



## THÈSE

Pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Biodiversité-Ecologie-Environnement

Arrêté ministériel : 25 mai 2016

Présentée par

**Charlotte HER**

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

préparée au sein du **Laboratoire d'ECologie Alpine** dans l'**École Doctorale Chimie et Sciences du Vivant**

**Apports de la datation d'évènements de mutation et de sélection à partir de génomes contemporains dans la compréhension d'un processus évolutif : la domestication du mouton**

**Contribution of the dating of mutation and selection events from contemporary genomes to the understanding of an evolutionary process: sheep domestication**

Thèse soutenue publiquement le **16 décembre 2021**, devant le jury composé de :

**Monsieur François POMPANON**

PROFESSEUR DES UNIVERSITÉS, Université Grenoble Alpes,  
Directeur de thèse

**Madame Emmanuelle VILA**

CHARGE DE RECHERCHE, CNRS délégation Rhône-Auvergne,  
Examinatrice

**Monsieur Olivier FRANÇOIS**

PROFESSEUR DES UNIVERSITÉS, Université Grenoble Alpes,  
Président

**Monsieur Simon BOITARD**

CHARGE DE RECHERCHE HDR, INRAE centre Occitanie-Montpellier,  
Rapporteur

**Monsieur Ludovic ORLANDO**

DIRECTEUR DE RECHERCHE, CNRS délégation Occitanie Ouest,  
Rapporteur

**Geneticists have (almost) created a time machine. It is now possible to work backwards from contemporary observations of genetic variation to inferences about past processes.**

Slatkin, M., Rannala, B., 2000. Estimating Allele Age. *Annual review of genomics and human genetics* 1, 225–49.

# Table des matières

<b>Introduction générale .....</b>	<b>10</b>
<b>1 La domestication animale.....</b>	<b>11</b>
Définition.....	11
Les différentes voies de domestication .....	11
Un modèle de processus évolutif .....	12
<b>2 Le cas de la domestication du mouton .....</b>	<b>13</b>
Les origines du mouton domestique .....	13
<b>Europe .....</b>	<b>21</b>
2.1.1 Description générale .....	21
2.1.2 La diffusion des moutons en Europe.....	22
2.1.3 Les moutons d'Europe du Nord en partie épargnés par le second épisode de migration .....	23
2.1.4 Les mouflons des îles méditerranéennes, reliques de la migration initiale .....	25
2.1.5 Les moutons ibériques .....	26
2.1.6 Les moutons des Balkans .....	29
2.1.7 Les moutons européens : bilan .....	30
<b>Afrique.....</b>	<b>31</b>
2.3.1. La diffusion des moutons en Afrique .....	32
2.3.2. Homogénéité génétique en Afrique du Nord.....	35
2.1.8 Les moutons africains : bilan .....	37
<b>Asie .....</b>	<b>37</b>
2.4.1. Chine .....	37
2.4.2. Inde .....	40
2.4.3. Népal .....	41
2.4.4. Russie .....	41
2.4.5. Mongolie .....	41
2.4.6. Introgression sauvage .....	42
2.1.9 Les moutons asiatiques : bilan .....	42
<b>Les gènes de la domestication.....</b>	<b>43</b>
<b>3 Objectifs et plan de la thèse.....</b>	<b>44</b>
<b>Présentation et préparation des données de génomes complets.....</b>	<b>47</b>
<b>1 Introduction .....</b>	<b>48</b>
<b>2 Composition des jeux de données .....</b>	<b>48</b>
2.1. Jeu de données de base.....	48
2.2. Jeu de données étendu.....	51
<b>3 Préparation des jeux de données .....</b>	<b>51</b>
Jeu de données de base.....	51

3.1.1.	Alignement .....	51
3.1.2.	Création des fichiers de variants génomiques .....	52
3.1.3.	Structure génétique neutre.....	56
3.1.4.	Phasage .....	60
3.1.5.	Etat ancestraux et dérivés .....	61
<b>Jeu de données étendu.....</b>		<b>64</b>
3.1.6.	Construction .....	64
3.1.7.	Structure.....	64
<b><i>Datation des évènements de mutations dans le génome du mouton.....</i></b>		<b>66</b>
<b>1</b>	<b><i>Introduction .....</i></b>	<b>67</b>
<b>2</b>	<b><i>Choix et description de la méthode de datation .....</i></b>	<b>68</b>
2.1.	Choix de la méthode.....	68
2.2.	Description de la méthode.....	70
<b>3</b>	<b><i>Mise en place de l'analyse .....</i></b>	<b>71</b>
3.1.	Adaptation de la méthode GEVA au cas du mouton.....	71
3.2.	Echantillonnage des individus .....	77
3.3.	Sélection des variants à dater .....	77
3.4.	Paramétrage des analyses de datation .....	77
3.4.1.	Paramètres GEVA .....	77
3.4.2.	Ressources computationnelles.....	78
3.4.3.	Architecture des répertoires de travail .....	78
3.5.	Filtration des estimations .....	78
<b>4</b>	<b><i>Distribution des âges des mutations dans le génome du mouton.....</i></b>	<b>79</b>
4.1.	Nombre de paires d'haplotypes analysées et de variants datés .....	79
4.2.	Comparaison des différentes horloges et choix de celle utilisée pour les analyses subséquentes.....	80
4.2.1.	Taux de recouvrement entre horloges.....	80
4.2.2.	Corrélations et différences entre les horloges avant le filtre qualité.....	81
4.2.3.	Choix de l'horloge pour les analyses .....	86
4.3.	Distributions des âges après choix de l'horloge et filtre qualité.....	88
4.3.1.	Nombre de variants datés le long du génome .....	88
4.3.2.	Distribution globale des âges des variants .....	88
4.4.	Disponibilité de l'Atlas.....	89
<b>5</b>	<b><i>Diffusion et connexions des populations au cours du temps .....</i></b>	<b>89</b>
5.1.	Formations des populations .....	89
5.1.1	Principes, objectifs et questions.....	89
5.1.2	Choix des populations .....	91
5.1.3	Approche stricte .....	91
5.1.4	Approche FST .....	100
5.1.5	Comparaison des approches stricte et FST .....	106
5.2.	Connexions des populations au cours du temps .....	108

5.2.1. Objectif.....	108
5.2.2. Méthode.....	109
5.2.3. Résultats.....	109
<b>6 Discussion .....</b>	<b>111</b>
6.1. Diffusion.....	111
6.2. Connexions entre les groupes au cours du temps.....	114
<b>7 Conclusion.....</b>	<b>115</b>
<b>Chronologie des évènements de sélection liés à la domestication du mouton.....</b>	<b>117</b>
<b>1. Introduction .....</b>	<b>118</b>
<b>2. Méthode de datation de sélection .....</b>	<b>119</b>
<b>3. Datations de variants potentiellement impliqués dans la domestication. ....</b>	<b>121</b>
3.1. Choix des variants .....	121
3.1.1. Groupe A : Signaux cosmopolites.....	121
3.1.2. Groupe B : Signaux intra-groupes .....	123
3.2. Datations de l'apparition et de la sélection des variants des groupes A et B .....	126
3.2.1. Groupe A .....	126
3.2.2. Groupe B .....	127
3.2.3. Paramètres de datation de sélection par startmrc .....	127
<b>4. Résultats .....</b>	<b>128</b>
4.1. Groupe A .....	128
4.1.1. Ages de la sélection des 25 régions.....	128
4.1.2. Comparaison des âges d'apparition et de sélection .....	132
4.2. Groupe B : .....	133
<b>5. Discussion .....</b>	<b>138</b>
5.1. Groupe A : Signaux cosmopolites .....	138
5.2. Groupe B : Signaux intragroupes .....	141
<b>6. Conclusion.....</b>	<b>144</b>
<b>Discussion générale .....</b>	<b>145</b>
<b>1. Récapitulatif des principaux résultats obtenus .....</b>	<b>146</b>
<b>2. Informations apportées par la combinaison des résultats des différents chapitres ...</b>	<b>149</b>
2.1. Origines de la sélection.....	149
2.2. Périodes de sélection versus périodes de flux de gènes .....	149
2.3. Précisions sur les causes possibles de rupture des connexions .....	151
2.4. Diffusion des empreintes de rupture de connexions .....	151
2.5. Recommandation d'application des méthodes de datations .....	152
<b>3. Perspectives .....</b>	<b>152</b>

3.1. Paléogénomique .....	152
3.2. Elargissement du jeu de données .....	153
3.3. Datation de signaux cosmopolites en Asie.....	154
3.4. Application à la datation de signatures d'adaptation.....	154
3.5. Application à la chèvre .....	155
<b>Références .....</b>	<b>158</b>
<b>Annexes.....</b>	<b>193</b>
<b>Appendix S1 : Material and methods (Article section Introduction générale - 2.1) .....</b>	<b>194</b>
<b>Modern samples .....</b>	<b>194</b>
Sampling .....	194
DNA extraction .....	194
DNA amplification and sequencing .....	194
<b>Ancient samples .....</b>	<b>195</b>
Sampling .....	195
Radiocarbon dates.....	195
DNA extraction .....	195
DNA amplification and sequencing .....	196
<b>Additional data .....</b>	<b>197</b>
<b>Genetic analyses .....</b>	<b>197</b>
Alignment .....	197
Phylogenetic analysis.....	197
The short CR fragment can discriminate haplogroups .....	197
Genetic parameters .....	198
<b>Supplementary tables &amp; figures .....</b>	<b>198</b>
<b>Figures et tableaux supplémentaires.....</b>	<b>202</b>

# Table des illustrations

## Figures

Figure 1 : Voies de diffusion des moutons vers l'Europe depuis le centre de domestication. ....	22
Figure 2 : Mouflons de Chypre ( <i>Ovis gmelini ophion</i> ).....	22
Figure 3 : Exemple de races de moutons à queue courte d'Europe du Nord. ....	23
Figure 4 : Mouton estonien Kihnu.....	24
Figure 5 : Mouflons Européens ( <i>Ovis aries musimon</i> ) .....	25
Figure 6 : Mouton de type ibérique (race Ojalada).....	27
Figure 7 : Moutons Mérinos .....	27
Figure 8 : Moutons Churro .....	28
Figure 9 : Mouton Entrefino (race Castellana) .....	29
Figure 10 : Comparaison des types de cornes entre (A) une représentation antique égyptienne (3900 ans AP) et (B) un mouton de type Zackel .....	30
Figure 11 : Mouton de type Sahel .....	31
Figure 12 : Mouton Djallonké.....	31
Figure 13 : Illustration d'un mouton à queue grasse datant de 1684.....	32
Figure 14 : Voies de diffusion des moutons en Afrique .....	34
Figure 15 : Mouton de race Ouled-Djellal .....	36
Figure 16 : Voies de diffusion des moutons en Asie.....	39
Figure 17 : Aspects des mouflons présents dans le jeu de données de base .....	50
Figure 18 : Etapes nécessaires à la construction des fichiers de variants selon les meilleures pratiques recommandées par GATK.....	52
Figure 19 : Comparaison des distributions des annotations de variants avant et après le filtre VQSR utilisée pour fixer les seuils de filtration fixes en cas d'impossibilité d'utilisation du VQSR. ....	54
Figure 20 : Distributions des annotations des variants issus du génotypage joint et avant l'étape de filtration GATK. ....	55
Figure 21 : Distribution des individus sur le premier plan de l'ACP réalisée sur le jeu de données de base. ....	57
Figure 22 : Apparementement entre les Mouflons Asiatiques du jeu de données de base.....	58
Figure 23 : Pourcentage d'ascendance inférés pour chaque individu du jeu de données de base grâce au programme ADMIXTURE pour un nombre de populations ancestrales égal à 6.....	60
Figure 24 : Structure génétique des groupes utilisés pour le phasage des individus domestiques .....	61
Figure 25 : Représentation des relations entre le groupe cible (moutons) et les groupes externes nécessaire à l'inférence des états ancestraux et dérivés des variants. ....	62
Figure 26 : Comparaison des âges d'apparition des variants après inférence des états ancestraux et dérivés avec différents groupes externes : AM (mouflons américains : <i>O. dalli</i> et <i>O. canadensis</i> ) ; IR (mouflons iraniens : <i>O. gmelini</i> et <i>O. vignei</i> ). ....	62
Figure 27 : Distributions des individus du jeu de données étendu sur le premier plan d'une ACP réalisées sur le jeu de données de base.....	63

Figure 28 : Principe de la méthode GEVA.....	71
Figure 29 : Probabilités d’observation des trois état génotypiques .....	73
Figure 30 : Probabilités d’émissions nécessaires à la Chaîne de Markov Cachée utilisées dans la méthode GEVA pour définir le segment partagé par deux haplotypes, établies pour l’humain dans Albers et al., 2020 et établies pour le mouton au cours de cette étude. ....	75
Figure 31 : Sous-échantillonnage des individus marocains utilisés pour l’analyse GEVA. ....	76
Figure 32 : Nombre de variants effectivement datés .....	79
Figure 33 : Comparaison des trois horloges de GEVA. ....	81
Figure 34 : Distribution des âges des variants estimés par l’horloge jointe avant application du filtre qualité.....	83
Figure 35 : Distribution des âges des variants estimés par l’horloge de recombinaison avant application du filtre qualité.....	84
Figure 36 : Distribution des âges des variants estimés par l’horloge de mutation avant application du filtre qualité .....	85
Figure 37 : Distribution des âges des variants estimés par l’horloge J après application du filtre qualité .....	87
Figure 38 : Nombre de variants datés le long du génome .....	88
Figure 39 : Nombre de variants spécifiques par groupe au sein du groupe de niveau supérieur. ....	93
Figure 40 : Nombre de variants spécifiques par groupes et par gammes de fréquence allélique .....	94
Figure 41 : Distributions des âges des variants spécifiques aux groupes continentaux, régionaux et aux races .....	96
Figure 42 : Distributions comparées des âges des variants continentaux et des variants continentaux cosmopolites. ....	97
Figure 43 : Distribution des âges des variants spécifiques aux groupes continentaux, régionaux et de races divisés par gammes de fréquence alléliques .....	99
Figure 44 : Nombre de variant caractéristiques par gamme de FST et par groupe continentaux, régionaux et de races. ....	101
Figure 45 : Recouvrement entre les variants caractéristiques des groupes continentaux, régionaux et locaux .....	102
Figure 46 : Distribution des âges des variants caractéristiques pour chaque groupe continental, régional et de race et par gamme de FST.....	105
Figure 47 : Comparaison des distributions des âges des variants caractéristiques et spécifiques pour chaque groupe continental, régional et local. ....	107
Figure 48 : Evolution de la CCF (Cumulative Coalescent Function), représentant la fraction de génome partagée au cours du temps entre cinq génomes cibles africains (Tableau 15) et les génomes comparateurs issus du monde entier.....	110
Figure 49 : Description du modèle utilisé pour la datation de la sélection de variant génomique par la méthode startmrca. ....	120
Figure 50 : : Datations réalisées pour chacun des groupes de variants associés à la domestication du mouton. ....	127
Figure 51 : Topologie permettant l’identification des haplotypes porteurs ou non porteurs des régions sous sélections du groupe A1.....	128



Figure 52 : Traces des chaînes de MCMC de l'inférence startmrca qui montre une répliquabilité entre chaînes indépendantes issues ou non d'un même échantillon. ....	130
Figure 53 : Traces des chaînes de MCMC de l'inférence startmrca qui montrent des différences d'estimations entre les chaînes issues d'échantillons différents. ....	131
Figure 54 : Âges de la sélection des signaux de sélection cosmopolites (groupe A1) .....	134
Figure 55 : Âges d'apparition des variants autour des signaux de sélection les plus anciens .....	135
Figure 56 : Comparaison des âges d'apparition et de sélection des signaux de sélection intragroupe (groupe B).....	136
Figure 57 : Âges de sélection des signaux de sélection intra-groupe (groupe B) .....	137
Figure 58 : Aspect de la laine des agneaux de race Hu .....	143
Figure 59 : Récapitulatif des principaux résultats obtenus. ....	148

## Tableaux

Tableau 1 : Origine des génomes complets de moutons domestiques constituant le jeu de données de base .....	48
Tableau 2 : Origine des génomes complets de mouflons sauvages constituant le jeu de données de base .....	50
Tableau 3 : origine des génomes complets ajoutés au jeu de données de bases pour former le jeu de données étendu .....	51
Tableau 4 : seuils imposés pour l'étape de filtration des SNPs par VariantFiltration de GATK .....	53
Tableau 5 : Nombre de variants par chromosomes avant et après le filtre GATK.....	56
Tableau 6 : Groupes utilisés pour phaser les variants du jeu de données de bases.....	61
Tableau 7 : Paramètres utilisés pour la simulations de 10 000 haplotypes par msprime sous un modèle démographique typique du mouton .....	72
Tableau 8 : Corrélations entre les probabilités d'émissions nécessaires à la Chaîne de Markov Cachée utilisé dans la méthode GEVA pour définir le segment partagé par deux haplotypes, établies pour l'humain dans Albers et al., 2020 et celles établies pour le mouton au cours de cette étude.....	74
Tableau 9 : Nombre de variants effectivement datés par chromosome et par horloge moléculaire par la méthode GEVA comparé au nombre de variants soumis à la datation .....	80
Tableau 10 : Résumé statistiques des âges de variants estimés par les trois horloges de GEVA et avant application du filtre qualité. ....	82
Tableau 11 : Organisation des groupes continentaux, régionaux et locaux utilisés pour l'analyse de la distribution des variants spécifiques et caractéristiques .....	92
Tableau 12 : Résumé statistiques des distributions des âges des variants spécifiques aux groupes continentaux, régionaux et aux races .....	95
Tableau 13 : Résumés statistiques des âges des variants caractéristiques. ....	103
Tableau 14 : Tableau Z : Recouvrement entre les variants spécifiques et caractéristiques .....	106
Tableau 15 : Individus choisis pour constituer les génomes cibles nécessaires à l'analyse CCF.....	108
Tableau 16 : Description des signaux de sélection cosmopolites (groupe A). ....	121
Tableau 17 : Description des signaux de sélection intra-groupe (groupe B) .....	124
Tableau 18 : Description des gènes associées aux signaux de sélection intra-groupe (groupe B) .....	125



# Chapitre 1

# Introduction générale

---

Dans cette thèse, nous explorons le processus évolutif de la domestication animale par la caractérisation des variations de l'ADN et au travers de l'exemple du mouton. Nous verrons dans ce chapitre introductif comment on peut définir la domestication et en quoi elle est un cas d'école de la biologie évolutive. Comme ce travail vise à étudier des aspects temporels du processus via des datations d'évènements de sélections et de mutations, nous décrirons ici ce que l'on sait des origines et des étapes de la diffusion des moutons dans le monde. Le but de ce chapitre est de répertorier les informations nécessaires à l'interprétation des résultats des analyses que nous avons menés, et de les replacer dans un contexte plus général. Nous illustrons également l'apport de la génomique dans la compréhension du processus de domestication par une série d'exemples non exhaustifs mais variés.

# 1 La domestication animale

---

La domestication des animaux et des plantes est l'un des changements majeurs de l'Histoire humaine, qui a profondément changé les modes de vie. Le chien a été la première espèce à être domestiquée il y a environ 15 à 17 milliers d'années (Pionnier-Capitan et al., 2011). Il s'en suivra la domestication des plantes et du bétail à partir de 14 000 ans avant le présent (AP) (Grasteau et al., 2005) puis celle de plusieurs autres animaux, comme le cheval, le chameau, le lapin etc. Dans ce travail nous nous concentrons sur la domestication animale.

## Définition

---

La définition de la domestication n'est pas évidente, tout comme il peut être difficile d'assigner un statut domestique ou sauvage à un animal, à partir du moment où il a une interaction avec l'humain. Il n'y a pas toujours de dichotomie stricte entre sauvages et domestiques surtout dans les étapes initiales de la domestication (Dobney and Larson, 2006; Larson and Burger, 2013). L'interaction humain-animal impose la sélection d'animaux plus dociles, ou répondant à certains besoins humains. Ils doivent aussi être suffisamment adaptables pour suivre les humains au fil de leurs migrations. C'est pourquoi dans cette thèse nous définirons la domestication comme **un continuum de processus de sélection des animaux les plus adaptés aux besoins et aux environnements humains** (Scott, Demoule and Saint-Upéry, 2019). **On considère donc qu'elle se poursuit tant que l'interaction humain-animal est maintenue.**

## Les différentes voies de domestication

---

De manière générale, la relation entre l'humain et l'animal s'intensifie au cours du temps (Vigne, 2011) mais pas toujours selon les mêmes voies. Zeder (2012) caractérisa la voie commensale, la voie « proie » et la voie dirigée.

La **voie commensale** débute par une attraction des animaux vers les campements humains notamment à cause de leurs déchets et de petites proies qui s'y concentrent. Ces animaux finissent par devenir plus dociles et moins peureux. Ce n'est que dans un second temps que

les humains agissent activement pour établir une relation réciproque. Les chiens sont un bon exemple d'espèce ayant suivi ce chemin.

La **voie « proie »** concerne plutôt les animaux initialement chassés par les humains. Ici le but premier est d'augmenter la disponibilité d'une ressource : les proies. Cela peut commencer par la modification des pratiques de chasse, par exemple en ciblant les jeunes mâles et en conservant le potentiel reproducteur des femelles. Au cours du temps cette gestion de gibier peut se transformer en gestion de troupeau par la capture et la rétention d'un groupe d'animaux. Le dernier stade est le contrôle complet de l'alimentation et de la reproduction des animaux. C'est le cas d'espèces d'élevage telles que les moutons, les chèvres ou les vaches.

Enfin, la **voie dirigée** est la plus directe des trois. C'est l'humain qui initie la relation avec un animal qu'il souhaite domestiquer pour exploiter une ressource. Cette voie repose donc sur des connaissances déjà acquises avec d'autres espèces domestiquées (bétail, chiens, chat...). Les chevaux, les chameaux pour le transport ou plus récemment les hamsters pour la compagnie ont suivi ce chemin.

## Un modèle de processus évolutif

---

D'un point de vue évolutif, la domestication reproduit un processus de spéciation sous l'influence principale de facteurs d'origine anthropique. C'est l'humain qui fonde des populations séparées des sauvages et qui finissent par évoluer indépendamment. Cette divergence résulte à la fois de la dérive engendrée par la fondation des populations et des nombreuses pressions de sélection qu'impose la cohabitation avec les humains.

L'avantage du processus de domestication est qu'il a lieu sur des échelles de temps accessibles via l'archéologie et les connaissances historiques rendues disponibles par l'apparition de l'écriture. Il est donc possible, dans une certaine mesure, de savoir quand il a commencé, d'identifier les pressions de sélections appliquées et de savoir quelles espèces ont été impliquées.

Dans cette thèse, nous nous focalisons sur le mouton. C'est une des premières espèces à avoir été domestiquée, ce qui permet d'étudier les premiers développements de l'élevage. C'est aussi un petit ruminant facilement transportable et donc répandu dans le monde entier. De plus, la domestication du mouton a déjà été bien étudiée et les données sont nombreuses, tant au niveau génomique qu'archéologique, ce qui est nécessaire au type d'étude que nous menons. Nous avons en effet besoin de données de génomes complets diversifiées mais aussi de repères chronologiques qui nous permettront de valider ou non nos datations d'évènement de mutations et de sélections.

## 2 Le cas de la domestication du mouton

---

### Les origines du mouton domestique

---

Le mouton (*Ovis aries*) a été domestiqué il y a environ 10 500 ans dans une région proche du croissant fertile (Ryder, 1983). Avec la chèvre, ce sont les premiers animaux d'élevage. Ses origines sont illustrées dans le papier suivant que nous avons soumis chez Animal Genetics et qui est actuellement en cours de révision. Cet article explore l'origine controversée du mouton grâce à l'ADN mitochondrial issu d'un large échantillonnage de moutons modernes et anciens et de leur plus proche parent sauvage : le Mouflon asiatique.

---

## Article : Broad maternal geographic origin of domestic sheep in Anatolia and the Zagros

---

Charlotte Her<sup>1\*</sup>, Hamid-Reza Rezaei<sup>1,2\*</sup>, Sandrine Hughes<sup>3\*</sup>, Saeid Naderi<sup>1,4</sup>, Marilyne Duffraisse<sup>3</sup>, Marjan Mashkour<sup>5</sup>, Hamid-Reza Naghash<sup>1</sup>, Adrian Bălăşescu<sup>6</sup>, Gordon Luikart<sup>7</sup>, Steve Jordan<sup>8</sup>, Deniz Özüt<sup>9</sup>, Aykut Kence<sup>9†</sup>, Michael W. Bruford<sup>10</sup>, Anne Tresset<sup>5†</sup>, Jean-Denis Vigne<sup>5</sup>, Pierre Taberlet<sup>11</sup>, Catherine Hänni<sup>1</sup>, François Pompanon<sup>1</sup>

**\*: these authors contributed equally and should be considered as co-first authors**

<sup>1</sup> LECA, Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, Grenoble, F-38000, FRANCE

<sup>2</sup> Environmental Sciences Department, Gorgan University of Agricultural Sciences and Natural Resources, Beheshti St. Gorgan, IRAN

<sup>3</sup> Institut de Génomique Fonctionnelle de Lyon, UMR 5242, Université de Lyon, Université Claude Bernard Lyon 1, ENSL, CNRS, Lyon, F-69007, FRANCE

<sup>4</sup> Natural Resources Faculty, Department of Environment, University of Guilan, Guilan, IRAN

<sup>5</sup> Département d'Ecologie et Gestion de la Biodiversité, CNRS, UMR 7209, Muséum National d'Histoire Naturelle, "AASPE" Archéozoologie, Archéobotanique, Sociétés, Pratiques et Environnements, Paris Cedex 05, F-75005. FRANCE

<sup>6</sup> Department of Bioarchaeology, "Vasile Pârvan" Institute of Archaeology, Romanian Academy, Bucharest, 10667, Romania

<sup>7</sup> Flathead Lake Biological Station, Montana Conservation Genomics Laboratory, Division of Biological Sciences, University of Montana, Polson, MT 59860, USA.

<sup>8</sup> Biology Department, Bucknell University, Lewisburg, PA, 17837, USA

<sup>9</sup> Biology Department, Middle East Technical University, Inonu Bulv. Ankara, 6531, TURKEY

<sup>10</sup> School of Biosciences, Cardiff University, Cardiff, CF10 3AX, UNITED KINGDOM

<sup>11</sup> The Arctic University Museum of Norway, UiT the Arctic University of Norway, Tromsø, 9010, NORWAY

### Summary

We investigated the controversial origin of domestic sheep (*Ovis aries*) using large samples of contemporary and ancient domestic individuals and their closest wild relatives: the Asiatic mouflon (*O. orientalisgmelini*), the urial (*O. vignei*) and the argali (*O. ammon*). A phylogeny based on mitochondrial DNA, including 213 new cytochrome-b sequences of wild *Ovis* confirmed suggested that *Ovis orientalisgmelini* is the maternal ancestor of sheep and precluded mtDNA contributions from *O. vignei* (and *O. orientalisgmelini* × *O. vignei* hybrids) to domestic lineages, as confirmed by an Approximate Bayesian Computation. We also produced 54 new control region sequences showing Sshared haplogroups (A, B, C and E) between domestic sheep and wild *O. orientalisgmelini* geographically which localized the domestication center in Eastern Anatolia and Central Zagros, excluding regions further East where exclusively wild haplogroups were found. This overlaps with the geographic distribution of *Ovis orientalisgmelini gmelini*, further suggesting the maternal origin of domestic sheep derives from this subspecies. Additionally, we produced 57 new CR sequences of sampled

Neolithic sheep remains across from a large area covering Anatolia to Europe, and discovered showing the early presence of at least three mitochondrial haplogroups (A, B and D) in Western colonization routes. This confirmed that sheep domestication was a large-scale process that captured diverse maternal lineages (haplogroups).

**Keywords** : Animal domestication – mitochondrial DNA – ancient DNA - sheep ancestor - area of early domestication - wild sheep – phylogeny – maternal genetic origins

The earliest archaeological evidence for the presence of domesticated sheep has been found in the Taurus mountains, Southeastern Anatolia and dates to 10,500 years B.P. (International Council for Archaeozoology et al. (eds.), 2005; Zeder, 2008) making this region a putative area of sheep domestication. Sheep later spread over the old world during the Neolithic, following human migration routes. Towards Europe, they took two major trajectories: the Danubian route across continental Europe, and the Mediterranean route crossing the sea to large islands, such as Cyprus and Sardinia from the 9th to the 6th millennia B.P. (Tresset and Vigne, 2011, 2007; Vigne et al., 2012). Towards Asia, the most likely hypothesis is colonization through Turkmenistan and Mongolia (Flad, Yuan and Li, 2007; Lv et al., 2015; Vigne, 2015). In the eastern part of the spread, the earliest evidence of sheep are to be found around 9,500-9,000 years B.P. in central and southern Zagros around (Daujat and Mashkour, 2017; Kamjan et al., 2018; Roustaei and Mashkour, 2016) and about the same dates in the north-east of Iran (Roustaei, Mashkour and Tengberg, 2015) and southern central Asia (Dobney and Jaques, 2010). The presence of domestic sheep is documented much later in China around 4,500-4,000 years B.P. (Flad, Yuan and Li, 2007).

Based on genetic studies, several wild Asiatic species have been proposed as the ancestor of domestic sheep (Nadler, Hoffmann and Woolf, 1973; Pedrosa et al., 2005a). These are the Argali *Ovis ammon*, the Asiatic Mouflon *O. gmelini* (often previously named *orientalis* following the nomenclature of Nadler et al. 1973), and the Urial *O. vignei*. Cytogenetic (Nadler, Hoffmann and Woolf, 1973) and archaeozoological (Clutton-Brock, 1981; Uerpmann, 1987) data have suggested the elimination of *O. ammon* and *O. vignei* as potential ancestors, leaving *O. gmelini* as the best candidate wild species from which sheep were domesticated. However, fluctuations in the nomenclature (Rezaei et al., 2010), the occurrence of a hybrid zone between *O. gmelini* and *O. vignei*, and putative domestication events in the lower Indus valley (Meadows, 1984) raise the possibility of genetic contribution from other wild species. Thus, we aimed to confirm the hypothesis of a unique maternal *O. gmelini* origin, and to localize the domestication area by finding the present wild populations representing the most direct descendants of domesticated matrilineages. We also sought to identify the first maternal lineages of domestic sheep entering Europe during the Neolithic spread.

Mitochondrial DNA (mtDNA) has been extensively and successfully used to describe the genetic diversity of domestic animals and to study their origins and histories. Indeed, it

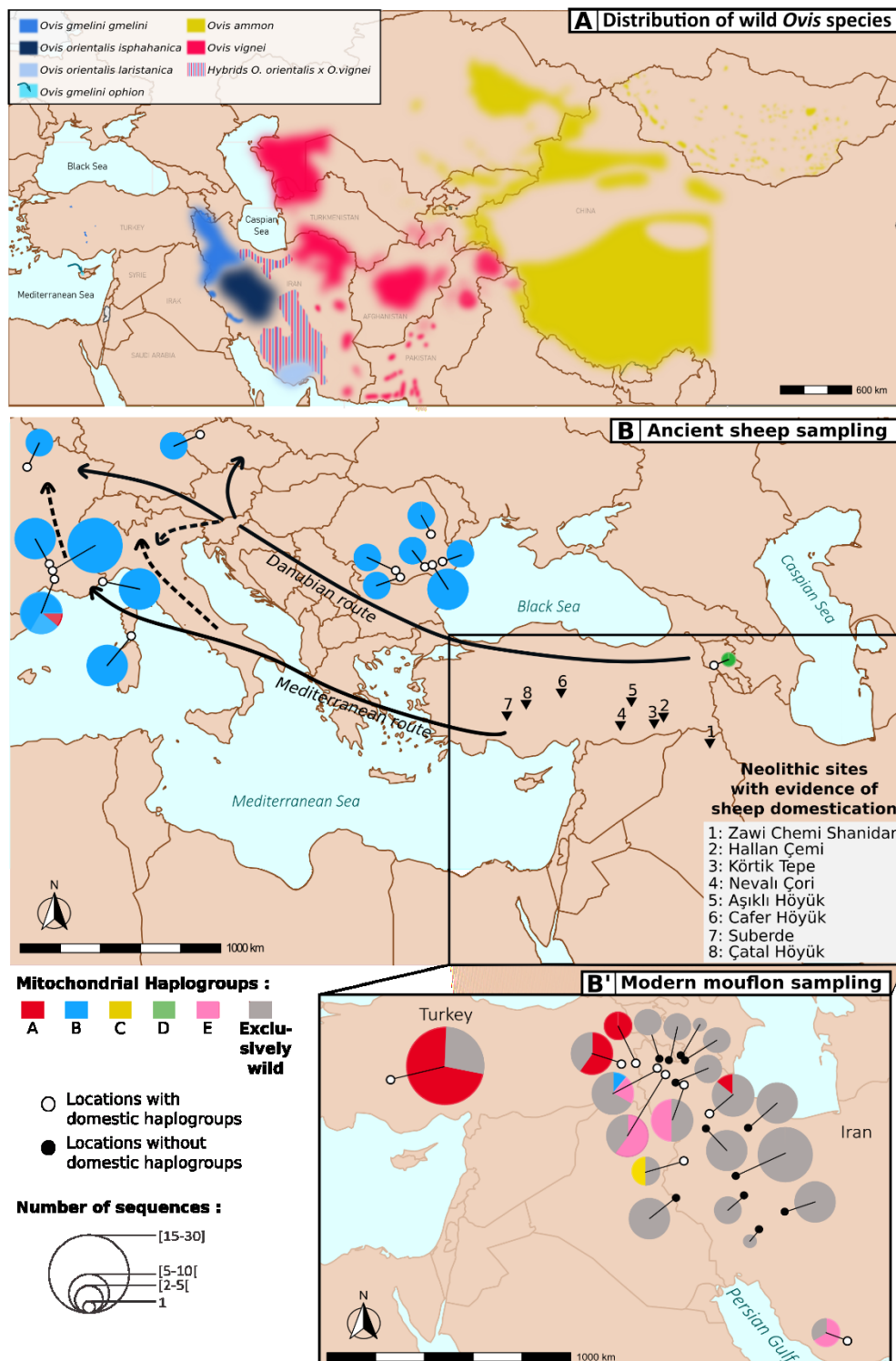


allows identifying the maternal evolutionary relationships with wild relatives because it is evolutionarily well conserved, yet sufficiently variable to capture genetic structure (Bruford, Bradley and Luikart, 2003; Zeder, M.A. (ed.), 2006). Fragments of the non-coding Control Region (CR) or Cytochrome b gene (*Cytb*) have been used to show the existence of five different haplogroups in domestic sheep, designated as A through E (Arora, Yadav and Mishra, 2013; Chen et al., 2006; Demirci et al., 2013; Guo et al., 2005; Hermes et al., 2020; Hiendleder et al., 2002; Horsburgh and Rhines, 2010; Ibrahim et al., 2020; Kolosov et al., 2021; Meadows et al., 2007; Pedrosa et al., 2005a; Resende et al., 2016; Rezaei et al., 2010; Singh et al., 2013; Tapio et al., 2006a; Zhao et al., 2011). The analysis of complete mitochondrial genomes (Meadows, Hiendleder and Kijas, 2011) confirmed the validity of such mtDNA fragments for discriminating among haplogroups and tracing the history of maternal lineages.

We built two datasets, one to compare modern domestic and wild sheep and another to compare modern and ancient domestic sheep (see **Table S1**). For the first one, we carried out an extensive non-invasive sampling of wild *Ovis* corresponding to 213 wild sheep including 3 *O. ammon*, 85 *O. gmelini*, 15 *O. vignei*, 1 *O. musimon*, 28 hybrids *gmelini* × *vignei*. Samples originated from 55 geographic localities representing most of their distribution area and encompassing the putative region of domestication according to archaeological data (see **Table S2, Appendix S1**). Moreover, we collected samples from 81 domestic sheep from 28 different countries as representative of the mtDNA worldwide domestic diversity (**Table S2, Appendix S1**). We sequenced 1,217 bp of the *Cytb* locus for these samples, and retrieved 421 supplementary sequences from GenBank (**Table S2, Appendix S1**). We built the second dataset on 120 bp of the more variable CR sequence to analyze intra-specific diversity (corresponding to the most recent divergences). Data for 1283 present-day sheep were collected from GenBank, and we sequenced 54 modern *O. gmelini* (**Table S3**) and 57 ancient domestic samples mostly distributed in Europe and dating back around 8,000 and 5,000 YBP (**Fig. 1, Tables S4, S5 Appendix S1**).

We built maximum likelihood and neighbour-joining (Saitou and Nei, 1987) trees using IQTREE (Nguyen et al., 2015) and the R package APE (Paradis, Claude and Strimmer, 2004), respectively (**Appendix S1**), to infer the phylogenetic relationships between the Middle East wild species and 362 *O. aries* representative of worldwide domestic diversity. Both phylogenies showed the monophyly of each of the three wild species (i.e. *O. gmelini*, *O. vignei* and *O. ammon*). All *O. aries* fall within the *O. gmelini* clade, confirming that domestic sheep evolved from ancestral matrilineages of the Asiatic mouflon (Bunch et al., 2006; Demirci et al., 2013; Hiendleder et al., 2002, 1998; Meadows, Hiendleder and Kijas, 2011; Meadows et al., 2007; Tapio et al., 2006a). This latter clade was divided into two clusters, both containing haplotypes from wild and domestic sheep (**Fig. 2**), showing the early divergence between A/B haplogroups and C/E haplogroups (Meadows et al., 2007). Interestingly, individuals from the *O. gmelini* × *vignei* hybrid zone in Central Iran had either an *O. gmelini* or an *O. vignei* maternal origin, as already shown by Demirci et al. 2013. None of them fall into a clade containing domestic sheep (**Fig. 2**) precluding the contribution of hybrid populations to domestics.

However, Urial Y-chromosome haplotypes found in several Asiatic mouflons (Deng et al., 2020) support a partial paternal contribution of the Urial.

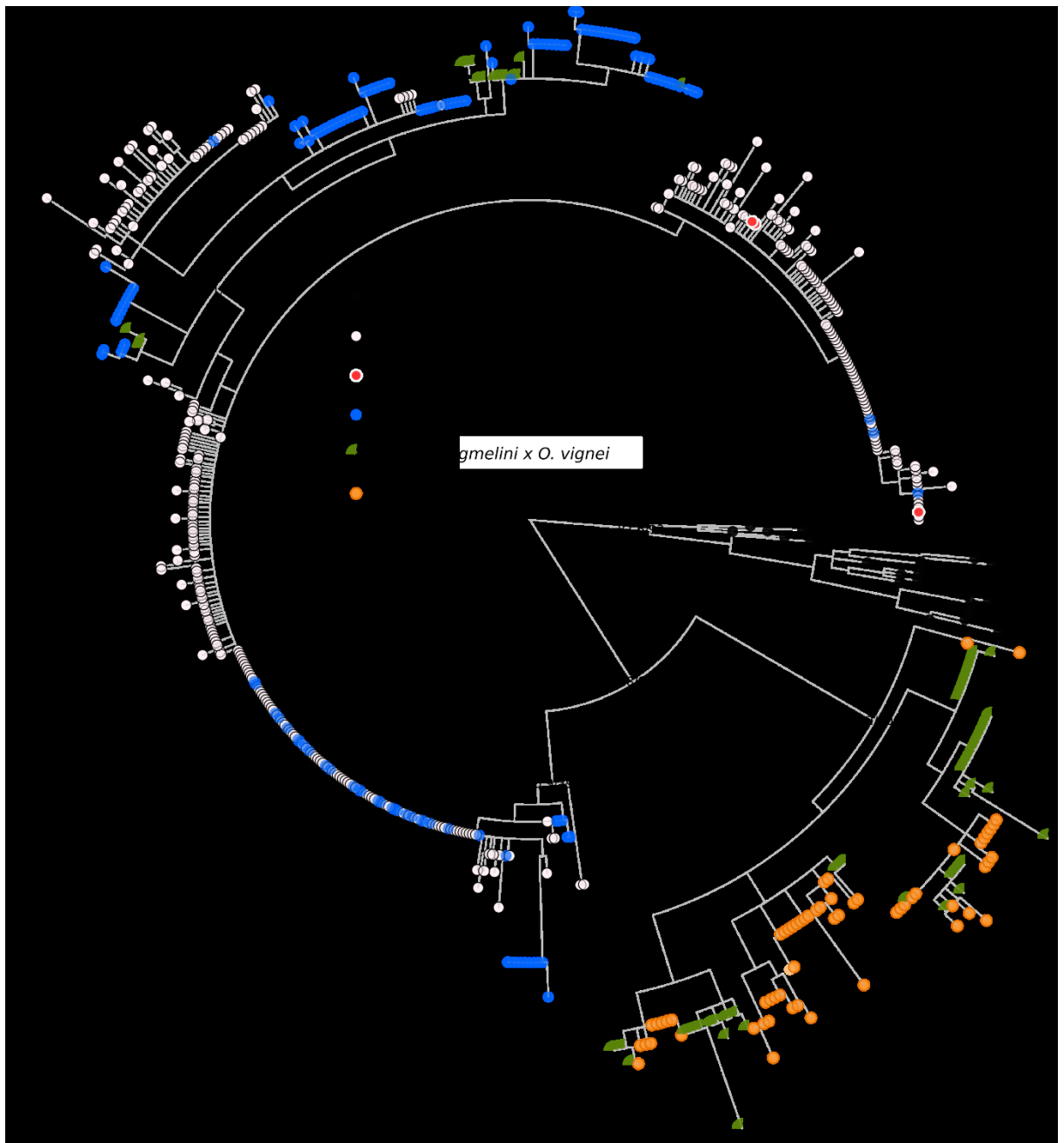


**Figure 1: Geographic distribution of ancient domestic and modern wild haplogroups.** The haplogroups of ancient sheep (*Ovis aries*) were determined based on CR sequences (120 bp). The related haplogroups of modern mouflon (*Ovis gmelini*) were determined based on CR sequences and confirmed by comparison to *Cytb* sequences (1217 bp). The Neolithic sites with evidence of sheep domestication were taken from the literature (International Council for Archaeozoology et al. (eds.), 2005; Zeder, 2008, 2011).

Assuming that the boundaries of the current distribution of wild species are similar to those at the time of domestication (Uerpmann, 1987), the exclusion of *O. ammon*, *O. vignei* and *O. vignei* × *O. gmelini* hybrids as possible ancestors precludes hypotheses that the Central Iranian Plateau, the Lower Indus Valley and even more eastern locations contributed to domestication. Based on archaeozoological data, the cradle for sheep domestication would extend from Central Anatolia to Northern and Central Zagros (International Council for Archaeozoology et al. (eds.), 2005; Zeder, 2008, 2011). This is confirmed by mtDNA data, as the highest nucleotide diversity was found in this region (**Table S6**) where, moreover, the mtDNA haplogroups carried by modern Asiatic mouflon were the closest to those of the domestics (**Fig. 1**). This region overlaps the current distribution of the subspecies *O. gmelini gmelini*, whereas *O. g. isphahanica* and *O. g. laristanica* are spread further east and south, respectively (**Fig. S3**) (Michel and Ghoddousi, 2020). The strong bottleneck experienced (even recently) by the Cyprus mouflon (*O. ophion*) which is close to *O. g. gmelini* (Sanna et al., 2015) and result from the early transportation of individuals from the continent (Vigne et al., 2014) makes it very difficult to assess its potential contribution to the domestic diversity. Besides that, the presence of some haplotypes similar to domestic sheep in modern *O. g. gmelini* populations offers evidence that the latter gave rise to the domestic haplogroups without excluding later introgression of the domestic stock into wild populations.

We used *CR* sequences from 1283 modern and 57 ancient domestic sheep to investigate global sheep mitochondrial diversity. Most modern sheep (56%) were from the B haplogroup, followed by the A (34%), C (9%) and D and E (<0.5%) (**Fig. S2**). These proportions are similar to those of other studies (Demirci et al., 2013; Kandoussi et al., 2020; Lv et al., 2015; Meadows et al., 2007; Sanna et al., 2015). As already shown (Handley et al., 2007; Kijas et al., 2009; Meadows et al., 2005), sheep showed very weak genetic differentiation at the continental scale (no between-continent *F<sub>ST</sub>* greater than 0.02). A and B haplogroups have a worldwide distribution that is not homogeneous, with more than 78% of the A haplogroup located in Asia. In North and South Europe, A and B haplogroups were detected with a high preponderance of the B haplotypes (above 90%) and some rare C haplotypes (**Table S6**). (Demirci et al., 2013; Deng et al., 2020; Lv et al., 2015).

The genotyping of 57 ancient samples showed that there were at least three haplogroups present in Neolithic European domestic sheep, with the B haplogroup already predominant (i.e. 55 B, 1 A and 1 D haplotypes; **Appendix S1, Table S5,S6**). We probably did not detect other low-frequency haplogroups from that area. However, this result indicates that the A and D haplogroups has been present since the beginning of western sheep diffusion during the Neolithic. This predates by several millennia the first evidence of the A haplogroup in Europe, as it was already detected in medieval sheep from Finland but not in the oldest samples from the Iron period (Demirci et al., 2013; Niemi et al., 2013; Rannamäe et al., 2016). Thus we confirm the hypothesis of a very early presence of this haplogroup in Europe, and support the idea that a later introduction of A haplotypes from Spanish populations during the 16th century is unlikely as these importations concerned mainly rams (Niemi et al., 2013).



**Figure 2: Phylogenetic relationships between the domestic sheep (*Ovis aries*) and its potential wild ancestors: the argali (*Ovis ammon*), the urial (*Ovis vignei*) and the Asiatic mouflon (*Ovis gmelini*).**

The phylogenetic tree was built using maximum likelihood methods implemented in IQTREE with *cytb* sequences (1217 pb). The numbers on the principal nodes correspond to support values (Ultrafast Bootstrap value/ SH-aLRT value). The black circle arcs represent the different mitochondrial haplogroups.

Combining genetic data from *O. aries* and its closest wild relatives with archaeozoological data makes it possible to propose a sub-species of the Asiatic mouflon, *O. gmelini gmelini*, as the maternal origin of domestic sheep. The current distribution of the Asiatic mouflon and the geographic pattern of its genetic diversity are consistent with domestication first taking place in Eastern Anatolia and in Northern and Central Zagros, as

supported by archaeozoological data (Clutton-Brock, 1981; International Council for Archaeozoology et al. (eds.), 2005; Uerpmann, 1987). They preclude more eastern independent domestication centers, especially in the Indus Valley, as well as the contribution of *O. gmelini* × *O. vignei* hybrids. Up to now, studies on the diversity of chromosome Y gave valuable insights in the demographic history of sheep (Deng et al., 2020) but did not provide information of the geographic origin. However, studies on chromosome Y (Li et al., 2020b) and whole genome data (Barbato et al., 2017) brought evidence for the occurrence of post-domestication introgressions from wilds towards domestics. At least three different haplogroups (B, A and D) diverged before domestication and underwent early diffusion towards Europe at the beginning of the Neolithic. Thus, our results globally indicate that sheep domestication was a process on a large geographical scale. It involved the capture of a wide genetic diversity over a large area, as already shown for other livestock species such as goats (Daly et al., 2018; Naderi et al., 2008), pigs (Bosse, 2018; Larson et al., 2007, 2010) and cattle (Götherström et al., 2005; McTavish et al., 2013).

### Acknowledgments

We would like to thank Benjamin Gillet and Bénédicte Bertrand for help with the ancient DNA experiments. We also are grateful to all the archaeozoologists that helped to obtain the ancient sheep samples, in particular to Daniel Helmer, Lenka Kovacicova, Rose-Marie Arbogast and Joris Peters. We would also like to thank our fellow archaeologists who allowed us access to the archaeozoological materials from Armenia (Ruben Badalyan and Christine Chataigner) and Romania (Radian Andreescu, Pavel Mirea, Dragomir Popovici, Valentin Parnic, Cătălin Bem and Florin Vlad).

This study has been supported by a grant from the French Agence Nationale de la Recherche (Project Chronobos, ANR 05-GANI-004-02).

The authors declare no conflict of interests.

### Availability of data

All *Cytb* and *CR* sequences produced in this study have been published on GenBank under the accession numbers: MW684398 - MW684451, MW684452 - MW684508, and MW677233 - MW677445.

Ainsi, contrairement, à d'autres espèces domestiques comme la vache ou le cheval, dont les lignées sauvages descendant des populations domestiquées sont éteintes, les descendants directs de l'ancêtre du mouton subsistent toujours aujourd'hui à l'état sauvage. C'est un avantage considérable pour l'étude des mécanismes biologiques sous-jacents au processus de domestication, puisque l'on dispose de deux espèces proches ayant évolué parallèlement, à la différence que l'une a été domestiquée et pas l'autre. Cet article montre comment l'étude de la diversité des mouflons asiatiques permet d'identifier le berceau de

domestication du mouton. Dans les parties suivantes nous parcourons les différents continents pour décrire la diffusion des moutons hors de ce centre de domestication au Moyen-Orient, ainsi que les processus plus récents à l'origine de leur diversité actuelle.

## Europe

---

### 2.1.1 Description générale

En Europe, 771 races sont enregistrées sur 850 au niveau mondial. Cependant le continent est particulièrement touché par l'extinction : 20 % sont éteintes et 40% sont en danger, contre 32% dans le reste du monde (FAO, 2000, 1998; Signorello and Pappalardo, 2003). Les races concernées ne conviennent pas aux besoins de productions intensives actuelles, ou sont des groupes minoritaires qui se perdent via des introgressions avec des races commerciales. Les pratiques d'élevage sont hétérogènes. Au Sud, comme au Moyen-Orient, les races ne sont pas homogènes et les croisements sont fréquents. C'est l'inverse au Nord où la gestion intensive, et le faible nombre de béliers reproducteurs a causé une perte de diversité génétique importante (FAO, 2000, 1998).

Au niveau génétique, la structuration globale est faible mais présente un gradient du Nord-Ouest au Sud-Est. La majorité de la variation se trouve au sein des races (Handley et al., 2007; Kijas et al., 2012a; Peter et al., 2007). Il existe un gradient de diversité inverse (du Sud-Est au Nord-Ouest) (Kijas et al., 2012a; Peter et al., 2007). Les trois raisons invoquées pour expliquer ce gradient sont les suivantes :

- La **succession d'effets de fondation** lors de l'expansion des moutons depuis leur centre de domestication au Moyen-Orient.
- L'**interruption des flux de gènes** entre les races d'Europe de l'Ouest et du Sud-est pendant la Guerre Froide entre 1950 et 1990.
- L'**influence Turc** sur la plupart des races du Sud-Est de l'Europe due à l'occupation de la région par les Turcs jusqu'en 1923 (Ryder 1983).

Peter et ses collaborateurs (2007) mettent aussi en évidence une structuration reflétant différents types phénotypiques comme le type queue grasse présent dans les races du Sud-Est et les types Merino et Alpin au Nord et à l'Ouest.

De manière générale, les populations ovines du sud de l'Europe sont plus diversifiées génétiquement (Handley et al., 2007; Peter et al., 2007; Tapio et al., 2010) mais moins différenciées les unes des autres que celles du nord (Tapio et al., 2010) où certaines races sont restées très pures. Ce sont souvent des races dites primitives qui sont très isolées géographiquement, sur des îles par exemple (Moutons de Soay, North Ronaldsay, Heidschnuke par exemple).

Les types de production diffèrent du Nord au Sud : Le sud et la région méditerranéenne sont plutôt tournés vers la production laitière tandis qu'au Nord, c'est la production de viande et de laine qui prédomine. On retrouve déjà des preuves de cette

répartition au Néolithique dans les profil d'abattages : les moutons étaient tués plus jeunes au sud, ce qui indiquerait une production de lait même si on ne peut pas écarter une préférence culturelle pour une viande plus tendre (Gillis et al., 2019).

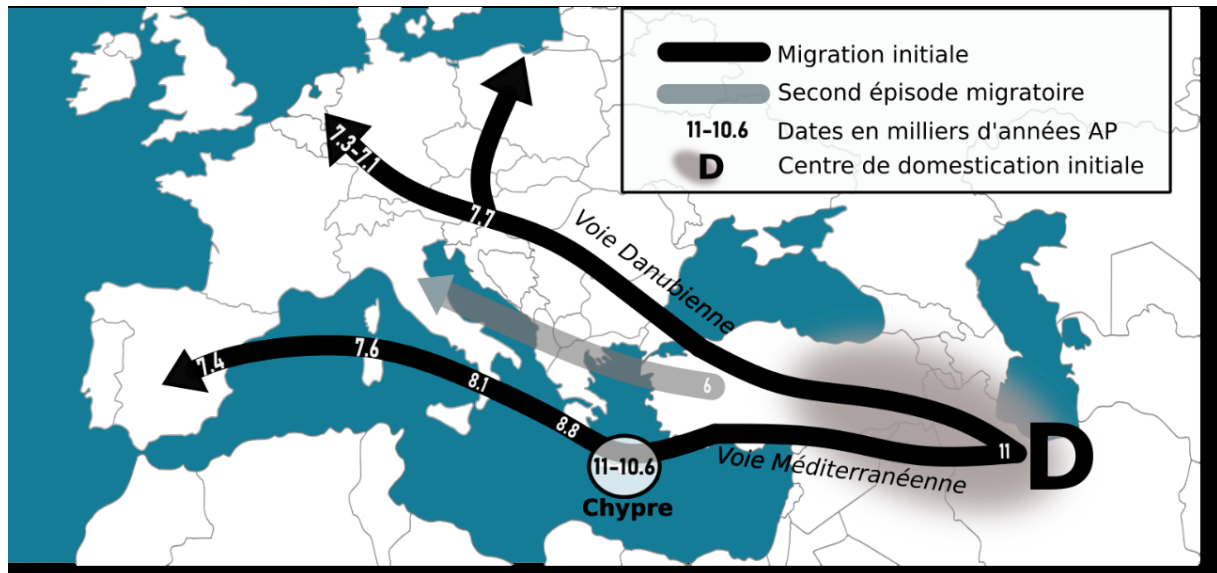


Figure 1 : Voies de diffusion des moutons vers l'Europe depuis le centre de domestication. Les dates sont exprimées en années avant le présent et correspondent aux plus anciennes preuves (archéologiques ou génétiques) de la présence de moutons à l'endroit où elles sont indiquées

## 2.1.2 La diffusion des moutons en Europe

Les premières preuves de diffusion du mouton vers l'Europe et hors du centre de domestication sont trouvées à **Chypre** vers la seconde moitié du 10<sup>ème</sup> millénaire (Figure 1). Sur l'île, les pratiques vont de la chasse d'animaux libres, au contrôle d'animaux domestiques mais d'apparence (archéologique) similaire aux ancêtres sauvages, en passant par la gestion de populations férales (Vigne et al., 2012, 2014). Les descendants de ces premières introductions à Chypre sont les mouflons de Chypre (*Ovis gmelini ophion*) (Figure 2) et constituent une preuve de dispersion par l'Homme. En effet, Chypre est une île isolée du continent depuis plus de 5 millions d'années (Vigne, Buitenhuis and Davis, 1999) qui n'a pu être gagnée que par bateau pour des espèces du genre *Ovis*. Chypre a donc constitué un tremplin vers l'Europe lors de la première vague de colons marins dispersant les animaux domestiques vers l'ouest (Peters, von den Driesch and Helmer, 2005; Vigne et al., 2012; Zeder, 2008).



Figure 2 : Mouflons de Chypre (*Ovis gmelini ophion*)

Crédit photo : <http://www.wilddocu.de/cyprus->

Le reste de l'Europe est colonisé via deux routes, la route continentale danubienne et la route méditerranéenne qui traverse les mers en passant par les grandes îles (Ryder, 1991; Tresset and Vigne, 2007). Les données archéologiques et les datations radiocarbone montrent une arrivée plus rapide à l'Ouest de l'Europe via la route méditerranéenne (Zilhao, 2001), il y a environ 7400 ans (Pereira et al., 2006; Tapio et al., 2006b) sûrement grâce au transport maritime.

L'analyse de marqueurs génomiques rétroviraux et de restes archéologiques montrent que la diffusion du mouton vers l'Ouest a eu lieu en deux vagues. La première il y a environ 11 000 ans et la deuxième plus tardivement autour de 6000 à 3000 ans AP (Chessa et al., 2009a; Deng et al., 2020; Helmer, Gourichon and Vila, 2007; Vigne, 1992) Cette seconde vague a remplacé les races primitives par des races plus modernes plutôt à laine. De nos jours, la standardisation et la sélection de races compétitives amène à un processus similaire à ce remplacement et donc à l'extinction de nombreuses races natives, surtout depuis l'apparition du concept de race il y a environ 200 ans. Seules quelques races aux marges de ce front de colonisation et les mouflons européens féraux<sup>1</sup> semblent avoir gardé des traces du premier épisode migratoire (Chessa et al., 2009a; Ryder, 1984).

### 2.1.3 Les moutons d'Europe du Nord en partie épargnés par le second épisode de migration

Les moutons à queue courte d'Europe du Nord constituent un groupe de races dont l'introduction remonte à 6000 ans environ (Ryder, 1991; Trow-Smith, 2013). Elles ont ensuite été dispersées par les Vikings entre le 8<sup>ème</sup> et le milieu du 11<sup>ème</sup> siècle sur une aire allant de la Russie à l'Islande (Dýrmundsson and Niznikowski, 2010; Ryder, 1983). Des études génétiques ont montré que ce groupe a été fondé par une population qui s'est ensuite fragmentée (Tapio et al., 2005). Avec le temps, ces populations isolées au Nord se sont différenciées en races et ont perdu leur diversité (Handley et al., 2007;



Figure 3 : Exemple de races de moutons à queue courte d'Europe du Nord.

Orkney : <https://www.theorkneysheepfoundation.org.uk/the-sheep/>  
Hebridean : Par Jim Champion from Southampton, UK — The three-horned one again, CC BY-SA 2.0  
Gute : Par Oskari Löytynoja - Own work, CC BY 3.0

<sup>1</sup> **Féral** : se dit d'une espèce ou d'une population qui descend d'individus domestiqués mais qui est retournée à l'état sauvage.



Tapio et al., 2005). Ce sont des races locales dont la répartition actuelle est proche de celle d'origine (Dýrmundsson and Niznikowski, 2010; Ryder, 1983; Tapio et al., 2006b) à l'exception de la Finnsheep et de la Romanov qui ont été beaucoup utilisées pour augmenter la fertilité d'autres races depuis les années 60 (Fahmy, 1989; Jakubec, 1977; Oltenacu and Boylan, 1981; Ricordeau et al., 1978; Thomas, 2010). Un peu plus récemment, les races Islandaise, Gotland et Shetland ont été utilisées hors d'Europe, particulièrement en Amérique du Nord. Les productions sont plutôt axées sur la laine et la viande alors que la production laitière reste marginale. Certaines races interviennent aussi dans l'entretien du paysage (Dýrmundsson and Niznikowski, 2010).

Ces moutons nordiques sont caractérisés par leur queue courte et effilée qui comporte 8 à 10 vertèbres contre 16 à 18 pour les moutons à queue longue (Dýrmundsson and Niznikowski, 2010; Frandson, Wilke and Fails, 2009). Ce sont des moutons rustiques et robustes très diversifiés en termes de taille, de productivité et de performances mais qui présentent des caractéristiques primitives telles qu'une laine grossière à double couche de couleurs variées et plutôt foncées, qui peut muer, ainsi que la présence de cornes même chez les femelles (Adalsteinsson, 1970; Dýrmundsson and Niznikowski, 2010; Fahmy, 1989; Jakubec, 1977; Maijala, 1997; Ricordeau et al., 1978) (**Figure 3**).

Des analyses génétiques ont confirmé l'origine primitive de ces races aux marges du front de diffusion européen grâce à l'étude d'une famille de rétrovirus endogènes (enJSRV). Chacun des rétrovirus connus de la famille enJSRV résulte de l'intégration stable (transmise de génération en génération) et irréversible, du génome du rétrovirus exogène Jaagsiekte à un endroit différent du génome à un moment donné de l'histoire de cette espèce. Ainsi, les races de moutons qui partagent la présence d'un élément enJSRV ont une origine commune postérieure à l'insertion de ce rétrovirus. L'analyse de la présence ou absence de chaque élément enJSRV montre que la plupart des races ovines européennes modernes sont caractérisées par la présence de l'élément enJSRV-18 qui est absent des populations primitives. Ces dernières présentent soit l'élément enJSRV-7 soit aucun (Chessa et al., 2009a). Cela montre que ces moutons primitifs sont issus du premier épisode migratoire et qu'elles ont été épargnées par le remplacement du second.

Plus récemment, des études portant sur des races de moutons Estoniennes montre que la race Kihnu<sup>2</sup> (**Figure 4**) présente aussi des rétrotypes primitifs. La comparaison d'individus anciens et modernes indique



Figure 4 : Mouton estonien Kihnu  
(Rannamäe et al., 2016)

---

<sup>2</sup> La race Kihnu provient de la découverte de troupeaux subsistant sur des îles au sud-ouest de l'Estonie.

cependant que la fréquence de ces rétrotypes diminue avec le temps. Cela montre que le remplacement a aussi touché les Kihnu mais de manière partielle par des croisements avec des races modernes. On retrouve d'ailleurs un nouveau marqueur dans les moutons Kihnu modernes, identiques à celui présent chez des races modernes, mais absent des Kihnu anciens (Rannamäe et al., 2020).

Les races d'Europe du Nord ne sont pas les seules à avoir gardé des empreintes de la première diffusion Néolithique. C'est aussi le cas des mouflons des îles méditerranéennes.

#### 2.1.4 Les mouflons des îles méditerranéennes, reliques de la migration initiale

Le mouflon européen (*Ovis aries musimon*) (Figure 5), est le fruit de la féralisation<sup>3</sup> de moutons primitifs issus de la première vague de diffusion Néolithique en Corse et en Sardaigne (Poplin, 1979). Il n'y avait alors aucun *Ovis* sauvage en Europe. Les rares incursions d'*Ovis* remontent au Pléistocène moyen, plusieurs centaines de milliers d'années auparavant et concernent des espèces beaucoup plus grosses dont des ossements ont été retrouvés dans les montagnes de France (Crégut, 1979; Delpech et al., 1978; Pommerol, 1879) et de Moravie (Nehring, 1891).



Figure 5 : Mouflons Européens (*Ovis aries musimon*)

Crédit photo : CC-BY-2.5; Released under the GNU Free Documentation License.

Les mouflons européens sont aujourd'hui les seuls *Ovis* sauvages d'Europe et coexistent avec des troupeaux de moutons domestiques. Des documents attestent de croisements dès l'antiquité romaine (Cetti, 1774) qui donnent des animaux plus gros qui ont un meilleur succès reproducteur (Barbato et al., 2017; Pipia et al., 2008). De tels croisements mènent à des introgressions génétiques<sup>4</sup>. Il a été montré que des introgressions du mouflon vers le mouton pouvaient avoir un avantage adaptatif : les régions génomiques concernées sont impliquées dans l'immunité et la reconnaissance des goûts (Barbato et al., 2017). On constate d'ailleurs que la contribution génétique des mouflons européens, et donc les traces de la première diffusion Néolithique, augmente le long d'un gradient allant du sud-ouest au nord-ouest de l'Europe (Ciani et al., 2020). Cela confirme les résultats apportés par l'étude des marqueurs rétroviraux et coïncide avec des climats plus rudes auxquelles les introgressions adaptatives pourraient permettre de faire face.

Les mouflons de Chypre (**Figure 2**) sont aussi les descendants des toutes premières incursions ovines hors du centre de domestication. Une analyse récente a montré une très

---

<sup>3</sup> **Férialisation** : retour à l'état sauvage d'un animal domestiqué non autochtone.

<sup>4</sup> **Introgression génétique** : transferts de matériel génétique d'une espèce à une autre

faible diversité génétique (Guerrini et al., 2015) due au long isolement qu'a subi cette population depuis le Néolithique mais aussi à la chasse intensive perpétrée dans les années 1930 qui a réduit brusquement son effectif à seulement 20 individus (Forestry Department., 2012; Maisels, 1988) dont les 3000 actuellement présents descendent (Sanna et al., 2015).

L'Europe du Nord et les îles Méditerranéennes constituent donc un réservoir de populations qui descendent de lignées plus anciennes ayant diffusé précocement (Chessa et al., 2009a), mais qui restent minoritaires. Le remplacement par des races plus productives a touché la majorité du territoire européen. Des populations primitives ont été préservées en retournant à la vie sauvage sur des îles où les prédateurs et les hommes sont peu nombreux, d'autres ont pu être conservées du fait de leur rusticité leur conférant un avantage dans des milieux contraignants. Cependant l'isolement de ces races bien adaptées induit un risque d'extinction via l'érosion de leur diversité génétique (Handley et al., 2007; Peter et al., 2007; Tapio et al., 2005). Comme nous allons le voir, dans le Sud de l'Europe on observe une plus faible structuration génétique traduisant des flux de gènes plus importants, et une plus grande diversité génétique (Handley et al., 2007; Peter et al., 2007; Tapio et al., 2010).

## 2.1.5 Les moutons ibériques

### Résultats de multiples influences

Les moutons arrivent dans la péninsule Ibérique depuis l'Asie il y a 7000 ans (Ryder, 1984) probablement par la voie méditerranéenne. Les races ibériques se forment à la suite de multiples invasions. Après la vague initiale d'immigration asiatique, les Celtes envahissent la région il y a environ 2700 ans par les Pyrénées. Suivront les Phéniciens, les Grecs, les Romains et plus tard, les Maures. Ces derniers atteignent la péninsule il y a 1300 ans et y resteront près de 800 ans (Pedrosa et al., 2007; Pereira et al., 2006). La péninsule est donc un carrefour commercial et culturel intégré dans une zone de libre-échange couvrant tout le bassin méditerranéen (Goitein, 1967; Picard, 1997). Les colons qui se sont succédés ont pu améliorer le bétail local ou en importer. Les changements les plus nets interviennent au moment de la période musulmane, comme le prouvent les changements ostéométriques importants après cette période (Davis, 2006).

Plusieurs études génétiques suggèrent que les races Ibériques ont subi un flux de gènes continu via la Méditerranée et plus particulièrement du Moyen-Orient (Meadows et al., 2005). Pereira et al. (2006) ont par exemple montré que sept races portugaises présentaient des niveaux de diversité beaucoup plus élevés que les autres races européennes. On retrouve la même observation dans l'étude d'autres races Ibériques telles que les Mérinos (Arranz, Bayon and Primitivo, 1998; Diez-Tascón et al., 2000; Rendo et al., 2004). Cela va à l'encontre de l'hypothèse classique d'une diminution de diversité avec la distance au centre de domestication. On retrouve également dans ces races l'haplogroupe mitochondrial C (d'ordinaire seulement présent en Asie et au Moyen-Orient) et le A (majoritaire en Asie) dans des proportions plus importantes qu'ailleurs en Europe (Pedrosa et al., 2007; Pereira et al., 2006).

La présence d'haplotypes rares dans les populations ibériques seraient probablement le résultat d'introductions anciennes. En effet, aucune preuve n'atteste d'échanges récents entre Moyen-Orient et Ibérie et les morphologies des moutons du moyen Orient sont très différentes (queue grasse) (Pereira et al., 2006). De plus, des représentations de moutons de morphologie similaire aux moutons ibériques et datant de 2300-2200 ans attestent de leur présence précoce dans la péninsule.

On compte traditionnellement quatre types principaux de moutons Ibériques (Sánchez and Sánchez, 1986) définis selon leur laine : le type **Merino** à laine fine ; le type **Churro** à laine grossière, réparti plutôt au nord et au centre de la péninsule ; le type **Entrefino** à laine intermédiaire ; le type **Ibérique (Figure 6)** qui rassemble quelques races très rustiques à laine grossière isolées dans les régions montagneuses. Il n'y a pas d'opinion unanime sur leur origine mais elles dériveraient des premières colonisations asiatiques et font partie des races qui présentes des haplogroupes mitochondriaux C.



Figure 6 : Mouton de type ibérique (race Ojalada)

<https://www.mapa.gob.es/fr/ganaderia/temas/zootecnia/razas-ganaderas/razas/catalogo-razas/ovino/ojalada/>

Cette classification n'est pas supportée par la génétique qui ne permet pas de séparer ces groupes mais seulement les races qui les composent (Arranz, Bayón and Primitivo, 2001; Pedrosa et al., 2007).

## Les mérinos

La présence de moutons à laine fine en Espagne est ancienne. Des outils de tonte, des fibres et des sculptures ont été retrouvés dans des couches préhistoriques (Sánchez and Sánchez, 1986). Les écrits d'un auteur Grec installé à Rome, Strabo, font référence à une race de mouton à laine noire d'une qualité exceptionnelle. C'est sûrement l'une des races à l'origine de Mérinos actuels (**Figure 7**) (Ryder, 1983; Sánchez and Sánchez, 1986). Columelle, un agronome romain de la même époque, et Pline l'ancien célèbrent la qualité de laine fine et blanche des moutons



Figure 7 : Moutons Mérinos

Par fir0002flagstaffotos [at] gmail.com Canon 20D + Tamron 28-75mm f/2.8 — Travail personnel, GFDL 1.2, <https://commons.wikimedia.org/w/index.php?curid=322406>

d'Apulia dans le sud de l'Italie. Ces derniers furent introduits en Espagne au premier siècle av. JC. et sont peut-être aussi des ancêtres des Mérinos. La couleur blanche a été sélectionnée par les Romains via des croisements avec des béliers d'Apulia et d'autres races d'Afrique du Nord. L'expansion de la race se fera grâce aux Visigoths et aux Arabes au fil des

transhumances. Son développement se poursuit de manière plus intensive avec la création de la « Mesta », une association d'éleveurs créée en 1273 par Alphonse X de Castille (Sorre, 1923). A partir de cette date, les Mérinos seront très protégés et leur export strictement interdit pendant plusieurs siècles (Sánchez and Sánchez, 1986).

Il faudra attendre le 18<sup>ème</sup> siècle pour voir les Mérinos sortir largement d'Espagne, même si quelques excursions ont été relevées avant cette période. C'est notamment le cas en Italie où des béliers Mérinos sont importés pour améliorer les races locales dès 1435 (Lancioni et al., 2013). Il formeront ainsi la race actuelle Gentile du Puglia (Maria Sarti et al., 2006). L'expansion de la race démarre donc réellement et massivement au 18<sup>ème</sup>. Vers le Nord, ils atteignent l'Allemagne en 1748, la France en 1767, et la Grande-Bretagne en 1787 (Mason, 1988). A l'est, en Hongrie, c'est l'impératrice Marie-Thérèse qui achète 300 mérinos espagnols en 1774 (Fésüs et al., 2002). Leur expansion atteint aussi la Chine où des races à laine fines sont produites à partir de croisements avec des Rambouillet et des Mérinos soviétiques (Philippine Council for Agriculture, Forestry and Natural Resources, Research and Development; International Development Research Centre, 1988). On les trouve finalement dans le monde entier dès le début du 19<sup>ème</sup> où ils colonisent l'Afrique du Sud, l'Australie, la Nouvelle-Zélande et les Amériques (Diez-Tascón et al., 2000) si bien qu'on considère la race comme le facteur qui a le plus influencé le développement des races modernes (Kijas et al., 2012a). Les Merionolandschaf en Allemagne, les Rambouillet en France ou encore les Sopravissana en Italie ne sont que quelques exemples des races dérivées du Mérino (Lasagna et al., 2011).

Aujourd'hui, l'industrie de la laine Merino est majoritairement australienne tandis que les mérinos européens subissent un grave déclin qui les classe dans les races en danger d'extinction (Ciani et al., 2015). En Australie, une sélection intensive a commencé en 1870, pour la laine évidemment (Fogarty et al., 2003; Gardner et al., 2010) mais aussi pour la résistance aux maladies et parasites (Kijas et al., 2012a; Raadsma, Gray and Woolaston, 1998).

## Les Churro

Le type Churro (**Figure 8**) a pu être influencé par des migrations venues d'Europe Centrale avant les invasions Celtes, comme le prouverait sa pigmentation et sa laine grossière, longue et non ondulée au contraire des types Merino et Entrefino (Arranz, Bayon and Primitivo, 1998; Sánchez and Sánchez, 1986). C'est une race autochtone du nord-ouest de la région Castille-et-León en Espagne. Elle est utilisée pour la viande et sa sélection pour la production laitière a commencé en 1986 (Baro, Primitivo and Fuente, 1995). C'est une race de ce type qui a été la première introduite en Amérique (Ryder, 1984).



Figure 8 : Moutons Churro  
By Fernando García from Madrid, España

On trouve dans ce groupe la race Latxa qui compte deux variétés principales qui diffèrent par la pigmentation des extrémités : la Latxa tête-noire et la Latxa tête rousse (Urarte, 1989). Elles présentent une bonne production laitière et sont adaptées à la région d'élevage dans l'ouest de Pyrénées. Un programme d'amélioration utilisant les méthodes de génétique quantitative a été mis en place en 1981 pour améliorer la production, la composition du lait, la morphologie des mamelles et la résistance à certains parasites (Ruiz-Larrañaga et al., 2020).

On compte aussi dans ce groupe une race semi-férale reconnue récemment et classée en danger d'extinction depuis 2007: la Sasi Ardi (brebis des broussailles). Comme la Latxa, on la trouve dans l'Ouest des Pyrénées mais dans des zones plus broussailleuses et montagneuses auxquelles elle est très bien adaptée (Lasarte, Lazkanoteg and Pérez de Muniain, 2007). Elle joue ainsi un rôle important dans la prévention des incendies. Il n'y a pas de programme d'élevage officiel mais les Sasi Ardi sont élevées principalement pour leur viande en élevage extensif (Lasarte, Lazkanoteg and Pérez de Muniain, 2007).

## Entrefino

Les moutons de type Entrefino (**Figure 9**) présentent une laine de type intermédiaire. Il existe ainsi deux hypothèses pour l'origine du type Entrefino : Un croisement Merino x Churro (Ryder, 1984) ou une origine indépendante (Sánchez and Sánchez, 1986). Les études qui se sont intéressées à la question sont assez anciennes et portent sur un petit nombre de marqueurs qui n'ont pas permis d'apporter une réponse claire (Arranz, Bayón and Primitivo, 2001; Pedrosa et al., 2007).



Figure 9 : Mouton Entrefino (race Castellana)

<https://www.mapa.gob.es/fr/ganaderia/temas/zootecnia/razas-ganaderas/razas/catalogo-razas/ovino/castellana/>

Les moutons de la péninsule ibérique ne sont pas les seuls à présenter une diversité génétique importante. Plus à l'Est, dans les Balkans, se trouvent de nombreuses races qui constituent également un réservoir de diversité.

### 2.1.6 Les moutons des Balkans

Les Balkans présentent de nombreuses races ovines autochtones. Malheureusement, la plupart sont en danger d'extinction, notamment à cause de croisements incontrôlés avec des races industrielles importées, comme les Lacaunes et les Assaf<sup>5</sup> (Gelasakis et al., 2012; Savic et al., 2011). Seules deux races sont soumises à des programmes d'amélioration : la Chios, et la Karagouniko (Michailidou et al., 2018).

---

<sup>5</sup> Assaf : croisement entre les races Awassi et Assaf

Dans cette région, l'origine des races ou des types de races est souvent difficile à retracer d'autant plus que leur nom change fréquemment d'un pays à l'autre. C'est par exemple le cas du type de race Zackel qui se nomme aussi Ratska pour les Hongrois et les Roumains et parfois Valaska Vitorioga par les Serbes qui tentent d'ailleurs de changer le nom en Vitoroga Zackel (Draganescu C., 2007).

L'origine des moutons Zackel est controversée mais leur cornes en hélice permettent de formuler deux hypothèses (Dudu et al., 2016) (**Figure 10**): (i) Une origine antique Egyptienne où l'on retrouve des moutons à cornes en hélice chez les mâles et les femelles ; (ii) Une origine mésopotamienne où les cornes en hélice sont présentes uniquement chez les béliers

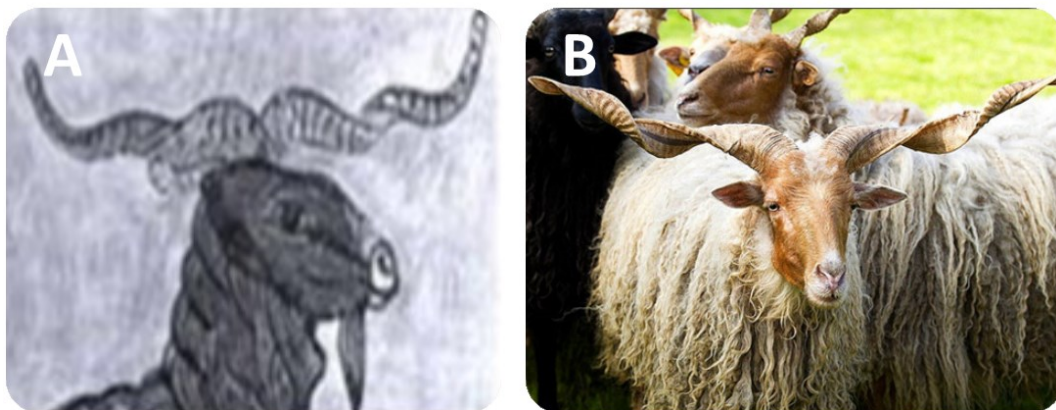


Figure 10 : Comparaison des types de cornes entre (A) une représentation antique égyptienne (3900 ans AP) et (B) un mouton de type Zackel

A : (Gootwine, 2018)

B : <http://bib.ge/sheep/pictures2.php?id=5434>

Dudu et al. (2016) montrent la présence d'au moins deux haplogroupes mitochondriaux dans ces populations à effectifs pourtant très réduits. On y retrouve le B, majoritaire en Europe mais aussi le A majoritaire en Asie et dont la présence peut être reliée à l'influence Turc dans cette région de l'Europe (voir partie 2.1.1).

## 2.1.7 Les moutons européens : bilan

Les moutons ont été introduits en Europe via deux voies géographiquement distinctes. La voie Méditerranéenne et la voie Danubienne. De plus, au moins deux épisodes migratoires se sont succédés et ont provoqué le remplacement partiel ou total des races les plus primitives par des races plus modernes, à l'exception de certaines populations en Europe du Nord et sur les îles méditerranéennes. L'histoire des moutons Européens est marquée par de nombreux échanges au sein du continent mais aussi avec les continents voisins, notamment via le bassin Méditerranéen. Ces multiples influences ont particulièrement marqué les races du Sud-Est qui sont plus diversifiées que celles du Nord. Il en résulte aujourd'hui une structure génétique marquée par un gradient Nord-Sud.

Toujours à l'ouest du centre de domestication, les moutons ont aussi été importés en Afrique. Cela ne s'est pas fait complètement indépendamment des événements qui ont eu lieu en Europe.

## Afrique

On trouve en Afrique plus de 400 races (International Livestock Research Institute, 2020) réparties en deux groupes : les races à queue fine et celle à queue grasse (Gifford-Gonzalez and Hanotte, 2011; Rege, Yapi-Gnaore and Tawah, 1996; Wilson, 1991). Le type queue fine est celui que l'on retrouve majoritairement dans le monde alors que le type queue grasse ne représente que 25% du cheptel mondial. Les moutons à queue grasse présentent une accumulation de graisse dans la queue ou la croupe et sont présents en Asie, au Moyen-Orient, marginalement dans la partie méditerranéenne de l'Europe et comme nous allons le voir, dans certaines régions d'Afrique (Mason, 1988).

Le **type queue fine** se retrouve en Afrique du nord (du Maroc à l'Égypte et au Soudan), en Afrique de l'ouest (du Sénégal au Nigéria). Ils se divisent en deux sous-groupes. Le **type Sahel (Figure 11)** regroupe des races à poils, avec de longues pattes et des oreilles tombantes. Les mâles présentent des cornes en hélice (Mason, 1988) qui ressemblent à celles de représentations datant de l'ère Thinite (Gifford-Gonzalez and Hanotte, 2011). Ce type ovin se retrouve dans les zones semi-arides du Sahel et de l'Afrique de l'Ouest. Les **moutons tropicaux nains, ou Djallonké (Figure 12)**, ont des poils, des cornes recourbées vers le bas, une collerette ou une crinière et sont considérés trypanotolérant (Animal Genetics Training Resource, 2020). Ils sont particulièrement adaptés aux milieux tropicaux humides d'Afrique centrale et de l'Ouest (Álvarez et al., 2020a; Muigai and Hanotte, 2013).

Les **moutons à queue grasse (Figure 13)** se distribuent en Afrique du Nord de l'Algérie à l'Égypte, et de l'Est, de l'Érythrée à l'Afrique du sud. Ils se séparent en deux sous-groupes à poils ou à laine (Rege, Yapi-Gnaore and Tawah, 1996). Les races à laine se retrouvent au Nord du continent tandis que celles à poils se concentrent majoritairement en Afrique sub-saharienne (Mason, 1988).



Figure 11 : Mouton de type Sahel (Fitzhugh et al., 1983)



Figure 12 : Mouton Djallonké Meyer C., ed. sc., 2021, Dictionnaire des Sciences Animales. Montpellier, France, Cirad.



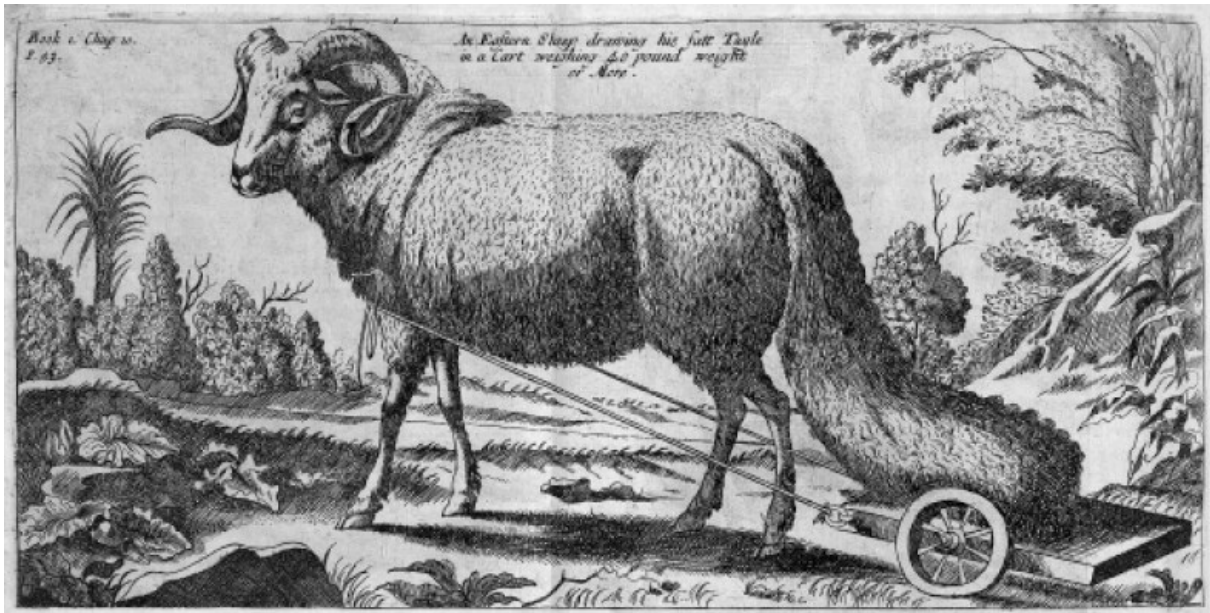


Figure 13 : Illustration d'un mouton à queue grasse datant de 1684  
 Par Job Ludolphus — A New History of Ethiopia (1684), Domaine public,  
<https://commons.wikimedia.org/w/index.php?curid=42159786>

Les moutons africains ont été moins étudiés au niveau génétique que les Asiatiques et Européens. Cependant, les quelques études les concernant permettent tout de même de compléter les informations archéologiques et historiques.

## 2.3.1. La diffusion des moutons en Afrique

### 2.3.1.1. Une double influence au Nord

D'après les relevés archéologiques, on pense que le mouton arrive en Afrique autour de 7 à 8000 ans AP (**Figure 14**) où on le trouve souvent associé aux sociétés égyptiennes (Blench and Donald, 1999). Des restes de cette espèce datant de 7000-7500 ans AP dans le delta du Nil, à l'Est du Sahara et dans les collines de la Mer Rouge (Close, 1992; Gautier, 1987; Smith, 1992; Vermeersch et al., 1996) ont été trouvés. Il est probable que les moutons soient arrivés depuis le Levant et le Sinaï puisque des restes de moutons de la même période y ont été retrouvés (Oren dans (Muigai and Hanotte, 2013). La dispersion a sûrement suivi le Nil et la côte méditerranéenne. On retrouve, vers 6100 BP, des restes ovins à Merimda Beni Salama en Egypte et au centre de la vallée du Nil un siècle plus tard (Gautier, 2002). Ils atteignent très vite la Lybie en 6800-6500, peut-être grâce au transport maritime (Close, 1992; Smith, 1992; Holl, 1998), le centre du Sahara il y a 6000 ans puis la péninsule tingitane au Nord du Maroc il y a 7400 ans AP environ (Martínez-Sánchez et al., 2018).

Comme en Europe on relève plusieurs épisodes migratoires. Au Nord, des analyses génétiques basées sur de l'ADN mitochondrial ont mis en évidence deux vagues d'arrivées de moutons dans cette région. La plus ancienne daterait d'il y a 8600 ans et la seconde d'il y a 7100 ans AP, d'après l'estimation de dates d'expansions basées sur les données génétiques et calibrées par les relevés archéologiques (Kandoussi et al., 2020). La première pourrait être

associée à une migration des premiers Berbères en réponse à l'aridification du Sahara du Nord à la fin de la période subpluviale Néolithique africaine. La seconde, quant à elle, viendrait plutôt du Sud de l'Europe, de l'Ibérie précisément. C'est d'ailleurs cette influence qui semble prédominer dans les génomes des races marocaines aujourd'hui (Kandoussi et al., 2020). Une influence dont on trouve aussi des empreintes dans les génomes anciens humains de l'époque (Fregel et al., 2018), ce qui indique qu'il y a bien eu des migrations entre Europe et Afrique et pas seulement des échanges de bétail. Deux autres importations indépendantes d'individus en provenance d'Ibérie et d'Italie ont été enregistrées, contribuant au plus à 10% de la diversité génétique chacune. Les liens entre Europe et Afrique se poursuivent jusque récemment avec des croisements entre races Italiennes et Nord africains recensés au 18<sup>ème</sup> siècle (Ciani et al., 2014).

Plus globalement, l'étude de l'ADN mitochondrial montre que l'Europe et l'Afrique présentent une composition haplotypique similaire avec une majorité d'haplogroupe de type B contre une faible fréquence d'haplogroupe A (Álvarez et al., 2020a; Gornas et al., 2011; Horsburgh and Rhines, 2010). Cela est vrai pour les populations actuelles mais c'était aussi déjà le cas il y a 2000 ans, comme le montre l'étude de spécimens archéologiques (Horsburgh and Rhines, 2010). La prépondérance de l'haplogroupe B en Europe et en Afrique contrastée avec celle de l'haplogroupe A en Asie suggèrent une origine commune et/ou de nombreux échanges entre les populations européennes et africaines.

### 2.3.1.2. Deux voies de migrations vers le Sud

On sait que les moutons domestiques sont présents en Afrique du Sud depuis 2000 à 2500 ans grâce à des données archéologiques (Albrecht et al., 2001; Kinahan, 1991; Sealy and Yates, 1996, 1994; Smith, 2009, 2008, 2006). Comme en Europe, il semble qu'il existe au moins deux voies de diffusion (**Figure 14**) (Gifford-Gonzalez and Hanotte, 2011; Smith, 2009) : Une route par l'Ouest via les côtes et à travers le bassin du Congo et une autre par l'Est, le long des régions montagneuses de la vallée du Rift.

La première est supportée par plusieurs découvertes archéologiques. D'abord des restes de moutons au Botswana datant de l'âge du fer pourrait faire de ce pays une porte d'entrée vers l'Afrique du Sud (Robbins et al., 2008). Ensuite, des os et des dents de moutons ont été trouvées dans la Grotte du Léopard, en Namibie (Pleurdeau et al., 2012) et datent de 2270 ans AP. De plus, cette région a pu constituer un corridor de migration à cette époque plus humide facilitant ainsi la migrations d'humains et de bétail à travers le désert du Kalahari entre 2500 et 2000 ans AP (Shaw et al., 2003; Shi et al., 2001) pour atteindre le fleuve Orange, le Namaqualand et l'extrême sud de l'Afrique.

La seconde route est supportée par des preuves génétiques et archéologiques. En effet il existe un partage haplotypique non négligeable entre moutons Kenyans et Sud-Africains (Resende et al., 2016). Au niveau archéologique, la diffusion vers l'Afrique du Sud par l'Est du continent est supportée par des restes ovins dans le bassin du lac Turkana datant d'il y a 4500-6500 ans (Barthelme, 1985; MacDonald and MacDonald, 2006). De plus, des connexions

commerciales entre l’Érythrée et la péninsule Arabique ont pu constituer un point d’entrée secondaire vers l’Afrique et la route de colonisation Est avec notamment des races à queue grasse (Boivin and Fuller, 2009; Chessa et al., 2009a; Marshall and Hildebrand, 2002; Muigai and Hanotte, 2013; Smith, 1992). Il n’y a en revanche pas de preuves d’introduction massive de moutons domestiques depuis le Sud de l’Asie via l’océan Indien (Resende et al., 2016) si ce n’est la présence d’un port de commerce important de la vallée de l’Indus : le port de Lothal.

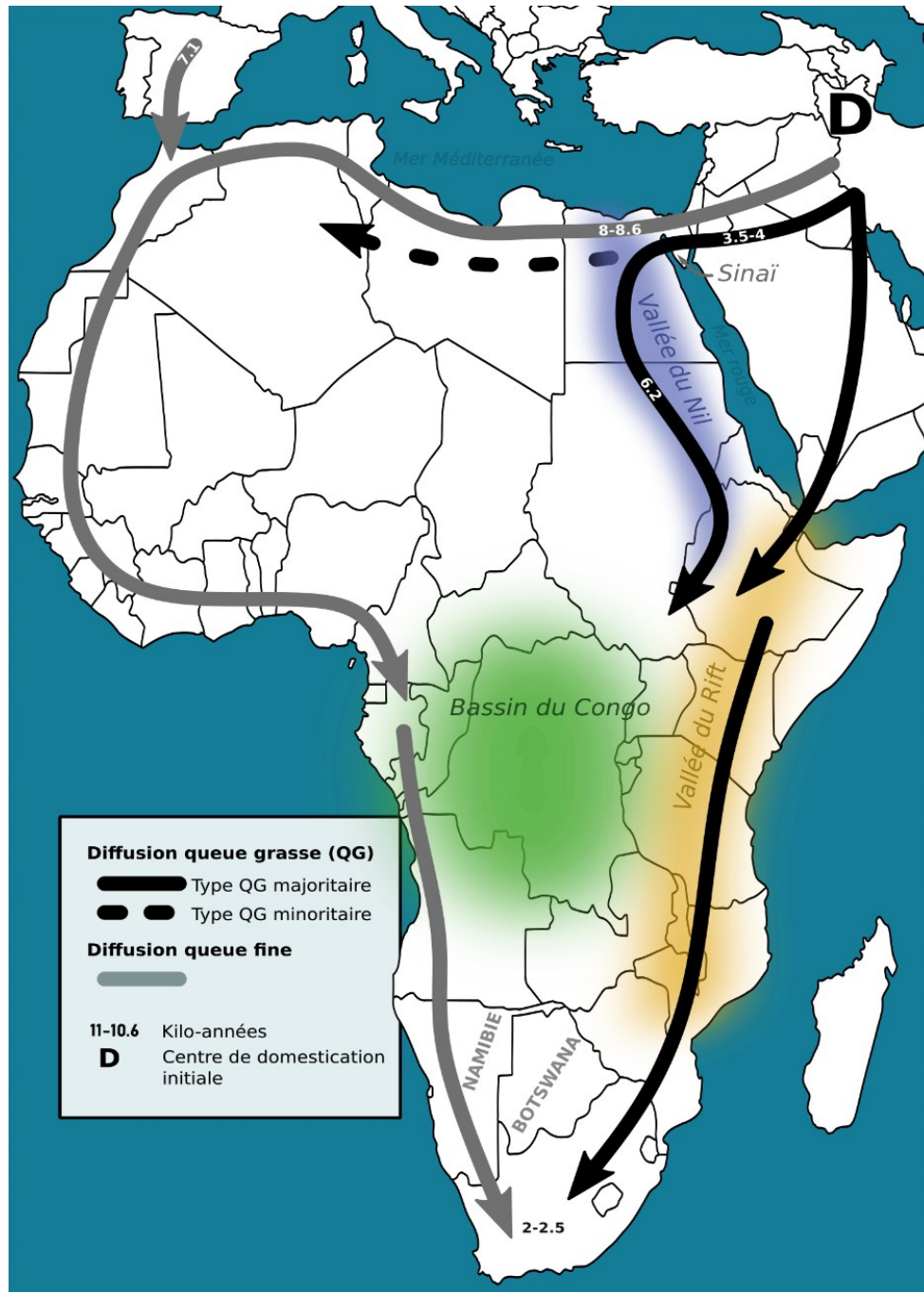


Figure 14 : Voies de diffusion des moutons en Afrique

Les dates sont exprimées en années avant le présent et correspondent aux plus anciennes preuves (archéologiques ou génétiques) de la présence de moutons à l’endroit où elles sont indiquées

### 2.3.1.3. Dispersions dissociées des races à queue fin et à queue grasse

La voie de colonisation Ouest vers le Sud a probablement concerné seulement des races à queue fine, alors que celle de l'Est comprenait sûrement des races à queue grasse (**Figure 14**). En effet, les moutons à queue grasse et à queue fine se distribuent respectivement majoritairement à l'Est et à l'Ouest de l'Afrique. Les deux voies se rencontrent au sud du continent, où l'on trouve les deux types ovins (Rocha, Chen and Beja-Pereira, 2011).

La diffusion des moutons à queue grasse et queue fine ne s'est probablement pas faite au même moment. On retrouve des représentations de moutons à queue fine datant de 5150-4700 ans AP dans des tombes égyptiennes, alors que celles de moutons à queue grasse ne font leur apparition qu'à partir de 4055-3620 ans AP environ (Clutton-Brock, 1993). En Afrique de l'Ouest, on ne compte aucune représentation de moutons à queue grasse (Muigai and Hanotte, 2013). Il existe aussi une séparation nette entre les stocks génétiques des races africaines à queue grasse de l'Est et du Sud et les moutons d'Afrique de l'Ouest (Aswani, 2007; Muigai, A. W. T, 2003). De plus, on trouve dans le génome du mouton de Djallonké à queue fine les mêmes marqueurs rétroviraux que ceux présents chez les moutons primitifs d'Europe du nord indiquant ainsi une origine plus ancienne et commune avec l'Europe (Chessa et al., 2009a) (voir partie **2.1.3**). Enfin, Deng et al. 2020 ont montré à partir de la combinaison de données génétiques et archéologiques que l'expansion des moutons à queue grasse a eu lieu entre 5300 et 1700 ans AP. Tout cela suggère une première vague de colonisation du continent par des races à queue fine, puis une deuxième par des races à queue grasse. Ces dernières ont remplacé où se sont hybridées avec les races à queue fine déjà présentes (Muigai and Hanotte, 2013).

### 2.3.2. Homogénéité génétique en Afrique du Nord

Le cheptel de moutons au Maghreb date d'il y a 7000 ans (Hassan, Blench and MacDonald, 2006) et contient une grande diversité de populations maintenues en élevage traditionnel depuis des millénaires (Taberlet et al., 2008). Ce système a permis l'émergence de races adaptées localement dans les divers territoires du Maghreb. En Afrique du Nord, les plus gros producteurs de moutons sont l'Algérie et le Maroc avec plus de 20 millions de tête chacun. Cette espèce constitue 80% de la production animale (MADRDSASI, 2010).

Les principales races reconnues par les autorités publiques sont classées par caractérisation morphologiques (Boujenane and Petit, 2016) et des protéines du sang (Boujenane et al., 2008). Comme les races maghrébines n'ont pas été vraiment standardisées au 20<sup>ème</sup> siècle comme les races européennes, il existe une grande variabilité intra-race. Cependant, depuis quelques décennies, la préférence des éleveurs va à la race Ouled-Djellal (**Figure 15**). Elle possède une conformation plus large ce qui en fait la race la plus productive. Elle envahit peu à peu le cheptel ovin et constituait déjà plus de 60 % de la population en 2010 (MADRDSASI, 2010). Les croisements s'intensifient de manière anarchique pour augmenter la productivité d'autres races (Djaout et al., 2017; Madani, Yakhlef and Abbache, 2003; MATET, 2009) comme les Rembi, Taâdmit et Berber (Gaouar et al., 2015, 2017). D'ailleurs la dilution

génétique des Rembi est dénoncée depuis plus de 20 ans (Gaouar et al., 2015; Madani, Yakhlef and Abbache, 2003). Elle a notamment provoqué la perte de leur couleur baie-fauve en faveur du blanc des Ouled-Djellal (Chelling, 1992; Madani, Yakhlef and Abbache, 2003). La période coloniale, le développement des transports mécanisés, l'intensification de l'élevage et la sédentarisation ont sûrement joué un rôle important dans l'expansion des Ouled-Djellal (Flamand et al., 1893; Viger, 2016). On peut noter que la race Beni-Guil est également privilégiée pour augmenter la productivité (Bechchari et al., 2014).



Figure 15 : Mouton de race Ouled-Djellal  
<https://agronomie.info/fr/la-race-arabe-ouled-djellal-2/>

Quelques races anciennes résistent tout de même à ce phénomène. C'est le cas de la Berber, peut-être la plus ancienne race du Maghreb (Mason, 1988) dont quelques populations isolées dans les montagnes restent préservées (Couput, 1900). Elle a évolué pendant des siècles dans les montagnes de Kabylie et représentait encore 20% du cheptel Algérien il y a 30 ans. Aujourd'hui c'est une race minoritaire (FAO, 2014) qui subit le remplacement par les Ouled-Djellal depuis 1950 (Sagne and d'Arces, 1950). La Barbarine est une autre race ancienne provenant du Proche-Orient. Il s'agit de la seule race à queue grasse de la région. Elle est également menacée de dilution par les Ouled-Djellal mais dans une moindre mesure que les Berber (Gaouar et al., 2017). Elle subit un déclin important : -20% entre 1964 et 2000 en Tunisie (Aw-Hassan, Shomo and Iniguez, 2008) et jusqu'à -60% entre 1990 et 2000 en Algérie (Laaziz, 2005).

Il semble que l'invasion des Ouled-Djellal soit une des principales causes de l'homogénéisation des races au Maroc. Gaouar et al. (2015, 2017) ont montré qu'il n'existe aucune différenciation génétique entre Berber, Ouled-Djellal, Rembi et Taâdmit. Seules les Sidaoun, D'man, Tazegzawth, et Beni-guil semblent plus structurées dans ces études. Belabdi *et al.* (2019) montrent, quant à eux, que l'homogénéisation ne touche pas de la même manière tous les types d'élevage et toutes les régions pour une seule et même race. Ainsi les Ouled-Djellal, Timahdite, Sardi, et Beni-Guil sont plus touchées dans les fermes privées algériennes que dans les fermes d'Etat. Les D'man, elles, s'homogénéisent au Maroc mais pas en Algérie. On ne peut cependant pas écarter l'hypothèse d'une homogénéité due à un pool génétique de base déjà peu différencié pour certaines races.

La diversité génétique présente dans les races locales et traditionnelles constitue le potentiel adaptatif d'une population. Ce phénomène d'homogénéisation, dont on ignore encore l'ampleur globale, interroge déjà sur la capacité d'adaptation des populations ovines futures dans le contexte actuel de changement climatique.

## 2.1.8 Les moutons africains : bilan

Comme en Europe, la diffusion des moutons dans le continent africain s'est faite par plusieurs routes et plusieurs épisodes. C'est le Nord du continent qui a été gagné en premier par des races à queue fine, puis le Sud par deux voies différentes à l'Est et à l'Ouest du continent. Les moutons à queue grasse sont vraisemblablement arrivés plus tard en Afrique et ont diffusé majoritairement à l'Est du continent. Le continent est ainsi caractérisé par une grande diversité phénotypique de mouton (queue grasse ou fine, poils ou laine etc.). La structure génétique est faible car il n'y a pas vraiment eu de standardisation des races et à cause de nombreux croisements incontrôlés. Le Nord du continent est touché par un processus d'invasion génétique par la race Ouled-Djellal, qui est utilisée pour accroître la productivité des autres races. Cependant, l'ampleur de ce processus dépend des pratiques d'élevage et indique que la diversité initiale de races anciennes pourrait être conservée dans des élevages traditionnels.

Le suivi de la diffusion vers l'ouest pourrait s'étendre au continent américain, mais l'histoire du mouton y est très récente (1493) et résulte de l'introduction de race issues du vieux continent. C'est pourquoi nous partons dès maintenant vers l'Est sur les routes de diffusion asiatiques.

## Asie

### 2.4.1. Chine

La Chine continentale compte 67 millions de moutons et 42 races indigènes (China National Commission of Animal Genetic Resources, 2011). Plus de la moitié de cet effectif (34 millions) se trouve dans la province de Xinjiang au Nord-Ouest du pays. Cependant, les races indigènes ne représentent qu'une faible proportion du total avec, par exemple, 4 millions d'individus sur cette province (enquête gouvernementale de 2007). Elles sont en fort déclin depuis une trentaine d'années (Chen and Chen, 2002; Gao et al., 2008; Jiang, 2006) pour plusieurs raisons : d'abord l'introduction et le croisement avec de nombreuses races exotiques dans la deuxième moitié du siècle 20<sup>ème</sup> dans le cadre d'une campagne nationale d'amélioration des races natives (Ma et al., 2006). Beaucoup de diversité génétique a alors été perdue à cause de corrélations non favorables entre traits économiques et adaptatifs (Rao and Notter, 2000). L'industrialisation massive, l'absence de plan de conservation et les changements de professions sont aussi des facteurs importants dans le déclin de ces races (E et al., 2016; Ma et al., 2006; Zhong et al., 2010). Aujourd'hui, la majorité des moutons chinois résultent de croisements et 15% des races indigènes chinoises sont en danger d'extinction (Zhong et al., 2010).

#### 2.4.1.1. Arrivée et établissement des moutons en Chine

Plusieurs études ont daté des restes de caprinés en Chine : des fragments d'os et de dents datant d'environ 7000-5000 ans AP ont été trouvés dans des sites Néolithiques des provinces de Qinghai, Gansu et Shaanxi au centre de la Chine (Li and Han, 1959; Sang, 1979;

Zhou, 1983, 1999). Plus tard des mandibules furent datées aux alentours de 5600-5000 ans AP (Yuan, 2007). Enfin, en 2014, Dodson et al. (2014) déclare avoir trouvé la plus ancienne trace de moutons domestiques identifiés avec certitude en Chine par datation directe à environ 5700 ans AP, toujours dans des provinces au centre et au Nord du pays (Mongolie Intérieure, Ningxia, Shaanxi) (**Figure 16**). Le problème de toutes ces études est l'incertitude quant à l'identification de l'espèce dont sont issus les os/dents. Par exemple dans l'étude la plus récente, elle a été faite à partir de séquences ADN de 78nt à 100% identiques à une séquence de mouton (*Ovis aries*) mais aussi 100% identiques à celle de l'Urial (*Ovis vignei*) et de l'Argali (*Ovis ammon*). Cependant, des traces de millet dans l'alimentation, et le contexte du site sont des arguments pour une identification domestique.

D'une manière plus générale, on observe une augmentation des os de moutons dans les sites des plaines centrales à partir de 4500 ans AP et des enterrements volontaires de moutons dès 5000 ans AP dans la région de Gansu et dès 4000 ans plus à l'Est (**Figure 16**) (Jing, 2008). A partir de la Dynastie Zhou (3046-2771 AP), le mouton prend une grande importance dans la culture chinoise et est même utilisé lors de cérémonie rituelles (Sun, 1987). On peut donc dire que le mouton est largement établi en Chine à la fin de l'âge du Bronze. Les données archéologiques concernant la présence de moutons/chèvres en Chine sont beaucoup plus abondantes dans le Nord du pays. Pour un revue complète des sites où des restes caprins ont été retrouvés, on peut se référer à Wang 2017 – Partie 2 & Appendice A.

Les moutons domestiques sont donc probablement arrivés en Chine via le Nord du pays aux alentours de 6000 ans AP. Leur dispersion (et leur divergence) a pu se faire grâce aux migrations massives du peuple Di-Qiang<sup>6</sup> d'abord vers les provinces de Henan et Shanxi à l'Est vers 3400 AP puis plus au Sud, via le corridor « Tibeto-Yi »<sup>7</sup> entre 2000 et 2600 ans AP lors des guerres d'unification de Qin<sup>8</sup> (**Figure 16**). C'est à la suite de ces mouvements massifs de populations que se forment plusieurs groupes ethniques tels que les Yangshao<sup>9</sup>, les Bo-Qiang<sup>10</sup>, les Yi, les Lahu etc. ((Fan, 1131; He, 2000; Ou, 1936; Ran, Li and Zhou, 1985, 1985; Si, 1131) dans Zhao et al., 2017b). Certains migrent vers le plateau Tibétain, d'autres vers le Sichuan et le plateau du Yunnan Kweichow. Ces différentes étapes de colonisation sont retrouvées dans une étude génétique menée par Zhao et al., (2017b). Les résultats montrent que le meilleur scénario consiste en une séparation des moutons nord-chinois il y a environ

---

<sup>6</sup> **Peuple Di-Qiang** : Tribus peuplant les hautes vallées du Fleuve Jaune au nord-Ouest de la Chine il y a 5000 ans. Ils vivent principalement de l'élevage ovin. Le mouton est leur totem et le sigle chinois les désignant est composé des signes « homme » et « mouton ». Fondateurs de plusieurs autres cultures

<sup>7</sup> **Corridor Tibeto-Yi** : passage naturel entre le Nord-Ouest et le Sud-ouest de la Chine, situé sur la limite orientale du Tibet.

<sup>8</sup> **Guerres d'unification de Qin** : Série de campagnes militaires lancées vers la fin du IIIe siècle av. J.-C. contre les autres Royaumes combattants — Han, Zhao, Yan, Wei, Chu et Qi — dans les territoires formant la Chine moderne. ([https://fr.wikipedia.org/wiki/Guerres\\_d%27unification\\_de\\_Qin](https://fr.wikipedia.org/wiki/Guerres_d%27unification_de_Qin))

<sup>9</sup> **Yangshao** : première culture néolithique en Chine (sud-ouest), considérée comme ancêtre des Tibétains et des Yi.

<sup>10</sup> **Bo-qiang** : peuple descendant des Di-Qiang et constituant la population ancestrale des proto-tibétains (ancêtre des Tibétains)

4000 ans et donc antérieure à la divergence des moutons des plateaux du Tibet et du Yunnan Kweichow il y a 2500 ans. Ils indiquent également un mélange plus tardif entre les populations du Nord et des moutons du Moyen-Orient il y a 1600 ans. D'autres éléments viennent appuyer ce scénario. Les populations du Nord présentent une diversité génétique plus élevée qu'au Sud. Cela correspond bien à une divergence antérieure et à la longue histoire d'isolement et d'effets fondateurs des populations des plateaux du Tibet et du Yunnan Kweichow (Lv et al., 2015; Zhao et al., 2017b) (Figure 16). Ensuite, la présence de certains haplogroupes mitochondriaux vient étayer la thèse d'une introgression tardive de moutons du Moyen-Orient. On retrouve l'haplogroupe C, typique des moutons à queue grasse et un gradient d'haplogroupe B, typique de L'Europe et du Moyen-Orient, décroissant vers l'Est (Deniskova et al., 2018; Lv et al., 2015). Cela résulterait probablement des mouvements de populations du Moyen-Orient vers l'Est lors des grandes migrations turques provoquées par les invasions Mongoles (1219-1260 ap JC)<sup>11</sup> (Grousset, 1941) , et sûrement aussi par les échanges commerciaux, notamment via la Route de la Soie .

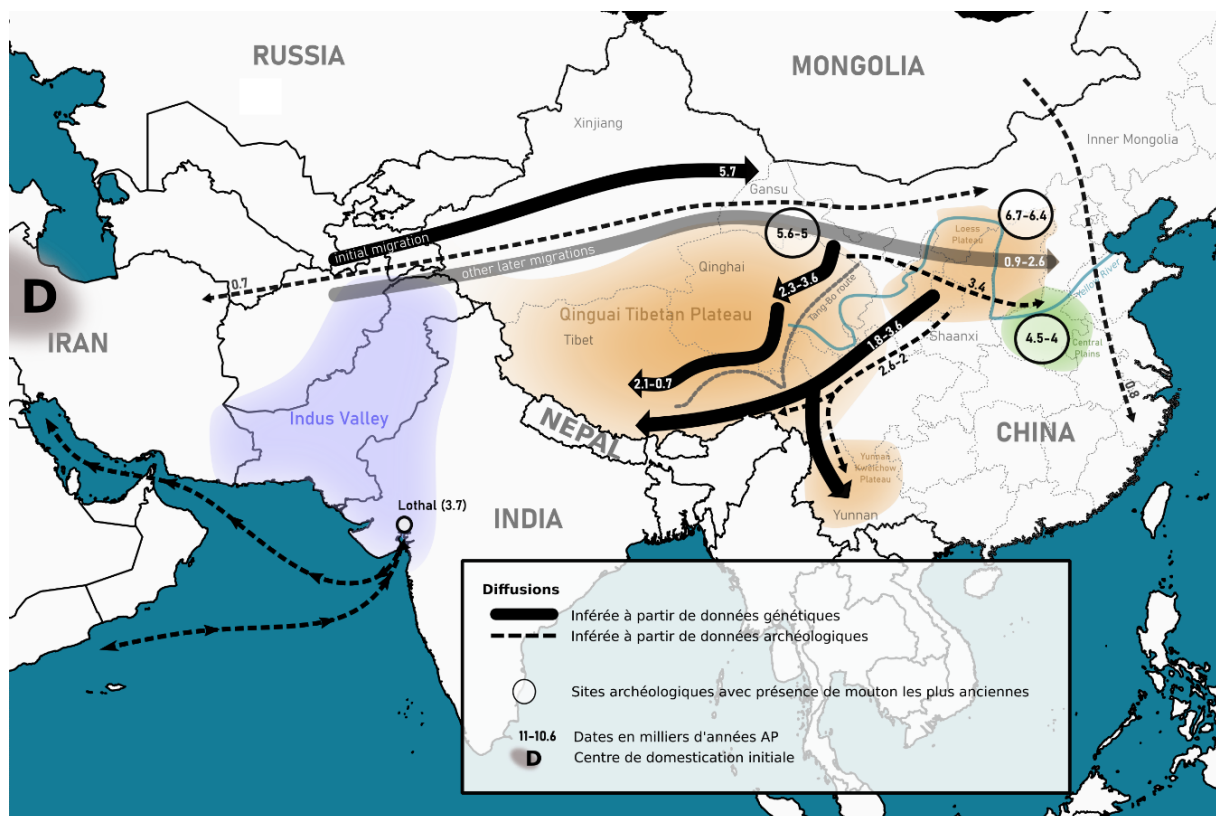


Figure 16 : Voies de diffusion des moutons en Asie

Les dates sont exprimées en années avant le présent et correspondent aux plus anciennes preuves (archéologiques ou génétiques) de la présence de moutons à l'endroit où elles sont indiquées

<sup>11</sup> **Précisions sur les invasions Mongoles :** Au début du 12<sup>ème</sup> siècle, les Mongoles, menés par Gengis Khan, envahissent la Chine par l'est et arrivent finalement dans l'actuelle province de Zhejiang. Un peu plus tard, au 13<sup>ème</sup> siècle, les Mongoles continuent leurs invasions vers l'ouest via la province de Xinjiang et la route de la Soie du Nord. Ces invasions provoquent la migration de nombreux musulmans du Moyen-Orient vers l'est, toujours le long de la route de la Soie.



#### 2.4.1.2. Cas du plateau Tibétain

Sur le plateau Tibétain (englobant les provinces du Tibet et du Qinghai), la culture nomade prédomine. Le mouton tibétain est l'une des races indigènes décrit par le (China National Commission of Animal Genetic Resources, 2011) et vit à plus de 4000m d'altitude. Cette race prédomine sur le plateau avec environ 20 millions d'individus (Du 2011 in (Hu et al., 2019)). Ils fournissent une grande variété de matières brutes, autant pour les besoins primaires (alimentaire, vestimentaire...) que culturels (totem religieux, ornement...) (Chen et al., 2006; Liu et al., 2016; Müller et al., 2012; Peng, 2010; Xian, 2005). Les moutons de ce plateau ont été plus isolés, et moins croisés avec des races exotiques que dans le nord de la chine.

La colonisation humaine de ce plateau remonte à au moins 20 000 ans (Qiu, 2015). Entre 5200 et 3600 AP, des sociétés humaines s'établissent sur les marges du plateau, vivent de la culture du millet et parfois de l'élevage de quelques animaux domestiques selon des données archéologiques (d'Alpoim Guedes et al., 2014). Plus tard, entre 3600 et 2300 AP, les plus hautes altitudes sont colonisées et les cultures changent pour des espèces plus résistantes au froid comme l'orge et le blé (Chen et al., 2015; D'Alpoim Guedes and Butler, 2014; d'Alpoim Guedes et al., 2015; Han et al., 2016). Plus généralement on passe à un système agro-pastoral avec d'avantage de moutons parmi les animaux d'élevage. Bien que les voies de diffusion précises soient encore mal connues, certains éléments montrent que son occupation permanente commence il y a 3600 ans environ (Chen et al., 2015) et même dès 7400 grâce à la ré-analyse du site de Chusang (quart sud-est du plateau) (Meyer et al., 2017).

Une étude récente (Hu et al., 2019) précise cette colonisation via l'étude génétique de moutons tibétains. Il semblerait que le processus se soit décomposé en deux phases. D'abord avec l'arrivée d'une population ancestrale au Nord du Qinghai il y a environ 3100 ans, probablement lié à des changements climatiques (deMenocal and Stringer, 2016; Han et al., 2016; Parmesan and Yohe, 2003). Dans un second temps, une migration au Sud-Ouest vers le Tibet qui résulte en une divergence des lignées de moutons du Tibet et du Qinghai vers 1300 AP et sûrement liée à la croissance démographique humaine (Han et al., 2016). Cette migration aurait d'ailleurs pu suivre l'ancienne route du Tang-Bo qui relie notamment le Nord du Qinghai et le centre du Tibet en passant par le tripoint Tibet-Qinghai-Sichuan (Hu et al., 2019).

#### 2.4.2. Inde

L'Inde compte une quarantaine de races identifiées par le NBAGR (National Bureau of Animal Genetic Ressources, Karnal) dont certaines sont surreprésentées comme la Deccani, la Bannur et la Garole (Pardeshi et al., 2007). On les retrouve dans le sud du pays qui concentre plus de 50% de la population ovine indienne. Cette zone concentre plutôt des races à poils tandis que le Nord présente une majorité de races adaptées au climat aride ou semi-aride à laine grossière et queue fine (Arora, Yadav and Mishra, 2013). Il n'y a que très peu de

mouvements entre les états indiens et chaque communauté suit des routes de transhumance fixes. Ce refus d'abandonner les pratiques traditionnelles a sûrement contribué au maintien de la diversité des races en Inde et à la forte structure qui en résulte (Acharya, 1982).

Les haplogroupes mitochondriaux en Inde, comme dans le reste de l'Asie, sont majoritairement du type A excepté dans une région côtière située aux abords de l'ancien port de Lothal<sup>12</sup> (Singh et al., 2013), qui fut un centre actif d'échanges avec la Mésopotamie, l'Égypte et la Perse. A cet endroit on observe une haute fréquence de l'haplogroupe mitochondrial B, typique de l'Afrique et de l'Europe mais très peu présent en Asie (Singh, Kumar and Kolte, 2013).

### 2.4.3. Népal

Au Népal, quatre races locales représentent 90% du cheptel national et se répartissent sur un gradient altitudinal. Il n'y a que peu de croisements entre ces races (Gorkhali, Han and Ma, 2015) et les groupes de hautes et basse altitude ont un fond génétique et une histoire différentes (Gorkhali, Han and Ma, 2015). On y retrouve donc une structure génétique forte comme en Inde.

### 2.4.4. Russie

Le climat Russe est favorable à une forte demande en laine. D'ailleurs, dans les années 90, la Russie était parmi les premiers producteurs de laine au monde. On comptait alors 58 millions de moutons sur l'ensemble du territoire. Ce chiffre est passé à 24 millions en 2014 avec la baisse de la demande. De même, le nombre de races est passé de 45 à 27 et la proportion de races à laine de 90% à 50% (Deniskova et al., 2018; Lescheva, Ivolska and Cvijanovic, 2015). Ces dernières se divisent en deux catégories : les races à laine grossière et celles à laine semi-fine. Les races à laine grossière sont bien adaptées à leur environnement. Elles sont légèrement améliorées par croisement et présentent une grande diversité de morphologie de queues (queue grasse longue, croupe grasse ...). Les races à laine semi-fine sont aussi des races locales mais ont été fortement influencées par des races européennes comme les Lincoln, les Romney ou les Mérinos pendant la période soviétique pour améliorer leur productivité ((Sel'kin II and Aboneev VV, 2007; Sel'kin II and Sokolov AN, 2002) dans Deniskova et al., 2018). Quasiment toutes les races ont subi une baisse démographique pendant la restructuration soviétique (Pérestroïka) à cause de la crise du secteur agricole et le manque de ressources (Deniskova et al., 2018).

### 2.4.5. Mongolie

#### 2.4.5.1. Histoire récente

L'élevage de mouton/chèvre est la principale activité économique en Mongolie qui produit notamment 30 % du cachemire mondial (Takahashi et al., 2008). Les moutons représentent à eux seuls 60 % des animaux domestiques du pays et 45% du bétail, soit 30.1 millions

---

<sup>12</sup> Port important de la civilisation de l'Indus datant d'environ 3700 ans AP et découvert par S.R. RAO.

d'individus. Cela en fait la 3<sup>ème</sup> population mondiale (NSO 2017). Le cheptel se compose de races rustiques bien adaptées aux conditions rudes du pays, mais aussi moins productives. Les petits éleveurs composent 80% de la Mongolie (156 ha) et il n'y a pas vraiment de programmes d'élevage officiel. Quelques races étrangères ont été introduites pendant la période soviétique (1950-1990) depuis les Pays socialistes dans le cadre d'un programme d'élevage contrôlé par les vétérinaires Russes pour améliorer les performances des animaux (Ganbold et al., 2019; Ganbold, Munkhbayar and Paek, 2018). D'un autre côté, dans les années 70s, des races Mongoles ont été sélectionnées par rapport à leur robes et leur capacités d'adaptation à l'environnement (Takahashi et al., 2008). On note aussi une réduction dans les possibilités de dispersion à partir de 1949 et la création des districts (Takahashi et al., 2008).

#### **2.4.5.2. Origine des moutons Mongoles**

En Mongolie, on ne dispose que de peu d'informations. Quelques rapports archéologiques ne donnent qu'une large fourchette pour les débuts de l'élevage dans le Pays entre 10 000-7000 avant JC (Ganbold, Munkhbayar and Paek, 2018). D'autre part, un fort flux de gènes entre les populations suggère une longue histoire du nomadisme et une courte période de sélection artificielle intensive (Takahashi et al., 2008).

#### **2.4.6. Introgression sauvage**

Les moutons domestiques peuvent se reproduire avec différentes espèces de mouflons sauvages. Ce phénomène est parfois important comme dans certaines populations tibétaines où le retrouve environ 5% du génome introgressé depuis l'Argali (Hu et al., 2019). Ceci résulte d'une pratique commune dans les communautés nomades (Zhu, 2010) et a permis à ces moutons de s'adapter à la vie en haute altitude (Hu et al., 2019). La reproduction entre brebis domestiques et béliers sauvages est aussi documenté en Asie centrale et au Nord de la Chine par Carruthers (1949) et Man (2007). Comme en Europe avec les Mouflons Européens, des introgressions génétiques parfois adaptatives ont été identifiées. Elles concernent notamment les gènes HBB impliqué dans le transport de l'oxygène et donc potentiellement liée à l'adaptation à la haute altitude. Mais aussi des gènes en lien avec la morphologie comme RXFP2 qui est impliqué dans la présence, la forme et la taille des cornes (Hu et al., 2019; Kijas et al., 2016).

#### **2.1.9 Les moutons asiatiques : bilan**

Les moutons ont été introduits en Asie via le Nord de la Chine et se sont ensuite dispersés plus au Sud. Des introductions maritimes via la Péninsule arabique et l'Afrique sont aussi possibles mais plus tardivement. On y retrouve plusieurs épisodes migratoires au fil des millénaires tout comme en Europe et en Afrique. L'Asie présente des problématiques de perte de diversité similaires aux autres continents, surtout au Nord de la Chine à cause de nombreux croisements visant à augmenter la productivité des races locales. Cependant, les races de régions isolées comme le plateau tibétain ont pu être épargnées et conserve une importante différenciation génétique. D'autre part, l'Asie est le seul continent qui recouvre l'aire de distribution

d'espèces de mouflons sauvages comme l'Argali et l'Urial qui sont interféconds avec le mouton. Des introgressions adaptatives entre Argali et mouton semblent essentielles à l'adaptation des moutons à la vie sur les hauts plateaux.

Finalement, la diffusion des moutons dans les différents continents de l'Ancien Monde présente plusieurs similarités. Dans tous les continents, la diversité actuelle des moutons résulte de plusieurs épisodes migratoires et de diverses influences. Globalement, cela amène à une structuration faible des populations ovines mondiales. Nous avons vu que de multiples races existent et qu'elles sont caractérisés par divers phénotypes. Cela résulte de processus de sélections que nous abordons dans la partie suivante.

## Les gènes de la domestication

Une empreinte importante de la domestication dans le génome des animaux, est celle laissée par les sélections humaines. Volontairement ou non, l'homme a sélectionné les animaux les plus adaptés à ses besoins et à son environnement. Ces phénomènes de sélection laissent des traces dans le génome que l'on peut détecter notamment en comparant les moutons et leurs équivalents sauvages. En effet, les descendants sauvages de l'espèce domestiquée ont évolué parallèlement aux domestiques. On peut alors s'attendre à ce que les régions génomiques les plus différenciées entre mouflons et moutons soient liées au processus de domestication. Une seconde manière de retrouver des parties de génome liées à la domestication consiste à repérer au sein même de populations domestiques, les régions génomiques présentant une faible diversité génétique, la sélection ne favorisant que quelques variants favorables. C'est ce que l'on appelle un balayage sélectif. Lorsqu'un variant génétique est ciblé par de la sélection, sa fréquence augmente dans la population. La sienne, mais aussi celle des variants alentours à cause du déséquilibre de liaison. Il s'agit de l'association préférentielle d'une combinaison de variants et c'est ce qui cause ce patron de baisse de diversité. Une dernière manière de trouver des signaux de sélection est de tester les corrélations qui peuvent exister un génotype et une variable environnementale ou un phénotype.

On peut distinguer trois groupes de caractères sélectionnés au cours de la domestication. Il y a d'abord ceux liés au syndrome de domestication et qui regroupent un ensemble de traits morphologiques, physiologiques ou comportementaux, partagés par la plupart des espèces de mammifères domestiquées. On trouve dans ce groupe des particularités comme les oreilles tombantes ou les couleurs variées. Ces caractéristiques ont surement été sélectionnés involontairement, lors des étapes initiales de la domestication, via la sélection des animaux les plus dociles (Price, 1999). Une théorie expliquant l'ensemble des traits du syndrome de domestication repose sur un déficit des cellules de la crête neurale pendant le développement embryonnaire (Wilkins, Wrangham and Fitch, 2014). Plusieurs études ont permis d'identifier des gènes en lien avec ces caractéristiques (Alberto et al., 2018; Li et al., 2020c).

Le deuxième groupe de caractères sélectionnés par l'homme est celui relié à l'augmentation de la valeur agronomique. Dans ce cas, chaque espèce, et même chaque race, est volontairement sélectionnée en fonction de besoins spécifiques comme le lait chez les vaches Prim'Holstein, ou la laine chez les mérinos. Chez le moutons, les études recherchant les régions génomiques impliqués dans ce type de trait sont très nombreuses, que ce soit pour des phénotypes liés à la production de viande (Cinar et al., 2016; Hickford et al., 2010; Mahrous et al., 2016; Moradian et al., 2013; Yang et al., 2014; Zhang et al., 2013), la déposition de gras dans la queue (Ma et al., 2018; Wang et al., 2014b; Xu et al., 2017), le type et la couleur de la toison (Adelson et al., 2004; Bolormaa et al., 2017; Ebrahimi et al., 2017; Gutiérrez-Gil et al., 2017; Hamadani et al., 2019; Han et al., 2015; Koseniuk and Słota, 2016; Ma et al., 2017; Nazari-Ghadikolaei et al., 2018; Rather et al., 2019; Sulayman et al., 2017), le lait (Cesarani et al., 2019b, 2019b; García-Gámez et al., 2012; Jiang, 2006; Li et al., 2020a; Moioli et al., 2013; Ozmen, Kul and Gok, 2020; Sallam, 2021; Sutera et al., 2019), les cornes (Duijvesteijn et al., 2018; Greyvenstein et al., 2016; He et al., 2016; Kijas et al., 2016; Pan et al., 2018a) ou encore la fertilité (Bodin et al., 2007; Bowles, Carson and Isaac, 2014; Chu et al., 2007; Galloway et al., 2000).

Enfin, le dernier groupe de caractères est lié à l'adaptation des moutons aux différents environnements imposés par les migrations humaines bien au-delà de leur aire native et notamment sous des conditions particulièrement rudes (altitude, zones arides, pluvieuses etc.) (Álvarez et al., 2020b; Edea et al., 2019; Kim et al., 2016; Mastrangelo et al., 2018; Mwacharo et al., 2017; Yang et al., 2016b). Certaines de ces adaptations peuvent provenir d'espèces sauvages ou férales avec lesquelles se croisent les moutons (voir parties 2.1.4 et 2.4.6).

Ainsi, de nombreuses études ont cherché à identifier les gènes de la domestication en utilisant ces techniques. Un premier constat intéressant est qu'un grand nombre de signaux de sélection se trouvent en dehors des gènes (Alberto et al., 2018; Li et al., 2020c). Ce peuvent être des variants entraînés par la sélection d'un variant génique proche, ou encore impliqués dans la régulation de l'expression de gènes plus ou moins distants. De nombreux signaux indiquent cependant des gènes impliqués dans des processus biologiques multiples : olfaction, reproduction, immunité, métabolisme des graisses, système nerveux, production de lait ou de viande, pigmentation, différenciation sexuelle, morphologie (etc.) (Alberto et al., 2018; Fariello et al., 2014; Kijas et al., 2012a; Naval-Sanchez et al., 2018).

### 3 Objectifs et plan de la thèse

---

Finalement, la domestication du mouton est marquée par de nombreux échanges et mélanges. Leur diversité actuelle a été modélisé par différentes vagues de migrations autant spatialement que temporellement, et par de nombreux croisements entre domestiques mais aussi avec des espèces sauvages. Divers processus de sélection ont aussi été à l'œuvre, dans le but de les apprivoiser, d'établir une cohabitation viable, de maximiser la production de

certaines denrées ou même de s'adapter à de nouveaux environnements. L'ensemble de ces processus constitue le cadre de travail de cette thèse.

Nous avons vu que l'étude de la domestication implique plusieurs disciplines complémentaires comme l'archéologie, la paléogénomique et la génomique. Ici nous utiliserons des données de génomes complets contemporaines pour explorer les informations temporelles qu'elles contiennent et qui restent encore sous exploitées.

Le second chapitre est destiné à la description des données utilisées dans ce travail. Le troisième aborde des questions liées à la diffusion et aux processus démographiques neutres expliquant la diversité observée chez les moutons modernes. Enfin, le quatrième se concentre sur la chronologie d'évènements de sélections au cours de la domestication du mouton.



## Chapitre 2

# Présentation et préparation des données de génomes complets

---



# 1 Introduction

---

Dans ce chapitre nous présenterons les données utilisées dans le cadre de cette thèse. Nous exposerons leurs origines, la manière dont nous avons préparé les données de génomes complets pour les analyses et nous étudierons leur structure neutre. L'ensemble des données utilisées se divise en deux groupes : un jeu de données de base, avec lequel nous avons réalisé la majorité des analyses et un jeu de données étendu. Ce dernier a été établi dans un second temps, pour des analyses de détection de sélection intra-groupe présentées dans le chapitre 4, ce qui explique les différences de construction avec le jeu de données de base.

## 2 Composition des jeux de données

---

### 2.1. Jeu de données de base

---

Notre jeu de données de base résulte de la fusion des données suivantes :

- **Jeu de données N** : 218 génomes complets issus de l'échantillonnage de moutons domestiques iraniens et marocains ainsi que de mouflons sauvages dans le cadre du projet Nextgen (PRJEB7436).
- **Jeu de données I** : 57 génomes complets issus d'un panel de races mondiales fourni par l'ISGC<sup>13</sup>
- **Jeu de données P** : 99 génomes complets produits par Pan et al. (2018b) (PRJEB23437), qui se composent de races asiatiques.

Les jeux de données N et I comportent des génomes ayant une couverture de 12 à 14X contre 5X pour ceux du jeu de données P. Le jeu de données de base comporte ainsi 347 moutons domestiques représentatifs de la diversité actuelle mondiale et 55 races (**Tableau 1**).

Il comporte également 27 individus sauvages de différentes espèces (**Tableau 2**). Le Mouflon asiatique (*Ovis gmelini*) est le descendant des ancêtres sauvages des moutons domestiques, il provient d'Iran. L'Urial (*Ovis vignei*) est aussi un mouflon du Moyen-Orient qui constitue un groupe phylogénétique frère du mouflon asiatique. Leur divergence date de 1.26 millions d'année (Rezaei et al., 2010). Le mouflon canadien (*Ovis canadensis*) et le mouflon de Dall (*Ovis dalli*) sont des mouflons d'Amérique du Nord qui ont divergé du mouflon asiatique et de l'Urial il y a plus de deux millions d'années (Rezaei et al., 2010). L'aspect de ces différentes espèces est présenté dans la **Figure 17**.

**Tableau 1** : Origine des génomes complets de moutons domestiques constituant le jeu de données de base

Origine N : données issues du projets Nextgen

Origine P : données issues de Pan et al. 2018

Origine I : Données issues de l'ISGC

La classification des races dans des zones géographique (Continent, région) fait référence au berceau des races. Cette classification n'implique pas que l'on trouve les races seulement dans la zone qui leur est assignée ici.

---

<sup>13</sup> International Sheep Genome Consortium

Continent	Région	Races	Origine	Effectif
Asie	Nord de la Chine	Bayinbuluke	P	10
		Cele Black Sheep	P	9
		Hu	P	10
		Small Tail Han	P	10
		Tan	P	10
		Wuzhumuqin	P	10
	Plateau Tibétain	Oula Sheep	P	10
		Valley Tibetan Sheep	P	10
		Prairie Tibetan Sheep	P	10
	Sud de l'Asie	Autre	N	2
		Garole	I	2
		Sumatra	I	2
		Garut	I	2
		Bangladeshi	I	2
		Changthangi	I	2
Moyen-Orient	-	Afshari	I	2
		Awassi	I	3
		Cine Capari	I	1
		Karakul	I	2
		Karya	I	1
		Norduz	I	2
		Sakiz	I	2
		Autre	N	20
Europe	Nord de l'Europe (Voie Danubienne)	Finnsheep	I	2
		Texel	I	1
		Swiss Mirror	I	1
		Swiss White Alpine	I	4
		Valais Blacknose	I	1
		Cheviot	I	2
		Dollgellau Welsh	I	1
		Moutain	I	1
		Poll Dorset	I	1
		Romney	I	1
		Scottish Blackface	I	1
		Tregaon Welsh mountain	I	1
		Welsh Hardy Speckled Face	I	1
		Lacaune	I	2
	Sud de l'Europe (Voie méditerranéenne)	Ojalada	I	2
Mérinos		I	3	
Mérinos Australien		P	10	
Churra		I	2	
Salz		I	3	
Castellana	I	2		

<b>Afrique</b>	<b>Afrique du Nord</b>	Beni-Guil	N	6
		Boujaad	N	1
		D'man	N	30
		Ouled-Djellal	N	8
		Rahalya	N	4
		Sardi	N	27
		Timahdite	N	16
		Autres	N	69
<b>Amériques</b>	<b>Afrique du Sud</b>	African White Dorper	I	2
		Ethipian Menz	I	1
		Namaqua Afrikaner	I	1
		Ronderib Afrikaner	I	2
		Brazilian Creole	I	1
		Gulf Coast Native	I	2
		Morada Nova	I	2
		Santa Inês	I	2
<b>Total</b>				<b>347</b>

Tableau 2 : Origine des génomes complets de mouflons sauvages constituant le jeu de données de base

<b>Espèce</b>		<b>Effectif</b>
<b>Nom binomial</b>	<b>Nom commun</b>	
<i>Ovis gmelini</i>	Mouflon asiatique	18
<i>Ovis vignei</i>	Urial	4
<i>Ovis canadensis</i>	Mouflon canadien	3
<i>Ovis dalli</i>	Mouflon de Dall	2

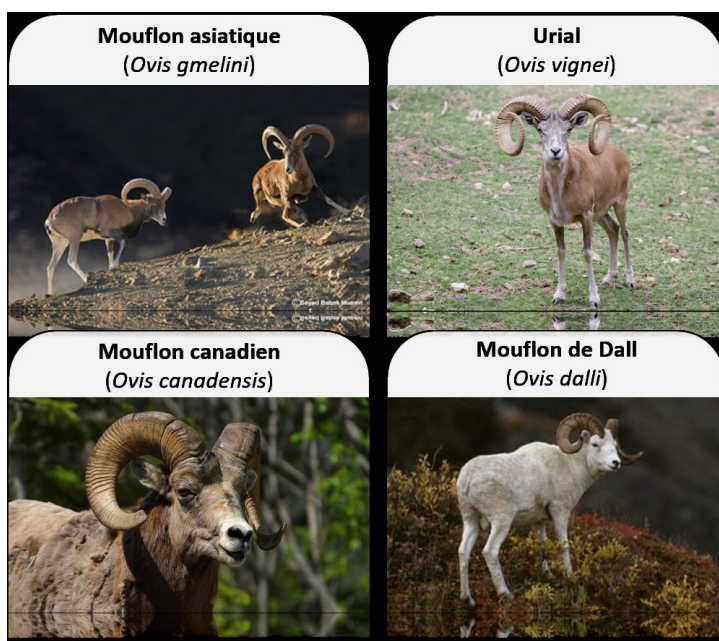


Figure 17 : Aspects des mouflons présents dans le jeu de données de base

La photo des mouflons asiatiques a été prise par Babak MOUSAVI, celle du mouflon canadien par Jon Sullivan, celle de l'Urial par le parc animalier d'Auvergne et celle du mouflon de Dall est libre de droit.

## 2.2. Jeu de données étendu

Pour les analyses de détection de sélection intra-groupe du chapitre 4, nous avons utilisé un jeu de données étendu en ajoutant des individus domestiques issus de 17 races supplémentaires (**Tableau 3**) ou complétant deux races déjà présentes dans le jeu de données de bases (Menz, et Finnsheep) mais en effectif réduit (moins de 3 individus). Cet ajout permet, lorsque cela est possible, d’avoir une meilleure représentativité de la diversité génétique de chaque race. Ces nouvelles données concernent 10 races africaines (BioProject [PRJNA523711](#)) et neuf races issues des autres continents (BioProject [PRJNA624020](#)), soit au total 185 génomes complets.

Tableau 3 : origine des génomes complets ajoutés au jeu de données de bases pour former le jeu de données étendu

Continent	Région	Pays	Races	Code race	BioProject	Effectif	
Afrique	Afrique du Nord	Lybie	Lybian Barbary	BRB	<a href="#">PRJNA523711</a>	10	
			Amhara-Adane	ADN	<a href="#">PRJNA523711</a>	8	
			Amhara-Arabo	ARB	<a href="#">PRJNA523711</a>	9	
	Afrique de l'Est	Ethiopie	Doyogena	DYG	<a href="#">PRJNA523711</a>	9	
			Gumuz Gesses	GSS	<a href="#">PRJNA523711</a>	9	
			Dibata				
			Afar-Kefis-Dulecha	KEF	<a href="#">PRJNA523711</a>	12	
			Loya	LYA	<a href="#">PRJNA523711</a>	9	
			Menz	MEN	<a href="#">PRJNA523711</a>	8	
Asie	Afrique de l'Ouest	Cameroun	Cameroon Sheep	CAM	<a href="#">PRJNA523711</a>	10	
			Nigéria	Dalljonke	DJK	<a href="#">PRJNA523711</a>	14
	Asie du Nord	Chine	Altay-Sheep	ALT	<a href="#">PRJNA624020</a>	10	
			Bashibai	BSB	<a href="#">PRJNA624020</a>	10	
Europe	Europe du Nord	Finlande	Finnsheep	FIN	<a href="#">PRJNA624020</a>	10	
			Suède	Gotland	GOT	<a href="#">PRJNA624020</a>	10
		Royaume-Uni	France	Ouessant	OUE	<a href="#">PRJNA624020</a>	10
			Suffolk Sheep	SFK	<a href="#">PRJNA624020</a>	10	
		Ecosse	Shetland	SHE	<a href="#">PRJNA624020</a>	7	
		France	Solognote	SOL	<a href="#">PRJNA624020</a>	10	
Moyen-Orient		Afghanistan	Waggir Sheep	WGR	<a href="#">PRJNA624020</a>	10	
<b>TOTAL</b>						<b>185</b>	

## 3 Préparation des jeux de données

### Jeu de données de base

#### 3.1.1. Alignement

Toutes les données ont été alignées sur la version 4 du génome de référence d’*Ovis aries* (GCA\_000298735.2) à l’aide du programme BWA (Li and Durbin, 2009). Nous avons utilisé l’outil `FixMateInformation` pour s’assurer que toutes les informations relatives aux paires

de lectures soient bien appariées, puis identifiées les lectures dupliquées avec `MarkDuplicates`.

### 3.1.2. Création des fichiers de variants génomiques

Pour créer les fichiers de variants, nous avons suivi le processus recommandé de GATK (version 4). Il consiste en plusieurs étapes : l'appel des variants par échantillon en mode GVCF, la combinaison des GVCF ainsi générés, le génotypage joint et la filtration des données (**Figure 18**).

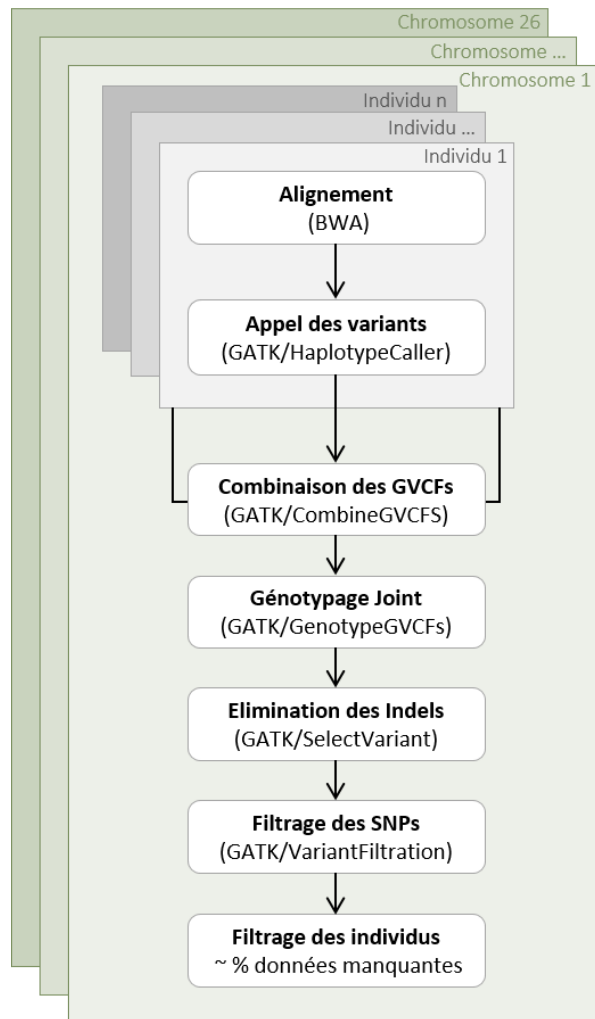


Figure 18 : Etapes nécessaires à la construction des fichiers de variants selon les meilleures pratiques recommandées par GATK.

L'appel des variants a été réalisé grâce à l'outil `HaplotypeCaller` de GATK en mode GVCF pour chaque échantillon et chaque chromosome indépendamment. Les GVCF individuels ainsi générés ont été combinés par chromosome grâce à l'outil `CombineGVCFs` de GATK. Le génotypage joint a ensuite été réalisé grâce à l'outil `GenotypeGVCFs` de GATK pour chaque chromosome. Nous n'avons ensuite conservé que les SNPs et éliminé les indels pour la filtration. La méthode recommandée par GATK est le VQSR qui utilise la modélisation par

« machine learning ». Cependant, notre jeu de données présente des différences de couverture qui pourraient créer des confusions lors de la construction du modèle nécessaire au VQSR, c'est pourquoi nous ne l'avons pas utilisé. L'alternative est une filtration par VariantFiltration de GATK qui utilise des seuils imposés pour différentes annotations (Tableau 4, Figure 19).

Tableau 4 : seuils imposés pour l'étape de filtration des SNPs par VariantFiltration de GATK

Annotations		Seuil
QD	QualByDepth	> 5
FS	FisherStrand	< 50
SOR	StranOddsRatio	< 1.5
MQ	RMSMappingQuality	> 57
MQRankSum	MappingQualityRankSumTest	> -2.5
ReadPosRankSum	ReadPosRankSumTest	> -2.5

Ces annotations reflètent des propriétés qui décrivent les variants et le contexte dans lequel ils se situent. Pour choisir les seuils à appliquer, nous avons suivi les recommandations GATK en nous basant sur les distributions des annotations issues du jeu de données (Figure 20). Ces recommandations s'appuient sur la comparaison des distributions d'annotations de variants filtrés avec la méthode VQSR et la distribution des annotations des variants non filtrés (Figure 19). La distribution des variants qui passent le filtre VQSR donne une idée de la distribution attendue pour des variants fiables. Par exemple, pour l'annotation QD, il est conseillé de filtrer les variants correspondants à la première « épaule » de distribution. Les détails de fixation des seuils de chaque annotation sont présentés dans la documentation de GATK (<https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants>). L'application de ces filtres a éliminé 37% des SNPs.

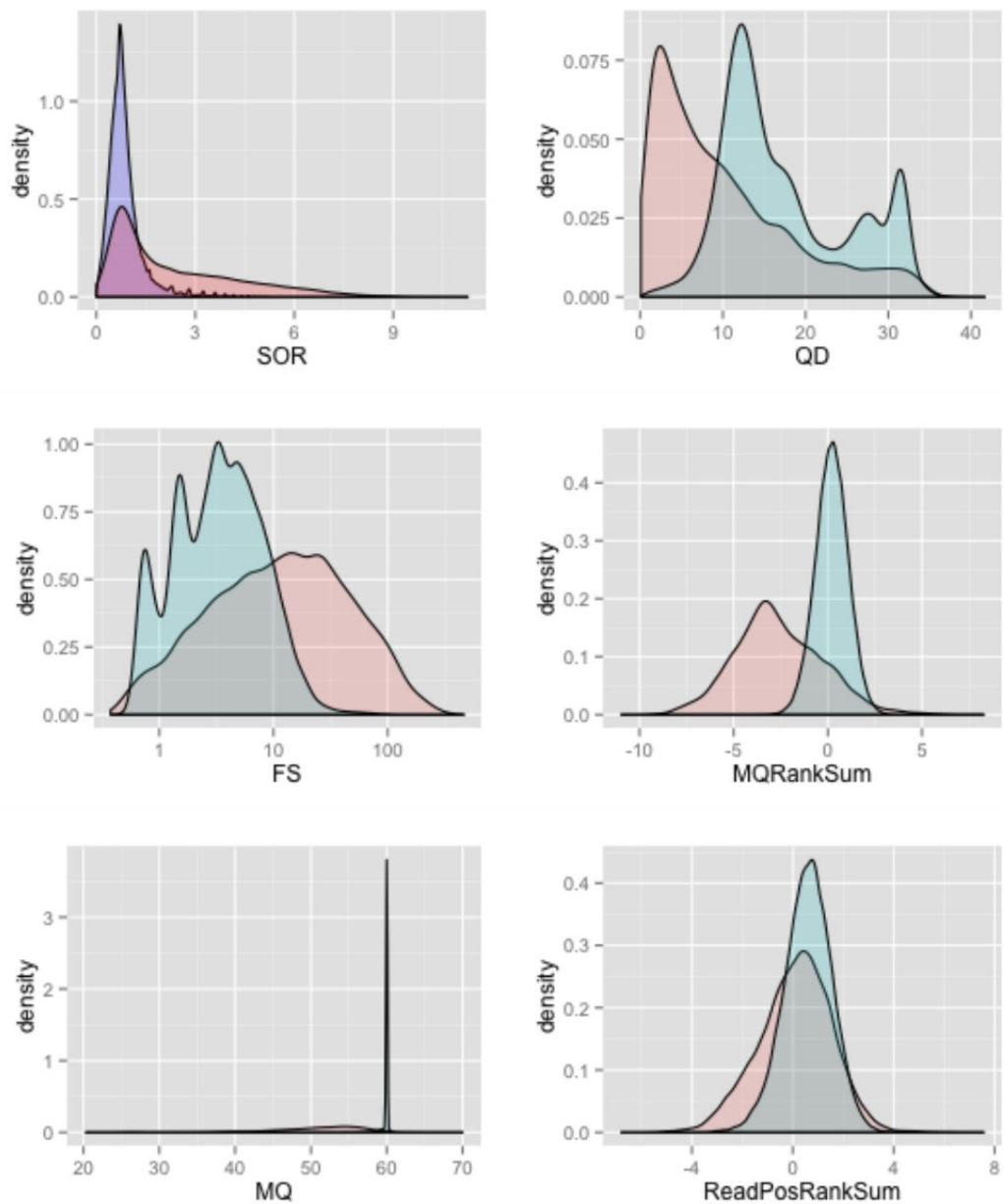
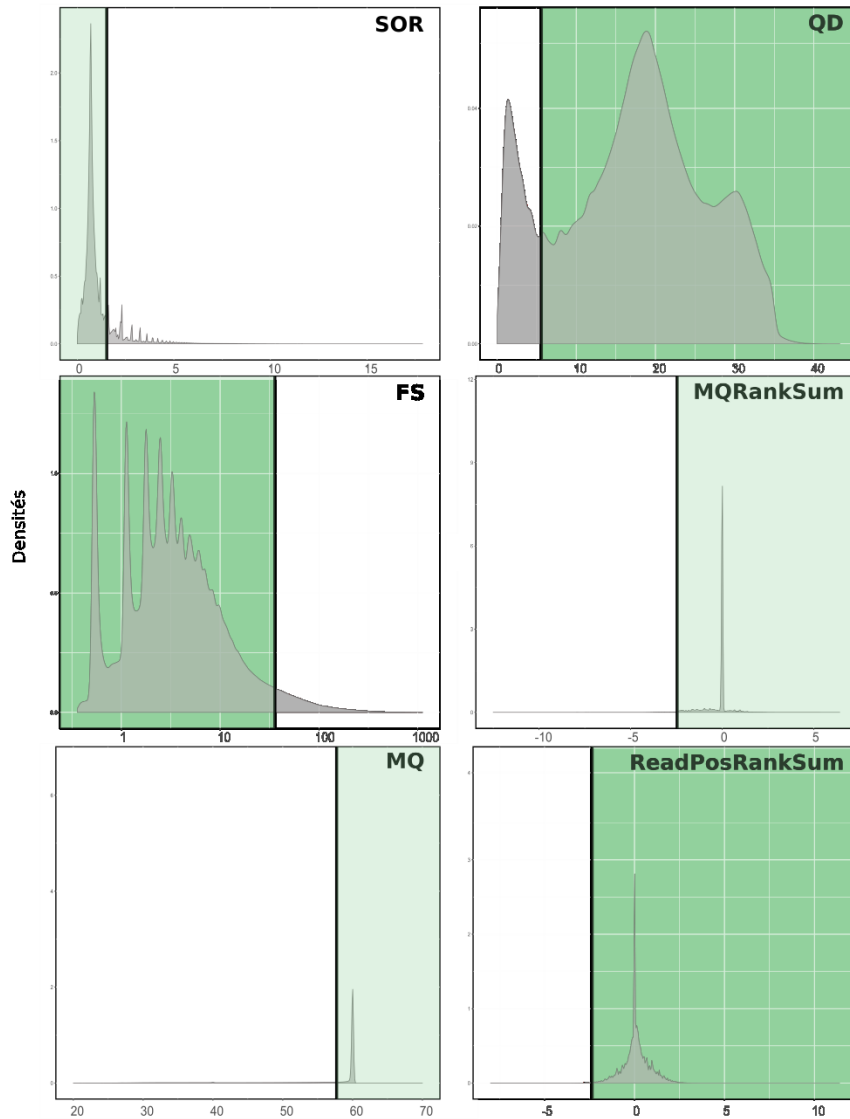


Figure 19 : Comparaison des distributions des annotations de variants avant et après le filtre VQSR utilisée pour fixer les seuils de filtration fixes en cas d'impossibilité d'utilisation du VQSR. Les différents graphiques proviennent de la documentation GATK disponible à l'adresse suivante : <https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants>



**Figure 20 : Distributions des annotations des variants issus du génotypage joint et avant l'étape de filtration GATK.**

Les variants présentant des annotations dans les gammes indiquées en vert ont été conservés. Les distributions présentées correspondent à celle du chromosome 1 uniquement.

Nous avons aussi supprimé deux individus qui présentaient plus de 50 % de données manquantes, un de race Bayinbuluke (CNOA-BY7) et un mouflon asiatique (IROO-C3-2743). Tous les autres individus présentent en moyenne 4% de données manquantes, avec un minimum de 3.6% et un maximum de 6.4%. La filtration des sites en fonction de la proportion de données manquantes est spécifique à chaque analyse et est indiquée ultérieurement dans les descriptions méthodologiques de chacune. Le jeu données de bases se compose finalement de 376 génomes complets, représentés par plus de 64 millions de SNPs (**Tableau 5**) avec une densité de 5 SNPs par kilobase (kb) en moyenne.



Tableau 5 : Nombre de variants par chromosomes avant et après le filtre GATK

chr	Avant filtre			Après filtre	
	SNPs	Indels	Autres	Tous types	SNPs
1	11296257	1695930	179987	12824347	7011589
2	10096704	1505251	154461	11454301	6285670
3	8951012	1367675	142420	10184475	5679786
4	4890402	746870	82583	5563416	3084298
5	4279113	656822	69320	4872442	2739527
6	5042700	772565	85259	5738030	3132550
7	3988596	595351	62740	4526047	2545676
8	3679608	551856	57739	4177726	2342571
9	3955169	604010	68447	4497672	2547374
10	3805249	598184	69221	4337619	2346159
11	2350660	371580	39613	2685962	1612296
12	3185815	485205	51291	3623326	2075265
13	3204801	478165	50229	3636242	2108061
14	2474480	391113	40132	2826433	1617903
15	3395694	519840	59633	3863005	2155672
16	3035823	453095	51700	3442518	1979169
17	2968045	468088	51613	3389439	1934889
18	2811267	429766	46797	3198562	1823675
19	2358438	355837	38874	2679635	1592022
20	2140079	331041	36092	2436629	1396164
21	2166096	325365	34803	2458457	1393990
22	2125280	316810	34745	2409823	1388136
23	2617767	390291	44301	2968089	1731037
24	1657588	263494	28505	1895490	1163222
25	1946485	283894	31101	2200827	1260574
26	1885628	282906	32791	2139264	1262827
<b>total</b>	<b>100308756</b>	<b>15241004</b>	<b>1644397</b>	<b>114029776</b>	<b>64210102</b>

### 3.1.3. Structure génétique neutre

Nous avons réalisé une ACP sur ce jeu de données grâce à l'outil `-pca` du programme PLINK (version 1.9) (Purcell et al., 2007). Les SNPs ont été filtrés pour ne garder que ceux en équilibre de liaison entre eux grâce à la fonction `-indep-pairwise` de PLINK avec comme paramètre une taille de fenêtre de 50kb, un nombre de variants conservé par fenêtre de 10 et un coefficient de corrélation seuil de 0.1.

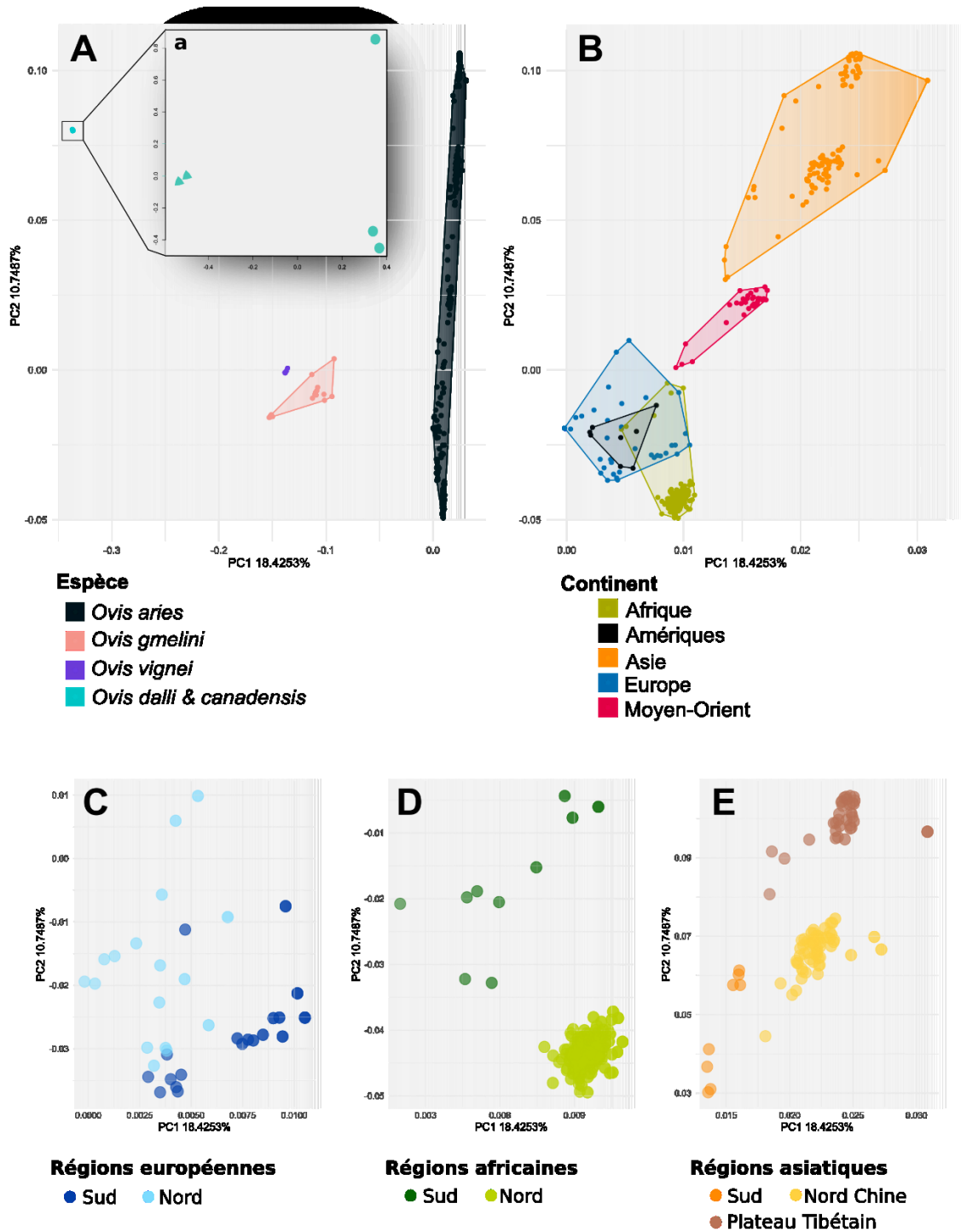


Figure 21 : Distribution des individus sur le premier plan de l'ACP réalisée sur le jeu de données de base.

A : Ensemble du jeu de données

a : Focus sur les mouflons américains (les rond correspondent à *O. canadensis* et les triangles à *O. dalli*)

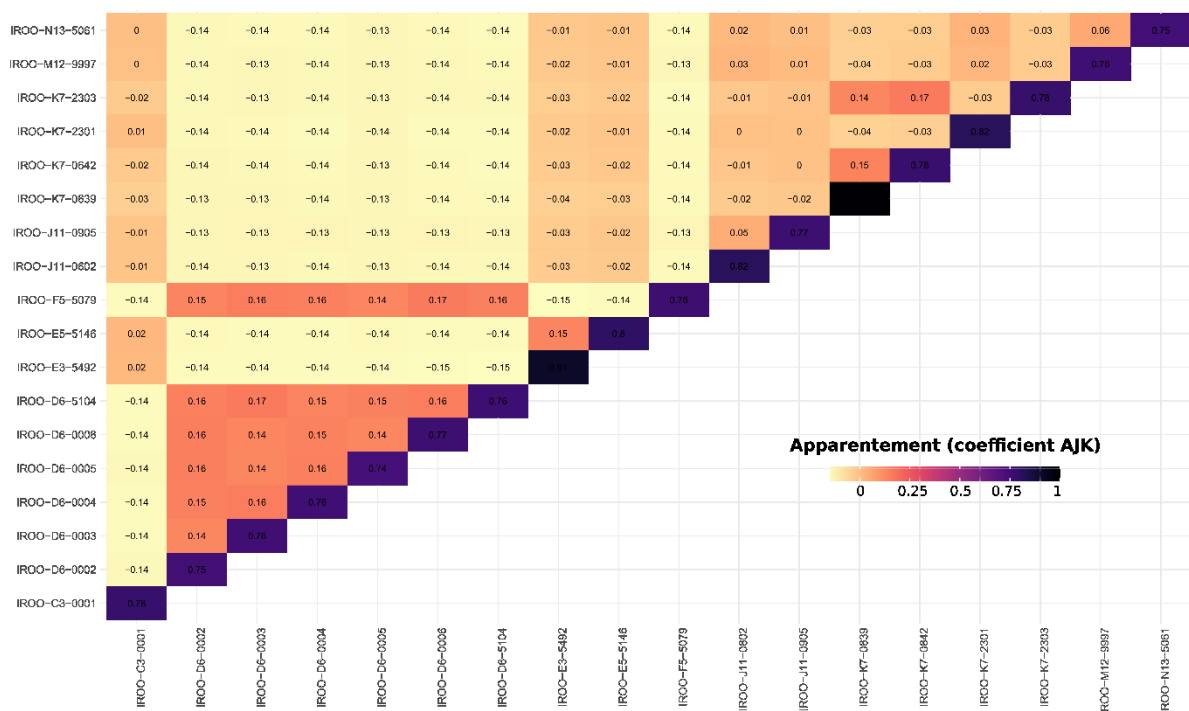
B : Focus sur le groupe de moutons domestiques

C : Focus sur le groupe de moutons domestiques européens

D : Focus sur le groupe de moutons domestiques africains

E : Focus sur le groupe de moutons domestiques asiatiques

L'ACP montre que les différentes espèces du jeu de données sont bien séparées génétiquement, à l'exception des mouflons américains (**Figure 21.A**). On discerne quatre groupes correspondant aux espèces *O. aries*, *O. gmelini*, *O. vignei* et un dernier groupe avec les mouflons américains *O. dalli* et *O. canadensis*. Le faible effectif de ces deux espèces contre le fort effectif de moutons et de mouflons asiatiques explique leur proximité sur une ACP réalisée sur l'ensemble des données et dont le but est d'en maximiser la variance. Si l'on réalise la même analyse seulement sur les mouflons américains, les deux espèces sont bien séparées génétiquement dès le premier plan de l'ACP (**Figure 21.a**). On remarque que six individus du groupe *O. gmelini* sont excentrés par rapport au reste du groupe. Ils correspondent à des individus échantillonnés sur une île et présentant un apparentement supérieur à 0.14 (coefficient AJK (Yang et al., 2010) calculé avec *vcftools* (Danecek et al., 2011)) (**Figure 22**). Nous sous-échantillonnerons un seul individu de ce groupe aléatoirement pour les analyses des chapitres 3 et 4.



**Figure 22 : Apparentement entre les Mouflons Asiatiques du jeu de données de base**

L'apparentement est représenté par le coefficient AJK calculé sur l'ensemble des données filtrées avec une fréquence allélique minimum de 0.02 et 70% de données manquante par site maximum. Un coefficient négatif montre une absence de relation et peut même indiquer de la structure entre les deux individus concernés par le coefficient négatif.

Au niveau des moutons, on observe une forte structuration géographique avec séparation entre l'Asie et les continents Occidentaux (Afrique, Europe, Amérique) (**Figure 21.B**). Les individus du Moyen-Orient se situent au milieu. Les groupes africains et européens sont proches, ce qui peut s'expliquer par les nombreux échanges qui ont eu lieu entre les deux continents (**Voir Chapitre 1, partie 0**). La distribution des individus américains est incluse dans celle des européens, ce qui est attendu puisque leur origine est européenne. Au niveau des groupes continentaux aussi on retrouve une structure marquée distinguant de grandes régions. En Europe (**Figure 21.C**) et en Afrique (**Figure 21.D**) on observe une séparation des

rares issues du Nord et du Sud. En Asie (**Figure 21.E**), on distingue les trois groupes correspondant au Sud de l'Asie, aux races du Nord de la Chine et à celles du Plateau Tibétain. Cela confirme une structuration déjà démontrée par (Kijas et al., 2012b; Peter et al., 2007). On note la présence de deux individus originaires du Sud de l'Europe dans le groupe issu du Nord. Ce sont des moutons Salz qui est une race composite formée à partir de l'Aragonese et la Romanov qui fait partie des races à queue courte d'Europe du Nord (Rasali, Shrestha and Crow, 2006).

L'apparementement entre individus domestiques (calculé selon la méthode de (Yang et al., 2010) grâce à `vcftools` sur l'ensemble des données filtrées avec une fréquence allélique minimum de 0.02 et 70% de données manquante par site maximum) se situe entre 0 et 0.37 mais ne dépasse pas 0.25 sauf pour deux paires d'individus de races identiques (Cheviot et Salz).

Nous avons aussi réalisé une analyse `ADMIXTURE` (Alexander and Lange, 2011) pour compléter l'étude de la structure neutre de notre jeu de données de base (**Figure 23**). En prenant six groupes ancestraux, nous mettons en évidence des groupes structurés et homogènes. Les individus nord-africains ont une ascendance unique sauf pour six individus qui présentent moins de 17% d'ascendance européenne. Pour le groupe de races européennes (auxquelles nous avons rajouté les races américaines), presque tous les individus présentent une ascendance unique européenne sauf 14 sur 49 qui présentent moins de 25% d'ascendance extérieure (africaine, asiatique ou moyen-orientale). Les races du Plateau Tibétains ont une ascendance unique. Les races nord Chinoises présentent une double ascendance : 75% asiatique et 24% moyen-orientale environ. Les races sud-asiatiques présente une ascendance Moyen-Orientale. Les races sud-africaines se divisent en deux. Les African White Dorper présentent 67% d'ascendance européenne et 33% d'ascendance moyen-orientale. Cette race résulte du croisement d'une race anglaise (la Dorper) et d'une race originaire de Somalie à queue grasse (la Persan à tête noire qui descend du mouton de Somalie) et donc probablement liée aux races de Moyen-Orient où ce phénotype est très présent (Dohner, 2001; Mason, 1988). Les Ronderib Afrikaner, Namaqua Afrikaner et les Menz ont une ascendance majoritaire moyen-oriental. Les deux premières sont originaires de l'Est de Asie (Epstein, 1960). Enfin les mouflons américains (*O. dalli* et *O. canadensis*) présentent une ascendance unique tout comme les mouflons d'Iran (*O. vignei* et *O. gmelini*) à l'exception de huit individus qui présentent moins de 7% d'ascendance moyen-orientale ou européenne.

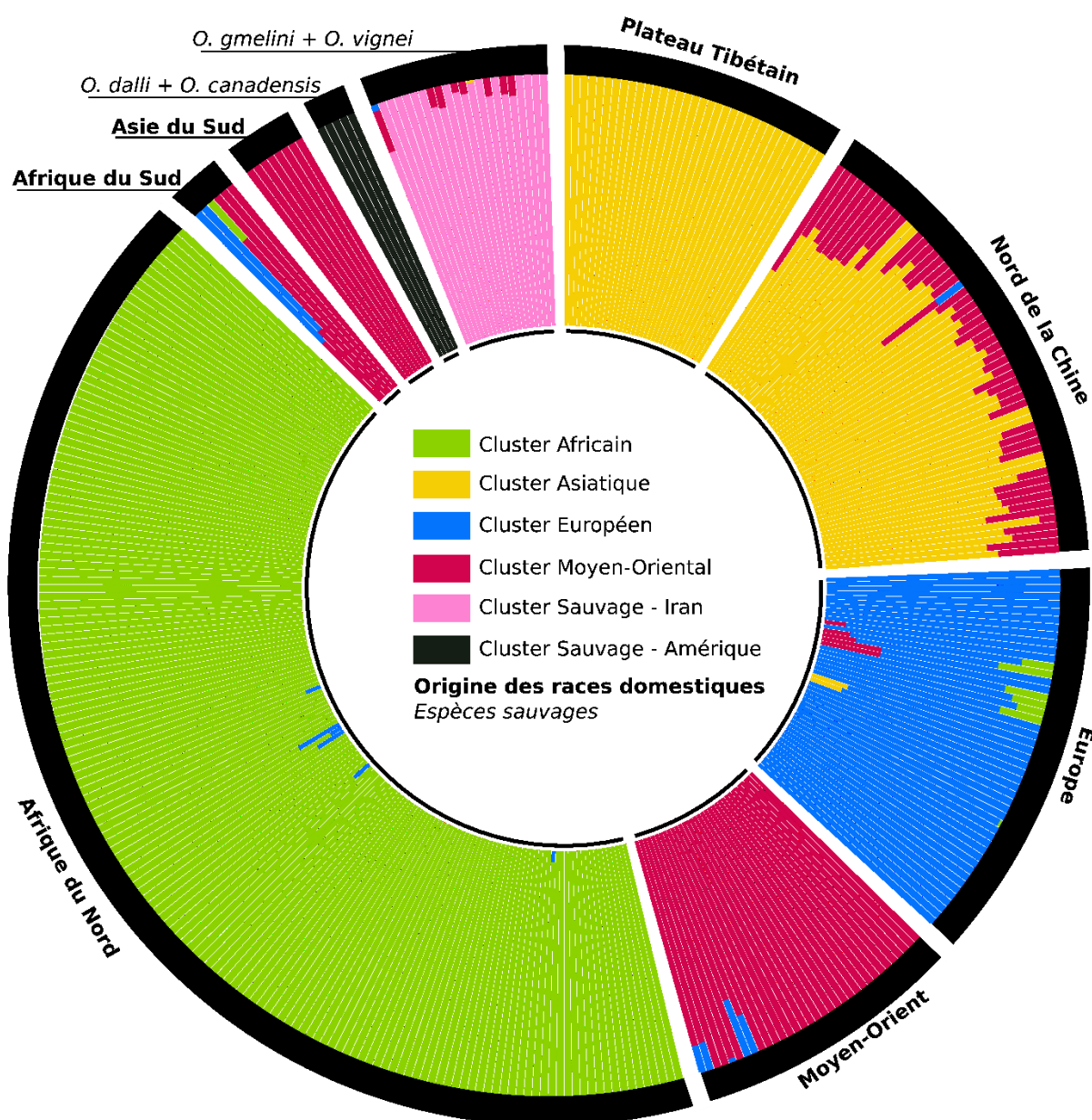


Figure 23 : Pourcentages d'ascendance inférés pour chaque individu du jeu de données de base grâce au programme ADMIXTURE pour un nombre de populations ancestrales égal à 6

### 3.1.4. Phasage

A l'issue des étapes de filtration et de l'étude de la structure génétique des données, nous avons phasé les variants et imputé les données manquantes grâce au programme BEAGLE (version 5.0) (Browning and Browning, 2007). Nous avons effectué le phasage séparément par groupe d'espèces ou de races cohérents avec la structure génétique révélée par l'ACP et l'analyse ADMIXTURE. Le but étant d'avoir la plus grande homogénéité génétique possible au sein du groupe utilisé pour faire le phasage. Les groupes constitués pour le phasage sont listés dans le **Tableau 6** et représentés sur les six premiers axes de l'ACP dans la **Figure 24** pour les individus domestiques. Nous avons inclus les moutons américains dans le groupe Européen étant donné leur similarité génétique. Quand les effectifs le permettaient (effectif supérieur à

6), nous avons constitué des groupes de races. Cela a été fait dans les régions Chine du Nord et Plateau tibétain.

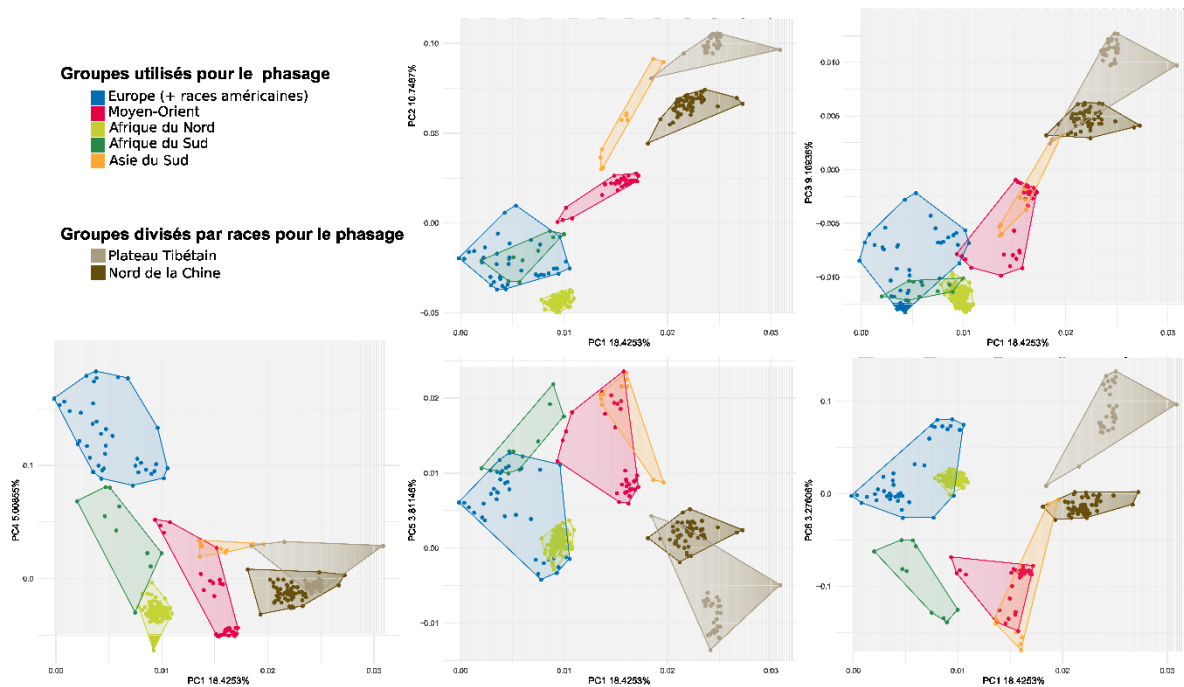


Figure 24 : Structure génétique des groupes utilisés pour le phasage des individus domestiques  
Le 5 panneaux correspondent à différents plans de l'ACP sur données génétiques.

Tableau 6 : Groupes utilisés pour phaser les variants du jeu de données de bases

Espèce	Groupe de races	Races
Mouflons américains	Moyen-Orient	Bayinbuluke
Urial	Europe	Cele black Sheep
Mouflons asiatique	Afrique du nord	Hu
	Afrique du Sud	Oula sheep
	Asie du Sud	Prairie Tibetan
		Valley Tibetan
		Wuzhumuqin

### 3.1.5. Etat ancestraux et dérivés

L'inférence des états ancestraux et dérivés des sites a été faite grâce à la méthode *EST-SFS* (Keightley and Jackson, 2018) sur les données phasées ne contenant que les sites bi-alléliques et avec 25% de données manquantes au maximum. Nous avons fait des inférences séparées pour les sites CpG et non CpG qui ont des taux d'évolution très différents (Keightley and Jackson, 2018).

Cette méthode infère les états ancestraux et dérivés grâce à plusieurs groupes externes. Dans notre cas, comme le mouton est la cible de nos analyses et nous avons utilisé le Mouflon Asiatique et l'Urial en tant que premier et second groupe externe respectivement (**Figure 25**) en nous basant sur la phylogénie du genre *Ovis* (Rezaei et al., 2010). A l'issue de la

procédure EST-SFS, nous avons fixé l'état ancestral ou dérivé en choisissant celui qui présentait la meilleure vraisemblance pour chaque site.

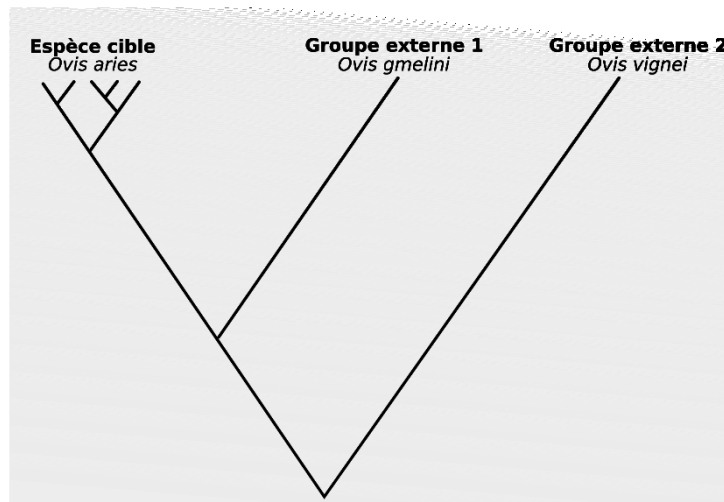


Figure 25 : Représentation des relations entre le groupe cible (moutons) et les groupes externes nécessaire à l'inférence des état ancestraux et dérivés des variants.

Nous avons aussi montré que le choix des groupes externes n'a que peu d'impact sur les analyses ultérieures. Nous avons utilisé la même méthode d'inférence des état ancestraux et dérivés mais en prenant les mouflons américains comme groupes externes. Nous avons ensuite estimé l'âge d'apparition de tous les variants du chromosome 26 avec ces états et les avons comparés avec ceux inférés à partir des variants dont les état ancestraux et dérivés ont été établis avec l'Urial et le Mouflon Asiatique en tant que groupe externe. La procédure de datation est expliquée dans le chapitre 3.

Les âges d'apparitions inférés à partir de ces deux différentes approches sont fortement corrélés ( $p$ -value <  $2.2e-16$ , test de Pearson) exceptés pour les variants présents

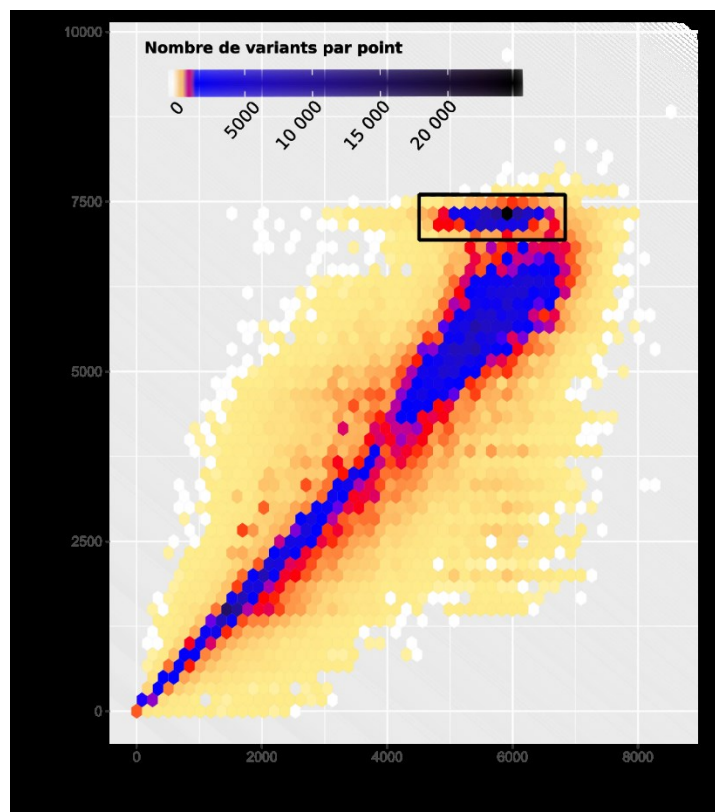


Figure 26 : Comparaison des âges d'apparition des variants après inférence des états ancestraux et dérivés avec différents groupes externes : AM (mouflons américains : *O. dalli* et *O. canadensis*) ; IR (mouflons iraniens : *O. gmelini* et *O. vignei*).

Les variants encadrés correspondent à des variants présents uniquement chez les mouflons américains

uniquement dans les espèces de mouflons américains (encadrés dans la **Figure 26**) et qui ne nous seront donc pas utiles dans les analyses.

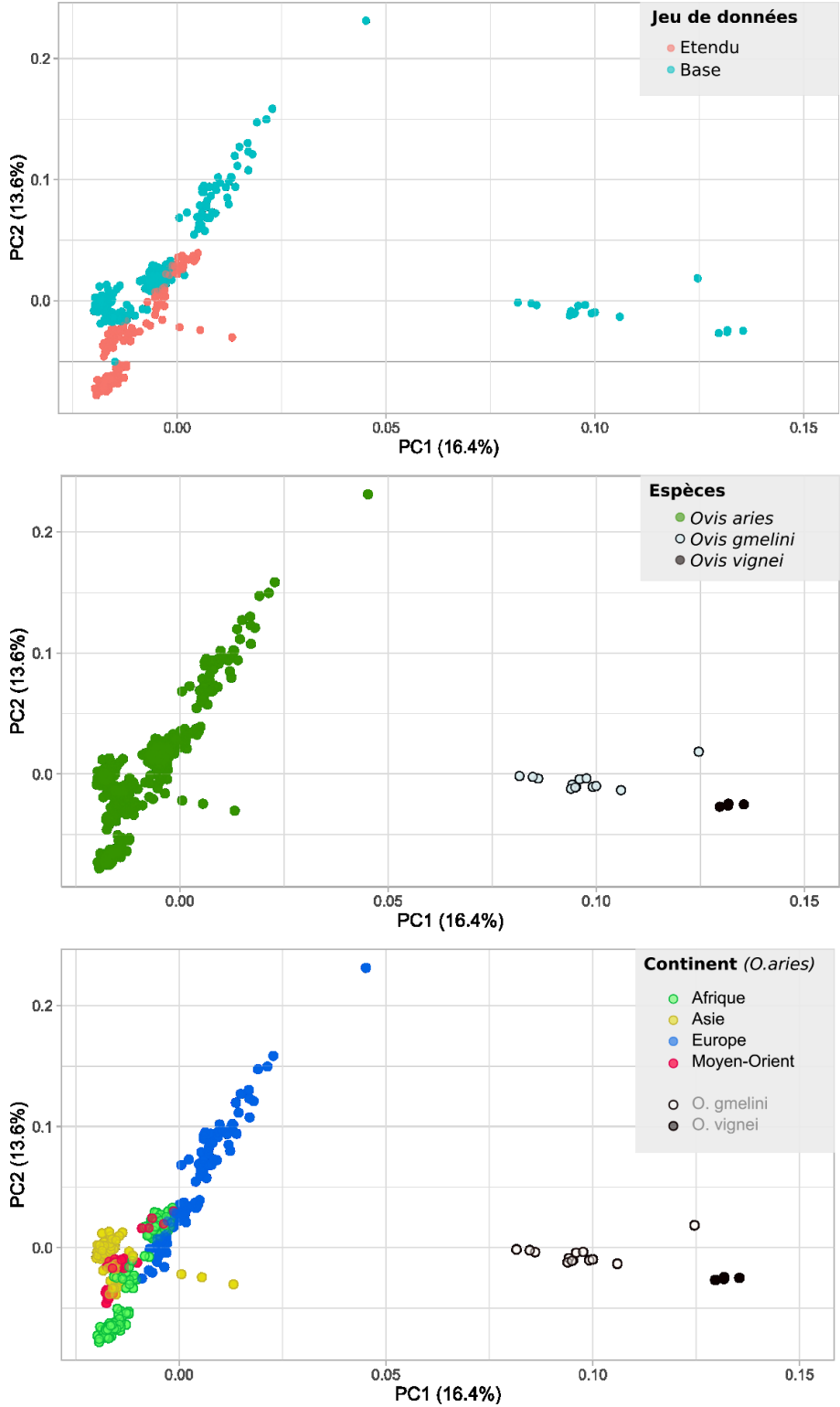


Figure 27 : Distributions des individus du jeu de données étendu sur le premier plan d’une ACP réalisées sur le jeu de données de base.



## Jeu de données étendu

---

### 3.1.6. Construction

Les 185 génomes supplémentaires ont été alignés de la même manière que ceux du jeu de données de base (voir **partie 0**). Nous avons ensuite appelé les variants identifiés dans le jeu de données de base (sites bi-alléliques avec 25% de données manquantes maximum) grâce au programme `mpileup` de `SAMTOOLS` (Li, 2011). Le phasage et l'imputation ont été réalisés avec `BEAGLE` (version 5.0) (Browning and Browning, 2007) indépendamment pour chaque race. Les états ancestraux et dérivés ont été établis à partir de ceux inférés sur le jeu de données de base (voir **partie 3.1.5**).

Nous avons observé un taux de données manquantes de 40 à 54% pour les génomes du jeu de données étendu par rapport aux génomes du jeu de données de bases. Cela pourrait provenir de la qualité des données ajoutées (informations non disponibles), ou des différences des algorithmes d'appel des variants de `GATK` et `SAMTOOLS`. Ils diffèrent à plusieurs niveaux comme le prétraitement des alignements, le modèle d'appel des variants, ou les filtres. Une différence majeure réside dans le traitement de erreurs. `GATK` suppose que les erreurs de séquençage sont indépendantes alors que `mpileup` de `SAMTOOLS` suppose qu'une seconde erreur est plus probable. L'effet de cette différence peut être important sur des données avec des taux de couvertures très importants et/ou hétérogènes (communications personnelles : Pamela Bretscher). Dans notre cas, le jeu données de base présentent des taux de couvertures de 5 à 14X alors qu'ils sont en moyenne de 27.5 à 54X pour le jeu de données étendu.

La meilleure façon de combiner ces jeux de données aurait été d'utiliser la même méthode d'appel des variants pour tous les génomes. Ici nous privilégierions le *caller* de `GATK` mais les durées de calculs nécessaires ne nous ont pas permis de le faire dans le cadre de cette thèse. Aussi ce jeu de données étendu n'a été utilisé que dans des méthodes de détection de sélection intra-groupe pour le chapitre 4 et jamais de manière combinée avec le jeu de données de base excepté pour vérifier leur structure génétique à l'aide d'une ACP sur données génétiques.

### 3.1.7. Structure

Nous avons analysé la structure génétique du jeu de données étendu en effectuant une ACP grâce à l'outil `-pca` du programme `PLINK` (version 1.9) (Purcell et al., 2007) sur les SNPs sans donnée manquante et en équilibre de liaison les uns avec les autres grâce à la fonction `-indep-pairwise` de `PLINK` avec comme paramètres une taille de fenêtre de 50kb, un nombre de variants à garder par fenêtre de 10 et un coefficient de corrélation seuil de 0.1.

Sur le premier plan de l'ACP (**Figure 27**), nous observons une structure similaire à celle du jeu de données de base seul. Les différentes espèces forment des groupes distincts. Au niveau des domestiques, on retrouve les races du Moyen-Orient en position intermédiaire entre celles d'Afrique et d'Europe. On remarque une plus grande diversité des races

européennes avec les 57 individus supplémentaires issus de ce continent en plus des 38 du jeu de données de base. Les individus africains se séparent en trois groupes : Le premier entre -0.05 et -0.1 de l'axe 2 est constitué par les races sub-sahariennes (Ethiopie, Cameroun, Niger). Le deuxième entre -0.05 et 0 de l'axe 2 est constitué par des Races sud-Africaines, la Namaqua Afrikaner et la Ronderib Afrikaner, deux races importées du Moyen-Orient il y a environ 2500 ans (Snyman, 2014). Cela explique leur proximité avec le groupe Moyen-Oriental sur le premier plan de l'ACP. Le troisième entre 0 et 0.05 de l'axe 2 est constitué par les races nord-africaines, essentiellement Marocaines.

## Chapitre 3

# Datation des évènements de mutations dans le génomome du mouton

---

# 1 Introduction

---

L'étude de la domestication des animaux passe par deux disciplines complémentaires : l'archéologie et la génétique. La seconde se divise également en deux domaines, celui de l'ADN moderne et celui de l'ADN ancien. Le premier a permis beaucoup d'avancées dans le domaine par exemple en localisant les centres de domestication ou en identifiant des loci impliqués dans différents traits liés à la domestication (Pang et al., 2009; Shannon et al., 2015; Wang et al., 2016; Wright, 2015). A priori, l'étude de génomes actuels ne donne qu'un aperçu contemporain d'une histoire bien plus longue et complexe alors que les génomes anciens sont une porte directe sur le passé. Avec les avancées technologiques de ces dernières décennies, de plus en plus de génomes anciens sont disponibles, notamment pour le cheval, le cochon, la vache, le chien et la chèvre (Frantz et al., 2020). Pour le mouton ce n'est pas encore le cas, même si cela ne saurait tarder. Malgré tout, l'échantillonnage de génomes anciens est limité à quelques points du temps et de l'espace. Les génomes modernes représentent, eux, seulement le présent mais leur échantillonnage peut être bien plus exhaustif. Ainsi, c'est la complémentarité des analyses d'ADN ancien et moderne qui permet au mieux de comprendre les processus à l'œuvre durant la domestication.

En effet, les données modernes permettent aussi d'étudier les processus ayant modelé la diversité actuelle de nos populations ovines. Il est par exemple possible de reconstruire les variations démographiques du passé grâce à des génomes modernes (Li and Durbin, 2011; Schiffels and Durbin, 2014). De récentes techniques permettent aussi de retracer l'histoire des mutations qui composent actuellement les populations. Montgomery Slatkin et Bruce Rannala écrivaient en 2000 à propos de la datation de variants génomiques : « Les généticiens ont (presque) créé une machine à remonter le temps. » (Slatkin and Rannala, 2000). Selon eux, l'intérêt pour la datation de variants est dû à la curiosité, mais aussi au désir d'utiliser davantage des données recueillies à d'autres fins. C'est donc cette « machine à remonter le temps » que nous utiliserons ici pour répondre à trois objectifs.

- **Caractériser et explorer la distribution des âges des variants actuellement présents dans le génome du mouton.**
- **Utiliser ces âges pour apporter des informations sur l'histoire démographique des populations ovines depuis le centre de domestication.**
- **Créer une nouvelle base de données disponible pour la communauté scientifique.**

Les âges des millions de variants que nous allons estimer auront un intérêt qui pourra dépasser le cadre de ce travail et permettre d'approfondir les études d'autres équipes. D'autre part, la comparaison à venir de nos résultats avec ceux apportés par la génomique ancienne permettra de valider ce type d'approche pour l'appliquer à des espèces dont nous n'avons pas de données anciennes. De plus, l'application de méthodes de datation sur des espèces domestiques pour lesquelles nous avons des repères chronologiques sûrs permettra aussi de

valider les méthodes pour les appliquer éventuellement à d'autres espèces pour lesquelles nous en savons moins.

## 2 Choix et description de la méthode de datation

---

### 2.1. Choix de la méthode

---

L'âge d'un variant peut-être estimé via sa fréquence ou grâce à la variation des sites adjacents.

L'estimation de l'âge d'une mutation par sa fréquence est assez intuitive. Très simplement, plus une mutation est fréquente dans une population plus elle a eu de temps pour se propager depuis son apparition. Ainsi, la fréquence et l'âge sont positivement corrélés. C'est en somme la théorie que posent Kimura et Ohta en 1973 (Kimura and Ohta, 1973). Ils intègrent aussi dans leur calcul un paramètre important, la taille efficace en le supposant constant. Les populations réelles, qu'elles soient humaines ou ovines répondent rarement à cette hypothèse puisqu'elles ont subi des variations démographiques parfois importantes. Or, ces variations agissent comme des contractions ou des expansions du temps, puisque la dérive génétique est plus forte dans les petites populations que dans les grandes.

La dérive n'est pas la seule force à agir sur la fréquence d'une mutation. En réalité, les trois autres grandes forces l'impactent également. La sélection peut accélérer la propagation d'une mutation bénéfique ou au contraire ralentir celle d'une mutation délétère. Ainsi l'âge calculé sera sur ou sous-estimé. L'apparition d'autres variants par mutation ou flux de gènes peut aussi modifier la fréquence d'un allèle. Finalement, l'estimation de l'âge d'un variant à partir de sa fréquence n'est pas impossible mais nécessite d'établir un modèle d'évolution précis et juste, et de bien évaluer l'incertitude. Les théories sur lesquelles reposent ce type d'estimation sont anciennes et plutôt ignorées dans les discussions actuelles sur le sujet (Griffiths and Tavaré, 1998; Kimura and Ohta, 1973; Li, 1975; Slatkin and Rannala, 2000; Watterson, 1976) ;

Une autre manière d'estimer l'âge d'un variant est d'utiliser la variation des sites adjacents. Cette approche est basée sur le déclin du déséquilibre de liaison (DL) au cours du temps. Le DL décrit l'association non aléatoire de différents variants qui peuvent être liés physiquement ou non. Dans le cas qui nous occupe ici, c'est la liaison des sites proches qui nous intéresse. Autrement dit, des segments de chromosomes qui présentent des variants que l'on dit liés et dont la combinaison est préférentiellement transmise à la génération suivante. Ainsi, lorsqu'un variant apparaît, il est transmis aux générations suivantes avec le segment auquel il appartient. On parlera de segment ancestral. Ce segment est brisé au fil des générations par des événements de recombinaisons et de mutations. De ce fait, plus le segment ancestral est long plus la mutation est récente et plus le segment est court plus elle est ancienne. L'estimation de l'âge d'un variant grâce au déclin du DL repose donc sur la

connaissance des taux de recombinaison et de mutation. Ces taux sont très faibles et donc difficiles à estimer. De plus, ils sont intrinsèquement imprévisibles au sein de chaque lignée en particulier (Slatkin and Rannala, 2000). La précision d'une telle estimation dépend aussi de la généalogie puisque le nombre de lignées qui portent le variant à chaque génération représente le nombre de fois où la recombinaison a pu briser la combinaison du segment ancestral (Slatkin and Rannala, 2000).

Malgré ces imprécisions nous privilégions des méthodes basées sur le déclin de DL plutôt que sur la fréquence allélique. La première reflète ce qui est réellement arrivé à un variant alors que la seconde reflète ce qu'un généticien pense être arrivé à ce variant (Slatkin and Rannala, 2000). Dans un second temps, nous devons choisir une méthode qui répond à plusieurs critères en lien avec notre objectif qui est de dater tous les variants du génome. Ainsi elle ne doit pas être impactée par la sélection puisque nous voulons dater l'ensemble des variants du génome indépendamment de leur statut bénéfique, délétère ou neutre. Elle ne doit pas être limitée à une gamme de fréquence. En raison du type de données utilisé, la méthode doit être suffisamment rapide et peu gourmande en ressources computationnelles pour traiter des millions de variants ainsi que l'information contenue dans plusieurs centaines de génomes complets. De plus, elle doit être assez robuste aux erreurs caractéristiques de ce type de données, que ce soient des erreurs de séquençage ou de phasage. Enfin, ce doit être une méthode applicable à toutes les gammes de temps avec une haute résolution, y compris les plus récents car la domestication est un événement jeune à l'échelle des temps évolutifs.

La plupart des méthodes permettant la datation de variants à partir de génomes complets sont basées sur de la modélisation coalescente et infèrent en réalité le temps à l'ancêtre commun le plus récent des haplotypes portant le variant (que nous noterons TMRCA dans la suite en référence à son abréviation anglophone). Le TMRCA des porteurs du variant est forcément plus récent que l'âge du variant. La méthode PSMC (Li and Durbin, 2011) permet d'inférer le TMRCA des variants tout au long du génome et son échelle de temps est discrète. La méthode de Platt et al. (2019) présente plusieurs avantages. Elle ne dépend pas de l'histoire démographique ce qui permet de comparer les âges de variants dans plusieurs populations qui n'ont pas forcément la même histoire ou une histoire inconnue. Elle peut aussi dater des variants très rares et supporte le traitement de gros jeu de données. En revanche elle utilise seulement une fraction de ces données. D'autres méthodes telles que MSMC (Schiffels and Durbin, 2014), SMC++ (Terhorst, Kamm and Song, 2017) et ASMC (Palamara et al., 2018) nécessitent des approximations à cause de la complexité computationnelle et ne sont pas applicables à la datation de millions de variants. Elles ne sont d'ailleurs pas destinées à cette utilisation mais à l'inférence de variations démographiques. D'autres méthodes sont spécifiquement pensées pour la datation de variants. Parmi elles, on trouve des approches basées sur des chaînes de Markov cachées qui permettent d'inférer la longueur du segment ancestral (Chen and Slatkin, 2013; Smith et al., 2018) ou sur des ABC (Approximate Bayesian Computation) (Nakagome, Hudson and Di Rienzo, 2019; Ormond et al., 2016). Cependant, toutes ces méthodes ciblent plutôt la datation d'événements de sélection et non d'apparition

d'un variant et nécessitent des hypothèses fortes sur la généalogie. Les approches ABC ont aussi l'inconvénient de ne pas utiliser l'ensemble de l'information contenue dans les données puisqu'elles reposent sur des statistiques résumées.

Notre choix s'est arrêté sur la méthode GEVA (Albers and McVean, 2020). Cette méthode ne requiert aucune hypothèse sur les processus sélectifs qui pourraient agir sur une partie des variants à dater, ni sur leur généalogie sous-jacente. Elle est applicable à tout type de variants même à fréquence faible et se veut robuste aux erreurs typiques des génomes complets. Elle date bien l'apparition des variants et non leur sélection. Enfin, elle est assez rapide et demande peu de mémoire ce qui rend son utilisation envisageable pour la datation de millions de variants. Elle a d'ailleurs été développée spécifiquement pour ce but et a permis la création d'un atlas des âges des mutations du génome humain (<https://human.genome.dating/>). Son principe est expliqué dans la section suivante.

RELATE (Speidel et al., 2019) est une autre méthode répondant à nos critères. Elle permet d'inférer la généalogie au niveau de chaque site variable. Ainsi, l'âge d'un variant est compris entre les extrémités de la branche sur laquelle il a été placé. L'état de développement du logiciel au moment de ce travail n'a pas permis de l'explorer pleinement. Il serait cependant intéressant de comparer les estimations obtenues avec GEVA à celle de RELATE. Les tests menés au cours de la thèse permettent de dire qu'il nécessite des temps d'analyse et un nombre d'étapes bien plus importants pour la datation que GEVA.

## 2.2. Description de la méthode

---

La méthode GEVA permet de dater l'apparition d'un variant. Elle repose sur deux principes : le déclin du déséquilibre de liaison au cours du temps (voir **partie 2.1**) et la généalogie de l'échantillon. L'estimation est basée sur la comparaison de plusieurs paires d'haplotypes. Le segment d'ADN partagé par les deux haplotypes de la paire permet d'estimer le TMRCA de cette paire. L'estimation repose soit sur la longueur du segment partagé (horloge de recombinaison), soit sur le nombre de mutation contenues dans le segment (horloge de mutation) soit sur les deux critères (horloge jointe).

Deux types de paires sont ainsi datées : des paires concordantes et des paires discordantes. Les paires concordantes correspondent à deux haplotypes qui portent la mutation à dater. Les discordantes, elles, se composent d'un haplotype portant la mutation à dater et d'un autre qui ne la porte pas. L'hypothèse est que le TMRCA d'une paire concordante est forcément postérieur à l'apparition de la mutation tandis que le TMRCA d'une paire discordante est forcément plus ancien. Ainsi en comparant les âges d'un grand nombre de paires de chaque type, l'âge d'apparition est estimé à l'intersection des âges des paires concordantes et discordantes (**Figure 28**).

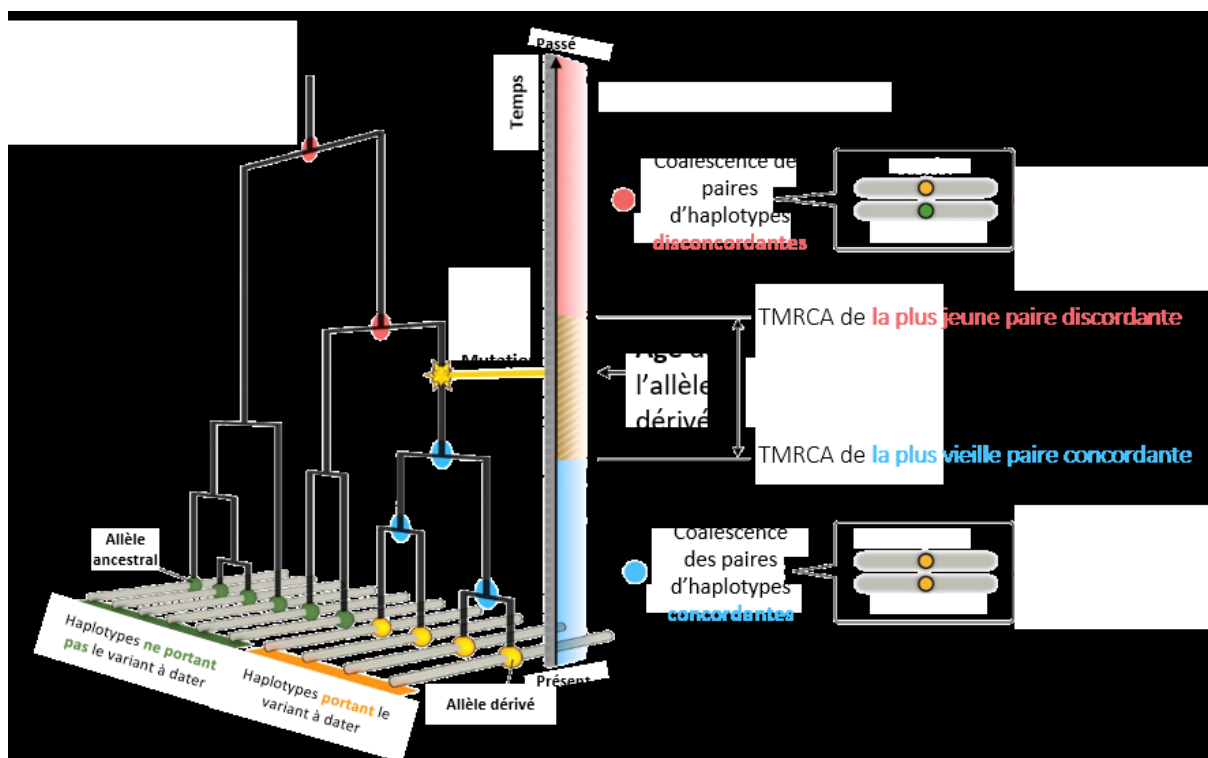


Figure 28 : Principe de la méthode GEVA.

D'après <https://human.genome.dating/info/about>  
©P. Albers 2019.

La description détaillée du fonctionnement de GEVA est disponible dans la publication correspondante : Albers, P.K., McVean, G., 2020. Dating genomic variants and shared ancestry in population-scale sequencing data. *PLOS Biology* 18, e3000586. <https://doi.org/10.1371/journal.pbio.3000586>.

## 3 Mise en place de l'analyse

### 3.1. Adaptation de la méthode GEVA au cas du mouton

La méthode GEVA repose sur la détermination de la longueur d'un segment d'ADN partagé entre deux haplotypes et qui dépend des mutations et des recombinaisons qui ont lieu dans la région concernée. GEVA détermine cette longueur grâce à une chaîne de Markov cachée (CMC) qui segmente la paire d'haplotype en deux états : l'état local et l'état périphérique. **L'état local** correspond au segment partagé entre les deux haplotypes aux abords immédiats de la mutation à dater. **L'état périphérique** constitue le reste de la séquence, au-delà des points de recombinaison. Les probabilités d'émissions nécessaires à la construction de la CMC ont été établies empiriquement à partir de simulations d'haplotypes résultant d'un scénario démographique de l'espèce humaine. Il a donc fallu vérifier que ces probabilités étaient applicables au mouton afin de les modifier le cas échéant.



Tableau 7 : Paramètres utilisés pour la simulations de 10 000 haplotypes par msprime sous un modèle démographique typique du mouton

Paramètres	Valeurs	
<b>Variations démographiques</b>	<b>Temps (années)</b>	<b>Taille efficace des moutons</b>
	1 000 000	40 000
	398 107	35 000
	158 489	60 000
	50 118	25 000
	12 589	30 000
	3 162	2 000
<b>Taux de mutation</b>	$2^{-8}$ /générations/pb	
<b>Temps de génération</b>	2 ans	
<b>Carte de recombinaison</b>	(Petit et al., 2017)	
<b>Longueur</b>	82 951 069 pb	

Une chaîne de Markov cachée permet de modéliser une séquence par plusieurs modèles qui alternent le long de la séquence selon une chaîne de Markov. C'est un moyen de la segmenter. Dans le cas qui nous occupe, elle se compose de deux processus :

- **Le processus caché** : la suite des états le long de la séquence, local ou périphérique. C'est celui que l'on cherche à identifier pour segmenter la séquence.
- **Le processus observable** : les états génotypiques directement observables : homozygote ancestral (G0), homozygote dérivé (G2) ou hétérozygotes (G1).

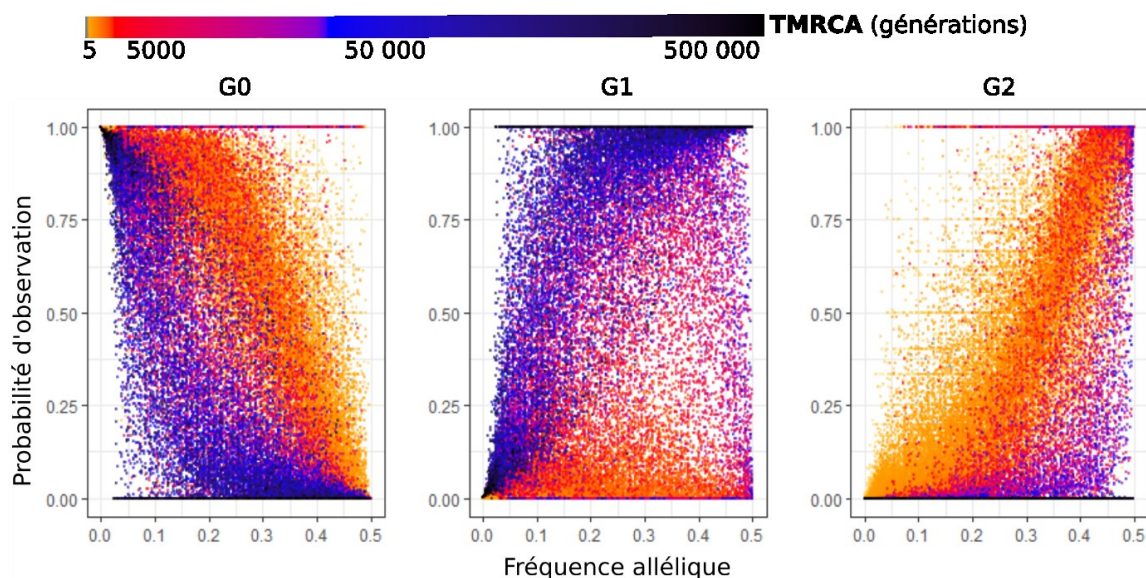
Le modèle d'émission des états observés dépend des états cachés sous-jacents. Il correspond aux probabilités d'observer chaque état génotypique pour chacun des deux états cachés : l'état local et l'état périphérique. Pour vérifier que ces probabilités d'émission sont applicables au mouton, nous avons repris les mêmes étapes de construction que les auteurs de la méthode GEVA.

Nous avons commencé par simuler 10 000 haplotypes sous un modèle démographique typique de l'évolution du mouton grâce au programme MSPRIME (Kelleher, Etheridge and McVean, 2016). Les paramètres définissant le modèle ainsi que les changements de taille de populations au cours du temps sont résumés dans le **Tableau 7**. Les variations démographiques sont tirées de Alberto et al. (2018). Nous avons simulé des haplotypes de 82 951 069 paires de bases (pb), ce qui correspond à la longueur du chromosome n°13. C'est un bon compromis entre la taille des données simulées et les ressources nécessaires pour les traiter. Nous avons également appliqué un taux de recombinaison variable suivant la carte de recombinaison correspondant à ce chromosome (Petit et al., 2017).

Ensuite, nous avons tiré au hasard environ 30 000 paires d'haplotypes, que nous avons découpés en segments d'après les points de recombinaison. Au total, 150 millions de segments ont ainsi été générés. Pour environ 600 000 segments tirés au hasard nous avons calculé la fréquence de chaque variant dans l'échantillon global, l'état génotypique observé (G0, G1 ou G2), ainsi que l'âge de l'ancêtre commun le plus récent de la paire à laquelle

appartient le segment à chaque position du segment. Cela a constitué un ensemble d'environ un milliard de sites.

Tous ces segments ont été répartis dans 100 gammes de TMRCA répartis sur une échelle logarithmique allant de 1 à 500 000 générations. Puis, tous les génotypes appartenant à ces segments ont été répartis dans des gammes de fréquence allélique allant de 0 à 1 par pas de 0.2%. Pour chaque combinaison de fréquence-TMRCA nous avons calculé la probabilité d'émettre chaque état génotypique. La **Figure 29** montre la relation entre les probabilités d'émissions des états G0, G1, et G2 en fonction de la fréquence de l'allèle dans la population totale et du TMRCA. Nous observons des tendances différentes entre les gammes de TMRCA récents (en ton chaud sur la **Figure 29**) et anciens (en tons froids sur la **Figure 29**).



**Figure 29 : Probabilités d'observation des trois états génotypiques**

(G0 : homozygote référence, G1 : hétérozygote, G2 : homozygote alternatif) mesurées dans 100 gammes de TMRCA équitablement réparties sur une échelle logarithmique allant de 0 à 500 000 générations et dans 500 gammes de fréquences allélique. Pour chaque combinaison de fréquence-TMRCA les probabilités d'observations ont été normalisées telle que leur somme soit égale à 1. Les données génétiques ont été simulées avec MSPRIME sous un scénario démographique caractéristique du mouton.

Nous décidons de diviser nos probabilités d'émission selon le même seuil que Albers et McVean 2020 : en dessous de 100 générations nous classons les segments dans l'état local, au-dessus de 100 générations, dans l'état périphérique. Nous avons généré les probabilités d'émission spécifiques à chaque état caché en calculant la moyenne des probabilités d'observation de chaque état génotypique pour chaque gamme de fréquence allélique.

Les probabilités d'émissions ainsi construites sont très fortement corrélées avec cette établies pour l'Homme par les auteurs de GEVA à l'exception de celles correspondant à l'état caché local pour l'émission de l'état génotypique G1 (**Tableau 8** et **Figure 30**). Ici, bien que la corrélation soit significative, le coefficient de corrélation est faible. Cela est dû à une différence de méthode de simulations des données. Nous n'avons en effet pas introduit d'erreur de phasage au contraire des auteurs de GEVA. Or, ces derniers montrent que les probabilités d'observation de l'état G0 sont non nulles seulement pour des sites ayant des

fréquences inférieures avant 1.5% avant l'ajout d'erreur, comme c'est aussi notre cas, alors qu'elles sont non nulles partout après l'ajout d'erreur.

Malgré cette différence, il semble que les probabilités d'émissions établies pour le mouton soient comparables à celles de l'Humain. Nous avons donc décidé d'utiliser les probabilités d'émissions disponibles pour l'humain pour nos datations.

Tableau 8 : Corrélations entre les probabilités d'émissions nécessaires à la Chaîne de Markov Cachée utilisé dans la méthode GEVA pour définir le segment partagé par deux haplotypes, établies pour l'humain dans Albers et al., 2020 et celles établies pour le mouton au cours de cette étude.

	G0		G1		G2	
	p-valeur	R <sup>2</sup>	p-valeur	R <sup>2</sup>	p-valeur	R <sup>2</sup>
Etat caché périphérique	< 2,2.10 <sup>-16</sup>	99.8	< 2,2.10 <sup>-16</sup>	84.7	< 2,2.10 <sup>-16</sup>	99.9
Etat caché local	< 2,2.10 <sup>-16</sup>	99.5	< 26,5.10 <sup>-6</sup>	28	< 2,2.10 <sup>-16</sup>	99.5

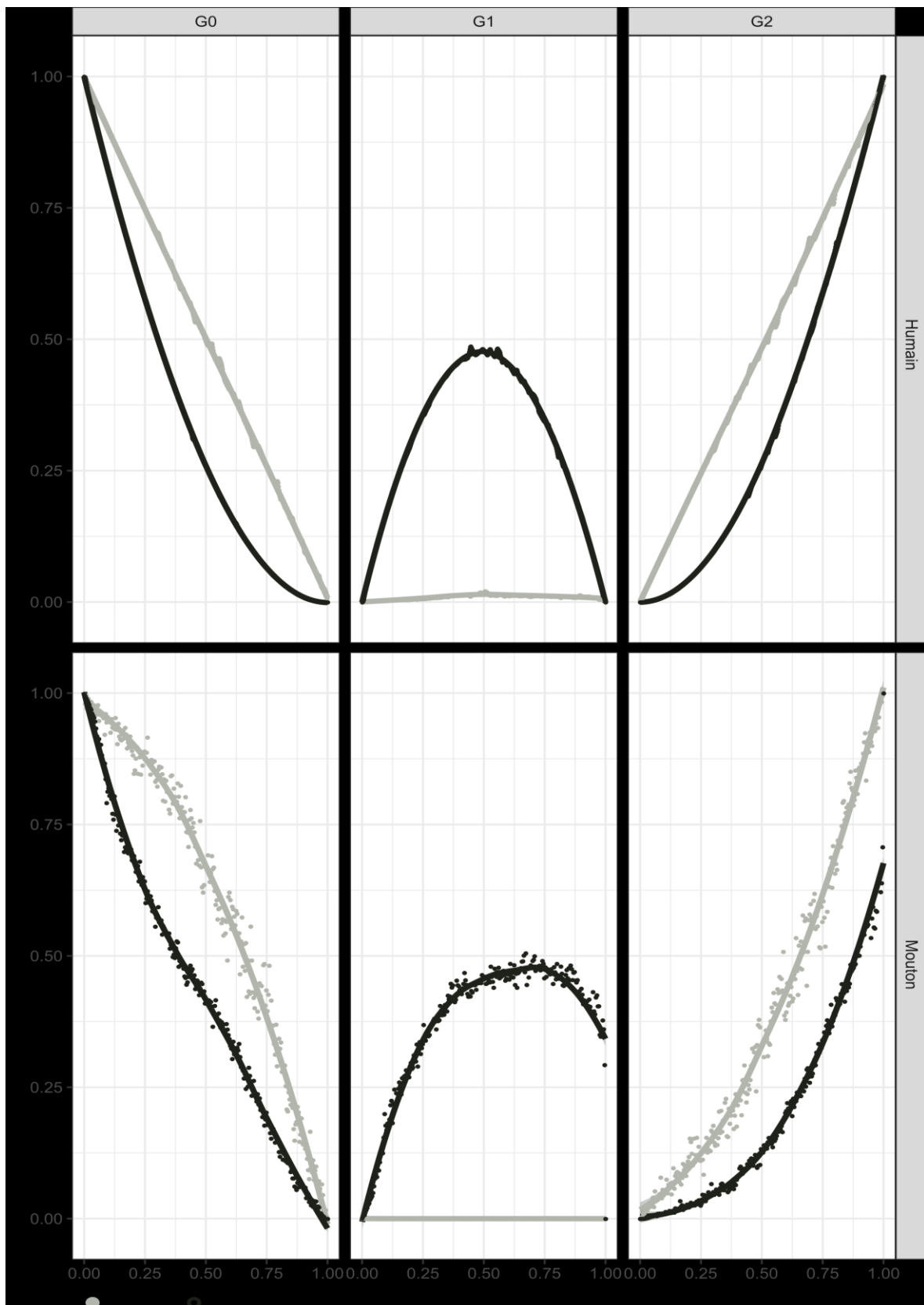


Figure 30 : Probabilités d'émissions nécessaires à la Chaîne de Markov Cachée utilisées dans la méthode GEVA pour définir le segment partagé par deux haplotypes, établies pour l'humain dans Albers et al., 2020 et établies pour le mouton au cours de cette étude.

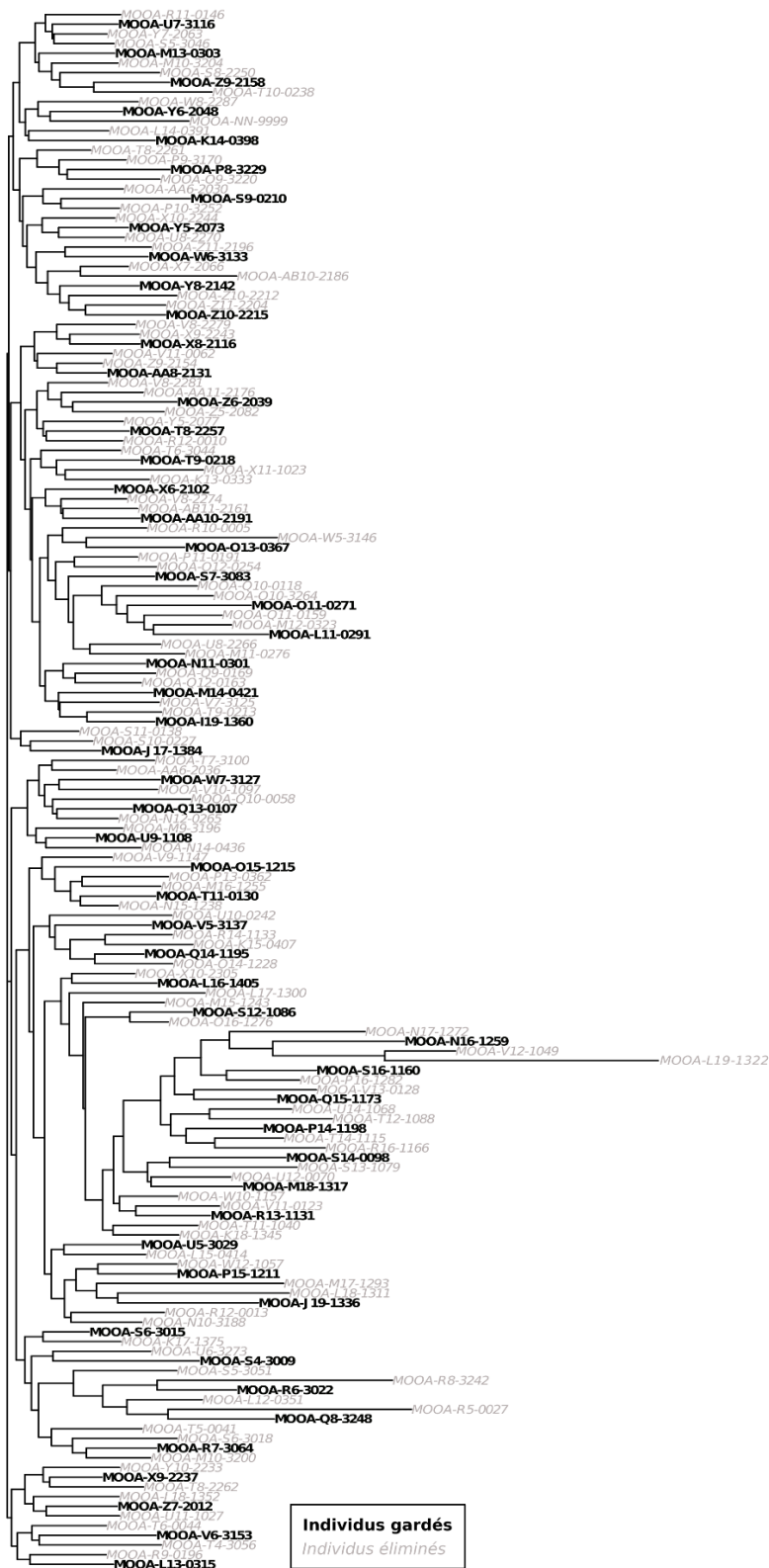


Figure 31 : Sous-échantillonnage des individus marocains utilisés pour l'analyse GEVA.

L'arbre a été construit selon une méthode de distance NJ. Les distances génétiques brutes ont été calculées sur l'ensemble du génome.

## 3.2. Echantillonnage des individus

---

D'après les tests de la méthode GEVA réalisés sur l'humain, les âges de mutations humaines estimés à partir de différents échantillons sont bien corrélés entre eux. Cependant, les estimations peuvent gagner en précision par l'élargissement du nombre d'individus dans le jeu de données et donc du nombre de paires concordantes et discordantes comparées. C'est pourquoi nous avons voulu garder un maximum d'individus, et donc de diversité génétique, pour nos datations. Cependant, notre jeu de données est hétérogène et contient une grande proportion (161/376, 42%) d'individus marocains. Nous avons décidé de sous-échantillonner les individus de ce groupe pour diminuer la probabilité de ne conserver que des haplotypes marocains lors du tirage aléatoire des paires nécessaires à la datation. Le risque serait alors de ne pas avoir assez de paires, notamment discordantes. Nous avons donc gardé 54 individus marocains sur 161. La probabilité de piocher seulement des marocains sur 100 tirages aléatoires est alors de  $10^{-11}$ . Pour les choisir tout en gardant un maximum de leur diversité, nous avons construit un arbre de distance génétique et sélectionné un individu sur trois tout le long de l'arbre (**Figure 31**).

## 3.3. Sélection des variants à dater

---

Nous avons daté les variants bi-alléliques dont nous avons inféré les états ancestraux et dérivés. Nous avons aussi filtré les données pour ne pas dépasser 25% de données manquantes. Nous avons ainsi daté plus de 61 millions de variants répartis sur les chromosomes 1 à 26 (**Tableau 9**).

## 3.4. Paramétrage des analyses de datation

---

### 3.4.1. Paramètres GEVA

Les datations ont été réalisées grâce au programme exécutable disponible à cette adresse (<https://github.com/pkalbers/geva>). Les fichiers d'entrée ont été générés à partir des fichiers VCF grâce à l'option `-vcf` du programme GEVA. La carte de recombinaison utilisée est celle produite par Petit *et al.* 2017. Les coordonnées de cette carte ont été modifiées pour correspondre à la version 4 du génome de référence du mouton (GCF\_000298735.2) sur laquelle sont alignées nos données grâce à l'outil de correspondance du NCBI (<https://www.ncbi.nlm.nih.gov.insee.bib.cnrs.fr/genome/tools/remap>). Nous avons fixé le taux de mutation à  $2 \cdot 10^{-8}$  mutations par génération et par paire de base. Pour la taille efficace nous avons choisi la valeur de 1000. Les moutons présentent une diversité génétique importante comparée aux autres espèces de bétail. De nombreuses races présentent des tailles efficaces comprises entre 500 et 1000 (Kijas *et al.*, 2012a). Comme notre jeu de données est composé de multiples races, nous avons choisi la borne supérieure de ces estimations. Le paramètre de taille efficace est un paramètre d'échelle, sa valeur n'a donc pas à être précise mais simplement réaliste. Sa modification n'entraîne pas d'effets forts sur les datations (Albers and McVean, 2020).

### 3.4.2. Ressources computationnelles

Le temps de calcul nécessaire à la datation d'un variant est en moyenne de 15 secondes et nécessite 4 à 12 GB de mémoire vive en fonction du nombre de paires échantillonnées et de la taille des chromosomes. Pour exploiter au mieux les ressources de calcul fournies par GRICAD, nous avons construit une campagne de datations par chromosome. Chaque campagne compte un nombre variable de jobs (15 000 à 70 000 environ), chacun correspondant à la datation de 100 variants. Ces campagnes ont été lancées sur deux clusters de calculs partageant une base de stockage commune et sur un mode best-effort. Ce mode permet d'exploiter toutes les ressources disponibles correspondant aux besoins de la campagne. Le système de partage équitable des ressources des clusters peut amener à l'élimination de jobs par des utilisateurs prioritaires. Dans ce cas les jobs concernés sont resoumis automatiquement. Le blocage complet des ressources pour éviter ces problèmes aurait été moins efficace car l'attribution des jobs aurait été plus lente.

### 3.4.3. Architecture des répertoires de travail

L'architecture des répertoires de travail est aussi un élément important dans la construction des campagnes. En effet, chaque campagne génère des milliers de fichiers. Il est donc important de bien répartir les fichiers de sorties dans plusieurs répertoires au risque de ralentir ou de faire tomber la plateforme de stockage. Nous avons choisi de limiter à environ 500 fichiers la capacité de chaque répertoire de sortie. Tous les fichiers générés ont ensuite été concaténés par chromosome à la fin des campagnes.

## 3.5. Filtration des estimations

Le programme `GEVA` fonctionne selon une approche composite. Pour rappel, il estime l'âge d'une mutation grâce à la datation de multiples paires d'haplotypes portant ou non la mutation (voir partie 2.2). Certaines de ces estimations peuvent être faussées par des erreurs d'appel des variants, des données manquantes, des pertes par mutation reverse (etc.). C'est pourquoi le programme filtre ces estimations qu'il considère correspondre aux 'outliers' : En pratique, les paires concordantes et discordantes sont triées indépendamment par TMRCA croissant. Ensuite, les paires dont les distributions se recouvrent sont éliminées.

Pour ne garder que les estimations d'âge de bonne qualité, nous nous sommes basés sur la proportion de paires rejetées par ce filtre. Nous avons calculé pour chaque variant un score de qualité (QS) :

$$QS = 1 - \max \left\{ \frac{\text{nombre de paires concordantes rejetées}}{\text{nombre de paires concordantes}}, \frac{\text{nombre de paires discordantes rejetées}}{\text{nombre de paires discordantes}} \right\}$$

Ce score varie de 0 à 1. Une valeur nulle indique une qualité d'estimation faible. Nous avons choisi de ne garder que les variant avec un QS supérieur à 0.7.

## 4 Distribution des âges des mutations dans le génome du mouton

### 4.1. Nombre de paires d'haplotypes analysées et de variants datés

La campagne de datation résumée en quelques chiffres

Après environ **30 années** de calculs cumulés réalisées sur une moyenne de 500 à 1000 cœurs de calculs en simultanés, et l'analyse de **22,850,969,870 paires** d'haplotypes, environ 40 millions de variants ont été datés par chacune des trois horloges. Cela représente entre 66 et 75% du nombre de variants initial soumis à la datation suivant le type d'horloge. Le nombre de variants conservés pour les analyses passe à environ **35 millions** après le filtre basé sur le score de qualité fixé à 70%. Cela représente environ 50% du nombre de variants initial (**Tableau 9**).

Dans le détail, le nombre de variants datés diffère légèrement en fonction de l'horloge moléculaire. L'horloge de recombinaison est celle qui en date le plus, suivi par l'horloge jointe et l'horloge de mutation. Après le filtre de qualité 56% des âges estimés sont conservés pour l'horloge jointe, 50% pour celle de recombinaison et 46% pour celle de mutation. Le nombre de variants datés dépend aussi du chromosome, et est proportionnel à la taille de ces derniers. Le détail des nombres de variants datés pour chaque chromosome et chaque horloge est résumé dans le **Tableau 9** et la **Figure 32**.

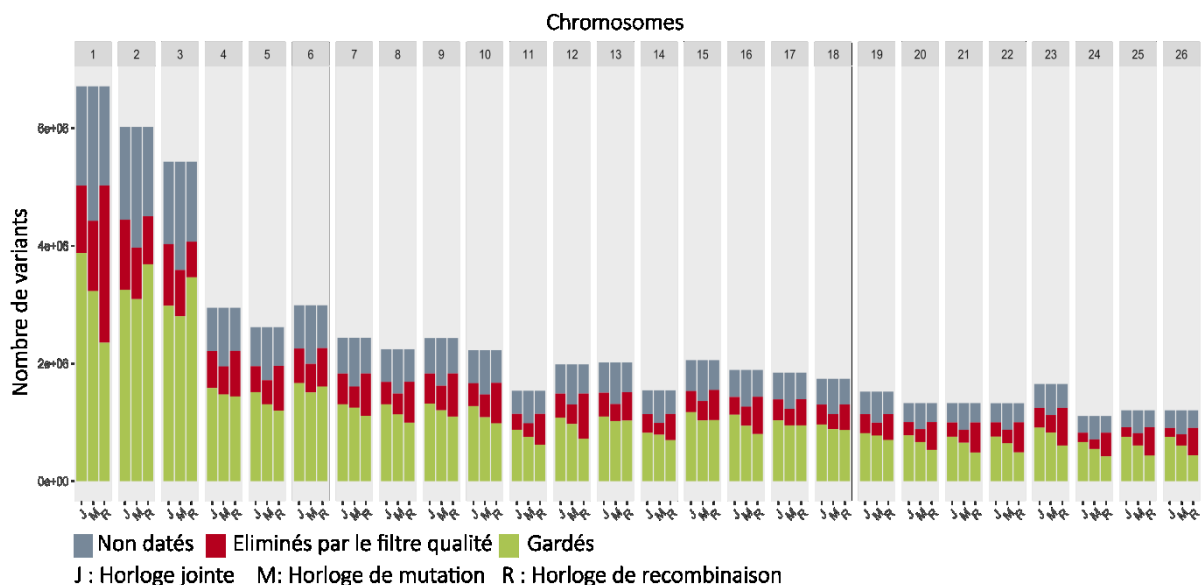


Figure 32 : Nombre de variants effectivement datés

Par chromosomes et par horloge moléculaire par la méthode GEVA comparé aux nombre de variants soumis à la datation avant et après le filtre de qualité.



Tableau 9 : Nombres de variants effectivement datés par chromosome et par horloge moléculaire par la méthode GEVA comparé au nombre de variants soumis à la datation

Chromosome	Nombre de variants initial	Nombre de variants datés avant filtre qualité			Nombre de variants datés après filtre qualité		
		J	M	R	J	M	R
1	6711485	5023063	4425162	5024715	3880574	3231573	2354793
2	6021769	4445835	3968559	4505585	3252290	3094526	3684462
3	5432416	4031226	3588976	4072697	2980675	2804738	3467434
4	2947038	2209989	1951740	2214853	1582570	1475263	1442882
5	2619035	1954306	1715558	1960040	1511162	1308504	1199792
6	2990581	2256288	1995400	2262214	1669087	1507332	1609800
7	2436373	1829178	1609959	1830103	1308108	1250266	1110100
8	2242723	1688327	1489347	1691444	1306241	1136500	994785
9	2432206	1829229	1622191	1830324	1318143	1205095	1096596
10	2225762	1666277	1473725	1672412	1277327	1087726	983931
11	1538245	1144991	984751	1145353	875537	752276	618949
12	1984275	1488446	1301553	1488969	1079039	973691	721861
13	2017513	1504744	1312224	1511332	1098013	1020770	1030855
14	1544010	1143186	992162	1144117	822645	789728	697169
15	2056591	1532488	1361826	1550452	1174939	1032543	1037464
16	1890532	1431259	1266988	1433489	1129634	941662	803529
17	1845503	1391196	1226956	1392981	1033706	948652	949304
18	1740237	1300861	1143348	1302325	961489	884088	869239
19	1522779	1142653	994745	1143276	818031	776797	701309
20	1327726	1006061	884037	1006464	780858	668165	532336
21	1328052	997506	874525	998137	754031	654481	486739
22	1325420	999387	876881	999803	757399	645626	489209
23	1653178	1244535	1125776	1245087	912804	826371	604345
24	1108190	824218	711496	824383	666214	546105	423839
25	1204695	918807	816224	919016	752530	604995	436716
26	1204540	904185	799370	904369	752110	603486	443239
<b>Total</b>	<b>61350874</b>	<b>45908241</b>	<b>40513479</b>	<b>46073940</b>	<b>34455156</b>	<b>30770959</b>	<b>30770959</b>
<b>% du nombre initial</b>	<b>100</b>	<b>75</b>	<b>66</b>	<b>75</b>	<b>56</b>	<b>50</b>	<b>46</b>

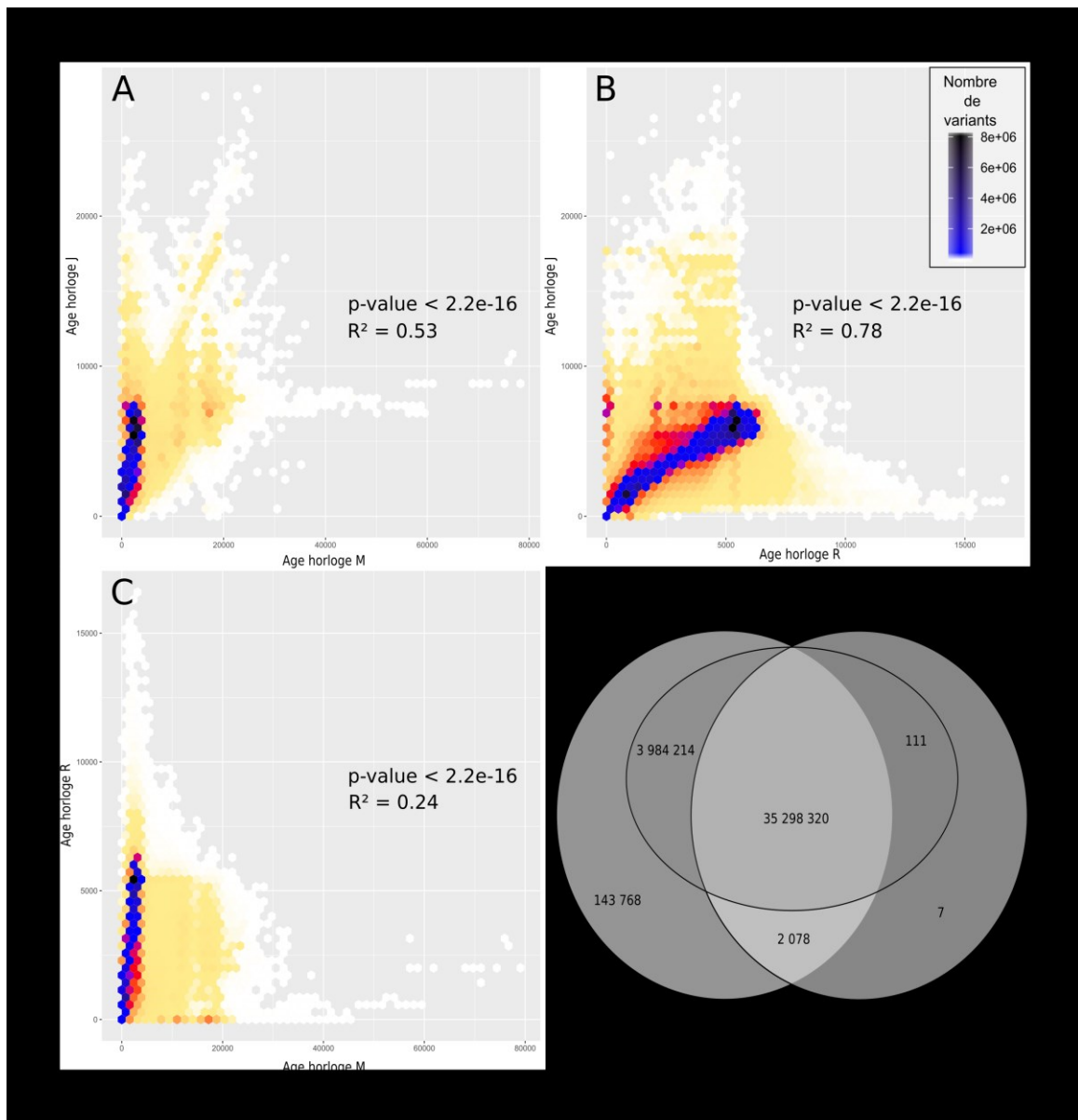
## 4.2. Comparaison des différentes horloges et choix de celle utilisée pour les analyses subséquentes

### 4.2.1. Taux de recouvrement entre horloges

La majorité des variants, 89%, effectivement datés le sont par les trois horloges. Quelques-uns ne le sont que par une ou deux des trois : 0.4% des variants ne sont datés que par l'horloge de recombinaison, 10% ne sont datés que par les horloges jointe et recombinaison, et moins de 1/10000 n'est datée que par l'horloge de mutation ou par elle et une des deux autres (**Figure 33.D**).

## 4.2.2. Corrélations et différences entre les horloges avant le filtre qualité

On observe des corrélations significatives entre les âges des variants datés par les trois horloges (**Figure 33.A,B et C**). L'horloge de recombinaison et l'horloge jointe sont les plus corrélées entre elles avec un coefficient de 0.78 (**Figure 33.B**). Les différences entre les horloges résident donc surtout dans les distributions des estimations. Les datations par l'horloge de mutation sont globalement plus récentes avec un âge moyen de 2208 générations contre 3235 et 4857 pour les horloges jointe et de recombinaison respectivement. L'horloge de mutation est aussi celle qui présente les âges les plus anciens, environ 3 à 4 fois plus ancien que pour les autres horloges (**Tableau 10**).



**Figure 33 : Comparaison des trois horloges de GEVA.**

Les panneaux A, B et C représentent les corrélations des datations issues des trois horloges. Les âges sont donnés en générations. Le panneau D représente le nombre de variant datés par une, deux ou trois horloges

Tableau 10 : Résumé statistiques des âges de variants estimés par les trois horloges de GEVA et avant application du filtre qualité.

M : horloge de mutation ; R : horloge de recombinaison ; J : horloge jointe

	M	R	J
Minimum	7	0.03	6.061
Moyenne	2208	3235	4857
Médiane	2413	3799	5368
Maximum	78289	16516	28350

Au niveau des distributions globales des âges estimés par les différentes horloges, on remarque qu'elles présentent toute une distribution quasi normale autour de l'âge moyen, ainsi qu'un pic dans une période plus récente. Globalement, les distributions ne diffèrent presque pas entre les chromosomes (**Figure 34**, **Figure 35**, **Figure 36**).

On remarque la présence d'un pic très marqué dans la distribution des âges estimés par l'horloge de recombinaison (cadres rouges dans la **Figure 35**). Ce pic est retrouvé dans les distributions des chromosomes 2 à 10 et 13 à 19 (**Figure 35**). Il correspond en fait à des variants situés dans des zones non couvertes par la carte de recombinaison.

Dans les distributions de l'horloge jointe, on trouve quelques pics qui rompent l'allure gaussienne de la distribution globale (cadres bleus dans la **Figure 34**). Les variants correspondants sont absents des populations domestiques, et sauvages iraniennes, mais présent chez les espèces d'*Ovis* américaines de notre jeu de données. Ils ne nous seront pas utiles pour les analyses ultérieures.

Dans les distributions de l'horloge de mutation (**Figure 36**), on trouve des variants très anciens en queue de distribution et que l'on ne retrouve pas datés aussi précocement par les autres horloges.

Le filtre basé sur le score de qualité à 70% ne change globalement pas les distributions des âges. Il supprime les très vieilles estimations de l'horloge de mutation. Il supprime également quelques estimations correspondant à des variants en fréquence nulle chez les domestiques et sauvages iraniens. En revanche, le filtre qualité n'élimine pas les pics dus à l'absence de carte recombinaison pour les estimations de l'horloge de recombinaison.

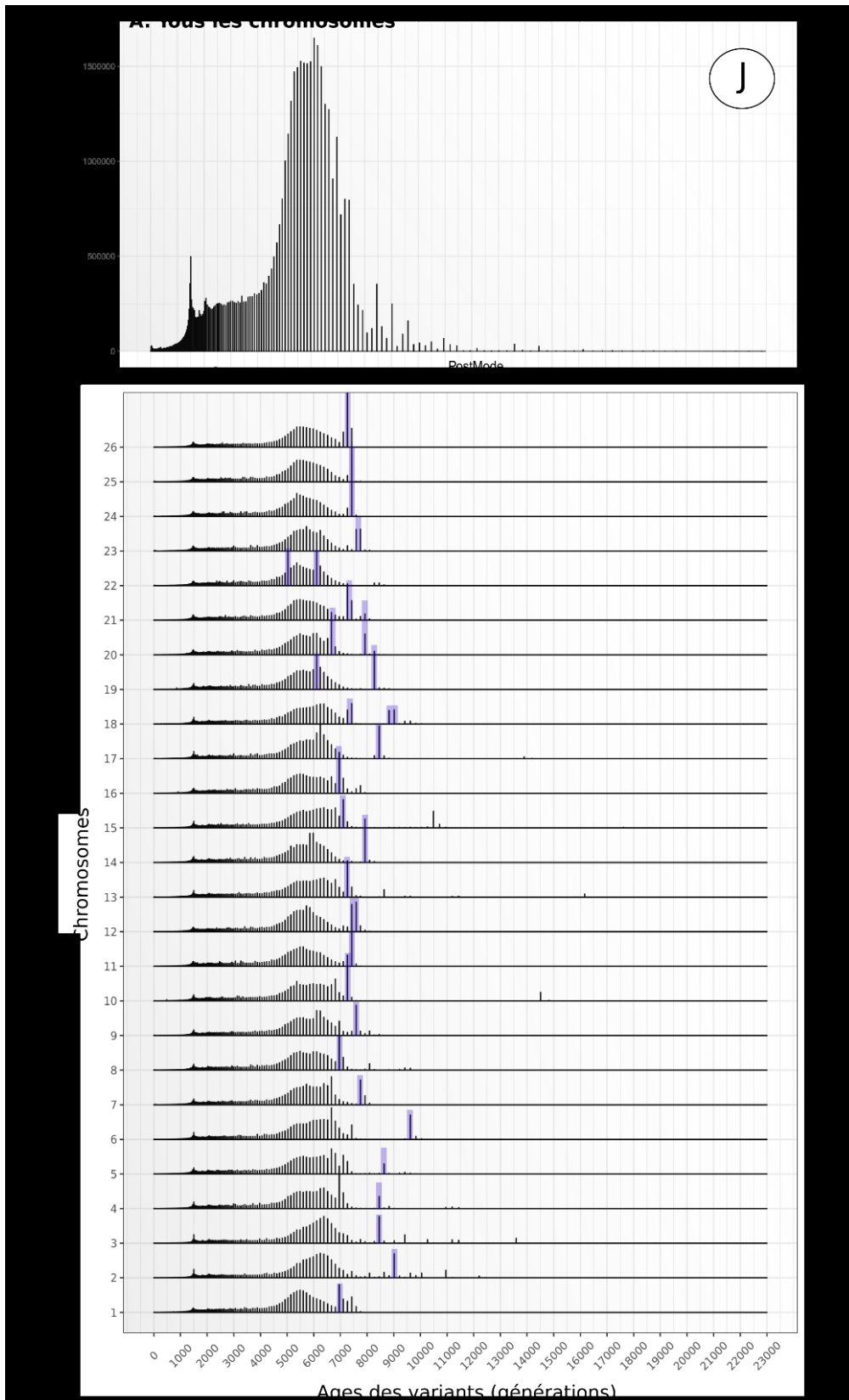
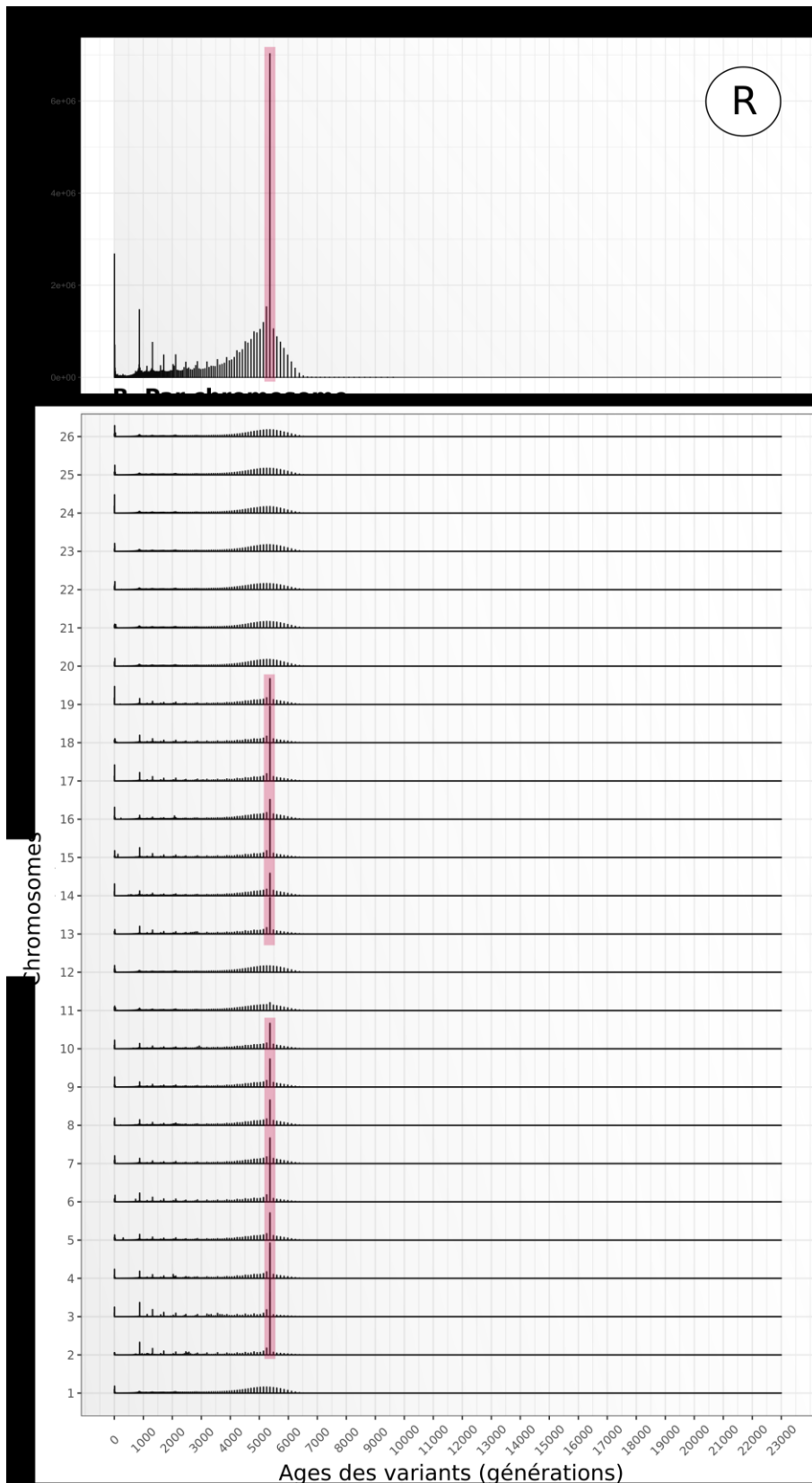


Figure 34 : Distribution des âges des variants estimés par l'horloge jointe avant application du filtre qualité.

Les cadres bleus indiquent des pics de distributions correspondant à des variants absents des populations domestiques et présents uniquement chez les *Ovis* américains.



**Figure 35 : Distribution des âges des variants estimés par l'horloge de recombinaison avant application du filtre qualité.**

Les cadres rouges indiquent des pics de distributions correspondant à des variants situés dans des zones hors carte de recombinaison.

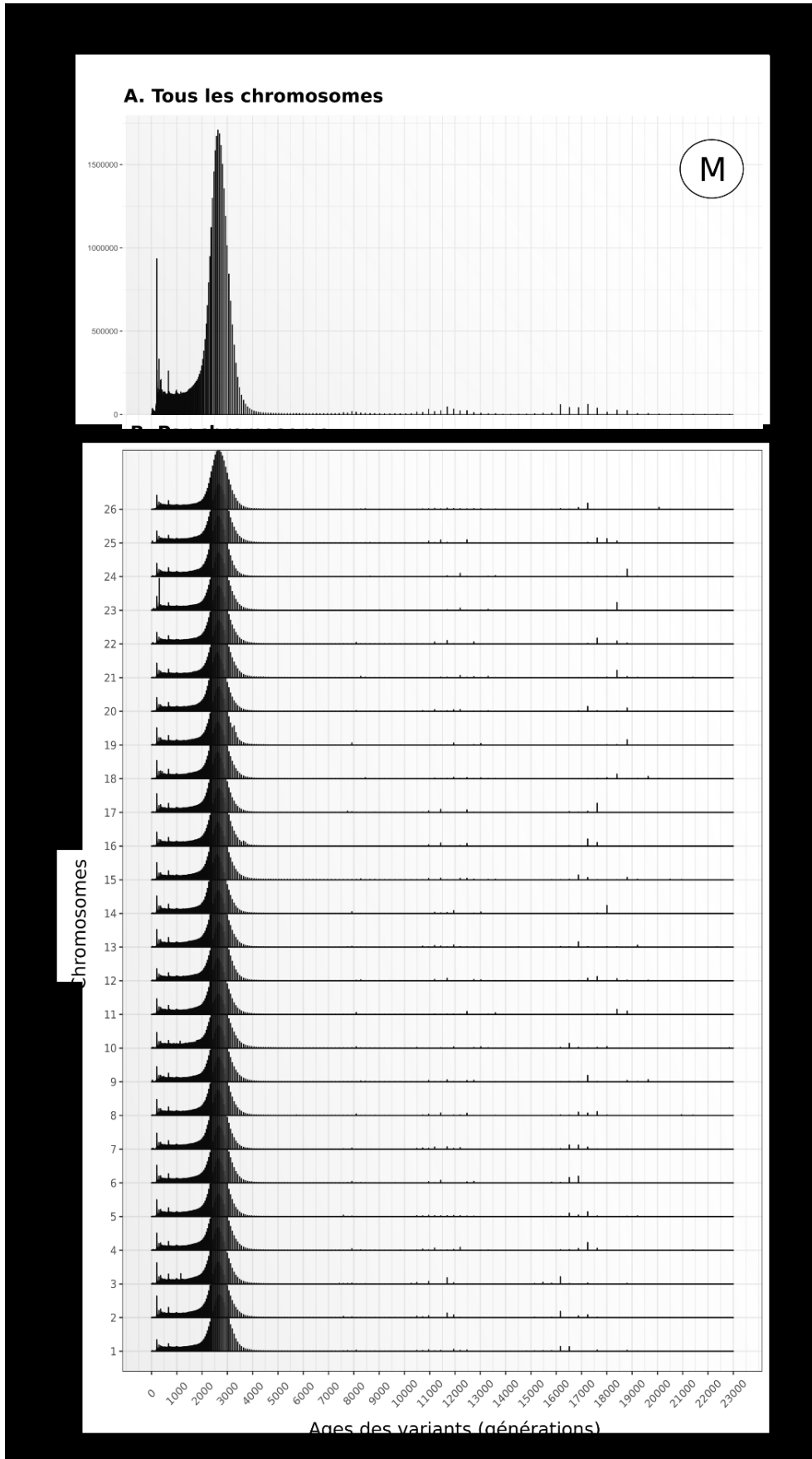


Figure 36 : Distribution des âges des variants estimés par l'horloge de mutation avant application du filtre qualité

### 4.2.3 Choix de l'horloge pour les analyses

L'horloge de recombinaison pose problème pour un grand nombre de variants dans des zones hors carte de recombinaison et peut varier de manière importante si la longueur de l'haplotype partagé est sur ou sous-estimée. L'horloge de mutation est plus stable dans le temps et notamment pour les variants les plus anciens (Albers and McVean, 2020) mais elle repose ici sur un taux de mutation constant. En pratique ce taux peut varier fortement dans le génome mais nous ne disposons pas de cette information. Aussi, nous préférons ne pas baser nos datations sur cette unique information. De plus, des simulations montrent que l'horloge jointe est la plus précise puisqu'elle prend en compte deux sources d'informations (mutation et recombinaison) et permet de réduire le biais d'estimation (P.K. Albers, communication personnelle). Nous choisissons donc de nous baser sur les estimations de l'horloge jointe pour la suite. Les distributions des âges qui passe le filtre qualité pour cette horloge sont représentées dans la **Figure 37**.

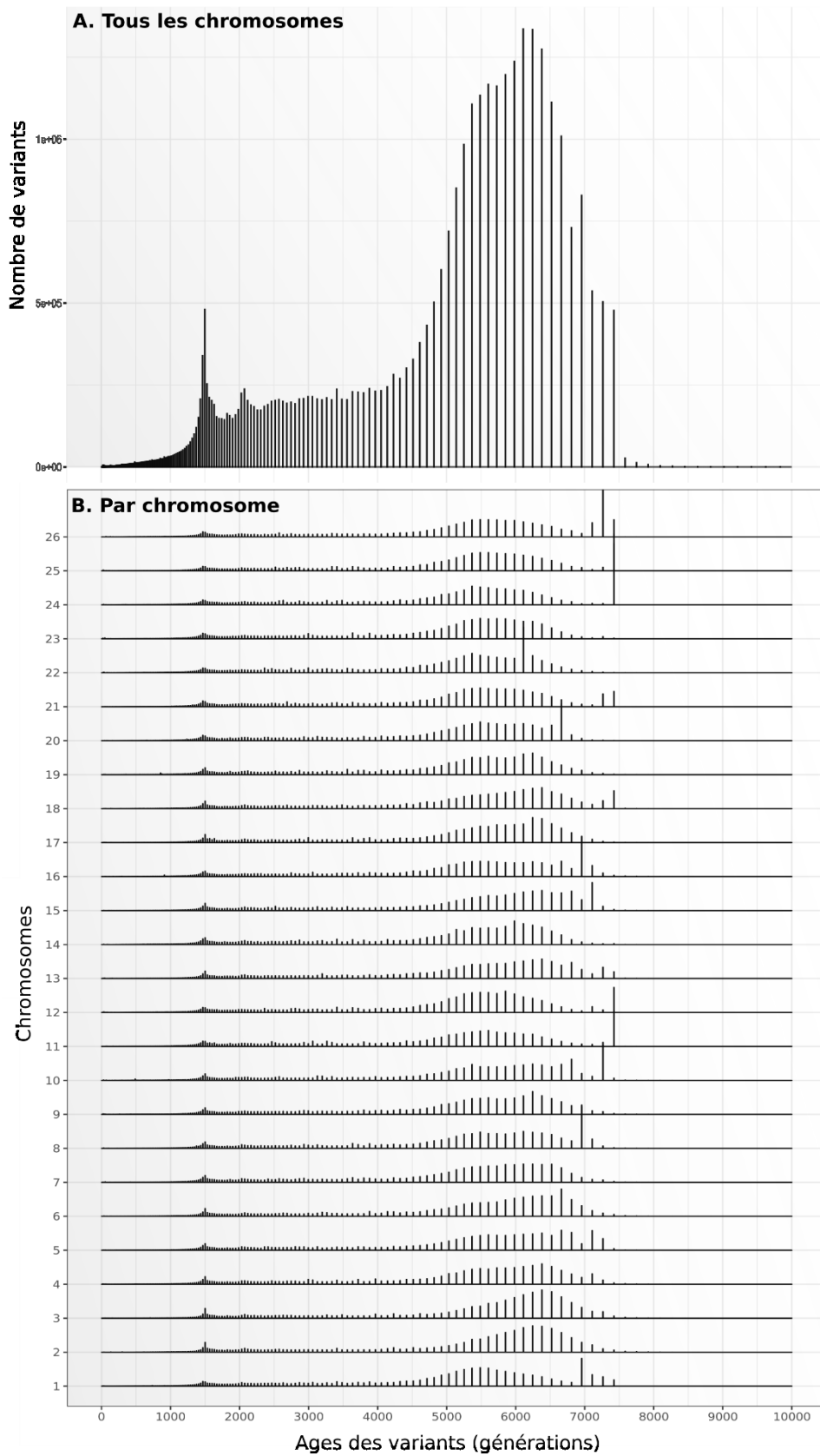


Figure 37 : Distribution des âges des variants estimés par l'horloge J après application du filtre qualité



## 4.3. Distributions des âges après choix de l'horloge et filtre qualité

### 4.3.1. Nombre de variants datés le long du génome

Les variants datés se répartissent tout le long du génome (**Figure 38**). Le nombre de variants datés par fenêtre de 250 kb varie avec la densité de variants (p-valeur =  $2.10^{-16}$ ,  $R^2 = 85\%$ , test de Pearson). Les fenêtres comptent en moyenne 3520 variants datés et 8 au minimum.

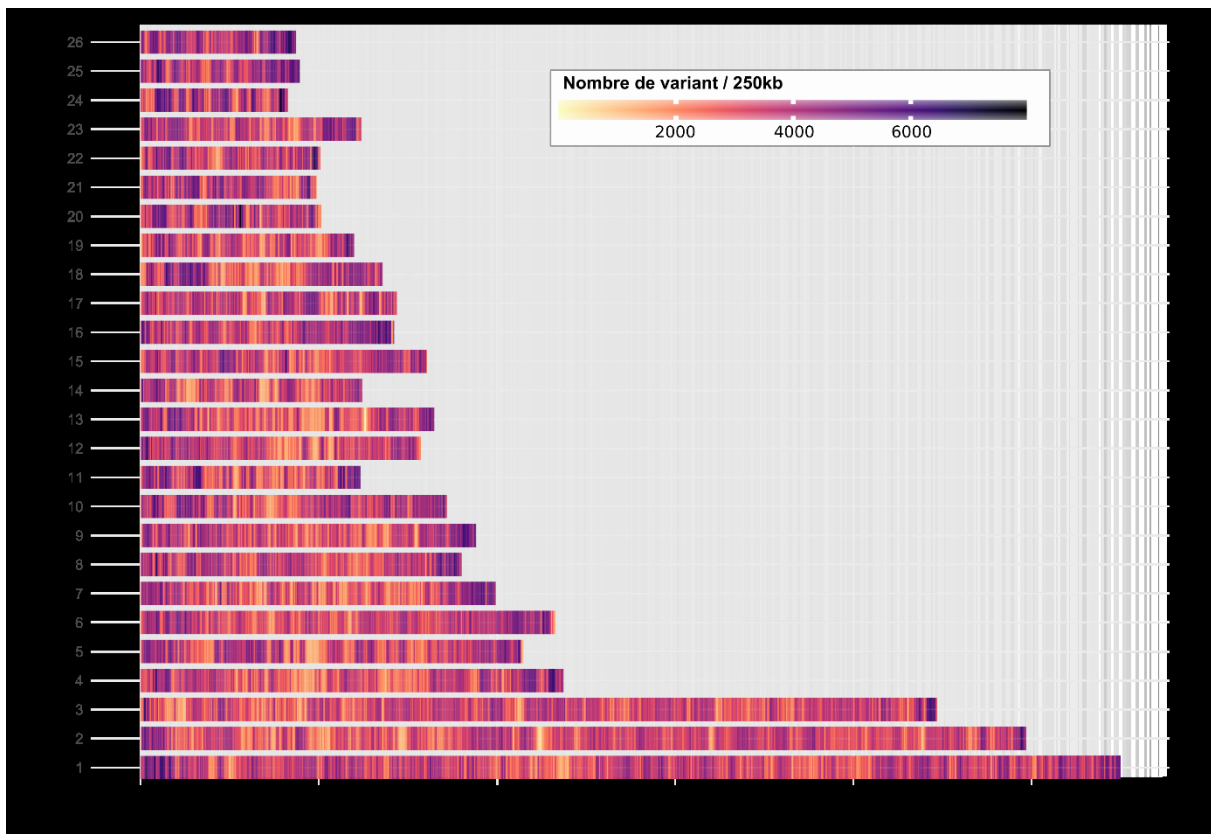


Figure 38 : Nombre de variants datés le long du génome

### 4.3.2. Distribution globale des âges des variants

La distribution des âges estimés par l'horloge jointe et ayant un score de qualité supérieur à 70% est représentée dans la **Figure 37**. La majorité des variants se concentrent entre 4500 et 7000 générations avant le présent dans une distribution centrée sur 6000 générations.

Le nombre de mutation apparues diminue continuellement jusqu'à environ 1500 générations (3000 ans avec un temps de génération de 2 ans) où l'on observe un pic correspondant à plus de 1.6 millions de variants compris entre 1400 et 1600 générations. La majorité (plus de 90%) des variants de ce pic présente des fréquences très faibles. Environ 60% de ces variants sont absents des populations domestiques et uniquement présents chez

les mouflons. Le reste est constitué de variants rares (avec une fréquence moyenne inférieure à 1%) et donc spécifiques à un groupe (race ou région géographique).

## 4.4. Disponibilité de l'Atlas

---

Nous avons conçu une interface web pour rendre disponible toutes les datations brutes (non filtrées) pour chacune des horloges. Le développement a été fait par Clément Lionnet et Lucas Moreau. L'interface permet d'explorer les âges de variants à travers des visualisations graphiques et des tableaux. Les utilisateurs peuvent également télécharger les données de datations des variants de leur choix pour compléter leurs analyses. Le site est aussi pensé pour évoluer. Il sera possible d'ajouter de nouvelles datations grâce à l'étude d'échantillons plus larges ou même celles d'une autre espèce comme la chèvre. Il sera prochainement disponible au public.

# 5 Diffusion et connexions des populations au cours du temps

---

## 5.1. Formations des populations

---

### 5.1.1 Principes, objectifs et questions

#### Hypothèse de base

Si une population est suffisamment isolée des autres en terme de flux de gènes alors les mutations qui s'y accumulent sont spécifiques à cette population. Ainsi, on peut faire l'hypothèse que l'âge des variants spécifiques à une population reflète l'âge de son isolement.

Cette hypothèse pourrait nous permettre de retracer la chronologie de la formation et de la diffusion des différentes populations de moutons de notre échantillon. Elle souffre cependant de plusieurs limites potentielles.

#### Limite 1 : Le nombre de mutations apparues et maintenues depuis la domestication

Les mutations sont des événements rares. A un nucléotide donné,  $2N_e\mu^{14}$  mutations apparaissent à chaque génération. Si l'on prend une population avec une taille efficace  $N_e = 500$  et taux de mutation  $\mu = 2.10^{-8}/\text{pb}/\text{génération}$ , des valeurs usuelles pour une espèce comme le mouton, on obtient  $2.10^{-5}$  mutations par génération pour un seul nucléotide dans la population. Le génome du mouton compte environ trois milliards de paires de bases. On a donc 60 000 nouvelles mutations qui apparaissent à chaque génération dans la population. Finalement, depuis la domestication il y a environ 10 500 ans, soit 5250 générations si l'on prend un temps de génération de deux ans, on totalise 315 millions de mutations. Beaucoup des nouveaux variants ainsi créés ne seront pourtant pas maintenus. Selon la théorie neutraliste de l'évolution (Kimura et al., 1986), la probabilité de fixation d'une nouvelle

---

<sup>14</sup>  $N_e$  = taille efficace,  $\mu$  = taux de mutation

mutation est égale à sa fréquence initiale, soit  $1/2Ne$ . Par conséquent sa probabilité d'élimination est de  $1-1/2Ne$ . Si on reprend la valeur de 500 pour  $Ne$ , une nouvelle mutation a 1 chance sur 1000 de se fixer et 99,9 % de chance d'être éliminée. Entre ces deux extrêmes on peut aussi compter le polymorphisme transitoire. De plus, le devenir d'une mutation dépend aussi de son type. Si elle est neutre, elle peut aussi bien être conservée qu'éliminée en fonction de la dérive génétique. Si elle est délétère, la sélection naturelle augmente sa probabilité d'élimination. A l'inverse, si elle est bénéfique, elle est favorisée par la sélection. Au final, nous pouvons dire que malgré le caractère récent de la domestication à l'échelle des temps évolutifs, il y a matière à utiliser l'âge des mutations pour préciser la chronologie des formations de populations.

### **Limite 2 : les effets fondateurs**

La domestication, et la diffusion des populations qui s'en suit, peuvent créer des goulots d'étranglements plus ou moins importants. Les hommes ont capturé seulement une fraction de la diversité des ancêtres sauvages lors des étapes initiales de la domestication. De la même manière, ils ont migré dans le monde entier en emportant des troupeaux représentant une petite partie de la diversité des animaux domestiques présents au niveau du centre de domestication (Larson and Burger, 2013).

Selon notre hypothèse de base, on cherche des mutations population-spécifiques dont l'âge nous renseignerait sur l'âge de formation de ladite population. Or, certaines de ces mutations peuvent provenir de la capture de diversité initiale, et donc être plus vieilles que la formation de la population. Cependant, ce cas suppose non seulement une conservation de la mutation dans la population considérée mais surtout son élimination dans toutes les autres. Ou au moins un différentiel de fréquence important entre la population et toutes les autres. Ceci est d'autant plus probable dans les petites populations issues d'un effet fondateur où la dérive est forte. On pourrait aussi observer cela à cause de notre échantillonnage.

### **Limite 3 : les flux de gènes**

Notre hypothèse de base suppose aussi que la séparation des populations a duré dans le temps. L'isolement est nécessaire à l'accumulation de nouvelles mutations spécifiques. A priori, les espèces domestiques ne répondent pas forcément à cette supposition. Elles sont, et ont été soumises à de nombreux échanges commerciaux, à des croisements, volontaires ou non et même à des remplacements. Mais, ce n'est peut-être pas le cas pour toutes, tout au long de leur histoire. On peut notamment penser aux peuples nomades, qui vivent parfois en autarcie depuis des siècles, ou aux limites de front d'expansion qui sont moins propices aux mélanges.

### **Limite 4 : les convergences**

Il existe aussi une possibilité que des mutations identiques surviennent indépendamment dans plusieurs populations pourtant bien isolées. Mais les mutations sont déjà des

événements rares alors on peut raisonnablement considérer les convergences comme des événements négligeables.

## Objectif

Enfinement en partant de cette hypothèse et en ayant conscience des limites qu'elle implique, nous avons étudié la distribution des âges de variants spécifiques à chacune de nos populations. L'objectif étant d'inférer une chronologie des formations de populations depuis les étapes initiales de la domestication des moutons ou de discerner quels processus évolutifs ont été à l'œuvre lors de la séparation des populations ovines.

La définition des variants "**spécifiques**" à chaque population est délicate. Au sens strict, il s'agit de mutations présentes dans la population cible mais absente dans toutes les autres. Etant donné les limites énoncées précédemment comme de potentiels échanges entre populations, cette définition stricte pourrait nous faire passer à côté de nombreux variants à l'origine des populations que l'on étudie. Aussi, nous avons également développé en parallèle une seconde approche moins catégorique basée sur des variants "**caractéristiques**" d'une population cible définis sur la base de valeurs de FST.

### 5.1.2 Choix des populations

Pour les deux approches, nous avons construit des groupes à différentes échelles : Échelle continentale, régionale ou locale (Races). Ces groupes recouvrent géographiquement une grande partie des voies de diffusion connues pour le mouton **Tableau 11**.

### 5.1.3 Approche stricte

#### 5.1.3.1 Méthode

Pour cette approche, nous avons considéré comme **spécifiques** les variants présents à fréquence non nulle dans le groupe cible et absents dans tous les autres groupes au sein du groupe de niveau supérieur (i.e., critère évalué au sein de chaque niveau : continental, régional et local). Voici trois exemples pour comprendre comment sont sélectionnés les variants spécifiques à chaque échelle :

- **Continentale** : les variants spécifiques à L'Asie seront ceux présent en fréquence non nulle dans ce continent mais absents dans tous les autres.
- **Régionale** : les variants spécifiques au Nord de la Chine seront ceux présents en fréquence non nulle dans cette région mais absents dans les autres régions du continent (Le Sud et le Plateau Tibétain).
- **Locale** : les variants spécifiques à la race Sardi seront ceux présent en fréquence non nulle dans cette race mais absents dans toutes les autres races de la région Nord-Africaine.

Les fréquences de chaque SNPs ont été calculées grâce à l'option --freq des VCFT00LS v0.1.16 (Danecek et al., 2011).

Tableau 11 : Organisation des groupes continentaux, régionaux et locaux utilisés pour l'analyse de la distribution des variants spécifiques et caractéristiques

CONTINENT	REGION		RACES	
	Nom	Code	Nom	Code
AFRIQUE	Afrique du Nord	NAF	Beni-Guil	BENI
			Sardi	SARDI
			D'man	DMAN
			Timahdite	TIMAD
			Ouled-Djellal	OUDJ
	Afrique du Sud	SAF	Mélange de races	-
ASIE	Nord de la Chine	NCHI	Bayinbuluke	BY
			Cele-black sheep	CB
			Hu	HU
			Tan	TA
			Small Tail Han sheep	STH
			Wuzhumuqin	WZ
			Altay	ALT
			Bashibai	BSB
			Oula sheep	OL
			Prairie Tibetan	PT
			Valley Tibetan	VT
	Asie du Sud	SAS	Mélange de races	-
EUROPE	Europe du Nord (voie Danubienne)	DAN	Mélange de races	-
	Europe du Sud (voie Méditerranéenne)	MED	Mélange de race	-
MOYEN-ORIENT	-	MIDE	Mélange de races	-

### 5.1.3.2. Résultats

#### Nombre de variants spécifiques par groupe (Figure 39)

Ce sont les groupes régionaux qui comptent de plus de variants spécifiques : plus de 6 millions pour l'Afrique du Nord, plus de 3 millions pour le Nord de la Chine et plus de 2 millions pour les groupes Danubien et Méditerranéen. Ces groupes comptent plus de variants spécifiques que les races qui les composent. Le groupe Plateau Tibétain fait exception avec seulement 632 124 variants spécifiques, soit environ deux fois moins que les races qui le compose. Les groupes Asie du Sud et Afrique du Sud présentent également peu de variants spécifiques (418 893 et 566 316 respectivement).

Les groupes continentaux présentent moins de variants spécifiques que les groupes régionaux qui les composent. Les groupes régionaux Afrique du Nord et Nord de la Chine comptent plus de variants spécifiques que les races qui les composent. C'est l'inverse pour la région tibétaine et les races qui la compose. L'Asie et l'Afrique comptent approximativement un million de variants spécifiques alors que l'Europe et le Moyen-Orient n'en présente environ que 300 000.

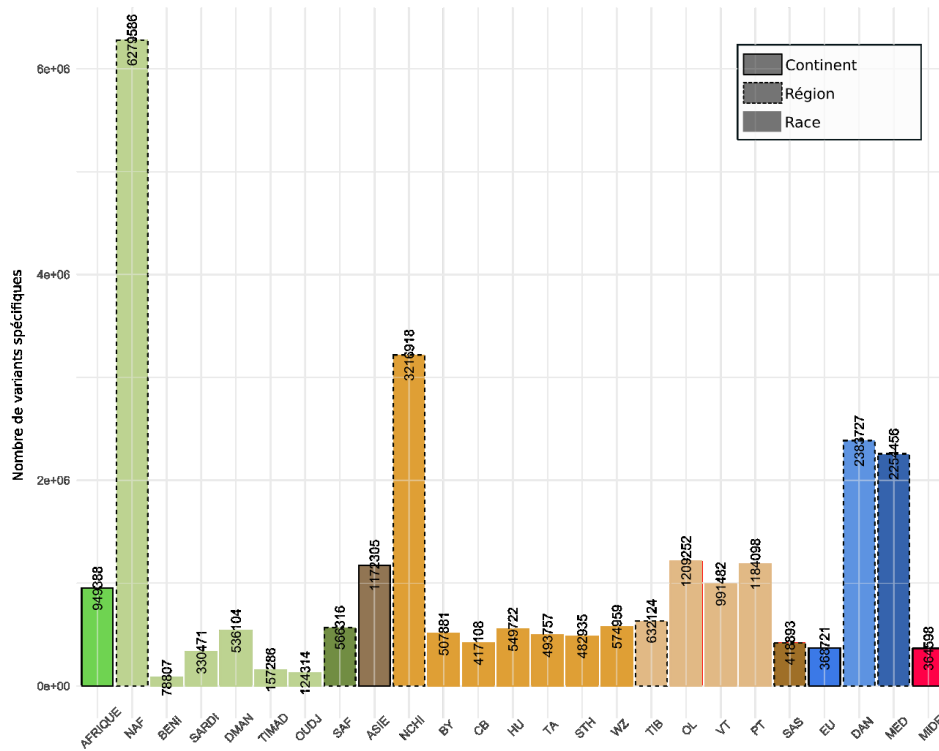


Figure 39 : Nombre de variants spécifiques par groupe au sein du groupe de niveau supérieur.

Au niveau des races, le nombre de variants spécifiques est assez hétérogène en Afrique du Nord. Les races Sardi et D'man présentent 330 471 et 536 104 variants spécifiques respectivement alors que les races Beni-Guil, Timahdite et Ouled-djellal en présentent seulement 78 807, 157 266 et 124 314 respectivement. Les races nord chinoises présentent toutes autour de 50 000 variants spécifiques. Les races tibétaines en comptent environ deux fois plus **Figure 39**.

### Fréquence des variants spécifiques

La majorité des variants spécifiques des groupes continentaux présentent des fréquences inférieures à 5 %. Plus de 75% des variants spécifiques à l'Afrique et à l'Asie ont des fréquences inférieures à 2.5% et le reste ne dépasse pas les 10%. En Europe et au Moyen-Orient, plus de 80% des variants spécifiques ont une fréquence comprise entre 1 et 5%. Le reste se situe globalement entre 5 et 10% avec quelques variants à 10-25% (**Figure 40**).

C'est au niveau des races que l'on trouve le plus de variants spécifiques en fréquence élevée. Plusieurs patrons se dessinent :

- Fréquence des variants spécifiques comprises entre 10 et 50% : Beni-Guil, Timahdite, Ouled-Djellal
- Fréquence majoritaire de 5-10%, au moins 20% de fréquence entre 10 et 25% et quelques-uns au-dessus : Sardi, Bayinbuluke, Cele black sheep
- Fréquence majoritaire de 2.5-5% : D'man, Hu, Tan, Small Tail Han, Wuzhumuqin, Oula, Prairie Tibetan, Valley Tibetan

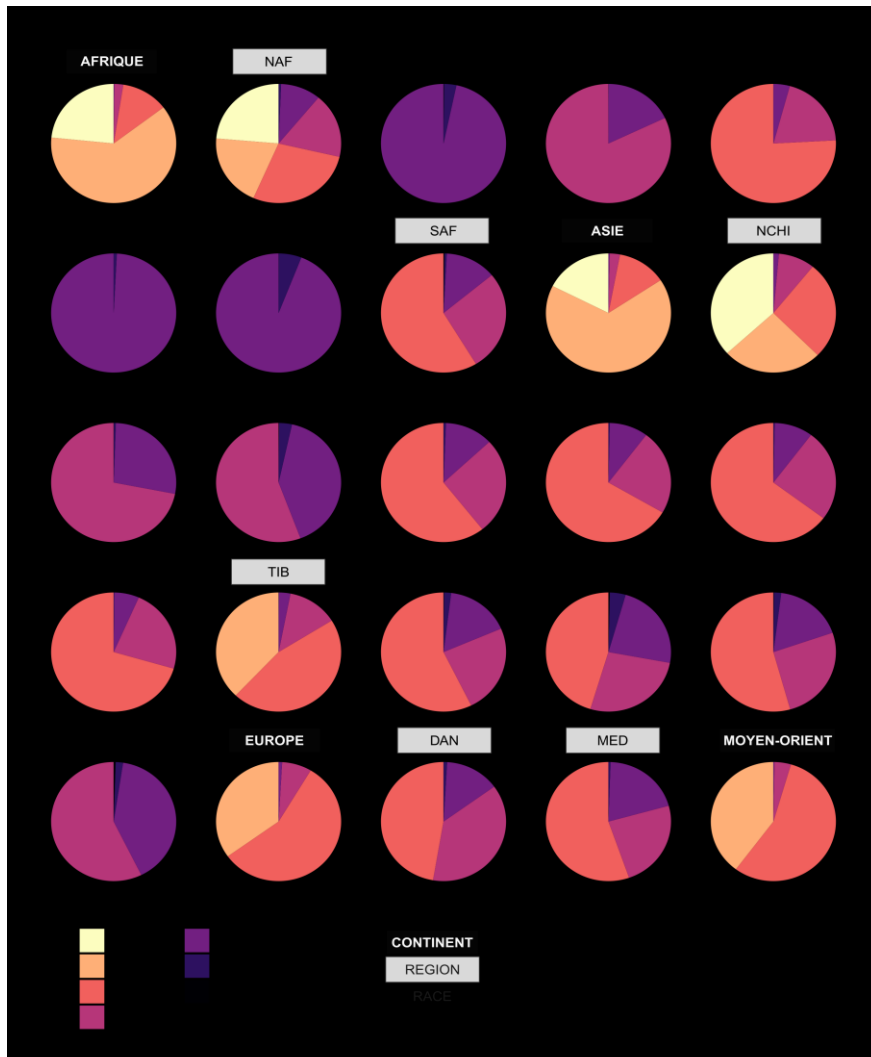


Figure 40 : Nombre de variants spécifiques par groupes et par gammes de fréquence allélique

Au niveau des régions, on retrouve les deux cas précédents. Le Nord de l’Afrique et le Nord de la Chine sont composés majoritairement de variants en fréquence inférieure à 5%. Les variants tibétains ont des fréquences majoritairement comprises entre 1 et 5%. Environ la moitié des variants des groupes sud-africain, Danubien et Méditerranéen ont des fréquences entre 2.5 et 5%, le reste correspond à des variants encore plus fréquents. Le groupe sud-asiatique, quant à lui, ne présente que des variants ayant une fréquence supérieure à 5% dont presque la moitié ont une fréquence supérieure à 10% (**Figure 40**).

## Profils de distributions des âges des variants spécifiques

Tableau 12 : Résumé statistiques des distributions des âges des variants spécifiques aux groupes continentaux, régionaux et aux races

groupe	moyenne	médiane	minimum	maximum	q25	q75
AFRIQUE	1903	1533	6	15148	855	2466
ASIE	2324	1902	8	16516	1406	2995
BENI	3651	3484	63	15148	2166	5141
BY	3934	4053	14	14197	2519	5368
CB	3798	3882	9	15479	2261	5253
DAN	4415	4818	17	18007	3060	5853
DMAN	3646	3560	6	15148	2119	5253
EU	2020	1600	16	18007	1108	2519
HU	3865	3967	12	15148	2361	5368
MED	4418	4818	16	16163	3060	5853
MIDE	2083	1635	10	15479	1182	2574
NAF	4226	4614	6	19213	2747	5728
NCHI	3995	4142	9	19213	2519	5485
OL	4902	5253	8	16516	4142	5981
OU DJ	3667	3484	64	10490	2119	5253
PT	4881	5253	10	15148	4142	5981
SAF	3351	3060	20	12204	1707	4923
SARDI	3629	3560	9	15148	2074	5253
SAS	3253	3060	18	14824	1671	4715
STH	3846	3882	17	14197	2361	5368
TA	3896	3967	13	19213	2413	5368
TIB	3240	2931	8	13306	1671	4614
TIMAD	3663	3560	12	14824	2119	5141
VT	4787	5253	16	16163	3967	5981
WZ	3966	4053	9	16516	2519	5368

On retrouve différents profils de distribution des âges des variants spécifiques : (**Figure 41** et **Tableau 12**). Les distributions des âges des variants spécifiques aux différents niveaux continentaux présentent des profils quasiment similaires. On observe une vague d'apparition de mutations qui commence aux alentours de 4-5000 générations. A partir de là, le nombre d'apparition de mutations spécifiques augmentent régulièrement et continuellement jusqu'à 1500 générations. Les moyennes de leur distribution se situent autour de 2000 générations et les troisièmes quartiles autour de 2500-3000 générations. La queue de distribution est assez longue et s'étend jusqu'à 15 à 18 000 générations. Les quatre groupes continentaux présentent un pic à environ 1500 générations. L'Afrique présente une deuxième vague d'apparition de variants spécifiques à partir de 1000 générations.

Les mutations spécifiques aux régions d'Afrique du Nord et aux régions Européennes Danubienne et Méditerranéenne commencent à s'accumuler il y a 6600 générations environ. Le nombre maximal de mutations spécifiques est atteint peu de temps après et forme un plateau jusqu'à 5000 générations environ. A partir de là, le nombre de mutation spécifiques



qui apparait diminue faiblement mais de manière continue jusqu'au présent. On devine un petit pic autour de 1500 générations.

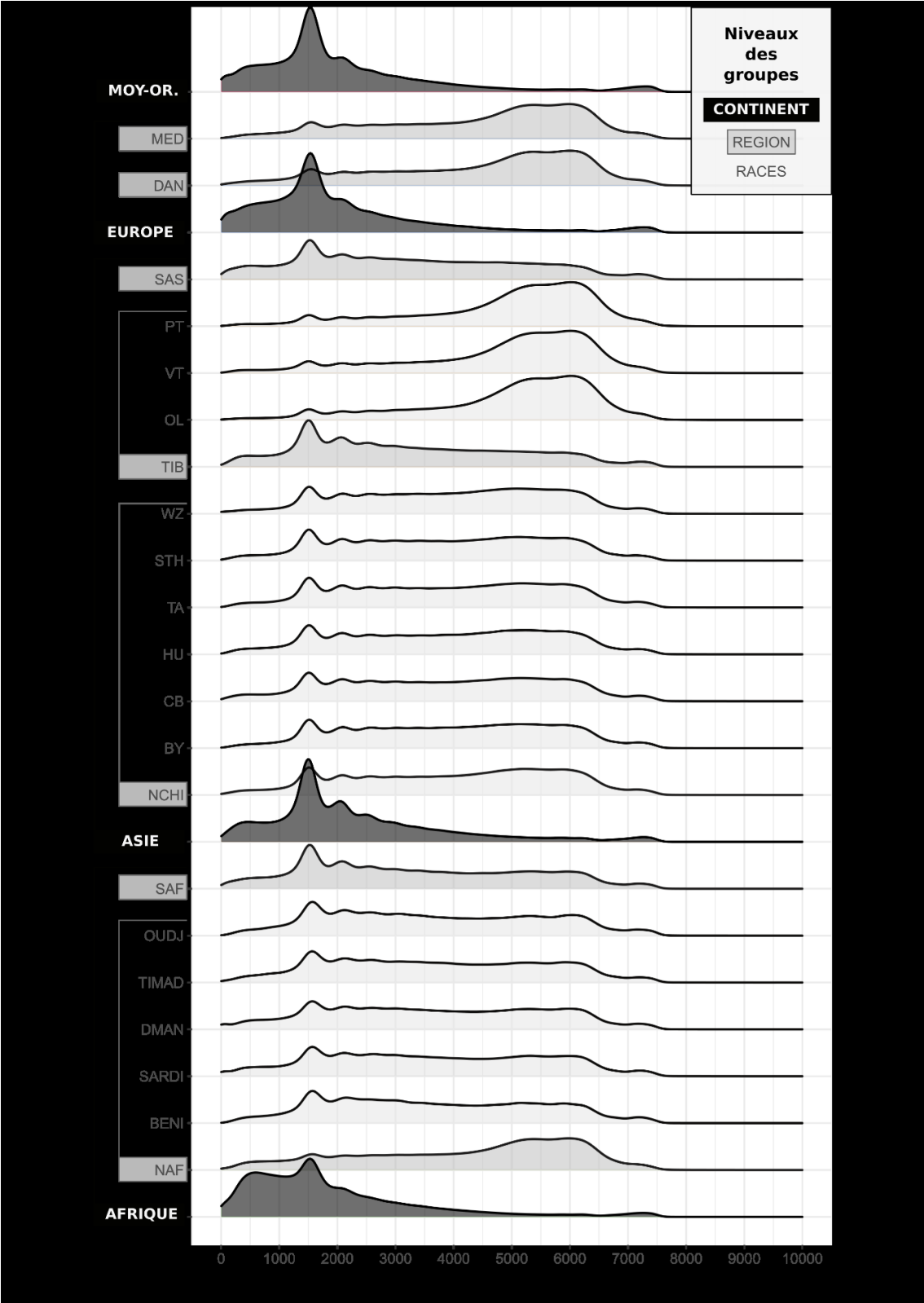


Figure 41 : Distributions des âges des variants spécifiques aux groupes continentaux, régionaux et aux races

Le groupe régional nord chinois ainsi que les groupes de races nord-africaines et nord-chinoises présentent une distribution uniforme entre 6500 et 200 générations suivie d'un petit pic autour de 1500 générations.

Les groupes régionaux tibétains et sud-asiatiques montrent un nombre d'apparition de mutation croissant légèrement entre 6500 et 2000 générations suivi par un pic marqué à 1500 générations et un léger épaulement autour de 500 générations.

Chez les groupes de races tibétaines, la majorité des variant spécifiques ont un âge situé entre 6500 et 4500 générations. Le nombre de mutation qui apparait diminue ensuite jusqu'à 2000 générations et un très léger pic à 1500 générations.

## Recouvrement

Il existe des recouvrements entre les variants spécifiques de différents niveaux. Par exemple un variant spécifique d'un continent peut aussi être spécifique d'une région et d'une race. En Afrique, Asie et Europe, 80, 89, et 78% des variants spécifiques aux continents sont aussi spécifiques d'une race ou d'une région respectivement. Nous qualifierons les 20% restant dans chaque continent de variants continentaux *partagés* car ils sont présents dans au moins deux régions ou races du continent.

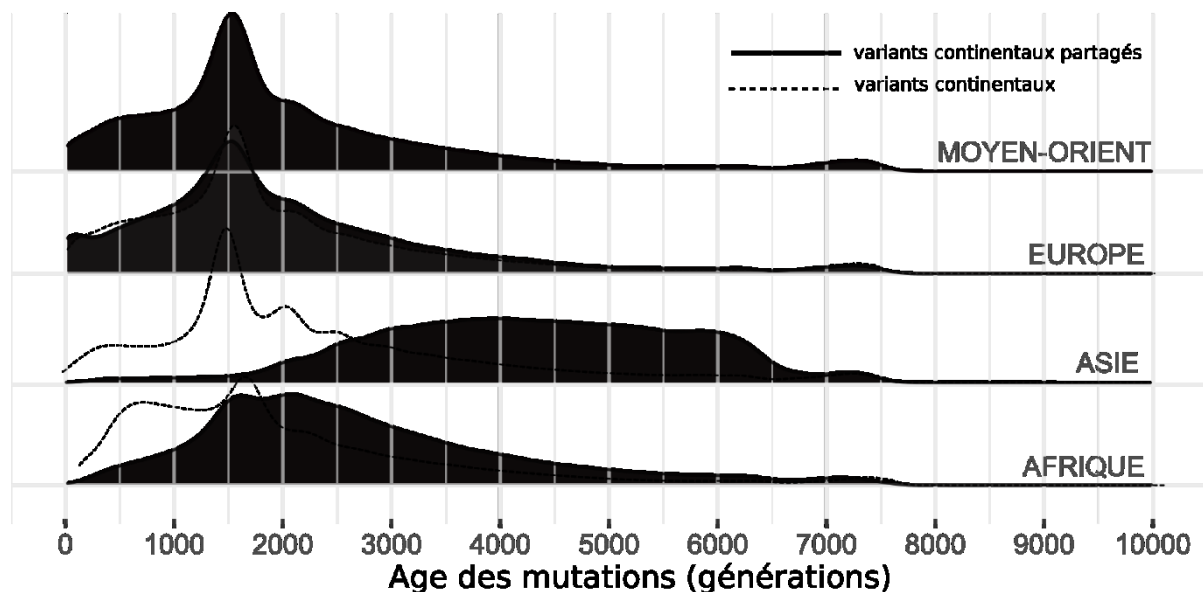


Figure 42 : Distributions comparées des âges des variants continentaux et des variants continentaux cosmopolites.

Les variants continentaux correspondent aux variant seulement présents dans un continent. Les variants continentaux cosmopolites sont les variants présent seulement dans un continent et dans au moins deux régions de ce continent.

La distribution des variants continentaux partagés reste semblable à celle de l'ensemble des variants spécifiques aux continents pour l'Europe. En Afrique, il y a peu de changements si ce n'est la diminution de variants âgés de moins de 1000 générations. Pour l'Asie, on ne retrouve quasiment aucun variant continental cosmopolite en dessous de 1500 générations (**Figure 42**). La distribution des variants continentaux aussi spécifiques à une race ou une région se concentre entre 1000 et 2000 générations.

## Distribution des âges en fonction de la fréquence des variants

De manière générale, les distributions d'âges de variants spécifiques diffèrent peu en fonction de leur fréquence. Les distributions par gamme de fréquence allélique sont très étalées (**Figure 43**), ce qui montre qu'on observe des variants de tous âges quelle que soit leur fréquence. On observe tout de même une tendance avec les variants les plus fréquents qui présentent des distributions légèrement décalées dans les temps plus anciens.

Deux cas viennent rompre cette tendance. D'abord, les variants rares (0.5-2.5%) spécifiques des groupes continentaux montrent des pics de distributions autour de 1500 générations. Ensuite, dans les groupes Africain et Nord chinois, on retrouve des variants dont la fréquence est élevée (10-25% et 25-50% respectivement) au niveau de 1500 et 500 générations respectivement (**Figure 43**). Ces catégories de fréquence sont rares (moins de 0.005%) mais représentent tout de même 31 140 variants en Afrique et 16 085 variants au Nord de la Chine.

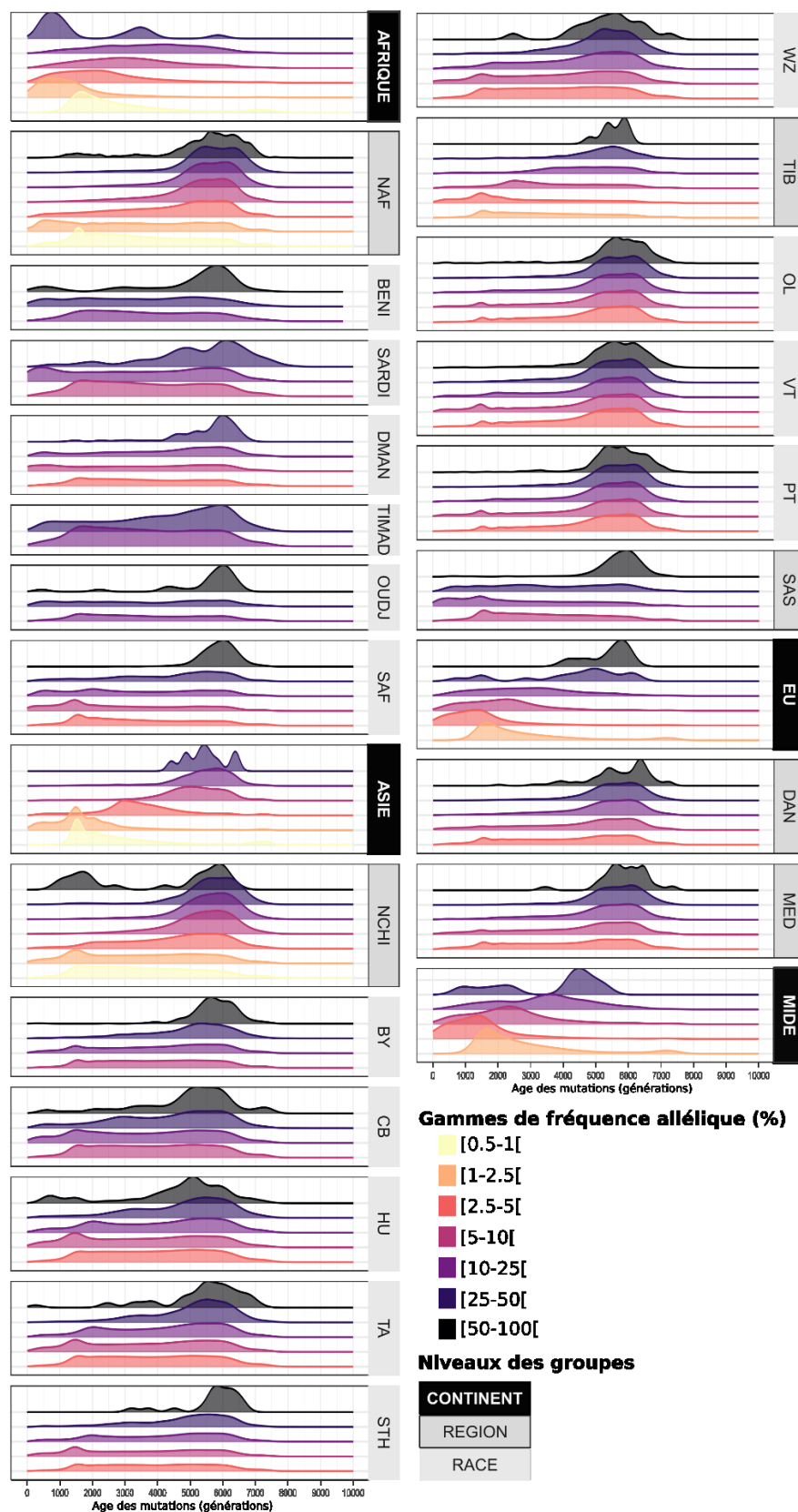


Figure 43 : Distribution des âges des variants spécifiques aux groupes continentaux, régionaux et de races divisés par gammes de fréquence alléliques

## 5.1.4 Approche FST

### 5.1.4.1. Méthode

Ici, nous avons défini un variant comme **caractéristique** d'un groupe s'il existe un différentiel de fréquence allélique fort entre le groupe cible et les autres et si la fréquence du variant est plus élevée dans le groupe cible que dans tous les autres. La différenciation génétique peut se mesurer à l'aide du FST. Il existe plusieurs variantes de cet indice. Nous avons choisi celui de Weir et Cockerham (Weir and Cockerham, 1984).

Nous avons calculé cet indice pour chaque variant entre les différents groupes et à différents niveaux.

- Au **niveau continental**, nous avons contrasté le continent cible avec l'ensemble des autres
- Au **niveau régional**, nous avons contrasté la région cible avec l'ensemble des régions du continent.
- Au **niveau local**, nous avons contrasté la race cible avec l'ensemble des races de la région.

Nous avons ensuite sélectionné, pour chaque groupe, les 10 000 variants dont le FST était le plus fort et dont la fréquence était supérieure dans le groupe cible. Nous avons divisé les variants caractéristiques ainsi sélectionnés en quatre lots correspondant à des intervalles de FST allant de 0.2 à 1 par pas de 0.2.

### 5.1.4.2. Résultats

#### Distribution des FST au sein des variants caractéristiques

Les groupes présentant les FST les plus faibles (compris entre 0.2 et 0.4 environ) sont l'Afrique, le Moyen-Orient, les D'man, l'Asie et le Nord de la Chine. Ceux qui présentent les FST les plus importants (supérieurs à 0.7) sont les Beni-Guil, les Ouled-Djellal, et les Sud-asiatiques. Les autres groupes se situent entre ces deux extrêmes (**Figure 44**).

#### Recouvrement

Il n'y a que très peu de recouvrement entre les variants caractéristiques des différents groupes et la plupart sont intra-continentaux ou intra-régionaux. On note quand même l'existence d'un partage de variants relativement important entre de Sud de l'Asie et le Sud de l'Afrique et entre l'Europe et le sud de l'Afrique (**Figure 45**).

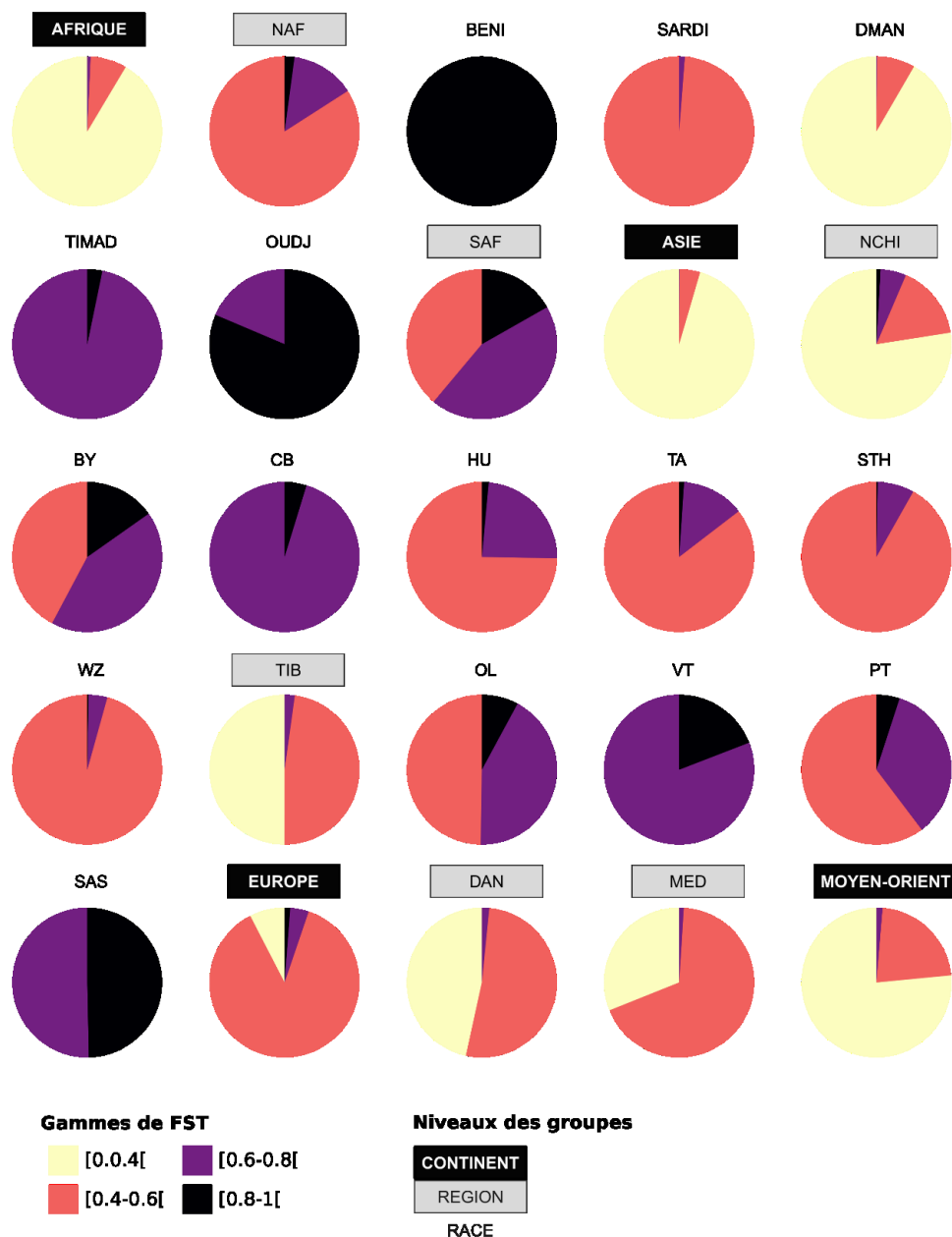


Figure 44 : Nombre de variant caractéristiques par gamme de FST et par groupe continental, régionaux et de races.

### Distribution des âges des variants caractéristiques

Au niveau continental, plus des trois quarts des variants caractéristiques ont un FST compris entre 0.2 et 0.4 sauf en Europe où plus de 80% des FST sont supérieurs à 0.4. Ces derniers présentent des âges similaires avec des distributions centrées sur 5500 générations. On ne compte que peu de variants caractéristiques dans les gammes de FST plus élevées pour ces groupes mais c'est à ce niveau que l'on trouve des différences (**Figure 46, Tableau 13**): En Afrique, 70 variants caractéristiques ayant un FST compris entre 0.6 et 0.8 se répartissent dans deux gammes d'âges. La première récente, autour de 2000 générations et la seconde plus

ancienne, autour de 6500 générations. En Asie on observe la même chose mais uniquement pour 9 variants pour cette gamme de FST.

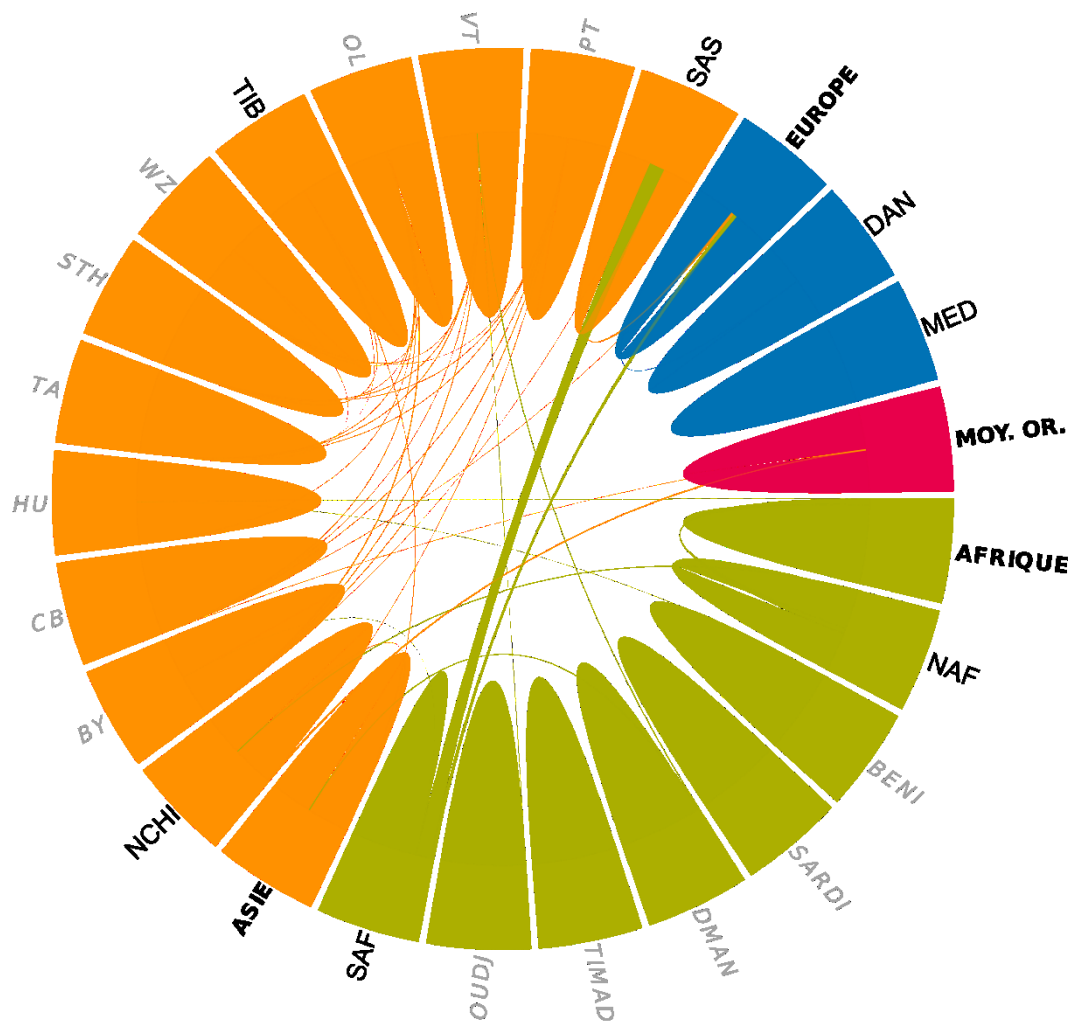


Figure 45 : Recouvrement entre les variants caractéristiques des groupes continentaux, régionaux et locaux

Dans les régions européennes (Danubienne et Méditerranéenne), quelle que soit la gamme de FST considérée, les distributions d'âge sont centrées sur 6000 générations environ.

En Asie, sauf exception, on retrouve des distributions d'âge centrées sur 6000 générations environ également, quels que soient le groupe ou la gamme de FST. L'exception se trouve dans le groupe nord-chinois. Il présente 89 variants à FST supérieur à 0.8 distribués plus récemment entre 2000 et 5000 générations. Il contient également 553 variants ayant des FST compris entre 0.6 et 0.8 et dont la distribution présente un pic à 6000 générations et un autre plus récent autour de 1000 générations.

En Afrique, la tendance générale est la même avec la majorité des distributions centrées sur environ 6000 générations. Les différences se situent dans les gammes de FST les plus élevées : Dans la région nord-africaine 211 variants à FST supérieur à 0.8 présentent des âges plus récents à environ 3500 et 2000 générations. Dans les races Beni-guil et Ouled-Djellal, pour des FST supérieurs à 0.8 et pour les Timahdites pour des FST entre 0.6 et 0.8, on observe des distributions qui s'étalent dans les temps très récents.

**Tableau 13 : Résumés statistiques des âges des variants caractéristiques.**

Les correspondances entre les gammes de FST et leurs valeurs dans la colonne FST sont : 1 : FST [0.2-0.4[, 2 : FST [0.4-0.6[, 3 : FST [0.6-0.8[, 4 : FST [0.8-1[.

GROUPE	Fst	Nombre	Moyenne	Médiane	Minimum	Maximum	q25	q75
AFRIQUE	1	9145	5048	5368	428	13596	4419	5981
AFRIQUE	2	785	4623	4923	467	11438	3560	5853
AFRIQUE	3	70	4419	5369	704	8094	2119	6245
NAF	2	8408	5586	5728	784	13596	5253	6245
NAF	3	1381	5901	6245	674	8094	5728	6381
NAF	4	211	3398	3410	531	6957	2030	4097
BENI	4	10000	4424	4923	70	7921	3196	5853
SARDI	2	9878	4768	5253	173	8094	3967	5981
SARDI	3	122	5189	5253	2074	7109	4516	5981
DMAN	1	9165	5278	5485	142	13021	4818	6111
DMAN	2	821	5161	5368	632	7586	4715	5981
DMAN	3	14	4098	4525	520	6663	2316	6211
TIMAD	3	9677	3928	4419	122	10266	2261	5605
TIMAD	4	323	4967	5141	579	7424	4325	5853
OUDJ	3	1861	5233	5485	229	7586	4818	5981
OUDJ	4	8139	3723	3967	64	7921	1944	5485
SAF	2	3891	4957	5368	310	7751	4232	5981
SAF	3	4432	5308	5605	205	13021	4818	6245
SAF	4	1677	5919	5981	316	7751	5605	6245
ASIE	1	9548	5450	5605	316	8825	5031	6111
ASIE	2	443	5301	5485	646	13021	4767	6111
ASIE	3	9	4399	5728	1671	7265	2074	6111
NCHI	1	7748	5540	5605	531	10490	5141	6245
NCHI	2	1610	4681	5031	428	7586	3967	5605
NCHI	3	553	3618	4419	316	6809	1289	5253
NCHI	4	89	3441	3410	1346	5368	2631	4232
BY	2	4224	5576	5728	316	8636	5141	6245
BY	3	4254	5730	5853	447	11438	5253	6381
BY	4	1522	5900	5853	201	12470	5485	6381
CB	3	9527	5191	5485	169	11193	4715	6111
CB	4	473	5189	5368	249	8451	4715	5981
HU	2	7472	5214	5485	255	8451	4715	6111
HU	3	2382	5438	5605	177	7921	4923	6245



HU	4	146	4594	4923	509	7586	4053	5697
TA	2	8543	5448	5605	173	9215	5031	6245
TA	3	1352	5667	5853	330	9017	5253	6245
TA	4	105	5824	5981	255	7751	5485	6381
STH	2	9184	5274	5485	214	8636	4818	6111
STH	3	788	5412	5605	419	7586	5031	6111
STH	4	28	5787	5981	3196	7109	5728	6279
WZ	2	9569	5467	5605	169	9416	5031	6245
WZ	3	405	5559	5728	1208	7751	5141	6245
WZ	4	26	5713	5666	4142	6957	5169	6111
TIB	1	4998	5581	5728	410	7751	5141	6245
TIB	2	4784	5315	5605	498	9416	4923	6111
TIB	3	217	4755	5253	819	9622	4053	5728
TIB	4	1	874	874	874	874	874	874
OL	2	4976	5781	5853	498	8094	5368	6381
OL	3	4236	5780	5853	284	8094	5368	6381
OL	4	788	5808	5981	353	7921	5485	6381
VT	3	8088	5727	5853	393	12204	5253	6381
VT	4	1912	5461	5728	520	8636	5141	6381
PT	2	6030	5743	5853	579	8271	5253	6381
PT	3	3470	5774	5853	509	8451	5368	6381
PT	4	500	5835	5853	2361	7751	5456	6381
SAS	3	5032	4751	5141	224	8451	3882	5981
SAS	4	4968	5235	5605	290	7751	4923	6111
EU	1	757	4707	4923	509	7751	3967	5728
EU	2	8722	4831	5031	393	12470	4232	5728
EU	3	401	5039	5368	660	7265	4614	5981
EU	4	120	5893	5981	1822	7265	5605	6245
DAN	1	4658	5834	5853	719	7921	5368	6381
DAN	2	5183	5857	5981	660	8271	5485	6521
DAN	3	159	6190	6245	4818	7424	5853	6663
MED	1	3107	5782	5981	1108	7921	5368	6381
MED	2	6794	5831	5981	973	9017	5368	6381
MED	3	99	6183	6245	3337	7424	5981	6521
MIDE	1	7653	5058	5253	297	7921	4614	5853
MIDE	2	2215	4818	5141	330	7424	4325	5605
MIDE	3	131	4306	4614	704	6663	3967	5253
MIDE	4	1	855	855	855	855	855	855

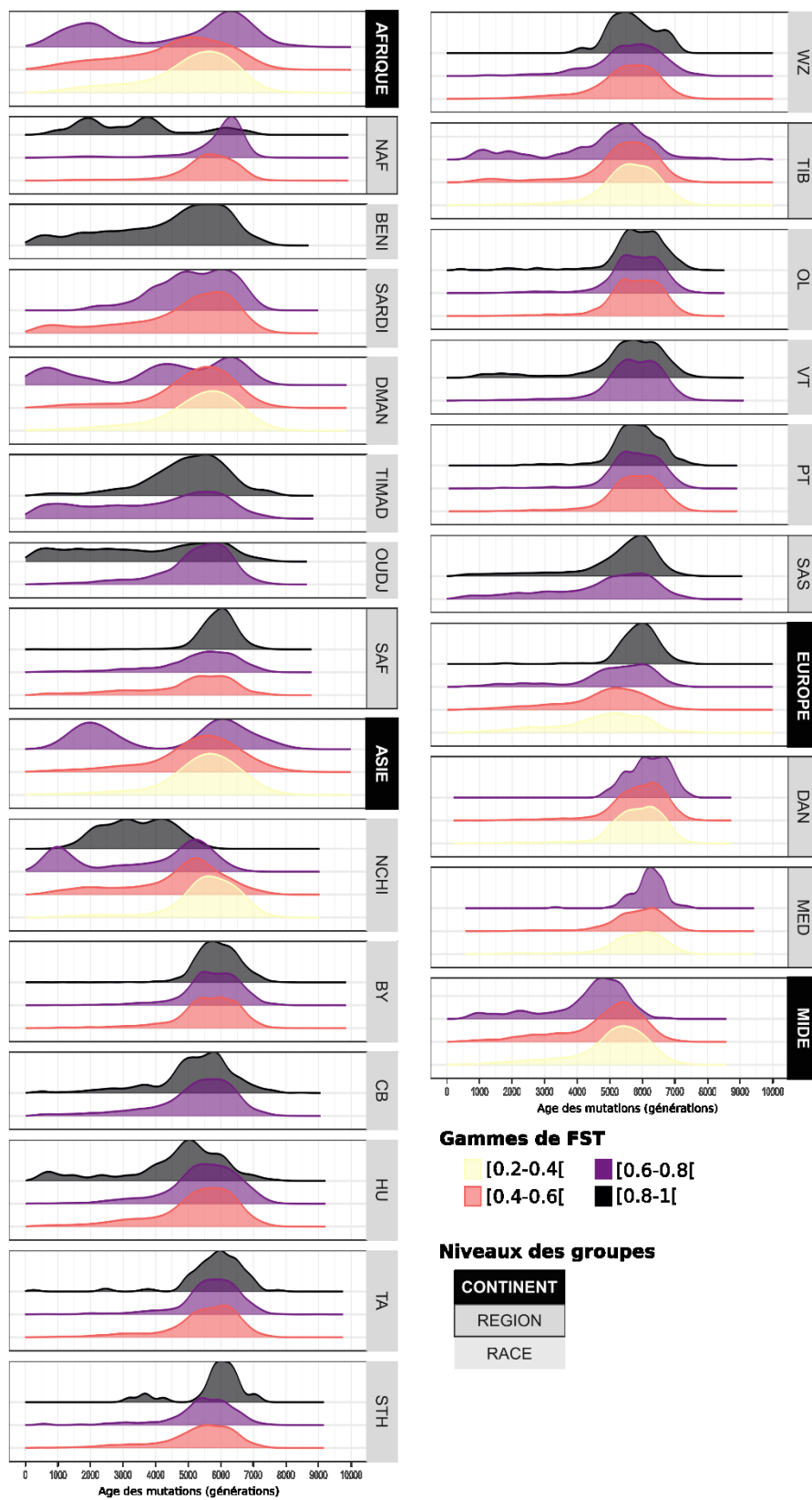


Figure 46 : Distribution des âges des variants caractéristiques pour chaque groupe continental, régional et de race et par gamme de FST.

### 5.1.5 Comparaison des approches stricte et FST

La proportion de variants spécifiques qui sont aussi caractéristiques ne dépasse pas 6% quel que soit le groupe. En revanche les proportions de variants caractéristiques qui sont aussi spécifiques sont globalement plus élevées (25% en moyenne) et varient davantage. Les proportions de variants caractéristiques qui sont aussi spécifiques ne dépassent pas 5% pour le niveau continental. Au niveau régional, moins de 15% des variants caractéristiques des groupes Afrique du Nord, Plateau Tibétain, Danubien et Méditerranéen sont aussi spécifiques. Pour l’Afrique du Sud et l’Asie du Sud, au contraire, plus de 40% des variants caractéristiques sont aussi spécifiques. Au niveau local, la proportion de variants caractéristiques aussi spécifiques varie de 19 à 73%.

**Tableau 14 : Recouvrement entre les variants spécifiques et caractéristiques**

Variants à la fois spécifiques et caractéristiques de chaque groupe représenté en nombre et en pourcentage du nombre total de variants caractéristiques ou spécifiques de chaque groupe. Les lignes grises foncées correspondent au niveau continental, les grises claires au niveau régional et les blanches au niveau local.

Groupe	Variants caractéristiques ET spécifiques		
	Nombre	% variants caractéristiques	% variants spécifiques
<b>AFRIQUE</b>	71	0,7	0,01
<b>NAF</b>	693	6,9	0,01
<b>BENI</b>	2651	26,5	3,36
<b>SARDI</b>	4286	42,9	1,30
<b>DMAN</b>	3019	30,2	0,56
<b>TIMAD</b>	6385	63,9	4,06
<b>OUDJ</b>	7369	73,7	5,93
<b>SAF</b>	4268	42,7	0,75
<b>ASIE</b>	25	0,3	0,00
<b>NCHI</b>	223	2,2	0,01
<b>BY</b>	1958	19,6	0,39
<b>CB</b>	3374	33,7	0,81
<b>HU</b>	3365	33,7	0,61
<b>TA</b>	2963	29,6	0,60
<b>STH</b>	3612	36,1	0,75
<b>WZ</b>	2808	28,1	0,49
<b>TIB</b>	461	4,6	0,07
<b>OL</b>	2232	22,3	0,18
<b>VT</b>	2570	25,7	0,26
<b>PT</b>	3413	34,1	0,29
<b>SAS</b>	4823	48,2	1,15
<b>EU</b>	464	4,6	0,13
<b>DAN</b>	910	9,1	0,04
<b>MED</b>	1431	14,3	0,06
<b>MIDE</b>	461	4,6	0,13

Les distributions des âges des variants caractéristiques sont globalement plus anciennes que celles des variants spécifiques avec très peu de variants récents sauf pour les

groupes Beni-Guil, Timahdite, et Nord de la Chine. Les différences les plus marquées sont au niveau continental : Les distributions des variants caractéristiques et spécifiques pour les groupes continentaux sont centrées autour de 6000 et autour de 1500 générations respectivement.

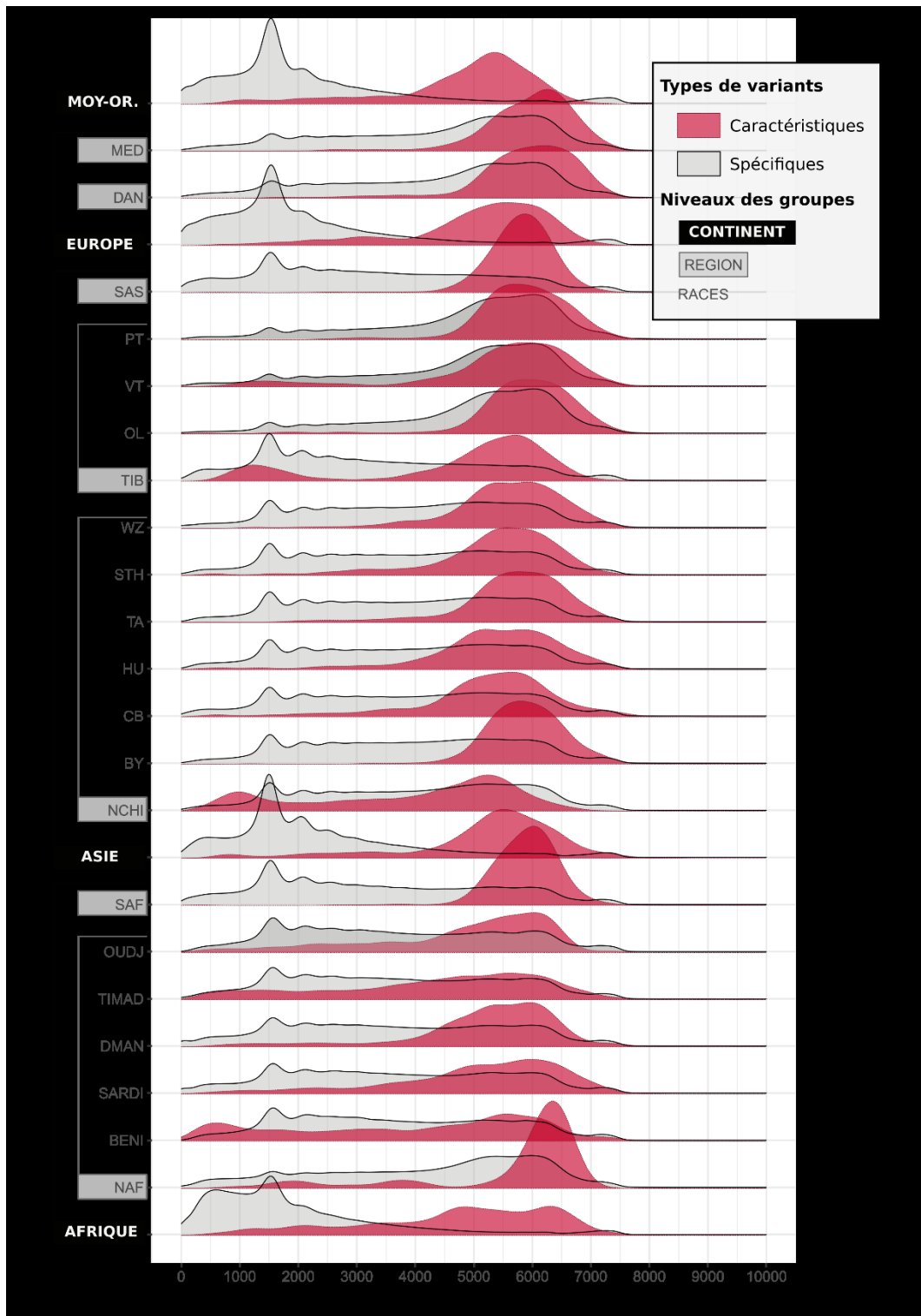


Figure 47 : Comparaison des distributions des âges des variants caractéristiques et spécifiques pour chaque groupe continental, régional et local.

## 5.2. Connexions des populations au cours du temps

### 5.2.1. Objectif

La structure génétique des moutons est globalement faible (voir **Introduction générale**), et il a déjà été montré qu'il y a eu beaucoup de flux de gènes au cours de leur histoire (Kijas et al., 2012a). C'est pourquoi nous allons tenter de retracer les connexions qui ont eu lieu entre les différentes populations de moutons au cours du temps. Grâce à la datation des plusieurs millions de mutations dans le génome des moutons nous pouvons comparer les partages de variants entre différents groupes au cours du temps.

Tableau 15 : Individus choisis pour constituer les génomes cibles nécessaires à l'analyse CCF

Haplotype cible	Groupe		
	Continent	Région	Race
CNOA-BY1	Asie	Nord Chine	Bayinbuluke
CNOA-BY5	Asie	Nord Chine	Bayinbuluke
CNOA-CB10	Asie	Nord Chine	Cele Black Sheep
CNOA-CB2	Asie	Nord Chine	Cele Black Sheep
CNOA-HU7	Asie	Nord Chine	Hu
CNOA-HU9	Asie	Nord Chine	Hu
CNOA-OL1	Asie	Plateau Tibétain	Oula sheep
CNOA-OL2	Asie	Plateau Tibétain	Oula sheep
CNOA-PT3	Asie	Plateau Tibétain	Prairie Tibetan
CNOA-PT6	Asie	Plateau Tibétain	Prairie Tibetan
CNOA-STH5	Asie	Nord Chine	Small Tail Han sheep
CNOA-STH8	Asie	Nord Chine	Small Tail Han sheep
CNOA-TA1	Asie	Nord Chine	Tan
CNOA-TA5	Asie	Nord Chine	Tan
CNOA-VT10	Asie	Plateau Tibétain	Valley Tibetan
CNOA-VT8	Asie	Plateau Tibétain	Valley Tibetan
CNOA-WZ3	Asie	Nord Chine	Wuzhumuqin
CNOA-WZ7	Asie	Nord Chine	Wuzhumuqin
MOOA-O13-0367	Afrique	Afrique Nord	Sardi
MOOA-P14-1198	Afrique	Afrique Nord	D'man
MOOA-R13-1131	Afrique	Afrique Nord	D'man
MOOA-R6-3022	Afrique	Afrique Nord	-
MOOA-S9-0210	Afrique	Afrique Nord	Timahdite
OARI_AWT1	Moyen-Orient	-	Awassi
OARI_FIN1	Europe	Danube	Finnsheep
OARI_KRS5	Moyen-Orient	-	Karakul
OARI_MERC1	Europe	Méditerranée	Merino
OARI_OJA4	Europe	Méditerranée	Ojalada
OARI_SBF454	Europe	Danube	Scottish Blackface
OARI_VBS2	Europe	Danube	Valais Blacknose

## 5.2.2. Méthode

La méthode employée pour étudier les connexions des populations au cours du temps permet d'inférer la fonction de coalescence cumulative (CCF). Cette fonction exprime la fraction d'un génome (cible) qui a coalescé avec un autre génome (comparateur) à un moment donné du passé (Albers and McVean, 2020, sup. text). Cette méthode est disponible en ligne (<https://github.com/pkalbers/ccf>). Nous avons choisi de comparer au moins deux haplotypes cibles par groupe (**Tableau 15**) à ceux de tous les autres individus. Dans tous les cas, cible, ou comparateur, nous n'avons pris qu'un haplotype sur deux, au hasard. Les cibles ont été choisies pour représenter au mieux les différents groupes du jeu de données. Nous avons ainsi sélectionné deux individus de chaque race en Asie, trois races différentes en Afrique, deux races au Moyen-Orient et cinq en Europe dont trois originaires du Nord (voie Danubienne) et deux du Sud (voie Méditerranéenne). Nous avons donc exécuté 8432 comparaisons au total. L'analyse sur l'ensemble du génome aurait été trop longue, c'est pourquoi nous avons choisi de ne l'exécuter que sur le chromosome 1. C'est le plus grand et donc celui qui contient le plus d'information et comme tous les chromosomes présentent des distributions d'âges similaires nous n'attendons pas d'effet spécifique à chacun d'entre eux.

## 5.2.3. Résultats

Pour tous les génomes cibles on observe une évolution de la CCF commune entre cibles domestiques et comparateurs domestiques et une évolution différente entre les cibles domestiques et les comparateurs sauvages :

### 5.2.3.1. Evolution de la CCF entre domestiques et sauvages :

Quand on compare un génome cible de mouton domestique avec un génome comparateur de mouflon Asiatique, on obtient une courbe exponentielle croissante. La fraction de génome partagée décolle aux alentours de 2000 générations et augmente ensuite de manière continue jusqu'à environ 5100 générations. A partir de ce moment-là, elle marque un pallier jusqu'à 7500 générations environ et augmente finalement brusquement. Notons que les variants datés au-delà de 7000 générations sont rares et donc que les résultats de cette analyse pour cette période ne seront pas pris en compte dans l'interprétation (**Figure 48** et **Figure S1**).

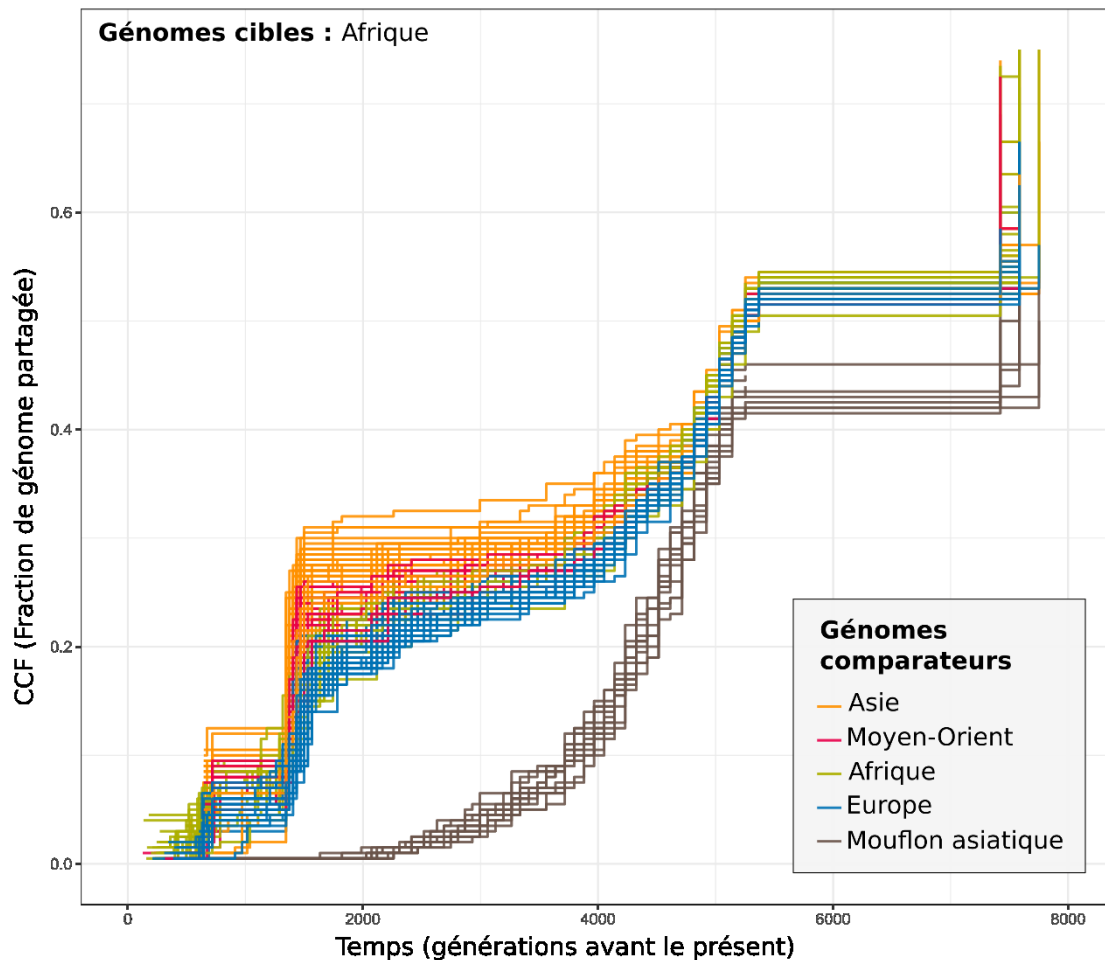


Figure 48 : Evolution de la CCF (Cumulative Coalescent Function), représentant la fraction de génome partagée au cours du temps entre cinq génomes cibles africains (Tableau 15) et les génomes comparateurs issus du monde entier.

Chaque ligne représente la CCF entre un comparateur et un génome cible.

### 5.2.3.2. Evolution de la CCF entre domestiques:

La comparaison de génomes cibles et comparateurs domestiques donne des profils similaires. Ces profils sont marqués par des variations brutales de la CCF en très peu de temps (**Figure 48** et **Figure S1**). On peut discerner quatre grandes phases :

- **Phase 1** (0 à 700 générations environ) : la CCF est faible (inférieure à 0.1)
- **Phase 2** (700 à 1500 générations environ) : la CCF reste faible mais on discerne une augmentation à environ 0.2 pour certains groupes comparateurs.
- **Phase 3** : (1500 à 4700 générations environ) : la CCF se situe autour de 0.3
- **Phase 4** (4700 à 5500 générations environ) : la CCF augmente régulièrement tout au long de la période.
- **Phase 5** (au-delà de 5500 générations environ) : la CCF atteint 0.5

Comme pour les paires de cibles domestiques et comparateur sauvage, la CCF atteint 1 au-delà de 7000 générations, mais avec peu de variant datés dans cette période les résultats ne sont pas considérés dans l'interprétation.

## 6 Discussion

---

### 6.1. Diffusion

---

La majorité des âges des variants présents dans le génome du mouton sont compris entre 4500 et 7000 générations soit 9 000 et 14 000 ans si l'on applique un temps de génération de deux ans. Cela correspond approximativement au début de la domestication. D'après plusieurs études génétiques, (Alberto et al., 2018; Frantz et al., 2020; Taberlet et al., 2008, 2011) il semble que les moutons n'aient pas subi un fort goulot d'étranglement au début de la domestication. Ce fut un processus à large échelle géographique impliquant la capture d'une diversité génétique importante (voir Article en **Introduction générale, partie 0**). Cela pourrait s'expliquer par la distribution restreinte des ancêtres des moutons. Dans ce cas, seul un processus de domestication impliquant de nombreux individus issus d'une aire géographique large peut permettre aux populations domestiques de se maintenir et de ne pas subir de dépression de consanguinité lors de leur dispersion en dehors du centre de domestication (Taberlet et al., 2011). De plus, les flux de gènes entre sauvages et domestiques et entre populations domestiques ont été importants durant les premières étapes de la domestication (Frantz et al., 2020; Hare, 2021). Les premières populations domestiques présentaient donc une taille efficace importante et par conséquent une diversité génétique élevée. Alberto et al. (2018) ont inféré, avec les données utilisées pour cette thèse, des tailles efficaces d'environ 20 000 à cette période. Or, plus la taille efficace d'une population est grande plus le nombre de mutations qui apparaît à chaque génération est élevé. Cela pourrait expliquer pourquoi on observe tant de variants datant des premiers stades de la domestication.

Ces résultats pourraient aussi indiquer que les premiers stades ne sont pas caractérisés par une sélection humaine intensive. Ainsi la taille efficace des premières populations domestiques a pu se maintenir à un niveau permettant l'apparition de nombreuses mutations. D'autres part, même si cela est sans doute quantitativement négligeable, le relâchement de la sélection pour certains caractères clés chez les sauvages, comme ceux liés à la prédation, a pu permettre l'accumulation de mutations délétères dans le génome (Moyers, Morrell and McKay, 2018). Pour le confirmer nous envisageons de comparer les âges des mutations synonymes et non-synonymes et d'étudier l'âge d'apparition du fardeau génétique de différents groupes, notamment ceux élevés traditionnellement ou de manière plus industrielle. Il a en effet été montré que l'accumulation de mutation délétères est différente dans ces deux types d'élevage (Alberto et al., 2018). Il a aussi été montré que ce type de mutations s'accumulent au niveau des fronts d'expansion (Peischl et al., 2013) ce qui a pu se produire



dans cette période initiale de la domestication étant donné la dispersion précoce des moutons hors du centre de domestication (Vigne et al., 2012).

Au contraire, durant les étapes plus tardives, la sélection de certains traits (par exemple agronomiques) et des premières races a diminué fortement la taille efficace des populations ovines domestiques. L'analyse des variants spécifiques montre que plusieurs dizaines de milliers de mutations ayant des fréquences élevées sont apparues il y a moins de 1000 générations en Afrique et au Nord de la Chine. Nous pouvons supposer que ces variants ont été soumis à des pressions de sélection pour atteindre de telles gammes de fréquence en une période relativement courte. Cette hypothèse devra être confirmée par des analyses de détection de sélection et par l'étude de la localisation des mutations concernées dans le génome.

L'étude des variants spécifiques aux différents continents où ont été introduits les moutons nous apporte aussi des informations sur cette dispersion. L'accumulation de ces variants forme une vague progressive depuis 8 à 10000 ans AP et atteint son maximum autour de 3000 ans AP. Elle reflèterait donc l'isolement progressif des populations ovines au cours de leur conquête du monde depuis le centre de domestication. Nous n'observons pas le même résultat pour les variants caractéristiques des continents. Alors que les variants spécifiques aux continents se concentrent autour de 3000 ans, les variants caractéristiques sont eux bien plus anciens, autour de 6000 générations environ. L'approche stricte suppose l'absence de la mutation dans tous les autres continents. L'approche FST impose seulement un différentiel de fréquence important. La première permet donc d'identifier des variants qui ont diffusé ou ne sont apparus que dans un continent tandis que la deuxième met en évidence des évolutions de fréquence différentes dans chacun des continents. L'approche stricte montre ainsi que de nombreuses mutations sont apparues après la séparation des populations ovines dans différents continents. L'approche FST indique, elle, que les mutations qui caractérisent les continents aujourd'hui sont un échantillonnage de mutations apparues tout au long de l'histoire de la domestication du mouton car leur distribution est semblable à celle de l'ensemble des mutations du génome.

Environ 80% des variants spécifiques aux continents sont aussi spécifiques d'une race ou d'une région. Ils ne sont donc pas présents partout dans le continent. Leur âge est compris entre 1000 et 2000 générations. Les 20% restants sont partagés par plusieurs régions du continent considéré. Ils caractérisent donc la diversité génétique capturée au début de la diffusion et donc avant la séparation des différentes populations. Ces variants continentaux partagés présentent des distributions différentes à l'Est et l'Ouest du centre de domestication. En Europe et en Afrique, les trois quarts apparaissent à partir de 3000 générations environ. En Asie, ils sont en moyenne plus vieux et leur âge varie de 8000 à 2000 générations environ. De plus, l'absence de variants anciens spécifiques aux continents occidentaux montre aussi indirectement que ces variants sont partagés entre ces continents. Il semble donc qu'il y ait eu une séparation précoce des groupes à l'ouest (Afrique, Europe, Moyen-Orient) et à l'est

(Asie) du centre de domestication. Nous ne pouvons cependant pas exclure que ces différences entre continents soient dues à l'hétérogénéité de l'échantillonnage. Nous n'avons en effet pas suffisamment d'individus pour construire des groupes de races en Europe et au Moyen-Orient. En Afrique, nous avons une surreprésentation des races du Nord. Ainsi, il est possible qu'avec un échantillonnage plus étendu on retrouve le même phénomène qu'en Asie pour les autres continents.

Une autre différence entre les groupes à l'Est et à l'Ouest du centre de domestication réside dans les distributions des âges des variants spécifiques aux continents comparées à celles des régions qui les composent. En Europe, les variants spécifiques aux régions danubienne et méditerranéenne commencent à s'accumuler dès 7000 générations environ. Les variants européens partagés entre ces deux régions ont un âge moyen bien plus récent (entre 1000 et 2000 générations). Ainsi, il semble que l'isolement des populations de ces deux régions soit antérieur à l'établissement d'une population européenne globale. Ce résultat est attendu puisque l'on sait que la diffusion en Europe s'est faite suivant deux voies qui ne se sont rejointes que plus tard il y a environ 7000 ans (Tresset and Vigne, 2007). Pour chacune de ces régions, les variants spécifiques s'accumulent continuellement de 7000 générations à aujourd'hui. Cependant, plus de la moitié ont un âge compris entre 4500 et 7000 générations, soit 9000 à 14000 ans AP. La domestication ayant eu lieu il y a 10 500 ans environ, on peut penser que ces variants anciens sont le résultat de la capture de diversité initiale au niveau du centre de domestication. Le reste des variants, plus récents, s'est sûrement accumulé au fil de la dispersion. On peut tenir le même raisonnement pour les variants spécifiques d'Afrique du Nord et d'Afrique du Sud, qui s'accumulent bien avant ceux communs à tout le continent. La diffusion des moutons en Afrique a probablement suivi deux routes distinctes, l'une passant par le nord et longeant la côte Atlantique, l'autre suivant la côte Est du continent avec possiblement une entrée par la péninsule Arabique. Ces routes ont pu se rejoindre il y a 2500 ans environ au Sud, date à laquelle on trouve les plus anciens restes ovins dans cette région (Muigai and Hanotte, 2013) (voir **Introduction générale, partie 1**).

Curieusement, le Sud de l'Afrique présente environ 10% de variants caractéristiques communs avec ceux du Sud de l'Asie. Cela implique que ces variants qui distinguent le nord et le sud de ces deux continents sont identiques. Ce résultat vient étayer une hypothèse formulée précédemment sur la diffusion des moutons à l'Est et au Sud de l'Afrique qui aurait pu se faire directement par la péninsule Arabique (Boivin and Fuller, 2009). En effet un important port de commerce de la vallée de l'Indus, le port de Lothal, connectait l'actuelle Inde avec la péninsule arabique et l'ouest de l'Afrique (McIntosh, 2007). Cela expliquerait un partage de variants entre l'Asie du Sud et l'Afrique du Sud dont une des voies de colonisation est la côte Ouest de l'Afrique. Une étude précédente a d'ailleurs montré que les actuels moutons indiens de la région du port de Lothal présentaient une haute fréquence d'haplogroupe mitochondrial B, un haplogroupe typique de l'Afrique et de l'Europe mais très peu présent en Asie où le A prédomine (Singh, Kumar and Kolte, 2013). Les échanges de moutons devaient donc être importants entre les deux continents.

## 6.2. Connexions entre les groupes au cours du temps

---

Le partage haplotypique entre des génomes cibles domestiques et des comparateurs sauvages décrit une courbe exponentielle croissante qui se distingue de celle correspondant à la comparaison de paires de génomes domestiques. Dans ce cas, la fraction de génome partagée augmente plus rapidement et par pallier du présent vers le passé. La séparation des trajectoires de partages haplotypiques domestique-sauvage et domestique-domestiques intervient autour de 5000 générations soit environ 10 000 ans. Cela coïncide bien avec le début de la domestication du mouton il y a environ 10 500 ans et donc la séparation des populations domestiques des populations sauvages.

L'évolution du partage haplotypique entre domestique décrit plusieurs phases au cours du temps. Pour résumer, nous avons depuis la domestication il y a 10 500 ans :

- Une période d'environ 1000 ans AP de diminutions des connexions entre les populations domestiques et sauvages et entre les différents groupes domestiques
- Une période de plus de 6 millénaires (de 9000 à 3000 ans AP environ) où les connexions entre populations domestiques diminuent peu. Celles entre domestiques et sauvages continuent de décroître au même rythme.
- Une diminution brutale des connexions entre populations domestiques autour de 3000 ans AP en moins de 500 ans
- Une période de 1300 ans où les connexions ne diminuent plus
- Une diminution brutale des connexions entre populations il y a 1300 ans.

Il semble donc que les populations ovines domestiques aient subi plusieurs phases de rupture de connexions. La première fait suite à l'initiation de la domestication il y a 10 500 ans et correspond probablement à la capture de diversité des troupeaux qui constitueront le cheptel domestique et à leur départ du centre de domestication. Les deux suivantes sont plus brutales mais aussi communes à toutes les paires de populations comparées.

La rupture brutale de 3000 ans AP coïncide avec les migrations massives de populations qui ont eu lieu en Asie suite aux guerres d'unification de Qin. Ces mouvements ont mené à la formation de plusieurs groupes ethniques (Fan, 1131; He, 2000; Ou, 1936; Ran, Li and Zhou, 1985, 1985; Si, 1131) . Zhao et al. (2017b) confirme d'ailleurs la divergence des moutons nord chinois et des plateaux Tibétains et du Yunnan Kweichow à cette période. Cela pourrait expliquer la disparition de connexions entre les populations ovines de la région mais pas de celles du reste du monde.

Le caractère généralisé de cette chute de connexions, pourrait nous inciter à chercher des causes plus globales. Or, il y a 3200 ans a eu lieu l'effondrement de l'âge du bronze. Cette période décrit une phase chaotique de l'histoire où les principales civilisations méditerranéennes se désagrègent et presque toutes les cités importantes et prospères sont détruites. La dévastation et le nombre de morts atteignent des niveaux inégalés auparavant. Les causes de ces désastres sont encore débattues mais certainement multiples. Elles sont à

la fois d'ordre politique et sociale avec de nombreuses rébellions et invasions mais aussi d'ordre naturel avec plusieurs tremblements de terre, des sécheresses et des vagues de froid causant de graves famines. C'est cette combinaison d'évènements qui aurait conduit à un effondrement global. De fait, les relations diplomatiques et commerciales sont complètement rompues. Les communautés se centrent alors autour de villages isolés et le monde tombe dans les Ages sombres (Cline, 2016; Drews, 1993; Liverani, 1987). Ce désastre pourrait donc bien expliquer la rupture brutale des connexions entre populations que nous avons montré.

La dernière rupture franche des connexions entre les populations ovines à lieu il y a 1400 ans environ soit autour de l'an 600 EC (ère commune). Cette période est marquée par la fin de l'optimum Climatique romain qui avait permis un développement important de toute la région méditerranéenne. Il s'en suit un petit âge glaciaire pendant près de deux siècles. Plusieurs éruptions volcaniques entre 530 et 540 EC provoquent des nuées de cendre qui cachent le soleil et bouleversent ainsi le climat du monde entier. De plus, en 542 EC survient la Peste de Justinien, dont la propagation a d'ailleurs pu être facilitée par les conditions climatiques. Les conséquences de ce changement de climat comme la famine, ou le déplacement des populations de rongeurs ont pu rendre les humains plus vulnérables. Le déclin des populations humaines est énorme, Rome passe de près d'un million d'habitants à seulement 20 000, à Constantinople, 10 000 personnes meurent tous les jours et dans les villages il n'est pas rare de voir plus de la moitié des habitants décimés en un été. Tout l'Ancien Monde est touché, depuis la source de l'épidémie en Asie à l'Ouest de l'Europe (Harper, Rossignol and Pignarre, 2019). Certains auteurs contestent l'intensité des conséquences de cette épidémie (Biraben, 1976; Mordechai et al., 2019) alors que d'autres en font une des causes majeures de la chute de Rome (Meier, 2016; Rosen, 2007). Les éléments que nous apportons ici, plaident pour une interruption des connexions entre les groupes ovins, inévitablement liés aux groupes humains, et donc sûrement pour des conséquences importantes sur le commerce et les échanges mondiaux.

## 7 Conclusion

---

Nous avons daté l'apparition de plus de 35 millions de variants du génome du mouton. L'analyse de leur distribution montre que la majorité datent des étapes initiales de la domestication. Nous avons aussi identifié des variants spécifiques et caractéristiques de différents groupes. Cela nous a permis d'étudier l'accumulation de mutation en lien avec la diffusion et l'isolement des groupes ovins qui ont été introduit dans le monde entier. Nous avons également quantifié le partage d'haplotype entre ces groupes au cours du temps et mis en évidence des périodes de flux de gènes et de ruptures brutales des connexions. Ces ruptures coïncident avec de grandes crises de l'Humanité.

Ce chapitre montre que par la datation d'évènements de mutation, nous pouvons préciser les évènements passés qui ont modelé le génome des moutons actuels. Dans le prochain

chapitre nous nous intéressons à la datation d'évènements de sélection en lien avec leur domestication.

## **Chapitre 4**

# **Chronologie des évènements de sélection liés à la domestication du mouton**

---

# 1. Introduction

---

La domestication est un continuum de processus de fondation de populations et de sélection des animaux les plus adaptés aux besoins et aux environnements humains. Ces processus entraînent un ensemble de différences comportementales, morphologiques et physiologiques entre les espèces ainsi domestiquées et leurs parents sauvages descendants des mêmes ancêtres. Une partie de ces changements sont d'ailleurs communs à de multiples espèces domestiquées. C'est Darwin, en 1868, qui recense en premier ces convergences chez les mammifères: variations de couleur de la robe, oreilles tombantes, forme du crâne et des dents, docilité, altération des niveaux de certaines hormones en sont quelques exemples.

D'autres caractéristiques sélectionnées chez les animaux domestiques sont plus spécifiques à chaque espèce ou même à chaque race ou variété. Elles dépendent de l'environnement dans lequel l'humain a introduit les animaux, et de besoins particuliers. Les Mérinos ont par exemple été sélectionnés pour la production de laine et les Frisonnes pour le lait. Ce type de caractéristique agronomique a récemment été intensivement sélectionné avec l'apparition du concept de race au 18<sup>ème</sup> siècle, la diminution des croisements inter-races et la généralisation de l'insémination artificielle (Taberlet et al., 2011).

Les bases génétiques de la sélection chez le mouton sont largement étudiées, pour des aspects historiques et surtout agronomiques (Bai et al., 2019; Cesar et al., 2014; Cesarani et al., 2019b, 2019a; Demars et al., 2017; Dong et al., 2020; Gong et al., 2016; Gratten et al., 2010; Gutiérrez-Gil et al., 2017, 2014; Han et al., 2019; Kijas et al., 2012b; Li, Tiirikka and Kantanen, 2014; Mastrangelo et al., 2019; Megdiche et al., 2019; Moradi et al., 2012; Mwacharo et al., 2017; Norris and Whan, 2008; Ozerov et al., 2019, 2020; Pan et al., 2018b; Rocha, Chen and Beja-Pereira, 2011; Rochus et al., 2019; Sutera et al., 2019). De nombreuses régions cibles de la sélection ont ainsi été identifiées et localisées dans les génomes. Cependant, l'origine des variants sélectionnés reste inconnue. Pourtant, situer dans le temps et l'espace les évènements de sélections pourrait nous renseigner sur le processus de domestication et les motivations des sociétés passées qui les ont mis en œuvre (leurs besoins ou leurs croyances), en complément des informations apportées par l'archéologie. L'objectif de ce chapitre sera donc d'apporter des éléments de réponses via la datation de variants génomiques. Les questions abordées seront les suivantes :

- **Quand sont apparus les variants impliqués dans des processus de sélection liés à la domestication du mouton et peut-on en déduire leur origine géographique ?**
- **Quand ont-ils été sélectionnés, et combien de temps s'est-il écoulé entre leur apparition et leur sélection ?**
- **Subsiste-t-il des traces des sélections réalisées lors des étapes initiales de la domestication, malgré les phénomènes de remplacement et des épisodes de sélections successives ?**

Pour répondre, nous établirons une liste de gènes candidats et de variants potentiellement impliqués dans le processus de domestication et nous caractériserons les âges d'apparition et de sélection de ces variants.

## 2. Méthode de datation de sélection

---

Pour dater la sélection de variants génomiques nous emploierons une méthode basée, comme **GEVA**, sur le déclin du déséquilibre de liaison (voir **chapitre 3, partie 2**). Parmi ces méthodes basées sur le LD, certaines utilisent des statistiques résumantes. Elles peuvent s'appuyer sur des estimations ponctuelles mais ignorent alors l'incertitude liée à l'estimation de la longueur des haplotypes (Coop et al., 2008; Hudson, 2007; Meligkotsidou and Fearnhead, 2005; Tang et al., 2002; Thomson et al., 2000). Les méthodes ABC (Approximate Bayesian Computation) utilisent des modèles de simulation pour identifier les âges des variants qui soient cohérents avec les données observées. Elles prennent en compte les incertitudes liées aux événements aléatoires de recombinaisons, mutations et/ou à l'histoire démographique de la population mais n'utilisent pas toutes les caractéristiques des données puisqu'elles reposent également sur des statistiques résumantes (Beleza et al., 2013; Nakagome, Hudson and Di Rienzo, 2019; Ormond et al., 2016; Peter, Huerta-Sanchez and Nielsen, 2012). De plus, les données de génomes complets rendent difficile la définition de statistiques qui résument correctement les patrons de DL (Smith et al., 2018).

Au sein des méthodes permettant de prendre en compte la vraisemblance complète des données on trouve celles de Chen, Slatkin 2013; Coop, Griffiths 2004 et Slatkin 2001. Elles restent cependant infaisables informatiquement ou ne considèrent pas la recombinaison entre de nombreux loci. Dans les années 2000, plusieurs autres méthodes modélisant la recombinaison autour de variants sélectionnés ont été développées mais se révèlent aussi trop lourdes à mettre en place avec des données actuelles (McPeck and Strahs, 1999; Morris, Whittaker and Balding, 2000; Rannala and Reeve, 2003, 2001).

Finalement, si l'on veut pouvoir prendre en compte l'incertitude de nos estimations et exploiter l'ensemble de l'information contenue dans des données de génomes complets, le type de méthode le plus approprié est basé sur des chaînes de Markov cachées (CMC) (Chen, Hey and Slatkin, 2015; Smith et al., 2018). C'est pourquoi nous utiliserons la méthode **STARTMRCA** (Smith et al., 2018). Son principe est différent de celui de **GEVA** (voir **chapitre 3, partie 2**) qui date l'apparition d'une mutation. **STARTMRCA** estime le TMRCA des haplotypes portant le variant sélectionné et l'utilise comme proxy de l'âge de la sélection. Elle utilise un MCMC (Monte Carlo Markov Chain) pour estimer la distribution a posteriori du TMRCA des haplotypes portant le variant sous sélection. A chaque itération un TMRCA est échantillonné, puis chaque haplotype portant le variant sélectionné (haplotypes sélectionnés) est scanné par une CMC qui identifie les points de recombinaison en segmentant la séquence en deux états (**Figure 49**):



- **L'état ancestral**, qui correspond au reste du segment d'ADN entraîné par DL au moment de la sélection.
- **L'état de fond**, qui est modélisé par une mosaïque imparfaite (Li and Stephens, 2003) des haplotypes d'un panel de référence ne portant pas le variant sélectionné et recombinant avec l'haplotype ancestral lors de l'augmentation de sa fréquence au cours du temps.

Les deux états peuvent présenter des mutations dont l'accumulation donne aussi une information sur le temps écoulé depuis la sélection et qui sont donc aussi prises en compte dans le modèle. Suite à la détection des points de recombinaison de chaque haplotype, une vraisemblance composite est calculée sur l'ensemble des haplotypes puis comparée à la précédente pour ajuster la fonction de proposition<sup>15</sup> et sélectionner le TMRCA de la prochaine itération. Après le nombre d'itérations choisi, le TMRCA est inféré grâce à sa distribution de probabilité a posteriori.

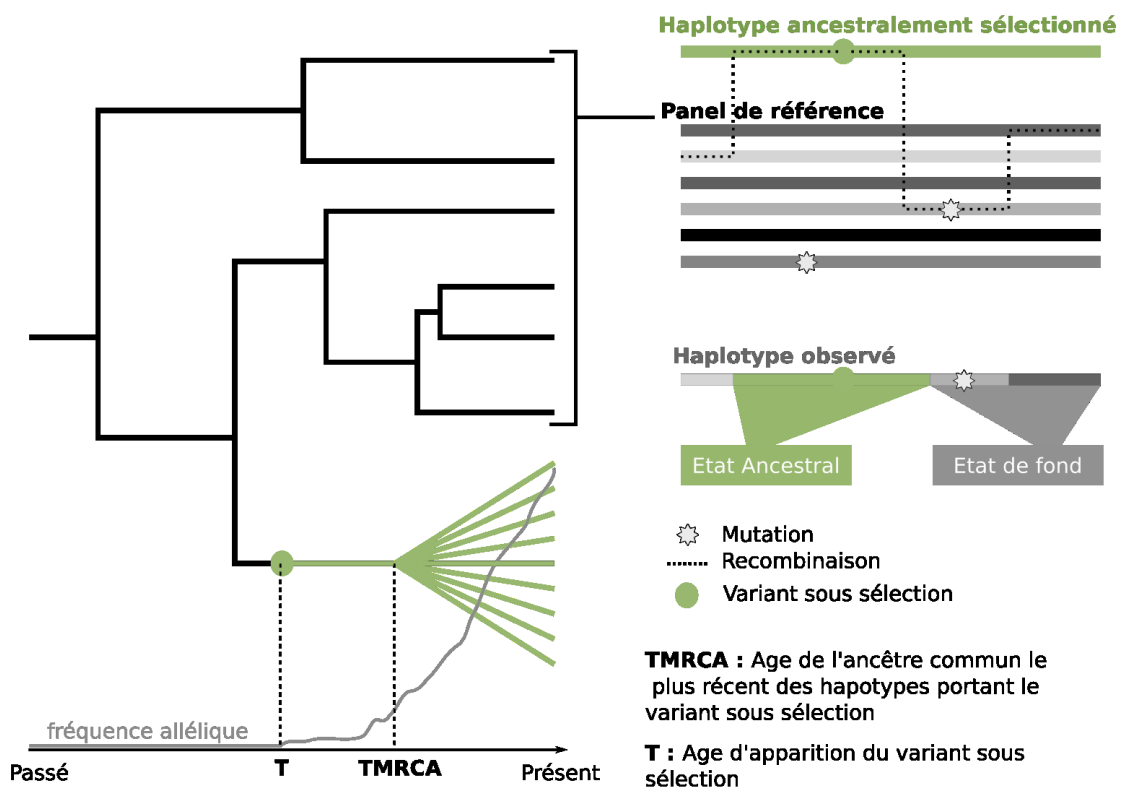


Figure 49 : Description du modèle utilisé pour la datation de la sélection de variant génomique par la méthode *startmrcr*.

D'après la Figure 1 de Smith *et al.* 2018

<sup>15</sup> La **fonction de proposition** permet de choisir la valeur du paramètre à inférer de la prochaine itération d'une chaîne de MCMC.

L'analyse `STARTMRCA` nécessite jusqu'à 15 cœurs de calculs de 4GB et jusqu'à 10 jours de calculs suivant la taille des haplotypes à analyser (variable dans le groupe A1) et le nombre d'haplotypes sélectionnés. Nous ne pourrions donc pas avoir une approche génome-complet comme nous l'avons fait avec `GEVA` mais plutôt une approche ciblée sur des régions candidates à la sélection.

## 3. Datations de variants potentiellement impliqués dans la domestication.

### 3.1. Choix des variants

Nous avons identifié des régions génomiques qui présentent des signaux de sélection et constitué deux groupes de variants sélectionnés a priori lors d'étapes différentes du processus de domestication (signaux cosmopolites et signaux intra-groupes).

#### 3.1.1. Groupe A : Signaux cosmopolites

Ce groupe est composé de 46 régions précédemment identifiées par Alberto et al., (2018) (**Tableau 16**). Ces régions présentent des signaux de sélection communs aux populations occidentales du jeu de données de bases (voir **Chapitre 2, partie 2.1**) : deux groupes de races traditionnelles en Iran (**IR**) au niveau du centre de domestication, et au Maroc (**MO**) à l'extrémité d'une des voies de diffusion, ainsi qu'un panel de races mondial (**RI**). Les signaux de sélection ont été détectés par la méthode `HAPFLK` (Fariello et al., 2013) en contrastant les moutons de chaque groupe avec des mouflons asiatiques et en ne conservant que les signaux communs au trois groupes (MO, IR et RI). Certaines de ces régions sont orthologues avec d'autres identifiées de la même manière chez la chèvre.

Leur ubiquité et leur partage avec une espèce proche soutient l'hypothèse que ces régions ont été sélectionnées dans les phases initiales de la domestication. En effet, le scénario le plus parcimonieux pour expliquer la présence de signaux de sélection dans différents groupes est une sélection avant leur séparation.

**Tableau 16 : Description des signaux de sélection cosmopolites (groupe A).**

Ces signaux correspondent à des régions génomiques liées à la domestication du mouton identifiées par Alberto et al. 2018. Les coordonnées des régions correspondent à la version 4 du génome de référence d'*Ovis aries*. Le groupe A1 correspond aux régions dont l'apparition et la sélection ont été estimées. Le groupe A2 correspond aux régions dont seuls les âges d'apparition ont été estimés. Les régions dont le code se termine par 'c' sont celles retrouvées également chez la chèvre.

GROUPE A1						
Codes régions	Coordonnées			Gène	Phénotype relié	Référence
	chr	Début	Fin			
R1c	6	36112003	36199650	HERCS	Composition du lait	(Cohen-Zinder, 2005)
R2	4	101691203	101696601	GLI3	Poids du corps –	(García-Gámez et al., 2011; Huang et

					Pigmentation bigarrée	al., 2013)
<b>R3</b>	13	49772631	49893512	HAO1	Composition du lait	(Baik et al., 2009)
<b>R4</b>	15	21887888	21935826	BCO2	Couleur du lait et de la graisse	(Bartonova et al., 2012; Våge and Boman, 2010)
<b>R5</b>	18	45479788	45500762	NKX2-1	Fertilité	(Nayeri et al., 2016)
<b>R6</b>	3	129600886	129630128	SOCS2	Poids du corps – croissance – production de lait – réponse inflammatoire aux infections mammaires	(Horvat and Medrano, 2001; Rupp et al., 2015)
<b>R7c</b>	3	124473834	124563229	KITLG	Couleur de la robe – taille de la portée – développement neuronal	(An et al., 2015; Conti and Cattaneo, 2010; Reissmann and Ludwig, 2013)
<b>R8c</b>	3	124627827	124826809			
<b>R9c</b>	1	109857674	109910789	SLAMF1	Immunité humorale – réponse des anticorps	(Wang et al., 2015)
<b>R10c</b>	16	70884098	70962549	TRIP13	Susceptibilité aux mammites	(Durán Aguilar et al., 2017)
<b>R11</b>	11	24708912	24725858	FBXO39	Voies pro inflammatoires	(Tollenaere et al., 2013)
<b>R12</b>	11	40281714	40303212	CCR7	Réponse immunitaire innée	(Galindo et al., 2008)
<b>R13c</b>	2	227432538	227448029	Intergénique	-	-
<b>R14</b>	2	115069690	115078874	Intergénique	-	-
<b>R15</b>	2	115250650	115253056	Intergénique	-	-
<b>R16</b>	4	78931960	78946506	Intergénique	-	-
<b>R17</b>	17	59461950	59495329	Intergénique	-	-
<b>R18</b>	17	52389734	52414267	Intergénique	-	-
<b>R19c</b>	1	106102777	106161764	ENSOARG0000006800	Inconnu	-
<b>R20c</b>	11	51202814	51294278	RNF213	Saisonnalité de la reproduction	(Martinez-Royo et al., 2017)
<b>R21</b>	9	30879974	30899480	HAS2	Matrice extra-cellulaire	<a href="https://www.ncbi.nlm.nih.gov/gene">https://www.ncbi.nlm.nih.gov/gene</a>
<b>R22</b>	11	10787165	10790775	BRIP1	Structure de l'ADN	<a href="https://www.ncbi.nlm.nih.gov/gene">https://www.ncbi.nlm.nih.gov/gene</a>
<b>R23</b>	13	50408792	50540720	ENSOARG00000023728	Inconnu	-
<b>R24</b>	18	32214373	32363532	NEIL1	Voie de réparation de l'ADN	<a href="https://www.ncbi.nlm.nih.gov/gene">https://www.ncbi.nlm.nih.gov/gene</a>
<b>R25</b>	20	14983419	15136347	ADCY10	Spermatogenèse – taux de testostérone	(Open Targets Genetics Portal, n.d.; Ruth et al., 2020; Serrano et al., 2021)
<b>Groupe A2</b>						
<b>R26c</b>	10	35999396	36231879	CRYL1	Teneur en acide gras du lait	(Yang et al., 2016a)
<b>R27c</b>	20	50782908	50936215	EXOC2	Reconnaissance et activation de l'immunité innée	(Paludan et al., 2011)
<b>R28c</b>	24	139671	145137	HBM	Transport de l'oxygène	<a href="https://www.ncbi.nlm.nih.gov/gene">https://www.ncbi.nlm.nih.gov/gene</a>
<b>R29c</b>	3	153598149	153706966	HMG1-C	Taille du corps, développement des cellules de crête neural	(Nishino et al., 2008; Ruyter-Spira, 1998)
<b>R30c</b>	1	239246357	239298321	Intergenic	-	-

R31c	7	28421856	28491599	Intergenic	-	-
R32c	26	18597538	18618884	MTMR7	Composition des acides gras musculaire, système nerveux central	(Ramayo-Caldas et al., 2012; Sanchez-Juan et al., 2012)
R33c	10	26008999	26046997	NBEA	Système nerveux central, ondulation de la laine	(Nair et al., 2012; Wang et al., 2014c)
R34c	9	78537804	78540976	POP1	Développement de l'épithélium	<a href="https://www.ncbi.nlm.nih.gov/gene">https://www.ncbi.nlm.nih.gov/gene</a>
R35c	6	44863756	44904665	SLC34A2	Système respiratoire, adaptation à l'altitude, tolérance à la Trypanosomose	(Serranito et al., 2021b, 2021a; Verma et al., 2018)
R36c	20	18251995	18371246	SUPT3H	Teneur en acide gras du lait	(Buitenhuis et al., 2014)
R37c	15	47305536	47562113	U1	Epissage de l'ARN	<a href="https://www.ncbi.nlm.nih.gov/gene">https://www.ncbi.nlm.nih.gov/gene</a>
R38	1	266547595	266569727	CLDN14	Perméabilité de l'épithélium	<a href="https://www.ncbi.nlm.nih.gov/gene">https://www.ncbi.nlm.nih.gov/gene</a>
R39	13	50235959	50350428	ENSOARG0000023728	Inconnu	-
R40	24	41741571	41749858	FAM20C	Composition de la caséine du lait	(Bijl et al., 2014)
R41	20	50006463	50027388	GMDS	Teneur en acide gras du lait	(Ibeagha-Awemu et al., 2016)
R42	21	49572100	49770628	IFITM5	Inhibition de virus pathogène	(Perreira et al., 2013)
R43	1	129001917	129011554	Intergenic	-	-
R44	17	33062008	33077721	KNTC1	Division cellulaire	<a href="https://www.ncbi.nlm.nih.gov/gene">https://www.ncbi.nlm.nih.gov/gene</a>
R45	11	18246976	18454063	NF1	Développement des tissus dérivés de la crête neurale	(Brannan et al., 1994; Fariello et al., 2014)
R46	5	106942568	107181105	SLC25A46	Transport du fer	<a href="https://www.ncbi.nlm.nih.gov/gene">https://www.ncbi.nlm.nih.gov/gene</a>

### 3.1.2. Groupe B : Signaux intra-groupes

Le groupe B cible des variants a priori plus récents puisque provenant de détection intra-groupe et non partagés entre tous les moutons domestiques. La méthode `NSL` (Ferrer-Admetlla et al., 2014) a été employée pour identifier des variants sous sélection au sein de chaque groupe de notre jeu de données étendu (voir **Chapitre 2, partie 2.2**). Cette méthode utilise des données phasées et permet de détecter à la fois des signatures de sélections issues de mutation de novo mais aussi de variations préexistantes, ce qui doit être le cas majoritaire dans un contexte de domestication. Cette méthode est aussi plus robuste aux variations des taux de recombinaison et de mutations que d'autres utilisant des statistiques similaires telles que l'iHS (Voight et al., 2006). Elle utilise l'augmentation de l'homozygotie des haplotypes induite par la sélection positive pour identifier les signaux qui en résultent.

Nous avons calculé et normalisé les statistiques `nsI` pour les variants sans données manquantes au sein de chacun des groupes grâce à l'option `-nsI` et à l'outil `norm` du programme `SELSCAN` (Szpiech and Hernandez, 2014). Nous avons sélectionné les variants dont la statistique est comprise dans 1% des valeurs les plus extrêmes ( $p$ -valeur  $< 10^{-5}$ ). Parmi ces variants nous n'avons conservé que ceux dont la position est comprise dans les intervalles élargi de 2500 pb correspondant à une liste de gènes que nous avons pré-établie. Cette liste a été construite à partir de la littérature et se compose de gènes associés à différents traits phénotypiques d'intérêt chez le mouton : La laine, le lait, la viande, la queue grasse, et les cornes.

Enfin, nous avons sélectionné parmi ces variants ceux dont l'effet prédit était le plus fort. Leur effet a été inféré grâce au Variant Effect Predictor (McLaren et al., 2016). Il en résulte 44 variants compris dans 23 gènes dont les caractéristiques sont résumées dans les (**Tableau 17 et Tableau 18**).

**Tableau 17 : Description des signaux de sélection intra-groupe (groupe B)**

Ces signaux correspondent au meilleurs candidats d'une analyse nsl effectuée au sein de chaque groupe du jeu de données étendu . Les codes des groupes utilisés pour la détection de sélection sont disponibles dans xxx . Les coordonnées des SNP impliquées correspondent à la version 4 du génome de référence de Ovis aries.

Chromosome	Position	Gène	Conséquence (VEP)	Trait	Identifiant	Groupe
6	92644785	FRAS1	missense_variant	laine	laine1	WZ
13	9534439	KIF16B	downstream_gene_variant	laine	laine2	ADN
4	52424434	TFEC	downstream_gene_variant	laine	laine4	VT
6	92555004	FRAS1	intron_variant	laine	laine5	HU
14	14184612	MC1R	intron_variant	laine	laine9	MEN
10	26547852	NBEA	intron_variant	laine	laine10	EU
11	54490144	RNF157	intron_variant	lait	lait_1	EU
5	5329038	SLC27A1	intron_variant	lait	lait_2	DAN
11	13100970	ACACA	intron_variant	lait	lait_3	BSB
11	13156851	ACACA	intron_variant	lait	lait_4	BRB
11	13076170	ACACA	intron_variant	lait	lait_5	ALT
11	13009214	ACACA	intron_variant	lait	lait_6	TA
11	13183990	ACACA	intron_variant	lait	lait_7	DAN
7	33316874	DNAJC17	downstream_gene_variant	lait	lait_8	BSB
21	29070445	KIRREL3	upstream_gene_variant	lait	lait_9	PT
26	12326528	TENM3	upstream_gene_variant	lait	lait_10	BSB
21	29050239	KIRREL3	intron_variant	lait	Lait_11	BSB
21	29044292	KIRREL3	intron_variant	lait	Lait_12	MED
12	59415849	LHX4	intron_variant	viande	viande_1	EU
24	11609448	SHISA9	intron_variant	viande	viande_5	DYG
24	11819895	SHISA9	intergenic_variant	viande	viande_7	GOT
5	93371948	CAST	intron_variant	viande	viande_8	FIN
8	70323757	GRM1	intron_variant	viande	viande_9	OU_DJ
1	261134106	CBS	intron_variant	Queue grasse	queuegrasse_2	ALT
11	60579293	WIPI1	intron_variant	Queue grasse	queuegrasse_3	MED

15	3951054	PDGFD	intron_variant	Queue grasse	queuegrasse_5	SAS
15	4084865	PDGFD	intron_variant	Queue grasse	queuegrasse_6	CB
21	23082465	NELL1	intron_variant	Queue grasse	queuegrasse_8	WGR
21	23171264	NELL1	intron_variant	Queue grasse	queuegrasse_9	OL
21	23290662	NELL1	intron_variant	Queue grasse	queuegrasse_10	SFK
17	4834112	TMEM154	intron_variant	Fertilité	fertilite_1	DYG
17	4869264	TMEM154	intron_variant	Fertilité	fertilite_2	DYG
17	4881722	TMEM154	intron_variant	Fertilité	fertilite_3	GSS
17	4866683	TMEM154	intron_variant	Fertilité	fertilite_4	KEF
7	78764920	SLC8A3	intron_variant	Fertilité	fertilite_5	MEN
7	78697340	SLC8A3	intron_variant	Fertilité	fertilite_6	STH
7	78674519	SLC8A3	intron_variant	Fertilité	fertilite_7	MED
7	78701697	SLC8A3	intron_variant	Fertilité	fertilite_8	TA
7	78778221	SLC8A3	intron_variant	Fertilité	fertilite_9	FIN
7	78687817	SLC8A3	intron_variant	Fertilité	fertilite_10	DJK
10	29502304	RXFP2	intergenic_variant	cornes	cornes_1	ADN
10	29488218	RXFP2	intergenic_variant	cornes	cornes_2	OU_DJ
2	132726209	MTX2	intron_variant	cornes	cornes_3	BRB
2	132687303	MTX2	intron_variant	cornes	cornes_5	IROOb

**Tableau 18 : Description des gènes associées aux signaux de sélection intra-groupe (groupe B)**  
 Les coordonnées des gènes correspondent à la version 4 du génome de référence d'*Ovis aries*.

Gènes	Chromosome		Trait	Rôle	Références
	Début	Fin			
<b>MTX2</b>	2	132672 354	132748 945	cornes	Associé au phénotype quatre cornes (He et al., 2016; Kijas et al., 2016)
<b>RXFP2</b>	10	294349 33	295091 59	cornes	Contrôle de la forme et de la taille des cornes (Pan et al., 2018b; Wang et al., 2014a; Wiedemar and Drögemüller, 2015)
<b>SLC8A3</b>	7	786483 40	788103 53	fertilité	Développement des ovocytes (Wang et al., 2015)
<b>TMEM154</b>	17	483600 4	488768 3	fertilité	Associé à la susceptibilité au lentivirus Maedi-Visna (Bowles, Carson and Isaac, 2014; Gebreselassie et al., 2020; Heaton et al., 2012)
<b>FRAS1</b>	6	922979 08	928152 12	laine	Adhésion dermique et épidermique – Biologie du Follicule pileux (Bolormaa et al., 2017; Clements et al., 2012)
<b>KIF16B</b>	13	953619 4	984867 9	laine	Impliqué dans le diamètre de la fibre (Wang et al., 2014c)
<b>KIRREL3</b>	21	289334 29	295479 95	lait	Impliqué dans le pourcentage de graisse et le rendement du lait (Sutera et al., 2019)
<b>MC1R</b>	14	141846	141854	laine	Impliqué dans la couleur (Gebreselassie et al., 2020; Rochus et

		66	86			al., 2019)
NBEA	10	259797 18	266540 48	laine	Ondulation de la laine	(Wang et al., 2014c)
TFEC	4	522007 87	524301 71	laine	Diamètre de la fibre	(Wang et al., 2014c)
ACACA	11	129873 06	132320 20	lait	Code pour une enzyme clé dans la synthèse des acide gras dans la grande mammaire	(Moioli et al., 2013)
DNAJC17	7	333181 92	333484 33	lait	Impliqué dans le pourcentage de graisse et le rendement du lait	(Sutera et al., 2019)
KIRREL3	21	289334 29	295479 95	lait	Impliqué dans le pourcentage de graisse et le rendement du lait	(Sutera et al., 2019)
RNF157	11	544246 84	544885 86	lait	Impliqué dans le pourcentage de graisse et le rendement du lait	(Sutera et al., 2019)
SLC27A1	5	529790 8	533501 5	lait	Impliqué dans le pourcentage de graisse et le rendement du lait	(Sutera et al., 2019)
TENM3	26	974623 7	125108 53	lait	Impliqué dans le pourcentage de graisse et le rendement du lait	(Sutera et al., 2019)
CBS	1	261133 034	261153 763	queue grasse	Déposition de graisse dans les tissus adipeux du mouton	(Wang et al., 2014b)
COL15A1	2	485250 47	486273 23	queue grasse	Déposition de graisse dans les tissus adipeux du mouton	(Wang et al., 2014b)
NELL1	21	225514 76	235673 62	queue grasse	Déposition de graisse dans les tissus adipeux du mouton	(X. Wang et al., 2014b)
PDGFD	15	385293 5	413738 2	queue grasse	Configuration de la queue, dépôt de gras dans la queue	(Li et al., 2020c)
RBM11	1	141182 795	141199 646	queue grasse	Métabolisme du gras	(Zhang et al., 2019)
WIPI1	11	605754 92	606045 77	queue grasse	Déposition de graisse dans les tissus adipeux du mouton	(Wang et al., 2014b)
CAST	5	933241 71	934141 16	viande	Qualité, tendreté et teneur en acide gras	(Aali et al., 2017)
FTO	14	214694 00	218972 67	viande	indice de masse corporelle, taux de croissance et masse grasseuse	(Wang et al., 2015)
GRM1	8	701691 98	706114 33	viande	Croissance et production de viande	(Zhang et al., 2013)
LHX4	12	594152 29	594661 80	viande	croissance	(Zhao et al., 2017a)
MBD5	2	159922 229	160113 386	viande	gain post sevrage	(Zhang et al., 2013)
SHISA9	24	115919 62	119389 92	viande	circonférence du tibia	(Zhang et al., 2013)

## 3.2. Datations de l'apparition et de la sélection des variants des groupes A et B

### 3.2.1. Groupe A

Ce groupe est constitué de 46 régions génomiques de 2407 à 254 100 pb (**Tableau 16**). Nous avons divisé ce groupe de régions en deux sous-groupes : le sous-groupe A1 avec des régions dont on datera l'apparition et la sélection et le groupe A2 dont on ne datera que l'apparition (**Figure 50**).

Le groupe A1 est constitué de seulement 25 régions sur les 46 disponibles. Nous avons d'une part dû restreindre le nombre de régions analysées en raison des contraintes computationnelles inhérentes à la méthode *STARTMRCA*. D'autre part nous n'avons conservé que des régions dont l'arbre de distance présentait une topologie similaire à celle présentée dans la **figure 51** et qui répond le mieux à l'hypothèse de généalogie en étoile du modèle utilisé dans *startmrca*. Ces arbres ont été construits à partir distances génétique brutes en

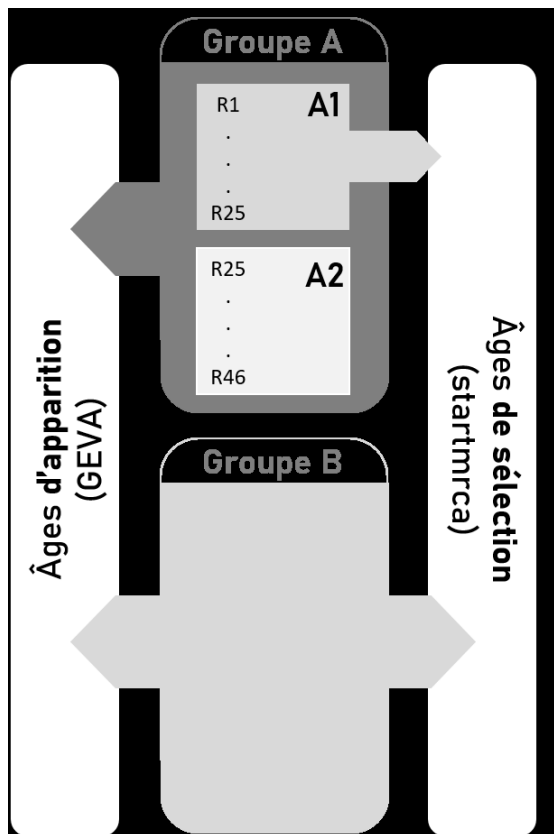


Figure 50 : : Datations réalisées pour chacun des groupes de variants associés à la domestication du mouton.

La description des groupes A et B est disponible dans la section 3.1. de ce chapitre

Neighbour-Joining localement sur l'intervalle correspondant à chaque région sélectionnée et avec l'ensemble des individus du jeu de données de base (partie N) (Figure S2).

Le groupe A1 comprend 18 régions tombant dans des gènes et 6 en zones inter-géniques. Parmi les régions géniques nous avons pris soin de sélectionner des gènes aux fonctions diverses, impliquées dans l'immunité, le système nerveux et la production de lait, de laine ou de viande. Cet ensemble contient également 8 régions dont les signaux de sélection sont aussi retrouvés chez la chèvre (Tableau 16). Nous avons identifié les haplotypes sélectionnés à partir des arbres de distances (Figure S2). Les haplotypes sélectionnés se trouvent ainsi dans une multifurcation en « râteau », signature de la sélection. Le reste des haplotypes, qui ne portent donc pas la région sélectionnée, ont constitué le panel de référence nécessaire à la datation par STARTMRCA. Cette topologie caractéristique est représentée dans la Figure 51.

Startmrca identifie le type d'haplotype (sélectionné ou référence) sur la base de l'état allélique en une position donnée. Nous avons donc tagué les haplotypes en insérant un variant fictif au milieu de la région sous sélection. Nous avons donné à ce variant l'état ancestral pour les haplotypes du panel de référence et l'état dérivé pour les haplotypes sélectionnés.

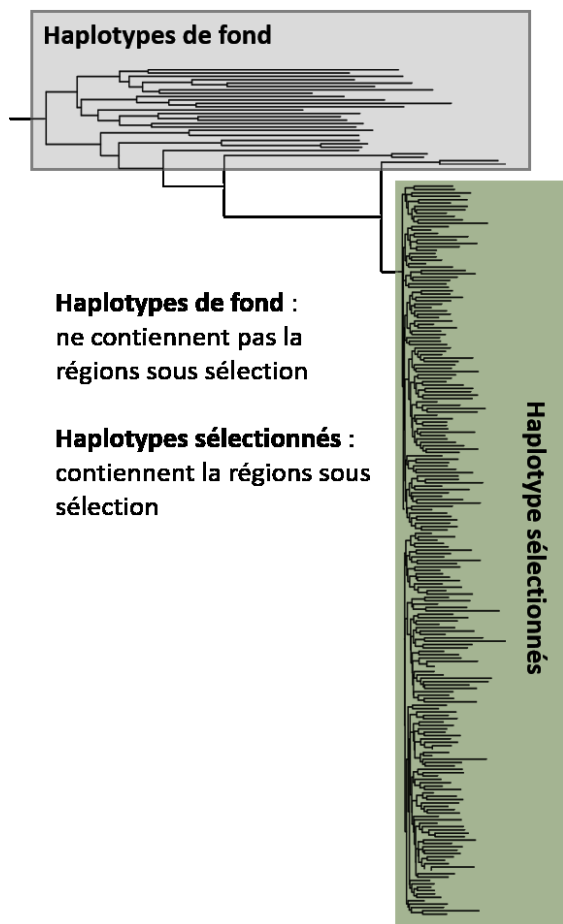
### 3.2.2. Groupe B

Pour les variants de ce groupe nous avons également rassemblé les âges d'apparition à partir de notre atlas et daté la sélection de chacun d'entre eux grâce à STARTMRCA (Figure 49).

### 3.2.3. Paramètres de datation de sélection par startmrca

Les datations de sélection ont été effectuées avec 10 000 itérations, un taux de mutation de  $2 \cdot 10^{-8}$  (Kumar and Subramanian, 2002; Nachman and Crowell, 2000) et en utilisant des taux de recombinaison variables le long du génome (Petit et al., 2017). Nous avons limité la valeur initiale du TMRCA à 20 000 générations. On s'attend à ce que les sélections soient postérieures à la domestication, soit à environ 5500 générations mais nous avons laissé une marge étant donné la fluctuation possible du temps de génération et du taux de mutation au cours du temps. Le nombre d'haplotypes sélectionnés a été limité à 50 et celui du panel de référence à 60. Nous avons scanné les haplotypes sur 1 Mégabase (Mb) de chaque





côté des variants à dater pour le groupe B et sur 1 Mb à partir des extrémités des régions sélectionnées pour le groupe A1.

Nous avons vérifié la convergence vers une même valeur de quatre chaînes indépendantes pour une même région sur les 25 régions du groupe A. Nous avons testé l'impact du choix de l'échantillonnage des haplotypes sélectionnés sur l'estimation de l'âge de la sélection en datant les mêmes régions avec des haplotypes sélectionnés provenant soit du groupe de races iraniennes (**IR**) uniquement, soit du groupe de races marocaines (**MO**) uniquement, soit en échantillonnant aléatoirement des individus du panel mondial (**RI**). Les distributions de probabilités des âges de sélection a posteriori ont été établies à partir des 3-5000 dernières itérations en fonction de la stabilisation des chaînes.

Figure 51 : Topologie permettant l'identification des haplotypes porteurs ou non porteurs des régions sous sélections du groupe A1

## 4. Résultats

### 4.1. Groupe A

#### 4.1.1. Ages de la sélection des 25 régions

##### 4.1.1.1. Etudes de la stabilisation et de la convergence des chaînes

Les différentes chaînes de MCMC se stabilisent autour d'une valeur au plus tard après 8000 itérations (**Figure 52** et **Figure 53**). Les estimations issues de différents échantillons sont presque toujours similaires (**Figure 52**) sauf pour les régions R1c, R7c, R8c, R9c et R20c (**Figure 53**). Ce sont toutes des régions qui sont aussi sous sélection en lien avec la domestication chez la chèvre :

- **Pour la R1c**, trois chaînes sur quatre de l'analyse sur l'échantillon IR se stabilisent à un TMRCA de 2000 générations au lieu de 55 pour toutes les autres chaînes. Cependant, la quatrième chaîne du groupe IR rejoint l'estimation des deux autres groupes. De plus, toutes les chaînes des groupes RI et MO restent stabilisées sur cet optimum local autour de 2000 générations pendant plusieurs milliers d'itérations avant d'atteindre

l'optimum final autour de 55 générations. Nous garderons donc cette dernière estimation pour les trois groupes.

- **Pour la régions R7c**, les quatre chaînes de l'analyse sur l'échantillon RI se stabilisent autour de 35 générations alors que les chaînes des autres échantillons se stabilisent autour de 1800 générations.
- **Pour la région R8c**, les quatre chaînes de l'analyse sur l'échantillon MO se stabilisent autour de 4500 générations alors que les chaînes des autres échantillons se stabilisent autour de 30 générations.
- **Pour la région R9c**, les quatre chaînes de l'analyse sur l'échantillon MO se stabilisent autour de 14 000 générations alors que les chaînes des autres échantillons se stabilisent entre 100 et 500 générations.
- **Pour la régions R20c**, les quatre chaînes de l'analyse sur l'échantillon RI se stabilisent autour de 70 générations alors que les chaînes des autres échantillons se stabilisent autour de 1200 générations.

Pour un même échantillon d'haplotypes sélectionnés (IR, MO ou RI) les 4 chaînes indépendantes se stabilisent autour d'une même valeur pour toutes les régions sauf la région R1c où nous avons vu qu'une chaîne sur quatre ne correspond pas. On a donc un pourcentage de réplicabilité pour un même échantillon de 99% et de 80% avec des échantillons différents.

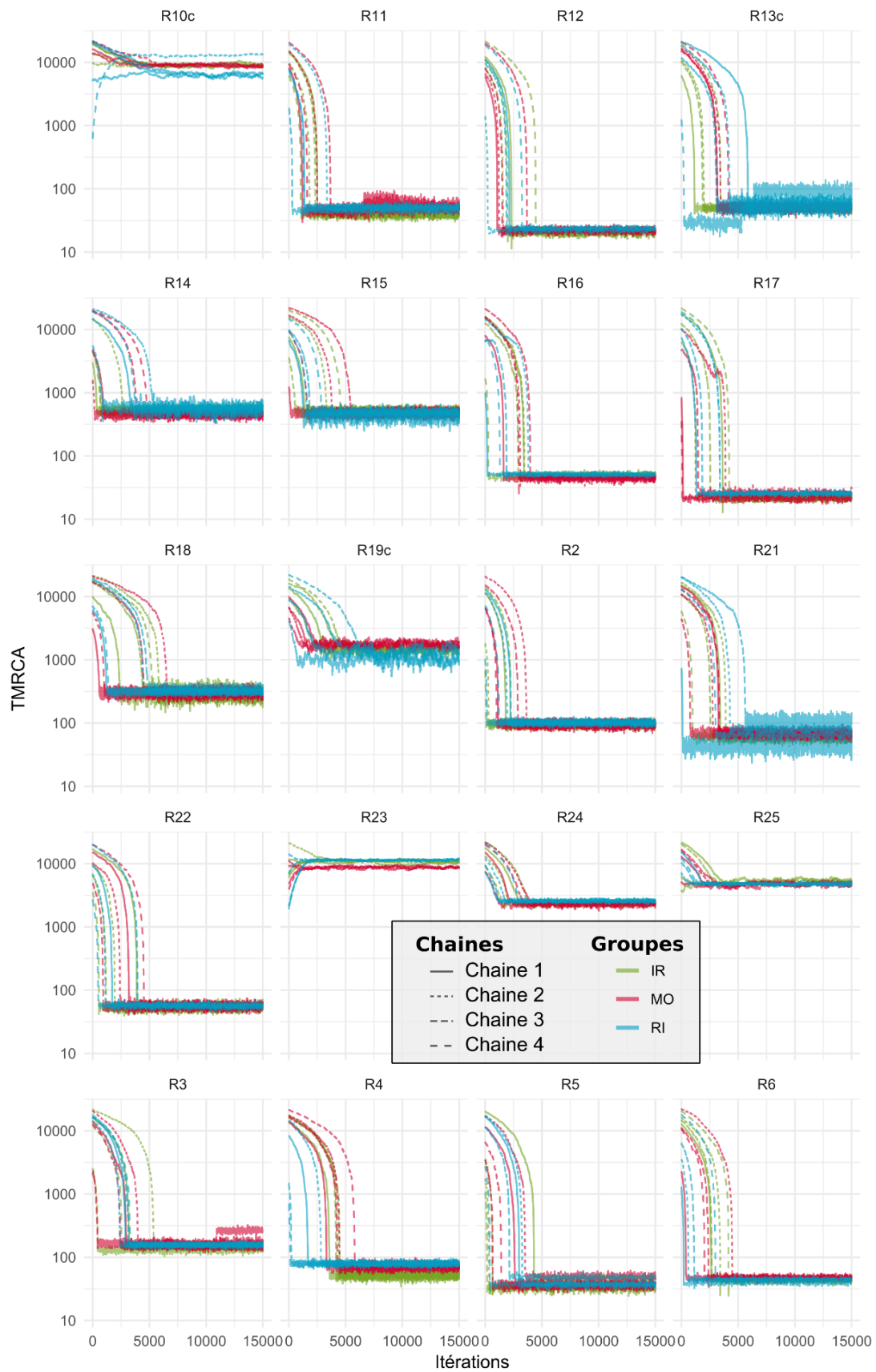


Figure 52 : Traces des chaines de MCMC de l'inférence startmrca qui montre une réplabilité entre chaines indépendantes issues ou non d'un même échantillon.

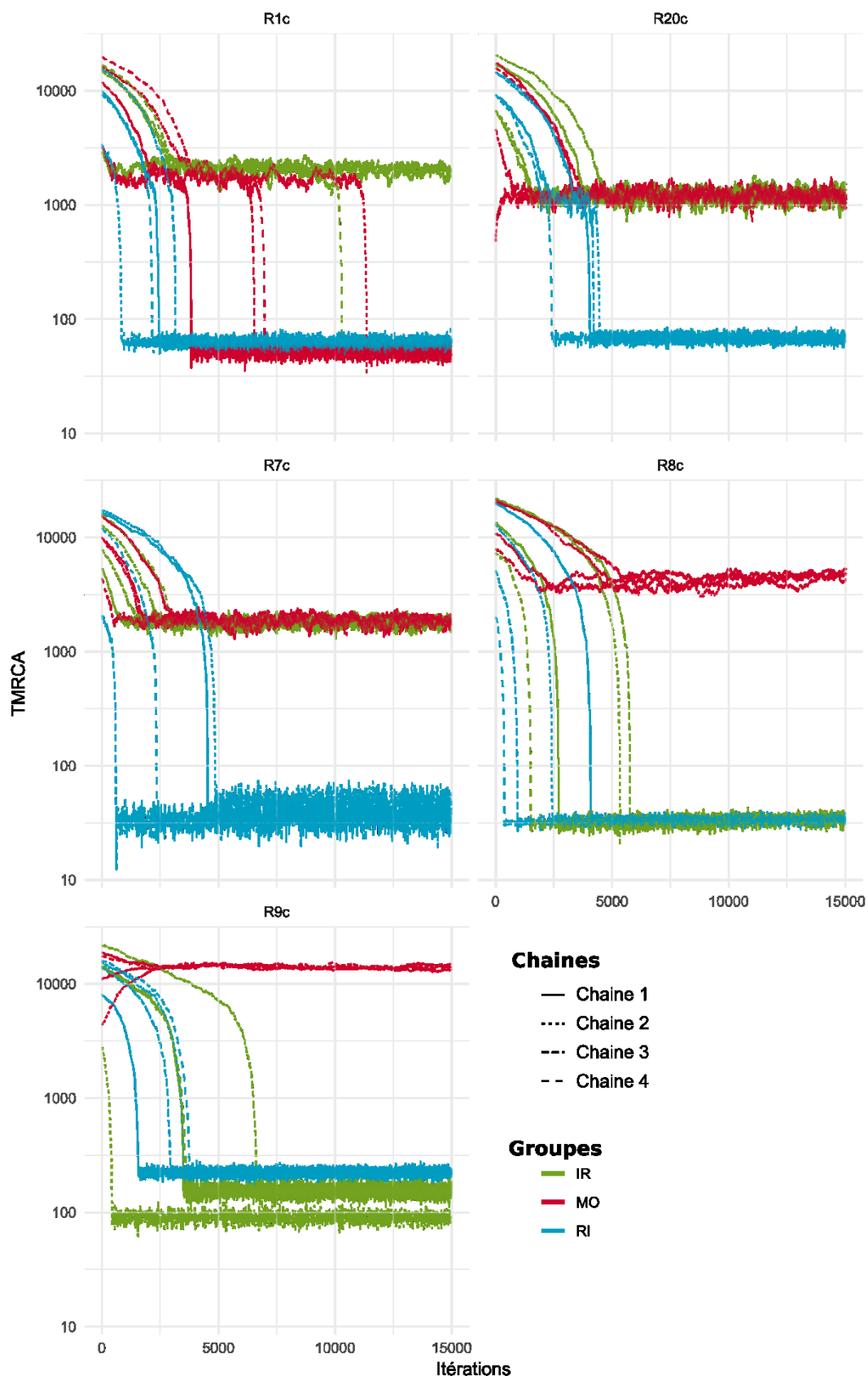


Figure 53 : Traces des chaînes de MCMC de l'inférence startmrca qui montrent des différences d'estimations entre les chaînes issues d'échantillons différents.

#### 4.1.1.2. Age de la sélection des régions du groupe A

Les âges de sélection des 25 régions du groupe A1 varient de 20 à 14 500 générations (**Figure 54**). Les régions les plus anciennes sont aussi celles dont l'estimation présente la plus grande variance. Neuf régions présentent des âges de sélection supérieurs à 1000 générations, toutes les autres ont été sélectionnées plus récemment. Parmi les neuf plus anciennes, six sont aussi retrouvées sous sélection en lien avec la domestication chez la chèvre. La surreprésentation de ces régions parmi le groupe le plus ancien n'est pas significative ( $p$ -value = 0.9). Parmi les 19 régions les plus récentes, 16 ont un âge de sélection inférieur à 250 générations.

Les régions géniques les plus anciennes (sélectionnées avant 1000 générations) sont impliquées dans la couleur, le développement du système nerveux, la reproduction et l'immunité. Les régions géniques les plus récentes (sélectionnées après 1000 générations), sont impliquées dans l'immunité, la couleur, le développement du système nerveux, la production de lait ou de viande et la fertilité. Aucune catégorie n'est surreprésentée dans ces deux classes d'âges. Notons que le développement neuronal n'est représenté que par KITLG, un gène pléiotrope, qui présente des âges de sélection très différents en fonction du groupe que l'on considère.

Les deux régions localisées dans le gène KITLG présentent des âges de sélection différents pour les échantillons marocains et iraniens. Pour l'échantillon marocain, la sélection date de 4500 générations pour la région R8c alors qu'elle ne date que de 1800 générations pour la R7c. Pour l'échantillon iranien, la sélection de la région R8c date de 35 générations alors que celle de la R7c remonte à 1800 générations. Pour l'échantillon correspondant au panel mondial la sélection de ces deux régions datent de 35 générations.

La région R9c, présente un âge de sélection de 13 900 générations environ pour le groupe marocain alors qu'il ne dépasse pas les 30 générations pour les deux autres groupes. De la même manière la région R20c présente des âges de sélection supérieurs à 1000 générations pour les groupes de races traditionnelles marocaines et iraniennes alors qu'il se situe à environ 500 générations pour le panel mondial.

#### 4.1.2. Comparaison des âges d'apparition et de sélection

D'après l'étude de notre atlas des âges de variants dans le génome du mouton (**voir Chapitre 3**), nous avons vu que les variants apparus il y a plus de 7000 générations sont rares. Aussi nous avons vérifié que les régions dont les sélections sont datées au-delà de cette limite présente bien des variants dans cette gamme d'âge (**Figure 55**).

- **Pour la régions R23** dont la sélection est datée entre 8000 et 12000 générations, on retrouve bien un pic de variants apparus il y a plus de 8000 générations au niveau de la région.

- **Pour la régions R10c** (sélection entre 6 et 10 000 générations) on ne retrouve pas de variants de plus de 7500 générations dans la région précisément mais quelques-uns dans une fenêtre de 2 Mb autour.
- **Pour la régions R9c** dont la sélection est datée à plus de 13 500 générations avec le groupe marocain (moins de 500 générations dans les deux autres groupes), on ne trouve aucun variant de plus de 8000 générations dans la région et dans une fenêtre de deux Mb autour.

Pour toutes les autres régions dont les dates de sélection sont inférieures à 7000 générations, on trouve une multitude de variants apparus dans cette gamme d'âge.

## 4.2. Groupe B :

Pour le groupe B nous avons fait des détections de sélection en intra-groupe, aussi nous avons réalisé les datations de chaque variant avec un seul groupe à chaque fois. Nous n'avons exécuté qu'une chaîne par variant au vu des 99% de répliquabilité donnés par l'analyse du groupe A, ce qui a permis d'étudier plus de variants compte-tenu des fortes contraintes computationnelles.

Les âges d'apparition et de sélection des variants ne sont pas corrélés ( $p$ -valeur = 0.33,  $R^2$  = 0.16, Test de Pearson) mais les âges d'apparition sont toujours plus anciens que ceux de la sélection (**Figure 56**). Tous les variants du groupe B présentent un âge de sélection inférieur à 350 générations et un âge d'apparition compris entre 4000 et 7000 générations sauf pour deux variants (**Figure 57**) :

- Le variant *cornes\_1* du gène *RXFP2* est apparu il y a 2074 générations. Sa fréquence est supérieure à 50% dans les races éthiopiennes et inférieure à 5% partout ailleurs.
- Le variant *fertilite\_2* du gène *TMEM154* est apparu il y a 660 générations. Sa fréquence est supérieure à 40% en Afrique et au Moyen-Orient et inférieure à 4% partout ailleurs.

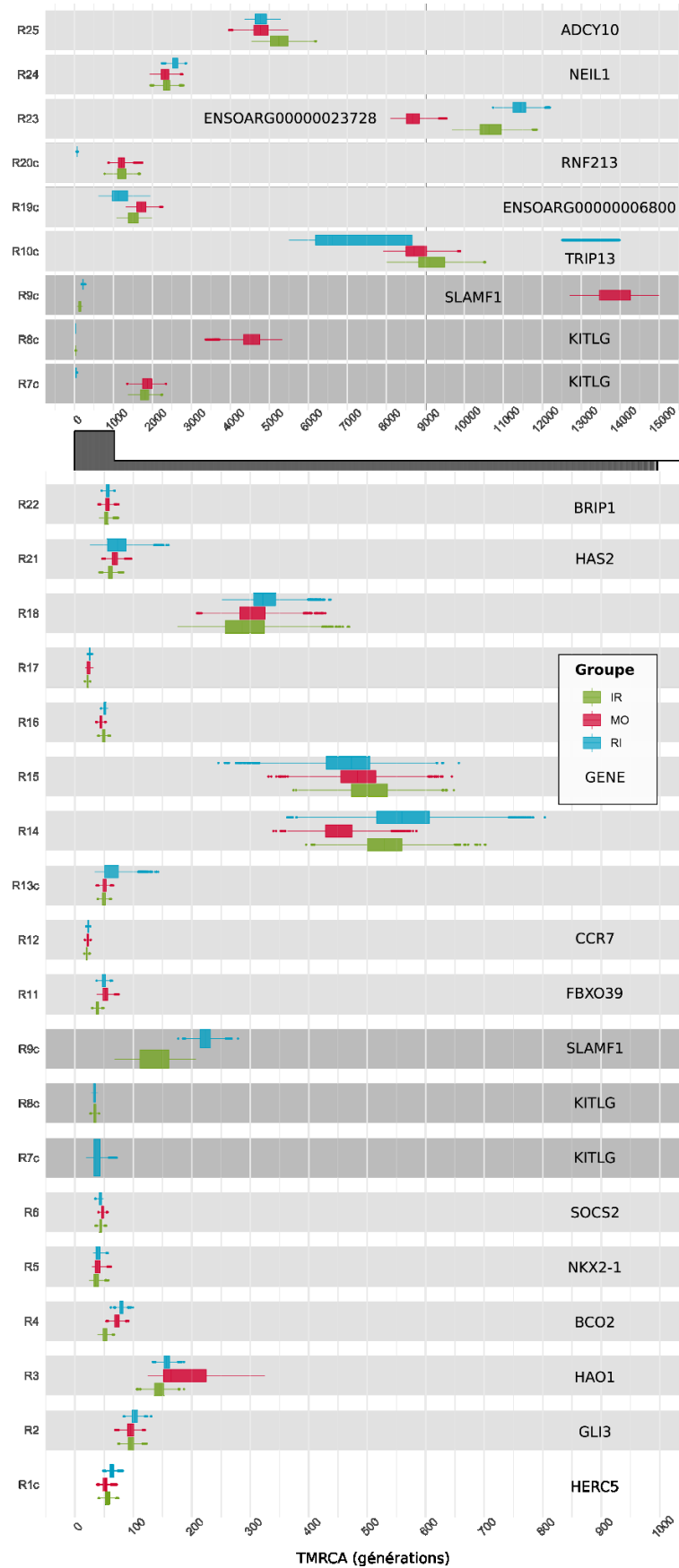
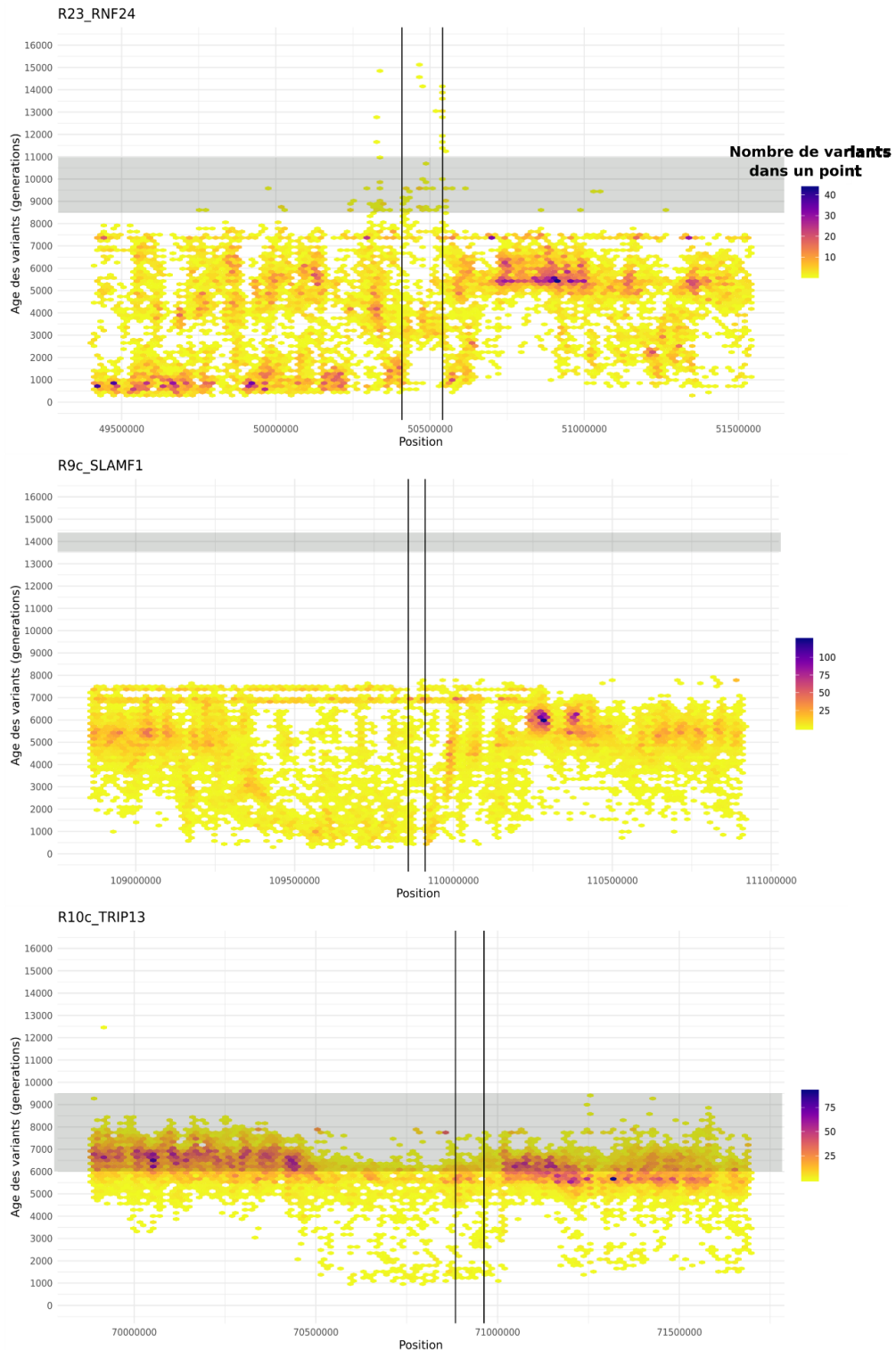


Figure 54 : Ages de la sélection des signaux de sélection cosmopolites (groupe A1)



**Figure 55 : Ages d'apparition des variants autour des signaux de sélection les plus anciens**  
 Les lignes verticales indiquent les bornes des régions identifiées sous sélection par Alberto et al. 2018. La zone représentée s'étend sur 1Mb de chaque côté des bornes de la région



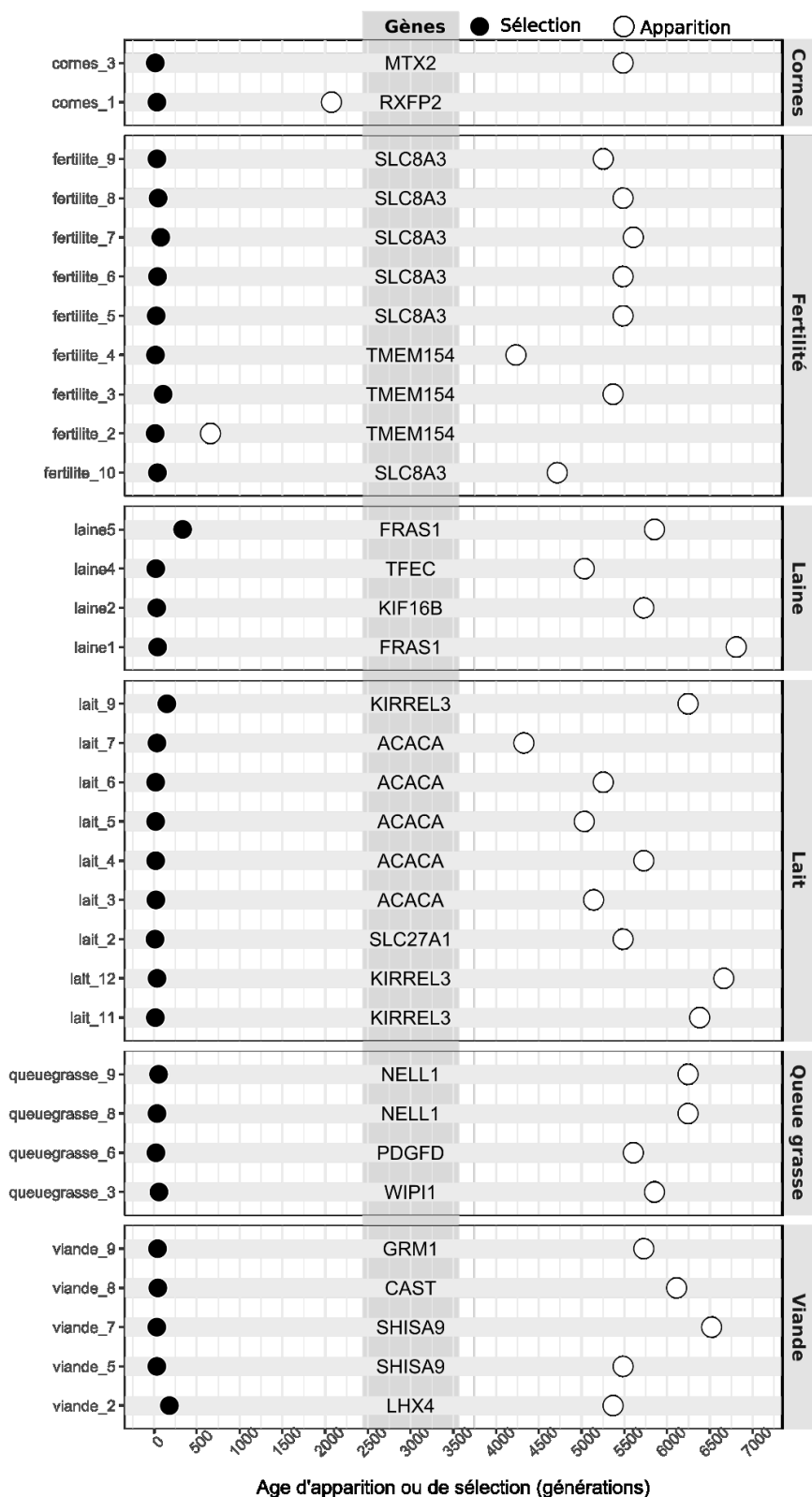


Figure 56 : Comparaison des âges d'apparition et de sélection des signaux de sélection intragroupe (groupe B)

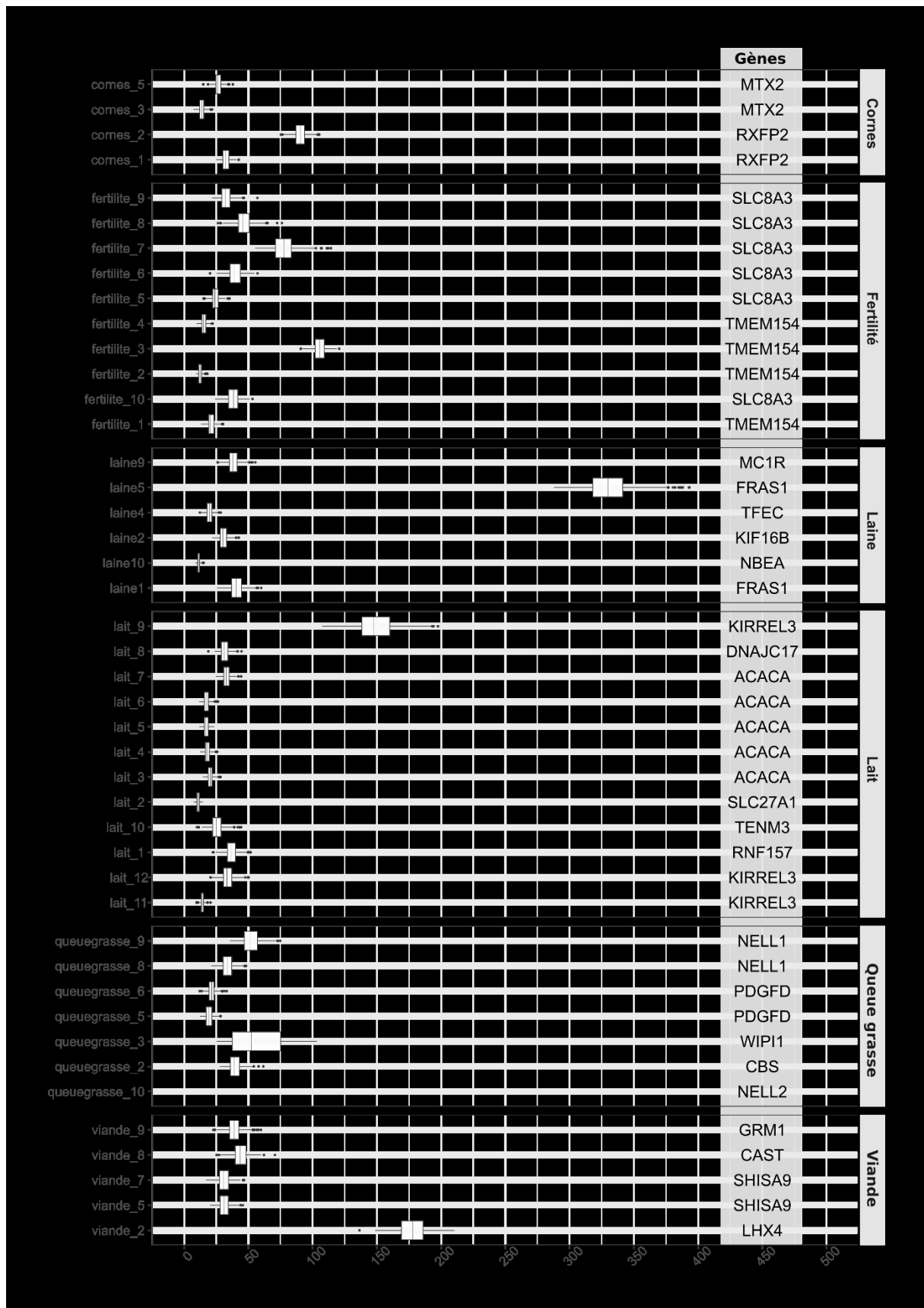


Figure 57 : Ages de sélection des signaux de sélection intra-groupe (groupe B)

## 5. Discussion

---

### 5.1. Groupe A : Signaux cosmopolites

---

La majorité des signaux de sélection cosmopolites (groupe A) présentent des âges de sélection très récents et donc largement postérieurs à la dispersion des moutons dans le monde. Plusieurs régions sont impliquées dans l'expression de traits agronomiques comme le lait, la laine ou la viande (R1c à R8). Ces régions ont pu être sélectionnées de manière indépendante dans les différentes populations à des temps identiques (ex : régions R1c à R6) ou différents (ex : régions R7c et R8c). Il en résulterait alors une convergence phénotypique basées sur des haplotypes différents en fonction de leur origine géographique.

La convergence phénotypique pourrait s'expliquer par l'intensification des communications, migrations et échanges entre les communautés humaines depuis le Néolithique jusqu'à l'Antiquité. On observe déjà à cette période une forme de mondialisation notamment via les transports maritimes et la place croissante de l'écriture. Ceci a certainement homogénéisé les besoins des hommes et mener à la sélection de traits similaires chez leurs animaux d'élevage en différents points géographiques.

Par ailleurs, la dispersion des moutons s'est faite par vagues successives ayant remplacé au fil du temps des races primitives par des races plus modernes (Chessa et al., 2009b; Pedrosa et al., 2005b; Tapio et al., 2006a). Il est possible que certaines régions aient été sélectionnées à un moment donné dans un endroit donné et qu'elles aient été propagées ensuite (I. Tunca and Taşkin, 2016; Meiri et al., 2017; Ollivier et al., 2016). C'est un scénario plus parcimonieux que l'hypothèse de plusieurs sélections indépendantes, puisqu'il ne nécessite qu'un événement de sélection au lieu de plusieurs dans différentes populations. Il pourrait être intéressant d'étudier ces régions chez certaines races primitives du nord de l'Europe comme les races Klövsjö, Värmland, Soay, Gute ou Roslag. Ces dernières sont situées aux marges des fronts de diffusion et présentent des caractéristiques morphologiques et génétiques (Chessa et al., 2009b; Rochus, 2017) qui indiquent qu'elles ont pu être épargnées par les vagues de remplacement. Il est donc possible qu'elles aient gardé des traces des sélections les plus anciennes. Si l'hypothèse de remplacement est validée, nos résultats montrent que c'est un phénomène à large échelle qui touche aussi bien des races traditionnelles que plus industrielles puisque les trois groupes (traditionnels iranien ou marocain et le panel mondial) et la majorité des régions de notre étude sont concernées.

Pour discriminer ces deux scénarios (convergence phénotypique ou remplacement), il faudra étudier les haplotypes de chacun des groupes pour savoir s'ils sont identiques ou non. Des haplotypes identiques témoigneraient d'une origine commune et soutiendraient plutôt l'hypothèse de remplacement. Tandis que des haplotypes différents indiqueraient des origines distinctes et plaideraient donc pour l'hypothèse de convergence phénotypique ou de sélection sur variation génétique préexistante. Les arbres de distances construits à partir des régions sous sélection nous indiquent que les haplotypes sélectionnés des différents groupes

sont très proches mais il n'est pas exclu que les différences se situent plus loin sur les haplotypes. Nous les étudierons donc sur au moins la totalité de l'intervalle fourni pour la datation par *startmrca*.

Des âges de sélection récents pourraient aussi être dus à des pressions successives s'intensifiant avec les nouveaux besoins de performances des sociétés humaines actuelles. De tels phénomènes auraient refixé des combinaisons génétiques aux abords des régions sous sélection. Autrement dit, les signaux de sélection initiaux sont effacés et ceux dont nous estimons l'âge sont les derniers.

Sur les huit régions qui contiennent des signaux de sélection convergents avec la chèvre, six sont parmi les plus anciennes que nous ayons daté. Trois présentent même des signaux antérieurs ou égaux à la date de la domestication. Les régions R10, R23 et R25 ont été sélectionnées il y a plus de 4500 générations, soit 9000 ans, dans les trois groupes (marocain, iranien et panel mondial). Le scénario le plus parcimonieux les concernant est celui d'une sélection au début de la domestication, au Moyen-Orient, et avant la diffusion des différentes populations dans les autres continents. Ces régions auraient ensuite été conservées durant la colonisation du reste du monde. Deux d'entre elles sont impliquées dans l'immunité, le comportement et la fertilité, ce qui indique que ces traits ont été la cible d'évènements de sélection précoces.

La région R25 est associée au gène *ADCY10* qui est impliqué dans la spermatogenèse et le taux de testostérone (Ruth et al., 2020; Serrano et al., 2021). La testostérone est notamment liée au comportement agressif des mâles (Illius, Haynes and Lamming, 1976; Ruiz-de-la-Torre and Manteca, 1999). La sélection de cette région pourrait donc être liée à la sélection des animaux les plus dociles lors des étapes initiales de la domestication.

Les régions R23 et R10 présentent des estimations d'âge de sélection très anciens situés entre 6000 et 15 000 générations soit 12 000 à 30 000 ans. Or, on s'attend à ce que les sélections de ces régions soient au moins aussi récentes que la domestication il y a 10 500 ans. Nous ne remettons pas en cause le statut ancien de ces sélections mais il est probable que la précision des estimations soit altérée pour les évènements les plus anciens. En effet, *STARTMRCA* tire parti du déclin du DL et du nombre de mutations qui s'accumulent autour d'un variant pour estimer la date de sa sélection. Or, plus cette dernière est ancienne plus l'estimation repose sur le taux de mutation étant donné le raccourcissement extrême de l'haplotype ancestralement sélectionné. Le taux de mutation que nous avons appliqué est constant et son estimation n'est pas forcément exacte. Ainsi les estimations les plus anciennes peuvent fortement varier en fonction de ce taux. Pour le vérifier nous pourrions relancer l'analyse *STARTMRCA* sur des régions d'âges différents en faisant varier le taux de mutation et ainsi estimer l'influence de ce paramètre en fonction des âges de sélection. En attendant que de telles vérifications soient faites nous affirmons simplement que les régions R23 et R10c sont des régions sélectionnées très précocement dans l'histoire de la domestication du mouton.

La région R10c est associée au gène TRIP3 qui est impliqué dans la susceptibilité aux mammites (Durán Aguilar et al., 2017) et donc à la traite. Cela pourrait indiquer une utilisation précoce du lait dès les premières étapes de la domestication confirmant ainsi d'autres preuves biochimiques, isotopiques et paléogénétiques de l'utilisation du lait dans les premières civilisations néolithiques (Vigne and Helmer, 2007). Si la datation très ancienne de la sélection de cette région était confirmée, elle pourrait aussi appuyer l'hypothèse du lait comme une des raisons potentielles de la domestication du mouton au lieu de la viande qui était déjà accessible via la chasse (Vigne, 2017).

La région R9c est aussi une région présentant un âge de sélection très ancien mais seulement pour le groupe marocain. Elle est associée au gène SLAMF1 qui est impliqué dans l'immunité. L'immunité pourrait donc être une des premières cibles de sélection due à la domestication. Les troupeaux formés par les humains ont été forcés à une promiscuité accrue, créant ainsi des conditions propices au développement et à la transmission des maladies. L'étude de l'âge d'apparition des variants autour de cette région n'a révélé aucun variant aussi vieux que la sélection de la région. Il est possible que nous n'ayons pas pu dater un tel variant, d'autant plus que nous avons sous-échantillonné le panel marocain et donc aussi la diversité génétique de ce groupe pour la création de l'Atlas des âges de variants (voir **Chapitre 3, partie 3.2**).

D'autres régions présentent des âges de sélection différents en fonction des groupes. C'est notamment le cas des régions R7c et R8c, toutes deux associées au gène KITLG. Dans les deux cas, un ou deux groupes de races traditionnelles (Iraniennes et/ou Marocaines) présentent un âge de sélection bien plus ancien que le groupe de races du panel mondial. On peut émettre l'hypothèse que ces races traditionnelles portent l'empreinte d'une sélection précoce sur le gène KITLG alors que les races du panel mondial ont subi d'autres pressions de sélection ou des remplacements au cours des deux derniers siècles. De manière intéressante, les sélections plus anciennes diffèrent de plus de 4000 ans alors qu'elles sont situées dans un même gène. Elles se comportent donc de manière indépendante malgré leur proximité physique. La sélection de la région R8c date d'environ 9000 ans pour le groupe marocain alors que celle de la région R7c ne remonte qu'à 3600 ans environ pour les groupes marocains et iraniens.

La région R8c a donc pu être sélectionnée au cours des stades initiaux de la domestication au Moyen-Orient. KITLG est d'ailleurs un gène qui pourrait être impliqué dans des caractéristiques comme la couleur ou le comportement (An et al., 2015; Conti and Cattaneo, 2010; Reissmann and Ludwig, 2013). La sélection de la région R7c a pu survenir en dehors du centre de domestication puisqu'il y a 3600 ans, le mouton est présent sur tous les continents de l'Ancien Monde. Cependant, l'estimation de sa sélection à 3600 ans est vraie à la fois pour des races Iraniennes, donc au niveau du centre de domestication et pour des races marocaines, à l'extrémité d'une des voies de colonisation méditerranéenne. Des sélections indépendantes à la fois convergentes et simultanées semblent peu probables, ce qui nous

amène à l'hypothèse d'une origine unique et commune. Cette supposition peut être appuyée par les faits suivants : Il semble que le Nord de l'Afrique ait subi, comme en Europe, deux vagues de diffusion. La première concernait des races à queue fine et la deuxième plutôt des races à queue grasse venues du Moyen-Orient qui auraient pu remplacer, ou au moins s'hybrider avec les moutons à queue fine déjà présents (Muigai and Hanotte, 2013). Notre groupe d'individus marocains ne contient pas de race à queue grasse mais cela n'empêche pas l'occurrence d'hybridation avec de telles races dans le passé. L'apparition des moutons à queue grasse est attestée il y a environ 4000 ans en Egypte par des représentations sur des fresques de la XII<sup>ème</sup> dynastie (Epstein, 1971). Il est donc possible que le signal de sélection que nous observons chez les moutons marocains, similaire à celui de races Iranienne et datant de 3600 ans ait été apporté par cette vague de diffusion de moutons à queue grasse à la même période.

La région 20c présente le même phénomène que les régions de KITLG. Le signal de sélection des groupes de races traditionnelles marocaines et iraniennes est bien plus vieux que celui du panel mondial : 2400 et 140 ans AP respectivement. Cette région est associée au gène RNF213 qui est impliqué dans la saisonnalité de la reproduction (Martinez-Royo et al., 2017) qui apparaît tôt dans le processus de domestication chez les mammifères en général (Trut, Oskina and Kharlamova, 2009; Trut, Plyusnina and Oskina, 2004) et chez le mouton en particulier (Tornerio et al., 2020, 2016). Ce n'est pas le cas pour cette région dont la sélection date de près de 8000 ans après les premières traces de domestication. La saisonnalité de la reproduction peut être un facteur important limitant la production, et donc un paramètre cible de sélection volontaire par l'homme à un stade déjà avancé de domestication dans les groupes de races traditionnelles. La datation très récente de la sélection de la région R20c (68 générations) dans le panel de races mondial pourrait résulter d'un autre processus de sélection : la saisonnalité des naissances peut être contrôlée par l'administration d'hormones, mais la demande de produits sans hormones augmente et mène à la recherche de marqueurs génétiques et la sélection subséquente des animaux les plus adaptés à ces besoins de productivité (Martinez-Royo et al., 2017). D'autre part, le gène RNF213 a aussi été retrouvé associé à des adaptations au climat désertique (Yang et al., 2016b). Or, il se trouve que la région MENA (Middle East and North Africa) qui couvre le nord de l'Afrique et le Moyen-Orient a subi une terrible sécheresse entre 3200 et 2100 ans AP environ, soit en même temps que la sélection de la région du gène RNF213 dans les groupes Iranien et Marocain. Cette sélection pourrait donc être due à une adaptation à ce nouveau climat. De plus, une étude montre que certaines sociétés se sont préparées à cette sécheresse en sélectionnant des vaches plus résistantes aux conditions sèches (Finkelstein et al., 2017). Peut-être en a-t-il été de même pour les moutons.

## 5.2. Groupe B : Signaux intragroupes

L'étude des signaux de sélection intragroupes (groupe B) nous a permis de montrer l'absence de corrélation entre les âges d'apparition et de sélection de variants bien précis (et

non des régions comme dans le groupe A. Ainsi, GEVA et STARTMRCA datent bien des événements différents, apparition et sélection respectivement. Cela est cohérent avec le fait qu'aucun âge de sélection n'est antérieur à l'apparition du variant concerné.

Les signaux de sélection des variants du groupe B sont tous très récents et aucun n'atteint d'âge supérieur à 350 générations. Cela indique que la sélection de variants génomiques au sein de groupes local ou de races s'est intensifiée récemment et qu'elle a pu effacer la trace de signaux plus anciens, surtout dans des races standardisées où la dérive et la sélection sont fortes. La rationalisation de l'élevage et la sélection systématique de certains traits dans un but agronomique se sont fortement développés au 18<sup>ème</sup> siècle en même temps que la révolution industrielle. La notion de race comme groupe standardisé sur la base de caractéristiques précises apparaît à ce moment-là (Verrier, Brives and Heams, 2011). Cependant, la création de groupes particuliers peut aussi remonter à plus de 200 ans, même si le terme de « races » n'était pas encore d'actualité. Un bon exemple est celui de la création de la Mesta en 1273 pour le développement de ce qui deviendra la race Mérinos que l'on connaît aujourd'hui (voir **Introduction générale, partie 0**).

Si les âges de la sélection des variants du groupe B ne varient que de quelques centaines de générations, ceux de leur apparition s'étendent sur des milliers. La majorité des âges d'apparition se trouvent entre 4500 et 7000 générations, soit au début de la domestication et probablement avant la diffusion des moutons dans le monde. Deux seulement sont apparus bien plus tard et nous permettent d'estimer le lieu de leur origine en fonction de leur fréquence actuelle dans les différents continents où les moutons ont été introduits.

Nous avons daté l'apparition et la sélection d'un variant du gène RXFP2 (cornes\_1) montrant un signal de sélection dans la race éthiopienne Amhara-Adane qui est caractérisée par l'absence de cornes. RXFP2 est un gène bien connu pour son rôle dans la forme, la taille et la présence ou l'absence des cornes chez le mouton (Dominik, Henshall and Hayes, 2012; Pan et al., 2018b; Wang et al., 2014a; Wiedemar and Drögemüller, 2015) et chez des espèces sauvages ou férales d'*Ovis* (Johnston et al., 2013, 2013, 2011; Kardos et al., 2015). La sélection de ce variant date de 33 générations soit environ 60 ans et son apparition d'environ 4000 ans soit après la diffusion des populations ovines dans le monde. Il est probablement apparu en Afrique où on le retrouve en fréquence élevée surtout dans les races éthiopiennes (supérieure à 50%). En Asie, Europe et Moyen-Orient, sa fréquence ne dépasse pas 5%.

Le variant fertilité\_2 apparaît encore plus récemment. Il ne date que de 660 générations, soit environ 1200 ans AP. Ce variant est donc apparu dans des populations domestiques bien après la diffusion des moutons dans le monde entier, puisqu'il est absent chez les mouflons. Il est probablement originaire d'Afrique ou du Moyen-Orient où il est présent à une fréquence supérieure à 40% alors qu'il est rare en Europe et en Asie (fréquence inférieure à 4%). Ce variant est associé au gène TMEM154 qui est impliqué dans la sensibilité au virus Visna/Maedi (VMV) (Heaton et al., 2012). Ce virus impacte la santé, la productivité et

la survie des animaux via de multiples symptômes respiratoires, articulaires, neurologiques (etc.) (Thormar, 2004). Cette maladie a vraisemblablement été introduite via l'importation de moutons Karakul en Islande depuis l'Allemagne en 1933 (Straub, 2004) puis dans une grande partie du cheptel mondial (Pépin et al., 1998). Dans les années 70, des plans de lutte contre le virus se mettent en place (Dion, 1994). Plusieurs techniques ont été envisagées. Les agneaux peuvent être retirés de leur mère si elle est positive au virus étant donné qu'il se transmet par le lait et le colostrum (Houwens et al., 1983). Des projets de vaccins ont aussi vu le jour mais sans succès notamment à cause de la variation génétique du virus et son interaction complexe avec son hôte (Montelaro et al., 1989; Pearson, Poss and Demartini, 1989). Une autre approche envisagée est la sélection de moutons présentant une résistance naturelle au VMV (Perk et al., 1996). Plusieurs études ont montré l'existence de loci impliqués dans la sensibilité à ce virus, et notamment dans le gène TMEM154 (Heaton et al., 2013, 2012; Molaee, Eltanany and Lühken, 2018; Yaman et al., 2019) dans lequel nous avons trouvé des signaux de sélection forts dans des races éthiopiennes. Les informations concernant la prévalence du virus en Afrique sont rares. Cependant, on reporte en Ethiopie des foyers de maladies respiratoires avec une mortalité élevée et des symptômes similaires à ceux provoqués par le VMV à la fin des années 90 (Mukasa-Mugerwa et al., 2000; Tibbo, Woldemeskel and Gopilo, 2001). Il a été montré que plus de 50% de la mortalité ovine étaient due à ce type de maladie qui constitue un problème important depuis 20-30 ans dans ce pays. De plus, l'incidence et le taux de mortalité dus aux maladies respiratoires ont été multipliés par 10 entre 1996 et 1999 en Ethiopie (Woldemeskel, Tibbo and Potgieter, 2002). Ayelet et al. (2001) ont montré que le VMV est effectivement présent dans la région. La sélection de trois variants du gène TMEM154 que nous avons trouvés dans les races éthiopiennes est très récente et date seulement de 12, 15 et 20 générations, soit environ 24 à 40 ans. Elle pourrait donc coïncider avec l'émergence du VMV.

Le signal de sélection le plus ancien que nous ayons trouvé dans ce groupe concerne le variant laine5 dans la race de mouton chinoise Hu. Ce variant est associé au gène FRAS1 qui est impliqué dans l'adhésion dermique et épidermique et la formation du follicule pileux (Bolormaa et al., 2017; Clements et al., 2012). La race Hu est réputée mondialement pour sa prolificité mais aussi pour la production d'objets de luxe et traditionnels confectionnés avec la toison des agneaux Hu qui est très particulière et rare. En effet, si elle est prélevée le jour de la naissance, les peaux de ces agneaux présentent des stries bien distinctives qui se conservent même après leur traitement (**Figure 58**) (Chen, 1990).



Figure 58 : Aspect de la laine des agneaux de race Hu  
D'après la figure 1 de Wu et al. 2021



Il a été montré que ce motif particulier est relié au développement du follicule pileux dont un des composant, la papille dermique joue un rôle important (Wu et al., 2021). FRAS1 pourrait donc bien être impliqué dans ce phénotype et sa sélection daterait du 15<sup>ème</sup> siècle. Pour le confirmer, il serait utile d'étudier et de dater les signaux de sélection des gènes PAPP, BMP7, MMP2, SNAI1, SFXN1, CDKNIC, MT3, et POU1F12 qui présentent des différentiels d'expression lié au phénotype des toisons d'agneaux Hu (Wu et al., 2021). L'étude des variants impliqués dans ce phénotype rare est d'autant plus importante que la race Hu est en danger d'extinction. Récemment, des croisements et le changement de priorité de production vers la viande détériore peu à peu la qualité des toisons d'agneaux (Chen, 1990; Sun et al., 2013). Un phénotype, qui, par son âge de sélection, pourrait être un héritage de la dynastie Ming est donc un élément important du patrimoine culturel chinois.

## 6. Conclusion

---

Les datations de signaux de sélection cosmopolites montrent que des traces de sélection anciennes persistent dans certaines populations. La datation de la sélection de signaux intra-groupes liés à différents caractères nous montre qu'ils résultent de processus récents masquant ou non des signaux plus anciens.

Un point intéressant qui pourrait être exploré par ce type d'analyse de datation est l'étude de l'apparition de phénotypes propres au mouton, c'est-à-dire inexistant chez les mouflons. Ce peut être le cas de phénotypes en lien avec trois caractères : les cornes, la toison et la queue. Pour les cornes les moutons peuvent présenter des cornes en hélice, avoir quatre cornes ou aucune (même chez les mâles). Pour la toison, les variations de couleurs et la présence de laine au lieu de poil sont uniquement retrouvées chez les moutons. Les mouflons ont du poil et une robe fauve qui leur permet de se camoufler des prédateurs. Le phénotype queue grasse est uniquement présent chez certaines races domestiques, alors que les mouflons présentent une queue fine. La question serait de remonter au plus proche des premières sélections, voire des apparitions de ces phénotypes. Cela permettrait de préciser les contextes dans lesquels se sont développés ces traits ainsi que les raisons de leur développement (climatique, culturel, agronomique etc.).

L'Approche pour y répondre ressemble à celles que nous avons utilisé dans ce chapitre mais en sélectionnant des régions génomiques différemment. Pour la laine par exemple, nous pourrions cibler des régions qui présentent des signaux de sélections communs entre de multiples races à laine d'origine géographique différentes et dans des populations primitives qui auraient pu garder des traces des sélections initiales comme nous l'avons montré pour certaines régions du groupe A dans des groupes de races traditionnelles iraniennes et marocaines.

# Chapitre 5

## Discussion générale

---

Dans ce dernier chapitre, nous récapitulons les principaux résultats obtenus grâce à ce travail. Nous mettons en lien les résultats des différentes parties et démontrons-en quoi ils permettent de valider l'utilisation de génomes actuels dans la datation d'évènements de mutation et de sélection passés.

## 1. Récapitulatif des principaux résultats obtenus

---

L'ensemble de ce travail s'est attaché à explorer les informations temporelles contenues dans des génomes d'animaux contemporains. Nous avons ainsi estimé l'âge de différents évènements correspondant à deux forces évolutives : la mutation et la sélection, et qui ont pu nous renseigner plus globalement sur l'impact des deux autres, la dérive et les flux de gènes. Il s'agit d'un travail inédit sur une espèce autre que l'humain, dont les résultats fournissent une base de données conséquente et disponible pour aider à la compréhension de processus évolutifs tels que la domestication mais aussi pour des applications plus larges notamment en agronomie.

La datation de plus de 35 millions de variants génomiques (SNPs) montre qu'ils sont apparus en majorité entre 9000 et 14000 ans AP, soit autour des étapes initiales de la domestication (**Figure 59.C**). Les variants spécifiques aux différents continents s'accumulent progressivement depuis ces premières étapes jusqu'à environ 3000 ans où la distribution de leurs âges présente un pic (**Figure 59.B**). Parmi ces variants spécifiques, certains sont présents dans toutes les régions de ces continents, ce sont les variants continentaux partagés. La distribution de leurs âges est similaire à celle de l'ensemble des variants continentaux sauf pour l'Asie. Dans ce cas, on ne trouve pas de variants partagés entre toutes les régions asiatiques autour de 3000 ans (**Figure 59.B**).

Nous avons aussi étudié les proportions de partage de génomes entre les différents groupes de notre jeu de données au cours du temps. Nous avons ainsi mis en évidence plusieurs ruptures brutales et généralisées des connexions entre les groupes ovins de l'Ancien Monde à des périodes coïncidant avec de grandes crises de l'Humanité telles que l'effondrement de l'âge du Bronze et la Peste de Justinien (**Figure 59.A**). Ces crises ont des causes multiples (évènements climatiques extrêmes, guerres, invasions, épidémies etc.) et généralisées qui ont provoqué l'arrêt du commerce et des échanges internationaux (Cline, 2016; Meier, 2016).

D'autre part, nous avons inféré la chronologie d'évènements de sélection liés à la domestication sur un ensemble de gènes candidats prédéfinis (**Figure 59. D & E**). La plupart des signaux sont très récents (moins de 1000 ans AP) et suggèrent l'importance des processus de remplacements et de sélections répétées dans l'effacement des empreintes de sélection initiales. Cependant, certaines populations conservent les traces de signaux anciens. Ces résultats encouragent l'étude de races primitives (dans des régions peu ou pas affectées par

les vagues de diffusion successives, au Nord de l'Europe ou à l'Ouest de l'Afrique) ou même dans des populations férales (i.e. mouflons méditerranéens) pour retrouver les traces de sélection précoces disparues dans les races modernes. Parmi les signaux anciens identifiés, certains sont impliqués dans le comportement et la couleur, ce qui est attendu dans les premières étapes de domestication, mais aussi dans la traite et l'immunité. Nous avons aussi montré que certains signaux de sélection pourraient coïncider avec des modifications de l'environnement (événements climatiques ou des épidémies).

Nous avons démontré que les âges de sélection sont bien toujours postérieurs à l'âge d'apparition des variants et que les deux ne sont pas corrélés, ce qui indique que les deux méthodes (*STARTMRCA* et *GEVA*) datent bien des événements distincts. De plus, la majorité des variants sous sélection sont apparus dès le début de la domestication sauf pour deux d'entre eux (**Figure 59.D**).

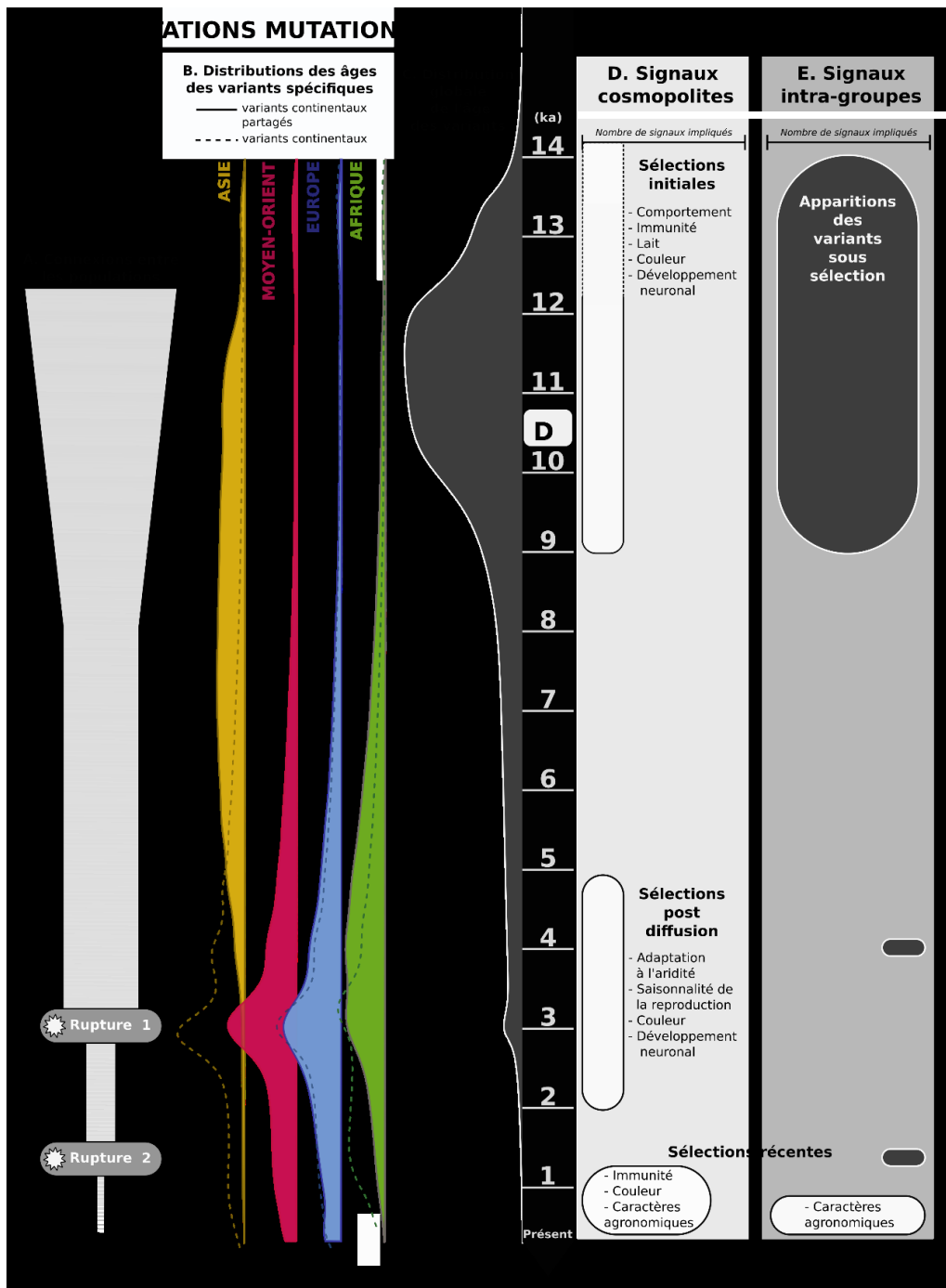


Figure 59 : Récapitulatif des principaux résultats obtenus.

L'axe central représente le temps en milliers d'années. Le « D » indique la période de domestication initiale du mouton. Chaque partie de la figure schématise les résultats obtenus dans les différentes parties du manuscrit : **A. Chapitre 3, 5.2.** Détermination de la fraction de génome partagé entre un génome cible et un génome comparateur via l'analyse CCF (Albers et McVean 2020) . **B. Chapitre 3, 5.1.** Les variant continentaux sont des variants présents uniquement dans un continent. Les variants continentaux partagés sont des variants continentaux présents dans toutes les régions du continent **C. Chapitre 3, 4.3.2.** L'âge de plus de 35 millions de variants a été inféré avec la méthode GEVA (Albers et McVean 2020) **D. Chapitre 4, 3.1.1.** Les signatures de sélection partagées entre deux groupes de races traditionnelles marocaines et iraniennes et un panel de races mondiales ont été datées avec la méthode startmrc (Smith et Coop 2019). Les pointillés de la période de sélection initiale indiquent une incertitude dans les estimations. **E. Chapitre 4, 3.1.2.** Les signatures de sélection de signaux intra-groupes détectés par la méthode nsl ont aussi été datés startmrc. Leurs âges d'apparition dérivent de ceux inférés dans la partie C de la figure.

## 2. Informations apportées par la combinaison des résultats des différents chapitres

---

### 2.1. Origines de la sélection

---

Plusieurs méthodes ont été développées pour déterminer si un variant a été sélectionné à partir de variation préexistante ou d'une mutation de novo (Barrett and Schluter, 2008), mais selon Stapley et al. (2010), la clé pour y parvenir est de comparer l'âge de la mutation et l'âge du changement (environnemental ou autre) qui impose une pression de sélection. C'est ce que nous avons mis en place grâce aux analyses des chapitres 3 et 4 et qui permettent la comparaison des âges d'apparition (âge de la mutation) et des âges de sélection (proxy de l'âge du début d'application de la pression de sélection) de variants génomiques.

Nous avons montré que la plupart des variants ont été sélectionnés des millénaires après leur apparition (**Figure 59.E**). Cela indique que la majorité de la sélection s'est faite sur de la variation préexistante. Ce résultat n'est pas nouveau et a été démontré de multiples fois (Barrett and Schluter, 2008; Karasov, Messer and Petrov, 2010). Il n'est pas non plus inattendu dans un modèle de domestication qui s'étend sur une échelle de temps courte. En effet, la sélection à partir de variations génétiques préexistantes présente l'avantage d'être plus rapide puisqu'elle se fait à partir d'allèles déjà disponibles dans la population (Tigano and Friesen, 2016).

Deux variants présentent néanmoins une apparition tardive autour de 4000 et 1300 ans AP (**Figure 59.D**). Dans ce cas nous pouvons préciser l'origine géographique de la mutation sélectionnée. En effet, leur apparition intervient alors que les moutons ont déjà diffusé dans le monde. Si les flux de gènes existants entre populations ovines (Kijas et al. 2012) sont suffisamment modérés depuis ces apparitions récentes, on peut donc supposer que la fréquence allélique du variant est encore plus élevée dans la région où il est apparu, (Kijas et al., 2012a). Si nous reprenons l'exemple du variant du gène *TMEM145* apparu il y a environ 1300 ans (**Figure 59.E, Figure 57**), nous constatons qu'il est présent à plus de 40% dans les populations d'Afrique et du Moyen-Orient alors que sa fréquence ne dépasse pas les 4% en Europe et en Asie. Selon notre hypothèse, nous pouvons supposer que son origine est africaine ou moyen-orientale.

Finalement la comparaison des âges d'apparition et de sélection d'un variant génomique permet d'identifier son origine géographique mais aussi la source de la sélection (mutation de novo ou variation préexistante).

### 2.2. Périodes de sélection versus périodes de flux de gènes

---

Globalement, nous avons mis en évidence trois grandes périodes de sélection (**Figure 59 D & E**). La première concerne les sélections pendant les phases initiales de la domestication, entre 14 000 et 9000 ans AP. La seconde intervient dans une phase postérieure à la diffusion des moutons dans le monde, entre 5000 et 2000 ans AP. Et la dernière concerne des sélections récentes datant de moins de 1000 ans AP. De manière intéressante, on remarque que ces périodes de sélection sont séparées par des périodes de connexion entre les populations ovines que l'on identifie dans nos analyses par l'existence de fractions partagées de génome en proportion stable entre ces populations

Rappelons qu'une de nos hypothèses pour expliquer la présence majoritaire de signaux récents ainsi que d'âges de sélection différents en fonction des populations, est le remplacement. Le remplacement de populations par des vagues de diffusion consécutives est un phénomène connu chez les animaux domestiques (Chessa et al., 2009a; Librado et al., 2021). Chez le mouton, au moins deux vagues ont été identifiées en Europe et en Afrique (Chessa et al., 2009a; Deng et al., 2020; Muigai and Hanotte, 2013). Or, une vague de diffusion implique l'arrivée de migrants au sein de populations déjà installées et donc des flux de gènes entre les deux, qui aboutissent à un remplacement plus ou moins complet de la diversité originelle.

La présence de périodes distinctes de flux de gènes soutient cette hypothèse. (**Figure 59.A**). De plus, la période de connexion entre les sélections initiales et post-diffusion (**Figure 59 A et D**) s'étend de 8000 à 3000 ans AP. Cela comprend les périodes des différents seconds épisodes migratoires qui ont eu lieu dans les continents occidentaux : de 6000 à 3000 ans AP environ en Europe, et entre 5300 et 1700 ans AP en Afrique (Deng et al., 2020).

Le fait que nous détectons moins de sélection durant les périodes de flux de gènes n'implique pas obligatoirement l'absence de sélection à ces moments. Elle pourrait impliquer plutôt le masquage de cette sélection par le flux de gènes. En effet, ces épisodes migratoires semblent avoir été de grande ampleur puisqu'on en retrouve les traces dans la majorité des populations actuelles (Chessa et al., 2009a; Rannamäe et al., 2016, 2020). On peut donc en déduire que le taux de migration a pu être supérieur aux coefficients de sélection des variants de populations locales (présentes avant le remplacement), ce qui induit une dilution de ces derniers dans le pool génétique migrant (Feder et al., 2012; Lenormand, 2002; Wright, 1931; Yeaman and Otto, 2011).

Nous ne pouvons cependant pas écarter l'hypothèse de sélections répétées pour expliquer l'abondance de signaux récents. La période de sélection la plus récente (moins de 1500 ans) comprend l'apparition du concept de race au 18<sup>ème</sup> siècle (Davidson, 2009; Derry, 2020). D'ailleurs, l'âge de sélection des signaux intragroupes se situe majoritairement après cette date (**Figure 57**). Dans ces cas-là, plutôt que le remplacement, nous pourrions invoquer le début de sélections intensives comme facteur d'effacement de signaux précédents.

## 2.3. Précisions sur les causes possibles de rupture des connexions

---

On observe à l'échelle globale, une rupture brutale des connexions intergroupes il y a environ 3000 ans (**Figure 59.A**). On observe aussi à cette période un pic dans la distribution des variants continentaux (variants présents uniquement dans un continent) (**Figure 59.B**). Cela confirme bien la rupture des échanges entre les continents et donc l'isolement génétique des groupes ovins qui les composent.

Pour les variants continentaux partagés, (présents uniquement dans un continent et dans toutes les régions qui le composent) on observe le même pic de distribution dans les âges des variants occidentaux. Comme ces variants sont présents dans toutes les régions des continents Africain et Européen, cela indique que les connexions n'ont pas été rompues au sein de ces continents mais surtout entre eux. En Asie, la situation est différente puisque l'apparition des variants continentaux partagés (présents uniquement en Asie à la fois au Nord de la Chine, sur le Plateau Tibétain et au Sud de l'Asie) diminue fortement autour de la période de crise de 3000 ans AP. Cela indique donc une rupture des connexions entre les régions d'Asie.

Dans notre discussion du chapitre 3, nous avons émis plusieurs hypothèses pour expliquer les ruptures de connexions à cette période. La première était celle des mouvements massifs de populations en Asie qui ont amené à la formation de plusieurs groupes ethniques (Fan, 1131; He, 2000; Ou, 1936; Ran, Li and Zhou, 1985, 1985; Si, 1131). Cette hypothèse était confirmée par l'étude de Zhao et al. (2017b) qui montre une divergence des moutons nord chinois et des plateaux Tibétains et du Yunnan Kweichow à cette période. Cette hypothèse n'expliquait pas le caractère généralisé des ruptures que nous observons. La seconde hypothèse était celle de l'effondrement global de toutes les grandes civilisations méditerranéennes et orientales qui a eu lieu il y a environ 3000 ans (Cline, 2016). En analysant conjointement les résultats des analyses de connexions entre les groupes au fil du temps et la distribution des âges de variants spécifiques, nous pouvons conclure que les deux hypothèses ne sont pas exclusives. La première explique une rupture entre les différents continents notamment via un arrêt du commerce international (Cline, 2016). La seconde explique la rupture des connexions entre les régions d'Asie.

## 2.4. Diffusion des empreintes de rupture de connexions

---

Nous avons vu que l'analyse des variants spécifiques continentaux partagés révèle qu'il n'y a pas forcément eu de rupture des connexions à l'intérieur des continents occidentaux (Afrique et Europe) puisqu'on observe aussi un pic dans la distribution des âges de variants partagés entre différentes régions d'un continent autour de 3000 ans (**Figure 59.B**). Or les analyses de connexions intergroupes (CCF) montrent bien une rupture entre toutes les paires d'individus analysées et donc entre les régions des continents (**Figure 59.A, Figure 48, Figure S1**).



L'analyse CCF permet de quantifier le nombre de variants dérivés que partage deux génomes à un temps donné. Elle prend donc en compte tous types de variants, et pas seulement les variants spécifiques au continent. Or, d'après cette analyse, on sait que les périodes de ruptures sont suivies de périodes de flux de gènes (**Figure 59.A**). De plus, il a été montré que les croisements ont été fréquents notamment durant le développement des races (Kijas et al., 2012a). Ainsi, les individus sont composés de variants formant une mosaïque des différentes régions du monde. Même si les ruptures de connexions n'ont pas eu lieu entre les régions d'un même continent au moment de la crise, les individus actuels, résultats de mélanges plus récents, en portent les traces car le nombre de variants cosmopolites (présents dans plusieurs régions ou continents) est bien supérieur au nombre de variants spécifiques. Les analyses CCF et de distributions des âges des variants spécifiques apportent donc des informations complémentaires.

## 2.5. Recommandation d'application des méthodes de datations

---

La précision des datations, que ce soient celles d'évènements de sélection ou de mutation, dépend de paramètres comme le temps de génération, le taux de mutation et le taux de recombinaison. Dans notre cas, nous disposons d'une carte de recombinaison permettant de prendre en compte la variation de ce taux le long du génome. Pour le taux de mutation, nous avons utilisé une valeur unique (correspondant à un taux typique des mammifères) (Kumar and Subramanian, 2002; Nachman and Crowell, 2000) bien que ce taux puisse varier dans le temps et le long du génome (Gonzalez-Perez, Sabarinathan and Lopez-Bigas, 2019; Lynch, 2010). Pour le temps de génération, nous l'avons fixé à 2 ans mais il est probable qu'il ait varié au cours du temps ou en fonction des races.

Malgré les incertitudes sur les valeurs attribuées à ces paramètres, l'étude d'un processus comme la domestication permet de vérifier la calibration des datations. Par exemple, dans notre analyse CCF (**Chapitre 3 partie 5.2**) nous avons estimé une séparation des trajectoires des sauvages et des domestiques autour de 10 000 ans, ce qui donne confiance dans les estimations des datations absolues. De la même manière, les âges de sélections sont cohérents avec l'apparition des races et des épisodes épidémiques ou climatiques. Dans le cas d'un modèle où aucune calibration n'est possible, nous conseillons de limiter l'interprétation des résultats aux datations relatives, autrement dit à l'ordre des évènements de sélection uniquement.

# 3. Perspectives

---

## 3.1. Paléogénomique

---

Les analyses que nous avons menées infèrent des évènements passés à partir de données actuelles. La suite logique est donc la combinaison de ces analyses avec celles portant sur de l'ADN ancien. D'autant que les données génétiques anciennes commencent à être

disponibles pour le mouton via le travail de l'équipe de Dan Bradley à Dublin (Hare, 2021), et de celle de Ludovic Orlando à Toulouse dans le cadre du projet EVOSHEEP (Vila et al., 2021). Ce projet combine des compétences archéologiques, génétiques et paléogénétiques de plusieurs experts pour étudier des assemblages de mouton datant de 8000 à 3000 ans AP issus d'Afrique et du Moyen-Orient.

L'inclusion de données anciennes aura d'abord un intérêt pour des questions de validation. En effet les âges d'apparition de variants pourraient être validés par vérification de leur présence dans des données issues de périodes postérieures aux âges inférés ici ou réfutés en cas de présence avant l'âge d'apparition estimé par nos analyses. Il serait également possible de réaliser des analyses de détection de sélection dans les données anciennes pour corroborer ou non les âges de sélection que nous avons estimés. Nous pourrions ainsi vérifier que les régions génomiques identifiées dans les données modernes sont retrouvées dans les données anciennes datant de périodes égales ou postérieures aux âges de sélection que nous avons estimés.

L'étude de l'ADN ancien pourrait aussi nous permettre de nous affranchir des effacements d'empreintes de sélections dus aux épisodes de remplacements et de sélections répétées. Ainsi, après avoir retrouvé certains signaux dans des données génomiques anciennes nous pourrions les dater directement à partir de ce point du passé et donc avant la disparition des empreintes initiales. Les données anciennes permettraient également de préciser l'origine géographique des variants sous sélection en étudiant leur distribution de fréquence au cours du temps et dans l'espace.

## 3.2. Elargissement du jeu de données

---

Un autre point qui pourrait être exploré est l'impact de l'élargissement du jeu de données. Albers et McVean (2020) ont montré, sur l'humain, que l'élargissement du jeu de données avait peu d'influence sur les estimations d'âge des mutations. Ils ont pour cela utilisé deux jeux de données indépendants, l'un comprenant 278 individus et l'autre plus de 2500 recouvrant tous deux une grande partie de la diversité mondiale mais avec des régions non représentées dans le premier (Océanie et Asie Centrale). Pour nos estimations nous avons utilisé 376 génomes, donc davantage que l'un des ensembles humains mais avec des manques de représentativité de certaines régions comme le Sud des continents asiatique et africain et le nord de l'Europe. L'élargissement du jeu de données à ce niveau pourrait (i) permettre de confirmer les estimations déjà faites et (ii) estimer l'âge de nouveaux variants absents de notre jeu de données de base. Ce deuxième point pourrait être utile ponctuellement pour la datation de variants d'intérêts, mais nous pensons que l'information générale portée par la datation de plus de 35 millions de variants n'en sera que peu affectée. D'ailleurs, avec une densité moyenne de 5 SNPs/kb, il sera possible d'approximer l'âge d'un variant non daté par l'âge des variants voisins étant donné l'ordre de grandeur du déséquilibre de liaison chez le mouton (plusieurs kb) (Al-Mamun et al., 2015; Liu et al., 2017).

### 3.3. Datation de signaux cosmopolites en Asie

---

Pour les datations de signaux de sélection cosmopolites, nous n'avons pour l'instant utilisé que la partie occidentale de notre jeu de données. Nous avons vu que les voies de diffusion occidentales (africaines et européennes) n'étaient pas indépendantes alors que peu de liens ont été trouvés avec les voies Asiatiques (Ciani et al., 2014; Fregel et al., 2018; Kandoussi et al., 2020). Une indication de cette séparation se retrouve notamment au niveau de la distribution des haplogroupes mitochondriaux, majoritairement de type B et A en Occident et en Orient respectivement (Gornas et al., 2011; Horsburgh and Rhines, 2010; Tapio et al., 2006a). Les différences entre ces deux grandes voies de diffusion depuis le centre de domestication pourraient donc aussi se retrouver au niveau des processus de sélection. Nous pourrions dans un premier temps, confirmer la présence de ces signaux en Asie pour savoir si les mêmes régions génomiques ont été sélectionnées à l'Est et à l'Ouest ou au Moyen-Orient avant la diffusion. Si nous les retrouvons, nous pourrions les dater pour voir si leur sélection date de la même période ou non.

D'autres part, nous disposons de populations élevées de manière traditionnelle par des peuples nomades isolées des échanges internationaux sur le Plateau Tibétain. Nos analyses ont montré que des signaux de sélection anciens persistaient dans des races élevées de manière traditionnelles. Comme c'est aussi le cas des races du Plateau Tibétain, il faudrait tester la présence de sélection anciennes dans ce groupe.

### 3.4. Application à la datation de signatures d'adaptation

---

Nos analyses de détection de sélection se sont limitées à des gènes liés à des caractéristiques agronomiques, physiologiques ou comportementales sous l'influence d'une sélection humaine directe. Les signatures de sélection liées à des adaptations aux facteurs environnementaux non anthropiques n'ont pas été recherchées. Plusieurs études ont d'ailleurs déjà identifié des régions génomiques impliquées dans des adaptations à l'altitude, au climat désertique etc. (Serranito et al., 2021a; Wei et al., 2016; Yang et al., 2016b). Pourtant elles pourraient présenter des dynamiques différentes notamment dans le cas d'adaptation à des environnements extrêmes.

En effet, les populations qui se maintiennent dans un environnement extrême y sont adaptées. Ainsi s'il y a remplacement, on peut supposer qu'il ne sera que partiel puisque la pression environnementale est très forte. Dans ce cas, les régions génomiques impliquées dans l'adaptation à l'environnement seraient conservées (Feder et al., 2012; Lenormand, 2002; Yeaman and Otto, 2011). Si c'est le cas nous pourrions trouver des signaux d'adaptation locale à l'environnement antérieurs à la sélection des gènes d'intérêt agronomique étudiés ici.

De plus, certaines régions impliquées dans ces adaptations proviennent d'introgessions issues de croisement avec d'autres races ou espèces. C'est le cas pour l'adaptation à l'altitude chez les moutons Tibétains. Hu et al., (2019) montrent que des régions génomiques

introgressées dans les génomes des moutons depuis les Argalis (mouflons *Ovis ammon*) sont enrichis en gènes impliqués dans l'hypoxie et le transport de l'oxygène. La méthode que nous avons utilisée pour dater le début de la sélection (*STARTMRCA*) (Smith et al., 2018) utilise comme proxy l'âge de l'ancêtre commun le plus récent des haplotypes portant le variant sous sélection (voir **Chapitre 4, partie 2**). Elle pourrait également être utilisée pour dater l'âge d'une introgression (Smith, 2019). Nos analyses pourraient donc permettre de dater l'introgression de ces régions adaptatives chez le mouton.

Cela a déjà été fait pour traiter une question similaire chez le loup. Les loups gris nord-américains présentent une fréquence élevée de robe noire dont l'allèle causal semble aussi être impliqué dans la réponse immunitaire microbienne. Cet allèle dérive d'une hybridation avec le chien. Schweizer et al. (2018) ont daté la sélection de ce locus, grâce à *STARTMRCA*, entre 1600 et 7200 ans. Cela leur a permis de conclure que l'allèle impliqué dans la couleur noire des loups provenait d'une introgression depuis les chiens des américains natifs.

### 3.5. Application à la chèvre

---

Il pourrait être utile de réitérer ce type d'analyse sur une autre espèce domestiquée. La chèvre est un bon candidat, puisque c'est une espèce proche du mouton qui a été domestiquée à peu près au même endroit et au même moment. D'une part, son étude permettrait de tester certains attendus que nous avons pour confirmer la validité des méthodes de datations employées. Par exemple, si les ruptures de connexions que nous avons observées chez le mouton sont effectivement dues à de grandes perturbations dans les sociétés humaines, on s'attend à les retrouver chez la chèvre. En effet, même si les niveaux de bases de connectivité entre les populations de chèvres et de moutons étaient différents, une rupture globale devrait être marquée dans les deux cas.

D'autre part, les convergences où les singularités qui seront révélées pourront nous renseigner sur les caractères sélectionnés chez ces deux espèces domestiquées selon la « voie proie » (Zeder, 2012). Alberto et al. (2018) ont déjà mis en évidence un certain nombre de signaux de sélection communs à la chèvre et au mouton. Est-ce que des signaux différents mais impliquant un même gène chez les deux espèces peuvent avoir le même âge ? a-t-on sélectionnés les mêmes caractéristiques chez le mouton et chez la chèvre au même moment, ou l'un a-t-il précédé l'autre ? Autant de questions qui pourraient être traitées par des analyses de datations génomiques.





# Références

---

## A

- Aali, M. et al. (2017) Association of the calpastatin genotypes, haplotypes, and SNPs with meat quality and fatty acid composition in two Iranian fat- and thin-tailed sheep breeds. *Small Ruminant Research*. [online]. 149, pp.40–51.
- Acharya, R.M. (1982) Sheep and goat breeds of India. *Sheep and goat breeds of India*. [online]. Food and Agriculture Organization of the United Nations. Available from: <https://www.cabdirect.org/cabdirect/abstract/19830171979> [Accessed 2 April 2020].
- Adalsteinsson, S. (1970) Colour inheritance in Icelandic sheep and relation between colour, fertility and fertilization. *Journal of Agricultural Research, Iceland*. 2 (1), pp.3–135.
- Adelson, D.L. et al. (2004) Gene expression in sheep skin and wool (hair). *Genomics*. [online]. 83 (1), pp.95–105.
- Albers, P.K. and McVean, G. (2020) Dating genomic variants and shared ancestry in population-scale sequencing data. *PLOS Biology*. [online]. 18 (1), Public Library of Science, p.e3000586.
- Alberto, F.J. et al. (2018) Convergent genomic signatures of domestication in sheep and goats. *Nature Communications*. [online]. 9 (1), p.813.
- Albrecht, M. et al. (2001) Oruwanje 95/1: a late holocene stratigraphy in Northwestern Namibia. *Cimbebasia*. 17, pp.1–22.
- Alexander, D.H. and Lange, K. (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*. [online]. 12 (1), p.246.
- Al-Mamun, H.A. et al. (2015) Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. *Genetics Selection Evolution*. [online]. 47 (1), p.90.
- Álvarez, I. et al. (2020a) Ancient Homozygosity Segments in West African Djallonké Sheep Inform on the Genomic Impact of Livestock Adaptation to the Environment. *Animals*. [online]. 10 (7), Multidisciplinary Digital Publishing Institute, p.1178.
- Álvarez, I. et al. (2020b) Genomic scan of selective sweeps in Djallonké (West African Dwarf) sheep shed light on adaptation to harsh environments. *Scientific Reports*. [online]. 10 (1), pp.1–13.
- An, X.P. et al. (2015) Two mutations in the 5'-flanking region of the *KITLG* gene are associated with litter size of dairy goats. *Animal Genetics*. [online]. 46 (3), pp.308–311.
- Animal Genetics Training Resource (2020) *Djallonke | Animal Genetics Training Resources2020* [online]. Available from: <http://agtr.ilri.cgiar.org/djallonke> [Accessed 3 June 2021].
- Arora, R., Yadav, H.S. and Mishra, B.P. (2013) Mitochondrial DNA diversity in Indian sheep. *Livestock Science*. [online]. 153 (1), pp.50–55.
- Arranz, J.-J., Bayón, Y. and Primitivo, F.S. (2001) Differentiation among Spanish sheep breeds using microsatellites. *Genetics Selection Evolution*. [online]. 33 (5), p.529.



Arranz, J.J., Bayon, Y. and Primitivo, F.S. (1998) Genetic relationships among Spanish sheep using microsatellites. *Animal Genetics*. [online]. 29 (6), pp.435–440.

Aswani, O.B. (2007) *Genetic diversity and relationships of East African sheep: A Y chromosome perspective* Kenya, Egerton University.

Aw-Hassan, A., Shomo, F. and Iniguez, L. (2008) Small ruminant production: challenges and opportunities for poverty alleviation in West Asia and North Africa.

Ayelet, G. et al. (2001) Survey of Maedi-Visna (MV) in Ethiopian Highland Sheep. *The Veterinary Journal*. [online]. 161 (2), pp.208–210.

## B

Bai, J.Y. et al. (2019) Analysis on correlation between polymorphism of MyoG gene exon I and body size traits of sheep. *Indian Journal of Animal Research*. [online]. (of). Available from: <http://arccjournals.com/journal/indian-journal-of-animal-research/B-1081> [Accessed 11 March 2020].

Baik, M. et al. (2009) Gene Expression Profiling of Liver and Mammary Tissues of Lactating Dairy Cows. *Asian-Australasian Journal of Animal Sciences*. 22 (6), Asian-Australasian Association of Animal Production Societies, pp.871–884.

Barbato, M. et al. (2017) Genomic signatures of adaptive introgression from European mouflon into domestic sheep. *Scientific Reports*. [online]. 7 (1), p.7623.

Baro, J.A., Primitivo, F.S. and Fuente, L.D. la (1995) Breeding programme for the Spanish Churra sheep breed.

Barrett, R.D.H. and Schluter, D. (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*. [online]. 23 (1), pp.38–44.

Barthelme, J.W. (1985) *Fisher-hunters and Neolithic pastoralists in East Turkana, Kenya* 1985

Bartonova, P. et al. (2012) Association between CSN3 and BCO2 gene polymorphisms and milk performance traits in the Czech Fleckvieh cattle breed. *Genetics and Molecular Research*. [online]. 11 (2), pp.1058–1063.

Bechchari, A. et al. (2014) Analyse de l'évolution du système pastoral du Maroc oriental. *Revue d'élevage et de médecine vétérinaire des pays tropicaux*. [online]. 67 (4), pp.151–162.

Belabdi, I. et al. (2019) Genetic homogenization of indigenous sheep breeds in Northwest Africa. *Scientific Reports*. [online]. 9 (1), Nature Publishing Group, p.7920.

Beleza, S. et al. (2013) The Timing of Pigmentation Lightening in Europeans. *Molecular Biology and Evolution*. [online]. 30 (1), pp.24–35.

Belyaev, D.K. (1979) Destabilizing selection as a factor in domestication. *Journal of Heredity*. [online]. 70 (5), pp.301–308.

Bijl, E. et al. (2014) Phosphorylation of  $\alpha$ S1-casein is regulated by different genes. *Journal of Dairy Science*. [online]. 97 (11), pp.7240–7246.

- Biraben, J.-N. (1976) *Les hommes et la peste en France et dans les pays européens et méditerranéens*. [online]. Paris, Persée - Portail des revues scientifiques en SHS. Available from: [https://www.persee.fr/doc/adh\\_0066-2062\\_1976\\_num\\_1976\\_1\\_1333](https://www.persee.fr/doc/adh_0066-2062_1976_num_1976_1_1333) [Accessed 7 September 2021].
- Blench, R. and Donald, K. (1999) *The Origins and Development of African Livestock: Archaeology, Genetics, Linguistics and Ethnography* Routledge & CRC Press.1999 [online]. Available from: <https://www.routledge.com/The-Origins-and-Development-of-African-Livestock-Archaeology-Genetics/Blench-MacDonald/p/book/9780415515771> [Accessed 10 May 2021].
- Bodin, L. et al. (2007) A Novel Mutation in the Bone Morphogenetic Protein 15 Gene Causing Defective Protein Secretion Is Associated with Both Increased Ovulation Rate and Sterility in Lacaune Sheep. *Endocrinology*. [online]. 148 (1), pp.393–400.
- Boivin, N. and Fuller, D.Q. (2009) Shell Middens, Ships and Seeds: Exploring Coastal Subsistence, Maritime Trade and the Dispersal of Domesticates in and Around the Ancient Arabian Peninsula. *Journal of World Prehistory*. [online]. 22 (2), pp.113–180.
- Bolormaa, S. et al. (2017) Multiple-trait QTL mapping and genomic prediction for wool traits in sheep. *Genetics Selection Evolution*. [online]. 49 (1), BioMed Central, pp.1–22.
- Bosse, M. (2018) A Genomics Perspective on Pig Domestication. *Animal Domestication*. [online]. IntechOpen. Available from: <https://www.intechopen.com/books/animal-domestication/a-genomics-perspective-on-pig-domestication> [Accessed 23 February 2021].
- Boujenane, I. et al. (2008) Variation at post-albumin, transferrin and haemoglobin proteins in Moroccan local sheep. *Small Ruminant Research*. [online]. 79 (2), pp.113–117.
- Boujenane, I. and Petit, D. (2016) Between- and within-breed morphological variability in Moroccan sheep breeds. *Animal Genetic Resources*. [online]. 1, pp.91–100.
- Bowles, D., Carson, A. and Isaac, P. (2014) Genetic Distinctiveness of the Herdwick Sheep Breed and Two Other Locally Adapted Hill Breeds of the UK. *PLoS ONE*. [online]. 9 (1). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3906253/> [Accessed 10 March 2020].
- Brannan, C.I. et al. (1994) Targeted disruption of the neurofibromatosis type-1 gene leads to developmental abnormalities in heart and various neural crest-derived tissues. *Genes & Development*. [online]. 8 (9), pp.1019–1029.
- Breunig, P., Neumann, K. and Van Neer, W. (1996) New Research on the Holocene Settlement and Environment of the Chad Basin in Nigeria. *Afr. Archaeol. Rev.* [online]. 13, pp.111–145.
- Browning, S.R. and Browning, B.L. (2007) Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *The American Journal of Human Genetics*. [online]. 81 (5), pp.1084–1097.
- Bruford, M.W., Bradley, D.G. and Luikart, G. (2003) DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics*. [online]. 4 (11), pp.900–910.
- Buitenhuis, B. et al. (2014) Genome-wide association and biological pathway analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. *BMC Genomics*. [online]. 15 (1), BioMed Central, pp.1–11.

Bunch, T.D. et al. (2006) Phylogenetic Analysis of Snow Sheep (*Ovis nivicola*) and Closely Related Taxa. *Journal of Heredity*. [online]. 97 (1), pp.21–30.

## C

Carruthers, D. (1949) *Beyond the Caspian: a naturalist in Central Asia* 1st ed. édition. OLIVER & BOYD LTD.

Cesar, A.S. et al. (2014) Genome-wide association study for intramuscular fat deposition and composition in Nellore cattle. *BMC Genetics*. [online]. 15, p.39.

Cesarani, A. et al. (2019a) Genomic selection of milk fatty acid composition in Sarda dairy sheep: Effect of different phenotypes and relationship matrices on heritability and breeding value accuracy. *Journal of Dairy Science*. [online]. 102 (4), pp.3189–3203.

Cesarani, A. et al. (2019b) Investigation of genetic diversity and selection signatures between Sarda and Sardinian Ancestral black, two related sheep breeds with evident morphological differences. *Small Ruminant Research*. [online]. 177, pp.68–75.

Cetti, F. (1774) *Storia Naturale di Sardegna* Chemistry & biodiversity 1.

Chelling, R. (1992) *Les Races Ovines Algériennes* Alger, Office des Publications Universitaires.

Chen, F.H. et al. (2015) Agriculture facilitated permanent human occupation of the Tibetan Plateau after 3600 B.P. *Science*. [online]. 347 (6219), American Association for the Advancement of Science, pp.248–250.

Chen, H., Hey, J. and Slatkin, M. (2015) A hidden Markov model for investigating recent positive selection through haplotype structure. *Theoretical Population Biology*. [online]. 99, pp.18–30.

Chen, H. and Slatkin, M. (2013) Inferring Selection Intensity and Allele Age from Multilocus Haplotype Structure. *G3 Genes/Genomes/Genetics*. [online]. 3 (8), pp.1429–1442.

Chen, R. (1990) Creation of conservation areas for Hu sheep in Jiangsu province. pp.19–29.

Chen, S.-Y. et al. (2006) Origin, genetic diversity, and population structure of Chinese domestic sheep. *Gene*. [online]. 376 (2), pp.216–223.

Chen, W.D. and Chen, X.T. (2002) *Xinjiang sheep industry at present and countermeasure Grass-feeding Livestock* 117:7-9.2002 [online]. Available from: [http://en.cnki.com.cn/Article\\_en/CJFDTOTAL-CQJQ200204002.htm](http://en.cnki.com.cn/Article_en/CJFDTOTAL-CQJQ200204002.htm) [Accessed 10 April 2020].

Chessa, B. et al. (2009a) Revealing the History of Sheep Domestication Using Retrovirus Integrations. *Science*. [online]. 324 (5926), pp.532–536.

Chessa, B. et al. (2009b) Revealing the History of Sheep Domestication Using Retrovirus Integrations. *Science*. [online]. 324 (5926), pp.532–536.

China National Commission of Animal Genetic Resources (2011) *Animal genetic resources in China: sheep and goats* Chinese Agricultural Press, Beijing.

- Chu, M.X. et al. (2007) Mutations in BMPR-IB and BMP-15 genes are associated with litter size in Small Tailed Han sheep (*Ovis aries*). *Journal of Animal Science*. [online]. 85 (3), pp.598–603.
- Ciani, E. et al. (2014) Genome-wide analysis of Italian sheep diversity reveals a strong geographic pattern and cryptic relationships between breeds. *Animal Genetics*. [online]. 45 (2), pp.256–266.
- Ciani, E. et al. (2015) Merino and Merino-derived sheep breeds: a genome-wide intercontinental study. *Genetics Selection Evolution*. [online]. 47 (1), p.64.
- Ciani, E. et al. (2020) On the origin of European sheep as revealed by the diversity of the Balkan breeds and by optimizing population-genetic analysis tools. *Genetics Selection Evolution*. [online]. 52 (1), p.25.
- Cinar, M.U. et al. (2016) Ovar-DRB1 haplotypes \*2001 and \*0301 are associated with sheep growth and ewe lifetime prolificacy. *Gene*. [online]. 595 (2), pp.187–192.
- Clements, S.E. et al. (2012) Mutations in AEC syndrome skin reveal a role for p63 in basement membrane adhesion, skin barrier integrity and hair follicle biology. *The British Journal of Dermatology*. [online]. 167 (1), pp.134–144.
- Cline, E. (2016) *1177 avant J.-C. Le jour où la civilisation s'est effondrée* [online]. Available from: <https://livre.fnac.com/a9537260/Eric-H-Cline-1177-avant-J-C-Le-jour-ou-la-civilisation-s-est-effondree> [Accessed 7 September 2021].
- Close, A.E. (1992) Holocene occupation of the Eastern SaharaIn: *New light on the Northeast African past*. F. Klees&R. Kuper. Koln: Heinrich Barth Institute, 155–183.
- Clutton-Brock, J. (1981) Domesticated animals from early times. *Domesticated animals from early times*. [online]. Available from: <https://www.cabdirect.org/cabdirect/abstract/19820166164> [Accessed 2 December 2019].
- Clutton-Brock, J. (1993) The spread of domestic animals in AfricaIn: *The archaeology of Africa: Food, metals and towns*. 61–70.
- Cohen-Zinder, M. (2005) Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Research*. [online]. 15 (7), pp.936–944.
- Conti, L. and Cattaneo, E. (2010) Neural stem cell systems: physiological players or in vitro entities? *Nature Reviews Neuroscience*. [online]. 11 (3), pp.176–187.
- Coop, G. et al. (2008) The timing of selection at the human FOXP2 gene. *Molecular Biology and Evolution*. [online]. 25 (7), pp.1257–1259.
- Coop, G. and Griffiths, R.C. (2004) Ancestral inference on gene trees under selection. *Theoretical Population Biology*. [online]. 66 (3), pp.219–232.
- Couput, M. (1900) *Espèce Ovine, Laine et Industrie Lainière*. Ed. Pour L'exposition Universelle De 1900, Section Algérie de M. Couput, directeur du Service des bergeries.
- Crégut, E. (1979) *La faune de Mammifères du Pléistocène moyen de la Caune de l'Arago à Tautavel (Pyrénées-Orientales)* Université de Provence.

## D

D'Alpoim Guedes, J. and Butler, E.E. (2014) 'Modeling constraints on the spread of agriculture to Southwest China with thermal niche models' *Quaternary International. Quaternary of East Asia and the Western Pacific: Part 2* [online]. 349, pp.29–41.

Daly, K.G. et al. (2018) Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science*. [online]. 361 (6397), pp.85–88.

Danecek, P. et al. (2011) The variant call format and VCFtools. *Bioinformatics*. [online]. 27 (15), pp.2156–2158.

Darwin, C. (1868) *The Variation of Animals and Plants Under Domestication* Cambridge University Press.

Daujatz, J. and Mashkour, M. (2017) *Faunal Remains from Middle Neolithic Site of Qaleh Rostam*In: *Proceedings of the 2008 Al Ain-Abu Dhabi conference*. M. Mashkour and M.J. Beech eds. Archaeozoology of the Near East 9 Oxbow Books, 41–58.

Davidson, J. (2009) *Breeding: A Partial History of the Eighteenth Century* Columbia University Press.

Davis, S.J.M. (2006) *Agricultural improvements in gharb al-Andaluz—zoo-archaeological evidence from the sheep bones*. *Proceedings of the fourth conference on Iberian Archaeology; 2004 Sept 14-19; Faro, Portugal. Promontoria Monogr 3:219–39*

Delpech, F. et al. (1978) Un nouveau gisement du Pléistocène moyen : Camp-de-Peyre à Sauveterre-la-Lémance. 286, pp.1101–1103.

Demars, J. et al. (2017) Genome-Wide Identification of the Mutation Underlying Fleece Variation and Discriminating Ancestral Hairy Species from Modern Woolly Sheep. *Molecular Biology and Evolution*. [online]. 34 (7), pp.1722–1729.

deMenocal, P.B. and Stringer, C. (2016) Climate and the peopling of the world. *Nature*. [online]. 538 (7623), Nature Publishing Group, pp.49–50.

Demirci, S. et al. (2013) Mitochondrial DNA Diversity of Modern, Ancient and Wild Sheep (*Ovis gmelinii anatolica*) from Turkey: New Insights on the Evolutionary History of Sheep. *PLOS ONE*. [online]. 8 (12), p.e81952.

Deng, J. et al. (2020) Paternal Origins and Migratory Episodes of Domestic Sheep. *Current Biology*. [online]. p.S0960982220311362.

Deniskova, T.E. et al. (2018) Population structure and genetic diversity of 25 Russian sheep breeds based on whole-genome genotyping. *Genetics Selection Evolution*. [online]. 50 (1), p.29.

Derry, M.E. (2020) Theory and Method: An Analysis of European and American Animal Breeding Practices, from the Eighteenth to the Twenty-First Century. *Agricultural History*. [online]. 94 (3), Agricultural History Society, pp.324–361.

Diez-Tascón, C. et al. (2000) Genetic variation within the Merino sheep breed: analysis of closely related populations using microsatellites. *Animal Genetics*. [online]. 31 (4), pp.243–251.

- Dion, F. (1994) Evolution et résultats du plan national de lutte contre le Visna-Maëdi. 1, pp.51–58.
- Djaout, A. et al. (2017) Etat de la biodiversité des «races» ovines algériennes. *Genetics and Biodiversity Journal*. 1, pp.1–16.
- Dobney, K. and Larson, G. (2006) Genetics and animal domestication: new windows on an elusive process. *Journal of Zoology*. [online]. 0 (0), pp.060222013030001-???
- Dobney, K.M. and Jaques, D. (2010) The vertebrate assemblage from excavations at Jeitun, 1993 and 1994In: *Origins of Agriculture in Western Central Asia: an Environmental-Archaeological Study*. Harris D. R. university of Pennsylvania, Museum of Archaeology and Anthropology, Philadelphia, 174–181.
- Dodson, J. et al. (2014) Oldest Directly Dated Remains of Sheep in China. *Scientific Reports*. [online]. 4 (1), pp.1–4.
- Dohner, J.V. (2001) *The Encyclopedia of Historic and Endangered Livestock and Poultry Breeds* Yale University Press.
- Dominik, S., Henshall, J.M. and Hayes, B.J. (2012) A single nucleotide polymorphism on chromosome 10 is highly predictive for the polled phenotype in Australian Merino sheep. *Animal Genetics*. [online]. 43 (4), pp.468–470.
- Dong, K. et al. (2020) Genomic analysis of worldwide sheep breeds reveals PDGFD as a major target of fat-tail selection in sheep. *BMC Genomics*. [online]. 21 (1), p.800.
- Draganescu C. (2007) A note on Balkan sheep breeds origin and their taxonomy.
- Draws, R. (1993) *The End of the Bronze Age: Changes in Warfare and the Catastrophe ca. 1200 B.C. - Third Edition* 3rd edition. [online]. Princeton University Press. Available from: <https://www.jstor.org/stable/j.ctvx5wbmc> [Accessed 7 September 2021].
- Dudu, A. et al. (2016) Origin and genetic diversity of Romanian Racka sheep using mitochondrial markers. *Small Ruminant Research*. [online]. 144, pp.276–282.
- Duijvesteijn, N. et al. (2018) Genomic prediction of the polled and horned phenotypes in Merino sheep. *Genetics Selection Evolution*. [online]. 50 (1), BioMed Central, pp.1–11.
- Durán Aguilar, M. et al. (2017) Genome-wide association study for milk somatic cell score in holstein cattle using copy number variation as markers. *Journal of Animal Breeding and Genetics*. [online]. 134 (1), pp.49–59.
- Dýrmundsson, Ó.R. and Niżnikowski, R. (2010) North European short-tailed breeds of sheep: a review. *animal*. [online]. 4 (08), pp.1275–1282.

## E

- E, G. et al. (2016) Conservation genetics in Chinese sheep: diversity of fourteen indigenous sheep (*Ovis aries*) using microsatellite markers. *Ecology and Evolution*. [online]. 6 (3), pp.810–817.
- Ebrahimi, F. et al. (2017) Detection of QTL for greasy fleece weight in sheep using a 50 K single nucleotide polymorphism chip. *Tropical Animal Health and Production*. [online]. 49 (8), pp.1657–1662.

Edea, Z. et al. (2019) Genomic signatures of high-altitude adaptation in Ethiopian sheep populations. *Genes & Genomics*. [online]. 41 (8), pp.973–981.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. [online]. 32 (5), pp.1792–1797.

Epstein, H. (1960) History and Origin of the Ronderib and Namaqua Afrikaner Sheep. *Zeitschrift für Tierzüchtung und Züchtungsbiologie*. [online]. 74 (1–4), pp.267–292.

Epstein, H. (1971) *The origin of the domestic animals of Africa* [online]. Africana Publishing Corporation. Available from: <https://cgspace.cgiar.org/handle/10568/70619> [Accessed 19 February 2019].

## F

Fahmy, M.H. (1989) Reproductive performance, growth and wool production of Romanov sheep in Canada. *Small Ruminant Research*. [online]. 2 (3), pp.253–264.

Fan, Y. (1131) *Hou Han Shu*

FAO (2014) *Characterization and value addition to local breeds and their products in the Near East and North Africa* Regional Workshop, Rabat, Morocco, 19–21 November 2012 Animal Production and Health Report No. 3, Rome, 48.

FAO (1998) *Secondary Guidelines for Development of National Farm Animal Genetic Resources Management Plans: Management of Small Populations at Risk* Food and Agriculture Organization, Rome.

FAO (2000) *World Watch List for Domestic Animal Diversity* Food and Agriculture Organization, Rome.

Fariello, M.I. et al. (2013) Detecting Signatures of Selection Through Haplotype Differentiation Among Hierarchically Structured Populations. *Genetics*. [online]. 193 (3), Genetics, pp.929–941.

Fariello, M.-I. et al. (2014) Selection Signatures in Worldwide Sheep Populations. *PLoS ONE*. [online]. 9 (8). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4134316/> [Accessed 10 March 2020].

Feder, J.L. et al. (2012) Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences*. [online]. 367 (1587), Royal Society, pp.461–474.

Ferrer-Admetlla, A. et al. (2014) On Detecting Incomplete Soft or Hard Selective Sweeps Using Haplotype Structure. *Molecular Biology and Evolution*. [online]. 31 (5), pp.1275–1291.

Fésüs, L. et al. (2002) *The determining role of Merino in Hungarian sheep breeding* Proceedings of the 6th Merino World Conference Budapest.

Finkelstein, I. et al. (2017) Egyptian Imperial Economy in Canaan: Reaction to the Climate Crisis at the End of the Late Bronze Age. *Ägypten und Levante*. [online]. 27, Verlag der Österreichischen Akademie der Wissenschaften, pp.249–260.

Flad, R.K., Yuan, J.袁靖 and Li, S.李水城 (2007)'Zooarcheological evidence for animal domestication in northwest China'In: *Developments in Quaternary Sciences*.Late Quaternary Climate Change and Human Adaptation in Arid China [online]. 9, Elsevier, 167–203. Available from: <http://www.sciencedirect.com/science/article/pii/S1571086607090124> [Accessed 12 December 2019].

Flamand, G.B.M. (1861-1919) A. du texte et al. (1893) *Le pays du mouton : des conditions d'existence des troupeaux sur les hauts-plateaux et dans le sud de l'Algérie / ouvrage publié par ordre de M. Jules Cambon,... ; [rédigé sous la direction de M. E. Du Champ, par MM. A. Turlin, F. Accardo, G.-B.-M. Flamand]* [online]. Available from: <https://gallica.bnf.fr/ark:/12148/bpt6k3064041d> [Accessed 13 May 2021].

Fogarty, N. et al. (2003) Genetic parameters for meat quality and carcass traits and their correlation with wool traits in Australian Merino sheep. *Australian Journal of Agricultural Research*. [online]. 54, pp.715–722.

Forestry Department. (2012) *The Cyprus mouflon* . Republic of Cyprus Press and Information Office, Ref no. 128/2012. Nicosia, Cyprus: Othon Press Ltd.

Frandsen, R.D., Wilke, W.L. and Fails, A.D. (2009) *Anatomy and Physiology of Farm Animals* 7th Edition. Ames, Iowa, Wiley–Blackwell.

Frantz, L.A.F. et al. (2020) Animal domestication in the era of ancient genomics. *Nature Reviews Genetics*. [online]. 21 (8), Nature Publishing Group, pp.449–460.

Fregel, R. et al. (2018) Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. *Proceedings of the National Academy of Sciences*. [online]. 115 (26), National Academy of Sciences, pp.6774–6779.

## G

Galindo, R.C. et al. (2008) Differential expression of inflammatory and immune response genes in sheep infected with *Anaplasma phagocytophilum*. *Veterinary Immunology and Immunopathology*. [online]. 126 (1), pp.27–34.

Galloway, S.M. et al. (2000) Mutations in an oocyte-derived growth factor gene ( BMP15 ) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nature Genetics*. [online]. 25 (3), pp.279–283.

Ganbold, O. et al. (2019) Genetic diversity and the origin of Mongolian native sheep. *Livestock Science*. [online]. 220, pp.17–25.

Ganbold, O., Munkhbayar, M. and Paek, W.K. (2018) A review of population genetics research on domestic animals in Mongolia and recommendations for the improvements. *Journal of Animal Breeding and Genomics*. [online]. 2 (1). Available from: <http://www.jabng.org/article.php?num=N0270020102> [Accessed 11 March 2020].

Gao, Z.Y. et al. (2008) *Current status and prospect for selection of Qira Black sheep*.

Gaouar, S.B.S. et al. (2015) Admixture and Local Breed Marginalization Threaten Algerian Sheep Diversity. *PLOS ONE*. [online]. 10 (4), Public Library of Science, p.e0122667.



- Gaouar, S.B.S. et al. (2017) Genome-wide analysis highlights genetic dilution in Algerian sheep. *Heredity*. [online]. 118 (3), pp.293–301.
- García-Gámez, E. et al. (2012) GWA Analysis for Milk Production Traits in Dairy Sheep and Genetic Support for a QTN Influencing Milk Protein Percentage in the LALBA Gene. *PLOS ONE*. [online]. 7 (10), Public Library of Science, p.e47782.
- García-Gámez, E. et al. (2011) Using Regulatory and Epistatic Networks to Extend the Findings of a Genome Scan: Identifying the Gene Drivers of Pigmentation in Merino Sheep. *PLoS ONE*. [online]. 6 (6). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3119053/> [Accessed 8 April 2020].
- Gardner, G. et al. (2010) Using Australian Sheep Breeding Values to increase lean meat yield percentage. [online].
- Gautier, A. (1987) Fishing, fowling and hunting in Late Palaeolithic times in the Nile Valley in Upper Egypt. 18, pp.429–440.
- Gautier, A. (2002) The Evidence for the Earliest Livestock in North Africa: or Adventures with Large Bovids, Ovicaprids, Dogs and Pigs In: Hassan, F.A. (ed.) *Droughts, Food and Culture: Ecological Change and Food Security in Africa's Later Prehistory*. [online]. Boston, MA, Springer US, 195–207. Available from: [https://doi.org/10.1007/0-306-47547-2\\_12](https://doi.org/10.1007/0-306-47547-2_12) [Accessed 17 May 2021].
- Gebreselassie, G. et al. (2020) Review on Genomic Regions and Candidate Genes Associated with Economically Important Production and Reproduction Traits in Sheep (*Ovis aries*). *Animals*. [online]. 10 (1), Multidisciplinary Digital Publishing Institute, p.33.
- Gelasakis, A.I. et al. (2012) Description and typology of intensive Chios dairy sheep farms in Greece. *Journal of Dairy Science*. [online]. 95 (6), pp.3070–3079.
- Gifford-Gonzalez, D. and Hanotte, O. (2011) Domesticating Animals in Africa: Implications of Genetic and Archaeological Findings. *Journal of World Prehistory*. [online]. 24 (1), pp.1–23.
- Gillis, R.E. et al. (2019) A Species Specific Investigation Into Sheep and Goat Husbandry During the Early European Neolithic. *Environmental Archaeology*. [online]. pp.1–12.
- Goitein, S. (1967) *A Mediterranean society* Los Angeles, CA: University of California Press. Economic foundations.
- Gong, H. et al. (2016) Wool Keratin-Associated Protein Genes in Sheep—A Review. *Genes*. [online]. 7 (6), Multidisciplinary Digital Publishing Institute, p.24.
- Gonzalez-Perez, A., Sabarinathan, R. and Lopez-Bigas, N. (2019) Local Determinants of the Mutational Landscape of the Human Genome. *Cell*. [online]. 177 (1), pp.101–114.
- Gorkhali, N.A., Han, J.L. and Ma, Y.H. (2015) Mitochondrial DNA variation in indigenous sheep (*Ovis aries*) breeds of Nepal. *Tropical Agricultural Research*. [online]. 26 (4), p.632.
- Gornas, N. et al. (2011) Genetic characterization of local Sudanese sheep breeds using DNA markers. *Small Ruminant Research*. [online]. 95 (1), pp.27–33.

Götherström, A. et al. (2005) Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. *Proceedings of the Royal Society B: Biological Sciences*. [online]. 272 (1579), pp.2345–2351.

Grasteau, S. et al. (2005) Genetics of adaptation and domestication in livestock. *Livestock Production Science*. [online]. 93, pp.3–14.

Gratten, J. et al. (2010) The genetic basis of recessive self-colour pattern in a wild sheep population. *Heredity*. [online]. 104 (2), pp.206–214.

Greyvenstein, O.F.C. et al. (2016) Polyceraty (multi-horns) in Damara sheep maps to ovine chromosome 2. *Animal Genetics*. [online]. 47 (2), pp.263–266.

Griffiths, R. and Tavaré, S. (1998) The age of a mutation in a general coalescent. *Stochastic Models - STOCH MODELS*. [online]. 14, pp.273–295.

Grousset, R. (1941) *The History of the Mongol Empire*

d'Alpoim Guedes, J. et al. (2014) Moving agriculture onto the Tibetan plateau: the archaeobotanical evidence. *Archaeological and Anthropological Sciences*. [online]. 6 (3), pp.255–269.

d'Alpoim Guedes, J.A. et al. (2015) Early evidence for the use of wheat and barley as staple crops on the margins of the Tibetan Plateau. *Proceedings of the National Academy of Sciences*. [online]. 112 (18), pp.5625–5630.

Guerrini, M. et al. (2015) Molecular DNA identity of the mouflon of Cyprus (*Ovis orientalis ophion*, Bovidae): Near Eastern origin and divergence from Western Mediterranean conspecific populations. *Systematics and Biodiversity*. [online]. 13 (5), Taylor & Francis, pp.472–483.

Guindon, S. et al. (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*. [online]. 59 (3), pp.307–321.

Guo, J. et al. (2005) A novel maternal lineage revealed in sheep (*Ovis aries*). *Animal Genetics*. [online]. 36 (4), pp.331–336.

Gutiérrez-Gil, B. et al. (2014) Application of Selection Mapping to Identify Genomic Regions Associated with Dairy Production in Sheep. *PLoS ONE*. [online]. 9 (5). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4006912/> [Accessed 10 March 2020].

Gutiérrez-Gil, B. et al. (2017) High-resolution analysis of selection sweeps identified between fine-wool Merino and coarse-wool Churra sheep breeds. *Genetics Selection Evolution*. [online]. 49 (1), p.81.

## H

Hamadani, A. et al. (2019) Estimation of genetic, heritability, and phenotypic trends for weight and wool traits in Rambouillet sheep. *Small Ruminant Research*. [online]. 177, pp.133–140.

Han, J. et al. (2019) Two linked TBXT (brachyury) gene polymorphisms are associated with the tailless phenotype in fat-rumped sheep. *Animal Genetics*. [online]. 50 (6), pp.772–777.

- Han, J.L. et al. (2015) Analysis of agouti signaling protein (ASIP) gene polymorphisms and association with coat color in Tibetan sheep (*Ovis aries*). *Genetics and Molecular Research*. [online]. 14 (1), pp.1200–1209.
- Han, M. et al. (2016) 'Climate, Vegetation and Human Land-Use Interactions on the Qinghai–Tibet Plateau Through the Holocene' In: Brierley, G.J. et al. (eds.) *Landscape and Ecosystem Diversity, Dynamics and Management in the Yellow River Source Zone*. Springer Geography [online]. Cham, Springer International Publishing, 253–274. Available from: [https://doi.org/10.1007/978-3-319-30475-5\\_12](https://doi.org/10.1007/978-3-319-30475-5_12) [Accessed 22 April 2020].
- Handley, L.-J.L. et al. (2007) Genetic structure of European sheep breeds. *Heredity*. [online]. 99 (6), pp.620–631.
- Hare, A.J. (2021) *Ancient Genomic Analysis of Wild and Domestic Sheep* Dublin, University of Dublin.
- Harper, K., ROSSIGNOL, B. and PIGNARRE, P. (2019) *Comment l'Empire romain s'est effondré* Paris, La Découverte.
- Hassan, Blench, R. and MacDonald, K. (2006) *The Origins and Development of African Livestock: Archaeology, Genetics, Linguistics and Ethnography* Routledge.
- Hassan, F.A. (2000) Holocene environmental change and the origins and spread of food production in the Middle East. 1, pp.7–28.
- He, G. zhu (2000) *Di Qiang yuan liu shi* Nanchang: Jiangxi Education Press.
- He, X. et al. (2016) Mapping the four-horned locus and testing the polled locus in three Chinese sheep breeds. *Animal Genetics*. [online]. 47 (5), pp.623–627.
- Heaton, M.P. et al. (2013) Genetic Testing for TMEM154 Mutations Associated with Lentivirus Susceptibility in Sheep. *PLOS ONE*. [online]. 8 (2), Public Library of Science, p.e55490.
- Heaton, M.P. et al. (2012) Reduced Lentivirus Susceptibility in Sheep with TMEM154 Mutations. *PLOS Genetics*. [online]. 8 (1), Public Library of Science, p.e1002467.
- Helmer, D., Gourichon, L. and Vila, E. (2007) The development of the exploitation of products from *Capra* and *Ovis* (meat, milk and fleece) from the PPNB to the Early Bronze in the northern Near East (8700 to 2000 BC cal.). p.31.
- Hermes, T.R. et al. (2020) Mitochondrial DNA of domesticated sheep confirms pastoralist component of Afanasievo subsistence economy in the Altai Mountains (3300–2900 cal BC). *Archaeological Research in Asia*. [online]. 24, p.100232.
- Hickford, J.G.H. et al. (2010) Polymorphisms in the ovine myostatin gene (MSTN) and their association with growth and carcass traits in New Zealand Romney sheep. *Animal Genetics*. [online]. 41 (1), pp.64–72.
- Hiendleder, S. et al. (2002) Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proceedings of the Royal Society of London. Series B: Biological Sciences*. [online]. 269 (1494), pp.893–904.

Hiendleder, S. et al. (1998) The Complete Mitochondrial DNA Sequence of the Domestic Sheep (*Ovis aries*) and Comparison with the Other Major Ovine Haplotype. *Journal of Molecular Evolution*. [online]. 47 (4), pp.441–448.

Hoang, D.T. et al. (2018) UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*. [online]. 35 (2), pp.518–522.

Holl, A.F.C. (1998) The Dawn of African Pastoralisms: An Introductory Note. *Journal of Anthropological Archaeology*. [online]. 17 (2), pp.81–96.

Horsburgh, K. and Rhines, A. (2010) Genetic characterization of an archaeological sheep assemblage from South Africa's Western Cape. *Journal of Archaeological Science*. [online]. 11, pp.2906–2910.

Horvat, S. and Medrano, J.F. (2001) Lack of *Socs2* Expression Causes the High-Growth Phenotype in Mice. *Genomics*. [online]. 72 (2), pp.209–212.

Houwers, D.J. et al. (1983) Maedi-visna control in sheep. I. Artificial rearing of colostrum-deprived lambs. *Veterinary Microbiology*. [online]. 8 (2), pp.179–185.

Hu, X.-J. et al. (2019) The Genome Landscape of Tibetan Sheep Reveals Adaptive Introgression from Argali and the History of Early Human Settlements on the Qinghai–Tibetan Plateau. Nielsen, R. (ed.) *Molecular Biology and Evolution*. [online]. 36 (2), pp.283–303.

Huang, Y.-Z. et al. (2013) Haplotype distribution in the *GLI3* gene and their associations with growth traits in cattle. *Gene*. [online]. 513 (1), pp.141–146.

Hudson, R.R. (2007) The Variance of Coalescent Time Estimates from DNA Sequences. *Journal of Molecular Evolution*. [online]. 64 (6), pp.702–705.

I

I. Tunca, R. and Taşkin, A. (2016) A comparison of genetic variations among native and some local chicken populations in Turkey. *Indian Journal of Animal Research*. [online]. 50, pp.652–657.

Ibeagha-Awemu, E.M. et al. (2016) High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits. *Scientific Reports*. [online]. 6 (1), p.31109.

Ibrahim, A. et al. (2020) The genetic profiles and maternal origin of local sheep breeds on Java Island (Indonesia) based on complete mitochondrial DNA D-loop sequences. *Veterinary World*. [online]. 13 (12), pp.2625–2634.

Illius, A.W., Haynes, N.B. and Lamming, G.E. (1976) Effects of ewe proximity on peripheral plasma testosterone levels and behaviour in the ram. *Reproduction*. [online]. 48 (1), Bioscientifica Ltd, pp.25–32.

International Council for Archaeozoology et al. (eds.) (2005) *The first steps of animal domestication: new archaeozoological approaches* Oxford, Oxbow.

International Livestock Research Institute (2020) *Breeds By Region | DAGRIS2020* [online]. Available from: [http://dagris.ilri.cgiar.org/reports/breeds\\_by\\_region](http://dagris.ilri.cgiar.org/reports/breeds_by_region) [Accessed 7 June 2021].

## J

Jakubec, V. (1977) Productivity of crosses based on prolific breeds of sheep. *Livestock Production Science*. [online]. 4 (4), pp.379–392.

Jiang, W.S. (2006) *Current status and prospect for selection of Duolang sheep China Herbivores* 26(2):28-30.

Jing, Y. (2008) The Origins and Development of Animal Domestication in China. *Chinese Archaeology*. [online]. 8 (1), pp.1–7.

Johnston, S.E. et al. (2011) Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology*. [online]. 20 (12), pp.2555–2566.

Johnston, S.E. et al. (2013) Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature*. [online]. 502 (7469), pp.93–95.

## K

Kalyaanamoorthy, S. et al. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*. [online]. 14 (6), pp.587–589.

Kamjan, S. et al. (2018) *The archaeozoology of the Pre-Pottery and Pottery Neolithic site of Qasr-e Ahmad (Iran)*In: *Proceedings of the Groningen 2015 ASWA conference*.Archaeozoology of the Near East 12. (Çakirlar C., Chahoud J., Berthon R., Pilaar Birch S. Eds.) Groningen Archaeological Series, 27–40.

Kandoussi, A. et al. (2020) The origin of sheep settlement in Western Mediterranean. *Scientific Reports*. [online]. 10 (1), Nature Publishing Group, p.10225.

Karasov, T., Messer, P.W. and Petrov, D.A. (2010) Evidence that Adaptation in *Drosophila* Is Not Limited by Mutation at Single Sites. *PLOS Genetics*. [online]. 6 (6), Public Library of Science, p.e1000924.

Kardos, M. et al. (2015) Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. *Molecular Ecology*. [online]. 24 (22), pp.5616–5632.

Katoh, K. and Standley, D.M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*. [online]. 30 (4), pp.772–780.

Keightley, P.D. and Jackson, B.C. (2018) Inferring the Probability of the Derived vs. the Ancestral Allelic State at a Polymorphic Site. *Genetics*. [online]. 209 (3), pp.897–906.

Kelleher, J., Etheridge, A.M. and McVean, G. (2016) Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLOS Computational Biology*. p.22.

Kijas, J.W. et al. (2009) A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. *PLoS ONE*. [online]. 4 (3). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2652362/> [Accessed 10 March 2020].

Kijas, J.W. et al. (2012a) Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLOS Biology*. [online]. 10 (2), p.e1001258.

Kijas, J.W. et al. (2012b) Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLOS Biology*. [online]. 10 (2), Public Library of Science, p.e1001258.

Kijas, J.W. et al. (2016) Genome-wide association reveals the locus responsible for four-horned ruminant. *Animal Genetics*. [online]. 47 (2), pp.258–262.

Kim, E.-S. et al. (2016) Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity*. [online]. 116 (3), pp.255–264.

Kimura, M. et al. (1986) DNA and the neutral theory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*. [online]. 312 (1154), Royal Society, pp.343–354.

Kimura, M. and Ohta, T. (1973) THE AGE OF A NEUTRAL MUTANT PERSISTING IN A FINITE POPULATION. *Genetics*. [online]. 75 (1), pp.199–212.

Kinahan, J. (1991) *Pastoral Nomads of the Central Namib Desert, by John Kinahan vorgestellt im Namibiana Buchdepot* [online]. Windhoek, Namibia: Namibia Archaeological Trust. Available from: <https://www.namibiana.de/namibia-information/literaturauszuege/titel/pastoral-nomads-of-the-central-namib-desert-john-kinahan.html> [Accessed 10 May 2021].

Kolosov, Y.A. et al. (2021) Investigation of the D-loop sequence of mitochondrial DNA of the Volgograd sheep breed. *IOP Conference Series: Earth and Environmental Science*. [online]. 677 (5), p.052111.

Koseniuk, A. and Słota, E. (2016) Mitochondrial control region diversity in Polish sheep breeds. *Archives Animal Breeding*. [online]. 59 (2), pp.227–233.

Kruska, D.C.T. (2005) On the evolutionary significance of encephalization in some eutherian mammals: effects of adaptive radiation, domestication, and feralization. *Brain, behavior and evolution*. [online]. 65 (2), Switzerland, Karger, pp.73–108.

Kumar, S. et al. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*. [online]. 35 (6), pp.1547–1549.

Kumar, S. and Subramanian, S. (2002) Mutation rates in mammalian genomes. *Proceedings of the National Academy of Sciences*. [online]. 99 (2), National Academy of Sciences, pp.803–808.

## L

Laaziz, D. (2005) Small ruminant breeds of AlgeriaIn: *niguez L(ed.) Characterization of Small Ruminant Breeds in West Asia and North Africa. Vol. 2: North Africa*. International Center for Agricultural Research in Dry Areas (ICARDA): Aleppo, Syria.

Lancioni, H. et al. (2013) Phylogenetic Relationships of Three Italian Merino-Derived Sheep Breeds Evaluated through a Complete Mitogenome Analysis. *PLOS ONE*. [online]. 8 (9), Public Library of Science, p.e73712.

Larson, G. et al. (2007) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences*. [online]. 104 (39), pp.15276–15281.

- Larson, G. et al. (2010) Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proceedings of the National Academy of Sciences*. [online]. 107 (17), pp.7686–7691.
- Larson, G. and Burger, J. (2013) A population genetics view of animal domestication. *Trends in Genetics*. [online]. 29 (4), pp.197–205.
- Lasagna, E. et al. (2011) Genetic relationships and population structure in three Italian Merino-derived sheep breeds. *Small Ruminant Research*. [online]. 96 (2), pp.111–119.
- Lasarte, J., Lazkanoteg, P. and Pérez de Muniain, A. (2007) *Sasi Ardi. Una raza autóctona en peligro de extinción* 2007 [online]. Available from: <https://www.navarraagraria.com/categories/item/644-sasi-ardi-una-raza-autoctona-en-peligro-de-extincion> [Accessed 23 April 2021].
- Leach, H.M. (2003) Human Domestication Reconsidered. *Current Anthropology*. [online]. 44 (3), The University of Chicago Press, pp.349–368.
- Lenormand, T. (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*. [online]. 17 (4), pp.183–189.
- Lescheva, M., Ivolga, A. and Cvijanovic, D. (2015) *Current state and perspectives of sheep breeding development in Russian modern economic conditions* *Economics of Agriculture*, 62, 2.
- Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford, England)*. [online]. 27 (21), pp.2987–2993.
- Li, H. et al. (2020a) Genome-wide association study of milk production traits in a crossbred dairy sheep population using three statistical models. *Animal Genetics*. [online]. 51 (4), pp.624–628.
- Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. [online]. 25 (14), pp.1754–1760.
- Li, H. and Durbin, R. (2011) Inference of human population history from individual whole-genome sequences. *Nature*. [online]. 475 (7357), pp.493–496.
- Li, M.-H., Tiirikka, T. and Kantanen, J. (2014) A genome-wide scan study identifies a single nucleotide substitution in ASIP associated with white versus non-white coat-colour variation in sheep (*Ovis aries*). *Heredity*. [online]. 112 (2), pp.122–131.
- Li, N. and Stephens, M. (2003) Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics*. 165 (4), pp.2213–2233.
- Li, R. et al. (2020b) A Hu sheep genome with the first ovine Y chromosome reveal introgression history after sheep domestication. *Science China Life Sciences*. [online]. Available from: <https://doi.org/10.1007/s11427-020-1807-0> [Accessed 9 February 2021].
- Li, W.H. (1975) The first arrival time and mean age of a deleterious mutant gene in a finite population. *American Journal of Human Genetics*. 27 (3), pp.274–286.
- Li, X. et al. (2020c) Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nature Communications*. [online]. 11 (1), Nature Publishing Group, p.2815.

Li, Y.D. and Han, D.F. (1959) Animal remains excavated from Beishouling neolithic site at Baoji. pp.173–185.

Librado, P. et al. (2021) The origins and spread of domestic horses from the Western Eurasian steppes. *Nature*. [online]. pp.1–7.

Liu, J. et al. (2016) Genetic Diversity and Phylogenetic Evolution of Tibetan Sheep Based on mtDNA D-Loop Sequences. *PLOS ONE*. [online]. 11 (7), Public Library of Science, p.e0159308.

Liu, S. et al. (2017) Estimates of linkage disequilibrium and effective population sizes in Chinese Merino (Xinjiang type) sheep by genome-wide SNPs. *Genes & Genomics*. [online]. 39 (7), pp.733–745.

Liverani, M. (1987) The collapse of the Near Eastern regional system at the end of the Bronze Age: the case of SyriaIn: *Centre and Periphery in the Ancient World*. M. Rowlands, M.T. Larsen et K. Kristiansen. Cambridge University Press.

Lv, F.-H. et al. (2015) Mitogenomic Meta-Analysis Identifies Two Phases of Migration in the History of Eastern Eurasian Sheep. *Molecular Biology and Evolution*. [online]. 32 (10), pp.2515–2533.

Lynch, M. (2010) Evolution of the mutation rate. *Trends in genetics : TIG*. [online]. 26 (8), pp.345–352.

## M

Ma, G.-W. et al. (2017) Polymorphisms of FST gene and their association with wool quality traits in Chinese Merino sheep. *PLOS ONE*. [online]. 12 (4), Public Library of Science, p.e0174868.

Ma, L. et al. (2018) Genetic variants in fat- and short-tailed sheep from high-throughput RNA-sequencing data. *Animal Genetics*. [online]. 49 (5), pp.483–487.

Ma, Y.-H. et al. (2006) Phylogeography and Origin of Sheep Breeds in Northern China. *Conservation Genetics*. [online]. 7 (1), pp.117–127.

MacDonald, R.H. and MacDonald, K. (2006)In: *The origins and development of domesticated animals in arid West Africa*. Routledge.

Madani, T., Yakhlef, H. and Abbache, N. (2003) Les races bovines, ovines, caprines et camélinesIn: *Abdelguerfi, A. & S.A. Ramdane. Evaluation des besoins en matière de renforcement des capacités nécessaires à la conservation et l'utilisation durable de la biodiversité importante pour l'agriculture en Algérie*. Alger.

MADRDSASI, (Ministère de l'Agriculture et du Développement Rural /Direction des Statistiques Agricoles et des Systèmes d'Information) (2010) *Statistiques Élevages, cheptels E2009*. Alger, Algérie.

Mahrous, K.F. et al. (2016) Association between single nucleotide polymorphism in ovine Calpain gene and growth performance in three Egyptian sheep breeds. *Journal of Genetic Engineering and Biotechnology*. [online]. 14 (2), pp.233–240.

Maijala, K. (1997) Genetic aspects of domestication, common breeds and their origin. *The genetics of sheep*. CAB INTERNATIONAL, pp.13–49.

Maisels, F.G. (1988) *The feeding ecology of the Cyprus mouflon Ovis orientalis Gmelin 1774, in the Paphos Forest*, Cyprus University of Edinburgh, UK.



- Man, D. (2007) Study on Hybridization of Wild Argali and Bashibai Sheep. *undefined*. [online]. Available from: [/paper/Study-on-Hybridization-of-Wild-Argali-and-Bashibai-Man/04d8ab98667a6ea353eb877e0c04a9bbea89fc25](#) [Accessed 13 April 2021].
- Maria Sarti, F. et al. (2006) Wool quality in Gentile di Puglia sheep breed as measure of genetic integrity. *Italian Journal of Animal Science*. [online]. 5 (4), pp.371–376.
- Marshall, F. and Hildebrand, E. (2002) Cattle Before Crops: The Beginnings of Food Production in Africa. *Journal of World Prehistory*. [online]. 16 (2), pp.99–143.
- Martinez-Royo, A. et al. (2017) Genome-wide association studies for reproductive seasonality traits in Rasa Aragonesa sheep breed. *Theriogenology*. [online]. 99, pp.21–29.
- Martínez-Sánchez, R.M. et al. (2018) 'The beginning of the Neolithic in northwestern Morocco' *Quaternary International. The Neolithic expansion in the Western Mediterranean: Understanding a global phenomenon from regional perspectives* [online]. 470, pp.485–496.
- Mason, I.L. (1988) *Mason's World Dictionary of Livestock Breeds, Types and Varieties* C.A.B. International, Wallingford.
- Mastrangelo, S. et al. (2018) Genome-wide scan of fat-tail sheep identifies signals of selection for fat deposition and adaptation. *Animal Production Science*. [online]. 59 (5), CSIRO PUBLISHING, pp.835–848.
- Mastrangelo, S. et al. (2019) Novel and known signals of selection for fat deposition in domestic sheep breeds from Africa and Eurasia. *PLoS ONE*. [online]. 14 (6). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6568386/> [Accessed 11 March 2020].
- MATET (2009) *Quatrième rapport national sur la mise en oeuvre de la convention sur la diversité biologique au niveau national. Alger, Algérie*
- McIntosh, J.R. (2007) *The Ancient Indus Valley: New Perspectives* 1er édition. Santa Barbara, Calif, ABC-CLIO.
- McLaren, W. et al. (2016) The Ensembl Variant Effect Predictor. *Genome Biology*. [online]. 17 (1), p.122.
- McPeck, M.S. and Strahs, A. (1999) Assessment of linkage disequilibrium by the decay of haplotype sharing, with application to fine-scale genetic mapping. *American Journal of Human Genetics*. [online]. 65 (3), pp.858–875.
- McTavish, E.J. et al. (2013) New World cattle show ancestry from multiple independent domestication events. *Proceedings of the National Academy of Sciences*. [online]. 110 (15), pp.E1398–E1406.
- Meadows, J.R.S. et al. (2007) Five Ovine Mitochondrial Lineages Identified From Sheep Breeds of the Near East. *Genetics*. [online]. 175 (3), pp.1371–1379.
- Meadows, J.R.S. et al. (2005) Mitochondrial Sequence Reveals High Levels of Gene Flow Between Breeds of Domestic Sheep from Asia and Europe. *Journal of Heredity*. [online]. 96 (5), pp.494–501.
- Meadows, J.R.S., Hiendleder, S. and Kijas, J.W. (2011) Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. *Heredity*. [online]. 106 (4), pp.700–706.

Meadows, R.H. (1984) 'Early Herders and their Flocks' In: *Animals and Archaeology 3. Early Herders and their Flocks*, British Archaeological Reports Juliet Clutton-Brock, Caroline Grigson, International Council for Archaeozoology, University of London. Institute of Archaeology, 309–337.

Megdiche, S. et al. (2019) A Combined Multi-Cohort Approach Reveals Novel and Known Genome-Wide Selection Signatures for Wool Traits in Merino and Merino-Derived Sheep Breeds. *Frontiers in Genetics*. [online]. 10. Available from: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01025/full> [Accessed 10 March 2020].

Meier, M. (2016) The 'Justinianic Plague': the economic consequences of the pandemic in the eastern Roman empire and its cultural and religious effects. *Early Medieval Europe*. [online]. 24 (3), pp.267–292.

Meiri, M. et al. (2017) Eastern Mediterranean Mobility in the Bronze and Early Iron Ages: Inferences from Ancient DNA of Pigs and Cattle. *Scientific Reports*. [online]. 7 (1), p.701.

Meligkotsidou, L. and Fearnhead, P. (2005) Maximum-Likelihood Estimation of Coalescence Times in Genealogical Trees. *Genetics*. [online]. 171 (4), pp.2073–2084.

Meyer, M.C. et al. (2017) Permanent human occupation of the central Tibetan Plateau in the early Holocene. *Science*. [online]. 355 (6320), American Association for the Advancement of Science, pp.64–67.

Michailidou, S. et al. (2018) Genomic diversity and population structure of three autochthonous Greek sheep breeds assessed with genome-wide DNA arrays. *Molecular Genetics and Genomics*. [online]. 293 (3), pp.753–768.

Michel, S. and Ghoddousi (2020) *IUCN Red List of Threatened Species: Ovis gmelini IUCN Red List of Threatened Species*. 18 March 2020 [online]. Available from: <https://www.iucnredlist.org/en> [Accessed 2 September 2020].

Moioli, B. et al. (2013) The ACACA gene is a potential candidate gene for fat content in sheep milk. *Animal Genetics*. [online]. 44 (5), pp.601–603.

Molae, V., Eltanany, M. and Lühken, G. (2018) First survey on association of TMEM154 and CCR5 variants with serological maedi-visna status of sheep in German flocks. *Veterinary Research*. [online]. 49 (1), p.36.

Montelaro, R. et al. (1989) Antigenic variation during persistent lentivirus infections and its implications for vaccine development. *Advances in Experimental Medicine and Biology*. [online]. 251, pp.251–272.

Moradi, M.H. et al. (2012) Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC Genetics*. [online]. 13 (1), p.10.

Moradian, C. et al. (2013) Effects of genetic polymorphism at the growth hormone gene on growth traits in Makooei sheep. *European Journal of Experimental Biology*. [online]. 3 (3), iMedPub. Available from: <https://www.imedpub.com/abstract/effects-of-genetic-polymorphism-at-the-growth-hormone-gene-on-growth-traitsrnin-makooei-sheep-14114.html> [Accessed 24 June 2021].

Mordechai, L. et al. (2019) The Justinianic Plague: An inconsequential pandemic? *Proceedings of the National Academy of Sciences*. [online]. 116 (51), National Academy of Sciences, pp.25546–25554.

Morris, A.P., Whittaker, J.C. and Balding, D.J. (2000) Bayesian Fine-Scale Mapping of Disease Loci, by Hidden Markov Models. *The American Journal of Human Genetics*. [online]. 67 (1), pp.155–169.

Moyers, B.T., Morrell, P.L. and McKay, J.K. (2018) Genetic Costs of Domestication and Improvement. *Journal of Heredity*. [online]. 109 (2), pp.103–116.

Muigai, A. W. T (2003) *Characterisation and conservation of indigenous animal genetic resources: The fat-tailed and thin-tailed sheep of Africa* Juja, Jomo Kenyatta University of Agriculture and Technology.

Muigai, A.W.T. and Hanotte, O. (2013) The Origin of African Sheep: Archaeological and Genetic Perspectives. *African Archaeological Review*. [online]. 30 (1), pp.39–50.

Mukasa-Mugerwa, E. et al. (2000) Between and within breed variation in lamb survival and the risk factors associated with major causes of mortality in indigenous Horro and Menz sheep in Ethiopia. *Small Ruminant Research*. [online]. 37 (1–2), pp.1–12.

Müller, K. et al. (2012) Effect of continuous v. daytime grazing on feed intake and growth of sheep grazing in a semi-arid grassland steppe. *animal*. [online]. 6 (3), Cambridge University Press, pp.526–534.

Mwacharo, J.M. et al. (2017) Genomic footprints of dryland stress adaptation in Egyptian fat-tail sheep and their divergence from East African and western Asia cohorts. *Scientific Reports*. [online]. 7 (1), pp.1–10.

## N

Nachman, M.W. and Crowell, S.L. (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics*. [online]. 156 (1), pp.297–304.

Naderi, S. et al. (2008) The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proceedings of the National Academy of Sciences*. [online]. 105 (46), pp.17659–17664.

Nadler, C.F., Hoffmann, R.S. and Woolf, A. (1973) G-band patterns as chromosomal markers, and the interpretation of chromosomal evolution in wild sheep (*Ovis*). *Experientia*. [online]. 29 (1), pp.117–119.

Nair, R. et al. (2012) Neurobeachin regulates neurotransmitter receptor trafficking to synapses. *Journal of Cell Biology*. [online]. 200 (1), pp.61–80.

Nakagome, S., Hudson, R.R. and Di Rienzo, A. (2019) Inferring the model and onset of natural selection under varying population size from the site frequency spectrum and haplotype structure. *Proceedings of the Royal Society B: Biological Sciences*. [online]. 286 (1896), Royal Society, p.20182541.

Naval-Sanchez, M. et al. (2018) Sheep genome functional annotation reveals proximal regulatory elements contributed to the evolution of modern breeds. *Nature Communications*. [online]. 9 (1), pp.1–13.

Nayeri, S. et al. (2016) Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. *BMC Genetics*. [online]. 17 (1), BioMed Central, pp.1–11.

Nazari-Ghadikolaei, A. et al. (2018) Genome-Wide Association Studies Identify Candidate Genes for Coat Color and Mohair Traits in the Iranian Markhoz Goat. *Frontiers in Genetics*. [online]. 9, Frontiers. Available from: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00105/full> [Accessed 8 June 2021].

Nehring, A. (1891) Diluviale Reste von Cuon, Ovis, Saiga, Ibex und Rupicapra aus Mahren. 2, pp.107–155.

Nguyen, L.-T. et al. (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*. [online]. 32 (1), pp.268–274.

Niemi, M. et al. (2013) Mitochondrial DNA and Y-chromosomal diversity in ancient populations of domestic sheep (*Ovis aries*) in Finland: comparison with contemporary sheep breeds. *Genetics Selection Evolution*. [online]. 45 (1), p.2.

Nishino, J. et al. (2008) Hmga2 Promotes Neural Stem Cell Self-Renewal in Young but Not Old Mice by Reducing p16Ink4a and p19Arf Expression. *Cell*. [online]. 135 (2), Elsevier, pp.227–239.

Norris, B.J. and Whan, V.A. (2008) A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome Research*. [online]. 18 (8), pp.1282–1293.

## O

Ollivier, M. et al. (2016) Amy2B copy number variation reveals starch diet adaptations in ancient European dogs. *Royal Society Open Science*. [online]. 3 (11), p.160449.

Oltenu, E. and Boylan, W.J. (1981) Productivity of purebred and crossbred finnsheep. I. Reproductive traits of ewes and lamb survival. *Journal of animal science*. [online].

Open Targets Genetics Portal (no date) *Evidence for ADCY10 in testosterone measurement* [online]. Available from: [https://platform.opentargets.org/evidence/ENSG00000143199/EFO\\_0004908](https://platform.opentargets.org/evidence/ENSG00000143199/EFO_0004908) [Accessed 30 September 2021].

Ormond, L. et al. (2016) Inferring the age of a fixed beneficial allele. *Molecular Ecology*. [online]. 25 (1), pp.157–169.

Ou, Y.X. (1936) *Xin Tang Shu* Taipei: The Commercial Press of Taiwan.

Ozerov, M.Y. et al. (2019) Genetic factors affecting genetic variance in coarse-wool sheep. *Rossiiskaia selskokhoziaistvennaia nauka*. [online]. 0 (6), pp.40–44.

Ozerov, M.Yu. et al. (2020) Environmental Factors Affecting Genetic Variation in Coarse-Wool Sheep. *Russian Agricultural Sciences*. [online]. 46 (1), pp.65–70.

Ozmen, O., Kul, S. and Gok, T. (2020) Determination of genetic diversity of the Akkaraman sheep breed from Turkey. *Small Ruminant Research*. [online]. 182, pp.37–45.

## P

Palamara, P.F. et al. (2018) High-throughput inference of pairwise coalescence times identifies signals of selection and enriched disease heritability. *Nature Genetics*. [online]. 50 (9), pp.1311–1317.

- Paludan, S.R. et al. (2011) Recognition of herpesviruses by the innate immune system. *Nature Reviews Immunology*. [online]. 11 (2), pp.143–154.
- Pan, Z. et al. (2018a) Whole-genome sequences of 89 Chinese sheep suggest role of RXFP2 in the development of unique horn phenotype as response to semi-feralization. *GigaScience*. [online]. 7 (4).
- Pan, Z. et al. (2018b) Whole-genome sequences of 89 Chinese sheep suggest role of RXFP2 in the development of unique horn phenotype as response to semi-feralization. *GigaScience*. [online]. 7 (4). Available from: <https://academic-oup-com.docelec.univ-lyon1.fr/gigascience/article/7/4/giy019/4924504> [Accessed 16 January 2019].
- Pang, J.-F. et al. (2009) mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Molecular Biology and Evolution*. [online]. 26 (12), pp.2849–2864.
- Paradis, E. (2010) pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*. [online]. 26 (3), pp.419–420.
- Paradis, E., Claude, J. and Strimmer, K. (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*. [online]. 20 (2), pp.289–290.
- Pardeshi, V.C. et al. (2007) Mitochondrial haplotypes reveal a strong genetic structure for three Indian sheep breeds. *Animal Genetics*. [online]. 38 (5), pp.460–466.
- Parmesan, C. and Yohe, G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*. [online]. 421 (6918), Nature Publishing Group, pp.37–42.
- Pearson, L.D., Poss, M.L. and Demartini, J.C. (1989) ‘Animal lentivirus vaccines: Problems and prospects’ *Veterinary Immunology and Immunopathology. Special Issue: Advances in Veterinary Immunology 1988* [online]. 20 (3), pp.183–212.
- Pedrosa, S. et al. (2005a) Evidence of three maternal lineages in near eastern sheep supporting multiple domestication events. *Proceedings of the Royal Society B: Biological Sciences*. [online]. 272 (1577), pp.2211–2217.
- Pedrosa, S. et al. (2005b) Evidence of three maternal lineages in near eastern sheep supporting multiple domestication events. *Proceedings of the Royal Society B: Biological Sciences*. [online]. 272 (1577), pp.2211–2217.
- Pedrosa, S. et al. (2007) Mitochondrial diversity and the origin of Iberian sheep. *Genetics Selection Evolution*. [online]. 39 (1), p.91.
- Peischl, S. et al. (2013) On the accumulation of deleterious mutations during range expansions. *Molecular Ecology*. [online]. 22 (24), pp.5972–5982.
- Peng, W. (2010) A textual research on “sheep cart.” (10), pp.71–75.
- Pépin, M. et al. (1998) Maedi-visna virus infection in sheep: a review. *Veterinary Research*. 29 (3–4), BioMed Central, pp.341–367.
- Pereira, F. et al. (2006) Genetic Signatures of a Mediterranean Influence in Iberian Peninsula Sheep Husbandry. *Molecular Biology and Evolution*. [online]. 23 (7), pp.1420–1426.

- Perk, K. et al. (1996) Evaluation of vaccines for ovine lentivirus infection. *AIDS research and human retroviruses*. [online]. 12 (5), pp.425–426.
- Perreira, J.M. et al. (2013) 'IFITMs Restrict the Replication of Multiple Pathogenic Viruses' *Journal of Molecular Biology. Antiviral Innate Immunity (Part I)* [online]. 425 (24), pp.4937–4955.
- Peter, B.M., Huerta-Sanchez, E. and Nielsen, R. (2012) Distinguishing between Selective Sweeps from Standing Variation and from a De Novo Mutation. *PLOS Genetics*. [online]. 8 (10), Public Library of Science, p.e1003011.
- Peter, C. et al. (2007) Genetic diversity and subdivision of 57 European and Middle-Eastern sheep breeds. *Animal Genetics*. [online]. 38 (1), pp.37–44.
- Peters, J., von den Driesch, A. and Helmer, D. (2005) The upper Euphrates-Tigris basin: cradle of agropastoralism? In: *The first steps of animal domestication: new archaeobiological approaches*.
- Petit, M. et al. (2017) Variation in Recombination Rate and Its Genetic Determinism in Sheep Populations. *Genetics*. [online]. 207 (2), pp.767–784.
- Philippine Council for Agriculture, Forestry and Natural Resources, Research and Development; International Development Research Centre (1988) *Sheep production in Asia* Proceedings of the Workshop on Sheep Production in Asia Los Baños, C. Devendra and P.S. Faylon; Book Series. No. 80/1989.
- Picard, C. (1997) *L'océan Atlantique musulman; De la conquête arabe à l'époque almohade; Navigation et mise en valeur des côtes d'al-Andalus et du Maghreb occidental (Portugal- Espagne-Maroc)* Paris: Maisonneuve et Larose.
- Pionnier-Capitan, M. et al. (2011) 'New evidence for Upper Palaeolithic small domestic dogs in South-Western Europe' *Journal of Archaeological Science. Satellite remote sensing in archaeology: past, present and future perspectives* [online]. 38 (9), pp.2123–2140.
- Pipia, A. et al. (2008) Influence of sex, season, temperature and reproductive status on daily activity patterns in Sardinian mouflon (*Ovis orientalis musimon*). *Behaviour*. [online]. 145, pp.1723–1745.
- Platt, A. et al. (2019) An estimator of first coalescent time reveals selection on young variants and large heterogeneity in rare allele ages among human populations. Lohmueller, K.E. (ed.) *PLOS Genetics*. [online]. 15 (8), p.e1008340.
- Pleurdeau, D. et al. (2012) "Of Sheep and Men": Earliest Direct Evidence of Caprine Domestication in Southern Africa at Leopard Cave (Erongo, Namibia). *PLoS ONE*. [online]. 7 (7). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3394794/> [Accessed 10 March 2020].
- Pommerol, F. (1879) *Le mouflon quaternaire*. Ass. Franç. Avancement des Sciences. C. R. 8e Sess. Montpellier, 600–609.
- Poplin, F. (1979) Origine du Mouflon de Corse dans une nouvelle perspective paléontologique: par marronnage. 11.
- Price, E.O. (1999) Behavioral development in animals undergoing domestication. *Applied Animal Behaviour Science*. [online]. 65 (3), pp.245–271.

Purcell, S. et al. (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics*. 81 (3), pp.559–575.

## Q

Qiu, J. (2015) Who are the Tibetans? *Science*. [online]. 347 (6223), American Association for the Advancement of Science, pp.708–711.

## R

Raadsma, H., Gray, G. and Woolaston, R. (1998) Breeding for disease resistance in Merino sheep in Australia. *Revue Scientifique et Technique (International Office of Epizootics)*. [online]. 17 (1), pp.315–328.

Ramayo-Caldas, Y. et al. (2012) Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition. *BMC Genomics*. [online]. 13 (1), BioMed Central, pp.1–18.

Ran, G.R., Li, S.M. and Zhou, X.Y. (1985) *Qiang Zu Shi*. Chengdu: Sichuan Nationalities.

Rannala, B. and Reeve, J.P. (2003) Joint Bayesian estimation of mutation location and age using linkage disequilibrium. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*. [online]. pp.526–534.

Rannala and Reeve (2001) High-resolution multipoint linkage-disequilibrium mapping in the context of a human genome sequence. *American journal of human genetics*. [online]. 69 (1), Am J Hum Genet. Available from: <https://pubmed.ncbi.nlm.nih.gov/11410841/> [Accessed 25 October 2021].

Rannamäe, E. et al. (2016) Maternal and paternal genetic diversity of ancient sheep in Estonia from the Late Bronze Age to the post-medieval period and comparison with other regions in Eurasia. *Animal Genetics*. [online]. 47 (2), pp.208–218.

Rannamäe, E. et al. (2020) Retroviral analysis reveals the ancient origin of Kihnu native sheep in Estonia: implications for breed conservation. *Scientific Reports*. [online]. 10 (1), Nature Publishing Group, p.17340.

Rao, S. and Notter, D.R. (2000) Genetic analysis of litter size in Targhee, Suffolk, and Polypay sheep. *Journal of Animal Science*. [online]. 78 (8), pp.2113–2120.

Rasali, D., Shrestha, J.N.B. and Crow, G.H. (2006) Development of composite sheep breeds in the world: A review. *Canadian Journal of Animal Science*. 86, pp.1–24.

Rather, M.A. et al. (2019) Genetic evaluation of wool traits of Kashmir Merino sheep in organized farms. *Small Ruminant Research*. [online]. 177, pp.14–17.

Rege, J.E.O., Yapi-Gnaore, C.V. and Tawah, C.L. (1996) The indigenous domestic ruminant genetic resource. [online]. South African Society of Animal Science. Available from: <https://agris.fao.org/agris-search/search.do?recordID=QT2016100541> [Accessed 4 May 2021].

Reissmann, M. and Ludwig, A. (2013) Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals' *Seminars in Cell & Developmental Biology. Genetic basis and evolutionary causes of colour variation in vertebrates* [online]. 24 (6), pp.576–586.

- Rendo, F. et al. (2004) Tracking diversity and differentiation in six sheep breeds from the North Iberian Peninsula through DNA variation. *Small Ruminant Research*. [online]. 52 (3), pp.195–202.
- Resende, A. et al. (2016) Mitochondrial DNA variation of domestic sheep (*Ovis aries*) in Kenya. *Animal genetics*. [online]. 47.
- Rezaei, H.R. et al. (2010) Evolution and taxonomy of the wild species of the genus *Ovis* (Mammalia, Artiodactyla, Bovidae). *Molecular Phylogenetics and Evolution*. [online]. 54 (2), pp.315–326.
- Ricordeau, G. et al. (1978) First survey of results obtained in France on reproductive and maternal performance in sheep, with particular reference to the Romanov breed and crosses with it. *Livestock Production Science*. [online]. 5 (2), pp.181–201.
- Robbins, L. et al. (2008) Recent Archaeological Research at Toteng, Botswana: Early Domesticated Livestock in the Kalahari. *Journal of African Archaeology*. [online]. 6, pp.131-149.
- Rocha, J., Chen, S. and Beja-Pereira, A. (2011) Molecular evidence for fat-tailed sheep domestication. *Tropical Animal Health and Production*. [online]. 43 (7), pp.1237–1243.
- Rochus, C.M. (2017) *Genetic diversity of domestic sheep* Doctoral thesis 2017 [online]. Available from: <https://pub.epsilon.slu.se/14322/> [Accessed 23 May 2018].
- Rochus, C.M. et al. (2019) Mutations in ASIP and MC1R: dominant black and recessive black alleles segregate in native Swedish sheep populations. *Animal Genetics*. [online]. 50 (6), pp.712–717.
- Rosen, W. (2007) *Justinian's Flea: The First Great Plague and the End of the Roman Empire* Penguin.
- Roustaei, K. and Mashkour, M. (2016) The Neolithic of the Iranian Plateau. Recent Research and Prospects. Studies in Early Near Eastern Production, Subsistence, and Environment 18 (SENEPSE series) In: K. Roustaei and M. Mashkour eds. ex oriente, Berlin, i–vii.
- Roustaei, K., Mashkour, M. and Tengberg, M. (2015) Tappeh Sang-e Chakhmaq and the beginning of the Neolithic in north-east Iran. *Antiquity*. [online]. 89 (345), Cambridge University Press, pp.573–595.
- Ruiz-de-la-Torre, J.L. and Manteca, X. (1999) Effects of testosterone on aggressive behaviour after social mixing in male lambs. *Physiology & Behavior*. [online]. 68 (1), pp.109–113.
- Ruiz-Larrañaga, O. et al. (2020) Genetic structure of Iranian indigenous sheep breeds: insights for conservation. *Tropical Animal Health and Production*. [online]. Available from: <https://doi.org/10.1007/s11250-020-02252-3> [Accessed 10 March 2020].
- Rupp, R. 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*. [online]. 11 (12), Public Library of Science, p.e1005629.
- Ruth, K.S. et al. (2020) Using human genetics to understand the disease impacts of testosterone in men and women. *Nature Medicine*. [online]. 26 (2), pp.252–258.
- Ruyter-Spira, C.P. (1998) The *HMGI-C* gene is a likely candidate for the autosomal dwarf locus in the chicken. *Journal of Heredity*. [online]. 89 (4), pp.295–300.



Ryder, M.L. (1991) Domestication, history and breed evolution in sheep. In: World Animal Science, B 8. Genetic Resources of Pig, Sheep and Goat (ed. Majjala K), 157–177.

Ryder, M.L. (1984) Evolution of domesticated animals. In: *Sheep*,. Mason I.L. (Ed.). Longman, London and New York, 63–85.

Ryder, M.L. (1983) *Sheep and Man* London, Duckworth.

## S

Sagne, J. and d'Arces, P.J. (1950) *Jean Sagne. L'Algérie pastorale, ses origines, sa formation, son passé, son présent, son avenir. Préface de... Pierre Jore d'Arces*

Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*. [online]. 4 (4), pp.406–425.

Sallam, A.M. (2021) A missense mutation in the coding region of the toll-like receptor 4 gene affects milk traits in Barki sheep. *Animal Bioscience*. [online]. 34 (4), pp.489–498.

Sánchez, A. and Sánchez, M.C. (1986) *Razas Ovinas Españolas, Ministerio de Agricultura, Pesca y Alimentación* [online]. Madrid. Available from: <https://www.mapa.gob.es/es/ganaderia/temas/zootecnia/razas-ganaderas/razas/catalogo-razas/> [Accessed 15 April 2021].

Sanchez-Juan, P. et al. (2012) Genome-wide study links MTMR7 gene to variant Creutzfeldt-Jakob risk. *Neurobiology of Aging*. [online]. 33 (7), p.1487.e21-1487.e28.

Sang, G.B. (1979) Tomb No.1 of the Majiayao-relic type, Hetaozhuang of Minhe County, Qinghai. pp.29–32.

Sanna, D. et al. (2015) The First Mitogenome of the Cyprus Mouflon (*Ovis gmelini ophion*): New Insights into the Phylogeny of the Genus *Ovis*. Achilli, A. (ed.) *PLOS ONE*. [online]. 10 (12), p.e0144257.

Savic, M. et al. (2011) *Autochthonous breeds as livestock potential in sustainable farming in Serbia* [online]. International Congress of Mediterranean Federation of Health and Production of Ruminants, 19, Belgrade (Serbia), 25-28 May 2011 University of Belgrade, Faculty of Veterinary Medicine. Available from: <https://agris.fao.org/agris-search/search.do?recordID=RS2011002612> [Accessed 20 April 2021].

Schiffels, S. and Durbin, R. (2014) Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*. [online]. 46 (8), pp.919–925.

Schweizer, R.M. et al. (2018) Natural Selection and Origin of a Melanistic Allele in North American Gray Wolves. *Molecular Biology and Evolution*. [online]. 35 (5), pp.1190–1209.

Scott, J.C., Demoule, J.-P. and Saint-Upéry, M. (2019) *Homo Domesticus* Paris, La Découverte.

Sealy, J. and Yates, R. (1996) Direct radiocarbon dating of early sheep bones: Two further results. 51, pp.109–110.

Sealy, J. and Yates, R. (1994) The chronology of the introduction of pastoralism to the Cape, South Africa - ProQuest. 68, pp.58–67.

- Sel'kin II and Aboneev VV (2007) *Severokavkazskay myaso-sherstnaya poroda* Stavropol: BI.
- Sel'kin II and Sokolov AN (2002) *Sozdanie i sovershenstvovanie polytonkorunnykh porod ovets*. Ovtsty, kosal, sherstyanoe delo.
- Serranito, B. et al. (2021a) Local adaptations of Mediterranean sheep and goats through an integrative approach. *bioRxiv*. [online]. Cold Spring Harbor Laboratory, p.2021.01.22.427461.
- Serranito, B. et al. (2021b) Search for Selection Signatures Related to Trypanosomosis Tolerance in African Goats. *Frontiers in Genetics*. [online]. 12, p.1305.
- Serrano, M. et al. (2021) Genome-wide association studies for sperm traits in Assaf sheep breed. *Animal*. [online]. 15 (2), p.100065.
- Shannon, L.M. et al. (2015) Genetic structure in village dogs reveals a Central Asian domestication origin. *Proceedings of the National Academy of Sciences*. [online]. 112 (44), National Academy of Sciences, pp.13639–13644.
- Shaw, P.A. et al. (2003) 'Holocene fluctuations of Lake Ngami, Middle Kalahari: chronology and responses to climatic change' *Quaternary International*. *Late Quaternary environmental change in African Drylands* [online]. 111 (1), pp.23–35.
- Shi, N. et al. (2001) Southeast trade wind variations during the last 135 kyr: evidence from pollen spectra in eastern South Atlantic sediments. *Earth and Planetary Science Letters*. [online]. 187 (3), pp.311–321.
- Si, M.Q. (1131) *Shi Ji, Sections on Yin Ben Ji and Qin Ben Ji* Yangzhou: Huainan Road Transport Department.
- Signorello, G. and Pappalardo, G. (2003) 'Domestic animal biodiversity conservation: a case study of rural development plans in the European Union' *Ecological Economics*. *Valuing Animal Genetic Resources* [online]. 45 (3), pp.487–499.
- Singh, S. et al. (2013) Extensive Variation and Sub-Structuring in Lineage A mtDNA in Indian Sheep: Genetic Evidence for Domestication of Sheep in India. *PLOS ONE*. [online]. 8 (11), p.e77858.
- Singh, S., Kumar, S. and Kolte, A.P. (2013) Extensive Variation and Sub-Structuring in Lineage A mtDNA in Indian Sheep: Genetic Evidence for Domestication of Sheep in India. *PLoS ONE*. [online]. 8 (11). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3823876/> [Accessed 10 March 2020].
- Slatkin, M. (2001) Simulating genealogies of selected alleles in a population of variable size. *Genetics Research*. [online]. 78 (1), Cambridge University Press, pp.49–57.
- Slatkin, M. and Rannala, B. (2000) Estimating Allele Age. *Annual review of genomics and human genetics*. [online]. 1, pp.225–49.
- Smith, A. (2008) Early Herders in Southern Africa: a synthesis In: *Animals and People: Archaeozoological Papers in Honour of Ina Plug*. Badenhorst S, Mitchell P, Driver J. Oxford: B BAR International Series 1849, 94–103.
- Smith, A. (2009) Pastoralism in the Western Cape Province, South Africa: A Retrospective Review. *Journal of African Archaeology*. [online]. 7, pp.239–252.

Smith, A.B. (2006) *Excavations at Kasteelberg and the Origins of the Khoekhoen in the Western Cape, South Africa* Oxford, England, BAR Publishing.

Smith, A.B. (1992) *Pastoralism in Africa: Origins and development ecology* London: Hurst.

Smith, J. et al. (2018) Estimating Time to the Common Ancestor for a Beneficial Allele. *Molecular Biology and Evolution*. [online]. 35 (4), pp.1003–1017.

Smith, J. (2019) *Leveraging Haplotype-Based Inference to Describe Adaptation and Speciation* [online].

Snyman, M.A. (2014) South African sheep breeds : Namaqua Afrikaner.

Sorre, M. (1923) 'La Mesta', d'après le livre de J. Klein. *Bulletin hispanique*. [online]. 25 (3), Persée - Portail des revues scientifiques en SHS, pp.237–252.

Speidel, L. et al. (2019) A method for genome-wide genealogy estimation for thousands of samples. *Nature Genetics*. [online]. 51 (9), Nature Publishing Group, pp.1321–1329.

Stapley, J. et al. (2010) Adaptation genomics: the next generation. *Trends in Ecology & Evolution*. [online]. 25 (12), pp.705–712.

Straub, O.C. (2004) Maedi–Visna virus infection in sheep. History and present knowledge. *Comparative Immunology, Microbiology and Infectious Diseases*. [online]. 27 (1), pp.1–5.

Sulayman, A. et al. (2017) Identification of Polymorphisms and Association of Five KAP Genes with Sheep Wool Traits. *Asian-Australasian Journal of Animal Sciences*. [online]. Asian-Australasian Association of Animal Production Societies. Available from: <https://www.animbiosci.org/journal/view.php?doi=10.5713/ajas.16.0632> [Accessed 22 June 2021].

Sun, W. et al. (2013) Genome Array of Hair Follicle Genes in Lambskin with Different Patterns. *PLOS ONE*. [online]. 8 (7), Public Library of Science, p.e68840.

Sun, Y. (1987) 夏官·羊人 [Summer Office. Sheep Man], in 周禮正義 [Zhou Li Zheng Yi], ed. Chen, J. 陈金生. Beijing: Zhong Hua Shu Ju 中华书局, 2393-5.

Sutera, A.M. et al. (2019) Genome-wide association studies for milk production traits in Valle del Belice sheep using repeated measures. *Animal Genetics*. [online]. 50 (3), pp.311–314.

Szpiech, Z. and Hernandez, R. (2014) selscan: An Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection. *Molecular biology and evolution*. [online]. 31.

## T

Taberlet, P. et al. (2008) Are cattle, sheep, and goats endangered species? *Molecular Ecology*. [online]. 17 (1), pp.275–284.

Taberlet, P. et al. (2011) 'Conservation genetics of cattle, sheep, and goats' *Comptes Rendus Biologies. On the trail of domestications, migrations and invasions in agriculture* [online]. 334 (3), pp.247–254.

Taberlet, P. and Luikart, G. (1999) Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society*. [online]. 68 (1–2), pp.41–55.

- Takahashi, H. et al. (2008) Genetic Structure of Mongolian Goat Populations Using Microsatellite Loci Analysis. *Asian-Australasian Journal of Animal Sciences*. [online]. 21 (7), pp.947–953.
- Tang, H. et al. (2002) Frequentist Estimation of Coalescence Times From Nucleotide Sequence Data Using a Tree-Based Partition. *Genetics*. 161 (1), Genetics, pp.447–459.
- Tapio, M. et al. (2010) Microsatellite-based genetic diversity and population structure of domestic sheep in northern Eurasia. *BMC Genetics*. [online]. 11 (1), p.76.
- Tapio, M. et al. (2005) Native breeds demonstrate high contributions to the molecular variation in northern European sheep. *Molecular Ecology*. [online]. 14 (13), pp.3951–3963.
- Tapio, M. et al. (2006a) Sheep Mitochondrial DNA Variation in European, Caucasian, and Central Asian Areas. *Molecular Biology and Evolution*. [online]. 23 (9), pp.1776–1783.
- Tapio, M. et al. (2006b) Sheep Mitochondrial DNA Variation in European, Caucasian, and Central Asian Areas. *Molecular Biology and Evolution*. [online]. 23 (9), pp.1776–1783.
- Terhorst, J., Kamm, J.A. and Song, Y.S. (2017) Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*. [online]. 49 (2), pp.303–309.
- Thomas, D.L. (2010) Performance and utilization of Northern European short-tailed breeds of sheep and their crosses in North America: a review. *animal*. [online]. 4 (8), Cambridge University Press, pp.1283–1296.
- Thomson, R. et al. (2000) Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. *Proceedings of the National Academy of Sciences*. [online]. 97 (13), National Academy of Sciences, pp.7360–7365.
- Thormar, H. (2004) Maedi-visna virus and its relationship to human immunodeficiency virus. *AIDS reviews*. 7, pp.233–45.
- Tibbo, M., Woldemeskel, M. and Gopilo, A. (2001) An Outbreak of Respiratory Disease Complex in Sheep in Central Ethiopia. p.11.
- Tigano, A. and Friesen, V.L. (2016) Genomics of local adaptation with gene flow. *Molecular Ecology*. [online]. 25 (10), pp.2144–2164.
- Tollenaere, C. et al. (2013) Beyond an AFLP genome scan towards the identification of immune genes involved in plague resistance in *Rattus rattus* from Madagascar. *Molecular Ecology*. [online]. 22 (2), pp.354–367.
- Tornero, C. et al. (2020) Early evidence of sheep lambing de-seasoning in the Western Mediterranean in the sixth millennium BCE. *Scientific Reports*. [online]. 10 (1), Nature Publishing Group, p.12798.
- Tornero, C. et al. (2016) Seasonal reproductive patterns of early domestic sheep at Tell Halula (PPNB, Middle Euphrates Valley): Evidence from sequential oxygen isotope analyses of tooth enamel. *Journal of Archaeological Science: Reports*. [online]. 6, pp.810–818.
- Tresset, A. and Vigne, J.-D. (2007) *Going Over: The Mesolithic-Neolithic Transition in North-West Europe* British Academy.

Tresset, A. and Vigne, J.-D. (2011) Last hunter-gatherers and first farmers of Europe. *Comptes Rendus Biologies*. [online]. 334 (3), pp.182–189.

Trow-Smith, R. (2013) *A History of British Livestock Husbandry, 1700-1900* Routledge.

Trut, L., Oskina, I. and Kharlamova, A. (2009) Animal evolution during domestication: the domesticated fox as a model. *BioEssays*. [online]. 31 (3), pp.349–360.

Trut, L.N., Plyusnina, I.Z. and Oskina, I.N. (2004) An Experiment on Fox Domestication and Debatable Issues of Evolution of the Dog. *Russian Journal of Genetics*. [online]. 40 (6), pp.644–655.

## U

Uerpmann, H.-P. (1987) *The ancient distribution of Ungulate Mammals in the Middle East*, Beihefte zum Tübingen Atlas des Vorderen Orients, Reihe, No. 27 (L.Reichert Verlag, Wiesbaden, 1987), pp. 173.

Urarte, E.U. (1989) *La raza latxa: sistemas de producción y características reproductivas*<http://purl.org/dc/dcmitype/Text> [online]. Universidad de Zaragoza. Available from: <https://dialnet.unirioja.es/servlet/tesis?codigo=205201> [Accessed 23 April 2021].

## V

Våge, D.I. and Boman, I.A. (2010) A nonsense mutation in the beta-carotene oxygenase 2 (BCO2) gene is tightly associated with accumulation of carotenoids in adipose tissue in sheep (*Ovis aries*). *BMC Genetics*. [online]. 11 (1), BioMed Central, pp.1–6.

Verma, P. et al. (2018) Transcriptome Analysis of Circulating PBMCs to Understand Mechanism of High Altitude Adaptation in Native Cattle of Ladakh Region. *Scientific Reports*. [online]. 8 (1), p.7681.

Vermeersch, P. et al. (1996) Neolithic occupation of the Sodmein area, Red Sea Mountains, EgyptIn: 411–420.

Verrier, E., Brives, H. and Heams, T. (2011) *Races animales, éleveurs et sociétés : évolution d'un concept et des pratiques de sélection*CGAAER/OIE « L'évolution des relations entre l'homme et l'animal : une approche transdisciplinaire » Paris.

Viger, A. (2016) *Étude sur la question ovine en Algérie* Hachette Livre BNF.

Vigne, J.-D. (2015) Early domestication and farming: what should we know or do for a better understanding? *Anthropozoologica*. [online]. 50 (2), pp.123–150.

Vigne, J.-D. et al. (2012) First wave of cultivators spread to Cyprus at least 10,600 y ago. *Proceedings of the National Academy of Sciences*. [online]. 109 (22), pp.8445–8449.

Vigne, J.-D. (2017) *Les débuts de l'élevage* Humensis.

Vigne, J.-D. (2011)'The origins of animal domestication and husbandry: A major change in the history of humanity and the biosphere' *Comptes Rendus Biologies*.*On the trail of domestications, migrations and invasions in agriculture* [online]. 334 (3), pp.171–181.

Vigne, J.-D. et al. (2014) The transportation of mammals to Cyprus sheds light on early voyaging and boats in the Mediterranean Sea. *Eurasian Prehistory*. 10.

Vigne, J.-D. (1992) Zooarchaeology and the biogeographical history of the mammals of Corsica and Sardinia since the last Ice Age. *Mamm. Rev.* 22.

Vigne, J.-D., Buitenhuis, H. and Davis, S.J.M. (1999) Les premiers pas de la domestication animale à l'Ouest de l'Euphrate : Chypre et l'Anatolie centrale. *Paléorient*. [online]. 25 (2), Persée - Portail des revues scientifiques en SHS, pp.49–62.

Vigne, J.-D. and Helmer, D. (2007) Was milk a “secondary product” in the Old World Neolithisation process? Its role in the domestication of cattle, sheep and goats. p.32.

Vila, E. et al. (2021) EVOSHEEP: the makeup of sheep breeds in the ancient Near East. *Antiquity*. [online]. 95 (379), Cambridge University Press. Available from: <https://www.cambridge.org/core/journals/antiquity/article/evosheep-the-makeup-of-sheep-breeds-in-the-ancient-near-east/9C2492EC4DD3481D2D43C63902FCD184> [Accessed 15 February 2021].

Voight, B.F. et al. (2006) A Map of Recent Positive Selection in the Human Genome. *PLOS Biology*. [online]. 4 (3), Public Library of Science, p.e72.

## W

Wang, G.-D. et al. (2016) Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Research*. [online]. 26 (1), pp.21–33.

Wang, H. et al. (2015) Genome-Wide Specific Selection in Three Domestic Sheep Breeds. *PLOS ONE*. [online]. 10 (6), Public Library of Science, p.e0128688.

Wang, X. et al. (2014a) Discovery of SNPs in RXFP2 related to horn types in sheep. *Small Ruminant Research*. [online]. 116 (2), pp.133–136.

Wang, X. et al. (2014b) Transcriptome profile analysis of adipose tissues from fat and short-tailed sheep. *Gene*. [online]. 549 (2), pp.252–257.

Wang, Y. (2017) *Identifying the beginnings of Sheep Husbandry in Western China* Thesis [online]. University of Cambridge. Available from: <https://www.repository.cam.ac.uk/handle/1810/270027> [Accessed 10 March 2020].

Wang, Z. et al. (2014c) Genome-Wide Association Study for Wool Production Traits in a Chinese Merino Sheep Population. *PLOS ONE*. [online]. 9 (9), Public Library of Science, p.e107101.

Watterson, null (1976) Reversibility and the age of an allele. I. Moran's infinitely many neutral alleles model. *Theoretical Population Biology*. [online]. 10 (3), pp.239–253.

Wei, C. et al. (2016) Genome-wide analysis reveals adaptation to high altitudes in Tibetan sheep. *Scientific Reports*. [online]. 6. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4882523/> [Accessed 7 June 2017].

Weir, B.S. and Cockerham, C.C. (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*. [online]. 38 (6), [Society for the Study of Evolution, Wiley], pp.1358–1370.

Wiedemar, N. and Drögemüller, C. (2015) A 1.8-kb insertion in the 3'-UTR of RXFP2 is associated with polledness in sheep. *Animal Genetics*. [online]. 46 (4), pp.457–461.

Wilkins, A.S., Wrangham, R.W. and Fitch, W.T. (2014) The “Domestication Syndrome” in Mammals: A Unified Explanation Based on Neural Crest Cell Behavior and Genetics. *Genetics*. [online]. 197 (3), pp.795–808.

Wilson, R.T. (1991) *Small ruminant production and the small ruminant genetic resource in tropical Africa* [online]. Animal production and health paper No. 88. Available from: <http://www.fao.org/3/t0376e/T0376E00.htm> [Accessed 4 May 2021].

Woldemeskel, M., Tibbo, M. and Potgieter, L. (2002) Ovine progressive pneumonia (Maedi-Visna): An emerging respiratory disease of sheep in Ethiopia. *DTW. Deutsche tierärztliche Wochenschrift*. 109, pp.486–8.

Wright, D. (2015) Article Commentary: The Genetic Architecture of Domestication in Animals. *Bioinformatics and Biology Insights*. [online]. 9S4, SAGE Publications Ltd STM, p.BBI.S28902.

Wright, S. (1931) Evolution in Mendelian Populations. *Genetics*. 16 (2), pp.97–159.

Wu, T. et al. (2021) PAPP2 Promote the Proliferation of Dermal Papilla Cells in Hu Sheep (*Ovis aries*) by Regulating IGFBP5. *Genes*. [online]. 12 (10), Multidisciplinary Digital Publishing Institute, p.1490.

## X

Xian, B. (2005) *A rustic opinion on the natural-harmoniously traditional energy culture of the Zang (Tibetan) nationality*

Xu, S.-S. et al. (2017) Genome-wide association analysis identifies the genetic basis of fat deposition in the tails of sheep (*Ovis aries*). *Animal Genetics*. [online]. 48 (5), pp.560–569.

## Y

Yaman, Y. et al. (2019) Association of TMEM154 variants with visna/maedi virus infection in Turkish sheep. *Small Ruminant Research*. [online]. 177, pp.61–67.

Yang, G. et al. (2014) Genetic variation in the ovine uncoupling protein 1 gene: association with carcass traits in New Zealand (NZ) Romney sheep, but no association with growth traits in either NZ Romney or NZ Suffolk sheep. *Journal of Animal Breeding and Genetics*. [online]. 131 (6), pp.437–444.

Yang, J. et al. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*. [online]. 42 (7), pp.565–569.

Yang, J. et al. (2016a) Differential expression of genes in milk of dairy cattle during lactation. *Animal Genetics*. [online]. 47 (2), pp.174–180.

Yang, J. et al. (2016b) Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. *Molecular Biology and Evolution*. [online]. 33 (10), pp.2576–2592.

Yeaman, S. and Otto, S.P. (2011) Establishment and Maintenance of Adaptive Genetic Divergence Under Migration, Selection, and Drift. *Evolution*. [online]. 65 (7), pp.2123–2129.

Yuan, J. (2007) Dongwu Kaoguxue Jiemi Gudai Renlei he Dongwu de Xianghu Guanxi.

## Z

- Zeder, M.A. (ed.) (2006) *Documenting domestication: new genetic and archaeological paradigms* Berkeley, Calif, University of California Press.
- Zeder, M.A. (2008) Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences*. [online]. 105 (33), pp.11597–11604.
- Zeder, M.A. (2012) Pathways to Animal Domestication In: Gepts, P. et al. (eds.) *Biodiversity in Agriculture*. [online]. Cambridge, Cambridge University Press, 227–259. Available from: [https://www.cambridge.org/core/product/identifier/CBO9781139019514A027/type/book\\_part](https://www.cambridge.org/core/product/identifier/CBO9781139019514A027/type/book_part) [Accessed 10 March 2020].
- Zeder, M.A. (2011) The Origins of Agriculture in the Near East. *Current Anthropology*. [online]. 52 (S4), pp.S221–S235.
- Zhang, L. et al. (2013) Genome-Wide Association Studies for Growth and Meat Production Traits in Sheep. *PLOS ONE*. [online]. 8 (6), Public Library of Science, p.e66569.
- Zhang, T. et al. (2019) Genome-wide association studies revealed candidate genes for tail fat deposition and body size in the Hulun Buir sheep. *Journal of Animal Breeding and Genetics*. [online]. 136 (5), pp.362–370.
- Zhao, H. et al. (2017a) A novel 29 bp insertion/deletion (indel) variant of the *LHX3* gene and its influence on growth traits in four sheep breeds of various fecundity. *Archives Animal Breeding*. [online]. 60 (2), Copernicus GmbH, pp.79–85.
- Zhao, Y. et al. (2011) Mitochondrial DNA diversity, origin, and phylogenetic relationships of three Chinese large-fat-tailed sheep breeds. *Tropical Animal Health and Production*. [online]. 43 (7), p.1405.
- Zhao, Y.-X. et al. (2017b) Genomic Reconstruction of the History of Native Sheep Reveals the Peopling Patterns of Nomads and the Expansion of Early Pastoralism in East Asia. *Molecular Biology and Evolution*. [online]. 34 (9), pp.2380–2395.
- Zhong, T. et al. (2010) Genetic diversity of Chinese indigenous sheep breeds inferred from microsatellite markers. *Small Ruminant Research*. [online]. 90 (1), pp.88–94.
- Zhou, B.X. (1983) Animal remains excavated from Beishouling neolithic site at Baoji.
- Zhou, B.X. (1999) Animal remains excavated from Shizhaocun and Xishanping site.
- Zhu, QWG. (2010) *Study on hybridization breeding of Xinjiang wild argali* Shihezi University, Shihezi (China).
- Zilhao, J. (2001) Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *Proceedings of the National Academy of Sciences*. [online]. 98 (24), pp.14180–14185.





# Annexes

---

# Appendix S1 : Material and methods (Article section Introduction générale - 0)

---

## Modern samples

---

### Sampling

We sampled 273 wild sheep corresponding to 7 *O. ammon*, 138 *O. gmelini*, 56 *O. vignei*, 2 *O. musimon*, and 70 hybrids *gmelini* x *vignei* from 55 geographic localities representing most of their distribution area [Table S2]. Most of samples were obtained using a non-invasive method between 2000 and 2004 (Taberlet and Luikart, 1999). Fresh faeces were collected in the field, after observation of the animal from a distance to ensure the identification of the sample. Two samples were collected for each individual and preserved with two methods (silica gel and ethanol 96%). Some samples comprised skin and muscle obtained from hunter kills and carcasses. Because of possible hybridization in captivity, no individual from zoos was considered in this study. In addition, we collected 82 domestic sheep tissue samples from different countries as representative of the worldwide domestic diversity [Table S2].

### DNA extraction

Whole genomic DNA was extracted from faecal samples following the method described in (Naderi et al., 2008) beginning with 20 min in washing buffer (Tris-HCl 0.1 M, EDTA 0.1 M, NaCl 0.1 M, N-lauroyl sarcosine 1%, pH 7.5– 8.0) and using the DNAeasy extraction blood kit (Qiagen®) following the manufacture's protocol except for the incubation with protease (2 h at 56 °C with 55 µL of protease).

### DNA amplification and sequencing

#### CytB

The *Cytb* locus was amplified in two parts (741 and 765 bp) with the (CYTB\_F: CCCCACAAAACCTATCACAAA - CYTB\_IN\_R: CCTGTTTCGTGGAGGAAGAG) and (CYTB\_IN\_F: ACCTCCTTTCAGCAATTCCA - CYTB\_R: AGGGAGGTTGGTTGTTCTCC) primer pairs, respectively. PCR reactions were performed in a final volume of 25 µL containing 2 µL of DNA, 0.5 µM of each primer, 1x PCR buffer, 200 µM of each dNTP, 1.5 mM MgCl<sub>2</sub>, and one unit of AmpliTaq Gold polymerase (Applied Biosystems®). The PCR conditions were: initial denaturation, 95°C, 10 min; then for 35-40 cycles, denaturation, 95°C, 30 s; annealing temperature 55°C and 60°C respectively, 30 s; extension, 72°C, 1 min; final extension, 7 min, 72°C. PCR products were purified using the Qiaquick kit (Qiagen®) following the manufacturer's instructions. Then, purified PCR products were used as template in 20 µL sequencing reactions involving the BigDye Terminator Cycle Sequencing kit version 3.1 (Applied Biosystems®) and were analysed on an ABI Prism 3700 semi-automated DNA analyser (Applied Biosystems®) using the POP 7 polymer. SeqScape 2.5 (Applied Biosystems®) was used to reconcile chromatograms of complementary fragments and to align sequences across taxa.

## CR

PCRs were performed to amplify the Control Region fragment using the primer pairs (Ovis-F3: 5' AGA CATTATATGTATAAAGTACATTAA 3' – Ovis-R3: 5' GTGAACAAGCTCGTGATCTA 3') following the method described in (Rezaei et al., 2010) except for the annealing temperature (50°C). The sequences obtained have been checked (base composition, aberrant phylogenetic position) to make sure those sequences were not numts. 54 sequences, segregating in 16 different haplotypes were obtained (**Table S3**) to compare with ancient samples (**see Table S1**).

## Ancient samples

---

### Sampling

We sampled sheep remains from 22 archaeological sites from the Neolithic period (**Tables S4, Figure S1**): six sites within or close to the domestication area, nine sites covering the Danubian route and seven sites covering the Mediterranean route (**Table S4**). In order to ensure analysing bones from different individuals, several parameters were considered together to select samples, such as the type of bone, laterality, adult or juvenile status, or layer of excavation (**Table S5**). Since it might be difficult to distinguish between sheep and goat bones on morphological traits (Clutton-Brock et al. 1990), we used genetic data to determine dubious samples. We ultimately analysed 123 bone samples.

### Radiocarbon dates

To reinforce the contextual dating of archaeological sites inferred from human artefacts (e.g., pottery), radiocarbon dates were obtained either for 12 sheep bones for nine sites, or for contemporaneous goats and cattle of the sheep bones for the others sites (**Table S4**).

### DNA extraction

The experiments were performed in ancient DNA dedicated rooms (Palgene, National French Platform of Paleogenetics, Lyon, France) and taking usual precautions for ancient DNA treatment (Hughes et al. 2006). After removing the exterior part of the bone with a scalpel, we reduced 100 to 500 mg of material to powder with a hammer. A buffer composed of 4.7 mL of EDTA 0.5 M, 250 µL of N-Lauryl-Sarcosine 10% and 50 µL of proteinase K (25 mg/ml), was added to the powder and the mixture was incubated for 18-24h under constant agitation at 37°C-55°C. We performed DNA extractions following either a phenol-chloroform protocol (Loreille et al. 2001) or a silica purification (Orlando et al. 2009). Depending on the samples, one or both of the protocols (when samples were independently extracted a second time) were applied. A supplementary purification step was performed on the phenol/chloroform extracts showing a high level of PCR inhibition using the Qiaquick purification kit (Qiagen, kit used following manufacturer's recommendations). We performed a mock control for each DNA extraction session in order to follow possible contaminations in the products used.

Moreover, when extracting more than two sheep samples together, a sample from a different species was co-extracted at the same time in order to control for potential cross-contamination.

## DNA amplification and sequencing

We targeted a short but highly variable fragment of the mitochondrial Control Region in the HVI region (120 bp without primers). A pair of primers was first designed (Ovis-F2: 5' CATTATATG TATAAAGTACATTAA 3' – Ovis-R2: 5' GCTCGTGATCTARTGGAC 3'), but was modified (Ovis-F3: 5' AGACATTATATGTATAAAGTACATTAA 3' – Ovis-R3: 5' GTGAACAAGCTCGTGATCTA 3') to better take into account the variability observed between divergent sheep haplogroups in the literature. Finally two supplementary primer pairs were designed for the Control Region (covering a part of the HVI fragment, Ovis-F4: 5' GCRTATARGCATGTACATT3' – Ovis-R4: 5' GGA TAYGCATGTTGACT 3') to amplify shorter fragments for difficult samples (respectively 69 bp and 66 bp without primers).

Mixes for PCR amplifications were prepared in specific cleaned rooms where no DNA was handled. Each mix (25µl) contained: 2.5 µL AmpliTaq Gold™ buffer 10x (Applied Biosystems®), 2 µL MgCl<sub>2</sub> (25 mM), 1 µL BSA (20 mg/ml Roche®), 1.25 U of AmpliTaq Gold™ DNA Polymerase (Applied Biosystems®), 0.25 µL dNTPs (100 mM Sigma®), 0.5 µL of each primer (25 µM Eurogentech®). Different quantity of DNA extracts were tested and PCR amplifications were usually performed with the following program: 94°C 10 min, 50 to 60 cycles (94°C 30 sec, 45-55°C 30 sec, 72°C 45 sec), 72°C 10 min. For each session of amplification, the mock extract, the PCR reagents and the air (aerosol contamination) were monitored and when two sheep samples were amplified in a same session, a DNA extract of another species was also amplified to test for cross-contaminations.

Several independent amplifications were performed per sample and positive amplifications were cloned using Invitrogen™ TOPO™ TA Cloning™ for sequencing. Bacterial colonies were picked-up and plasmid insert amplified by PCR (Hughes et al. 2006).

PCR products were then Sanger sequenced using universal primers. Each sequence of each clone was manually checked and verified. Then a consensus of the clones was done for each amplicon and a final consensus of the different amplicons was performed to generate the final sequence of the sample. Only a small number of additional amplicons for 6 samples were sequenced by 454 to confirm the final consensus already obtained by Sanger (**see Table S5**). In this case, and using a dedicated Galaxy pipeline, the low quality reads were first removed, then reads were sorted by barcode and primers were trimmed on both ends. The cleaned reads were then collapse allowing visualizing possible degradation signatures and confirming that the Sanger consensus was the most frequent among all fragments. As the protocol used to construct the libraries was not orientated, sequences were generated in both directions, forward and reverse, avoiding possible sequencing bias.

The sequences generated in this study were deposited in GenBank under accession numbers MW684398 - MW684451, MW684452 - MW684508, and MW677233 - MW677445 (**See Table S1 for a summary of the data used**).

## Additional data

---

We obtained 421 Cytb sequences (**Table S2**) and 1283 CR sequences (**Table S3**) of present-day sheep from GenBank after discarding sequences of unknown geographic origin, showing gaps or highly divergent (i.e., likely numts).

## Genetic analyses

---

### Alignment

The 634 Cytb modern sequences were aligned using MEGA version 10.0.5 (Kumar et al., 2018) by MUSCLE (Edgar, 2004). The 1337 CR modern (wild and domestic) and ancient sequences were aligned with MAFFT version 7 (Kato and Standley, 2013) with default parameters.

### Phylogenetic analysis

Maximum likelihood phylogenies were performed using IQtree (Nguyen et al., 2015). The most appropriate model of sequence evolution (HKY+F+R2) was determined using ModelFinder (Kalyaanamoorthy et al., 2017). 1000 bootstrap replicates were done using ultrafast bootstrap approximation (`-bb` option) (Hoang et al., 2018) and a NNI branch swapping algorithm (`-bnni` option). Branch supports were also assessed by 1000 bootstrap replicates for SH-like approximate likelihood ratio test (`-alrt` option) (Guindon et al., 2010). Neighbor-joining (Saitou and Nei, 1987) trees were constructed with the R package ape (Paradis, Claude and Strimmer, 2004).

### The short CR fragment can discriminate haplogroups

To ensure that the assignment to haplogroups was reliable on the short CR fragment sequenced for ancient sheep, we used CR sequences obtained for modern wild and domestic sheep (**Table S3**). We retrieved 1283 sequences of modern sheep from GenBank (*O. aries*, **Table S3**) and assigned them to one of the five haplogroups (A, B, C, D, E) either on the basis of the information available in the database or related publication, or on the basis of similarities between sequences. Haplotypes carried by wild individuals and absent in domestic sheep it was classified in a group of haplotypes called “exclusively wild”.

Because mismatch pairwise distribution is a way to differentiate haplogroups (Naderi et al. 2007), such analysis was performed with DnaSP v5.10 (Librado & Rozas 2009) on all haplotypes, but also on each haplogroup specific dataset, to control *a posteriori* for possible errors in the assignment step. In our global dataset, a clear bimodal distribution was observed with a lowest value between both modes at nine mismatches considered thus as the limit

between intra- and inter- haplogroup diversity for the CR fragment of interest. Finally, when the geographic origin was known, sequences were attributed to one of the five large geographical areas defined as Northern Europe, Southern Europe, Northeastern Europe, Eastern Asia and Western Asia (**Table S3**).

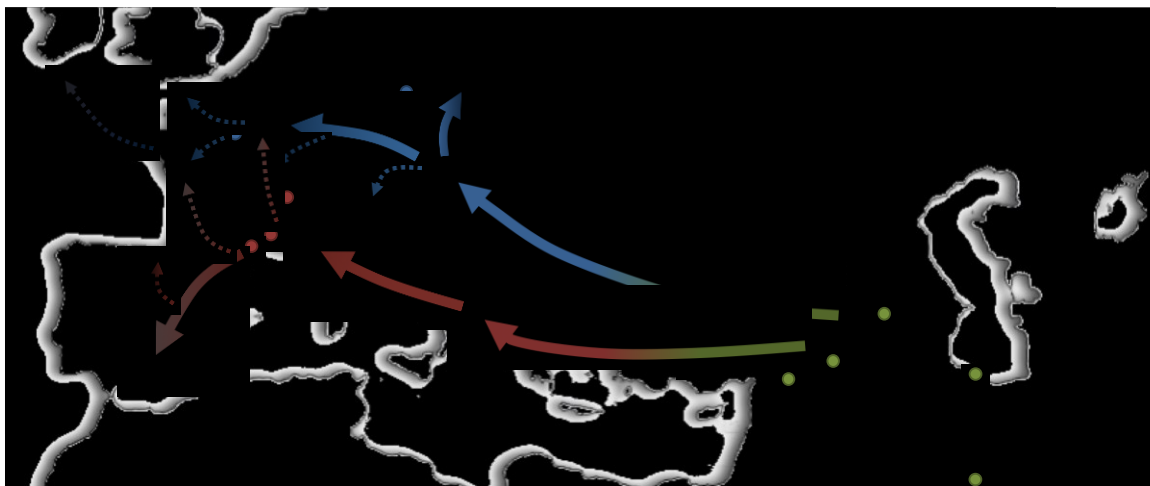
A genetic network analysis was performed to assess the relationships between haplotypes from different ages (ancient or modern) and statuses (domestic or wild) with the R package `pegas` (Paradis, 2010). The network showed three major haplotypes within the B haplogroup (Figure S3). The Neolithic haplotypes were found in all three haplogroups independent of their geographic origin (i.e., Danubian or Mediterranean colonization routes). The majority haplogroups for modern and ancient sequences were the same (i.e. A and B). 11 (out of 55) wild haplotypes were from either the B or E haplogroups.

### Genetic parameters

Nucleotide diversities were calculated for wild sheep, modern and Neolithic domestic sheep from the CR dataset. For domestic individuals, they were calculated for each haplogroup and for modern ones within each of five global geographic regions: Eastern Asia, Northern Europe, Northeastern Europe, Southern Europe and Western Asia. The latter contains the domestication centre. For haplogroup B (the most represented), pairwise genetic distances and  $F_{st}$  were calculated between CR sequences from archaeological sites (for ancient haplotypes) or from global geographic regions (for modern haplotypes).

### Supplementary tables & figures

---



**Figure S1.** Location of the 22 archaeological sites identified by the number given in Table S3  
Green indicates the domestication centre, blue is for the Danubian route and red for the Mediterranean route.

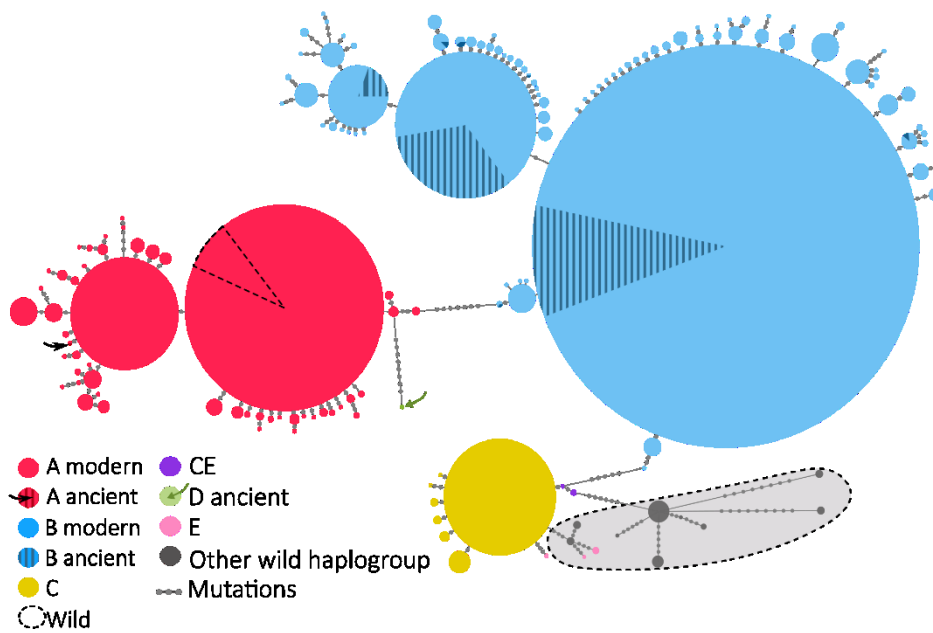


Figure S2: Network obtained for the short CR fragment for present-day and Neolithic domestic sheep and wild mouflon.

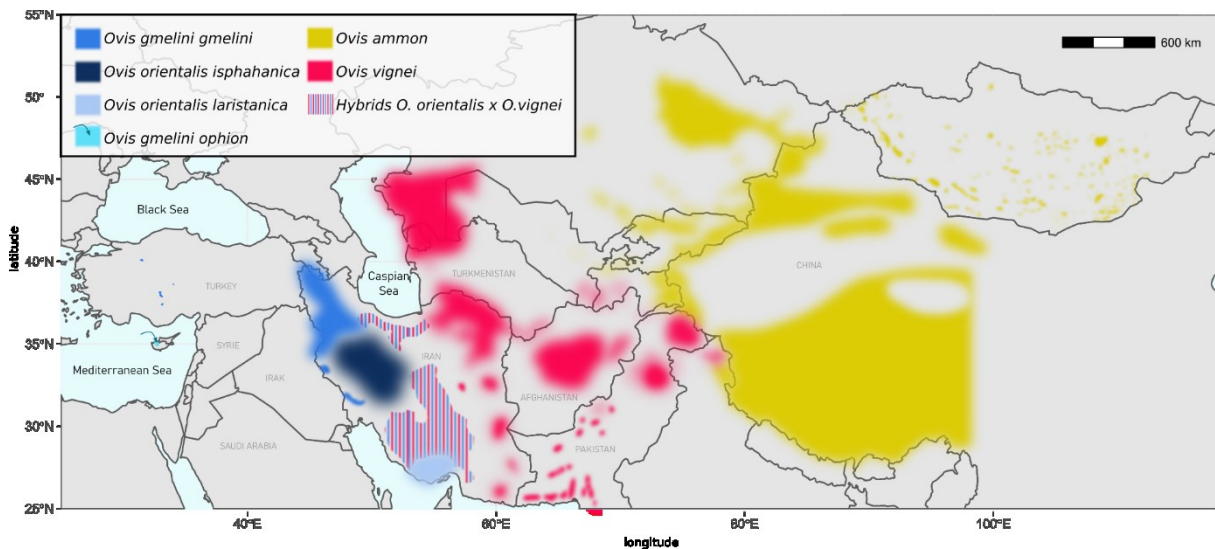


Figure S3: Distribution of wild *Ovis* species.

We obtained species distributions from the IUCN database (see references below). For *Ovis gmelini* subspecies, we combined information from sites described in the detailed report and general distribution of the species to draw approximate ranges of continental subspecies.

***Ovis ammon***: Reading, R., Michel, S. & Amgalanbaatar, S. 2020. *Ovis ammon*. The IUCN Red List of Threatened Species 2020: e.T15733A22146397. <https://dx.doi.org/10.2305/IUCN.UK.2020-2.RLTS.T15733A22146397.en>.

***Ovis gmelini***: Michel, S. & Ghoddousi, A. 2020. *Ovis gmelini*. The IUCN Red List of Threatened Species 2020: e.T54940218A22147055. <https://dx.doi.org/10.2305/IUCN.UK.2020-2.RLTS.T54940218A22147055.en>.

***Ovis vignei***: Michel, S. & Ghoddousi, A. 2020. *Ovis vignei*. The IUCN Red List of Threatened Species 2020: e.T54940655A54940728. <https://dx.doi.org/10.2305/IUCN.UK.2020-2.RLTS.T54940655A54940728.en>.



Table S1 : Summary of data used in the study

dataset	analyses	Number of new sequences	Accession number of new sequences	Number of sequences from GenBank	Corresponding supplementary table
cyt-b	comparison of modern wild and domestic sheep	213	MW677233 - MW677445	421	Table S2
CR	comparison of ancient and modern domestic sheep	ancient : 57	MW684452 - MW684508	0	Table S3
		modern : 54	MW684398 - MW684451	1283	Table S4 & S5

Table S2: Origin of sheep and mouflon sequences used for analysing cytb diversity

.xlsx file

Table S3: Origin of sheep and mouflon sequences used for analysing CR diversity

.xlsx file

Table S4: Detailed information on archaeological sites studied

.xlsx file

Table S5: Origin of samples used for analysing ancient CR sequences

.xlsx file

Table S6: Geographic distribution of CR diversity in domestic sheep

Area	Haplogroup	Number of sequences	Nucleotidic diversity $\pi$
Western Asia	A	11	0.022± 2.118875e-04
	B	21	0.032± 3.569125e-04
	C	3	0.011± 1.303155e-04
	D	0	-
	E	2	-
	All	37	0.075±0.0015
Eastern Asia	A	342	0.017± 1.143038e-04
	B	197	0.024± 1.990014e-04
	C	85	0.010± 5.330696e-05
	D	2	-
	E	2	-
	All	628	0.076 ±0.0015
North-eastern Europe	A	66	0.026± 2.263941e-04
	B	217	0.020± 1.490183e-04
	C	18	0.007± 3.454335e-05
	D	1	-
	E	1	-
	All	303	0.056±0.0009
North Europe	A	6	0.019± 2.016461e-0
	B	44	0.018±1.307756e-04
	C	0	-
	D	0	-
	E	0	-
	All	50	0.032 ± 0.0003
South Europe	A	11	0.012± 7.911994e-05

	B	224	0.014±7.941595e-05
	C	0	-
	D	0	-
	E	-	-
	All	235	0.023±0.0002
<b>Neolithic</b>	A	1	-
	B	55	0.007± 3.383544e-05
	C	-	-
	D	1	-
	E	-	-
	All	57	0.013 ± 7.843952e-05

# Figures et tableaux supplémentaires

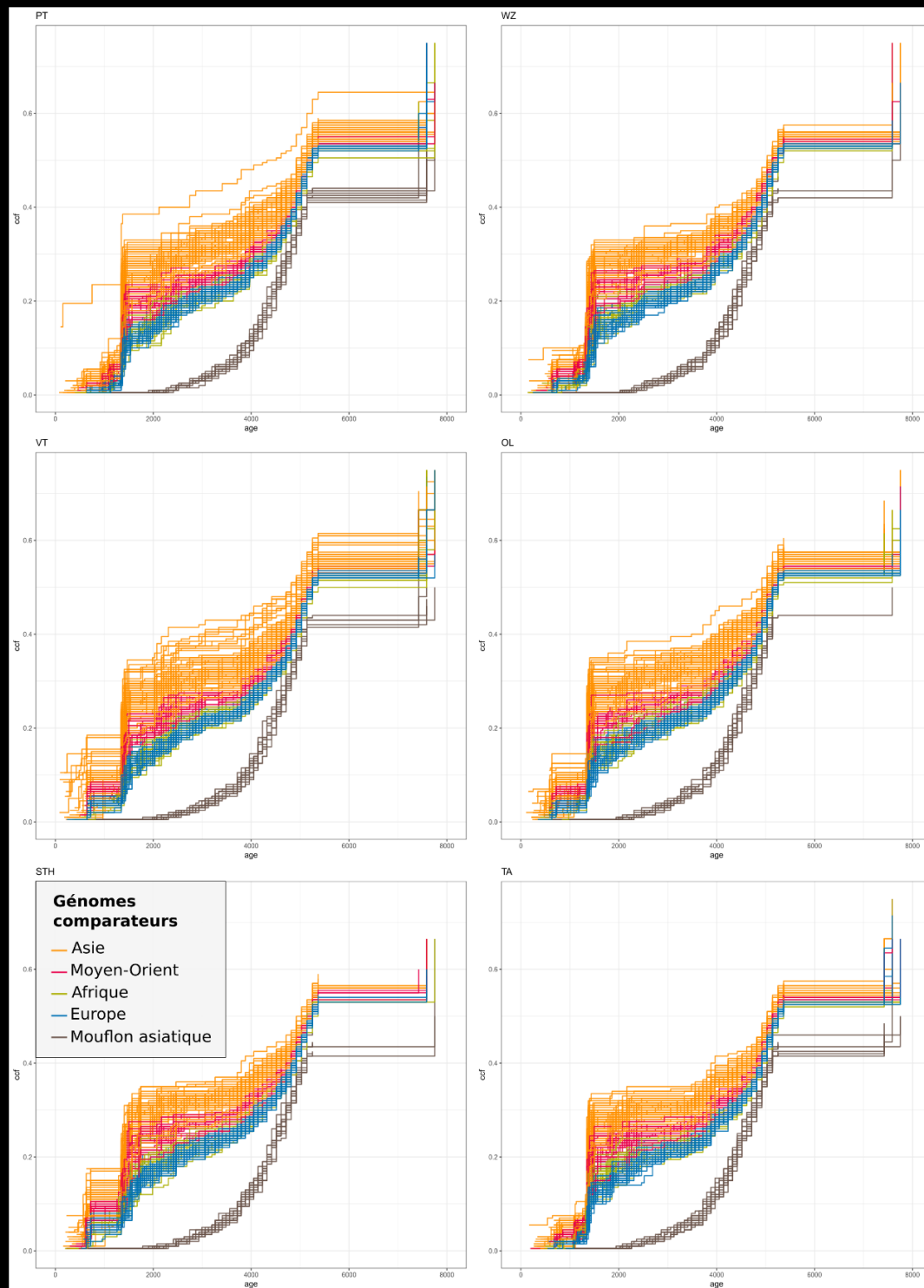


Figure S1 : Evolution de la CCF (Cumulative Coalescent Function), représentant la fraction de génome partagée au cours du temps entre des génomes cibles (Tableau 15) et les génomes comparateurs issus du monde entier. L'origine des génomes cibles est indiquée dans le coin supérieur gauche de chaque figure. Les correspondances entre les codes inscrits sur les figures et le groupe sont présentées dans le **Tableau 11**

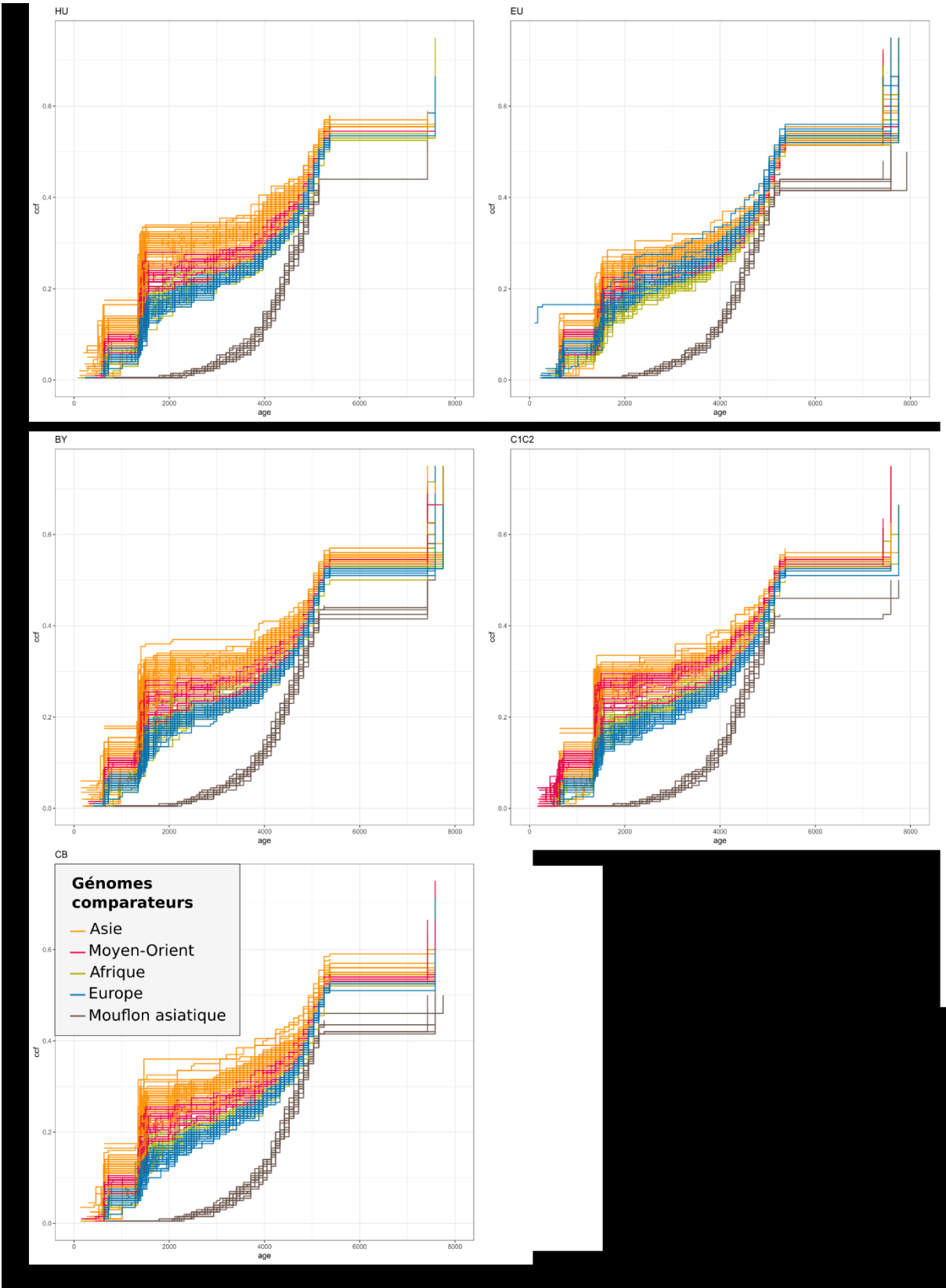
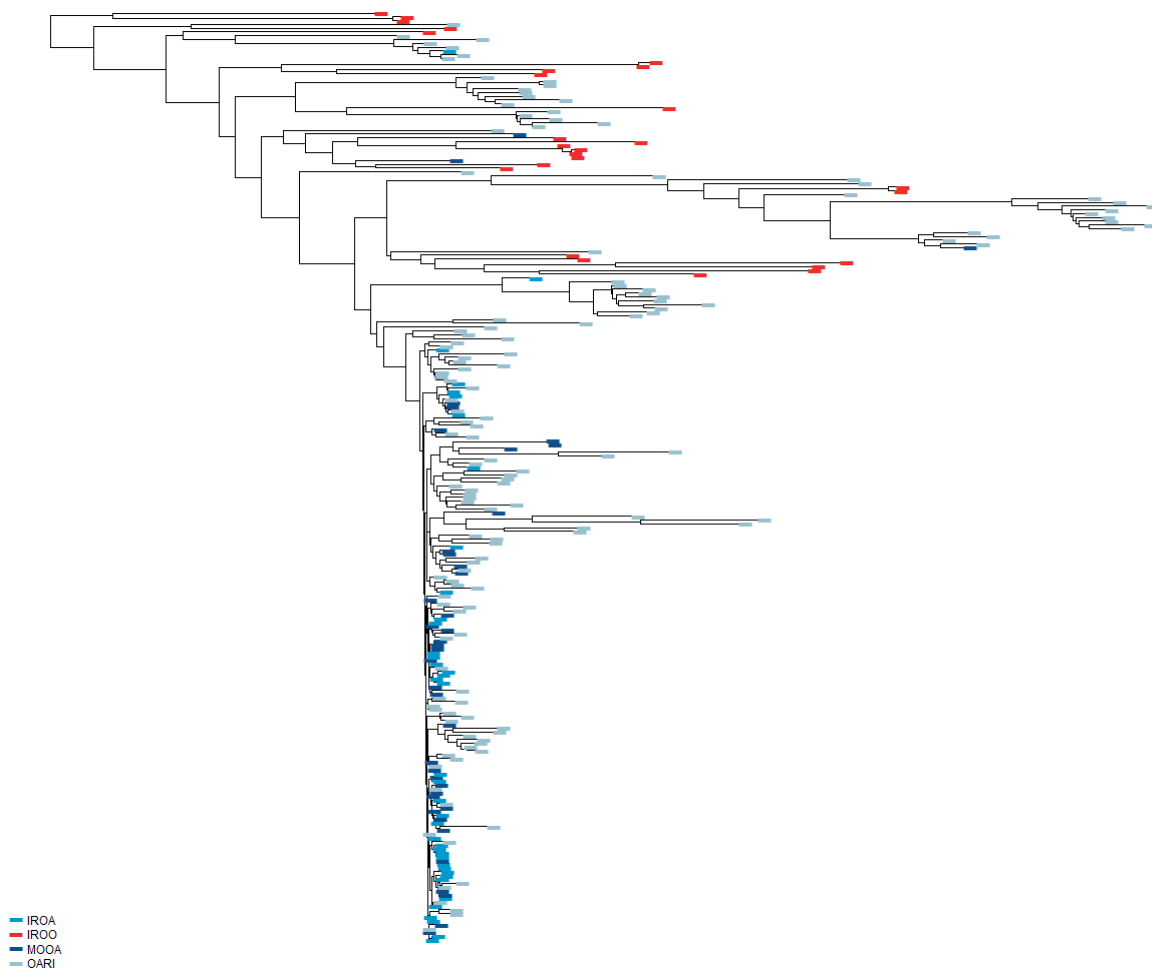


Figure S1 (suite)

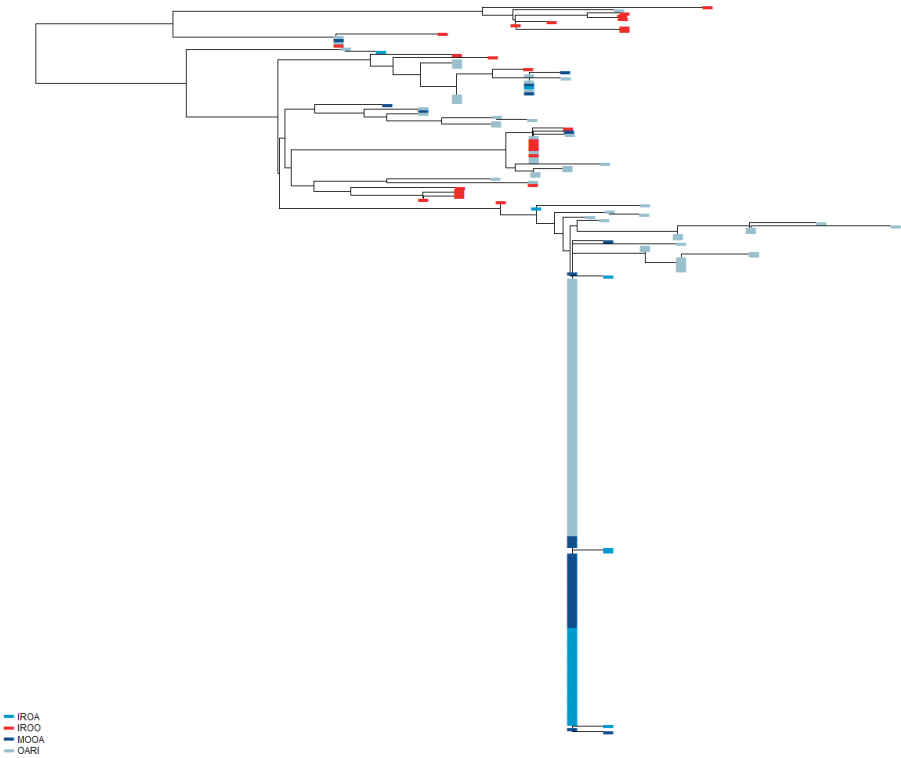


**Figure S2 : Arbre de distance représentant les relations entre les haplotypes du jeu de données Nextgen pour la région génomique correspondant au signal de sélection cosmopolite R1c**

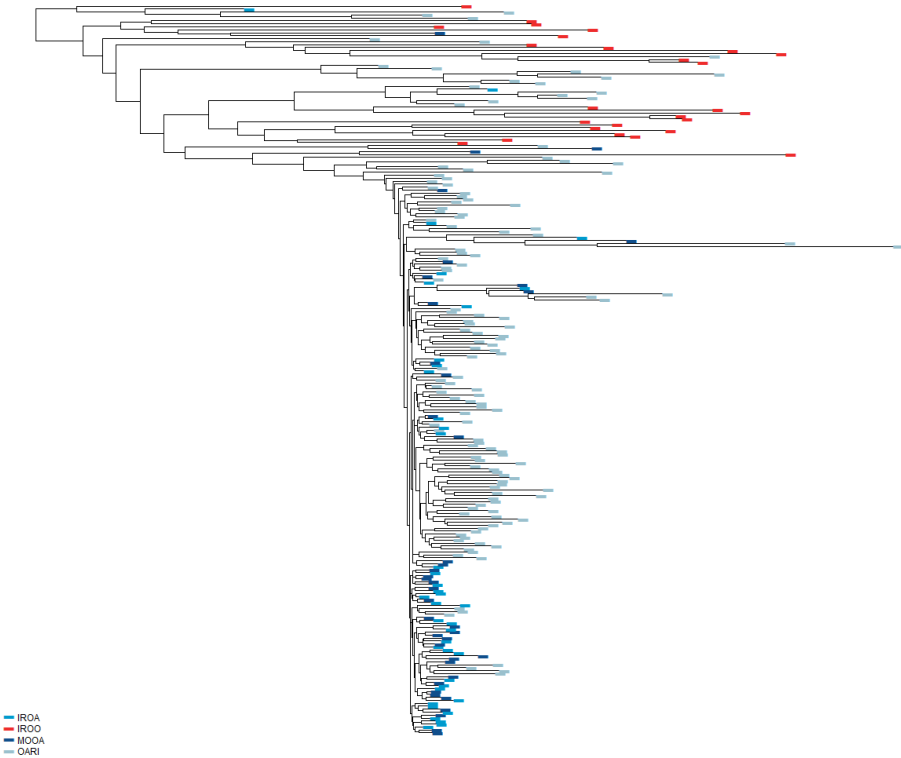
- IROA = races traditionnelles iraniennes
- IROO = mouflons asiatiques
- MOOA = races traditionnelles marocaines
- OARI = races du panel mondial

Les figures suivantes ont été construites de la même manière sur les autres régions du groupe de signaux cosmopolite A1. Les identifiants de régions sont inscrits en haut à gauche de chaque arbre.

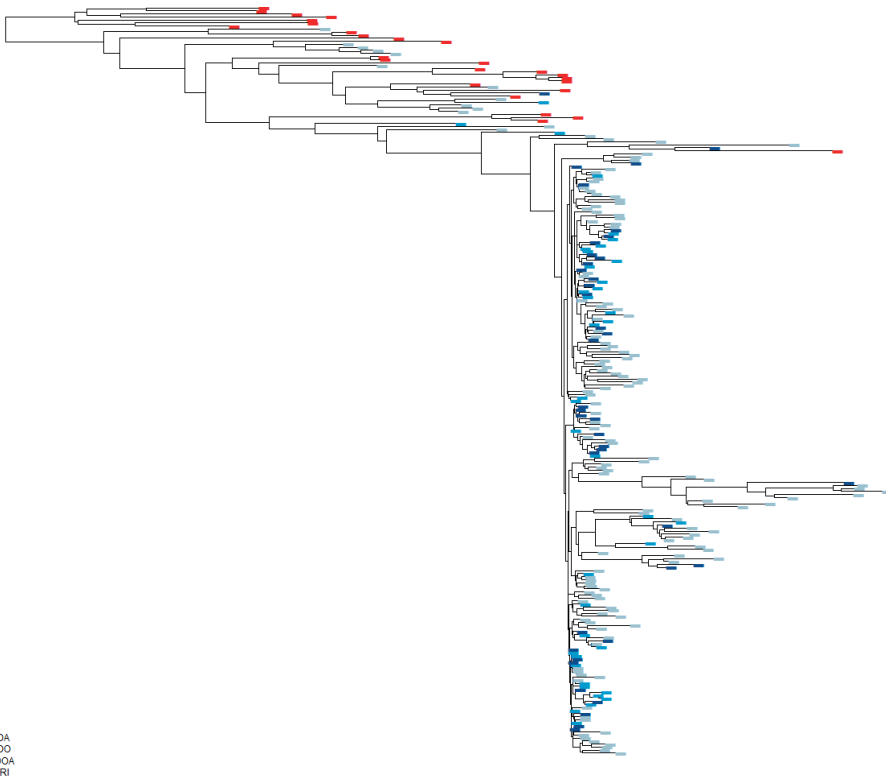
R2



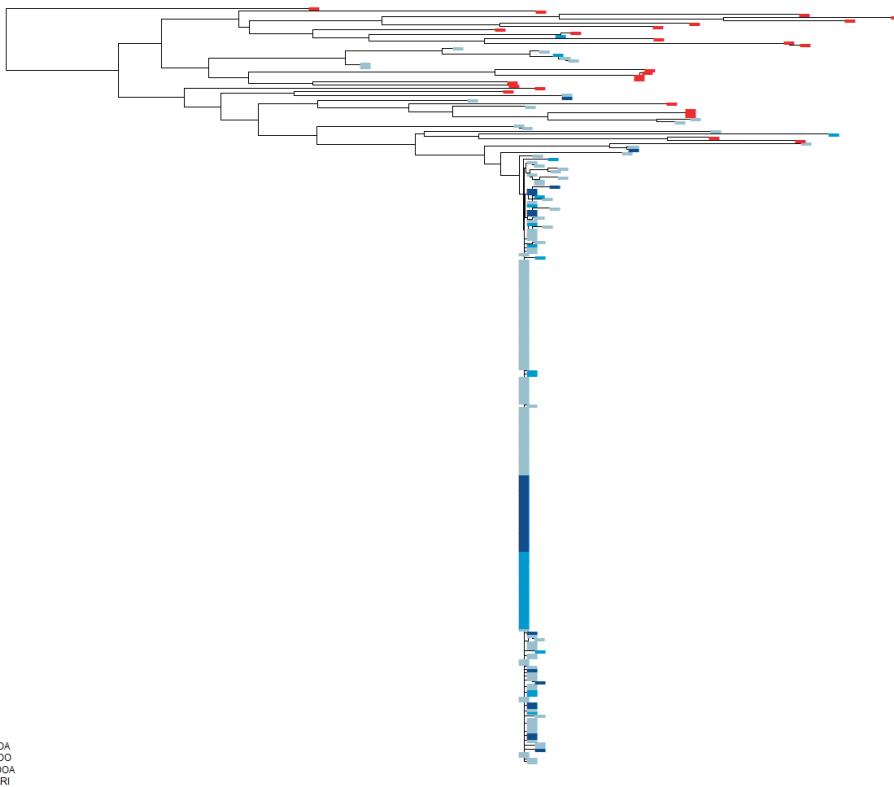
R3



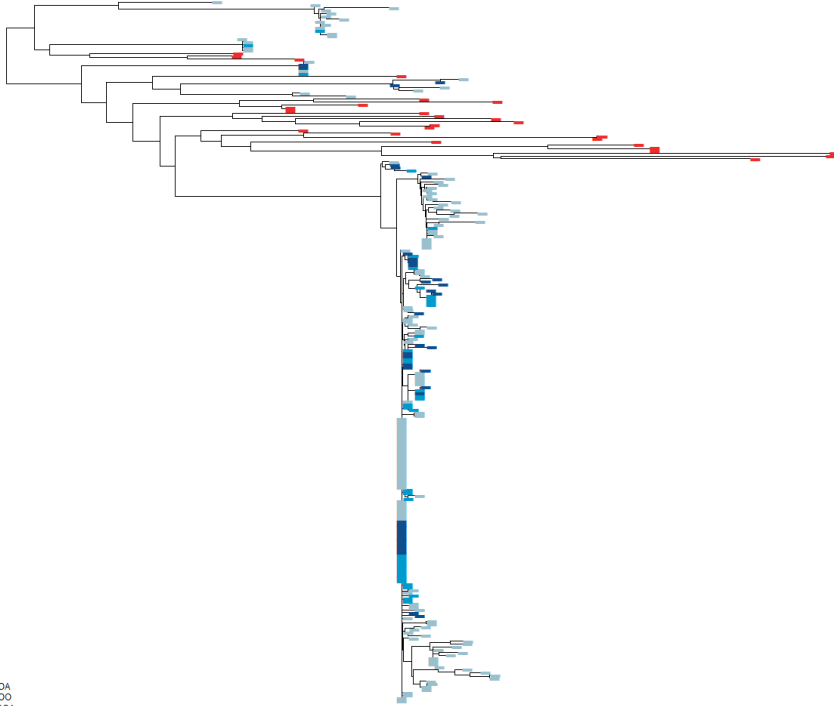
R4



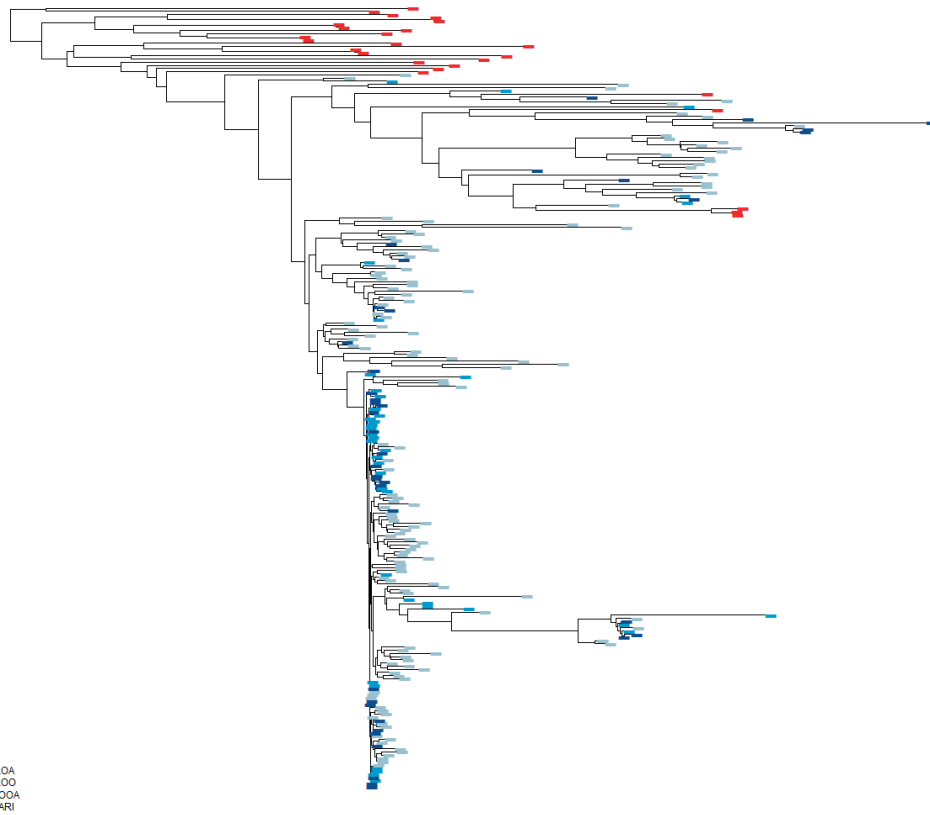
R5



R6

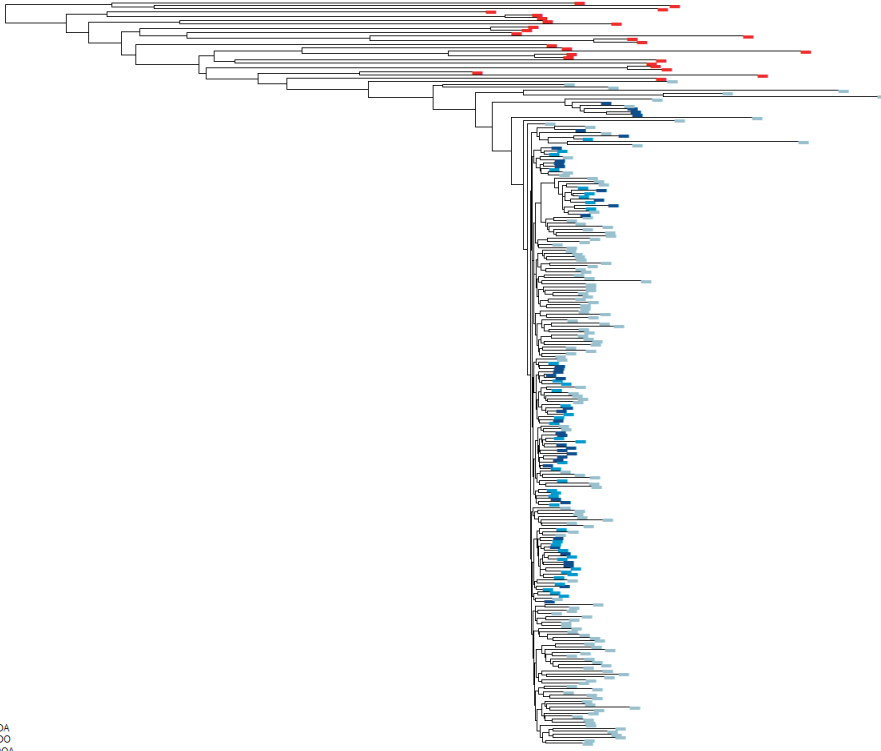


R7c

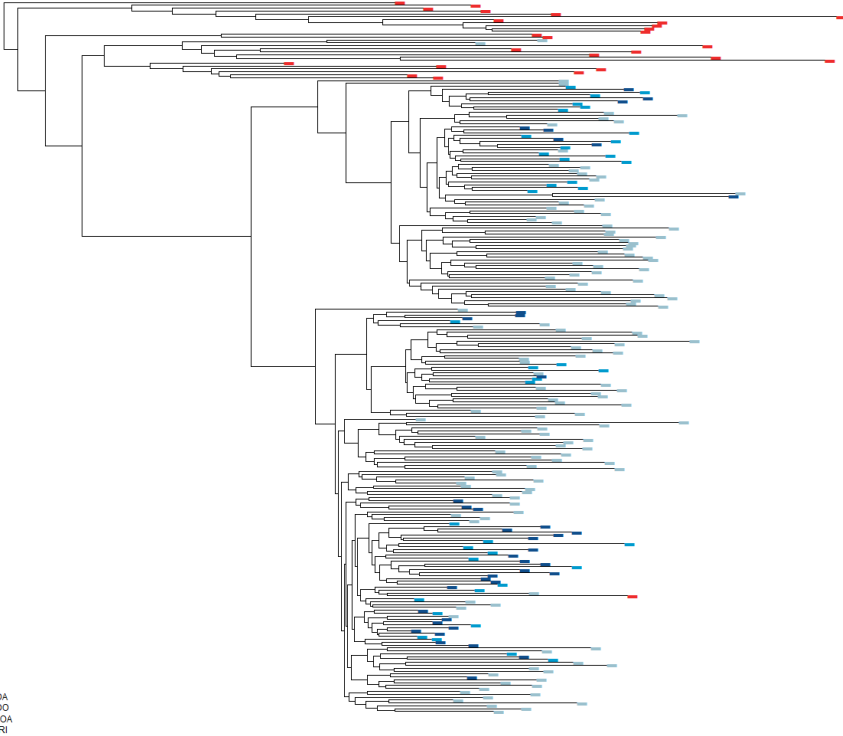


R8c

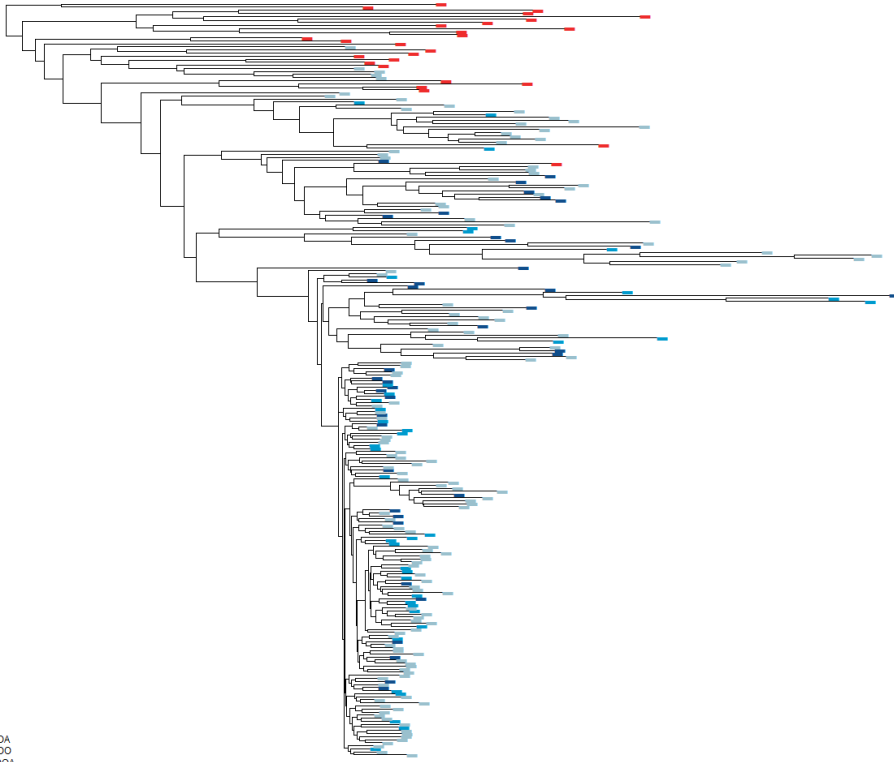




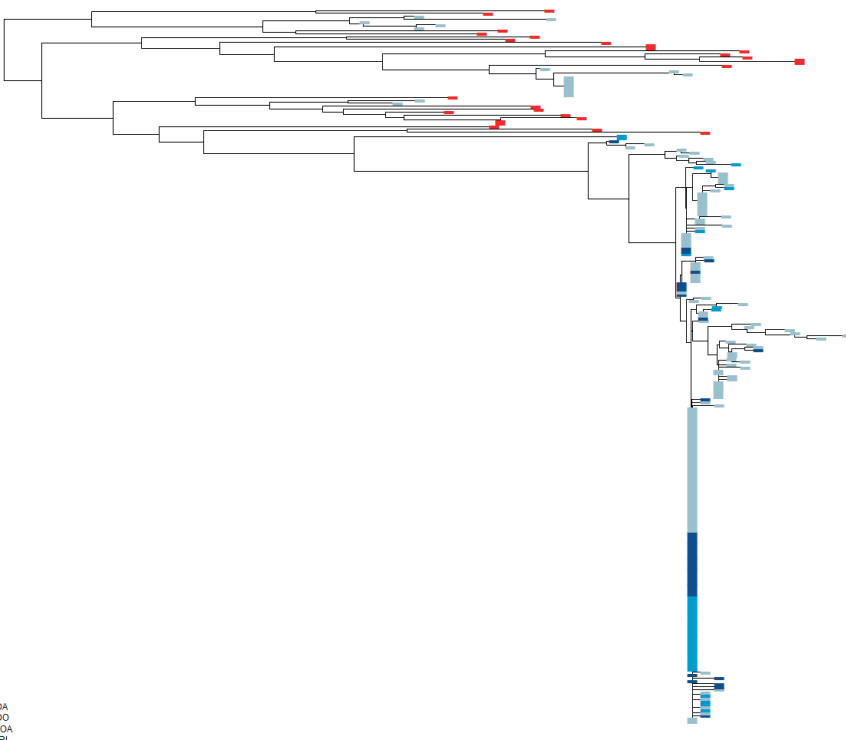
R9c



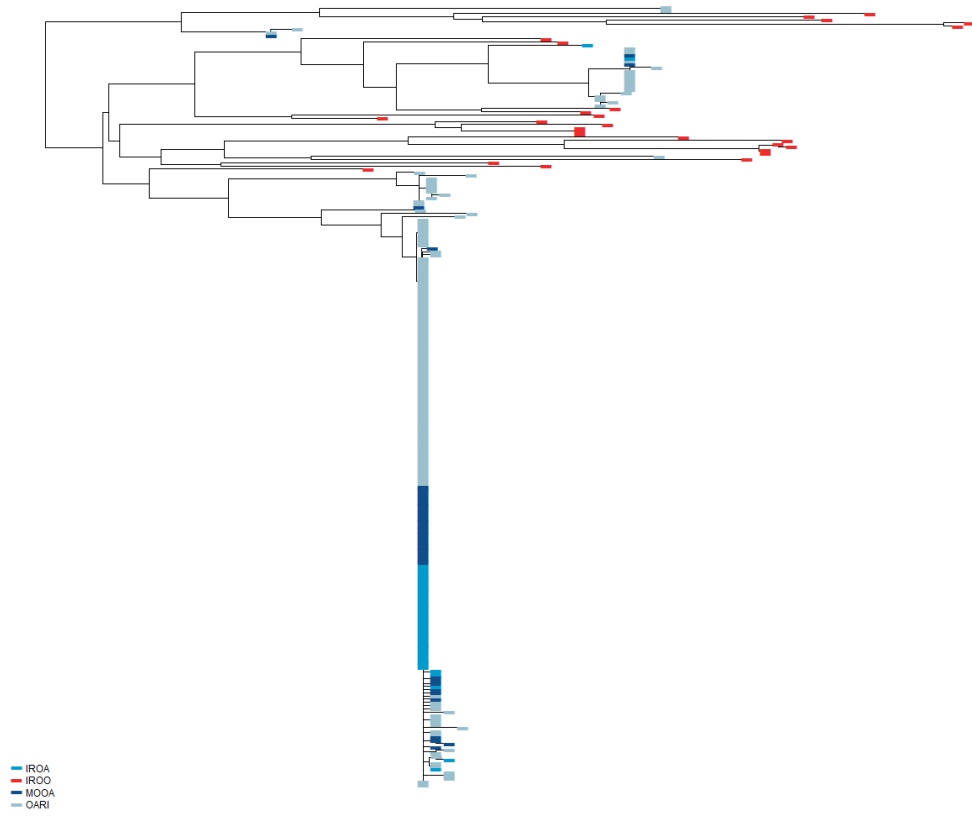
R10c



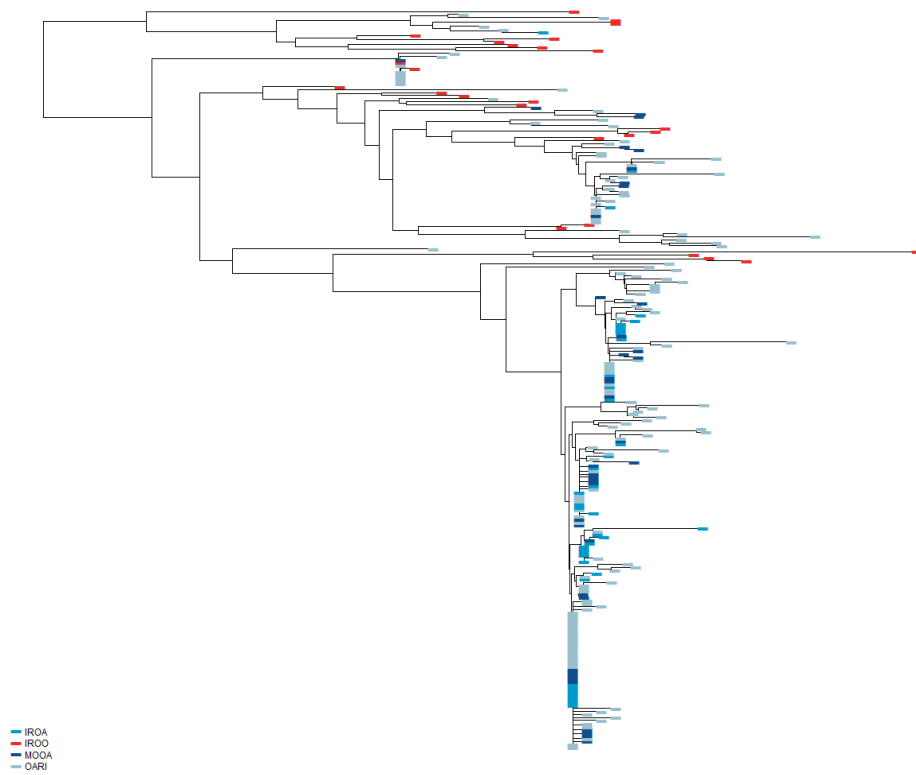
**R11**



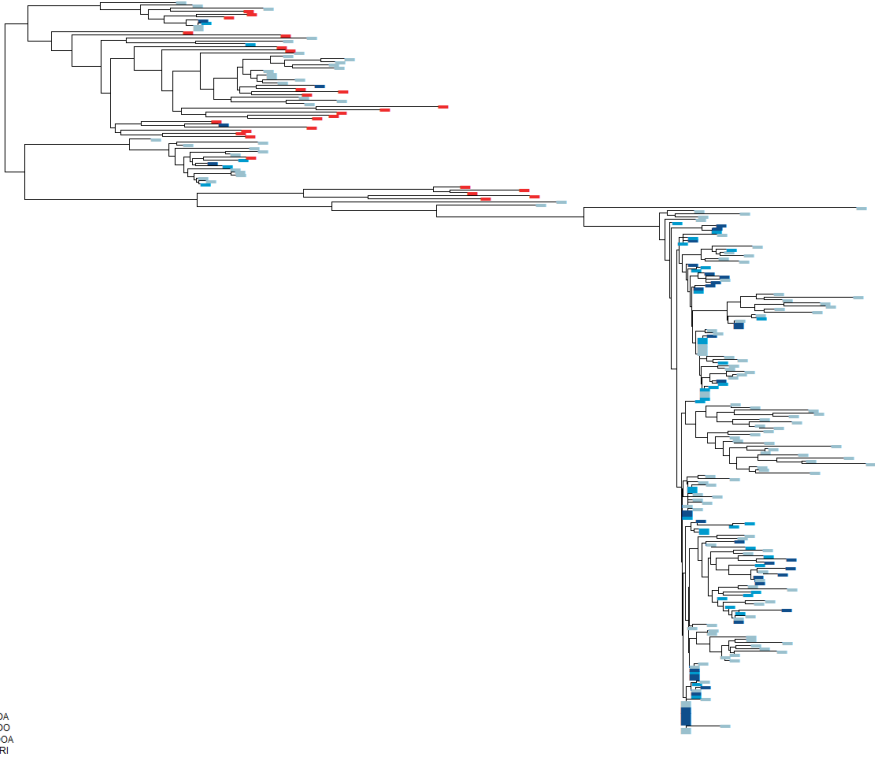
**R12**



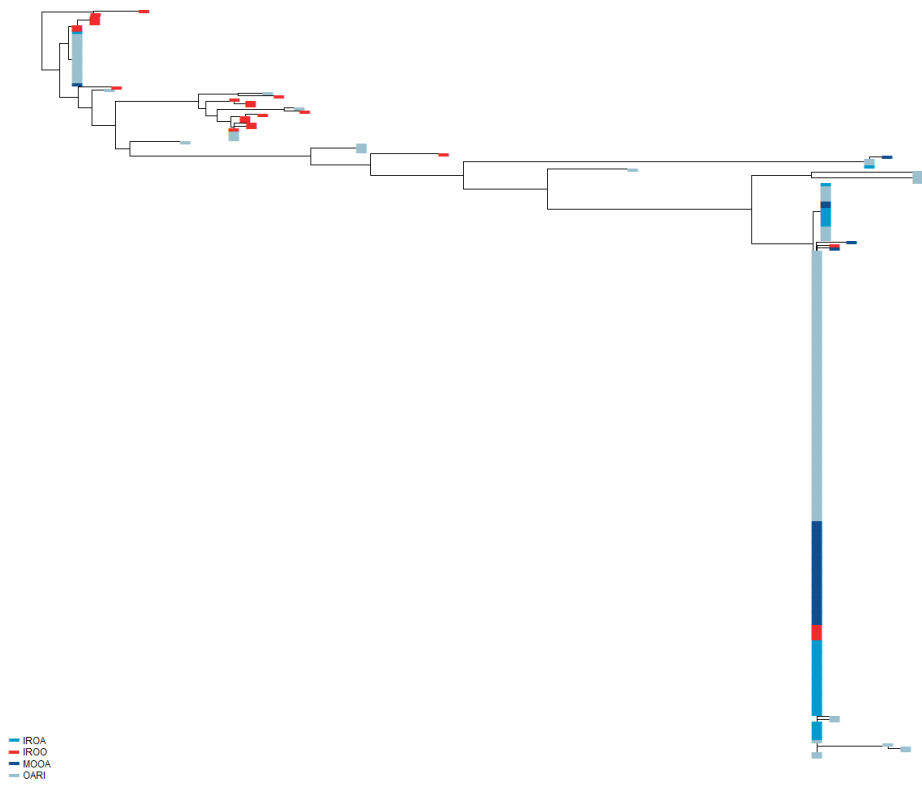
R13



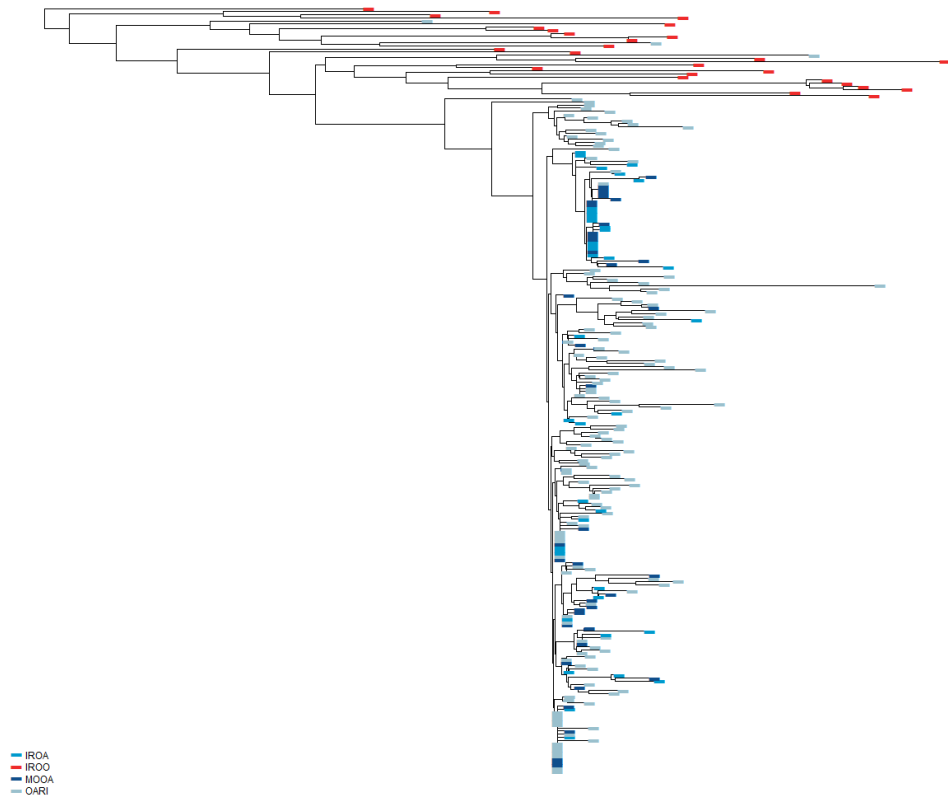
R14



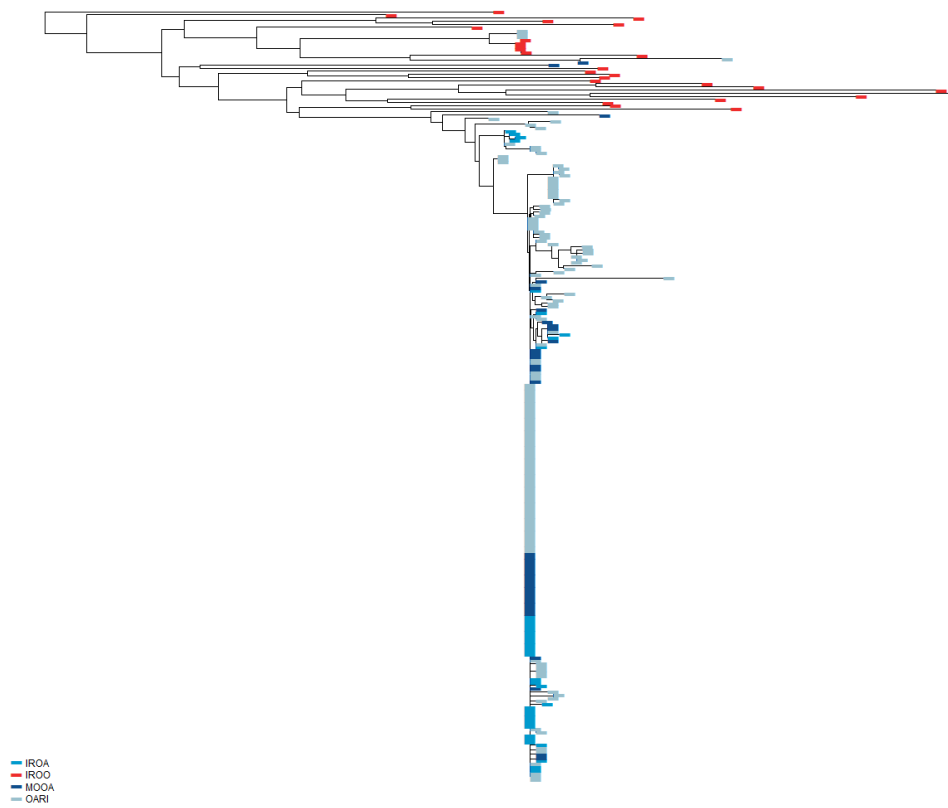
R15



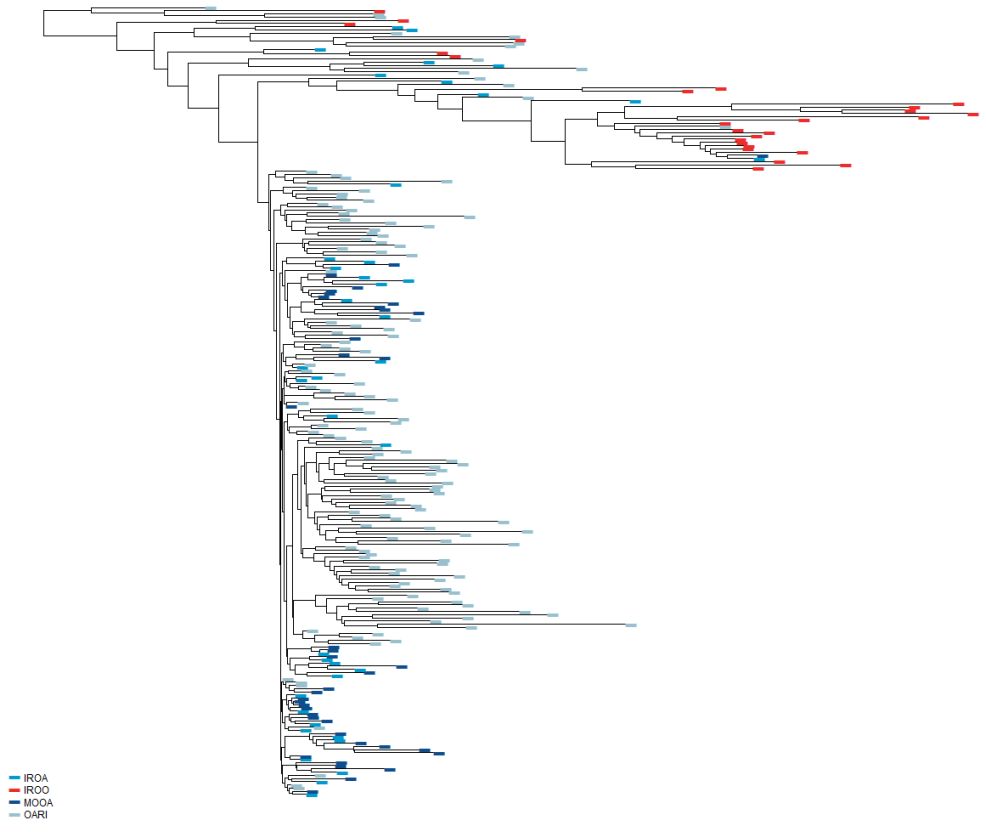
R16



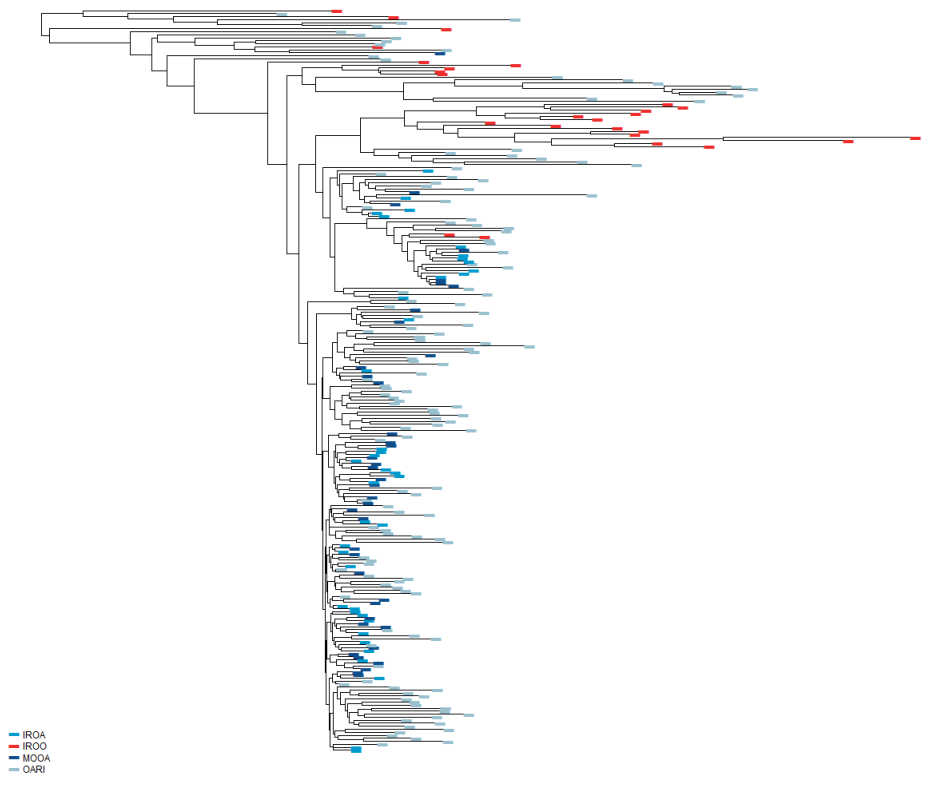
R17



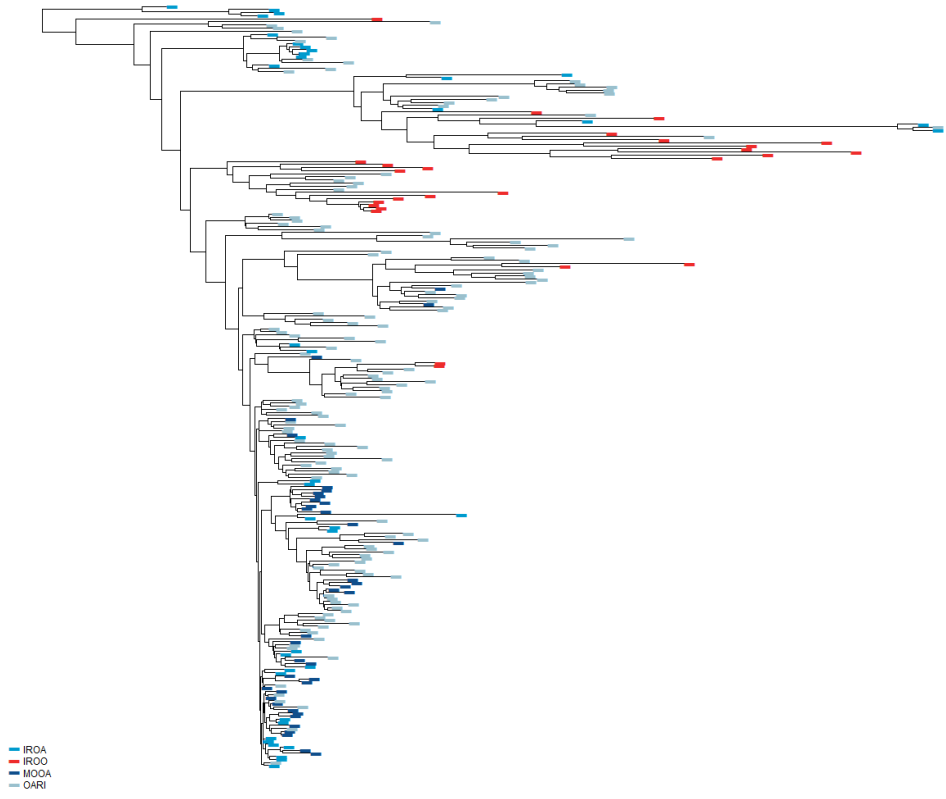
R18



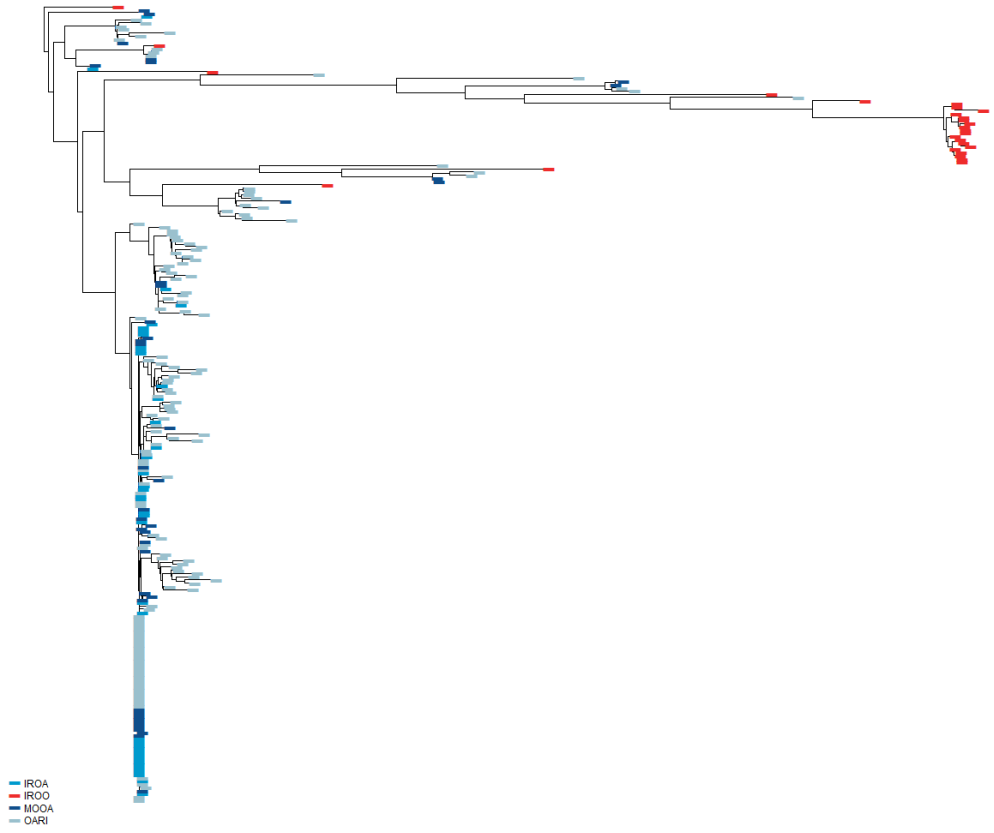
R19c



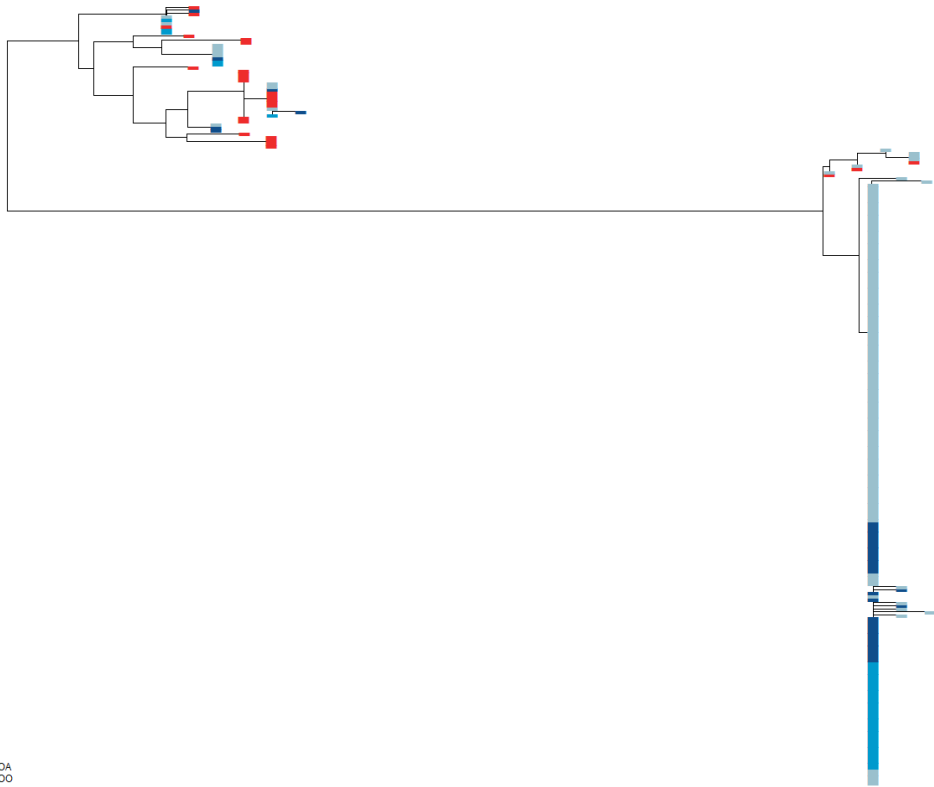
R20c



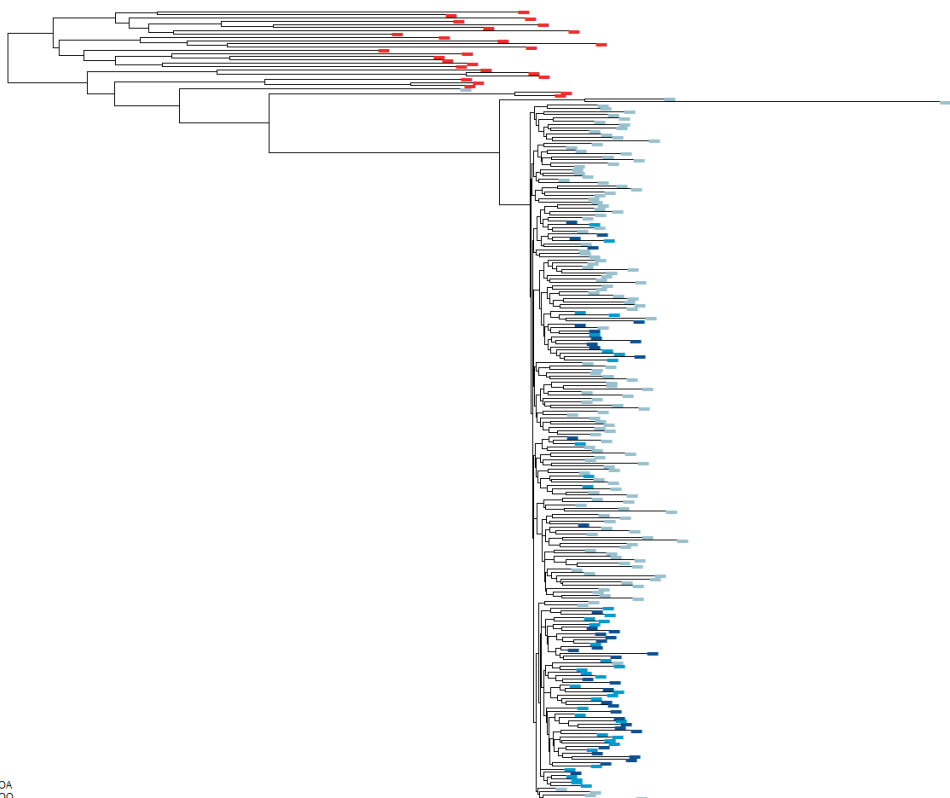
R21



R22

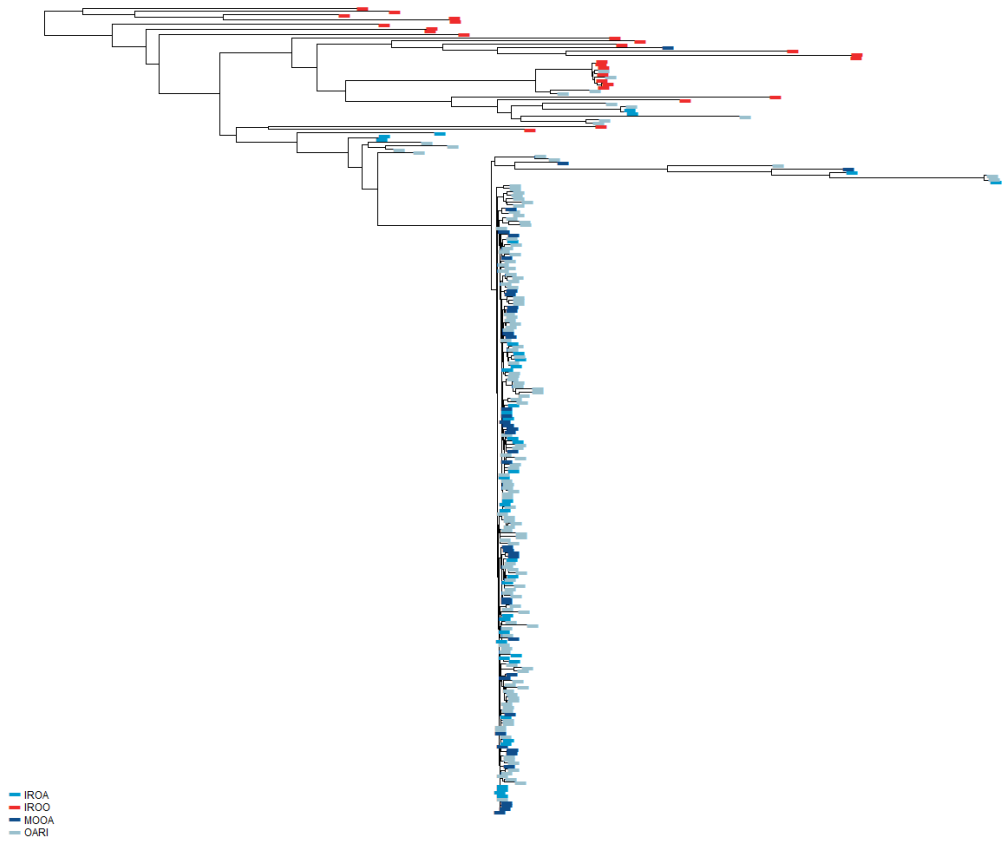


R23

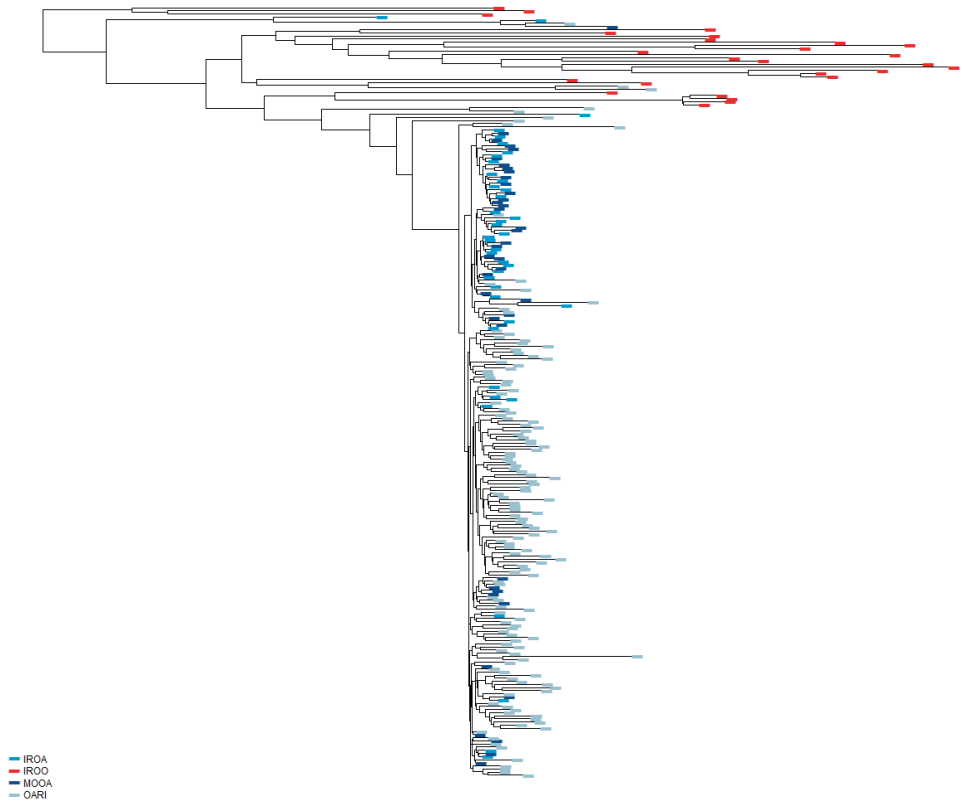


R24





R25



## Résumé

La domestication représente un changement majeur de l'histoire de l'humanité, menant à l'émergence de l'agriculture durant le Néolithique. En offrant une série d'expériences évolutives indépendantes où les plantes et les animaux ont été sélectionnés pour des traits spécifiques, ce processus intéresse depuis longtemps les biologistes évolutionnistes. Le mouton (*Ovis aries*) est l'une des premières espèces à avoir été domestiquée, ce qui en fait aussi un modèle de choix pour étudier les premières tentatives de l'homme dans le développement de l'élevage. L'étude de ce processus implique plusieurs disciplines complémentaires comme l'archéologie, la paléo-génétique ou la génétique moderne. Les deux premières procurent des témoins directs des processus passés, mais reposent sur des données fragmentées dans le temps et l'espace. L'ADN moderne est au contraire une source de données abondantes dont l'échantillonnage peut être bien plus exhaustif au niveau géographique, mais n'est pas exempt d'incertitudes liées à la nécessité d'inférer les processus passés ayant conduit à la diversité génétique actuelle. Ces inférences sont possibles car les génomes actuels gardent les traces des processus qui les ont façonnés au fil des millénaires. Les variations génétiques le long des génomes et à travers les régions géographiques ont été largement étudiées chez le mouton, permettant d'inférer les processus démographiques (migrations, variations de tailles de populations au cours du temps) et adaptatifs (régions génomiques répondant à la sélection naturelle). Cependant, les informations temporelles contenues dans la variabilité des génomes sont encore largement sous-exploitées. Grâce à 376 génomes complets représentatifs de la diversité actuelle mondiale des moutons et de leurs proches parents sauvages, les mouflons asiatiques, nous avons précisé l'histoire de leur domestication. Les âges d'apparition de plus de 30 millions de variants génomiques ont été estimés et constituent un Atlas disponible pour la communauté scientifique via un site web. La distribution de ces âges montre que l'histoire des moutons est intimement liée à la nôtre, et que leur ADN porte la trace de grandes crises de l'Humanité telles que l'effondrement de l'âge du Bronze ou la peste de Justinien. Les variants génomiques datant de ces périodes indiquent des chutes du partage d'haplotypes entre toutes les populations domestiques. La datation des traces de sélection présentes dans les génomes souligne l'importance des remplacements de populations liés à différentes vagues de diffusion, et de l'intensification récente des sélections de caractères agronomiques. Ces processus ont effacé une partie des empreintes des sélections anciennes. Cependant, certaines races épargnées par ces phénomènes témoignent de l'utilisation précoce probable du lait dès le début de la domestication et de la sélection de caractéristiques potentiellement impliqués dans le syndrome de domestication (couleur, baisse de l'agressivité, immunité). Ce syndrome désigne un ensemble de traits morphologiques, physiologiques et comportementaux partagés par de multiples espèces domestiques sélectionnés probablement involontairement via la sélection des animaux les plus dociles. Cette datation de variants génomiques à large échelle est la première du genre mis en œuvre sur une espèce autre que l'humain. Son utilisation ouvre de nouvelles perspectives pour préciser la chronologie des événements constituant les scénarios évolutifs (par exemple synchronisme entre migrations et événements adaptatifs) sur d'autres espèces, dans un contexte de domestication ou plus largement.

---

## Summary

Domestication represents a major change in the history of humankind, leading to the emergence of agriculture during the Neolithic. By providing a series of independent evolutionary experiments in which plants and animals were selected for specific traits, this process has long interested evolutionary biologists. The sheep (*Ovis aries*) was one of the first species to be domesticated, which also makes it a prime model for studying early human attempts to develop livestock. The study of this process involves several complementary disciplines such as archaeology, palaeogenetics and modern genetics. The first two provide direct evidence of past processes, but are based on fragmented data in time and space. Modern DNA, on the other hand, is an abundant source of data that can be sampled much more exhaustively at the geographical level, but is not free of uncertainties related to the need to infer the past processes that led to the present genetic diversity. Such inferences are possible because present-day genomes retain traces of the processes that shaped them over the millennia. Genetic variation along genomes and across geographic regions has already been extensively studied in sheep, allowing inference of demographic (migrations, variations in population sizes over time) and adaptive (genomic regions responding to natural selection) processes. However, the temporal information contained in the variability of genomes is still largely under-exploited. Using 376 complete genomes representative of the current global diversity of sheep and their close wild relatives, the Asiatic mouflon, we have clarified their domestication history. The ages of appearance of more than 30 million genomic variants have been estimated and constitute an Atlas available to the scientific community via a website. The distribution of these ages shows that the history of sheep is intimately linked to our own and that their DNA bears the signs of major human crises such as the collapse of the Bronze Age or the Justinian plague. Genomic variants from these periods indicate drops in haplotype sharing across all domestic populations. The dating of selection signals in the genomes underlines the importance of population replacements linked to different waves of diffusion, and of the recent intensification of selections for agronomic traits. These processes have erased some of the traces of ancient selections. However, some breeds spared by these phenomena attest the probable early use of milk from the beginning of domestication and the selection of characteristics potentially involved in the domestication syndrome (colour, reduced aggressiveness, and immunity). This syndrome refers to a set of morphological, physiological and behavioural traits shared by multiple domestic species, probably selected unintentionally through the selection of the most docile animals. This large-scale dating of genomic variants is the first of its kind to be carried out in a species other than humans. Its use opens up new perspectives for specifying the chronology of events constituting evolutionary scenarios (e.g. synchronism between migrations and adaptive events) in other species, in a context of domestication or more widely.