pedigrees. Si nous prenons l'exemple de la race Angora, celle-ci a le deuxième coefficient de consanguinité le plus élevé (5,5%) et une augmentation de 4,13% de sa consanguinité sur 10 ans et pourtant nous ne la trouvons pas significativement différentes des autres races si nous nous basons sur son F_{IS} ou sa proportion de R_{OH} .

Nous avons trouvé que les chèvres de Lorraine avaient un coefficient de consanguinité plus bas que 6 des 10 autres races pourtant c'est celle qui a la consanguinité moyenne la plus élevée d'après le rapport de l'IDELE (5,8%). De plus nous trouvons des proportions des R_{OH} plus important dans cette race par rapport aux autres, ce qui semble indiquer des croisements consanguins récents pourtant le rapport indique que la consanguinité est en baisse (-1,21%) sur une période de 10 ans. Il est donc possible que les croisements consanguins que nous détectons soient antérieurs à la période étudiée dans le rapport.

Tout comme pour la race Poitevine chez laquelle nous détectons une proportion de R_{OH} plus importante que chez les autres races, alors que l'accroissement de la consanguinité est très limité (0,14%) sur 10 ans.

Signature de sélection

Analyse globale

L'exécution du programme hapFLK nous a permis d'identifier des haplotypes sélectionnés au sein des races caprines métropolitaines (Figure 42).

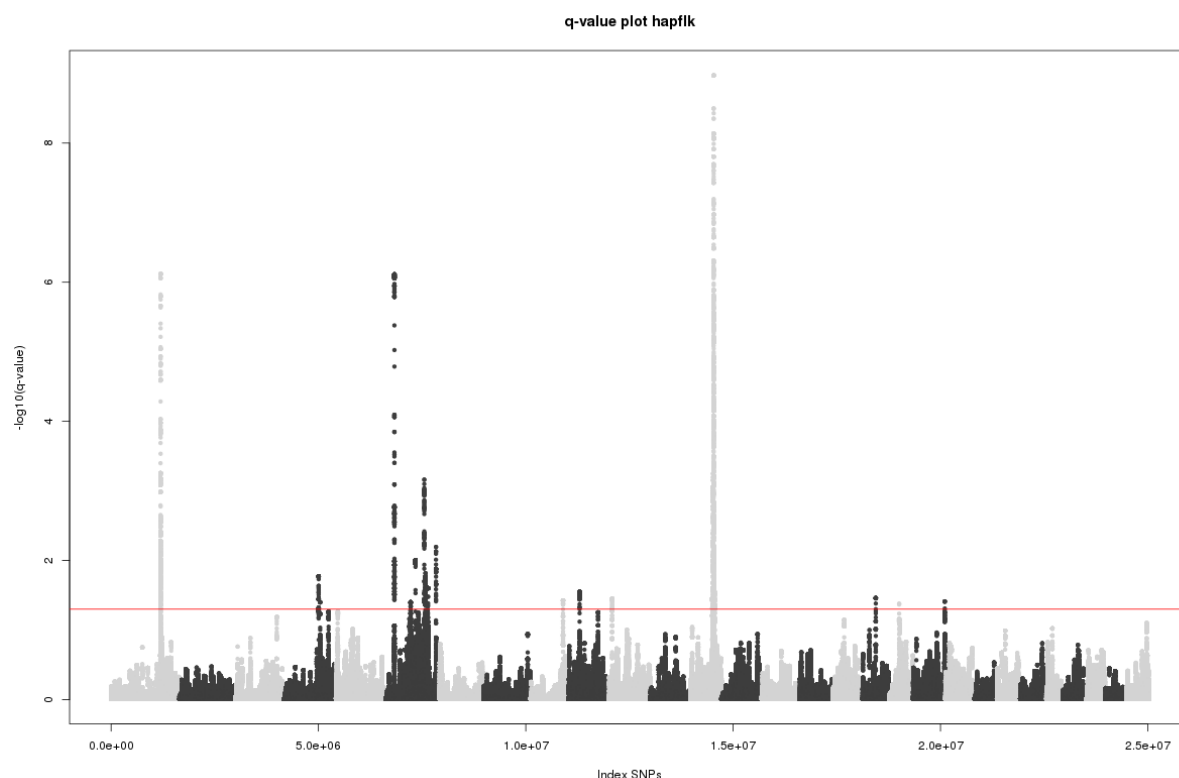


Figure 42 : Résultats du test hapFLK sur l'ensemble du génome caprin. L'axe des ordonnées représente la significativité des SNPs et l'axe des abscisses la position sur le génome avec une alternance de couleurs en fonction des différents chromosomes. La ligne horizontale correspond au seuil de q-valeur 0,05.

Les SNPs composant les haplotypes sous sélection ont été regroupés en 29 régions comportant 33 gènes. Parmi ceux-ci, nous en comptabilisons 21 candidats impliqués dans la reproduction (6 d'entre eux), la production laitière (6 gènes), l'immunité (4 gènes) et des caractères morphologiques (6 gènes). Ceux-ci sont présentés ainsi que les races porteuses des haplotypes sélectionnés dans le Tableau XI.

Le tableau de résultats complets, les arbres locaux et les haplotypes présents dans les régions sous sélection sont disponibles en Annexe (Tableau S3 et Figures S6 à S33).

Tableau XI : Régions sélectionnées mises en évidence chez les races caprines métropolitaines par le programme hapFLK ainsi que les gènes candidats et les races associés à celles-ci.

Chromosome	Début région	Fin région	Gènes	Fonction	Races
1	114496500	114631500	P2RY1	MORPHOLOGIE Croissance	ALP LOR PTV ROV
4	87189300	87217500	DMTF1	MORPHOLOGIE Musculature	ALP PTV SAV
6	85778600	85814000	SULT1B1	LACTATION Proche des gènes de la caséine	LOR PTV ALP PYR
6	90866000	90893500	RCHY1	LACTATION Production de lait	FSS LOR PTV
6	91804000	91829000	SHROOM3	IMMUNITÉ Résistance aux mammites et parasites	LOR PTV
9	79278800	79301900	NOX3	IMMUNITÉ Immunité innée	ALP SAV PYR
10	27600000	27628300	RHOJ	REPRODUCTION Cycle ovarien LACTATION Glande mammaire	LOR PTV SAV
11	7043000	7052500	IL18R1	IMMUNITÉ Résistance aux mammites	LOR PTV SAA ALP
13	59512000	59583000	ANGPT4	LACTATION Vitesse de traite	SAA SAV
13	59512000	59583000	FAM110A	LACTATION Glande mammaire	SAA SAV
13	61260000	61302000	NOL4L	REPRODUCTION Cycle ovarien	LOR SAA SAV
13	62258000	62319000	BPIFA1	MORPHOLOGIE Musculature	ROV LOR PTV

13	62382000	62461500	CDK5RAP1	MORPHOLOGIE Croissance	FSS SAA ROV LOR PTV
13	62461500	62514500	SNTA1	MORPHOLOGIE Musculature	ROV PTV FSS
13	63160000	63288000	ASIP	MORPHOLOGIE Couleur de la robe	SAA ALP SAV PTV
13	64051000	64099000	MMP24	IMMUNITÉ Susceptibilité aux tumeurs induites par le papillomavirus	FSS LOR PTV PVC SAV
13	64394000	64433000	SPAG4	REPRODUCTION Spermatogenèse	LOR PTV
13	64394000	64433000	CPNE1	REPRODUCTION Joue un rôle au niveau du placenta	LOR PTV
13	64708000	64737000	CNBD2	REPRODUCTION Spermatogenèse	SAA PTV FSS SAV LOR
13	64776000	64845000	EPB41L1	REPRODUCTION Fertilité	ALP LOR PTV FSS SAV SAA
13	65132000	65151000	DLGAP4	LACTATION Taux de cholestérol du lait	ALP SAA PTV LOR FSS ROV SAV

- Les caractères de reproduction

Les caractères en lien avec la reproduction sont principalement liés avec le cycle ovarien (RHOJ et NOL4L chez les vaches et les cochons, Roberts & McLean, 2011; Li et al., 2016; Talebi et al., 2018) et la spermatogénèse (gènes SPAG4 et CNBD2, Bai et al., 2017; Frischknecht et al., 2017; Khayatzaadeh et al., 2019). La protéine produite par le gène CPNE1 est retrouvée dans le placenta (Frischknecht et al., 2017; Mei et al., 2019). Le rôle du gène EPB41L1 n'est pas détaillé mais l'étude de Gaddis et al. (2016) montre que ce gène est impliqué dans la fertilité des bovins.

Les races portant les haplotypes sélectionnés associés à ces gènes sont la Lorraine, la Poitevine, la Savoie, la Fossés et la Saanen.

- La production laitière

Nous trouvons 2 gènes en relation avec la glande mammaire (RHOJ et FAM110A, Chomwisarutkun et al., 2012; Gutiérrez-Gil et al., 2014), et la vitesse de traite ANGPT4 (Jardim et al., 2018). La production laitière est influencée par le gène RCHY1 (Du et al., 2020) et 2 autres ont une influence sur la composition du lait (DLGAP4 et SULT1B1 qui peuvent réguler les taux de cholestérol ou de caséine dans le lait, Do et al., 2018; Bertolini et al., 2018).

Toutes les races, excepté les Provençales, portent au moins un haplotype sélectionné en relation avec ce phénotype.

- L'immunité

Les gènes SHROOM et IL18R1 sont impliqués dans la résistance des chèvres aux mammites (Schulman et al., 2009; Cai et al., 2018), le gène NOX3 joue un rôle dans l'immunité innée (Lambeth, 2004) et le gène MMP24 impliqué dans la susceptibilité aux tumeurs induites par le papillomavirus chez les chevaux (Staiger et al., 2016).

Les signatures de sélection relatives à ces gènes sont principalement présentes chez les races Lorraine, Poitevine et Alpine ; ainsi que Savoie, Saanen et Pyrénées.

- Caractères morphologiques

Nous trouvons 5 gènes directement liés à la musculature et la croissance des animaux (P2RY1, CDK5RAP1, DMTF1, SNTA1 et BPIFA1, Fernández et al., 2012; Keel et al., 2018; Ahbara et al., 2019; Lee et al., 2019; An et al., 2020) dont les haplotypes sélectionnés sont présents majoritairement chez les chèvres du Rove, Poitevines, Alpines, Lorraines, Fossés et Provençales.

Le dernier gène code pour la couleur de la robe (ASIP, Henkel et al., 2019) et est sélectionné chez les races Saanen, Alpine, Savoie et Poitevine.

- Prédications de l'effet des variants

Les résultats obtenus avec l'analyse « Variant Effect Predictor » d'Ensembl sur l'ensemble du jeu de données ou seulement les variants sous sélection sont présentés dans le Tableau XII.

Tableau XII : résultats de l'analyse « Variant Effect Predictor » d'Ensembl sur l'ensemble des variants présents dans notre jeu de données sur les races caprines françaises et sur les variants détectés sous sélections par la programme hapFLK.

Position	Tous les variants	Pourcentage	Variants candidats	Pourcentage
Dans un intron	22 801 176	47%	917 169	45%
Inter-génique	20 519 097	42%	894 301	44%
En amont du gène	2 449 768	5%	100 454	5%
En aval du gène	2 309 928	5%	102 186	5%
Dans un exon	424 580	1%	16 066	1%
Autre	70 150	< 1%	2 682	< 1%
Total	48 574 699	100%	2 032 858	100%

Tout d'abord, 61% de l'ensemble des variants présents dans les jeux de données des chèvres françaises correspondent à des variants connus par la base Ensembl et 39% sont de nouveaux variants.

Le test du χ^2 ne nous a pas permis de montrer une distribution des variants différente entre les deux jeux de variants (p-valeur = 0,2243). Nous pouvons remarquer que dans les deux jeux

de données environ 43% des variants sont positionnés entre les gènes et 10 % dans des régions proches (en amont ou en aval). Quand on prend en compte les variants présents dans des gènes, 46% sont positionnés dans des introns c'est-à-dire une région non codante. Seulement 1% des variants de nos jeux de données sont localisés dans des régions codantes appelées exon avec 0,49% des variants de notre jeu de données impliqués dans des mutations synonymes, (pas de modification de la protéine produite), 0,38% dans des mutations faux sens (modification de la protéine) et 0,01% qui entraîne la perte ou l'ajout de codons « start » ou « stop » pouvant tronquer la protéine.

Analyse ciblée

Pour cette analyse nous contrastons une race par rapport à toutes les autres pour trouver les signatures de sélection qui lui sont propres. Les résultats, obtenus avec le programme hapFLK_{adapt}, sont présentés dans le Tableau XIII. Les « manhattan plots » et les tableaux de résultats complets par race sont disponibles en Annexe (Tableaux S4 à S10 et Figures S34 à S42).

Tableau XIII : Nombre de signatures de sélection spécifique à chaque race caprine métropolitaine et mises en évidence par le programme hapFLK_{adapt} avec des q-valeurs inférieures ou égales à 1.10^{-4} .

Races	ALP	FSS	LOR	PTV	PVC	PYR	ROV	SAA	SAV
Nombre de régions	29	15	393	145	0	15	0	53	3

Nous pouvons noter qu'avec un seuil de q-valeurs inférieur ou égal à 1.10^{-4} les races Provençale et Rove n'ont aucune signature spécifique donc les haplotypes sélectionnés chez ces races le sont aussi dans les autres races françaises. Pour les autres races nous avons recherché les gènes associés au top 20 régions.

▪ Chèvres Alpines

L'analyse hapFLK_{adapt} nous permet de trouver 29 régions comportant des SNPs ayant une q-valeur inférieure ou égale à 1.10^{-4} . Nous avons cherché les gènes associés aux 20 régions les plus significatives (q-valeur la plus significative par région) et nous répertorions 16 gènes en relation avec celles-ci. Ces gènes sont impliqués dans la reproduction (1 d'entre eux), la lactation (6), des caractères morphologiques (7) et d'immunité (3).

Le gène PARM1 est lié à différents caractères comme la fertilité mais aussi le rendement laitier et le taux de protéines dans le lait (Cochran et al., 2013). Deux autres gènes sont impliqués dans des modifications de la composition du lait : CSN1S1 et CSN2 impactent le taux de caséine dans le lait et ainsi sa coagulation permettant la production fromagère. Le premier gène code pour la caséine α S1 et le second pour la caséine β (Martin & Leroux, 2000). Le gène CPT1A joue un rôle dans l'absorption et le métabolisme des acides gras par le foie (McGarry & Brown, 1997) et a été trouvé différenciellement exprimé dans cet organe chez des vaches laitières en fonction des périodes de lactation (Wiedemann et al., 2013). Enfin, les gènes AHCYL2 et ADAMTS3 sont en relation avec la santé et la conformation de la mamelle (Pausch et al., 2014; Choudhary et al., 2018).

Nous trouvons 7 gènes en relation avec des caractères morphologiques. Le gène CNTNAP2 est en relation avec le comportement alimentaire et donc peut avoir une influence sur la masse corporelle chez les souris (Buchner et al., 2012). Trois autres gènes sont associés à la musculature chez les bovins élevés pour leur viande (SNX24, SNX2, Tizioto et al., 2015). Le gène CALD1 est associé à la contraction des muscles lisses vasculaires en relation avec une variable d'altitude des chèvres sud-africaines (Mdladla et al., 2018). Concernant le gène FMN1 il est associé à la présence de pampilles chez la chèvre (Reber et al., 2015). Pour finir avec les caractères physiques, les gènes RALY et EIF2S2 portent des signatures de sélection en lien la couleur de la robe chez les chiens (Dreger et al., 2013) et les chèvres (Guo et al., 2018).

Si nous nous attardons sur les gènes associés à des caractères d'immunité nous trouvons le gène MTPN qui joue un rôle dans la résistance aux tiques et nématodes mais aussi dans la croissance et le développement musculaire (Makina et al., 2015). Concernant le gène ZBP1, il est connu pour avoir un rôle dans l'immunité innée (Bhuiyan et al., 2017). LUZP6 joue aussi un rôle dans le système immunitaire sans que celui-ci ne soit très bien connu (Evangelou et al., 2019).

- Chèvres des Fossés

Les 15 régions sous sélection dans cette race sont associées à 6 gènes codant pour des caractères physiques, de reproduction, de lactation, d'immunité ou autre.

Concernant la reproduction, le gène SNRPB est trouvé en relation avec la folliculogénèse chez le chat (Lee et al., 2018). En ce qui concerne la lactation, le gène LMX1A joue un rôle dans le

rendement laitier et de protéine et de graisse dans le lait (Du et al., 2020). Pour les caractères morphologiques, nous trouvons le gène NDFIP2 qui joue un rôle dans la masse corporelle chez le mouton d'après Pasandideh et al. (2018). Le gène TMEM117 est associé à la résistances aux mammites (Cai et al., 2018).

Les derniers gènes sélectionnés sont associés à des caractères sensitifs avec le gène TMC2 qui est associé aux cellules ciliées de l'oreille interne des mammifères (Pan et al., 2013) et le gène KIF21A avec des anomalies au niveau des yeux (Fink et al., 2012).

▪ Chèvres de Lorraine

Cette race est celle qui possède le plus de régions sélectionnées qui lui sont propres avec 49% des SNPs significatifs présents sur le chromosome 6. Si l'on s'intéresse aux 20 régions les plus significatives nous en trouvons 18 sur une région de 13Mb sur le chromosome 6. Celle-ci est associée à 33 gènes dont le gène KIT qui est impliqué dans l'apparition de taches blanches sur le pelage (Bai et al., 2011; Jivanji et al., 2019).

Les deux dernières régions sont positionnées sur le chromosome 14 dont une est en lien avec le gène HNF4G impliqué dans la musculature (Ramayo-Caldas et al., 2014).

▪ Chèvres Poitevines

L'analyse hapFLK_{adapt} nous permet de trouver 145 régions comportant des SNPs ayant une q-valeur inférieure ou égale à 1.10^{-4} . Nous avons cherché les gènes associés aux 20 régions les plus significatives et en avons répertoriés 24.

Nous trouvons par exemple 6 gènes en relation avec la reproduction. Les gènes TLN2 et FAS jouent un rôle dans le cycle d'œstrus (Slot et al., 2006; Dai et al., 2014). Les gènes CNBD2 et EPB41L1, associés respectivement à la spermatogénèse et à la fertilité, étaient déjà détectés chez cette race avec la première analyse hapFLK. Nous détectons également le gène UBTD1 qui joue un rôle dans la mise bas prématurée (Tiensuu et al., 2019) ainsi que le gène PIK3R1 qui régule la puberté chez les chèvres (Fortes et al., 2013).

Deux gènes sont liés à la production laitière : LRBA qui influence le taux de matière grasse dans le lait (Laodim et al., 2017) et SHTN1 le rendement laitier (Cai et al., 2020). Deux gènes sélectionnés sont en relation avec la morphologie mammaire : NEIL3 qui joue un rôle dans l'apparition de mamelon surnuméraire chez le mouton (Peng et al., 2017) et PYROXD2 qui influence la hauteur de la mamelle arrière (Gonzalez et al., 2020).

Si nous nous focalisons sur les régions portant des gènes associés à des caractères physiques, nous en comptabilisons 11. Parmi ceux-ci 7 sont liés à la musculature et au taux de graisses musculaires (ACTA2, ANKRD2, EXOSC1, FRAT1, FRAT2, SLIT1, ZDHHC16, Sun et al., 2011; Yuan, 2015; Komolka et al., 2016; Wang et al., 2017; Borowska et al., 2017; de Oliveira et al., 2019; Dall’Olio et al., 2020). Le gène PGAM1 est en relation avec le taux de graisse car il joue un rôle dans la glycolyse (Li et al., 2019). Le gène DGKH, quant à lui, régule la croissance des animaux (Widmann et al., 2013).

Nous trouvons un signal de sélection correspondant aux gènes ARHGAP19 et CHMP4B qui se situent juste avant la grande région sélectionnée lors de l’analyse hapFLK pour le gène ASIP. Il est donc fort probable que cette région soit associée à la couleur de la robe.

Nous trouvons une région sous sélection associée au gène GPM6A qui joue un rôle dans le développement du système nerveux (Alfonso et al., 2005) et SCAND1 qui est un répresseur de tumeur (Eguchi et al., 2019). Pour finir, la recherche bibliographique n’a pas permis de déterminer le phénotype qui était influencé par le gène RRP12.

▪ Chèvres des Pyrénées

Nous trouvons 15 régions sous sélection spécifiques à cette race, et associées à 7 gènes. Un premier CDH9 joue un rôle dans la reproduction car il est trouvé associé à la gestation chez le mouton (Li et al., 2014). Le gène SRGAP2 est associé au rendement laitier (da Costa Barros et al., 2018). Le gène GALNTL6 influence la croissance chez les bovins (Seabury et al., 2017).

Les 4 derniers gènes sont associés à des résistances aux mammites (TMEM117, Cai et al., 2018), à la maladie de Johne qui est une infection bactérienne répandue chez les bovins (PRICKLE1 et GXYLT1, Pant et al., 2010; Brito et al., 2018) et à la maladie de Huntington qui se traduit par une dégénérescence neurologique (Aigner et al., 2010). Benjelloun et al. (2015) trouve une signature de sélection concernant ce dernier gène chez les chèvres marocaines. Ils spéculent que ce gène joue un rôle dans l’adaptation du réticulum endoplasmique à des conditions physiologiques ou pathologiques.

▪ Chèvres Saanen

Quand nous observons les « top 20 » régions présentes chez cette race nous en trouvons une sur le chromosome 26 qui est relation avec le gène PCDH15 qui influence la concentration de neurotransmetteurs (Chen et al., 2020). Sur le chromosome 13 nous retrouvons des gènes

déjà présents dans l'analyse hapFLK : NOL4L jouant un rôle dans la reproduction et le gène ASIP en relation avec la couleur de la robe.

Les régions portées par le chromosome 19 se situent 21Mb et 29Mb et sont associées à 108 gènes. Cette région correspond à un QTL avec des effets pléiotropiques comme par exemple une implication dans la morphologie mammaire ; la production laitière ; les rendements de graisse, protéines et cellules somatiques dans le lait ; ainsi que le volume de sperme dans les éjaculats de boucs (Martin et al., 2018; Talouarn et al., 2020; Scholtens et al., 2020).

- Chèvres de Savoie

La dernière race présentant des régions sous sélection spécifiques est la Savoie. Nous comptabilisons 3 régions chacune associée à un gène. Il s'agit du gène OPTC sur le chromosome 16 qui semble impliqué dans la taille de la portée chez le cochon (Guo et al., 2016). Sur le chromosome 13 nous trouvons un premier gène nommé PREX1 qui influence la qualité de la viande chez les bovins (Xia et al., 2016). Le second gène positionné sur le chromosome 13 est AH CY, qui est situé en aval du gène ASIP, et influence la couleur de la robe.

IV. Discussion

Diversité neutre

Structure des races

Tout d'abord nous avons vu avec l'analyse de structure que les animaux échantillonnés se regroupent bien en fonction de leur race, et que celles-ci forment des groupes homogènes et différenciés.

Nous pouvons noter que les 5 chèvres Pyrénéennes et la chèvre de Lorraine échantillonnées en dehors de l'aire de répartition historique de leur race (Figure 33) s'agrègent avec le reste de leur groupe exceptée une chèvre Pyrénéenne (FRCH-PYR-0007) qui est échantillonnée en Aquitaine (comme 3 autres chèvres) et qui est proche des chèvres Fossés sur l'ACP. La présence de chèvres des Pyrénées en Aquitaine et en Bretagne est concordante avec la route de migration Océan-Nord des chevriers béarnais (Carte de Jean-Noël Passal, Figure S3 en Annexe) qui amenaient leurs chèvres jusqu'à Paris et au Nord de la France pour vendre leur

lait ainsi que les chèvres à la fin de la période de lactation (Passal, 2013). Cette pratique a disparu progressivement après la première guerre mondiale avec l'intensification de la circulation automobile et les progrès pour conserver les produits laitiers.

Les chèvres vendues à la fin de la période de lactation ont pu se trouver en contact avec des chèvres Poitevines et des Fossés élevées dans les régions traversées par la voie de migration. Nous avons trouvé un croisement entre ces 3 races avec l'analyse TreeMix, ce qui pourrait indiquer qu'effectivement ces races ont été en contact et que des reproductions ont pu avoir lieu entre elles.

Les croisements entre les chèvres Alpines et les Lorraines sont décrites dans l'histoire des chèvres de Lorraine (cf. partie 'Description des races'). Ils ont entraîné une « alpinisation » de la race, et l'étude des pedigrees sur la population actuelle estime qu'environ 40,5% du génome Lorrain est d'origine Alpine (IDELE, 2019). Cela montre que même si les croisements ont eu lieu dans les années 1970, le processus d'élimination des allèles alpins n'est pas encore terminé.

Malgré la présence d'allèles alpins dans son génome, la Lorraine est celle qui présente la plus grande différence génétique avec les autres races métropolitaines.

Il existe deux voies de colonisation de l'Europe occidentale : soit une voie méditerranéenne longeant la mer et passant par le détroit de Gibraltar, soit une route danubienne suivant ce fleuve. Une étude sur l'origine des moutons français effectuée par Rochus et al. (2018) a montré que les races de moutons au Nord de la France ont une origine danubienne et non méditerranéenne comme les races plus au Sud. De plus, l'étude de Nijman et al. (2020) montre que la diversité du chromosome Y, transmis par voie paternelle, sépare les races de chèvres du Sud-Ouest et du Nord-Est de l'Europe. Il serait possible que la race Lorraine, présente au Nord de la France, ait une origine danubienne alors que les autres races françaises auraient une origine méditerranéenne. Cette dernière a aussi pu être croisée, au moment des guerres, avec des animaux d'origine germanique donc ayant probablement une origine danubienne. Dans ce cas, la différence de la Lorraine ne viendrait pas de son origine, mais des croisements qu'il y a pu avoir avec des races d'origine différente. Le jeu de données VarGoats même complet ne devrait pas permettre de répondre pleinement à cette question car seulement 4 races du Nord-Ouest de l'Europe ont été échantillonnées (3 races de Landrace et une race Irlandaise).

Démographie des races

L'inférence démographique ne nous a pas permis de différencier une race ayant une histoire particulière. En effet, toutes présentent les mêmes variations au cours du temps, y compris les races non métropolitaines. Comme les races Créole et Angora sont issues d'importations, nous aurions pu nous attendre à observer une chute de la taille efficace due à un effet fondateur. Le fait que ces races soient un mélange de populations provenant de différents pays (Afrique, France et Inde pour la Créole, Canada, États-Unis, Australie et Afrique du Sud pour l'Angora), cet effet a pu être diminué par la diversité génétique apportée par chacune de ces populations.

La décroissance observée tout au long des 500 dernières générations peut être l'effet de la sélection des animaux pour améliorer leurs caractères de production. Cela a donné lieu, dans les années 1990, au concept de race (Food and Agriculture Organization of the United Nations, 1999) et à amener les éleveurs à reproduire les animaux portant le plus de caractères d'intérêt de la race. Au départ, les chèvres n'étaient pas considérées comme des animaux d'intérêt donc la sélection des différentes races c'est fait plus récemment (1948 pour la Poitevine, Vinatier Roche, n.d.); années 1970 pour les Alpines et Saanen et 1974 pour les Angoras (Allain, 1994). Les races locales sont aussi soumises à de la sélection afin que les animaux soient proches des standards de la race (décrits dans la présentation des races). Nous ne voyons pas d'effet de la sélection des races sur la taille efficace des races au niveau de l'inférence démographique.

De plus, les données historiques présentées dans la partie « description des races » font toutes référence à une forte chute des effectifs de ces races dans les années 70-80. Celle-ci n'est pas non plus retrouvée dans notre analyse d'inférence démographique où nous ne voyons pas de forte baisse entre les effectifs aux alentours de 50 générations (correspondant à 100 ans si nous prenons un temps de génération de 2 ans) et ceux à 10 générations (20 ans si nous prenons un temps de génération de 2 ans).

Si nous nous intéressons spécifiquement aux valeurs actuelles des tailles efficaces nous avons vu que les estimations du programme PopSizeABC se situent presque toujours entre les valeurs calculées par la méthode de Wright ou de Cervantès. La méthode de Wright se base seulement sur les effectifs des mâles et femelles adultes dans la population alors que Cervantès prend en compte l'accroissement de la parenté pondéré par le niveau de qualité des individus. C'est pour cela que nous trouvons systématiquement des effectifs plus élevés

avec la méthode de Wright qu'avec celle de Cervantès (Leroy et al., 2013). Les résultats obtenus avec le programme PopSizeABC permettant d'inférer les changements de tailles efficaces des races françaises semblent cohérentes avec les valeurs calculées sur les données réelles des populations, excepté pour les chèvres de Savoie où la taille efficace obtenue est plus importante que la taille de la population. Les erreurs de séquençage ont pour effet d'augmenter la taille efficace récente de manière artificielle donc nous avons regardé les données de couverture de séquençage et nombre d'allèles détectés chez cette race sans que cela nous ait permis de comprendre ce résultat.

Après discussion avec Coralie Danchin (IDELE) sur les résultats obtenus avec la méthode de Cervantès, il semblerait que le résultat des Pyrénées ($N_e = 249$ selon la méthode de Cervantès) ne soit pas représentatif de la race. En effet, les pedigrees ne sont pas bien renseignés (50% des femelles ont des pedigrees n'excédant pas 3 générations alors qu'en moyenne 7 générations sont connues pour 50% des chèvres testées par l'IDELE) donc cela entraîne un biais dans le calcul de la consanguinité. L'estimation obtenue avec PopSizeABC pourrait aussi être impactée par une couverture de séquençage significativement moins bonne (test de Wilcoxon par paire avec les autres races) chez cette race par rapport aux Lorraines, Poitevines Rove et Savoie. Il pourrait être intéressant de refaire notre analyse en augmentant le seuil de MAF de 5% à 10% afin de voir si les tailles efficaces sont modifiées et si les résultats obtenus sont plus cohérents avec les données obtenues sur les pedigrees des animaux.

Compte tenu que nos résultats peuvent être biaisés nous allons utiliser les résultats obtenus par l'analyse des pedigrees pour discuter certains résultats.

La race Lorraine est celle qui a la taille efficace la plus faible ($N_e = 32$). Cela peut être la conséquence de deux phénomènes : le fort goulot d'étranglement dans les années 70 quand la race a été délaissée au profit des races en sélection ; et la purge des allèles alpins dans la population. Dans ce cas, les animaux présentant les standards de la race sont utilisés pour la reproduction afin que leurs allèles lorrains se répandent dans la population au détriment des allèles alpins. Cela peut aussi expliquer la forte proportion de fragments homozygotes chez cette race. Ces fragments sont les témoins de croisements consanguins récents, ce qui est concordant avec la purge des allèles alpins. Cela se mesure aussi sur la consanguinité globale de la race car c'est la race ayant la plus forte valeur de consanguinité (5,8%).

La race ayant la seconde taille efficace la plus faible est la race Provençale ($N_e = 62$). Nous avons vu précédemment que c'est une des races qui possède le plus de diversité de phénotypes. Les données historiques collectées sur la race nous indiquent que dans les années 1990 on ne dénombrait pas plus de 200 animaux présentant les standards de la race. Ce goulot d'étranglement peut être à l'origine de la taille efficace observée actuellement. Cependant, nous avons vu dans la présentation de la race que celle-ci présente de nombreux phénotypes, malgré un goulot d'étranglement. Cela s'explique car les éleveurs de cette race appliquent une sélection « diversifiante » afin de maintenir la diversité génétique de la race (nombre limité de saillies par boucs, échanges fréquents des animaux, Association pour la Sauvegarde de la Chèvre Provençales, 2017).

La deuxième race ayant une forte proportion de R_{OH} dans son génome est la race Poitevine. Même si des chèvres Alpines ont été utilisées pour reconstituer les troupeaux après l'épidémie de fièvre, cette race n'est pas dans une démarche de purge des allèles alpins comme la Lorraine. De plus, récemment certains éleveurs ont mis en place des élevages par croisement d'absorption à partir d'Alpine. Nous ne voyons pas ces croisements dans notre jeu de données car les animaux échantillonnés ont été choisis pour avoir la plus grande diversité de génotypes, ne pas être apparentés et ne pas être issus de croisements récents.

De manière général notre analyse de variabilité génétiques se base sur un faible nombre d'individus analysés ne représentant qu'une petite partie de la race ce qui peut expliquer pourquoi nos résultats sont si différents (en particulier pour la consanguinité) de ceux estimés par l'IDELE. Certaines tendances sont retrouvées mais il faudrait que nous analysions plus d'animaux tirés aléatoirement dans la race afin d'avoir de meilleur résultat concernant la variabilité génétique globale de chaque race.

Sélection

Après nous être intéressés à l'histoire des races, nous avons cherché les signatures de sélection chez celles-ci. Nos analyses ont permis de mettre en évidence des gènes candidats en relation avec des voies d'intérêt agronomique impliquées dans la reproduction, la lactation,

l'immunité et la présence de certains caractères morphologiques, ce que nous allons détailler ci-après.

La reproduction

Quand nous comparons les races entre elles nous trouvons des haplotypes sélectionnés associés à cette fonction les races Lorraine, Poitevine Savoie, Fossés et Saanen. Avec les analyses ciblée nous trouvons au moins un gène candidat dans chaque race, exceptées les Provençales et les Rove chez qui nous n'avons pas trouvé de signature spécifique.

Étant donné qu'aucun relevé de prolificité n'est effectué sur les races locales il est difficile de savoir quels sont les effets des gènes sélectionnés chez les différentes races et en particulier les Lorraines, Poitevines et Savoie qui pourraient avoir un phénotype différent des autres races.

Concernant les Lorraines, un élément de réponse peut venir de l'intervalle de génération moyen qui est plus bas dans cette race (2,9 ans) par rapport aux autres races étudiés dans les rapports d'indicateurs de variabilité génétique de l'IDELE. Ce temps pourrait venir d'une pratique d'élevage mais aussi de meilleures aptitudes maternelles (capacité à s'occuper des chevreaux) chez cette race par rapport aux autres.

La production de lait

Les races portant les haplotypes sélectionnés sont principalement les Saanen, Lorraines, Poitevines et Savoie. Parmi ces 4 races, 2 sont sélectionnées pour améliorer leur qualité laitière : la race Saanen et la race Poitevine.

Connaissant l'histoire commune entre les races Savoie et Lorraine avec l'Alpine qui est sélectionnée pour améliorer ses qualités laitières, il serait possible que des haplotypes en relation avec la production laitière soient partagés entre ces races.

Lors de l'analyse ciblée, 6 gènes sont sélectionnés spécifiquement chez les Alpines. Parmi ces gènes nous retrouvons les gènes CSN1S1 et CSN2. Sept variants sont connus pour le gènes CSN1S1 associés à 4 niveaux de synthèse de la caséine α S1. Les allèles A, B et C sont associés à un fort taux de caséine α S1 dans le lait ; l'allèle E à un taux moyen ; les allèles D et F à un taux faible ; et pour finir le lait des chèvres homozygotes porteuses de l'allèle O (nul) est dépourvu de caséine α S1 (Martin & Leroux, 2000).

Les allèles associés à de fort taux été fortement sélectionné chez les races Alpine et Saanen dans les années 1990 et l'allèle A est présent à 49% chez les boucs en sélection (Manfredi et al., 1995). De nos jours cet allèle est présent à 50% en Saanen et 80% en Alpine (Isabelle Palhière, communication personnelle).

Concernant la race Poitevine, l'étude de Ricordeau et al. (1996) montrait aussi une forte fréquence des allèles B chez cette race mais nous ne trouvons pas de signature spécifique à cette race dans cette région.

L'immunité

Les signatures de sélection relatives à ces gènes sont principalement présentes chez les Lorraines, Poitevines et Alpines ; ainsi que les Savoie, Saanen et Pyrénées. Des signatures spécifiques sont retrouvées chez les Alpines et les Pyrénées.

Les gènes associés à cette voie sont majoritairement en relation avec la résistance aux mammites. Celles-ci se présentent sous la forme d'une réaction inflammatoire de la glande mammaire souvent due à des bactéries (Wilson et al., 1997) et une augmentation du nombre de cellules somatiques dans le lait. Cette maladie peut entraîner des pertes économiques pour les éleveurs car la production est affectée, le lait est payé moins cher et les éleveurs ont des coûts pour soigner leurs bêtes (Seegers et al., 2003). La résistance aux mammites fait l'objet de nombreuses recherches (Rupp et al., 2009; Palhière et al., 2015) et programmes partenariaux (De Cremoux et al., 2018).

Les caractères morphologiques

- La carrure

Les haplotypes sélectionnés sont présents chez les chèvres du Rove, Poitevines, Alpines, Lorraines, Fossés et Provençales.

Ces 6 races présentent plusieurs types de carrure : de grandes chèvres avec un poids important (Alpine et Rove), de grandes chèvres avec un poids plus faible (Poitevine), des chèvres de taille moyenne avec un poids important (Lorraine), des chèvres de taille moyenne (Provençale) ou de petite taille avec un faible poids (Fossés). Nous ne pouvons donc pas faire de lien direct entre les gènes candidats détectés et la carrure des animaux.

La carrure des animaux peut aussi être reliée à son environnement. En effet, les chèvres Rove transhumaient avec les troupeaux de mouton donc la sélection d'animaux musclés, pouvant répondre à ces contraintes peut expliquer la présence de ces gènes sélectionnés. De plus, cette race produit aussi des cabris pour la consommation donc des gènes en relation avec la musculature peuvent avoir une influence sur la qualité de la viande.

Quatre signatures de sélection sont trouvées spécifiquement chez les Alpines. Celles-ci peuvent être en relation avec son origine localisée dans les Alpes qui a pu favoriser sa musculature, mais également avec la production laitière. En effet, Rupp et al. (2015) ont trouvé une corrélation entre la taille des moutons et leur production laitière liée à l'action du gène SOCS2.

Pour finir, 7 gènes sont associés à la musculature et à la proportion de graisse dans les muscles chez la chèvre Poitevine qui pourraient être liés à la taille de cette race. Quatre de ces gènes sont détectés chez des cochons et deux chez des vaches qui sont élevés pour leur viande. Ce n'est pas le cas de la race Poitevine qui est sélectionnée sur des caractères laitiers, donc nous ne comprenons pas à quel(s) phénotype(s) peuvent être liés ces gènes.

- La couleur de la robe

La région sur le chromosome 13 entre 62,8 Mb et 63,5Mb portent 13 gènes dont ASIP, est sélectionnée chez les races Poitevine, Alpine, Saanen et Savoie dans l'analyse globale. Elle est également retrouvée chez ces mêmes 4 races dans l'analyse ciblée, ce qui signifie que chacune d'elle porte une signature spécifique sur cette région. Nous avons représenté les q-valeurs obtenues avec l'analyse hapFLK_{adapt} pour les 4 races dans cette région (Figure 43).

Nous voyons effectivement bien sur la Figure 43 que les signaux de chaque race sont positionnés à des endroits différents de la région : la Poitevine est la plus en amont, puis vient l'Alpine, la Saanen avec un pic dans le gène ASIP, puis la Savoie pour finir. Ces 4 races ont des couleurs de robes très contrastées : cape de maure pour la Poitevine, chamoisée pour l'Alpine, blanche pour la Saanen et on retrouve 4 types de coloris allant du noir au blanc chez les chèvres de Savoie.

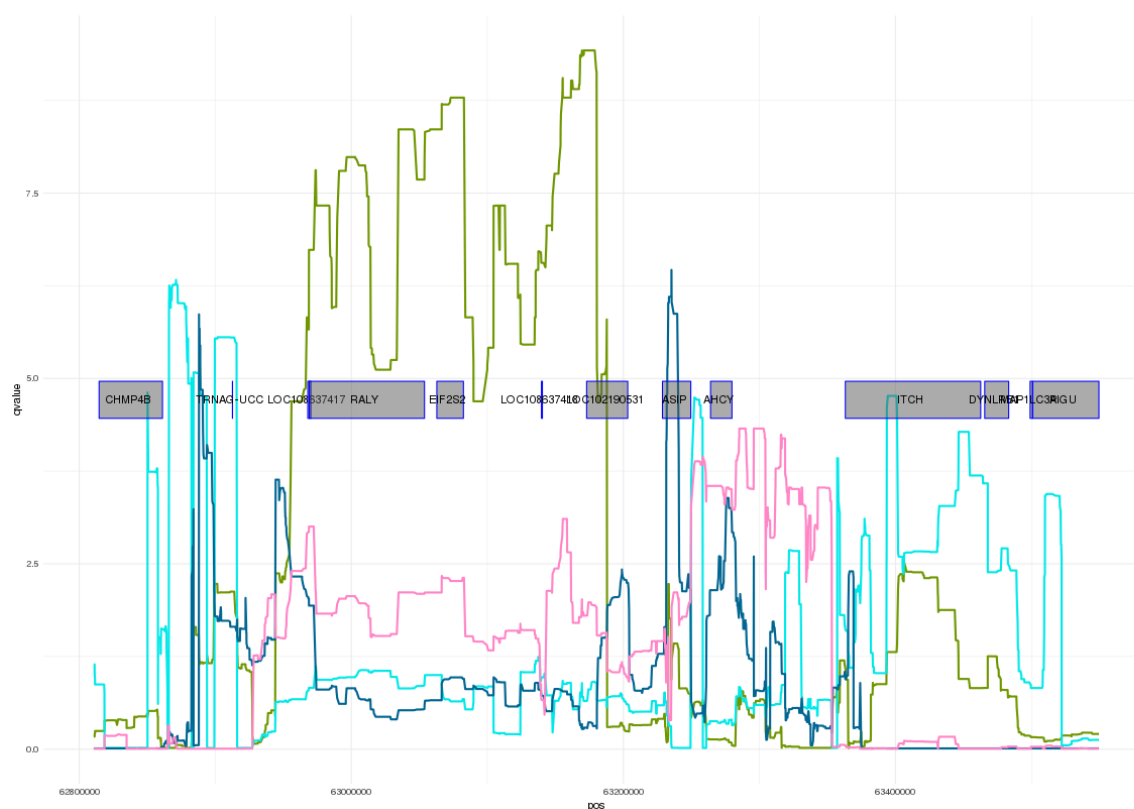


Figure 43 : Représentation des q-valeurs obtenues avec l'analyse hapFLKadapt pour les races Alpine, Poitevine, Saanen et Savoie chacune représentée par une couleur. L'axe des ordonnées représente la significativité des SNPs et l'axe des abscisses la position sur le génome.

Nous nous sommes alors posé la question si ces différents signaux pouvaient être directement reliés à la couleur des animaux pour les 3 races ayant des patrons de couleur fixes. Henkel et al. (2019) ont mis en relation des CNVs à différents endroits de la région ASIP avec la couleur de la robe des animaux. Ainsi des CNVs en amont du gène ASIP (position proche de 63,15 Mb) sont en relation avec la couleur noire ; des CNVs au début du gène sont liés à la couleur chamoisée ; et des CNVs au niveau de la fin du gène sont reliés à la couleur blanche. Nous retrouvons le même ordre dans les pics significatifs obtenus pour les races Poitevine, Alpine et Saanen mais les positions ne sont pas les mêmes que celles présentées dans l'article. La région étudiée dans l'article s'étend de 63,10 Mb à 63,40 Mb alors que les haplotypes détectés sous sélection sont situés entre 62,80 Mb et 63,25 Mb si nous excluons le signal de la race Savoie.

L'étude de Cumer et al. (*Submitted to GBE*) a montré que des différenciations génétiques entre SNPs pouvaient être associées à des variants structuraux sous sélection. Les SNPs étant à moins de 10 Kb de variants structuraux (SV) sous sélection sont plus différents que ceux étant à la même distance de SV non sélectionnés, et cette différence s'amointrit au fur et à mesure que la distance entre les SNPs et le SV augmente. Si nous comparons la fin des pics significatifs

obtenus avec les SNPs et les débuts des régions CNVs pour les différents patrons de couleur nous obtenons des signaux chevauchants pour la couleur blanche (17 Kb en commun) et chamoisée (33 Kb en commun), ainsi qu'une différence de 210 Kb entre les deux signaux pour la couleur cape de maure. Une détection des CNVs sur notre jeu de données nous permettrait de vérifier la position de ceux-ci et ainsi vérifier s'il peut y avoir un signal concordant entre les SNPs et les CNV sélectionnés dans cette région.

Effets des variants

Lorsque nous nous intéressons à l'effet potentiel des variants et donc à leur position par rapport aux gènes, nous ne pouvons pas différencier la distribution des positions des variants sélectionnés de celle de tous les variants du génome. Dans les deux cas, les variants présents dans des exons représentent 1% du jeu de données et 0,01% de ce dernier est en relation avec la perte ou le gain de codons « start » ou « stop » entraînant donc une modification de la protéine produite. La majorité des variants (49%) sont dans des régions inter-géniques ou en amont d'un gène. Ces régions peuvent contenir des régions régulatrices mais l'annotation du génome ne nous permet pas de savoir si des variants sélectionnés sont localisés dans ce genre de région.

De plus, la caractérisation des phénotypes autres que morphologiques n'est pas réalisée systématiquement chez les races locales ce qui ne nous permet pas de mettre en lien les variations alléliques que nous avons détectées sous sélection avec un phénotype quantifié chez ces races. Il n'est pas prévu, dans le projet VarGoats, de faire des analyses de corrélations entre le génotype et les phénotypes des animaux donc il n'y a pas eu d'effort d'échantillonnage sur les phénotypes de chaque race. L'objectif du projet est d'analyser un jeu de données plus global afin de comprendre les différences qu'il existe entre les races de différentes origines géographiques, leurs relations et les sélections qu'elles partagent ou qui sont propres à chacune. Cela permettant d'avoir une vision d'ensemble de la diversité des races caprines dans le monde et d'étudier l'histoire évolutive des chèvres domestiques.

MECANISMES ÉPIGENETIQUES

Résumé de l'Article

Cet article présente l'analyse des mécanismes génétiques et épigénétiques chez des petits ruminants marocains. Nous avons constitué deux groupes de chèvres (*Capra hircus*) et de moutons (*Ovis aries*) de part et d'autre d'un gradient de variation de températures. Pour chaque espèce nous avons identifié les régions candidates sous sélection (séquençage tout génome, 12X) et les régions différentiellement méthylées (données de MeDIP-Seq) entre les deux groupes.

L'analyse des différences génétiques (F_{ST}) entre les deux groupes, pour chaque espèce, nous a permis de localiser des régions sous sélection en relation avec des gènes impliqués dans la perception de l'environnement (5 gènes), l'immunité (4 gènes), la reproduction (8 gènes) et la production (11 gènes). L'analyse des régions différentiellement méthylées (DMR) entre les deux groupes nous a permis de trouver 2 DMRs (une chez chaque espèce) en relation, entre autre, avec la production et la qualité du lait. La comparaison entre les deux mécanismes et les deux espèces suggère des voies biologiques impactées similaires, mais aucun gène n'est trouvé en commun.

État d'avancement :

Cet article doit être relu par les co-auteurs et mis au format « Full Paper » de la revue *Animal Genetics* avant soumission.

Genetic and epigenetic factors to face environmental variations in small ruminants

Laure Denoyelle, Pierre de Villemereuil, Frédéric Boyer, Meidhi Kelifi, Clément Gaffe, Florian Alberto, Badr Benjelloun, François Pompanon

Introduction

The evolution and distribution of species is driven by the variations of their environment. Optimization of the individual's phenotype to fit the environment may occur at long term, as environmental drivers lead across generations to locally adapted populations (i.e., native populations having a higher fitness than any other introduced population, Savolainen et al., 2013), by the mean of natural selection. The access to whole genome sequences now allows to study the genomic bases of local adaptation by identifying genes and genomic regions under selection and the environmental parameters responsible for their selection. In the meantime, the context of global changes lead to an increasing interest on the ability of populations to develop short-term responses to face e.g., climate variations, as well as the interaction between short and long-term mechanisms. Indeed, populations might quickly react to an environmental change through migration towards more conducive environments, which is well documented in the global warming context (Chen et al., 2011; McDonald et al., 2012).

In the short-term, phenotypic plasticity can also increase the ability of organisms to cope with environmental changes. The ability of a genotype to produce under different conditions different phenotypes that are best fitted to the environment involves non-genetic mechanisms by which favorable variations can be acquired and transmitted. They consist in epigenetic variations, parental effects, ecological and cultural variations (Danchin et al., 2011).

Among the diversity of epigenetic mechanisms, more and more studies show the important role of chromatin modifications (including e.g., DNA methylation or histone modification) by affecting gene expression (Gibney & Nolan, 2010). Until now, most of the studies pointing out the role of molecular epigenetics mechanisms were performed in stress-controlled conditions on plant or animal models. In this case, the effect of one varying parameter on the physiological or epigenetic responses of the organisms is assessed. For example, the

glyphosate herbicide injury on *Arabidopsis thaliana* (Kim et al., 2017), or the maternal diet in mice (Cooney et al., 2002) were shown to affect the methylome.

Since the end of the 2000s, the role of these epigenetic processes in ecology and evolution is increasingly being studied (Bossdorf et al., 2008). In this context, epigenetic marks were analyzed in natural populations. For example, the search for epigenetic differentiation by contrasting natural populations of *Lilium bosniacum* living in different habitat conditions (Zoldoš et al., 2018), or in wild baboons (*Papio cynocephalus*) populations with different food resources (natural fodder in a savanna environment or human food scraps, Lea et al., 2016).

Like wild species, farm animals will have to face future climate change, and understanding the genetic and epigenetic mechanisms responding to environmental changes, especially for free-ranging populations, is a key issue for the conservation of Farm Animal Genomic REsources (Bruford et al., 2015). Our study focuses on small domestic ruminants, namely sheep (*Ovis aries*) and goats (*Capra hircus*), which are a very good model for studying local adaptation. Indeed they spread all over the world for about 10,000 years from the domestication center through human migration and commercial trades (Zeder, 2008). Until recent centuries, they were traditionally managed in a sustainable way leading to populations well adapted to the large variety of environments worldwide (Taberlet et al., 2008). About 200 years ago, the situation started to change dramatically, with the rise of the concept of breed. All animals from the same breed began to be selected for the same phenotypic characteristics, and reproduction among breeds was seriously reduced. This corresponded to a strong fragmentation of the initial populations. A few decades ago, the selection pressures were increased again in order to further improve productivity, without enough emphasis on the preservation of the overall genetic diversity. The efficiency of modern selection methods successfully increased the production, but with a dramatic loss of genetic variability. Many industrial breeds now suffer from inbreeding, with effective population sizes falling below 50. With the development of these industrial breeds came economic pressure on farmers to abandon their traditional breeds, and many of these have recently become extinct as a result. This means that genetic resources in cattle, sheep, and goats are highly endangered, particularly in developed countries. It is therefore important to take measures that promote a sustainable management of these genetic resources; first, by in situ preservation of endangered breeds; second, by using selection programs to restore the genetic diversity of industrial breeds; and finally, by protecting the wild relatives that might provide useful genetic

resources (Taberlet et al., 2008; Yang et al., 2016). Nowadays, the effect of the environment on farm animals strongly depends on breeding conditions. Intensively farmed animals are preserved from environmental variations because they are generally kept indoors with access to quality supply, in contrast to extensive livestock which are raised outdoors with little additional water and feed. Until now, epigenetic analyses on livestock have been performed in stress-controlled conditions in relation with agronomic traits such as reproduction (Lan et al., 2013), production (Peñagaricano et al., 2014; Wang et al., 2015), and disease sensibility (Doherty et al., 2014). To our knowledge, only the study of Sevane et al. (2018) searched DNA methylation in relation with climate variation by comparing cattle adapted to tropical and temperate climates.

Our study focused on sheep and goats sampled in Morocco where farming is mainly carried out in traditional ways with a strong impact of environmental conditions upon animals. Within each species, our goal was to identify candidate genomic regions involved in both genetic (selection on Single Nucleotide Polymorphisms) and epigenetic (differential DNA methylation) mechanisms related to environmental variations, by whole genome comparisons of groups of individuals from locations with contrasted temperature annual range.

Material & Methods

Sampling

Morocco has very contrasted climates including hot/cold desert and semi-arid, warm-summer mediterranean, dry-summer subarctic (Born et al., 2008). In this context, we sampled sheep (*Ovis aries*) and goats (*Capra hircus*) from locations with contrasted annual temperature variations (Figure 44, bioclimatic variable BIO7 from the WorldClim dataset, www.worldclim.org/bioclim, Hijmans et al., 2005). This variable is the difference between the maximum temperature of warmest month and the minimum temperature of coldest month. It is correlated with several temperature variables, sunshine and elevation (Tableau S12). Twenty-two unrelated goats (11 and 11 from low and high annual temperature ranges, respectively) and 25 unrelated sheep (12 and 13 from low and high annual temperature ranges, respectively) were selected from the samples collected between January 2008 and March 2012 in the frame of the Nextgen European project (Grant Agreement no. 244356), in accordance with ethical regulations of the European Union Directive 86/609/EEC, and for

which Whole Genome Sequences were already available (<http://projects.ensembl.org/nextgen/>, Benjelloun, 2015).

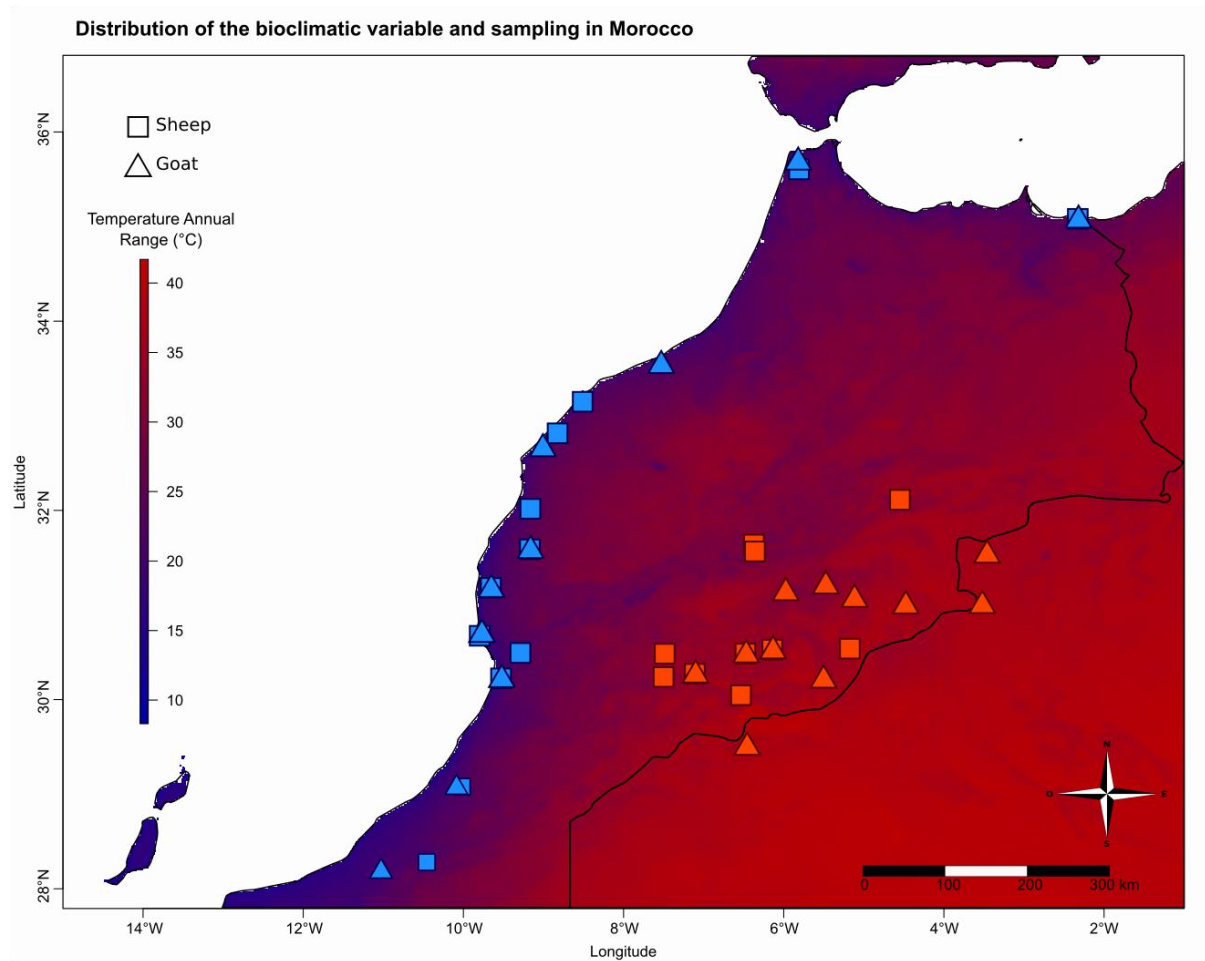


Figure 44: distribution of sheep and goats sampling at each end of the temperature gradient in Morocco. The color gradient correspond to the variation of annual temperature (BIO7).

Whole Genome Sequences

Dataset

Whole Genome Sequences (WGS) at 12-14X coverage corresponding to the studied individuals were retrieved from the European Nucleotide Archive (<https://www.ebi.ac.uk/ena>, accession code PRJEB7436). The sequences, produced following the protocol in Alberto et al. (2018), were aligned on the reference genomes CHIR v1.0 (GenBank assembly GCA_000317765.147, Dong et al., 2013) for goats and OAR v3.1, (GenBank assembly GCA_000298735.1, Jiang et al., 2014) for sheep. We extracted the states of SNPs from the variants files generated by the NextGen project.

- Detection of selection signatures

We kept SNPS with a missing data rate superior to 30% and a minor allele frequency of 1%. Then we calculated the F_{st} value (Weir & Cockerham, 1984) between the two groups for each SNP using *vcftools* version 0.1.16 (Danecek et al., 2011). We have then implemented two approaches: a pathway approach aiming at identifying the GO-terms enriched in candidate genes under selection, and a gene approach aiming at identifying the candidate genes under selection. Candidate genomic regions were defined by merging SNPs with the highest F_{st} values (i.e., top 0.025% and top 0.001% for the pathway and gene approaches, respectively) closer than 5,000 bases and with 2,500 bases added after the last outlier SNP.

Two enrichment analyses were performed using the statistical over-representation test in PANTHER (version 15.0, Mi et al., 2019).

(i) Pathway approach: we compared the list of gene extracted with the 0.025% threshold, with those present in the *Bos taurus* reference. Each biological process with a significant Fisher's exact test and an False Discovery Rate (FDR) lower than 0.05 was annotated with the GO Ontology database released on 2020-08-10 (Carbon & Mungall, 2018). We kept the more basal significant biological pathway and searched the GO terms with the highest hierarchical level (i.e. child term of “biological process”) associated.

(ii) Gene approach: the same annotation database and protocol were used to extract the GO term corresponding to the top genes (0.001% threshold). In addition, a specific gene bibliography was also done to determine what could be their effect on livestock phenotypes. These were grouped into large phenotype categories like “environmental perception”, “immunity”, “production” (genes involved in milk or meat production), “reproduction” and “other”.

DNA Methylated Regions

- DNA extraction, library preparation and sequencing

DNA was extracted from ear biopsies collected by the Nextgen consortium using the isolation of genomic DNA from tissues protocol from QIAamp® DNA micro kit (QIAGEN, Germantown, MD, USA).

Methylated DNA immunoprecipitation sequencing (MeDIP-Seq) was performed at the GeT-PlaGe core facility, INRAE Toulouse (<http://www.get.genotoul.fr>). WGBS libraries were prepared according to Bioo Scientific's protocol using the Bioo Scientific NEXTflex™ Methyl-seq Library Prep Kit for Illumina Sequencing. Briefly, DNA was fragmented by sonication on a

covaris M220, size selection was performed using AMPure XP beads and adapters were ligated to be sequenced. 3µg of each library was diluted in 450µl of TE buffer, denatured in boiling water for 10 minutes and immediately cooled in ice for 10 minutes. Then 50µL of 10X concentrated IP buffer was added to the mixture as well as 1µg of anti-5-methylcytosine monoclonal antibody (clone 33D3, Diagenode®). After overnight incubation at 4°C with agitation, the DNA-antibody complexes were purified using 40µL of "Dynabeads-ProteinG" previously washed in PBS. The DNA-antibody complexes were washed twice in 700ul of IP buffer and then the DNA was purified using the iPure kit (Diagenode®). Control or input DNA samples were not coupled to the antibodies and were therefore not enriched in epigenetic markers. Then, 12 cycles of PCR were performed. Library quality was assessed using an Advanced Analytical Fragment Analyzer and libraries were quantified by QPCR using the Kapa Library Quantification Kit. MeDIP-Seq was performed on an Illumina® HiSeq 2500 with the Illumina Reagent Kits.

- Dataset preparation

Illumina paired-end reads (mean size of 125 bp) from all animals were aligned on the same reference genomes as WGS data by using the default parameters of BWA-MEM version 0.7.12 (Li & Durbin, 2009). The function *multicov* from bedtools version 2.29.2 (Quinlan & Hall, 2010) was used to report the MeDIP-Seq coverage of alignment files inside windows of 100 non-overlapping base pairs along the genome. The same function was used to extract the WGS coverage. In these windows we also extracted the number of CpG from the reference genomes by using the *MEDIPS* package (Lienhard et al., 2014) on R software version 3.5.3 (R Core Team, 2019).

We compared genome wide coverage profiles of MeDIP-Seq by using the *MEDIPS.correlation* function from the *MEDIPS* package with the *uniq*, *extent*, and *shift* parameters equal to 0 and the *window size* parameter equal to 100, to keep all MeDIP-Seq fragments in each 100 base window on the genome. We removed one goat and three sheep from the dataset for which the pairwise Pearson correlation coefficients were less than 0.6 with all other individuals of the same species. Thus, our final dataset contained 21 goats (10 and 11 at each end of the BIO7 gradient) and 22 sheep (12 and 10 at each end) (Figure 44 and Tableau S11 en Annexe). The numbers of WGS and MeDIP-Seq reads in each window were normalized between individuals with default parameters of *calcNormFactors* function (*edgeR* package, Robinson et al., 2010). Then, normalized MeDIP-Seq fragment counts were divided by normalized WGS

fragment counts in order to get rid of the effect of sequencing coverage variations (e.g., due to repeated regions). Count values have been rounded to two decimals. Four outlying windows (2 for each species) with 10 times more MeDIP-Seq fragments than the others were removed as this may be due to PCR or alignment errors.

- Whole epigenome correlations

For each species, correlations between the number of CpG and that of MeDIP-Seq fragments were tested using a Pearson's correlation test. We used the same test to test the correlation between species for each of these two variables. For that we used the correspondence between *Capra* and *Ovis* homologous regions resulting from the cross-alignment between the two reference genomes CHIR_1.0 and OAR v3.1 performed by Alberto et al. (2018). Results were represented as density plots performed with the *hexbin* package (Carr et al., 2019).

- Detection of Differentially Methylated Regions

We used the *edgeR* package to detect DMRs. Starting from non-standardized MeDIP-Seq fragment counts, we kept the 100 bp windows for which at least 10% of the samples had at least 1 count-per-million (1 fragment per million windows, this accommodates differences in library sizes between sample) and we normalized the fragment counts with the *calcNormFactors* function with default parameters. We estimated common dispersion with the *estimateGLMRobustDisp* function which mitigates the influence of outliers (Zhou et al., 2014) and we tested the occurrence of differential methylation between groups with the default parameters of the *exactTest* function. The p-values obtained were converted into q-values with the *Bioconductor* package *qvalue* (Storey et al., 2015) and the ones showing q-values inferior to 0.01 were selected for the rest of the analysis. Finally, we merged significant windows distant than less than 5,000 bases into regions.

In order to delineate the peaks of methylation occurring for some of the individuals within these regions, we used the *changepoint* package (Killick & Eckley, 2014) which detects breakpoints in time series. We looked for all possible breakpoints within a region using the *cpt.meanvar* function that identify change in mean and variance of the sum of the animal counts for each group. We kept the breakage pairs that contained at least one window detected as significant with *edgeR*, thus, the delimited regions were merged with the *bedr* package (Waggott et al., 2017) when they overlapped. Then, we used a Wilcoxon test to detect

peaks with different methylation levels between the two groups. Among those peaks, we discarded those where the number of MeDIP-Seq fragments was correlated to the DNA coverage in the WGS data for which high MeDIP-Seq counts would result from the high DNA coverage (e.g., due to repeated sequences). The regions containing the remaining peaks were mapped on the reference genomes and included or overlapping genes were identified. We used the Functional classification viewed in gene list analysis in PANTHER to extract the Biological Pathways annotated in the Gene Ontology Database associated to these genes.

Results

Detection of selection signatures



Figure 45: number of occurrences of biological pathways for each analysis and species: epigenetics and top genes on the left and enrichment analysis on the right. One color corresponds to one analysis for one species.

- Pathway approach

The 5.000 SNPs for goat and 6.000 SNPs for sheep which harbored the highest 0.025% F_{st} values (Figures S43 and S44 en Annexe) corresponded to 442 and 489 genes, respectively, putatively under selection. Twenty-six and 116 GO terms were significantly enriched/depleted in candidate genes in goat and sheep, respectively (Figures S45 et S46, and Tableaux S13 and S14 en Annexe). They corresponded to 10 clusters (all parent and child terms representing the same biological pathway, 9 with gene number overrepresented and 1 underrepresented) in goat and 25 clusters (23 overrepresented and 2 underrepresented) in sheep.

The GO term associated with these 10 clusters in goat were “biological adhesion”, “biological regulation”, “cellular process”, “developmental process”, “localization”, “multicellular organismal process”, “response to stimulus”, and “signaling”. For sheep, the 25 clusters were associated with the following GO terms: “biological adhesion”, “biological regulation”, “cellular process”, “developmental process”, “localization”, “locomotion”, “metabolic process”, “multicellular organismal process”, “response to stimulus”, “signaling”, and unknown.

Among the 0.025% top genes (i.e., 931 genes selected for the enrichment analyses), 41 were homologous between sheep and goats. When we did the same enrichment analysis that we described above we obtained 21 GO terms corresponding to 7 overrepresented clusters. Four of them had the “nervous system development” in their parental terms; and the last three were associated with cellular processes, one of which was related to synapses organization.

- Gene approach

The top 0.001% F_{st} values corresponded to 194 SNPs for goat and 146 SNPs for sheep (Figures S43 and S44 en Annexe) distributed in 56 and 42 genomic regions, respectively (Tableaux S15 and S16 en Annexe). In goats, 34 of them were intergenic and the others were related to 22 genes. For sheep, we obtained 24 intergenic regions and 18 genes, which differed between the two environmental groups. No genes were common between the two species.

The GO terms linked with these 22 genes in goat were “behavior”, “biological adhesion”, “biological regulation”, “cellular process”, “developmental process”, “immune system process”, “localization”, “locomotion”, “metabolic process”, “multi-organism process”, “multicellular organismal process”, “reproduction”, “response to stimulus”, “signaling”, and unknown. For sheep, the 18 genes were associated with: “behavior”, “biological regulation”, “biomineralization”, “cellular process”, “developmental process”, “growth”, “immune system

process”, “localization”, “locomotion”, “metabolic process”, “multicellular organismal process”, “response to stimulus”, “signaling”, and unknown.

In livestock, several of these genes are known to be involved in phenotypic changes in relation with the perception of environment (3 genes in goat, 2 in sheep), immunity (1 in goat, 3 in sheep), production (9 in goat, 2 in sheep), reproduction (5 in goat and 3 in sheep) (Table XIV). The 'Other' category brings together the genes whose effect on the phenotype was not documented in livestock species. We also detected genes that were uncharacterized.

Table XIV: phenotype categories in which the top genes are involved.

Species	Phenotype categories	Genes
Goat	Environmental perception	TMTC2, EDIL3, SASH1
	Immunity	PLCG2
	Production	CP, SLC9A9, PPFIA2, TMTC2, GALNTL6, CTSB, SASH1, PLCG2, MYADM
	Reproduction	USH1C, ADGRB3, PLCG2, MYADM, SERPINB7
	Other	TRABD2B, MVB12B, KLF12, TNFSF9, CALHM3, SORCS1, RAB30
	Unknown	LOC106503718, SERPINB3-like
Sheep	Environmental perception	NOX3, KSR2
	Immunity	SAMD12, SEMA5A, SEZ6L
	Production	BMPER, PTPRE
	Reproduction	GCSAML, KSR2, BMPER
	Other	AGO3, IQCJ, KCNG3, ANO6, CASP8AP2, CNKSR3, ENGASE, KCNA4
	Unknown	LOC101107868, OR2W1-like

Comparison of Methylated Regions

Among more than 22 million windows, 1% and 2% didn't carry methylated fragment in goat and sheep, respectively. The majority of windows (96% for each species) had between 0 or 1 methylated fragment. Finally 3% and 2% of the windows, in goat and sheep, contained more than 1 methylated fragment. In both species, 47% of the windows had no CpG, 50% had between 1 and 4, lastly 3% had more than 4 in the two species.

Genomic windows of 100 bp that were homologous between *Capra* and *Ovis* showed a correlation between species for the number of methylated fragment ($R=0.81$, $p\text{-value} < 2.10^{-16}$, Figure 46 C.) and the number of CpGs ($R=0.67$, $p\text{-value} < 2.10^{-16}$, Figure 46 D.) A large number of windows had few CpG and methylated fragments (25% of the data set carried less than 3 CpG and 1 methylated fragment). Within species, we found a correlation between the number of CpGs and that of methylated fragments, $R=0.49$ ($p\text{-value} < 2.10^{-16}$, Figure 46 A.) for goats and $R=0.44$ ($p\text{-value} < 2.10^{-16}$, Figure 46 B.) for sheep.

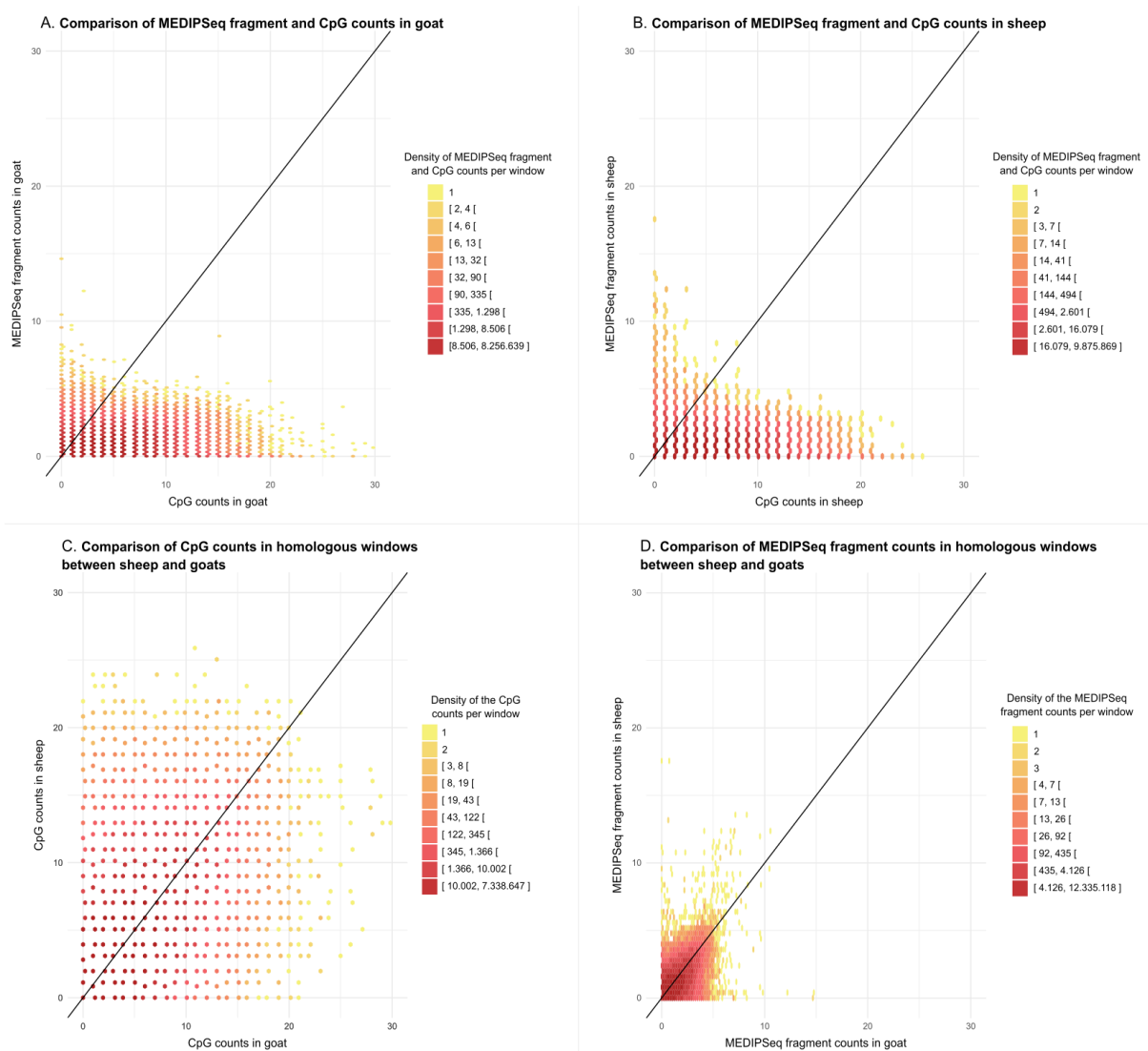


Figure 46: details and comparison of epigenomes. Relation between the number of MEDIP-Seq fragments and the number of CpG for A. goat, and B. sheep. Relationship between the number C. CpG, or D. MEDIP-Seq fragments per 100 bp window between the two species along the genome. The color represent the number of points.

For the detection of DMRs, 20.256 windows in goat and 16.281 windows in sheep were kept out of more than 24 million. The edgeR analysis revealed 20 and 7 windows with significant differences in methylation levels between the two groups (i.e., low vs high annual

temperature variations) in goat and sheep, respectively. Merging candidate windows closest than 5,000 bp, we obtained 8 and 4 regions in which we detected 8 and 5 methylation peaks for goat and sheep, respectively. From these, 4 and 2 peaks respectively showed significant differential methylation MeDIP-Seq coverage between groups. For 4 peaks, the differential MeDIP-Seq coverage was concomitant to a differential WGS coverage, indicating that the differentiation would probably result from repetition polymorphism. Then only one region per species was kept as showing differential methylation. They were associated to AGPTA4 in goat and SLIT3 in sheep (Figure 47).

The GO terms associated with these genes were “cellular process”, and “metabolic process” for AGPAT4; and “biological regulation”, “cellular process”, “developmental process”, “growth”, “locomotion”, “metabolic process”, “multicellular organismal process”, “reproduction”, “reproductive process”, “response to stimulus”, “rhythmic process”, and “signaling” for SLIT3.

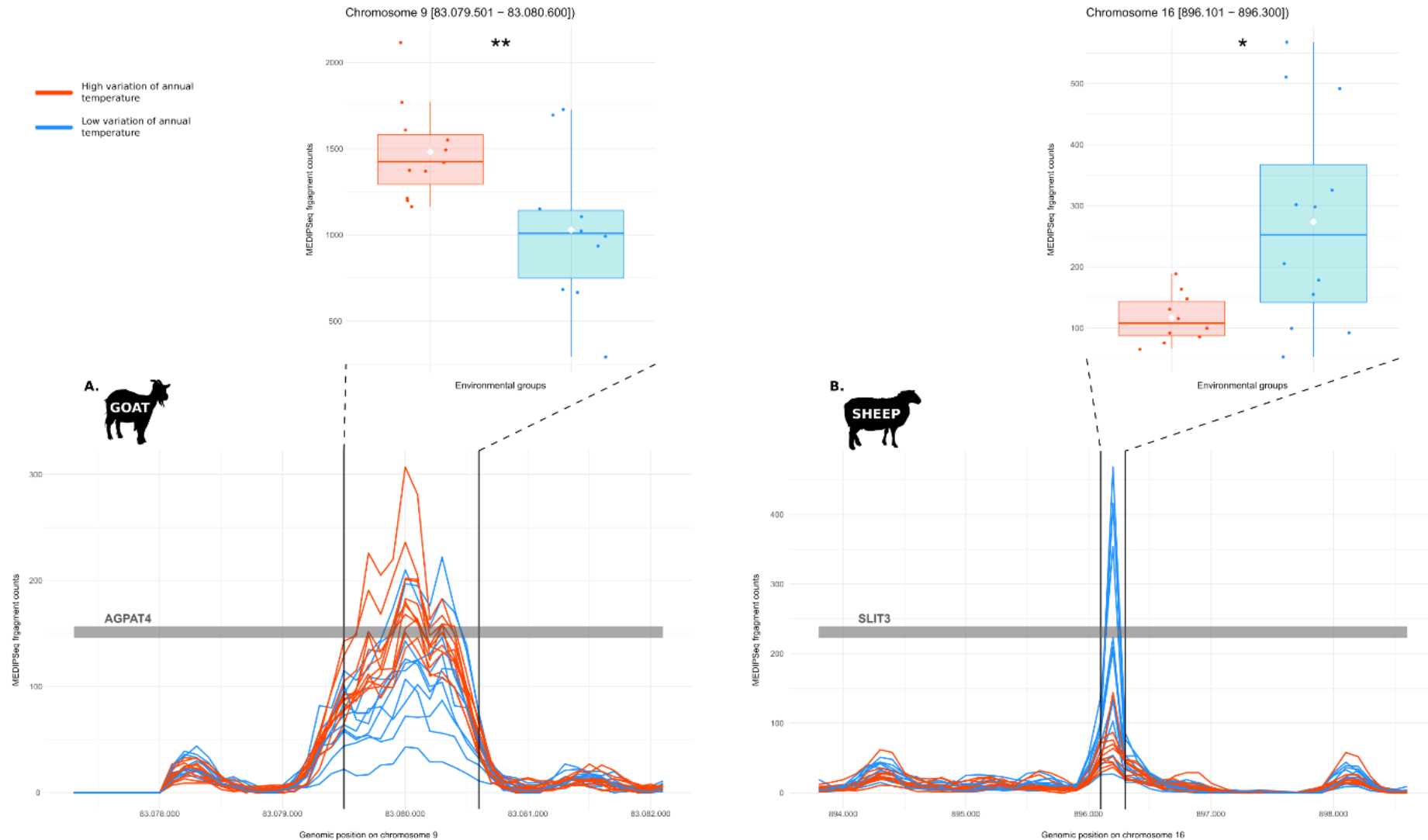


Figure 47: DMRs between genomes from locations with low (blue) and high (red) temperature annual range for A. goat or B. sheep. Each line represents the variations of the level of DNA methylation along a gene for one individual; the grey bar represents part of the gene; the two black lines delineate the peak on which the differential level of methylation was tested. Boxplots represent the level of methylation (MeDIP-Seq coverage) for each group compared with a Wilcoxon test.

Discussion

To our knowledge, our study is the first to investigate simultaneously the genetic and epigenetic variations involved in the adjustment of phenotypes to field conditions in small ruminants. Until now, few publications jointly addressed both mechanisms. Some studies highlighted significant association between local genetic variation and the presence of DMRs like Eichten et al. (2013) that found 51% of the DMRs between maize populations associated with local SNPs. Foust et al. (2016) and Herrera & Bazaga (2010) find a correlation between AFLP and MS-AFLP marks in two salt marsh perennials (*Spartina alterniflora* and *Borrchia frutescens*) and violet populations (*Viola cazorlensis*) respectively.

The sampling was carried out in order to avoid a breed effect as much as possible. However, given the specific distribution of some breeds, several Ghazalia goats and D'man sheep were part of the groups submitted to high annual temperature ranges. However, they represented less than half of the individuals (3/11 and 4/10 in goats and sheep, respectively) and did not behave differently from other breeds. Thus, the mechanisms inferred would be shared by the variety of Moroccan breeds represented in our sampling. The sheep and goats sampled originated from two contrasted climates with regards to the temperature regime i.e., deserts with a high temperature annual range (c.a. 39.5°C) towards coastal regions with low temperature annual range (c.a. 20°C). Our study points out several genes which polymorphism or methylation level is contrasted between the two groups and may contribute to the adaptation and acclimatization to these specific environments. The candidate genes are primarily involved in environmental perception, immunity, production and reproduction (see Table XIV and Tableaux S15 and S16 en Annexe), and their variations could be driven by different environmental factors.

First, a direct effect of temperature might operate through thermoregulation. Moving away from a thermo-neutral range where least effort is required to regulate body temperature, the energetic and physiological costs of thermoregulation increase until death (Silanikove, 2000). To reduce their body temperature, animals decrease metabolic heat production and enhance heat dissipation. For that they take refuge in the shade to avoid solar radiations (Al-Tamimi, 2007), increase their respiration or sweating rates and reduce food intake (Marai et al., 2007; Sejian et al., 2010). These adjustments related result in a loss of energy for other biological processes such as productivity (Tao et al., 2011; Dado-Senn et al., 2019) or reproduction (Meyerhoeffer et al., 1985; Krishnan et al., 2017; Monteiro et al., 2016),

which may be affected by the temperature rimes at least by this way. In relation with temperature annual range, we found genes involved in production, including milk production for CP, MYAMD, PTPRE, SASH1 and SLC9A9 (Hussein & Staufenbiel, 2012; Dong et al., 2015; Shin et al., 2017; Xiang et al., 2017; Laodim et al., 2017), mammary gland development for PPFIA2 (Mota et al., 2020) and morphology for TMTC2 (Abo-Ismaïl et al., 2017; Fang & Pausch, 2019). In addition, CTSB, GALNTL6, and KSR2 genes were related to meat production (Russo et al., 2002; Doran et al., 2014; Puig-Oliveras et al., 2014; Lukić et al., 2020), and BMPER and KSR2 with body size (Puig-Oliveras et al., 2014; Zhao et al., 2015; Lukić et al., 2020). Genes involved in reproduction were also impacted. ADGRB3, GCSAML, and KSR2 affect fertility (Pimentel et al., 2011; Wang et al., 2015; Mohammadi et al., 2020), while PLCG2 and SERPINB7 affect gestation (Vallée et al., 2003; Ponsuksili et al., 2012). Moreover offspring survival can also be impacted through litter size via BMPER and USH1C (Kwon et al., 2016; Bolormaa et al., 2017; Mota et al., 2020) and offspring weight with MYADM (Gonzalez et al., 2020). Also, we found two genes related to environmental perception. NOX3 is involved in the response to stresses in general (Lambeth, 2004) and has been related with heat stress in chicken (Aggrey et al., 2018), and KSR2 plays a role in the energy balance by adapting feeding behaviour and thermogenesis in mice (Guo et al., 2017). Second, the temperature regime may have indirect impacts on phenotypes by driving other environmental factors. Especially, it determines plant resources (i.e., food quality and quantity) and is related to water availability, and thus strongly impacts the diet of free ranging livestock. The lack of resources have a negative impact on, e.g., reproduction (Martin et al., 2004; Robinson et al., 2006), lactation (Razzaghi et al., 2016). In contrast, the supplementation have a positive impact on yield and composition of milk in sheep (Caroprese et al., 2011). As a candidate gene, we found PLCG2 which is downregulated in the cattle liver according to a high-residual feed intake (difference between the actual and predicted feed intake, Salleh et al., 2017). Likewise, genes involved in immune functions may be selected by multi-factorial drivers related to nutrition or the pathogenic context in relation with climate (Bett et al., 2017). This could be the case for SMAD12 playing a role in pancreatitis in human (Giri et al., 2016), SEZ6L associated with foot-and-mouth disease (Lee et al., 2015) or SEMA5A with mastitis (Sugimoto et al., 2006) in cattle. Third, environmental parameters covarying with the annual temperature range may be prominent in driving the variation of the genes which polymorphism was contrasted between the two groups. TMTC2 and EDIL3 are involved in adaptation to high altitude in sheep and cattle (Ji Yang et al., 2016; Pierce et al., 2020), this later variable being

correlated with the annual temperature variation (BIO7) in Morocco. In addition, BIO7 variable is correlated with sunshine which impact directly skin pigmentation, what would explain the detection of SASH1, which regulates melanocytes migration in relation with the production and transfer of melanin to protect skin from UV radiations (Zhou et al., 2013).

Besides the candidate genes putatively involved in local adaptation through allelic selection, we identified two strong candidates in relation with the epigenetic response of the organisms, namely AGPAT4 (1-acylglycerol-3-phosphate O-acyltransferase 4) in goats, and SLIT3 (Slit Guidance Ligand 3) in sheep. These two genes were differentially methylated according to the provenance of the individual, in a global epigenomic context where the level of methylation of homologous genomic regions was correlated between sheep and goats, and more than 98% of the 100 bp genomic windows characterized were methylated at a level correlated with their CpG content. This is consistent with what is known about mammalian epigenomes where the CG methylation generally covers the entire nuclear genome, with the exception of CpG rich regions near the promoters of active genes (Suzuki & Bird, 2008; Feng et al., 2010).

AGPAT4 is part of the 1-acyl-sn-glycerol 3-phosphate acyltransferases family which is involved in triglyceride synthesis (Lu et al., 2005; Takeuchi & Reue, 2009), with a role in the milk composition of ruminants (Bionaz & Looor, 2008; He et al., 2011; Jie Yang et al., 2016). It has an organ-related level of expression in human (Lu et al., 2005; Prasad et al., 2011) and plays a role in muscle development and meat quality in cattle (Fang et al., 2017). In addition, the variation of AGPAT genes expression in the mammary gland has been related to milk fat depression in dairy cows (Peterson et al., 2003).

On its side, SLIT3 takes place into the SLIT-ROBO interactions which are involved in the development of the nervous system (Wang et al., 1999; Andrews et al., 2007), cardiac development (Liu et al., 2003; Mommersteeg et al., 2015) and reproduction, especially menstrual cycle (Dickinson & Duncan, 2010) and fertility (Amundson et al., 2015). SLIT3 was also shown to be differentially methylated and upregulated in goat mammary glands during lactation compared to the dry period (Zhang et al., 2017). Such DMRs related to milk quality and lactation might result directly from heat stress. Indeed, it was shown that heifers submitted to a thermic shock during fetal development had smaller mammary gland alveoli, resulting in lower milk production (Skibiél et al., 2018; Dahl et al., 2019). The DMRs observed could also result from indirect effects of the climate through a variation of the diet resulting from different plant resources. Peterson et al. (2003) showed that the high concentrate/low forage diet decrease the mRNA abundance of AGPAT genes in mammary

glands in dairy cows, involving a milk fat depression. The regulation of the SLIT-ROBO pathway is also affected by the diet (Amundson et al., 2015). Indeed, the diet can directly influence the levels of DNA methylation (Anderson et al., 2012). Such variations of DNA methylation may result from direct exposure of the individuals to environmental variations (at different stages including the fetus) or from parental effects. For example, a heat-stress-induced perturbation of mammary glands development in the fetus, related to differential methylation and differential gene expression (Skibiél et al., 2018), results in effects persisting over the lactations and in the next generation too (Ouellet et al., 2020). DNA methylation and gene expression of lambs is also affected by maternal nutrition, especially in muscle and adipose tissues (Lan et al., 2013; Peñagaricano et al., 2014; Namous et al., 2018).

We found that candidate genes for local adaptation in response to different temperature regimes were related to a large variety of biological pathways which were mainly involved in functions such as immunity, perception of the environment and reproduction, or related to productivity traits. While these functions were affected in both sheep and goats, the best candidate genes (i.e., top 0.001%) identified were different, confirming that alternative genes might be selected under the same environmental constraints for close species (e.g., (Benjelloun et al., *Submitted to MBE*; Raeymaekers et al., 2017) or even populations of the same species (Manceau et al., 2010; Benjelloun et al., 2015). However, out of the 0.025% top genes (i.e., 931 genes selected for the enrichment analyses), 41 were homologous between sheep and goats and mainly related to the nervous system development.

Interestingly, the present study also investigated epigenetic variations in the individuals displaying candidate genes for local adaptation. Our deliberately stringent approach to avoid false positive, which identified only one candidate DMR per species, prevents a global view but already points out similarities between traits impacted by genetic and epigenetic mechanisms. In sheep, genes impacted by selection such as BMPER, GCSAML, KSR2, PTPRE (Tableau S16 en Annexe) affect embryo quality, litter size and lactation, which are also influenced by SLIT3 that was differentially methylated. Likewise, in goats, AGPAT4 that was differentially methylated has an effect on lipid metabolism, meat quality and the presence of lipids in milk, these traits being influenced by several candidate genes such as CTSB, MYADM, SERPINB7, USH1C, GALNTL6 and CALHM3 (Tableau S15 en Annexe).

Conclusion

This study showed candidate differentially methylated genes related to a differential regime of temperatures underwent by sheep and goats from extensive breeding. These genes have not been selected, but several other sharing the same target functions are putatively driven by selection and thus involved in local adaptation. Among the target functions, lactation and milk composition emerged especially from our study. These adjustments of phenotypes to the two contrasted environments may result from drivers related to temperature either directly or indirectly through e.g., the variation of plant resources conditioning the diet. Deciphering the relative impact of climate and e.g., differential diets on the epigenetically-induced variations of phenotypes in such a context will be a major issue for developing future breeding strategies in the context of global change.

DISCUSSION GENERALE

Comme présentés dans l'introduction, deux mécanismes, l'un génétique et l'autre épigénétique, peuvent être mis en place pour permettre l'adéquation d'un phénotype à l'environnement. Mon travail de thèse a permis de mettre en évidence leur rôle à travers la recherche de signatures de sélection et de régions différentiellement méthylées chez les petits ruminants.

Dans cette partie nous allons replacer les résultats que nous avons obtenus dans un contexte plus global permettant de répondre aux problématiques que nous avons présentées en introduction, qui portaient sur la détection de ces deux mécanismes et de leurs effets sur l'adaptation ou l'acclimatation des animaux à leur environnement.

I. Marques génétiques : témoins de l'histoire des espèces

Les différents SNPs portés par chaque race caprine française nous ont permis d'explorer l'histoire de chacune d'elles (analyse démographique et de consanguinité) et les relations qu'elles pouvaient avoir entre elles (distances génétiques et recherche de croisements). Ces résultats sont pour la plupart concordants avec les données historiques collectées et les conditions d'élevage connues pour chacune d'elles.

Nous avons aussi pu mettre en évidence des signatures de sélection en rapport avec des caractères d'intérêt agronomique chez les races caprines françaises et en relation avec l'environnement chez des petits ruminants marocains.

De manière générale, les études s'intéressant aux SNPs sous sélection chez les petits ruminants détectent des signatures en relation avec la domestication et des caractères d'intérêt agronomique pour tous, ainsi qu'avec l'environnement pour les races élevées de manière traditionnelle.

Début de la domestication

Le système nerveux

Les signatures de sélection associées à la domestication des animaux sont souvent en relation avec la docilité de ceux-ci. L'étude d'Alberto et al. (2018) a comparé des chèvres et des

moutons marocains et iraniens avec leurs parents sauvages *Capra aegagrus* et *Ovis orientalis* respectivement. Des marques de sélection associées à des gènes impliqués dans le système nerveux ont été mises en évidence avec les gènes MTMR7 liés à la maladie de Creutzfeldt-Jakob, caractérisée par une dégénérescence du système nerveux central (Sanchez-Juan et al., 2012) et KITLG qui parmi ses nombreux effets, joue un rôle dans la différenciation des cellules neurales (Pollard et al., 2008). De plus, des gènes en relation avec le comportement ont été mis en évidence, comme NBEA qui pourrait être impliqué dans la docilité chez les vaches (Ramey et al., 2013).

Il est très peu probable que nous ayons mis en évidence une sélection liée à la docilité (trois gènes en relation avec le système nerveux chez les Fossés, Poitevine et Saanen) car tous les animaux que nous contrastons ont une histoire commune pendant et post-domestication comme nous avons pu le voir sur notre analyse d'inférence démographique. L'impact de la sélection sur le comportement les a donc affectés de la même manière, donc il n'est pas possible de trouver des régions en relation avec ce début de la domestication en contrastant les races de notre étude. Une étude intra-race permettrait de détecter ces signatures de sélection commune à toutes les races.

Sélection agronomique

Les caractères ayant un intérêt agronomique ont aussi été sélectionnés afin d'améliorer la rentabilité des élevages. Les voies principalement sélectionnées sont l'immunité, la reproduction, les caractères de production, et certaines spécificités morphologiques comme la couleur de la robe, la présence ou non de cornes, la morphologie de la queue ou des oreilles.

L'immunité

L'analyse des régions différenciées entre animaux sauvages et domestiques dans les études d'Alberto et al. (2018) et Zheng et al. (2020) ont permis de détecter des gènes sous sélection en relation avec l'immunité chez les petits ruminants. Les gènes sous sélection peuvent être associés au système immunitaire inné, comme les gènes MUC6 associé à la résistance aux parasites gastro-intestinaux (Rinaldi et al., 2011) et SLAMF1 jouant un rôle dans la migration des macrophages (Wang et al., 2015) ; ou au système immunitaire acquis (exemple des gènes SLAP et DUSP22 qui régulent la signalisation des récepteurs des lymphocytes T (Sosinowski et al., 2000; J.-P. Li et al., 2014). Ce caractère a pu être sélectionné de façon intentionnelle ou

non au début de la domestication car seuls les animaux en bonne santé ont pu être élevés. Par la suite une sélection dirigée a été mise en place, c'est par exemple le cas avec la résistance aux mammites dans les races laitières françaises avec une sélection sur les taux de cellules somatiques dans le lait (Rupp et al., 2009) ou à la tremblante chez les races ovines françaises (Ministre des Finances et des Comptes publics & Ministre de l'Agriculture, de l'agroalimentaire et de la forêt, 2014). Nous retrouvons d'ailleurs trois gènes sous sélection en rapport avec la résistance aux mammites chez les races sélectionnées pour leur lait : Poitevine, Alpine, Saanen, mais également chez les races Lorraine, Fossés et Pyrénées. Nous trouvons 8 autres gènes associés au système immunitaire et à des maladies.

La reproduction

Nous avons détecté 15 gènes sous sélection associés à des caractères de reproduction répartis dans toutes les races exceptées les Rove et Provençales. De manière générale, les gènes associés à la reproduction sont en lien avec l'ensemble du processus, de la formation des gamètes jusqu'à la survie des descendants. Comme par exemple avec des signatures de sélection trouvées dans notre étude sur le gène SPAG4 qui est impliqué dans la spermatogenèse chez les bovins et les ovins (Bai et al., 2017; Frischknecht et al., 2017; Khayatzaheh et al., 2019). Chez les femelles, des gènes en relation avec le cycle ovarien et l'œstrus peuvent être détectés sous sélection comme les gènes RHOJ, NOL4L et TLN2 chez les vaches et les cochons (Roberts & McLean, 2011; Dai et al., 2014; Li et al., 2016; Talebi et al., 2018). Des gènes peuvent porter des signatures de sélection pour leur rôle dans la gestation comme SERPINB7 chez la truie (Vallée et al., 2003). Pour finir, des impacts sur la taille de la portée (gènes BMPER et USH1C par exemple, Kwon et al., 2016; Bolormaa et al., 2017; Mota et al., 2020) ou le poids des descendants chez les vaches Holstein (par exemple le gène MYADM, Gonzalez et al., 2020) pouvant impacter leur survie peuvent être détectés.

La production

Les petits ruminants sont élevés pour fournir du lait, de la viande et des fibres (Bencini & Pulina, 1997; Safari et al., 2005; Skapetas & Bampidis, 2016), et nous trouvons effectivement des traces de sélection en rapport avec ces trois types de production.

- Production de lait

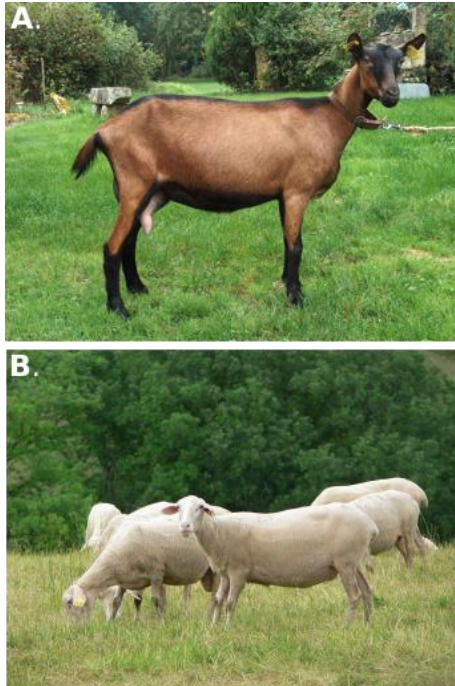


Figure 48 : exemple de races élevées pour leur lait : A. l'Alpine (© Capgènes) et B. les Lacaunes (© Deviers.fabien)

Les races caprines européennes sont principalement élevées pour leur lait (Miller & Lu, 2019). C'est le cas des races françaises que nous avons étudiées. Les gènes sous sélections peuvent être impliqués dans la composition du lait avec l'exemple du gène CSN1S1 responsable de la coagulation du lait (Manfredi & Ådnøy, 2012) et trouvé sous sélection chez d'autres races européennes dans plusieurs études (Bertolini et al., 2018; Oget et al., 2019). Nous trouvons également sur ce gène un signal de sélection spécifique de la race Alpine (Figure 48 A.) qui est sélectionnée depuis les années 1970 pour améliorer ses qualités laitières. Des caractères liés à la morphologie mammaire (exemple du gène TMTC2, Abo-Ismaïl et al., 2017; Fang & Pausch,

2019) ou à la vitesse de la traite (exemple du gène ANGPT4, Jardim et al., 2018) peuvent aussi être détectés sous sélection.

Certaines races de moutons sont élevées pour leur lait, c'est par exemple le cas de la race Lacaune (Figure 48 B.) dont le lait sert à la fabrication du fromage AOC roquefort. Chez cette race les gènes SUCNR et PPARGC1A ont été mis en évidence pour leur effet sur la production laitière et le pourcentage de matière grasse dans le lait (Yuan et al., 2019).

■ Production de fibres

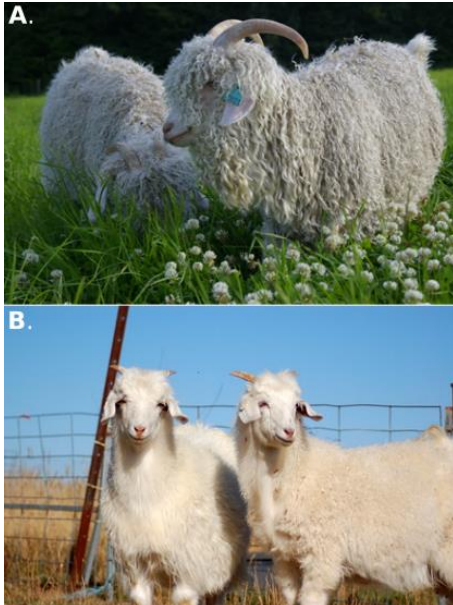


Figure 49 : exemples de races caprines élevées pour leurs poils : A. Angora (© Capgènes) et B. Cashmere (© Paul Esson)

Nous avons retiré la race Angora de notre analyse de signature de sélection et n'avons pas trouvé de signatures de sélection impliquées dans ce phénotype. Les trois races principalement élevées pour ce type de production sont les Angoras (Figure 49 A.), les Ankaras et les Cashmeres (Figure 49 B.). En se focalisant sur ces trois races, l'étude de Bertolini et al. (2018) a pu mettre en évidence deux gènes sous sélection et en relation avec le poil. En effet le gène CUX1 est, entre autre, associé aux poils ondulés et aux moustaches frisées chez la souris (Sansregret & Nepveu, 2008) et le gène PLOD3 peut jouer un rôle dans la formation et la texture des poils (Salo et al., 2008).

■ Production de viande

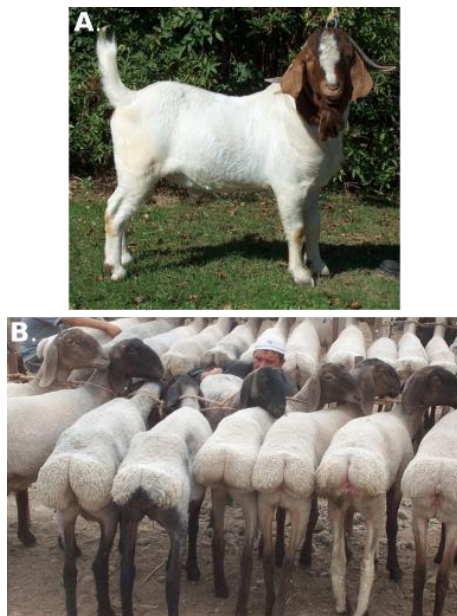


Figure 50 : exemple de petits ruminants élevés pour leur viande : A. race Boer (© Capgènes) et B. moutons portant le phénotype queue grasse (© erector)

Nous n'avons pas de race spécifiquement élevée pour la viande dans notre jeu de données car nous avons retiré la race Créole lors de l'analyse des signatures de sélection. Des traces de sélections en rapport avec la musculature avec des gènes comme SNTA1 (Keel et al., 2018), CDK5RAP1 et P2RY1 impliqués dans la croissance des animaux (Fernández et al., 2012; An et al., 2020) ont tout de même été détectés. Le projet ADAPTmap a cherché les marques de sélection spécifiques à six races élevées pour leur viande : les Boers (Figure 50 A.), Baris, Murciano-Granadina, Pothoharis, Teddi et Thari. Les principaux gènes sélectionnés chez ces races sont AMPD1 et THSB influençant la masse corporelle (Reinehr, 2011; C.-B. Wei et al., 2015).

Chez les moutons il existe un phénotype appelé queue grasse (Figure 50 B.) qui est en relation avec une meilleure qualité de la viande car la graisse est stockée préférentiellement dans la queue et non dans le reste du corps, dont les muscles (Kashan et al., 2005).

Les gènes liés à la musculature ne sont pas forcément en relation avec la production de viande. Ils peuvent également être en relation avec la carrure des races associée à leur origine (montagnarde pour l'Alpine par exemple), mais aussi potentiellement avec la production laitière. En effet, chez le mouton, il a été montré que le gène *SOCS2* joue un rôle dans la production du lait et la taille et le poids des animaux (Rupp et al., 2015).

Caractères morphologiques spécifiques aux races

D'autres caractères ont pu être sélectionnés pour des raisons esthétiques et sont maintenant décrits dans les standards des races. C'est par exemple le cas de la couleur du pelage, de la présence ou non de cornes ou de pampilles, et de la morphologie des oreilles (Figure 51).

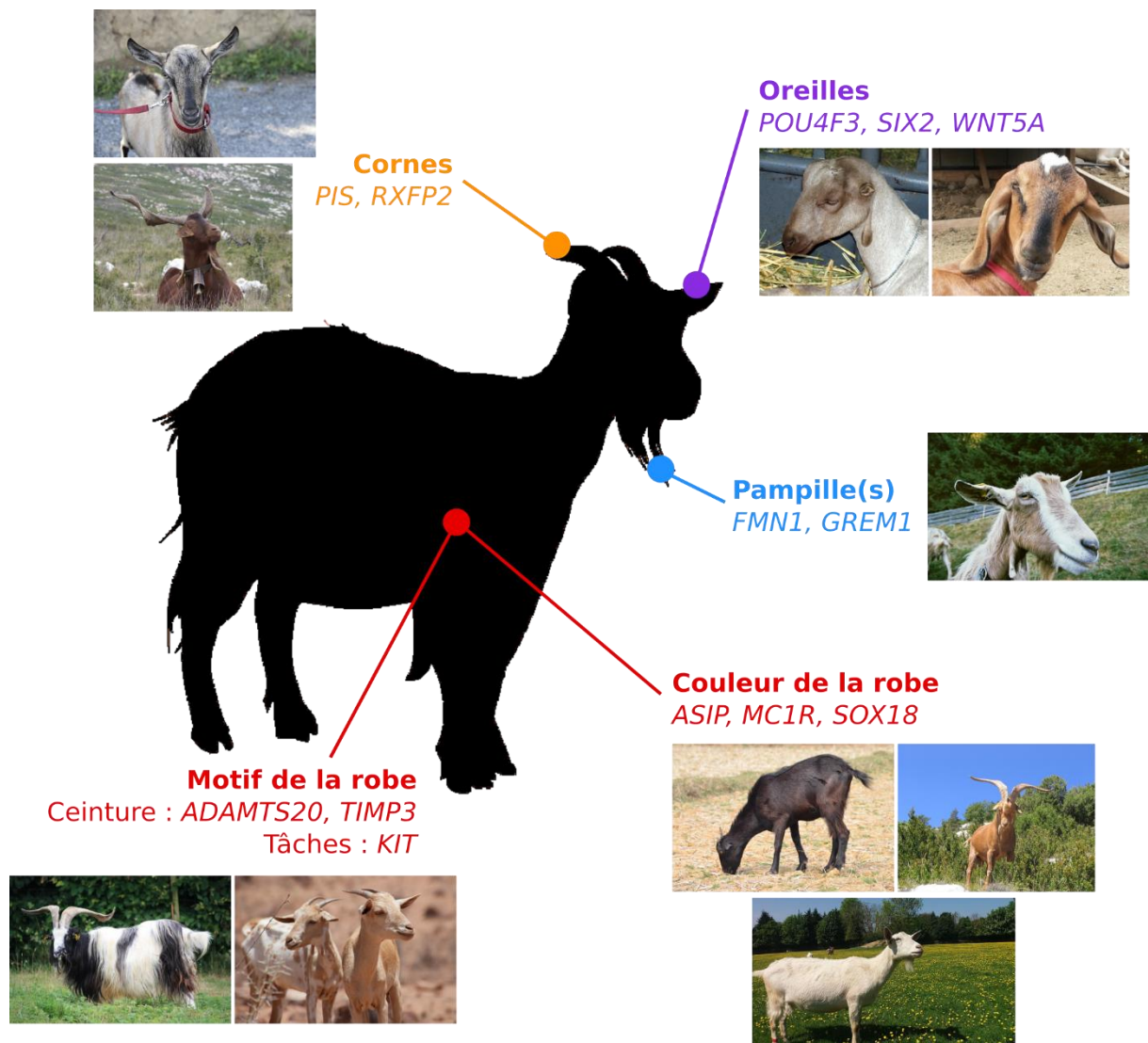


Figure 51 : caractéristiques morphologiques pouvant être spécifiques à certaines races de petits ruminants. Le pelage peut se décliner entre les couleurs noir (race Bengal, © ILRI/Jules Mateo), rouge (race Rove, © Dirk Beye) et blanc (race Saanen, © ChristianNawroth), et être uni, tacheté (race Abergelle, © ILRI\Zerihun Sewunet) ou avec une ceinture (race Fossés, © Tamsin Cooper). La présence de pampille(s) (race Toggenburger, © Baph) ou de cornes (race Rove avec des cornes, © Capgènes ; et chèvre motte, © JACLOU-DL) sont des caractères variants suivant les races. La morphologie des oreilles peut aussi varier suivant les races : oreilles courtes pour la race Lamancha (© jean) et oreilles longues pour la race Nubienne (© macphysto).

▪ La couleur du pelage

Les principaux gènes connus pour avoir un effet sur la couleur de la robe sont MC1R (Switonski et al., 2013), TYRP1 (Gratten et al., 2007; Becker et al., 2015), et SOX18 (Fitch, 2003) étant en relation avec les couleurs de robes sombres, et le gène ASIP avec la couleur blanche (Martin et al., 2016; Henkel et al., 2019). Excepté le gène TYRP1, ces gènes ont été détectés lors du projet ADAPTmap et en particulier le gène ASIP sélectionné chez les races du groupe Alpine et d'Asie Centrale. Nous trouvons ce gène sélectionné chez plusieurs de nos races françaises avec des régions sous sélections spécifiques aux races Poitevine, Alpine, Saanen et Savoie. Nous

trouvons une signature de sélection associée au gène KIT chez la race Lorraine. Ce gène est impliqué dans l'apparition de tâches sur la robe des chevaux (Bai et al., 2011) et est également sous sélection chez les races de chèvres Abergelle, et Kacchan (Bertolini et al., 2018). Pour finir, la présence d'une ceinture blanche sur le pelage de certaines races peut être due au gène TIMP3 qui interagit avec les gènes ADAMTS (Jackson et al., 2016) et ADAMTS20 (Rao et al., 2003) responsables de l'apparition de cette ceinture chez les souris. L'étude d'Oget et al. (2019) détecte le gène ADAMTS20 sous sélection chez les races Pyrénées et Fossés mais nous ne le retrouvons pas dans notre étude.

▪ La morphologie des oreilles

Les races françaises ne se différencient pas sur ce caractère, mais de nombreux gènes ont été trouvés associés à la morphologie des oreilles des chèvres dans l'étude de Brito et al. (2017). En particulier les gènes POU4F3 (Masuda et al., 2017) et WNT5A (Yamaguchi et al., 1999) en relation avec la morphogenèse des oreilles et le gène SIX2 avec leur développement (Guan et al., 2016).

Concernant les moutons, l'étude de C. Wei et al. (2015) détecte le gène MSRB3 sous sélection chez des races chinoises, qui est un gène connu pour avoir un effet sur la morphologie des oreilles chez les chiens (Vaysse et al., 2011) et les cochons (Ren et al., 2014).

Le phénotype « lobes d'oreilles » a été trouvé associé à des animaux ayant une masse corporelle plus élevée chez la race chinoise Longling (Leng et al., 2010), élevée pour sa viande.

▪ La présence de cornes

Les animaux mottes sont acceptés dans plusieurs races françaises, même si ce n'est pas un caractère recherché dans les races. Deux gènes en relation avec ce phénotype ont été détectés. Il s'agit du gène RXFP2 dont l'étude de Wiedemar et Drögemüller (2015) a montré qu'une insertion dans la partie 3' UTR du gène était en relation avec l'absence de cornes chez le mouton ; et de locus PIS (« polled intersex syndrome ») où une délétion d'environ 11 700 bases et la présence de plusieurs mutations dans cette région du génome entraînent l'apparition du phénotype motte chez la chèvre (Pailhous et al., 2001; Kijas et al., 2013). Comme son nom l'indique, le locus PIS est également associé à l'intersexualité chez les chèvres (Szatkowska et al., 2014). Ce point est détaillé dans la partie « Différentes cibles de la sélection ».

- La présence de pampilles

Les pampilles sont des appendices cutanés situés sur le cou des animaux, en particulier les chèvres et parfois les moutons (par exemple les races vivant dans la région du Sahel, Traoré et al., 2008). On ne connaît pas l'intérêt de ces appendices. Les gènes candidats concernant ce phénotype chez la chèvre sont FMN1 et GREM1 (Reber et al., 2015) dont le premier que nous trouvons spécifiquement sélectionné dans la race Alpine, chez qui ce phénotype est accepté. Ce sont des INDELs et non des SNPs qui seraient responsables de leur apparition (C. Drögemüller, communication personnelle). On ne connaît pas bien l'intérêt de ces appendices, mais ils ont été trouvés en lien avec des caractères laitiers ou de fertilité (Ricordeau et al., 1967; Shongjia et al., 1992), des productions bouchères (Leng et al., 2010), ou aussi la résistance des animaux à la chaleur (Adedeji et al., 2011).

Sélection induite par le climat

Si nous reprenons l'exemple du phénotype de la queue grasse chez le mouton, en plus d'être associé à de meilleures qualités bouchères, il permet aussi aux animaux de vivre dans des environnements et conditions extrêmes (Atti et al., 2004). Ces réserves lipidiques dans la queue du mouton peuvent faire penser aux bosses des chameaux chez qui il a été montré que le glucose avait un effet dans la rétention d'eau en réponse au climat désertique (Wu et al., 2014). Les SNPs qui sont détectés sous sélection peuvent donc aussi être la conséquence d'une adaptation des petits ruminants à l'environnement (climat, altitude).

Bertolini et al. (2018) ont effectué des analyses de génomique du paysage (« landscape genomic ») en comparant des races de chèvres vivant dans des climats tropicaux, secs tempérés ou continentaux ; ou en cherchant des associations entre les génotypes et ces climats. La comparaison entre un climat sec et chaud avec des climats continentaux ou tempérés a mis en évidence des gènes impliqués dans la réponse au stress oxydatif avec le gène GPR37L1 qui est impliqué dans la protection des cellules neuronales contre ce stress (Meyer et al., 2013) et dans la régulation du métabolisme avec le gène IGF2.

L'article de Benjelloun et al. (*Submitted to MBE*) se focalise sur l'adaptation locale chez les petits ruminants marocains en utilisant l'échantillonnage du projet NextGen pour réaliser différentes analyses d'adaptation à l'environnement. Si nous comparons les gènes que nous trouvons sélectionnés par rapport à ceux détectés dans ces articles nous ne trouvons aucun gène en commun avec l'étude de Bertolini et al (2018); deux gènes en commun (CP et KLF12)

pour les chèvres et un pour les moutons (SEMA5A) avec l'étude de Benjelloun et al. (*Submitted to MBE*). L'échantillonnage ainsi que les méthodes de détection ne sont pas les mêmes que celles que nous avons utilisées donc cela peut expliquer les différentes signatures de sélection trouvées.

Influence des données de génomes complets vs puce ADN

Comme il a été présenté dans le premier article de ce manuscrit décrivant les données du projet VarGoats, ce dernier prend la suite du projet ADAPTmap qui avait pour but d'étudier la diversité des chèvres à partir de 148 populations réparties sur cinq continents (Europe, Afrique, Asie, Amérique et Océanie, Stella et al., 2018). Les 4 653 animaux du projet ADAMTmap ont été génotypés avec la puce Illumina GoatSNP50 (Tosser-Klopp et al., 2014). Les résultats portant sur les signatures de sélection détectées au sein de ces populations ont été publiés par Bertolini et al. (2018). En parallèle, l'étude de Oget et al., (2019) s'est intéressée uniquement aux races françaises présentes dans le projet, c'est-à-dire les races Alpine, Angora, Corses, Fossés, Poitevine, Provençale, Pyrénées et Saanen. L'analyse hapFLK de ces races a permis de détecter cinq régions sous sélection dont quatre associées à des gènes en relation avec la production laitière (CSN1S1), la couleur de la robe (ADAMTS20 et ASIP) et la reproduction (CDH9).

Notre propre étude des signatures de sélection chez les races caprines françaises au sein du projet VarGoats comprend trois races en plus (Lorraine, Rove et Savoie) et une en moins (Corse). L'analyse hapFLK nous a permis de détecter 142 régions sous sélection associées à 273 gènes. Trente de nos régions sont comprises dans deux trouvées par Oget et al. (2019), correspondant au chromosome 6 ayant pour gène candidat CSN1S1, et au chromosome 13 avec le gène ASIP. L'étude des arbres locaux nous indique que la région sous le chromosome 6 est sélectionnée chez les Lorraines, Poitevines, Pyrénées et Alpines qui portent en plus une signature spécifique par rapport aux autres. L'étude précédente avait détecté cette région chez les races Alpine, Corse et Pyrénées. Pour l'haplotype en relation avec ASIP, l'étude précédente la trouvait en relation avec les races Fossés, Poitevine et Saanen. Nous retrouvons bien une signature de sélection chez les races Poitevine et Saanen auquel nous ajoutons les Alpines et Savoie.

La signature inter-génique trouvée sur le chromosome 11 chez les Fossés dans l'étude précédente est aussi détectée dans notre étude spécifique de la race. La signature sur le

chromosome 5 associée au gène ADAMTS20 chez les Fossés et les Pyrénées dans l'étude de Oget et al. (2019) est finalement détectée dans l'analyse spécifique à ces deux races. Cependant les pics significatifs sont plus proches des gènes TMEM117, PRICKLE1 et GXYLT1 qui jouent un rôle dans la résistances aux mammites et à la maladie de Johne (infection bactérienne répandue chez les bovins, Pant et al., 2010; Brito et al., 2018; Cai et al., 2018). Pour finir, la sélection sur le chromosome 20 chez les Pyrénées en relation avec le gène CDH9 est aussi retrouvée dans l'analyse spécifique de la race.

Pour conclure, nous retrouvons toutes les signatures de sélection mises en évidence par Oget et al. (2019) si nous cumulons nos analyses du jeu de données global et des signaux spécifiques à chaque race. Cependant en prenant uniquement l'étude globale nous ne détectons que deux régions en commun avec l'étude précédente. La résolution des régions sous sélection n'étant pas les mêmes, 250 Kb en moyenne dans notre étude contre 10 Mb de moyenne pour l'étude précédente, les interprétations des signaux de sélection ne sont donc pas identiques.

Le fait de réduire le nombre de SNPs (52 000 sur la puce contre plus de 34 millions sur le génome entier) peut influencer les résultats. L'étude de Benjelloun et al. (2019) s'est intéressée à ces biais en analysant les résultats de calculs de diversité nucléotidique ou de différenciation génétique (F_{st}) entre différentes densités de SNPs tirées au hasard dans le génome (de 1 000 SNPs à 5 millions), les puces illumina 50K et 600K ovines et 50K caprine et un séquençage entier du génome. Il en ressort qu'un tirage aléatoire de 10 000 SNPs dans le génome permet d'obtenir d'un spectre de fréquence des allèles alternatifs fortement corrélé à celui obtenu avec un séquençage complet. Cependant, quand on compare ce spectre entre les puces à ADN (aussi bien 50K que 600K) et le génome entier on ne trouve pas de corrélation. De plus, en ce qui concerne les signatures de sélection, certains signaux sont exclusivement détectés avec un séquençage entier du génome car la probabilité d'avoir un SNP dans la région sélectionnée est plus importante. Cette probabilité dépend de la densité de SNPs séquencés, le biais de sélection des SNPs, mais aussi de la taille du balayage sélectif. L'exemple pris dans l'étude est une région de 20 Mb sur le chromosome 10 du mouton qui porte trois signatures de sélection sur les gènes NBEA, MAB21L1 et RXFP2 avec des données de génome entier. Le signal de RXFP2 est quasiment détecté par tous les jeux de données (sauf le 50 000 SNPs aléatoires). Les puces détectent la sélection de RXFP2 car celles-ci sont biaisées, en effet la puce ovine 50K possède quatre SNPs au niveau du gène, alors qu'en tirant aléatoirement 50 000 SNPs cette région n'est pas toujours représentée. Les SNPs présents sur la puce caprine

ont été choisis pour avoir une fréquence de 0,2 principalement chez les Alpines et les Saanen ainsi qu'une distance d'environ 25 Kb d'écart ce qui permet de ne pas surreprésenter une région. Ces choix impliquent des biais de représentation du génome par rapport à un tirage aléatoire du même nombre de SNPs dans le génome ce qui peut induire des différences de résultat au niveau des calculs de diversité neutre ou de détection de signatures de sélection. En résumé, comme il a été dit dans l'introduction, les régions sous sélection ne représentent qu'une faible part du génome. Un effectif de 10 000 SNPs pris au hasard dans le génome peut suffire à estimer des paramètres neutres tels que la diversité nucléotidique, la consanguinité et la différenciation génétique entre groupes d'individus. Pour arriver à détecter les balayages sélectifs, il faut une densité de SNPs plus élevée afin qu'au minimum un SNP soit localisé au niveau de l'haplotype sélectionné. Sachant qu'un haplotype sélectionné peut être plus ou moins grand suivant l'âge et la force de la sélection (les recombinaisons et les mutations vont réduire la taille de cet haplotype). Le choix des SNPs présents sur la puce influence donc aussi les signatures de sélection qu'il est possible de détecter. Si nous reprenons les résultats que nous avons obtenus sur l'analyse globale nous obtenons 18 régions supplémentaires par rapport à l'étude d'Oget et al. (2019), ainsi que des interprétations des signaux de sélection différents car nous avons une résolution plus fine d'où sont situées les régions candidates.

Différentes cibles pour la sélection

La sélection peut se baser sur des valeurs de phénotypes. Dans ce cas les éleveurs vont se reproduire entre eux les animaux portant les phénotypes d'intérêt. C'est par exemple le cas de la sélection des chèvres Alpines et Saanen qui est basée sur un index de synthèse caprin, prenant en compte des caractères laitiers (matière protéique, matière grasse ...), la santé de la mamelle (cellules somatiques) et la morphologie mammaire (Palhière et al., 2015). La construction de cet index de synthèse caprin a pris en compte l'héritabilité de chaque caractère ainsi que le fait qu'il ne soit pas lié à un autre caractère défavorable.

C'est par exemple ce qui est arrivé lors de la sélection des chèvres mottes car le même locus génétique (PIS) code pour l'absence de cornes mais aussi l'inversion de sexe chez les femelles. Cela se présente sous la forme de différentes anomalies de développement du système sexuel allant de changements subtils à une inversion complète du sexe chez les femelles (Szatkowska et al., 2014) et ayant des répercussions sur le système reproducteur. Une sélection basée sur des critères de reproduction permet de ne pas fixer ce caractère dans la race.

L'exemple le plus connu concernant la sélection involontaire de phénotype est l'expérience de Dmitri Belyaev (1917-1985) qui a entrepris de domestiquer des renards argentés. Il les a sélectionnés sur la docilité (faible distance de fuite face aux humains) et a remarqué des changements dans leur morphologie comme une queue enroulée, des oreilles tombantes ou un pelage tacheté (Belyaev, 1979). Il est possible que certains haplotypes sous sélection que nous avons mis en évidence soient des sélections involontaires, comme par exemple les pampilles qui sont associées à des caractères de production mais ne sont pas un phénotype sélectionné chez les races françaises.

Une autre façon de sélectionner les phénotypes d'intérêt est de détecter les SNPs ayant un impact sur le phénotype observé (mutation causale). Nous pouvons citer l'exemple de la mutation R396W dans le gène DGAT1 qui explique 46% de la variance génétique concernant la teneur en matière grasse du lait chez les races Saanen et Alpine (Martin et al., 2017). La découverte de ce type de mutation est cependant assez rare, comme nous l'avons vu avec l'analyse de l'effet des variants prédit par Ensembl (VEP), seulement 1% des SNPs présents dans notre jeu de données sont positionnés dans des régions codantes et aucun dans des régions régulatrices annotées. Il ne suffit pas qu'un variant soit dans une région codante ou régulatrice pour qu'il ait un impact sur un phénotype, cela va dépendre du changement qu'il implique dans la synthèse de la protéine. Cependant avec l'analyse de prédiction des effets des variants nous avons vu que 0,01% des variants sous sélection sont associés à la perte ou au gain de codons « stop » et la perte de codons « start » qui ont une influence directe sur la synthèse de protéine. De plus, un phénotype peut être codé par plusieurs gènes (polygénie), donc une mutation sur un seul gène ne modifiera pas forcément le caractère observé.

Pour finir, un gène peut avoir un effet pléiotropique (sur différents phénotypes) pouvant avoir des impacts différents sur les phénotypes impactés. C'est par exemple le cas du gène SOCS2 chez le mouton dont l'allèle p.R96C réduit le taux de cellules somatiques dans le lait, mais implique également un moins bon rendement laitier et réduit la taille des animaux (Rupp et al., 2015). La sélection des moutons pour réduire le taux de cellules somatiques dans le lait doit donc s'effectuer en prenant en compte d'autres gènes afin de ne pas réduire la production laitière.

II. Marques épigénétiques : témoin de l'acclimatation

Chez les petits ruminants marocains nous avons trouvé des régions différentiellement méthylées pouvant être impliquées dans des mécanismes d'acclimatation à la variation de température. Ces mécanismes peuvent permettre un changement rapide de phénotype pour s'acclimater à de nouveaux environnements comme l'a montré l'étude de Hu et al. (2019) qui a trouvé des cytosines différentiellement méthylées en relation entre autre avec l'immunité et le métabolisme chez des lézards qui sont restés quatre jours dans un nouvel environnement. Concernant les animaux d'élevage, le projet CLIMGEN est le premier à analyser des ruminants élevés de manière traditionnelle et donc à mettre en évidence des marques de méthylation de l'ADN en relation avec l'environnement chez ces animaux. En effet, en plus de notre étude sur les chèvres et les moutons, Sevane et al. (2018) se sont intéressés aux races bovines adaptées aux tropiques et des races ibériques élevées dans l'environnement de leur ancêtre. Ils ont ainsi pu identifier des DMR dans des gènes directement ou indirectement impliqués dans les processus d'acclimatation au climat tropical, tels que des processus en lien avec le système nerveux, l'immunité, la gestion de l'énergie, la résistance à la chaleur et les attributs de la peau et du pelage. Nous ne trouvons aucun gène (portant des marques génétiques ou épigénétiques) en commun avec cette étude cependant les voies métaboliques impactées par l'environnement sont en partie identiques, avec par exemple des marques épigénétiques en relation avec l'immunité, la gestion de l'énergie et la résistance à la chaleur. Nous trouvons, en plus, des relations avec des caractères d'intérêt agronomique tels que la reproduction et la production de lait, qui peuvent être impactées de manière directe par la température ou de manière indirecte par l'intermédiaire de la qualité ou quantité de ressources en eau ou nourriture.

III. Relation entre les deux mécanismes

Nous avons vu en introduction que les mécanismes génétiques pouvaient être mis en relation avec le processus d'adaptation et les mécanismes épigénétiques avec l'acclimatation. Cela a souvent donné lieu à des publications spécifiques à l'une ou l'autre des approches mais sans les mettre en relation. Cependant, ces deux mécanismes impactent l'expression des gènes et *a fortiori* le phénotype des organismes, leur permettant une adéquation à leur environnement

(Figure 52). Il paraît donc intéressant d'étudier les relations qu'il peut y avoir entre les mécanismes génétiques et épigénétiques.

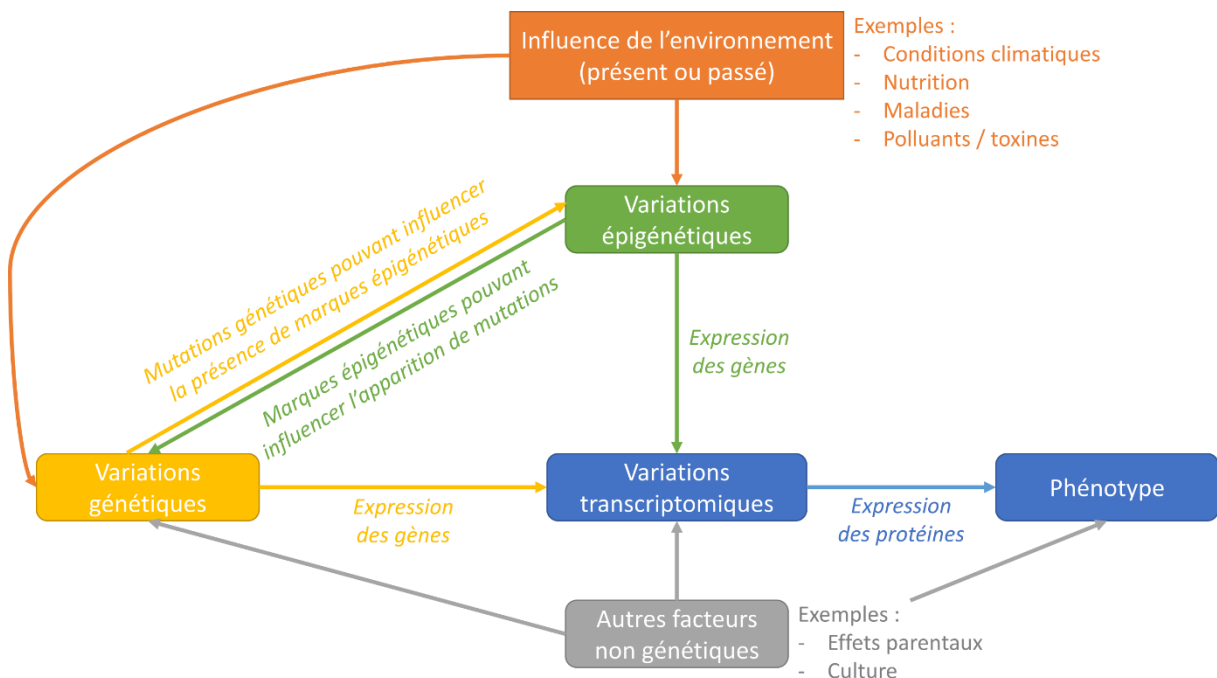


Figure 52 : Schéma mettant en relation les mécanismes génétiques et épigénétiques ainsi que leurs effets sur un phénotypique d'après Ibeagha-Awemu & Khatib (2017).

Localisation des variants

La localisation d'un allèle ou d'une cytosine méthylée sur le génome peut être liée à la présence de l'autre marqueur dans la région.

Richards (2006) propose trois cas de figure théoriques :

1. Le premier comprenant une corrélation entre la présence d'une marque épigénétique et d'un allèle (Figure 53 A.)
2. Le cas intermédiaire où la présence de la marque de méthylation est facilitée par la présence d'un allèle quel qu'il soit, mais avec également un effet de l'environnement (Figure 53 B.)
3. Le cas où la marque de méthylation est seulement corrélée à l'environnement (Figure 53 C.)

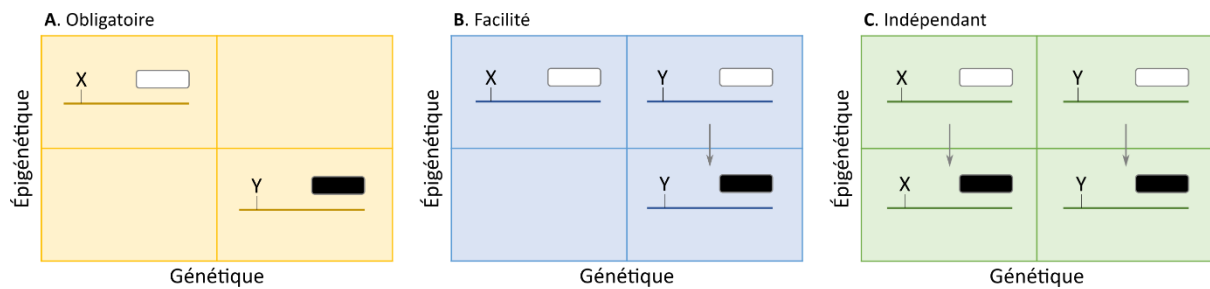


Figure 53 : classe de variation épigénétique. L'axe horizontal distingue deux génotypes qui sont représentés par les allèles X et Y et l'axe vertical des états épigénétiques sont représentés par des cases vides ou remplies (d'après Richards, 2006).

En pratique, Eichten et al. (2013) ont étudié les profils de méthylation de l'ADN chez 20 lignées consanguines de maïs (*Zea mays*) et détectent 3 720 DMRs quand ils les comparent à la lignée référente B73. Parmi les 1 966 DMRs communs à plusieurs lignées, 51% sont associés à des SNPs locaux (effet *cis*), environ 2% sont localisés dans des régions sans SNPs et pour les 47% DMRs restants, les auteurs proposent que ceux-ci peuvent ne pas être corrélés à des SNPs ou que des effets de SNPs plus éloignés (effet *trans*) peuvent influencer leur présence.

Une perspective de notre étude sur les différences de méthylation de l'ADN chez les petits ruminants marocains, pourrait être de chercher si les DMRs que nous avons détectées sont associées à des SNPs particuliers.

À l'inverse, il est aussi possible que des marques épigénétiques finissent par transformer la séquence ADN et donc devenir des marqueurs génétiques. En effet, le taux de mutation des cytosines méthylées est environ 20 000 fois plus élevé que celui des cytosines non méthylées (Gorelick, 2003). Jones et al. (1992) estiment que les cytosines méthylées pourraient provoquer 30 à 40 % des mutations ponctuelles de la lignée germinale chez l'humain.

Effet des variants

Meng et al. (2016) ont montré que chez *Arabidopsis thaliana* l'expression des gènes était régulée par les mécanismes génétiques et épigénétiques. Cependant la proportion de variance expliquée par la méthylation de l'ADN est très faible par rapport à l'effet des SNPs locaux ou même globaux sur l'expression des gènes.

Une analyse transcriptomique pourrait être effectuée afin de détecter les gènes différentiellement exprimés entre nos groupes environnementaux de petits ruminants

marocains. Cela nous permettrait de voir si les gènes portant les variants génétiques et épigénétiques que nous avons détectés, sont régulés.

Différents processus évolutifs pour répondre à l'environnement

Comme peu d'études s'intéressent à la relation entre les marques génétiques et épigénétiques, il est compliqué de trouver des articles comparant l'adaptation et l'acclimatation dans différentes populations. Certaines études s'intéressent à l'évolution de la plasticité phénotypique (ensemble des mécanismes permettant la variation du phénotype à partir d'un génotype fixe) dans des populations et son lien avec l'adaptation.

Il peut être intéressant de se demander pourquoi ces deux processus existent si tous les deux permettent l'ajustement du phénotype des organismes à leur environnement. Ceux-ci n'ont pas le même mode d'action comme nous avons vu tout au long de cette thèse et ils ont également des temps de réponse face à changement d'environnement qui ne sont pas les mêmes.

Les mécanismes de plasticité phénotypiques sont mis en place rapidement en réponse à un changement environnemental et le changement de phénotype peut être réversible si le changement disparaît. Cela a été montré par Sentis et al. (2018) qui ont utilisé une population de pucerons clonés à partir d'un seul individu (pas de variabilité génétique) afin d'observer la réponse de celle-ci à des prédateurs. Pour se protéger des coccinelles, la fréquence des pucerons ailés a augmenté rapidement (deux générations) avant de se stabiliser durant les 25 génération suivantes. Ce résultat montre que la plasticité phénotypique permet une réponse rapide de l'organisme à un stress et que ce phénotype peut être transmis à la descendance. Pour trois groupes d'animaux testés, les prédateurs ont été retirés à 3, 13 et 22 générations. Dans ces trois cas, les auteurs ont observé une fréquence de pucerons ailés revenant à la valeur des témoins. Cela montre que la plasticité phénotypique a un coût assez important pour que, en absence du stress, le phénotype initial réapparaisse. Ce coût serait principalement lié à la mise en place du mécanisme de régulation et à l'expression du caractère plastique (Auld et al., 2010). Cependant les études essayant de quantifier le coût de la plasticité dans différents environnements ne permettent pas de tirer de conclusion claire. En effet, Auld et al. (2010) concluent que 62 % des 227 estimations de coût de plasticité dans deux environnements n'ont pas permis de détecter de coût significatif ; 21 % ont trouvé soit

un coût soit un bénéfice dans un environnement et rien dans l'autre ; 8 % ont détecté un coût dans un environnement et un bénéfice dans l'autre ; 5 % ont détecté un coût dans les deux environnements ; et 4 % ont détecté un bénéfice dans les deux environnements.

Dans le cas où le stress persiste assez longtemps, comme par exemple lors de la colonisation des nouveaux environnement par des espèces invasives, Lande (2015) propose un scénario d'évolution de la plasticité et de la génétique au cours du temps. Tout d'abord une augmentation de la plasticité phénotypique permet l'apparition d'un nouveau phénotype optimal dans l'environnement colonisé ; ensuite, une assimilation génétique lente du nouveau phénotype entraînant une réduction de la plasticité apparaissent. Ce scénario est cohérent avec le taux de mutation plus élevé dans les régions méthylées du génome qui permettent la transformation de la séquence génétique par des mécanismes épigénétiques.

Dans ce cas-là, la plasticité phénotypique devient alors de l'adaptation locale dont on peut chercher les signatures de sélection suivant les techniques présentées dans cette thèse. Cette adaptation locale permet le maintien du phénotype optimal en réduisant certaines contraintes liées à la plasticité comme la mise en place du mécanisme de régulation.

Le scénario proposé par Lande (2015) pourrait être étendu à n'importe quel organisme devant faire face à une nouvelle contrainte environnementale. Cependant, il peut être compliqué de tester la vraisemblance de ce scénario car pour ça il faut pouvoir élever les organismes durant plusieurs générations (entre 1 000 et 10 000 d'après l'auteur) afin de commencer à voir des changements génétiques en relation avec la plasticité phénotypique.

Enfin, comme il a été dit en introduction, l'adaptation provient principalement de la variabilité résidente dans le génome que de nouvelles mutations (Savolainen et al., 2013) donc celle-ci peut aussi être un processus mis en place indépendamment de la plasticité phénotypique. Suivant la force de la sélection il faut plus ou moins de générations avant de le phénotype optimal se fixe dans la population subissant le changement environnemental. L'adaptation est donc plutôt un mécanisme à long terme, là où la plasticité phénotypique permet une réponse à court terme des organismes à un changement d'environnement.

IV. Conclusion

Cette thèse a permis de mettre en évidence des mécanismes d'adaptation et d'acclimatation des petits ruminant en fonction de leur conditions d'élevage et/ou de leur environnement.

Tout d'abord une analyse de la diversité neutre des races caprines françaises nous a permis de décrire la diversité génétique de ces races et comment celle-ci pouvait être liée à leur histoire (démographie, croisements entre races). Dans un second temps, nous avons cherché les signatures de sélection contrastant ces différentes races. Nous avons ainsi pu détecter des SNPs sous sélection en relation avec des gènes impliqués dans des caractères d'intérêt agronomique tels que la reproduction, la production laitière, l'immunité et la morphologie. La relation entre les gènes sous sélection et le phénotype des races n'a pas toujours pu être faite, car il y a peu de relevés chez les races locales. Chez les chèvres la majorité de la sélection s'effectue sur le phénotype des animaux, même si certains allèles commencent à être recherchés chez les mâles. Dans ce cas-là une sélection intentionnelle sur le caractère d'intérêt est effectuée, comme par exemple le gène de la caséine CN1S1 chez les Alpines, en relation avec la coagulation du lait permettant la fabrication de fromage. D'autres SNPs ont pu être sélectionné de façon involontaire, comme le gène FMN1 que nous trouvons sélectionné dans la race Alpine et qui est impliqué dans l'apparition de pampilles chez les chèvres, elles-mêmes associées à des caractères de production et de fertilité.

Notre étude a été effectuée sans a priori, en contrastant les races présentes dans notre jeu de données. Une caractérisation des phénotypes de chaque race caprine française permettrait de réaliser des analyses d'association entre ceux-ci et les génotypes (GWAS). Nous pourrions ainsi détecter des gènes spécifiquement impliqués dans l'expression d'un phénotype. Cela nous permettrait également de chercher des signatures de sélection chez les races Angora et Créole qui avaient été retirées de notre jeu de données pour la comparaison entre race car leur origine hors de la France métropolitaine (Moyen-Orient et Afrique respectivement) les différenciait trop des autres races.

Le séquençage de différentes races de chèvres originaires d'Europe du Nord et de l'Est (issues de la voie de migration danubienne) nous permettrait de comparer leur génome avec celui de la Lorraine. Nous saurions alors si celle-ci a également une origine danubienne et/ou si des croisements ont pu avoir lieu entre la Lorraine et ces autres races. La voie de migration

danubienne est moins étudiée que la voie méditerranéenne donc nous n'avons pas trouvé de données publiques correspondantes à des races ayant cette origine pour les ajouter aux données VarGoats. De plus, aucun partenaire n'a été trouvé afin de nous fournir des échantillons que nous aurions pu analyser.

La seconde étude a permis de mettre en évidence deux régions différentiellement méthylées entre deux groupes de chèvres et moutons marocains, contrastés en fonction de variations annuelles de températures plus ou moins importantes. Les gènes associés à ces régions jouent un rôle dans le métabolisme, la production et la qualité du lait, et la reproduction. En plus de cette analyse, nous avons effectué une autre analyse des différences génétiques entre les deux groupes de petits ruminants. Des gènes sous sélection, impliqués dans la perception de l'environnement, l'immunité, la production et la reproduction ont été trouvés.

Nous n'avons trouvé aucun gène en commun entre les analyses génétiques et épigénétiques, cependant les voies biologiques impactées sont similaires. Cela suggère une complémentarité entre les deux mécanismes.

Deux analyses complémentaires pourraient être effectuées afin de valider la présence de cytosines différentiellement méthylée dans les DMR, et vérifier si celles-ci ainsi que les SNPs sous sélection induisent une régulation des gènes les portant.

La première analyse consisterait à effectuer un séquençage bisulfite des DMRs. Cette méthode permet de convertir les cytosines non méthylées en uracile et ainsi détecter exactement la position des marques de méthylation.

Une étude transcriptomique permettrait de trouver les gènes différentiellement exprimés entre les deux groupes de petits ruminants définis dans notre étude. En théorie nous pourrions retrouver les gènes associés aux DMRs et à la sélection, mais également d'autres gènes soumis à d'autres mécanismes de régulation.

BIBLIOGRAPHIE

A

- ABO-ISMAIL, M.K., BRITO, L.F., MILLER, S.P., SARGOLZAEI, M., GROSSI, D.A., MOORE, S.S., ET AL. (2017) Genome-wide association studies and genomic prediction of breeding values for calving performance and body conformation traits in Holstein cattle. *Genetics Selection Evolution*, 49, 82.
- ADEDEJI, T.A., OZOJE, M.O., PETERS, S.O., SANUSI, O.A., OJEDAPO, L.O. & IGE, A.O. (2011) Coat pigmentation and Wattle Genes' Effect on some Haematological Characteristics of Heat Stressed and Extensively Reared West African Dwarf Goats. *World Journal of Life Sciences and Medical Research*, 1, 48–55.
- AGGREY, S.E., HABASHY, W.S., MILFORT, M.C., ADOMAKO, K., FULLER, A.L. & REKAYA, R. (2018) Molecular and cellular mechanisms that underlie genes and antioxidant enzyme activities in meat-type birds during heat stress. In p. 4.
- AHBARA, A., BAHBAHANI, H., ALMATHEN, F., AL ABRI, M., AGOUB, M.O., ABEBA, A., ET AL. (2019) Genome-Wide Variation, Candidate Regions and Genes Associated With Fat Deposition and Tail Morphology in Ethiopian Indigenous Sheep. *Frontiers in Genetics*, 9. Frontiers.
- AI, H., YANG, B., LI, J., XIE, X., CHEN, H. & REN, J. (2014) Population history and genomic signatures for high-altitude adaptation in Tibetan pigs. *BMC Genomics*, 15, 834.
- AIGNER, B., RENNER, S., KESSLER, B., KLYMIUK, N., KUROME, M., WÜNSCH, A. & WOLF, E. (2010) Transgenic pigs as models for translational biomedical research. *Journal of Molecular Medicine*, 88, 653–664.
- ALBERTO, F.J., BOYER, F., OROZCO-TERWENGEL, P., STREETER, I., SERVIN, B., VILLEMEREUIL, P. DE, ET AL. (2018) Convergent genomic signatures of domestication in sheep and goats. *Nature Communications*, 9, 813.
- ALBRECHTSEN, A., NIELSEN, F.C. & NIELSEN, R. (2010) Ascertainment Biases in SNP Chips Affect Measures of Population Divergence. *Molecular Biology and Evolution*, 27, 2534–2547.
- ALEXANDER, D.H., NOVEMBRE, J. & LANGE, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655–1664.
- ALEXANDRE, G., ASSELIN DE BEAUVILLE, S., SHITALOU, E. & ZEBUS, M.F. (2008) An overview of the goat meat sector in Guadeloupe: conditions of production, consumer preferences, cultural functions and economic implications. *Livestock Research for Rural Development*, 20.
- ALFONSO, J., FERNANDEZ, M.E., COOPER, B., FLUGGE, G. & FRASCH, A.C. (2005) The stress-regulated protein M6a is a key modulator for neurite outgrowth and filopodium/spine formation. *Proceedings of the National Academy of Sciences*, 102, 17196–17201.
- ALLAIN, D. (1994) The genetic improvement of angora goats in France. In *Genetic Improvement of Fine Fibre Producing Animals* (eds J.P. Laker & S.C. Bishop), pp. 47–60 European Fine Fibre Network. Macaulay Land Use Research Institute, Aberdeen, Scotland.
- AL-TAMIMI, H.J. (2007) Thermoregulatory response of goat kids subjected to heat stress. *Small Ruminant Research*, 71, 280–285.
- AMARAL, A.J., BRESSAN, M.C., ALMEIDA, J., BETTENCOURT, C., MOREIRA, O., SA, J., ET AL. (2019) Combining genome-wide association analyses and gene interaction networks to reveal new genes associated with carcass traits, meat quality and fatty acid profiles in pigs. *Livestock Science*, 220, 180–189.

- AMUNDSON, O.L., FOUNTAIN, T.H., LARIMORE, E.L., RICHARDSON, B.N., MCNEEL, A.K., WRIGHT, E.C., ET AL. (2015) Postweaning nutritional programming of ovarian development in beef heifers. *Journal of Animal Science*, 93, 5232–5239.
- AN, B., XU, LEI, XIA, J., WANG, X., MIAO, J., CHANG, T., ET AL. (2020) Multiple association analysis of loci and candidate genes that regulate body size at three growth stages in Simmental beef cattle. *BMC Genetics*, 21, 32.
- ANDERSON, O.S., SANT, K.E. & DOLINOY, D.C. (2012) Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and DNA methylation. *The Journal of Nutritional Biochemistry*, 23, 853–859.
- ANDREWS, W.D., BARBER, M. & PARNAVELAS, J.G. (2007) Slit–Robo interactions during cortical development. *Journal of Anatomy*, 211, 188–198.
- ASSOCIATION POUR LA SAUVEGARDE DE LA CHEVRE PROVENÇALE (2017) Schéma de sélection. La Chèvre Provençale - Schéma de sélection. <https://lachevreprovencale.wordpress.com/donnees-genetiques/> [accessed 12 October 2020].
- ATTI, N., BOCQUIER, F. & KHALDI, G. (2004) Performance of the fat-tailed Barbarine sheep in its environment: adaptive capacity to alternation of underfeeding and re-feeding periods. A review. *Animal Research*, 53, 165–176.
- AULD, J.R., AGRAWAL, A.A. & RELYEA, R.A. (2010) Re-evaluating the costs and limits of adaptive phenotypic plasticity. *Proceedings of the Royal Society B: Biological Sciences*, 277, 503–511. Royal Society.

B

- BAI D.-Y., YANG L.-H., UN ERHU, ZHAO Y.-P., ZHAO Q.-N., HASI GAOWA & DUGARJAVIIN M. (2011) Effects of Kit gene on coat depigmentation in white horses: Effects of Kit gene on coat depigmentation in white horses. *Hereditas (Beijing)*, 33, 1171–1178.
- BAI, M., SUN, L., ZHAO, J., XIANG, L., CHENG, X., LI, J., ET AL. (2017) Histological analysis and identification of spermatogenesis-related genes in 2-, 6-, and 12-month-old sheep testes. *The Science of Nature*, 104, 84.
- BARCHUK, A.R., CRISTINO, A.S., KUCHARSKI, R., COSTA, L.F., SIMÕES, Z.L. & MALESZKA, R. (2007) Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *BMC Developmental Biology*, 7, 70.
- BECKER, D., OTTO, M., AMMANN, P., KELLER, I., DRÖGEMÜLLER, C. & LEEB, T. (2015) The brown coat colour of Coppernecked goats is associated with a non-synonymous variant at the TYRP1 locus on chromosome 8. *Animal Genetics*, 46, 50–54.
- BÉLTEKY, J., AGNVALL, B., BEKTIC, L., HÖGLUND, A., JENSEN, P. & GUERRERO-BOSAGNA, C. (2018) Epigenetics and early domestication: differences in hypothalamic DNA methylation between red junglefowl divergently selected for high or low fear of humans. *Genetics Selection Evolution*, 50.
- BELYAEV, D.K. (1979) Destabilizing selection as a factor in domestication. *The Journal of Heredity*, 70, 301–308.
- BENCINI, R. & PULINA, G. (1997) The quality of sheep milk: a review. *Australian Journal of Experimental Agriculture*, 37, 485.

- BENJELLOUN, B. (2015) Diversité des génomes et adaptation locale des petits ruminants d'un pays méditerranéen : le Maroc. Grenoble Alpes.
- BENJELLOUN, B., ALBERTO, F.J., STREETER, I., BOYER, F., COISSAC, E., STUCKI, S., ET AL. (2015) Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (*Capra hircus*) using WGS data. *Frontiers in Genetics*, 6.
- BENJELLOUN, B., BOYER, F., STREETER, I., ZAMANI, W., ENGELEN, S., ALBERTI, A., ET AL. (2019) An evaluation of sequencing coverage and genotyping strategies to assess neutral and adaptive diversity. *Molecular Ecology Resources*, 19, 1497–1515.
- BENJELLOUN, B., LEEMPOEL, K., BOYER, F., STUCKI, S., STREETER, I., OROZCO-TERWENGEL, P., ET AL. (Submitted to MBE) Different patterns of adaptive genomic variation in two close species (sheep and goats) coexisting along the same environmental gradients.
- BERTOLINI, F., CARDOSO, T.F., MARRAS, G., NICOLAZZI, E.L., ROTHSCCHILD, M.F., AMILLS, M. & ADAPTMAP CONSORTIUM (2018) Genome-wide patterns of homozygosity provide clues about the population history and adaptation of goats. *Genetics Selection Evolution*, 50, 59.
- BERTOLINI, F., SERVIN, B., TALENTI, A., ROCHAT, E., KIM, E.S., OGET, C., ET AL. (2018) Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genetics Selection Evolution*, 50, 57.
- BETT, B., KIUNGA, P., GACHOHI, J., SINDATO, C., MBOTHA, D., ROBINSON, T., ET AL. (2017) Effects of climate change on the occurrence and distribution of livestock diseases. *Preventive Veterinary Medicine*, 137, 119–129.
- BHUIYAN, A., LI, J., WU, Z., NI, P., ADETULA, A., WANG, H., ET AL. (2017) Exploring the Genetic Resistance to Gastrointestinal Nematodes Infection in Goat Using RNA-Sequencing. *International Journal of Molecular Sciences*, 18, 751.
- BICKHART, D.M., ROSEN, B.D., KOREN, S., SAYRE, B.L., HASTIE, A.R., CHAN, S., ET AL. (2017) Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics*, 49, 643–650.
- BIONAZ, M. & LOOR, J.J. (2008) ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 Are the Most Abundant Isoforms in Bovine Mammary Tissue and Their Expression Is Affected by Stage of Lactation. *The Journal of Nutrition*, 138, 1019–1024.
- BOITARD, S., RODRÍGUEZ, W., JAY, F., MONA, S. & AUSTERLITZ, F. (2016) Inferring Population Size History from Large Samples of Genome-Wide Molecular Data - An Approximate Bayesian Computation Approach. *PLOS Genetics*, 12, e1005877.
- BOITARD, S., SCHLÖTTERER, C. & FUTSCHIK, A. (2009) Detecting Selective Sweeps: A New Approach Based on Hidden Markov Models. *Genetics*, 181, 1567–1578.
- BOLLATI, V. & BACCARELLI, A. (2010) Environmental epigenetics. *Heredity*, 105, 105–112. Nature Publishing Group.
- BOLORMAA, S., BROWN, D.J., SWAN, A.A., WERF, J.H.J. VAN DER, HAYES, B.J. & DAETWYLER, H.D. (2017) Genomic prediction of reproduction traits for Merino sheep. *Animal Genetics*, 48, 338–348.
- BONHOMME, M., CHEVALET, C., SERVIN, B., BOITARD, S., ABDALLAH, J., BLOTT, S. & SANCRISTOBAL, M. (2010) Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended. *Genetics*, 186, 241–262.

- BORN, K., CHRISTOPH, M., FINK, A.H., KNIPPERTZ, P., PAETH, H. & SPETH, P. (2008) Moroccan Climate in the Present and Future: Combined View from Observational Data and Regional Climate Scenarios. In *Climatic Changes and Water Resources in the Middle East and North Africa* (eds F. Zereini & H. Hötzl), pp. 29–45. Springer Berlin Heidelberg, Berlin, Heidelberg.
- BOROWSKA, A., REYER, H., WIMMERS, K., VARLEY, P.F. & SZWACZKOWSKI, T. (2017) Detection of pig genome regions determining production traits using an information theory approach. *Livestock Science*, 205, 31–35.
- BOSSDORF, O., RICHARDS, C.L. & PIGLIUCCI, M. (2008) Epigenetics for ecologists. *Ecology Letters*, 11, 106–115.
- BRENNAN, R.S., HEALY, T.M., BRYANT, H.J., LA, M.V., SCHULTE, P.M. & WHITEHEAD, A. (2018) Integrative Population and Physiological Genomics Reveals Mechanisms of Adaptation in Killifish. *Molecular Biology and Evolution*, 35, 2639–2653.
- BRITO, L.F., KIJAS, J.W., VENTURA, R.V., SARGOLZAEI, M., PORTO-NETO, L.R., CÁNOVAS, A., ET AL. (2017) Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC Genomics*, 18.
- BRITO, L.F., MALLIKARJUNAPPA, S., SARGOLZAEI, M., KOECK, A., CHESNAIS, J., SCHENKEL, F.S., ET AL. (2018) The genetic architecture of milk ELISA scores as an indicator of Johne's disease (paratuberculosis) in dairy cattle. *Journal of Dairy Science*, 101, 10062–10075. Elsevier.
- BRUFORD, M.W., GINJA, C., HOFFMANN, I., JOOST, S., OROZCO-TERWENGEL, P., ALBERTO, F.J., ET AL. (2015) Prospects and challenges for the conservation of farm animal genomic resources, 2015–2025. *Frontiers in Genetics*, 6, 314.
- BRUSH, S., KESSELI, R., ORTEGA, R., CISNEROS, P., ZIMMERER, K. & QUIROS, C. (1995) Potato Diversity in the Andean Center of Crop Domestication. *Conservation Biology*, 9, 1189–1198.
- BUCHNER, D.A., GEISINGER, J.M., GLAZEBROOK, P.A., MORGAN, M.G., SPIEZIO, S.H., KAIYALA, K.J., ET AL. (2012) The juxtaparanodal proteins CNTNAP2 and TAG1 regulate diet-induced obesity. *Mammalian Genome*, 23, 431–442.

C

- CAI, Z., DUSZA, M., GULDBRANDTSEN, B., LUND, M.S. & SAHANA, G. (2020) Distinguishing pleiotropy from linked QTL between milk production traits and mastitis resistance in Nordic Holstein cattle. *Genetics Selection Evolution*, 52.
- CAI, Z., GULDBRANDTSEN, B., LUND, M.S. & SAHANA, G. (2018) Prioritizing candidate genes post-GWAS using multiple sources of data for mastitis resistance in dairy cattle. *BMC Genomics*, 19.
- CAILLAT, H., BOUVIER, F., RAINARD, P., MARTIN, P., CLEMENT, V., LARROQUE, H., ET AL. (2011) Réponse à une sélection divergente sur la concentration de cellules somatiques du lait chez la chèvre Alpine. *Rencontres autour des Recherches sur les Ruminants*, 18, 405–408.
- CARBON, S. & MUNGALL, C. (2018) Gene Ontology Data Archive. Zenodo. <https://zenodo.org/record/3980761> [accessed 14 September 2020].
- CARDOSO, T.F., AMILLS, M., BERTOLINI, F., ROTHSCCHILD, M., MARRAS, G., BOINK, G., ET AL. (2018) Patterns of homozygosity in insular and continental goat breeds. *Genetics Selection Evolution*, 50, 56.

- CARILLIER, C., LARROQUE, H., PALHIÈRE, I., CLÉMENT, V., RUPP, R. & ROBERT-GRANIÉ, C. (2013) A first step toward genomic selection in the multi-breed French dairy goat population. *Journal of Dairy Science*, 96, 7294–7305.
- CARILLIER-JACQUIN, C., LARROQUE, H. & ROBERT-GRANIÉ, C. (2016) Including α 1casein gene information in genomic evaluations of French dairy goats. *Genetics Selection Evolution*, 48, 54.
- CAROPRESE, M., ALBENZIO, M., BRUNO, A., FEDELE, V., SANTILLO, A. & SEVI, A. (2011) Effect of solar radiation and flaxseed supplementation on milk production and fatty acid profile of lactating ewes under high ambient temperature. *Journal of Dairy Science*, 94, 3856–3867.
- CARR, D., LEWIN-KOH, N., MAECHLER, M. & SARKAR, D. (2019) hexbin: Hexagonal Binning Routines.
- CERVANTES, I., GOYACHE, F., MOLINA, A., VALERA, M. & GUTIERREZ, J.P. (2011) Estimation of effective population size from the rate of coancestry in pedigreed populations. *Journal of Animal Breeding and Genetics*, 128, 56–63.
- CHARMET, G. (2011) Wheat domestication: Lessons for the future. *Comptes Rendus Biologies*, 334, 212–220.
- CHEN, I.-C., HILL, J.K., OHLEMÜLLER, R., ROY, D.B. & THOMAS, C.D. (2011) Rapid range shifts of species associated with high levels of climate warming. *Science (New York, N.Y.)*, 333, 1024–1026.
- CHEN, Q., QU, K., MA, Z., ZHAN, J., ZHANG, F., SHEN, J., ET AL. (2020) Genome-Wide Association Study Identifies Genomic Loci Associated With Neurotransmitter Concentration in Cattle. *Frontiers in Genetics*, 11. *Frontiers*.
- CHOMWISARUTKUN, K., MURANI, E., PONSUKSILI, S. & WIMMERS, K. (2012) Microarray analysis reveals genes and functional networks relevant to the predisposition to inverted teats in pigs¹. *Journal of Animal Science*, 90, 1–15.
- CHOUDHARY, R.K., CHOUDHARY, S. & VERMA, R. (2018) In vivo response of xanthosine on mammary gene expression of lactating Beetal goat. *Molecular Biology Reports*, 45, 581–590.
- CHRISTMAS, M.J., WALLBERG, A., BUNIKIS, I., OLSSON, A., WALLERMAN, O. & WEBSTER, M.T. (2019) Chromosomal inversions associated with environmental adaptation in honeybees. *Molecular Ecology*, 28, 1358–1374.
- CINGOLANI, P., PLATTS, A., WANG, L.L., COON, M., NGUYEN, T., WANG, L., ET AL. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w 1118 ; iso-2; iso-3. *Fly*, 6, 80–92.
- COCHRAN, S.D., COLE, J.B., NULL, D.J. & HANSEN, P.J. (2013) Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. *BMC Genetics*, 14, 49.
- COLLI, L., MILANESI, M., TALENTI, A., BERTOLINI, F., CHEN, M., CRISA, A., ET AL. (2018) Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. *Genetics Selection Evolution*, 50, 58.
- COONEY, C.A., DAVE, A.A. & WOLFF, G.L. (2002) Maternal Methyl Supplements in Mice Affect Epigenetic Variation and DNA Methylation of Offspring. *The Journal of Nutrition*, 132, 2393S-2400S.
- CUMER, T. (2017) Etude des variants structuraux génomiques pour comprendre les processus démographiques et adaptatifs impliqués dans la domestication des petits ruminants. Université Grenoble Alpes.

CUMER, T., BOYER, F. & POMPANON, F. (Submitted to GBE) Genome-wide detection of structural variations reveals new regions associated to domestication in small ruminants.

D

DA COSTA BARROS, C., DE ABREU SANTOS, D.J., ASPILCUETA-BORQUIS, R.R., DE CAMARGO, G.M.F., DE ARAUJO NETO, F.R. & TONHATI, H. (2018) Use of single-step genome-wide association studies for prospecting genomic regions related to milk production and milk quality of buffalo. *Journal of Dairy Research*, 85, 402–406.

DADO-SENN, B., SKIBIEL, A.L., FABRIS, T.F., DAHL, G.E. & LAPORTA, J. (2019) Dry period heat stress induces microstructural changes in the lactating mammary gland. *PLOS ONE*, 14, e0222120.

DAHL, G.E., SKIBIEL, A.L. & LAPORTA, J. (2019) In Utero Heat Stress Programs Reduced Performance and Health in Calves. *Veterinary Clinics of North America: Food Animal Practice*, 35, 343–353.

DAI, L., XU, J., LIU, S., MA, T., ZHU, Y., XU, F., ET AL. (2014) Characterization of miR-126-3p and its Target Talin2 in the Bovine Corpus Luteum during the Oestrus Cycle. *Reproduction in Domestic Animals*, 49, 913–919.

DALL’OLIO, S., SCHIAVO, G., GALLO, M., BOVO, S., BERTOLINI, F., BUTTAZZONI, L. & FONTANESI, L. (2020) Candidate gene markers associated with production, carcass and meat quality traits in Italian Large White pigs identified using a selective genotyping approach. *Livestock Science*, 240, 104145.

DANCHIN, É., CHARMANTIER, A., CHAMPAGNE, F.A., MESOUDI, A., PUJOL, B. & BLANCHET, S. (2011) Beyond DNA: integrating inclusive inheritance into an extended theory of evolution. *Nature Reviews Genetics*, 12, 475–486.

DANCHIN-BURGE, C., DANVY, S., LALOE, D. & VERRIER, E. (2017) Création d’un observatoire de la VARIabilité génétique des RUMinants et des Equidés (VARUME). *Innovations Agronomiques*, 55, 235–245.

DANECEK, P., AUTON, A., ABECASIS, G., ALBERS, C.A., BANKS, E., DEPRISTO, M.A., ET AL. (2011) The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.

DE CREMOUX, R., LAGRIFFOUL, G., ALLAIN, C., ALAOUI-SOSSE, L., ASTRUC J-M., BATUT, E., ET AL. (2018) MAMOVICAP - Vers des outils innovants d’intervention et d’aide à la décision pour la maîtrise des mammites en élevage de petits ruminants laitiers. *Innovations Agronomiques*, 63, 99–114. INRA.

DE OLIVEIRA, P.S.N., COUTINHO, L.L., CESAR, A.S.M., DINIZ, W.J. DA S., DE SOUZA, M.M., ANDRADE, B.G., ET AL. (2019) Co-Expression Networks Reveal Potential Regulatory Roles of miRNAs in Fatty Acid Composition of Nelore Cattle. *Frontiers in Genetics*, 10. *Frontiers*.

DELANEAU, O., HOWIE, B., COX, A.J., ZAGURY, J.-F. & MARCHINI, J. (2013) Haplotype Estimation Using Sequencing Reads. *The American Journal of Human Genetics*, 93, 687–696.

DICKINSON, R.E. & DUNCAN, W.C. (2010) The SLIT–ROBO pathway: a regulator of cell function with implications for the reproductive system. *Reproduction*, 139, 697–704.

DO, D.N., SCHENKEL, F.S., MIGLIOR, F., ZHAO, X. & IBEAGHA-AWEMU, E.M. (2018) Genome wide association study identifies novel potential candidate genes for bovine milk cholesterol content. *Scientific Reports*, 8, 13239. *Nature Publishing Group*.

DOHERTY, R., FARRELLY, C.O. & MEADE, K.G. (2014) Comparative epigenetics: relevance to the regulation of production and health traits in cattle. *Animal Genetics*, 45, 3–14.

DOLINOY, D.C., HUANG, D. & JIRTLE, R.L. (2007) Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proceedings of the National Academy of Sciences*, 104, 13056–13061.

DONG, Y., XIE, M., JIANG, Y., XIAO, N., DU, X., ZHANG, W., ET AL. (2013) Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nature Biotechnology*, 31, 135–141.

DONG, Y., ZHANG, X., XIE, M., AREFNEZHAD, B., WANG, Z., WANG, WENLIANG, ET AL. (2015) Reference genome of wild goat (*capra aegagrus*) and sequencing of goat breeds provide insight into genic basis of goat domestication. *BMC Genomics*, 16.

DORAN, A.G., BERRY, D.P. & CREEVEY, C.J. (2014) Whole genome association study identifies regions of the bovine genome and biological pathways involved in carcass trait performance in Holstein-Friesian cattle. *BMC Genomics*, 15, 837.

DREGER, D.L., PARKER, H.G., OSTRANDER, E.A. & SCHMUTZ, S.M. (2013) Identification of a Mutation that Is Associated with the Saddle Tan and Black-and-Tan Phenotypes in Basset Hounds and Pembroke Welsh Corgis. *Journal of Heredity*, 104, 399–406. Oxford Academic.

DU, C., DENG, T.X., ZHOU, Y., GHANEM, N. & HUA, G.H. (2020) Bioinformatics analysis of candidate genes for milk production traits in water buffalo (*Bubalus bubalis*). *Tropical Animal Health and Production*, 52, 63–69.

E

EGUCHI, T., PRINCE, T.L., TRAN, M.T., SOGAWA, C., LANG, B.J. & CALDERWOOD, S.K. (2019) MZF1 and SCAND1 Reciprocally Regulate CDC37 Gene Expression in Prostate Cancer. *Cancers*, 11, 792.

EICHTEIN, S.R., BRISKINE, R., SONG, J., LI, Q., SWANSON-WAGNER, R., HERMANSON, P.J., ET AL. (2013) Epigenetic and Genetic Influences on DNA Methylation Variation in Maize Populations. *The Plant Cell*, 25, 2783–2797.

ELENA, S.F. & LENSKI, R.E. (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics*, 4, 457–469. Nature Publishing Group.

EVANGELOU, P., GROLL, M., OPPERMAN, H., GAUNITZ, F., EISENLÖFFEL, C., MÜLLER, W., ET AL. (2019) Assessment of ApoC1, LuzP6, C12orf75 and OCC-1 in cystic glioblastoma using MALDI-TOF mass spectrometry, immunohistochemistry and qRT-PCR. *Medical Molecular Morphology*, 52, 217–225.

F

FANG, X., ZHAO, Z., YU, H., LI, G., JIANG, P., YANG, Y., ET AL. (2017) Comparative genome-wide methylation analysis of longissimus dorsi muscles between Japanese black (Wagyu) and Chinese Red Steppes cattle. *PLoS ONE*, 12.

FANG, Z.-H. & PAUSCH, H. (2019) Multi-trait meta-analyses reveal 25 quantitative trait loci for economically important traits in Brown Swiss cattle. *BMC Genomics*, 20, 695.

FARIELLO, M.I., BOITARD, S., NAYA, H., SANCRISTOBAL, M. & SERVIN, B. (2013) Detecting Signatures of Selection Through Haplotype Differentiation Among Hierarchically Structured Populations. *Genetics*, 193, 929–941.

- FENG, S., COKUS, S.J., ZHANG, X., CHEN, P.-Y., BOSTICK, M., GOLL, M.G., ET AL. (2010) Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences*, 107, 8689–8694.
- FERNANDEZ, A. L., PEREZ-MONTARELO, D., BARRAGAN, C., RAMAYO-CALDAS, Y., IBAÑEZ-ESCRICHE, N., CASTELLO, A., ET AL. (2012) Genome-wide linkage analysis of QTL for growth and body composition employing the PorcineSNP60 BeadChip. *BMC Genetics*, 13, 41.
- FEUK, L., CARSON, A.R. & SCHERER, S.W. (2006) Structural variation in the human genome. *Nature Reviews Genetics*, 7, 85–97.
- FINK, S., MÖMKE, S. & DISTL, O. (2012) PLXNC1 and RDH13 associated with bilateral convergent strabismus with exophthalmus in German Brown cattle. *Molecular Vision*, 18, 2229–2240.
- FITAK, R.R. (submitted) optM: an R package to optimize the number of migration edges using threshold models. *Journal of Heredity*.
- FITCH, K.R. (2003) Genetics of dark skin in mice. *Genes & Development*, 17, 214–228.
- FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS (ed) (1999) The global strategy for the management of farm animal genetic resources: executive brief. Food and Agricultural Organization of the United Nations, Rome, Italy.
- FORTES, M.R.S., LI, Y., COLLIS, E., ZHANG, Y. & HAWKEN, R.J. (2013) The IGF1 pathway genes and their association with age of puberty in cattle. *Animal Genetics*, 44, 91–95.
- FOUST, C.M., PREITE, V., SCHREY, A.W., ALVAREZ, M., ROBERTSON, M.H., VERHOEVEN, K.J.F. & RICHARDS, C.L. (2016) Genetic and epigenetic differences associated with environmental gradients in replicate populations of two salt marsh perennials. *Molecular Ecology*, 25, 1639–1652.
- FRICHOT, E., SCHOVILLE, S.D., BOUCHARD, G. & FRANÇOIS, O. (2013) Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. *Molecular Biology and Evolution*, 30, 1687–1699.
- FRISCHKNECHT, M., BAPST, B., SEEFRIED, F.R., SIGNER-HASLER, H., GARRICK, D., STRICKER, C., ET AL. (2017) Genome-wide association studies of fertility and calving traits in Brown Swiss cattle using imputed whole-genome sequences. *BMC Genomics*, 18, 910.

G

- GADDIS, K.L.P., NULL, D.J. & COLE, J.B. (2016) Explorations in genome-wide association studies and network analyses with dairy cattle fertility traits. *Journal of Dairy Science*, 99, 6420–6435. Elsevier.
- GARRISON, E. & MARTH, G. (2012) Haplotype-based variant detection from short-read sequencing. *arXiv:1207.3907 [q-bio]*.
- GAUTIER, M. & VITALIS, R. (2012) reh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics*, 28, 1176–1177.
- GIBNEY, E.R. & NOLAN, C.M. (2010) Epigenetics and gene expression. *Heredity*, 105, 4–13. Nature Publishing Group.
- GIRI, A.K., MIDHA, S., BANERJEE, P., AGRAWAL, A., MEHDI, S.J., DHINGRA, R., ET AL. (2016) Common Variants in CLDN2 and MORC4 Genes Confer Disease Susceptibility in Patients with Chronic Plu. *PLOS ONE*, 11, e0147345. Public Library of Science.

GONZALEZ, M., VILLA, R., VILLA, C., GONZALEZ, V., MONTANO, M., MEDINA, G. & MAHADEVAN, P. (2020) Inspection of real and imputed genotypes revealed 76 SNPs associated to rear udder height in Holstein cattle. *Journal of Advanced Veterinary and Animal Research*, 7, 234.

GORELICK, R. (2003) Evolution of dioecy and sex chromosomes via methylation driving Muller's ratchet: EVOLUTION OF DIOECY and SEX CHROMOSOMES. *Biological Journal of the Linnean Society*, 80, 353–368.

GRATTEN, J., BERALDI, D., LOWDER, B.V., MCRAE, A.F., VISSCHER, P.M., PEMBERTON, J.M. & SLATE, J. (2007) Compelling evidence that a single nucleotide substitution in TYRP1 is responsible for coat-colour polymorphism in a free-living population of Soay sheep. *Proceedings of the Royal Society B: Biological Sciences*, 274, 619–626.

GROSS, B.L. & ZHAO, Z. (2014) Archaeological and genetic insights into the origins of domesticated rice. *Proceedings of the National Academy of Sciences*, 111, 6190–6197.

GUAN, J., WANG, D., CAO, W., ZHAO, Y., DU, R., YUAN, H., ET AL. (2016) SIX2 haploinsufficiency causes conductive hearing loss with ptosis in humans. *Journal of Human Genetics*, 61, 917–922. Nature Publishing Group.

GUO, J., TAO, H., LI, P., LI, L., ZHONG, T., WANG, L., ET AL. (2018) Whole-genome sequencing reveals selection signatures associated with important traits in six goat breeds. *Scientific Reports*, 8, 10405. Nature Publishing Group.

GUO, L., COSTANZO-GARVEY, D.L., SMITH, D.R., NEILSEN, B.K., MACDONALD, R.G. & LEWIS, R.E. (2017) Kinase Suppressor of Ras 2 (KSR2) expression in the brain regulates energy balance and glucose homeostasis. *Molecular Metabolism*, 6, 194–205.

GUO, X., SU, G., CHRISTENSEN, O.F., JANSS, L. & LUND, M.S. (2016) Genome-wide association analyses using a Bayesian approach for litter size and piglet mortality in Danish Landrace and Yorkshire pigs. *BMC Genomics*, 17.

GUTIÉRREZ-GIL, B., ARRANZ, J.J., PONG-WONG, R., GARCÍA-GÁMEZ, E., KIJAS, J. & WIENER, P. (2014) Application of Selection Mapping to Identify Genomic Regions Associated with Dairy Production in Sheep. *PLOS ONE*, 9, e94623. Public Library of Science.

H

HE, C., WANG, C., CHANG, Z.H., GUO, B.L., LI, R., YUE, X.P., ET AL. (2011) AGPAT6 polymorphism and its association with milk traits of dairy goats. *Genetics and molecular research: GMR*, 10, 2747–2756.

HENKEL, J., SAIF, R., JAGANNATHAN, V., SCHMOCKER, C., ZEINDLER, F., BANGERTER, E., ET AL. (2019) Selection signatures in goats reveal copy number variants underlying breed-defining coat color phenotypes. *PLOS Genetics*, 15, e1008536. Public Library of Science.

HERRERA, C.M. & BAZAGA, P. (2010) Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet *Viola cazorlensis*. *New Phytologist*, 187, 867–876.

HIJMANS, R.J., CAMERON, S.E., PARRA, J.L., JONES, P.G. & JARVIS, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978.

HOLLISTER, J.D. & GAUT, B.S. (2009) Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research*, 19, 1419–1428.

HOSSAINI-HILALI, J. & MOUSLIH, Y. (2002) La chèvre Draa. Potentiel de production et caractéristiques d'adaptation aux contraintes de l'environnement aride. *Animal Genetic Resources Information*, 32, 49–56.

HU, J., ASKARY, A.M., THURMAN, T.J., SPILLER, D.A., PALMER, T.M., PRINGLE, R.M. & BARRETT, R.D.H. (2019) The Epigenetic Signature of Colonizing New Environments in Anolis Lizards. *Molecular Biology and Evolution*, 36, 2165–2170.

HUSSEIN, H.A. & STAUFENBIEL, R. (2012) Variations in Copper Concentration and Ceruloplasmin Activity of Dairy Cows in Relation to Lactation Stages with Regard to Ceruloplasmin to Copper Ratios. *Biological Trace Element Research*, 146, 47–52.

I

IBEAGHA-AWEMU, E.M. & KHATIB, H. (2017) Epigenetics of Livestock Breeding. In *Handbook of Epigenetics: The New Molecular and Medical Genetics*, Second edition pp. 441–463 Trygve O. Tollefsbol. Academic Press, University of Alabama at Birmingham, Birmingham, AL, United States.

IDELE (2019) Indicateurs de variabilité génétique - races caprines - Edition 2019. [idele.fr. http://idele.fr/no_cache/recherche/publication/idelesolr/recommends/indicateurs-de-variabilite-genetique-races-caprines-edition-2019.html](http://idele.fr/no_cache/recherche/publication/idelesolr/recommends/indicateurs-de-variabilite-genetique-races-caprines-edition-2019.html) [accessed 1 July 2020].

J

JACINTO, F.V., BALLESTAR, E. & ESTELLER, M. (2008) Methyl-DNA immunoprecipitation (MeDIP): Hunting down the DNA methylome. *BioTechniques*, 44, 35–43.

JACKSON, H.W., DEFAMIE, V., WATERHOUSE, P. & KHOKHA, R. (2016) TIMPs: versatile extracellular regulators in cancer. *Nature Reviews Cancer*, 17, 38–53.

JANOWITZ KOCH, I., CLARK, M.M., THOMPSON, M.J., DEERE-MACHEMER, K.A., WANG, J., DUARTE, L., ET AL. (2016) The concerted impact of domestication and transposon insertions on methylation patterns between dogs and grey wolves. *Molecular Ecology*, 25, 1838–1855.

JARDIM, J.G., GULDBRANDTSEN, B., LUND, M.S. & SAHANA, G. (2018) Association analysis for udder index and milking speed with imputed whole-genome sequence variants in Nordic Holstein cattle. *Journal of Dairy Science*, 101, 2199–2212. Elsevier.

JIANG, Y., XIE, M., CHEN, W., TALBOT, R., MADDOX, J.F., FARAUT, T., ET AL. (2014) The sheep genome illuminates biology of the rumen and lipid metabolism. *Science*, 344, 1168–1173.

JIVANJI, S., WORTH, G., LOPDELL, T.J., YEATES, A., COULDREY, C., REYNOLDS, E., ET AL. (2019) Genome-wide association analysis reveals QTL and candidate mutations involved in white spotting in cattle. *Genetics Selection Evolution*, 51, 62.

JONES, P.A., RIDEOUT, W.M., SHEN, J.-C., SPRUCK, C.H. & TSAI, Y.C. (1992) Methylation, mutation and cancer. *BioEssays*, 14, 33–36.

K

- KASHAN, N.E.J., AZAR, G.H.M., AFZALZADEH, A. & SALEHI, A. (2005) Growth performance and carcass quality of fattening lambs from fat-tailed and tailed sheep breeds. *Small Ruminant Research*, 60, 267–271.
- KEEL, B.N., ZAREK, C.M., KEELE, J.W., KUEHN, L.A., SNELLING, W.M., OLIVER, W.T., ET AL. (2018) RNA-Seq Meta-analysis identifies genes in skeletal muscle associated with gain and intake across a multi-season study of crossbred beef steers. *BMC Genomics*, 19.
- KHAYATZADEH, N., MESZAROS, G., UTSUNOMIYA, Y.T., SCHMITZ-HSU, F., SEEFRIED, F., SCHNYDER, U., ET AL. (2019) Genome-wide mapping of the dominance effects based on breed ancestry for semen traits in admixed Swiss Fleckvieh bulls. *Journal of Dairy Science*, 102, 11217–11224. Elsevier.
- KIJAS, J.W., ORTIZ, J.S., MCCULLOCH, R., JAMES, A., BRICE, B., SWAIN, B., ET AL. (2013) Genetic diversity and investigation of polledness in divergent goat populations using 52 088 SNPs. *Animal Genetics*, 44, 325–335.
- KILLICK, R. & ECKLEY, I.A. (2014) changepoint: An R Package for Changepoint Analysis. *Journal of Statistical Software*, 58.
- KIM, G., CLARKE, C.R., LAROSE, H., TRAN, H.T., HAAK, D.C., ZHANG, L., ET AL. (2017) Herbicide injury induces DNA methylome alterations in *Arabidopsis*. *PeerJ*, 5, e3560. PeerJ Inc.
- KOMOLKA, K., PONSUKSILI, S., ALBRECHT, E., KÜHN, C., WIMMERS, K. & MAAK, S. (2016) Gene expression profile of *Musculus longissimus dorsi* in bulls of a Charolais×Holstein F2-cross with divergent intramuscular fat content. *Genomics Data*, 7, 131–133.
- KRISHNAN, G., BAGATH, M., PRAGNA, P., VIDYA, M.K., ALEENA, J., RAVINDRANATHANARCHANA, P., ET AL. (2017) Mitigation of the Heat Stress Impact in Livestock Reproduction. *Theriogenology*.
- KUCHARSKI, R., MALESZKA, J., FORET, S. & MALESZKA, R. (2008) Nutritional Control of Reproductive Status in Honeybees via DNA Methylation. *Science*, 319, 1827–1830.
- KWON, S.G., HWANG, J.H., PARK, D.H., KIM, T.W., KANG, D.G., KANG, K.H., ET AL. (2016) Identification of Differentially Expressed Genes Associated with Litter Size in Berkshire Pig Placenta. *PLOS ONE*, 11, e0153311. Public Library of Science.

L

- LAMBETH, J.D. (2004) NOX enzymes and the biology of reactive oxygen. *Nature Reviews Immunology*, 4, 181–189. Nature Publishing Group.
- LAN, X., CRETNEY, E.C., KROPP, J., KHATEEB, K., BERG, M.A., PEÑAGARICANO, F., ET AL. (2013) Maternal Diet during Pregnancy Induces Gene Expression and DNA Methylation Changes in Fetal Tissues in Sheep. *Frontiers in Genetics*, 4.
- LANDE, R. (2015) Evolution of phenotypic plasticity in colonizing species. *Molecular Ecology*, 24, 2038–2045.
- LAODIM, T., ELZO, M.A., KOONAWOOTRITRIRON, S., SUWANASOPEE, T. & JATTAWA, D. (2017) Identification of SNP markers associated with milk and fat yields in multibreed dairy cattle using two genetic group structures. *Livestock Science*, 206, 95–104.

- LARSON, G. & FULLER, D.Q. (2014) The Evolution of Animal Domestication. *Annual Review of Ecology, Evolution, and Systematics*, 45, 115–136.
- LEA, A.J., ALTMANN, J., ALBERTS, S.C. & TUNG, J. (2016) Resource base influences genome-wide DNA methylation levels in wild baboons (*Papio cynocephalus*). *Molecular ecology*, 25, 1681–1696.
- LEE, B.-Y., LEE, K.-N., LEE, T., PARK, J.-H., KIM, S.-M., LEE, H.-S., ET AL. (2015) Bovine Genome-wide Association Study for Genetic Elements to Resist the Infection of Foot-and-mouth Disease in the Field. *Asian-Australasian Journal of Animal Sciences*, 28, 166–170.
- LEE, KANG & KIM (2019) Bayes Factor-Based Regulatory Gene Network Analysis of Genome-Wide Association Study of Economic Traits in a Purebred Swine Population. *Genes*, 10, 293.
- LEE, P.-C., WILDT, D.E. & COMIZZOLI, P. (2018) Proteomic analysis of germinal vesicles in the domestic cat model reveals candidate nuclear proteins involved in oocyte competence acquisition. *Molecular Human Reproduction*, 24, 14–26. Oxford Academic.
- LENG, J., ZHU, R., ZHAO, G., YANG, Q. & MAO, H. (2010) Quantitative and Qualitative Body Traits of Longling Yellow Goats in China. *Agricultural Sciences in China*, 9, 408–415.
- LEROY, G., MARY-HUARD, T., VERRIER, E., DANVY, S., CHARVOLIN, E. & DANCHIN-BURGE, C. (2013) Methods to estimate effective population size using pedigree data: Examples in dog, sheep, cattle and horse. *Genetics Selection Evolution*, 45, 1.
- LI, H. & DURBIN, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25, 1754–1760.
- LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., ET AL. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078–2079.
- LI, J.-P., YANG, C.-Y., CHUANG, H.-C., LAN, J.-L., CHEN, D.-Y., CHEN, Y.-M., ET AL. (2014) The phosphatase JKAP/DUSP22 inhibits T-cell receptor signalling and autoimmunity by inactivating Lck. *Nature Communications*, 5, 3618. Nature Publishing Group.
- LI, L., ZHANG, D., REN, Y., YE, S., ZHENG, B., LIU, S., ET AL. (2019) The modification of mitochondrial energy metabolism and histone of goat somatic cells under small molecules compounds induction. *Reproduction in Domestic Animals*, 54, 138–149.
- LI, Y., KIJAS, J., HENSHALL, J.M., LEHNERT, S., MCCULLOCH, R.M. & REVERTER, A. (2014) Using random forests (RF) to prescreen candidate genes: A new prospective for GWAS. 10th World Congress for Genetics Applied to Livestock Production, British Columbia, Vancouver.
- LI, Z., CHEN, J., WANG, ZHEN, PAN, Y., WANG, Q., XU, N. & WANG, ZHENG GUANG (2016) Detection of selection signatures of population-specific genomic regions selected during domestication process in Jinhua pigs. *Animal Genetics*, 47, 672–681.
- LIENHARD, M., GRIMM, C., MORKEL, M., HERWIG, R. & CHAVEZ, L. (2014) MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics*, 30, 284–286.
- LIU, J., ZHANG, L., WANG, D., SHEN, H., JIANG, M., MEI, P., ET AL. (2003) Congenital diaphragmatic hernia, kidney agenesis and cardiac defects associated with Slit3-deficiency in mice. *Mechanisms of Development*, 120, 1059–1070.
- LIU, M., ZHOU, Y., ROSEN, B.D., VAN TASSELL, C.P., STELLA, A., TOSSER-KLOPP, G., ET AL. (2018) Diversity of copy number variation in the worldwide goat population. *Heredity*, 122, 636–646. Nature Publishing Group.

LU, B., JIANG, Y.J., MAN, M.Q., BROWN, B., ELIAS, P.M. & FEINGOLD, K.R. (2005) Expression and regulation of 1-acyl- sn -glycerol- 3-phosphate acyltransferases in the epidermis. *Journal of Lipid Research*, 46, 2448–2457.

LUKIĆ, B., FERENČAKOVIĆ, M., ŠALAMON, D., ČAČIĆ, M., OREHOVAČKI, V., IACOLINA, L., ET AL. (2020) Conservation Genomic Analysis of the Croatian Indigenous Black Slavonian and Turopolje Pig Breeds. *Frontiers in Genetics*, 11.

LUU, K., BAZIN, E. & BLUM, M.G.B. (2017) pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17, 67–77.

M

MAKINA, S.O., MUCHADEYI, F.C., VAN MARLE-KÖSTER, E., TAYLOR, J.F., MAKGAHLELA, M.L. & MAIWASHE, A. (2015) Genome-wide scan for selection signatures in six cattle breeds in South Africa. *Genetics Selection Evolution*, 47, 92.

MANCEAU, M., DOMINGUES, V.S., LINNEN, C.R., ROSENBLUM, E.B. & HOEKSTRA, H.E. (2010) Convergence in pigmentation at multiple levels: mutations, genes and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 2439–2450.

MANFREDI, E. & ÅDNØY, T. (2012) Génétique des caprins laitiers, 11.

MANFREDI, E., RICORDEAU, G., BARBIERI, M.E., AMIGUES, Y. & BIBÉ, B. (1995) Genotype at the α s1-casein locus and selection of bucks on progeny test in the Alpine and Saanen breeds. *Genetics Selection Evolution*, 27, 451–458.

MARAI, I.F.M., EL-DARAWANY, A.A., FADIEL, A. & ABDEL-HAFEZ, M.A.M. (2007) Physiological traits as affected by heat stress in sheep—A review. *Small Ruminant Research*, 71, 1–12.

MARTIN, G.B., RODGER, J. & BLACHE, D. (2004) Nutritional and environmental effects on reproduction in small ruminants. *Reproduction, Fertility and Development*, 16, 491.

MARTIN, P. & LEROUX, C. (2000) Le gène caprin spécifiant la caséine α s1 : un suspect tout désigné aux effets aussi multiples qu'inattendus. *INRAE Productions Animales*, 13, 125–132.

MARTIN, P., PALHIÈRE, I., MAROTEAU, C., BARDOU, P., CANALE-TABET, K., SARRY, J., ET AL. (2017) A genome scan for milk production traits in dairy goats reveals two new mutations in *Dgat1* reducing milk fat content. *Scientific Reports*, 7, 1872. Nature Publishing Group.

MARTIN, P., PALHIÈRE, I., MAROTEAU, C., CLEMENT, V., DAVID, I., KLOPP, G.T. & RUPP, R. (2018) Genome-wide association mapping for type and mammary health traits in French dairy goats identifies a pleiotropic region on chromosome 19 in the Saanen breed. *Journal of Dairy Science*, 101, 5214–5226. Elsevier.

MARTIN, P.M., PALHIÈRE, I., RICARD, A., TOSSER-KLOPP, G. & RUPP, R. (2016) Genome Wide Association Study Identifies New Loci Associated with Undesired Coat Color Phenotypes in Saanen Goats. *PloS one*, 11, e0152426.

MASUDA, M., LI, Y., PAK, K., CHAVEZ, E., MULLEN, L. & RYAN, A.F. (2017) The promoter and multiple enhancers of the *pou4f3* gene regulate expression in inner ear hair cells. *Molecular neurobiology*, 54, 5414–5426.

MCCRACKEN, K.G., BARGER, C.P. & SORENSON, M.D. (2010) Phylogenetic and structural analysis of the HbA (α A/ β A) and HbD (α D/ β A) hemoglobin genes in two high-altitude waterfowl from the Himalayas

- and the Andes: Bar-headed goose (*Anser indicus*) and Andean goose (*Chloephaga melanoptera*). *Molecular Phylogenetics and Evolution*, 56, 649–658.
- MCDONALD, K.W., MCCLURE, C.J.W., ROLEK, B.W. & HILL, G.E. (2012) Diversity of birds in eastern North America shifts north with global warming. *Ecology and Evolution*, 2, 3052–3060.
- MCGARRY, J.D. & BROWN, N.F. (1997) The Mitochondrial Carnitine Palmitoyltransferase System — From Concept to Molecular Analysis. *European Journal of Biochemistry*, 244, 1–14.
- MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., ET AL. (2010) Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation. *Genome Research*, 20, 1297–1303.
- MDLADLA, K., DZOMBA, E.F. & MUCHADEYI, F.C. (2018) Landscape genomics and pathway analysis to understand genetic adaptation of South African indigenous goat populations. *Heredity*, 120, 369–378.
- MEI, C., WANG, HONGCHENG, LIAO, Q., KHAN, R., RAZA, S.H.A., ZHAO, C., ET AL. (2019) Genome-wide analysis reveals the effects of artificial selection on production and meat quality traits in Qinchuan cattle. *Genomics*, 111, 1201–1208.
- MENG, D., DUBIN, M., ZHANG, P., OSBORNE, E.J., STEGLE, O., CLARK, R.M. & NORDBORG, M. (2016) Limited Contribution of DNA Methylation Variation to Expression Regulation in *Arabidopsis thaliana*. *PLoS Genetics*, 12.
- MEUWISSEN, T.H.E. & LUO, Z. (1992) Computing inbreeding coefficient in large populations. *Genetics, Selection, Evolution*, 24, 305–313.
- MEYER, R.C., GIDDENS, M.M., SCHAEFER, S.A. & HALL, R.A. (2013) GPR37 and GPR37L1 are receptors for the neuroprotective and glioprotective factors prosaptide and prosaposin. *Proceedings of the National Academy of Sciences*, 110, 9529–9534.
- MEYERHOEFFER, D.C., WETTEMANN, R.P., COLEMAN, S.W. & WELLS, M.E. (1985) Reproductive Criteria of Beef Bulls during and after Exposure to Increased Ambient Temperature. *Journal of Animal Science*, 60, 352–357.
- MI, H., MURUGANUJAN, A., EBERT, D., HUANG, X. & THOMAS, P.D. (2019) PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research*, 47, D419–D426.
- MILLER, B.A. & LU, C.D. (2019) Current status of global dairy goat production: an overview. *Asian-Australasian Journal of Animal Sciences*, 32, 1219–1232.
- MILLS, R.E., WALTER, K., STEWART, C., HANDSAKER, R.E., CHEN, K., ALKAN, C., ET AL. (2011) Mapping copy number variation by population-scale genome sequencing. *Nature*, 470, 59–65. Nature Publishing Group.
- MINISTRE DE L'AGRICULTURE ET DE L'ALIMENTATION (2020) Arrêté du 31 juillet 2020 modifiant l'arrêté du 5 juillet 2019 relatif à l'agrément des organismes de sélection des ruminants et des porcins. In p. 24.
- MINISTRE DE L'AGRICULTURE, DE L'AGROALIMENTAIRE ET DE LA FORET (2015) Arrêté du 29 avril 2015 fixant la liste des races des espèces bovine, ovine, caprine et porcine reconnues et précisant les ressources zoogénétiques présentant un intérêt pour la conservation du patrimoine génétique du cheptel et l'aménagement du territoire. In p. 29.
- MINISTRE DES FINANCES ET DES COMPTES PUBLICS & MINISTRE DE L'AGRICULTURE, DE L'AGROALIMENTAIRE ET DE LA FORET (2014) Arrêté du 22 décembre 2014 fixant les mesures

techniques et financières relatives au programme national d'amélioration génétique pour la résistance à la tremblante classique. In p. 86.

MOHAMMADI, A., ALIJANI, S., RAFAT, S.A. & ABDOLLAHI-ARPANAHI, R. (2020) Genome-Wide Association Study and Pathway Analysis for Female Fertility Traits in Iranian Holstein Cattle. *Annals of Animal Science*, 20, 825–851. Sciendo.

MOMMERSTEEG, M.T., YEH, M.L., PARNAVELAS, J.G. & ANDREWS, W.D. (2015) Disrupted Slit-Robo signalling results in membranous ventricular septum defects and bicuspid aortic valves. *Cardiovascular research*, 106, 55–66.

MONTEIRO, A.P.A., TAO, S., THOMPSON, I.M.T. & DAHL, G.E. (2016) In utero heat stress decreases calf survival and performance through the first lactation. *Journal of Dairy Science*, 99, 8443–8450.

MOTA, L.F.M., LOPES, F.B., FERNANDES JÚNIOR, G.A., ROSA, G.J.M., MAGALHÃES, A.F.B., CARVALHEIRO, R. & ALBUQUERQUE, L.G. (2020) Genome-wide scan highlights the role of candidate genes on phenotypic plasticity for age at first calving in Nellore heifers. *Scientific Reports*, 10, 6481. Nature Publishing Group.

MUCHA, S., MRODE, R., COFFEY, M., KIZILASLAN, M., DESIRE, S. & CONINGTON, J. (2018) Genome-wide association study of conformation and milk yield in mixed-breed dairy goats. *Journal of Dairy Science*, 101, 2213–2225.

N

NAMOUS, H., PEÑAGARICANO, F., DEL CORVO, M., CAPRA, E., THOMAS, D.L., STELLA, A., ET AL. (2018) Integrative analysis of methylomic and transcriptomic data in fetal sheep muscle tissues in response to maternal diet during pregnancy. *BMC Genomics*, 19.

NICOGLIOU, A. & MERLIN, F. (2017) Epigenetics: A way to bridge the gap between biological fields. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 66, 73–82.

NIELSEN, R. (2005) Molecular Signatures of Natural Selection. *Annual Review of Genetics*, 39, 197–218.

NIJMAN, I.J., ROSEN, B.D., ZHENG, Z., JIANG, Y., CUMER, T., DALY, K.G., ET AL. (2020) Phylogeny and distribution of Y-chromosomal haplotypes in domestic, ancient and wild goats. preprint, *Genetics*.

O

OGET, C., SERVIN, B. & PALHIÈRE, I. (2019) Genetic diversity analysis of French goat populations reveals selective sweeps involved in their differentiation. *Animal Genetics*, 50, 54–63.

OUELLET, V., LAPORTA, J. & DAHL, G.E. (2020) Late gestation heat stress in dairy cows: Effects on dam and daughter. *Theriogenology*.

P

PAILHOUX, E., VIGIER, B., CHAFFAUX, S., SERVEL, N., TAOURIT, S., FURET, J.P., ET AL. (2001) A 11.7-kb deletion triggers intersexuality and polledness in goats. *Nature Genetics*, 29, 453–458.

- PALHIERE, I., CLEMENT, V., CROUE, I. & PINARD, D. (2015) Un objectif de sélection qui augmente le profit des éleveurs caprins, 5.
- PAN, B., GELEOC, G.S., ASAI, Y., HORWITZ, G.C., KURIMA, K., ISHIKAWA, K., ET AL. (2013) TMC1 and TMC2 Are Components of the Mechanotransduction Channel in Hair Cells of the Mammalian Inner Ear. *Neuron*, 79, 504–515. Elsevier.
- PANT, S.D., SCHENKEL, F.S., VERSCHOOR, C.P., YOU, Q., KELTON, D.F., MOORE, S.S. & KARROW, N.A. (2010) A principal component regression based genome wide analysis approach reveals the presence of a novel QTL on BTA7 for MAP resistance in holstein cattle. *Genomics*, 95, 176–182.
- PASANDIDEH, M., RAHIMI-MIANJI, G. & GHOLIZADEH, M. (2018) A genome scan for quantitative trait loci affecting average daily gain and Kleiber ratio in Baluchi Sheep. *Journal of Genetics*, 97, 493–503.
- PASSAL, J.-N. (2013) 1880-1930 L'aventure des chevriers béarnais sur les routes de France Des Pyrénées en Belgique en passant par Paris.... Site officiel de l'Association de la chèvre de race pyrénéenne. <http://www.chevredespirenees.org/wordpress/wp-content/uploads/2011/09/Chevriers-Bearnais-2013.pdf> [accessed 8 July 2020].
- PAUSCH, H., WURMSER, C., EDEL, E., EMMERLING, R., GÖTZ, K.U. & FRIES, R. (2014) Exploiting Whole Genome Sequence Data for the Identification of Causal Trait Variants in Cattle. Unpublished.
- PAVLIDIS, P., ŽIVKOVIĆ, D., STAMATAKIS, A. & ALACHIOTIS, N. (2013) SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. *Molecular Biology and Evolution*, 30, 2224–2234.
- PEÑAGARICANO, F., WANG, X., ROSA, G.J., RADUNZ, A.E. & KHATIB, H. (2014) Maternal nutrition induces gene expression changes in fetal muscle and adipose tissues in sheep. *BMC Genomics*, 15, 1034.
- PENG, W.-F., XU, S.-S., REN, X., LV, F.-H., XIE, X.-L., ZHAO, Y.-X., ET AL. (2017) A genome-wide association study reveals candidate genes for the supernumerary nipple phenotype in sheep (*Ovis aries*). *Animal Genetics*, 48, 570–579.
- PICKRELL, J.K. & PRITCHARD, J.K. (2012) Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLOS Genetics*, 8, e1002967.
- PETERSON, D.G., MATITASHVILI, E.A. & BAUMAN, D.E. (2003) Diet-Induced Milk Fat Depression in Dairy Cows Results in Increased trans-10, cis-12 CLA in Milk Fat and Coordinate Suppression of mRNA Abundance for Mammary Enzymes Involved in Milk Fat Synthesis. *The Journal of Nutrition*, 133, 3098–3102. Oxford Academic.
- PIERCE, C.F., SPEIDEL, S.E., COLEMAN, S.J., ENNS, R.M., BAILEY, D.W., MEDRANO, J.F., ET AL. (2020) Genome-wide association studies of beef cow terrain-use traits using Bayesian multiple-SNP regression. *Livestock Science*, 232, 103900.
- PIMENTEL, E.C.G., BAUERSACHS, S., TIETZE, M., SIMIANER, H., TETENS, J., THALLER, G., ET AL. (2011) Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling: Relationships between production and fertility at the genomic level. *Animal Genetics*, 42, 251–262.
- PIPERNO, D.R., RANERE, A.J., HOLST, I., IRIARTE, J. & DICKAU, R. (2009) Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Sciences*, 106, 5019–5024.
- POLLARD, S.M., WALLBANK, R., TOMLINSON, S., GROTEWOLD, L. & SMITH, A. (2008) Fibroblast growth factor induces a neural stem cell phenotype in foetal forebrain progenitors and during embryonic stem cell differentiation. *Molecular and Cellular Neuroscience*, 38, 393–403.

- PONSUKSILI, S., MURANI, E., SCHWERIN, M., SCHELLANDER, K., TESFAYE, D. & WIMMERS, K. (2012) Gene Expression and DNA-Methylation of Bovine Pretransfer Endometrium Depending on Its Receptivity after In Vitro-Produced Embryo Transfer. *PLOS ONE*, 7, e42402. Public Library of Science.
- PRASAD, S.S., GARG, A. & AGARWAL, A.K. (2011) Enzymatic activities of the human AGPAT isoform 3 and isoform 5: localization of AGPAT5 to mitochondria. *Journal of Lipid Research*, 52, 451–462.
- PUIG-OLIVERAS, A., BALLESTER, M., COROMINAS, J., REVILLA, M., ESTELLÉ, J., FERNÁNDEZ, A.I., ET AL. (2014) A Co-Association Network Analysis of the Genetic Determination of Pig Conformation, Growth and Fatness. *PLOS ONE*, 9, e114862. Public Library of Science.
- PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M.A.R., BENDER, D., ET AL. (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics*, 81, 559–575.

Q

- QUINLAN, A.R. & HALL, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841–842.

R

- R CORE TEAM (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- RAEYMAEKERS, J.A.M., CHATURVEDI, A., HABLÜTZEL, P.I., VERDONCK, I., HELLEMANS, B., MAES, G.E., ET AL. (2017) Adaptive and non-adaptive divergence in a common landscape. *Nature Communications*, 8, 267. Nature Publishing Group.
- RAMAYO-CALDAS, Y., FORTES, M.R.S., HUDSON, N.J., PORTO-NETO, L.R., BOLORMAA, S., BARENDSE, W., ET AL. (2014) A marker-derived gene network reveals the regulatory role of PPARGC1A, HNF4G, and FOXP3 in intramuscular fat deposition of beef cattle. *Journal of Animal Science*, 92, 2832–2845. Oxford Academic.
- RAMEY, H.R., DECKER, J.E., MCKAY, S.D., ROLF, M.M., SCHNABEL, R.D. & TAYLOR, J.F. (2013) Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics*, 14, 382.
- RAO, C., FOERNZLER, D., LOFTUS, S.K., LIU, S., MCPHERSON, J.D., JUNGERS, K.A., ET AL. (2003) A defect in a novel ADAMTS family member is the cause of the belted white-spotting mutation. *Development*, 130, 4665–4672.
- RAZZAGHI, A., VALIZADEH, R., NASERIAN, A.A., MESGARAN, M.D., CARPENTER, A.J. & GHAFARI, M.H. (2016) Effect of dietary sugar concentration and sunflower seed supplementation on lactation performance, ruminal fermentation, milk fatty acid profile, and blood metabolites of dairy cows. *Journal of Dairy Science*, 99, 3539–3548.
- REBER, I., KELLER, I., BECKER, D., FLURY, C., WELLE, M. & DRÖGEMÜLLER, C. (2015) Wattles in goats are associated with the FMN1/GREM1 region on chromosome 10. *Animal Genetics*, 46, 316–320.
- REINEHR, T. (2011) Thyroid function in the nutritionally obese child and adolescent: Current Opinion in Pediatrics, 23, 415–420.

- REN, J., CHEN, C., MA, J. & HUANG, L.S. (2014) New Strategies Accelerating Progress from QTL to QTN: Example of the Identification of Causal Variants Underlying Three Complex Traits in Pigs. 10th World Congress of Genetics Applied to Livestock Production, Asas.
- RICHARDS, E.J. (2006) Inherited epigenetic variation — revisiting soft inheritance. *Nature Reviews Genetics*, 7, 395–401.
- RICORDEAU, G., BOUILLON, J. & LAJOURS, A. (1967) Hérité des pendeloques en race Saanen. Différences de fécondité entre les génotypes avec et sans pendeloques. *Annales de Zootechnie*, 16, 263–270.
- RICORDEAU, G., MAHE, M.F., AMIGUES, Y., GROSCLAUDE, F. & MANFREDI, E. (1996) Fréquence des allèles de la caséine alpha S1 en race Poitevine. *Animal Genetic Resources Information*, 17, 95–99.
- RINALDI, M., DREESEN, L., HOORENS, P.R., LI, R.W., CLAEREBOUT, E., GODDEERIS, B., ET AL. (2011) Infection with the gastrointestinal nematode *Ostertagia ostertagi* in cattle affects mucus biosynthesis in the abomasum. *Veterinary Research*, 42, 61.
- ROBERTS, A.J. & MCLEAN, D.J. (2011) Differential gene expression in anterior pituitary glands from anestrous and cycling postpartum beef cows. *Journal of Animal Science*, 89, 1035–1041.
- ROBINSON, J.J., ASHWORTH, C.J., ROOKE, J.A., MITCHELL, L.M. & MCEVOY, T.G. (2006) Nutrition and fertility in ruminant livestock. *Animal Feed Science and Technology*, 126, 259–276.
- ROBINSON, M.D., MCCARTHY, D.J. & SMYTH, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139–140.
- ROCHUS, C.M., TORTEREAU, F., PLISSON-PETIT, F., RESTOUX, G., MORENO-ROMIEUX, C., TOSSER-KLOPP, G. & SERVIN, B. (2018) Revealing the selection history of adaptive loci using genome-wide scans for selection: an example from domestic sheep. *BMC Genomics*, 19, 71.
- RUPP, R., FRITZ, S. & BARBAT, A. (2009) La sélection génétique des ruminants laitiers pour des numérations cellulaires faibles et les conséquences sur leur sensibilité aux mammites. In *Journées Nationales des GTV* pp. 935–942. Société Nationale des Groupements Techniques Vétérinaires (SNGTV). FRA., Nantes, France.
- RUPP, R., SENIN, P., SARRY, J., ALLAIN, C., TASCA, C., LIGAT, L., ET AL. (2015) A Point Mutation in Suppressor of Cytokine Signalling 2 (*Socs2*) Increases the Susceptibility to Inflammation of the Mammary Gland while Associated with Higher Body Weight and Size and Higher Milk Production in a Sheep Model. *PLOS Genetics*, 11, e1005629.
- RUSSO, V., FONTANESI, L., DAVOLI, R., COSTA, L.N., CAGNAZZO, M., BUTTAZZONI, L., ET AL. (2002) Investigation of candidate genes for meat quality in dry-cured ham production: the porcine cathepsin B (CTSB) and cystatin B (CSTB) genes. *Animal Genetics*, 33, 123–131.

S

- SAFARI, E., FOGARTY, N.M. & GILMOUR, A.R. (2005) A review of genetic parameter estimates for wool, growth, meat and reproduction traits in sheep. *Livestock Production Science*, 92, 271–289.
- SALLEH, M.S., MAZZONI, G., HÖGLUND, J.K., OLIJHOEK, D.W., LUND, P., LØVENDAHL, P. & KADARMIDEEN, H.N. (2017) RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high- and low-residual feed intake in Nordic dairy cattle. *BMC Genomics*, 18, 258.

- SALO, A.M., COX, H., FARNDON, P., MOSS, C., GRINDULIS, H., RISTELI, M., ET AL. (2008) A Connective Tissue Disorder Caused by Mutations of the Lysyl Hydroxylase 3 Gene. *The American Journal of Human Genetics*, 83, 495–503.
- SANCHEZ, M.-P., GOVIGNON-GION, A., CROISEAU, P., FRITZ, S., HOZÉ, C., MIRANDA, G., ET AL. (2017) Within-breed and multi-breed GWAS on imputed whole-genome sequence variants reveal candidate mutations affecting milk protein composition in dairy cattle. *Genetics Selection Evolution*, 49, 68.
- SANCHEZ-JUAN, P., BISHOP, M.T., AULCHENKO, Y.S., BRANDEL, J.-P., RIVADENEIRA, F., STRUCHALIN, M., ET AL. (2012) Genome-wide study links MTMR7 gene to variant Creutzfeldt-Jakob risk. *Neurobiology of Aging*, 33, 1487.e21-1487.e28.
- SANSREGRET, L. & NEPVEU, A. (2008) The multiple roles of CUX1: Insights from mouse models and cell-based assays. *Gene*, 412, 84–94.
- SAVOLAINEN, O., LASCOUX, M. & MERILÄ, J. (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics*, 14, 807–820.
- SCHEET, P. & STEPHENS, M. (2006) A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. *The American Journal of Human Genetics*, 78, 629–644.
- SCHOLTENS, M., JIANG, A., SMITH, A., LITTLEJOHN, M., LEHNERT, K., SNELL, R., ET AL. (2020) Genome-wide association studies of lactation yields of milk, fat, protein and somatic cell score in New Zealand dairy goats. *Journal of Animal Science and Biotechnology*, 11, 55.
- SCHULMAN, N.F., SAHANA, G., ISO-TOURU, T., LUND, M.S., ANDERSSON-EKLUND, L., VIITALA, S.M., ET AL. (2009) Fine mapping of quantitative trait loci for mastitis resistance on bovine chromosome 11. *Animal Genetics*, 40, 509–515.
- SEABURY, C.M., OLDESCHULTE, D.L., SAATCHI, M., BEEVER, J.E., DECKER, J.E., HALLEY, Y.A., ET AL. (2017) Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics*, 18, 386.
- SEEGERS, H., FOURICHON, C. & BEAUDEAU, F. (2003) Production effects related to mastitis and mastitis economics in dairy cattle herds. *Veterinary Research*, 34, 475–491.
- SEJIAN, V., MAURYA, V.P. & NAQVI, S.M.K. (2010) Adaptability and growth of Malpura ewes subjected to thermal and nutritional stress. *Tropical Animal Health and Production*, 42, 1763–1770.
- SERÃO, N.V.L., GONZÁLEZ-PEÑA, D., BEEVER, J.E., BOLLERO, G.A., SOUTHEY, B.R., FAULKNER, D.B. & RODRIGUEZ-ZAS, S.L. (2013) Bivariate Genome-Wide Association Analysis of the Growth and Intake Components of Feed Efficiency. *PLoS ONE*, 8, e78530.
- SEVANE, N., MARTÍNEZ, R. & BRUFORD, M.W. (2018) Genome-wide differential DNA methylation in tropically adapted Creole cattle and their Iberian ancestors. *Animal Genetics*, 50, 15–26.
- SHIN, D., LEE, C., PARK, K.-D., KIM, H. & CHO, K. (2017) Genome-association analysis of Korean Holstein milk traits using genomic estimated breeding value. *Asian-Australasian Journal of Animal Sciences*, 30, 309–319.
- SHONGJIA, L., XINAGMO, L., GANGY, X. & SHENOV, C. (1992) Relationship between physical traits, litter size and milk yield in Saanen, Guanzhong and crossbred goats. In *Proceedings of the International Conference on Goats* p. 83. Book of Abstracts New Delhi, India.
- SILANIKOVE, N. (2000) Effects of heat stress on the welfare of extensively managed domestic ruminants. *Livestock Production Science*, 67, 1–18.

- SKAPETAS, B. & BAMPIDIS, V. (2016) Goat production in the World: present situation and trends. *Livestock Research for Rural Development*, 28, 200.
- SKIBIEL, A. L., DADO-SENN, B., FABRIS, T.F., DAHL, G.E. & LAPORTA, J. (2018) In utero exposure to thermal stress has long-term effects on mammary gland microstructure and function in dairy cattle. *PLOS ONE*, 13, e0206046.
- SKIBIEL, A.L., PEÑAGARICANO, F., AMORÍN, R., AHMED, B.M., DAHL, G.E. & LAPORTA, J. (2018) In Utero Heat Stress Alters the Offspring Epigenome. *Scientific Reports*, 8.
- SLOT, K.A., VOORENDT, M., BOER-BROUWER, M. DE, VUGT, H.H. VAN & TEERDS, K.J. (2006) Estrous cycle dependent changes in expression and distribution of Fas, Fas ligand, Bcl-2, Bax, and pro- and active caspase-3 in the rat ovary. *Journal of Endocrinology*, 188, 179–192. BioScientifica.
- SOSINOWSKI, T., PANDEY, A., DIXIT, V.M. & WEISS, A. (2000) Src-like Adaptor Protein (Slap) Is a Negative Regulator of T Cell Receptor Signaling. *Journal of Experimental Medicine*, 191, 463–474.
- STELLA, A., NICOLAZZI, E.L., VAN TASSELL, C.P., ROTHSCILD, M.F., COLLI, L., ROSEN, B.D., ET AL. (2018) AdaptMap: exploring goat diversity and adaptation. *Genetics Selection Evolution*, 50.
- STOREY, J.D., BASS, A.J., DABNEY, A. & ROBINSON, D. (2015) qvalue: Q-value estimation for false discovery rate control. R package version 2.8.0. <http://github.com/jdstorey/qvalue>.
- SUGIMOTO, M., FUJIKAWA, A., WOMACK, J.E. & SUGIMOTO, Y. (2006) Evidence that bovine forebrain embryonic zinc finger-like gene influences immune response associated with mastitis resistance. *Proceedings of the National Academy of Sciences*, 103, 6454–6459. National Academy of Sciences.
- SUN, L.S., DONG, X., FAN, B. & LIU, B. (2011) The Association of ANKRD2 with Loin Depth and Muscle Firmness in Pigs. *Journal of Animal and Veterinary Advances*, 10, 1462–1468.
- SUZUKI, M.M. & BIRD, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics*, 9, 465–476.
- SUZUKI, S., ONO, R., NARITA, T., PASK, A.J., SHAW, G., WANG, C., ET AL. (2007) Retrotransposon Silencing by DNA Methylation Can Drive Mammalian Genomic Imprinting. *PLOS Genetics*, 3, e55.
- SWITONSKI, M., MANKOWSKA, M. & SALAMON, S. (2013) Family of melanocortin receptor (MCR) genes in mammals—mutations, polymorphisms and phenotypic effects. *Journal of Applied Genetics*, 54, 461–472.
- SZATKOWSKA, I., ZABORSKI, D., PROSKURA, W.S. & TABOR, S. (2014) Polledness intersex syndrome in goats – molecular and histological aspects. *TURKISH JOURNAL OF VETERINARY AND ANIMAL SCIENCES*, 38, 612–617.

T

- TABERLET, P., VALENTINI, A., REZAEI, H.R., NADERI, S., POMPANON, F., NEGRINI, R. & AJMONE-MARSAN, P. (2008) Are cattle, sheep, and goats endangered species? *Molecular Ecology*, 17, 275–284.
- TAKEUCHI, K. & REUE, K. (2009) Biochemistry, physiology, and genetics of GPAT, AGPAT, and lipin enzymes in triglyceride synthesis. *American Journal of Physiology-Endocrinology and Metabolism*, 296, E1195–E1209.

- TALEBI, R., AHMADI, A. & AFRAZ, F. (2018) Analysis of protein-protein interaction network based on transcriptome profiling of ovine granulosa cells identifies candidate genes in cyclic recruitment of ovarian follicles. *Journal of Animal Science and Technology*, 60, 11.
- TALENTI, A., PALHIÈRE, I., TORTEREAU, F., PAGNACCO, G., STELLA, A., NICOLAZZI, E.L., ET AL. (2018) Functional SNP panel for parentage assessment and assignment in worldwide goat breeds. *Genetics Selection Evolution*, 50, 55.
- TALOUARN, E., BARDOU, P., PALHIÈRE, I., OGET, C., CLEMENT, V., TOSSER-KLOPP, G., ET AL. (2020) Genome wide association analysis on semen volume and milk yield using different strategies of imputation to whole genome sequence in French dairy goats. *BMC Genetics*, 21, 19.
- TAO, S., BUBOLZ, J.W., DO AMARAL, B.C., THOMPSON, I.M., HAYEN, M.J., JOHNSON, S.E. & DAHL, G.E. (2011) Effect of heat stress during the dry period on mammary gland development. *Journal of Dairy Science*, 94, 5976–5986.
- TATTINI, L., D'AURIZIO, R. & MAGI, A. (2015) Detection of Genomic Structural Variants from Next-Generation Sequencing Data. *Frontiers in Bioengineering and Biotechnology*, 3. Frontiers.
- TIENSUU, H., HAAPALAINEN, A.M., KARJALAINEN, M.K., PASANEN, A., HUUSKO, J.M., MARTTILA, R., ET AL. (2019) Risk of spontaneous preterm birth and fetal growth associates with fetal SLIT2. *PLOS Genetics*, 15, e1008107. Public Library of Science.
- TIZIOTO, P.C., TAYLOR, J.F., DECKER, J.E., GROMBONI, C.F., MUDADU, M.A., SCHNABEL, R.D., ET AL. (2015) Detection of quantitative trait loci for mineral content of Nelore longissimus dorsi muscle. *Genetics Selection Evolution*, 47, 15.
- TOSSER-KLOPP, G., BARDOU, P., BOUCHEZ, O., CABAU, C., CROOIJMANS, R., DONG, Y., ET AL. (2014) Design and characterization of a 52K SNP chip for goats. *PloS one*, 9, e86227.
- TRAORE, A., TAMBOURA, H.H., KABORE, A., ROYO, L.J., FERNANDEZ, I., ÁLVAREZ, I., ET AL. (2008) Multivariate characterization of morphological traits in Burkina Faso sheep. *Small Ruminant Research*, 80, 62–67.

V

- VALLÉE, M., BEAUDRY, D., ROBERGE, C., MATTE, J.J., BLOUIN, R. & PALIN, M.-F. (2003) Isolation of Differentially Expressed Genes in Conceptuses and Endometrial Tissue of Sows in Early Gestation. *Biology of Reproduction*, 69, 1697–1706. Oxford Academic.
- VAYSSE, A., RATNAKUMAR, A., DERRIEN, T., AXELSSON, E., PIELBERG, G.R., SIGURDSSON, S., ET AL. (2011) Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds using Selection Mapping. *PLOS Genetics*, 7, e1002316. Public Library of Science.
- VERRIER, E., AUDIOT, A., BERTRAND, C., CHAPUIS, H., CHARVOLIN, E., DANCHIN-BURGE, C., ET AL. (2015) Assessing the risk status of livestock breeds: a multi-indicator method applied to 178 French local breeds belonging to ten species. *Animal Genetic Resources*, 57, 105–118.
- VIGNE, J.D. (1988) Les grandes étapes de la domestication de la chèvre : une proposition d'explication de son statut en Europe occidentale. *Société d'Ethnozootechnie*, 41, 1–13.
- VINATIER ROCHE, C. (undated) La chèvre poitevine : son histoire et sa renaissance. Association de Défense et de Développement de la Chèvre Poitevine. <http://www.chevre-poitevine.org/wp-content/uploads/2010/12/HistoireEtRenaissanceDeLaPoitevine.pdf> [accessed 12 October 2020].



- WAGGOTT, D., HAIDER, S. & BOUTROS, P.C. (2017) bedr: Genomic Region Processing using Tools Such as 'BEDTools', 'BEDOPS' and 'Tabix'. R package version 1.0.4. <https://CRAN.R-project.org/package=bedr>.
- WANG, G., DRIEL, B.J. VAN, LIAO, G., O'KEEFFE, M.S., HALIBOZEK, P.J., FLIPSE, J., ET AL. (2015) Migration of Myeloid Cells during Inflammation Is Differentially Regulated by the Cell Surface Receptors Slamf1 and Slamf8. *PLOS ONE*, 10, e0121968. Public Library of Science.
- WANG, K.H., BROSE, K., ARNOTT, D., KIDD, T., GOODMAN, C.S., HENZEL, W. & TESSIER-LAVIGNE, M. (1999) Biochemical Purification of a Mammalian Slit Protein as a Positive Regulator of Sensory Axon Elongation and Branching. *Cell*, 96, 771–784.
- WANG, M., YAN, J., ZHAO, J., SONG, W., ZHANG, X., XIAO, Y. & ZHENG, Y. (2012) Genome-wide association study (GWAS) of resistance to head smut in maize. *Plant Science*, 196, 125–131.
- WANG, X., DING, R., QUAN, J., YANG, L., YANG, M., ZHENG, E., ET AL. (2017) Genome-wide association analysis reveals genetic loci and candidate genes associated with intramuscular fat in Duroc pigs. *Frontiers of Agricultural Science and Engineering*, 4, 335–341.
- WANG, X., LAN, X., RADUNZ, A.E. & KHATIB, H. (2015) Maternal nutrition during pregnancy is associated with differential expression of imprinted genes and DNA methyltransferases in muscle of beef cattle offspring. *Journal of Animal Science*, 93, 35–40.
- WEI, C., WANG, H., LIU, G., WU, M., CAO, J., LIU, Z., ET AL. (2015) Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. *BMC Genomics*, 16, 194.
- WEI, C.-B., WANG, J.-Q., CHEN, F.-Y., NIU, H. & LI, K. (2015) DNA sequence polymorphism within the bovine adenosine monophosphate deaminase 1 (AMPD1) is associated with production traits in Chinese cattle. *Genetics and Molecular Research*, 14, 1025–1033.
- WEIR, B.S. & COCKERHAM, C.C. (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38, 1358–1370.
- WIDMANN, P., REVERTER, A., FORTES, M.R.S., WEIKARD, R., SUHRE, K., HAMMON, H., ET AL. (2013) A systems biology approach using metabolomic data reveals genes and pathways interacting to modulate divergent growth in cattle. *BMC Genomics*, 14, 798.
- WIEDEMANN, S., SIGL, G., SCHMAUTZ, C., KASKE, M., VITURRO, E. & MEYER, H.H.D. (2013) Omission of dry period or milking once daily affects metabolic status and is reflected by mRNA levels of enzymes in liver and muscle of dairy cows. *Livestock Science*, 154, 193–203.
- WIEDEMAR, N. & DRÖGEMÜLLER, C. (2015) A 1.8-kb insertion in the 3'-UTR of RXFP2 is associated with polledness in sheep. *Animal Genetics*, 46, 457–461.
- WILSON, D.J., GONZALEZ, R.N. & DAS, H.H. (1997) Bovine Mastitis Pathogens in New York and Pennsylvania: Prevalence and Effects on Somatic Cell Count and Milk Production. *Journal of Dairy Science*, 80, 2592–2598.
- WOLC, A., ARANGO, J., JANKOWSKI, T., DUNN, I., SETTAR, P., FULTON, J.E., ET AL. (2014) Genome-wide association study for egg production and quality in layer chickens. *Journal of Animal Breeding and Genetics*, 131, 173–182.
- WRIGHT, S. (1931) Evolution in Mendelian Populations. *Genetics*, 16, 97–159.

WU, H., GUANG, X., AL-FAGEEH, M.B., CAO, J., PAN, S., ZHOU, H., ET AL. (2014) Camelid genomes reveal evolution and adaptation to desert environments. *Nature Communications*, 5.

X

XIA, J., QI, X., WU, Y., ZHU, B., XU, L., ZHANG, L., ET AL. (2016) Genome-wide association study identifies loci and candidate genes for meat quality traits in Simmental beef cattle. *Mammalian Genome*, 27, 246–255.

XIANG, R., MACLEOD, I.M., BOLORMAA, S. & GODDARD, M.E. (2017) Genome-wide comparative analyses of correlated and uncorrelated phenotypes identify major pleiotropic variants in dairy cattle. *Scientific Reports*, 7, 9248. Nature Publishing Group.

Y

YAMAGUCHI, T.P., BRADLEY, A., MCMAHON, A.P. & JONES, S. (1999) A Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. *Development*, 126, 1211–1223.

YANG, J., BENYAMIN, B., MCEVOY, B.P., GORDON, S., HENDERS, A.K., NYHOLT, D.R., ET AL. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, 42, 565–569.

YANG, J., LI, W.-R., LV, F.-H., HE, S.-G., TIAN, S.-L., PENG, W.-F., ET AL. (2016) Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. *Molecular Biology and Evolution*, 33, 2576–2592.

YANG, JI, LI, W.-R., LV, F.-H., HE, S.-G., TIAN, S.-L., PENG, W.-F., ET AL. (2016) Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. *Molecular Biology and Evolution*, 33, 2576–2592.

YANG, JIE, JIANG, J., LIU, X., WANG, H., GUO, G., ZHANG, Q. & JIANG, L. (2016) Differential expression of genes in milk of dairy cattle during lactation. *Animal Genetics*, 47, 174–180.

YUAN, S.-M. (2015) α -Smooth Muscle Actin and ACTA2 Gene Expressions in Vasculopathies. *Brazilian Journal of Cardiovascular Surgery*, 30, 644–649. *Revista Brasileira de Cirurgia Cardiovascular*.

YUAN, Z., LI, W., LI, F. & YUE, X. (2019) Selection signature analysis reveals genes underlying sheep milking performance. *Archives Animal Breeding*, 62, 501–508.

Z

ZEDER, M. (2017) Out of the fertile crescent: The dispersal of domestic livestock through Europe and Africa. In *Human Dispersal and Species Movement: From Prehistory to the Present* (eds N. Boivin, R. Crassard & M. Petraglia), pp. 261–303. Cambridge University Press, Cambridge, UK.

ZEDER, M.A. & HESSE, B. (2000) The Initial Domestication of Goats (*Capra hircus*) in the Zagros Mountains 10,000 Years Ago. *Science*, 287, 2254–2257.

ZEDER, M.A. (2008) Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences*, 105, 11597–11604.

- ZHANG, X., ZHANG, S., MA, L., JIANG, E., XU, H., CHEN, R., ET AL. (2017) Reduced representation bisulfite sequencing (RRBS) of dairy goat mammary glands reveals DNA methylation profiles of integrated genome-wide and critical milk-related genes. *Oncotarget*, 8, 115326–115344.
- ZHAO, C., GUI, L., LI, Y., PLATH, M. & ZAN, L. (2015) Associations between allelic polymorphism of the BMP Binding Endothelial Regulator and phenotypic variation of cattle. *Molecular and Cellular Probes*, 29, 358–364.
- ZHENG, Z., WANG, XIHONG, LI, M., LI, Y., YANG, Z., WANG, XIAOLONG, ET AL. (2020) The origin of domestication genes in goats. *Science Advances*, 6, eaaz5216.
- ZHOU, D., WEI, Z., DENG, S., WANG, T., ZAI, M., WANG, H., ET AL. (2013) SASH1 regulates melanocyte transepithelial migration through a novel Gαs–SASH1–IQGAP1–E-Cadherin dependent pathway. *Cellular Signalling*, 25, 1526–1538.
- ZHOU, X. & STEPHENS, M. (2012) Genome-wide Efficient Mixed Model Analysis for Association Studies. *Nature genetics*, 44, 821–824.
- ZHOU, X., LINDSAY, H. & ROBINSON, M.D. (2014) Robustly detecting differential expression in RNA sequencing data using observation weights. *Nucleic Acids Research*, 42, e91.
- ZOLDOŠ, V., BIRUŠ, I., MURATOVIĆ, E., ŠATOVIĆ, Z., VOJTA, A., ROBIN, O., ET AL. (2018) Epigenetic Differentiation of Natural Populations of *Lilium bosniacum* Associated with Contrasting Habitat Conditions. *Genome Biology and Evolution*, 10, 291–303.

ANNEXES

I. Mécanismes génétiques et sélection

Diversité neutre

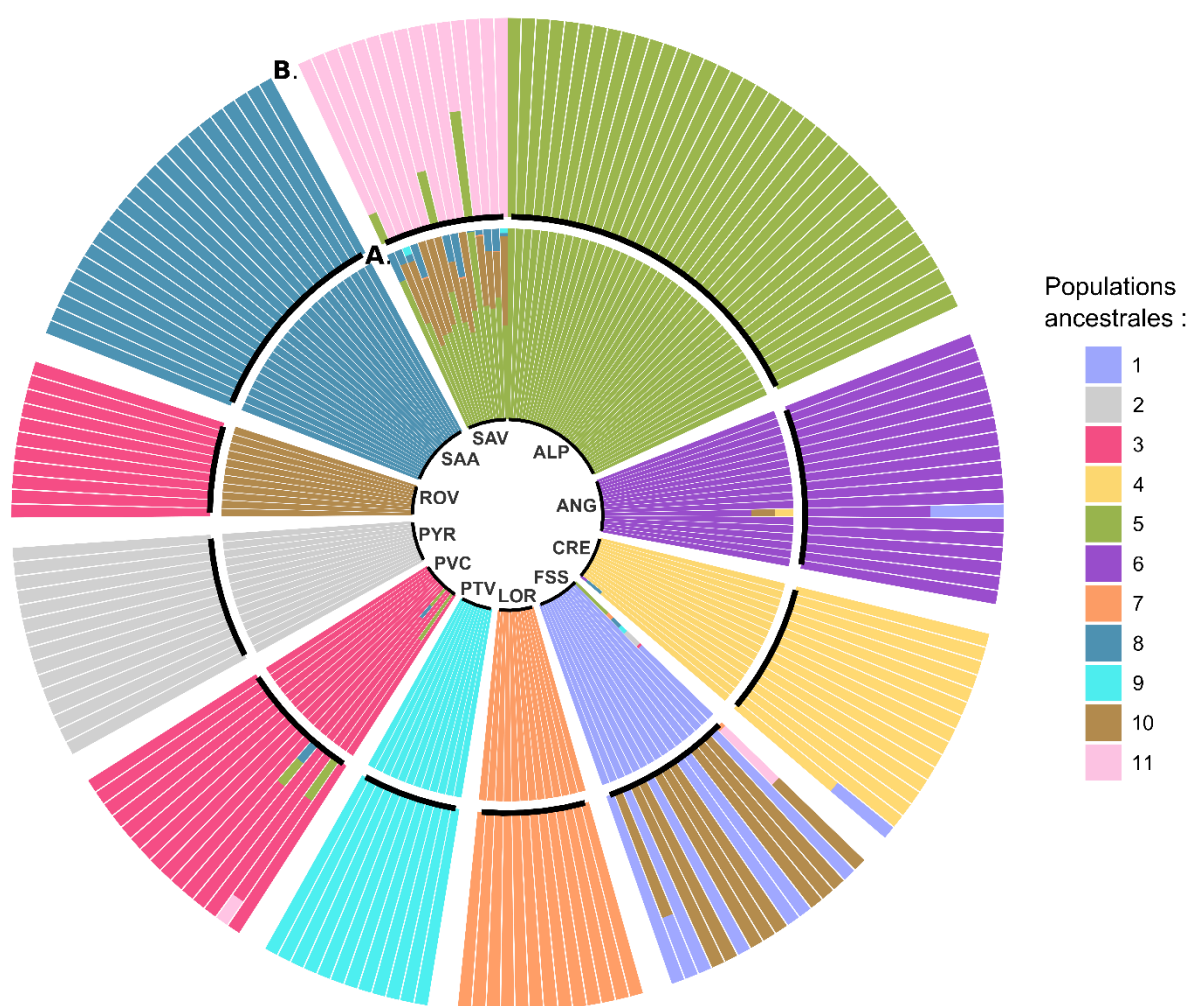


Figure S1 : Analyse de structure avec les 195 chèvres françaises non apparentées (chacune représentée par une barre) afin d'éliminer les animaux non représentatif de leur race. Chaque cercle correspond à une analyse avec un nouveau nombre de population ancestrale (K) : A. K = 10, et B. K = 11. Chaque couleur correspond à une population ancestrale.

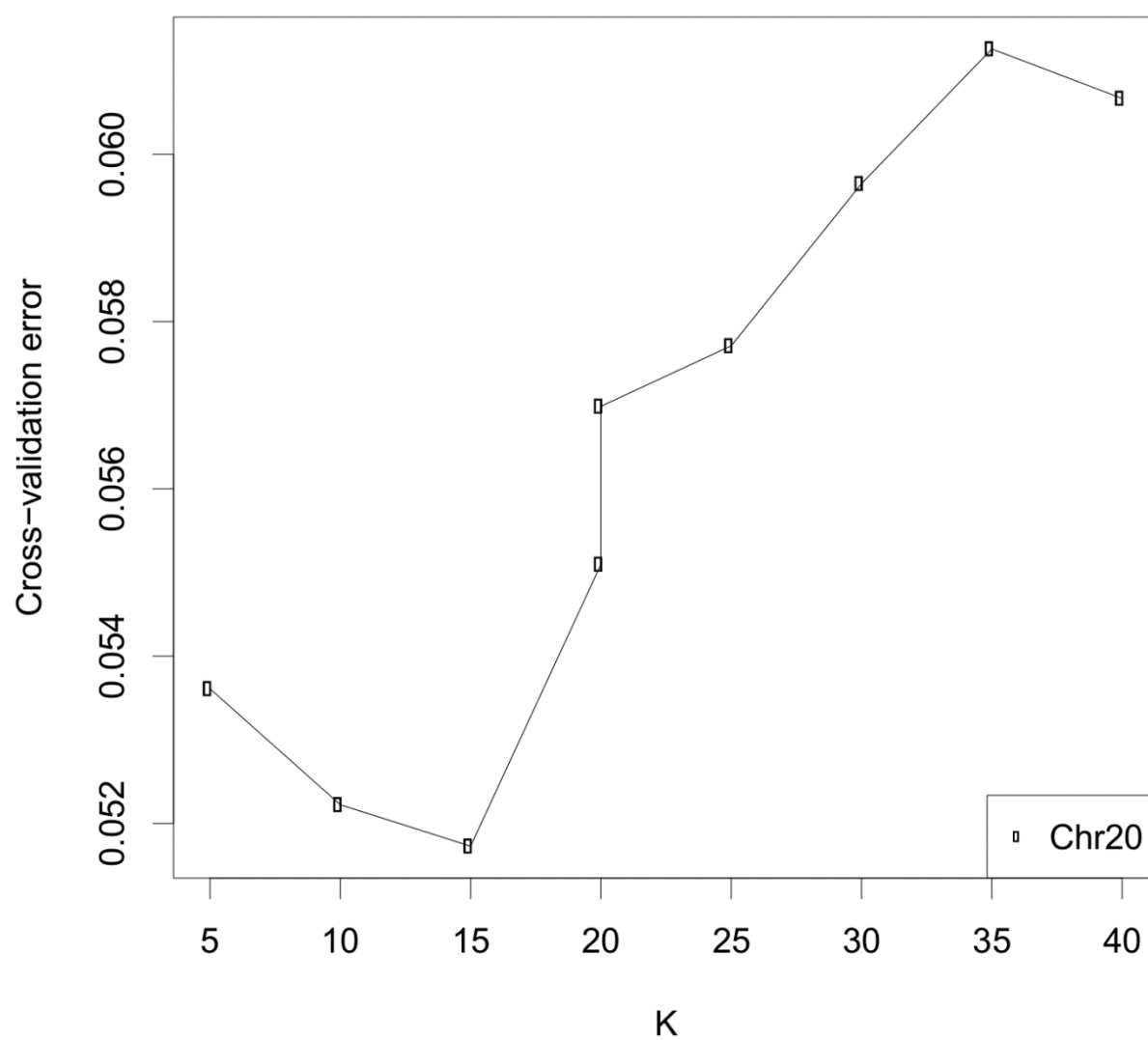


Figure S2 : Étape de validation croisée du logiciel fastPHASE afin d'estimer le nombre d'haplotypes présents dans notre jeu de données sur les chèvres françaises

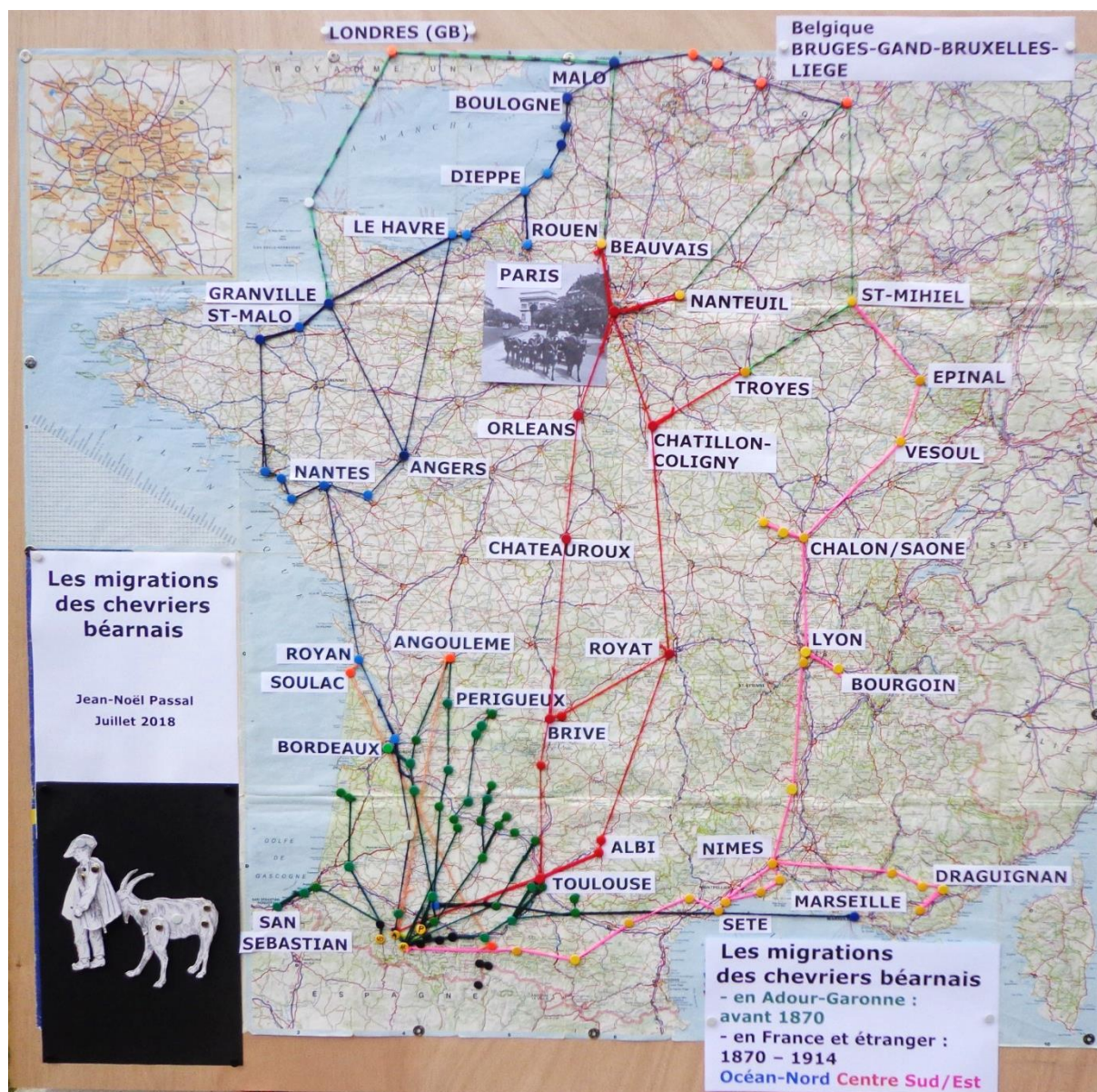


Figure S3 : Carte représentant les différentes voies de migration des chevriers béarnais avec leurs chèvres des Pyrénées (© Jean-Noël Passal, 2018).

A.**II. Résultats de l'inventaire 2017***A- Effectifs totaux*

118 éleveurs ont été recensés comme possédant un cheptel de chèvres du Rove et faisant reproduire tout ou une partie du troupeau en race pure. Deux listes référençant les éleveurs et leurs coordonnées et les éleveurs et leur principales caractéristiques d'élevage figurent en fin du document (parties IV et V).

Au total il a été recensé : **9 398 animaux dont 8 904 femelles et 494 mâles.**

B.**2- Type de production**

Tableau 6: Répartition des élevages de chèvre du Rove en allaitant ou fromager

		Nbre éleveurs	% éleveurs	% éleveurs 2013	Nbre femelles	% femelles	Nbre mâles	% mâles	Taille cheptel femelle
Allaitants	Cabris/débroussaillage	23			1 431		58		
	Ovins+Cabris	34			1 327		132		
	Reproducteurs	6			271		34		
	TOTAL	63	53%	60%	3 029	34%	224	45%	48
Fromagers	Fromage+ Cabris	28			3 231		164		
	Fromage	27			2 644		111		
	TOTAL	55	47%	40%	5 875	66%	275	55%	102

Figure S4 : Extraits du Répertoire 2017 des éleveurs de la Race caprine du Rove (document de travail de l'IDELE). **A.** Inventaires des effectifs des chèvres et boucs de la race en et **B.** Répartition des chèvres du Rove entre les élevages allaitants et fromagers.

Bilan Inventaire 2018 de la chèvre des Savoie

I. Paramètres démographiques

Tableau 1: Evolution des effectifs du répertoire (catégorie génétique A, B, C = 25% autres races maximum)

	2004	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Nombre de mâles	51	77	77	70	55	63	58	62	87	95	63	74	69	55
Nombre de femelles	509	757	738	728	896	766	806	793	1024	1044	957	1003	1001	871
Nombre d'éleveurs	39	39	32	24	24	28	31	34	39	39	35	33	30	24
Taille moyenne des troupeaux	13	21	25	30	37	28	26	24	26	27	27	28,5	33	36

Figure S5 : Extraits de l'Inventaire 2018 de la chèvre de Savoie (document de travail de l'IDELE) représentant les effectifs des chèvres et boucs de la race.

Signatures de sélection

Analyse globale

Tableau S3 : Présentation de toutes les régions détectées sous sélection chez les races caprines métropolitaines par le programme hapFLK ainsi que les gènes et les races associés à celles-ci.

Chromosome	Début région	Fin région	Début pic	Fin pic	Gènes	Races
1	114323058	114683573	114368000	114449500	NA	ALP LOR PTV SAA
			114496500	114631500	P2RY1 LOC106502324	ALP LOR PTV PVC ROV
4	83615398	83733021	83665500	83684000	NA	ALP PTV SAA SAV
4	87150587	87251884	87189300	87217500	DMTF1 LOC102179731	ALP PTV SAV
6	14007660	14164400	14059000	14114600	NA	FSS PYR
6	51554424	51690957	51571800	51614000	NA	ALP LOR PTV SAV
			51614000	51644600	NA	LOR PTV SAA
6	63279790	63400033	63329000	63373500	NA	LOR
6	83260323	83381550	83309900	83337000	NA	ALP LOR PTV PYR ROV
6	85597941	85702865	85645000	85655000	LOC102183682	ALP LOR PTV PYR SAA SAV
6	85731399	85839941	85778600	85814000	SULT1B1 LOC102178524	ALP LOR PTV PYR
6	90829178	90930186	90866000	90893500	RCHY1 THAP6	FSS LOR PTV
6	91759972	91865741	91804000	91829000	CCDC158 SHROOM3	LOR PTV
6	109745976	109850921	109795000	109808000	NA	LOR PYR
9	79229134	79333095	79278800	79301900	NOX3	ALP PYR SAV
10	27568543	27675459	27600000	27628300	RHOJ	LOR PTV SAV
11	6995989	7099330	7043000	7052500	IL18R1	ALP LOR PTV SAA
13	57062336	57169248	57081000	57126000	NA	FSS LOR
13	59467159	59607028	59512000	59583000	ANGPT4 FAM110A LOC108637404	SAA SAV
					SLC52A3	
13	61217573	61350799	61260000	61302000	LOC102190706 NOL4L	LOR SAA SAV
13	61974350	62082904	62017000	62047000	NA	FSS PYR SAV

			62258000	62319000	BPIFA1 BPIFA3	LOR PTV ROV
13	62210588	62568730	62382000	62461500	LOC102187691 LOC108637413 LOC102171893 CDK5RAP1	FSS LOR PTV ROV SAA
			62461500	62514500	SNTA1	FSS PTV ROV
			62881000	62925000	TRNAG-UCC	ALP LOR PTV SAA
			62945000	63160000	LOC108637417 RALY EIF2S2 LOC108637418	ALP LOR PTV SAA SAV
13	62833529	63539347	63160000	63288000	LOC102190531 ASIP AHCY	ALP PTV SAA SAV
			63288000	63400000	ITCH	LOR PTV ROV SAA SAV
			63400000	63525000	DYNLRB1 MAP1LC3A PIGU	ALP FSS PTV ROV
13	64000715	64228651	64051000	64099000	MMP24	FSS LOR PTV PVC SAV
			64141000	64180000	FAM83C LOC102172278	PTV ROV SAA
13	64376554	64483576	64394000	64433000	LOC102173760 SPAG4 CPNE1	LOR PTV
13	64660439	64892421	64708000	64737000	CNBD2	FSS LOR PTV SAA SAV
			64776000	64845000	EPB41L1	ALP FSS LOR PTV SAA SAV
13	65094529	65199534	65132000	65151000	DLGAP4	ALP FSS LOR PTV ROV SAA SAV
13	66093183	66217223	66139000	66168000	NA	FSS LOR PTV PYR SAA SAV
18	33766811	33868720	33810000	33823000	NA	ALP FSS LOR PTV PYR SAA SAV
19	19341398	19442589	19390000	19406000	NA	FSS LOR PYR SAA
20	70334783	70435949	70382000	70400000	NA	PTV PVC SAA

■ Région 1

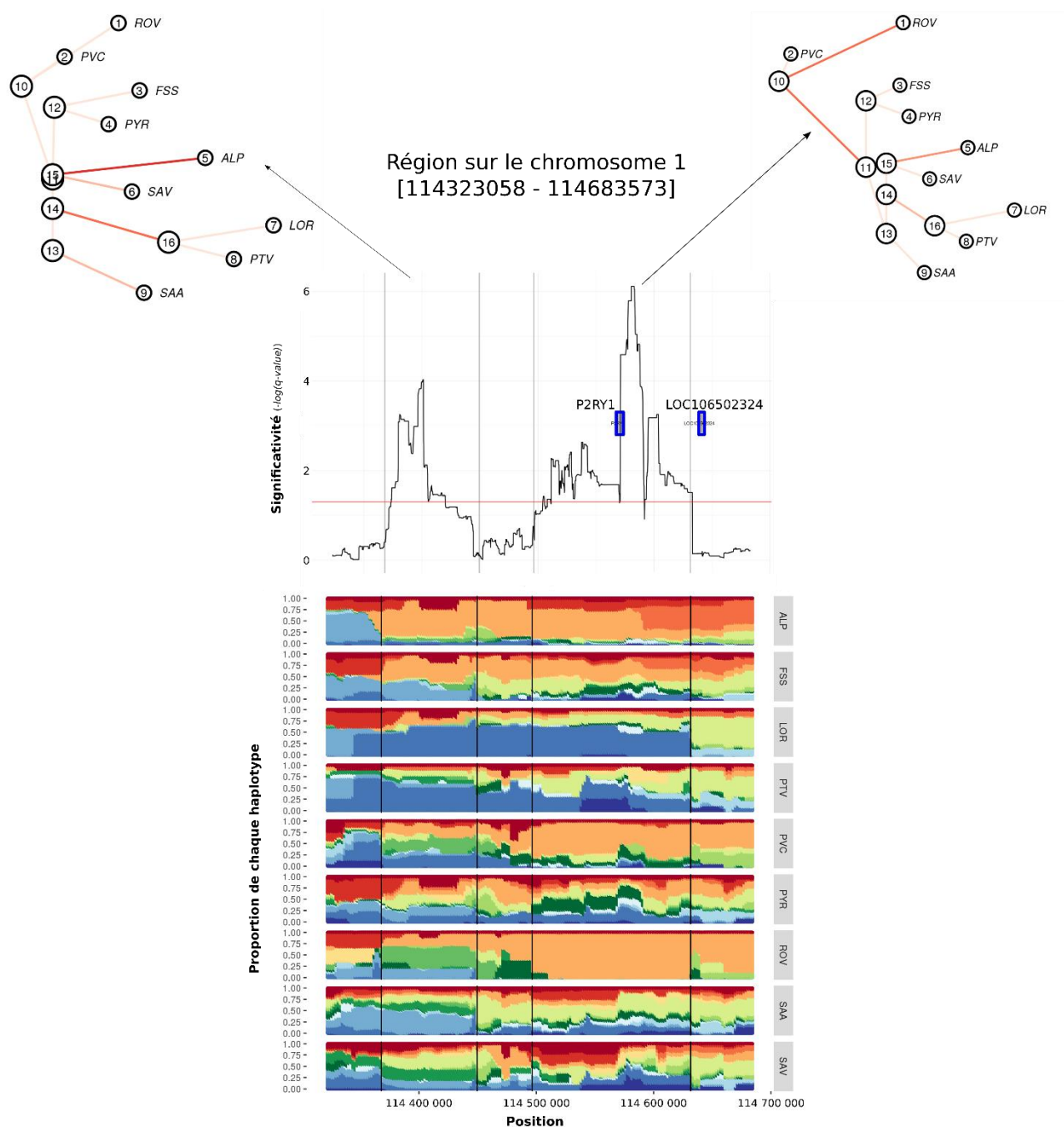


Figure S6 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 2

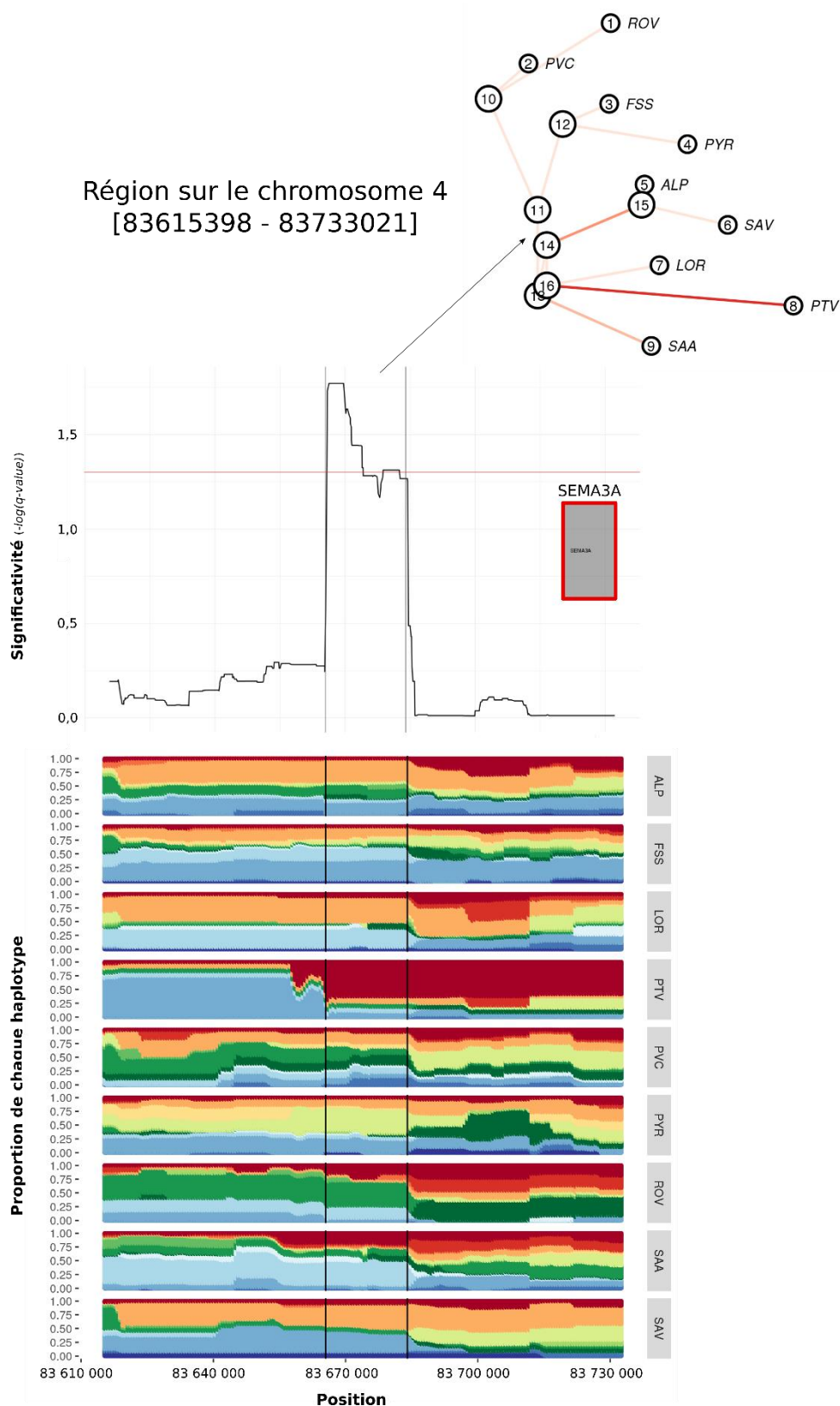


Figure S7 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

- Région 3

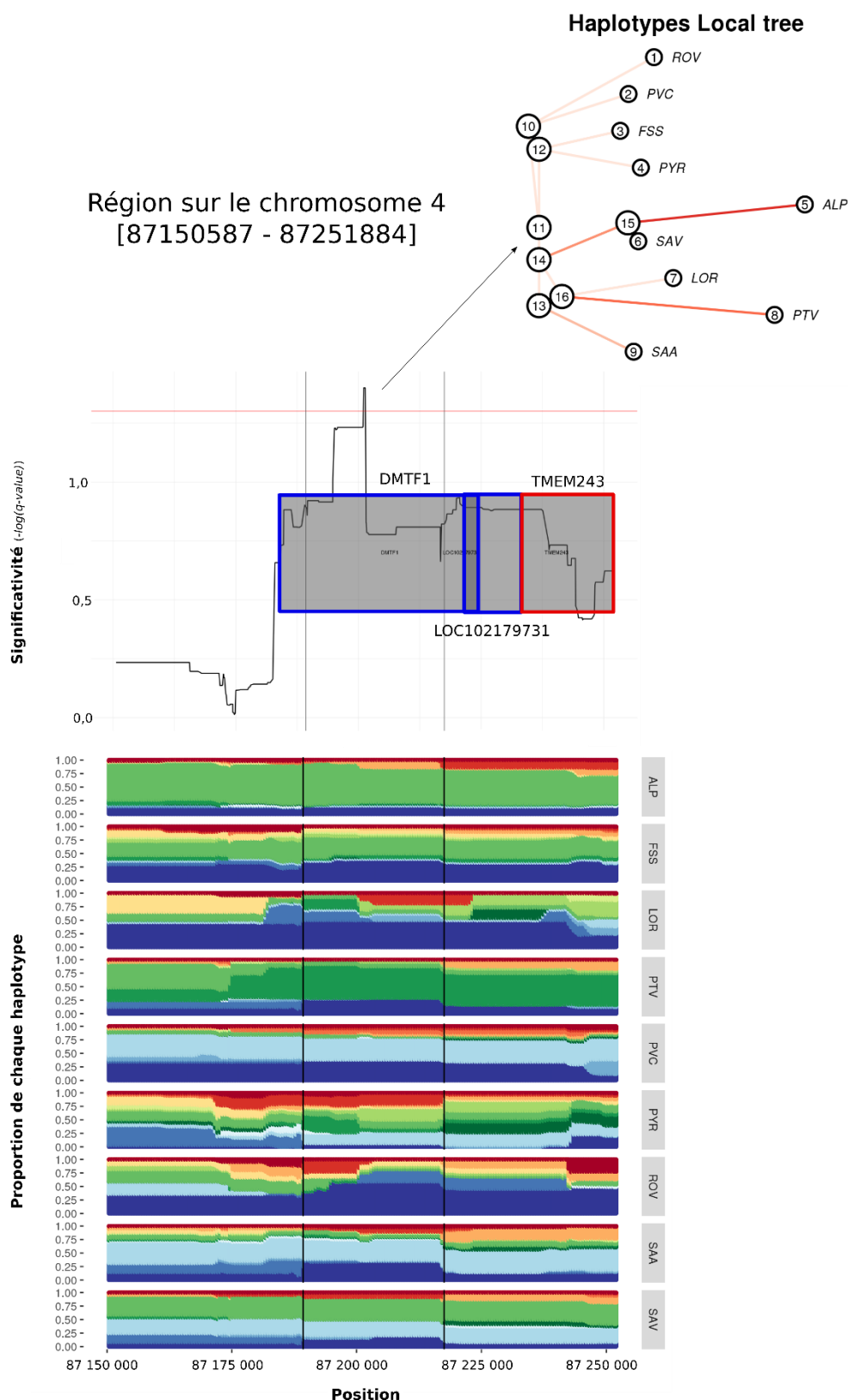


Figure S8 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 4

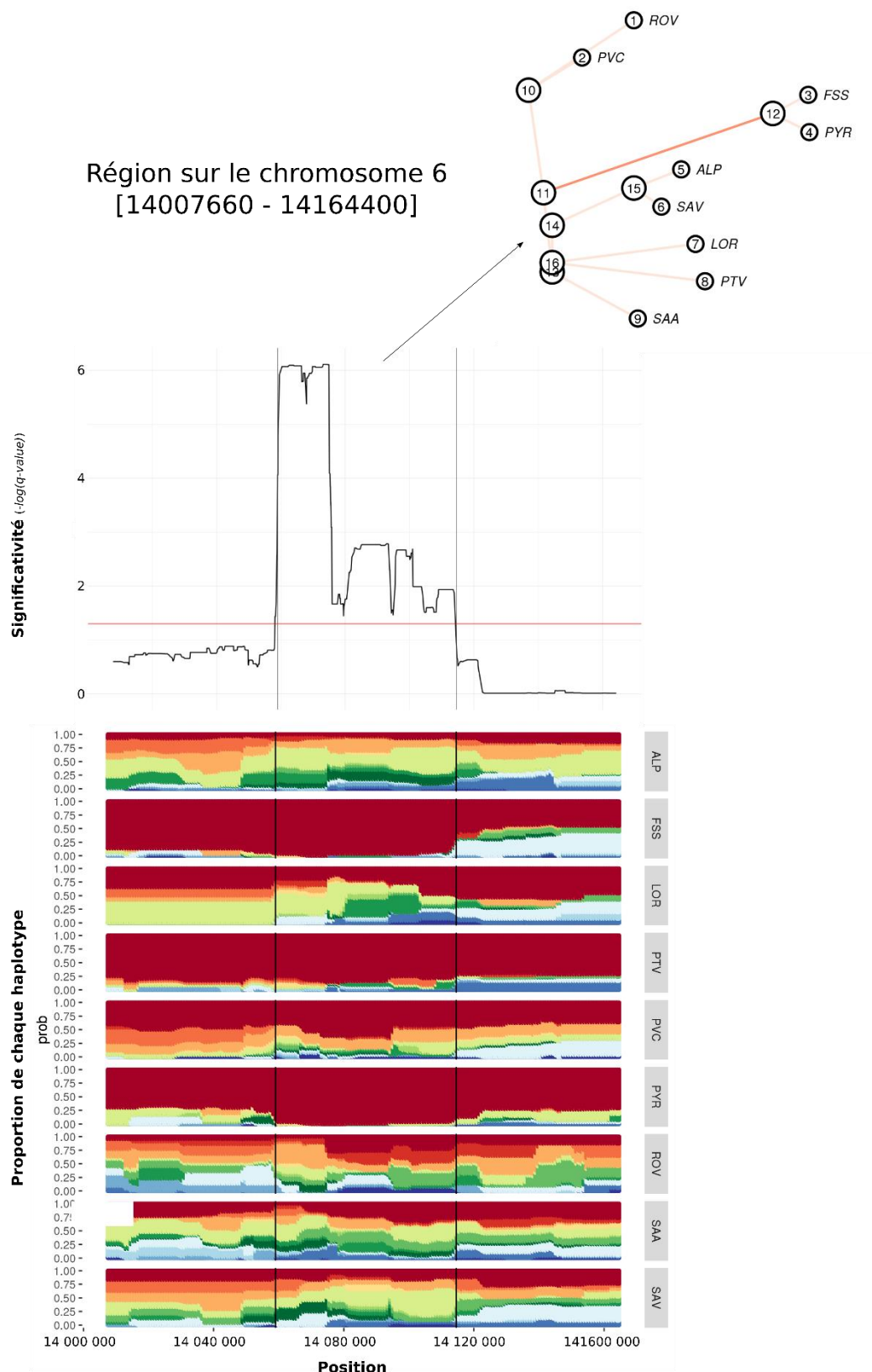


Figure S9 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 5

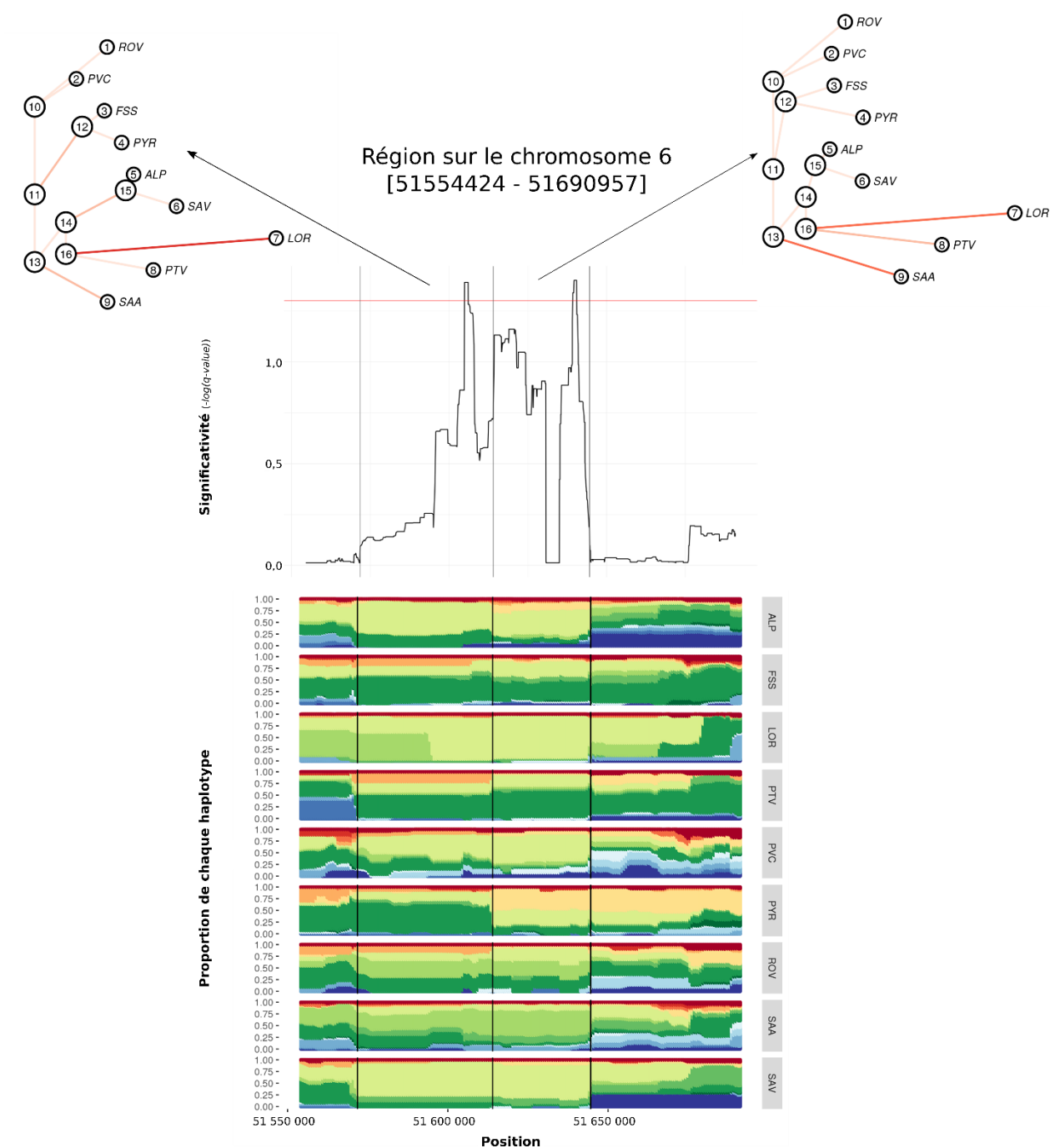


Figure S10 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 6

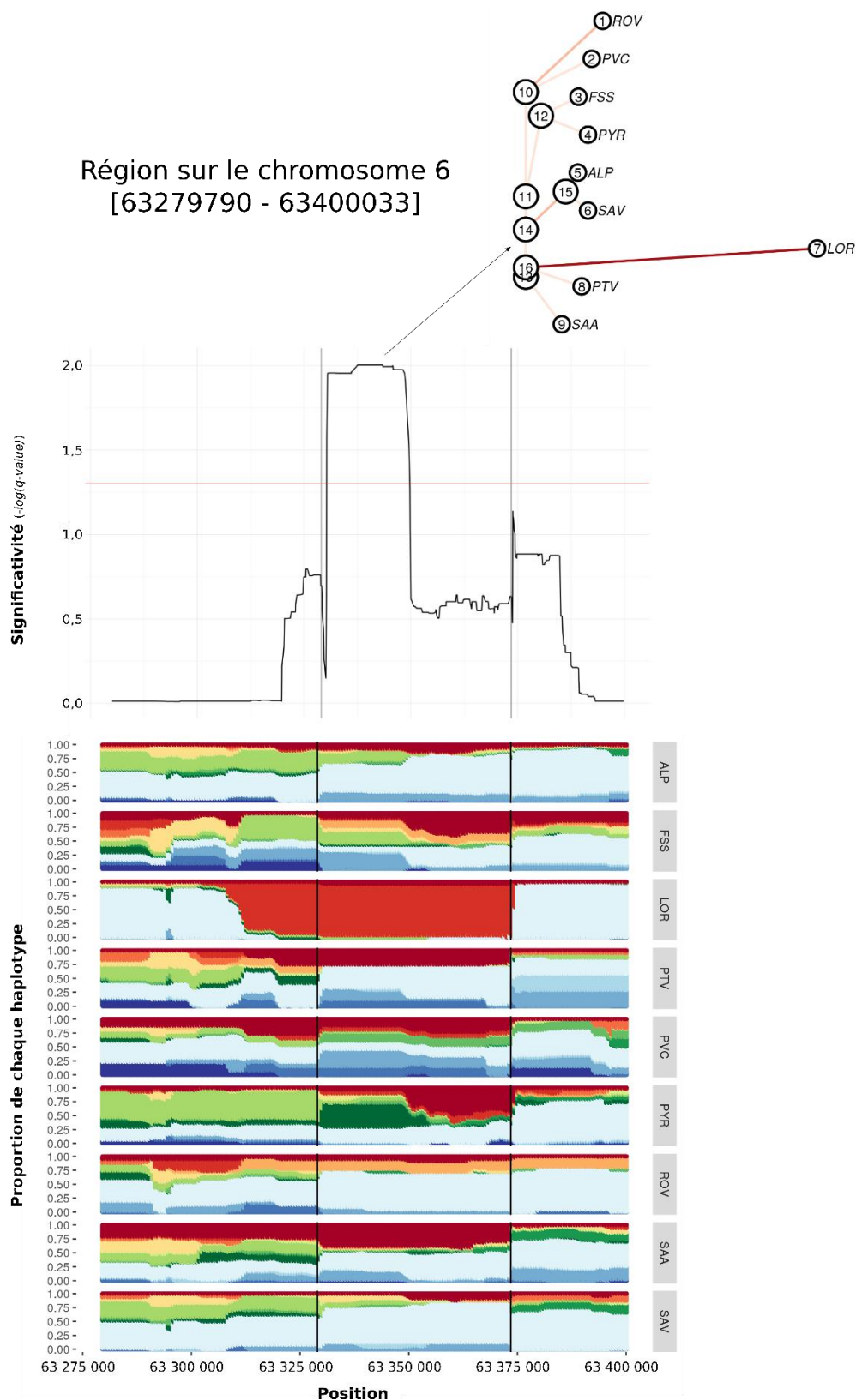


Figure S11 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 7

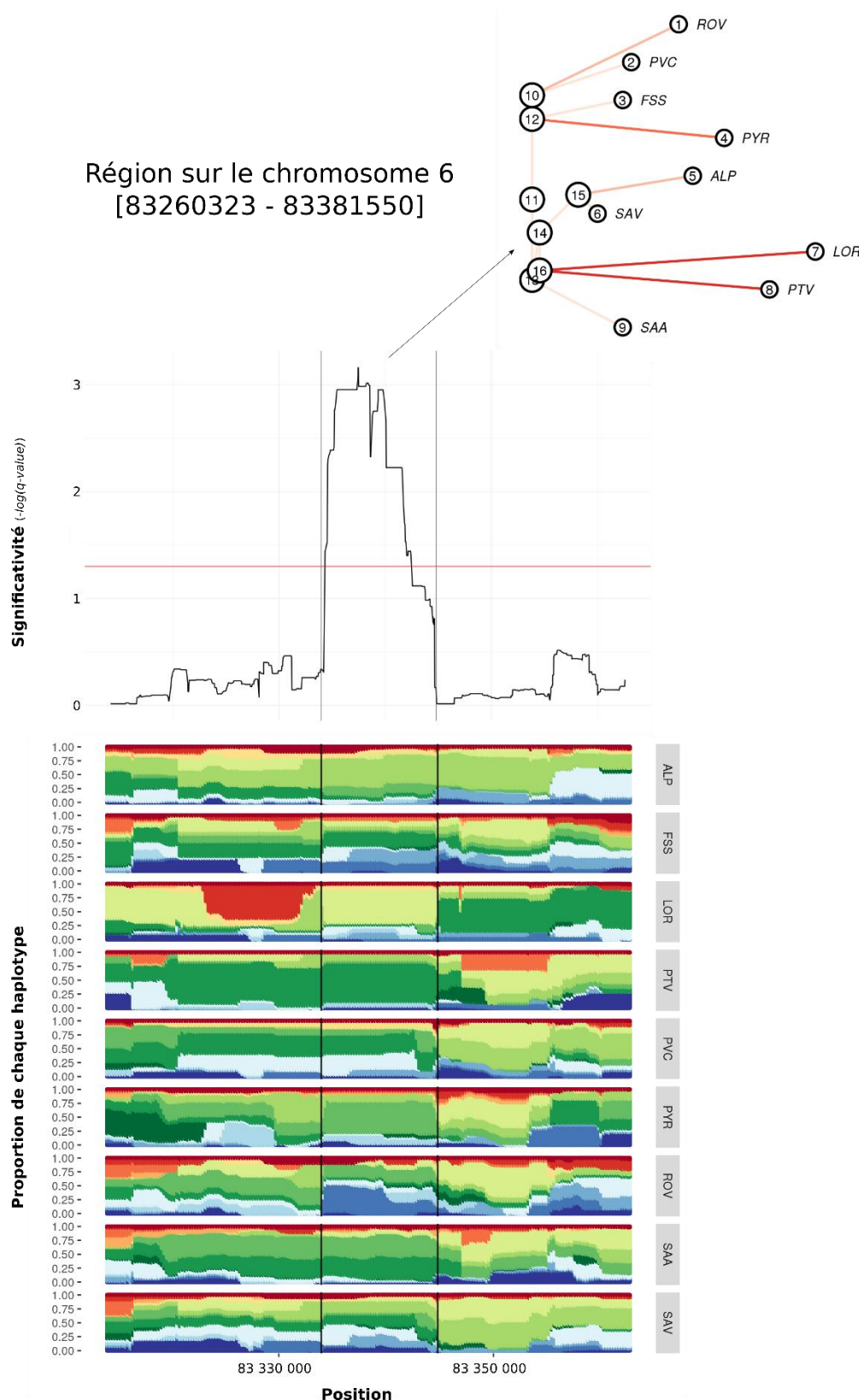


Figure S12 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleu s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 8

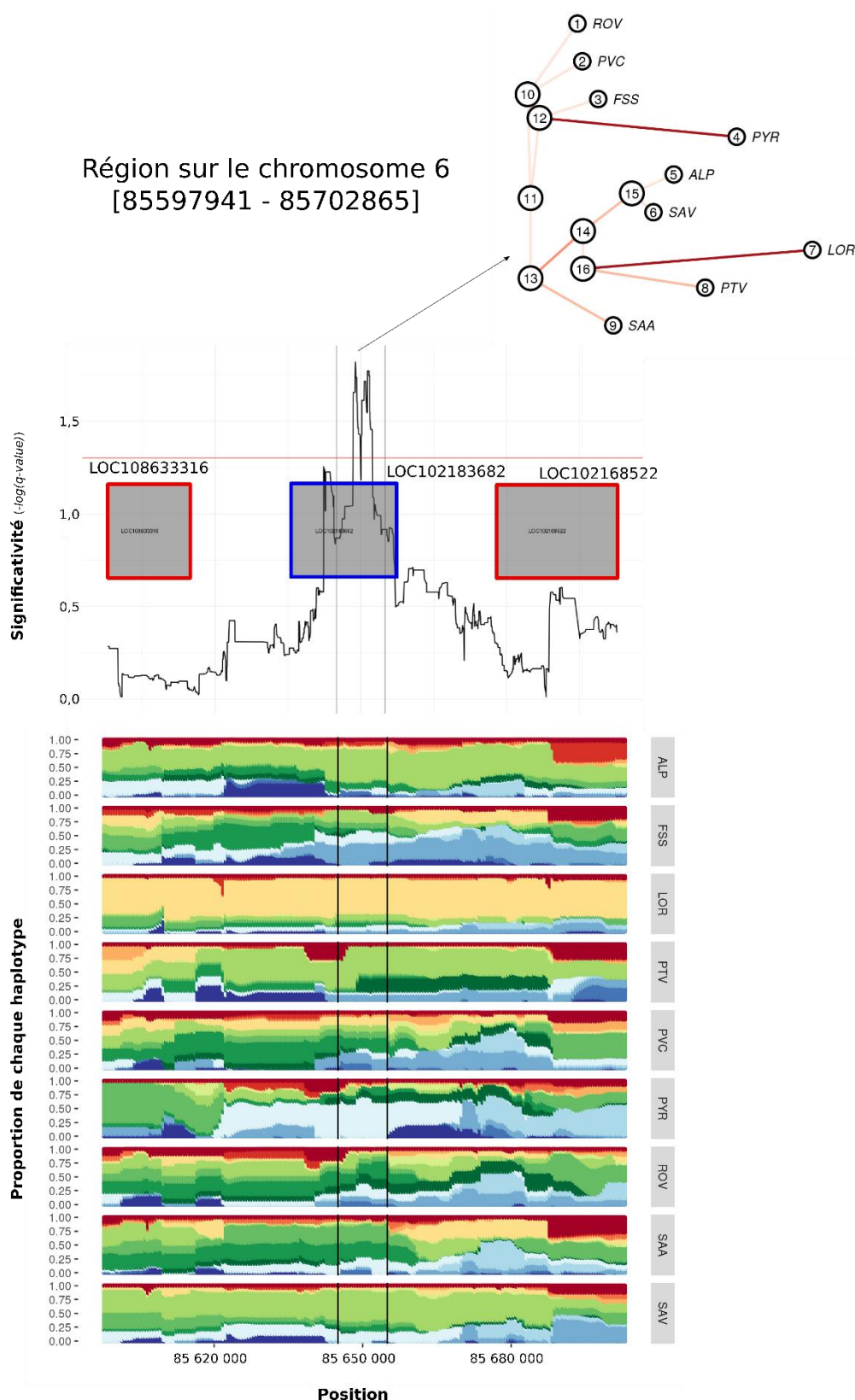


Figure S13 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 9

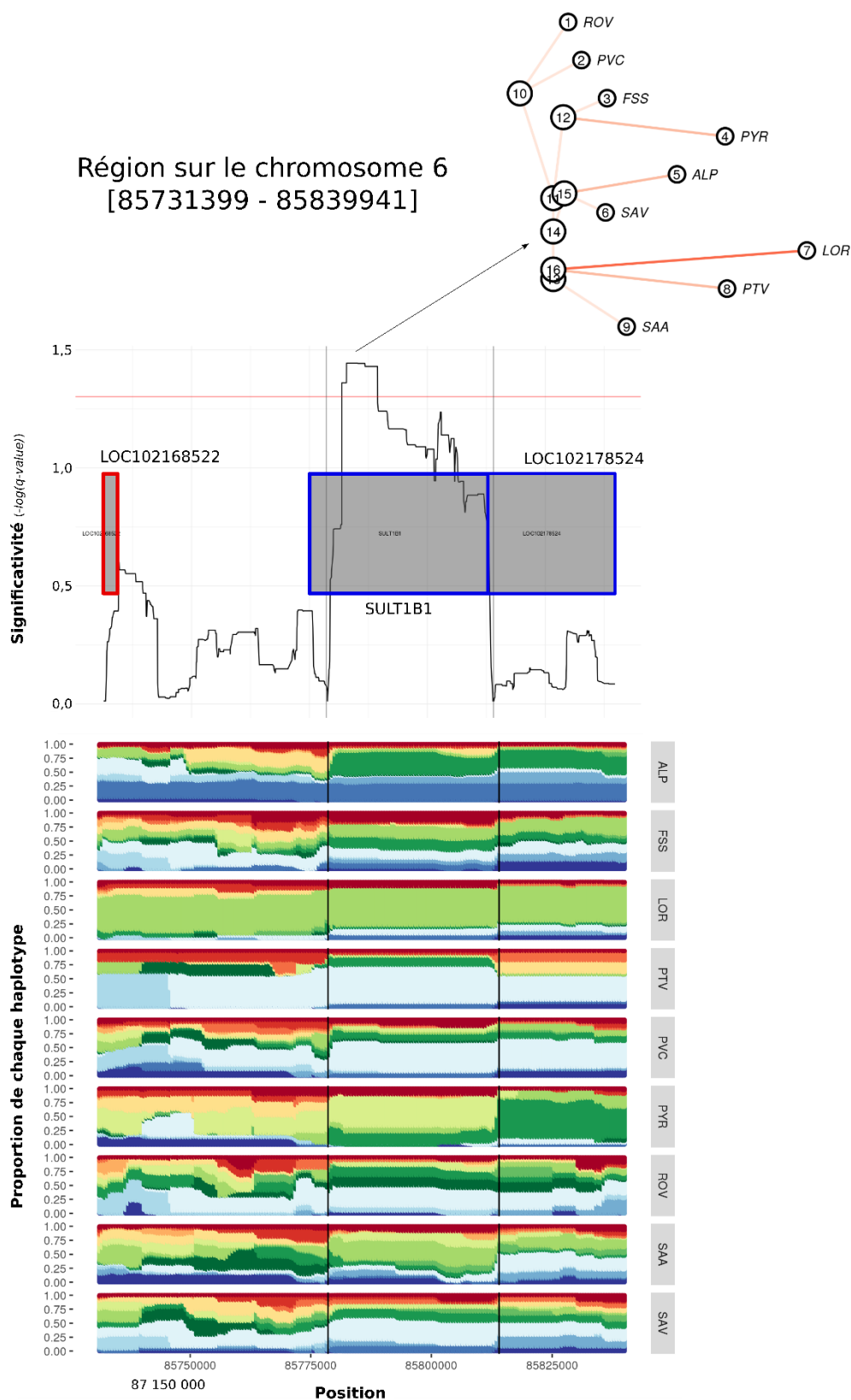


Figure S14 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 10

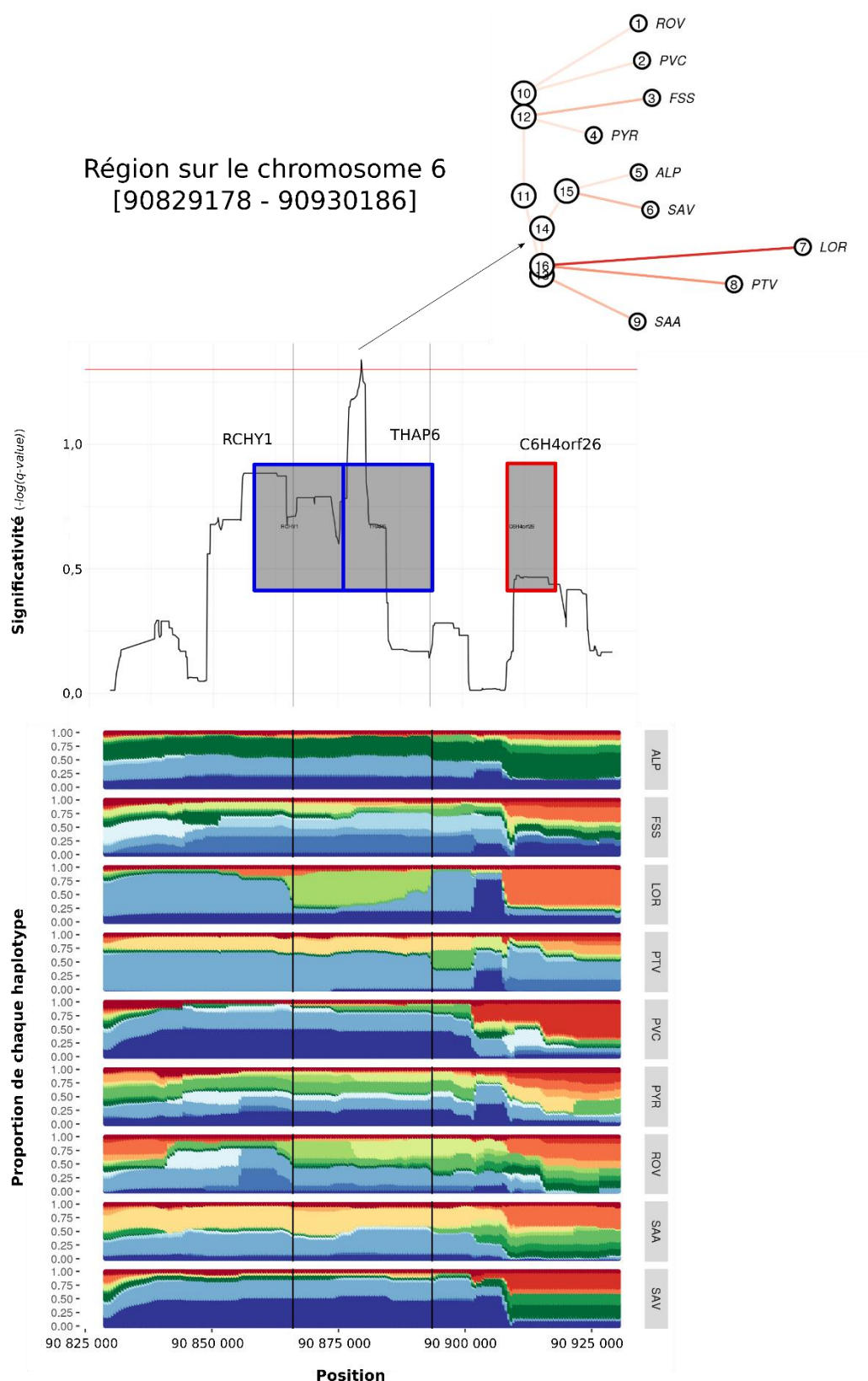


Figure S15 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 11

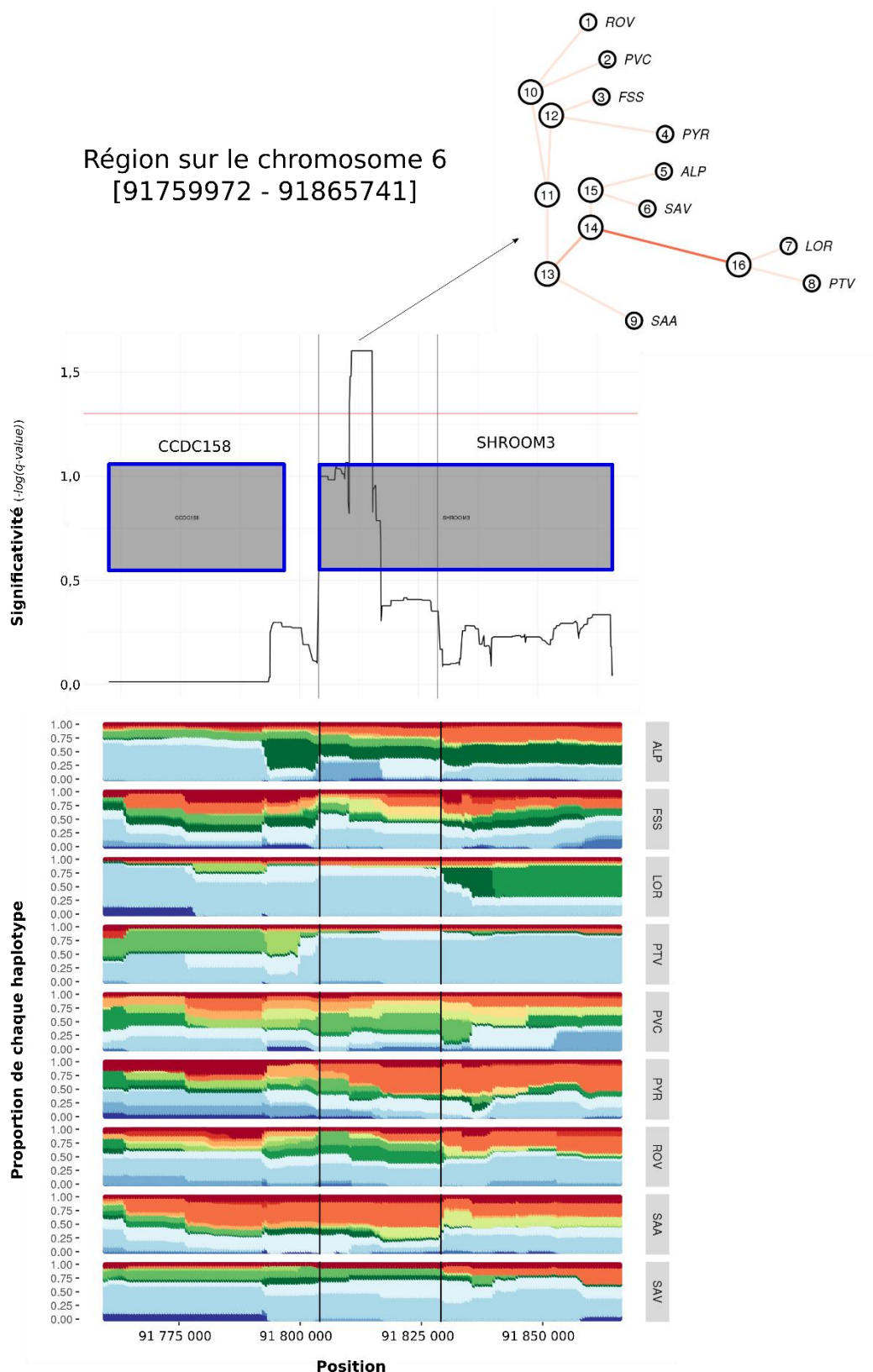


Figure S16 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 12

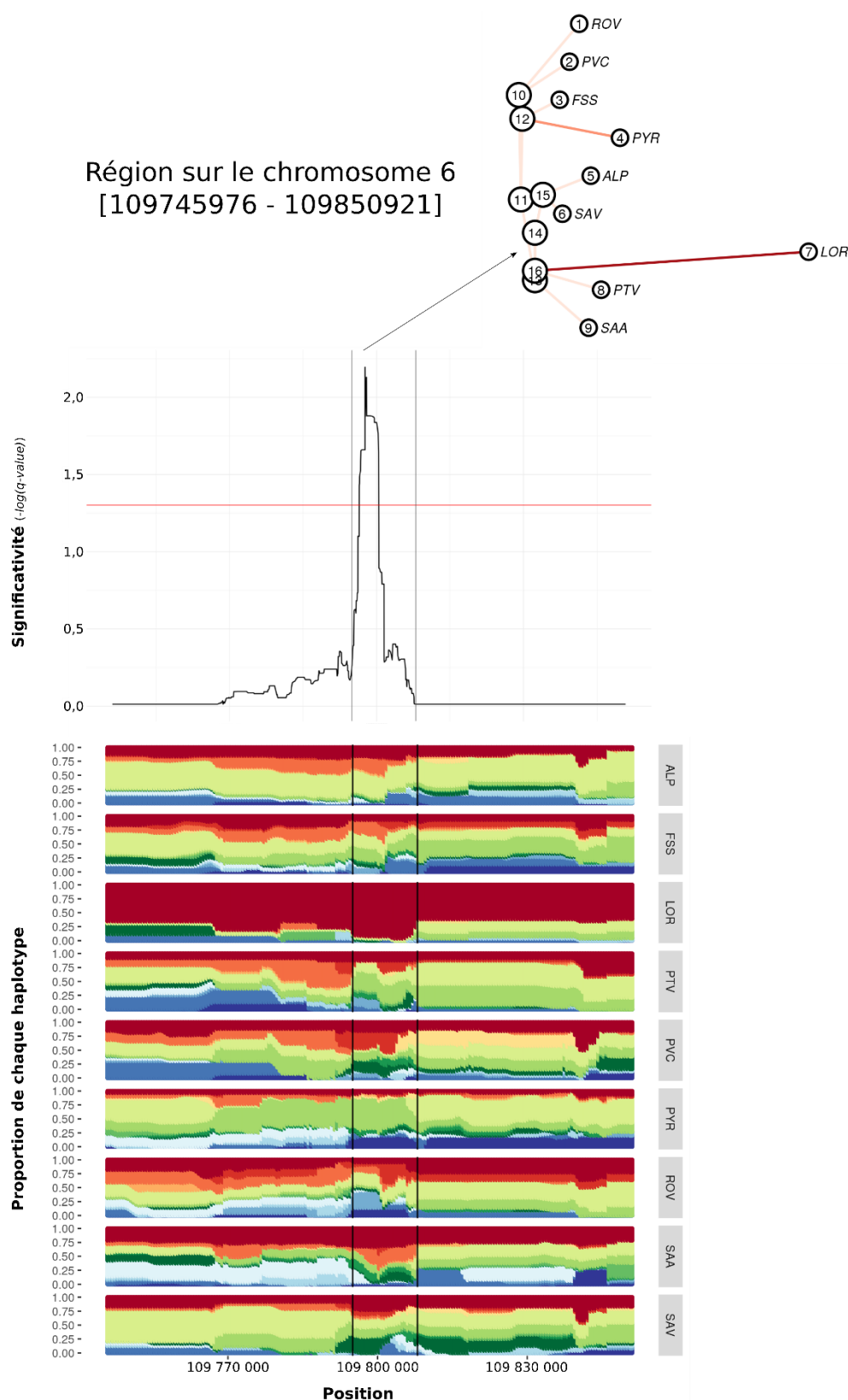


Figure S17 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 13

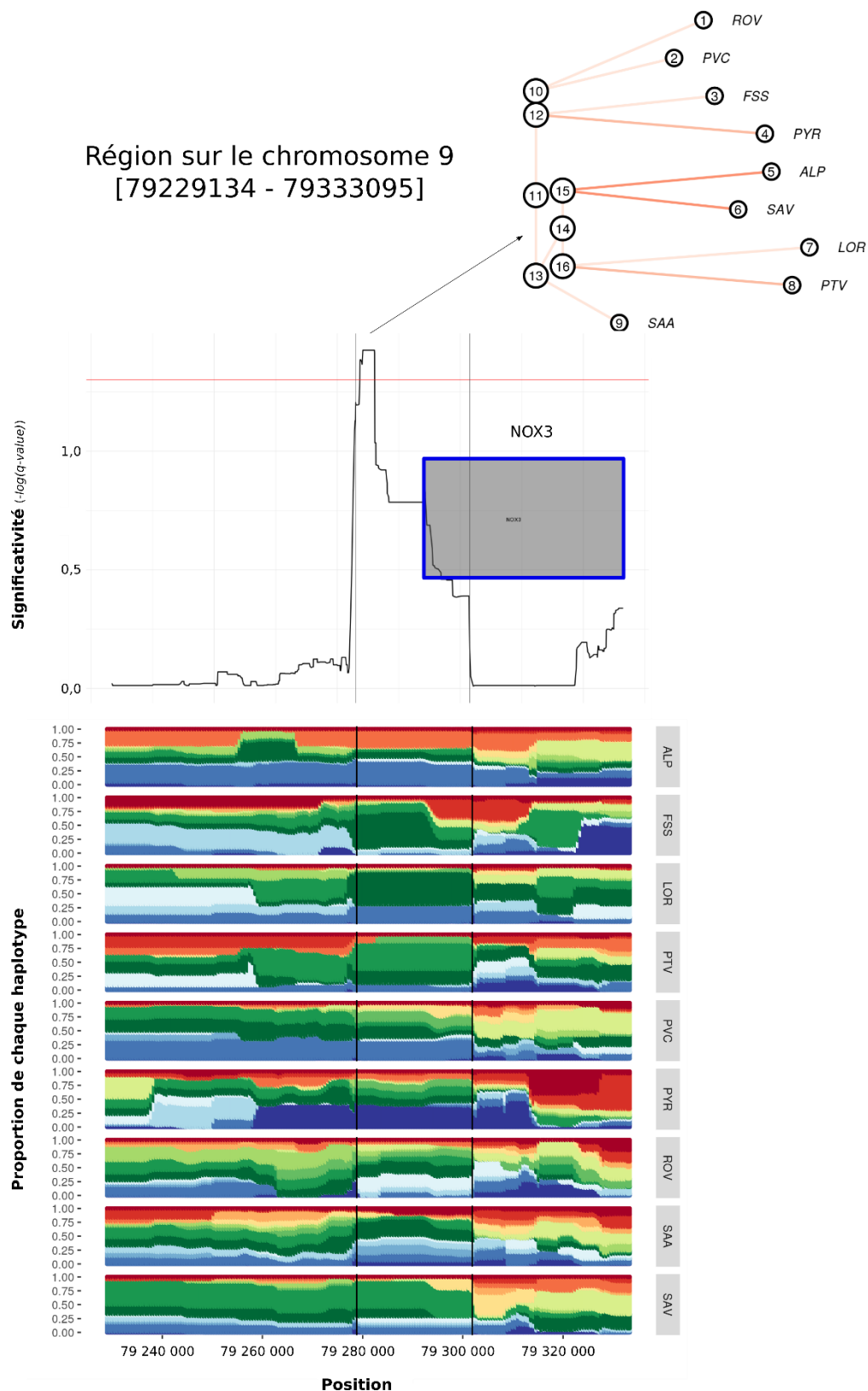


Figure S18 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 14

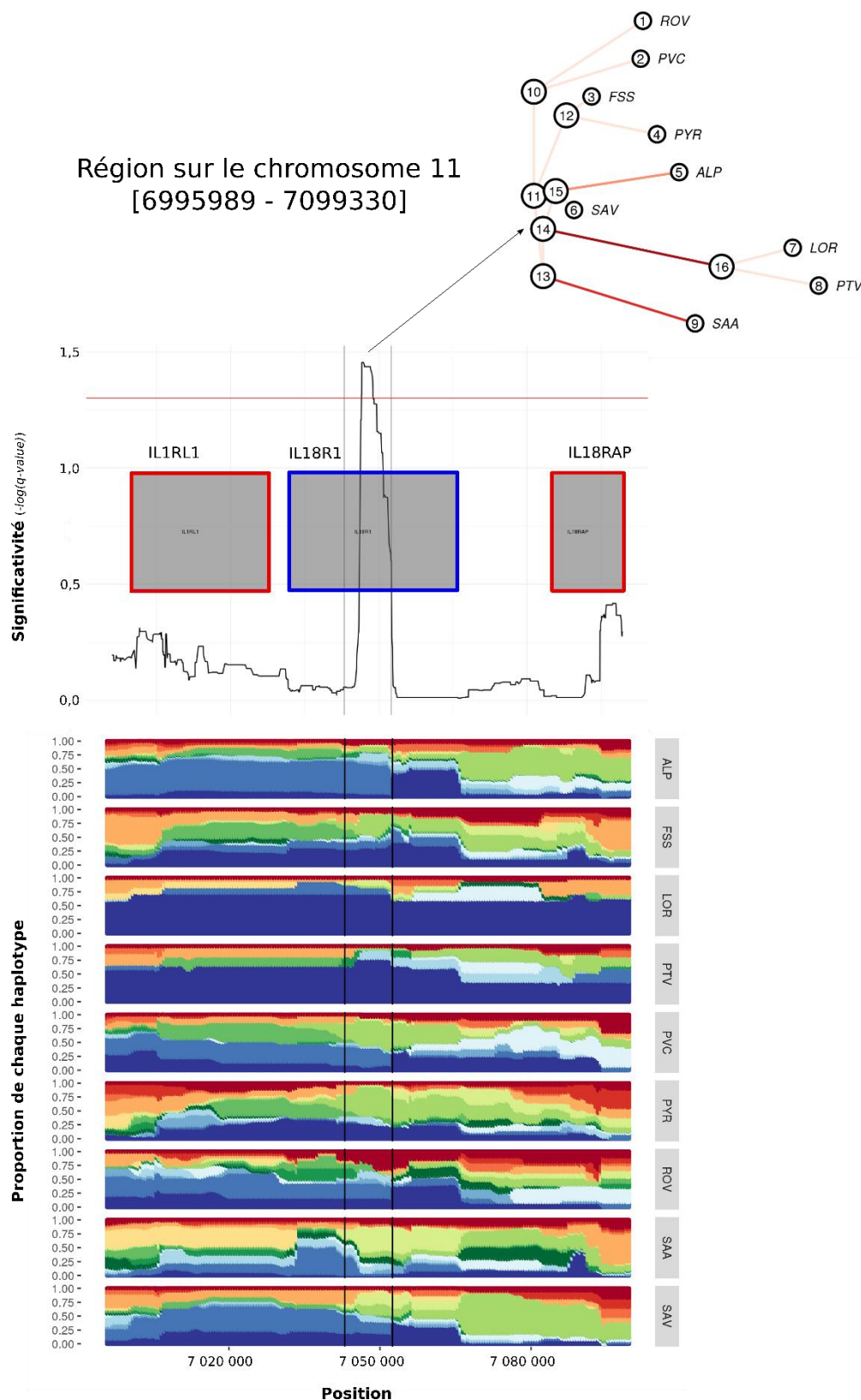


Figure S19 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 15

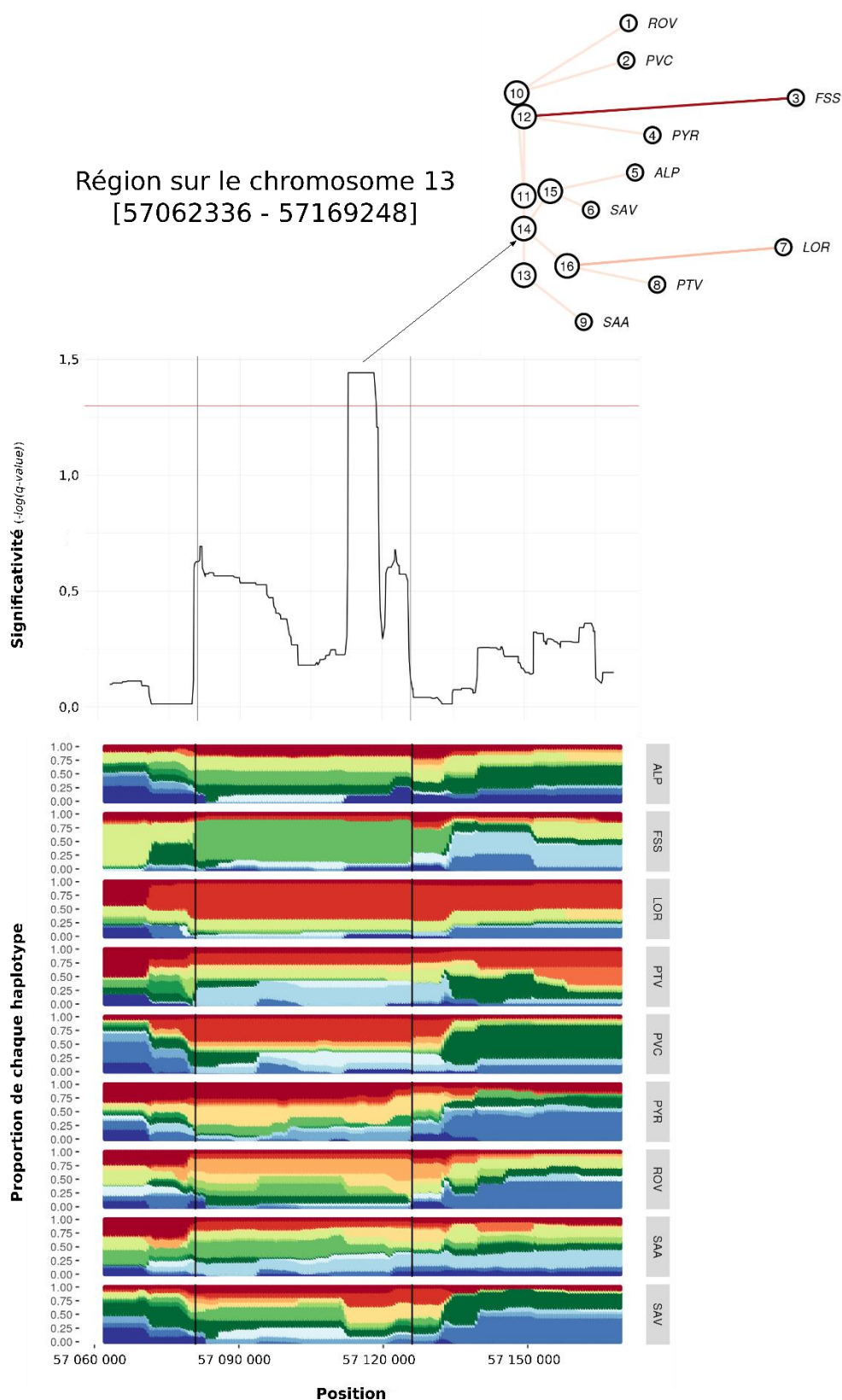


Figure S20 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 16

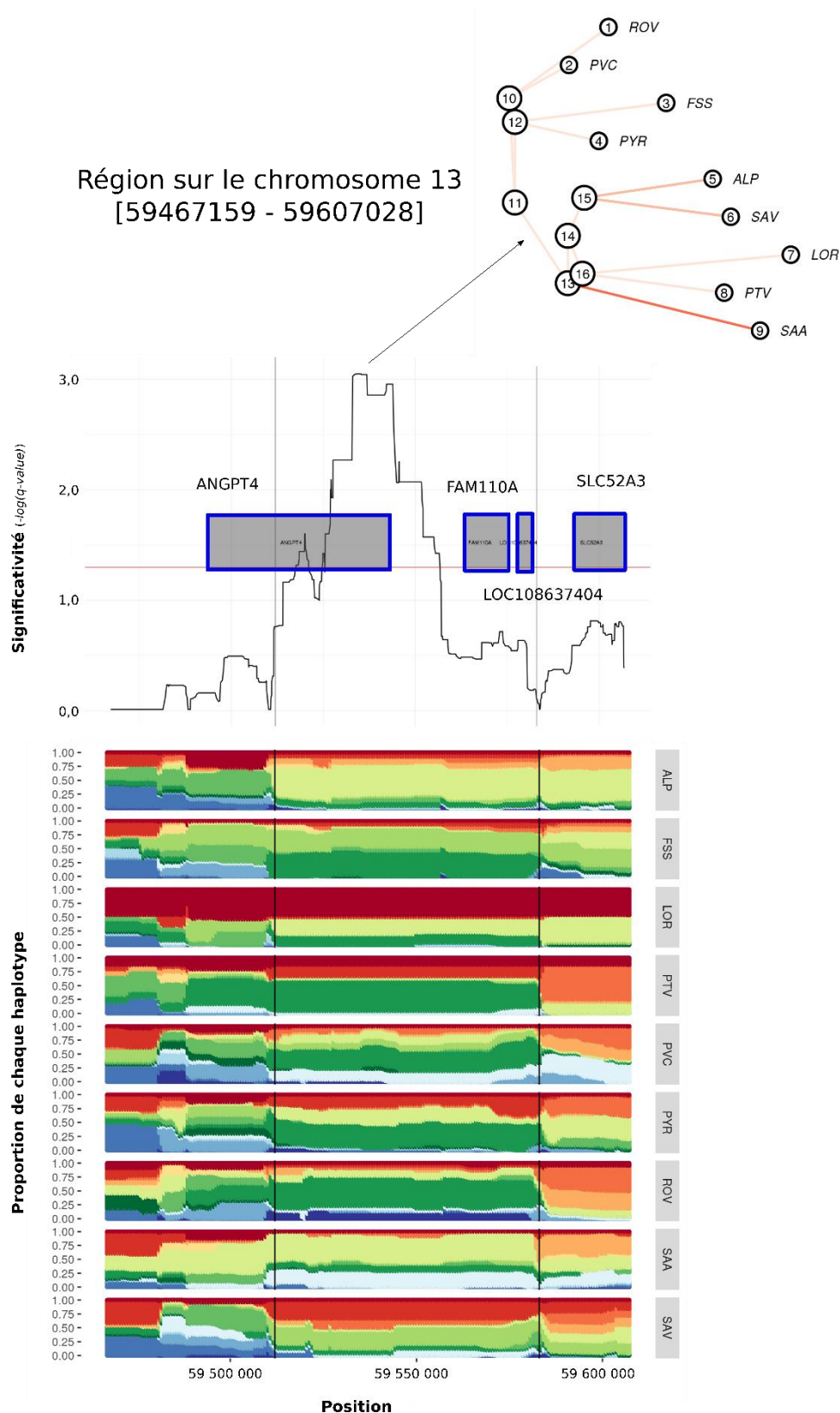


Figure S21 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 17

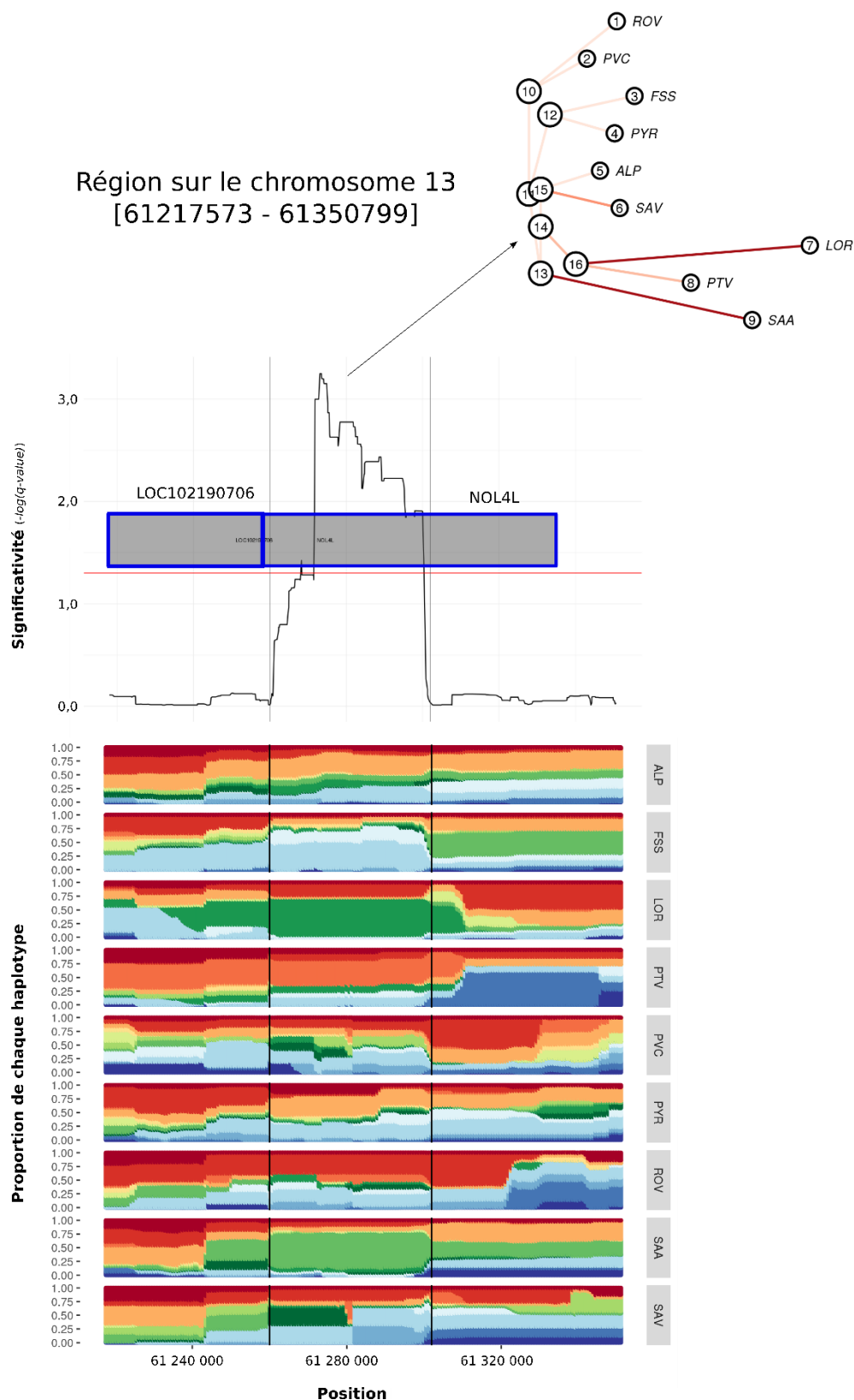


Figure S22 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 18

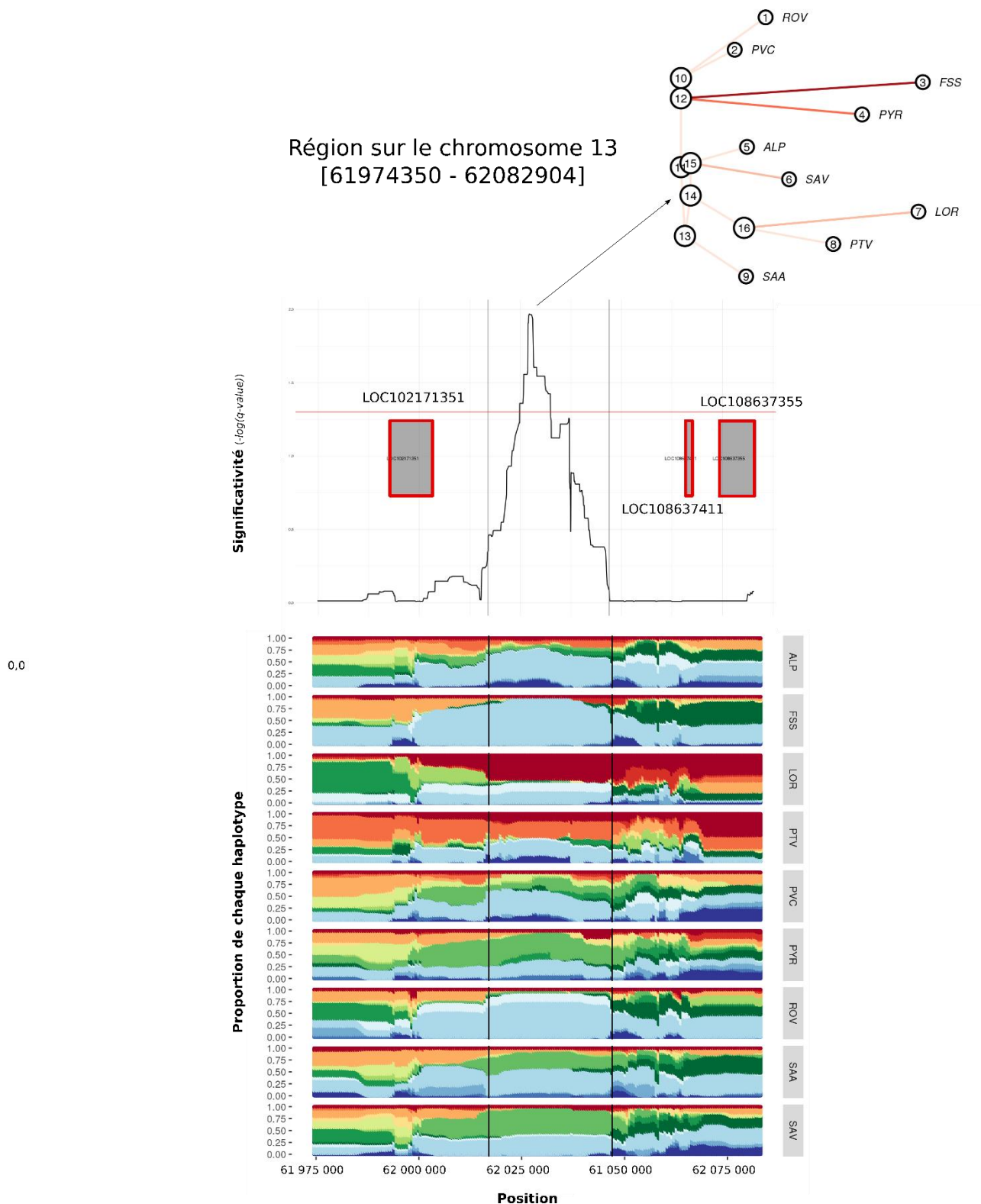


Figure S23 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 20

Région sur le chromosome 13
[62210588 - 62568730]

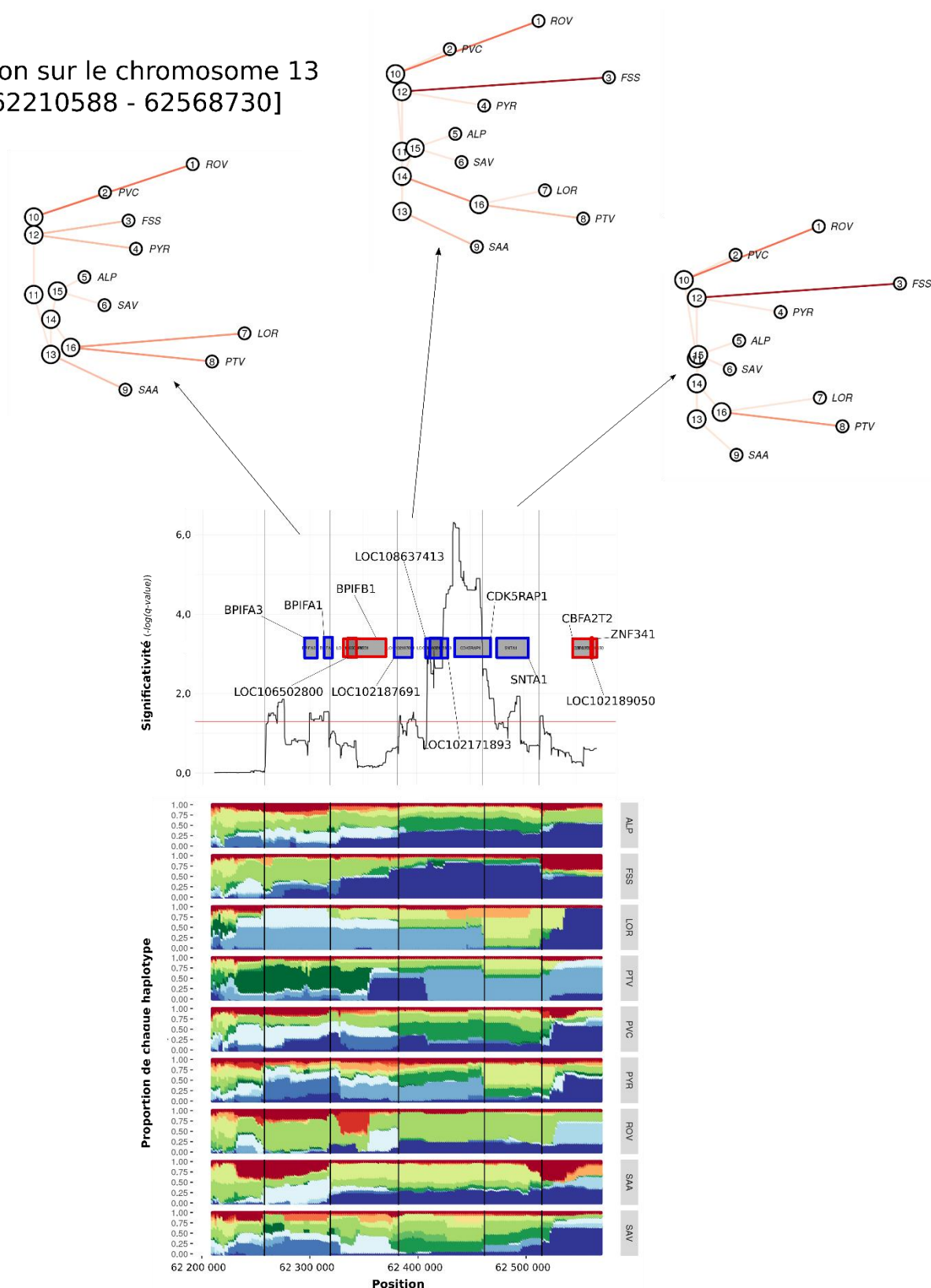


Figure S24 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 21

Région sur le chromosome 13
[62833529 - 63539347]

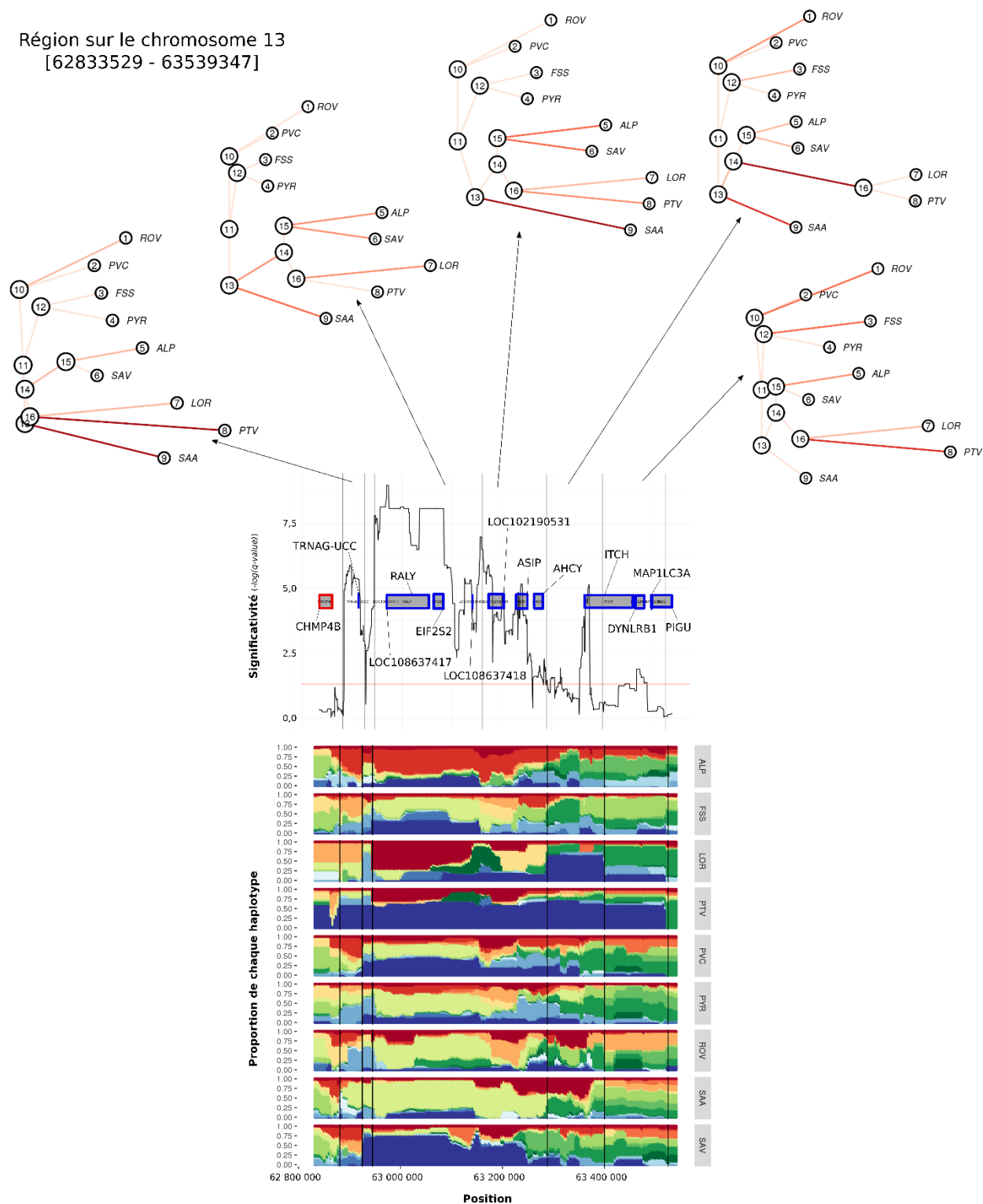


Figure S25 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 22

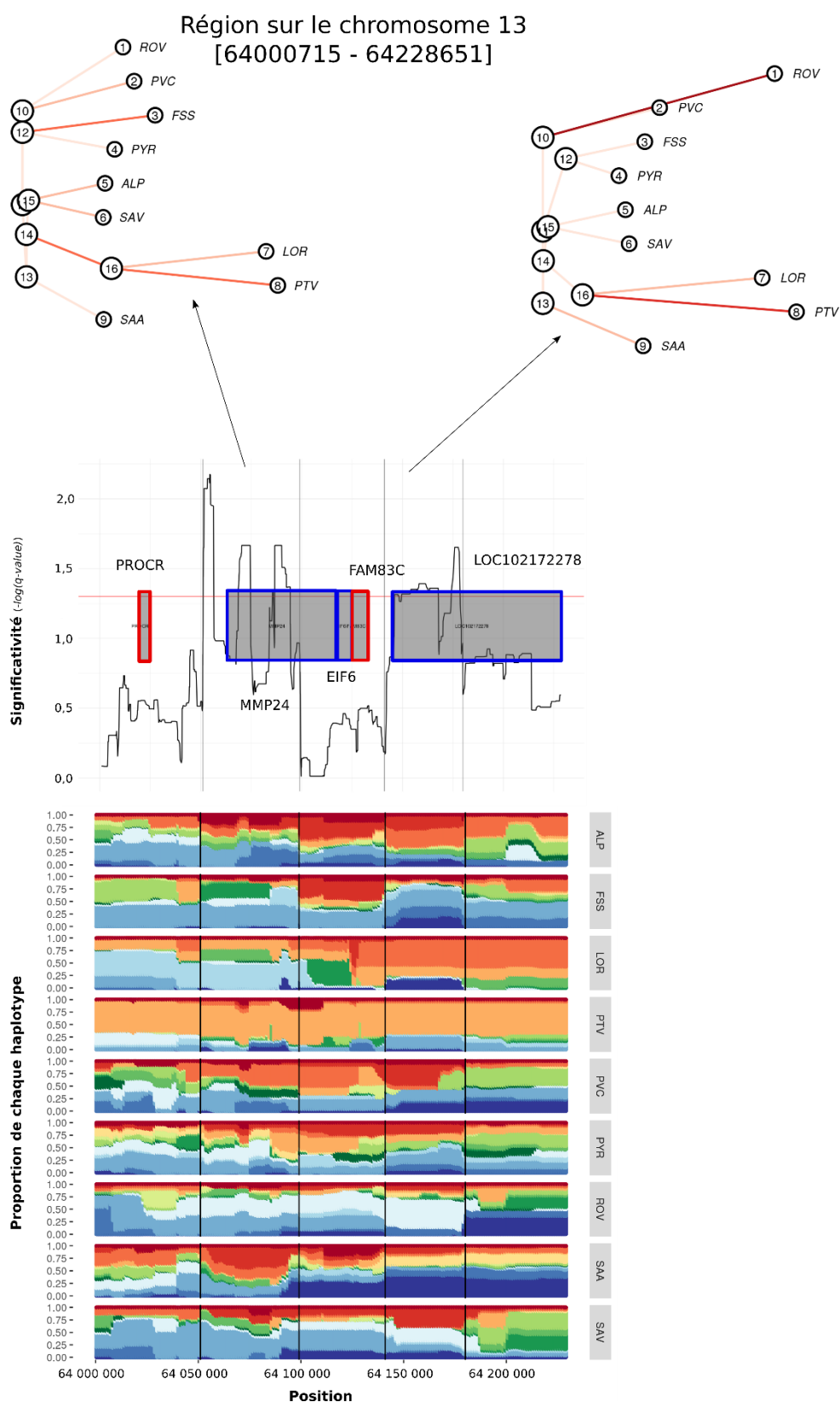


Figure S26 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 23

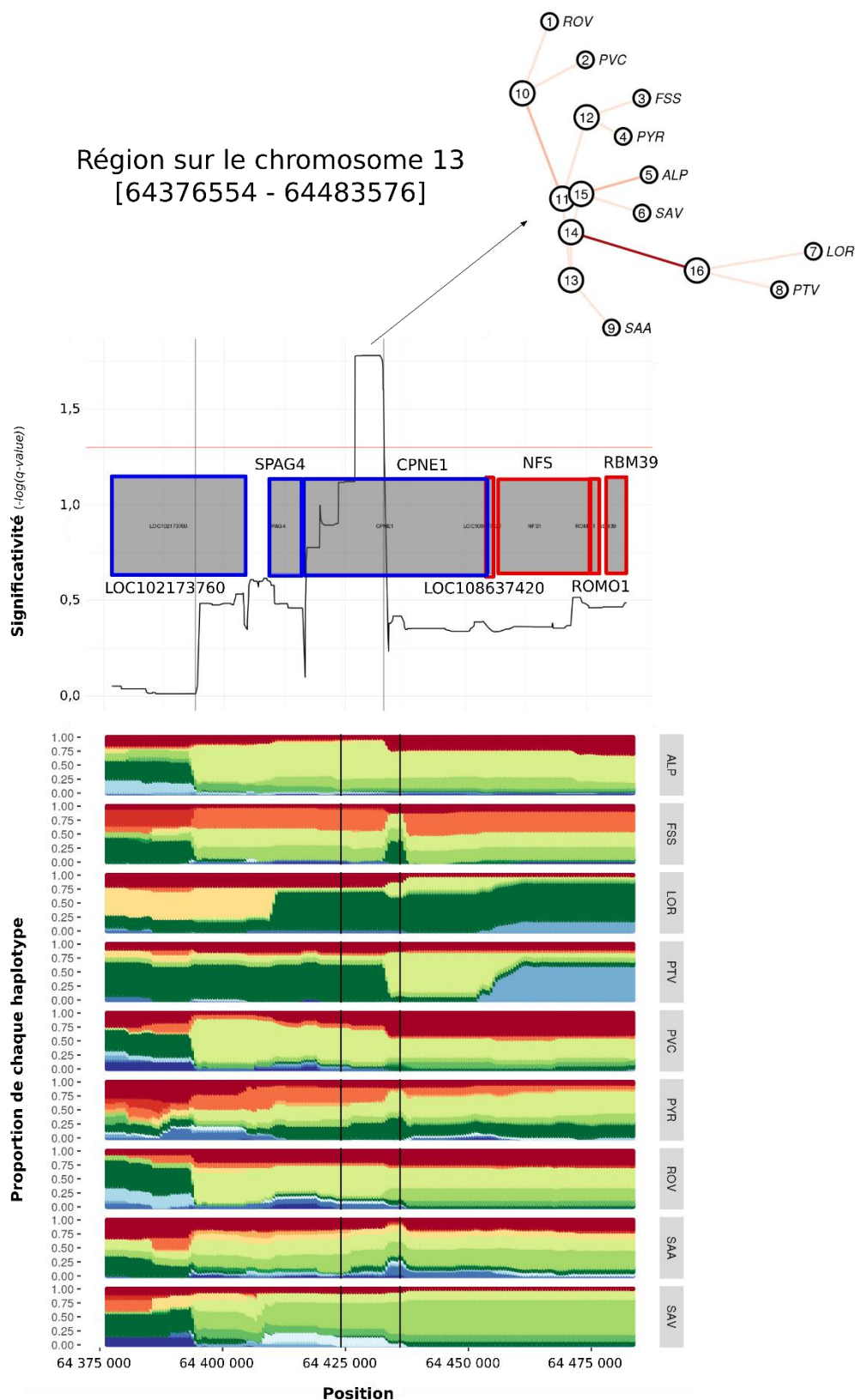


Figure S27 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 24

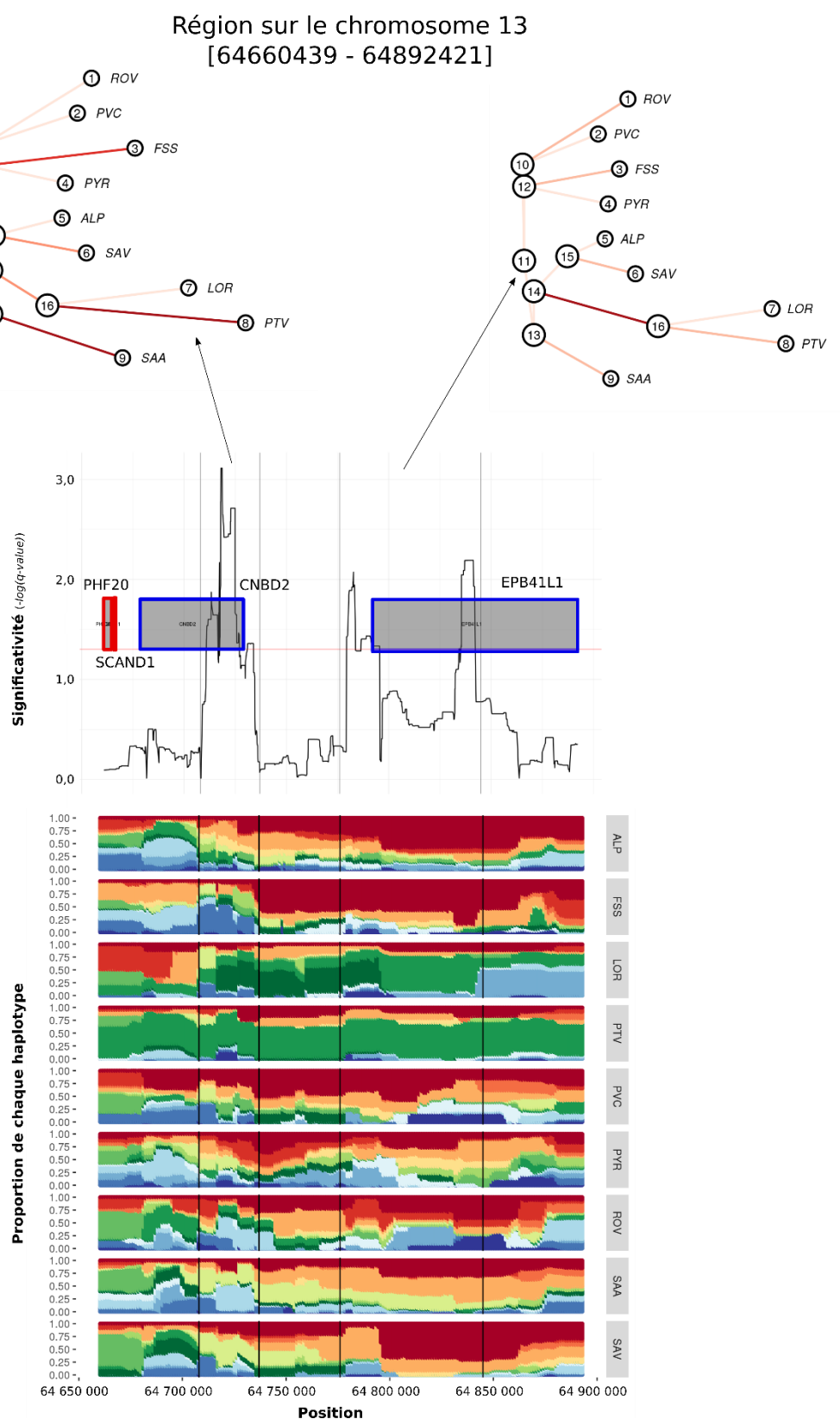


Figure S28 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 25

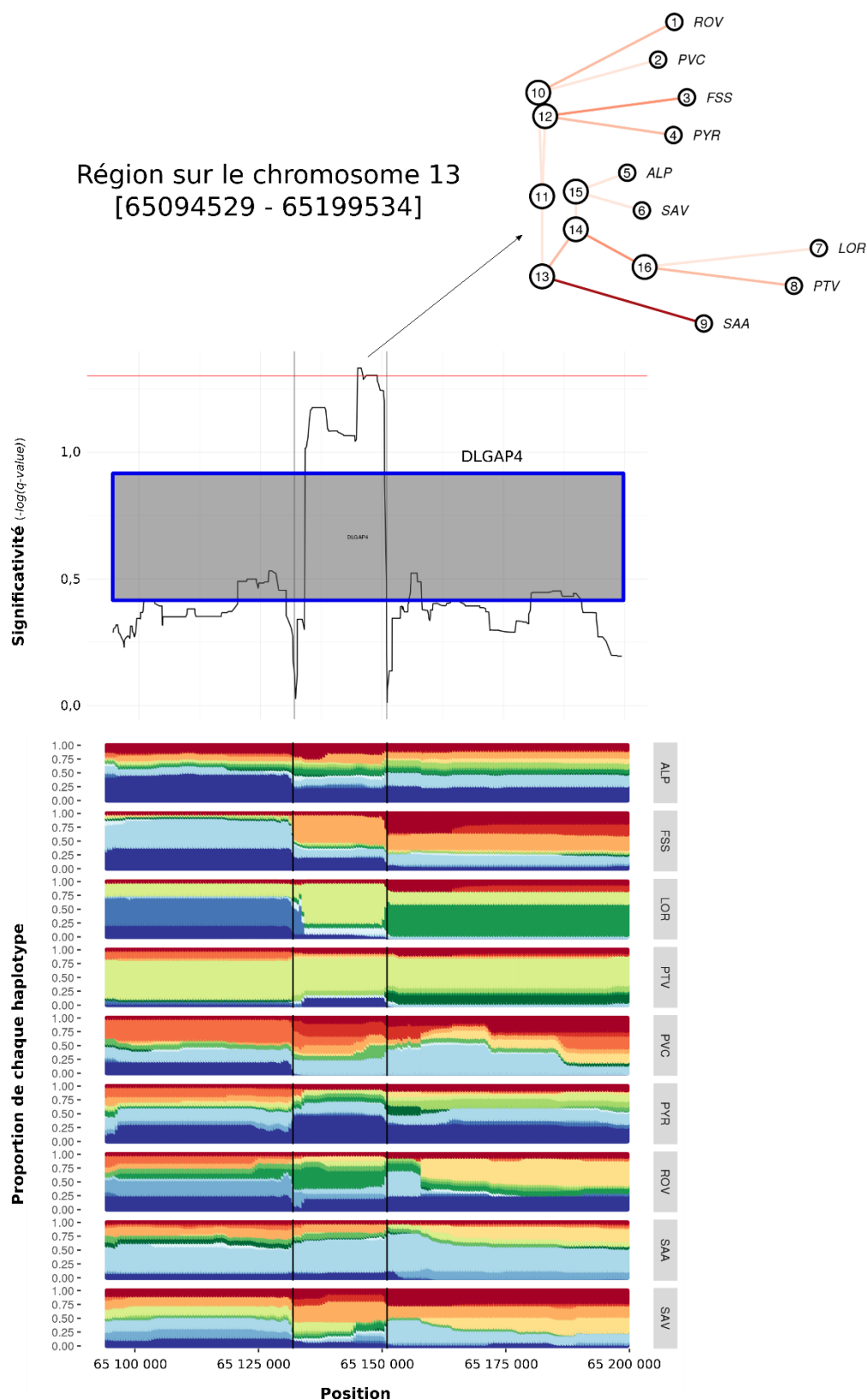


Figure S29 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 26

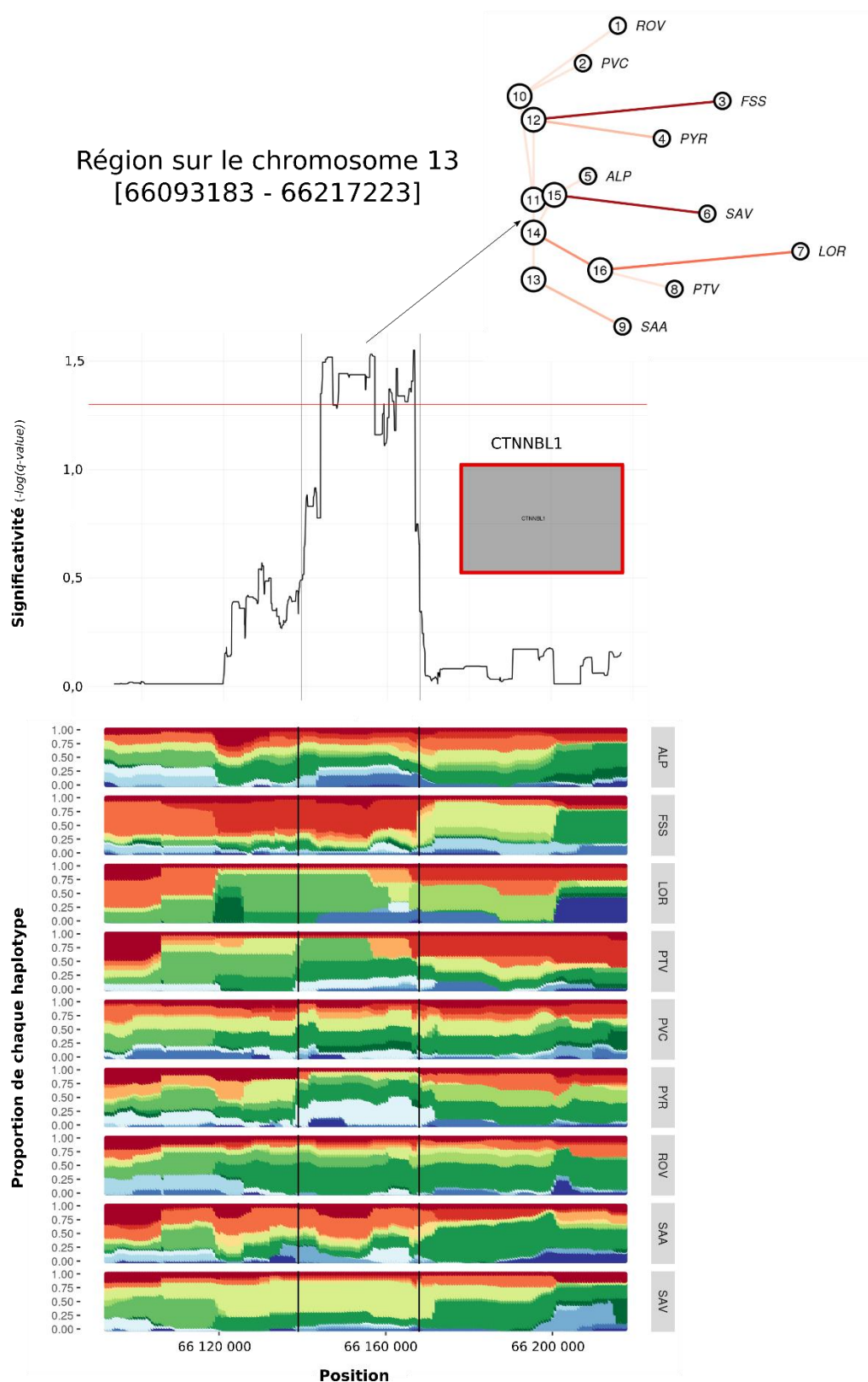


Figure S30 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 27

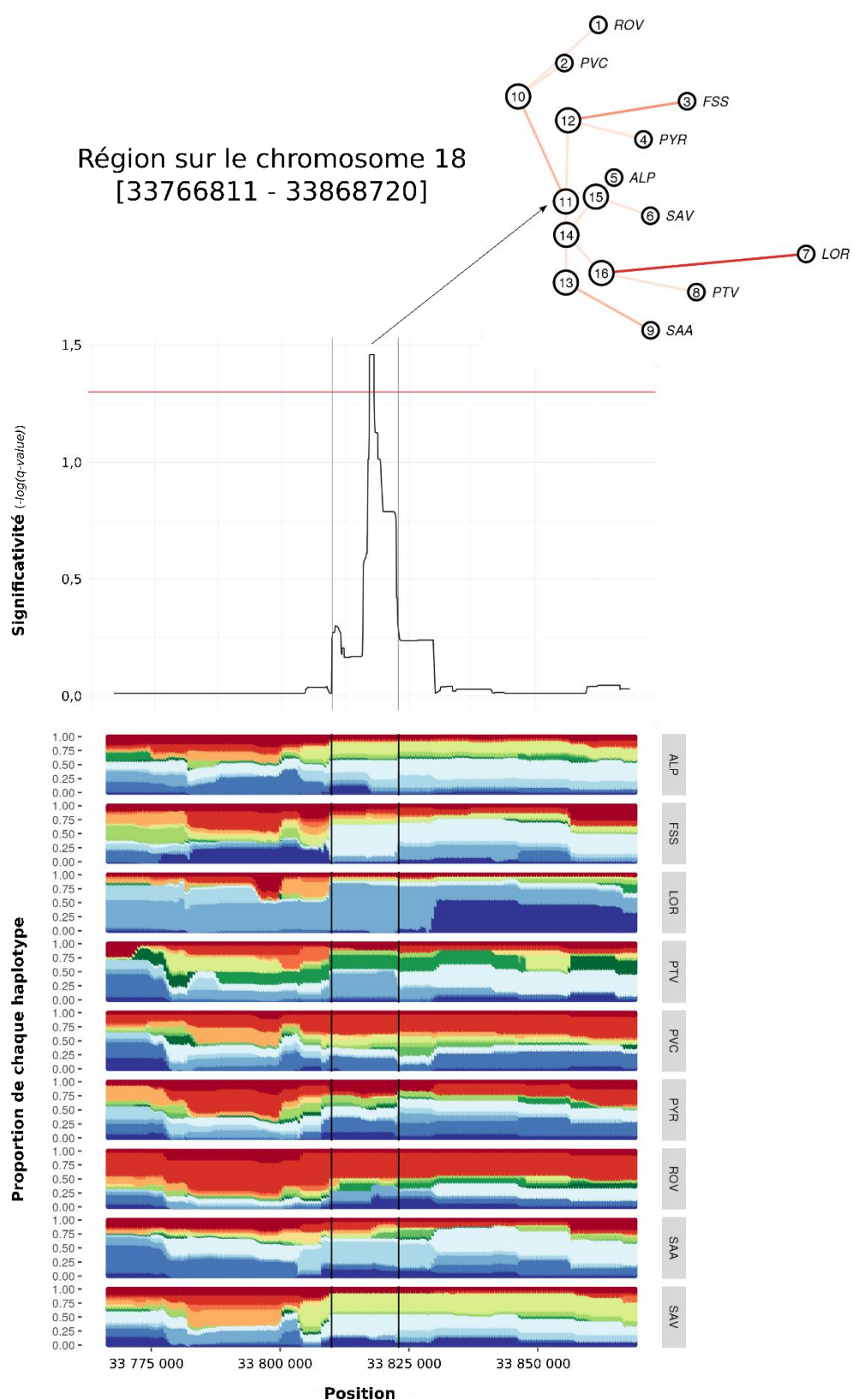


Figure S31 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 28

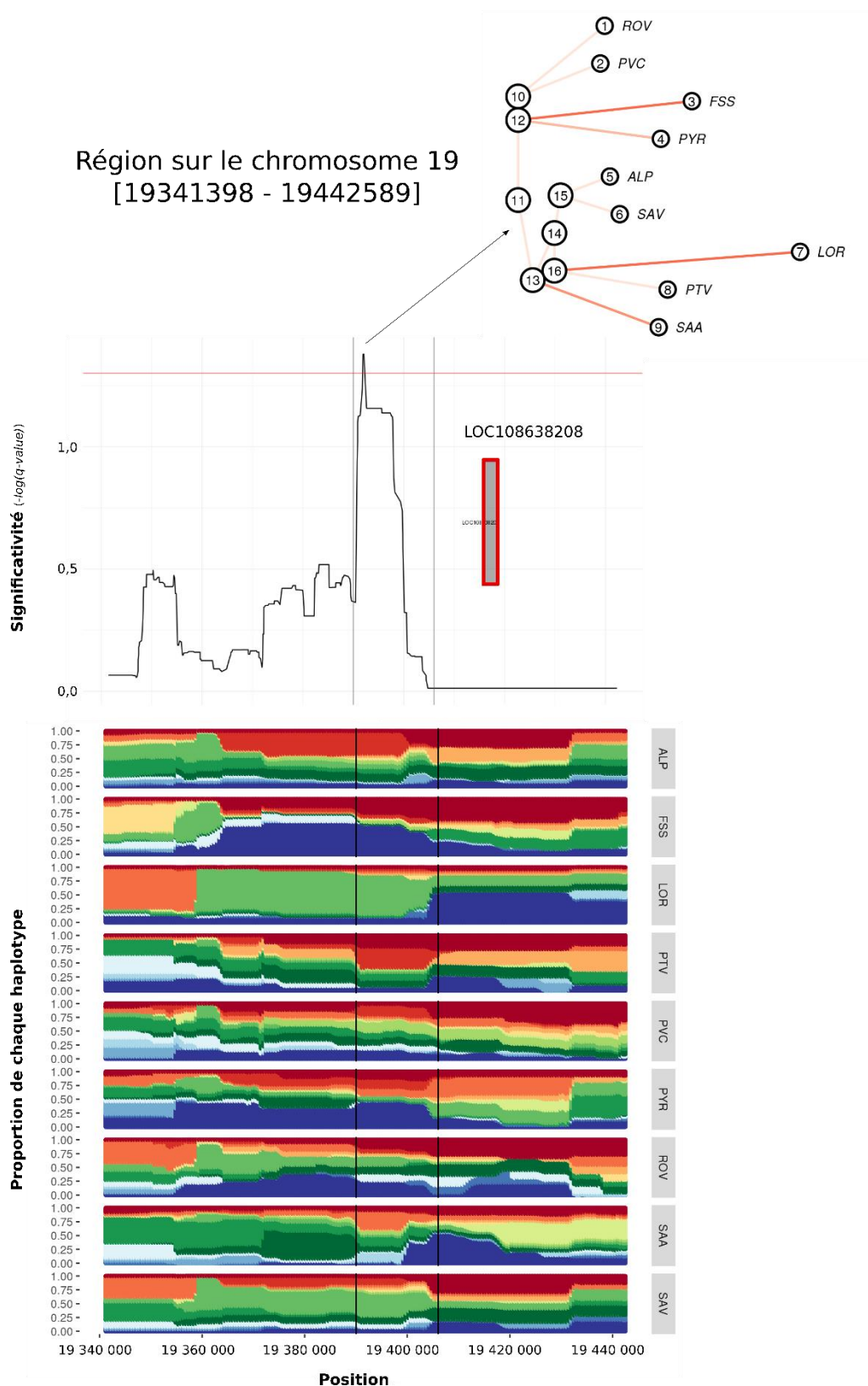


Figure S32 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

■ Région 29

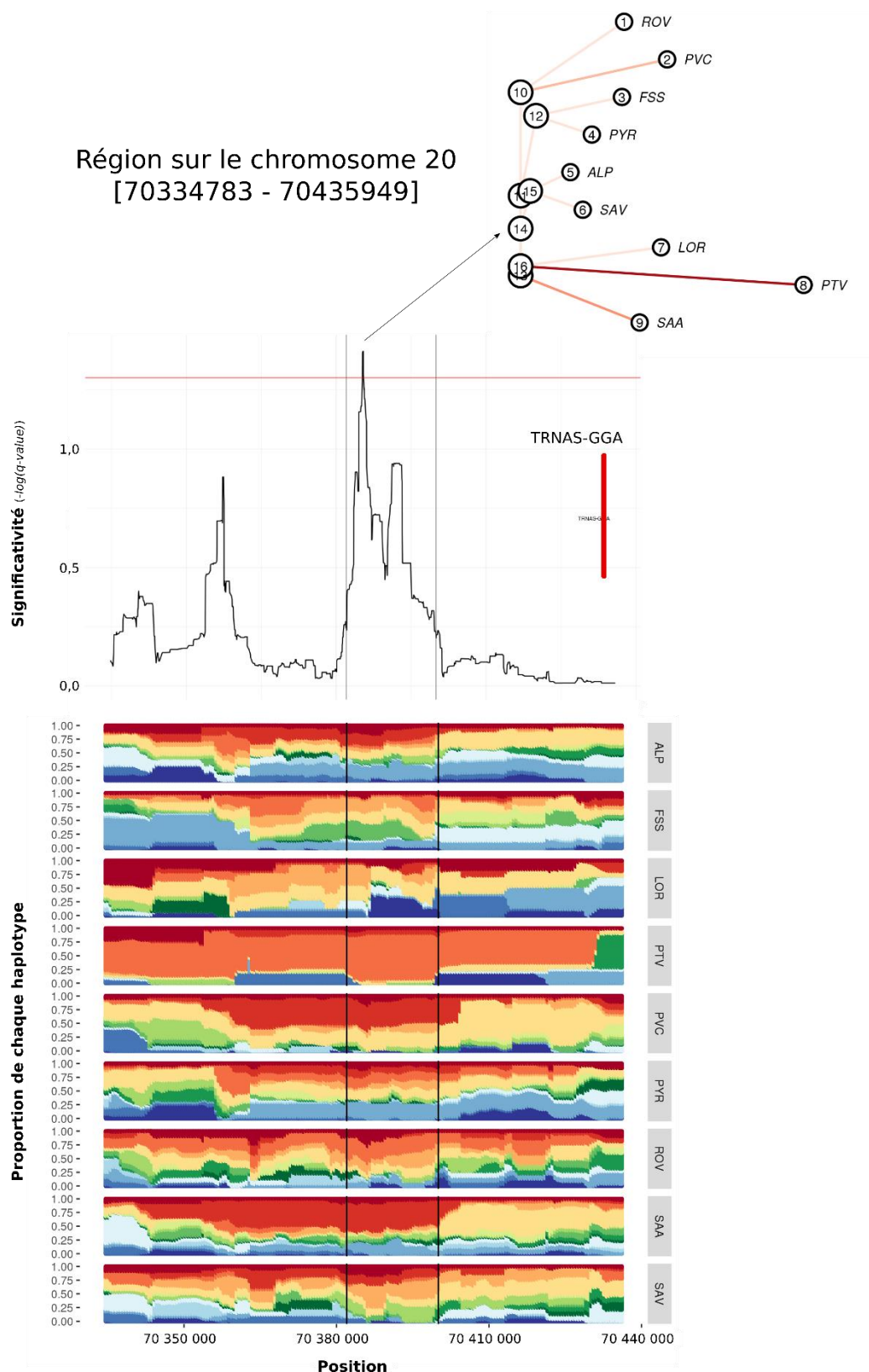


Figure S33 : Région détectée sous sélection par le logiciel hapFLK. Les haplotypes présents dans chaque race sont représentés par différentes couleurs ; les q-valeurs du test hapFLK sont présentées au-dessus avec la ligne horizontale représentant le seuil q-valeur égale 0,05 ; ensuite nous retrouvons les arbres locaux associés à chaque pic significatif. Les gènes présents dans la région sont encadrés en bleus s'ils sont associés au pic (distance de moins de 10 Kb) ou en rouge si ce n'est pas le cas.

Analyse ciblée

■ Chèvres Alpines

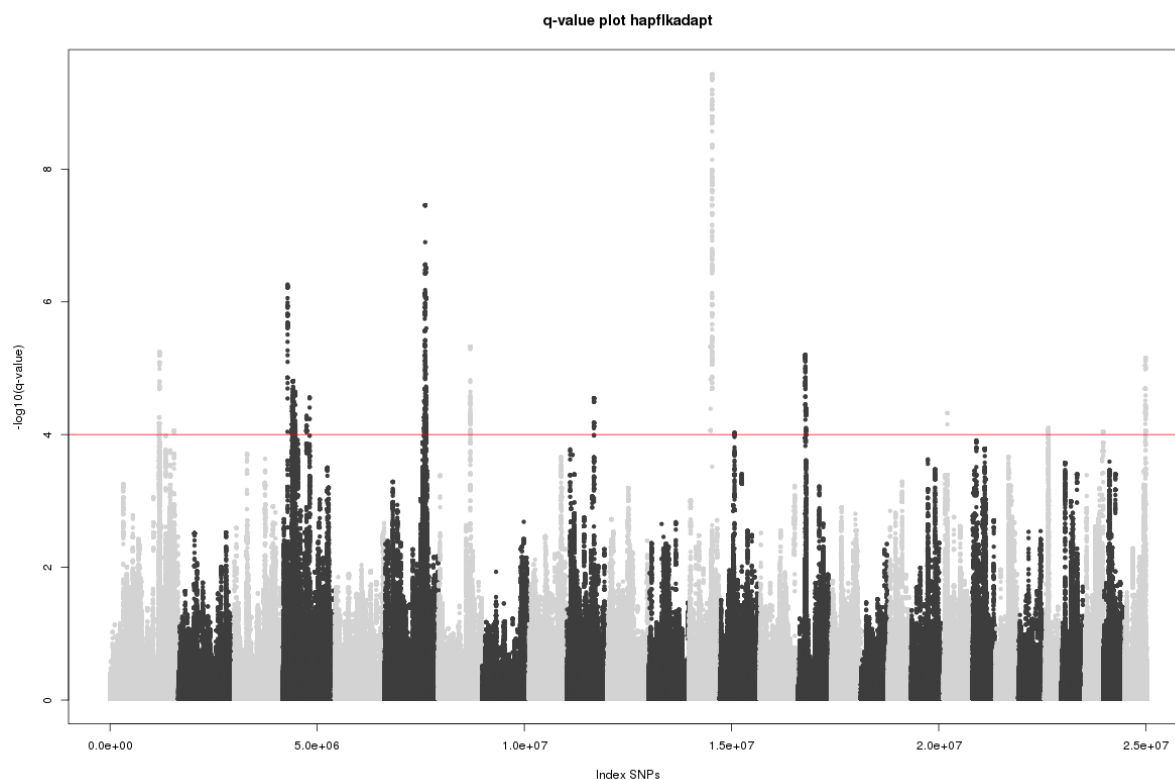


Figure S34 : Résultats du test hapFLK_{adapt} sur l'ensemble du génome de la race Alpine. L'axe des ordonnées représente la significativité des SNPs et l'axe des abscisses la position sur le génome avec une alternance de couleurs en fonction des différents chromosomes. La ligne horizontale correspond au seuil de q-valeur 1.10^{-4} .

Tableau S4 : Régions sélectionnées mises en évidence chez la race Alpine par le programme hapFLK_{adapt} ainsi que les gènes associés.

Chr	Début région	Fin région	Gènes	Chr	Début région	Fin région	Gènes
1	114541516	114654331	NA	7	79465943	79575765	LOC102177155
4	9164792	9280240	CNTNAP2	7	79575933	79750714	SNX2
4	18233776	18381675	NA	7	79575933	79750714	SNX24
4	20521734	20838915	TRNAQ-CUG	10	73478008	73582869	FMN1
4	20521734	20838915	LUZP6	13	57815438	57918469	ZBP1
4	20521734	20838915	MTPN	13	62905564	63238297	LOC108637417
4	21599968	21721538	CALD1	13	62905564	63238297	RALY
4	26783560	26960727	LOC102179265	13	62905564	63238297	EIF2S2
4	26783560	26960727	AHCYL2	13	62905564	63238297	LOC108637418
4	61888067	61989333	NA	13	62905564	63238297	LOC102190531
6	85920854	86060367	CSN1S1	16	12954914	13068423	NA
6	85920854	86060367	CSN2	29	47202555	47315554	CPT1A
6	86886106	86999866	NA	29	47832293	47941353	NA
6	87889832	88007030	ADAMTS3				
6	89911396	90081039	NA				
6	90431173	90533268	PARM1				

■ Chèvres des Fossés

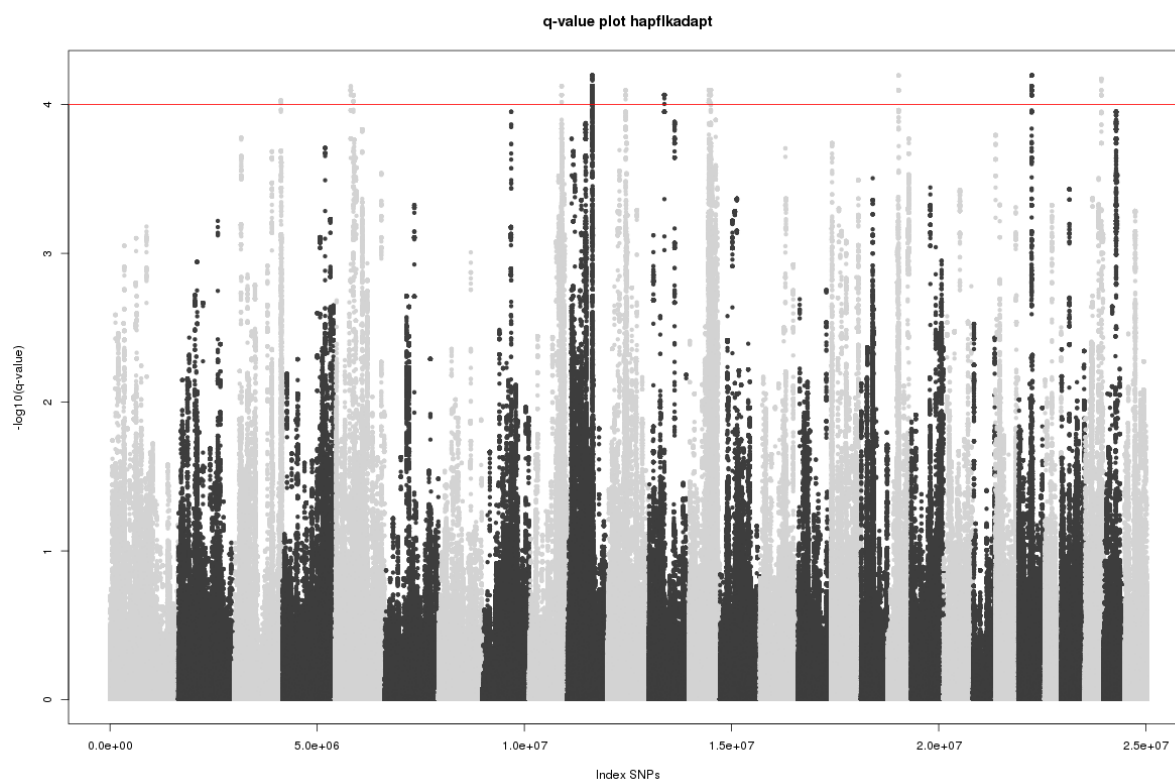


Figure S35 : Résultats du test hapFLK_{adapt} sur l'ensemble du génome de la race des Fossés. L'axe des ordonnées représente la significativité des SNPs et l'axe des abscisses la position sur le génome avec une alternance de couleurs en fonction des différents chromosomes. La ligne horizontale correspond au seuil de q-valeur 1.10^{-4} .

Tableau S5 : Régions sélectionnées mises en évidence chez la race des Fossés par le programme hapFLK_{adapt} ainsi que les gènes associés.

Chromosome	Début région	Fin région	Gènes
3	116356886	116458797	LMX1A
5	35641955	35745094	TMEM117
5	41221991	41328592	KIF21A
9	79335338	79440987	NA
10	68355101	68470534	NA
10	68608359	68806686	NA
10	69202115	69329617	NA
10	69473793	69583347	NA
11	45710375	45814513	NA
12	31444062	31548746	NDFIP2
13	51942794	52049912	TMC2
13	51942794	52049912	SNRPB
13	58448958	58630158	NA
19	23826129	23935422	LOC102183999
19	23826129	23935422	LOC108638211
19	23826129	23935422	LOC102184552
24	28733702	28866126	NA
27	39929391	40032270	NA

■ Chèvres de Lorraine

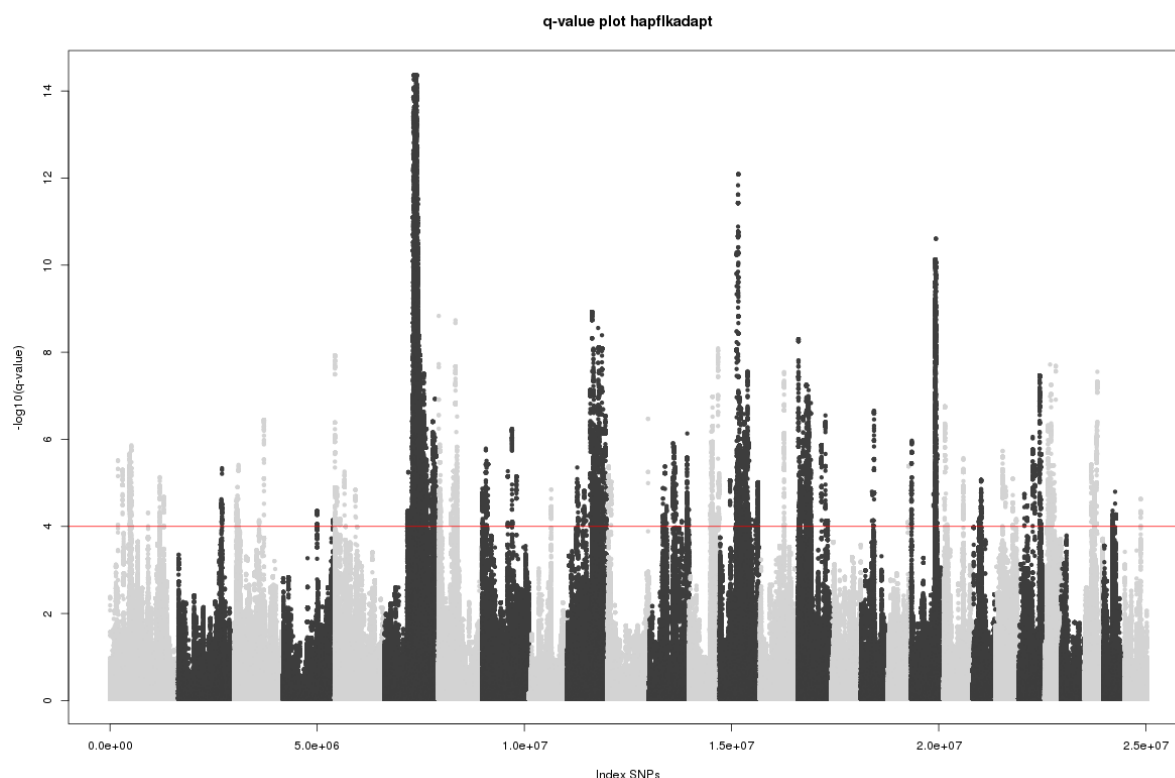


Figure S36 : Résultats du test hapFLK_{adapt} sur l'ensemble du génome de la race Lorraine. L'axe des ordonnées représente la significativité des SNPs et l'axe des abscisses la position sur le génome avec une alternance de couleurs en fonction des différents chromosomes. La ligne horizontale correspond au seuil de q-valeur 1.10^{-4} .

Tableau S6 : Régions sélectionnées mises en évidence chez la race Lorraine par le programme hapFLK_{adapt} ainsi que les gènes associés.

Chromosome	Début région	Fin région	Gènes
6	61146581	61393817	KIT
6	61465725	62143294	
6	62244279	62896389	
6	63278725	63425095	
6	63733296	64542731	
6	64619855	64906685	
6	64914924	65340431	
6	65758751	65888359	
6	66025006	66443997	
6	66499602	67038829	
6	67159376	67853176	
6	68083156	68290808	
6	68369902	69338965	
6	69382803	69502265	
6	69540085	69793544	
6	70434786	70589041	
6	70695626	71198978	
6	73327898	74204553	
14	42506670	42781175	
14	43320222	43571251	
			PHOX2B
			SLC30A9
			BEND4
			SHISA3
			ATP8A1
			GRXCR1
			YIPF7
			GUF1
			GNPDA2
			TRNAW-CCA
			GABRG1
			GABRA4
			GABRB1
			COMMD8
			ATP10D
			CORIN
			NFXL1
			CNGA1
			NIPAL1
			TXK
			TEC
			SLAIN2
			SLC10A4
			ZAR1
			FRYL
			CWH43
			LOC108636223
			DCUN1D4
			SGCB
			LOC106502204
			SPATA18
			LOC102169845
			LOC102179181
			USP46
			SCFD2
			LNK1
			LOC102174283
			KIT
			LOC102174549
			LOC108636277
			LOC102180298
			LOC106502205
			KDR
			TRNAC-GCA
			NA
			HNF4G

■ Chèvres Poitevines

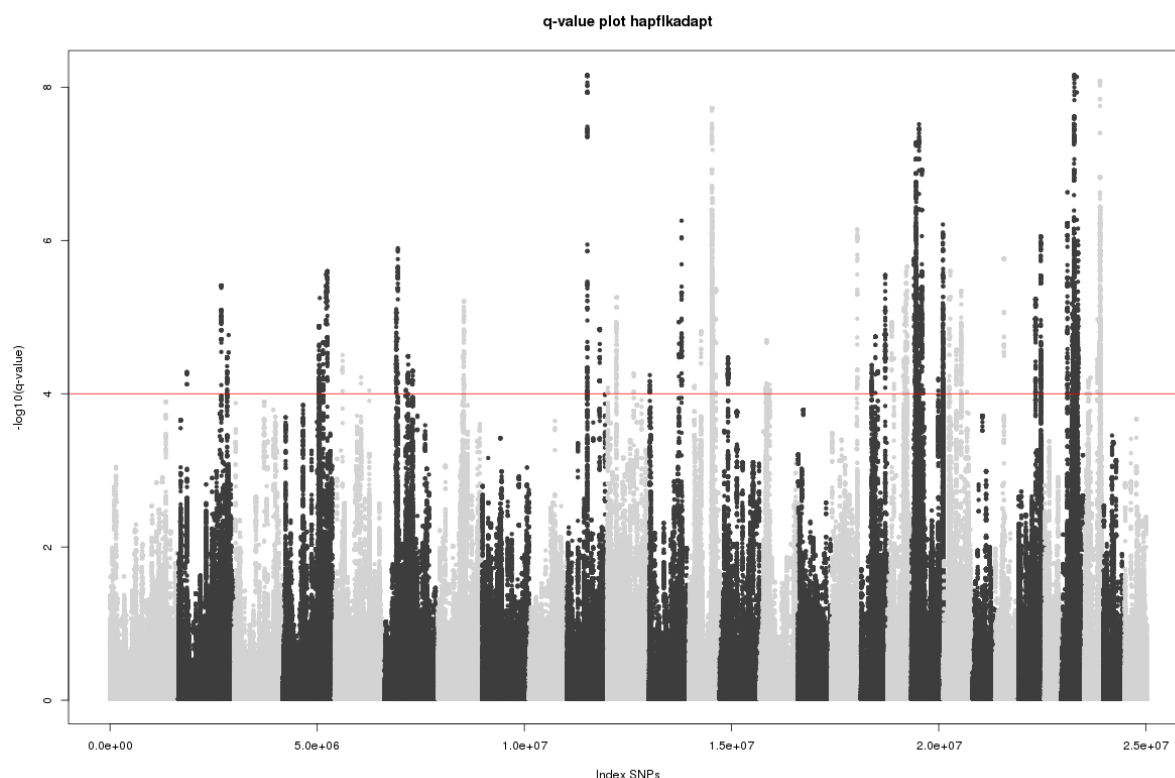


Figure S37 : Résultats du test hapFLK_{adapt} sur l'ensemble du génome de la race Poitevine. L'axe des ordonnées représente la significativité des SNPs et l'axe des abscisses la position sur le génome avec une alternance de couleurs en fonction des différents chromosomes. La ligne horizontale correspond au seuil de q-valeur 1.10^{-4} .

Tableau S7 : Régions sélectionnées mises en évidence chez la race Poitevine par le programme hapFLK_{adapt} ainsi que les gènes associés.

Chr	Début région	Fin région	Gènes	Chr	Début région	Fin région	Gènes
10	55465495	55599653	TLN2	26	14543131	14690205	SHTN1
12	74283482	74394998	DGKH	26	32113217	32242729	PYROXD2
13	62030547	62327871	LOC102170799	26	32813244	33292373	ANKRD2
13	62030547	62327871	LOC108637355	26	32813244	33292373	ARHGAP19
13	62030547	62327871	LOC108637412	26	32813244	33292373	EXOSC1
13	62799733	62966173	CHMP4B	26	32813244	33292373	FRAT1
13	62799733	62966173	TRNAG-UCC	26	32813244	33292373	FRAT2
13	64628434	64938065	CNBD2	26	32813244	33292373	LOC102170110
13	64628434	64938065	EPB41L1	26	32813244	33292373	PGAM1
13	64628434	64938065	SCAND1	26	32813244	33292373	RRP12
17	63694068	63856646	LRBA	26	32813244	33292373	SLIT1
20	9639943	9744637	NA	26	32813244	33292373	UBTD1
20	10612116	11090469	LOC102188442	26	32813244	33292373	ZDHHC16
20	11124144	11496056	PIK3R1	26	38972848	39076641	NA
20	18005723	18121867	NA	26	40597826	40755492	ACTA2
20	25684871	25889167	LOC102171377	26	40597826	40755492	FAS
20	70062260	70262020	NA	27	36819482	37075769	LOC106503660
24	53604263	53821760	LOC102185135	27	36819482	37075769	NEIL3
				27	37876217	38061394	GPM6A

■ Chèvres des Pyrénées

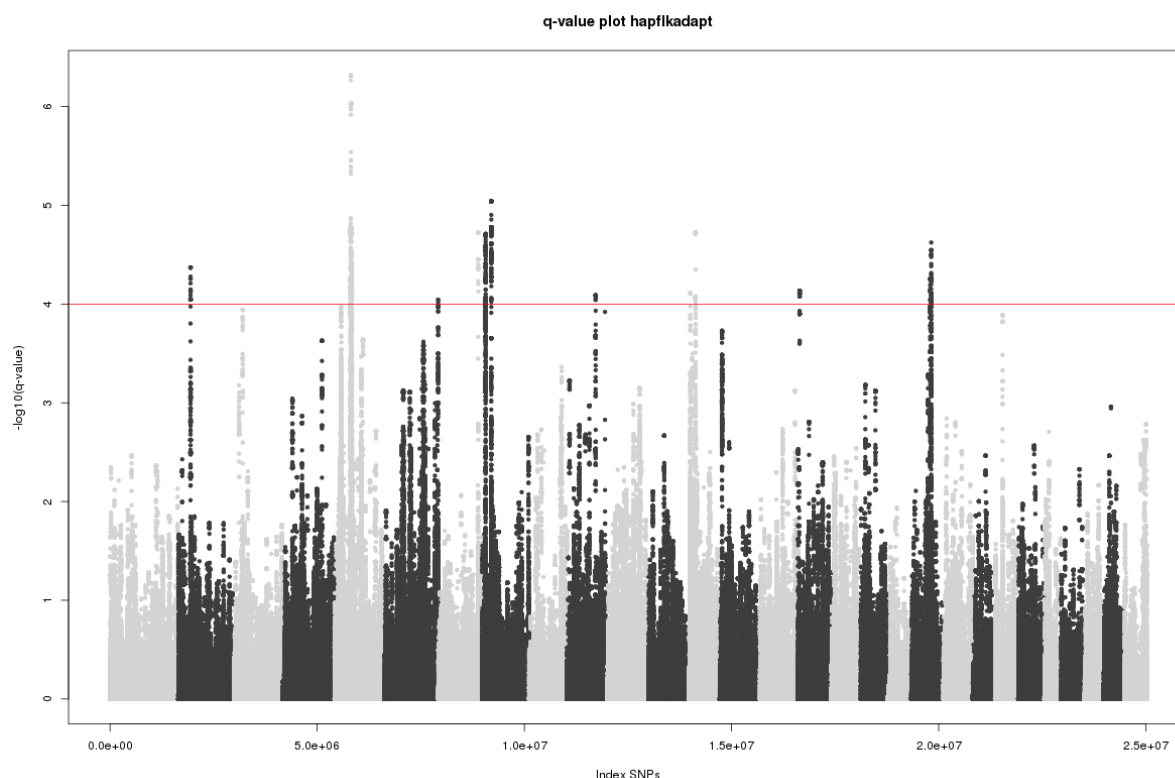


Figure S38 : Résultats du test $\text{hapFLK}_{\text{adapt}}$ sur l'ensemble du génome de la race des Pyrénées. L'axe des ordonnées représente la significativité des SNPs et l'axe des abscisses la position sur le génome avec une alternance de couleurs en fonction des différents chromosomes. La ligne horizontale correspond au seuil de $q\text{-valeur } 1.10^{-4}$.

Tableau S8 : Régions sélectionnées mises en évidence chez la race des Pyrénées par le programme $\text{hapFLK}_{\text{adapt}}$ ainsi que les gènes associés.

Chromosome	Début région	Fin région	Gènes
2	32236797	32340564	NA
5	35395229	35597782	TMEM117
5	35641362	35916923	TMEM117
5	37538880	37641322	PRICKLE1
5	38026079	38189534	GXYLT1
5	38026079	38189534	LOC102185633
6	116869636	116971276	HTT
7	100257878	100359824	LOC108636448
7	100257878	100359824	LOC102170309
8	5036182	5192518	GALNTL6
8	15525737	15684144	NA
10	76581440	76684977	LOC102175026
10	76581440	76684977	LOC102174398
13	3108389	3210073	LOC102181967
13	14030824	14172676	NA
16	3528631	3641878	SRGAP2
20	46523664	46635968	CDH9
20	48432898	48569725	NA

■ Chèvres Saanen

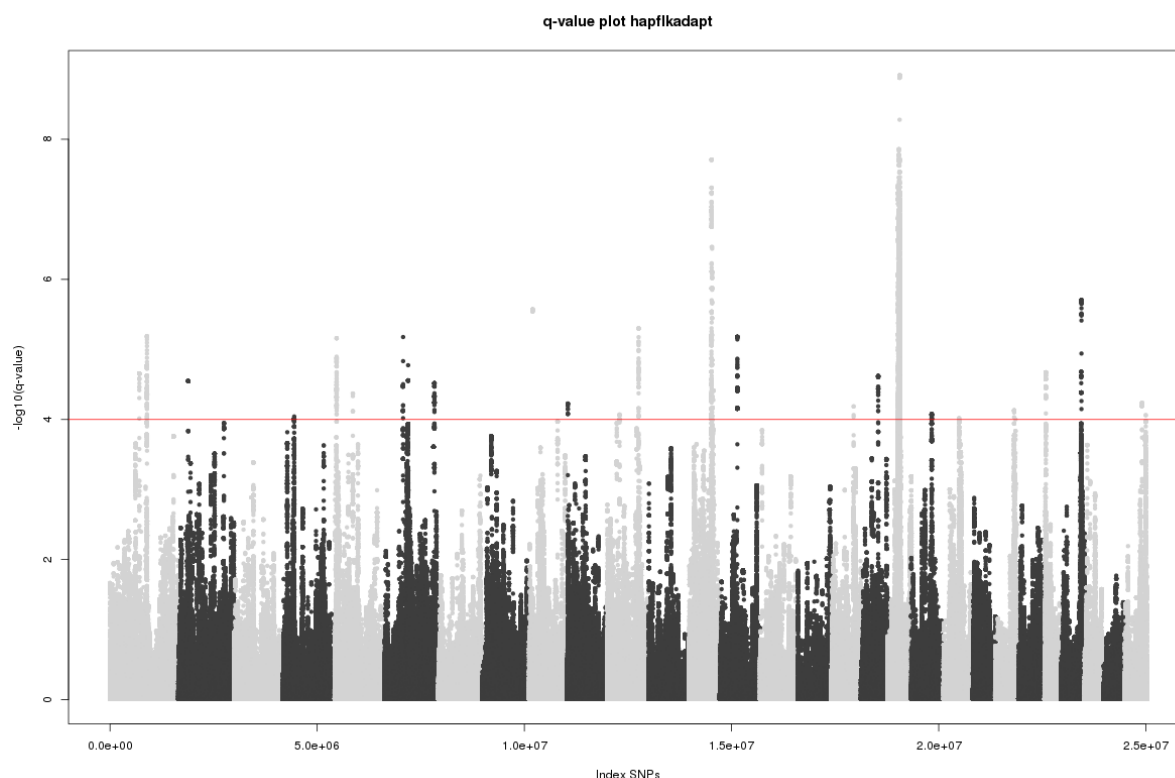


Figure S39 : Résultats du test hapFLK_{adapt} sur l'ensemble du génome de la race Saanen. L'axe des ordonnées représente la significativité des SNPs et l'axe des abscisses la position sur le génome avec une alternance de couleurs en fonction des différents chromosomes. La ligne horizontale correspond au seuil de q-valeur 1.10^{-4} .

Tableau S9 : Régions sélectionnées mises en évidence chez la race Saanen par le programme hapFLK_{adapt} ainsi que les gènes associés.

Chromosome	Début région	Fin région	Gènes
9	6000923	6108405	NA
13	61221190	61350799	NOL4L
13	61528458	61635983	NA
13	62837560	62944981	NA
13	63181002	63291187	ASIP
19	21189939	21535631	
19	21962997	22275999	
19	22880724	23040249	
19	23059182	23321726	
19	23644997	23977093	
19	24059174	24537314	
19	24646485	24847200	
19	25092857	25316873	
19	25376298	26382767	
19	26483389	27352361	
19	27365518	27467804	
19	27665915	28179883	
19	28289110	29115623	
19	29162682	29420662	
26	46666235	46775292	PCDH15

QTL
pléiotrope

TMIGD1
CPD
GOSR1
TUSC5
BHLHA9
ABR
VPS53
TRNAC-GCA
C19H17orf97
RPH3AL
SMG6
TSR1
SGSM2
TRNAG-CCC
METTL16
LOC108638230
LOC102178481
LOC102178764
LOC102179041
LOC108638210
LOC102179591
LOC102179759
LOC102181809
LOC102182341
LOC102183177
LOC102183450
LOC102183723
LOC102183999
LOC108638211
LOC102184552

LOC102182625
LOC108638237
LOC102182904
LOC102191033
TRPV1
SHPK
CTNS
TAX1BP3
EMC6
P2RX5
ITGAE
NCBP3
CAMKK1
P2RX1
ATP2A3
ZZEF1
TRNAK-CUU
UBE2G1
SPNS3
SPNS2
PITPNM3
WSCD1
NLRP1
DHX33
C1QBP
RPAIN
NUP88
RABEP1
SCIMP
LOC102184193

ZNF232
ZFP3
TRNAS-UGA
KIF1C
RNF167
SLC25A11
GP1BA
LOC108638265
CHRNE
MINK1
LOC102178853
LOC102186996
ALOX12
RNASEK
C19H17orf49
MIR195
MIR497
BCL6B
LOC102180151
SLC16A11
LOC106503212
ASGR2
LOC106503213
ASGR1
DLG4
GABARAP
CTDNEP1
ELP5
CLDN7
SLC2A4

YBX2
EIF5A
GPS2
NEURL4
LOC102184370
ACAP1
KCTD11
TMEM95
TNK1
PLSCR3
TMEM256
NLGN2
POLR2A
TNFSF12
TNFSF13
SENP3
EIF4A1
CD68
MPDU1
SOX15
FXR2
SAT2
SHBG
TP53
WRAP53
EFNB3
DNAH2
KDM6B
TMEM88
NAA38

LOC102170907
LOC108638268
CHD3
GUCY2D
KRBA2
RPL26
RNF222
NDEL1
MYH10
CCDC42
MFSD6L
PIK3R6
PIK3R5
NTN1
STX8
CFAP52
LOC106503225
USP43
TRNAW-CCA
GLP2R
RCVRN
GAS7
MYH13
LOC102181426
LOC102181155
LOC108638272
LOC106503236
LOC102174428
ADPRM

■ Chèvres de Savoie

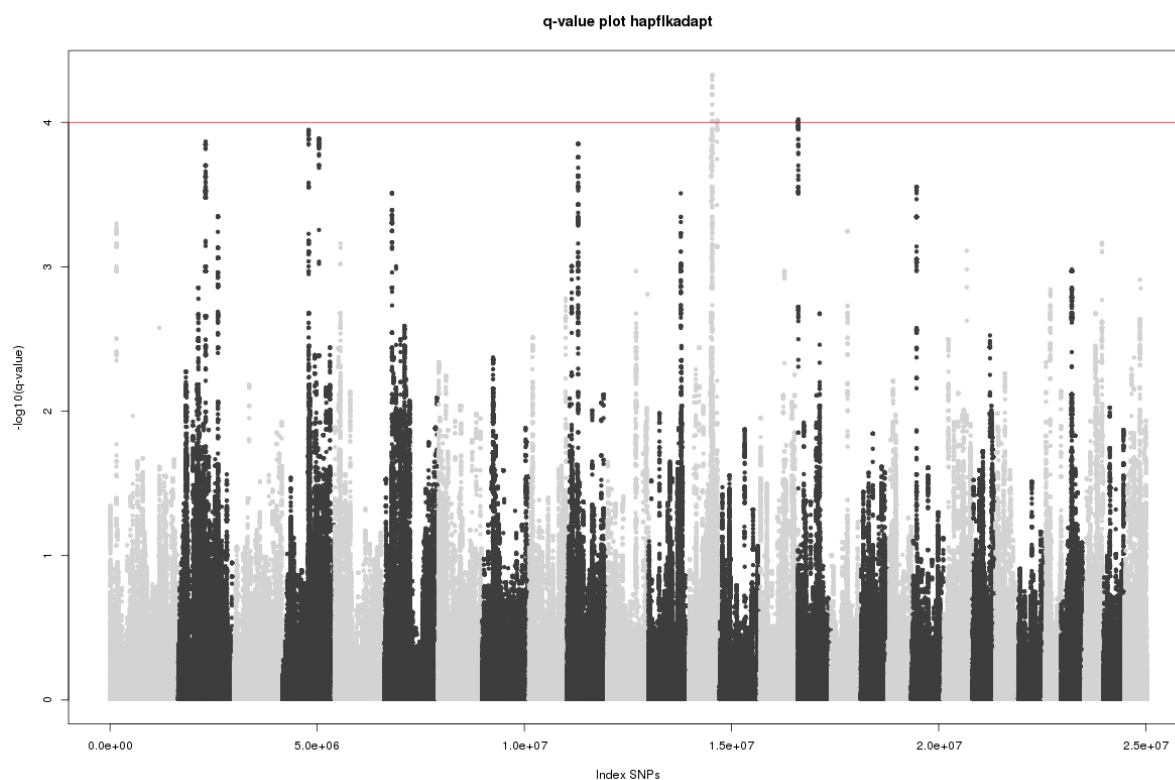


Figure S40 : Résultats du test $\text{hapFLK}_{\text{adapt}}$ sur l'ensemble du génome de la race de Savoie. L'axe des ordonnées représente la significativité des SNPs et l'axe des abscisses la position sur le génome avec une alternance de couleurs en fonction des différents chromosomes. La ligne horizontale correspond au seuil de $q\text{-valeur } 1.10^{-4}$.

Tableau S10 : Régions sélectionnées mises en évidence chez la race de Savoie par le programme $\text{hapFLK}_{\text{adapt}}$ ainsi que les gènes associés.

Chromosome	Région_start	Région_stop	Gènes
13	63234334	63369329	AHCY
13	76456335	76566451	PREX1
16	606511	710361	OPTC

■ Chèvres Provençales

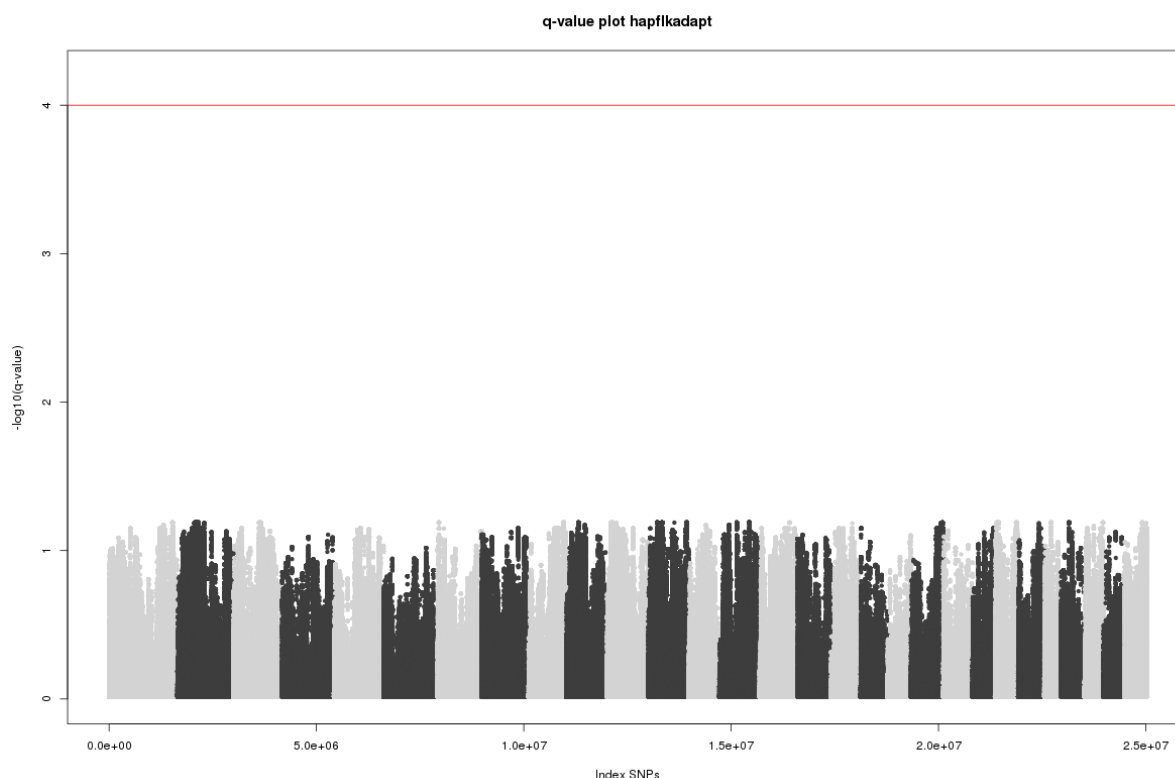


Figure S41 : Résultats du test $\text{hapFLK}_{\text{adapt}}$ sur l'ensemble du génome de la race Provençale. L'axe des ordonnées représente la significativité des SNPs et l'axe des abscisses la position sur le génome avec une alternance de couleurs en fonction des différents chromosomes. La ligne horizontale correspond au seuil de q -valeur 1.10^{-4} .

■ Chèvres du Rove

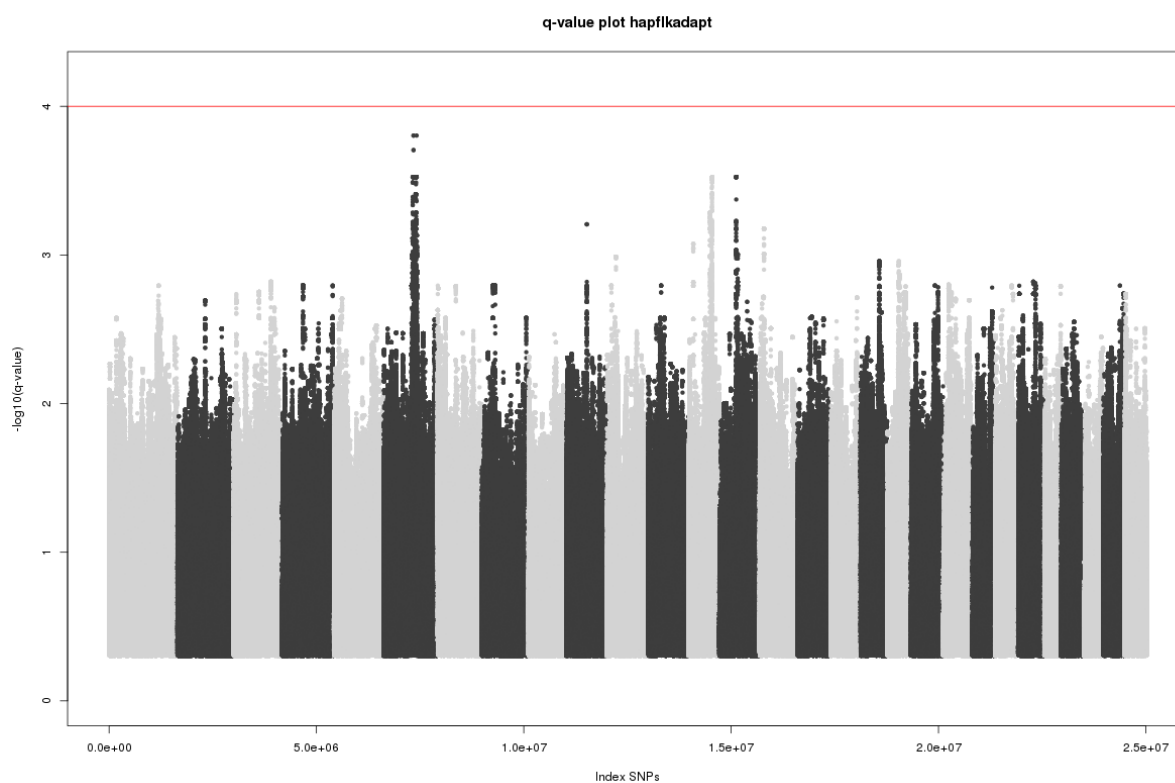


Figure S42 : Résultats du test $\text{hapFLK}_{\text{adapt}}$ sur l'ensemble du génome de la race Saanen. L'axe des ordonnées représente la significativité des SNPs et l'axe des abscisses la position sur le génome avec une alternance de couleurs en fonction des différents chromosomes. La ligne horizontale correspond au seuil de q -valeur 1.10^{-4} .

II. Mécanismes épigénétiques et acclimatation

Tableau S11 : Informations détaillées pour chaque chèvre et mouton marocain analysés comportant l'espèce, le nom de l'individu, sa race, ses coordonnées GPS ainsi que la valeur de la variation de température annuelle associée.

Species	Sample name	Breeds	Longitude	Latitude	BIO7	Species	Sample name	Breeds	Longitude	Latitude	BIO7
Goat	MOCH-K13-0366	Other	-9.70159	31.21258	158	Sheep	MOOA-K13-0333	Other	-9.69874	31.21548	159
Goat	MOCH-K14-0425	Noire-Atlas	-9.8197	30.7366	164	Sheep	MOOA-K14-0398	Other	-9.83761	30.70932	163
Goat	MOCH-J17-1355	Other	-10.11372	29.09516	185	Sheep	MOOA-J17-1384	Other	-10.1135	29.09419	185
Goat	MOCH-K15-0440	Other	-9.57139	30.25939	198	Sheep	MOOA-K15-0407	Other	-9.57139	30.25939	198
Goat	MOCH-H19-1343	Draa	-11.05555	28.26907	202	Sheep	MOOA-M9-3196	Other	-8.5516	33.183	207
Goat	MOCH-S4-0026	Nord	-5.87	35.73	203	Sheep	MOOA-L11-0291	Sardi	-9.20446	32.05161	212
Goat	MOCH-O8-3153	Other	-7.58	33.58	207	Sheep	MOOA-M10-3200	Other	-8.865	32.85	213
Goat	MOCH-L10-3100	Other	-9.058	32.7	209	Sheep	MOOA-S4-3009	Other	-5.85	35.64	215
Goat	MOCH-Z5-2083	Other	-2.369167	35.11667	220	Sheep	MOOA-I19-1360	Other	-10.52004	28.35465	215
Goat	MOCH-L12-0379	Others	-9.21121	31.62354	220	Sheep	MOOA-L14-0391	Other	-9.32916	30.52825	220
Goat	MOCH-R13-1104	Draa	-6.02545	31.17796	390	Sheep	MOOA-Z5-2082	Other	-2.369167	35.11667	220
Goat	MOCH-P15-1186	Other	-7.15083	30.30812	391	Sheep	MOOA-L12-0351	Sardi	-9.21121	31.62354	220
Goat	MOCH-W12- 1051	Other	-3.51098	31.57536	392	Sheep	MOOA-U11-1027	Rahalya	-4.59069	32.14685	373
Goat	MOCH-T13-0128	Noire-Atlas	-5.1657	31.1077	392	Sheep	MOOA-O14-1228	Other	-7.52765	30.52013	374
Goat	MOCH-S13-1064	Ghazalia	-5.52492	31.2501	395	Sheep	MOOA-R12-0010	Rahalya	-6.4135	31.6739	377
Goat	MOCH-U13-1059	Draa	-4.52598	31.03844	398	Sheep	MOOA-O15-1215	Other	-7.54106	30.27412	378
Goat	MOCH-W13-1009	Ghazalia	-3.57073	31.04187	401	Sheep	MOOA-R12-0013	Other	-6.401	31.601	389
Goat	MOCH-R14-1105	Ghazalia	-6.18389	30.56283	402	Sheep	MOOA-P15-1211	Other	-7.15144	30.30759	391
Goat	MOCH-Q14-1167	Draa	-6.51941	30.52099	404	Sheep	MOOA-R14-1133	Dman	-6.18382	30.56283	402
Goat	MOCH-Q16-1147	Other	-6.50726	29.5435	406	Sheep	MOOA-Q14-1195	Dman	-6.51935	30.5211	404
Goat	MOCH-S15-1165	Draa	-5.55291	30.25376	412	Sheep	MOOA-Q15-1173	Dman	-6.57552	30.08149	405
						Sheep	MOOA-T14-1115	Dman	-5.21574	30.57319	406

Tableau S12 : Corrélation entre BIO7 et les différentes variables climatiques récupérées sur le site WordClim (www.worldclim.org)

Env	bio_7	Env	bio_7
bio_7	1,00	Du216	-0,03
bio_4	0,98	Plan Curvature	-0,07
bio_2	0,97	tmean_10	-0,09
tmax_7	0,91	tmax_11	-0,10
bio_5	0,90	prec_5	-0,12
tmax_8	0,89	tmin_9	-0,15
tmax_6	0,84	tmin_5	-0,17
tmean_7	0,76	tmean_3	-0,19
tmax_9	0,75	prec_10	-0,19
tmean_8	0,73	Curvature	-0,20
tmax_5	0,71	tmax_1	-0,20
bio_10	0,69	Du2112	-0,21
Altitude	0,64	Sky View Factor	-0,26
TI216	0,64	Aspect	-0,29
tmean_6	0,62	Profile Curvature	-0,30
DI216	0,57	tmax_12	-0,31
prec_8	0,55	prec_4	-0,36
tmax_4	0,53	tmin_4	-0,37
tmean_9	0,40	tmean_2	-0,38
tmean_5	0,38	tmean_11	-0,39
prec_7	0,37	prec_3	-0,39
bio_14	0,37	bio_12	-0,41
tmin_7	0,34	prec_11	-0,42
bio_9	0,34	tmin_10	-0,46
tmin_8	0,31	bio_11	-0,51
tmax_10	0,30	bio_16	-0,51
TI2112	0,29	bio_15	-0,52
prec_9	0,29	bio_13	-0,52
bio_17	0,25	tmin_3	-0,53
tmax_3	0,22	tmean_1	-0,55
DI2112	0,18	prec_12	-0,56
bio_18	0,14	bio_19	-0,57
bio_1	0,14	prec_2	-0,57
Slope	0,13	tmean_12	-0,57
tmean_4	0,13	prec_1	-0,57
Vector Terrain Ruggedness	0,11	tmin_11	-0,60
bio_8	0,08	tmin_2	-0,66
tmin_6	0,07	bio_3	-0,66
prec_6	0,02	tmin_12	-0,75
tmax_2	0,00	bio_6	-0,75
		tmin_1	-0,75

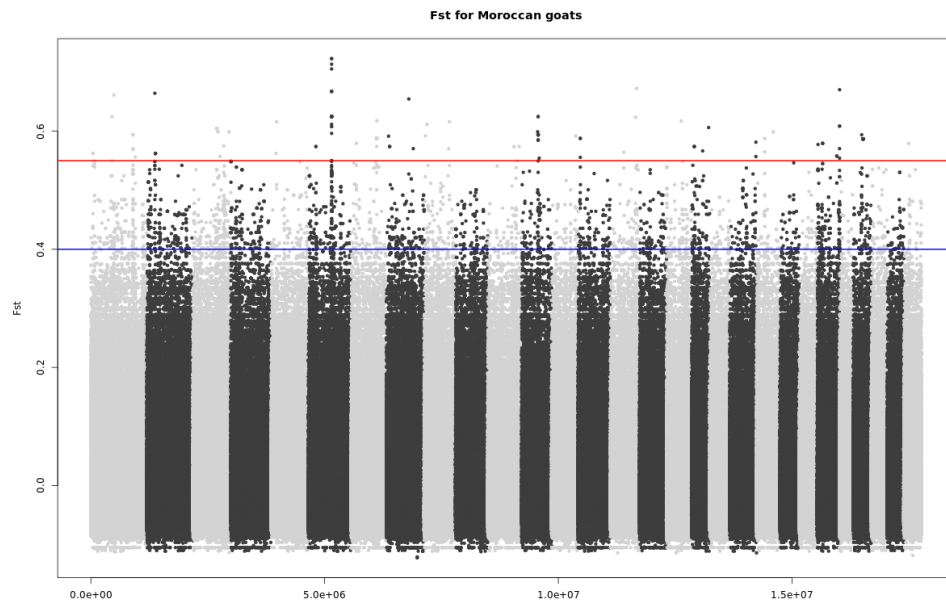


Figure S43 : Résultats de l'analyse F_{ST} contrastant les chèvres marocaines suivant la variation de température annuelle. L'axe des ordonnées représente la valeur de la différence génétique et l'axe des abscisses la position sur le génome avec une alternance de couleurs en fonction des différents chromosomes. La ligne horizontale bleue correspond au seuil pour l'analyse « pathway approach » (top 0.025%) et la rouge pour l'analyse « gene approach » (top 0.001%).

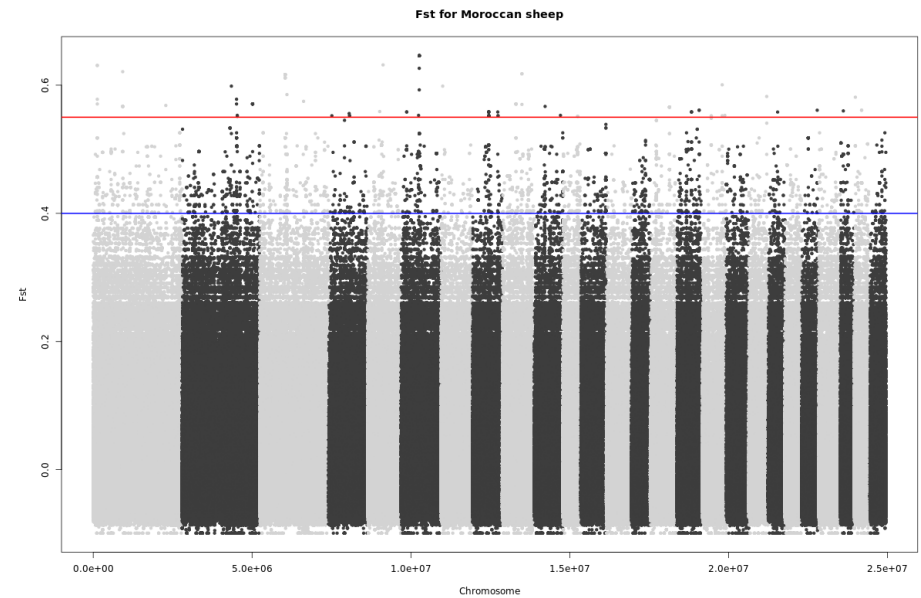


Figure S44 : Résultats de l'analyse F_{ST} contrastant les moutons marocains suivant la variation de température annuelle. L'axe des ordonnées représente la valeur de la différence génétique et l'axe des abscisses la position sur le génome avec une alternance de couleurs en fonction des différents chromosomes. La ligne horizontale bleue correspond au seuil pour l'analyse « pathway approach » (top 0.025%) et la rouge pour l'analyse « gene approach » (top 0.001%).

Pathway approach

Résultats des chèvres marocaines

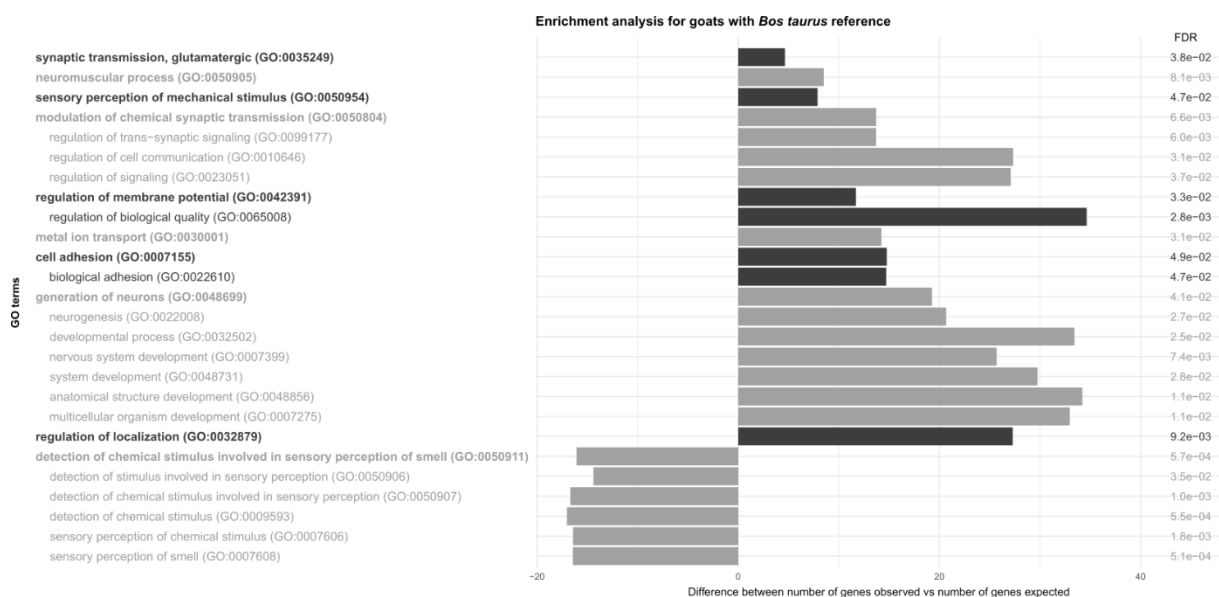


Figure S45 : GO termes significativement enrichis en comparant les gènes extraits chez les chèvres (top gènes au seuil 0,001%) par rapport à la référence *Bos taurus*. Les GO termes les plus basaux sont en gras.

Tableau S13 : Voies biologiques associées au GO termes le plus basaux des voies enrichies en comparant les gènes extraits chez les chèvres (top gènes au seuil 0,001%) par rapport à la référence *Bos taurus*.

GO terms	FDR	Parental biological process
synaptic transmission, glutamatergic	$3,81.10^{-2}$	cellular process ; signaling
neuromuscular process	$8,07.10^{-3}$	multicellular organismal process
sensory perception of mechanical stimulus	$4,73.10^{-2}$	multicellular organismal process
modulation of chemical synaptic transmission	$6,64.10^{-2}$	biological regulation ; cellular process
regulation of membrane potential	$3,27.10^{-2}$	biological regulation
metal ion transport	$3,11.10^{-2}$	localization
cell adhesion	$4,86.10^{-2}$	cellular process ; biological adhesion
generation of neurons	$4,08.10^{-2}$	cellular process ; multicellular organismal process ; developmental process
regulation of localization	$9,24.10^{-2}$	biological regulation ; localization
detection of chemical stimulus involved in sensory perception of smell	$5,75.10^{-2}$	response to stimulus ; multicellular organismal process

Résultats des moutons marocains

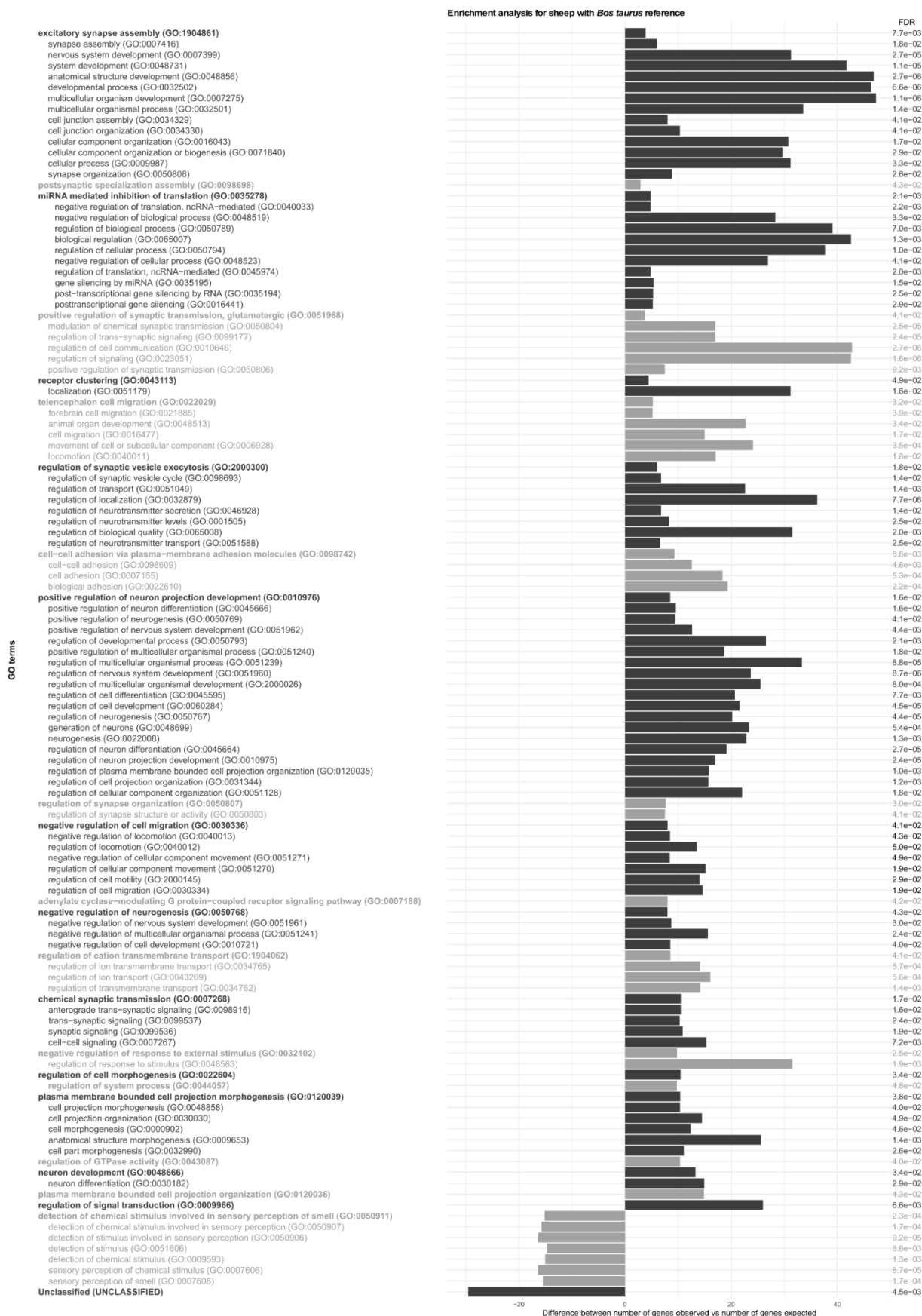


Figure S46 : GO termes significativement enrichis en comparant les gènes extraits chez les moutons (top gènes au seuil 0,001%) par rapport à la référence *Bos taurus*. Les GO termes les plus basaux sont en gras.

Tableau S14 : Voies biologiques associées au GO termes le plus basaux des voies enrichies en comparant les gènes extraits chez les moutons (top gènes au seuil 0,001%) par rapport à la référence Bos taurus.

GO_term	FDR	Biological_process
excitatory synapse assembly	$7,74.10^{-3}$	cellular process ; multicellular organismal process ; developmental process
postsynaptic specialization assembly	$4,28.10^{-2}$	cellular process ; multicellular organismal process ; developmental process
miRNA mediated inhibition of translation	$2,09.10^{-3}$	metabolic process ; cellular process ; biological regulation
positive regulation of synaptic transmission, glutamatergic receptor clustering	$4,08.10^{-2}$	biological regulation ; cellular process ; signaling
telencephalon cell migration	$4,94.10^{-2}$	cellular process ; localization
regulation of synaptic vesicle exocytosis	$3,20.10^{-2}$	developmental process ; multicellular organismal process ; cellular process ; localization ; locomotion
cell-cell adhesion via plasma-membrane adhesion molecules	$1,82.10^{-2}$	biological regulation ; cellular process ; localization ; signaling
positive regulation of neuron projection development	$8,55.10^{-3}$	cellular process ; biological adhesion
regulation of synapse organization	$1,57.10^{-2}$	biological regulation ; developmental process ; multicellular organismal process ; cellular process
negative regulation of cell migration	$2,98.10^{-2}$	biological regulation ; cellular process
adenylate cyclase-modulating G protein-coupled receptor signaling pathway	$4,13.10^{-2}$	biological regulation ; cellular process ; localization
negative regulation of neurogenesis	$4,22.10^{-2}$	biological regulation ; cellular process ; signaling ; response to stimulus
regulation of cation transmembrane transport	$4,27.10^{-2}$	biological regulation ; developmental process ; multicellular organismal process ; cellular process
chemical synaptic transmission	$4,06.10^{-2}$	biological regulation ; cellular process ; localization
negative regulation of response to external stimulus	$1,67.10^{-2}$	cellular process ; signaling
regulation of cell morphogenesis	$2,48.10^{-2}$	biological regulation ; response to stimulus
regulation of system process	$3,41.10^{-2}$	biological regulation ; developmental process
plasma membrane bounded cell projection morphogenesis	$4,79.10^{-2}$	biological regulation ; multicellular organismal process
regulation of GTPase activity	$3,79.10^{-2}$	developmental process ; cellular process
neuron development	$3,99.10^{-2}$	biological regulation
plasma membrane bounded cell projection organization	$3,44.10^{-2}$	developmental process ; multicellular organismal process ; cellular process
regulation of signal transduction	$4,27.10^{-2}$	cellular process
detection of chemical stimulus involved in sensory perception of smell	$6,55.10^{-3}$	biological regulation ; cellular process ; response to stimulus
Unclassified	$2,26.10^{-4}$	biological regulation
	$4,48.10^{-3}$	unknown

Gene approach

Résultats des chèvres marocaines

Tableau S15 : Voies biologiques et phénotypes associés aux gènes extraits chez les chèvres (top gènes au seuil 0,025%) par rapport à la référence Bos taurus.

Chromosome	Region start	Region stop	Gene	Biological pathway	Phenotype category
1	5286019	5291254	NA		
1	53132888	53137888	NA		
1	58696209	58701209	NA		
1	116640037	116656446	CP	biological regulation ; cellular process ;	Production
1	116663633	116668633		localization ; metabolic process	
1	123089669	123094669	SLC9A9	biological regulation ; cellular process ; localization	Production
2	19444789	19449789	NA		
2	21481517	21492023	NA		
3	71155644	71162075	NA		
3	74958486	74963506	NA		
3	81724175	81729175	NA		
3	94930069	94935069	TRABD2B	biological regulation ; cellular process ; response to stimulus ; signaling ; metabolic process	Other
3	111116952	111121952	NA		
5	10526701	10531701	PPFIA2	biological regulation ; cellular process ; developmental process ; multicellular organismal process ; localization	Production
5	11943652	11948652	TMTC2	biological regulation ; cellular process ; metabolic process	Environmental perception ; Production
6	17048237	17053556	NA		
6	60786586	60802588	NA		
6	60810915	60816982	NA		
6	60820541	60825631	NA		
7	14423240	14428240	TNFSF9	biological regulation ; cellular process ; immune system process ; developmental process ; biological adhesion ; response to stimulus ; signaling ; multicellular organismal process	Other
7	79951464	79965821	EDIL3	biological regulation ;	Environmental perception
7	79970611	79976237		cellular process ;	
7	80325262	80330262		biological adhesion	

8	4656423	4661423	GALNTL6	No Biological Process information available	Production
8	7336391	7343638	CTSB	developmental process ; cellular process ; response to stimulus ; biological regulation ; metabolic process ; multiorganism process ; reproduction ; multicellular organismal process	Production
8	65759151	65764151	NA		
8	80441599	80446599	NA		
9	2965763	2970763	NA		
9	6490988	6495988	NA		
9	71411538	71416538	SASH1	response to stimulus ; cellular process ; metabolic process ; biological regulation ; signaling ; multicellular organismal process ; localization ; locomotion ; developmental process	Environmental perception ; Production
9	73645306	73650306	NA		
11	80150734	80155735	NA		
11	95730618	95735618	MVB12B	biological regulation ; response to stimulus ; cellular process ; signaling	Other
12	45274673	45279673	NA		
12	46688101	46696577	KLF12	biological regulation ; cellular process ; metabolic process	Other
12	49573144	49578144	NA		
13	74454261	74459261	NA		
14	4998305	5004299	ADGRB3	developmental process ; multicellular organismal process ; biological regulation ; response to stimulus ; cellular process ; signaling ; behavior	Reproduction
15	33302961	33307961	USH1C	developmental process ; multicellular organismal process ; cellular process ; localization ; biological regulation	Reproduction
15	68423693	68428693	NA		
15	70961905	70966905	NA		
17	40321099	40326099	NA		

18	7798511	7805266	PLCG2	metabolic process ; cellular process ; biological regulation ; localization ; multicellular organismal process ; response to stimulus ; immune system process ; developmental process ; locomotion ; signaling	Reproduction
18	36708024	36713024	NA		
18	57689477	57694477	MYADM	metabolic process ; biological regulation ; cellular process ; signaling ; response to stimulus ; developmental process ; multicellular organismal process ; biological adhesion ; localization ; locomotion	Reproduction
20	68971744	68981508	NA		
21	18928512	18933512	NA		
21	18933945	18938945	NA		
21	47741784	47746784	NA		
24	1944889	1949889	NA		
24	12789771	12795697	NA		
24	54399735	54405063	NA		
24	61552341	61557341	LOC102184299		
24	61640196	61645232	SERPINB7	metabolic process ; cellular process ; biological regulation	Reproduction
24	61700913	61705913	NA		
26	22334608	22339608	CALHM3	multicellular organismal process ; cellular process ; localization	Other
26	26210390	26219685	SORCS1	cellular process ; localization	Other
29	11738281	11743281	LOC106503718		
29	11738281	11743281	RAB30	biological regulation ; cellular process ; signaling ; response to stimulus	Other

Résultats des moutons marocains

Tableau S16 : Voies biologiques et phénotypes associés aux gènes extraits chez les moutons (top gènes au seuil 0,025%) par rapport à la référence Bos taurus.

Chromosome	Region start	Region stop	Gene	Biological pathway	Phenotype category
1	10687323	10692323	AGO3	metabolic process ;	Other
1	10696625	10708822		cellular process ; biological regulation	
1	92082650	92094108	NA		
1	226301475	226306475	IQCJ	No Biological Process information available	Other
2	157311694	157316694	NA		
2	173348060	173355100	NA		
2	175798932	175803932	NA		
2	175805660	175810660	NA		
2	224924298	224932214	NA		
3	81528550	81540796	KCNG3	localization ; cellular process ; biological regulation	Other
3	87341773	87346773	NA		
3	141099996	141104996	ANO6	localization ; cellular process ; biological regulation ; multicellular organismal process ; developmental process ; biomineralization ; response to stimulus ; signaling ; immune system process ; locomotion	Other
4	7017993	7022993	NA		
4	61497407	61502407	NA		
4	62672561	62677733	BMPEP	cellular process ; biological regulation ; response to stimulus ; signaling ; multicellular organismal process ; localization ; locomotion ; developmental process ; metabolic process	Production ; Reproduction
5	38381091	38386091	GCSAML	cellular process ; response to stimulus ; signaling ; immune system process ; biological regulation ; locomotion ; localization	Reproduction
5	53053705	53058705	NA		
6	15120664	15125690	NA		
6	51546154	51551154	LOC101107868		
6	53274965	53286912	NA		
7	4331512	4336512	NA		
8	47493701	47498701	NA		

8	47500909	47529293	CASP8AP2	metabolic process ; cellular process ; biological regulation ; response to stimulus ; signaling	Other
8	77772322	77777322	CNKS3	metabolic process ; cellular process ; biological regulation ; response to stimulus ; signaling ; localization	Other
8	78653039	78658050	NOX3	response to stimulus ; developmental process ; multicellular organismal process ; biological regulation ; metabolic process ; cellular process	Environmental perception
9	38392866	38402045	NA		
9	58487130	58492269	SAMD12	No Biological Process information available	Production ; Immunity
10	30967239	30972239	NA		
10	80059589	80064589	NA		
11	52324554	52329555	ENGASE	metabolic process ; cellular process	Other
15	58913215	58924042	KCNA4	localization ; cellular process ; biological regulation	Other
16	45028297	45033631	NA		
16	64022455	64028612	SEMA5A	locomotion ; response to stimulus ; developmental process ; multicellular organismal process ; cellular process ; localization ; growth ; biological regulation ; signaling	Immunity
17	24296308	24301336	NA		
17	56816440	56821440	KSR2	biological regulation ; response to stimulus ; cellular process ;	Production ; Reproduction
17	56824292	56829292		signaling ; metabolic process	
17	65753866	65758866	SEZ6L	developmental process ; multicellular organismal process ; biological regulation ; response to stimulus ; cellular process ; signaling ; behavior	Immunity
19	55972020	55977020	NA		
20	28601629	28606629	LOC101119001		
22	46336039	46341039	PTPRE	cellular process ; metabolic process	Production
24	8374897	8379897	NA		

25	3860306	3865306	NA
25	21600748	21605748	NA
25	21619196	21624196	NA

Résumé

Les organismes vivants peuvent être confrontés à des variations plus ou moins importantes de leur environnement (qu'il soit climatique, sanitaire, etc...). En réponse à ces variations, trois types de mécanismes peuvent être mis en place pour ajuster leur phénotype et l'environnement : le choix d'un habitat favorable, l'adaptation et l'acclimatation. Dans le cas des animaux domestiques, le choix de l'habitat n'est pas possible mais les deux derniers mécanismes peuvent être mis en évidence en étudiant respectivement la présence de marques génétiques et épigénétiques dans le génome des individus. Le but de cette thèse a été de mettre en évidence ces deux types de mécanismes chez les petits ruminants (chèvres et moutons) en rapport avec l'histoire des races et leur environnement.

Dans un premier temps, nous avons analysé des données de génomes complets pour chercher des signatures de sélection chez onze races caprines françaises. Cette analyse nous a d'abord permis d'explorer leurs histoires, de comprendre les croisements possibles entre elles et de confronter ces résultats aux données historiques collectées. De plus des gènes sous sélection en rapport avec des caractères d'intérêt agronomique tels que la production laitière (21 gènes), la reproduction (14 gènes), l'immunité (11 gènes), ainsi que des caractères morphologiques spécifiques aux races étudiées (28 gènes) ont été mis en évidence.

Dans un second temps, nous avons étudié les deux types de mécanismes chez des chèvres et des moutons marocains qui ont été choisis pour former deux groupes à chaque extrémité d'un gradient de variation de températures. L'analyse des différences génétiques entre les deux groupes, pour chaque espèce, nous a permis de localiser des régions sous sélection en relation avec des gènes impliqués dans la perception de l'environnement (5 gènes), l'immunité (4 gènes), la reproduction (8 gènes) et la production (11 gènes). Nous avons aussi séquencé les régions du génome portant des groupements méthyles chez ces mêmes animaux. L'analyse des régions différenciellement méthylées (DMR) entre les deux groupes nous a permis de trouver 2 DMRs (une chez chaque espèce) en relation, entre autre, avec la production et la qualité du lait. Cette étude des mécanismes d'adaptation et d'acclimatation chez les petits ruminants marocains est la première à chercher des marques épigénétiques en relation avec l'environnement chez des animaux d'élevage et à les comparer avec les marques génétiques présentes chez ces mêmes animaux. Au vu de nos résultats nous supposons que la variation de températures pourrait avoir deux types d'effet sur les animaux pouvant impacter les voies biologiques que nous avons détectées. Un premier effet, direct, qui influencerait les mécanismes de thermorégulation, et un second effet, indirect, en relation avec la quantité et la qualité des ressources alimentaires disponibles. La comparaison entre les deux mécanismes et les deux espèces nous a permis de trouver des voies biologiques impactées similaires, mais aucun gène en commun.

Ces résultats montrent le rôle des mécanismes génétiques et épigénétiques dans l'ajustement des phénotypes à l'environnement. Dans un contexte de changements climatiques, il semble important de les prendre en compte pour développer des stratégies d'élevage en lien avec ces variations.

Abstract

Organisms can be confronted with more or less important variations in their environment (climate, health, etc...). In response to these variations, three types of mechanisms can be put in place to adjust their phenotype with the environment: choice of a favorable habitat, adaptation and acclimatization. In the case of domestic animals, the choice of habitat is not possible, but the last two mechanisms can be revealed by studying the presence of respectively genetic and epigenetic markers in the genome of individuals. The aim of this thesis was to highlight these two types of mechanisms in small ruminants (sheep and goats) in relation to the breeds history and their environment.

First, we analyzed whole genome data to search selection signatures in 11 French goat breeds. On one hand, this analysis allowed us to explore their histories, to understand possible crossbreeding between them and to compare these results with the historical data collected. On the other hand, selected genes related to agronomic traits of interest such as milk production (21 genes), reproduction (14 genes), immunity (11 genes), as well as morphological traits specific to the breeds studied (28 genes) were highlighted.

In a second step, we studied the two types of mechanisms in Moroccan goats and sheep which were chosen to form two groups at each end of a temperature variation gradient. The analysis of genetic differences between the two groups, for each species, allowed to locate selected regions in relation to genes involved in environmental perception (5 genes), immunity (4 genes), reproduction (8 genes) and production (11 genes). We have also sequenced regions of the genome bearing methyl groups in these same animals. Analysis of the differentially methylated regions (DMRs) between the two groups allowed to find 2 DMRs (one in each species) in relation, among other things, to milk quality and production. This study of adaptation and acclimatization mechanisms in Moroccan small ruminants is the first to look for epigenetic marks in relation to the environment in farm animals and to compare them with genetic marks present in these same animals. Based on our results, we hypothesize that temperature variation could have two types of effects on animals that could impact the biological pathways we detected. A first effect, direct, which would influence the thermoregulation mechanisms, and a second effect, indirect, in relation to the quantity and quality of available food resources. The comparison between the two mechanisms, and the two species, allowed to find similar impacted biological pathways, but no gene in common.

These results show the role of genetic and epigenetic mechanisms in the adjustment of phenotypes to the environment. In a context of climate change, it seems important to take them into account to develop breeding strategies related to these variations.