THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE MONTPELLIER SUPAGRO

En Génétique et Génomique

École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau

Portée par

Unité de recherche UMR SELMET

Contribution à l'étude génomique des histoires démographique et adaptative des bovins et ovins de l'Océan Indien

Présentée par Jessica MAGNIER

Le 26 novembre 2021

Sous la direction de Laurence FLORI
et Matthieu LESNOFF, co-encadré par Emmanuel TILLARD

Devant le jury composé de

Joëlle RONFORT, Directrice de recherche, INRAE	Présidente du Jury
Nathalie MANDONNET, Directrice de recherche, INRAE	Rapporteure
Xavier ROGNON, Professeur, AgroParisTech	Rapporteur
Simon BOITARD, Chercheur, INRAE	Examinateur
Denis LALOE, Ingénieur de recherche, INRAE	Examinateur
Gwenola TOSSER-KLOPP, Directrice de recherche, INRAE	Examinatrice
Laurence FLORI, Chercheure, INRAE	Directrice de Thèse

UNIVERSITÉ DE MONTPELLIER

l'institut Agro
agriculture • alimentation • environnement

Montpellier SupAgro

Abstract

The Indian Ocean represents a central migration zone for bovine *(Bos taurus)* and ovine *(Ovis aries)* populations accompanying human populations between East Africa, the Middle East and Southeast Asia, but their genetic diversity as well as their demographic and adaptive histories remain poorly known.

The main objective of the thesis was to study the demographic and adaptive histories of cattle and sheep in the Indian Ocean, based on medium and high density genotyping data, with particular attention to the breeds of the western part of the zone located on the island of Mayotte, in the Comoros (never before phenotypically and genetically characterized) and on the island of Madagascar. For this purpose, a first phenotypic characterization was conducted on the sheep and cattle populations of Mayotte, allowing to highlight their phenotypic heterogeneity and to contribute to the recognition of the Zebu breed of Mayotte as a local French breed. This phenotypic characterization revealed a malformation of the external ear in 20% of the studied sheep, which present reduced or totally atrophied ears (microtia).

The study of SNP-HD data of current cattle from Mayotte and Madagascar by comparing them to those of a larger panel of breeds representative of the bovine genetic diversity of the world has allowed to highlight a great genetic proximity (predominantly indicine ancestry and weak African taurine ancestry) between these two breeds, having diverged in the 16th century, at the time of the arrival of the Europeans. Their common ancestral population is the result of a crossbreeding event involving an African zebu population (*African taurine x zebu*) and an Indian index population dating back to the 12th century. A sharp increase in the size of the estimated populations is observed between the 16th and 17th centuries and coincides with the expansion of the livestock trade to supply European ships in particular. The origin of the cattle populations of the western Indian Ocean islands reflects the complex history of human migration and trade in this area.

The Mayotte sheep population, which is close to the African fat-tailed sheep, is genetically homogeneous and would have undergone two peaks of inbreeding in the first half of the 20th century and another later one about ten years ago. An association analysis allowed us to locate a candidate mutation responsible for the microtia phenotype in the vicinity of the HMX1 gene, which is associated with the same type of malformation in Awassi and Altay sheep, and Highland cattle.

For each population, the search for selection signatures and functional annotation on candidate genes allowed us to determine the main biological functions impacted by natural

and artificial selection.

At the Indian Ocean scale, we completed the search for selection signatures in the genome of 17 cattle breeds and 9 sheep breeds of the coast (SNP50K data), by GEA analyses, which facilitate the identification of genetic variants associated with population environmental covariates. Functional annotation of candidate genes associated with 6 climatic covariates in both species identified key biological functions associated with the climate of the area.

Of the regions under selection identified in both species, only two regions contain genes associated with a climate variable (mean annual precipitation). Taken together, these results highlight the genetic originality of the Indian Ocean cattle and sheep breeds and specify the chromosomal regions, genes and biological functions/pathways involved in the adaptation of these breeds to the specific environmental conditions of this geographical area.

Résumé

L'Océan Indien représente une zone centrale de migration des populations bovines *(Bos taurus)* et ovines *(Ovis aries)* accompagnant les populations humaines entre l'Afrique de l'Est, le Moyen-Orient et l'Asie du Sud-Est mais leur diversité génétique ainsi que leurs histoires démographiques et adaptative demeurent mal connues.

L'objectif principal de la thèse a été d'étudier les histoires démographique et adaptative des bovins et ovins de l'Océan Indien, à partir de données de génotypages à moyenne et haute densité, en accordant une attention plus particulière aux races de l'Ouest de la zone localisées sur l'île de Mayotte, aux Comores (encore jamais caractérisée phénotypiquement et génétiquement jusqu'à présent) et sur l'île de Madagascar. Pour cela, une première caractérisation phénotypique a été conduite sur les populations ovine et bovine de Mayotte, permettant de mettre en évidence leur hétérogénéité phénotypique et de contribuer à la reconnaissance de la race Zébu de Mayotte comme race locale française. Cette caractérisation phénotypique a mis en évidence une malformation de l'oreille externe chez 20% des ovins étudiés, qui présentent des oreilles de taille réduite ou totalement atrophiée (microtie).

L'étude des données SNP-HD des bovins actuels de Mayotte et de Madagascar en les comparant à celles d'un panel de races plus large représentatif de la diversité génétique mondiale bovine a permis de mettre en évidence une grande proximité génétique (ascendance indicine prédominante et faible ascendance taurine africaine) entre ces deux races, ayant divergé au 16^{ème}siècle, au moment de l'arrivée des Européens. Leur population ancestrale commune résulte d'un évènement de métissage impliquant une population zébuine africaine (métisse *taurin africain x zébu*) et une population indicine Indienne remontant au 12^{ème}siècle. Une forte augmentation de la taille des populations estimée est observée entre le 16^{ème}et le 17^{ème}siècle et coincide avec l'expansion du commerce du bétail pour approvisionner notamment les navires européens. L'origine des populations bovines des îles de l'océan Indien occidental reflète l'histoire complexe des migrations humaines et du commerce dans cette zone.

La population ovine de Mayotte proche des moutons africains à queue grasse présente une homogénéité génétique et aurait subit deux pics de consanguinité dans la première moitié du 20^{ème}siècle et un autre plus tardif il y a une dizaine d'années. Une analyse d'association a permis de localiser une mutation candidate responsable du phénotype microtie à proximité du gène HMX1, associé au même type de malformation chez les moutons Awassi, Altay et les bovins Highland.

Pour chaque population, la recherche des signatures de sélection et l'annotation fonc-

tionnelle sur les gènes candidats a permis de déterminer les principales fonctions biologiques impactées par la sélection naturelle et artificielle.

A l'échelle de l'Océan Indien, nous avons complété la recherche de signatures de sélection dans le génome de 17 races bovines et 9 races ovines du littoral (données SNP50K), par des analyses GEA, qui facilitent l'identification de variants génétiques associés avec des covariables environnementales populationnelles. L'annotation fonctionnelle des gènes candidats associés à 6 covariables climatiques chez les deux espèces a permis d'identifier les principales fonctions biologiques associées au climat de la zone.

Parmi les régions sous sélection identifiées chez les deux espèces, seules deux régions contiennent des gènes associés à une variable climatique (précipitations moyennes annuelles). Pris globalement, ces résultats soulignent l'originalité génétique des races bovine et ovine de l'Ocean Indien et précisent les régions chromosomiques, les gènes et les fonctions/voies biologiques impliqués dans l'adaptation de ces races aux conditions environnementale spécifiques de cette zone géographique.

Liste des communications

Publiées

Magnier J., Ouvrard M., Raoul S., Janelle J., Mkadara A., Attoumani H., Saindou O., Moussa K., Naves M., Gautier M., Flori L., Tillard E. (2019) - Phenotypic and genetic characterization of zebu cattle population of Mayotte Island : on path to breed recognition - poster – 1p- *IIIe symposium Harlan* – juin 2019

Soumis

Magnier J., Druet T., Naves M., Ouvrard M., Raoult S., Janelle J., Moazami-Goudarzi K., Lesnoff M., Tillard E., Gautier M., Flori L. (2021) -The genetic of local cattle breeds from Madagascar and Comoros mirrors the complex pattern of human migrations and trade in Western Indian Ocean islands *Genes Genomes Genetics*,15 p.

En préparation

Magnier J., Ouvrard M., Raoul S., Janelle J., Mkadara A., Attoumani H., Saindou O., Moussa K., Naves M., Gautier M., Flori L., Tillard E.- Phenotypic characterization of Zebu from Mayotte.

Magnier J., Janelle J., Saindou O., Ouvrard M., Raoul S., Naves M., Gautier M., Flori L., Tillard E. - Phenotypic characterization of sheep from Mayotte.

Magnier J., Tillard E., Ouvrard M., Raoult S., Janelle J., deux collaborateurs de la plateforme MGX (Montpellier), Naves M., Gautier M. and Flori L. - Genetic characterization of the local sheep breed of Mayotte island and identification of a locus associated with microtia near the HMX1 gene.

Magnier J., Tillard E., Ouvrard M., Raoult S., Janelle J., Naves M., Gautier M. and Flori L. - Climate association test and comparison between bovine and ovine populations.

Remerciements

Enfin, çà y est. Après avoir postulé à plusieurs concours pour décrocher une bourse doctorale, c'est au moment où je commençais à lâcher l'affaire, qu'en 2018 Laurence et Emmanuel m'ont proposé un sujet de recherche en continuité avec la caractérisation phénotypique et génétique du zébu mahorais, projet pour lequel j'avais transpiré au sens propre et figuré à Mayotte.

Je remercie Laurence Flori pour son travail d'accompagnement quotidien, de soutien (surtout dans cette dernière phase de rédaction). Ton exigence a permis d'arriver à ce résultat. Je remercie également Matthieu Lesnoff et Emmanuel Tillard qui ont également participé à rendre possible cette thèse. Emmanuel, je te remercie tout particulièrement d'avoir parié sur moi il y a un peu plus de 5 ans quand tu m'as embauché pour travailler sur la caractérisation des races de ruminants de Mayotte. Tu as su me faire confiance et c'est ce dont j'avais le plus besoin à ce moment-là.

Je remercie les membres du jury, Joëlle Ronfort, Simon Boitard, Gwenola Tosser-Klopp, Denis Laloë, Nathalie Mandonnet, Xavier Rognon, qui ont eu la gentillesse d'évaluer ce travail de recherche. J'espère que les rapporteurs Nathalie et Xavier auront plaisir à décortiquer ce document.

Je remercie également les membres de mon comité de suivi pour leurs précieux conseils : Tom Druet, Stéphanie Manel, Alexandre Ickowizc et Mathieu Gautier. Merci à Vous Tom et Mathieu, ainsi que Michel Naves pour les précieux ajouts au premier article issu de ce projet de thèse.

Ce travail n'aurait pas vu le jour sans toute la phase de terrain qui s'est déroulé à Mayotte pendant mon VSC entre 2016 et 2018. Je remercie donc mes collègues de terrain les plus proches : Mélissa Ouvard, ma binôme technicienne génétique qui m'a initié aux organisations des instances raciales et avec qui le travail était fluide. Bien évidemment je n'oublie pas les techniciens de la CAPAM, Mkadara Anlidine, Oussoufi Saindou, Hidachi Attoumani et Moussa Kamardine qui ont eu la gentillesse de se décarcasser pour trouver les 400 zébus nécessaires et par la-même de nous faire découvrir l'île de Mayotte à travers leur regard. Je remercie également Jérôme Janelle dont le rôle a été essentiel, surtout pour la caractérisation phénotypique du mouton mahorais à un moment où la situation était critique. Je remercie également Solène Raoult, Laure Dommergues, Toufa Moussa, et tous les autres personnes de la Coopadem qui de près ou de loin ont participé au succès de la caractérisation phénotypique.

Retour à Montpellier. Ces 3 ans, en l'absence de terrain, se sont déroulés la majeure

partie du temps derrière l'écran. Si au début, j'envisageai d'aller au moins voir des zébus à Madagascar, une pandémie en a décidé autrement... Je remercie donc mes collègues de bureau : Bérénice qui était déjà là à mon arrivée (dont j'ai évité le mauvais oeil, enfin je crois...), Lina qui a commencé le même jour que moi et m'a accompagné durant la dernière année de thèse, Ilyes, Alioune et Annaël. Votre présence était indispensable pour sortir la tête de l'eau. Je remercie également les collègues de Baillarguet : Samantha, Johann, Marie-Cécile, Lionel, Jean-Marie, Marie-Jo, Corine, Isabelle, les deux Christian, les deux Matthieu, Vincent, Alexandre le DU à mon arrivée, Guillaume le DU à mon départ, Xavier, Samir, Laurent, Elodie, Denis, Ibra, Eric pour leur gentillesse, leur humour décapant et leurs discussions qui me permettaient d'aller sur le terrain par procuration.

Enfin un parcours de thèse n'est réussi que si on est bien accompagné personnellement : je remercie mes amis de partout, Amandine, Fanny, Arielle, les Siranais, les mahoraises, Juliette, Mano, Diane, ainsi que mes cousins, mes oncles et tantes qui ont été là pour me changer les idées ou pour m'écouter et me relancer quand le doute s'immisçait en moi. Je remercie spécialement ma meilleure amie Amélie qui m'a fait l'honneur d'être sa témoin de mariage et qui en revenant à Aurillac avec son Benjamin m'a redonné le coup de boost (quelle bonne idée ce mariage 15 jours avant de rendre le manuscrit ! Idéal pour recharger les batteries !) qu'il me fallait pour la dernière ligne droite.

Je remercie ma belle-famille avec qui j'ai fait connaissance au début de la thèse pour leur accueil. Les courts séjours en Bretagne m'ont permis un temps de penser à autre chose. Normalement après la thèse, vous devriez désormais faire connaissance avec la Jessica moins stressée et un peu plus sur Terre. Vos encouragements de la dernière ligne droite m'ont particulièrement aidé. Je remercie ma petite soeur, que j'ai eu le bonheur de retrouver après être partie 2 ans à l'autre bout de la planète. L'éloignement a fini par nous rapprocher et je suis bien contente d'avoir cette relation avec toi. Merci pour ton soutien moral et logistique de ces derniers mois. Je remercie bien évidemment mes parents qui ont eu la bonté de m'héberger dans leur chalet pendant les 4 + 8 mois de confinement. Ils ont été les plus heureux de voir revenir leur grande fille de 30 ans à la maison (pas forcément DANS la maison mais à 50m). Je sais que je vous ai fait revivre une autre période type "prépa" où j'étais là mais en fait pas vraiment là. Et pour finir, mon Gaël, toi qui a été le plus heureux de me voir arriver me confiner au Bousquin, même si tu as dû apprendre à vivre avec une thésarde, et j'imagine que cela n'a pas dû être simple. Je te remercie pour le soutien que tu m'as apporté dans ces 3 ans malgré la distance. Une nouvelle aventure nous attend ...

Table des matières

Introduction **19**

I Synthèse bibliographique 21

1 Origine de la diversité des races locales bovines et ovines et zoom sur les races de l'Ouest de l'Océan Indien **22**

1.1 Domestication des bovins et ovins 23
1.2 Histoire post domestication 25
1.2.1 Expansion européenne 25
1.2.2 Arrivée en Asie orientale 27
1.2.3 Arrivée sur le continent Africain 31
1.3 Histoire des populations bovines et ovines de l'Ouest de l'Océan Indien . . . 33
1.4 Conclusion du chapitre 38

2 Caractérisation de la structure de la diversité génétique des races locales **42**

2.1 De la caractérisation phénotypique à la caractérisation génétique des races . 43
2.2 Outils et méthodes de caractérisation de la diversité génétique 45
2.2.1 Les outils de production de données génomiques 45
2.2.2 Estimation de la diversité génétique 47
2.2.3 Inférence démographique 52
2.3 Application aux populations tropicales 60

3 Etudes des histoires adaptatives des populations bovines et ovines tropicales **65**

3.1 Recherche de signatures de sélection 66
3.1.1 Différentes signatures de sélection 66
3.1.2 Méthodes de détection 71
3.1.3 Identification des gènes candidats 75
3.1.4 Annotation fonctionnelle des gènes candidats 75
3.2 Association Genome Environnement 77
3.3 Adaptation des bovins et des ovins à l'environnement tropical 78
3.3.1 Adaptation aux conditions climatiques - tolérance à la chaleur . . . 78

3.3.2 Adaptation à la disponibilité des ressources alimentaires 81
3.3.3 Adaptation aux agents pathogènes et à leurs vecteurs 82

II Justification du projet et axes de recherche 84

III Résultats 108

4 Description phénotypique des bovins et ovins de Mayotte 109
4.1 Synthèse des résultats 110
4.2 Article 1 : Phenotypic characterization of Zebu from Mayotte 112
4.2.1 Abstract 112
4.2.2 Introduction 113
4.2.3 Material and methods 114
4.2.4 Results 120
4.2.5 Discussion 123
4.3 Article 2 : Phenotypic characterization of sheep from Mayotte 127
4.3.1 Abstract 127
4.3.2 Introduction 128
4.3.3 Material and methods 129
4.3.4 Results 133
4.3.5 Discussion 136
4.4 Discussion du chapitre 140

5 Caractérisation des histoires démographiques et adaptatives des populations bovines locales de l'Ouest de l'Océan Indien 141
5.1 Synthèse des résultats 142
5.2 Article 3 : Demographic and adaptive histories of local cattle breeds from Mayotte and Madagascar 143
5.3 Résultats supplémentaires obtenus avec la puce moyenne densité 159
5.3.1 Analyse exploratoire 159
5.3.2 Un évènement de métissage supplémentaire détecté 162
5.3.3 Un clustering des ROH variable 162
5.3.4 Conclusion 163
5.4 Introgression du boeuf de Bali 165
5.5 Discussion du chapitre 169

6 Caractérisation génétique du Mouton mahorais et localisation de la mutation responsable de microtie 174
6.1 Synthèse des résultats 175
6.2 Point méthodologique sur les analyses d'association cas/contrôles 176

6.3 Article 4 : Genetic characterization of sheep from Mayotte . . . 178
6.3.1 Abstract . . . 178
6.3.2 Introduction . . . 179
6.3.3 Material and Methods . . . 180
6.3.4 Results . . . 186
6.3.5 Discussion . . . 192
6.3.6 Supplementary Tables and Figures . . . 197
6.4 Résultats supplémentaires . . . 201
6.4.1 Identification de la mutation candidate . . . 201
6.4.2 Détection de signatures de sélection dans le génome des ovins mahorais et annotation fonctionnelle des gènes candidats identifiés . . . 202
6.5 Discussion du Chapitre . . . 208

7 Identification des régions génomiques et des gènes associés avec le climat chez les populations bovines et ovines de l'Océan Indien 216
7.1 Synthèse des résultats . . . 217
7.2 Article 5 : Climate association tests for bovine and ovine populations . . . 218
7.2.1 Abstract . . . 218
7.2.2 Introduction . . . 219
7.2.3 Material and methods . . . 221
7.2.4 Results . . . 228
7.2.5 Discussion . . . 237
7.2.6 Supplemental Figures and Tables . . . 243
7.2.7 Supplemental Results : Exploratory analysis . . . 251
7.3 Discussion du chapitre . . . 259

IV Discussion et perspectives 268

8 Discussion 269
D.1 Résumé des travaux . . . 269
D.2 Perspectives . . . 271
D.2.1 Intégration de nouvelles données et recherches de nouveaux gènes candidats sous pression de sélection . . . 271
D.2.2 Caractérisation des mutations candidates impliquées dans l'adaptation 274
D.2.3 Prédiction génomique de l'adaptation . . . 274
D.3 Application dans les pays en développement . . . 275
D.3.1 Utilisation des outils génomiques dans les pays en développement . . . 275
D.3.2 Utilisation des résultats pour les populations de Mayotte . . . 277
D.4 Conclusion . . . 279

V Annexes **284**

A Monographies **285**

A.1 Phylogénie 286
A.2 Monographie des différentes populations étudiées 289
A.2.1 Bovines 289
A.2.2 Ovines 294
A.3 Histoire des épizooties principales dans la zone 299

B Méthodes et outils **301**

B.1 Récapitulatifs des analyses des jeux de données 302
B.2 Pipelines 303
B.2.1 Conversion des positions des SNPs 303
B.2.2 Analyse exploratoire 303
B.2.3 Inférence démographique 304
B.2.4 Recherche des signatures de sélection -Rehh 304
B.2.5 Tests d'association génome-environnement 306
B.2.6 Tests d'association microtie 311

Table des figures

1.1 Evènements de domestication des animaux d'élevage 24
1.2 Schéma des intéractions spatio-temporelle entre auroch et les races bovines européennes 27
1.3 Utilisation du zébu à Mayotte comme outil de traction (Carte postale - Société des Comores - 1910) 36
1.4 Utilisation du zébu Moka de la Réunion lors de la récolte de la canne à sucre. 37
1.5 Bêlier photographié à Madagascar en 1920. 38
1.6 Synthèse des voies de migration des genres *Bos taurus taurus* et *Bos taurus indicus* 40
1.7 Synthèse des voies de migrations du mouton domestique *Ovis aries* 41

2.1 Quelques mesures corporelles effectuées sur des zébus et des moutons de Mayotte. 44
2.2 Distribution de la MAF et de l'hétérozygotie observée des puces SNP bovine 47
2.3 Méthode de l'ACP 50
2.4 Example de clustering ADMIXTURE 51
2.5 Schéma des Runs Of Homozygoty 52
2.6 Principales sorties graphiques du logiciel RZooroh ayant calculé les segments HBD pour le bison Européen 54
2.7 Arbres phylogéniques reliant les populations A, B, C et D 55
2.8 Principe du logiciel malder 58
2.9 Résultats de l'analyse exploratoire (ACP et clustering) du jeu de données SNPs regroupant un panel de races mondiales [Decker et al., 2014] 61
2.10 Comparaison du premier plan d'ACP entre [Kijas et al., 2012] et [Cao et al., 2021] 62
2.11 Résultats du calcul de la taille efficace Ne pour des populations bovines tropicales 64

3.1 Modes de sélection naturelle 66
3.2 Effet des différents types de balayage sélectifs sur la diversité génétique locale. 67
3.3 Les différents types de signature de sélection positive 69
3.4 Méthodes de détection des signatures de sélection de populations 70
3.5 Représentation du processus de phasage 73
3.6 Représentation des EHH 73
3.7 Représentation graphique des approches de recherche de gènes candidats . . 76

3.8 Résumé graphique de la stratégie globale d'étude des données de génotypage 88

4.1 Zébus mahorais femelle **a** et mâle **b**. Moutons mahorais femelle **c** et mâle **d**. 111
4.2 Geographic position of the 128 farms where 400 local zebu cattle were sampled in Mayotte island 115
4.3 Proportion of horn type in function of age class between male and female individuals. 122
4.4 Geographic positions of the 20 farms where the 100 sheep individuals were phenotypically sampled. 129
4.5 Violin plots describing relationship between Body score condition and Medium tail circumference 135

S5.1 Résultats de l'ACP et du clustering sur puce bovine 54K 160
S5.2 Résultats du clustering pour la population de Mayotte 54K 160
S5.3 Comparaison de la distribution des ROH pour MAY et ZMA en fonction du type de données 164
S5.4 Clustering hierarchique issus du logiciel ADMIXTURE [Alexander et al., 2009]. 167

6.1 Les méthodes GWAS 177
6.2 Picture of the three external ear phenotypes in the ovine population from Mayotte 182
6.3 Neighbor-joining tree based on pairwise *Fst* between each worldwide sheep breeds of the dataset. For the significance of breed abreviations, see Table S6.3 188
6.4 PCA sheep worldwide populations 189
6.5 Unsupervised hierarchical clustering for sheep worldwide dataset 190
6.6 Characterization of individuals inbreeding levels in ovine population from Mayotte island 195
6.7 Focus on the OAR6 region concerned by microtia mutation. 196
S6.8 Heatmap representation of Fst values between the 81 ovine populations each other. 198
S6.9 Manhattan plot des signatures de sélection obtenues par le calcul des scores EHH pour le mouton de Mayotte 207

7.1 Localization of the 17 cattle and 9 sheep populations used in differentiation and climatic association tests. 222
7.2 Geographic pattern of the annual mean values of the five main climatic variables 225
S7.3 Supplemental figures for Genome Environment Analyses with bovine dataset 245
S7.4 Supplemental figures for Genome Environment Analyses with ovine dataset . 246
S7.5 Principal Components analysis (PCA) and unsupervised clustering results for bovine populations. 254
S7.6 Principal Components analysis (PCA) and unsupervised clustering results for ovine populations. 255

D.1 Illustration des principales approaches de prédiction génomique basée sur le décalage génomique 276
D.2 Schéma des processus de décision pour la gestion des ressources zoogénétiques d'élevage. 279

A.1 Arbre phylogénétique des bovidés 287

B.1 Schéma récapitulatif des analyses des différents jeux de données utilisés dans les sections du Chapitre 5. 302
B.2 Schéma récapitulatif des analyses des différents jeux de données étudiés dans le Chapitre 6. 302
B.3 Schéma récaputilatif des analyses des différents jeux de données étudiés dans le Chapitre 7. 303

Liste des tableaux

4.1 Sampling objective of zebu individuals among township of Mayotte island. . . 114
4.2 Description of quantitative variables and calculated index from the variables for zebu from Mayotte . . . 117
4.3 Overview of the frequency of each level for the qualitative traits recorded in Zebu from Mayotte . . . 118
4.4 Overview of the quantitative variables measured on 400 zebu individuals . . . 120
4.5 Relationship between body score condition and hump size . . . 121
4.6 Description of the quantitative variables stadied for sheep from Mayotte . . . 132
4.7 Overview of the quantitative variables measure on 100 sheep individuals . . . 133
4.8 Overview of the frequency of each level for the qualitative traits recorded for local sheeps from Mayotte . . . 134

S5.1 Description des populations utilisées dans l'analyse exploratoire du jeu de données moyenne densité (532 individus génotypés sur 39'729 SNPs) . . . 161
S5.2 Description du jeu de données utilisé pour détecter une introgession indonésienne dans les zébus de l'Ouest de l'Océan Indien. . . . 165
S5.3 D-statistiques pour l'introgression de Bali cattle dans AFZ . . . 168
S5.4 f3-stats pour l'introgression de Bali cattle dans AFZ . . . 168
S5.5 f4-ratio pour l'introgression de Bali cattle dans AFZ . . . 168

6.1 Fis and Fst coefficient calculated for worldwide sheep breeds dataset . . . 183
S6.2 The 48 individuals genotyped on Illumina ovine 50K SNP bead chip in sampled population of sheep from Mayotte. . . . 197
S6.3 Description of the 81 breeds used in exploratory analysis. . . . 199
S6.4 Populations used for the identification of footprints of selection based on EHH-derived tests . . . 203
S6.5 Candidate genes associated with EHH score for sheep from Mayotte . . . 206

7.1 List of the 17 cattle and 9 sheep populations integrated into the genotyping dataset after quality filtering. . . . 221
7.2 Positions and candidate genes associated with climatic covariable for bovine populations . . . 232

7.3 Positions and candidate genes associated with climatic covariable for ovine populations . . . 233
7.4 Regions harbouring footprints of selection based on the XtX score . . . 234
7.5 Regions harbouring footprints of selection based on the XtX score for ovine populations . . . 236
7.6 Top five significant functions of the putative candidate genes associated with climatic covariables for Zebus (**a**) and Sheep (**b**) . . . 237
S7.7 Description of climate covariable from Climond website. . . . 243
S7.8 Climatic covariable matrix for bovine dataset . . . 244
S7.9 Climatic covariable matrix for ovine dataset . . . 244
S7.10 Networks bovine BF . . . 247
S7.11 Networks ovine BF . . . 248
S7.12 5 first lines of annotation table of candidate genes for bovine dataset . . . 249
S7.13 8 first lines of annotation table of candidate genes for ovine dataset . . . 250
S7.14 Description of 31 cattle and 29 sheep breeds used for exploratory analysis. . . . 253
S7.15 Matrix pairwise Fst for 31 cattle breeds used for exploratory analysis . . . 256
S7.16 Matrix pairwise Fst for 29 sheep breeds used for exploratory analysis . . . 257
S7.17 Results of clustering for parameter value K = 3 on individuals from bovine MAY and ONG populations. . . . 258

Introduction

Face au contexte de changement climatique global, la place des ressources génétiques animales prend tout son sens en étant un des moyens d'adaptation et de résilience des élevages. Cependant, huit principales menaces (croisements incontrôlés, introduction/ utilisation accrue de races exotiques, politiques ou institutions faibles, manque de rentabilité/compétitivité, intensification des systèmes de production, maladies/ contrôle des maladies, perte des pâturages ou de l'envionnement de production et un mauvais contrôle de la consanguinité) sont à l'origine de l'érosion génétique observé par la FAO.

Les pays membres de la FAO ont entrepris en 2007 le Plan d'Action Mondial pour les Ressources Zoogénétiques consistant en une caractérisation phénotypique des populations d'animaux d'élevages, un inventaire démographique, la mise en place de programme d'utilisation et de conservation des ressources génétiques animales présentes sur la surface du globe. Avec plus ou moins de succès, car la politique de gestion de ces ressources est fortement dépendante de la volonté et des moyens des pays membres.

Le deuxième rapport sur l'état des ressources génétiques animales [Scherf et al., 2015] dresse le constat que 14 espèces sur 30 de mammifères et d'oiseaux assurent la fourniture de 90 % des produits d'élevage destinés à la consommation. A ce jour 8 800 races sont reconnues mais 7 % ont disparu et 17 % sont menacées d'extinction. Parmis ces races, environ 7 700 sont de types locales. Les races locales représentent un réservoir de la diversité des ressources génétiques et un facteur de résilience des systèmes [Dumont et al., 2013, Couix et al., 2016]. En France, le terme de race locale est définie dans le Code rural en tant que "race majoritairement liée par ses origines, son lieu et son mode d'élevage, à un territoire donné". Toujours selon la loi, une race est reconnue comme locale si 30 % des effectifs sont recensés dans un seul département (vache Bretonne Pie-Noire) ou 70 % dans 3 départements limitrophes (vache Salers). La définition établie par la FAO établie qu'une race locale est un "sous-groupe spécifique d'animaux domestiques présentant des caractéristiques externes définissables et qui sont présentes dans le pays depuis suffisamment longtemps pour être génétiquement adaptées à un ou plusieurs des systèmes de production ou environnements traditionnels du pays" [Scherf et al., 2015]. Actuellement, à l'échelle mondiale, peu d'informations remontent du terrain quant au statut des races locales : en 2019, seules 2 761 races ont pu être évaluées, et parmi elles 73 % ont été jugées menacées d'extinction.

Ainsi la caractérisation phénotypique et génétique des ressources génétiques animales constituent des étapes primordiales pour mettre en place les politiques nécessaires à leur

conservation. Les bovins et les ovins sont les espèces d'élevage les plus diverses en nombre de populations locales déclarées (respectivement 1 047 et 1 164), de populations "régionales-transfrontalières" (respectivement 98 et 130) et de populations internationales (respectivement 110 et 98) [FAO, 2018].

Première partie

Synthèse bibliographique

Chapitre 1

Origine de la diversité des races locales bovines et ovines et zoom sur les races de l'Ouest de l'Océan Indien

Les bovins, classés dans l'espèce *Bos taurus*, regroupent deux sous-espèces : les taurins ou *Bos taurus taurus* et les zébus ou *Bos taurus indicus* qui s'en distinguent au premier abord par la présence d'une bosse plus ou moins développée chez les zébus. D'autre animaux se rapportant au genre *Bos* côtoient les taurins et les zébus et sont interfertiles bien que les produits mâles résultant de ces accouplements soient stériles la plupart du temps [Zhang et al., 2020] : Banteng (*Bos javanicus*), Gaur (*Bos gaurus*), Gayal (*Bos frontalis*) et Yack (*Bos grunniens*). Concernant les ovins, seule l'espèce *Ovis aries aries* définit et rassemble toutes les races de mouton domestique. La classification des moutons peut se réaliser selon plusieurs critères : selon le type de pelage (laine ou poil), la présence ou non d'une queue ou croupe grasse, et la longueur des pattes. Il existe une grande diversité de races au sein de chaque espèce. Plus de 800 races bovines et plus de 1500 races ovines ont été répertoriées dans monde (http ://www.fao.org/dad-is/fr/). Le processus de domestication, les migrations des animaux depuis leurs foyers de domestication et leur répartition dans différents environnements auxquels ils ont dû s'adapter ont contribué à cette diversité intraspécifique.

1.1 Domestication des bovins et ovins

Les données archéozoologiques et génétiques suggèrent que les bovins auraient été domestiqués il y a 10 000 ans environ, dans au moins deux centres de domestication distincts [Loftus et al., 1994, Hanotte, 2002], à partir de l'auroch leur ancêtre sauvage (*Bos primigenius*) dont le dernier spécimen a disparu en 1627 en Pologne.

En effet, si des travaux semblaient questionner la présence d'un troisième site de domestication à l'origine du rameau des taurins Africains dans la région Nord-Est de l'Afrique il y a 8 000 - 9 000 ans [Gifford-Gonzalez and Hanotte, 2011], les récents travaux semblent écarter cette hypothèse et aboutir à un consensus pour deux sites de domestication : le Croissant Fertile aurait été le théâtre de la domestication du rameau "taurin" il y a 8-10 000 ans tandis que la Vallée de l'Indus (aire géographique correspondant au Baluchistan, région du Pakistan) celui du rameau "zébu", plus récemment autour d'il y a 7 500 ans [Loftus et al., 1994, Troy et al., 2001, Lewis et al., 2011, Decker et al., 2014, Pitt et al., 2019].

Les moutons, quant à eux, ont été la première espèce domestiquée et descendraient du mouflon asiatique (*Ovis orientalis*) [Bruford and Townsend, 2006, Meadows et al., 2011, Muigai and Hanotte, 2013, Zeder, 2017]. Leur bassin de domestication semble être situé au Proche Orient, dans la partie centrale du Croissant Fertile, dans les parties hautes du Tigre et de l'Euphrate [Lev-Yadun et al., 2000, Zeder, 2011, Muigai and Hanotte, 2013]. Le processus de domestication des ovins remonterait à 8 à 11 000 ans [Ryder, 1984].

Les ressources génétiques ovines actuelles peuvent être classifiées en fonction de leur type phénotypique : le type de queue (les moutons à queue fine "thin-tailed" ou à queue grasse "fat-tailed"), la présence ou non de croupe grasse "rump-fat" et pelage à laine ou à poil. Il semblerait que l'ancêtre sauvage des moutons aurait présenté le phénotype "queue fine", de longs poils de jarre et relativement peu de laine comme les premiers moutons domestiques

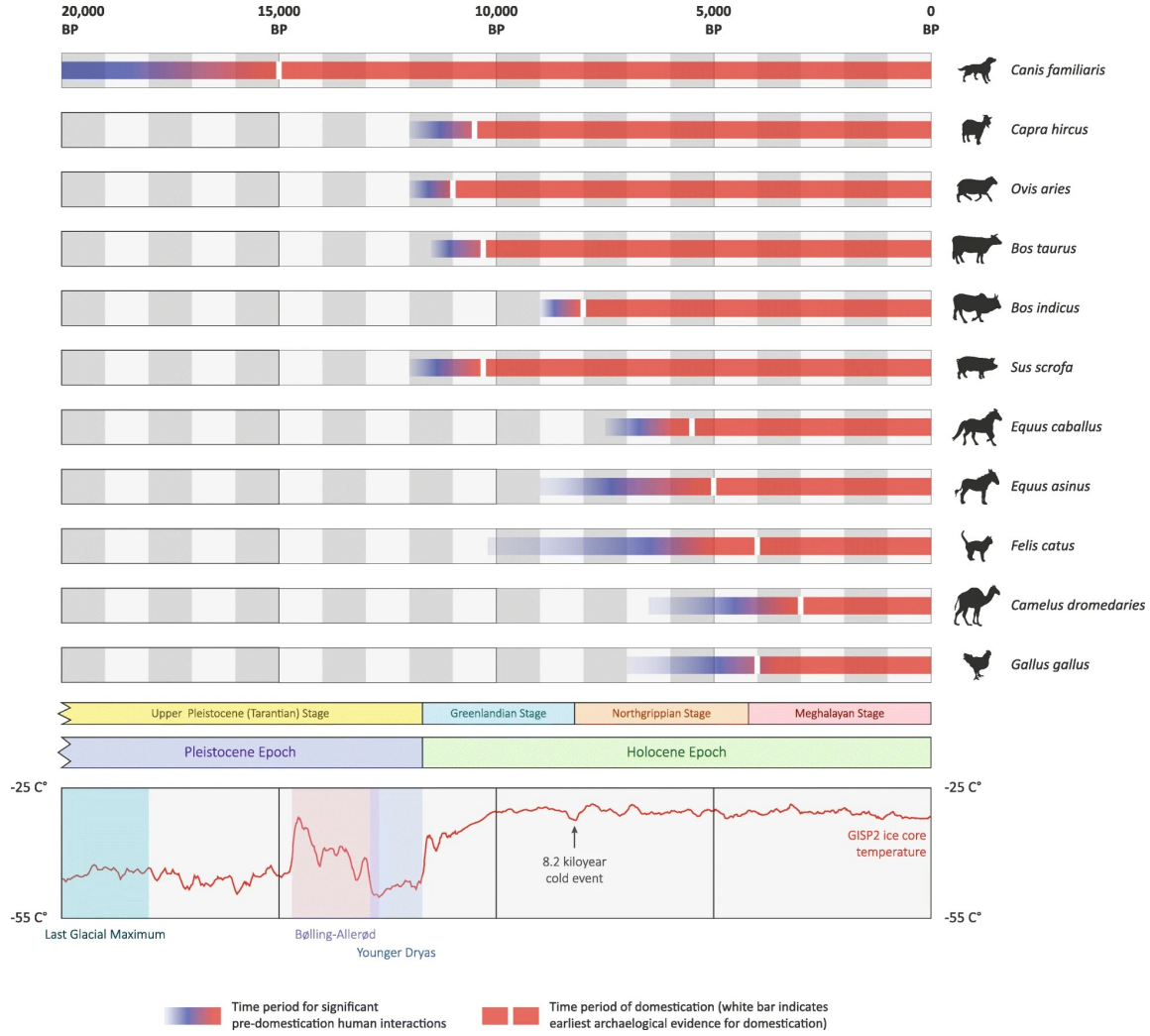


FIGURE 1.1 – Evènements de domestication de 11 espèces d'animaux d'élevage [Larson and Fuller, 2014]. Les périodes historiques (stratigraphie obtenue sur le site de la Commission Internationale de Statigraphie) ainsi que les chronologies climatiques sont présentées. Pour chaque espèce, les périodes temporelles pré-domestication où les intéractions entre l'homme et l'animal étaient supposées sont indiquées. Les courbes de température ont été générées à partir des données de température et d'accumulation des carottes de glace de GISP2 (Figure issue de [McHugo et al., 2019])

[Ryder, 1984, Sherratt, 1981, Fuller, 2006].

Le processus continu de domestication a impliqué de forts changements morphologiques distinguant les espèces domestiquées des espèces sauvages et contribuant à la diversité des populations bovines et ovines. Le syndrome de domestication observé par Darwin [Darwin, 1868] chez plusieurs espèces de mammifères et expérimenté un siècle plus tard sur le renard roux [Belyaev and al, 1979, Trut, 1999, Lord et al., 2020] englobe un ensemble de caractéristiques biologiques comprenant la pédomorphose (vitesse et durée du développement), la réduction du dimorphisme sexuel, les modifications de la morphologie crânio-faciale accompagnée d'une diminution de la taille du cerveau, une forte variation de la couleur du pelage et une dépigmentation, des oreilles non dressées, tombantes et petites ainsi que des altérations du système endocrinien avec des changements significatifs de la physiologie de la reproduction chez la femelle, notamment des cycles d'oestrus fréquents et non saisonniers [Dobney and Larson, 2006, Driscoll et al., 2009, Jensen, 2014, Wilkins et al., 2014].

Une hypothèse quant à l'origine de ce syndrome est que le développement de toutes les caractéristiques observées est étroitement lié aux cellules de la crête neurale (CCN)[1]. Ainsi, la sélection initiale pour la docilité entraîne une réduction des tissus dérivés de la crête neurale ayant une importance comportementale, par le biais de multiples variantes génétiques préexistantes qui affectent le nombre de cellules de la crête neurale aux sites finaux. Cette hypofonction de la crête neurale produit, en tant que sous-produit non sélectionné, les changements morphologiques de la pigmentation, des mâchoires, des dents, des oreilles, etc. présentés dans le Syndrome de Domestication [Wilkins et al., 2014].

1.2 Histoire post domestication

Après leur domestication, les bovins et les ovins ont emprunté différentes routes de migration depuis leurs foyers de domestication vers les différents continents en suivant les populations d'éleveurs. Leurs histoires démographiques sont en effet étroitement liées à celles des populations humaines. Ces migrations ont largement contribué à façonner la diversité des races locales actuelles.

1.2.1 Expansion européenne

Le bétail domestiqué *Bos taurus* s'est répandu depuis l'Anatolie dans toute l'Europe grâce aux migrations des premiers agriculteurs pendant la transition néolithique le long des côtes méditerranéennes et du Danube, impliquant un petit nombre d'individus transportés par bateau, respectivement vers - 8 000 ans et - 7 500 ans [Cymbron et al., 2005, Felius et al., 2014, Zeder,

1. Les CCN sont des cellules souches spécifiques aux vertébrés qui apparaissent pour la première fois au début de l'embryogenèse au niveau du bord dorsal ("crête") du tube neural, puis migrent ventralement dans tout le corps, à la fois dans le crâne et le tronc, donnant naissance aux précurseurs cellulaires de nombreux types de cellules et de tissus et favorisant indirectement le développement d'autres types de cellules [Hall, 1999, Gilbert, 2000, Trainor, 2013, Carlson, 2018]

2017]. Ceci expliquerait le gradient de diversité génétique observé de l'Europe de l'Est vers l'Europe centrale et occidentale, les bovins ibériques et du sud de la France étant ceux qui présentent la plus faible diversité génétique en Europe [da Fonseca et al., 2019]. Les races du Sud de l'Europe se caractérisent par une introgression taurine africaine, dont l'origine serait des bovins taurins africains qui ont probablement été introduits dans la péninsule ibérique à partir du Maghreb, d'où ils se sont répandus dans le sud de l'Europe [da Fonseca et al., 2019]. En outre, plusieurs études génomiques révèlent une influence indicine dans les races bovines italiennes [McTavish and Hillis, 2014, Decker et al., 2014, Upadhyay et al., 2017, Flori et al., 2019, Barbato et al., 2020] dont l'origine pourrait être la migration massive des bergers des steppes de Yamnaya [2] (entre la mer Noire et la mer Caspienne), via une migration maritime depuis la région de la Méditerranée orientale il y a 3 000 ans, ou en suivant la route de la Soie 2 000 ans plus tard [Pellecchia et al., 2007, Flori et al., 2019].

Après le Moyen Âge, le progrès culturel et technique et la demande croissante de nourriture liés à la Révolution industrielle ont incité à la rationalisation de l'élevage européen qui se caractérisait par une productivité faible, une utilisation dans l'agriculture de subsistance. A partir du 18^{iéme} siècle, une gestion organisée des races régionales se met en place par l'utilisation de livres généalogiques regroupant les pedigrees de chaque animal répondant au standard de la race défini au préalable et la rationalisation de l'utilisation des reproducteurs mâles. Ainsi, des groupes génétiquement isolés d'animaux phénotypiquement homogènes sont constitués. Cette révolution de la gestion raciale s'est produite dans toute l'Europe, où les expositions de bétail ont rapidement abouti à la fusion des races primitives entre régions voisines. La productivité des races européennes a ainsi été améliorée par la spécialisation, concomitante avec une modification de l'apparence des races [Felius et al., 2014].

Il est à noter cependant, que plusieures races bovines traditionnelles qualifiées de primitives constituent une ressource précieuse pour la variabilité génétique [Tapio et al., 2006]. Ces races seraient issues de croisements sporadiques ayant eu lieu entre le bétail domestique et les aurochs européens indigènes, qui ont persisté dans certaines régions jusqu'au Moyen Âge [Rokosz, 1995, McHugo et al., 2019] (Figure 1.2) et auraient conservé les caractéristiques primitives de leurs ancêtres sauvages, leur conférant une rusticité par rapport aux races commerciales. Néanmoins, plusieures de ces races présentent des tailles de population effectives déclinante, ce qui érode leur diversité génétique [Tapio et al., 2006].

Concomitamment aux déplacement des bovins, un groupe d'ovins primitif à "queue fine" élevés pour la viande et le lait aurait migré hors du centre de domestication vers le Nord et le long du bassin méditerranéen [Pereira et al., 2006, Zeder, 2017] suivi d'un autre évènement de migration de moutons sélectionnés pour leur laine [Ryder, 1983, Rannamäe et al., 2016] (au moins à partir du VI^{iéme} millénaire avant notre ère) [Epstein, 1971, Pereira et al., 2006, Zygoyiannis, 2006, Chessa et al., 2009, Muigai and Hanotte, 2013, Beaujard, 2019c, Deng et al., 2020].

2. source de dispersion des langues indo-européennes vers l'Europe centrale et septentrionale et l'Inde (Jones et al.,2015) and (Haak et al.,2015)

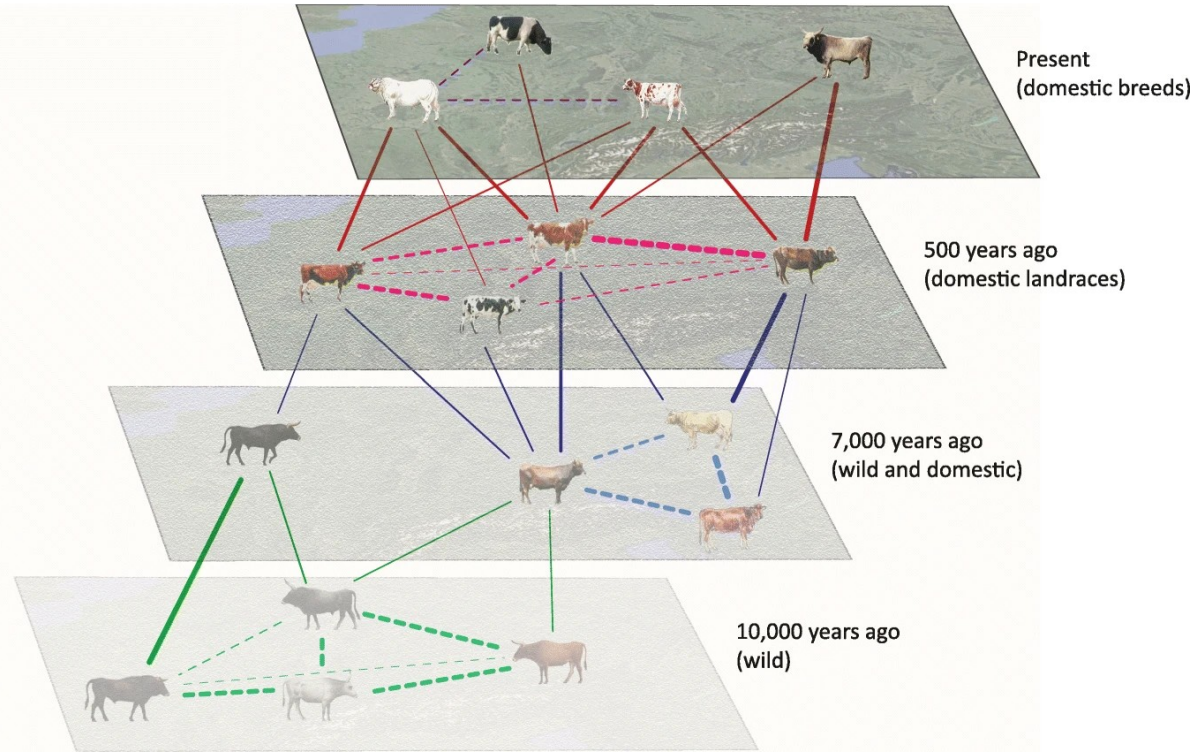


FIGURE 1.2 – Schéma des intéractions spatio-temporelle entre auroch et les races bovines européennes. Les intéractions temporelles sont représentées par des traits pleins entre les animaux, et les intéractions spatiales par des trains en pointillés (Figure issue de [McHugo et al., 2019]).

Plus récemment, au sein du Croissant Fertile, suite à un processus de sélection parmi les moutons domestiqués à "queue fine" datant d'il y a 3 000 ans [Ryder, 1984], le groupe des moutons à "queue grasse" est apparu et a persisté dans la région jusqu'à récemment puisqu'il réprésente le type dominant. iCe type de moutons a été sélectionné en réponse à des conditions climatiques extrêmes ou des besoins des éleveurs et se serait étendu vers l'Afrique de l'Est et l'Asie centrale et occidentale, bien qu'une migration vers l'Europe ne soit pas à exclure[3] [Epstein, 1971, Ryder, 1983, Ryder, 1984, Moradi et al., 2012, Muigai and Hanotte, 2013, Deng et al., 2020].

1.2.2 Arrivée en Asie orientale

Des bovins

Les preuves archéologiques suggèrent que des bovins relevant de l'espèce *Bos taurus* ont pu être importés dans un premier temps en Asie de l'Est depuis l'Asie de l'Ouest à la fin du Néolithique (entre 3 000 et 2 000 avant JC) [Cai et al., 2014]. Puis, le bétail taurin d'Asie de l'Est a pu se répandre dans tout le nord de la Chine et sur les marges du plateau

3. L'observation d'un haplotype du chromosome Y (y-HB) associé au phénotype de la queue grasse dans quatre populations européennes à queue fine (Leccese, White Mountain, Brown Mountain et Sumavska) suggère une introgression de moutons à queue grasse depuis l'Afrique du Nord via la route maritime méditerranéenne [Deng et al., 2020].

tibétain grâce à une expansion plus large de la culture du millet il y a 3 600 ans [Chen et al., 2015]. La population taurine située au nord-est de la Chine a pu migrer vers le Japon via la péninsule coréenne à un stade assez tardif, au deuxième siècle de notre ère [Mukai et al., 1989]. L'expansion du pastoralisme précoce et l'invasion récente au 13^{éme} siècle de Mongols [Grousset, 1941] dans des zones immenses et étendues de la Chine a pu également conduire à un flux supplémentaire de gènes à travers l'Eurasie orientale [Chen et al., 2018].

L'expansion de *Bos indicus* à partir de la Vallée de l'Indus vers l'Asie orientale a débuté il y a 5 500 - 4 000 ans comme l'en attestent les données archéozoologiques suggérant la présence de zébus domestiques dans le Gujarat et la région du Gange (au Nord-Ouest de l'Inde) [Fuller, 2006], suite à l'invasion des Rig Vedic Aryan, une tribu nomade pastoraliste, venant de l'Asie centrale [Felius et al., 2014]. Il y a 3 500 ans les zébus auraient été introduits dans la région du Bihar et du Bengale (Nord Est de l'Inde) [Payne and Hodges, 1997]. Il semblerait cependant que cette première expansion terrestre ait été stoppée au niveau du Bangladesh par la chaîne Montagneuse de l'Arakan [Yonesaka et al., 2016, Sudrajad et al., 2020]. Par la suite, les populations zébuines d'Inde ont été déplacées par la mer au cours des échanges commerciaux s'intensifiant dans la zone de l'Océan Indien durant les derniers siècles avant J-C ([Beaujard, 2019c], p.498) [Fuller, 2006]. En effet, à cette époque, les déplacements se faisaient par la mer entre l'Inde et les côtes occidentales de la Birmanie et de la péninsule Thai-Malay, puis par la terre en traversant l'isthme de Kra en Thaïlande ([Beaujard, 2019c], p.499). Des populations zébuines indiennes auraient été introduites il y a 2 500 ans dans la péninsule indochinoise [Siripan et al., 2019] et au Sud de la Chine [Higham, 1996], et se seraient croisées avec d'autres espèces de bovidés (banteng, gaur et gayal) qui étaient longtemps dominantes dans la région [Payne and Hodges, 1997, Decker et al., 2014, Felius et al., 2014, Wu et al., 2018] et qui ont persisté jusqu'à nos jours [Chen et al., 2018]. Un gradient de métissage avec des populations taurines présentes en Chine le long de l'axe Sud-Nord (*Zebu X Taurin*) a également été détecté [Chen et al., 2018]. Il résulterait de l'introduction zébuine plus tardive que celle des taurins en Asie Centrale il y a 5 000 ans [Flad et al., 2007].

En Thaïlande, les premières traces de bovins domestiques dateraient d'il y a 4 000 à 5 000 ans avant JC [MacHugh et al., 1997]. Néanmoins, les animaux de type *Bos indicus* semblent avoir atteint la Thaïlande il y a au moins 2 500 ans [Siripan et al., 2019]. Les zébus Thai actuels [Charoensook et al., 2013, Wangkumhang et al., 2015] récemment génotypés, distingués en fonction de leur phénotype et de leur localisation (au Nord connu sous le nom de White Lamphun ou Kho-Khaolumpoon, Kho-Isaan au Nord-Est, Kho-Lan en Thaïlande Central, et Kho-Chon au Sud) ont été identifiés comme appartenant à la sous-espèce *Bos indicus* et sont exempts de métissage avec des taurins [Charoensook et al., 2013, Wangkumhang et al., 2015] [4].

En Indonésie, le boeuf de Bali, forme domestiquée du Banteng (*Bos javanicus javanicus*)

4. Ces populations zébuines locales côtoient des taurins de races productives (Charolais, Hereford, Simmental, Shorthorn) et indicine (Brahman) que la Thaïlande a aussi importé afin de palier la faible production qu'offrait ces races locales. Néanmoins selon les dernières études génétiques [Wangkumhang et al., 2015], il semblerait que les races locales aient été préservées de flux génétiques exotiques récents.

il y a 5 000 ans [Payne and Hodges, 1997] sont les bovins les plus nombreux [Sutarno and Setyawan, 2016]. Ces bovins sont interfertiles avec les zébus mais seuls les produits femelles isssus de croisement *Zebu x Banteng - Bali cattle* peuvent se reproduire [Vadhanakul et al., 2004]. Ainsi, la particularité des races zébuines indonésiennes est leur structure génétique présentant un gradient d'introgression *Bos javanicus* [Mohamad et al., 2009]. Les dernières études génétiques basées sur des données de génotypage sur puce 50K [Sudrajad et al., 2020] suggère une arrivée de zébus indiens en Indonésie durant l'Age de Bronze par les routes maritimes empruntées par les commerçants arabes et hindous entre l'Inde et l'Asie du Sud-Est [Wormser, 2017], ainsi que peut-être plus récemment, il y a 1 000 ans par une population de Singhalais originaires du Sri Lanka et ayant probablement emmené des zébus [Felius, 1995, Felius et al., 2014]. Ces hypothèses sont confirmées par des données archéologiques [Felius, 1995, Sutarno and Setyawan, 2016]. Des importations récentes de zébus Ongole d'Inde durant le 19^{iéme} siècle sont survenues d'abord sur l'île de Sumatra, puis sur l'île de Java absorbant ainsi les éventuelles populations bovines locales [Porter et al., 2016, Sutarno and Setyawan, 2016]. Ces bovins Ongole, en se métissant avec les bovins locaux typés "zébus" ayant une ascendance maternelle Banteng, ont donné une autre population bovine, le *Peranakan Ongole* ou *Filial Ongole* dont 90 % des effectifs se situent sur l'île de Java, ainsi que la race Aceh qui serait le résultat d'une absorption de la race Madura et de 11% de Banteng [Mohamad et al., 2009]. Les îles de Madura et de Bali auraient été préservées des importations massives de zébu Ongole au 19^{éme} siècle : ainsi nous retrouvons des populations bovines indigènes présentant une proportion plus importante de Banteng [Mohamad et al., 2009]. La race Madura, présente dans la partie orientale de l'Indonésie, serait donc un témoignage direct du métissage ancien *Zebu X Banteng* d'il y a 1 500 ans [Porter and Tebbit, 2007, Sutarno and Setyawan, 2016].

Des ovins

L'histoire des ovins est moins documentée que pour les bovins. Il est probable que les ovins se soient déplacés vers l'Est à partir de l'Asie occidentale, en passant par l'Asie centrale vers -8 000 [Taylor et al., 2021] puis ont poursuivi vers l'Ouest autour de 5 000 - 5 700 ans av J-C emmenés par des éleveurs pastoraux [Aldenderfer, 2011, Cai et al., 2011, Dodson et al., 2014, Lv et al., 2015, Jeong et al., 2016, Zhao et al., 2017]. Ces populations se seraient étendues aux régions au Nord du plateau tibétain concomitant avec les invasions Mongols au Moyen Orient au 13^{iéme} siècle, obligeant des populations musulmanes de la péninsule Arabe, Perse, d'Asie Centrale à migrer vers l'Est de l'Asie [Grousset, 1941], (Song and Wand 1983 in [Zhao et al., 2017]). Ces dernières auraient introduit des moutons de type "long-fat-tailed" dans le Nord de la Chine.

Une hypothèse de deux vagues d'expansion directe vers l'Inde en passant par la terre et par la mer d'Arabie est émise et étayée par les analyses mitochondriales : l'hypothèse d'une origine directe à partir des espèces sauvages du genre *Ovis*, comme *O. vignei* et *O. orientalis* venues des régions montagneuses d'Afghanistan est avancée [Acharya, 1982] et confirmée par les récentes études génomiques datant l'arrivée des populations ancestrales d'il

y a environ 4 000 ans par voie terrestre [Pardeshi et al., 2007, Lv et al., 2015] et de type "longue queue" [Fuller, 2006] puis en passant par l'Arabie Saoudite ou la mer d'Arabie [Singh et al., 2013]. [5] Actuellement, on retrouve dans la région du Balouchistan des moutons à queue grasse, probablement issus de la deuxième vague [Fuller, 2006], encore non datée. Les moutons à poils sont encore les races prédominantes dans le sud de l'Inde [Ryder, 1984] et sur la côte sud et centrale de l'Orissa [Fuller, 2006].

Dans l'histoire récente, Bell (1928) in [Banerjee et al., 2011] raconte que des marchands venant du Bhoutan faisaient du commerce d'ovins et de laine (en sus de la soie et du sel) avec la Chine à travers le Tibet au cours du 19^{éme} siècle. Il est fort probable que cette migration ovine soit essentiellement constituée de béliers, potentiels meilleurs porteurs que les brebis, qui pouvaient ralentir les caravanes à cause des mises-bas. Une fois arrivée dans la région du Cooch Behar (Région Indienne entre le Bhoutan et le Bangladesh) ou de Rangpur (Nord du Bangladesh), les ovins étaient vendus aux marchands locaux, intéressés car l'élevage ovin était peu prégnant dans la région (Hunter 1875 in [Banerjee et al., 2011]), et transportés par bateau vers les régions méridionales du Bengal (correspondant au Bangladesh et au Bengal occidental) et de la région de Calcutta (William Milburn 1813 in [Banerjee et al., 2011]). Ces moutons étaient probablement abattus par les marins ou embarqués à bord des navires à destination de Java pour y être échangés, donnant ainsi le mouton Javanais à queue fine [Bradford and Inounu, 1996].

En Indonésie les deux types d'ovins à queue fine et grasse se côtoient. A l'origine, il semblerait que des ovins de type "queue fine" [Udo and Budisatria, 2011] originaires de la région indienne du Bengale et du Bangladesh [Banerjee et al., 2011] auraient été introduits probablement entre les 7^{iéme} et 12^{iéme} siècle, période durant laquelle le commerce entre l'Inde et l'Indonésie, était florissant [Wormser, 2017], notamment par la côte de Sumatra [Grant, 1972]. Plus récemment, les commerçants arabes auraient introduit des ovins de type "queue grasse" originaire de l'Asie du Sud-Ouest. En 1731, le gouvernement sous administration néerlandaise a importé des mâles de race Kiramni de Perse, une race de type "queue grasse" . Plus tard, en 1754 et 1779, bien que des subventions aient été proposées pour les importateurs d'ovins, il n'est pas certain que ces importations aient été effectives [Mason and others, 1980, Udo and Budisatria, 2011]. En 1802, des moutons Africander de type "queue grasse" originaires de l'Afrique du Sud ont également été importés (Mason 1978 in [Udo and Budisatria, 2011]). Des importations d'ovins de type européen notamment de race Mérinos venant d'Australie ont également eu lieu en 1860, 1897, 1903 et 1906. Ces ovins ont été croisés dans la région de Garut sur l'île occidentale de Java avec des Africanders et avec des moutons de race locale de type "queue fine". D'autres races de type européen comme la Romney Marsh (en 1912 et 1914) [Mason and others, 1980], la Texel, la Suffolk, et des hybrides Dormer (Dorset X Merino)

5. Les ADN mitochondriaux de la lignée A sont les plus fréquents en Asie : les fréquences des lignées présentes chez les moutons indiens s'élèvent pour A à 84%, pour B à 15% et pour C à 1% [Pardeshi et al., 2007, Singh et al., 2013]. Les auteurs proposent la présence d'un lieu de domestication de la lignée A en Inde et que les lignées B et une partie du A ont été introduites à partir du proche Orient par la mer plus tardivement.

et Suffmer (Suffolk X Merino) ont été importées massivement et ont eu inévitablement une influence sur la structure génétique des populations ovines locales [Udo and Budisatria, 2011]. Dans le contexte de crise économique des années 1920, les petits ruminants ont bénéficié d'un regain d'intérêt : les ovins à queue grasse sont devenus bien établis sur l'île de Madura depuis 1938, et ont ensuité été diffusés vers Java où ils se sont croisés avec les ovins locaux à queue fine donnant ainsi les ovins Javanais à queue grasse [Sutama, 1990, Sodiq and Tawfik, 2004, Udo and Budisatria, 2011]. Une autre race ovine, le Garut ou Priangan est également issue d'un croisement entre les ovins locaux à queue fine avec du Merinos et des moutons à queue grasse [Udo and Budisatria, 2011]. Ils présentent un dépôt de graisse à la base de la queue, ils sont toutefois classés comme mouton à queue fine [Porter et al., 2016]. Ceux présentant une robe noire sont utilisés lors de combat de béliers.

1.2.3 Arrivée sur le continent Africain

Des bovins

Les premières introductions de bovins en Afrique se seraient déroulées en 2 vagues principales d'animaux de type *Bos taurus* : les taurins à longues cornes (ou longhorns) sans bosse pourraient avoir été introduits en premier, il y a 7 000 ans, suivis par les shorthorns sans bosse environ 2 500 ans plus tard [Epstein, 1971, Mukasa-Mugerwa, 1989, Rege and Tawah, 1999, Rewe, 2013]. Cependant, ce scénario des deux vagues d'arrivée des taurins en Afrique n'est encore étayé par aucune preuve moléculaire. Le bétail africain *Bos taurus* se serait étendu du nord-est du continent vers l'ouest et l'est de l'Afrique [Hanotte, 2002, Mwai et al., 2015]. La composante de variation génétique supplémentaire habituellement détectée chez les bovins taurins africains peut s'expliquer par l'hybridation avec les aurochs locaux en Afrique après la domestication des bovins taurins dans le Croissant fertile [Pitt et al., 2019].

Les zébus ont été introduits en Afrique après leur domestication dans la vallée de l'Indus en Inde [Mwai et al., 2015] via une première vague datant d'il y a 4 000 ans via la Mésopotamie et l'Egypte [Payne and Wilson, 1999, Gifford-Gonzalez and Hanotte, 2011, Edea et al., 2017a]. La présence de zébus indiens à Oman la plus ancienne daterait d' il y a 2 600 - 2 000 ans avant notre ère (Uerpmann 2001 in [Beaujard, 2019c] puis les zébus auraient ensuite atteint progressivement la partie méridionale de l'Afrique et se serait métissés avec les taurins déjà présents [Mwai et al., 2015]. L'introduction de zébu indien en Afrique se serait déroulée à cette même période [Boivin and Fuller, 2009], ainsi que des figurines de zébus datant de 600 ans avant J-C trouvées à Sembel sur la côte de l'Erythrée - Yemen (Schmidt 2006 in [Beaujard, 2019c] p.575). L'ascension de la civilisation éthiopienne de Daamat et plus tard de Axum aurait contribué à la diffusion de plantes et d'animaux asiatiques (zébus, dromadaires, poules) en Afrique notamment par l'intermédiaire des marchands originaires de l'empire Perse qui contrôlait une partie de l'Asie de la vallée de l'Indus jusqu'au Levant ([Beaujard, 2019c] p.575). Des preuves supplémentaires de la présence de bovins à bosse à partir de la moitié du I^{ier} millénaire (an 500) dans la corne de l'Afrique de l'Est ont été rapportées (Russell

2007 in [Beaujard, 2019c]). Selon l'étude de données génomiques, l'origine des populations indigènes actuelles de zébu de la Corne de l'Afrique de l'Est serait le résultat d'un métissage entre des taurins africains et des zébus indiens, issus d'une deuxième vague [Mwai et al., 2015], qui ce serait déroulé il y a 750-1050 ans au début du II^{iéme} millénaire [Kim et al., 2020]. Ce métissage aurait été possible grâce à une vague supplémentaire de zébus indiens datant d'il y a 700 ans par les sources historiques, parallèlement à l'islamisation des la côte Est-Africaine et l'émergence de la culture Swahili [Epstein, 1971, Boivin et al., 2013, Mwai et al., 2015, Beaujard, 2019a]. Les zébus auraient ainsi été «confinés» au niveau du littoral des côtes Est-Africaines pour au moins 2-3 siècles avant de largement se métisser avec les taurins africains en s'enfonçant dans les terres aboutissant ainsi à la structure génétique actuelle de ces populations indigènes de la corne de l'Afrique de l'Est [Kim et al., 2020], jusqu'en Afrique de l'Ouest, le long de la ceinture sahélienne. Cette introgression est retrouvée chez les zébus ouest-africains actuels [Gautier et al., 2009, Flori et al., 2014]. Les derniers événements de métissage auraient eu lieu il y a environ 500 ans pour les zébus Fulani et 140 ans pour la race Borgou [Flori et al., 2014]. Ce laps de temps entre l'arrivée des zébus et les premiers métissages peut s'expliquer par le fait que l'activité d'élevage de bovins ne semble avoir pris une place prépondérante qu'à partir de la fin du Ier millénaire (Horton 1996, Kusimba 1999 in [Beaujard, 2019c]).

Des ovins

Concernant l'arrivée des populations ovines en Afrique, elle s'est produite en deux lieux : par la partie nord-Est de l'Afrique et la corne de l'Afrique en accompagnant les éleveurs pastoraux [Hanotte, 2002, Boivin and Fuller, 2009, Gifford-Gonzalez and Hanotte, 2011, Muigai and Hanotte, 2013, Deng et al., 2020]. Les moutons étaient présents dans la corne de l'Afrique bien avant l'arrivée et la dispersion des bovins à bosses [Clark and Williams, 1978], c'est-à-dire avant la moitié du 1er millénaire après JC (ou an 500) [Gifford-Gonzalez and Hanotte, 2011], côtoyant ainsi des bovins de type "taurins". Vraisemblablement, les ovins n'auraient pas suivi la dispersion des zébus le long de la ceinture saharienne vers l'Afrique de l'Ouest. Il semblerait que plusieurs vagues d'introduction se soient déroulées dans la partie Nord-est de l'Afrique puisque des représentations de mouton à queue grasse et à queue fine ont été retrouvées en Egypte : une seconde vague d'introduction de moutons à queue grasse datant d'il y a 3 002 - 2 644 ans [Clutton-Brock, 1993] et auraient remplacé ou se seraient introgressé dans les populations ovines de type "queue fine" déjà présentes. Les récents résultats génétiques ont montré une histoire différentes entre ces deux types de moutons [Aswani, 2007, Muigai and Hanotte, 2013, Deng et al., 2020]. Les moutons auraient atteint le Sud de l'Afrique à la fin du 3^{iéme} millénaire avant le présent [Sealy and Yates, 1994] et atteignant le Cap Occidental il y a 2 000 ans [Smith, 2009].

Un questionnement cependant reste quant à la route de migration vers la partie méridionale de l'Afrique : soit en partant de l'Afrique occidentale le long de la côte atlantique par le bassin du Congo ou plus sûrement, en partant de l'Afrique orientale et en passant par l'Afrique

australe par la vallée du Rift. Il semblerait d'ailleurs qu'il y ait eu une cohabitation des deux groupes d'ovins. La répartition géographique actuelle des ovins en Afrique se décline ainsi :

— les moutons à queue fine se trouvent principalement en Afrique du Nord et en Afrique occidentale : deux sous-groupes se distinguent entre les moutons associés à des systèmes pastoralistes où les animaux présentent de longues pattes, un phénotype adapté aux longs déplacement dans les régions arides et semi-arides du sahel et de l'Afrique occidentale, et les moutons "nains" mieux adaptés aux climat tropicaux des régions humides d'Afrique occidentale et centrale [Meyer et al., 2004].

— les moutons à queue grasse se trouvent également en Afrique du Nord puis dans les parties australes et méridionales du continent africain, de l'Erytrée à l'Afrique du Sud. Ce type de mouton est subdivisé en deux sous-types par rapport au type de la robe : les moutons à poils et à laine [Rege and Tawah, 1999, Meyer et al., 2004, Gifford-Gonzalez and Hanotte, 2011, Meyer, 2020].

1.3 Histoire des populations bovines et ovines de l'Ouest de l'Océan Indien

La zone autour de l'Océan Indien est particulièrement intéressante puisqu'elle a été le théâtre d'échanges commerciaux entre différents peuples et cultures [Beaujard, 2019c, Beaujard, 2019a]. Des périodes d'intensification des échanges alternent avec des périodes de régression dues bien souvent à des évènements politiques et environnementaux ayant des impacts à moyen et long terme et modifiant le système commercial [Beaujard, 2015, Brucato et al., 2017, Beaujard, 2019c, Beaujard, 2019a]. Les traces d'échanges entre l'Inde, l'Afrique et Madagascar les plus anciennes datent du II^{iéme} millénaire avant JC et concernent notamment des plantes bien que l'arrivée d'agriculteurs soit attestée dès le 1^{ier} siècle [Beaujard, 2015].

Des transferts de bétail aux Comores et à Madagascar pourraient avoir eu lieu suite aux mouvements de populations bantoues d'Afrique et à la navigation austronésienne d'Indonésie [Fuller et al., 2011, Brucato et al., 2017, Brucato et al., 2018, Brucato et al., 2019, Beaujard, 2019a]. Les premières introductions par des populations Bantous [Boivin et al., 2013, Beaujard, 2019c] de bovins appartenant au rameau *Bos taurus* [Porter et al., 2016] sur la zone des îles de l'Ouest de l'Océan Indien sont datées d'entre le 1^{éme} et 5^{éme} siècle à Madagascar dans un premier temps puis à partir du 7^{éme} aux Comores, bien que l'hypothèse de l'arrivée de populations bovines métissées *Taurin X Zébu* ne soient pas à exclure [Boivin et al., 2013]. Durant cette période, l'archipel de Lamu et Zanzibar occupait une place dominante dans les échanges vers des contrées lointaines notamment avec la Chine [Beaujard, 2015], en plus des contacts "locaux" entre les populations Bantoues des Comores et les Austronésiens venus directement d'Indonésie ([Beaujard, 2019c], p.642). Puis à partir du 7^{éme} siècle, la formation et l'interconnexion des deux empires Tang et musulman sont à l'origine d'un essor significatif des échanges par la terre et la mer, accompagnés d'une mise en place de

la culture Swahili sur la côte Est [Beaujard, 2015]. Les plus anciennes preuves archéologiques et squelettiques de la présence de bétail à Mayotte et à Madagascar remontent aux 8^{éme}-10^{éme} siècles après J.-C. [Allibert et al., 1989, Cheke, 2010, Rasamuel, 1988, Rafolo, 1989, Magnavita, 2006, Boivin et al., 2013, Anderson et al., 2018]. Cette période est marquée par un déclin du système de commerce de l'Océan Indien et par une première avancée des commerçants Arabes et Swahili vers les Comores et Madagascar aboutissant à la mise en place du corridor Swahili, connu pour être un réseau reliant la côte Tanzanienne, les Comores et le Nord Ouest de Madagascar [Beaujard, 2015, Beaujard, 2019a], et permettant la rencontre entre les locuteurs swahilis et austronésiens [Beaujard, 2015, Brucato et al., 2017, Brucato et al., 2018, Brucato et al., 2019], bien que les contacts entre les premières populations Bantoues et les Austronésiens ne soient pas à exclure ([Beaujard, 2019c] p.642). A partir du 10^{éme} siècle, les découvertes archéologiques confirment la présence de bovins type "zébu" à Madagascar et aux Comores [Boivin et al., 2013, Porter et al., 2016], probablement arrivés d'Inde avec les marchands Arabes, Shirazi [Blanchy, 2015] et Swahili et dans une moindre mesure Indiens et Chinois [Beaujard, 2019a] en passant par l'Afrique via le cabotage [Mwai et al., 2015]. L'hypothèse d'une arrivé de zébus "purs" est défendue [Mwai et al., 2015] mais il est possible que des individus déjà métissés *(Taurins africains X zébus)* aient été introduits dans la zone. Une autre origine hypothétique est par l'arrivée des populations Austronésiennes sur la côte Est d'Asie du Sud-Est mais semble écartée par les archéologues [Beaujard, 2019a]. L'arrivée des Austronésiens est supposée continue à partir de la fin du 1er millénaire avant J-C, notamment avec le développement de nouvelles techniques de navigation [Boivin et al., 2013], jusqu'à l'arrivée des Européens soit via l'Inde ou via les Maldives et Chagos [Beaujard, 2019c, Beaujard, 2019a]. Par la suite, l'élevage se serait développé comme le suppose l'augmentation des quantités d'os de bovins identifiés attestant d'une augmentation du cheptel bovin Mahorais du 14^{éme} au 15^{éme} siècle de notre ère [Pauly, 2014, Boivin et al., 2013]. A partir du 16^{éme} siècle, la zone est le théâtre de différents épisodes de colonisations : les Sakalavas, un peuple malgache, effectuent des raids vers l'archipel Comorien et raflent des esclaves. Les Portugais, premiers colons européens abordent l'archipel en 1505, puis les Français en 1529, faisant par la suite d'Anjouan un port de relâche sur la route maritime vers le Golfe Persique, les Indes ou l'Extrême Orient. Selon plusieurs écrits, le bétail domestique était mentionné aux Comores par la quasi-totalité des visiteurs européens des 17^{éme} et 18^{éme} siècles ; l'observation de bovins sauvages à Mayotte en 1646 et la "chasse au bétail à bosse" en 1729 ont également été rapportées (le journal de Robert Drury dans [Cheke, 2010]).

Vers la fin du 19^{éme} siècle, le continent Africain a connu une forte épidémie de peste bovine [Spinage, 2003] ayant décimé une large proportion du cheptel bovin. Concernant les îles de l'Ouest de l'Océan Indien, Madagascar serait restée exempte tandis que la Réunion aurait vu son cheptel décimé entre 1900 et 1902 [Mandret, 2000] ainsi qu'une des îles des Comores en 1901 [Barrett et al., 2006]. Cette épidémie a été suivi d'une nouvelle introduction massive de zébus en provenance d'Inde [Mwai et al., 2015] afin de reconstituer le cheptel en Afrique de l'Est.

L'histoire récente des populations bovines est beaucoup plus documentée surtout à Madagascar puisque durant la période coloniale des recherches ont été menées sur le zébu malgache [Ribot, 2005]. A Madagascar au début du 20^{éme} siècle, différentes races ont été importées dans des stations expérimentales afin de tester l'adaptabilité de race à haute valeur productive telle que la Holstein, le Brahman, la Brune des Alpes. Deux races ont ainsi été créées par croisement avec le zébu malgache : la Renitelo et la Manjan'i Boina [Porter et al., 2016].

La population de zébu mahorais n'a pas fait l'objet du même intérêt des vétérinaires coloniaux de Madagascar, Mayotte étant rattaché administrativement à Madagascar jusqu'en 1946 puis uniquement aux autres îles Comoriennes jusqu'en 1975. De plus, les archives départementales de Mayotte sont incomplètes du fait de la destruction d'une partie du fond documentaire en 1993. Néanmoins quelques éléments peuvent être rapportés : il semblerait qu'une prise de conscience de la préservation de la population bovine soit apparue en 1966 puisque une loi promulgait l'interdiction d'abattage des jeunes bovins et des vaches en état de se reproduire (JO des Comores du 16 Décembre 1966, p.519). En 1988, un rapport (Anonyme 1988) mentionnait l'origine des zébus locaux mahorais comme étant la résultante d'un croisement entre le zébu malgache, le zébu de l'Afrique de l'Est et de taurins métropolitains importés en petit nombre de race brune des Alpes et Pie noire. Plus récemment, à Mayotte, des flux d'importations de ressources génétiques bovines européennes ont eu lieu à partir des années 90 (Convention 1991) suite à l'instauration de l'insémination artificielle par le service élevage de l'île de Mayotte. Les taureaux proposés étaient de race Montbéliarde. Plus tard, en 2008, 80 vaches de race montbéliarde ont été importées en vif. En 2018, plusieurs vagues d'importations ont été organisées avec notamment des Jersiaises, des Brunes des Alpes et des Gasconnes.

Les îles de la Réunion, Maurice et d'Amsterdam présentent une histoire bovine plus récente. Aux 17^{éme} et 18^{éme}, des zébus malgaches auraient été introduits sur l'île de la Réunion [Porter et al., 2016], ainsi que 357 têtes de bétail de décembre 1845 à décembre 1846. Du fait d'une meilleure conformation, des zébus (au nombre de 590) venant de Bali ont été également introduits dans cette même période (Bavoux 1994 in [Mandret, 2000]). De même, des reproducteurs de races anglaise (Durham) et française ont été importés en 1862 afin d'améliorer le cheptel réunionnais [Mandret, 2000]. Ces efforts n'obtiendront pas le succés escompté [6], d'autant plus que la grande épizootie de peste bovine décima le cheptel bovin de 1900 à 1902 [Mandret, 2000]. A l'île Maurice, une première introduction de zébus venant de Madagascar a été effectuée par les Hollandais aux 16^{éme} - 17^{éme} siècle. Une deuxième vague d'introduction de zébu venant de Madagascar semble s'être déroulée à l'initiative cette fois des colons Français au début du 19^{éme} dans un premier temps pour un rôle de traction dans la culture de canne à sucre [Porter et al., 2016], en parallèle de taurins, dont l'origine est encore inconnue et destinés à la production laitière. Il semblerait que les individus typés zébus aient

6. Il faudra attendre la départementalisation de l'île en 1946 pour voir un réel développement de l'élevage basé sur des races européennes.

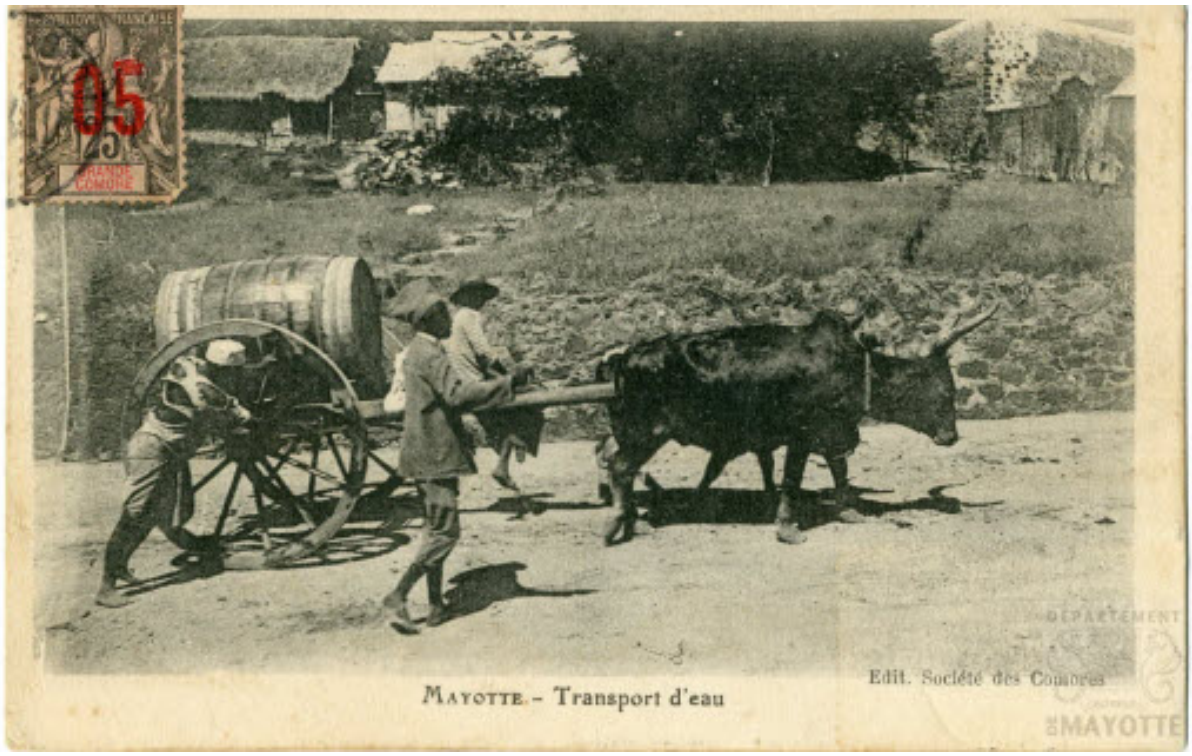


FIGURE 1.3 – Utilisation du zébu à Mayotte comme outil de traction (Carte postale - Société des Comores - 1910)

été éradiqués de l'île ne trouvant plus leur utilité dans la traction agricole au cours du 20^{éme} siècle. Les taurins de la race Créole de Maurice subistent encore mais en petits effectifs, ceci étant dû à l'absorption avec des races européennes productives [Meyer, 2020].

Les zébus de la Réunion, Madagascar, Maurice, Comores appartiennent au groupe "Malagasy zebu" selon [Porter et al., 2016]. Le zébu de Madagascar est classé dans le groupe "Small East African Zebu" selon la base de données DAGRIS de l'ILRI.

L'utilisation actuelle des races locales est dépendante du contexte socio- économique du territoire et de l'importance accordée à la valeur patrimoniale des animaux locaux. Ainsi, si auparavant les zébus à Mayotte servaient comme outil de traction, leur rôle s'est restreint autour de la constitution d'une épargne transmissible aux enfants, les mâles étant abattus pour les mariages religieux. A Madagascar, l'élevage de zébus dans la partie occidentale de l'île est de type extensif en grand troupeau [Souvenir, 1991], mais, compte-tenu du faible niveau de vie des agriculteurs malgaches, les zébus sont encore utilisés comme outil de traction surtout dans les zones de cultures rizicoles dans la région des Hautes Terres. A l'île de la Réunion par contre, avec la reconnaissance du zébu Moka en 2016 (arrêté du 13 avril 2016), l'APPER[7] met en place des opérations de communication et de promotion de la race et de son rôle passé d'animal de traction, mais aussi son rôle à venir pour la gestion des savanes réunionnaises [Rivière, 2021].

L'histoire des populations ovines autour des îles de la zone est encore moins documentée.

7. Association pour la Protection du Patrimoine et de l'Ecologie à la Réunion



FIGURE 1.4 – Utilisation du zébu Moka de la Réunion lors de la récolte de la canne à sucre.(source www.reunionnaisdumonde.com)

Toutefois quelques informations ont pu être synthétisées : aux Comores, la population ovine est de type "à poil" et "à queue grasse" et se rapprochant du type East African Fat Tailed [Meyer et al., 2004] ou de type Djallonké à queue fine et poil (DAGRIS) [Saido, 2005, Meyer et al., 2004]. La robe est assez variable : marron unie ou tachetée de blanc, pie noir. L'implantation de la graisse est uniforme de la base de la queue à la pointe. La tête est de type convexe, les oreilles pendantes [Saint Martin, 1988].

La présence la plus ancienne mentionnée des moutons remontent au 11^{éme} siècle à Mayotte [Wright et al., 1984]. Il est probable que, compte tenu des échanges réguliers entre les îles de la zone, des flux géniques entre l'archipel des Comores et Madagascar aient eu lieu.

La présence de moutons à queue grasse est mentionnée en 1950 [Anonymous, 1950]. L'origine de cette population était inconnue mais cependant contemporaine du zébu malgache. Présente dans les région du Sud et des Hauts plateaux, elle était confondue avec le type *Ovis dolichura* [Anonymous, 1950, Anonymous, 1952] qui caractérisait les races ovines de Crimée [Rozier, 1800]. Ce type était décrit comme intermédiaire entre une race d'Asie à queue courte et grosse *Ovis statopygea* et celle dite du Soudan à queue longue et tombante *Ovis africana*. La description phénotypique (chanfrein convexe, sans corne, robe pie-noir, pie-roux) rappelle néanmoins les moutons à tête noire de Tanganyika et d'Ouganda (Mason and Maule 1960, DAGRIS) ou les Black Head Persian [Anonymous, 1952] ou d'Egypte [Dechambre, 1951]. Dechambre [Dechambre, 1951] signale l'importation par un certain Laborde, probablement Jean Laborde, consul de Madagascar et à l'origine d'importation de races éxotiques sur l'île-



FIGURE 1.5 – Bêlier photographié à Madagascar en 1920. On note la robe pie avec la partie foncée concentrée sur l'avant de l'animal, le profil convexe de la tête ainsi que la queue grasse bien développée. (Source : Anonyme - Photothèque du Cirad)

continent [Deschamps, 1961], qui aurait également importé des moutons d'Egypte.

Des essais de croisement avec des Mérinos d'île de France pour les Hauts plateaux et avec des Merinos d'Afrique du Sud pour la région Sud ont également été entrepris dans la première partie du 20^{éme} siècle. Le succès n'aurait pas été au rendez-vous à part pour les merinos d'Arles qui auraient été plus résistants aux conditions environnementales de Madagascar [Anonymous, 1952]. Actuellement, les moutons sont présents dans les zones sud et occidentale de Madagascar avec une forte concentration (80% des effectifs) dans la Province de Tuléar (Sud).

1.4 Conclusion du chapitre

Les races locales bovines et ovines résultent d'une histoire démographique complexe, comme nous venons de le détailler. L'histoire adaptative de ces races a également eu un impact sur leur diversité génétique. En migrant sur les différents continents et au sein de chaque continent, les animaux ont en effet dû également s'adapter à des conditions environnementales différentes et être exposés à diverses contraintes environnementales (par exemple, l'exposition à de nouvelles conditions climatiques, à de nouveaux agents pathogènes ou la modification des ressources alimentaires et hydriques disponibles), auxquelles ils ont du s'adapter. D'autres contraintes exercées par l'homme se sont également ajoutées à ces contraintes environnementales. Dans l'Océan Indien, l'histoire des populations bovines et ovines étroitement

liée aux échanges et métissages de nombreuses populations humaines, a abouti à la présence de races locales originales dans cette zone, dont la diversité génétique reste à étudier.

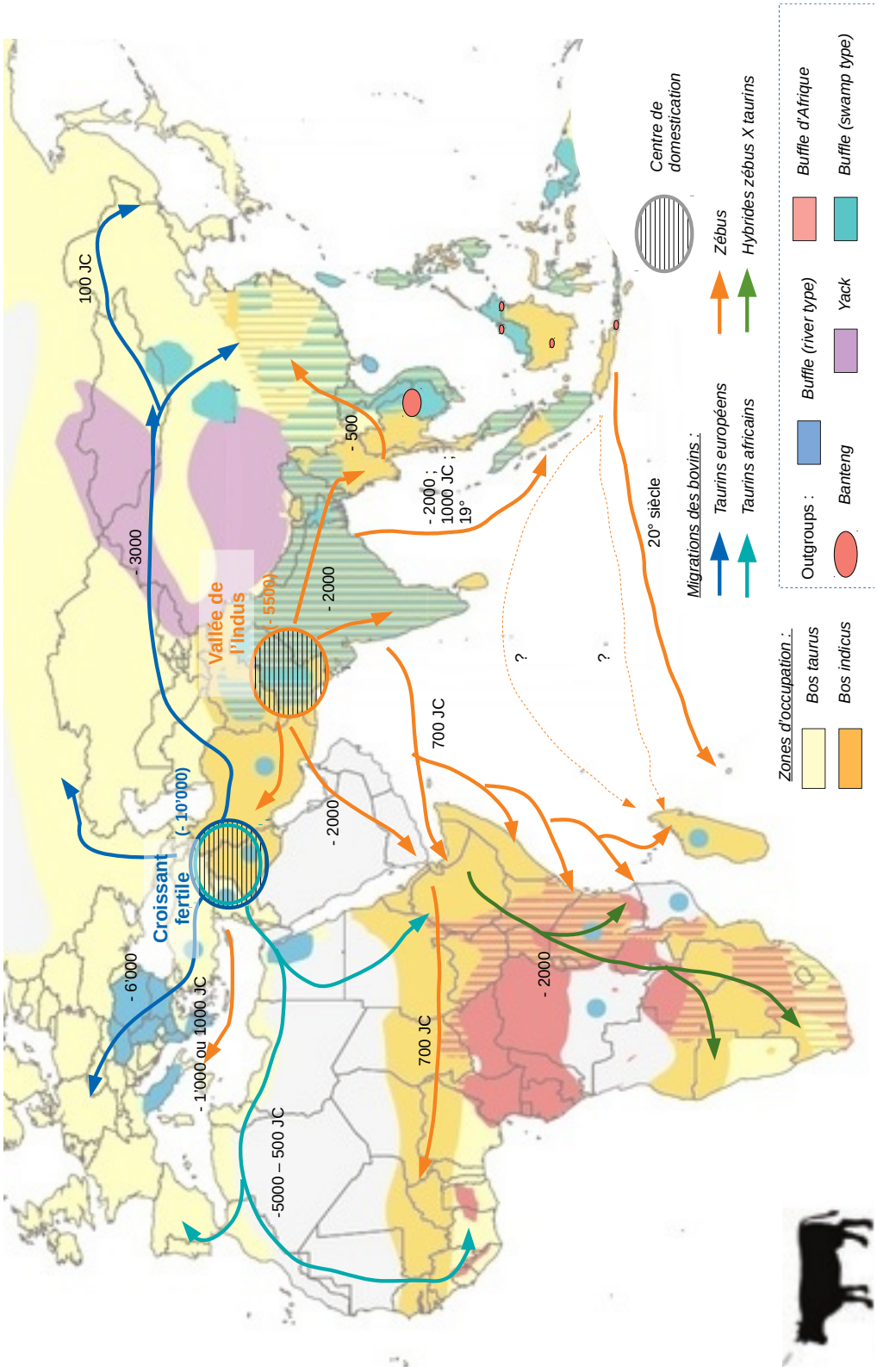


FIGURE 1.6 – Hypothèses d'expansion et de migration des genres *Bos taurus taurus* et *Bos taurus indicus* à partir des centres de domestications à partir de données archéologiques, historiques et génétiques. (Figure adaptée de [Zhang et al., 2020])

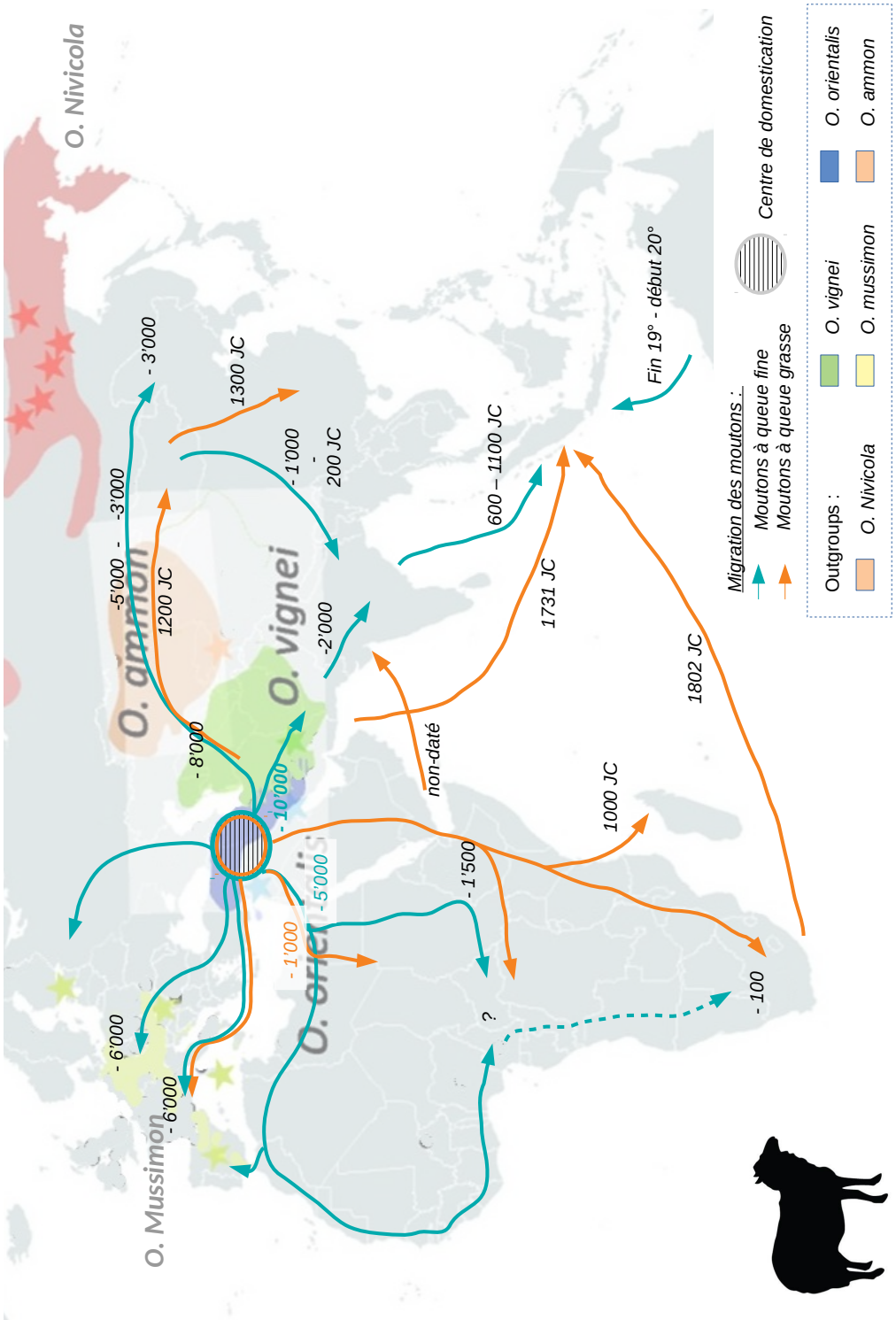


FIGURE 1.7 – Hypothèses d'expansion et de migration du genre *Ovis aries* à partir des centres de domestications à partir de données archéologiques, historiques et génétiques. (Figure issue de [Cao et al., 2021])

Chapitre 2

Caractérisation de la structure de la diversité génétique des races locales

Avant toute étape d'exploration de la diversité génétique, il est important de caractériser phénotypiquement les populations locales afin de sélectionner au mieux les animaux à génotyper. La caractérisation de la structure de la diversité génétique des races s'effectue par des approches exploratoires (multi-dimensionnelles de type ACP ou de classification hiérarchique non-supervisée ou supervisée) appliquées à des données de génotypage individuel. Cette caractérisation permet de suggérer ou d'éliminer des hypothèses sur l'origine des races. Pour étayer ces hypothèses, ces approches descriptives doivent ensuite être complétées par l'inférence de paramètres démographiques (caractérisation de la dynamique démographique et datation des principaux évènements historiques notamment) à l'aide de modèles démo-génétiques. Les caractérisations phénotypique et génétique participent au processus de conservation et de valorisation des races [Felius et al., 2011].

2.1 De la caractérisation phénotypique à la caractérisation génétique des races

Dans un cadre de conservation des ressources zoogénétiques, la caractérisation phénotypique permet tout d'abord de distinguer ou non une population animale en race en fonction de caractères observables et est une condition préalable avant d'évaluer la diversité des ressources et de déterminer le niveau d'érosion génétique. A partir de là, les inventaires nationaux des ressources zoogénétiques sont possibles, ainsi que la mise en place de projets de gestion des populations animales. La caractérisation phénotypique a été formalisée assez récemment, en 2012, par la FAO qui a produit un guide méthodologique afin de standardiser les procédures de collecte de données [FAO, 2012] dans le but d'alimenter un système d'information général des ressources génétiques animales, le DAD-IS (*worldwide Domestic Animal Diversity Information System*). Selon la FAO, "la caractérisation phénotypique des ressources zoogénétiques correspond à l'identification des différentes races et à la description de leurs caractéristiques externes et de production dans un environnement et un cadre de gestion donnés, en tenant compte des facteurs socio-économiques qui les affectent." [FAO, 2012].

Les données nécessaires à collecter durant l'étape de caractérisation phénotypique sont les suivantes :

- la position géographique des troupeaux,
- un inventaire numérique des individus,
- les caractéristiques phénotypiques (description des animaux, mesures de variation des caractères productifs notamment),
- les caractéristiques phénotypiques de réponse aux stress environnementaux (climat, ressources alimentaires, hydriques, pathogènes),
- le milieu naturel, les conditions d'élevage et toutes les connaissances locales sur les stratégies de conduite d'élevage utilisées pour utiliser la diversité génétique du bétail,
- des informations sur l'origine et le développement de la race. Cette étape peut deman-



FIGURE 2.1 – Quelques mesures corporelles effectuées sur des zébus et des moutons de Mayotte : Mesure **a** de la hauteur au garrot, **b** du périmètre thoracique, **c** de la longueur de la tête, **d** de la largeur aux épaules. (Crédits : **a,b** Mélissa Ouvrard, **c,d** Jessica Magnier)

der des recherches bibliographiques dans des archives, des enquêtes de territoire,

— les liens éventuels avec d'autres races à l'extérieur et à l'intérieur du pays afin de déterminer les risques ou non d'absorption génétique.

Une fois ces données collectées et les analyses effectuées, il est possible d'établir un standard de la race et de regrouper des éleveurs intéressés par une gestion plus formelle de la population.

L'abondance et la baisse des coûts de génotypage incite désormais à coupler caractérisation phénotypique et génétique, cette dernière pouvant être un outil de distinction des races. La caractérisation génétique permet aussi d'apprécier, par l'estimation le niveau d'hétérozygotie et de consanguinité, le niveau de risque de disparition de la population et de proposer des pistes d'amélioration de la gestion de la race. De plus elle permet d'éviter des erreurs de gestion de population. En effet, si on prend l'exemple de la Senepol [Flori et al., 2012], originaire des Caraïbes (île de St Croix) et introduite dans d'autres pays (Brésil, Australie notamment), la caractérisation génétique à partir de données SNP de moyenne densité a permis d'établir que ces animaux n'étaient pas issus d'un croisement entre taurins européens et africains, comme les données historiques le laissaient supposer. En effet, l'étude a mis en évidence un métissage *taurins européens X Zébus* avec une proportion d'ascendance indicine d'un peu plus de 10 % en moyenne. Or, compte tenu de la sensibilité des zébus (et des taurins européens) à la Trypanosomiase Animale Africaine (TAA), contrairement aux taurins africains qui sont trypanotolérants, l'introduction d'animaux de race Senepol dans des zones où la prévalence de la maladie TAA est élevée (en Afrique de l'Ouest) serait un facteur d'échec de l'élevage de cette race.

2.2 Outils et méthodes de caractérisation de la diversité génétique des races bovines et ovines

2.2.1 Les outils de production de données génomiques

La compréhension des histoires démographiques et des relations entre les populations est facilitée par l'accroissement récent de données génétiques disponibles dans les espèces d'élevage (bovine et ovine notamment) grâce aux nouvelles méthodes de génotypage et de séquençage à haut débit.

Séquençage

Le premier séquençage du génome de référence bovin (Btau_4.0) [Elsik et al., 2009] a été effectué en 2009 par le Bovine Genome Sequencing and Analysis Consortium en combinant deux assemblages préalablement obtenus par le séquençage Sanger [Consortium et al., 2009]. L'individu utilisé était une vache de race Hereford donc d'ascendance taurine. Depuis plusieurs assemblages ont été produits, les principaux étant UMD_3.1/bosTau6 [Zimin et al., 2009], et sa version suivante UMD_3.1.1/bosTau8, avec 138 contigs supprimés.

Avec les développement des méthodes Next Generation Sequencing permettant le séquençage d'un nombre plus conséquent d'échantillons, les côuts d'analyse d'ADN ont chuté drastiquement et de plus en plus d'espèces et d'individus ont pu être séquencés. Ainsi, les premiers génomes de référence pour le genre *Bos indicus* ont été produits en Février 2012, puis 2018. Ils étaient tous deux basés sur les assemblages taurins (Btau 4.0 et UMD 3.1.1) durant l'étape d'alignement des reads. Le séquençage NGS a aussi permis d'établir des assemblages *de novo*[1], permettant ainsi l'établissement d'un assemblage prenant en compte l'ascendance indicine [Koren et al., 2018], en séquençant un individu F1 issu d'un métissage entre un taureau de race Angus et une vache de race Brahman. Plus récemment, le séquence NGS a permis d'obtenir le dernier assemblage de référence ARS-UCD1.2/bosTau9 [Rosen et al., 2020].

Concernant les ovins, le premier assemblage ISGC Ovis_aries_1.0/oviAri1 obtenu après un séquençage avec la méthode Sanger a été produit en 2010 par l'International Sheep Genomics Consortium en se servant de l'assemblage bovin comme référence pour l'alignement des reads [Consortium et al., 2010]. Cet assemblage a été suivi de deux autres ISGC Oar_v3.1/oviAri3 et Oar_v4.0/oviAri4 [Jiang et al., 2014].

A noter que le développement et la mise au point des technologies de troisième génération, ou long read, ouvrent encore de nouvelles possibilités permettant de progresser en terme de qualité de séquençage des génomes.

Le séquençage des génomes complets de différentes races bovines et ovines s'est accompagné du développement d'outils de génotypage à faible, moyenne et haute densité.

1. Assemblage de novo : organisation des segments séquencés pour un individu en ensembles de segments successifs qui sont ensuite mis bout à bout pour constituer le génome entier, cette opération complexe suppose une grande quantité de séquences obtenues avec différentes méthodes.

Outils de génotypage

Les méthodes NGS développées dans les années 2000 ont permis aussi d'établir de nouvelles méthodes et outils d'analyse de variation nucléotidique pour des milliers de marqueurs répartis sur tout le génome, les SNPs (Single Nucleotide Polymorphism). Ces types de marqueurs représentent la variation (polymorphisme) d'une seule paire de bases du génome entre individus d'une même espèce et permettent une avancée, par rapport aux microsatellites utilisés jusque là en génétique des populations, dans la détermination des variants nucléotidiques en augmentant la densité de marqueurs afin de mieux caractériser la diversité génétique neutre. Les marqueurs SNPs ont été déterminés à partir de données de séquences haut débit. Ainsi, pour l'espèce bovine, différentes puces sont disponibles via la technologie Illumina (LD 7K (6 909 SNPs), medium 54K (54 609 SNPs), HD 777K (777 962 SNPs)) ainsi que la technologie Affymetrix (Axiom BOS 1 (648 851 SNPs), Axiom Bovine v1 (51 988 SNPs), Axiom Bovine v2 (67 641 SNPs), Axiom Bovine v3 (63 988 SNPs)). Pour l'espèce ovine, la puce Ovine Illumina SNP50 BeadChip (54 241 SNPs) est la plus couramment utilisée. Il existe aussi une puce HD de la technologie Illumina et une puce moyenne densité relevant de la technologie Affymetrix (Axiom Ovine (51 000 SNPs).

Des biais de sélection de marqueurs sont observés pour certaines puces de génotypage. Les marqueurs SNP inclus dans la première puce de génotypage, "Illumina Bovine SNP50 Genotyping BeadChip" ont été sélectionnés en se basant uniquement sur des marqueurs polymorphes pour les taurins européens [Matukumalli et al., 2009]. Par conséquent, les individus génotypés sur cette puce et relevant de la sous-espèce *Bos indicus* et issus de métissages *B. indicus* X *B. taurus* présentent des hétérozygoties plus faibles (Figure 2.2). Ce biais a également un impact dans la construction des arbres phylogénétiques [Decker et al., 2009] et des analyses multidimensionnelles [Consortium et al., 2009]. En ce qui concerne la puce "Illumina Bovine HD", le niveau de biais est plus modéré que pour les puces 50K de chez Illumina [Utsunomiya et al., 2019] (Figure 2.2). D'autres puces commerciales sont également disponibles et permettent de résoudre le biais pour les populations non-taurines. Leur particularités ont été comparées par [Utsunomiya et al., 2019]

La situation est différente pour la puce Ovine SNP50K développée en 2005 [Kijas et al., 2009] et en prenant en compte un large nombre de populations ovines présentes autour du globe. Sa validation a été opérée sur des populations ovines africaines. Une comparaison des différentes puces ovines pourrait aussi être conduite.

Les données disponibles

Pour les bovins, le dernier assemblage de référence est le ARS-UCD1.2/ Btau9 [Rosen et al., 2020]. Cependant nos données de génotypage stockées dans la base de données WIDDE [Sempéré et al., 2015] ont été issues du précédent assemblage UMD3.1/ Btau6 [Zimin et al., 2009]. Il sera donc nécessaire d'effectuer une conversion des positions. L'étude est basée sur des données SNP obtenues par le génotypage sur les puces Illumina Bovine SNP50K Genotyping

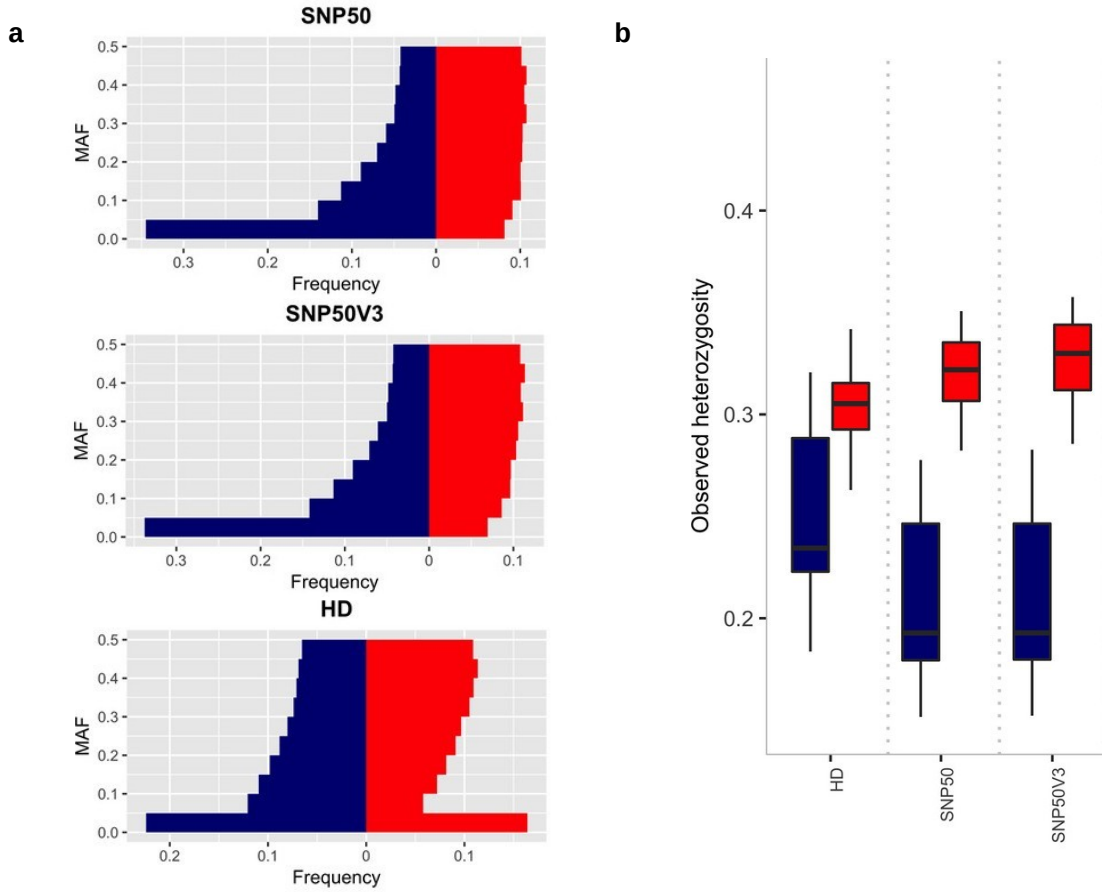


FIGURE 2.2 – **a** Distribution de la MAF pour les zébus (blue) et les taurins (rouge) au sein de 3 puces SNP disponibles commercialement ; **b** Hétérozygotie observée pour le zébu (bleu) et les taurins (rouge) (issus de [Utsunomiya et al., 2019])

BeadChip pour la moyenne densité et sur les puces Illumina Bovine SNP-HD Genotyping BeadChip. Au vu des biais identifiés pour ces deux puces, il s'agira dans la suite de l'étude de les prendre en compte dans nos conclusions.

Pour les ovins, le dernier assemblage de référence est le Oar_v4.0/oviAri4, une version publiée par l'International Sheep Genomics Consortium (`www.sheephapmap.org`, version améliorée de l'assemblage Oar_v3.1/ovAri3 [Jiang et al., 2014]. Néanmoins, nos données de génotypage présentes dans la base de données WIDDE relèvent de l'assemblage Oar_v3.1/ovAri3 [Jiang et al., 2014]. Une conversion sera nécessaire en mettant à jour les positions des SNPs obtenues sur le site SNPchimp [Nicolazzi et al., 2015].

2.2.2 Estimation de la diversité génétique

Mesure de la différentiation génétique

Les statistiques F proposées par Wright représentent un outil de description de la diversité génétique inter et intra-population [Wright, 1949].

Le F_{i_s} mesure l'écart à la panmixie (déficit d'hétérozygote) dû à la consanguinité du

régime de reproduction au sein de la population. Il est égal à la différence entre l'hétérozygotie observée Ho et de l'hétérozygotie attendue He (sous les conditions de Hardy Weinberg) sur He. Il varie entre 0 (si conformité à la panmixie) et 1

Le F_{S_T} ou indice de fixation mesure le degré de différentiation entre populations divergeant uniquement sous l'effet de la dérive. Il varie entre 0 (aucune différentiation génétique entre les populations) et 1 (différentiation extrême entre populations). Il mesure le déficit en hétérozygotes dû à la structuration de la population selon l'estimation proposée par [Weir and Cockerham, 1984]. Le calcul des Fst peut être utilisé pour la construction de matrice de distances génétiques.

Analyse exploratoire de l'admixture

Une première exploration de la structure génétique de la population est réalisée par des méthodes multi-dimentionnelles comme l'Analyse en Composantes Principales (ACP) et des méthodes de classification hiérachiques (non supervisées et supervisée) appliquées à des données de génotypage individuel.

Analyse en composantes principales (ACP) L'ACP est une technique algébrique qui utilise la décomposition des vecteurs propres en vue de réduire des jeux de données de plusieurs dimensions en un petit nombre de variables indépendantes, les composantes principales, tout en expliquant une bonne partie de la variation présente dans le jeu de données initiales. Le but est de résumer la variation de données génétiques, les génotypes, multidimensionnelles (=nombre de marqueurs). En règle générale, l'ACP est appliquée sur le jeu de données regroupant les génotypes de plusieurs individus. Les premières composantes sont ensuite représentées graphiquement afin de visualiser les distances génétiques entre les individus. Le logiciel utilisé fréquemment sur les données génétiques est le logiciel Smartpca [Patterson et al., 2006].

Classification Les algorithmes de clustering dit "model-based" fonctionnent en trouvant le meilleur ajustement pour les données génétiques (génotypes des individus constituant les populations) parmi un nombre arbitraire ou mathématiquement déterminé de clusters, de sorte que les différences au sein des clusters soient minimisées et que les différences entre les clusters soient maximisées. Cette méthode de regroupement est également appelée "admixture inference", car les génomes individuels (ou les individus au sein des populations) peuvent être caractérisés par les proportions d'allèles liées à chaque groupe. En d'autres termes, ces algorithmes génèrent des résultats qui supposent l'existence de populations ancestrales distinctes s'étant combinées au fil du temps pour former les populations métissés actuelles. Le logiciel ADMIXTURE [Alexander et al., 2009] estime simultanément les fractions individuelles d'ascendance Q et les fréquences alléliques des populations F. L'estimation de l'ascendance est plus précise dans les analyses supervisées puisqu'il y a moins d'incertitude par rapport aux fréquences alléliques. Les résultats issus des méthodes de clustering ne sont en partie qu'informatifs dans un cadre de recherche d'histoire des populations. Ces méthodes sont basées

sur des approximations qui peuvent être inappropriées dans certaines applications. Le K est souvent choisi arbitrairement, tout en privilégiant les valeurs pour lesquelles la différentiation génétique la plus élevée est capturée dans les clusters. Concernant les cas de métissage, les clusters ne doivent pas être interprétés comme reflétant les populations ancestrales.

En règle générale, les analyses exploratoires (ACP et clustering) ne permettent pas de conclure sur la réalité d'un scénario démographique. En effet, certains scénarios historico-démographiques très divers peuvent aboutir à des motifs de diversité similaires [Novembre and Stephens, 2008]. Pour caractériser la dynamique démographique et dater les principaux événements historiques, il est ensuite nécessaire d'utiliser des méthodes d'inférence de paramètres démographiques.

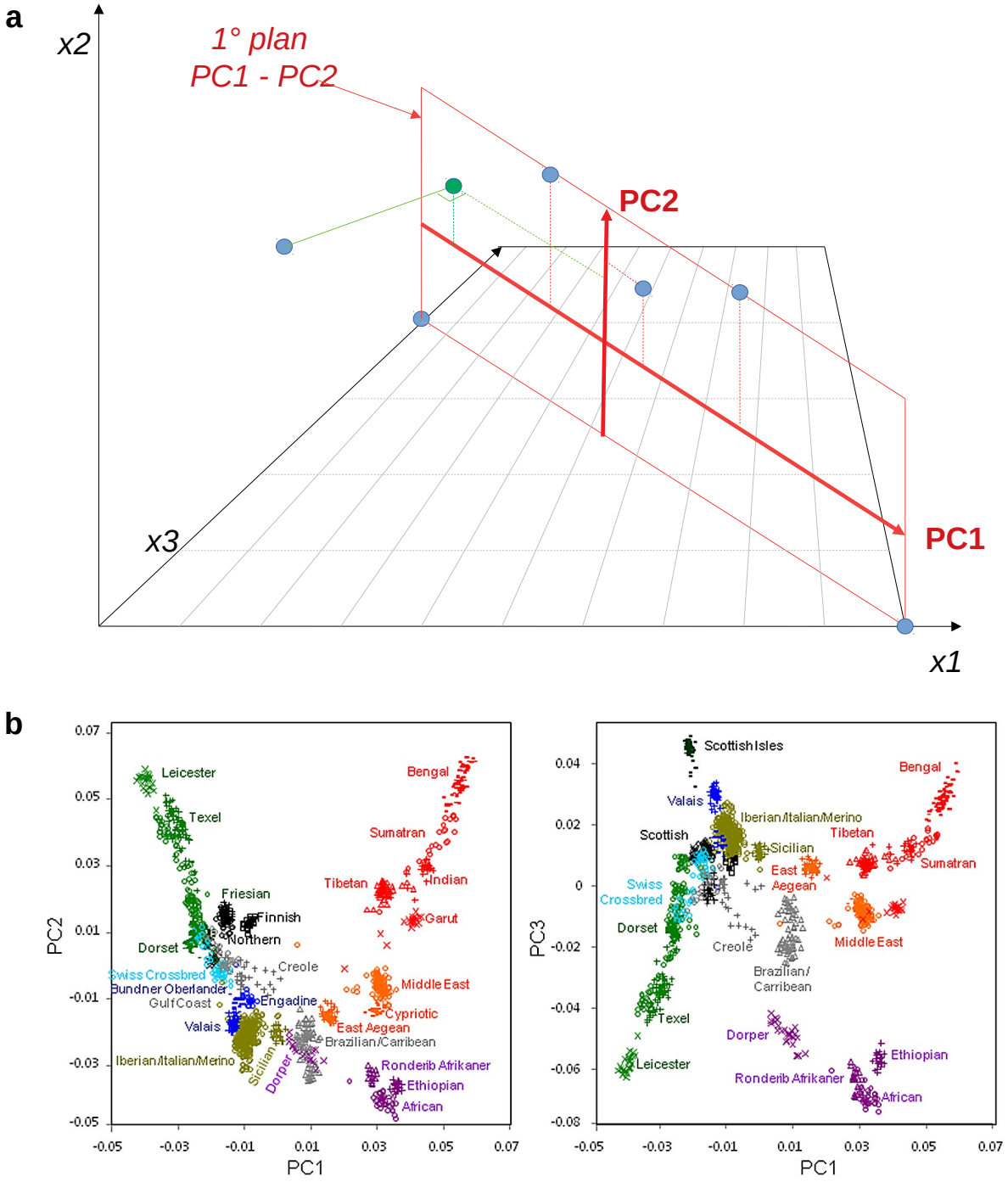


FIGURE 2.3 – Méthode de l'Analyse en Composante Principale. **a** Les points bleus représentent les individus représentés pour 3 variables (*x1, x2, x3*) Le point vert représente la projection de l'individu le plus à gauche sur le 1° plan PC1 - PC2 (en rouge). **b** Un exemple de résultats d'ACP pour les deux premiers plans PC1-PC2 et PC1-PC3 pour des populations ovines échantillonnées dans différentes régions et génotypées sur une puce Illumina ovine 50K (Figure issue de [Kijas et al., 2012]).

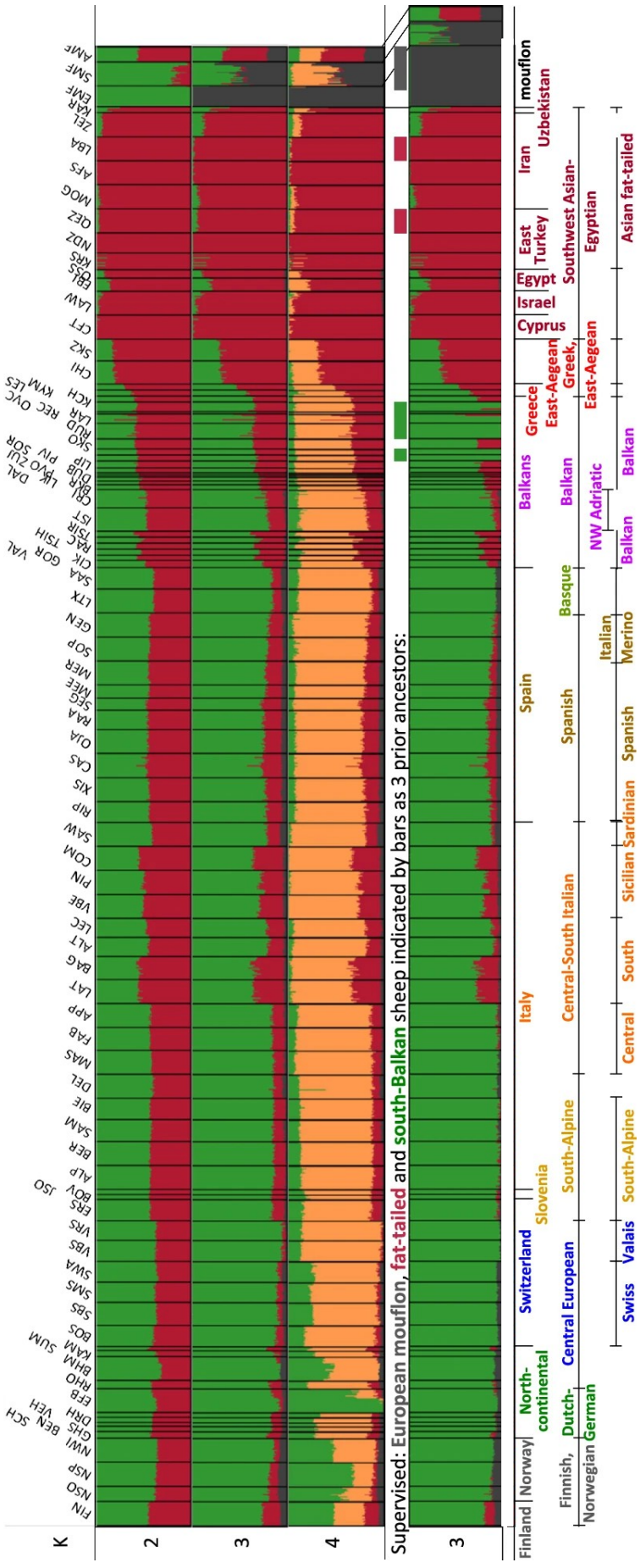


FIGURE 2.4 – Exemple de résultats graphiques obtenus après utilisation du logiciel ADMIXTURE sur un jeu de données constitué de plusieurs populations ovines européennes génotypées sur puce Illumina ovine 50K (21 960 SNPs restants après filtration) et un outgroup constitué de trois populations de mouflons. Pour les 3 premiers graphiques, l'analyse a été conduite en non-supervisé. Le dernier graphique correspond aux résultats obtenus en mode supervisé avec trois ascendances présumées (Mouflon européen, moutons à queue grasse, moutons des Balkans). La figure est tirée de [Ciani et al., 2020].

2.2.3 Inférence démographique

Consanguinité récente

Dans les ensembles de données de génotypage ou de séquençage, des segments contigus homozygotes d'une séquence d'ADN [Gibson et al., 2006] peuvent être présents. Ces segments contigus sont autozygotes ou homozygotes par filiation (HBD), résultant de l'héritage de deux copies d'un seul segment chromosomique provenant d'un ancêtre (Figure 2.5) [Wright, 1922] par filiation. Ces segments sont aussi appelés Runs of Homozygosity (ROH). Les méthodes d'étude des ROH permettent de déterminer l'inférence démographique [MacLEOD et al., 2009, Palamara et al., 2012, Druet and Gautier, 2017, Peripolli et al., 2017] de populations et sont basées sur la détermination de ces segments contigus, et plusieurs ont été proposées.

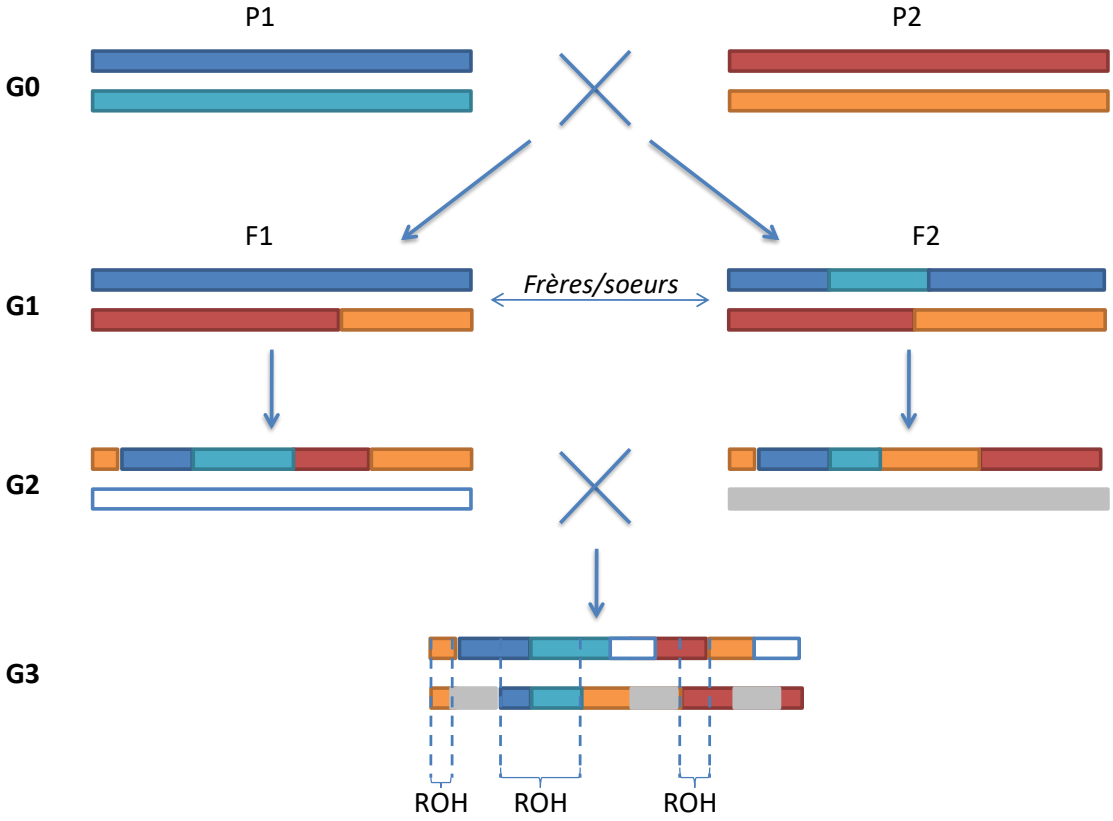


FIGURE 2.5 – Illustration de la formation de ROH après 3 générations successives.

Les premières méthodes basées sur des règles [McQuillan et al., 2008, Kirin et al., 2010] considèrent que seuls les tronçons de génotypes répondant à certains critères prédéfinis (par exemple, le nombre de marqueurs, le nombre de génotypes hétérozygotes et manquants, la densité de marqueurs, la longueur de la fenêtre, l'espacement des marqueurs) sont des segments HBD, comme cela est proposé par le logiciel Plink [Purcell et al., 2007]. Ces critères doivent être optimisés pour chaque ensemble de données en fonction des caractéristiques de la population étudiée ou de la technologie de génotypage utilisée.

Un autre type d'approches basées sur des modèles de Markov cachés (HMM) et reposant sur des tests de rapport de vraisemblance pour classer des fenêtres de marqueurs consécutifs

comme étant des segments HBD ou non HBD ont été proposées [Wang et al., 2008, Pemberton et al., 2012, Druet and Gautier, 2017]. Ces approches permettent de tenir compte des fréquences alléliques des marqueurs et des erreurs de génotypage, ce qui rend la décision moins sensible au biais de vérification des marqueurs. Dans le logiciel RZooroh [Druet and Gautier, 2017, Bertrand et al., 2019], le modèle utilisé décrit un génome individuel comme une succession de segments classés en classes multiples HBD, au nombre de K-1 et définies par la longueur attendue des segments, et une classe non-HBD, étiquetée K. La longueur des segments HBD de la classe k est distribuée exponentiellement avec un taux Rk, c'est-à-dire qu'elle est égale à 1/Rk Morgans. Les classes HBD à faible taux (segments HBD longs) correspondent aux segments HBD hérités d'ancêtres communs récents, tandis que celles à taux élevé correspondent à l'autozygotie associée à des ancêtres communs anciens. Le taux d'une classe est approximativement égal à deux fois le nombre de générations jusqu'aux ancêtres communs associés à cette classe.

Les modèles avec plusieurs classes de HBD offrent ainsi la possibilité de révéler l'histoire démographique récente en partitionnant les segments de HBD en différentes classes liées à l'âge et d'estimer la contribution des différentes générations passées à l'autozygotie actuelle [Druet and Gautier, 2017] : à l'échelle individuelle, les contributions élevées des classes de HBD très récentes suggèrent que les parents sont fortement apparentés (Figure 2.6.c). À l'échelle de la population, les contributions importantes d'une classe à l'autozygotie indiquent une taille de population effective *(Ne)*[2] plus faible à la période correspondante, et pourrait être éventuellement associée à un goulot d'étranglement ou à un effet fondateur. Inversement, une faible contribution à l'autozygotie suggère une grande *Ne* (Figure 2.6.b). Ces résultats pourront être confirmés par des méthodes d'estimation de variation de la *Ne*.

2. **Taille efficace d'une population *Ne*** : une mesure statistique du nombre de copies d'un gène en circulation dans une population (contribuant au patrimoine génétique), exprimée en termes de nombre d'individus contribuant génétiquement. Ce paramètre peut être défini comme le nombre d'individus qu'une population aurait eu si elle avait été une population idéale de taille constante N, présentant des générations ne se chevauchant pas, un sex-ratio équilibré, le même nombre de descendants pour chaque parent, pas de sélection et l'accouplement entre les individus serait aléatoire. Dans les populations réelles, les individus contribuent différemment au patrimoine génétique pour un certain nombre de raisons. Par conséquent, la taille effective de la population est généralement inférieure à la taille du recensement.

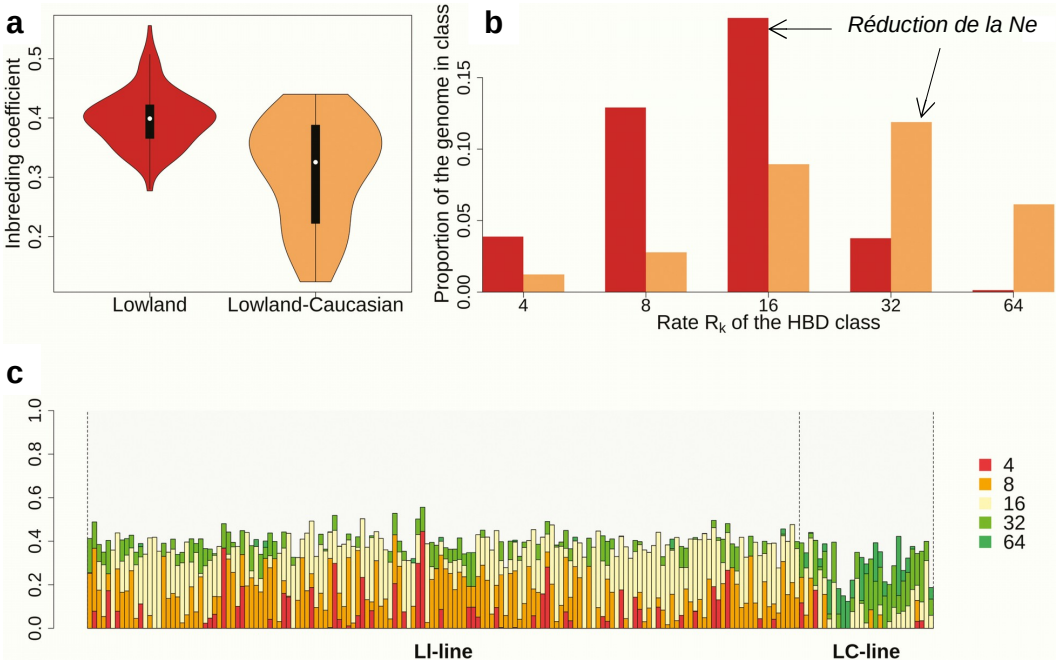


FIGURE 2.6 – Principales sorties graphiques du logiciel RZooroh ayant calculé les segments HBD pour le bison Européen [Druet et al., 2018]. **a** Distribution de la consanguinité globale sur la population. **b** Distribution des segments HBD en fonction de la classe HBD. **c** Partitionnement de la consanguinité en classe HBD pour chaque individus de la population.

f statistics

Les statistiques résumant le partage d'allèles ou statistiques f (f statistics ; [Reich et al., 2009, Patterson et al., 2012]) sont largement utilisées pour réaliser des inférences sur la phylogénie et le métissage des populations à partir de données génétiques. Les applications de ces statistiques permettent de tester si des populations sont phylogénétiquement liées, si une population résulte du métissage entre plusieurs populations ancestrales et dans ce cas estimer les contributions des différentes populations à la population cible, si les métissages et les divergences entre populations peuvent expliquer la démographie ou encore de déterminer la population la plus proche d'une population actuelle ou ancienne [Peter, 2016]. Ces statistiques mesurent les corrélations entre les fréquences alléliques parmi des ensembles de deux, trois ou quatre populations et l'espérance théorique des statistiques f peut être estimée en terme de longueur de branches et de proportions de métissage d'un arbre de métissage [Peter, 2016, Lipson, 2020]. La statistique f2 calculée pour deux populations A et B représente la différence de fréquence moyenne des allèles entre ces deux populations au carré ($F2(A,B) = E[(a'-b')]$) et correspond à la longueur de la branche de l'arbre phylogénétique reliant les deux populations A et B. Dans le cas de trois populations A, B et C, la longueur de la branche reliant les trois populations est égale à $F3(C;A,B) = E[(c'-a')(c'-b')]$, a', b' et c' étant les fréquences alléliques des trois populations A, B et C pour un SNP.

La statistique f4(A, B ; C, D) mesure la corrélation moyenne des différences de fréquences alléliques entre les populations A et B, d'une part et les populations C et D, d'autre part

(c'est à dire $a-b$ et $c-d$, pour les fréquences alléliques a, b, c et d généralement moyennées sur de nombreux SNP). Si les quatre populations sont phylogénétiquement liées (arbre non enraciné (A, B),(C, D)), alors la valeur attendue du f4(A, B ; C, D) sera nulle, alors que les valeurs attendues de f4(A, C ; B, D) et f4(A, D ; B, C) seront positives. Cette statistique est similaire à la statistique D (avec un facteur de normalisation ; voir plus loin).

Si une population C résulte d'un métissage entre les sources C' et C'' dans les proportions d'ascendance α et $(1-\alpha)$, alors en espérance $f4(A,B;C,D) = \alpha.f4(A,B;C',D) + (1-\alpha).f4(A,B;C'',D)$. Dans le cas du métissage, la valeur de f4 attendue est la longueur de la branche multipliée par la proportion du mélange qu'on peut estimer par le calcul du f4 ratio :

$$\alpha = \frac{f_4(A,B;C,D) - (1-\alpha).f_4(A,B;C'',D)}{f_4(A,B;C',D)}$$

en simplifiant :

$$\alpha = \frac{f_4(A,B;C,C'')}{f_4(A,B;C',C'')}$$

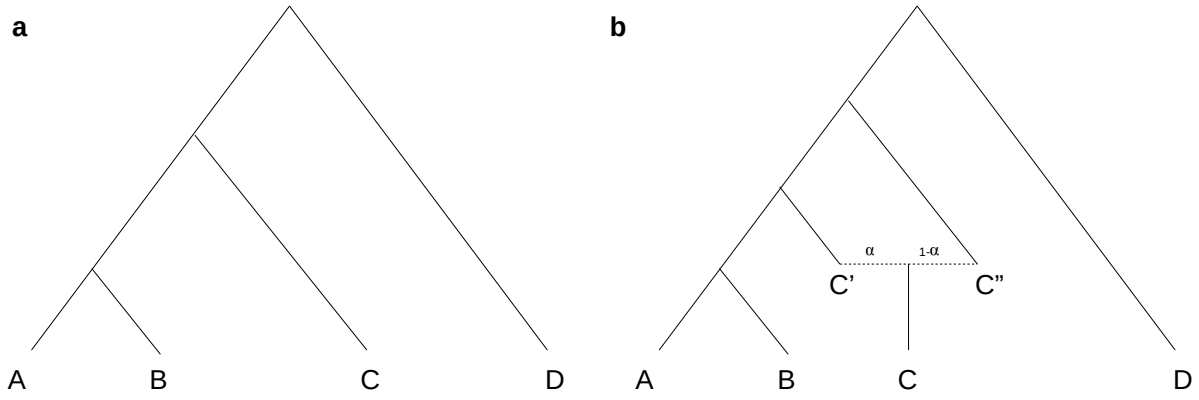


FIGURE 2.7 – Arbres phylogéniques reliant les populations A, B, C et D dans le cas où **a** aucun métissage n'a lieu, et dans le cas où **b** un évènement de métissage a lieu entre les populations C' et C'' résultant dans la population C.

Le test D-statistics également nommé le test ABBA-BABA ou "four-population test" est très similaire au test "f4-statistic" [Reich et al., 2009, Reich et al., 2010, Durand et al., 2011, Patterson et al., 2012] et permet de tester le métissage à partir de quatre populations [Reich et al., 2009, Reich et al., 2010, Durand et al., 2011]. L'objectif de cette statistique est de tester l'hypothèse nulle d'absence de flux génique entre les populations A et C (Figure 2.7.a), c'est-à-dire que pour un marqueur donné, les populations B et C portent l'allèle dérivé noté *B*, les populations A et D portent l'allèle ancestral noté *A* (schéma ABBA), ou, les populations A et C portent l'allèle *B* et les populations B et D portent l'allèle *A* (schéma BABA). Sous cette hypothèse d'absence d'admixture, les probabilités d'observer les deux schémas sont équivalentes. Ainsi la D-statistique évalue la corrélation moyenne des différences de fréquences alléliques en fonction des schémas ABBA et BABA selon le calcul suivant :

$$D = \frac{Pr(BABA) - Pr(ABBA)}{Pr(BABA) + Pr(ABBA)}$$

$$D = \frac{E[(c-d)(a-b)]}{E[(c+d-2c.d)(a+b-2a.b)]}$$

avec les fréquences alléliques a, b, c et d des populations A, B, C et D. Dans le logiciel ADMIXTOOLS [Patterson et al., 2012], si le Z-score (correspondant à la D statistiques standardisée par l'erreur standard issue du calcul blockjacknife [3]) est positif, un flux génique est détecté entre A et C ou B et D. Si le Z-score est négatif, un flux génique est détecté entre A et D ou B et C. La D-statistique est calculée sur des données à l'échelle du génome (séquences, SNP) et nécessite des données issues de 2 populations, de la population à tester et d'une population "outgroup".

Graphes d'admixture

Les statistiques f [Reich et al., 2009, Patterson et al., 2012] peuvent être utilisées pour reconstruire des graphes d'admixture, c'est à dire des arbres de plusieurs populations impliquant des événements de métissage, à partir des données de fréquences. Les valeurs observées reflètent les degrés d'ascendance partagée et peuvent servir de moyen de tester des hypothèses concernant la hiérarchie de séparation des populations et les événements passés de flux de gènes. La méthode de construction des graphes d'admixture consiste à estimer les longueurs de branches et les proportions de métissage en résolvant un système d'équations dans lequel les valeurs des f-stats réprésentent les paramètres du modèle, si le classement des populations et les positions des évènements de métissage sont connus. Les méthodes les plus utilisées sont basées sur la moyenne et la variance de la divergence des fréquences alléliques : par exemple, le logiciel *TreeMix* [Pickrell and Pritchard, 2012] produit l'arbre le plus vraisemblable pour les populations échantillonnées et ensuite tente d'inférer un nombre d'évènements de métissage. Le principal inconvénient est que le programme construit le modèle en commençant par un arbre optimal sans admixture, puis en ajoutant des événements de métissage pour tenir compte des écarts entre les prédictions du modèle d'arbre et les données observées. En fonction de l'histoire réelle des populations, cette approche peut être fructueuse, mais peut aussi augmenter le risque de tomber dans les optima locaux imposés par l'arbre initial (surtout si de nombreuses populations sont mélangées). De plus, le choix du nombre d'événements de métissage à inclure, qui peut parfois être difficile à estimer, dépend de l'utilisateur.

Une autre méthode basée sur le calcul de la divergence des fréquences alléliques est *qpGraph* faisant partie de la suite ADMIXTOOLS [Patterson et al., 2012]. Une autre package s'inspirant d'Admixtools vient d'être déposé sur le CRAN : poolfstat [Hivert et al., 2018, Gautier et al., 2021]. La méthode est basée sur la définition préalable d'une topologie et de

3. **Blockjacknife :** Méthode de rééchantillonnage permettant de calculer des statistiques descriptives (moyenne, intervalle de confiance, écart-type) d'un jeu de données dont la distribution des individus est inconnue ou non-normale.

l'estimation des *f2, f3 et f4* statistiques entre toutes les paires, triplés et quadruplés comparé à la corrélation des fréquences alléliques des différentes populations.

Néanmoins les statistiques *f* utilisent seulement une partie de l'information contenue dans les données génétiques. Ce type de statistiques réduisent les données à une estimation de la distance génétique pour chaque paire de population. Les informations supplémentaires pertinentes pour l'histoire démographique qui ne sont pas utilisées par les statistiques *f* comprennent la distribution conjointe du nombre d'allèles entre les populations (AFS [4] ou SFS [5]) et les modèles de déséquilibre de liaison.

Datation des évènements de métissage

Actuellement les évènements les plus récents (de l'ordre de plusieurs centaines d'années) peuvent être détectés par les méthodes d'inférence de l'ascendance au niveau local ("local ancestry inference"), c'est-à-dire au niveau d'un segment chromosomique qui dérive d'une population ancestrale [Pool and Nielsen, 2009, Price et al., 2009, Sankararaman et al., 2012, Corbett-Detig and Nielsen, 2017]. En effet après un métissage, les chromosomes de la génération "fille-F1" contiennent des segments chromosomiques hérités des populations parentales. Au fur et à mesure des nouvelles générations, ces blocs se cassent par le phénomène de recombinaison [6], la longueur des segments chromosomiques ancestraux serait donc inversement proportionnelle à la date du métissage. Sous l'hypothèse d'un évènement de métissage unique dans un intervalle temporel court, et qu'ensuite les individus de la population métissée se reproduisent de façon aléatoire sur plusieurs générations, la probabilité que deux loci soient issus du même ancêtre au moment du métissage décroît exponentiellement avec un taux *r*, qui correspond au nombre de générations après la date du métissage. Ainsi, l'estimation de la datation des évènements de métissage est possible soit par l'estimation de la longueur des segments chromosomiques ancestraux [Pool and Nielsen, 2009, Gravel, 2012, Medina et al., 2018], soit par l'évaluation du taux de décroissance du déséquilibre de liaison [7] des alleles ancestraux de chaque SNP [Moorjani et al., 2011, Loh et al., 2013]. Ce dernier principe sera utilisé pour dater les métissages dans nos sujets d'étude au moyen du logiciel *Alder* [Loh et al., 2013] et de son extension *malder* pour les évènements de métissages multiples.

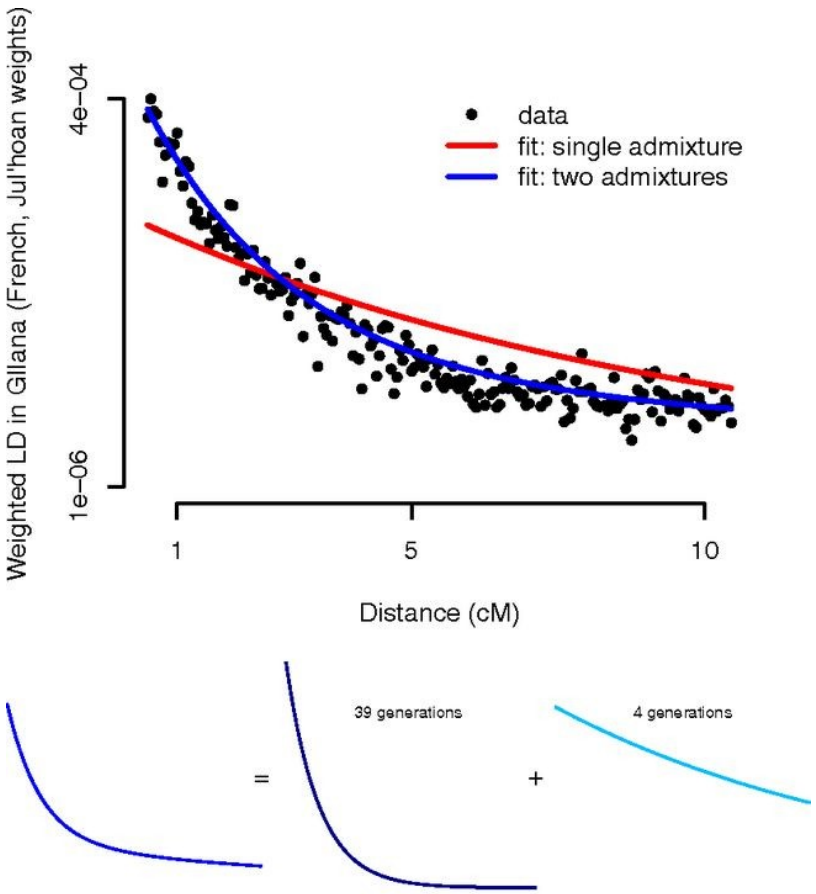
Alder Ce logiciel mesure la décroissance du Déséquilibre de liaison en fonction de la distance génétique (en Morgan) pour chaque paire de SNP bi-allélique, pondérée par la capacité de chaque locus de se différencier entre les deux populations ancestrales.

4. **AFS :** Allele Frequency Spectrum, distribution des fréquences alléliques de sites polymorphiques le long d'une séquence ADN au sein d'une population.

5. **SFS :** Site Frequency Spectrum = AFS

6. **Recombinaison :** Ensemble des processus qui se produisent au cours de la méiose et qui aboutissent à la production d'une progéniture présentant une combinaison de caractères différente de celle des deux parents. Les événements de recombinaison tels que les crossing over et la conversion de gènes peuvent entraîner des mutations.

7. **Déséquilibre de liaison :** situation où la fréquence des allèles de deux loci différents A et B est différente de ce que donnerait une association aléatoire de ces allèles. Il y a association préférentielle entre deux allèles.



Reference populations	Amplitude 1 (x10^{-4})	Amplitude 2 (x10^{-4})
Italian;Taa_West	**4.14**	0.954
Sardinian;Taa_West	**4.09**	0.97
Orcadian;Taa_West	**4.04**	0.962
Druze;Taa_West	**4.04**	0.957
Russian;Taa_West	**4.02**	0.972
Glui;Yoruba	2.99	**1.36**
Jul'hoan_North;Yoruba	2.81	**1.32**
Jul'hoan_South;Yoruba	2.81	**1.32**
Glui;Mandenka	2.92	**1.31**
Taa_North;Yoruba	2.93	**1.3**

FIGURE 2.8 – Principe du logiciel malder basé sur le calcul du LD pour plusieurs évènements d'admixture chez des populations humaines. Ici la courbe de DL a été calculé pour la population Gliana avec les populations Françaises et Jul'Hoan comme références. Deux modèles ont été testés : 1 avec un seul évènement d'admixture (rouge), le 2eme avec deux évènements d'admixture (bleu). La courbe bleue est constituée de deux autres courbes de LD, l'une en bleu foncé et la deuxième en bleu clair. Le tableau à droite montre les plus grandes amplitudes estimées sur chaque évènement de métissage et contribuant à chaque courbe en fonction de la couleur associée. L'illustration est issue de [Pickrell et al., 2014].

C'est-à-dire que ce logiciel calcule la covariance entre les 2 SNP en la pondérant par la différence de la fréquence allélique de ces 2 SNPs entre les deux populations ancestrales, les proportions α et $(1-\alpha)$ d'ADN issues des populations ancestrales, en lui appliquant une fonction exponentielle dépendante de la distance génétique et du nombre de générations depuis le dernier évènement de métissage. Puis, le logiciel recherche les meilleures paires de populations ancestrales en trouvant le meilleur modèle parmi toutes les populations ancestrales proposées.

malder Cette extension du logiciel *Alder* [Loh et al., 2013] permet d'inférer de multiples évènements de métissages avec plus de deux populations ancestrales. Dans cette situation, l'étendue du DL n'est plus seulement une seule courbe exponentielle fonction de la distance génétique mais un mélange de courbes exponentielles [Pickrell et al., 2014] (2.8).

Estimation des variations de taille efficace des populations

La taille efficace d'une population N_e [8] peut se déterminer à partir des données démographiques, des données de pedigree, ou des données génétiques. Les méthodes permettant de déterminer N_e à partir des données génétiques utilisant l'estimation du déséquilibre de liaison sont les plus efficaces.

L'un des enjeux des méthodes d'estimation de la taille efficace est la prise en compte des histoires complexes démographiques (goulot d'étranglement, réduction drastique des effectifs, combinés ou non, à plusieurs moments). Plusieurs approches permettant de les prendre en compte sont disponibles et utilisées pour le calcul de la *Ne* des espèces bovines (liste non-exhaustive) en se basant sur des modèles de coalescence [9] : le logiciel PSMC, par exemple [Li and Durbin, 2011] permet d'estimer les variations de *Ne* à partir de l'analyse de séquence d'un individu et a été utilisé dans la détermination de la taille efficace de populations taurines européennes et du bison européen [Gautier et al., 2016], tandis que le logiciel MSMC [Schiffels and Durbin, 2014], utilisé pour les populations bovines chinoises [Chen et al., 2018], permet d'estimer la *Ne* à partir de l'analyse de séquence de deux individus diploïdes.

D'autres méthodes basées sur la prise en compte de la totalité de l'échantillon ont été proposées : Le logiciel GONE [Santiago et al., 2020] détermine la taille efficace de la population à chaque génération N_g pour un temps récent ($<$ 100 générations) à partir du calcul du Déséquilibre de liaison LD pondéré par les variances génétiques des loci et est donc moins

8. **Taille efficace d'une population *Ne*** : une mesure statistique du nombre de copies d'un gène en circulation dans une population (contribuant au patrimoine génétique), exprimée en termes de nombre d'individus contribuant génétiquement. Ce paramètre peut être définie comme le nombre d'individus qu'une population aurait eu si elle avait été une population idéale de taille constante N, présentant des générations ne se chevauchant pas, un sex-ratio équilibré, le même nombre de descendants pour chaque parent, pas de sélection et l'accouplement entre les individus serait aléatoire. Dans les populations réelles, les individus contribuent différemment au patrimoine génétique pour un certain nombre de raisons. Par conséquent, la taille effective de la population est généralement inférieure à la taille du recensement.

9. **Coalescence** : modèle rétrospectif dont l'objectif est de suivre l'évolution de tous les allèles d'un gène donné jusqu'à l'ancêtre commun le plus récent

affecté par les erreurs de génotypage et les variants présentant une faible fréquence. Cette méthode permet d'estimer l'évolution de Ne sans hypothèse à fournir sur la magnitude et le type de changement dans la population. Une approche de block-jacknife [10] est utilisée pour estimer les intervalles de confiance des trajectoires de *Ne*. Un autre logiciel, PopSizeABC, est disponible [Boitard et al., 2016] et nécessite une conversion des jeux de données au format VCF et impose une étape de filtrage supplémentaire visant à éliminer les marqueurs pour lesquels la donnée est manquante. Il est basé sur le calcul par l'approche de calcul bayésien approximatif (ABC) de deux types de statistiques, le spectre des fréquences alléliques AFS et le déséquilibre de liaison moyen à différentes distances physiques, calculés à partir d'un échantillon et par des simulations de la variation de la taille de la population.

2.3 Application aux populations tropicales

Dans le cas de l'étude des populations bovines à l'échelle mondiale, l'utilisation des méthodes exploratoires (calcul de distances génétiques, ACP et classification hiérarchique) apporte des informations préliminaires sur la proximité génétique entre populations, permet d'émettre des hypothèse sur l'ascendance (taurine européenne, taurine africaine ou indicine par exemple pour les bovins) [Decker et al., 2014]. Ainsi, en ce qui concerne les bovins, le premier plan de l'ACP des données de génotypage (de plus de 40 000 SNPs) regroupant un panel de races représentatives de la diversité génétique mondiale permet de représenter la diversité génétique sous une forme triangulaire avec à chaque sommet du triangle les individus appartenant aux trois principaux groupes de races : les taurins européens (EUT : Hereford, Holstein, Angus), les taurins africains (AFT : Lagunaire, N'dama, Baoulé, Somba) et les zébus (ZEB : Gir, Nelore) (Figure 2.9). Sur chaque côté du triangle, entre chaque sommet se situent les populations métissées. Ainsi, les individus de race Santa Gertrudis (SGT), qui résulte d'un métissage synthétique créé dans les années 1910 entre des zébus Brahman et des taurins européens Shorthorn se positionnent entre les sommets EUT et ZEB, comme attendu. Entre les sommets EUT et AFT, nous retrouvons la population Oulmès Zaër dont la position intermédiaire est en adéquation le croisement entre le taurin Brun de l'Atlas et des N'Dama (AFT). Entre les sommets AFT et ZEB, se situent une grande majorité des zébus africains (en particulier en Afrique de l'ouest comme les zebu Fulani et Bororo), de l'est (East African Shorthorn zebu) et du sud (zébu de Madagascar) [Gautier et al., 2009, Mbole-Kariuki et al., 2014]. La population d'EASZ [Mbole-Kariuki et al., 2014], population autochtone de l'Ouest du Kenya, a été caractérisée génétiquement comme étant la résultante d'un metissage ancien entre des populations ancestrales proche des zébus indiens et des taurins Africains puis une introgression de taurins européens beaucoup plus récente et non-homogènement répartie entre les individus de la population (2% en moyenne).

10. **Blockjacknife :** Méthode de rééchantillonnage permettant de calculer des statistiques descriptives (moyenne, intervalle de confiance, écart-type) d'un jeu de données dont la distribution des individus estinconnue ou non-normale.

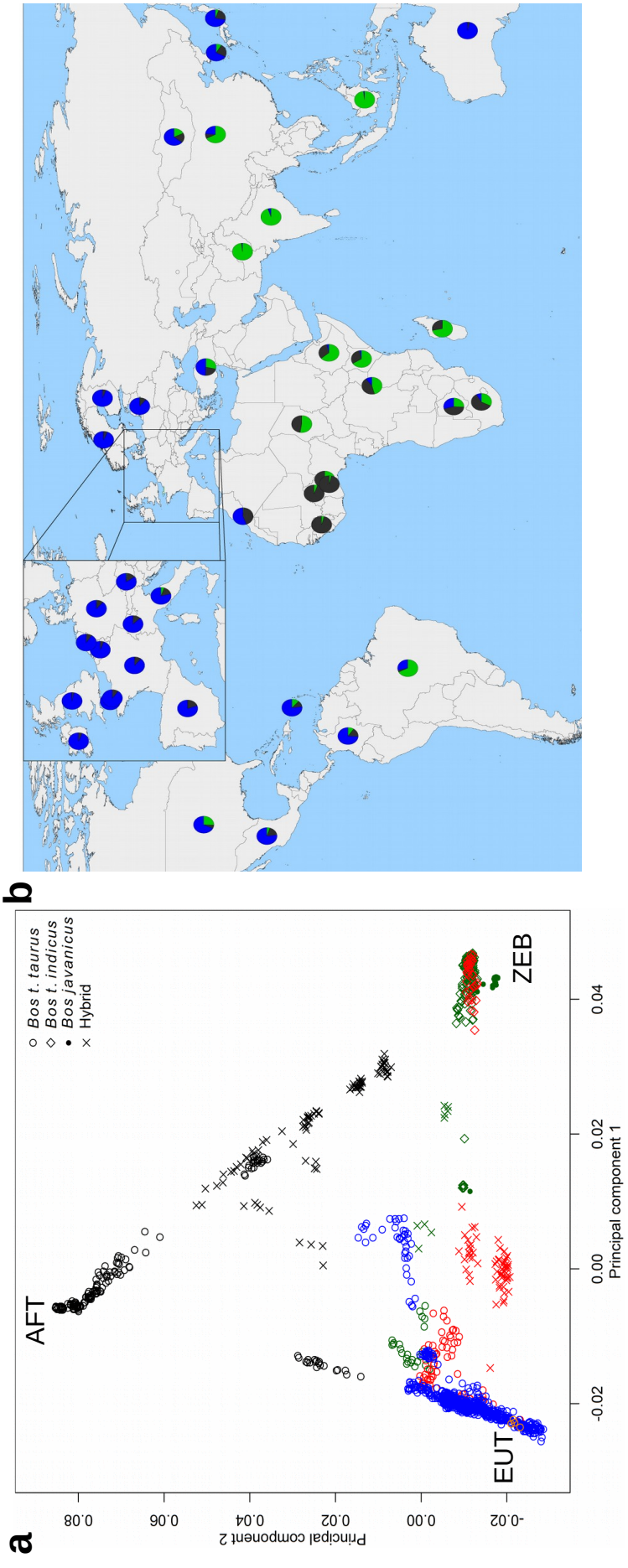


FIGURE 2.9 – **a** Premier plan de l'ACP conduite sur un jeu données regroupant 1 543 individus parmis 134 populations (43 043 SNPs) représentatifs de la diversité génétique bovine mondiale. Les points sont colorés en fonction de l'origine géographique (noir : Afrique, vert : Asie, rouge : Amerique, orange : Australie, and bleu : Europe). **b** Cartographie des proportions d'ascendance taurine européenne (bleu), taurine africaine (noir) et indicine (vert) estimées par le logiciel Admixture 1.21 pour $K = 3$ pour 39 populations bovines. Les figures sont issues de [Decker et al., 2014].

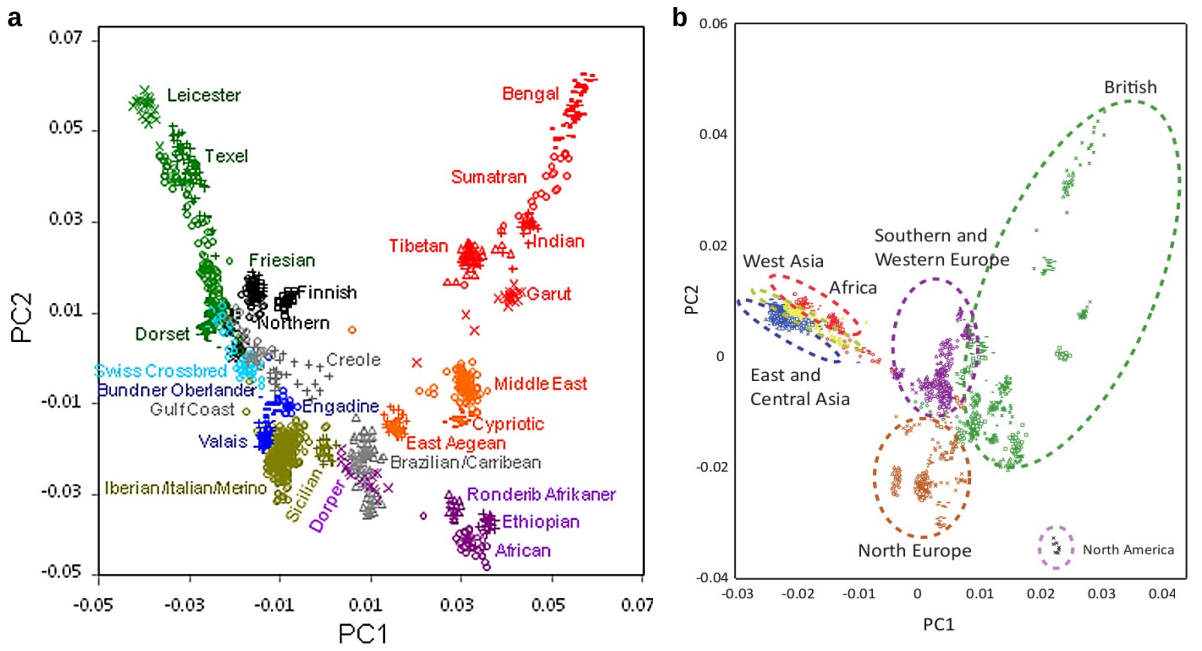


FIGURE 2.10 – Comparaison du premier plan d'ACP pour **a** le jeu de données utilisé dans [Kijas et al., 2012] et **b** celui utilisé dans [Cao et al., 2021], regroupant des individus génotypés sur la puce Illumina OvineSNP50 et représentant la diversité génétique ovine mondiale.

Toujours sur l'axe (AFT-ZEB) mais plus proche du sommet AFT, on retrouve les populations ouest-africaines métissées Borgou, caractérisée génétiquement récemment [Flori et al., 2014], et Kouri [Gautier et al., 2009]. Les populations bovines situées au centre du triangle sont issus d'un métissage des trois ascendances comme c'est le cas de la Créole de Guadeloupe [Gautier and Naves, 2011]. Concernant les ovins, les résultats d'analyse exploratoire sont à prendre avec précaution du fait de la présence d'introgression de Merino dans la plupart des populations ovines. La variation génétique est également très peu expliquée par les 2 premiers plans de l'ACP par rapport aux résultats observés chez les bovins (PC1 = 2.98%, PC2 = 1.44% [Kijas et al., 2012] vs PC1 = 15.34%, PC2 = 2.87% [Mastrangelo et al., 2020]). De plus, la projection des individus varie en fonction de l'effectif européen : entre les deux études de la diversité génétique mondiale des moutons domestiques à partir de données SNP, les moutons asiatiques et locaux africains (excepté la race Dorper) forment un groupe plus concentré sur un des sommets du triangle de projection (Figure 2.10).

Ainsi, compte tenu des biais d'interprétation possibles avec les analyses exploratoires, les méthodes d'inférence démographique basées sur les fréquences alléliques et le déséquilibre de liaison permettent d'estimer les proportions d'ascendance des évènements d'admixture, de les dater et les variations de taille efficace. En calculant les f-statistics, [Flori et al., 2014] a déterminé que le génome de la Borgou était constitué à 60 % d'ascendance taurine africaine, et à 40 % d'ascendance indicine. Chez les zébus Bororo et Fulani 40 % d'ascendance taurine africaine et 60 % d'ascendance indicine sont détectées. L'utilisation du logiciel Alder [Loh et al., 2013] a permis de dater l'admixture assez récente à l'origine de la race Borgou il y a 130 ans environ, c'est-à-dire autour des années 1880, au moment de la grande épizootie de

peste bovine de la fin du 19ème siècle. La datation du métissage à l'origine des zébus Bororo et Fulani est quant à elle plus ancienne, il y a 500 ans environ. En se basant toujours sur les fréquences alléliques, l'inférence de l'histoire démographique des populations bovines locales méditerranéennes [Flori et al., 2019] basé sur un arbre phylogénétique auquel 15 évènements de migration ont été ajoutés par le logiciel Treemix [Pickrell and Pritchard, 2012] a permis de confirmer les différents métissages supposés par l'ACP et le clustering. Le calcul des statistiques f a permis également de mettre en évidence des évènements d'introgression des espèces sauvages du genre *Ovis* vers le mouton domestique *Ovis aries* [Zhao et al., 2017, Cao et al., 2021].

Les populations bovines d'Afrique de l'Est ont été particulièrement étudiées : Ainsi, à partir de données de génotypages (45 000 SNPs), l'estimation de la taille efficace de la population d'EASZ [Mbole-Kariuki et al., 2014] par le calcul du LD a permis de déterminer des variations d'effectifs : après une période d'expansion débutant il y a 750 ans, la population a connu une forte décroissance d'effectif il y a 240 ans, probablement due à un fort épisode de sécheresse et une instabilité politique [Verschuren et al., 2000]. A partir de données de séquences, l'utilisation des logiciels MSMC [Li and Durbin, 2011] et PopsizeABC [Boitard et al., 2016] a permis de détecter les variations dans les effectifs de populations bovines d'Asie orientale et d'Afrique de l'Est mais à différentes échelles de temps. Ainsi alors que le logiciel MSMC a permis de mettre évidence une chute des tailles de populations à partir de 1 million d'années pour les taurins et à partir de 100 000 ans pour les zébus [Chen et al., 2018], le logiciel PopSizeAbc a détecté une chute globale d'effectifs à partir de 100 000 ans jusqu'à il y a 100 ans pour les populations locales Africaines [Kim et al., 2017]. La race taurine africaine N'Dama a particulièrement subi une chute d'effectifs plus importante que les Ankole les Boran et les races commerciales à partir de 1 000 ans. Les races Ogaden et Kenana ont connu un sursaut à partir de 1 000 ans (Figure 2.11).

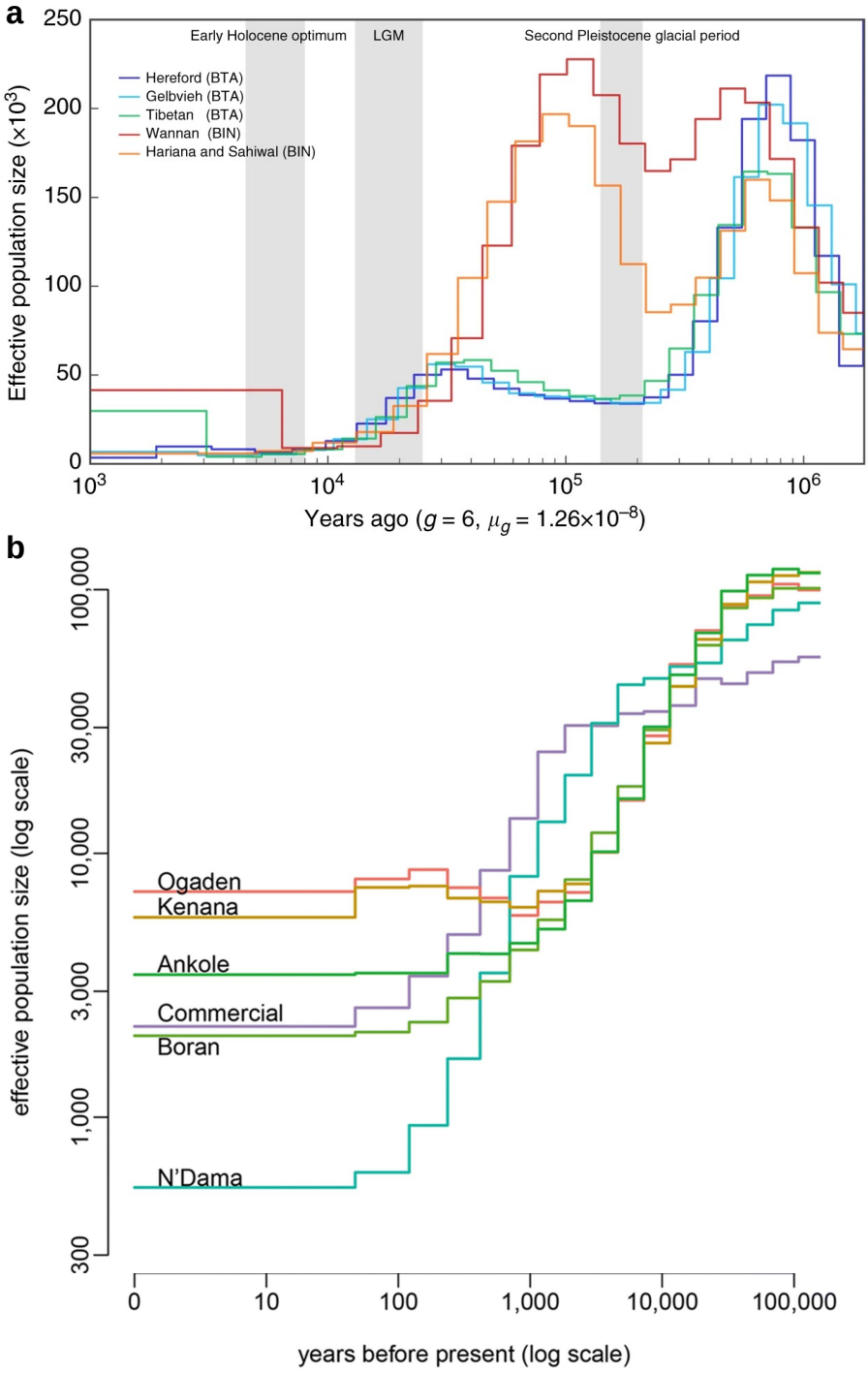


FIGURE 2.11 – Evolution de la taille efficace Ne pour des populations bovines tropicales **a** d'Asie orientale calculée à partir du logiciel MSMC [Li and Durbin, 2011] et **b** d'Afrique de l'Est calculée avec le logiciel PopSizeAbc [Boitard et al., 2016].
a : Figure issue de [Chen et al., 2018]. Le calcul est basé sur quatre haplotypes pour chaque lignées BTA. : *Bos taurus* européens (Hereford), eurasiens (Gelbvieh) et est-asiatiques (Tibétains) ; BIN : *Bos indicus* Zébus chinois (Wannan) et indiens (Hariana et Sahiwal). Les grandes cases grisées illustrent l'optimum de l'Holocène précoce, le dernier maximum glaciaire (LGM) et la deuxième période glaciaire du Pléistocène.
b : Figure issue de [Kim et al., 2017]. Estimation de la taille effective de la population de chaque race bovine africaine et d'un ensemble de races améliorés (Commercial) constitué de 53 individus Hanwoo, Jersey, Holstein et Angus.

Chapitre 3

Etudes des histoires adaptatives des populations bovines et ovines tropicales

L'histoire adaptative des races locales actuelles qui ont dû s'adapter à de nouveaux environnements et aux pressions de sélection associées (pressions de sélection naturelle), ainsi qu'aux contraintes exercées par l'homme (pressions de sélection artificielle) explique en partie leur diversité. Cette adaptation a laissé des traces (ou signatures de sélection) dans le génome que l'on peut détecter à l'aide de différentes méthodes développées ces dernières années, qui distinguent les effect locaux de la sélection des effets des autres forces évolutives (p.ex. dérive génétique, flux de gènes) qui affectent de la même manière tous les loci sur l'ensemble du génome.

3.1 Recherche de signatures de sélection

3.1.1 Différentes signatures de sélection

Différent types de sélection vont laisser des signatures distinctes dans le génome.

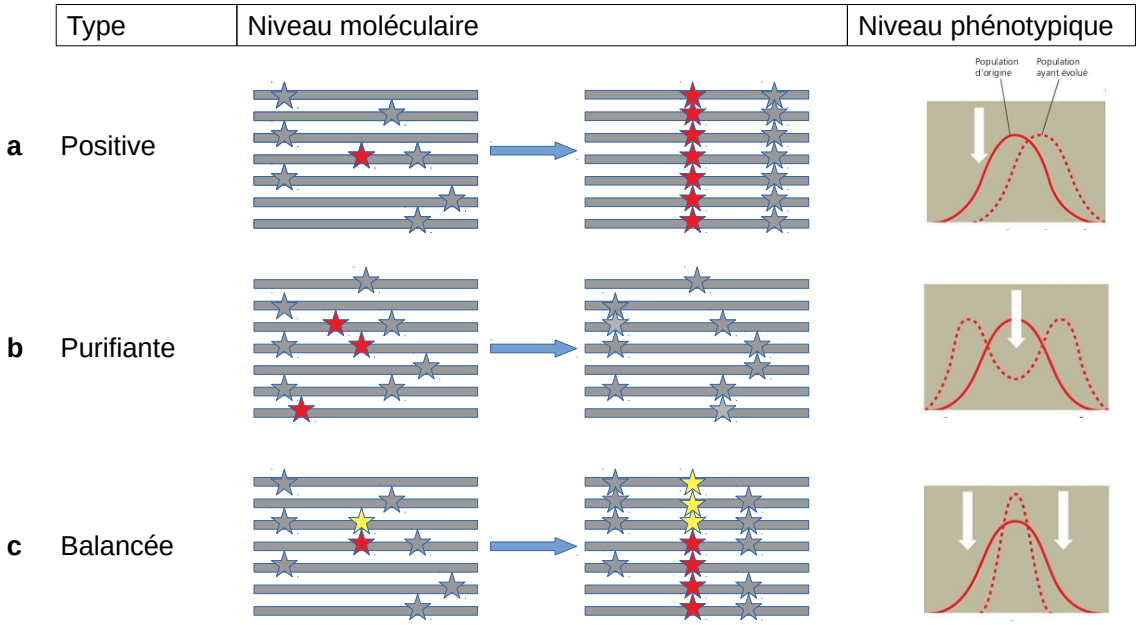


FIGURE 3.1 – Les différents modes de sélection naturelle au niveau moléculaire et phénotypique. Figure construite à partir de [Reece et al., 2012]

Ainsi, la **sélection négative ou sélection purifiante** correspond à l'élimination de mutations (ou variants) délétères dans une population et induit localement une réduction de la diversité, associée potentiellement à une augmentation du nombre d'allèles rares du marqueur concerné par la mutation (Figure 3.1.b). La **sélection positive** agit sur des variants alléliques intervenant favorablement dans la variation du caractère sélectionné (Figure 3.1.a). Ces variants peuvent apparaître spontanément dans la population (par mutation ou migration) (Figure 3.3.a) ou bien préexister et devenir avantageux à la faveur d'un changement de pression de sélection (Figure 3.3.b). Sous l'effet de la sélection positive, l'allèle favorable augmente

en fréquence dans la population et entraîne avec lui les variants proches qui lui sont associés (processus nommé «auto-stop génétique »). A l'échelle de la population, ce phénomène s'accompagne donc d'une baisse de la variabilité dont l'importance dépend de la proximité du site sous sélection (phénomène de «balayage» sélectif de la diversité) où elle est estimée, de la fréquence initiale du variant et de l'importance de la pression de sélection s'exerçant sur le site sous sélection. La **sélection balancée ou équilibrante** résulte du maintien d'un ou plusieurs allèles (favorables) à un seul locus dans une population, favorisant ainsi le maintien de la diversité dans la population (Figure 3.1.c). Elle agit par deux mécanismes principaux : la surdominance ou avantage de l'hétérozygote et la sélection dépendante de la fréquence. Si la sélection balancée n'est pas trop récente, elle conduit à un excès d'allèles de fréquence intermédiaire résultant en une augmentation du niveau de diversité génétique.

La fixation d'une mutation est régie par le phénomène de balayage sélectif réduisant la variation génétique autour de la mutation à divers degrés : si tous les allèles neutres sur le chromosome concerné par une nouvelle mutation avantageuse (Figure 3.2) ou préexistante (Figure 3.2) auront une fréquence de 100%, la variabilité sera éliminée dans cette région chromosomique par un **balayage sélectif dit "complet"** entraînant la fixation de la mutation. Si la fréquence des allèles est inférieure à 100%, le balayage sélectif est dit **"incomplet"**, tandis que dans le cas d'un balayage sélectif **"doux"**, plusieurs haplotypes abritant des mutations avantageuses peuvent augmenter simultanément en fréquence (Figure 3.2).

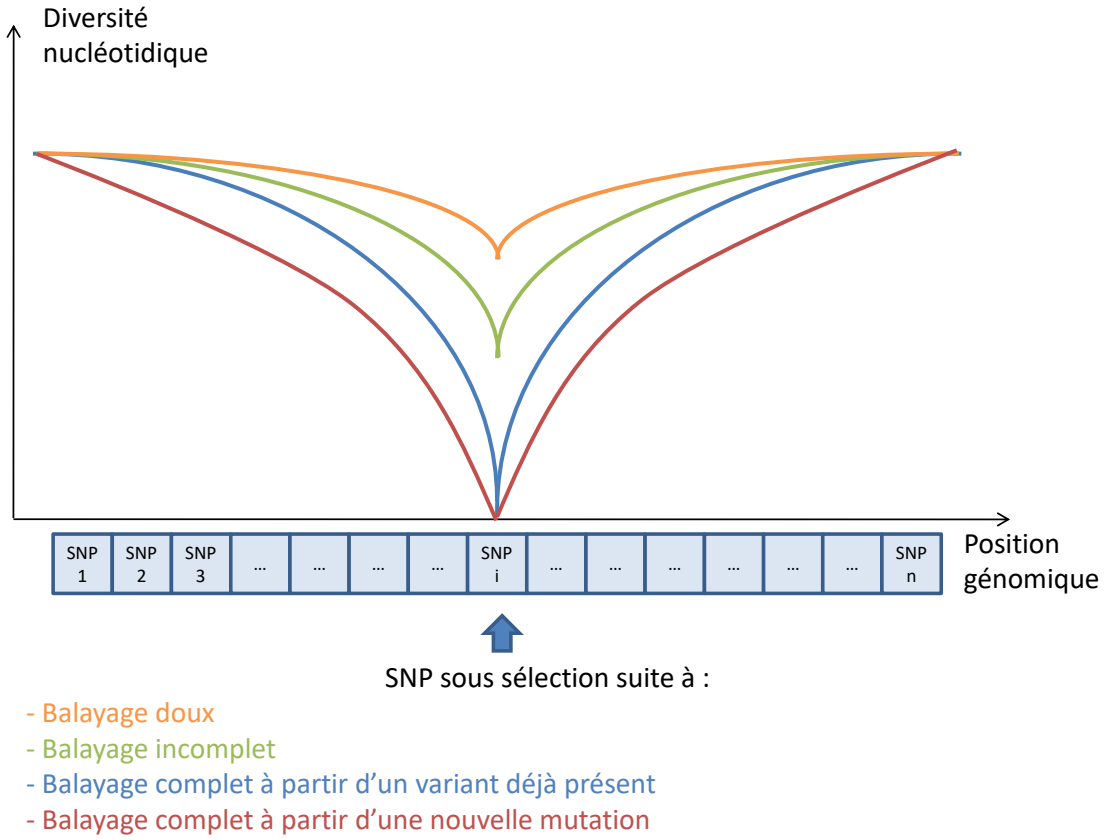


FIGURE 3.2 – Effet des différents types de balayage sélectifs sur la diversité génétique locale.

Le motif de diversité génétique autour du variant sous sélection peut aussi être modifié par le phénomène de recombinaison, permettant à certaines des mutations neutres liées à la mutation avantageuse de s'isoler après un balayage sélectif complet. Cela produit de nouveaux haplotypes[1]. Ceci a également pour effet de diminuer l'étendue du déséquilibre de liaison (Figure 3.4) dans la région génomique sous sélection. Comme le taux de recombinaison est dépendant de la distance physique (plus la distance augmente, la probabilité qu'un crossing over[2] se produise augmente), l'effet d'un balayage sélectif sur la variation dans les régions génomiques qui l'entourent diminue avec la distance au site qui est soumis à la sélection.

Ainsi, les variations génétiques et les structures d'haplotype sont fixées au sein des sous-populations, aboutissant ainsi à un grand nombre de races d'animaux domestiques et des populations génétiquement distinctes [Ma et al., 2015a]. Les signaux de sélection peuvent également renseigner sur l'histoire des populations. Pour résumer, la diversité nucléotidique décroit là où la sélection a amené à fixation un nouvel allèle avantageux. Le but des méthodes de détection des signatures de sélection est donc d'identifier les loci présentant une décroissance de la variabilité le long du génome.

1. **Haplotype** : combinaison d'allèles à deux ou plusieurs loci qu'un individu a hérité de l'un de ses parents, ou en d'autres termes, la combinaison d'allèles à deux ou plusieurs loci présents dans l'un des gamètes lors de la fécondation.

2. **Crossing over** : L'échange de matériel génétique entre les chromosomes homologues au cours de la méiose, qui donne lieu à des chromosomes filles recombinants dans les gamètes.

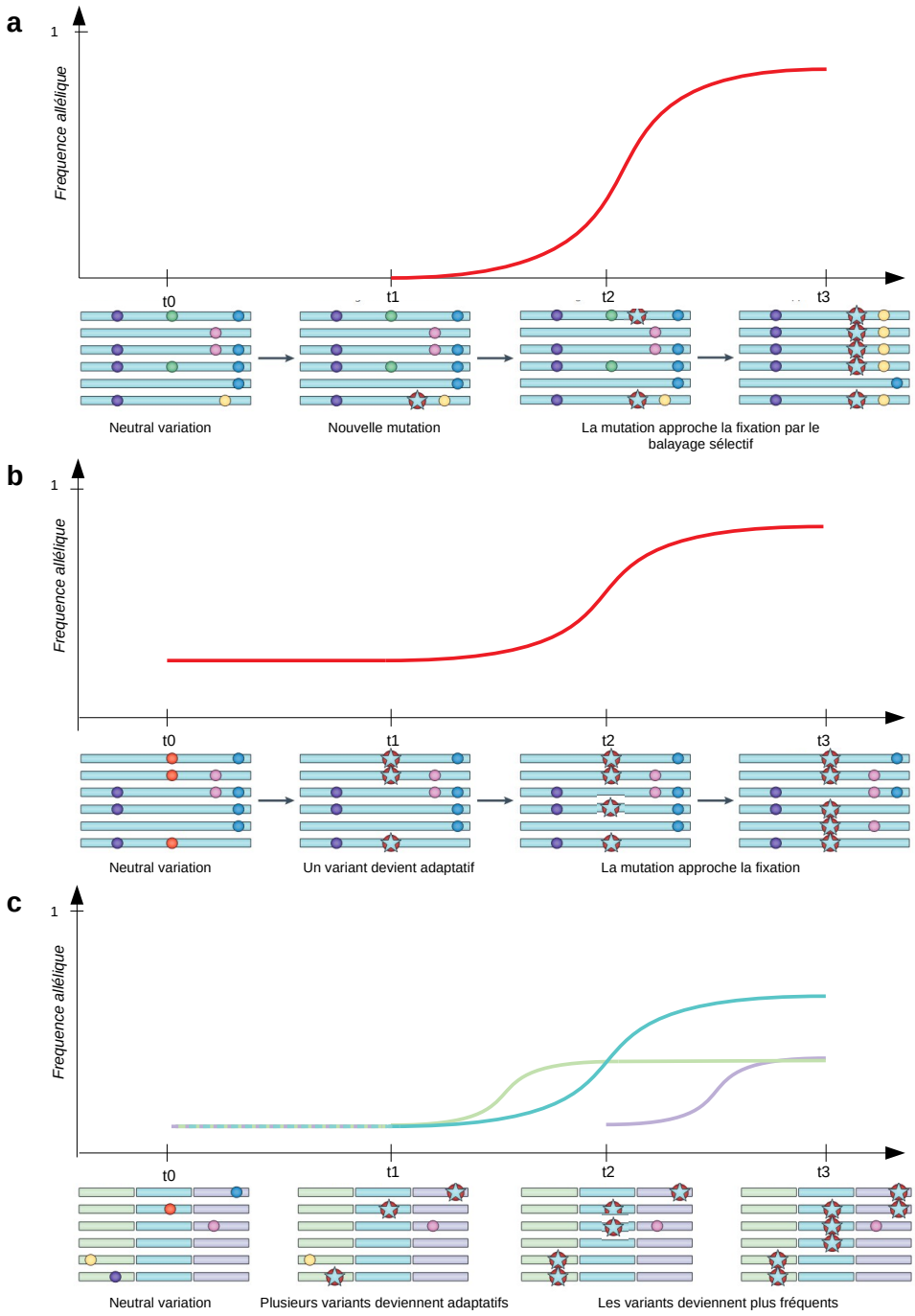


FIGURE 3.3 – Changements temporels de fréquence allélique dans le cas : **a** d'un balayage sélectif classique, dans lequel une nouvelle mutation apparaît dans la population et dont la fréquence augmente jusqu'à la fixation, aboutissant à un déséquilibre de liaison avec les variants voisins ; **b** de la sélection d'un variant déjà présent dans la population qui devient avantageux dans un nouvel environnement et augmente en fréquence jusqu'à être fixé ;**c** de la sélection sur un trait complexe impliquant plusieurs loci sur différents chromosomes (représentés par des segments de différentes couleurs) qui devient avantageux dans un nouvel environnement et dont la fréquence augmente en tant qu'ensemble, laissant une signature plus faible de l'adaptation qui peut inclure des changements plus faibles dans les fréquences alléliques sur plusieurs loci.
Les variants sont représentés par des cercles et les variants "avantageux" sont représentés par des étoiles. Les schémas ont été construits à partir de [Scheinfeldt and Tishkoff, 2013]

.

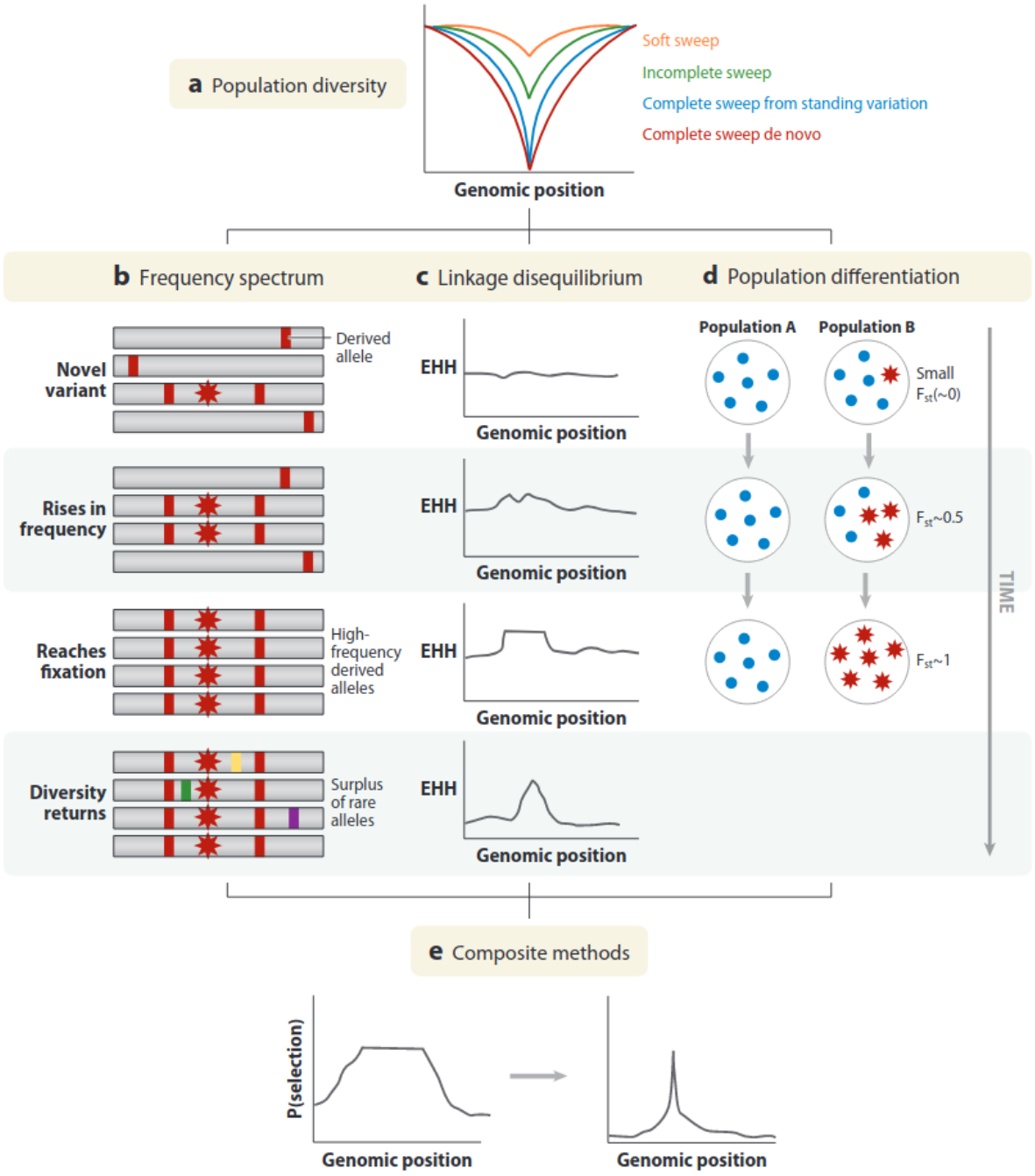


FIGURE 3.4 – Méthodes de détection des signatures de sélection dans le génome (schéma issu de [Vitti et al., 2013]). **a** Effet de plusieurs types de balayage sélectif sur la diversité génétique de la population autour d'un locus sous sélection ; **b** Dans le cas d'une mutation avantageuse (étoile rouge), le balayage sélectif amène les allèles dérivés proches de la mutation à une fréquence élevée ; **c** Un balayage sélectif entraîne une augmentation de l'étendue du déséquilibre de liaison (EHH) sur l'haplotype contenant le variant sous sélection. Le plateau de EHH élevée diminuera progressivement au fur et à mesure de l'apparition des nouvelles mutations et de la recombinaison, restaurant progressivement la diversité de la population. ; **d** Les différences de fréquences alléliques, reflètant l'action de la sélection spécifique à une population, font augmenter l'indice de différentiation entre deux populations. ; **e** Les méthodes composites, intègrant des informations provenant de plusieurs signaux de sélection peuvent fournir une résolution plus fine et aider à identifier les variantes causales.

3.1.2 Méthodes de détection

Plusieurs types de méthodes [Oleksyk et al., 2010, Vitti et al., 2013, Werren et al., 2021] sont utilisés pour détecter des signatures de sélection au sein d'une espèce. Ces méthodes sont basées sur la recherche à l'échelle du génome des écarts statistiques par rapport à l'équilibre génétique attendu des populations, qui sont un indicateur de l'occurrence du processus de sélection dans une région génomique donnée. Parmi celles ci, on peut citer :

— les méthodes basées sur la mesure de la diminution locale de la variation génétique (mesurée par l'hétérozygotie) en comparant les niveaux de polymorphisme au niveau et autour du locus sélectionné avec l'espérance neutre estimée,
— la détection du changement de forme de la distribution des fréquences alléliques au sein d'une population (Figure 3.4.b) : *le D de Tajima* [Tajima, 1989], la statistique *H de Fay et Wu* [Fay and Wu, 2000],
— la mesure de la différenciation entre plusieurs populations (Figure 3.4.d) : *Fst*, *hapFLK* [Fariello et al., 2013], *XtX* [Günther and Coop, 2013],
— la mesure de l'étendue du déséquilibre de liaison (extended haplotype homozygosity - EHH) [Sabeti et al., 2002] (Figure 3.4.c) et ses statistiques dérivées comme le score *iHS* au sein d'une population [Voight et al., 2006] et le score *Rsb* entre deux populations [Tang et al., 2007].

Ces méthodes ne détectent pas les mêmes types de signatures de fixation : ainsi les méthodes basées sur la mesure de l'hétérozygotie, la distribution des fréquences alléliques et la mesure de la différentiation permettent de détecter les allèles fixés. Pour les méthodes basées sur l'étendue du déséquilibre de liaison, une distinction se fait entre les scores Rsb et iHS. Le score Rsb présente une puissance de détection élevée lorsque la fixation de l'allèle sélectionné est atteinte, tandis que le score iHS a la puissance élevée lorsque la sélection est en cours [Ma et al., 2015b].

Le choix de score statistique implique de prendre en compte de l'histoire démographique qui est un effet confondant pouvant aboutir à des taux de faux positifs dans la détection de signatures de sélection [Nielsen et al., 2007, Scheinfeldt and Tishkoff, 2013].

L'échelle de temps sur laquelle la sélection s'est produite a un impact majeur sur la capacité de chaque méthode à détecter des preuves de sa présence, les méthodes basées sur la fréquence et la différenciation de la population étant les mieux adaptées à la détection d'événements survenus dans un passé plus lointain [Oleksyk et al., 2010]. En effet, ces méthodes reposent sur l'accumulation de mutations supplémentaires autour du variant causal. Dans les situations où l'avantage en termes de fitness du variant sélectionné est faible (en particulier s'il est récessif), le temps nécessaire pour que le variant sélectionné atteigne une fréquence détectable dans la population sera beaucoup plus long, ce qui réduit la puissance de ces méthodes. Dans les situations où une nouvelle mutation (ou un variant précédemment neutre rencontrant un nouvel environnement accompagnée d'une pression de sélection modifiée) fournit un avantage en termes de fitness et augmente en fréquence dans la population sans atteindre la fixation, les

approches basées sur le déséquilibre de liaison fournissent une puissance accrue pour détecter des preuves de sélection [Oleksyk et al., 2010, Cadzow et al., 2014]. Néanmoins, les méthodes basées sur le déséquilibre de liaison permettent de détecter les signatures de sélection récentes [Oleksyk et al., 2010] car les balayages sélectifs plus anciens ont été sujets à la recombinaison, diminuant ainsi le déséquilibre de liaison [Werren et al., 2021]. Cependant, certaines régions chromosomiques peuvent faire l'objet d'un niveau de recombinaison variable en fonction de la population [Shen et al., 2018]. Un signal basé sur un long segment haplotypique sur ces locis pourrait résulter d'un taux de recombinaison plus faible dans une population par rapport aux autres populations, au lieu d'un pic dans la fréquence des allèles dû à la sélection naturelle [Sabeti et al., 2002]. Pour tenir compte de cela, le niveau de décroissance de l'étendue du déséquilibre de liaison (EHH) des allèles non sélectionnés dans la région entourant l'allèle candidat peut être comparé à la décroissance de l'EHH de ces mêmes allèles non sélectionnés dans d'autres populations [Sabeti et al., 2002]. Si la décroissance EHH est similaire entre les populations, le taux de recombinaison variable peut être exclu comme source du signal, ce qui renforce l'hypothèse de sélection [Sabeti et al., 2002].

La mesure de l'étendue du désequilibre de liaison

Définition La détermination des signatures de sélection à partir de calcul de score basé sur l'étendue du déséquilibre de liaison (ou Etendue d'Homozygotie Haplotypique - EHH) pour chaque marqueur nécessite au préalable de réaliser un phasage des données de génotypage (Figure 3.5). Un haplotype est défini comme une combinaison d'allèles présents à un ou plusieurs loci sur un seul chromosome qui ont tous été hérités d'un seul parent. La statistique EHH permet de quantifier l'étendue de la diversité d'un haplotype potentiellement porteur d'une mutation sous sélection [Sabeti et al., 2002] est une méthode visant à estimer la probabilité que 2 haplotypes choisis au hasard soient identiques à une distance x au voisinage d'un SNP particulier ou *core SNP*. Plus précisément, il s'agit de déterminer le nombre de paires d'haplotypes identiques dans une fenêtre de longueur x divisée par le nombre total de paires d'haplotypes.

iHS Le score iHS (integrated Haplotype homozygosity Score) permet de comparer la décroissance de l'EHH entre l'allèle ancestral et l'allèle dérivé [Voight et al., 2006] au sein d'une population en calculant l'aire sous la courbe de l'EHH en fonction de la distance autour du SNP cible. Ainsi cette comparaison est effectuée en divisant l'aire sous la courbe EHH pour l'allèle ancestral (*iHHa* dans l'équation ci-dessous, en bleu dans la Figure 3.6) par celle calculée pour l'allèle dérivé (*iHHb* dans l'équation ci-dessous, en rouge dans la Figure 3.6) et en appliqant un logarithme au rapport :

$$uniHS(s) = ln\left(\frac{iHHa(s)}{iHHb(s)}\right)$$

Ensuite à partir de ce score, [Voight et al., 2006] propose une standardisation :

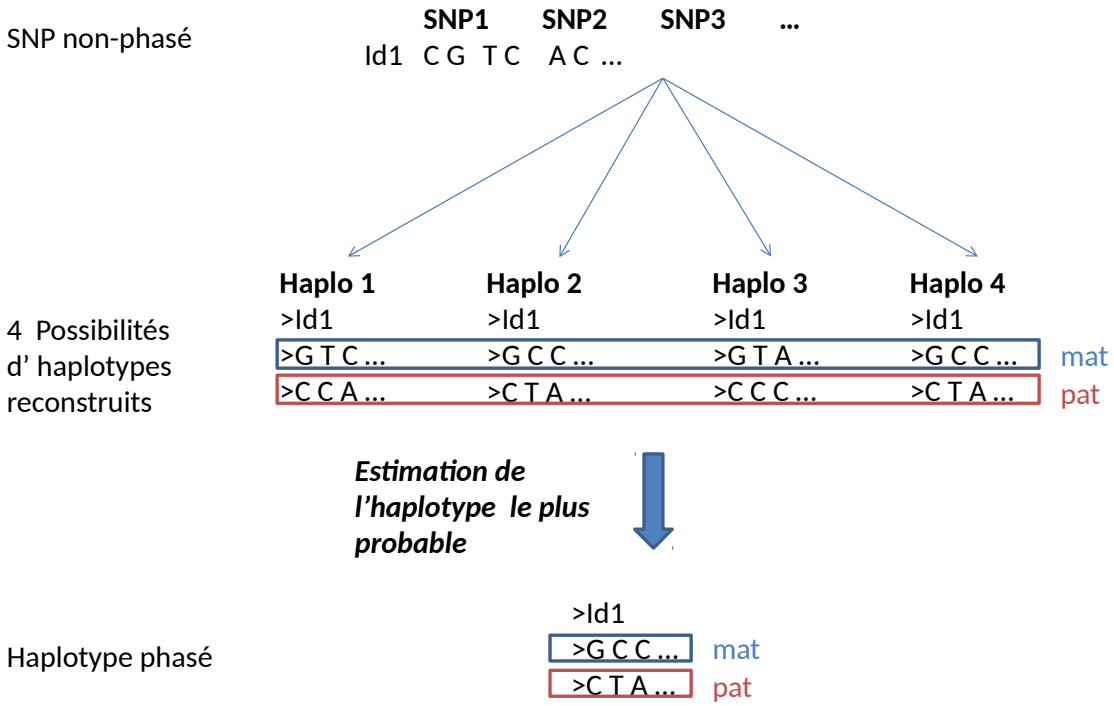


FIGURE 3.5 – Le processus de phasage consiste à déterminer les haplotypes parentaux à partir des données de génotypage individuels.

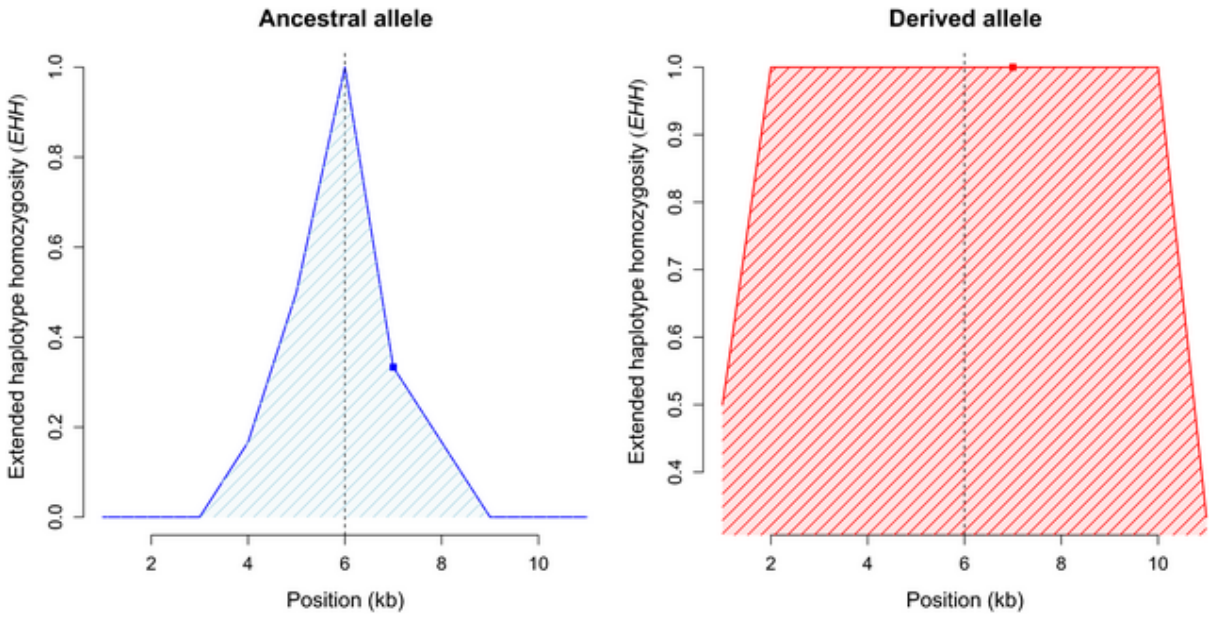


FIGURE 3.6 – Représentation des EHH à plusieurs distances du SNP cible (positionné à 6kb). L'EHH décroît beaucoup plus rapidement pour les haplotypes contenant le variant ancestral (en bleu) que pour ceux contenant le variant dérivé (en rouge). L'aire rayée, située sous les courbes de EHH, correspond au score iHH (EHH intégré) (Figure issue de [Gautier et al., 2017]).

$$iHS(s) = \left(\frac{uniHS(s) - moyenne(uniHS|ps)}{sd(uniHS|ps)} \right)$$

Ce score permet de détecter les balayages partiels ou incomplets.

Rsb Le score Rsb (ratio of EHH between populations) compare l'étendue du déséquilibre de liaison entre deux populations. Comme le score iHS, Rsb est un rapport de l'aire sous la courbe EHH pour la population 1 et de l'aire sous la courbe EHH pour la population 2 au niveau d'un marqueur donné.

Pour un marqueur noté *s* :

$$unRsb(s) = ln \left(\frac{inESpop1(s)}{inESpop2(s)} \right)$$

est le log-ratio des valeurs de *inES* calculée pour les populations *pop1* et *pop2*.

A partir de ce score, [Tang et al., 2007] propose une standardisation :

$$Rsb(s) = \left(\frac{unRsb(s) - median(unRsb)}{sd(unRsb)} \right)$$

Le score Rsb est adapté pour identifier les signatures de sélection où la balayage sélectif est proche de la fixation dans une population.

Méthodes basées sur la différenciation génétique

Ces méthodes identifient les marqueurs qui présentent un excès ou un défaut de variation génétique quantifié en estimant le paramètre F_{ST} qui quantifie la part de variation génétique globale qui est attribuable à la structuration de la population en sous-population relativement à un attendu neutre. Plusieurs outils sont disponibles pour calculer ce paramètre (Poolfstat [Hivert et al., 2018, Gautier et al., 2021] mais le défaut principal de l'utilisation du scan de F_{ST} pour déterminer la présence de signature de sélection est l'absence de prise en compte de la structure génétique des populations.

Le score XtX est équivalent à la statistique FST en prenant en compte la variance-covariance entre les populations [Günther and Coop, 2013] par le calcul des fréquences alléliques standardisées. Le logiciel BAYPASS 2.2 [Gautier, 2015] permet de calculer ces scores XtX de différentiation. L'identification de SNP très différentiés est basé sur les statistiques XtX [Gautier, 2015, Günther and Coop, 2013] estimés sous le "core model" de Baypass avec la matrice de covariance Ω résumant les fréquences alléliques d'une population tout en tenant compte de la structure de la covariance. Le modèle inclut aussi des hyperparamètres $a\pi$ et $b\pi$ pour estimer la distribution Beta des fréquences allèliques pour chaque population, ce qui permet de prendre en compte le biais associé à la puce SNP [Gautier, 2015]. L'utilisation de ce logiciel incite à effectuer des calculs de calibration des statistiques XtX en simulant un jeu de données, contenant 100 000 SNPS, sous le modèle inféré avec les hyperparamètres

équivalents à ceux estimés sur le vrai jeu de données, et en relançant le logiciel Baypass sur ce jeu de données pseudo-observé (POD). En particulier, pour s'assurer de la similarité entre l'estimation de Ω obtenu avec le jeu de données POD et le "vrai" jeu de données, la distance FMD [Förstner and Moonen, 2003] calculée entre les deux matrices de covariance doit être égale à 0.21 et la corrélation entre leurs éléments doit être proche de 1. De plus, les moyennes des 2 hyperparamètres a et b pour la distribution Beta des fréquences alléliques des populations obtenues avec le jeu de données simulé doivent être égales à celles obtenues pour le vrai jeu de données. A partir de ces deux vérifications indiquant que le jeu de donnée simulé imite assez vraisemblablement le vrai jeu de données, le seuil de significativité est égal à la valeur du quantile 0.1% de la distribution des XtX calculés sur le jeu de données POD. Les statistiques XtX sont ensuite utilisées pour identifier les signatures de sélection.

3.1.3 Identification des gènes candidats

Approche région candidate

Afin d'identifier les régions candidates (Figure 3.7.b), un criblage du génome est effectué en considérant des fenêtres chevauchantes de taille et de longueur de chevauchement définies (Pour la puce 50K : les fenêtres sont de 500 kb de longueur et de 25 kb de chevauchement [Gautier et al., 2009, Flori et al., 2014, Flori et al., 2019]). Pour chaque fenêtre, le nombre de SNP présentant une valeur de score ou de p-value supérieurs à un seuil défini est comptabilisé. Les régions sont considérées comme candidates si un nombre défini de SNP présentent un score supérieur à ce seuil. Les gènes localisés dans les régions candidates à proximité des SNP significatifs sont ensuite identifiés en utilisant les données d'annotation des génomes des sites UCSC (`https://genome.ucsc.edu/`) ou Ensembl (`https://www.ensembl.org/`) [Gautier and Naves, 2011, Flori et al., 2014].

Approche gènes candidats

Cette approche (Figure 3.7.c) consiste à localiser les SNP présentant des valeurs supérieures au seuil défini par rapport aux positions de l'ensemble des gènes localisés sur un assemblage donné de l'espèce étudiée (RefSeqGenes). Un SNP est considéré comme "représentatif" d'un des gènes s'il est localisé entre les positions de début et de fin du gène plus ou moins une distance définie (par exemple 15kb) [Gautier et al., 2007].

3.1.4 Annotation fonctionnelle des gènes candidats

L'annotation fonctionnelle des gènes candidats identifiés est basées sur l'utilisation d'outils de la biologie des systèmes (Ingenuity Pathway Analysis et Cytoscape par exemple) qui permettent de proposer une interprétation biologique des signaux de sélection. Ces outils permettent d"identifier les fonctions, les réseaux de gènes et les voie bioologiques dans lesquels les gèènes candidats sont impliqués. Ils intègrent notamment différentes informations comme

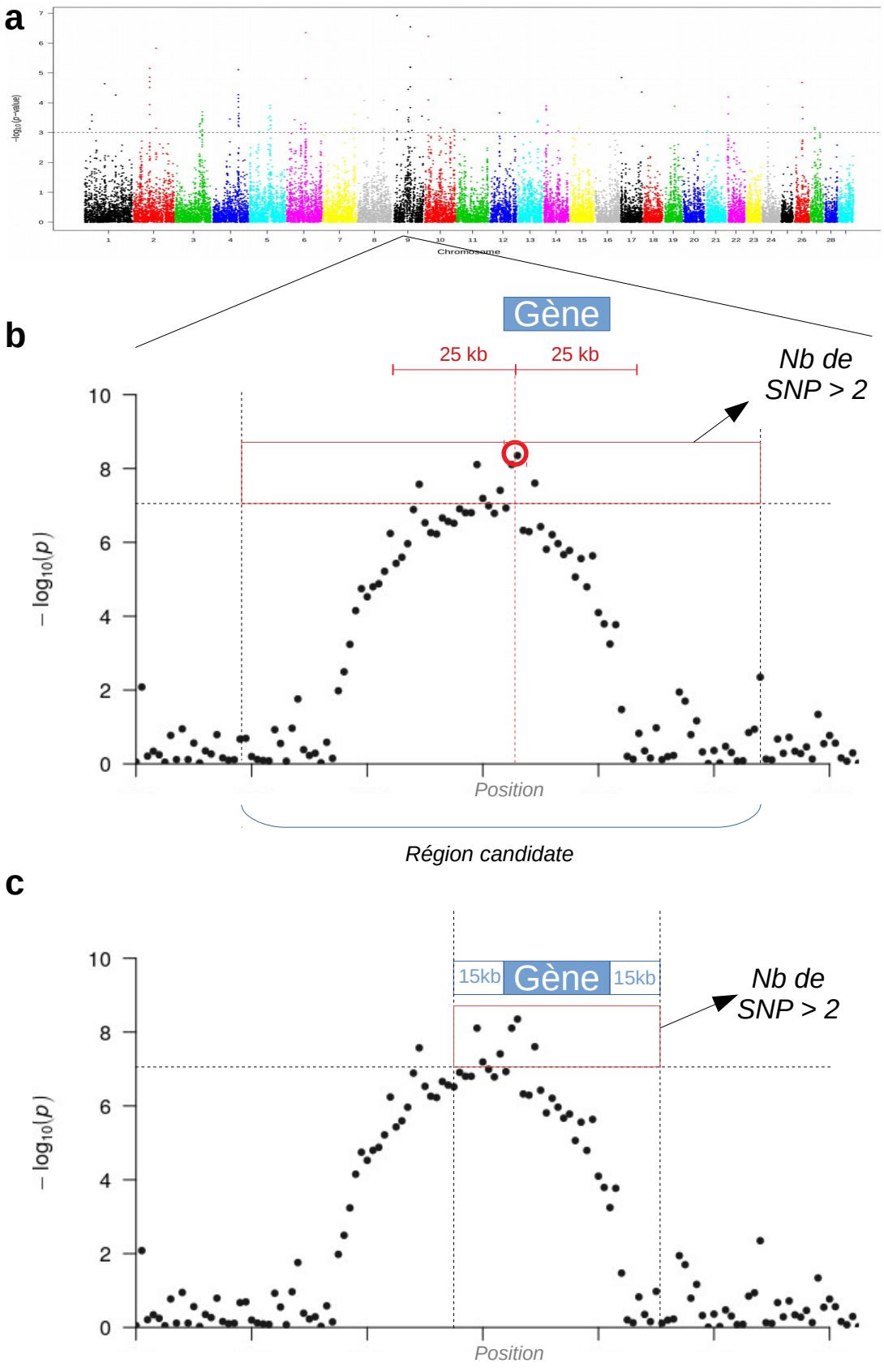


FIGURE 3.7 – Représentation graphique des valeurs de score calculés pour chaque SNP sur tout le génome : **a** Manhattan plot. Zoom sur une région chromosomique présentant un score élevé et visualisation des méthodes de détermination des gènes candidats : **b** approche régions candidates et **c** approche gènes candidats

des annotations fonctionnelles (ontologies de gènes), des annotations en termes de voies biologiques (KEGG pathways par exemple) et des informations d'interactions entre molécules. Les méthodes statistiques utilisées permettent d'identifier les fonctions ou les voies biologiques sur-représentés dans la liste de gène candidat gènes à l'aide d'un test Exact de Fisher unilatéral qui compare les fonctions associées aux gènes candidats sélectionnés avec les fonctions associées à tous les gènes de référence (tous ceux présents dans la base de données utilisée ou tous les gènes localisés à proximité d'un SNP testé).

Ces outils permettent également de construire des réseaux de gènes intégrant les gènes candidats identifiés. Avec le logiciel Ingenuity Pathway Analysis, les réseaux de gènes sont construits tout d'abord à partir des molécules "cibles" (c'est-à-dire les gènes associés à des signatures de sélection) qui présentent les plus fortes intéractions, puis les molécules "non-cibles" du dataset de référence (gènes à proximité des SNP génotypés mais considérés comme non significatifs) sont ensuite rajoutées puis les molécules de la base de données Ingenuity Knowledge Base (IKB) du logiciel jusqu'à obtenir le nombre maximal de molécules défini (35, 70 ou 140), afin de combler les manques de connectivité entre les molécules. Un score S basé sur un test exact de Fisher unilatéral de surreprésentation des gènes candidats ($S = -log(pvalue)$)est ensuite calculé pour chaque réseau. Un réseau est considéré comme significatif si $S > 3, (pvalue > 0.001)$. Chaque réseau peut ensuite être annoté en termes de fonctions ou de voies biologiques. Les fonctions les plus significatives sont obtenues en comparant les fonctions associées avec les molécules du réseau avec les fonctions associées avec toutes les molécules de la base d'Ingenuity Pathway Analysis ou avec un jeu de données de référence, en utilisant un test exact de Fisher unilatéral.

3.2 Association Genome Environnement

Les tests d'association entre le génome et des covariables environnementales (Association Génome-Environnement ou GEA) sont utilisés en génomique du paysage ("Landscape genomics"). Cette approche combine en effet la recherche de corrélations entre polymorphismes génomiques et variations environnementales aux zones d'origine des animaux et l'utilisation de tests de signature de sélection. Le but de ces approches est de trouver une indication de correlation des réponses évolutionnaires à l'hétérogénéité spatiale [Holderegger et al., 2010]. Les études de génomique du paysage se focalisent sur l'adaptation locale aux conditions environnementales entre et parmi différentes localisations géographiques [Rellstab et al., 2015, Hoban et al., 2016]. Comme les tests de différentiation génétique seuls ne permettent pas d'identifier les facteurs environnementaux ou les pressions de sélection conduisant à une adaptation locale, les analyses d'association génotype-environnement permettent d'identifier les loci associés à des facteurs environnementaux spécifiques et conduisant à l'adaptation locale [Haasl and Payseur, 2016].

Plusieurs types d'outils ont été développés pour mener ces études d'associations genome-environnement. Ils se différencient par le modèle et la procédure statistique utilisés et la

façon dont la structure de la diversité génétique neutre des populations due à leur histoire démographique (qui peut être un facteur confondant dans ce type d'étude) est prise en compte. C'est ce que proposent les méthodes GEA intégrées aux logiciels Bayenv [Coop et al., 2010, Günther and Coop, 2013], LFMM et LFMM2 [Frichot et al., 2013] et BayPass [Gautier, 2015].

Le logiciel BAYPASS [Gautier, 2015] qui s'appuie sur des modèles hiérarchiques bayésiens d'étude de la différenciation et permet également de réaliser des tests d'association avec des variables définies à l'échelle des populations, a été utilisé récemment sur des jeux de données SNP de populations bovines et ovines [Gao et al., 2017, Flori et al., 2019, Wiener et al., 2021] pour identifier les régions génomiques associées aux variables climatiques issues de la base de données Climond (`www.climond.org` ; [Kriticos et al., 2012]). L'association d'un SNP avec la covariable est estimée par un Bayes Factor (BF). La règle de Jeffrey [Jeffreys, 1998] est un outil permettant de déterminer le niveau de corrélation en deciban (dB) : "forte preuve" pour 10 < BF <15, "très forte preuve" pour 15 < BF < 20, "preuve déterminante" BF > 20. Le logiciel BAYPASS peut également être utilisé pour tester l'association de régions génomiques avec des phénotypes populationnels comme l'illustre les analyses d'association réalisées avec le patron de coloration et le poids chez les bovins [Gautier, 2015]. Les phénotypes peuvent être catégoriels ou quantitatifs.

3.3 Adaptation des bovins et des ovins à l'environnement tropical

Les bovins et ovins qui ont été introduits dans les zones tropicales ont dû s'adapter aux conditions environnementales spécifiques de ces zones. Ces animaux ont en effet dû faire face aux conditions climatiques tropicales caractérisées par des températures élevées, de fortes radiations solaires et des niveaux d'humidité importants, aux variations des ressources alimentaires et hydriques disponibles et à la présence d'agents pathogènes responsables de maladies tropicales. Ces différentes pressions de selection naturelle ont eu un impact sur le génome de ces animaux et y ont donc laissé des signatures de sélection.

3.3.1 Adaptation aux conditions climatiques - tolérance à la chaleur

Le climat tropical se caractérise par la succession de deux saisons, une sèche et une pluvieuse. Il se sub-divise en trois sous-types (equatorial, savane, mousson) selon le niveau des précipitations [Koppen, 1936, Peel et al., 2007].

Les zébus ont recours à une combinaison de six stratégies pour éviter l'augmentation de la température corporelle et maximiser la perte de chaleur : l'augmentation de la surface par unité de poids corporel ; l'augmentation du gradient de température entre l'animal et l'air ; l'augmentation de la conduction de la chaleur du noyau corporel vers la peau ; la diminution

de la réflexion des radiations solaires ; l'augmentation du taux métabolique et de la prise de nourriture, et l'ajustement des mécanismes cellulaires. Ainsi, la graisse de la bosse permet de diminuer le dépôt de graisse corporelle sur l'ensemble du corps, ce qui entraîne une augmentation de la dissipation de la chaleur. De plus, la bosse est un appendice supplémentaire qui contribue à l'augmentation de la surface corporelle [Hansen, 2004]. Au niveau cellulaire, la tolérance au stress thermique est directement reliée par la capacité de la cellule à maintenir un niveau élevé de protéines de stress de la famille des heat shock proteins. Les gènes codant ces protéines sont régulièrement identifiés dans des régions sous sélection chez des taurins africains comme la N'Dama (*HSPA4* ; [Kim et al., 2017]), des zébus africains (*HSPB9* ; [Chan et al., 2010, Bahbahani et al., 2015, Makina et al., 2015, Kim et al., 2020] ou associés à des variables climatiques (*HSPB3* ; [Flori et al., 2019]). D'autres gènes sont également impliqués dans la tolérance au stress thermique comme les gènes liés à la prolactine (*PRL, PRLR, PRLH*) : en effet la prolactine est non seulement associée à la production laitière mais a également un impact sur la thermorégulation via la morphologie des poils notamment chez la Sénépol [Littlejohn et al., 2014, Kim et al., 2017].

La peau et les poils interviennent en effet dans les mécanismes de diffusion de la chaleur et de réflexion des radiations solaires [Turner, 1980, Payne and Wilson, 1999, Hansen, 2004]. **La pigmentation de la peau et des poils** est un important caractère dans une stratégie de protection des tissus profonds contre les radiations solaires courtes caractéristiques des régions tropicales [Collier et al., 2012, Naskar et al., 2012]. Ainsi, les animaux au pelage foncé reçoivent plus de chaleur que les animaux au pelage clair [Collier et al., 2012], un phénotype qui serait plus souhaitable sous certaines latitudes. Néanmoins, la particularité des bovins tropicaux est bien souvent la présence d'une pigmentation foncée sous un pelage clair [Naskar et al., 2012]. Différents gènes, *KIT, MITF*, ont été mis en évidence dans plusieurs études de signatures de sélection chez les ovins et les bovins [Kijas et al., 2012, Edea et al., 2017b, Kim et al., 2017, Tijjani et al., 2019]. Ces gènes confèrent une résistance à la mélanose [3] et la capacité à résister aux radiations UV chez l'homme [Satzger et al., 2008] et les bovins [Pausch et al., 2012]. Le gène *KIT* est également associé à la robe "pie", c'est-à-dire constitué de grandes taches blanches sur une base foncé (noir ou marron) chez de nombreuses races taurines [Fontanesi et al., 2012, Qanbari et al., 2014, Jivanji et al., 2019]. Le gène *MC1R*, situé dans une région sous sélection chez la Normande et la Holstein [Flori et al., 2009] en milieu tempéré et chez l'Ankole [Kim et al., 2017] en milieu tropical, code pour récepteur membranaire qui contrôle la production de mélanine par les mélanocytes, les cellules responsables de la pigmentation de la peau. Ce gène contrôle en particulier la production des pigments eumelanine et de pheamelanin, responsable de la pigmentation de la peau, des yeux et du pelage [Ito and Wakamatsu, 2011]. **La texture, le type de poil et sa densité** jouent également un rôle dans la protection contre la chaleur et minimisent la perte d'eau due au refroidissement par transpiration. Globalement, une présence important de poils agit comme un piège à humidité,

3. Les mélanoses sont liées à la prolifération de certaines cellules cutanées spécialisées, les mélanocytes, se traduisant par une accumulation du pigment (la mélanine) que ceux-ci sécrètent dans le derme.

réduisant le gradient d'humidité entre la surface de la peau et la couche de poils située au-dessus, et par conséquent la perte de chaleur de la surface de la peau et le refroidissement par évaporation de la surface de la peau [Collier et al., 2012]. Ainsi, certaines populations tropicales présentant des poils plus brillants sont connus pour mieux réguler leur température corporelle durant un stress thermique comparé à des populations taurines [Turner, 1980, Hansen, 2004]. Un exemple particulièrement frappant de l'adaptation aux conditions tropicales est le phénotype "Slick" rencontré chez la Sénépol [Mariasegaram et al., 2007] et la Limonero, une race créole (du Venezuela), caractérisé par un poil très ras et brillant et une faible densité folliculaire. Cette caractéristique est le résultat de l'expression de la mutation d'un gène situé dans une région sous sélection sur le chromosome 6 chez la Sénépol [Flori et al., 2012]. Ce phénotype "Slick" est associé à un variant causal situé dans le 11eme exon du gène *PRLR*, codant pour le récepteur de la prolactine, résultant dans le décalage du cadre de lecture et aboutissant à une protéine tronquée. D'autres variants causaux présents dans le 11eme exon de ce récepteur sont également présents chez une race créole, la Limonero [Porto-Neto et al., 2018]. Chez les ovins, les moutons à poils sont mieux adaptés à des conditions de fortes températures et humidité [McManus et al., 2009] grâce à une meilleure efficience du mécanisme de respiration et de convection pour la dissipation de la chaleur contrairement aux mécanismes de refroidissement par transpiration observé chez les types laineux [McManus et al., 2011].

Le climat tropical et le stress thermique ont également une influence sur la physiologie de **la reproduction** des animaux qui le subissent. Ainsi, les zébus présentent une puberté retardée (anoestrus prépubertaire) et des périodes d'anoestrus post-partum prolongées par rapport aux bovins des climats tempérés. Des différences entre les intéractions hypothalamiques, hypophysaires et ovariennes expliqueraient les différences de fertilité entres les deux types de bovins [Abeygunawardena and Dematawewa, 2004]. Les taux ovulatoires seraient plus faibles chez les zébus que les taurins [Turner, 1980] mais des analyses comparatives ont indiqué que l'effet néfaste du stress thermique sur la fertilité et la mortalité embryonnaire [Paula-Lopes et al., 2013] et la qualité de la semence [Hansen, 2004, Rahman et al., 2018] est moins prononcé chez *B. indicus* que chez *B. taurus*.

Le gène *RXFP2*, bien que connu principalement pour être situé dans les régions génomiques voisines responsable du phénotype POLLED (sans corne) [Aldersey et al., 2020], est retrouvé associé à des signatures de sélection chez la Créole de Guadeloupe [Gautier and Naves, 2011], la Borgou [Flori et al., 2014], l'East African Shorthorn Zebu [Bahbahani et al., 2015] et la Sheko [Bahbahani et al., 2018]. Ce gène encode une protéine impliquée dans le développement descendant testiculaire qui peut être une adaptation pour maintenir une spermatogénèse adéquate sous des conditions de stress thermique.

Le comportement est aussi impliqué dans l'adaptation aux stress environnementaux. Ainsi on note une modification du comportement pour réduire l'impact du stress thermique qui s'accompagne d'augmentation de l'agressivité face à la limitation des ressources (ombre, eau) et à la reduction/augmentation de l'ingestion d'eau ou d'aliments. Le stress thermique réduit

également l'expression des différents comportements sexuels notamment chez les ovins. La modification du comportement peut être reliée aux gènes ayant une action dans les fonctions biologiques du système nerveux comme cela a pu être observé chez différentes populations tropicales [Gautier et al., 2009, Makina et al., 2015].

3.3.2 Adaptation à la disponibilité des ressources alimentaires

En zones tropicales, des animaux au gabarit plus réduit que dans les zones tempérées sont plus fréquemment rencontrés, en lien une meilleure thermorégulation et des dépenses énergétiques moindres (phénomène d'adaptation écologique) [Becker and Stewart, 2011]. Chez les animaux, le dépôt préférentiel de gras est un des méchanismes d'adaptation en réponse aux périodes où la ressource alimentaire est moins abondante [Mirkena et al., 2010]. Ainsi plusieurs gènes et régions liés à la croissance (*SOCS2* [Fernández-Pérez et al., 2016]), au métabolisme lipidique ont été mis en exergue chez plusieurs populations bovines adaptés aux conditions tropicales : Chez les ovins, l'allocation préférentielle du gras au niveau de la croupe ou de la queue ("Fat rump", "Fat-tailed") est un exemple concret de l'adaptation des animaux à des contextes climatiques difficiles. La localisation externe de la graisse dans la queue permet une meilleure dissipation de la chaleur du reste du corps [Degen and Shkolnik, 1978] en raison d'une moindre isolation du corps par le tissu adipeux. Cette graisse représente également une réserve d'énergie qui peut être mobilisée en période de déficit alimentaire [Chilliard et al., 2000, Atti et al., 2004]. La fourniture d'énergie par le tissu adipeux est accompagnée d'une certaine formation d'eau métabolique qui pourrait aider à combler les besoins en eau dans une certaine mesure [Sileshi et al., 2003, Sejian et al., 2017]. Les gènes associés à ce phénotype ne sont pas clairement définis bien que de récentes études sur les signatures de sélection de mouton à queue grasse aient été entreprises [Moradi et al., 2012, Mwacharo et al., 2017, Mastrangelo et al., 2019a, Mastrangelo et al., 2019b, Wiener et al., 2021]. Des récentes études on permis de trouver certains gènes associés de façon récurrente à des signatures de sélection : par exemple *PDGFRA*, lié au dépôt de gras chez des moutons lybiens à queue grasse [Mastrangelo et al., 2019b] sous sélection chez des moutons soudanais [Abied et al., 2021], ainsi que le gène *CAMK4* [Abied et al., 2021] sur le chromosome OAR5 est associé à la morphologie de la queue et le dépôt lipidique chez plusieurs races à queue grasse [Moradi et al., 2012, Ma et al., 2018, Abdel-Gader et al., 2015, Ahbara et al., 2019, Li et al., 2020] [Li et al., 2020]. Les gènes *CDS2* [Mastrangelo et al., 2019a], *PROKR1* [Yuan et al., 2017], et *BMP2* [Mastrangelo et al., 2019a, Pan et al., 2019, Lu et al., 2020] ont été trouvés sous sélection chez les moutons à queue grasse Tunisien [Baazaoui et al., 2021].

Il en résulte que les gènes candidats sont associés avec divers processus et voies physiologiques, moléculaires et cellulaires [4]

4. Concernant les niveaux d'expression, d'autres gènes sont impliqués dans le métabolisme lipidique chez les moutons : le gène *FABP4* est beaucoup plus exprimé dans la queue grasse chez la Lori-Bakhtiari, une race ovine iranienne [Bakhtiarizadeh et al., 2013, Bakhtiarizadeh and Alamouti, 2020]. Cependant une étude récente [Wang et al., 2021] basée sur les séquences d'ARN a permis de déterminer que les gènes *SCD, ESR1,*

3.3.3 Adaptation aux agents pathogènes et à leurs vecteurs

Les agents pathogènes et certains de leurs vecteurs exercent une pression de sélection majeure sur les races locales, notamment en milieu tropical. Les régions sous sélection associées aux gènes de la famille BOLA (bovine lymphocyte antigen) sont statistiquement significatives dans plusieurs études chez la Borgou, un métissage Zébu X Taurin africain [Flori et al., 2014], les bovins africains en général [Kim et al., 2017]. De manière générale, les gènes de la famille *BOLA*, reliée au complexe majeur d'histocompatibilité (CMH) spécifique des bovins, sont retrouvés associés à des signatures de sélection chez des races bovines africaines [Kim et al., 2017] dont la Borgou [Flori et al., 2014] ou encore des hybrides synthétiques *Taurins X Zébus* [Goszczynski et al., 2018]. Le CMH est un ensemble complexe de gènes participant aux réponses immunitaires, étant la clef de l'immunité cellulaire et de la communication entre les cellules qui travaillent à la protection de l'organisme contre des agressions d'organismes pathogènes. Une autre famille associé à l'immunité est constituée des gènes *TRIM* (tripartite motif-containing) exprimés en réponse aux interférons et sont impliqués dans un large éventail de processus biologiques associés à l'immunité innée. Ce type de gènes a également été trouvés sous sélection chez la Muturu, une race bovine africaine [Tijjani et al., 2019].

Plus particulièrement, les parasites gastro-intestinaux, les trypanosomes, agents de la Trypanosomose Animale Africaine et *Bacillus anthracis*, l'agent du charbon bactéridien ont eu un impact sur la diversité génétique adaptative des bovins de même que les tiques, vectrices de certaines maladies.

Les parasites gastro-intestinaux Concernant les parasites gastro-intestinaux, problématique récurrente chez les petits-ruminants, le gène *SOCS2* joue un rôle important pour des phénotypes adaptaifs chez les moutons, comme la réponse immunitaire à la suite d'une infection chez des races australiennes [Al Kalaldeh et al., 2019] et la resistance à des nématodes gastro-intestinaux (*Haemonchus contortus*) chez les races Dorper, Katahdin, et Sainte Croix [Estrada-Reyes et al., 2019] et les moutons soudanais du désert [Abied et al., 2021].

Les trypanosomes, agents de la Trypanosomose Animale Africaine La Trypanosomose Animale Africaine, causée par 3 espèces de trypanosome (*T.congolense*, *T.vivax* et *T. brucei*), est responsable d'une réduction de la productivité de certaines races bovines dans les aires de répartition des glossines, vectrices de la maladie, dans les zones humides à sub-humides de l'Afrique de l'Ouest. Les taurins Ouest-Africains (N'Dama, Baoulé, Somba, Lagunaire, Muturu) sont connus pour avoir la capacité de contrôler la parasitémie et l'anémie associées à la maladie. Ils ont une plus forte capacité à se développer et à produire dans les aires infestées par les glossines que les zébus africains ou les taurins européens qui sont sensibles à la maladie [Mattioli et al., 2000]. La proportion de métissage semble jouer dans le degré de

EMR1 et PHYH jouaient un rôle direct dans le dépôt de gras dans la queue et que *GPAM, STAT3, ALDH1A1* régule indirectement l'évolution des queues grasses chez les moutons chinois Bashby. soutenant l'hypothèse d'une origine multigénique de l'adaptation à des environnements difficiles.

tolérance à la Trypanosomose Animale Africaine [Freeman et al., 2004]. Les races taurines ouest-africaines dites trypanotolérantes [5] représentent une faible proportion (6%) parmi l'ensemble des populations bovines africaines et seulement 17% des bovins présents dans les aires infestées par les glossines [Agyemang, 2005]. Les gènes suivants ont été trouvés sous sélection chez la N'Dama [Kim et al., 2017] en mesurant l'étendue du déséquilibre de liaison par rapport à d'autres races bovines africaines présentant une introgression indicine : *HCRTR1* lié au comportement, *SLC40A1, STOM, SBDS, EPB42, RPS26* associés à l'anémie, un des symptômes de la maladie avec *STOM* et *EPB42* complètement fixés.

Tiques Les climats tropicaux créent des conditions climatiques idéales pour l'expansion des tiques, vecteurs principaux d'hémoparasites (anaplasmose, babésiose, cowdriose). Quelques tests d'associations avec des traits phénotypiques caractérisant la résistance aux tiques [Mapholi et al., 2016] et aux hémoparasitoses [Sollero et al., 2017, Cavani et al., 2020] ont permis de déterminer des gènes candidats.

***Bacillus anthracis*, l'agent du charbon bactéridien** (ou anthrax), maladie la plus ancienne présente mondialement, provoque de forts taux de mortalité chez les ruminants [Velimirovic, 1984]. La resurgence des épidémies est liée aux facteurs climatiques (temperature, précipitations, radiations solaires) du fait de la persistence des spores de *Bacillus anthracis* dans les sols [Dorofeev and Blagoveshchenskaia, 1978, Carlson et al., 2019]. Le gène *ANTXR2*, qui code pour le principal récepteur médiateur de la létalité de la toxine produite par *Bacillus anthracis* [Arévalo et al., 2014, Liu et al., 2012] responsable de l'anthrax a été trouvé sous sélection chez plusieurs populations bovines tropicales notamment celle de l'Afrique de l'Ouest [Gautier et al., 2009], et chez dans les populations ovines [Lv et al., 2014] et également chez les populations bovines du pourtour méditerranéen [Flori et al., 2019] . Il est aussi associé à des covariables climatiques représentatives du climat méditerranéen [Flori et al., 2019] suggérant une forte pression de sélection de ce pathogène sur les populations bovines locales de la zone.

5. La trypanotolérance (reduction des effets délétères de la charge du pathogène dans l'hôte) est un mécanimse différent de la trypanorésistance (mécanisme gérer par le système immunitaire dans le but de réduire la charge du pathogène suivant l'infection.

Deuxième partie

Justification du projet et axes de recherche

L'Océan Indien a joué un rôle de premier plan dans la migration des populations de bétail conduite par les populations humaines entre l'Afrique de l'Est, le Moyen-Orient et l'Asie du Sud-Ouest. Cependant, l'origine et la diversité génétique des populations bovines et ovines vivant dans les îles de l'Océan Indien restent mal étudiées et relativement peu claires. Leur meilleure compréhension peut permettre de compléter l'histoire récente des populations humaines dans cette région qui a longtemps affiché des niveaux importants d'interaction, de commerce, de migration et de translocation des espèces domestiques dans le temps [Boivin et al., 2013, Beaujard, 2019b, Beaujard, 2019c].

Notre objectif principal sera d'étudier les histoires démographique et adaptative des bovins et des ovins de l'Océan Indien à partir de données génétiques de nouvelle génération (génotypes SNP à moyenne et haute densité en particulier). Les races bovines Zébus Mahorais et de Madagascar et ovine de Mayotte vivant dans les régions méridionales au sud-ouest de la zone seront en particulier étudiées et participent à une analyse plus globale des bovins et ovins de l'Océan Indien intégrant les données génétiques déjà obtenues pour d'autres animaux vivant dans des écosystèmes distincts (sous-continent indien et est de l'Afrique). Les données génétiques d'un large panel de races bovines et ovines stockées dans la base de données WIDDE seront notamment exploitées pour explorer la diversité génétique de ces races ainsi que des séquences génomiques individuelles obtenues sur certaines populations. Plusieurs paramètres démographiques (p.e. proportions d'ascendances, dates d'évènements de métissage, consanguinité récente) seront inférés et les gènes sous sélection et associés à des variables climatiques et leurs principales fonctions seront identifiés.

Cet objectif principal se décline en plusieurs questions de recherche :

Quelles sont les caractéristiques phénotypiques des populations ovines et bovines autochtones présentes sur l'île de Mayotte ?

Pour cela, l'analyse de données morphologiques récoltées durant le projet permettra de déterminer l'inventaire des variables catégorielles (couleur de la robe, type de corne) et le gabarit moyen des races locales bovine et ovine présentes sur l'île de Mayotte. Cette question sera traitée dans le chapitre 4 (page 109).

Quelle est la structure génétique des populations bovines et ovines de l'île de Mayotte ?

La structure génétique des populations raciales locales bovines et ovines de l'île de Mayotte sera déterminée à partir des données de génotypage (Haute densité pour les bovins, moyenne densité pour les ovins) en comparaison avec des jeux de données globaux représentatifs de la diversité génétique mondiale. en utilisant des méthodes de calcul des indices de fixation, d'analyse multidimensionnelle (ACP, clustering), d'estimation de la consanguinité par le calcul des ROH. Les résultats seront présentés dans les articles présentés dans les sections 5.2 (page 143) pour les zébus et 6.3 (page 178) pour les ovins.

Quelle est l'origine des bovins locaux de Mayotte et Madagascar ?

A partir des données génomiques, il sera possible d'inférer les histoires démographiques de ces populations bovines en testant différents scénarios de métissage à partir des fréquences alléliques de chaque SNP et l'estimation du déséquilibre de liaison. Les résultats seront présentés dans les articles présentés dans le chapitre 5 (page 141) et dans l'article en section 5.2 (page 143).

Quelles sont les régions génomiques et les gènes sous sélection dans les génomes des populations locales bovines de Mayotte et Madagascar et ovine de Mayotte et les principales pressions de sélection qui se sont exercées sur ces populations ?

L'identification des signatures de sélection de ces populations pour mieux comprendre leur histoire adaptative sera réalisée à partir de l'estimation de l'étendue du déséquilibre de liaison (EHH) et des scores dérivés (iHs, Rsb) par le scan du génome sur les chromosomes autosomes. Les gènes candidats situés dans ces signatures seront ensuite annotés fonctionnellement à l'aide d'outils de la biologie des systèmes. Les résultats seront présentés dans l'article en section 5.2 (page 143) pour les bovines et dans la section 6.4.2 (page 202) du chapitre 6 pour les ovins.

Quelles sont les régions génomiques et les gènes impliqués dans l'adaptation des populations locales bovines et ovines aux différentes zones climatiques composant le littoral de l'Océan Indien ?

Pour cela, le but sera de réaliser une carte génétique/génomique d'adaptation au moyen d'un criblage des régions génomiques associées au climat du littoral de l'Océan Indien en utilisant des données de génotypage en moyenne densité de 17 populations bovines et 9 populations ovines disponibles. L'approche consistera à effectuer des tests d'association à l'échelle du génome entre les marqueurs autosomiques et les variables climatiques et déterminer les gènes candidats. Ensuite, il s'agira de déterminer si les gènes candidats sont sous sélection par le calcul de scores de différentiation *XtX*. Les résultats seront présentés dans le chapitre 7 (page 216).

La stratégie développée pour répondre à ces questions de recherche est résumée sur la Figure 3.8. Elle s'appuiera sur l'étude de plusieurs jeux de données issus de regroupements de différents dataset déjà publiés et disponibles sur le site WIDDE (`widde.toulouse.inra.fr`) [Sempéré et al., 2015]. Pour les bovins, les datasets publiés dans [Matukumalli et al., 2009, Gautier et al., 2009, Gautier et al., 2010, Decker et al., 2014, Flori et al., 2009], le projet PERSAfrica et les données concernant le zébu mahorais, issus du projet européen DEFI ANIMAL et dont le protocole d'obtention sera décrit dans les Chapitres 4 et 5, seront utilisées. Pour les ovins, les données de génotypage en moyenne densité de la population ovine mahoraise seront intégrées au dataset disponible présenté dans les publications de [Young et al., 2011, Kijas et al., 2012, Rochus et al., 2018].

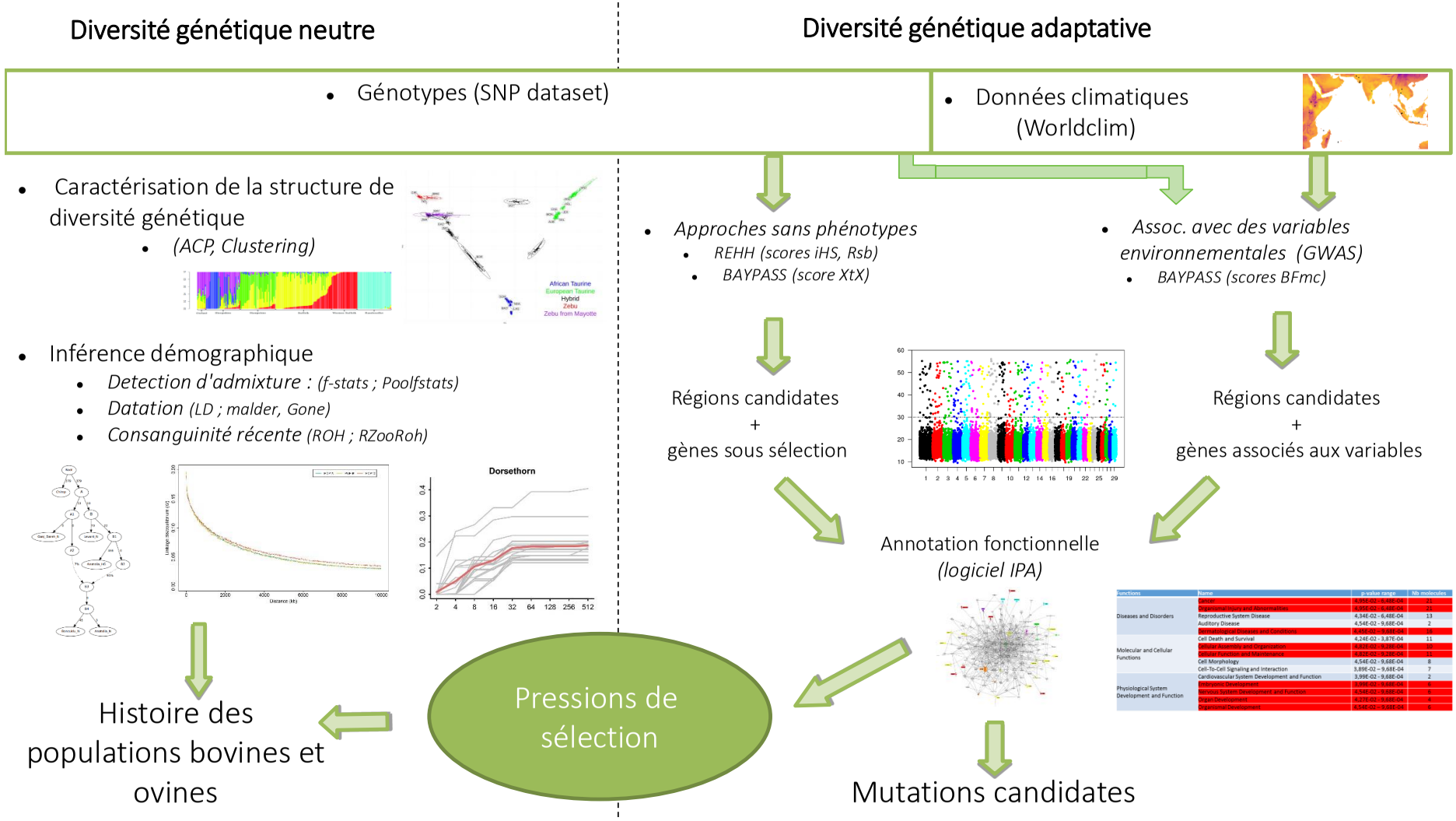


FIGURE 3.8 – Résumé graphique de la stratégie globale d'étude des données de génotypage

Bibliographie

[Abdel-Gader et al., 2015] Abdel-Gader, A. Z., Musa, A., Musa, L. M.-A., and others (2015). Milk yield and reproductive performance of Friesian cows under Sudan tropical conditions.

[Abeygunawardena and Dematawewa, 2004] Abeygunawardena, H. and Dematawewa, C. M. B. (2004). Prepubertal and postpartum anestrus in tropical Zebu cattle. *Animal Reproduction Science*, 82-83 :373–387.

[Abied et al., 2021] Abied, A., Ahbara, A. M., Berihulay, H., Xu, L., Islam, R., El-Hag, F. M., Rekik, M., Haile, A., Han, J.-L., Ma, Y., Zhao, Q., and Mwacharo, J. M. (2021). Genome Divergence and Dynamics in the Thin-Tailed Desert Sheep From Sudan. *Frontiers in Genetics*, 12 :1296.

[Acharya, 1982] Acharya, R. M. (1982). *Sheep and goat breeds of India*. Number 30 in FAO animal production and health paper. Food and Agriculture Organization of the United Nations, Rome.

[Agyemang, 2005] Agyemang, K. (2005). *Trypanotolerant Livestock in the Context of Trypanosomiasis Intervention Strategies*. Food & Agriculture Org. Google-Books-ID : KTF0wlUVD1YC.

[Ahbara et al., 2019] Ahbara, A., Bahbahani, H., Almathen, F., Al Abri, M., Agoub, M. O., Abeba, A., Kebede, A., Musa, H. H., Mastrangelo, S., Pilla, F., Ciani, E., Hanotte, O., and Mwacharo, J. M. (2019). Genome-Wide Variation, Candidate Regions and Genes Associated With Fat Deposition and Tail Morphology in Ethiopian Indigenous Sheep. *Frontiers in Genetics*, 9.

[Al Kalaldeh et al., 2019] Al Kalaldeh, M., Gibson, J., Lee, S. H., Gondro, C., and van der Werf, J. H. J. (2019). Detection of genomic regions underlying resistance to gastrointestinal parasites in Australian sheep. *Genetics Selection Evolution*, 51(1) :37.

[Aldenderfer, 2011] Aldenderfer, M. (2011). Peopling the Tibetan plateau : insights from archaeology. *High Altitude Medicine & Biology*, 12(2) :141–147.

[Aldersey et al., 2020] Aldersey, J. E., Sonstegard, T. S., Williams, J. L., and Bottema, C. D. K. (2020). Understanding the effects of the bovine POLLED variants. *Animal Genetics*, 51(2) :166–176.

[Alexander et al., 2009] Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9) :1655–1664.

[Allibert et al., 1989] Allibert, C., Argant, A., and Argant, J. (1989). Le site de Dembeni (Mayotte, Archipel des Comores). Mission 1984. *Etudes Océan Indien*, 11 :63–172.

[Anderson et al., 2018] Anderson, A., Camens, A., Clark, G., and Haberle, S. (2018). Investigating premodern colonization of the Indian Ocean : the remote islands enigma. *Connecting Continents : archaeology and history in the Indian Ocean*, pages 30–67.

[Anonymous, 1950] Anonymous, N. (1950). L'élevage du mouton à Madagascar. *Revue d'élevage et de médecine vétérinaire des pays tropicaux*, 4(4) :204.

[Anonymous, 1952] Anonymous, N. (1952). Rapport sur l'élevage à Madagascar. *Revue d'élevage et de médecine vétérinaire des pays tropicaux*, 5 :139.

[Arévalo et al., 2014] Arévalo, M. T., Navarro, A., Arico, C. D., Li, J., Alkhatib, O., Chen, S., Diaz-Arévalo, D., and Zeng, M. (2014). Targeted silencing of anthrax toxin receptors protects against anthrax toxins. *Journal of Biological Chemistry*, 289(22) :15730–15738.

[Aswani, 2007] Aswani, O. (2007). *Genetic diversity and relationships of African sheep : AY chromosome perspective.* PhD Thesis, Egerton University.

[Atti et al., 2004] Atti, N., Bocquier, F., and Khaldi, G. (2004). Performance of the fat-tailed Barbarine sheep in its environment : adaptive capacity to alternation of underfeeding and re-feeding periods. A review. *Animal Research*, 53(3) :165–176.

[Baazaoui et al., 2021] Baazaoui, I., Bedhiaf-Romdhani, S., Mastrangelo, S., and Ciani, E. (2021). Genome-wide analyses reveal population structure and identify candidate genes associated with tail fatness in local sheep from a semi-arid area. *Animal*, 15(4) :100193.

[Bahbahani et al., 2018] Bahbahani, H., Afana, A., and Wragg, D. (2018). Genomic signatures of adaptive introgression and environmental adaptation in the Sheko cattle of southwest Ethiopia. *PloS One*, 13(8) :e0202479.

[Bahbahani et al., 2015] Bahbahani, H., Clifford, H., Wragg, D., Mbole-Kariuki, M. N., Tassell, C. V., Sonstegard, T., Woolhouse, M., and Hanotte, O. (2015). Signatures of positive selection in East African Shorthorn Zebu : A genome-wide single nucleotide polymorphism analysis. *Scientific Reports*, 5(1) :1–13.

[Bakhtiarizadeh and Alamouti, 2020] Bakhtiarizadeh, M. R. and Alamouti, A. A. (2020). RNA-Seq based genetic variant discovery provides new insights into controlling fat deposition in the tail of sheep. *Scientific Reports*, 10(1) :13525.

[Bakhtiarizadeh et al., 2013] Bakhtiarizadeh, M. R., Moradi-Shahrbabak, M., and Ebrahimie, E. (2013). Underlying functional genomics of fat deposition in adipose tissue. *Gene*, 521(1) :122–128.

[Banerjee et al., 2011] Banerjee, S., Galloway, S. M., and Davis, G. H. (2011). Distribution of prolific Garole sheep in West Bengal, India. *Animal Genetic Resources/Resources génétiques animales/Recursos genéticos animales*, 48 :29–35.

[Barbato et al., 2020] Barbato, M., Hailer, F., Upadhyay, M., Del Corvo, M., Colli, L., Negrini, R., Kim, E.-S., Crooijmans, R. P. M. A., Sonstegard, T., and Ajmone-Marsan, P. (2020). Adaptive introgression from indicine cattle into white cattle breeds from Central Italy. *Scientific Reports*, 10(1) :1279.

[Barrett et al., 2006] Barrett, T., Pastoret, P.-P., and Taylor, W. P., editors (2006). *Rinderpest and peste des petits ruminants : virus plagues of large and small ruminants*. Biology of animal infections. Academic Press, London. OCLC : 249810303.

[Beaujard, 2015] Beaujard, P. (2015). East Africa and oceanic exchange networks between the first and the fifteenth centuries. *Afriques-Debats Methodes Et Terrains D Histoire*, 6. WOS :000420334300001.

[Beaujard, 2019a] Beaujard, P. (2019a). *The Worlds of the Indian Ocean : A Global History*. Cambridge University Press, 1 edition.

[Beaujard, 2019b] Beaujard, P. (2019b). *The Worlds of the Indian Ocean : A Global History : Volume 1 : From the Fourth Millennium BCE to the Sixth Century CE*, volume 1. Cambridge University Press, Cambridge.

[Beaujard, 2019c] Beaujard, P. (2019c). *The worlds of the Indian Ocean : A global history : Volume 2 : From the Seventh Century to the Fifteenth Century CE*, volume 2. Cambridge University Press, Cambridge, United Kingdom ; New York, NY.

[Becker and Stewart, 2011] Becker, J. A. and Stewart, L. K. (2011). Heat-Related Illness. *American Family Physician*, 83(11) :1325–1330.

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

[Bertrand et al., 2019] Bertrand, A., Kadri, N. K., Flori, L., Gautier, M., and Druet, T. (2019). RZooRoH : An R package to characterize individual genomic autozygosity and identify homozygous-by-descent segments. *Methods in Ecology and Evolution*, 10(6).

[Blanchy, 2015] Blanchy, S. (2015). Anjouan (Comores), un nøeud dans les réseaux de l'océan Indien. Émergence et rôle d'une société urbaine lettrée et marchande (XVIIe-XXe siècle). *Afriques. Débats, méthodes et terrains d'histoire*, (06).

[Boitard et al., 2016] Boitard, S., Rodríguez, W., Jay, F., Mona, S., and Austerlitz, F. (2016). Inferring Population Size History from Large Samples of Genome-Wide Molecular Data - An Approximate Bayesian Computation Approach. *PLOS Genetics*, 12(3) :e1005877.

[Boivin et al., 2013] Boivin, N., Crowther, A., Helm, R., and Fuller, D. Q. (2013). East Africa and Madagascar in the Indian Ocean world. *Journal of World Prehistory*, 26(3) :213–281.

[Boivin and Fuller, 2009] 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*, 22(2) :113–180.

[Bradford and Inounu, 1996] Bradford, G. and Inounu, I. (1996). Prolific breeds of Indonesia. *by MH Fahmy. CAB International : Wallingford, UK*, pages 137–145.

[Brucato et al., 2019] Brucato, N., Fernandes, V., Kusuma, P., Černý, V., Mulligan, C. J., Soares, P., Rito, T., Besse, C., Boland, A., Deleuze, J.-F., Cox, M. P., Sudoyo, H., Stoneking, M., Pereira, L., and Ricaut, F.-X. (2019). Evidence of Austronesian Genetic Lineages in East Africa and South Arabia : Complex Dispersal from Madagascar and Southeast Asia. *Genome Biology and Evolution*, 11(3) :748–758.

[Brucato et al., 2018] Brucato, N., Fernandes, V., Mazières, S., Kusuma, P., Cox, M. P., Ng'ang'a, J. W., Omar, M., Simeone-Senelle, M.-C., Frassati, C., Alshamali, F., Fin, B., Boland, A., Deleuze, J.-F., Stoneking, M., Adelaar, A., Crowther, A., Boivin, N., Pereira, L., Bailly, P., Chiaroni, J., and Ricaut, F.-X. (2018). The Comoros Show the Earliest Austronesian Gene Flow into the Swahili Corridor. *The American Journal of Human Genetics*, 102(1) :58–68.

[Brucato et al., 2017] Brucato, N., Kusuma, P., Beaujard, P., Sudoyo, H., Cox, M. P., and Ricaut, F.-X. (2017). Genomic admixture tracks pulses of economic activity over 2,000 years in the Indian Ocean trading network. *Scientific Reports*, 7(1) :2919.

[Bruford and Townsend, 2006] Bruford, M. W. and Townsend, S. J. (2006). Mitochondrial DNA diversity in modern sheep. *Documenting domestication : New genetic and archaeological paradigms*, pages 306–316.

[Cadzow et al., 2014] Cadzow, M., Boocock, J., Nguyen, H. T., Wilcox, P., Merriman, T. R., and Black, M. A. (2014). A bioinformatics workflow for detecting signatures of selection in genomic data. *Frontiers in Genetics*, 5 :293.

[Cai et al., 2014] Cai, D., Sun, Y., Tang, Z., Hu, S., Li, W., Zhao, X., Xiang, H., and Zhou, H. (2014). The origins of Chinese domestic cattle as revealed by ancient DNA analysis. *Journal of Archaeological Science*, 41 :423–434.

[Cai et al., 2011] Cai, D., Tang, Z., Yu, H., Han, L., Ren, X., Zhao, X., Zhu, H., and Zhou, H. (2011). Early history of Chinese domestic sheep indicated by ancient DNA analysis of Bronze Age individuals. *Journal of Archaeological Science*, 38(4) :896–902.

[Cao et al., 2021] Cao, Y.-H., Xu, S.-S., Shen, M., Chen, Z.-H., Gao, L., Lv, F.-H., Xie, X.-L., Wang, X.-H., Yang, H., Liu, C.-B., Zhou, P., Wan, P.-C., Zhang, Y.-S., Yang, J.-Q., Pi, W.-H., Hehua, E., Berry, D. P., Barbato, M., Esmailizadeh, A., Nosrati, M., Salehian-Dehkordi, H., Dehghani-Qanatqestani, M., Dotsev, A. V., Deniskova, T. E., Zinovieva, N. A., Brem, G., Štěpánek, O., Ciani, E., Weimann, C., Erhardt, G., Mwacharo, J. M., Ahbara, A., Han, J.-L., Hanotte, O., Miller, J. M., Sim, Z., Coltman, D., Kantanen, J., Bruford, M. W., Lenstra, J. A., Kijas, J., and Li, M.-H. (2021). Historical Introgression from Wild Relatives Enhanced Climatic Adaptation and Resistance to Pneumonia in Sheep. *Molecular Biology and Evolution*, 38(3) :838–855.

[Carlson, 2018] Carlson, B. M. (2018). *Human embryology and developmental biology*. Elsevier Health Sciences.

[Carlson et al., 2019] Carlson, C. J., Kracalik, I. T., Ross, N., Alexander, K. A., Hugh-Jones, M. E., Fegan, M., Elkin, B. T., Epp, T., Shury, T. K., Zhang, W., and others (2019). The global distribution of Bacillus anthracis and associated anthrax risk to humans, livestock and wildlife. *Nature microbiology*, 4(8) :1337–1343.

[Cavani et al., 2020] Cavani, L., Braz, C. U., Giglioti, R., Okino, C. H., Gulias-Gomes, C. C., Caetano, A. R., Oliveira, M. C. d. S., Cardoso, F. F., and de Oliveira, H. N. (2020). Genomic Study of Babesia bovis Infection Level and Its Association With Tick Count in Hereford and Braford Cattle. *Frontiers in Immunology*, 11 :1905.

[Chan et al., 2010] Chan, E. K. F., Nagaraj, S. H., and Reverter, A. (2010). The evolution of tropical adaptation : comparing taurine and zebu cattle. *ANIMAL GENETICS*, 41(5) :467–477.

[Charoensook et al., 2013] Charoensook, R., Knorr, C., Brenig, B., and Gatphayak, K. (2013). Thai pigs and cattle production, genetic diversity of livestock and strategies for preserving animal genetic resources. *Maejo International Journal of Science and Technology*, 7(1) :113–132.

[Cheke, 2010] Cheke, A. (2010). The timing of arrival of humans and their commensal animals on Western Indian Ocean oceanic islands. *Phelsuma*, (18) :38–69.

[Chen et al., 2015] Chen, F. H., Dong, G. H., Zhang, D. J., Liu, X. Y., Jia, X., An, C.-B., Ma, M. M., Xie, Y. W., Barton, L., Ren, X., and others (2015). Agriculture facilitated permanent human occupation of the Tibetan Plateau after 3600 BP. *science*, 347(6219) :248–250.

[Chen et al., 2018] Chen, N., Cai, Y., Chen, Q., Li, R., Wang, K., Huang, Y., Hu, S., Huang, S., Zhang, H., Zheng, Z., Song, W., Ma, Z., Ma, Y., Dang, R., Zhang, Z., Xu, L., Jia, Y., Liu, S., Yue, X., Deng, W., Zhang, X., Sun, Z., Lan, X., Han, J., Chen, H., Bradley, D. G., Jiang, Y., and Lei, C. (2018). Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nature Communications*, 9(1).

[Chessa et al., 2009] Chessa, B., Pereira, F., Arnaud, F., Amorim, A., Goyache, F., Mainland, I., Kao, R. R., Pemberton, J. M., Beraldi, D., Stear, M. J., and others (2009). Revealing the history of sheep domestication using retrovirus integrations. *Science*, 324(5926) :532–536.

[Chilliard et al., 2000] Chilliard, Y., Ferlay, A., Faulconnier, Y., Bonnet, M., Rouel, J., and Bocquier, F. (2000). Adipose tissue metabolism and its role in adaptations to undernutrition in ruminants. *Proceedings of the Nutrition Society*, 59(1) :127–134.

[Ciani et al., 2020] Ciani, E., Mastrangelo, S., Da Silva, A., Marroni, F., Ferenčaković, M., Ajmone-Marsan, P., Baird, H., Barbato, M., Colli, L., Delvento, C., and others (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*, 52 :1–14.

[Clark and Williams, 1978] Clark, J. D. and Williams, M. a. J. (1978). Recent Archaeological Research in Southeastern Ethiopia. 1974 - 1975. *Annales d'Éthiopie*, 11(1) :19–44.

[Clutton-Brock, 1993] Clutton-Brock, J. (1993). The spread of domestic animals in Africa. In *The archaelogy of Africa : Foods, metals and towns*, pages 43–60. T. Shaw, P. Sinclair, B. Andah, A. Okpoko, London : Routledge.

[Collier et al., 2012] Collier, R. J., Gebremedhin, K., Macko, A. R., and Roy, K. S. (2012). Genes Involved in the Thermal Tolerance of Livestock. In Sejian, V., Naqvi, S., Ezeji, T., Lakritz, J., and Lal, R., editors, *Environmental Stress and Amelioration in Livestock Production*, pages 379–410. Springer, Berlin, Heidelberg.

[Consortium et al., 2010] Consortium, I. S. G., Archibald, A., Cockett, N., Dalrymple, B., Faraut, T., Kijas, J., Maddox, J., McEwan, J., Hutton Oddy, V., Raadsma, H., et al. (2010). The sheep genome reference sequence : a work in progress. *Animal genetics*, 41(5) :449–453.

[Consortium et al., 2009] Consortium, T. B. G. S. a. A., Elsik, C. G., Tellam, R. L., and Worley, K. C. (2009). The Genome Sequence of Taurine Cattle : A Window to Ruminant Biology and Evolution. *Science*, 324(5926) :522–528.

[Coop et al., 2010] Coop, G., Witonsky, D., Di Rienzo, A., and Pritchard, J. K. (2010). Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics*, 185(4) :1411–1423.

[Corbett-Detig and Nielsen, 2017] Corbett-Detig, R. and Nielsen, R. (2017). A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy. *PLOS Genetics*, 13(1) :e1006529.

[Couix et al., 2016] Couix, N., Gaillard, C., Lauvie, A., Mugnier, S., and Verrier, E. (2016). Des races localement adaptées et adoptées, une condition de la durabilité des activités d'élevage. *Cahiers Agricultures*, 25(6) :650009.

[Cymbron et al., 2005] Cymbron, T., Freeman, A. R., Isabel Malheiro, M., Vigne, J.-D., and Bradley, D. G. (2005). Microsatellite diversity suggests different histories for Mediterranean and Northern European cattle populations. *Proceedings of the Royal Society B : Biological Sciences*, 272(1574) :1837–1843.

[da Fonseca et al., 2019] da Fonseca, R. R., Ureña, I., Afonso, S., Pires, A. E., Jørsboe, E., Chikhi, L., and Ginja, C. (2019). Consequences of breed formation on patterns of genomic diversity and differentiation : the case of highly diverse peripheral Iberian cattle. *BMC Genomics*, 20(1) :334.

[Darwin, 1868] Darwin, C. (1868). *The variation of animals and plants under domestication*, volume 2. Cambridge University Press.

[Dechambre, 1951] Dechambre (1951). Origine des animaux domestiques de Madagascar - CORE.

[Decker et al., 2014] Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Molina Alcalá, A., Sonstegard, T. S., Hanotte, O., Götherström, A., Seabury, C. M., Praharani, L., Babar, M. E., Correia de Almeida Regitano, L., Yildiz, M. A., Heaton, M. P., Liu, W.-S., Lei, C.-Z., Reecy, J. M., Saif-Ur-Rehman, M., Schnabel, R. D., and Taylor, J. F. (2014). Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *PLoS Genetics*, 10(3) :e1004254.

[Decker et al., 2009] Decker, J. E., Pires, J. C., Conant, G. C., McKay, S. D., Heaton, M. P., Chen, K., Cooper, A., Vilkki, J., Seabury, C. M., Caetano, A. R., Johnson, G. S., Brenneman, R. A., Hanotte, O., Eggert, L. S., Wiener, P., Kim, J.-J., Kim, K. S., Sonstegard, T. S., Tassell, C. P. V., Neibergs, H. L., McEwan, J. C., Brauning, R., Coutinho, L. L., Babar, M. E., Wilson, G. A., McClure, M. C., Rolf, M. M., Kim, J., Schnabel, R. D., and Taylor, J. F. (2009). Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proceedings of the National Academy of Sciences*, 106(44) :18644–18649.

[Degen and Shkolnik, 1978] Degen, A. A. and Shkolnik, A. (1978). Thermoregulation in Fat-Tailed Awassi, a Desert Sheep, and in German Mutton Merino, a Mesic Sheep. *Physiological Zoology*, 51(4) :333–339.

[Deng et al., 2020] Deng, J., Xie, X.-L., Wang, D.-F., Zhao, C., Lv, F.-H., Li, X., Yang, J., Yu, J.-L., Shen, M., Gao, L., Yang, J.-Q., Liu, M.-J., Li, W.-R., Wang, Y.-T., Wang, F., Li, J.-Q., Hehua, E., Liu, Y.-G., Shen, Z.-Q., Ren, Y.-L., Liu, G.-J., Chen, Z.-H., Gorkhali, N. A., Rushdi, H. E., Salehian-Dehkordi, H., Esmailizadeh, A., Nosrati, M., Paiva, S. R., Caetano, A. R., Štěpánek, O., Olsaker, I., Weimann, C., Erhardt, G., Curik, I., Kantanen, J., Mwacharo, J. M., Hanotte, O., Bruford, M. W., Ciani, E., Periasamy, K., Amills, M., Lenstra, J. A., Han, J.-L., Zhang, H.-P., Li, L., and Li, M.-H. (2020). Paternal Origins and Migratory Episodes of Domestic Sheep. *Current Biology*, 30(20) :4085–4095.e6.

[Deschamps, 1961] Deschamps, H. (1961). *Histoire de Madagascar*. Mondes d'Outre-Mer. Editions Berger-Levrault.

[Dobney and Larson, 2006] Dobney, K. and Larson, G. (2006). Genetics and animal domestication : new windows on an elusive process. *Journal of zoology*, 269(2) :261–271.

[Dodson et al., 2014] Dodson, J., Dodson, E., Banati, R., Li, X., Atahan, P., Hu, S., Middleton, R. J., Zhou, X., and Nan, S. (2014). Oldest directly dated remains of sheep in China. *Scientific Reports*, 4(1) :1–4.

[Dorofeev and Blagoveshchenskaia, 1978] Dorofeev, A. and Blagoveshchenskaia, G. (1978). Solar activity and anthrax epizootics. *Veterinariia*.

[Driscoll et al., 2009] Driscoll, C. A., Macdonald, D. W., and O'Brien, S. J. (2009). From wild animals to domestic pets, an evolutionary view of domestication. *Proceedings of the National Academy of Sciences*, 106(Supplement 1) :9971–9978.

[Druet et al., 2018] Druet, T., Flori, L., Olénski, K., Tokarska, M., Kaminski, S., and Gautier, M. (2018). On the path to recovery : characterizing inbreeding in the European Bison with a model-based approach.

[Druet and Gautier, 2017] Druet, T. and Gautier, M. (2017). A model-based approach to characterize individual inbreeding at both global and local genomic scales. *Molecular Ecology*, 26(20) :5820–5841.

[Dumont et al., 2013] Dumont, B., Fortun-Lamothe, L., Jouven, M., Thomas, M., and Tichit, M. (2013). Prospects from agroecology and industrial ecology for animal production in the 21st century. *animal*, 7(6) :1028–1043.

[Durand et al., 2011] Durand, E. Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular biology and evolution*, 28(8) :2239–2252.

[Edea et al., 2017a] Edea, Z., Dadi, H., Dessie, T., Kim, I.-H., and Kim, K.-S. (2017a). Association of MITF loci with coat color spotting patterns in Ethiopian cattle. *Genes & Genomics*, 39(3) :285–293.

[Edea et al., 2017b] Edea, Z., Dessie, T., Dadi, H., Do, K.-T., and Kim, K.-S. (2017b). Genetic Diversity and Population Structure of Ethiopian Sheep Populations Revealed by High-Density SNP Markers. *Frontiers in Genetics*, 8.

[Elsik et al., 2009] Elsik, C. G., Tellam, R. L., and Worley, K. C. (2009). The genome sequence of taurine cattle : A window to ruminant biology and evolution. the bovine genome sequencing and analysis consortium. *Science*, 324(5926) :522–528.

[Epstein, 1971] Epstein, H. (1971). *The origin of the domestic animals of Africa*. Africana Publishing Corporation.

[Estrada-Reyes et al., 2019] Estrada-Reyes, Z. M., Tsukahara, Y., Amadeu, R. R., Goetsch, A. L., Gipson, T. A., Sahlu, T., Puchala, R., Wang, Z., Hart, S. P., and Mateescu, R. G. (2019). Signatures of selection for resistance to Haemonchus contortus in sheep and goats. *BMC genomics*, 20(1) :1–14.

[FAO, 2012] FAO (2012). *Phenotypic characterization of animal genetic resources*. Number 11 in FAO animal production and health guidelines. Food and Agriculture Organization of the United Nations, Rome.

[FAO, 2018] FAO (2018). STATUS AND TRENDS OF ANIMAL GENETIC RESOURCES - 2018. Technical report, FAO, Rome.

[Fariello et al., 2013] Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M., and Servin, B. (2013). Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*, 193(3) :929–941.

[Fay and Wu, 2000] Fay, J. C. and Wu, C.-I. (2000). Hitchhiking Under Positive Darwinian Selection. *Genetics*, 155(3) :1405–1413.

[Felius, 1995] Felius, M. (1995). Cattle breeds : an encyclopedia. *Cattle breeds : an encyclopedia.*, (Ed. 1).

[Felius et al., 2014] Felius, M., Beerling, M.-L., Buchanan, D. S., Theunissen, B., Koolmees, P. A., and Lenstra, J. A. (2014). On the History of Cattle Genetic Resources. *Diversity*, 6(4) :705–750.

[Felius et al., 2011] Felius, M., Koolmees, P. A., Theunissen, B., Consortium, E. C. G. D., and Lenstra, J. A. (2011). On the Breeds of Cattle—Historic and Current Classifications. *Diversity*, 3(4) :660–692.

[Fernández-Pérez et al., 2016] Fernández-Pérez, L., Flores-Morales, A., Guerra, B., Díaz-Chico, J. C., and Iglesias-Gato, D. (2016). Growth hormone receptor signaling pathways and its negative regulation by socs2. *Restricted growth : clinical, genetic and molecular aspects. InTechOpen*, pages 125–45.

[Flad et al., 2007] Flad, R., Yuan, J., DB, L. S. M., Chen, F.-H., and Gao, X. (2007). Zooarchaeological evidence for animal domestication in northwest China. In *Late quaternary climate change and human adaptation in arid China*, pages 167–204. Elsevier, Amsterdam.

[Flori et al., 2009] Flori, L., Fritz, S., Jaffrézic, F., Boussaha, M., Gut, I., Heath, S., Foulley, J.-L., and Gautier, M. (2009). The genome response to artificial selection : a case study in dairy cattle. *PloS One*, 4(8) :e6595.

[Flori et al., 2012] Flori, L., Gonzatti, M. I., Thevenon, S., Chantal, I., Pinto, J., Berthier, D., Aso, P. M., and Gautier, M. (2012). A Quasi-Exclusive European Ancestry in the Senepol Tropical Cattle Breed Highlights the Importance of the slick Locus in Tropical Adaptation. *PLOS ONE*, 7(5) :e36133.

[Flori et al., 2019] Flori, L., Moazami-Goudarzi, K., Alary, V., Araba, A., Boujenane, I., Boushaba, N., Casabianca, F., Casu, S., Ciampolini, R., D'Acier, A. C., Coquelle, C., Delgado, J.-V., El-Beltagi, A., Hadjipavlou, G., Jousselin, E., Landi, V., Lauvie, A., Lecomte, P., Ligda, C., Marinthe, C., Martinez, A., Mastrangelo, S., Menni, D., Moulin, C.-H., Osman, M.-A., Pineau, O., Portolano, B., Rodellar, C., Saïdi-Mehtar, N., Sechi, T., Sempéré, G., Thévenon, S., Tsiokos, D., Laloë, D., and Gautier, M. (2019). A genomic map of climate adaptation in Mediterranean cattle breeds. *Molecular Ecology*, 28(5) :1009–1029.

[Flori et al., 2014] Flori, L., Thevenon, S., Dayo, G.-K., Senou, M., Sylla, S., Berthier, D., Moazami-Goudarzi, K., and Gautier, M. (2014). Adaptive admixture in the West African bovine hybrid zone : insight from the Borgou population. *Molecular Ecology*, 23(13) :3241–3257. WOS :000338014900009.

[Fontanesi et al., 2012] Fontanesi, L., Scotti, E., and Russo, V. (2012). Haplotype variability in the bovine MITF gene and association with piebaldism in Holstein and Simmental cattle breeds. *Animal Genetics*, 43(3) :250–256.

[Freeman et al., 2004] Freeman, A. R., Meghen, C. M., MacHUGH, D. E., Loftus, R. T., Achukwi, M. D., Bado, A., Sauveroche, B., and Bradley, D. G. (2004). Admixture and diversity in West African cattle populations. *Molecular Ecology*, 13(11) :3477–3487.

[Frichot et al., 2013] Frichot, E., Schoville, S. D., Bouchard, G., and François, O. (2013). Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. *Molecular Biology and Evolution*, 30(7) :1687–1699.

[Fuller et al., 2011] Fuller, D., Boivin, N., Hoogervorst, T., and Allaby, R. (2011). Across the Indian Ocean : the prehistoric movement of plants and animals. *Antiquity*, 85(328) :544–558.

[Fuller, 2006] Fuller, D. Q. (2006). Agricultural Origins and Frontiers in South Asia : A Working Synthesis. *Journal of World Prehistory*, 20(1) :1–86.

[Förstner and Moonen, 2003] Förstner, W. and Moonen, B. (2003). A metric for covariance matrices. *Geodesy-the Challenge of the 3rd Millennium*, pages 299–309.

[Gao et al., 2017] Gao, Y., Gautier, M., Ding, X., Zhang, H., Wang, Y., Wang, X., Faruque, M. O., Li, J., Ye, S., Gou, X., Han, J., Lenstra, J. A., and Zhang, Y. (2017). Species composition and environmental adaptation of indigenous Chinese cattle. *Scientific Reports*, 7(1) :1–14.

[Gautier, 2015] Gautier, M. (2015). Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. *Genetics*, 201(4) :1555–1579.

[Gautier et al., 2007] Gautier, M., Faraut, T., Moazami-Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., Boland, A., Garnier, J.-G., Boichard, D., Lathrop, G. M., Gut, I. G., and Eggen, A. (2007). Genetic and Haplotypic Structure in 14 European and African Cattle Breeds. *Genetics*, 177(2) :1059–1070.

[Gautier et al., 2009] Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., Moazami-Goudarzi, K., and Foulley, J.-L. (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC genomics*, 10 :550.

[Gautier et al., 2017] Gautier, M., Klassmann, A., and Vitalis, R. (2017). rehh 2.0 : a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular Ecology Resources*, 17(1) :78–90.

[Gautier et al., 2010] Gautier, M., Laloë, D., and Moazami-Goudarzi, K. (2010). Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PloS One*, 5(9).

[Gautier et al., 2016] Gautier, M., Moazami-Goudarzi, K., Levéziel, H., Parinello, H., Grohs, C., Rialle, S., Kowalczyk, R., and Flori, L. (2016). Deciphering the Wisent Demographic and Adaptive Histories from Individual Whole-Genome Sequences. *Molecular Biology and Evolution*, 33(11) :2801–2814.

[Gautier and Naves, 2011] Gautier, M. and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, 20(15) :3128–3143.

[Gautier et al., 2021] Gautier, M., Vitalis, R., Flori, L., and Estoup, A. (2021). f-statistics estimation and admixture graph construction with Pool-Seq or allele count data using the R package poolfstat. *bioRxiv*, page 2021.05.28.445945.

[Gibson et al., 2006] Gibson, J., Morton, N. E., and Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics*, 15(5) :789–795.

[Gifford-Gonzalez and Hanotte, 2011] Gifford-Gonzalez, D. and Hanotte, O. (2011). Domesticating Animals in Africa : Implications of Genetic and Archaeological Findings. *Journal of World Prehistory*, 24(1) :1–23.

[Gilbert, 2000] Gilbert, S. F. (2000). *Developmental Biology. Sunderland, MA*. Sinauer Associates, Inc.

[Goszczynski et al., 2018] Goszczynski, D. E., Corbi-Botto, C. M., Durand, H. M., Rogberg-Muñoz, A., Munilla, S., Peral-Garcia, P., Cantet, R. J. C., and Giovambattista, G. (2018). Evidence of positive selection towards Zebuine haplotypes in the BoLA region of Brangus cattle. *animal*, 12(02) :215–223.

[Grant, 1972] Grant, B. (1972). *Indonesia*. Penguin Books, Melbourne.

[Gravel, 2012] Gravel, S. (2012). Population Genetics Models of Local Ancestry. *Genetics*, 191(2) :607–619.

[Grousset, 1941] Grousset, R. (1941). The History of the Mongol Empire. *Shanghai : The Commercial Press.[Google Scholar]*.

[Günther and Coop, 2013] Günther, T. and Coop, G. (2013). Robust Identification of Local Adaptation from Allele Frequencies. *Genetics*, 195(1) :205–220.

[Haasl and Payseur, 2016] Haasl, R. J. and Payseur, B. A. (2016). Fifteen years of genomewide scans for selection : trends, lessons and unaddressed genetic sources of complication. *Molecular ecology*, 25(1) :5–23.

[Hall, 1999] Hall, B. K. (1999). *The neural crest in development and evolution*. Springer Science & Business Media.

[Hanotte, 2002] Hanotte, O. (2002). African Pastoralism : Genetic Imprints of Origins and Migrations. *Science*, 296(5566) :336–339.

[Hansen, 2004] Hansen, P. J. (2004). Physiological and cellular adaptations of zebu cattle to thermal stress. *Animal Reproduction Science*, 82(Supplement C) :349–360.

[Higham, 1996] Higham, C. (1996). *The Bronze Age of Southeast Asia*. Cambridge University Press. Google-Books-ID : syRlCA6a0AsC.

[Hivert et al., 2018] Hivert, V., Leblois, R., Petit, J., Gautier, M., and Vitalis, R. (2018). Measuring genetic differentiation from Pool-seq data. *Genetics*, 210(1) :315–330.

[Hoban et al., 2016] Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., Reed, L. K., Storfer, A., and Whitlock, M. C. (2016). Finding the Genomic Basis of Local Adaptation : Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist*, 188(4) :379–397.

[Holderegger et al., 2010] Holderegger, R., Buehler, D., Gugerli, F., and Manel, S. (2010). Landscape genetics of plants. *Trends in Plant Science*, 15(12) :675–683.

[Ito and Wakamatsu, 2011] Ito, S. and Wakamatsu, K. (2011). Diversity of human hair pigmentation as studied by chemical analysis of eumelanin and pheomelanin. *Journal of the European Academy of Dermatology and Venereology*, 25(12) :1369–1380.

[Jeffreys, 1998] Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.

[Jensen, 2014] Jensen, P. (2014). Behavior genetics and the domestication of animals. *Annu. Rev. Anim. Biosci.*, 2(1) :85–104.

[Jeong et al., 2016] Jeong, C., Ozga, A. T., Witonsky, D. B., Malmström, H., Edlund, H., Hofman, C. A., Hagan, R. W., Jakobsson, M., Lewis, C. M., Aldenderfer, M. S., and others (2016). Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proceedings of the National Academy of Sciences*, 113(27) :7485–7490.

[Jiang et al., 2014] Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J. F., Faraut, T., Wu, C., Muzny, D. M., Li, Y., Zhang, W., Stanton, J.-A., Brauning, R., Barris, W. C., Hourlier, T., Aken, B. L., Searle, S. M. J., Adelson, D. L., Bian, C., Cam, G. R., Chen, Y., Cheng, S., DeSilva, U., Dixen, K., Dong, Y., Fan, G., Franklin, I. R., Fu, S., Fuentes-Utrilla, P., Guan, R., Highland, M. A., Holder, M. E., Huang, G., Ingham, A. B., Jhangiani, S. N., Kalra, D., Kovar, C. L., Lee, S. L., Liu, W., Liu, X., Lu, C., Lv, T., Mathew, T., McWilliam, S., Menzies, M., Pan, S., Robelin, D., Servin, B., Townley, D., Wang, W., Wei, B., White, S. N., Yang, X., Ye, C., Yue, Y., Zeng, P., Zhou, Q., Hansen, J. B., Kristiansen, K., Gibbs, R. A., Flicek, P., Warkup, C. C., Jones, H. E., Oddy, V. H., Nicholas, F. W., McEwan, J. C., Kijas, J. W., Wang, J., Worley, K. C., Archibald, A. L., Cockett, N., Xu, X., Wang, W., and Dalrymple, B. P. (2014). The sheep genome illuminates biology of the rumen and lipid metabolism. *Science*, 344(6188) :1168–1173.

[Jivanji et al., 2019] Jivanji, S., Worth, G., Lopdell, T. J., Yeates, A., Couldrey, C., Reynolds, E., Tiplady, K., McNaughton, L., Johnson, T. J. J., Davis, S. R., Harris, B., Spelman, R., Snell, R. G., Garrick, D., and Littlejohn, M. D. (2019). Genome-wide association analysis reveals QTL and candidate mutations involved in white spotting in cattle. *Genetics Selection Evolution*, 51(1) :62.

[Kijas et al., 2012] Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., and Dalrymple, B. (2012). Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biology*, 10(2).

[Kijas et al., 2009] Kijas, J. W., Townley, D., Dalrymple, B. P., Heaton, M. P., Maddox, J. F., McGrath, A., Wilson, P., Ingersoll, R. G., McCulloch, R., McWilliam, S., Tang, D., McEwan, J., Cockett, N., Oddy, V. H., Nicholas, F. W., Raadsma, H., and Consortium, f. t. I. S. G. (2009). A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. *PLOS ONE*, 4(3) :e4668.

[Kim et al., 2020] Kim, K., Kwon, T., Dessie, T., Yoo, D., Mwai, O. A., Jang, J., Sung, S., Lee, S., Salim, B., Jung, J., Jeong, H., Tarekegn, G. M., Tijjani, A., Lim, D., Cho, S., Oh, S. J., Lee, H.-K., Kim, J., Jeong, C., Kemp, S., Hanotte, O., and Kim, H. (2020). The mosaic genome of indigenous African cattle as a unique genetic resource for African pastoralism. *Nature Genetics*, 52(10) :1099–1110.

[Kim et al., 2017] Kim, S.-J., Ka, S., Ha, J.-W., Kim, J., Yoo, D., Kim, K., Lee, H.-K., Lim, D., Cho, S., Hanotte, O., Mwai, O. A., Dessie, T., Kemp, S., Oh, S. J., and Kim, H. (2017). Cattle genome-wide analysis reveals genetic signatures in trypanotolerant N'Dama. *BMC Genomics*, 18(1) :371.

[Kirin et al., 2010] Kirin, M., McQuillan, R., Franklin, C. S., Campbell, H., McKeigue, P. M., and Wilson, J. F. (2010). Genomic Runs of Homozygosity Record Population History and Consanguinity. *PLOS ONE*, 5(11) :e13996.

[Koppen, 1936] Koppen, W. (1936). Das geographische system der klimat. *Handbuch der klimatologie*, page 46.

[Koren et al., 2018] Koren, S., Rhie, A., Walenz, B. P., Dilthey, A. T., Bickhart, D. M., Kingan, S. B., Hiendleder, S., Williams, J. L., Smith, T. P. L., and Phillippy, A. M. (2018). De novo assembly of haplotype-resolved genomes with trio binning. *Nature biotechnology*.

[Kriticos et al., 2012] Kriticos, D. J., Webber, B. L., Leriche, A., Ota, N., Macadam, I., Bathols, J., and Scott, J. K. (2012). CliMond : global high-resolution historical and future scenario climate surfaces for bioclimatic modelling : CliMond : climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution*, 3(1) :53–64.

[Larson and Fuller, 2014] Larson, G. and Fuller, D. Q. (2014). The Evolution of Animal Domestication. *Annual Review of Ecology, Evolution, and Systematics*, 45(1) :115–136.

[Lev-Yadun et al., 2000] Lev-Yadun, S., Gopher, A., and Abbo, S. (2000). The cradle of agriculture. *Science*, 288(5471) :1602–1603.

[Lewis et al., 2011] Lewis, J., Abas, Z., Dadousis, C., Lykidis, D., Paschou, P., and Drineas, P. (2011). Tracing Cattle Breeds with Principal Components Analysis Ancestry Informative SNPs. *PLOS ONE*, 6(4) :e18007.

[Li and Durbin, 2011] Li, H. and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357) :493–496.

[Li et al., 2020] Li, X., Yang, J., Shen, M., Xie, X.-L., Liu, G.-J., Xu, Y.-X., Lv, F.-H., Yang, H., Yang, Y.-L., Liu, C.-B., and others (2020). Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nature communications*, 11(1) :1–16.

[Lipson, 2020] Lipson, M. (2020). Applying f4-statistics and admixture graphs : Theory and examples. *Molecular Ecology Resources*, 20(6) :1658–1667.

[Littlejohn et al., 2014] Littlejohn, M. D., Henty, K. M., Tiplady, K., Johnson, T., Harland, C., Lopdell, T., Sherlock, R. G., Li, W., Lukefahr, S. D., Shanks, B. C., Garrick, D. J., Snell, R. G., Spelman, R. J., and Davis, S. R. (2014). Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. *Nature Communications*, 5(1) :1–8.

[Liu et al., 2012] Liu, T., Milia, E., Warburton, R. R., Hill, N. S., Gaestel, M., and Kayyali, U. S. (2012). Anthrax lethal toxin disrupts the endothelial permeability barrier through blocking p38 signaling. *Journal of cellular physiology*, 227(4) :1438–1445.

[Loftus et al., 1994] Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M., and Cunningham, P. (1994). Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences*, 91(7) :2757–2761.

[Loh et al., 2013] Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J. K., Reich, D., and Berger, B. (2013). Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium. *Genetics*, 193(4) :1233–1254.

[Lord et al., 2020] Lord, K. A., Larson, G., Coppinger, R. P., and Karlsson, E. K. (2020). The History of Farm Foxes Undermines the Animal Domestication Syndrome. *Trends in Ecology & Evolution*, 35(2) :125–136.

[Lu et al., 2020] Lu, Z., Liu, J., Han, J., and Yang, B. (2020). Association Between BMP2 Functional Polymorphisms and Sheep Tail Type. *Animals*, 10(4) :739.

[Lv et al., 2014] Lv, F.-H., Agha, S., Kantanen, J., Colli, L., Stucki, S., Kijas, J. W., Joost, S., Li, M.-H., and Ajmone Marsan, P. (2014). Adaptations to Climate-Mediated Selective Pressures in Sheep. *Molecular Biology and Evolution*, 31(12) :3324–3343.

[Lv et al., 2015] Lv, F.-H., Peng, W.-F., Yang, J., Zhao, Y.-X., Li, W.-R., Liu, M.-J., Ma, Y.-H., Zhao, Q.-J., Yang, G.-L., Wang, F., Li, J.-Q., Liu, Y.-G., Shen, Z.-Q., Zhao, S.-G., Hehua, E., Gorkhali, N. A., Farhad Vahidi, S. M., Muladno, M., Naqvi, A. N., Tabell, J., Iso-Touru, T., Bruford, M. W., Kantanen, J., Han, J.-L., and Li, M.-H. (2015). Mitogenomic Meta-Analysis Identifies Two Phases of Migration in the History of Eastern Eurasian Sheep. *Molecular Biology and Evolution*, 32(10) :2515–2533.

[Ma et al., 2018] Ma, J., Yu, M. K., Fong, S., Ono, K., Sage, E., Demchak, B., Sharan, R., and Ideker, T. (2018). Using deep learning to model the hierarchical structure and function of a cell. *Nature Methods*.

[Ma et al., 2015a] Ma, L., O'Connell, J. R., VanRaden, P. M., Shen, B., Padhi, A., Sun, C., Bickhart, D. M., Cole, J. B., Null, D. J., Liu, G. E., Da, Y., and Wiggans, G. R. (2015a). Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis. *PLOS Genetics*, page 24.

[Ma et al., 2015b] Ma, Y., Ding, X., Qanbari, S., Weigend, S., Zhang, Q., and Simianer, H. (2015b). Properties of different selection signature statistics and a new strategy for combining them. *Heredity*, 115(5) :426–436.

[MacHugh et al., 1997] MacHugh, D. E., Shriver, M. D., Loftus, R. T., Cunningham, P., and Bradley, D. G. (1997). Microsatellite DNA Variation and the Evolution, Domestication and Phylogeography of Taurine and Zebu Cattle (Bos taurus and Bos indicus). *Genetics*, 146(3) :1071–1086.

[MacLEOD et al., 2009] MacLEOD, I. M., Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2009). A novel predictor of multilocus haplotype homozygosity : comparison with existing predictors. *Genetics Research*, 91(6) :413–426.

[Magnavita, 2006] Magnavita, C. (2006). Ancient Humped Cattle in Africa : A View from the Chad Basin. *African Archaeological Review*, 23(3) :55–84.

[Makina et al., 2015] Makina, S. O., Muchadeyi, F. C., van Marle-Köster, E., Taylor, J. F., Makgahlela, M. L., and Maiwashe, A. (2015). Genome-wide scan for selection signatures in six cattle breeds in South Africa. *Genetics Selection Evolution*, 47(1) :92.

[Mandret, 2000] Mandret, G. (2000). *L'élevage bovin à la Réunion : Synthèse de quinze ans de recherche*. OCLC : 966066698.

[Mapholi et al., 2016] Mapholi, N. O., Maiwashe, A., Matika, O., Riggio, V., Bishop, S. C., MacNeil, M. D., Banga, C., Taylor, J. F., and Dzama, K. (2016). Genome-wide association study of tick resistance in South African Nguni cattle. *Ticks and Tick-borne Diseases*, 7(3) :487–497.

[Mariasegaram et al., 2007] Mariasegaram, M., Chase Jr, C., Chaparro, J., Olson, T., Brenneman, R., and Niedz, R. (2007). The slick hair coat locus maps to chromosome 20 in Senepol-derived cattle. *Animal Genetics*, 38(1) :54–59.

[Mason and others, 1980] Mason, I. L. and others (1980). *Prolific tropical sheep.*

[Mastrangelo et al., 2019a] Mastrangelo, S., Bahbahani, H., Moioli, B., Ahbara, A., Abri, M. A., Almathen, F., Silva, A. d., Belabdi, I., Portolano, B., Mwacharo, J. M., Hanotte, O., Pilla, F., and Ciani, E. (2019a). Novel and known signals of selection for fat deposition in domestic sheep breeds from Africa and Eurasia. *PLOS ONE*, 14(6) :e0209632.

[Mastrangelo et al., 2019b] Mastrangelo, S., Moioli, B., Ahbara, A., Latairish, S., Portolano, B., Pilla, F., and Ciani, E. (2019b). Genome-wide scan of fat-tail sheep identifies signals of selection for fat deposition and adaptation. *Animal Production Science*, 59(5) :835.

[Mastrangelo et al., 2020] Mastrangelo, S., Tolone, M., Ben Jemaa, S., Sottile, G., Di Gerlando, R., Cortés, O., Senczuk, G., Portolano, B., Pilla, F., and Ciani, E. (2020). Refining the genetic structure and relationships of European cattle breeds through meta-analysis of worldwide genomic SNP data, focusing on Italian cattle. *Scientific Reports*, 10(1) :14522.

[Mattioli et al., 2000] Mattioli, R. C., Pandey, V. S., Murray, M., and Fitzpatrick, J. L. (2000). Immunogenetic influences on tick resistance in African cattle with particular reference to trypanotolerant N'Dama (Bos taurus) and trypanosusceptible Gobra zebu (Bos indicus) cattle. *Acta Tropica*, 75(3) :263–277.

[Matukumalli et al., 2009] Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., O'Connell, J., Moore, S. S., Smith, T. P. L., Sonstegard, T. S., and Van Tassell, C. P. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PloS One*, 4(4) :e5350.

[Mbole-Kariuki et al., 2014] Mbole-Kariuki, M. N., Sonstegard, T., Orth, A., Thumbi, S. M., Bronsvoort, B. M. d. C., Kiara, H., Toye, P., Conradie, I., Jennings, A., Coetzer, K., Woolhouse, M. E. J., Hanotte, O., and Tapio, M. (2014). Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya. *Heredity*, 113(4) :297–305.

[McHugo et al., 2019] McHugo, G. P., Dover, M. J., and MacHugh, D. E. (2019). Unlocking the origins and biology of domestic animals using ancient DNA and paleogenomics. *BMC Biology*, 17(1) :98.

[McManus et al., 2011] McManus, C., Louvandini, H., Paim, T., Martins, R. S., Barcellos, J. O. J., Cardoso, C., Guimarães, R. F., and Santana, O. A. (2011). The challenge of sheep farming in the tropics : aspects related to heat tolerance. *Revista Brasileira de Zootecnia*, 40(Suppl. Esp) :107–120.

[McManus et al., 2009] McManus, C., Paludo, G. R., Louvandini, H., Gugel, R., Sasaki, L. C. B., and Paiva, S. R. (2009). Heat tolerance in Brazilian sheep : Physiological and blood parameters. *Tropical Animal Health and Production*, 41(1) :95–101.

[McQuillan et al., 2008] McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A. K., Farrington, S. M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S. H., Dunlop, M. G., Wright, A. F., Campbell, H., and Wilson, J. F. (2008). Runs of Homozygosity in European Populations. *The American Journal of Human Genetics*, 83(3) :359–372.

[McTavish and Hillis, 2014] McTavish, E. J. and Hillis, D. M. (2014). A Genomic Approach for Distinguishing between Recent and Ancient Admixture as Applied to Cattle. *JOURNAL OF HEREDITY*, 105(4) :445–456.

[Meadows et al., 2011] Meadows, J., Hiendleder, S., and Kijas, J. (2011). Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. *Heredity*, 106(4) :700–706.

[Medina et al., 2018] Medina, P., Thornlow, B., Nielsen, R., and Corbett-Detig, R. (2018). Estimating the Timing of Multiple Admixture Pulses During Local Ancestry Inference. *Genetics*, 210(3) :1089–1107.

[Meyer, 2020] Meyer, C. (2020). Dictionnaire des Sciences Animales.

[Meyer et al., 2004] Meyer, C., Faye, B., and Karembe, H. (2004). *Guide de l'élevage du mouton méditerranéen et tropical*. CEVA Santé animale.

[Mirkena et al., 2010] Mirkena, T., Duguma, G., Haile, A., Tibbo, M., Okeyo, A. M., Wurzinger, M., and Sölkner, J. (2010). Genetics of adaptation in domestic farm animals : A review. *Livestock Science*, 132(1) :1–12.

[Mohamad et al., 2009] Mohamad, K., Olsson, M., Tol, H. T. A. v., Mikko, S., Vlamings, B. H., Andersson, G., Rodríguez-Martínez, H., Purwantara, B., Paling, R. W., Colenbrander, B., and Lenstra, J. A. (2009). On the Origin of Indonesian Cattle. *PLOS ONE*, 4(5) :e5490.

[Moorjani et al., 2011] Moorjani, P., Patterson, N., Hirschhorn, J. N., Keinan, A., Hao, L., Atzmon, G., Burns, E., Ostrer, H., Price, A. L., and Reich, D. (2011). The History of African Gene Flow into Southern Europeans, Levantines, and Jews. *PLOS Genetics*, 7(4) :e1001373.

[Moradi et al., 2012] Moradi, M. H., Nejati-Javaremi, A., Moradi-Shahrbabak, M., Dodds, K. G., and McEwan, J. C. (2012). Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC Genetics*, 13(10).

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

[Mukai et al., 1989] Mukai, F., Tsuji, S., Fukazawa, K., Ohtagaki, S., and Nambu, Y. (1989). History and population structure of a closed strain of Japanese Black cattle. *Journal of Animal Breeding and Genetics*, 106(1-6) :254–264.

[Mukasa-Mugerwa, 1989] Mukasa-Mugerwa, E. (1989). *A review of reproductive performance of female Bos indicus (zebu) cattle*. International Livestock Centre for Africa, Addis Ababa, Ethiopia. Open Library ID : OL1848366M.

[Mwacharo et al., 2017] Mwacharo, J. M., Kim, E.-S., Elbeltagy, A. R., Aboul-Naga, A. M., Rischkowsky, B. A., and Rothschild, M. F. (2017). Genomic footprints of dryland stress adaptation in Egyptian fat-tail sheep and their divergence from East African and western Asia cohorts. *Scientific Reports*, 7(1) :17647.

[Mwai et al., 2015] Mwai, O., Hanotte, O., Kwon, Y.-J., and Cho, S. (2015). African Indigenous Cattle : Unique Genetic Resources in a Rapidly Changing World. *Asian-Australasian Journal of Animal Sciences*, 28(7) :911–921.

[Naskar et al., 2012] Naskar, S., Gowane, G. R., Chopra, A., Paswan, C., and Prince, L. L. L. (2012). Genetic Adaptability of Livestock to Environmental Stresses. In Sejian, V., Naqvi, S., Ezeji, T., Lakritz, J., and Lal, R., editors, *Environmental Stress and Amelioration in Livestock Production*, pages 317–378. Springer, Berlin, Heidelberg.

[Nicolazzi et al., 2015] Nicolazzi, E. L., Caprera, A., Nazzicari, N., Cozzi, P., Strozzi, F., Lawley, C., Pirani, A., Soans, C., Brew, F., Jorjani, H., Evans, G., Simpson, B., Tosser-Klopp, G., Brauning, R., Williams, J. L., and Stella, A. (2015). SNPchiMp v.3 : integrating and standardizing single nucleotide polymorphism data for livestock species. *BMC Genomics*, 16(1) :283.

[Nielsen et al., 2007] Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., and Clark, A. G. (2007). Recent and ongoing selection in the human genome. *Nature Reviews Genetics*, 8(11) :857–868.

[Novembre and Stephens, 2008] Novembre, J. and Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*, 40(5) :646–649.

[Oleksyk et al., 2010] Oleksyk, T. K., Smith, M. W., and O'Brien, S. J. (2010). Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 365(1537) :185–205.

[Palamara et al., 2012] Palamara, P. F., Lencz, T., Darvasi, A., and Pe'er, I. (2012). Length Distributions of Identity by Descent Reveal Fine-Scale Demographic History. *The American Journal of Human Genetics*, 91(5) :809–822.

[Pan et al., 2019] Pan, Z., Li, S., Liu, Q., Wang, Z., Zhou, Z., Di, R., An, X., Miao, B., Wang, X., Hu, W., Guo, X., Lv, S., Li, F., Ding, G., Chu, M., and Li, Y. (2019). Rapid evolution of a retro-transposable hotspot of ovine genome underlies the alteration of BMP2 expression and development of fat tails. *BMC Genomics*, 20(1) :261.

[Pardeshi et al., 2007] Pardeshi, V., Kadoo, N., Sainani, M., Meadows, J., Kijas, J., and Gupta, V. (2007). Mitochondrial haplotypes reveal a strong genetic structure for three Indian sheep breeds. *Animal genetics*, 38(5) :460–466.

[Patterson et al., 2006] Patterson, N., Price, A. L., and Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genetics*, 2(12).

[Patterson et al., 2012] Patterson, N. J., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient Admixture in Human History. *Genetics*, page genetics.112.145037.

[Paula-Lopes et al., 2013] Paula-Lopes, F. F., Lima, R. S., Satrapa, R. A., and Barros, C. M. (2013). PHYSIOLOGY AND ENDOCRINOLOGY SYMPOSIUM : Influence of cattle genotype (Bos indicus vs. Bos taurus) on oocyte and preimplantation embryo resistance to increased temperature,. *Journal of Animal Science*, 91(3) :1143–1153.

[Pauly, 2014] Pauly, M. (2014). La diffusion de l'islam à Mayotte à l' époque médiévale. *Taarifa (revue des archives départementales de Mayotte)*, (4) :69–113.

[Pausch et al., 2012] Pausch, H., Wang, X., Jung, S., Krogmeier, D., Edel, C., Emmerling, R., Götz, K.-U., and Fries, R. (2012). Identification of QTL for UV-Protective Eye Area Pigmentation in Cattle by Progeny Phenotyping and Genome-Wide Association Analysis. *PLOS ONE*, 7(5) :e36346.

[Payne and Hodges, 1997] Payne, W. J. A. and Hodges, J. (1997). *Tropical cattle : origins, breeds, and breeding policies*. Blackwell Science, Oxford ; Malden, MA, USA.

[Payne and Wilson, 1999] Payne, W. J. A. and Wilson, R. T. (1999). An introduction to animal husbandry in the tropics. *An introduction to animal husbandry in the tropics.*, (Edn 5).

[Peel et al., 2007] Peel, M. C., Finlayson, B. L., and McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, 11(5) :1633–1644.

[Pellecchia et al., 2007] Pellecchia, M., Negrini, R., Colli, L., Patrini, M., Milanesi, E., Achilli, A., Bertorelle, G., Cavalli-Sforza, L. L., Piazza, A., Torroni, A., and Ajmone-Marsan, P. (2007). The mystery of Etruscan origins : novel clues from Bos taurus mitochondrial DNA. *Proceedings of the Royal Society B : Biological Sciences*, 274(1614) :1175–1179.

[Pemberton et al., 2012] Pemberton, T. J., Absher, D., Feldman, M. W., Myers, R. M., Rosenberg, N. A., and Li, J. Z. (2012). Genomic Patterns of Homozygosity in Worldwide Human Populations. *The American Journal of Human Genetics*, 91(2) :275–292.

[Pereira et al., 2006] Pereira, F., Davis, S. J., Pereira, L., McEvoy, B., Bradley, D. G., and Amorim, A. (2006). Genetic signatures of a Mediterranean influence in Iberian Peninsula sheep husbandry. *Molecular Biology and Evolution*, 23(7) :1420–1426.

[Peripolli et al., 2017] Peripolli, E., Munari, D. P., Silva, M. V. G. B., Lima, A. L. F., Irgang, R., and Baldi, F. (2017). Runs of homozygosity : current knowledge and applications in livestock. *Animal Genetics*, 48(3) :255–271.

[Peter, 2016] Peter, B. M. (2016). Admixture, Population Structure, and F-Statistics. *Genetics*, 202(4) :1485–1501.

[Pickrell et al., 2014] Pickrell, J. K., Patterson, N., Loh, P.-R., Lipson, M., Berger, B., Stoneking, M., Pakendorf, B., and Reich, D. (2014). Ancient west Eurasian ancestry in southern and eastern Africa. *Proceedings of the National Academy of Sciences*, 111(7) :2632–2637.

[Pickrell and Pritchard, 2012] Pickrell, J. K. and Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLOS Genetics*, 8(11) :e1002967.

[Pitt et al., 2019] Pitt, D., Sevane, N., Nicolazzi, E. L., MacHugh, D. E., Park, S. D. E., Colli, L., Martinez, R., Bruford, M. W., and Orozco-terWengel, P. (2019). Domestication of cattle : Two or three events ? *Evolutionary Applications*, 12(1) :123–136.

[Pool and Nielsen, 2009] Pool, J. E. and Nielsen, R. (2009). Inference of Historical Changes in Migration Rate From the Lengths of Migrant Tracts. *Genetics*, 181(2) :711–719.

[Porter et al., 2016] Porter, V., Alderson, L., Hall, S. J. G., and Sponenberg, D. P. (2016). *Mason's World Encyclopedia of Livestock Breeds and Breeding, 2 Volume Pack*. CABI. Google-Books-ID : 2UEJDAAAQBAJ.

[Porter and Tebbit, 2007] Porter, V. and Tebbit, J. (2007). *Cattle : a handbook to the breeds of the world*. Crowood, Ramsbury, Marlborough, Wiltshire. OCLC : ocm84151515.

[Porto-Neto et al., 2018] Porto-Neto, L. R., Bickhart, D. M., Landaeta-Hernandez, A. J., Utsunomiya, Y. T., Pagan, M., Jimenez, E., Hansen, P. J., Dikmen, S., Schroeder, S. G., Kim, E.-S., Sun, J., Crespo, E., Amati, N., Cole, J. B., Null, D. J., Garcia, J. F., Reverter, A., Barendse, W., and Sonstegard, T. S. (2018). Convergent Evolution of Slick Coat in Cattle through Truncation Mutations in the Prolactin Receptor. *Frontiers in Genetics*, 9.

[Price et al., 2009] Price, A. L., Tandon, A., Patterson, N., Barnes, K. C., Rafaels, N., Ruczinski, I., Beaty, T. H., Mathias, R., Reich, D., and Myers, S. (2009). Sensitive Detection of Chromosomal Segments of Distinct Ancestry in Admixed Populations. *PLOS Genetics*, 5(6) :e1000519.

[Purcell et al., 2007] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., and Sham, P. C. (2007). PLINK : A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81(3) :559–575.

[Qanbari et al., 2014] Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T. M., Fries, R., Nielsen, R., and Simianer, H. (2014). Classic Selective Sweeps Revealed by Massive Sequencing in Cattle. *PLOS Genetics*, 10(2) :e1004148.

[Rafolo, 1989] Rafolo, A. (1989). Les fouilles de Lohavohitra. Contribution à l'étude des aménagements d'un site ancien fortifié et perché du Vonizongo. *Omaly sy Anio : revue d'etudes historiques*, (29-32) :57–70.

[Rahman et al., 2018] Rahman, M. B., Schellander, K., Luceño, N. L., and Van Soom, A. (2018). Heat stress responses in spermatozoa : Mechanisms and consequences for cattle fertility. *Theriogenology*, 113 :102–112.

[Rannamäe et al., 2016] Rannamäe, E., Lõugas, L., Niemi, M., Kantanen, J., Maldre, L., Kadõrova, N., and Saarma, U. (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*, 47(2) :208–218.

[Rasamuel, 1988] Rasamuel, D. (1988). Domenichini-Ramiaramanana (Bakoly) : Du ohabolana au hainteny. Langue, littérature et politique à Madagascar. *Outre-Mers. Revue d'histoire*, 75(281) :500–501.

[Reece et al., 2012] Reece, J. B., Faucher, J., and Lachaîne, R. (2012). *Campbell biologie*. OCLC : 947927225.

[Rege and Tawah, 1999] Rege, J. E. O. and Tawah, C. L. (1999). The state of African cattle genetic resources II. Geographical distribution, characteristics and uses of present-day breeds and strains. *Animal Genetic Resources/Resources génétiques animales/Recursos genéticos animales*, 26 :1–25.

[Reich et al., 2010] Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L., and others (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, 468(7327) :1053–1060.

[Reich et al., 2009] Reich, D., Thangaraj, K., Patterson, N., Price, A. L., and Singh, L. (2009). Reconstructing Indian population history. *Nature*, 461(7263) :489–494.

[Rellstab et al., 2015] Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., and Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, 24(17) :4348–4370.

[Rewe, 2013] Rewe, T. (2013). *Breeding For Profitable Beef Production In Sub-sahara Africa*. OCLC : 859346925.

[Ribot, 2005] Ribot, J. (2005). 1905-2005 : Cent ans de présence vétérinaire à Madagascar. *Médecine tropicale*, 65(3) :245.

[Rivière, 2021] Rivière, Q. (2021). D'un cheptel conservatoire de races locales à un outil de gestion de la savane : le cas de la chèvre Péï et de la vache Moka à La Réunion. *Essais. Revue interdisciplinaire d'Humanités*, (Hors-série 6) :125–146.

[Rochus et al., 2018] Rochus, C. M., Tortereau, F., Plisson-Petit, F., Restoux, G., Moreno-Romieux, C., Tosser-Klopp, G., and Servin, B. (2018). Revealing the selection history of adaptive loci using genome-wide scans for selection : an example from domestic sheep. *BMC Genomics*, 19(1) :71.

[Rokosz, 1995] Rokosz, M. (1995). History of the Aurochs (Bos taurus primigenius) in Poland. *Animal Genetic Resources Information*, 16 :5–12.

[Rosen et al., 2020] Rosen, B. D., Bickhart, D. M., Schnabel, R. D., Koren, S., Elsik, C. G., Tseng, E., Rowan, T. N., Low, W. Y., Zimin, A., Couldrey, C., Hall, R., Li, W., Rhie, A., Ghurye, J., McKay, S. D., Thibaud-Nissen, F., Hoffman, J., Murdoch, B. M., Snelling, W. M., McDaneld, T. G., Hammond, J. A., Schwartz, J. C., Nandolo, W., Hagen, D. E., Dreischer, C., Schultheiss, S. J., Schroeder, S. G., Phillippy,

A. M., Cole, J. B., Van Tassell, C. P., Liu, G., Smith, T. P. L., and Medrano, J. F. (2020). De novo assembly of the cattle reference genome with single-molecule sequencing. *GigaScience*, 9(3).

[Rozier, 1800] Rozier, J.-F. R. M. (1800). *Cours complet d'agriculture, théorique, pratique, économique et de médecine rurale et vétérinaire ; suivi d'une méthode pour étudier l'agriculture par principes : ou, Dictionnaire universel d'agriculture*. Serpente. Google-Books-ID : P4xYAAAAcAAJ.

[Ryder, 1983] Ryder, M. (1983). *Sheep and Man*. Geral Duckworth & Co., London, UK.

[Ryder, 1984] Ryder, M. (1984). Sheep. In *Evolution of domesticated animals*, pages 63–85. I.L.Mason, London : Longman.

[Sabeti et al., 2002] Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., Ackerman, H. C., Campbell, S. J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., and Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419(6909) :832–837.

[Saido, 2005] Saido (2005). Rapport National sur l'Etat des Ressources Génétiques Animales - Comores. Technical report, FAO.

[Saint Martin, 1988] Saint Martin, G. (1988). *Système d'élevage et productivité des ruminants en République Fédérale Islamique des Comores*. PhD thesis, Université de Paris-Val-de-Marne, Maisons-Alfort : CIRAD-IEMVT.

[Sankararaman et al., 2012] Sankararaman, S., Patterson, N., Li, H., Pääbo, S., and Reich, D. (2012). The Date of Interbreeding between Neandertals and Modern Humans. *PLOS Genetics*, 8(10) :e1002947.

[Santiago et al., 2020] Santiago, E., Novo, I., Pardiñas, A. F., Saura, M., Wang, J., and Caballero, A. (2020). Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. *Molecular Biology and Evolution*, 37(12) :3642–3653.

[Satzger et al., 2008] Satzger, I., Schaefer, T., Kuettler, U., Broecker, V., Voelker, B., Ostertag, H., Kapp, A., and Gutzmer, R. (2008). Analysis of c-KIT expression and KIT gene mutation in human mucosal melanomas. *British Journal of Cancer*, 99(12) :2065–2069.

[Scheinfeldt and Tishkoff, 2013] Scheinfeldt, L. B. and Tishkoff, S. A. (2013). Recent human adaptation : genomic approaches, interpretation and insights. *Nature Reviews Genetics*, 14(10) :692–702.

[Scherf et al., 2015] Scherf, B. D., Pilling, D., and others (2015). The second report on the state of the world's animal genetic resources for Food and Agriculture.

[Schiffels and Durbin, 2014] Schiffels, S. and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, 46(8) :919–925.

[Sealy and Yates, 1994] Sealy, J. and Yates, R. (1994). The chronology of the introduction of pastoralism to the Cape, South Africa. *Antiquity*, 68(258) :58–67.

[Sejian et al., 2017] Sejian, V., Hyder, I., Maurya, V., Bagath, M., Krishnan, G., Aleena, J., Archana, P., Lees, A. M., Kumar, D., Bhatta, R., and others (2017). Adaptive mechanisms of sheep to climate change. In *Sheep Production Adapting to Climate Change*, pages 117–147. Springer.

[Sempéré et al., 2015] Sempéré, G., Moazami-Goudarzi, K., Eggen, A., Laloë, D., Gautier, M., and Flori, L. (2015). WIDDE : a Web-Interfaced next generation database for genetic diversity exploration, with a first application in cattle. *BMC Genomics*, 16(1).

[Shen et al., 2018] Shen, B., Jiang, J., Seroussi, E., Liu, G. E., and Ma, L. (2018). Characterization of recombination features and the genetic basis in multiple cattle breeds. *BMC Genomics*, 19(1) :304.

[Sherratt, 1981] Sherratt, A. (1981). Plough and pastoralism : aspects of the secondary. *Pattern of the Past : Studies in the Honour of David Clarke*, page 261.

[Sileshi et al., 2003] Sileshi, Z., Tegegne, A., Tsadik, G. T., and others (2003). Water resources for livestock in Ethiopia : Implications for research and development. *Integrated water and land management research and capacity building priorities for Ethiopia*, 66.

[Singh et al., 2013] Singh, S., Jr, S. K., Kolte, A. P., and Kumar, S. (2013). Extensive Variation and Sub-Structuring in Lineage A mtDNA in Indian Sheep : Genetic Evidence for Domestication of Sheep in India. *PLOS ONE*, 8(11) :e77858.

[Siripan et al., 2019] Siripan, S., Wonnapinij, P., Auetrakulvit, P., Wangthongchaicharoen, N., and Surat, W. (2019). Origin of prehistoric cattle excavated from four archaeological sites in central and northeastern Thailand. *Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis*, pages 1–9.

[Smith, 2009] Smith, A. B. (2009). Pastoralism in the Western Cape Province, South Africa : A retrospective review. *Journal of African Archaeology*, 7(2) :239–252.

[Sodiq and Tawfik, 2004] Sodiq, A. and Tawfik, E. S. (2004). Productivity and breeding strategies of sheep in Indonesia : a review. *Journal of Agriculture and Rural Development in the Tropics and Subtropics (JARTS)*, 105(1) :71–82.

[Sollero et al., 2017] Sollero, B. P., Junqueira, V. S., Gomes, C. C. G., Caetano, A. R., and Cardoso, F. F. (2017). Tag SNP selection for prediction of tick resistance in Brazilian Braford and Hereford cattle breeds using Bayesian methods. *Genetics Selection Evolution*, 49(1) :49.

[Souvenir, 1991] Souvenir, Zafindrajaona, P. (1991). *Profils genetiques du zebu malgache*. These de doctorat, Paris 11.

[Spinage, 2003] Spinage, C. A. (2003). *Cattle Plague : A History*. Springer Science & Business Media. Google-Books-ID : t5QUSfS8FfYC.

[Sudrajad et al., 2020] Sudrajad, P., Subiharta, S., Adinata, Y., Lathifah, A., Lee, J. H., Lenstra, J. A., and Lee, S. H. (2020). An insight into the evolutionary history of Indonesian cattle assessed by whole genome data analysis. *PLOS ONE*, 15(11) :e0241038.

[Sutama, 1990] Sutama, I.-K. (1990). Production and reproductive performance of Javanese Fat Tail sheep. *PRODUCTION ASPECTS OF JAVANESE FAT TAIL*.

[Sutarno and Setyawan, 2016] Sutarno, S. and Setyawan, A. D. (2016). Review : The diversity of local cattle in Indonesia and the efforts to develop superior indigenous cattle breeds. *Biodiversitas Journal of Biological Diversity*, 17(1).

[Tajima, 1989] Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3) :585–595.

[Tang et al., 2007] Tang, K., Thornton, K. R., and Stoneking, M. (2007). A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *PLoS Biology*, 5(7) :e171.

[Tapio et al., 2006] Tapio, I., Värv, S., Bennewitz, J., Maleviciute, J., Fimland, E., Grislis, Z., Meuwissen, T. H. E., Miceikiene, I., Olsaker, I., Viinalass, H., Vilkki, J., and Kantanen, J. (2006). Prioritization for Conservation of Northern European Cattle Breeds Based on Analysis of Microsatellite Data. *Conservation Biology*, 20(6) :1768–1779.

[Taylor et al., 2021] Taylor, W. T. T., Pruvost, M., Posth, C., Rendu, W., Krajcarz, M. T., Abdykanova, A., Brancaleoni, G., Spengler, R., Hermes, T., Schiavinato, S., Hodgins, G., Stahl, R., Min, J., Alisher kyzy, S., Fedorowicz, S., Orlando, L., Douka, K., Krivoshapkin, A., Jeong, C., Warinner, C., and Shnaider, S. (2021). Evidence for early dispersal of domestic sheep into Central Asia. *Nature Human Behaviour*, pages 1–11.

[Tijjani et al., 2019] Tijjani, A., Utsunomiya, Y. T., Ezekwe, A. G., Nashiru, O., and Hanotte, O. (2019). Genome Sequence Analysis Reveals Selection Signatures in Endangered Trypanotolerant West African Muturu Cattle. *Frontiers in Genetics*, 10 :442.

[Trainor, 2013] Trainor, P. (2013). *Neural crest cells : evolution, development and disease*. Academic Press.

[Troy et al., 2001] Troy, C. S., MacHugh, D. E., Bailey, J. F., Magee, D. A., Loftus, R. T., Cunningham, P., Chamberlain, A. T., Sykes, B. C., and Bradley, D. G. (2001). Genetic evidence for Near-Eastern origins of European cattle. *Nature*, 410(6832) :1088–1091.

[Trut, 1999] Trut, L. N. (1999). Early Canid Domestication : The Farm-Fox Experiment : Foxes bred for tamability in a 40-year experiment exhibit remarkable transformations that suggest an interplay between behavioral genetics and development. *American Scientist*, 87(2) :160–169.

[Turner, 1980] Turner, J. W. (1980). Genetic and Biological Aspects of Zebu Adaptability. *Journal of Animal Science*, 50(6) :1201–1205.

[Udo and Budisatria, 2011] Udo, H. M. J. and Budisatria, I. G. S. (2011). Fat-tailed sheep in Indonesia ; an essential resource for smallholders. *Tropical Animal Health and Production*, 43(7) :1411–1418.

[Upadhyay et al., 2017] Upadhyay, M. R., Chen, W., Lenstra, J. A., Goderie, C. R. J., MacHugh, D. E., Park, S. D. E., Magee, D. A., Matassino, D., Ciani, F., Megens, H.-J., van Arendonk, J. A. M., Groenen, M. a. M., European Cattle Genetic Diversity Consortium, Marsan, P. A., Balteanu, V., Dunner, S., Garcia, J. F., Ginja, C., Kantanen, J., and Rpma Crooijmans (2017). Genetic origin, admixture and population history of aurochs (*Bosprimigenius*) and primitive European cattle. *Heredity*, 118(2) :169–176.

[Utsunomiya et al., 2019] Utsunomiya, Y. T., Milanesi, M., Fortes, M. R. S., Porto-Neto, L. R., Utsunomiya, A. T. H., Silva, M. V. G. B., Garcia, J. F., and Ajmone-Marsan, P. (2019). Genomic clues of the evolutionary history of Bos indicus cattle. *Animal Genetics*, 50(6) :557–568.

[Vadhanakul et al., 2004] Vadhanakul, N., Tunwattana, W., Sekkasiddhi, P., Tansatit, M., and Chavananikul, V. (2004). Karyotype of crossbred progeny born to the interspecies hybridization of wild and zebu cattle with reference to their fertility and growth performance. *Wetchasan Sattawaphaet*.

[Velimirovic, 1984] Velimirovic, B. (1984). Anthrax in Europe. *Revue Scientifique et Technique (International Office of Epizootics)*, 3(3) :527–559.

[Verschuren et al., 2000] Verschuren, D., Laird, K. R., and Cumming, B. F. (2000). Rainfall and drought in equatorial east Africa during the past 1,100 years. *Nature*, 403(6768) :410–414.

[Vitti et al., 2013] Vitti, J. J., Grossman, S. R., and Sabeti, P. C. (2013). Detecting Natural Selection in Genomic Data. *Annual Review of Genetics*, 47(1) :97–120.

[Voight et al., 2006] Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A Map of Recent Positive Selection in the Human Genome. *PLoS Biology*, 4(3) :e72.

[Wang et al., 2008] Wang, L., Sewell, W. F., Kim, S. D., Shin, J. T., MacRae, C. A., Zon, L. I., Seidman, J. G., and Seidman, C. E. (2008). Eya4 regulation of Na+/K+-ATPase is required for sensory system development in zebrafish. *Development*, 135(20) :3425–3434.

[Wang et al., 2021] Wang, X., Fang, C., He, H., Cao, H., Liu, L., Jiang, L., Ma, Y., and Liu, W. (2021). Identification of key genes in sheep fat tail evolution Based on RNA-seq. *Gene*, 781 :145492.

[Wangkumhang et al., 2015] Wangkumhang, P., Wilantho, A., Shaw, P. J., Flori, L., Moazami-Goudarzi, K., Gautier, M., Duangjinda, M., Assawamakin, A., and Tongsima, S. (2015). Genetic analysis of Thai cattle reveals a Southeast Asian indicine ancestry. *PeerJ*, 3 :e1318.

[Weir and Cockerham, 1984] Weir, B. S. and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, (19) :395–420.

[Werren et al., 2021] Werren, E. A., Garcia, O., and Bigham, A. W. (2021). Identifying adaptive alleles in the human genome : from selection mapping to functional validation. *Human Genetics*, 140(2) :241–276.

[Wiener et al., 2021] Wiener, P., Robert, C., Ahbara, A., Salavati, M., Abebe, A., Kebede, A., Wragg, D., Friedrich, J., Vasoya, D., Hume, D. A., Djikeng, A., Watson, M., Prendergast, J. G. D., Hanotte, O., Mwacharo, J. M., and Clark, E. L. (2021). Whole-Genome Sequence Data Suggest Environmental Adaptation of Ethiopian Sheep Populations. *Genome Biology and Evolution*, 13(3).

[Wilkins et al., 2014] 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*, 197(3) :795–808.

[Wormser, 2017] Wormser, P. (2017). Les relations maritimes entre l'Indonésie et l'Océan Indien au Moyen-Âge. In Buchet, C. and Balard, M., editors, *The Sea in History - The Medieval World*, pages 918–925. Boydell &amp ; Brewer.

[Wright et al., 1984] Wright, H. T., Sinopoli, C., Wojnaroski, L., Hoffman, E. S., Scott, S. L., Redding, R. W., and Goodman, S. M. (1984). Early Seafarers of the Comoro Islands : the Dembeni Phase of the IXth-Xth Centuries AD. *Azania : Archaeological Research in Africa*, 19(1) :13–59.

[Wright, 1922] Wright, S. (1922). Coefficients of Inbreeding and Relationship. *The American Naturalist*, 56(645) :330–338.

[Wright, 1949] Wright, S. (1949). THE GENETICAL STRUCTURE OF POPULATIONS. *Annals of Eugenics*, 15(1) :323–354.

[Wu et al., 2018] Wu, D.-D., Ding, X.-D., Wang, S., Wójcik, J. M., Zhang, Y., Tokarska, M., Li, Y., Wang, M.-S., Faruque, O., Nielsen, R., Zhang, Q., and Zhang, Y.-P. (2018). Pervasive introgression facilitated domestication and adaptation in the Bos species complex. *Nature Ecology & Evolution*, 2(7) :1139–1145.

[Yonesaka et al., 2016] Yonesaka, R., Sasazaki, S., Yasue, H., Niwata, S., Inayoshi, Y., Mukai, F., and Mannen, H. (2016). Genetic structure and relationships of 16 Asian and European cattle populations using DigiTag2 assay. *Animal Science Journal*, 87(2) :190–196.

[Young et al., 2011] Young, E., Kijas, J., McCulloch, R., Scobie, D., McRae, K., Pickering, N., Dodds, K., and McEwan, J. (2011). Arapawa : a novel New Zealand sheep breed of distinct origin. *Proceedings of the New Zealand Society of Animal Production*, 71 :248–250.

[Yuan et al., 2017] Yuan, Z., Liu, E., Liu, Z., Kijas, J. W., Zhu, C., Hu, S., Ma, X., Zhang, L., Du, L., Wang, H., and Wei, C. (2017). Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep. *Animal Genetics*, 48(1) :55–66.

[Zeder, 2011] Zeder, M. A. (2011). The origins of agriculture in the Near East. *Current Anthropology*, 52(S4) :S221–S235.

[Zeder, 2017] Zeder, M. A. (2017). Out of the Fertile Crescent : The dispersal of domestic livestock through Europe and Africa. In Boivin, N., Petraglia, M., and Crassard, R., editors, *Human Dispersal and Species Movement*, pages 261–303. Cambridge University Press, Cambridge.

[Zhang et al., 2020] Zhang, K., Lenstra, J. A., Zhang, S., Liu, W., and Liu, J. (2020). Evolution and domestication of the Bovini species. *Animal Genetics*, 51(5) :637–657.

[Zhao et al., 2017] Zhao, Y.-X., Yang, J., Lv, F.-H., Hu, X.-J., Xie, X.-L., Zhang, M., Li, W.-R., Liu, M.-J., Wang, Y.-T., Li, J.-Q., Liu, Y.-G., Ren, Y.-L., Wang, F., Hehua, E., Kantanen, J., Arjen Lenstra, J., Han, J.-L., and Li, M.-H. (2017). 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*, 34(9) :2380–2395.

[Zimin et al., 2009] Zimin, A. V., Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassell, C. P., Sonstegard, T. S., Marçais, G., Roberts, M., Subramanian, P., Yorke, J. A., and Salzberg, S. L. (2009). A whole-genome assembly of the domestic cow, Bos taurus. *Genome Biology*, 10(4) :R42.

[Zygoyiannis, 2006] Zygoyiannis, D. (2006). Sheep production in the world and in Greece. *Small ruminant research*, 62(1-2) :143–147.

Troisième partie

Résultats

Chapitre 4

Description phénotypique des bovins et ovins de Mayotte

L'objectif de ce chapitre est de présenter les résultats de la caractérisation phénotypique des zébus et des moutons locaux rencontrés sur l'île de Mayotte. Ces deux études de caractérisation phénotypiques faisaient partie d'un projet global, financé par les fonds FEADER (2015-2020), de caractérisation des ressources génétiques animales d'intérêt économique pour l'élevage mahorais (bovins, ovins et caprins locaux).

Les données morphobiométriques sont nécessaires dans le processus de reconnaissance des races bovine et ovine locales mahoraises afin d'établir un standard morphologique de la race à reconnaître afin de, par la suite, éliminer les individus ne correspondant pas à des critères morphologiques définis. A ce jour, le dossier de reconnaissance pour la population bovine a été présenté à la Commission Nationale d'Amélioration Génétique (CNAG) du ministère français de l'Agriculture en 2018. La race "zébu mahorais" a été reconnue comme race bovine française locale mais n'a pas été inscrite sur la liste des "races menacées d'être perdues pour l'agriculture", compte tenu des effectifs encore disponibles malgré des incertitudes quant à la véracité des informations rentrées dans la Base de Données Nationale d'Identification. Le processus de reconnaissance pour le mouton de Mayotte doit être engagé.

4.1 Synthèse des résultats

Pour les bovins, 400 animaux (2/3 de femelles, 1/3 de mâles) présentant un phénotype "zébu local" ont été échantillonnés sur les 14'027 animaux identifiés dans la Base de Données Nationale d'Identification (BDNI) au 10/03/2016 et sur les 17 communes de l'île de Mayotte. En résumé, les zébus mahorais ont présenté une hauteur au garrot moyenne de 105.7 cm (min = 90.0 cm, max = 126.0 cm), plus petit que le zébu de Madagascar. Ces zébus sont caractérisés par des motifs de coloration hétérogènes avec cependant une forte proportion pour la robe "unie noir" 32 %. Les cornes sont pour la plupart des individus en forme de coupe (70 %). En reliant la note d'Etat Corporel et la taille de la bosse (petit, moyenne, grande), le rôle de la bosse comme organe de réserve a été mis en relief. Par son phénotype, nous pouvons déduire de son appartenance au groupe des petits zébus de l'Est Africain (Small East African Zebus) reconnu par la FAO.

Pour les ovins, 100 animaux (2/3 de femelles, 1/3 de mâles) ont été échantillonnés sur les 1'000 animaux identifiés localement par la Chambre d'Agriculture de Mayotte. 27 élevages ont été visités sur 17 communes de l'île. La population de moutons mahorais présente une hauteur au garrot moyenne de 55 cm (min = 44.0 cm, max = 67.0 cm). Ces moutons sont caractérisés par une robe hétérogène de type "poil" avec cependant des motifs plus fréquents "pie-rouge clair" (30 %), "pie noir" (20 %) et "rouge-clair avec tache blanche isolée" (15 %). Les femelles sont sans cornes et seulement 1/4 des mâles sont cornus. En reliant l'état corporel avec la taille de la queue grasse, le rôle de la queue comme organe de réserve a été mis en évidence. De plus, 21 % des individus présentent une mutation auriculaire classifiée selon la forme "oreille courte" ou "vestigial". Cette population de moutons est très proche phénotypiquement des moutons présents à Madagascar qui appartiennent au groupe des moutons à queue grasse et

à poil (Fat tailed sheep) selon la classification proposée par la base de données DAGRIS.



FIGURE 4.1 – Zébus mahorais femelle **a** et mâle **b**.
Moutons mahorais femelle **c** et mâle **d**.

4.2 Article 1 : Phenotypic characterization of Zebu from Mayotte

(en préparation)

Jessica Magnier[1], Mélissa Ouvrard[2], Solène Raoul[2], Jérôme Janelle[3], Anlidine Mkadara[4], Hidachi Attoumani[4], Oussoufi Saindou[4], Kamardine Moussa[4], Michel Naves[5], Mathieu Gautier[6], Laurence Flori[1], Emmanuel Tillard[3]

[1] UMR SELMET, INRAE, CIRAD, Montpellier SupAgro, Univ. Montpellier, Montpellier, France

[2] Coopadem, Coconi, 97660 Ouangani, Mayotte

[3] UMR SELMET, INRAE, CIRAD, Montpellier SupAgro Univ., Saint Pierre, La Réunion, France

[4] CAPAM 976, Coconi, 97660 Ouangani, Mayotte

[5] URZ, INRAE, Petit-Bourg, France

[6] UMR CBGP, INRAE, CIRAD, IRD, Montpellier Supagro, Univ. Montpellier, Montferrier-sur-Lez, France

4.2.1 Abstract

In Mayotte island, three types of cattle are raised. We gave insight in local zebu cattle by sampling 400 individuals (1/3 males, 2/3 females) showing a "local zebu" phenotype from the 14,027 animals identified in the National Identification Database as of 10/03/2016 and from the 17 townships of the island of Mayotte. In summary, zebu of Mayotte have an average withers height of 105.7 cm (min = 90.0 cm, max = 126.0 cm), smaller than the zebu of Madagascar and small East African shorthorned zebu. These zebus were characterized by heterogeneous coloration patterns with, however, a high proportion for the "black plain" coat (32%). The horns were for the most part cup-shaped (70.25 %). By linking the Body Condition Score and the size of the hump (small, medium, large), the role of the hump as a reserve organ was highlighted and is part of a particularity of the zebu of Mayotte breed. By its phenotype, we can deduce that it belongs to the Small East African Zebus group recognized by the FAO.

4.2.2 Introduction

Documenting the diversity of livestock genetic resources by phenotype characterization of local livestock populations is a prerequisite to design strategies for their sustainable conservation.

Several populations of African cattle Small East african Zebu are raised (DAGRIS) in western part of Indian Ocean. In Mayotte island, three type of cattle co-exist : european type represented by Montbeliard individual, local zebu cattle, and the animals resulting of mating between the two first population. Mayotte island is part of geographic comorian archipelago. However, due to political context, few animal exchange occured between Mayotte and the three other islands but are not excluded. In Comoros, except imported european cattle and crossbred, several type of zebuine cattle remain : sahiwal breed imported in 2004, zebu from Madagascar and a primitive type almost disappeared [Saido, 2005]. Zebu from Madagascar was characterized with phenotype description in 1989 [Lauvergne and Zafindrajaona, 1992, Zafindrajaona, 1991] and by genotype characterization determined as a results of admixture between zebu and african taurine ancestry [Decker et al., 2014, Pitt et al., 2019] .

Previous study of bovine population of Mayotte was conducting and showed a weak level meat and milk production for local zebu cattle (2.5 kg of milk on average during 280 days ; male adult weight : 200 kg, for female : 193 kg). Nevertheless, some breeders were still interested in rearing this population because of a lack of infrastructure for water resources and a lack of space [Barthes, 2003] for cultivation of forage resources.

The aim of this study was to phenotypically characterize the local zebu cattle raised in Mayotte island. First, we presented inventory of the observed characteristics and the significant differences between male and female, then we investigated some relationships between variables.

TABLE 4.1 – Sampling objective of zebu individuals among township of Mayotte island.

Commune	Nb Farm	Nb Zebu	Prop. Zebu	Obj coll.	F	M
ACOUA	58	252	1,8%	7,27	5	2
BANDRABOUA	314	1628	11,6%	46,446	32	14
BANDRELE	190	793	5,7%	22,623	16	7
BOUENI	86	294	2,1%	8,48	6	2
CHICONI	151	542	3,9%	15,515	10	5
CHIRONGUI	224	874	6,2%	24,925	17	8
DEMBENI	294	1143	8,1%	32,633	23	10
DZAOUDZI	61	237	1,7%	6,87	5	2
KANI KELI	168	696	5,0%	19,820	14	6
KOUNGOU	185	808	5,8%	23,023	16	7
MAMOUDZOU	455	1957	14,0%	55,856	39	17
MTSANGAMOUJI	103	531	3,8%	15,115	10	5
MTZAMBORO	49	194	1,4%	5,56	4	2
OUANGANI	382	1513	10,8%	43,143	30	13
PAMANDZI	23	115	0,8%	3,33	2	1
SADA	174	677	4,8%	19,319	12	7
TSINGONI	342	1773	12,6%	50,651	35	16
TOTAL	3259	14027	100,0%	400	276	124

4.2.3 Material and methods

Sampling

The study was carried out in Mayotte island, which is a tiny island of 375 km² in the Mozambique Channel, in Indian Ocean. The climate of the island is characterized as tropical and the topography ranges form the sea level to 600 m. Temperature ranges from a minimum of 23 to a maximum of 30 degrees. Two seasons marked by rainfall differences : high rainfall during november to april (*Kashkasini*) and low to no rainfall during june to september (*Kussini*). Based on National Identification DataBase extraction of march 10th 2016, 400 individuals, 276 females and 124 males, identified as local zebu cattle from Mayotte were sampled and distributed among 17 townships of Mayotte Island (Table 4.1, Figure 4.2). The animals were chosen following the criteria of female/male proportion at the district level (2/3 female, 1/3 male), age superior to 18 month, minimizing family relationship, docility. Pregnant female were excluded from the sampling. On average, 3.1 animals were sampled in 128 farms.

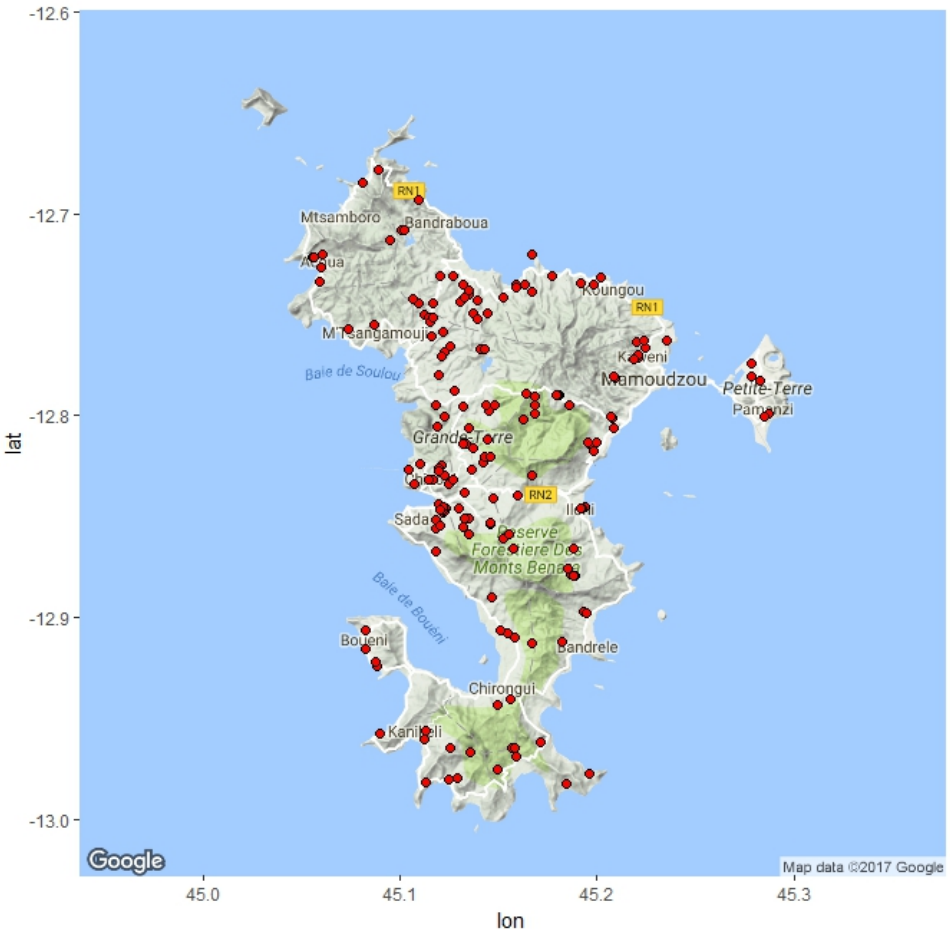


FIGURE 4.2 – Geographic position of the 128 farms where 400 local zebu cattle were sampled in Mayotte island

Morphological characterization

For each individuals, we collected the following meta-data : field login (if necessary, earring with a number was posed), database login (type DABVxxx), birth date, sex, field login mother and father if known, geographic position of farm. Every individual is described with different variables following FAO recommandations [FAO, 2012] : 20 qualitative indicators that concern hair color and pattern, skin color, conformation of several parts of body such as ears, horns shape and orientation, facial shape, tail size, back and rump profile. Body condition score is evaluated with scale from 1 (extremely thin) to 5 (extremely thick) following [Vall et al., 2002] guidelines. Measurements were recorded to the nearest centimetre using a cloth tape measure

for withers height, scapular-ischial lenght, thoracic girth, shoulders width, chest width, chest depth, hip length, hip width for body measure, and, horn lenght, distance between horn tips, distance between horn bases, arc length, horn girth for horn measurement as recommended by [FAO, 2012]. From measurements, 5 indexes were calculated (Table 4.2) : Balance, width-slope, RBI, pelvic, depth index.

Statistical analysis

Phenotypic dataset was analysed using 4.1.1 version of R software and classical methods of descriptive analysis to estimate sexual dimorphism. For qualitative variables, we calculated frequencies for each sex and performed χ^2 test. For quantitative variables, we calculated mean and standard deviation and performed analysis of variance with 1 factor (Sex) following the model : $Y_{ij} = \nu + S_j + E_{ij}$ with Y_{ij} measured value en each animal ; ν general mean ; S_j Sex effect and E_{ij} residual effect. In order to investigate the effect of age covariable and sex on the horn shape in the population, we compared for each sex the distribution of horn type in function of five age classes such as (1 year, 3 years], (3 years, 5 years], (5 years, 10 years], (10 years, 15 years], (15 years, 20 years]) and a χ^2 test was performed for female and male. For investigate the role of hump as body reserve, we performed a χ^2 test on each sex-specific contingency table relying hump size and body score condition.

TABLE 4.2 – Description of quantitative variables and calculated index from the variables for zebu form Mayotte. Reference "FAO guidelines" corresponds to [FAO, 2012].

Variable	**Description**	**Réferences**
Body		
Withers height	Vertical height from the bottom of front foot to the highest point of shoulder	FAO guidelines
Scapular-ischial length/body oblique length	Oblique distance from point of shoulder to hook bone of pelvis	FAO guidelines
Thoracic-parameter	Body circumference behind shoulder blades in a vertical plane	FAO guidelines
Shoulders width	Intercoastal distance just behind the withers	
Chest width	Distance between each point of shoulder	
Chest depth	Vertical distance from the top of the withers to the sternum	
Hip lenght	Distance between hip bone and hook bone of the pelvis	
Hip width	Distance between each hip bone	FAO guidelines
Horn		
Horn lenght	Distance btw. base and tip of the right horn foll. the curve side	FAO guidelines
Distance tip	Distance between the 2 horn tips	
Distance base	Distance between the 2 horn bases	
Arc lenght	Straight lenght btw. base and tip of the right horn	
Girth horn	Circumference of the right horn at the base	
Index		
Balance	(Hip lenght x Hip width) / (Chest depth x Shoulders width)	[Alderson, 1999, Salako, 2006]
Width-slope	Hip width - shoulders width	[Alderson, 1999]
RBI (Relative Body Index)	Scapular-ischial lenght / Withers height	[Chacón et al., 2011]
Pelvic	Hip width / Hip lenght	[Chacón et al., 2011]
Depth index	Chest depth / Withers height	[Alderson, 1999, Salako, 2006]

TABLE 4.3 – Overview of the frequency of each level for the qualitaive traits recorded in Zebu from Mayotte. Value of χ^2 and p-value to test sexual dimorphism is mentioned, a star indicates if test is significant.

Qualitative traits	**Tot** *(400)*	**Fem** *(276)*	**M** *(124)*	χ^2	**pval**
Coat Pattern				3.75	0.4408
1/2<pied<3/4	3.75	2	1.75		
1/4<pied<1/2	9.75	6.25	3.5		
3/4<pied	4	3.25	0.75		
Plain	60.5	42.75	17.75		
Spotted	22	14.75	7.25		
Color of coat				16.761	0.002151 (*)
Fawn	5.75	4.25	1.5		
Grey	3	1.75	1.25		
Black	58.5	39.25	19.25		
Dark red	9.75	19	4		
Light red	23	4.75	5		
Other color				12.621	0.01329 (*)
Absence	73.75	52.75	21		
Undefined	3.75	3.25	0.5		
Brindle	14.75	7.5	7.25		
Speckling	7.5	5.25	2.25		
Tiger type	0.25	0.25	-		
Rays of mulket				11.819	0.002713 (*)
Absence	82.25	59.75	22.5		
Light on dark coat	17	8.75	8.25		
Dark on light coat	0.75	0.5	0.25		
Spot on cephalic				5.7e-30	1
Absence	63.5	43.75	19.75		
Presence	36.5	25.25	11.25		
Type of poil				0.0009	0.9759
Medium	29.25	22.25	7		
Small (ras)	70.75	46.75	24		
Cephalic profile				1.0436	0.3069
Convex	1.25	1.25	-		
Straight	98.75	67.75	31		
Hump position				1.4631	0.2264
Cervico-thoracic	1.5	1.5	-		
Thoraco-thoracic	96	67.5	31		
Hump size				90.905	< 2.2e-16 (*)
High	23.25	7.25	16		
Medium	40	29	11		
Small	36.75	32.75	4		
Presence of horn				-	-
Presence	100	69	31		
Color of horn				16.643	0.0008368 (*)
White	1.5	1.25	0.25		
Brown	77	49.25	27.75		
Black	4.5	3.5	1		
Several colors	17	15	2		

TABLE 4.3 (suite)

Qualitative traits	**Tot** *(400)*	**Fem** *(276)*	**M** *(124)*	χ^2	**pval**
Horn shape				44.344	4.897e-07 (*)
Cup	70.25	43.5	26.75		
Folded back cup	1.5	1.5	-		
Crescent	12.75	12.5	0.25		
En fourche	5.25	5.25	-		
Straight	3.75	2.25	1.5		
Lyre	0.25	0.25	-		
Wheel	0.5	0.5	-		
V-shape	5.25	2.75	2.5		
Undefined	0.5	0.5	-		
Ear shape				0.29938	0.5843
Rounded	34.75	23.25	11.5		
Pointed	65.25	45.75	19.5		
Orientation of ear				-	-
Horizontal	100	69	31		
Dewlap size				29.811	3.363e-07 (*)
High	12	5.75	6.25		
Medium	62.25	40.5	21.75		
Small	25.75	22.75	3		
Navel size (for female				-	-
Absence	-	66.30	-		
High	-	-	-		
Medium	-	2.17	-		
Small	-	31.52	-		
Preputial sheath (for male)				-	-
Absence	-		-		
High	-	-	0.81		
Medium	-	-	80.65		
Small	-	-	18.55		
Tail length				2.0943	0.3509
Long	35.75	23.25	12.50		
Medium	48.75	35.25	13.50		
Small	15.50	10.50	5.00		
Backline profile				2.3776	0.4978
Straight	30.50	21.00	9.50		
Slopes up towards rump	64.00	45.00	19.00		
Slopes down from withers	0.50	0.25	0.25		
Dipped	5.00	2.75	2.25		
Rump profile				4.5227	0.1042
Flat	3.50	2.00	1.50		
Sloping	94.50	65.00	29.50		
Roofy	2.00	2.00	-		

4.2.4 Results

The quantitative results are shown in Table 4.4. Significant difference ($P < 0.05$) between sex was observed for hip, chest and shoulders width, balance, width-slope and pelvic indexes. The qualitative results are shown in Table 4.3. Body color, rays of mulket, hump size, dewlap size were significantly different between the sex ($P < 0.05$) and horn color and shape were concerned by sexual dimorphism.

TABLE 4.4 – Overview of the quantitative variables measured on 400 zebu individuals. Level of significance on the effects of sex on the linear body measurements for the whole data are indicated with **Pvalue** column.

Variables	**Mean**	**Mean**	**P-value**	**S-e**	**S-e**
	F (n = 276)	M (n = 124)		F	M
Withers height	105.3	106.5	-	6.3	7.2
Scapular-ischial length/body oblique length	118.9	119.1	-	8.1	10.5
Thoracic-parameter	139.8	141.6	0.1	8.9	12.3
Shoulders width	39.6	41.2	0.05	6.1	6.9
Chest width	33.2	35.3	0.001	4.9	5.5
Chest depth	54.6	54.8	-	5.5	6.0
Hip lenght	39.0	38.9	-	3.2	4.0
Hip width	35.2	33.4	0.001	3.9	4.8
Horn					
Horn lenght	23.95	14.99	0.001	8.40	5.04
Distance tip	34.97	36.94	-	10.55	7.78
Distance base	14.67	16.02	0.001	1.65	3.13
Arc lenght	17.11	12.63	0.001	5.46	4.36
Horn girth	16.18	17.88	0.001	2.48	2.96
Index					
Balance	0.65	0.59	0.001	0.12	0.10
Width-slope	0.90	0.83	0.001	0.13	0.12
RBI (Relative Body Index)	1.13	1.12	-	0.09	0.09
Pelvic	0.91	0.86	0.001	0.10	0.11
Depth index	0.52	0.51	-	0.06	0.05

The colour parameters of the local zebu cattle of Mayotte island are non homogeneous : black was the most colour found (58,4%) then light red (23%), dark red (9.5%), Fawn (5.75%) and grey (3.00%). The plain coat pattern was the most body pattern observed (60.5%), then spotted pattern (22.00%) and pied pattern (17.5%). Some color particularities were observed such as brindle, speckling and one animal presented tiger type color. 17.75% of the animals

have rays of mulket on the back with a difference between male and female. Almost one third of the population showed a spot on cephalic. The length of hair was at 70.75 % small and medium at 29.25 %.

The ears were all horizontal, almost one third was rounded shape and two third was pointed shape. Concerning horns, all the individuals had horns with several shapes : the major type was cup (70.25 %), then for less than 20%, crescent, en fourche, straight, lyre, wheel, V-shape. 1 animal presented undefined type horn shape. Some shapes were not encountered in male populations such as folded back up, en fourche, lyre, wheel. On average, horn lenght was higher for female (23.95 cm +/-8.40) than for male (14.99 cm +/-5.04). Interestingly, for female, the proportion of cup and straight shape decreased with age class while crescent and fourche shape were increased. Lyre shape was present only in female (15,20] age class.

Concerning body shape, dewlap was not very developped with a medium (62.25 %) or small (25.75 %) size. For female, the navel was absent (66.30%) or small (31.52%) and so for preputial sheath for male population with 80.65% and 18.55% respectively. Tail lenght was variable but in majority medium size (48.75 %). The back profile sloped up towards rump at 64.00 % or straight (30.50 %). The rump profile was sloping (94.50 %).

TABLE 4.5 – Relationship between body score condition and hump size. For male and female, the distribution in BDS score for one modalitie of hump size is compared to the two others.

		Hump Size			χ^2	**pvalue**
		High	Medium	Small		
Female					29.923	4.066e-05 (*)
	1	1	17	27		
BDS	2	17	80	99		
	3	10	17	5		
	4	1	2	-		
Male					14.924	0.004862 (*)
	1	3	2	-		
BDS	2	23	28	13		
	3	38	14	3		
	4	-	-	-		

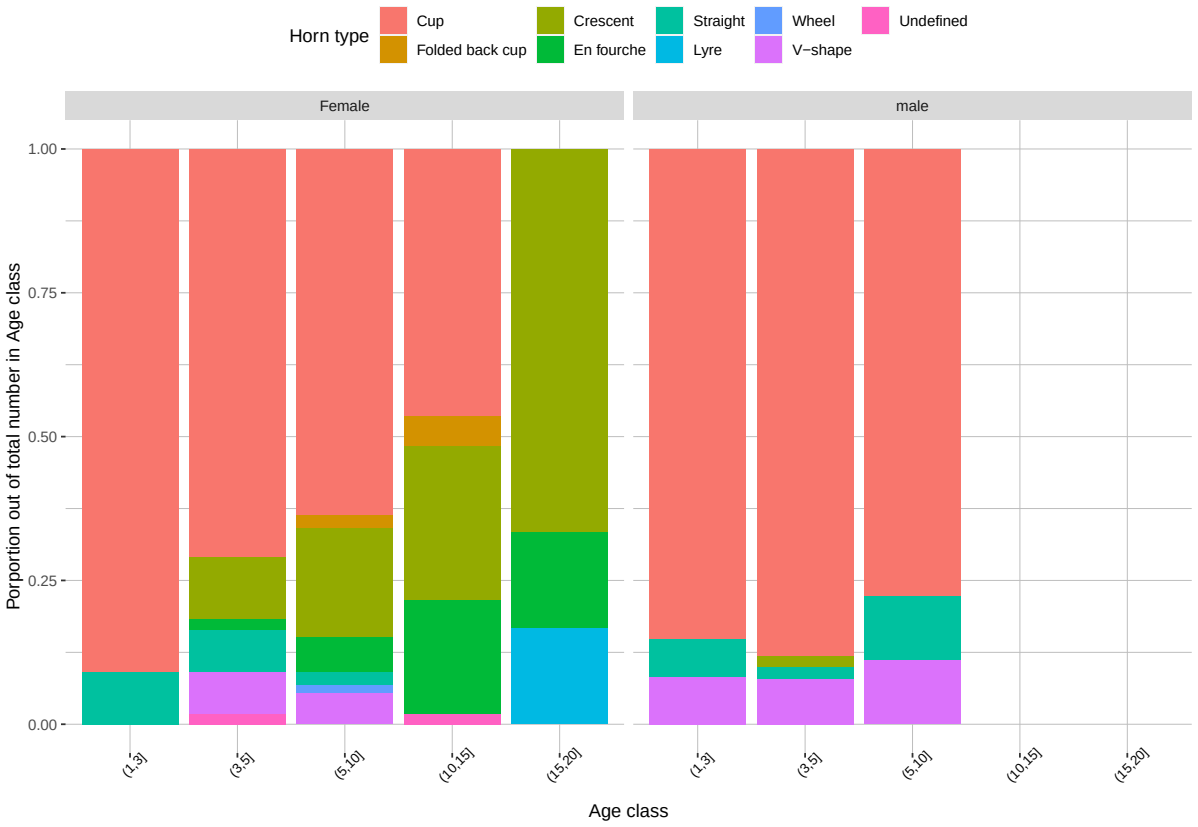


FIGURE 4.3 – Proportion of horn type in function of age class between male and female individuals. χ^2 = 106.64, df = 32, p-value = 5.892e-10 | χ^2 = 3.5023, df = 6, p-value = 0.7437

4.2.5 Discussion

Phenotypic characterization allows to reach conclusions about production ability and adaptability of individuals to harsh environment. Zebu from Mayotte was smaller and thinner than Zebu from Madagascar as the withers height and thoracic girth showed : for withers height, the average measurement for zebu from Madagascar are 121 cm for male and 111 cm for female [Lauvergne and Zafindrajaona, 1992] and for the small East African shorthorn zebu (SEAZ), 117cm for male and 109 cm for female [Mwacharo et al., 2006]. For thoracic girth, Zebu from Madagascar sample presented 166 cm for male and 130 cm for female [Lauvergne and Zafindrajaona, 1992] and for SEAZ, the average measurement was 157 cm for male and 143 cm for female. Zebu from Mayotte was on average shorter than zebu from Madagascar (139 cm for male, 126 cm for female) [Lauvergne and Zafindrajaona, 1992] and male SEAZ (129 cm) [Mwacharo et al., 2006]. Female MAY were longer than female SEAZ (116 cm) [Mwacharo et al., 2006].

In general, skeletal measurements relative to inherent size of the animals, such as body height and lenght, chest depth were less affected by nutrition, excepting during growing phase [Coleman and Evans, 1986, Coleman et al., 1993, Keogh et al., 2015] and width and girth measurements were related to muscle and fat deposition [Kamalzadeh et al., 1998] and so were more susceptible to vary in function of nutritional resources [Keogh et al., 2015]. Moreover, the calculus of conformation indices [Alderson, 1999, Chacón et al., 2011] give an insight in the assessment of conformation and functional aptitude of the individuals. Indeed, RBI [Lauvergne and Zafindrajaona, 1992], superior to 1, was an indice that give information about the compacity of the body conformation. Concomitant with the global small size of the population, the hip bone width was 10 cm less than in SEAZ [Mwacharo et al., 2006], which can be a indication of calving capacity. However, the inclination of the pelvis (94.5% of the

sample presents "sloping" modality for rump profile) and the pelvic indice (Hip width / Hip length) superior to 0.85 give clue of the good conformation for parturition capacity for Zebu from Mayotte.

Horn size and shape were significantly different between male and female, this was probably due to the difference on age than horn differentiation due to sex [Santiago-Moreno et al., 2012] : mean female age was higher than for male. Knowing that horns grow during cow life, there was more chance than horn were much longer and reach more often crescent shape, probably following cup shape for female (Figure 4.3).

The characterized relationship ($P < 0.05$) (Table 4.5) between hump size and body score condition was one clue of adaptation capacity of zebu from Mayotte to fluctuant nutrient and water resources and the role of hump as part of body reserve. It could be interesting to follow body condition score of several animals and measure hump size regularly during one year.

This population of zebu from Mayotte presented a heterogeneous phenotype probably due to the absence of selection on well-defined external traits such as coat color and pattern, horn shape. Due to that, its difficult to classify this population [Marshall et al., 2019]. As a consequence, genetic characterization is necessary to estimate breed composition and determine the pressions of selection that have shaped phenotypic traits in the particular environment and livestock system of Mayotte.

Bibliographie

[Alderson, 1999] Alderson, G. (1999). The development of a system of linear measurements to provide an assessment of type and function of beef cattle. *Animal Genetic Resources Information*, 25 :45–55.

[Barthes, 2003] Barthes, C. (2003). *L'état et le monde rural à Mayotte*. KARTHALA Editions.

[Chacón et al., 2011] Chacón, E., Macedo, F., Velázquez, F., Paiva, S. R., Pineda, E., and McManus, C. (2011). Morphological measurements and body indices for Cuban Creole goats and their crossbreds. *Revista Brasileira de Zootecnia*, 40(8) :1671–1679.

[Coleman and Evans, 1986] Coleman, S. W. and Evans, B. C. (1986). Effect of Nutrition, Age and Size on Compensatory Growth in Two Breeds of Steers. *Journal of Animal Science*, 63(6) :1968–1982.

[Coleman et al., 1993] Coleman, S. W., Evans, B. C., and Guenther, J. J. (1993). Body and carcass composition of Angus and Charolais steers as affected by age and nutrition. *Journal of Animal Science*, 71(1) :86–95.

[Decker et al., 2014] Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Molina Alcalá, A., Sonstegard, T. S., Hanotte, O., Götherström, A., Seabury, C. M., Praharani, L., Babar, M. E., Correia de Almeida Regitano, L., Yildiz, M. A., Heaton, M. P., Liu, W.-S., Lei, C.-Z., Reecy, J. M., Saif-Ur-Rehman, M., Schnabel, R. D., and Taylor, J. F. (2014). Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *PLoS Genetics*, 10(3) :e1004254.

[FAO, 2012] FAO (2012). *Phenotypic characterization of animal genetic resources*. Number 11 in FAO animal production and health guidelines. Food and Agriculture Organization of the United Nations, Rome.

[Kamalzadeh et al., 1998] Kamalzadeh, A., Koops, W. J., and van Bruchem, J. (1998). Feed quality restriction and compensatory growth in growing sheep : Modelling changes in body dimensions. *Livestock Production Science*, 53(1) :57–67.

[Keogh et al., 2015] Keogh, K., Waters, S. M., Kelly, A. K., and Kenny, D. A. (2015). Feed restriction and subsequent realimentation in Holstein Friesian bulls : I. Effect on animal performance ; muscle, fat, and linear body measurements ; and slaughter characteristics1. *Journal of Animal Science*, 93(7) :3578–3589.

[Lauvergne and Zafindrajaona, 1992] Lauvergne, J. and Zafindrajaona, P. S. (1992). Comparaison de deux populations de Zébu Malgache par les distances génétiques biométriques. 45(2) :167–174.

[Marshall et al., 2019] Marshall, K., Gibson, J. P., Mwai, O., Mwacharo, J. M., Haile, A., Getachew, T., Mrode, R., and Kemp, S. J. (2019). Livestock Genomics for Developing Countries – African Examples in Practice. *Frontiers in Genetics*, 10 :297.

[Mwacharo et al., 2006] Mwacharo, J. M., Okeyo, A. M., Kamande, G. K., and Rege, J. E. O. (2006). The small East African shorthorn zebu cows in Kenya. I : Linear body measurements. *Tropical Animal Health and Production*, 38(1) :65–74.

[Pitt et al., 2019] Pitt, D., Sevane, N., Nicolazzi, E. L., MacHugh, D. E., Park, S. D. E., Colli, L., Martinez, R., Bruford, M. W., and Orozco-terWengel, P. (2019). Domestication of cattle : Two or three events ? *Evolutionary Applications*, 12(1) :123–136.

[Saido, 2005] Saido (2005). Rapport National sur l'Etat des Ressources Génétiques Animales - Comores. Technical report, FAO.

[Salako, 2006] Salako, A. E. (2006). Application of Morphological Indices in the Assessment of Type and Function in Sheep. *International Journal of Morphology*, 24(1).

[Santiago-Moreno et al., 2012] Santiago-Moreno, J., Gómez-Brunet, A., Toledano-Díaz, A., Salas-Vega, R., Gómez-Guillamón, F., and López-Sebastián, A. (2012). Role of testosterone and photoperiod on seasonal changes in horn growth and sperm variables in the Iberian ibex : a model for polygynous wild bovids. *Journal of Endocrinology*, 214(2) :155–163.

[Vall et al., 2002] Vall, E., Meyer, C., Abakar, O., and Dongmo, A. L. (2002). Note d'état corporel des zébus de trait (NEC) : un outil simple pour piloter l'alimentation.

[Zafindrajaona, 1991] Zafindrajaona, P. S. (1991). *Profils genetiques du zebu malgache*. thesis, Paris 11.

4.3 Article 2 : Phenotypic characterization of sheep from Mayotte

(en préparation)

Jessica Magnier[1], Jérôme Janelle[2], Oussoufi Saindou[4], Mélissa Ouvrard[3], Solène Raoul[3], Michel Naves[5], Mathieu Gautier[6], Laurence Flori[1], Emmanuel Tillard[2]

[1] UMR SELMET, INRAE, CIRAD, Montpellier SupAgro, Univ. Montpellier, Montpellier, France

[2] UMR SELMET, INRAE, CIRAD, Montpellier SupAgro Univ., Saint Pierre, La Réunion, France

[3] Coopadem, Coconi, 97660 Ouangani, Mayotte

[4] CAPAM 976, Coconi, 97660 Ouangani, Mayotte

[5] URZ, INRAE, Petit-Bourg, France

[6] UMR CBGP, INRAE, CIRAD, IRD, Montpellier Supagro, Univ. Montpellier, Montferrier-sur-Lez, France

4.3.1 Abstract

In Mayotte island, 100 individuals (2/3 female, 1/3 male) were sampled out of the 1'000 individuals identified locally by the Chamber of Agriculture of Mayotte. 27 farms were visited in 17 townships of the island. The population of Mayotte sheep has an average height at the withers of 55.76 cm (min = 44.0 cm, max = 67.0 cm). These sheep are characterized by a heterogeneous coat of "hair" type with however more frequent patterns "light red piebald" (30%), "black piebald" (20%) and "light red with isolated white spot" (15%). Females are hornless and only 1/4 of the males are horned. By relating Body Condition Score with fat tail size, the role of the tail as a reserve organ was highlighted. In addition, 21% of the individuals have an ear mutation classified as "short ear" or "vestigial". This population of sheep is very close phenotypically to the sheep present in Madagascar which belong to the group of fat tailed sheep.

4.3.2 Introduction

A sheep population was observed in Mayotte Island but little information was available about the history of this local breed. Its earliest presence was estimated in Dembeni site dating back 11th-12th [Wright et al., 1984]. Other recent archeological findings in Acoua indicated skeletal evidence of small ruminant presence, tracing back to the 13th-15th centuries CE (Antsiraka Boira [Pauly and Ferrandis, 2018]). Effectif of sheeps in Mayotte Island was estimated around 1'000 individuals [France, 2011] and this population was considered as fat-tailed, hair sheep.

Sheep breeds are classified according to hair/wool type and tail type (DAGRIS, Dad-is). Fat-tailed sheep breeds were widely studied in several geographic localisations in Indonesia [Udo and Budisatria, 2011], Southest Asia Africa [Pourlis, 2011, Almeida, 2011] but were not well known in the islands of West Area of the Indian Ocean. The closest geographically sheep population from Mayotte is the Madagascar sheep which was classified in African Fat-tailed type, regarding the long fat tailed and coarse fibre of the hairy coat (Mason and Maule 1960 in DAGRIS). This breed is principally raised in the southern part of Madagascar. In the western part of the Indian Ocean, another sheep population was inventoried in Comoros, and considered as Djallonké sheep [Saido, 2005] which is a thin-tailed sheep type, a breed originated of West african area [Meyer et al., 2004].

A previous study conducted in Mayotte (not published) about livestock system, reproduction and production level was conducted by CIRAD research group : sheep are generally raised jointly with goat in weak proportion in small paddock. Transhumance is not conducted due to landed issues. Few exchange of male individuals, in order to mating, occured between the different farms. In an administrative way, although Mayotte Island is considered as French department, all livestock species are not identified. Moreover, for sheep and goat, only a local

database is managed by CAPAM (Agricultural and fisheries consulary organism from Mayotte).

We presented here a phenotypic characterization of the sheep population in Mayotte island in order to classify among the differents groups of african sheep.

4.3.3 Material and methods

Sampling

As an exhaustive inventory of sheep population in Mayotte island was not available, a preliminary visit on 27 husbandry was conducted in order to determine sample size, ratio female/male. Finally, 100 sheeps were described among 20 husbandry localized on 17 townships of Mayotte Island (Figure 4.4) with a proportion of 81 females and 19 males.

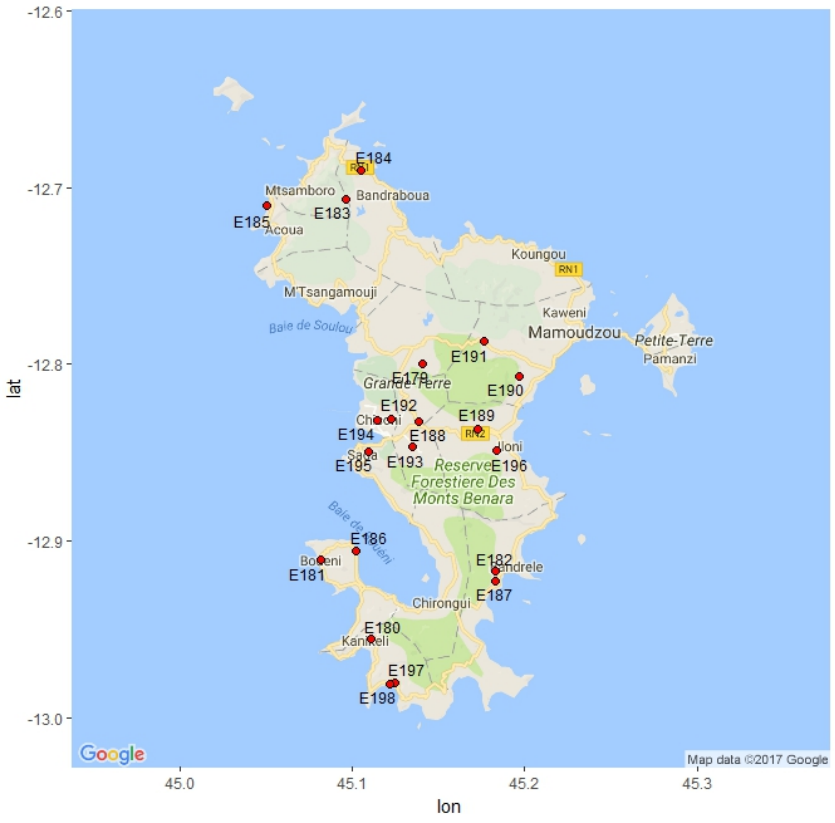


FIGURE 4.4 – Geographic positions of the 20 farms where the 100 sheep individuals were phenotypically sampled.

Morphological characterization

For each sheep, we collected the following meta-data : field login (if necessary, earring with a number was posed), database login (type DAOVxxx), birth date or estimated age, sex, field login mother, field login father. Every sheep is described with different indicators recommanded by FAO [FAO, 2012] and some supplementary indicators : 15 quantitative indicators (Table 4.6) which are bathymetry measures on body and horns in order to appreciate global size of animal. From measurements, 5 indexes were calculated (Table 4.6) : Balance, width-slope, RBI, pelvic, depth index. 14 qualitative indicators (Table 4.8) which are hair color and pattern, skin color, conformation of several parts of body such as ears, horns shape and orientation, facial shape, tail type and shape, back and rum profile. Body condition score is evaluated with scale from 1 (extremely thin) to 5 (extremely thick). To better describe ear status, we choose to estimate length of ears by both quantitative and qualitative variables. We measured ear length (in cm) and we classified ear according to the shape : if the end of the ear is round, the ear is considered as "normal", if the end of the ear described triangular shape, ear is classifed as "small".

Statistical analysis

Phenotypic dataset were analysed using 4.1.1 version of R software and classical methods of descriptive analysis : calcul of frequencies and χ^2 test for qualitative variables mean, standard deviation and for quantitative variables, analysis of variance with 1 factor (Sex) :

$Y_{ij} = \nu + S_j + E_{ij}$ with Y_{ij} measured value en each animal ; ν general mean ; S_i Sex effect and E_{ij} residual effect.

For investigate the role of tail as body reserve, we performed an analysis of variance, for each sex-specific sub-dataset, on medium tail circumference with body score condition (BDS)

as a factor. Pair comparison between (BDS) modalities were performed with HSD test.

TABLE 4.6 – Description of quantitative variables and calculated index from the variables for sheep from Mayotte. Reference "FAO guidelines" corresponds to [FAO, 2012].

Variable	Description	Réferences
Body		
Withers height	Vertical height from the bottom of front foot to the highest point of shoulder	FAO guidelines
Scapular-ischial length/body oblique length	Oblique distance from point of shoulder to hook bone of pelvis	FAO guidelines
Thoracic-parameter	Body circumference behind shoulder blades in a vertical plane	FAO guidelines
Shoulders width	Intercoastal distance just behind the withers	
Shin circumference	Circumference of shin bone	FAO guidelines
Chest depth	Vertical distance from the top of the withers to the sternum	
Hip lenght	Distance between hip bone and hook bone of the pelvis	
Hip width	Distance between each hip bone	FAO guidelines
Head lenght	Distance between top of the crane and top of the muzzle	
Head width	Distance between the two ears	
Horn lenght	Distance btw. base and tip of the right horn foll. the curve side	FAO guidelines
Ear lenght	Distance between base and tip of the ear	FAO guidelines
Tail lenght	Distance between base and tipe of the tail	
Tail circumference	Circumference of the tail at medium distance	
Hair lenght	Length of the hair in rump area	FAO guidelines
Index		
Balance	(Hip lenght x Hip width) / (Chest depth x Shoulders width)	[Alderson, 1999, Salako, 2006]
Width-slope	Hip width - shoulders width	[Alderson, 1999]
RBI (Relative Body Index)	Scapular-ischial lenght / Withers height	[Chacón et al., 2011]
Pelvic index	Hip width / Hip lenght	[Chacón et al., 2011]
Depth index	Chest depth / Withers height	[Alderson, 1999, Salako, 2006]

TABLE 4.7 – Overview of the quantitative variables measure on 100 sheep individuals. Level of significance (Pvalues) on the effects of sex on the linear body measurements for the whole data

Variables	**Mean**	**Mean**	**P-value**	**S-e**	**S-e**
	F (n = 81)	M (n = 19)		F	M
Withers height	55.4	57.3	-	3.7	5.4
Scapular-ischial length/body oblique length	56.5	58.8	-	4.6	5.6
Thoracic-parameter	73.9	78.5	0.05	6.7	9.2
Shoulders width	18.2	20.8	0.001	2.5	2.6
Shin circumference	3.7	5.1	0.001	1.1	1.6
Chest depth	30.0	31.7	0.05	3.1	3.4
Hip lenght	17.5	17.7	-	2.0	2.4
Hip width	17.9	19.3	-	2.9	3.1
Head lenght	6.8	8.3	0.05	2.6	3.0
Head width	12.7	13.6	0.001	1.1	1.2
Horn lenght	-	7.0	-	-	3.3
Ear lenght	5.3	10.21	0.001	4.9	6.5
Tail lenght	11.0	15.0	0.001	4.3	4.3
Tail circumference	16.5	27.0	0.001	5.2	7.0
Hair lenght	6.7	9.3	0.001	2.8	3.7
Index					
Balance	0.58	0.51	0.05	0.11	0.8
Width-slope	-0.25	-1.47	-	2.56	2.52
RBI (Relative Body Index)	1.02	1.03	-	0.07	0.08
Pelvic	1.03	1.13	-	0.21	0.25
Depth index	0.54	0.55	-	0.05	0.05

4.3.4 Results

Global phenotype The quantitative results are shown in Table 4.7. Significance difference ($P < 0.05$) between male and female was observed for thoracic parameter, shoulders width, shin circumference, chest depth, head length and width, ear length, tail length and circumference and hair length.

Concerning the qualitative observation, the χ^2 analysis did not support sexual dimorphism, except for horn presence (Table 4.8). Only 5 male individuals presented black ($n = 2$) or brown ($n = 3$) curved horns. These horns were 7 cm mean. The Mayotte population presented a diversified phenotype for the criterion of body color : the three most frequent observed body color are light red patchy (30%), black patchy (20%), light red spotted (15%). 24% presents

TABLE 4.8 – Overview of the frequency of each level for the qualitative traits recorded for local sheeps from Mayotte. Value of Khi2 and p-value to test sexual dimorphism is mentioned, a star indicates if test is significant.

Qualitative traits	**Tot** *(100)*	**Fem** *(81)*	**M** *(19)*	**Khi2**	**pval**
Hair coat colour				5.3627	0.2521
Fawn	8.00	8.00	-		
Grey	1.00	1.00	-		
Black	28.00	21.00	7.00		
Light red	52.00	44.00	8.00		
Dark red	11.00	7.00	4.00		
Hair coat colour pattern				2.5728	0.7655
Plain	18.00	14.00	4.00		
3/4<Pied	31.00	26.00	5.00		
1/2<Pied<3/4	12.00	8.00	4.00		
1/4<Pied<1/2	16.00	14.00	2.00		
Spotted	22.00	18.00	4.00		
Head patch				0.0012824	0.9714
Presence	24.00	20.00	4.00		
Absence	76.00	61.00	15.00		
Body skin colour				1.0183e-30	1
Pigmented	87.00	70.00	17.00		
Non-pigmented	13.00	11.00	2.00		
Fibre type				0.63074	0.4271
Hair	99.00	81.00	18.00		
Wool	1.00	-	1.00		
Horn presence				17.239	3.295e-05(*)
Presence	5.00	-	5.00		
Absence	95.00	81.00	14.00		
Color of horn				-	-
Brown	2.00	-	2.00		
Black	3.00	-	3.00		
Horn shape				-	-
Curved	5.00	-	5.00		
Facial profile				4.9708	0.08329
Straight	96.00	78.00	18.00		
Concave	1.00	-	1.00		
Convex	3.00	3.00	-		
Orientation of ear				0.63074	0.4271
Pendulous	99.00	81.00	18.00		
Semi-pendulous	1.00	-	1.00		
Ear size				2.3559	0.3079
Earless	6.00	5.00	1.00		
Small	15.00	10.00	5.00		
Normal	79.00	66.00	13.00		
Back profile				-	-
Straight	100.00	81.00	19.00		
Rump profile				6.7378e-29	1
Slopping	98.00	79.00	19.00		
Roofy	2.00	2.00	-		

a white spot hair in face.

Fat-tailed Linkage between Body condition score and tail circumference was tested by analysis of variance for each male and female sampled. We noticed a positive correlation between these two variables $(P < 0.001, F - stat)$, which confirmed role of tail as a storage organ. We confirmed fat-tailed type of Mayotte sheep (Figure 4.5).

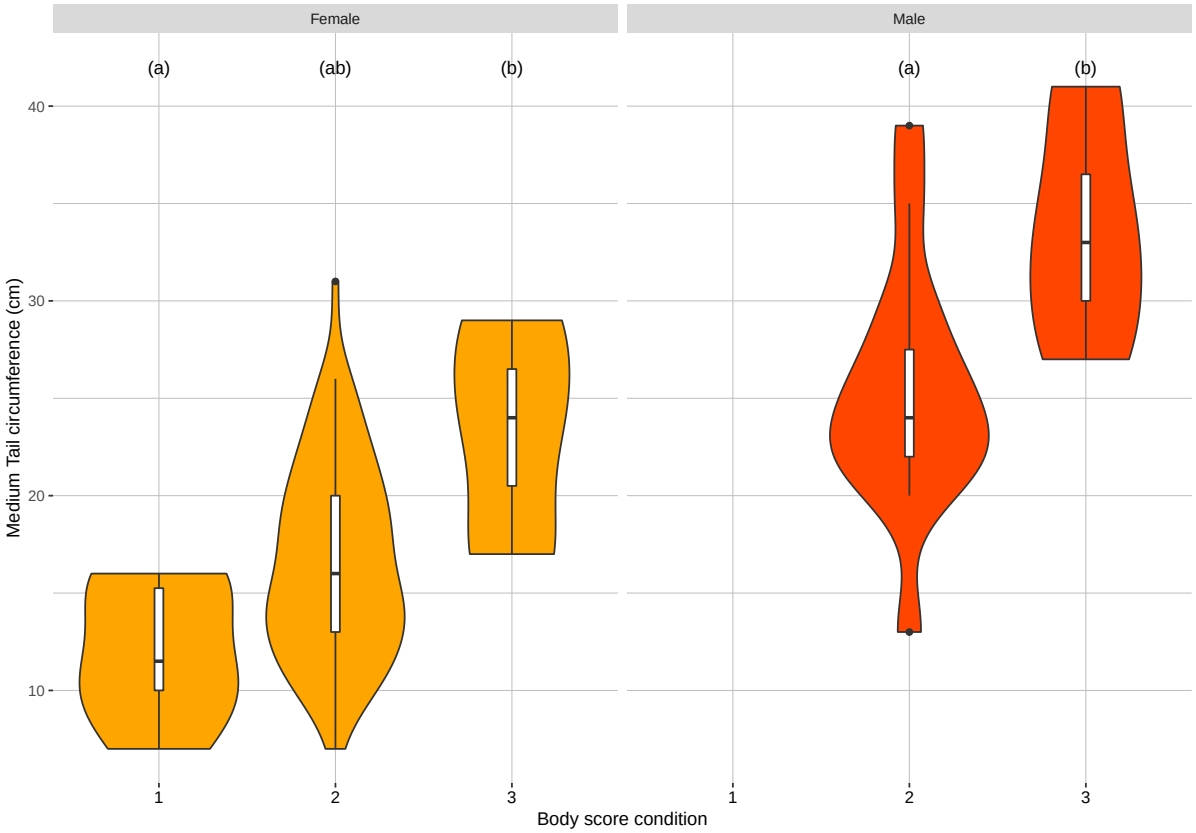


FIGURE 4.5 – Violin plots describing relationship between Body score condition and Medium tail circumference for Female (yellow) and Male (red) samples. Analysis of variance was performed with R-software and pair comparisons with HSD test.

Ear size Focusing on ear status, we noticed a strong correlation between quantitative and qualitative measurement of ear $(P < 0.001, F - stat)$. The frequency of "small" ear was estimated at 15% and for "earless" modality at 6%. As a reminder, microtia is considered as abnormality of ear development. So we considered modalities "small" and earless" together to estimate microtia prevalence, which was 21% in our population.

4.3.5 Discussion

Our studied sheep population was phenotypically close to sheep from Madagascar that was roughly described [Anonymous, 1950, Anonymous, 1952] : for sheep from Madagascar, the coat was red, brown or black with large white patches, the head profile was convex, the ears were long and pendent, the horns were small or absent, fat-tail. No studies were conducted anymore for this breed.

Sheep from Mayotte was on average as tall and long as breeds raised for meat production as Menz and Nguni sheep (Wither height equal to a range of value from 54.07 to 60.1 cm for Menz sheep and from 54.03 to 63.66 cm for Nguni sheep ; Scapulo-ischial length equal to a range of value from 47.91 to 60.1 cm for Menz sheep). However sheep from Mayotte presented a longer thoracic perimeter than Menz and Nguni breeds (from 62.8 to 67.7 cm for Menz sheep and from 62.13 to 77.2 cm for Nguni sheep) but they were as thick as the others African fat-tailed breeds raised for meat production [Pourlis, 2011].

Concerning tail measurements, the medium circonference was higher than for Menz sheep (13.47 cm), but the tail was thinner that Middle East fat-tailed breeds such as Mehraban, Ghezel, Makoei, Awassi that presented a medium circumference above 33 cm [Pourlis, 2011]. The external localization of the fat in the tail allowed better heat dissipation from the rest of the body [Degen and Shkolnik, 1978] due to a less isolation of the body by the fat tissue. This fat represented also an energy store that can be mobilized in times of dietary deficit [Chilliard et al., 2000, Atti et al., 2004]. The energy supply by the tailed fat was accompagnied by some metabolic water formation that could help in filling the water requirements to some extent [Sileshi et al., 2003, Sejian et al., 2017].

Our population presented a auricular mutation. The phenotype of ear microtia was encountered in several sheep populations. High prevalence of shortened ear ($> 50\%$) were reported

in local populations of Central Europe, Asia and Africa [Lauvergne et al., 1988]. In Africa, the fat-tailed breed where microtia mutation occured are zimbabwean Sabi, ethiopian Afar breed (prevalence of 80% in Tigray area), Red Masai in Kenya (7%), East African Blackheaded, Rwanda and Burundi sheep, Nguni sheep (12%) [Wilson, 1991]. Microtia was encountered in other mammals such as human [Schorderet et al., 2008, Luquetti et al., 2012], mice [Munroe et al., 2009], goat [Wilson, 1991] and cow [Koch et al., 2013]. In human population, it was associated with other deleterious phenotype of abnormalities of the face and jaw, kidney and heart problems, and vertebral deformities [Luquetti et al., 2012]. In sheep populations, it seemed that this mutation was not associated with default, nor in sheep from Mayotte. However the microtia, especially when no ear or only a piece of skin was present, did not facilitate identification by French official ear straps. Furthermore, in the context of climate adaptation, the ear could play a role in heat dissipation by the action of the arteriovenous anastomoses (AVAs). These bypass channels join arteries and veins in skin and act in transport of heat outside the body for dissipation by increasing total blood flow to the periphery [Sejian et al., 2017].

The sheep from Mayotte was a local breed morphologically closed to African Long-Fat-Tailed sheep group that was well adapted to tropical climate of the island. However, the estimated small size of this population should be taken into high consideration in the management of the breed. Because of the lack of information on the genealogy of animals, genotypic characterisation is necessary to estimate genetic diversity and inbreeding coefficient. Genotyping could be used to determine the genetic determinant of the auriculary mutation and to infer the history of this population.

Bibliographie

[Alderson, 1999] Alderson, G. (1999). The development of a system of linear measurements to provide an assessment of type and function of beef cattle. *Animal Genetic Resources Information*, 25 :45–55.

[Almeida, 2011] Almeida, A. M. (2011). The Damara in the context of Southern Africa fat-tailed sheep breeds. *Tropical Animal Health and Production*, 43(7) :1427–1441.

[Anonymous, 1950] Anonymous, N. (1950). L'élevage du mouton à Madagascar. *Revue d'élevage et de médecine vétérinaire des pays tropicaux*, 4(4) :204.

[Anonymous, 1952] Anonymous, N. (1952). Rapport sur l'élevage à Madagascar. *Revue d'élevage et de médecine vétérinaire des pays tropicaux*, 5 :139.

[Atti et al., 2004] Atti, N., Bocquier, F., and Khaldi, G. (2004). Performance of the fat-tailed Barbarine sheep in its environment : adaptive capacity to alternation of underfeeding and re-feeding periods. A review. *Animal Research*, 53(3) :165–176.

[Chacón et al., 2011] Chacón, E., Macedo, F., Velázquez, F., Paiva, S. R., Pineda, E., and McManus, C. (2011). Morphological measurements and body indices for Cuban Creole goats and their crossbreds. *Revista Brasileira de Zootecnia*, 40(8) :1671–1679.

[Chilliard et al., 2000] Chilliard, Y., Ferlay, A., Faulconnier, Y., Bonnet, M., Rouel, J., and Bocquier, F. (2000). Adipose tissue metabolism and its role in adaptations to undernutrition in ruminants. *Proceedings of the Nutrition Society*, 59(1) :127–134.

[Degen and Shkolnik, 1978] Degen, A. A. and Shkolnik, A. (1978). Thermoregulation in Fat-Tailed Awassi, a Desert Sheep, and in German Mutton Merino, a Mesic Sheep. *Physiological Zoology*, 51(4) :333–339.

[FAO, 2012] FAO (2012). *Phenotypic characterization of animal genetic resources*. Number 11 in FAO animal production and health guidelines. Food and Agriculture Organization of the United Nations, Rome.

[France, 2011] France, M. (2011). Synthèse illustrée du recensement agricole 2010. Technical report, Ministère de l'Agriculture, de l'Alimentation, de la Pêche, de la Ruralité et de l'Aménagement du Territoire, Mayotte.

[Kim et al., 2018] Kim, Y., Dommergues, L., M'sa, A. B., Mérot, P., Cardinale, E., Edmunds, J., Pfeiffer, D., Fournié, G., and Métras, R. (2018). Livestock trade network : potential for disease transmission and implications for risk-based surveillance on the island of Mayotte. *Scientific Reports*, 8(1) :11550.

[Koch et al., 2013] Koch, C. T., Bruggmann, R., Tetens, J., and Drögemüller, C. (2013). A Non-Coding Genomic Duplication at the HMX1 Locus Is Associated with Crop Ears in Highland Cattle. *PLoS ONE*, 8(10).

[Lauvergne and Zafindrajaona, 1992] Lauvergne, J. and Zafindrajaona, P. S. (1992). Comparaison de deux populations de Zébu Malgache par les distances génétiques biométriques. 45(2) :167–174.

[Lauvergne et al., 1988] Lauvergne, J. J., Djorbineva, M. K., and Aleksieva, S. A. (1988). Hereditary determinism of form and attitude of the external ear of sheep. A review. *Small Ruminant Research*, 1(3) :237–248.

[Luquetti et al., 2012] Luquetti, D. V., Heike, C. L., Hing, A. V., Cunningham, M. L., and Cox, T. C. (2012). Microtia : Epidemiology and genetics. *American Journal of Medical Genetics Part A*, 158A(1) :124–139.

[Meyer et al., 2004] Meyer, C., Faye, B., and Karembe, H. (2004). *Guide de l'élevage du mouton méditerranéen et tropical*. CEVA Santé animale.

[Munroe et al., 2009] Munroe, R. J., Prabhu, V., Acland, G. M., Johnson, K. R., Harris, B. S., O'Brien, T. P., Welsh, I. C., Noden, D. M., and Schimenti, J. C. (2009). Mouse H6 Homeobox 1 (Hmx1) mutations cause cranial abnormalities and reduced body mass. *BMC Developmental Biology*, 9(1) :27.

[Mustefa et al., 2020] Mustefa, A., Belayhun, T., Melak, A., Hayelom, M., Tadesse, D., Hailu, A., and Assefa, A. (2020). Phenotypic characterization of Raya cattle in northern Ethiopia. *Tropical Animal Health and Production*, 53(1) :48.

[Mwacharo et al., 2006] Mwacharo, J. M., Okeyo, A. M., Kamande, G. K., and Rege, J. E. O. (2006). The small East African shorthorn zebu cows in Kenya. I : Linear body measurements. *Tropical Animal Health and Production*, 38(1) :65–74.

[Pauly and Ferrandis, 2018] Pauly, M. and Ferrandis, M. (2018). Le site funéraire d'Antsiraka Boira (Acoua, Grande Terre) : Islamisation et syncrétisme culturel à Mayotte au XIIe siècle. *Afriques. Débats, méthodes et terrains d'histoire*.

[Pourlis, 2011] Pourlis, A. F. (2011). A review of morphological characteristics relating to the production and reproduction of fat-tailed sheep breeds. *Tropical Animal Health and Production*, 43(7) :1267–1287.

[Saido, 2005] Saido (2005). Rapport National sur l'Etat des Ressources Génétiques Animales - Comores. Technical report, FAO.

[Salako, 2006] Salako, A. E. (2006). Application of Morphological Indices in the Assessment of Type and Function in Sheep. *International Journal of Morphology*, 24(1).

[Schorderet et al., 2008] Schorderet, D. F., Nichini, O., Boisset, G., Polok, B., Tiab, L., Mayeur, H., Raji, B., de la Houssaye, G., Abitbol, M. M., and Munier, F. L. (2008). Mutation in the Human Homeobox Gene NKX5-3 Causes an Oculo-Auricular Syndrome. *The American Journal of Human Genetics*, 82(5) :1178–1184.

[Sejian et al., 2017] Sejian, V., Hyder, I., Maurya, V., Bagath, M., Krishnan, G., Aleena, J., Archana, P., Lees, A. M., Kumar, D., Bhatta, R., and others (2017). Adaptive mechanisms of sheep to climate change. In *Sheep Production Adapting to Climate Change*, pages 117–147. Springer.

[Sileshi et al., 2003] Sileshi, Z., Tegegne, A., Tsadik, G. T., and others (2003). Water resources for livestock in Ethiopia : Implications for research and development. *Integrated water and land management research and capacity building priorities for Ethiopia*, 66.

[Udo and Budisatria, 2011] Udo, H. M. J. and Budisatria, I. G. S. (2011). Fat-tailed sheep in Indonesia ; an essential resource for smallholders. *Tropical Animal Health and Production*, 43(7) :1411–1418.

[Wilson, 1991] Wilson, R. T. (1991). *Small ruminant production and the small ruminant genetic resource in tropical Africa*. Number 88 in FAO animal production and health paper. Food and Agriculture Organization of the United Nations, Rome.

[Wright et al., 1984] Wright, H. T., Sinopoli, C., Wojnaroski, L., Hoffman, E. S., Scott, S. L., Redding, R. W., and Goodman, S. M. (1984). Early Seafarers of the Comoro Islands : the Dembeni Phase of the IXth-Xth Centuries AD. *Azania : Archaeological Research in Africa*, 19(1) :13–59.

4.4 Discussion du chapitre

Les deux articles présentés dans ce chapitre sont les résultats d'un projet européen de caractérisation des ressources génétiques des ruminants élevés sur l'île de Mayotte. Plusieurs études de caractérisation phénotypique proposent de tester l'effet de la localisation sur les variables recueillies [Lauvergne and Zafindrajaona, 1992, Mwacharo et al., 2006, Mustefa et al., 2020]. Dans ce chapitre, la localisation n'a pas été prise en compte comme un facteur expliquant la variabilité observée compte tenu de l'étroitesse du territoire de Mayotte facilitant les échanges [Kim et al., 2018] et diluant l'effet du climat sur les caractéristiques du phénotype. De plus, compte-tenu de la structure foncière en construction des exploitations agricoles de l'île de Mayotte, il n'est pas rare que des éleveurs de bovins déplacent leur troupeau.

En ce qui concerne les troupeaux ovins, tous sont conservés dans des enclos fixes sans possibilité de brouter. Les ressources nutritives sont donc dépendantes du système d'élevage et si un effet localisation est détecté, il s'agit probablement d'un effet "élevage". Néanmoins, le système d'élevage n'a été décrit qu'avec peu de variables (nombre d'individus totals, type de production, type de reproduction, croisement ou non) lors de la campagne d'échantillonnage des individus et ne pourra pas nous permettre de caractériser l'effet élevage expliquant la variation phénotypique observée dans les populations étudiées.

Les observations ont été réalisées sur de courtes périodes, l'idée étant au départ du projet d'effectuer une caractérisation primaire des populations. Ce type d'étude ne permet pas de déterminer avec précision la réponse physiologique des individus au stress thermique, aux variations des conditions climatiques tropicales. Le rôle et la composition de la queue grasse chez le mouton a été largement étudié [Pourlis, 2011, Sejian et al., 2017] mais ce fut moins le cas pour la bosse chez le zébu. Notre étude a permis de déterminer une variation significative de la taille de la bosse en fonction de la note d'état corporel. Néanmoins, ces résultats ne permettent pas de conclure sur le mécanisme de stockage préférentiel et de remobilisation des réserves corporelles dans la bosse du zébu.

Chapitre 5

Caractérisation des histoires démographiques et adaptatives des populations bovines locales de l'Ouest de l'Océan Indien

L'étape de caractérisation phénotypique du zébu de Mayotte (chapitre 4, section 4.2) a mis en évidence un profil phénotypique diversifié basé sur la description des paramètres de coloration et de morphologie (bosse, cornes notamment) et marqué par un dimorphisme sexuel. Afin de caractériser génétiquement cette population, nous avons étudié sa diversité génétique en analysant les données de génotypage d'un échantillon d'animaux, représentatif de la diversité des systèmes d'élevage concernés par un schéma de reproduction impliquant les zébus mahorais. Ces animaux ont été choisis sur la base de leur phénotype, de leur localisation en minimisant l'apparentement. Cette caractérisation génétique a été réalisée dans le cadre du projet DEFI-ANIMAL financé par l'Europe, l'Etat français et le département de Mayotte. De plus, nous avons entrepris d'étudier l'histoire démographique et adaptative de cette population en comparaison avec le zébu de Madagascar.

5.1 Synthèse des résultats

Sur les 400 individus phénotypés, 150 ont été génotypés sur une puce Illumina BovineSNP50 (54K) et parmi ces individus, 32 ont été génotypés sur une puce Illumina BovineSNPHD (777K), puis 8 ont été séquencés. Les données HD ont été analysées dans l'article (paragraphe 5.2). Les résultats de l'analyse exploratoire issus des données 54K sont présentés au paragraphe 5.3.

L'analyse des données de génotypage SNP à haute densité a permis de préciser l'histoire démographique et adaptative des Zébus des îles de Mayotte (MAY) et de Madagascar (ZMA). Ces deux populations sont caractérisées par une étroite proximité génétique (avec une faible ascendance taurine africaine et une ascendance indicine prédominante) ainsi qu'une histoire démographique commune avant le 18^{éme} siècle, en accord avec les récentes découvertes du schéma complexe de la dynamique des populations humaines dans la région. Les deux populations résultent d'un premier événement de métissage impliquant des populations sources proches des taurins africains et des populations indicines (proches des zébus d'Afrique de l'Est), suivi d'un second métissage avec une source proche des zébus indiens, remontant au 12^{éme} siècle, à l'époque des premiers contacts entre les populations humaines africaines du Corridor Swahili et les populations austronésiennes d'Asie du Sud-Est aux Comores et à Madagascar. L'estimation du déséquilibre de liaison a permis de détecter une augmentation de la taille efficace *Ne* de MAY et ZMA du début du 16^{éme} siècle jusqu'au 17^{éme} siècle suggérant ainsi un métissage significatif des populations bovines avec une homogénéisation, concomitant avec les échanges importants entre les Comores, Madagascar et la côte Est africaine durant cette période. Les distributions des segments chromosomiques ROH sur plusieurs classes dépendantes de la longueur des segments sont comparables dans les deux populations avec des niveaux élevés de consanguinité récente provenant principalement du 13^{éme} au 17^{éme} siècle. La divergence entre les populations MAY et ZMA a été datée au 16^{éme} siècle, concomitante à l'arrivée des Européens et la modification du réseau commercial dans la région. L'analyse des signatures de sélection adaptative au sein des populations (score iHS) et entre elles (scores

Rsb) a permis d'identifier des 27 gènes candidats pour MAY et 47 pour ZMA qui sont pour la plupart différents (seulement 8 en commun) mais impliqués dans les mêmes fonctions biologiques principales (cancer, structure dermatologique et protection contre les UV, système nerveux et comportement, développement des organes, métabolisme et réponse immunitaire) largement représentatives de l'adaptation physiologique aux conditions tropicales.

5.2 Article 3 : The genetic history of Mayotte and Madagascar cattle breeds mirrors the complex pattern of human exchanges in Western Indian Ocean

(soumis)

The genetic history of Mayotte and Madagascar cattle breeds mirrors the complex pattern of human exchanges in Western Indian Ocean

Jessica Magnier[1,2], Tom Druet[3], Michel Naves[4], Melissa Ouvrard[5], Solene Raoul[5], Jérôme Janelle[1,6], Katayoun Moazami-Goudarzi[7], Matthieu Lesnoff[1,2], Emmanuel Tillard[1,6], Mathieu Gautier[8] and Laurence Flori[9,*]

[1]SELMET, Univ. Montpellier, CIRAD, INRAE, L'Institut Agro, 34398 Montpellier, France
[2]CIRAD, UMR SELMET, 34398 Montpellier, France
[3]Unit of Animal Genomics, GIGA-R, Faculty of Veterinary Medicine, University of Liège, 4000 Liège, Belgium
[4]URZ, INRAE, 97170 Petit-Bourg, Guadeloupe, France
[5]COOPADEM, 97670 Ouangani, Mayotte, France
[6]CIRAD, UMR SELMET, 97410 Saint-Pierre, Réunion, France
[7]University Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France
[8]CBGP, INRAE, CIRAD, IRD, L'Institut Agro, Univ. Montpellier, 34988 Montferrier sur Lez, France
[9]SELMET, INRAE, CIRAD, L'Institut Agro, Univ. Montpellier, 34398 Montpellier, France

*Corresponding author: laurence.flori@inrae.fr

Abstract

Despite their central economic and cultural role, the origin of cattle populations living in Indian Ocean islands still remains poorly documented. Here, we unravel the demographic and adaptive histories of the extant Zebus from the Mayotte and Madagascar islands using high-density SNP genotyping data. We found that these populations are very closely related and both display a predominant indicine ancestry. They diverged in the 16th century at the arrival of European people who transformed the trade network in the area. Their common ancestral cattle population originates from an admixture between an admixed African zebu population and an Indian zebu that occurred around the 12th century at the time of the earliest contacts between human African populations of the Swahili corridor and Austronesian people from Southeast Asia in Comoros and Madagascar. A steep increase of the estimated population sizes from the beginning of the 16th to the 17th century coincides with the expansion of the cattle trade. By carrying out genome scans for recent selection in the two cattle populations from Mayotte and Madagascar, we identified sets of candidate genes involved in biological functions (cancer, skin structure and UV-protection, nervous system and behavior, organ development, metabolism and immune response) broadly representative of the physiological adaptation to tropical conditions. Overall, the origin of the cattle populations from Western Indian Ocean islands mirrors the complex history of human migrations and trade in this area.

Keywords: cattle; genomics; demography; adaptation; Indian Ocean

Introduction

The Indian Ocean has played a prominent role in the human-mediated migration of cattle populations between East-Africa, Middle-East and South-West Asia. However, the origin and genetic diversity of cattle populations living in Indian Ocean islands remain poorly investigated and relatively unclear. Their understanding may provide insights into the recent history of human populations in this area that has long displayed important levels of interaction, trade, migration and domestic species translocation over time (Boivin *et al.* 2013; Beaujard 2019a,b).

Within the large Indian Ocean area, the Comoro Islands (including Ngazidja, Ndzuwani, Mwali and Mayotte) and Madagascar occupy a key position in the maritime trade routes that has linked the East-African coast, Middle East and Asia over the past two millenia. Interestingly, cattle may have been introduced early in this area and have subsequently represented an important domestic species. In both Mayotte and Madagascar where human have permanently settled from the 6th century CE, the most ancient archaeological and skeletal evidence of cattle presence traces back to the 9th-10th centuries CE but the quantities of identified cattle bones only increased from 14th - 15th centuries CE (Pauly 2013; Boivin *et al.* 2013). The first Portuguese eyewitnesses reported the presence of cattle on the Western Indian Ocean islands from the 16th century like almost all 17th and 18th century visitors in Comoros (Cheke 2010).

Nowadays, among the Comoros, the cattle population of Mayotte is the best characterized thanks to a recent effort that lead to the official recognition of a local humped breed named 'Zebu of Mayotte' (MAY) by the French government for conservation purpose (France 2011; Ouvrard *et al.* 2018). Overall this breed, traditionally and extensively raised in small herds of a handful head, represents about 70% of the 20,000 identified individuals identified in the island. Other individuals consist of recently imported individuals from European taurine (EUT) breeds (i.e., Montbeliarde, Jersey, French Brown Swiss, Gasconne) and MAYxEUT admixed individuals. In contrast, nine millions of cattle heads are identified in Madagascar (Ministère de l'Agriculture 2007), including about 85% of Zebu of Madagascar (ZMA) raised in large herds, imported taurine breeds (e.g. Holstein, Norvegian Red), several synthetic admixed breeds with ZMA (such as Renitelo, Manjan'i Boina and Rana) and Baria, a small wild and free-roaming population (Porter 2007, www.fao.org). Both MAY and ZMA populations are considered to be well adapted to their respective islands conditions. In particular, they are generally regarded as more resistant to heat stress and tick-borne diseases than imported European breeds, such as African zebu and N'Dama breeds (Hansen 2004; Mattioli *et al.* 1995; Bock *et al.* 1999; Glass *et al.* 2005).

If the MAY zebu population has not been genetically characterized yet, previous genetic characterization of ZMA populations

based on the analysis of a few tens of microsatellite markers or medium density SNP genotyping data suggested an hybrid composition between African taurine (AFT) and indicine (ZEB) ancestries, the latter being predominant (Zafindrajaona and Lauvergne 1993; Hanotte 2002; Gautier *et al.* 2009). However, these analyses remained mostly descriptive and only provide limited insights into the origins of these populations. In particular, from the history of human migration routes, cattle may have been introduced possibly repeatedly with subsequent exchanges between the 8th and the 13th centuries in the Comoro and Madagascar islands from several places including East Africa (as a result of movements of Bantu populations) or Indonesia (with Austronesian navigator populations) (Fuller and Boivin 2009; Fuller *et al.* 2011; Beaujard 2005, 2007, 2011, 2015).

The purpose of our study was to clarify the origin of the local Mayotte and Madagascar cattle breeds to better understand the demography of Western Indian Ocean cattle breeds based on their refined genetic characterization. To that end, we genotyped 32 MAY and 24 ZMA individuals on the bovineHD high density SNP genotyping assay (comprising >770,000 SNPs) together with 113 individuals belonging to the Somba (SOM, n=44) and the Lagune (LAG, n=44) West-African taurine breeds and to the Fulani West-African Zebu (ZFU, n=25). These newly generated data were combined to publicly available bovineHD genotyping data for 363 individuals belonging to 8 other breeds representative of EUT (Angus, Holstein, Jersey and Limousine), another African taurine (N'Dama), Indian Zebus (Gir and Nelore) and East African Zebus (Bahbahani *et al.* 2017). We first provide a refined analysis of the structuring of genetic diversity among the combined data sets and carried out a detailed inference of the demographic history of the MAY and ZMA populations with respect to the other extant populations by constructing admixture graphs (Patterson *et al.* 2012; Lipson 2020; Gautier *et al.* 2021) and estimating the timing of some admixture events using Linkage-Disequilibrium (LD) information (Loh *et al.* 2013). We further estimated their recent changes in effective population sizes using the recently developed method GONE (Santiago *et al.* 2020) and characterized the levels of genomic inbreeding in MAY and ZMA (Druet and Gautier 2017; Bertrand *et al.* 2019; Druet and Gautier 2021). We finally investigated the patterns of genetic adaptation of MAY and ZMA cattle breeds which recently diverged and live in slightly different tropical island conditions by searching for footprints of positive selection. The identified signals were subjected to a detailed functional analysis to identify putative physiological pathways and their possible underlying selective pressures (see e.g., Flori *et al.* 2019).

Materials and methods

Animal sampling and genotyping

For MAY individuals (Zebus from Mayotte), we selected a group of 32 presumably non-related individuals (based on the newly created French National Registration Database) representative of the phenotypic diversity of the local population. They each originated from 32 different farms located in 17 townships spread over the Grande and Petite terres of the Mayotte Island (Figure S1, Table S1). Blood samples were collected during year 2016 from the tail vein of the individuals using 10 ml EDTA vacutainer tubes, strictly following the recommendations of the directive 2010/63/EU for animal care. Genomic DNA was further extracted using the Wizard Genomic DNA purification kit (Promega, France) and stored at -20°C. MAY DNA samples were genotyped (Table S1) on the Illumina BovineHD genotyping beadchip at Labogena plateform (Jouy-en-Josas) using standard procedures (www.illumina.com) together with 24 Zebus of Madagascar (ZMA) and 113 West-African cattle DNA samples (i.e. 44 Lagune, 44 Somba and 25 Zebus Fulani), sampled in the 1990's and previously genetically characterized on the Illumina BovineSNP50 beadchip (Gautier *et al.* 2009). The genotyping data were then added to the WIDDE database and combined with other publicly available ones, using WIDDE utilities (Sempéré *et al.* 2015).

To obtain the position of the 777,962 SNPs included in the BovineHD genotyping assay onto the latest ARS-UCD1.2 (aka bosTau9) bovine genome assembly (Rosen *et al.* 2020), we first retrieved the sequences of each SNP (length ranging from 61 to 121 bp with only 227 sequences <121 bp long) from the Illumina manifest file `bovinehd_15013478_b.csv`. These sequences were further realigned onto the ARS-UCD1.2 assembly using `pblat` (Wang and Kong 2019) run with options `-out=pslx -minIdentity=98 -threads=4` and the resulting alignment file was parsed with a custom `awk` script. Unambiguous positions (i.e., with a single alignment hit of the underlying sequence) could then be obtained for 721,583 SNPs (92.8%). Among the 53,750 remaining SNPs, the number of alignment hits ranged from 2 to 1790 and followed a L-shaped distribution with a median and mean values of 3 and 20.2 respectively. As a matter expedience, we thus chose to discard from further analysis the 15,533 SNPs with more than 9 alignment hits (i.e., the third quartile of the aforementioned distribution) and to assign to the 40,846 other SNPs (with 2 to 9 hits) the position given by the alignment with the highest score. We indeed noticed that for these latter SNPs, multiple alignments mostly consisted of a single high scoring hit (score$>$ 100 for 99% of them), the other alternative hits being generated by partial alignment of only a few tens bases of the SNP sequence.

The complete genotyping dataset consisted of 532 animals including 363 individuals from 8 other bovine populations (Bahbahani *et al.* 2017, www.illumina.com) representative of the bovine worldwide genetic diversity (Table 1). The minimal individual genotyping call rate was set to 90% and the minimal SNP genotyping call rate to 90% over all populations and 75% within each population (i.e. SNPs genotyped for less than 75% of the animals from at least one population were discarded). SNPs with a MAF$<$0.01 or departing from Hardy-Weinberg equilibrium expectation (exact test P-value$<$ 10^{-3} in at least one breed) were also discarded. A total of 680,338 SNPs distributed throughout the 29 autosomes of the bosTau9 bovine genome assembly passed finally all our filtering criteria.

Inference of the population demographic history.

Characterization of the structuring of genetic diversity Principal Component Analysis (PCA) based on individual SNP genotyping data was performed with *smartpca* (Patterson *et al.* 2006) and visualized with the R package `ggplot2` (Wickham 2016). Unsupervised genotype-based hierarchical clustering of the individual animal samples was carried out using the maximum-likelihood method implemented in ADMIXTURE 1.06 (Alexander *et al.* 2009). Results were visualized with custom functions in R environment (www.r-project.org). Finally, the overall and pairwise-population F_{ST} were computed with version 2.0 of the R package `poolfstat` (Gautier *et al.* 2021) using the `computeFST` ran with default settings (i.e., `method=Anova`) and option `nsnp.per.bjack.block=5000` to estimate standard errors (and 95% CI as $\pm$1.96 s.e.) with block-jackknife.

f-statistics based demographic inference. f-statistics based demographic inference (Patterson *et al.* 2012) were carried out with the

Table 1 Sample description and origin of the Illumina BovineHD chip genotyping data (* before filtering).

Code	Population name	Country	Sampling Year	Nb	Reference
ANG	Angus	Scotland	n.a.	42	Illumina (Sempéré *et al.* 2015)
EAZ	East African Shorthorn Zebu	Kenya	n.a.	92	Bahbahani *et al.* (2017)
GIR	Gir	India	n.a.	27	Illumina (Sempéré *et al.* 2015)
HOL	Holstein	Netherlands	n.a.	60	Illumina (Sempéré *et al.* 2015)
JER	Jersey	Jersey	n.a.	38	Illumina (Sempéré *et al.* 2015)
LAG	Lagune	Benin	1996	44	*This study*
LMS	Limousine	France (mainland)	n.a.	50	Illumina (Sempéré *et al.* 2015)
MAY	Zebu from Mayotte	France (Mayotte)	2016	30 (32)*	*This study*
NDA	N'Dama	Guinea	n.a.	23	Illumina (Sempéré *et al.* 2015)
NEL	Nelore	India	n.a.	31	Illumina (Sempéré *et al.* 2015)
SOM	Somba	Togo	1996	44	*This study*
ZFU	Zebu Fulani	Benin	1996	24 (25)*	*This study*
ZMA	Zebu from Madagascar	Madagascar	1991	23 (24)*	*This study*

new functionalities of the version 2.0 of the R package `poolfstat` (Gautier *et al.* 2021). We used the `compute.fstats` function to estimate the different f-statistics (including F_3 and F_4 for all the population triplets and quadruplets respectively) and within-population heterozygosities. As for F_{ST}, standard-errors of the estimated statistics (and their corresponding Z-scores for f_3 and f_4) were estimated using block-jackknife defining blocks of 5,000 consecutive SNPs (i.e., option `nsnp.per.bjack.block=5000`). In addition, to mitigate SNP ascertainment bias by favoring SNPs of most remote ancestry (e.g., discard SNPs of exclusive European ancestry), we only kept for f-statistics based demographic inference the 497,949 polymorphic SNPs that were polymorphic (MAF>0.05) in both ZEB (GIR and NEL combined) and in AFT (NDA, SOM and LAG combined) populations. Following Patterson *et al.* (2012) (see also Lipson 2020), we then carried out formal tests of population admixture using the estimated f_3 statistics. A negative Z-score ($Z < -1.65$ at the 95% significance threshold) associated to an f_3 for a given population triplet A;B,C showing that the target population A is admixed between two source populations each related to B and C. To further provide insights into the origins of MAY and ZMA, we build an admixture graph with `poolfstat` utilities (Gautier *et al.* 2021). Briefly, we first build a scaffold tree of presumably unadmixed populations (as suggested by both exploratory analyses and f_3 and f_4 based tests) consisting of two AFT breeds (NDA and LAG), two ZEB breeds (GIR and NEL) and one EUT breed (HOL) with the `rooted.njtree.builder` function. We further relied on the *graph.builder* function (ran with default options) to jointly include EAZ, MAY and ZMA that showed clear evidence for admixture on the graph considering all the six possible orders of inclusion. The fit of the best fitting graph (displaying a BIC more than 8 units lower than all the other graphs explored in the graph building process) was further validated with the *compare.fitted.fstats* function that allows to compare to which extent the estimated f-statistics depart from their predicted values based on the fitted admixture graph parameters via a Z-score (Patterson *et al.* 2012; Lipson 2020; Gautier *et al.* 2021).

Estimation ot the timing of admixture events. We estimated the timing of admixture events (in generations) with the program `mALDER` (Pickrell *et al.* 2014) that implements a modified version of `ALDER` method originally described by Loh *et al.* (2013). This approach relies on the modelling of the exponential decay of admixture-induced LD in a target admixed population (here based on two-reference weighted LD curves, i.e., using a LD measure weighted by allele frequencies in two source population proxies) as a function of genetic distance. Genetic distances between pairs of SNPs were derived from physical distances assuming a cM to Mb ratio of 1 (Kadri *et al.* 2016). As commonly done in studies focusing on the population genetics of local cattle breeds, we here assumed a six-year generation time to convert the timing from generations to years (e.g., Gautier *et al.* 2007; Flori *et al.* 2014; Mbole-Kariuki *et al.* 2014; Bahbahani *et al.* 2015, 2017). Such a long-term (i.e., across several generations) average generation time is in agreement with estimation by Keightley and Eyre-Walker (2000) for auroch and is also consistent with early estimation by Mahadevan (1955) (cited in Mbole-Kariuki *et al.* 2014) in the Red Sindhi East-African breed. We also obtained a very similar estimate of 5.6 years ($\pm$0.3) by analyzing genotyping data publicly available for the synthetic Santa-Gertrudis breed (Matukumalli *et al.* 2009; Sempéré *et al.* 2015) which was established in 1910 in South-Texas by crossing taurine Shorthorns and Brahman zebus (Felius 2016). More precisely, to obtain this estimate of long-term generation time, we simply divided the presumed number of years separating birth of the genotyped samples from breed formation (ca. 95 years) by the timing of admixture estimated by `ALDER` (equal to 17.0 ± 0.99).

Inference of the recent population size histories. Historical effective population sizes (N_e) were inferred for the MAY and ZMA breeds with the program `GONE` that implements an approach recently developed by Santiago *et al.* (2020) to fit the observed spectrum of LD of pairs of loci over a wide range of recombination rates (which we derived from physical map distances assuming a cM to Mb ratio equal to 1, see above). In practice, we adopted a block-jackknife approach to estimate confidence intervals for

the inferred N_e trajectories by first identifying 55 non-overlapping blocks of 10,000 consecutive SNPs (block size ranging from 32.8 to 40.2 Mb) out of the 680,338 genotyped ones. We then analyzed 55 different data sets of 670,338 SNPs that were each formed by removing from the original data sets one block of 10,000 SNPs. The resulting 55 inferred N_e trajectories were then summarized by computing a mean trajectory and a 95% confidence envelope defined by the 2.5 and 97.5 percentiles from the 55 estimated N_e at each time point.

Age-based partitioning of individual inbreeding in MAY and ZMA breeds. The characterization of individual levels of inbreeding at a global and local scale in MAY and ZMA breeds was performed with the model-based approach implemented in the RZooRoH package (Bertrand *et al.* 2019; Druet and Gautier 2017, 2021). This method models each individual genomes as a mosaic of Homozygous-by-Descent (HBD) and non-HBD segments using a multiple HBD-classes Hidden Markov Model (Druet and Gautier 2017). Each HBD class is specified by a rate R_k related to the expected length (equal to $\frac{1}{R_k}$ Morgans) of the associated HBD segments and that is approximately equal to twice the number of generations to the common ancestor that transmitted the DNA segment. Given the density of the HD chip, we considered a model with 11 HBD classes (with $R_k = 2^k$ for $k = 1, \ldots, 11$) and one non-HBD class (with rate equal to R_{11}) allowing to capture the contribution to the overall individual inbreeding levels from each age-based classes of ancestors (living up to 1,024 generations in the past).

Whole-genome scan for footprints of selection

Computation of iHS and Rsb statistics. The genome-wide scan for footprints of positive selection within and between MAY and ZMA breeds was performed using extended haplotype homozygosity (EHH)-based tests. To that end we first jointly phased the genotyping data for MAY, ZMA, EAZ, GIR and NEL individuals with version 1.4 of the `fastPHASE` software (Scheet and Stephens 2006) for each chromosome in turn using population label information and options `-Ku40 -Kl10 -Ki10`. Genetic distances between consecutive SNPs were obtained from physical distances assuming a cM to Mb ratio equal to 1 (see above). Based on haplotype information, we further computed the *iHS* (Voight *et al.* 2006) and *Rsb* (Tang *et al.* 2007) statistics using the version 3.1.2 of the R package `rehh` (Gautier *et al.* 2017; Klassmann and Gautier 2020). These statistics are designed to detect regions with high level of haplotype homozygosity over an unexpected long distance (relative to neutral expectations), either within population (*iHS*) or across populations (*Rsb*). Note that for the computation of *iHS* within MAY and ZMA populations, we chose to not polarize the SNP alleles (i.e., `scan_hh` function was run with option `polarized=FALSE`) as discussed in section 7.6 of the online `rehh` package vignette (https://cran.r-project.org/web/packages/rehh/vignettes/rehh.html). The different SNP *iHS* and *Rsb* statistics were further transformed into $p_{iHS} = -\log_{10}(1 - 2 \mid \Phi(iHS) - 0.5 \mid)$ and $p_{Rsb} = -\log_{10}(\Phi(Rsb))$ where $\Phi(x)$ represents the Gaussian cumulative function. Assuming *iHS* and *Rsb* scores are normally distributed under neutrality, p_{iHS} and p_{Rsb} might thus be a two-sided $-\log_{10}$(P-value) associated with the neutral hypothesis.

Identification of candidate genes. We used the approach previously described in Flori *et al.* (2019) to identify candidate genes under positive selection from *iHS* and *Rsb* estimates. Briefly, all the genotyped SNPs were annotated using as a gene set reference a list of 14,562 RefSeqGenes anchored on the Btau9 bovine genome assembly (`refGene.txt.gz, 2019-06-07`; https://hgdownload.soe.ucsc.edu/goldenPath/bosTau9/database/). A SNP was then considered as representative of one of these RefSeq genes if localized within its boundary positions extended by 15 kb upstream and downstream to account for persistence of LD (e.g., Gautier *et al.* 2007). Among the 680,338 SNP of HD-dataset participating to the analysis, 293,527 SNPs mapped to 13,696 different RefSeq Genes, corresponding to 13,536 gene symbols. On average, each SNP mapped within 1.2 RefSeq genes (from 1 to 34, median= 1) and each RefSeq gene was represented by 21 SNPs (from 1 to 641, median=13). For a given statistic, genes with at least 2 representative SNPs with $-\log_{10}$(P-value) > 4 were considered as candidate genes.

Functional annotation of candidate genes under selection. Following the approach outlined in Flori *et al.* (2019), candidate genes were functionally annotated and submitted to gene network analysis using Ingenuity Pathway Analysis software (IPA, QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis). Among the 13,535 genes symbols taken into account in the analysis, 12,324 were mapped in the Ingenuity Pathway Knowledge Base (IPKB) and were considered as the reference data set. Among the candidate genes identified for MAY (N=27) and ZMA (N=47) breeds, 24 and 45 were respectively mapped to IPKB. The top significant functions and diseases (P-value< 0.05) were obtained by comparing functions associated with the candidate genes under selection against functions associated with all genes in the reference set, using the right-tailed Fisher exact test. In the network analysis, a score S was computed for each network that contained at most 140 molecules (including candidate genes under selection) based on a right-tailed Fisher exact test for the over-representation of candidate genes under selection ($S = -\log_{10}$(P-value)). A network was considered as significant when $S > 3$. The top significant functions and diseases associated with significant networks were also reported.

Results

Genetic relationships of Mayotte and Madagascar cattle breeds with other world cattle breeds

We first carried out an individual-based PCA of the SNP genotyping dataset consisting of 680,338 autosomal SNPs genotyped on individuals from MAY (n=32) and ZMA (n=23) populations together with 473 individuals belonging to 11 cattle breeds representative of the bovine genetic diversity (Table 1). The first factorial plan of the PCA is represented in Figure 1a. The first two components accounting for 21.09% and 7.85% of the total variation respectively. In agreement with previous studies (e.g., Gautier *et al.* 2010), the first two PCs revealed a clear structuring of individual genetic diversity according to their population of origins and highlighted a triangle-like 2-dimensional global structure of the cattle populations with apexes corresponding to three main groups: i) European taurines (EUT) represented by ANG, HOL, JER and LMS; ii) African taurines (AFT) represented by LAG, NDA and SOM; and iii) zebus (ZEB) represented by NEL and GIR. Similarly, as previously observed with data from medium density SNP genotyping data (on the same sample), ZMA and West-African hybrids (ZFU and EAZ) were located at intermediate positions on the AFT/ZEB segment, the ZMA individuals being closer to the ZEB apex (Gautier *et al.* 2010). Finally, the newly characterized MAY individuals appeared mostly confounded with ZMA individuals which is consistent with their close geographical proximity. Unsupervised hierarchical clustering analysis of the individuals

using K=3 predefined clusters provided a complementary view in close agreement with PCA results while allowing to provide first estimates of ancestry proportions (Figure 1b). Roughly interpreting the blue, red and green cluster in Figure 1b as representative of EUT, ZEB and AFT ancestries respectively suggested that ZMA and MAY individuals had similarly high levels of ZEB ancestry (0.83 ± 0.02 and 0.82 ± 0.01 on average, respectively), a weaker level of AFT ancestry (0.16 ± 0.01), and almost no detectable EUT ancestry except for three individuals (2 MAY and 1 ZMA) with more than 5% of EUT ancestry, which were discarded from further analyses.

The different analyses performed at an individual scale showed that the partitioning of cattle into distinct populations (and breeds) is relevant to study their genetic history at the population level (e.g., Gautier *et al.* 2010). Accordingly, the overall F_{ST} among the 13 populations was found equal to 0.271 (95% CI, $[0.262; 0.279]$). Population pairwise F_{ST} ranged from 2.40% ($\pm$0.01%) for the ZMA/MAY pair to 51.1% ($\pm$0.06%) for the LAG/NEL pair (Figure S2) confirming the close relatedness of this two populations. Among the 11 other populations, EAZ was found the most closely related to both MAY and ZMA ($F_{ST}^{may,eaz} = 7.69\%$ and $F_{ST}^{zma,eaz} = 8.07$) followed by ZFU and the two ZEB populations (GIR and NEL).

Inferring the history of Mayotte and Madagascar cattle populations

f_3-based tests shows clear evidence for admixture in the history of both MAY and ZMA. Out of the 66 ($= 12$) possible f_3-based tests when considering either MAY or ZMA as a target population, six were significantly negative (Z-score< -2.33 at the 99% significance threshold) providing strong evidence for admixture events in their history. They all involved the same pair of source proxies that consisted of one AFT population (LAG, NDA or SOM) and one ZEB population (NEL or GIR), the most significant signal being observed with the LAG/NEL pair for ZMA (f_3 (ZMA;LAG,NEL) associated Z-score equal to -3.92) and the NDA/NEL pair for MAY (f_3 (MAY;NDA,NEL) associated Z-score equal to -6.38). Among the 11 other populations, EAZ, ZFU and SOM also showed clear evidence for admixture with 14, 14 and 6 f_3-based tests with a significantly negative Z-score (at the 99% significance threshold), respectively. For each of these three target populations, the lowest Z-score was always associated with the LAG/NEL pair of source proxies (f_3 (EAZ;LAG,NEL), f_3 (ZFU;LAG,NEL) and f_3 (SOM;LAG,NEL) associated Z-scores being equal to -36.5, -33.6 and -9.13 respectively).

admixture graph construction. To infer the history of MAY and ZMA ancestry, we constructed an admixture graph relating these two populations with two out of the three AFT breeds (LAG and NDA, discarding SOM as it displayed strong signal of presumably recent admixture with ZEB ancestry, see above), the two ZEB breeds (GIR and NEL), one (HOL) out of four EUT breeds and the EAZ African hybrid. We did not seek to include ZFU as it was more remotely related to East-African populations with a recent origin not directly informative for the MAY and ZMA history (Flori *et al.* 2014). The inferred best fitting admixture graph is represented in Figure 2. It shows that MAY and ZMA breeds both derived from an admixed population we referred to Indian Ocean Zebu in Figure 2, although it should be noticed that we have no clear evidence about the actual geographic origin of this ancestral population. This Indian Ocean Zebu had an admixed origin with a predominant ancestry (88.6%) from an admixed Zebu, likely of African origin since the most closely related to extant EAZ, and the remaining ancestry (11.4%) originating from a presumably Indian Zebu related to extant GIR. The former African Zebu was itself admixed with a Zebu related to extant NEL and a Taurine related to African taurine ancestries (in 68.4% vs. 31.4% respective proportions). Note that these inferred origins lead to overall ZEB (resp. AFT) related ancestries for MAY and ZMA similar to the one obtained in Figure 1.

timing of admixture events. We estimated the timing of the two aforementioned admixture event t_α^{afz} (leading to African Zebus, closely related to EAZ) and t_α^{ioz} (leading to Indian Ocean Zebus, the direct ancestor of MAY and ZMA) with the program `mALDER` Pickrell *et al.* (2014). Based on the inferred admixture graph (Figure 2), we estimated t_α^{afz} by using EAZ as a population target and a pair of source population proxies consisting of one AFT (either NDA or LAG) and one ZEB (either GIR or NEL). In agreement with admixture graph fitting, the highest amplitude (i.e., y-intercept) estimate of the fitted weighted LD curve was obtained with NDA and NEL as source proxies which suggests that these populations are the closest (among the sampled ones) to the actual source population (Loh *et al.* 2013). The corresponding estimated timing for the admixture events leading to African Zebus was found equal to $t_\alpha^{afz} = 174 \pm 19$ generations (i.e., $1,044 \pm 114$ years). This admixture events thus traces back to the 10th century CE (ca. year 950$\pm$150). We similarly estimated t_α^{ioz} by considering either MAY or ZMA as a population target and a pair of source population proxies consisting of EAZ and one ZEB (either GIR or NEL). For both target populations, the amplitude was the highest (although only slightly) with GIR as the ZEB represent. The estimated timing for the admixture events leading to Indian Ocean Zebus was found equal to $t_\alpha^{ioz} = 153 \pm 6$ and $t_\alpha^{ioz} = 152 \pm 9$ generations with MAY and ZMA as a target population respectively. The (slightly) lower latter estimate being consistent with the difference in sampling time (ca. 4 generations) between MAY and ZMA samples. This second admixture events thus traces back to the early 12th century CE (ca. year 1,100$\pm$40).

recent population size history of MAY and ZMA. Figure 3 plots the recent effective population size histories of MAY and ZMA as inferred with `GONE` (Santiago *et al.* 2020). The two trajectories were found similar before the beginning of the 16th century (i.e., ca. 80 generations before MAY sampling time) when the two populations likely started to diverge. From that time onward, the two populations experienced a rapid growth in a few generations toward a mostly constant size (approximately twice higher for ZMA and MAY) followed by a new moderate increase from the beginning of the 20th century. As expected from the large differences in census cattle population sizes between Mayotte and Madagascar islands, the current estimated N_e was about twice lower in MAY compared to ZMA. More precisely, the harmonic mean of the (mean) historical N_e from generation 80 (76 for ZMA) onward was found equal to 2,160 and 1,045 for ZMA and MAY respectively. Surprisingly, a close inspection of the inferred admixed graph in Figure 2 seems to contradict these large differences in N_e since the estimated (leave) branch lengths was about twice higher for ZMA compared to MAY. This thus rather suggested increased drift in ZMA since these estimates (on a diffusion timescale) might be interpreted under a pure-drift model of divergence as inversely proportional to N_e (i.e., equal to $\frac{t}{2N_e}$ where t is the divergence time in generation). Population-specific F_{ST} estimates for MAY and ZMA essentially confirmed this trend (1.59% and 3.15% respec-

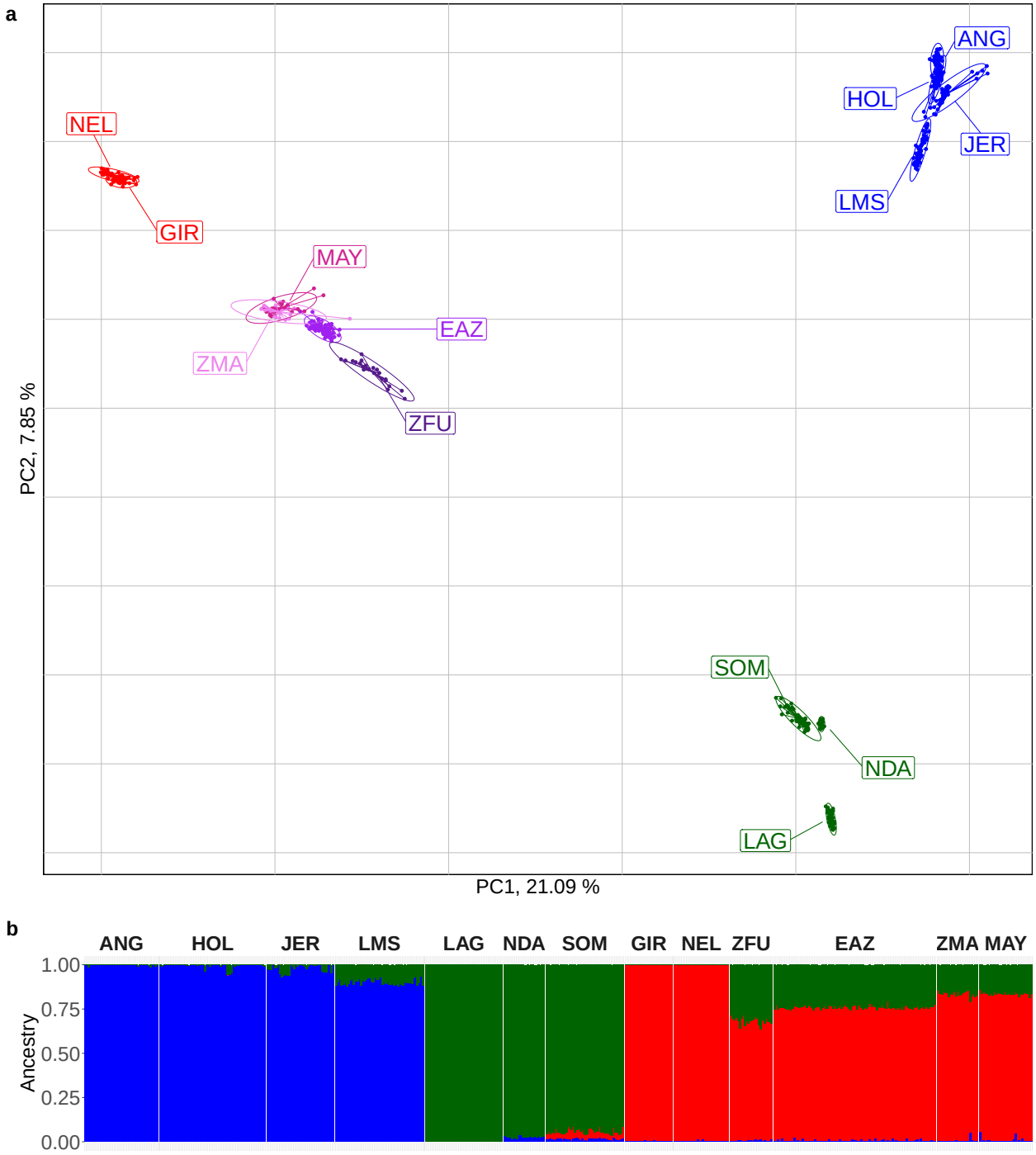


Figure 1 Results of the Principal Component Analysis (PCA) and unsupervised hierarchical clustering including HD genotyping data (528 individuals from 13 populations genotyped for 680,338 SNPs). **a** PCA results. The individuals are plotted on the first two principal components according to their coordinates. Ellipses characterize the dispersion of each population around its centre of gravity. MAY and ZMA individuals are plotted in dark-pink and pink and EAZ and ZFU individuals in purple and dark purple, respectively. EUT, AFT and ZEB individuals are plotted in blue, green and red, respectively. **b** Unsupervised hierarchical clustering results with K=3 predefined clusters. For each individual, the proportions of each cluster (y axis) which were interpreted as representative of EUT, AFT and ZEB ancestries are plotted in blue, green and red, respectively.

tively, Figure S2). Conversely, when considering only the 518,315 SNPs with a MAF>1% in MAY and ZMA (computed over MAY and ZMA combined individuals), the estimated heterozygosity was slightly higher in MAY (33.2%) than in ZMA (32.7%). As detailed in Figure S3, these apparent discrepancies may actually be explained by some residual asymmetric gene flow between

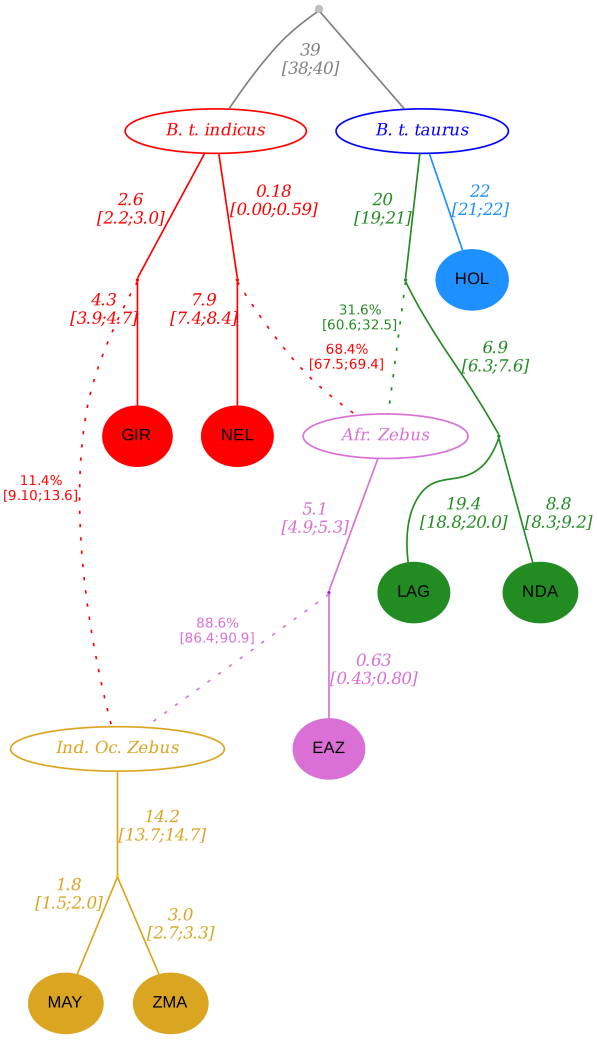


Figure 2 Inferred admixture graph connecting cattle breeds from Mayotte and Madagascar (MAY and ZMA, in yellow) with two Indian indicine breeds (GIR and NEL in red), one African zebu breed (EAZ in purple), two African taurine breeds (LAG and NDA in green) and one European taurine breed (HOL in blue). Admixture events are shown by dotted arrows. Estimates of branch lenghts (x10^3, in drift units of $\frac{t}{2Ne}$) and admixture rates are indicated next to the corresponding edges.The Z-score of the worst fitted *f*-statistics *f4*(MAY,ZMA;HOL,LAG) is equal to -1.04.

MAY and ZMA (or from a common gene pool related to some other continental population, e.g., from the East-African coast) after these two population split. More specifically, very similar level of differentiation and heterozygosities among MAY and ZMA could be obtained with data simulated with `msprime` (Kelleher *et al.* 2016) under a simplified scenario assuming MAY and ZMA split 80 generations ago and maintained a constant N_e (equal to the one estimated above) with MAY receiving twice as many migrants as ZMA from a common gene pool ($m_{\rightarrow MAY} = 4 \times 10^{-3}$ vs $m_{\rightarrow ZMA} = 2 \times 10^{-3}$) (Figure S3).

MAY and ZMA individuals levels of inbreeding. In order to characterize individual levels of recent inbreeding at both a global (genome-wide) and local (locus-specific) scale, we estimated the individual inbreeding coefficients in MAY and ZMA populations by applying the method implemented in RZooRoH software (Druet and Gautier 2017; Bertrand *et al.* 2019). The distribution of the genome-wide coefficients of individual inbreeding estimated with the R package `RZooRoH` are detailed in Figure 4a for the MAY and ZMA populations together with their age-based partitioning to assess the contribution of various class of ancestors (Figures 4b, 4c, Figure S4). Overall, the average individual inbreeding levels were found similar and equal to 0.308 ± 0.06 and 0.3136 ± 0.03 in MAY and ZMA respectively, MAY and ZMA individuals all displaying at least 16% and 25% of their genome in autozygous (HBD) segments. However, in agreement with the inferred trajectories of historical N_e (Figure 3, most HBD segments were concentrated in ancestor age classes associated with R_k equal to 128 and 256 (i.e. approximately 64 and 128 generations before sampling thus spanning the split time period between MAY and ZMA). Hence, most of the estimated inbreeding level originated from the foundation of the two populations. Consistently, only a few individuals displayed a high proportion of HBD segments associated with the most recent common ancestors, two generations ago. More precisely, two MAY individuals showed more than 49% of their genome in HBD classes and two ZMA individuals showed more than 40% of their genome classified has HBD.

Identification of selection footprints in the MAY and ZMA genomes

The genome of ZMA and MAY individuals was scanned for footprints of selection with the package *rehh* 3.1.2 (Gautier *et al.* 2017), following the approach described in Flori *et al.* (2019) to identify directly candidate genes. According to our criteria (see Material and Methods), the computation of the *iHS* statistic for each SNP over the whole MAY and ZMA genomes detected two (i.e. GRIK2 and RBM47) and 12 (i.e. AASS, ACTR6, ATP10D, COL15A1, DPP6, DPYSL5, EXOC4, ITGA7, MAPRE3, SLC17A8, TNFSF11, ZCCHC7) genes under selection located on two (i.e. BTA6 and 9) and six (i.e. BTA4, 5, 6, 8, 11 and 12) chromosomes, respectively (see details in Table S2; Figures S5 and S6), with no overlap between these two lists of genes. By computing different *Rsb* statistics comparing the local extend of haplotype homozygosities between MAY and ZMA, and between these two breeds and each of their different main ancestries (i.e. AFZ and ZEB), 3, 16, 12, 18, 11 and 11 candidate genes were identified with the $Rsb_{MAY/ZMA}$, $Rsb_{MAY/EAZ}$, $Rsb_{MAY/ZEB}$, $Rsb_{ZMA/MAY}$, $Rsb_{ZMA/EAZ}$ and $Rsb_{ZMA/ZEB}$ tests, respectively (Table S2, Figures S5 and S6). Most of the candidate genes identified by the *Rsb* involving MAY or ZMA populations were different (74 and 85%, respectively). However seven candidates genes (i.e. GRIK2, HMGSC2, LPCAT2, MIR2285BM, PHGDH, TRIM10 and TRIM15) were identified under selection in both breeds by some of these tests. Specifically, GRIK2, located on BTA9 was found under selection by five tests (iHS_{MAY}, $Rsb_{MAY/EAZ}$, $Rsb_{MAY/ZEB}$ and $Rsb_{ZMA/EAZ}$ and $Rsb_{ZMA/ZEB}$ tests), HMGSC2 and PHGDH located on BTA3 and LPCAT2 located on BTA18 by $Rsb_{MAY/ZEB}$ and $Rsb_{ZMA/ZEB}$ tests, MIR2285BM located on BTA16 by $Rsb_{MAY/ZEB}$ and $Rsb_{ZMA/EAZ}$ tests and TRIM10 and 15 located on BTA23 by $Rsb_{MAY/EAZ}$ and $Rsb_{ZMA/EAZ}$ tests. $Rsb_{MAY/ZMA}$ test detected three (i.e. LRRK2, GSTCD and GPC5) and 18 genes (i.e. ATP2C1, SCN2A, STPG1, KCND2, ANKS1B, NR1H4, MIR2434, ANO4, SPIC, KCTD8, TG, LRRC6, CACNA1E, KCNH1, SLC4A7, ZSCAN31, PGBD1, ZSCAN26) under selection in MAY and in ZMA, respectively. Considering each breed separately, four candidate genes (i.e. GBE1, GPC5, SLC6A2 and ZMAT4) were detected

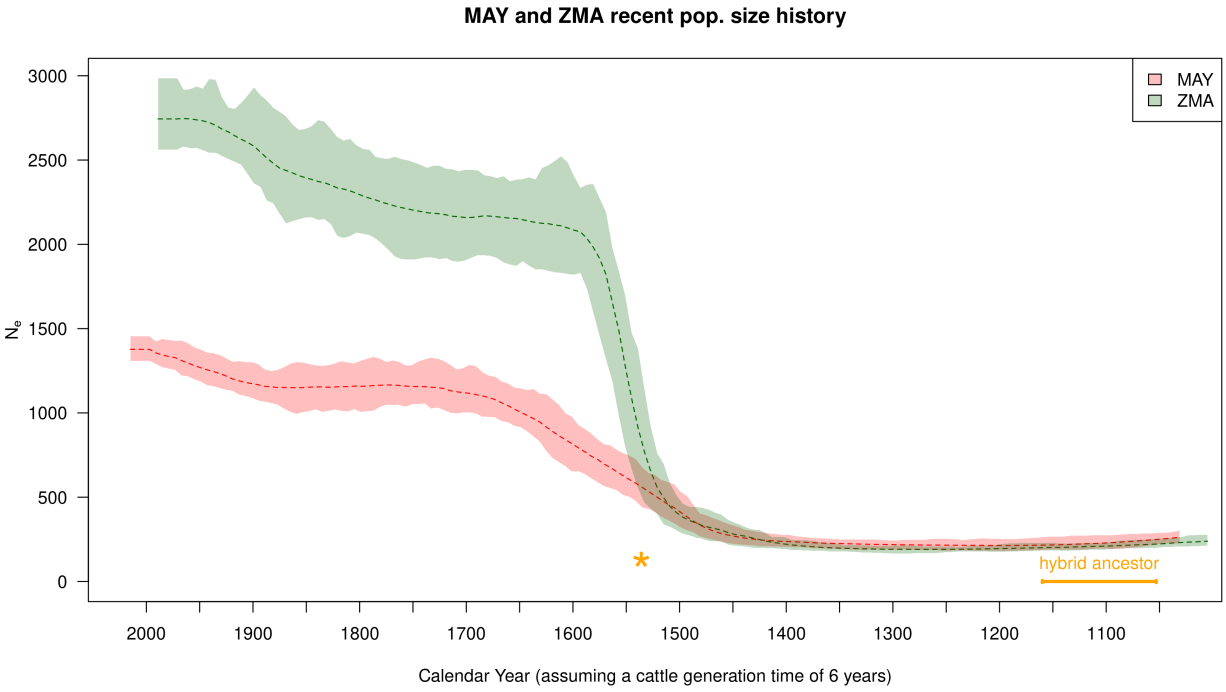


Figure 3 Population size history (N_e) of MAY and ZMA populations estimated with GONE (Santiago *et al.* 2020). The average N_e trajectories (dashed line) and 95% confidence envelope estimated from block-jackknife samples are plotted in green for ZMA and red for MAY. The time scale was transformed into calendar assuming a 6-year generation time for cattle and accounting for the difference in sampling time between ZMA (1990) and MAY (2016). The estimated timing of the admixture event (t_α^{ioz}) that led to the common hybrid ancestor (named Indian Ocean Zebus in the main text) of MAY and ZMA is given in orange. The orange asterisk gives the likely splitting time (beginning of the 16th century, i.e., $t_S \simeq 80$ generations before MAY sampling time) between MAY and ZMA roughly estimated from the separation of the two trajectories.

in MAY population and four (i.e. CPS1, KCND2, KCTD8 and KCNH1) in ZMA population by two *Rsb* tests.

Overall, a total of 27 and 47 candidate genes under selection were detected in at least one of the above tests for MAY and ZMA, respectively. To obtain a global view of the main gene functions under selection, these two lists of candidate genes were annotated using IPA software. The main functional categories, in which these genes are involved, are listed in Table S3 (see Tables S4 and S5 for an exhaustive list of their functional annotation). For MAY, the top five significant functions belonging to the three IPA main categories (i.e. Diseases and Disorders, Molecular and Cellular Functions and Physiological System Development and Function) were related to inflammation ('Inflammatory Disease'), cancer and cell death ('Cancer' and 'Cell Death and Survival'), gastrointestinal and hepatic diseases ('Hepatic Disease','Gastrointestinal disease'), nervous system ('Nervous System Development and Function'), organ and tissue development and morphology ('Organismal Injury and Abnormalities', 'Cell Morphology', 'Cellular development', 'Embryonic Development', 'Organ development', 'Organ morphology' and 'Organismal development') and inflammation ('Inflammatory Disease'). For ZMA, the top five significant functions were related to cancer, dermatology ('Dermatological Diseases and Conditions'), organ and embryonic development ('Organismal Injury and Abnormalities', 'Organismal Development', 'Embryonic Development', 'Organ Development'), metabolism and gastrointestinal disease ('Molecular Transport', 'Lipid Metabolism', 'Small Molecule Biochemistry', 'Gastrointestinal Disease'), respiratory ('Respiratory System Development and Function') and nervous ('Nervous System Development and Function') systems (Table S3). Interestingly, many functional categories or subcategories (9 among the top five significant functions) were found in both populations such as functions related to cancer, nervous system, organ development and gastrointestinal disease.

Moreover, candidate genes identified in each breed participated to one significant network (Table S6) which included 85% and 93% of candidates genes for MAY (score=54) and ZMA (score=108) breeds, respectively (Figures 5a and 5b). Ten molecules are in common between both networks (i.e. AR, GRIK2, HMGCS2, LPCAT2, MAP3K10, MAP3K11, PHGDH, TRIM10, TRIM15 and kynurenic acid) among which six were detected as candidate genes under selection (i.e. GRIK2, HMGCS2, LPCAT2, PHGDH, TRIM10, TRIM15). Candidate genes were located in periphery or sub-periphery of both networks and no main node were detected under selection, except TNFSR11, associated with a significant *iHS* for ZMA breed. The functional annotation of each network gave a more complete picture of the functions involved in adaptation of MAY and ZMA breeds (Tables S7, S8 and S9). Molecules integrated in the MAY network play a role in carbohydrate metabolism, development, nervous and respiratory systems function, dermatology, inflammation and behavior. The ZMA network in mainly involved in cancer, in organismal abnormalities, cell signaling growth and proliferation, reproductive system, dermatology, gastrointestinal disease, endocrine and nervous systems and inflammation.

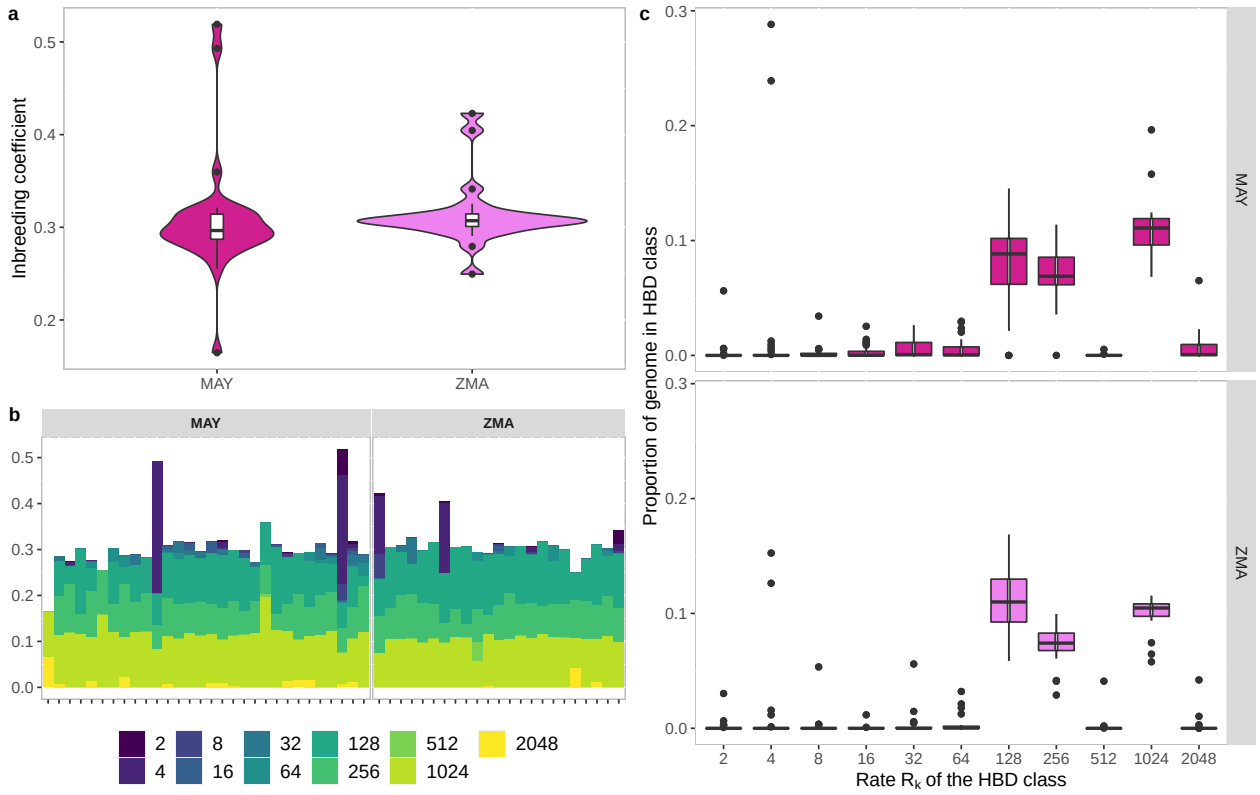


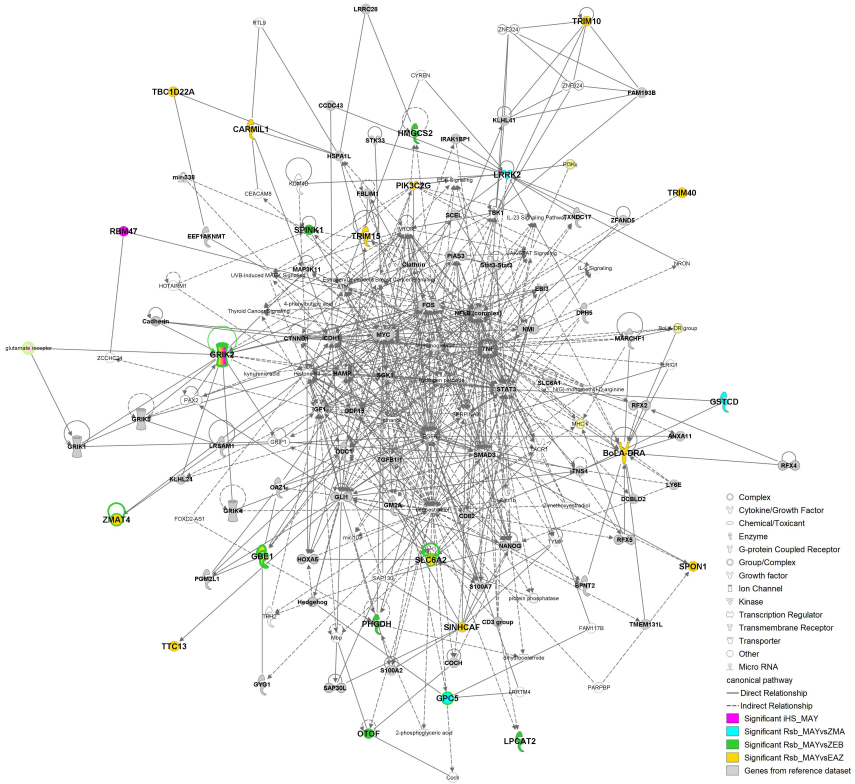
Figure 4 Characterization of individual inbreeding levels in Mayotte and Madagascar breeds. **a** Violin plots representing the distribution of inbreeding coefficients for the MAY (N=30) and ZMA (N=23) breeds, coloured in pink and purple, respectively. **b** Partitioning of individual genomes in different HBD classes. Each bar represent an individual and its total height the overall level of inbreeding. The height of the different stacks, which appear in different colors, represents the proportion of the genome associated with each HBD class, defined by their rate R_k. **c** Average proportion of individual genomes associated with different HBD classes for MAY (pink) and ZMA (purple) populations. Individual proportions of the genome associated with a specific HBD class are obtained by averaging the corresponding HBD-classes probabilities over all marker positions.

Discussion

The genetic characterization of Zebus from Mayotte and Madagascar (MAY and ZMA), two geographically closely related island cattle populations from the Western Indian Ocean, showed that they share a similar recent and rather complex demographic history. In particular, we detected in these two populations a weak African taurine ancestry and a predominant indicine ancestry. The indicine ancestry in MAY and ZMA was even higher than in the East-African zebu breed, the African continental breed with the highest estimated indicine ancestry proportion in our dataset. This pattern had already been reported for ZMA in some previous studies (Hanotte 2002; Gautier *et al.* 2009) but still remained difficult to explain from an historical point of view. Our demographic inference, in the form of admixture graph construction, demonstrated that the higher indicine ancestry in the MAY and ZMA populations, as compared to other African continental zebus, could actually be explained by a second pulse of Indian Zebu introgression, limited to the island populations, into an already admixed African taurine x Zebu population of likely East-African origin that traces back to the 12th century. These results are in agreement with the recent findings in human population dynamics in the area (Brucato *et al.* 2017, 2018) based on genetic data analysis, which report that Comorian and Malagasy communities resulted from admixture between Swahili people from East-Africa and individuals from islands of Southeast Asian (Austronesian) dating from the 8th-12th centuries (Brucato *et al.* 2018). Migrations of Austronesian populations was actually promoted by the significant development of sea borne trade, from the Arabian Peninsula and Indonesia to East Africa and Madagascar during this period (Choudhury *et al.* 2018; Brucato *et al.* 2018; Bahbahani *et al.* 2017; Brucato *et al.* 2016). These excellent navigators should have transported indicine cattle (and probably other livestock species) towards Western Indian Ocean that must have been crossed with AFZ individuals brought from African East coast by Swahili people.

Estimation of past effective population sizes for both MAY and ZMA populations showed similar trend and allowed clarifying the timing of their recent split to the 16th century which coincides with the arrival of Europeans in the area and the modification of the trade network (Newitt 1983; McPherson 1984; Beaujard 2005). Some residual cattle migration may have still remained after this split with either a differential contribution of an unknown continental population to MAY and ZMA or some weak and asymmetric gene flow between them, further discriminating between these two scenarios would require additional samples from the area that may be difficult to collect (e.g., ancient DNA). Looking forward in time, after the split of MAY and ZMA, we estimated a steep increase in their ancestral populations sizes starting from the beginning of the 16th century to the 17th century, which was

(a)



(b)

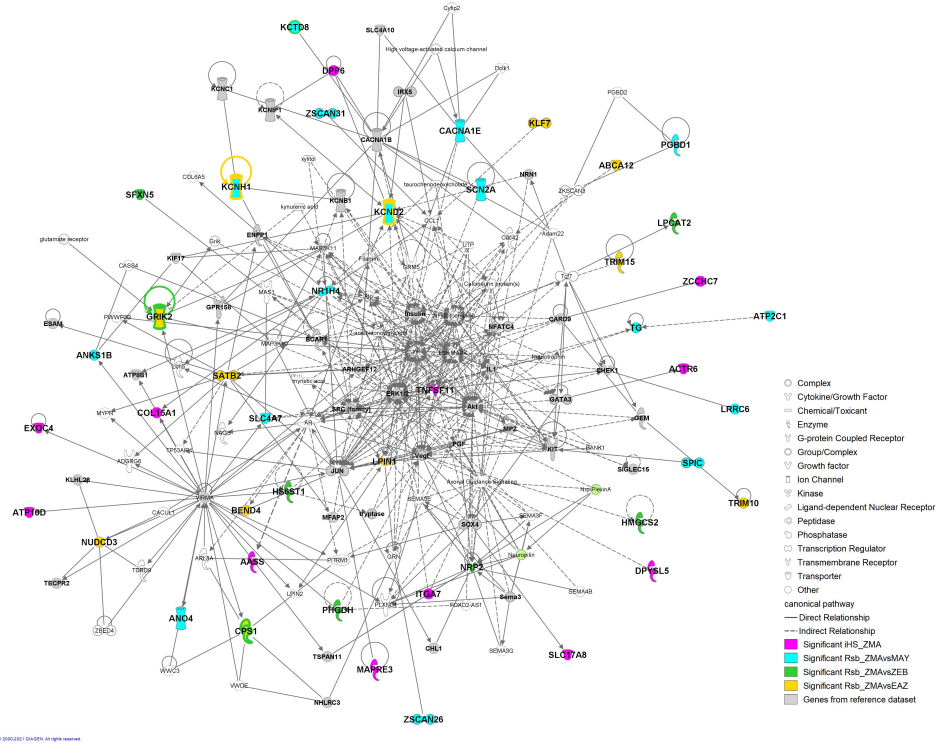


Figure 5 Gene networks including candidate genes under selection detected in at least one *iHS* or *Rsb* test in the genome of breeds from Mayotte (**a**) and Madagascar (**b**)

more pronounced for ZMA probably due to the far larger size of Madagascar. This observed cattle population expansion may be related to the increase of the livestock production by islands communities, in particular to meet the growing demand of European ships in fresh meat (Newitt 1983). Indeed, as the Portuguese retained control of the East African coast, the other Europeans (Dutch, English and French), who began to enter Indian Ocean, needed revictualling ports and bases that Madagascar and the Comoros islands could offer. Cattle then represented one of the main trading resources that contributed to the economic development in this area during the early 17th century. From the middle of the 17th century, the population sizes of both MAY and ZMA tended to stabilize suggesting a maintenance of cattle production until the beginning of the 20th century, when both population sizes started to increase again but at a more moderate rate. This period actually coincides with French colonization of both islands that possibly lead to a new development in the use of local bovine resources. As expected, the characterization and partitioning of inbreeding levels into age-related HBD classes for MAY and ZMA individuals was found consistent with the inferred recent history of their corresponding population N_e. More specifically, in both populations, the contribution to individual inbreeding levels of HBD classes of most recent origin (with rates $R_k < 64$, i.e., tracing back to ancestors living less than 32 generations ago) remained limited, except for a few individuals that may result from sporadic unintentional consanguineous matings. This is thus quite encouraging from a breed conservation perspective. Conversely, the highest contributing HBD classes pointed to ancestors living between 64 and 128 generations (i.e., HBD classes with rates R_k =128 and R_k =256). Assuming a 6 year generation time (see above), this corresponds to the middle of the 13th century and the middle of the 17th century. Accordingly, very similar profile of the most remote HBD classes contribution (with $R_k \geq$128) are observed for both MAY and ZMA individuals since these pertain to their common ancestral population. It should also be noticed, that considering a smaller (i.e., five years) or larger (i.e., seven years) generation time did not fundamentally challenge our interpretation of the demographic history of MAY and ZMA with respect to either the effective population size histories or timing of admixture events (Figure S7).

From the 12th century (i.e., ca. 150 generations ago) onward, cattle living in Comoros and Madagascar islands have likely experienced various environmental and artificial selection pressures. To identify their footprints in the MAY and ZMA breeds, we relied on EHH-based tests considering different contrasts that our detailed inference of the demographic history helped interpreting. First, it should be noticed that only a few candidate genes were found in common in the two populations when comparing the various population-specific signals. These include i) two genes (TRIM10 and TRIM15) with significant $Rsb_{MAY/EAZ}$ and $Rsb_{ZMA/EAZ}$signals; and ii) five genes (HMGSC2, PHGDH, GRIK2, MIR2285BM, LPCAT2) with significant $Rsb_{MAY/ZEB}$ and $Rsb_{ZMA/ZEB}$ signals. Note that among the latter, GRIK2 also displayed a significant *iHS* signal in the MAY breed. These different common footprints may be considered of older origin in response to environmental (e.g., tropical climate) and human-driven constraints imposed to the common ancestral population of MAY and ZMA breeds (i.e., predating their divergence). Among these genes, TRIM10 and TRIM15 are involved in innate immune response and were already detected under selection in Muturu cattle in Africa (Tijjani *et al.* 2019). In addition, missense mutations (with respect to the Taurine assembly) within TRIM10 were found fixed in Nelore suggesting a putative role in tropical adaptation (Júnior *et al.* 2020). HMGCS2, PHGDH and LPCAT2 are related to lipid or serine metabolism (Hegardt 1999; Vilà-Brau *et al.* 2011; Reid *et al.* 2018; Morimoto *et al.* 2014), the first two being reported as associated with some fatty acid in intramuscular fat of Nelore cattle (Cesar *et al.* 2014). GRIK2 that participates to the metabolism of glutamate, an important brain neurotransmitter (Purves *et al.* 2001), is involved in tropical cattle behavior. For instance, deficiency of this gene was found associated with a significant reduction in anxiety, fear memory and the flight time temperament phenotype in tropical cattle with varying amount of indicine ancestry (Ko *et al.* 2005; Porto-Neto *et al.* 2014).

Conversely, significant signals with $Rsb_{MAY/ZMA}$ based tests only, provided insights into footprints of selection of recent origin postdating the divergence of the two populations. These signals might be related to local differences between the tropical climate of Mayotte and Madagascar. In the small Mayotte island, the climate is homogeneous with a hot, humid and rainy season during the northeastern monsoon and a cooler dry season. In contrast, Madagascar island whose surface area is 15 times greater than that of Mayotte, presents a wider variety of climates being humid tropical on the East coast, dry tropical on the West coast and temperate on the Highlands. In addition, the differential evolution of husbandry practices in the two islands after the divergence of the two populations may also have contributed to leave different selection footprints. For instance, in Mayotte, Zebus have been reared in the recent past in traditional extensive systems with small herds (a few heads) with regular close contacts with breeders. Besides, in Madagascar, the sampled individuals originated from an area where breeding practices are based on larger herds (from 4 to few hundred heads) often dispersed on large pasture territory (Zafindrajaona 1991). A higher number of ZMA driven than MAY driven signals (n=18 vs. n=3) were found with this test, as expected from the twice highest N_e of ZMA over the period which may result in improved detection power. Among the three genes that display significant signals in the MAY breed (LRRK2, GSTCD and GPC5), GSTCD is involved in lung function and was identified as CNVR-harbored gene associated with adaptation to hypoxia in yak population (Wang *et al.* 2019). GPC5 was associated with feed efficiency in cattle (Buitenhuis *et al.* 2014; Serão *et al.* 2013). The 18 genes with significant $Rsb_{MAY/ZMA}$ signals driven by ZMA breed included ANOA4, which is located in a region under selection in New World Creole cattle and in Korean Brindle Hanwoo cattle (Gautier and Naves 2011; Kim *et al.* 2018). This gene belongs to the anoctamin protein family, whose members are expressed in sweat glands and are involved in thermal sweating, one of the mechanisms of heat tolerance (Cui and Schlessinger 2015; Jian *et al.* 2014). Within this gene list, KCTD8, encoding a potassium channel, was found associated with the Temperature Humidity Index in Mediterranean cattle breeds and also identified as a heat stress responsive gene in chicken (Flori *et al.* 2019; Sun *et al.* 2015). It is also associated with milk fat percentage in buffaloes (De Camargo *et al.* 2015) and located within selection signatures detected in dairy cattle breeds (Hayes *et al.* 2008; Flori *et al.* 2009). Potassium channels are also related to the production of prolactin, a key hormone for mammary development, milk production, hair development, thermoregulation and water balance during heat stress (Underwood and Suttle 1999; Silanikove *et al.* 2000; Czarnecki *et al.* 2003; Littlejohn *et al.* 2014). KCNH1 represents another remarkable gene since it is associated with a syndrome characterized by dysmorphism and hypertrichosis in human (Kortüm *et al.* 2015) and was associated with several climatic variables in local Mediterranean

cattle breeds (Flori *et al.* 2019). Finally, the thyroglobulin gene (TG), encodes a protein required for thyroid hormone synthesis and iodine storage and contains variants associated with carcass and meat quality traits in beef cattle (Hou *et al.* 2011).

The timing of the footprints underlying the other 18 and 26 candidate genes found specific to MAY (iHS_{MAY},$Rsb_{MAY/EAZ}$, $Rsb_{MAY/ZEB}$) or to ZMA (iHS_{ZMA},$Rsb_{ZMA/EAZ}$, $Rsb_{ZMA/ZEB}$) breeds is more difficult to interpret. For instance, they may either result from recent breed-specific selective constraints or older selective constraints further relaxed in one breed. In both cases, the fact that these genes were not identified with $Rsb_{MAY/ZMA}$ could be due to the overall limited power of the different tests (see above). Yet, some noticeable genes showed up in the list of these breed-specific candidate genes. For instance, among the genes specifically detected in MAY, SPINK1 promotes the progression of various types of cancer and is involved in abnormal morphology of thin skin (Lin 2021). PIK3C2G, under selection in Normande and Holstein breeds (Flori *et al.* 2009), is involved in the regulation of insulin-mediated activation of the PI3K/Akt pathway in the liver that controls metabolism (Braccini *et al.* 2015), PI3K and Akt corresponding to central nodes of the MAY network (Figure 5a). TRIM40, is involved in innate immune response and was already detected under selection in Muturu cattle in Africa (Tijjani *et al.* 2019) and BOLA-DRA, which encodes the alpha subunit of the class II Major Histocompatibility Complex, expressed at the Antigen-Presenting Cells surface, which presents peptides from foreign origin to the immune system. Among the genes specific to ZMA breed, DDP6 was associated with feed efficiency (de Camargo *et al.* 2015; Buitenhuis *et al.* 2014; Serão *et al.* 2013), COL15A1 with iris hypopigmentation in Holstein Friesan cattle (Hollmann *et al.* 2017) with a possible involvement in UV-protection and NUDCD3, with milk yield and components in Swedish cattle breeds (Ghoreishifar *et al.* 2020), this latter being under positive selection in dairy cattle breeds and across Holstein, Angus, Charolais, Brahman, and N'Dama breeds (Flori *et al.* 2009; Xu *et al.* 2015).

Overall, the functional annotation of all the candidate genes identified with at least one test and of their inferred underlying networks provided some insights into the main functions involved in adaptation of cattle to Western Indian Ocean tropical conditions and breeding practices. Some genes (e.g. SPINK1 in MAY and ANO4, KCTD8, COL15A1 and KCNH1 in ZMA) are related to cancer and skin properties and may have played a role in adaptation to Mayotte and Madagascar climates by increasing tolerance to thermal stress, humidity and UV exposition. Respiratory function in which some candidate genes are involved (e.g. GSTCD in MAY) may also be related to tropical environment adaptation, breeds from warm climates presenting lower respiration rates under heat stress to maintain their body temperature. Several genes carrying footprints of selection are involved in nervous system and cattle behavior (e.g. GRIK2 in MAY and ZMA) which indicates also a physiological adaptation to the livestock system and the proximity with breeders in both islands. Development of different organs or physiological structures, gastrointestinal and hepatic system function, metabolism and hormone production are also among the main functions targeted by selection in the two breeds (e.g. HMGCS2, PHGDH, MIR2285BM and LPCAT2, detected in MAY and ZMA breeds, GPC5 detected in MAY breed and NUDCD3, TG and DDP6 detected in ZMA breed). If these signatures of selection could be related to artificial selection on conformation and production traits, they could also be linked to the adaptation to tropical climate (e.g. through thermoregulation) and to available food and water resources. At last, several candidate genes involved in inflammatory and immune response (e.g. TRIM10, TRIM15, TRIM40, BOLA-DRA) play a role in adaptation of Mayotte and Madagascar breeds suggesting that tropical pathogens responsible for e.g. as East Coast Fever, Rift Valley Fever or blackleg documented outbreaks, may have had a deep impact on the adaptive genetic diversity of MAY and ZMA breeds (De Deken *et al.* 2007; Porter *et al.* 2016; Cêtre-Sossah *et al.* 2012; Dommergues *et al.* 2015).

This study displays for the first time a genetic characterization of the local cattle breed of Mayotte and presents a deep analysis of demographic and adaptive histories of cattle breeds from Mayotte and Madagascar islands. The genetic proximity between these two populations reflects their closely tied demographic history until the 16th century before the population divergence. Their demographic history mirrors the complex pattern of human migrations and trade in Western Indian Ocean islands. Their adaptive history has been probably overall conditioned by the same selective pressures and also by some differences between climates and breeding practices in the two islands. This study highlights the great originality of Zebus from Madagascar and Mayotte compared to the cattle populations from Africa mainland and their relevance in the context of global changes, which promotes specific measures of conservation especially for Mayotte breed.

Data availability

SNP genotyping data are available in the WIDDE database (http://widde.toulouse.inra.fr/widde/) and in the portal Data INRAE (https://data.inrae.fr/; https://doi.org/10.15454/IQL6GE).

Acknowledgments

We would like to thank Sitty Bahyat Chamassi et Abdou Achiraffi (Laboratoire Vétérinaire d'Analyses Départemental de Mayotte, Kaweni, Mayotte), Catherine Cêtre-Sossah (UMR ASTRE, CIRAD, Sainte-Clotilde, La Réunion) and Anlidine Mkadara, Hidachi Attoumani, Oussoufi Saindou, Kamardine Moussa and Adjibou Grimaldine (Chambre de l'Agriculture, de la Pêche et de l'Aquaculture de Mayotte, Coconi, Mayotte) for their technical assistance and all the breeders of the Mayotte cattle breed who participated to this study. Tom Druet is Senior Research Associate from the F.R.S.-FNRS.

Funding

This work was supported by the EU (FORWARD RITA DEFI-ANIMAL project, 2015-2020) and INRAE, France (SELTRYP project, AIP-Bioressources, 2010 and PERSAFRICA project, Animal Genetics Division, 2013).

Conflicts of interest

Authors declare that there is no conflict of interest.

Literature cited

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research. 19:1655–1664.

Bahbahani H, Clifford H, Wragg D, Mbole-Kariuki MN, Tassell CV, Sonstegard T, Woolhouse M, Hanotte O. 2015. Signatures of positive selection in East African Shorthorn Zebu: A genome-wide single nucleotide polymorphism analysis. Scientific Reports. 5:1–13.

Bahbahani H, Tijjani A, Mukasa C, Wragg D, Almathen F, Nash O, Akpa GN, Mbole-Kariuki M, Malla S, Woolhouse M *et al*. 2017. Signatures of Selection for Environmental Adaptation and Zebu x Taurine Hybrid Fitness in East African Shorthorn Zebu. Frontiers in Genetics. 8.

Beaujard P. 2005. The Indian Ocean in Eurasian and African world-systems before the sixteenth century. Journal of World History. 16:411–+. WOS:000235458400002.

Beaujard P. 2007. East Africa, the Comoros Islands and Madagascar before the sixteenth century: On a neglected part of the world system. Azania: Archaeological Research in Africa. 42:15–35.

Beaujard P. 2011. The first migrants to Madagascar and their introduction of plants: linguistic and ethnological evidence. Azania-Archaeological Research in Africa. 46:169–189. WOS:000293472700004.

Beaujard P. 2015. East Africa and oceanic exchange networks between the first and the fifteenth centuries. Afriques-Debats Methodes Et Terrains D Histoire. 6. WOS:000420334300001.

Beaujard P. 2019a. *The Worlds of the Indian Ocean: A Global History: Volume 1: From the Fourth Millennium BCE to the Sixth Century CE*. volume 1. Cambridge University Press. Cambridge.

Beaujard P. 2019b. *The worlds of the Indian Ocean: A global history: Volume 2 : From the Seventh Century to the Fifteenth Century CE*. volume 2. Cambridge University Press. Cambridge, United Kingdom ; New York, NY.

Bertrand AR, Kadri NK, Flori L, Gautier M, Druet T. 2019. Rzooroh: an r package to characterize individual genomic autozygosity and identify homozygous-by-descent segments. Methods in Ecology and Evolution. 10:860–866.

Bock R, Kingston T, De Vos A. 1999. Effect of breed of cattle on transmission rate and innate resistance to infection with babesia bovis and b bigemina transmitted by boophilus microplus.

Boivin N, Crowther A, Helm R, Fuller DQ. 2013. East Africa and Madagascar in the Indian Ocean world. Journal of World Prehistory. 26:213–281.

Braccini L, Ciraolo E, Campa CC, Perino A, Longo DL, Tibolla G, Pregnolato M, Cao Y, Tassone B, Damilano F *et al*. 2015. PI3K-C2Y is a Rab5 effector selectively controlling endosomal Akt2 activation downstream of insulin signalling. Nature Communications. 6:1–15.

Brucato N, Fernandes V, Mazières S, Kusuma P, Cox MP, Ng'ang'a JW, Omar M, Simeone-Senelle MC, Frassati C, Alshamali F *et al*. 2018. The Comoros Show the Earliest Austronesian Gene Flow into the Swahili Corridor. The American Journal of Human Genetics. 102:58–68.

Brucato N, Kusuma P, Beaujard P, Sudoyo H, Cox MP, Ricaut FX. 2017. Genomic admixture tracks pulses of economic activity over 2,000 years in the Indian Ocean trading network. Scientific Reports. 7:2919.

Brucato N, Kusuma P, Cox MP, Pierron D, Purnomo GA, Adelaar A, Kivisild T, Letellier T, Sudoyo H, Ricaut FX. 2016. Malagasy Genetic Ancestry Comes from an Historical Malay Trading Post in Southeast Borneo. Molecular Biology and Evolution. 33:2396–2400.

Buitenhuis B, Janss LL, Poulsen NA, Larsen LB, Larsen MK, Sørensen P. 2014. Genome-wide association and biological pathway analysis for milk-fat composition in danish holstein and danish jersey cattle. BMC genomics. 15:1–11.

Cesar AS, Regitano LC, Mourão GB, Tullio RR, Lanna DP, Nassu RT, Mudado MA, Oliveira PS, do Nascimento ML, Chaves AS *et al*. 2014. Genome-wide association study for intramuscular fat deposition and composition in nellore cattle. BMC genetics. 15:1–15.

Cheke A. 2010. The timing of arrival of humans and their commensal animals on Western Indian Ocean oceanic islands. Phelsuma. 18:38–69.

Choudhury A, Aron S, Sengupta D, Hazelhurst S, Ramsay M. 2018. African genetic diversity provides novel insights into evolutionary history and local adaptations. Human Molecular Genetics. 27:R209–R218.

Cui CY, Schlessinger D. 2015. Eccrine sweat gland development and sweat secretion. Experimental Dermatology. 24:644–650.

Czarnecki A, Dufy-Barbe L, Huet S, Odessa MF, Bresson-Bepoldin L. 2003. Potassium channel expression level is dependent on the proliferation state in the GH3 pituitary cell line. American Journal of Physiology-Cell Physiology. 284:C1054–C1064.

Cêtre-Sossah C, Pédarrieu A, Guis H, Defernez C, Bouloy M, Favre J, Girard S, Cardinale E, Albina E. 2012. Prevalence of Rift Valley Fever among Ruminants, Mayotte. Emerging Infectious Diseases. 18:972–975.

De Camargo G, Aspilcueta-Borquis RR, Fortes M, Porto-Neto R, Cardoso DF, Santos D, Lehnert S, Reverter A, Moore S, Tonhati H. 2015. Prospecting major genes in dairy buffaloes. BMC genomics. 16:1–14.

de Camargo GMF, Porto-Neto LR, Kelly MJ, Bunch RJ, McWilliam SM, Tonhati H, Lehnert SA, Fortes MR, Moore SS. 2015. Non-synonymous mutations mapped to chromosome x associated with andrological and growth traits in beef cattle. BMC genomics. 16:1–10.

De Deken R, Martin V, Saido A, Madder M, Brandt J, Geysen D. 2007. An outbreak of East Coast Fever on the Comoros: A consequence of the import of immunised cattle from Tanzania? Veterinary Parasitology. 143:245–253.

Dommergues L, Pannequin M, Cavalerie L, Cardinale E. 2015. Etude épidémiologique sur le charbon symptomatique à Mayotte (2014). Technical report. Coopadem - GDS Mayotte.

Druet T, Gautier M. 2017. A model-based approach to characterize individual inbreeding at both global and local genomic scales. Molecular Ecology. 26:5820–5841.

Druet T, Gautier M. 2021. An improved hidden markov model for the characterization of homozygous-by-descent segments in individual genomes. bioRxiv. .

Felius M. 2016. *On the breeds of cattle: Their history, classification and conservation*. Ph.D. thesis. University Utrecht.

Flori L, Fritz S, Jaffrézic F, Boussaha M, Gut I, Heath S, Foulley JL, Gautier M. 2009. The genome response to artificial selection: a case study in dairy cattle. PloS One. 4:e6595.

Flori L, Moazami-Goudarzi K, Alary V, Araba A, Boujenane I, Boushaba N, Casabianca F, Casu S, Ciampolini R, D'Acier AC *et al*. 2019. A genomic map of climate adaptation in Mediterranean cattle breeds. Molecular Ecology. 28:1009–1029.

Flori L, Thevenon S, Dayo GK, Senou M, Sylla S, Berthier D, Moazami-Goudarzi K, Gautier M. 2014. Adaptive admixture in the West African bovine hybrid zone: insight from the Borgou population. Molecular Ecology. 23:3241–3257. WOS:000338014900009.

France M. 2011. Synthèse illustrée du recensement agricole 2010. Technical report. Ministère de l'Agriculture, de l'Alimentation, de la Pêche, de la Ruralité et de l'Aménagement du Territoire. Mayotte.

Fuller D, Boivin N, Hoogervorst T, Allaby R. 2011. Across the Indian Ocean: the prehistoric movement of plants and animals. Antiquity. 85:544–558.

Fuller DQ, Boivin N. 2009. Crops, cattle and commensals across

the Indian Ocean. Etudes Ocean Indien. pp. 13–46.
Gautier M, Faraut T, Moazami-Goudarzi K, Navratil V, Foglio M, Grohs C, Boland A, Garnier JG, Boichard D, Lathrop GM *et al*. 2007. Genetic and Haplotypic Structure in 14 European and African Cattle Breeds. Genetics. 177:1059–1070.
Gautier M, Flori L, Riebler A, Jaffrézic F, Laloé D, Gut I, Moazami-Goudarzi K, Foulley JL. 2009. A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. BMC genomics. 10:550.
Gautier M, Klassmann A, Vitalis R. 2017. rehh 2.0 : a reimplementation of the R package rehh to detect positive selection from haplotype structure. Molecular Ecology Resources. 17:78–90.
Gautier M, Laloë D, Moazami-Goudarzi K. 2010. Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. PloS One. 5.
Gautier M, Naves M. 2011. Footprints of selection in the ancestral admixture of a New World Creole cattle breed. Molecular Ecology. 20:3128–3143.
Gautier M, Vitalis R, Flori L, Estoup A. 2021. f-statistics estimation and admixture graph construction with Pool-Seq or allele count data using the R package poolfstat. bioRxiv. 2021.05.28.445945.
Ghoreishifar SM, Eriksson S, Johansson AM, Khansefid M, Moghaddaszadeh-Ahrabi S, Parna N, Davoudi P, Javanmard A. 2020. Signatures of selection reveal candidate genes involved in economic traits and cold acclimation in five swedish cattle breeds. Genetics Selection Evolution. 52:1–15.
Glass EJ, Preston PM, Springbett A, Craigmile S, Kirvar E, Wilkie G, Brown CD. 2005. Bos taurus and bos indicus (sahiwal) calves respond differently to infection with theileria annulata and produce markedly different levels of acute phase proteins. International journal for parasitology. 35:337–347.
Hanotte O. 2002. African Pastoralism: Genetic Imprints of Origins and Migrations. Science. 296:336–339.
Hansen P. 2004. Physiological and cellular adaptations of zebu cattle to thermal stress. Animal reproduction science. 82:349–360.
Hayes BJ, Lien S, Nilsen H, Olsen HG, Berg P, Maceachern S, Potter S, Meuwissen THE. 2008. The origin of selection signatures on bovine chromosome 6. Animal Genetics. 39:105–111.
Hegardt FG. 1999. Mitochondrial 3-hydroxy-3-methylglutaryl-coa synthase: a control enzyme in ketogenesis. Biochemical Journal. 338:569–582.
Hollmann AK, Bleyer M, Tipold A, Neßler JN, Wemheuer WE, Schütz E, Brenig B. 2017. A genome-wide association study reveals a locus for bilateral iridal hypopigmentation in holstein friesian cattle. BMC genetics. 18:1–8.
Hou GY, Yuan ZR, Zhou HL, Zhang LP, Li JY, Gao X, Wang DJ, Gao HJ, Xu SZ. 2011. Association of thyroglobulin gene variants with carcass and meat quality traits in beef cattle. Molecular biology reports. 38:4705–4708.
Jian W, Duangjinda M, Vajrabukka C, Katawatin S. 2014. Differences of skin morphology in Bos indicus, Bos taurus, and their crossbreds. International Journal of Biometeorology. 58:1087–1094.
Júnior GAF, de Oliveira HN, Carvalheiro R, Cardoso DF, Fonseca LFS, Ventura RV, de Albuquerque LG. 2020. Whole-genome sequencing provides new insights into genetic mechanisms of tropical adaptation in nellore (bos primigenius indicus). Scientific reports. 10:1–7.
Kadri NK, Harland C, Faux P, Cambisano N, Karim L, Coppieters W, Fritz S, Mullaart E, Baurain D, Boichard D *et al*. 2016. Coding and noncoding variants in hfm1, mlh3, msh4, msh5, rnf212, and rnf212b affect recombination rate in cattle. Genome Research. .
Keightley PD, Eyre-Walker A. 2000. Deleterious mutations and the evolution of sex. Science. 290:331–3.
Kelleher J, Etheridge AM, McVean G. 2016. Efficient coalescent simulation and genealogical analysis for large sample sizes. PLoS Computational Biology. 12:e1004842.
Kim S, Cheong HS, Shin HD, Lee SS, Roh HJ, Jeon DY, Cho CY. 2018. Genetic diversity and divergence among Korean cattle breeds assessed using a BovineHD single-nucleotide polymorphism chip. Asian-Australasian Journal of Animal Sciences. 31:1691–1699.
Klassmann A, Gautier M. 2020. Detecting selection using extended haplotype homozygosity-based statistics on unphased or unpolarized data. Authorea. 10.22541/au.160405572.29972398/v1.
Ko S, Zhao MG, Toyoda H, Qiu CS, Zhuo M. 2005. Altered Behavioral Responses to Noxious Stimuli and Fear in Glutamate Receptor 5 (GluR5)- or GluR6-Deficient Mice. Journal of Neuroscience. 25:977–984.
Kortüm F, Caputo V, Bauer CK, Stella L, Ciolfi A, Alawi M, Bocchinfuso G, Flex E, Paolacci S, Dentici ML *et al*. 2015. Mutations in kcnh1 and atp6v1b2 cause zimmermann-laband syndrome. Nature genetics. 47:661–667.
Lin TC. 2021. Functional roles of spink1 in cancers. International Journal of Molecular Sciences. 22:3814.
Lipson M. 2020. Applying f-statistics and admixture graphs: Theory and examples. Molecular Ecology Resources. 20:1658–1667.
Littlejohn MD, Henty KM, Tiplady K, Johnson T, Harland C, Lopdell T, Sherlock RG, Li W, Lukefahr SD, Shanks BC *et al*. 2014. Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. Nature Communications. 5:1–8.
Loh PR, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, Berger B. 2013. Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium. Genetics. 193:1233–1254.
Mahadevan P. 1955. Population and production characteristics of red sindhi cattle in ceylon. Journal of Dairy Science. 38:1231–1241.
Mattioli R, Bah M, Kora S, Cassama M, Clifford D. 1995. Susceptibility to different tick genera in gambian n'dama and gobra zebu cattle exposed to naturally occurring tick infestations. Tropical animal health and production. 27:95–105.
Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TP, Sonstegard TS *et al*. 2009. Development and characterization of a high density snp genotyping assay for cattle. PloS one. 4:e5350.
Mbole-Kariuki MN, Sonstegard T, Orth A, Thumbi S, de C Bronsvoort B, Kiara H, Toye P, Conradie I, Jennings A, Coetzer K *et al*. 2014. Genome-wide analysis reveals the ancient and recent admixture history of east african shorthorn zebu from western kenya. Heredity. 113:297–305.
McPherson K. 1984. Cultural exchange in the indian ocean region. Westerly. 29:5–16.
Ministère de l'Agriculture dledlpdM. 2007. Recensement de l'agriculture (RA) Campagne agricole 2004-2005 - Tome 1 Généralités, méthodologies et principaux résultats. Technical report. Ministère de l'Agriculture, de l'élevage et de la pêche de Madagascar.
Morimoto R, Shindou H, Tarui M, Shimizu T. 2014. Rapid production of platelet-activating factor is induced by protein kinase cα-mediated phosphorylation of lysophosphatidylcholine acyltransferase 2 protein. Journal of Biological Chemistry. 289:15566–

15576.
Newitt M. 1983. The comoro islands in indian ocean trade before the 19th century (les comores et le commerce dans l'océan indien avant le xixe siècle). Cahiers d'études africaines. pp. 139–165.
Ouvrard M, Magnier J, Raoul S, Anlidine M, Attoumani H, Oussoufi A, Kamardine M, Janelle J, Naves M, Flori L *et al*. 2018. Caractérisation phénotypique et génétique : le cas du zébu mahorais.
Patterson N, Price AL, Reich D. 2006. Population Structure and Eigenanalysis. PLOS Genetics. 2:e190.
Patterson NJ, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient Admixture in Human History. Genetics. p. genetics.112.145037.
Pauly M. 2013. Acoua-Agnala M'kiri, Mayotte (976), archéologie d'une localité médiévale (XIe-XVe siècles), entre Afrique et Madagascar. NYAME AKUMA. pp. 73–90.
Pickrell JK, Patterson N, Loh PR, Lipson M, Berger B, Stoneking M, Pakendorf B, Reich D. 2014. Ancient west Eurasian ancestry in southern and eastern Africa. Proceedings of the National Academy of Sciences. 111:2632–2637.
Porter V. 2007. *Cattle : A Handbook to the Breeds of the World*. Crowood.
Porter V, Alderson L, Hall SJG, Sponenberg DP. 2016. *Mason's World Encyclopedia of Livestock Breeds and Breeding, 2 Volume Pack*. CABI.
Porto-Neto LR, Reverter A, Prayaga KC, Chan EK, Johnston DJ, Hawken RJ, Fordyce G, Garcia JF, Sonstegard TS, Bolormaa S *et al*. 2014. The genetic architecture of climatic adaptation of tropical cattle. Plos one. 9:e113284.
Purves D, Augustine GJ, Fitzpatrick D, Katz LC, Lamantia AS, McNamara JO, Williams SM. 2001. *Neuroscience*. Sinauer Associates. second edition.
Reid MA, Allen AE, Liu S, Liberti MV, Liu P, Liu X, Dai Z, Gao X, Wang Q, Liu Y *et al*. 2018. Serine synthesis through phgdh coordinates nucleotide levels by maintaining central carbon metabolism. Nature communications. 9:1–11.
Rosen BD, Bickhart DM, Schnabel RD, Koren S, Elsik CG, Tseng E, Rowan TN, Low WY, Zimin A, Couldrey C *et al*. 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. GigaScience. 9. giaa021.
Santiago E, Novo I, Pardiñas AF, Saura M, Wang J, Caballero A. 2020. Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. Molecular Biology and Evolution. 37:3642–3653.
Scheet P, Stephens M. 2006. A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. American Journal of Human Genetics. 78:629–644.
Sempéré G, Moazami-Goudarzi K, Eggen A, Laloë D, Gautier M, Flori L. 2015. WIDDE : a Web-Interfaced next generation database for genetic diversity exploration, with a first application in cattle. BMC Genomics. 16.
Serão NV, González-Peña D, Beever JE, Faulkner DB, Southey BR, Rodriguez-Zas SL. 2013. Single nucleotide polymorphisms and haplotypes associated with feed efficiency in beef cattle. BMC genetics. 14:1–20.
Silanikove N, Shamay A, Shinder D, Moran A. 2000. Stress down regulates milk yield in cows by plasmin induced B-casein product that blocks K+ channels on the apical membranes. Life Sciences. 67:2201–2212.
Sun L, Lamont SJ, Cooksey AM, McCarthy F, Tudor CO, Vijay-Shanker K, DeRita RM, Rothschild M, Ashwell C, Persia ME *et al*. 2015. Transcriptome response to heat stress in a chicken hepatocellular carcinoma cell line. Cell Stress and Chaperones. 20:939–950.
Tang K, Thornton KR, Stoneking M. 2007. A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. PLoS Biology. 5:e171.
Tijjani A, Utsunomiya YT, Ezekwe AG, Nashiru O, Hanotte O. 2019. Genome sequence analysis reveals selection signatures in endangered trypanotolerant west african muturu cattle. Frontiers in genetics. 10:442.
Underwood E, Suttle N. 1999. *The mineral nutrition of livestock 3rd edition*. CABI.
Vilà-Brau A, De Sousa-Coelho AL, Mayordomo C, Haro D, Marrero PF. 2011. Human hmgcs2 regulates mitochondrial fatty acid oxidation and fgf21 expression in hepg2 cell line. Journal of Biological Chemistry. 286:20423–20430.
Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A Map of Recent Positive Selection in the Human Genome. PLOS Biology. 4:e72.
Wang H, Chai Z, Hu D, Ji Q, Xin J, Zhang C, Zhong J. 2019. A global analysis of cnvs in diverse yak populations using whole-genome resequencing. BMC genomics. 20:1–12.
Wang M, Kong L. 2019. pblat: a multithread blat algorithm speeding up aligning sequences to genomes. BMC Bioinformatics. 20:28.
Wickham H. 2016. ggplot2 : Elegant Graphics for Data Analysis. Springer-Verlag New York. .
Xu L, Bickhart DM, Cole JB, Schroeder SG, Song J, Tassell CPV, Sonstegard TS, Liu GE. 2015. Genomic signatures reveal new evidences for selection of important traits in domestic cattle. Molecular biology and evolution. 32:711–725.
Zafindrajaona PS. 1991. *Profils genetiques du zebu malgache*. thesis. Paris 11.
Zafindrajaona PS, Lauvergne J. 1993. Comparaison de populations de zébu malgache à l'aide des distances génétiques. Genetics, Selection, Evolution : GSE. 25:373–395.

5.3 Résultats supplémentaires obtenus avec la puce moyenne densité

Les résultats présentés ci-après sont issus d'un travail préliminaire exploratoire à l'aide d'un jeu de données issu de génotypage sur une puce bovine Illumina BovineSNP50. Ce jeu de données regroupent 532 individus dispatchés dans 24 populations génotypés sur 39 729 SNPs. Les populations d'ascendance taurine européenne (EUT), taurine africaine (AFT), zébuine (ZEB), résultantes d'un métissage EUT X ZEB, métissage EUT X AFT et métissage ZEB X AFT sont au nombre de 8, 4, 3, 1, 1 et 7 respectivement (Table S5.1).

5.3.1 Analyse exploratoire

Le jeu de données moyenne densité (39'729 SNP) a fait l'objet d'une analyse exploratoire au moyen d'une acp (logiciel smartpca [Patterson et al., 2006]) et d'un clustering non-supervisé (logiciel Admixture [Alexander et al., 2009]. Les résultats du premier plan de l'ACP (PC1 et PC2) et du clustering (K=3) sont visibles à la Figure S5.1. Les résultats du clustering pour K=3 nous permettent d'estimer pour chaque individus la proportion indicine (ZEB en rouge), taurine africaine (AFT en vert) et taurine européenne (EUT en bleu). Ainsi, sur les 150 individus MAY, 29 présentent une proportion d'EUT de plus de 4% dont 21 présentent une proportion d'EUT de plus de 5%. La moyenne de la proportion d'EUT (ligne jaune sur la Figure S5.2) est de 2%, 38 individus se situent au-dessus de ce seuil, soit 25% de la population échantillonnée.

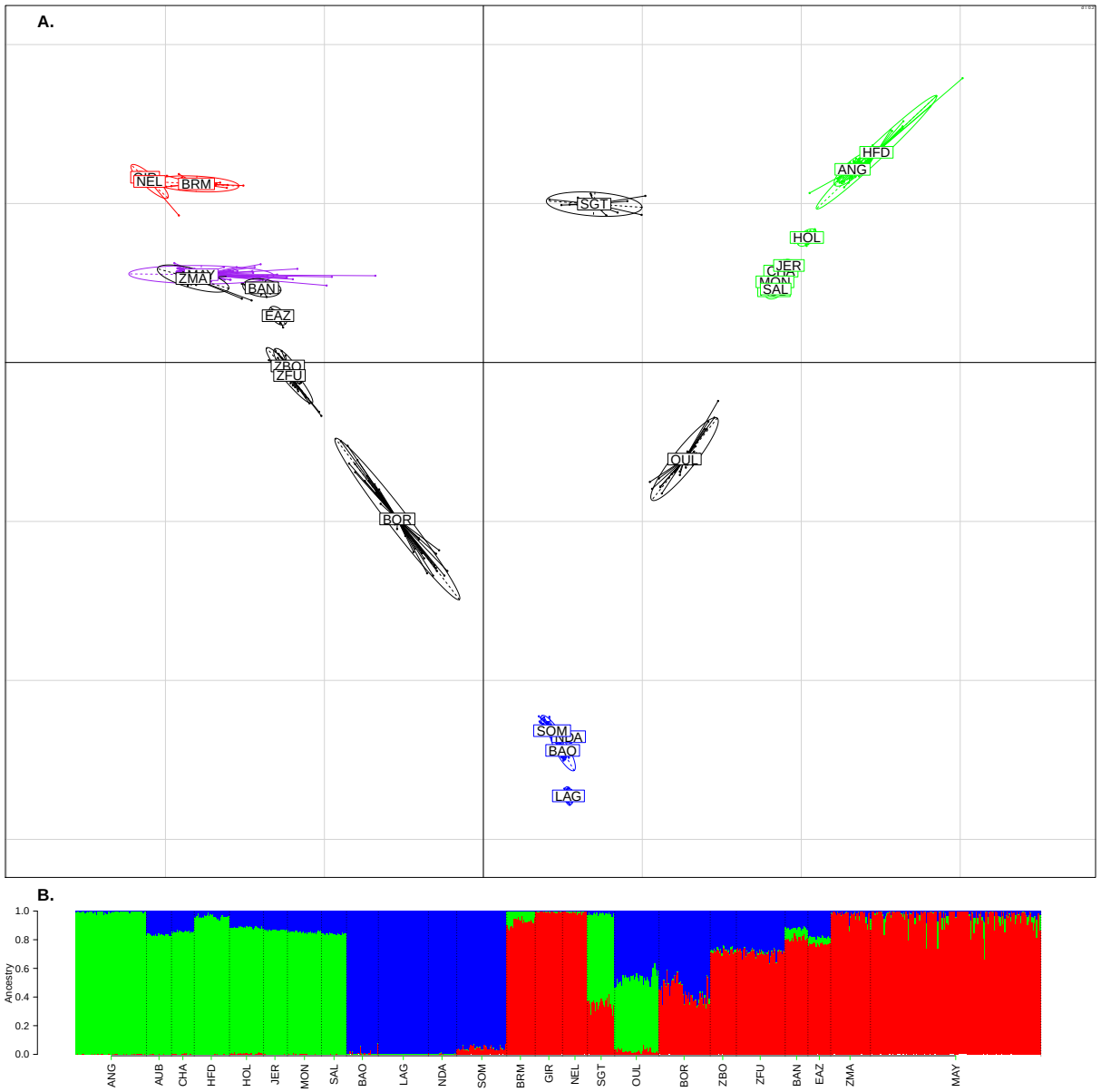


FIGURE S5.1 – Résultats de l'ACP **A** et du clustering **B** pour 532 individus génotypés sur puce bovine Illumina 54K (39'729 SNP). Les zébus mahorais MAY sont colorés en violet dans l'ACP. Les populations d'ascendance taurine européenne sont en vert, taurine africaine en bleu et zébu en rouge.

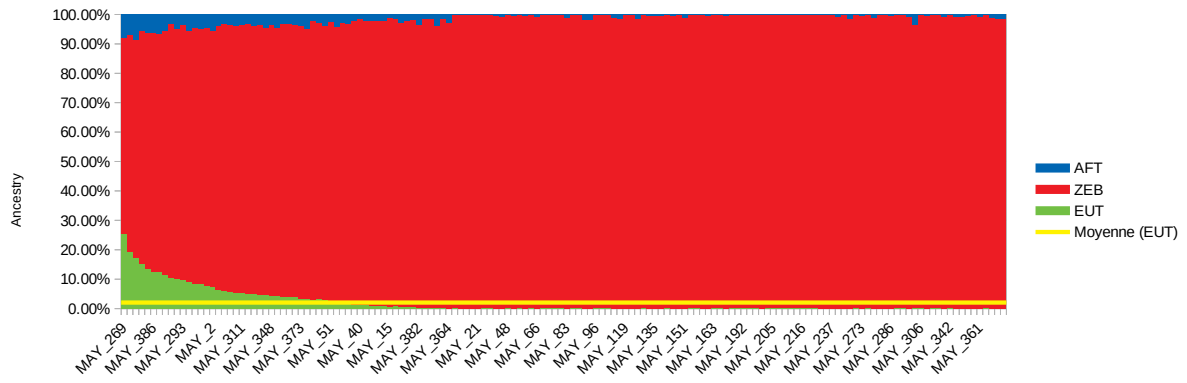


FIGURE S5.2 – Résultats du clustering pour 150 individus MAY génotypés sur puce bovine Illumina 54K (39'729 SNP).

TABLE S5.1 – Description des populations utilisées dans l'analyse exploratoire du jeu de données moyenne densité (532 individus génotypés sur 39'729 SNPs)

Population name	Population code	Land of origin	Nb-indiv-50K	Fis	Reference
Angus	ANG	Scotland	62	-0.00222 ± 0.009	Matukumalli et al, 2009
Aubrac	AUB	France	22	0.01057 ± 0.015	Gautier et al, 2010
Boran	BAN	Southern Ethiopia	20	-0.01539 ± 0.017	Decker et al., 2014
Baoule	BAO	Burkina Faso	28	-0.01324 ± 0.016	Gautier et al, 2009
Borgou	BOR	Benin	45	0.00408 ± 0.011	Gautier et al, 2009
Brahman	BRM	India (sampled in USA)	25	0.02684 ± 0.016	Matukumalli et al, 2009
Charolais	CHA	France	20	0.00128 ± 0.016	Gautier et al, 2010
East African Zebu	EAZ	Ethiopie	20	0.00317 ± 0.017	Decker et al., 2014
Gir	GIR	India (sampled in Brazil)	24	0.00691 ± 0.019	Matukumalli et al, 2009
Hereford	HFD	England	31	0.06264 ± 0.012	Matukumalli et al, 2009
Holstein	HOL	Holland	30	-0.00056 ± 0.013	Flori et al, 2009
Jersiaise	JER	Channel Islands	21	0.00076 ± 0.016	Gautier et al, 2010
Lagune	LAG	Benin	44	0.04014 ± 0.013	Gautier et al, 2009
Zebu Mahorais	MAY	Mayotte Island	150	0.04463 ± 0.008	This study
Montbéliarde	MON	France	30	-0.03652 ± 0.013	Flori et al, 2009
N'dama	NDA	Guinea	25	0.00722 ± 0.017	Matukumalli et al, 2009
Nelore	NEL	India (sampled in Brazil)	22	-0.02881 ± 0.020	Matukumalli et al, 2009
Oulmès Zaer	OUL	Morocco	39	-0.00219 ± 0.012	Gautier et al, 2009
Salers	SAL	France	22	0.01824 ± 0.015	Gautier et al, 2010
Santa Gertrudis	SGT	Texas	24	-0.00361 ± 0.014	Matukumalli et al, 2009
Somba	SOM	Togo	44	0.04068 ± 0.012	Gautier et al, 2009
Zebu Bororo	ZBO	Tchad	23	0.02018 ± 0.016	Gautier et al, 2009
Zebu Fulani	ZFU	West-Africa	43	0.00191 ± 0.011	Gautier et al, 2009
Zebu from Madagascar	ZMA	Madagascar	35	0.01164 ± 0.014	Gautier et al, 2009

5.3.2 Un évènement de métissage supplémentaire détecté

La datation des métissages par le logiciel *malder* a permis de détecter un métissage récent avec des populations européennes datant des années 1970 uniquement dans un jeu de données contenant les 129 animaux de la population de zébu mahorais présentant une proportion ascendance taurine européenne inférieure à 5% (il y a 6.9737 générations $\pm$ 1.00823, en considérant qu'une génération équivaut à 6 ans). Cet évènement pourrait correspondre à des métissages avec quelques individus taurins européens importés il y a 40 ans (Anonyme 1988). A Mayotte, les décideurs politiques encouragent le développement de l'insémination artificielle depuis au moins les années 1990 et soutiennent l'import de bétail sur pied d'origine européenne mais cela sans structuration des filières viande et lait et de l'accompagnement technique. Toutefois, une prise de conscience récente s'est engagée au niveau global [Scherf et al., 2015] et a permis de faire reconnaître la population de zébu mahorais en tant que race bovine locale française [noa, 2018a, noa, 2018b]. A la différence de la population de zébu mahorais, aucune introgression européenne n'a été détectée dans la population de zébu malgache. Ce résultat peut s'expliquer par une faible diffusion des ressources génétiques européennes sur la partie ouest de Madagascar, au moins jusqu'au début des années 1990 (date de l'échantillonnage de la population de zébu de Madagascar) malgré l'introduction de race européennes laitières et la création de la race hybride Renitelo au 20ème siècle [Porter and Tebbit, 2007, Porter et al., 2016]. Ce défaut de diffusion est certainement dû à l'étendue du territoire de Madagascar beaucoup plus important que l'île de Mayotte, qui a permis l'isolement des troupeaux. De plus, les conditions climatiques spécifiques de Madagascar et la pression pathogénique n'a pas encouragé les croisements [Ranaivoson et al., 1986]. La situation politique contrastée des deux territoires peut aussi être une raison à l'absence de métissage européen avec la population de zébu malgache. Néanmoins, du fait de la différence de date d'échantillonnage, une introgression européenne dans la population ZMA n'est pas à exclure et peut être intéressante à étudier.

5.3.3 Un clustering des ROH variable

Dans une première analyse exploratoire de la consanguinité récente à partir du calcul des Runs of Homozygosity au moyen du package RZooroh [Druet and Gautier, 2017, Bertrand et al., 2019], la distribution des ROH a été comparée pour les populations zébuine de Mayotte (MAY) et Madagascar (ZMA) sur les jeux de données issus du génotypage en moyenne (39'729 SNPs) et haute densité (644'327 SNPs). Pour rappel, chaque classe HBD est définie par son propre taux Rk qui détermine la longueur attendue (égale à 1/Rk Morgans) des segments concernés. Cela représente approximativement le double du nombre de générations entre l'ancêtre commun qui a transmis le segment d'ADN et la date d'échantillonnage. Pour les jeux de données issus des puces moyenne densité, nous avons considéré un modèle de 8 classes HBD (Rk égal à 2, 4, 8, 16, 32, 64, 128 et 256), pour ceux issus des puces HD, 11 classes HBD ont été prises en compte (Rk égal à 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024 et 2048). Les classes

non-HBD ont un taux égal à 256 pour le jeu de données moyenne densité et à 2048 pour le jeu de données haute densité. A noter que les segments HBD associés avec des classes dont les taux Rk sont supérieurs à 128 et 1024 ne seront pas capturés, dus aux densités respectives des puces de génotypage. Le modèle Rzooroh estime la contribution de chaque classe sur la consanguinité individuelle.

Le niveau moyen de la consanguinité individuelle est de 0.169 (SD = 0.08) pour MAY et 0.137 (SD = 0.03) pour ZMA pour les données SNP moyenne densité, et 0.313 (SD = 0.06) pour MAY et 0.315 (SD = 0.04) pour ZMA pour les données SNP haute densité. Pour les deux populations, les segments HBD sont concentrés dans les classes avec un taux Rk égal à 128 (64 generations, 384 ans) pour le jeux de données en moyenne densité, et dans les classes correspondant à un taux Rk de 128, 256 et 1024 (soit 64, 128 et 512 générations, 384, 782 et 3072 années avant l'échantillonnage)(Figure S5.3). Cela signifie que la consanguinité vient d'une origine ancienne, avec un potentiel ancêtre commun datant d'il y a 64, 128 et 512 générations. Cependant, les résultats sur moyenne densité révèlent une autre origine plus récente de la consanguinité pour certains individus MAY présentant de longs segments HBD qui appartiennent aux classes de faibles taux Rk (2 et 4, correspondant à 1 et 2 générations, soit 6 à 12 ans avant l'échantillonnage) bien que le niveau moyen de consanguinité individuelle soit faible pour ces classes.

5.3.4 Conclusion

Les résultats obtenus confirment une tendance à l'absorption récente des zébus locaux de Mayotte par des races taurines européennes. Concernant nos échantillons, cette ascendance vient de l'importation d'une centaine d'individus Montbéliards en 2008, et du service d'insémination proposé sur l'île de Mayotte essentiellement en race Montbéliarde. Les précédents résultats ont été intégrés dans le dossier de reconnaissance de la race Zébu mahorais. Il devient donc nécessaire d'engager un programme de gestion de la race Zebu Mahorais.

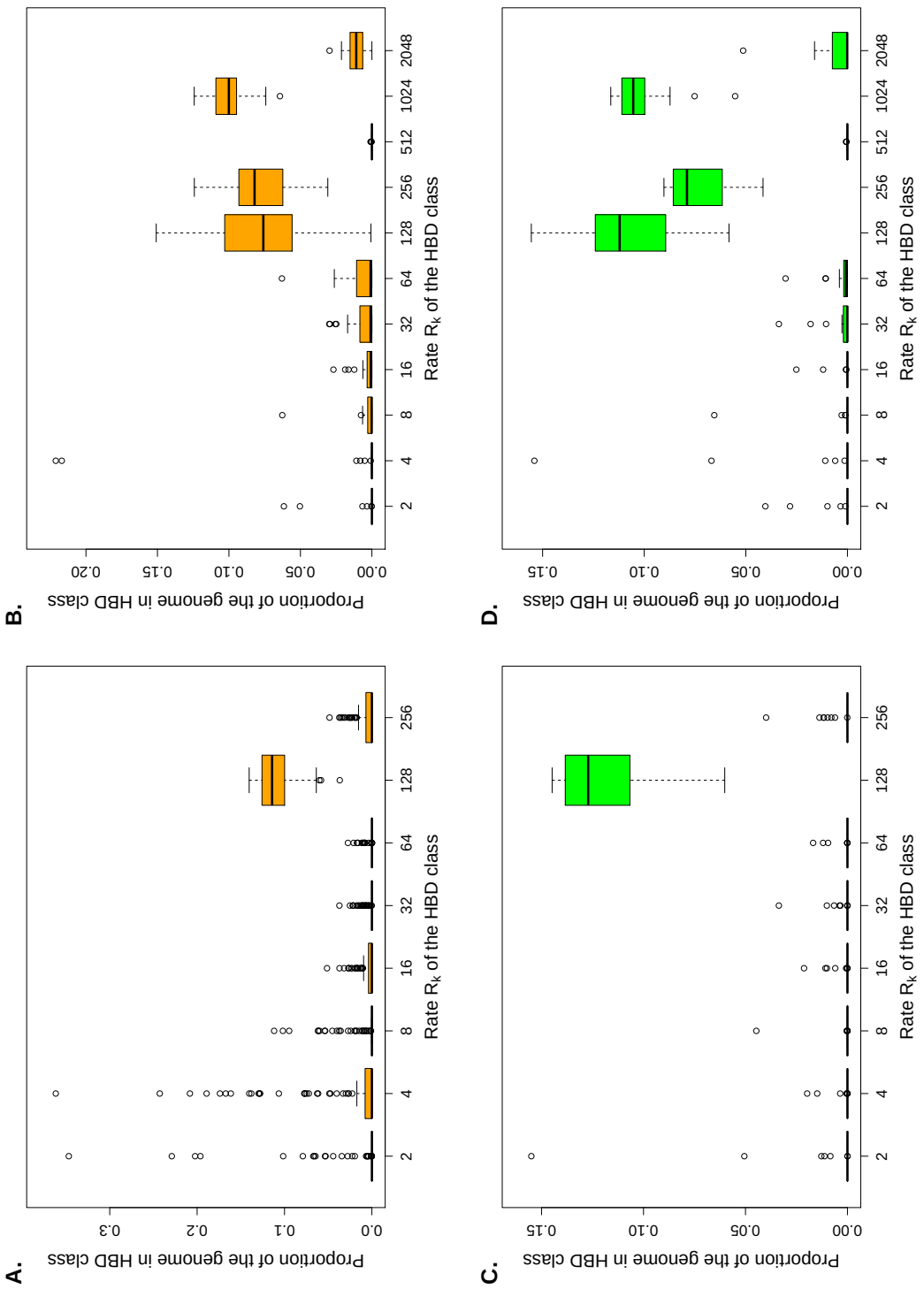


FIGURE S5.3 – Comparaison de la distribution des ROH pour MAY et ZMA en fonction du type de données. **A** : May en 54K ; **B** : May en HD ; **C** : Zma en 54K ; **D** : Zma en HD. La valeur de R_k correspond à 2 fois le nombre de générations écoulées à partir de la date d'échantillonnage. Par exemple pour les graphiques **A** et **B**, un pic de consanguinité est observé pour une valeur R_k de 128, soit à la 64ème génération.

5.4 Introgression du boeuf de Bali

L'étude bibliographique des études archéologiques [Beaujard, 2005, Beaujard, 2007] et génétiques récentes concernant les migrations des populations humaines dans l'Océan Indien ont permis d'émettre une nouvelle hypothèse concernant l'histoire démographique des populations bovines. Ces études montrent en effet que les populations humaines austronésiennes et africaines Swahili ont participé au peuplement des Comores et de Madagascar. Elles ont donc pu transporter par voie maritime des bovins sur pied [Fuller et al., 2011] provenant d'Indonésie et d'Afrique. Étant donné la présence en Indonésie d'une forme domestiquée de Banteng (Bos javanicus), appelée bœuf de Bali et de plusieurs populations issues du métissage *Bali x zébu* [Gao et al., 2017], il est possible que ces populations aient participé aux événements de métissage à l'origine des zébus de l'Océan Indien. Il semblerait également que des bovins indonésiens aient été introduits à la Réunion au milieu du 19ème siècle [Mandret, 2000]. Nous avons donc décidé de tenter de mettre en évidence un flux génique direct entre les bœufs de Bali ou les populations indonésiennes de zébus métissées avec le bœuf de Bali (Madura, Pesisir, Brebes) et les zébus de l'Océan Indien.

TABLE S5.2 – Description du jeu de données utilisé pour détecter une introgession indonésienne dans les zébus de l'Ouest de l'Océan Indien.

Pop	Name	Zone	Nb
ANG	Angus	Ecosse	62
ANK	Ankole	Afrique du Sud	22
AUB	Aubrac	France	22
BAO	Baoulé	Burkina Faso	28
BLI	Boeuf de Bali	Indonésie	20
BRE	Brebes	Indonésie	9
EAZ	East African Shorthone Zebu	Ethiopie	92
GIR	Gir	Inde	27
KAR	Zebu Karamajong	Kenya	16
LAG	Lagunaire	Bénin	44
MAD	Madura	Indonésie	7
MAY	Zebu Mahorais	Mayotte	150
MOK	Zebu Moka	La Réunion	31
NDA	N'Dama	Guinée	25
NEL	Nellore	Inde	31
NGA	Nganda	Ouganda	23
PES	Pesisir	Indonésie	6
SER	Serere	Afrique de l'Est	12
SIS	Sistani	Iran	34
SOM	Somba	Togo	44
ZMA	Zébu de Madagascar	Madagascar	35
OCB	Buffle Africain	-	2

Les résultats de clustering supervisés et non -supervisés obtenus par l'analyse du jeu de données constitué de 742 individus répartis dans 22 populations et 25'720 SNPs (Table S5.2)

avec le logiciel ADMIXTURE [Alexander et al., 2009] suggèrent une ascendance Bali chez les zébus Mahorais, de la Réunion et de Madagascar mais ne permettent pas de déterminer clairement une introgression de Bali cattle (Figure S5.4). Entre supervisé et non-supervisé, les résultats ne sont pas constants : non-supervisé, il semblerait qu'aucune introgression Bali ne se soit produit ; en mode supervisé, les populations Malgache et Mahoraise sont issues principalement d'une admixture entre Bali cattle et AFT. Il semblerait que la proportion de zébu ait été remplacée par le Bali.

Les D-statistiques laissent à supposer un flux génique entre les populations bovines indonésiennes (Bali cattle et zébus admixés) et les Zébu Africains (Z-score négatif) (Table S5.3). Un métissage entre les populations indicines indonésiennes et taurines (uniquement africaine pour MAY) est détecté pour MAY et MOK par les f3-stats (Table S5.4) mais n'est pas confirmé par le calcul des f4-ratio (Table S5.5). Concernant MAY, il est intéressant de noter que les résultats significatifs pour les f3-stats sont obtenus pour, le test du métissage *AFT X ZEB*, mais aussi uniquement pour le test du métissage *AFT X Ind* avec la population indonésienne Pesisir présentant une proportion zébuine beaucoup plus forte (environ 90 %) que les autres populations indonésiennes testées Brebes (80 %) et Madura (70 %), laissant à supposer que le métissage *ZEB X AFT* a été probablement mieux détecté que l'introgression indonésienne.

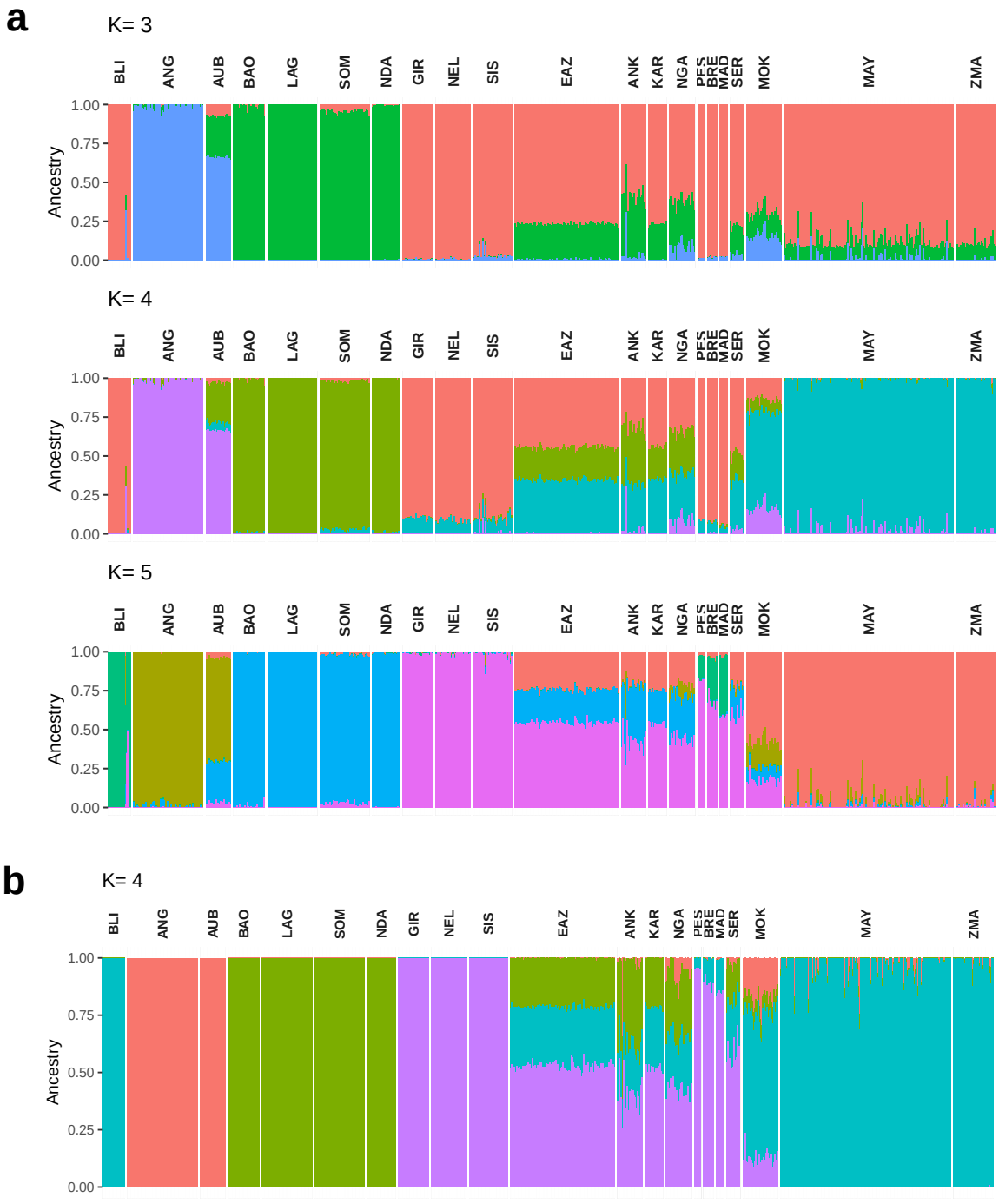


FIGURE S5.4 – Clustering hierarchique issus du logiciel ADMIXTURE [Alexander et al., 2009]. **a** Non-supervisé de K = 3 à 5. **b** Supervisé pour K = 4 avec comme populations ancestrales définies Bali cattle (BLI), Taurins européens (ANG, AUB), Taurins africains (BAO, LAG, NDA, SOM) et Zébus (GIR, NEL, SIS).

TABLE S5.3 – Résultats synthétiques du calcul des D-statistiques testant le flux génique entre les populations W, X, Y et Z. Si le Z-score est négatif, un flux génique a lieu entre W et Z ou entre X et Y. Si le Z-score est positif, le flux génique a lieu entre W et Y ou X et Z.

W	X	Y	Z	Z-score
MAD, BLI, PES, BRE	OCB	ANG, SOM	MOK, EAZ, ZMA, MAY	Négatif
MAD, BLI, PES, BRE	OCB	GIR	MOK, EAZ, ZMA, MAY	Positif

TABLE S5.4 – Résultats significatifs de la statistiques *f3* (Z-score <0) testant les métissages candidats *Pop1 X Pop2* à l'origine des populations de Zébu de Mayotte (MAY), Madagascar (ZMA) et de la Réunion (MOK).

Type	Pop (X)	Pop1	Pop2	f3_stat
AFT X Indo	MAY	SOM	PES	Négatif
AFT X ZEB	MAY	SOM	GIR	Négatif
Taurins X ZEB	MOK	ANG, SOM	GIR	Négatif
Taurin X AFZ	MOK	ANG, SOM	ZMA	Négatif
	MOK	ANG, SOM	MAY	Négatif
Taurin X Indonésie	MOK	ANG, SOM	PES	Négatif
	MOK	ANG, SOM	BRE	Négatif
	MOK	ANG, SOM	MAD	Négatif
	MOK	ANG, SOM	BLI	Négatif
Taurin X Outgroup	MOK	ANG, SOM	OCB	Négatif
AFT X ZEB	ZMA	SOM	GIR	Négatif

TABLE S5.5 – Résultats f4ratio avec Buffle Africain comme outgroup. Alpha indique la proportion estimée de la population B dans la population X.

Type	Pop (X)	Taurin (A)	Zebu (B)	Bali (C)	alpha	se	Z(null=0)
Indonesie	MAD	SOM	GIR	BLI	0.723948	0.014761	49.045
	BRE	SOM	GIR	BLI	0.788797	0.013716	57.511
	PES	SOM	GIR	BLI	0.913984	0.016688	54.768
	MAD	ANG	GIR	BLI	0.728956	0.015423	47.264
	BRE	ANG	GIR	BLI	0.792899	0.013979	56.722
	PES	ANG	GIR	BLI	0.902328	0.017385	51.902
AFZ	ZMA	ANG	GIR	BLI	1.544533	0.027437	56.293
	MAY	ANG	GIR	BLI	1.574389	0.026791	58.765
	EAZ	ANG	GIR	BLI	1.591333	0.025128	63.329
	MOK	ANG	GIR	BLI	1.876836	0.034049	55.122
	ZMA	SOM	GIR	BLI	1.668939	0.031048	53.754
	MAY	SOM	GIR	BLI	1.680974	0.030355	55.378
	EAZ	SOM	GIR	BLI	1.740186	0.028777	60.471
	MOK	SOM	GIR	BLI	1.838856	0.033365	55.114

5.5 Discussion du chapitre

Dans ce chapitre, l'histoire démographique des populations bovines locales de l'Ouest de l'Océan indien a été abordée. L'inférence démographique grâce à des méthodes de calcul basées sur la variation des fréquences alléliques et le déséquilibre de liaison a permis d'établir des scénarios de métissage pour les zébus locaux de Mayotte et de Madagascar. Ces deux populations auraient une origine commune issue d'un métissage ancien entre des populations sources proches des zébus indiens actuels et des populations sources proches des taurins africains actuels datée du 10^{iéme} siècle, suivi d'une introgression indicine datée du 12^{iéme} siècle, le lieu de ces métissages restant toutefois à déterminer :

(a) le premier métissage a-t-il eu lieu sur les îles suivi d'une introgression indicine au 12^{iéme} siècle ?

(b) le premier métissage a-t-il eu lieu sur le continent africain aboutissant aux zébus africains qui ont ensuite été introduits sur les îles puis ont été ensuite métissés avec des zébus indiens ?.

(c) les deux évènements de métissage ont eu lieu sur le continent africain puis les zébus de l'Océan Indien ont été ensuite introduit dans la zone ?

Après une période d'homogénéisation et à partir du 16^{iéme} siècle, les trajectoires d'évolution des tailles efficaces des deux populations ont divergé suggérant un isolement des deux populations bovines, concomitant à l'influence croissante des colons Européens dans la zone.

Bien que les logiciels Alder et malder aient permis d'améliorer la compréhension de la démographie et de l'histoire des populations bovines, le modèle de décroissance exponentielle du métissage, sur lequel les deux logiciels sont basés, a tendance à surestimer les dates de métissage les plus anciens [Chimusa et al., 2018]. En outre, ce modèle est limité à la détection d'au plus trois événements de mélange dans les événements de mélange multipoints. Dans le contexte des populations bovines et ovines de l'Ouest de l'Océan Indien, il n'est pas à exclure que des introductions régulières sur les îles et que des métissages complexes se soient déroulés au regard des intenses échanges commerciaux dans le canal du Mozambique. De plus le choix des populations bovines ancestrales est aussi à questionner : la population GIR estimée représentative de l'ascendance indicine d'origine indienne a été échantillonnée au Brésil. Bien qu'elle soit reconnue comme population de référence, une faible proportion (1 %) d'ascendance taurine européenne a été détectée [O'Brien et al., 2015].

Compte tenu des profils génétiques observés chez les populations humaines de la zone et suggérant une influence austronésienne, l'hypothèse d'une introgression indonésienne par l'intermédiaire du boeuf de Bali dans les populations bovines de l'Ouest de l'Océan Indien a été testée sur des données de génotypage moyenne densité. Les résultats confirment la présence de flux géniques (D-stats) entre les bovins indonésiens et les zébus locaux de la zone d'étude. L'introgression indonésienne des zébus Moka, auparavant documentée [Mandret, 2000], est confirmée timidement. Ces résultats peu significatifs quant à l'hypothèse d'introgression indonésienne dans les zébus locaux de l'Ouest de l'Océan Indien sont probablement dûs d'une

part à la faible densité des marqueurs SNP de la puce Illumina 50K et d'autre part au biais de recrutement des marqueurs SNP de cette puce, qui a été effectués sur jeux de données regroupant des bovins taurins. L'utilisation de la puce bovine haute densité, ou mieux des données de séquences pourraient permettre de mieux explorer cette hypothèse en s'affranchissant des biais de la puce bovine moyenne densité.

La divergence des histoires démographiques récentes des populations de zébus de Mayotte et Madagascar est liée aux contextes géographiques et des systèmes d'élevage spécifiques à chaque territoire (système au piquet/élevage en parc pour Mayotte, élevage extensif pour Madagascar). Cette divergence a abouti à des histoires adaptatives différentes caractérisés par des variations locales génomiques entre les deux populations. Hormis les 8 gènes trouvés en commun associés à des signaux de sélection déterminés par le calcul de scores basés sur l'étendue d'homozygotie haplotypique (iHS, Rsb), les réseaux de gènes sont différents mais sont associés aux grandes fonctions biologiques impliquées dans l'adaptation aux conditions tropicales : traits de la peau et pigmentation, fonctions respiratoires, système nerveux et comportement, métabolisme, système endocrinien, réponse inflammatoire et immunitaire. Ces fonctions biologiques ont été sélectionnées en réponse au climat qui constitue une pression de sélection conjointe avec celle exercée par les hommes, consciemment ou non. Parmi les gènes candidats identifiés et associés à au moins une signature de sélection, certains ont été trouvés également sous sélection chez d'autres populations bovines tropicales : *ANO4* pour la Créole de Guadeloupe [Gautier and Naves, 2011], *KCTD8, KCHN1* chez les races locales de méditerrannée [Flori et al., 2019], *TRIM10, TRIM15, TRIM40* chez la Muturu, une race taurine africaine [Tijjani et al., 2019]. Il reste cependant un pan de comparaison à mener sur les régions sous sélection. En effet, les analyses de signatures de sélection sur d'autres populations locales bovines ont été effectuées sur l'assemblage UMD3.1/ Btau6 contrairement à notre étude qui a été effectuée sur l'assemblage ARS-UCD1.2/Btau9. Une procédure de conversion des positions à l'aide du logiciel "liftover" est donc nécessaire afin d'effectuer des comparaisons fiables.

Le niveau de consanguinité de la population de zébu de Mayotte correct (30 % basé sur les proportions de segments chromosomiques homozygotes, F_{is} = 0.028 [0.013 - 0.041]) au moment de l'échantillonnage en 2017 donne de bonnes perspectives pour l'avenir de la race et la préservation de cette ressource génétique originale. Cependant cette population jusqu'à récemment n'avait pas d'existence officielle et l'augmentation des individus de race européenne et d'individus croisés identifiés officiellement par la Chambre d'Agriculture de Mayotte laissait à craindre un risque d'absorption et de disparition. La reconnaissance officielle de la race par l'Etat français effective depuis septembre 2018 est déjà une étape primordiale franchie qui a permis d'engager la requalification des animaux auparavant définis comme "autre race allaitante étrangère" dans la Base Nationale d'Identification en tant que zébu mahorais. Cette étape de recensement servira de base pour la constitution d'un noyau d'animaux fondateurs de la race. Néanmoins, le contexte difficile que rencontre les organisations professionnelles agricoles sur le territoire du département de Mayotte peut être un frein quant à la mise en

place des outils de gestion de la race. De plus, le zébu de Mayotte fait face à la concurrence des races taurines européennes importées ou introduites par l'insémination artificielle présentant des niveaux de productions laitières et bouchères supérieures. Il s'agira donc d'identifier les orientations stratégiques de commercialisation des produits issus de la race locale et d'organiser le schéma de conservation et/ou de sélection en fonction des objectifs.

Bibliographie

[noa, 2018a] (2018a). Arrêté du 25 septembre 2018 modifiant l'arrêté du 12 décembre 2013 relatif à l'enregistrement et à la certification de la parenté des bovins.

[noa, 2018b] (2018b). Arrêté du 25 septembre 2018 modifiant l'arrêté du 29 avril 2015 fixant la liste des races des espèces bovine, ovine, caprine et porcine reconnues et précisant les ressources zoogénétiques présentant un intérêt pour la conservation du patrimoine génétique du cheptel et l'aménagement du territoire - Légifrance.

[Alexander et al., 2009] Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9) :1655–1664.

[Beaujard, 2005] Beaujard, P. (2005). The Indian Ocean in Eurasian and African world-systems before the sixteenth century. *Journal of World History*, 16(4) :411–+. WOS :000235458400002.

[Beaujard, 2007] Beaujard, P. (2007). East Africa, the Comoros Islands and Madagascar before the sixteenth century : On a neglected part of the world system. *Azania : Archaeological Research in Africa*, 42(1) :15–35.

[Bertrand et al., 2019] Bertrand, A., Kadri, N. K., Flori, L., Gautier, M., and Druet, T. (2019). RZooRoH : An R package to characterize individual genomic autozygosity and identify homozygous-by-descent segments. *Methods in Ecology and Evolution*, 10(6).

[Chimusa et al., 2018] Chimusa, E. R., Defo, J., Thami, P. K., Awany, D., Mulisa, D. D., Allali, I., Ghazal, H., Moussa, A., and Mazandu, G. K. (2018). Dating admixture events is unsolved problem in multi-way admixed populations. *Briefings in Bioinformatics*.

[Druet and Gautier, 2017] Druet, T. and Gautier, M. (2017). A model-based approach to characterize individual inbreeding at both global and local genomic scales. *Molecular Ecology*, 26(20) :5820–5841.

[Flori et al., 2019] Flori, L., Moazami-Goudarzi, K., Alary, V., Araba, A., Boujenane, I., Boushaba, N., Casabianca, F., Casu, S., Ciampolini, R., D'Acier, A. C., Coquelle, C., Delgado, J.-V., El-Beltagi, A., Hadjipavlou, G., Jousselin, E., Landi, V., Lauvie, A., Lecomte, P., Ligda, C., Marinthe, C., Martinez, A., Mastrangelo, S., Menni, D., Moulin, C.-H., Osman, M.-A., Pineau, O., Portolano, B., Rodellar, C., Saïdi-Mehtar, N., Sechi, T., Sempéré, G., Thévenon, S., Tsiokos, D., Laloë, D., and Gautier, M. (2019). A genomic map of climate adaptation in Mediterranean cattle breeds. *Molecular Ecology*, 28(5) :1009–1029.

[Fuller et al., 2011] Fuller, D., Boivin, N., Hoogervorst, T., and Allaby, R. (2011). Across the Indian Ocean : the prehistoric movement of plants and animals. *Antiquity*, 85(328) :544–558.

[Gao et al., 2017] Gao, Y., Gautier, M., Ding, X., Zhang, H., Wang, Y., Wang, X., Faruque, M. O., Li, J., Ye, S., Gou, X., Han, J., Lenstra, J. A., and Zhang, Y. (2017). Species composition and environmental adaptation of indigenous Chinese cattle. *Scientific Reports*, 7(1) :1–14.

[Gautier and Naves, 2011] Gautier, M. and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, 20(15) :3128–3143.

[Mandret, 2000] Mandret, G. (2000). *L'élevage bovin à la Réunion : Synthèse de quinze ans de recherche*. OCLC : 966066698.

[O'Brien et al., 2015] O'Brien, A. M. P., Höller, D., Boison, S. A., Milanesi, M., Bomba, L., Utsunomiya, Y. T., Carvalheiro, R., Neves, H. H., da Silva, M. V., VanTassell, C. P., Sonstegard, T. S., Mészáros, G., Ajmone-Marsan, P., Garcia, F., and Sölkner, J. (2015). Low levels of taurine introgression in the current Brazilian Nelore and Gir indicine cattle populations. *Genetics Selection Evolution*, 47(1) :31.

[Patterson et al., 2006] Patterson, N., Price, A. L., and Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genetics*, 2(12).

[Porter et al., 2016] Porter, V., Alderson, L., Hall, S. J. G., and Sponenberg, D. P. (2016). *Mason's World Encyclopedia of Livestock Breeds and Breeding, 2 Volume Pack*. CABI. Google-Books-ID : 2UEJDAAAQBAJ.

[Porter and Tebbit, 2007] Porter, V. and Tebbit, J. (2007). *Cattle : a handbook to the breeds of the world*. Crowood, Ramsbury, Marlborough, Wiltshire. OCLC : ocm84151515.

[Ranaivoson et al., 1986] Ranaivoson, A., Ranaivoson, R., and Rambelomanana, D. (1986). Epizootiologie et incidence de la dermatophilose bovine à Madagascar. *Revue d'élevage et de médecine vétérinaire des pays tropicaux*, 39(3-4) :279–287.

[Scherf et al., 2015] Scherf, B. D., Pilling, D., and others (2015). The second report on the state of the world's animal genetic resources for Food and Agriculture.

[Tijjani et al., 2019] Tijjani, A., Utsunomiya, Y. T., Ezekwe, A. G., Nashiru, O., and Hanotte, O. (2019). Genome Sequence Analysis Reveals Selection Signatures in Endangered Trypanotolerant West African Muturu Cattle. *Frontiers in Genetics*, 10 :442.

Chapitre 6

Caractérisation génétique du Mouton mahorais et localisation de la mutation responsable de microtie

L'étape de caractérisation phénotypique du mouton mahorais a mis en évidence un gabarit globalement homogène et révélé un profil phénotypique plus diversifié basé sur la détermination des paramètres de coloration et de morphologie (queue et oreilles notamment). Cette étape a en particulier permis de mettre en évidence des individus (N=21) présentant des oreilles de taille réduite ou une atropie de l'oreille externe (microtie). Afin de caractériser génétiquement cette race, nous avons étudié sa diversité génétique en analysant les données de génotypage d'un sous-ensemble animaux, représentatifs de la diversité des élevages, choisis sur la base de leur phénotype, en minimisant l'apparentement. Cette caractérisation génétique a été réalisée dans le cadre du projet DEFI-ANIMAL financé par l'Europe, l'Etat français et le département de Mayotte. Nous avons également entrepris d'étudier le déterminisme génétique de la microtie dans cette population. Ce phénotype a déjà été observé chez certaines races de moutons de différentes zones géographiques : chez les moutons Awassi et Altay en Asie et chez la race italienne Valle del Belice [Jawasreh et al., 2016, Mastrangelo et al., 2018, He et al., 2020]

6.1 Synthèse des résultats

Sur les 100 moutons phénotypés, 48 ont été génotypés sur puce Illumina ovine 50K et 8 séquencés. L'analyse exploratoire (ACP, classification hiérachique, calcul de distances génétiques) des 48 individus a permis de montrer que la population était génétiquement homogène et de confirmer la proximité génétique des animaux échantillonnés avec le groupe des moutons africains à queue grasse. La consanguinité moyenne individuelle a été évaluée à 30%, avec un individu présentant une consanguinité très élevée (60%) et très récente (Rk = 2, soit 1 génération). Deux pics de consanguinité ont été détectés au début du 20ème siècle puis beaucoup plus récemment il y a une dizaine d'années (analyse de la consanguinité récente avec le logiciel RZooRoh [Druet and Gautier, 2017a, Bertrand et al., 2019]). Ces augmentations de consanguinité pourraient être des conséquences d'épisodes infectieux ou de l'introduction de moutons sur l'île de Mayotte.

Plusieurs tests d'association ont détecté un SNP associé avec le phénotype oreille réduite ou atrophiée (SNP OAR-115757345.1) à la position 114 296 273 sur le chromosome 6, à proximité du gène HMX1 qui est associé à la microtie chez les moutons chinois de race Altay [He et al., 2020].

La recherche des signatures de sélection à partir du calcul de l'étendue du déséquilibre de liaison (scores iHS et Rsb) et l'annotation fonctionnelle sur les gènes candidats a permis de déterminer les fonctions biologiques relatives au système nerveux, reproductif, la réponse aux agents pathogènes, le métabolisme, les fonctions dermatologiques.

6.2 Point méthodologique sur les analyses d'association cas/contrôles

Les analyses d'association pangénomiques (GWAS) sont largement utilisées en médecine humaine pour rechercher des régions candidates associées à des pathologies. L'approche la plus commune est l'étude cas-contrôle, qui compare deux grands groupes d'individus, un groupe contrôle en bonne santé et un groupe de cas affecté par une pathologie. En recherche agronomique appliquée aux animaux d'élevage ces analyses sont utilisées pour localiser les mutations associées à un phénotype particulier : microtie chez certaines populations ovines [Jawasreh et al., 2016, Mastrangelo et al., 2018] et bovines [Koch et al., 2013] ou encore le syndrome German White Fleckvieh [Philipp et al., 2011] pour ne citer que quelques exemples.

Le principe général des GWAS cas/contrôles repose sur la comparaison des fréquences alléliques d'un grand nombre de marqueurs répartis sur le génome chez les cas (présentant le phénotype) et les contrôles d'une population de sujets non apparentés. L'hypothèse du test étant que la fréquence d'un variant impliqué dans un phénotype est significativement différente dans le groupe des cas comparativement aux contrôles.

Les méthodes les plus classiques (test du χ^2, test Exact de Fisher, test de Cochran Armitage) sont basées sur l'analyse, pour chaque SNP, des tables de contingences croisant les fréquences alléliques ou génotypiques présentes dans le groupe des cas et le groupe des contrôles et testent l'hypothèse nulle H_0 d'absence d'association entre les lignes (phénotype cas vs controle) et les colonnes (les 3 génotypes ou les deux allèles) (Figure 6.1).

Compte tenu du nombre élevé de tests effectués (lié au nombre élevé de marqueurs), il est nécessaire d'ajuster le seuil de significativité pour limiter le taux de faux positifs en utilisant la correction de Bonferroni ($p_{Bonf} = p_{test}$/*nb de SNPs*) ou d'autres méthodes [Bush and Moore, 2012] comme la détermination du taux de faux positivité (FDR) proposé par Benjamini et Hochberg [Hochberg and Benjamini, 1990, Benjamini and Hochberg, 1995].

La détection d'une association entre un SNP et le phénotype d'intérêt suggère que le marqueur est soit directement impliqué dans le phénotype (variant causal), soit en déséquilibre de liaison avec un variant causal. En effet, en cas de déséquilibre de liaison entre ces deux loci, la variation génétique du marqueur est corrélée à celle du variant fonctionnel.

Chez les animaux d'élevage, les GWAS sont également utilisées pour localiser des mutations associées avec des traits quantitatifs présentant un intérêt économique pour les filières (p.ex. quantité de lait, taux protéique du lait, poids vif) à l'aide de différentes méthodes statistiques basées sur la mesure du déséquilibre de liaison entre le SNP et les variants causaux [Bush and Moore, 2012, Sharma et al., 2015].

Test alléliques

Allèles	D	d	Somme
Control	N(control- D)	N(control- d)	∑ Control
Case	N(case- D)	N(case- d)	∑ Case

Tests génotypiques

Génotypes	DD (.1)	Dd (.2)	dd (.3)	Somme
Control (1.)	N11	N12	N13	∑ Control
Case (2.)	N21	N22	N23	∑ Case

- Test exact de Fisher

$$p = \frac{\binom{\sum control}{N(control - D)}\binom{\sum case}{N(case - D)}}{\binom{\sum control + \sum case}{N(control - D) + N(Case - D)}}$$

Plink --fisher

- Test du χ^2

$$\chi^2 = \sum_{i=1}^{n} \frac{(Oi - Ei)^2}{Ei}$$

Plink --assoc
Plink –model-allelic

- Test de Cochran-Armitage

$$T = \sum_{i=1}^{k} t_i (N_{1i\cdot} \sum case - N_{2i\cdot} \sum control)$$

TREND		$t_1 = t_2 = t_3 = 1$	*--model-trend*
GENO	DD vs Dd vs dd	$t_1 = 0$ $t_2 = 1$ $t_3 = 2$	*–model-geno*
DOM	(DD, Dd) vs dd	$t_1 = t_2 = 1$ et $t_3 = 0$	*--model-dom*
REC	DD vs (Dd, dd)	$t_1 = 0$ et $t_2 = t_3 = 1$	*--model-rec*

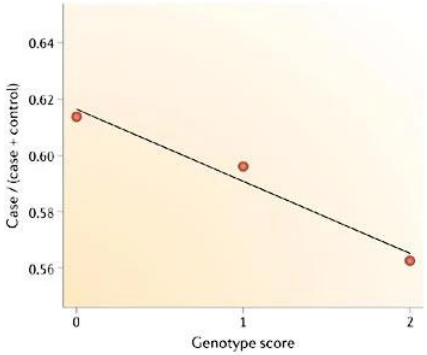


FIGURE 6.1 – Les trois principales méthodes de test d'association phénotype - génotype (GWAS) basés sur le calcul des fréquence allélique (Tests alléliques) et génotypique (Tests génotypiques).

6.3 Article : Genetic characterization of the local sheep breed of Mayotte island and identification of a locus associated with microtia near the HMX1 gene

(en préparation)

Jessica Magnier[1], Emmanuel Tillard[2] , Mélissa Ouvrard[3], Solène Raoul[3], Jérôme Janelle[2], *XXXXX*[4], *XXXX*[4] , Michel Naves[5], Mathieu Gautier[6], Laurence Flori[1]

[1] UMR SELMET, INRAE, CIRAD, Montpellier SupAgro, Univ. Montpellier, Montpellier, France

[2] UMR SELMET, INRAE, CIRAD, Montpellier SupAgro Univ., Saint Pierre, La Réunion, France

[3] Coopadem, Coconi, 97660 Ouangani, Mayotte

[4] MGX - Montpellier GenomiX, c/o Institut de Génomique Fonctionnelle, 141 rue de la cardonille, 34094 Montpellier Cedex 5, France

[5] URZ, INRAE, Petit-Bourg, France

[6] UMR CBGP, INRAE, CIRAD, IRD, Montpellier Supagro, Univ. Montpellier, Montferrier-sur-Lez, France

6.3.1 Abstract

We sampled 48 sheep belonging to the local sheep population on the island of Mayotte and genotyped on the ovine Illumina SNP50K bead chip. Exploratory analysis (PCA, hierarchical classification, calculation of genetic distances) of the 48 individuals showed that the population was genetically homogeneous and confirmed the genetic proximity of the sampled animals to the African long-fat-tailed sheep group. The average individual inbreeding was estimated at 30%, with one individual showing a very high inbreeding (60%) and very recent (Rk = 2, i.e. 1 generation). Two peaks of inbreeding were detected at the beginning of the 20th century and much more recently about ten years ago (analysis of recent inbreeding with the RZooRoh software). These increases in inbreeding could be consequences of infectious episodes or the introduction of sheep on the island of Mayotte. Several association tests detected a SNP associated with the reduced or atrophied ear phenotype (SNP OAR-115757345.1) at position

114 296 273 on chromosome 6, near the *HMX1* gene that is associated with microtia in Chinese Altay sheep.

6.3.2 Introduction

Documenting the diversity of livestock genetic resources by means of a genetic characterization of local livestock populations is a prerequisite to design strategies for their sustainable conservation. In addition, the follow-up of their recent inbreeding and the identification of mutations responsible for genetic disorders could help to limit and manage consanguinity. In Mayotte island, about a thousand of sedentary fat-tailed hair sheeps, well adapted to the tropical local conditions, are traditionally raised in small paddocks for meat purpose. Little is known about the history of this local breed. Its earliest presence was estimated in Dembeni site dating back to the 11-12th [Wright et al., 1984]. Other recent archeological findings in Acoua indicated skeletal evidence of small ruminant presence, tracing back to the 13th-15th centuries CE (Antsiraka Boira [Pauly and Ferrandis, 2018]). In addition, the few exchanges of males between farms suggest a potential high inbreeding, a major factor in emergence of defects, such as microtia. This latter that encompasses a spectrum of congenital abnormalities of the external ear with a wide range of severity, affects 20% of the Mayotte ovine population, that are short-eared or earless (Chapitre 4.3, [Jawasreh et al., 2016, Mastrangelo et al., 2018, He et al., 2020]. Microtia has also already been observed in several sheep populations from different geographical zones (i.e. in Awassi sheep from southwest Asia, originated in the Syro-Arabian desert, in Altay sheep from the Altay county in Xinjiang province and in Valle del Belice sheep from southern Italy), with a phenotype varying from ear atresia to different degrees of lenght reduction [Jawasreh et al., 2016, Mastrangelo et al., 2018, He et al., 2020]. Loci controlling microtia have been previously located on OAR23 with a significant association

near to *GATA-6* gene in Awassi breed [Jawasreh et al., 2016], and on OAR1 near the *CLRN1* gene in Valle del Belice breed [Mastrangelo et al., 2018]. If the underlying causal mutations have not been identified in these breeds, a 76 pb duplication in an evolutionary conserved region downstream of the *HMX1* gene on OAR6 has been described in Altay sheep [He et al., 2020]. A 76 pb duplication within the same conserved region has also been identified in Highland cattle with crop-ears [Koch et al., 2013]. Microtia was already observed in other species such as mouse, human and cattle [Koch et al., 2013, Gendron et al., 2016]. Genetic studies in murine models have identified different genes involved in microtia occurrence, e.g., *HOXA2, SIX 1* and *4*, *EYA1, TBX1, IRF6, TCOF1* and *HMX1* (see [Gendron et al., 2016] for review). The aim of this study is to explore the structure of the genetic diversity of Mayotte sheep breed, to estimate its recent inbreeding and to identify the mutation responsible for microtia by analysing 50K genotyping data and individual whole genome sequences.

6.3.3 Material and Methods

Animal sampling and ear phenotyping

A total number of 48 local sheeps of Mayotte (OMAY) including 9 males and 39 females, kept in 20 farms (2.4 animals per farm on average) located in 17 townships, spread over Grande Terre (Table S6.2), was considered in this study. The following meta-data were collected for each sheep, i.e. identification number, birth date or estimated age, sex, sire and dam identification numbers.

To define ear status, we measured the lenght of ears (in cm) and classified ears according to the shape. Ear types were defined based on quantitative (lenght of ears in cm) and qualitative (ear shape) criteria. More precisely, ears with a minimum length of 8 cm and a round end were considered as "normal" while ears with a minimum length of 5 cm and a triangular shape were

classified as "small" and complete ear atrophy as "atrophied". Figure 6.2 illustrates the three phenotypes differenciated in our sample. Animals with small ears or earless were considered as affected by microtia in genetic association analyses. A blood sample was collected on each of the 48 animals from tail vein using 10 ml EDTA vacutainer tubes, following recommendations of the directive 2010/63/EU, standard procedures and relevant guidelines of animal care. Genomic DNA were then extracted using the Wizard Genomic DNA purification kit (Promega, France) and stored at -20°C.

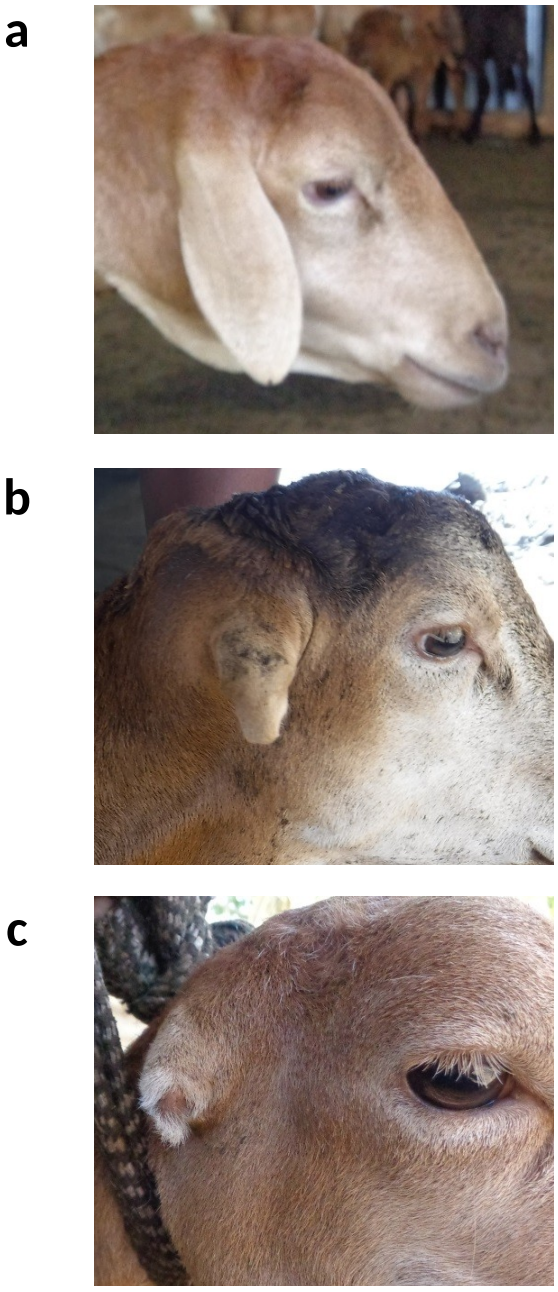


FIGURE 6.2 – Picture of the three external ear phenotypes encountered in the ovine population present in Mayotte island : **a** Normal ear, **b** Small ear, **c** Atrophied ear. (credits : Jessica Magnier)

TABLE 6.1 – Fixation and differentiation index calculated for the worldwide sheep breeds dataset. **Fst** correspond to calculated Fst between sheep from Mayotte (OMAY) and breeds of worldwide dataset. Abbreviations used in the column **Zone** correspond to the following geographic locations of the breeds : Europe (Eu), Asia (As), Africa (Af), America (Am), Oceania (Oc) and Mayotte island (May). Correspondance between abbreviations in the column **Pop** and the name of breeds is shown in Table S6.3

Pop	Zone	Fis	Fst	Pop	Zone	Fis	Fst	Pop	Zone	Fis	Fst	Pop	Zone	Fis	Fst
ADP	Af	-0.011	0.192	CHA	As	0.060	0.200	LIM	Eu	-0.001	0.240	RMN	Eu	0.006	0.230
AFS	ME	-0.020	0.183	CHI	ME	0.019	0.235	LME	Eu	0.016	0.202	ROM	Oc	0.023	0.253
ALT	Eu	0.027	0.210	CHL	Eu	0.003	0.247	OMAY	May	0.159	-	ROU	Eu	-0.002	0.263
APA	Oc	0.067	0.234	CHR	Eu	-0.001	0.267	MER	Eu	-0.003	0.209	ROV	As	0.016	0.255
APD	Oc	-0.003	0.253	COM	Eu	-0.009	0.198	MOG	ME	-0.001	0.170	RWE	Eu	-0.002	0.254
ASU	Oc	-0.001	0.209	COR	Eu	0.001	0.219	MOU	Eu	0.005	0.208	SAB	Eu	-0.049	0.248
AUM	Oc	0.031	0.194	CPW	Oc	-0.032	0.239	MTR	Eu	0.002	0.227	SBF	Eu	-0.002	0.226
BBB	Am	0.068	0.206	DSH	Eu	-0.066	0.320	NDZ	ME	-0.046	0.206	SBS	Eu	-0.007	0.238
BCS	Am	0.114	0.207	EFB	Eu	0.018	0.297	NQA	Af	-0.139	0.260	SKZ	Eu	-0.065	0.256
BDC	Eu	-0.007	0.300	EMZ	Af	0.013	0.146	NTX	Oc	0.008	0.264	SMS	Eu	-0.016	0.239
BGA	As	0.069	0.251	ERS	Eu	-0.010	0.216	NVE	Eu	-0.003	0.218	SPW	Eu	0.012	0.258
BGE	As	0.139	0.244	GAL	Eu	-0.003	0.266	OJA	Eu	0.000	0.197	STX	Eu	-0.053	0.265
BHM	Eu	0.029	0.260	GAR	As	0.023	0.266	OUE	Eu	0.175	0.352	SUF	Eu	-0.004	0.269
BMC	Eu	0.003	0.212	GTX	Eu	0.016	0.242	PAS	Eu	-0.004	0.211	SUM	As	0.037	0.241
BMN	Am	0.028	0.230	GUR	As	-0.004	0.207	QEZ	ME	0.015	0.161	SWA	Eu	-0.008	0.242
BRL	Eu	0.001	0.321	IDC	As	0.005	0.215	RAA	Eu	-0.002	0.189	TAR	Eu	0.016	0.213
BSI	Am	0.023	0.189	IDF	Eu	0.003	0.255	RAV	Eu	-0.005	0.224	TEX	Eu	0.005	0.268
CAS	Eu	-0.008	0.197	ISF	Eu	0.046	0.264	RDA	Af	-0.050	0.209	TIB	As	0.052	0.216
CDL	Eu	-0.003	0.245	KRS	ME	-0.021	0.196	RMA	Af	-0.004	0.113	VEN	Eu	0.005	0.265
CFT	ME	-0.035	0.234	LEC	Eu	0.070	0.205	RMB	Eu	-0.023	0.371	VRS	Eu	-0.030	0.278
												WIL	Eu	-0.007	0.344

Animal genotyping and genotyping data filtering

Fifty DNA samples selected based on geographic criteria, pedigree records and kinship were genotyped on the Illumina OvineSNP50 beadchip at Labogena plateform (Jouy-en-Josas) using standard procedures (www.illumina.com). Genotyping data were then included in the WIDDE database [Sempéré et al., 2015] to be combined with other publicly available ones using WIDDE utilities [Kijas et al., 2012]. A dataset including 2340 animals (with 48 OMAY individuals) from 81 sheep breeds (Table 6.1, See Table S6.3 for the correspondence between population acronymes and complete name of breeds), representative of the ovine worldwide genetic diversity was extracted from the database [Young et al., 2011, Kijas et al., 2012, Rochus et al., 2018] after applying several filters. A minimal individual genotyping call rate of 90% over all populations and 75% within each population (i.e. SNPs genotyped for less than 75% of the animals from at least one population were discarded) and a minimal SNP genotyping call rate set to 95% were applied. SNPs with a MAF below 0.01 or departing from Hardy-Weinberg equilibrium expectation (exact test p-value <0.001) were also discarded. A total of 36 902 SNPs distributed throughout the 26 autosomes passed all quality criteria, with the positions relative to OAR_v4.0/oviAri4 assembly [Jiang et al., 2014].

Exploratory analysis

The structure of the genetic diversity of OMAY sheep breed was explored by analyzing the filtered 50K dataset. Principal Component Analysis (PCA) based on individual SNP genotyping data was performed with the *smartpca* [Patterson et al., 2006] and visualized with the R package ggplot2 [Wickham, 2016]. Unsupervised genotype-based hierarchical clustering of the individual animal samples was carried out using the maximum-likelihood method implemented in ADMIXTURE 1.06 [Alexander et al., 2009]. Results were visualized with custom functions in

R environment (http ://www.r-project.org). Within population F_{IS} and between populations F_{ST} were estimated with custom R functions implementing the estimator derived by [Weir and Cockerham, 1984]. A neigbor-joining tree [Saitou and Nei, 1987] based on F_{ST} was computed with the APE [Paradis et al., 2006] and ggtree [Yu et al., 2017] R packages.

Recent inbreeding estimation

Individual levels of inbreeding were characterized at a global and local scale in OMAY sheep breed with the ZooRoH model [Druet and Gautier, 2017b] implemented in the RZooRoH package [Bertrand et al., 2019]. Using a multiple HBD-classes Hidden Markov Model, this method describes individual genomes as a mosaic of Homozygous-by-Descent (HBD) and non-HBD segments. Each HBD class is defined by its own rate Rk that determines the expected length (equal to 1/Rk Morgans) of the associated HBD segments and that is approximately equal to twice the number of generations to the common ancestor that transmitted the DNA segment. We considered a model with 8 HBD classes (with Rk equal to 2, 4, 8, 16, 32, 64, 128, 256) and one non-HBD class whose rate was equal to 256 for individuals genotyped on the 50K SNP chip. The model captured the contribution of ancestors living up to approximately 128 generations. HBD segments associated with HBD classes with higher rates were not captured with the marker density of the 50K SNP chip. The RZooRoH model estimated the contribution of each of the age-based classes of ancestors to the overall level of individual inbreeding.

Genome-wide association study with microtia

Genotyping data of 35 and 13 individuals with normal ear and microtia (small external ears or complete atrophia) respectively were included in GWAS. Several association tests were carried out using PLINK v1.90b6.6 software [Purcell et al., 2007], i.e. a standard case-control

association analysis was performed using the allelic Fisher's exact test and the allelic and genotypic χ^2 test and the genotypic Cochran-Armitage trend tests. A Bonferroni correction for multiple testing was also applied ($P_{Bonf} = pval * 39,570$, if $pval < 0.0001$). In addition, the association test based on mixed linear model (MLMA) implemented in GCTA software was performed after the calculation of the genetic relationship matrix (GRM) from all the autosomal SNPs. The Ovis aries v4.0 assembly OviAri4 was used to determine genes closed to SNP associated with microtia.

6.3.4 Results

Genetic characterization of the sheep breed of Mayotte

The OMAY sheep breed (N=48) was characterized by comparing the 50K genotyping data with genotyping data of 80 other sheep breeds representative of the *Ovis aries* genetic diversity (Table S6.3). The combined dataset consists of 2 340 animals genotyped for 36 902 SNPs. Using all available SNP information, a Principal Component Analysis (PCA) showed that the first three components (PC1, PC2 and PC3) explained 4.07, 1.93 and 1.50 % of the variation, respectively (Figure 6.4). A global organization of the sheep genetic diversity already described in [Kijas et al., 2012] was observed on the first factorial plan. Animals from the same breed clustered together. The combination of PC1, PC2, and PC3 separated individuals according to their geographic origin. PC1 separated European and Oceanian sheep apart from African, Asian, and South-West Asian animals. PC2 separated European Dorset populations from Texel European animals. PC3 identified admixed populations such as the African Dorper and breeds developed in South America and the Caribbean were positioned away from other clusters. The OMAY breed (in red) is positioned near several local African (i.e. RMA, NQA, EMZ, RDA), West-Asian (i.e. CFT, CHI, KRS, MOG, NDZ, QEZ) and Est-Asian (i.e. BGA, BGE, IDC,

GAR, GUR, IDC, SUM) breeds, far from European breeds (except SKZ, that is a Turkish breed) (Figure 6.4). Admixture results (Figure 6.5) revealed a geographic pattern separating European, American and Ocean breed on a first group, then Asian and populations from Middle East on a second group and African populations on a third group of which OMAY is one. Considering the F_{ST} estimation between OMAY and other breeds (Table 6.1, Figure S6.8) and of the neighbor joining (NJ) tree based on F_{ST} (Figure 6.3), the genetically closest sheep breeds (with F_{ST}=0.11 and 0.14, respectively) were Red Maasai (RMA) and Ethiopian Menz (EMZ). F_{IS} coefficient (i.e. 0.15) estimated for OMAY breed is among the highest F_{IS} across the sheep populations studied (Table 6.1). Its value is close to those estimated for Ouessant (F_{IS}=0.17), Bangladeshi East BGE (F_{IS}=0.13) and Brazilian Creole (F_{IS}=0.11) breeds (Table 6.1).

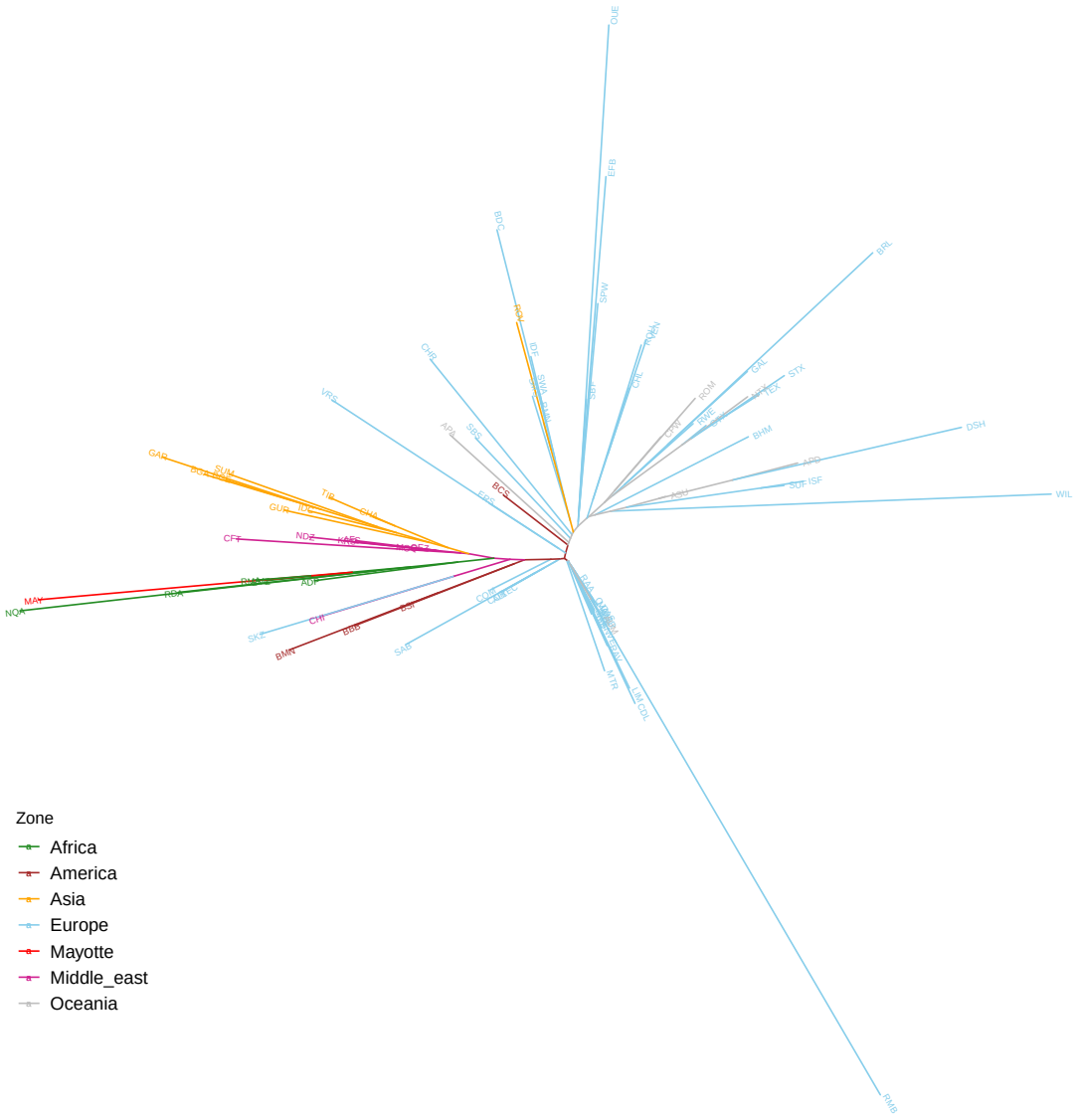


FIGURE 6.3 – Neighbor-joining tree based on pairwise *Fst* between each worldwide sheep breeds of the dataset. For the significance of breed abreviations, see Table S6.3

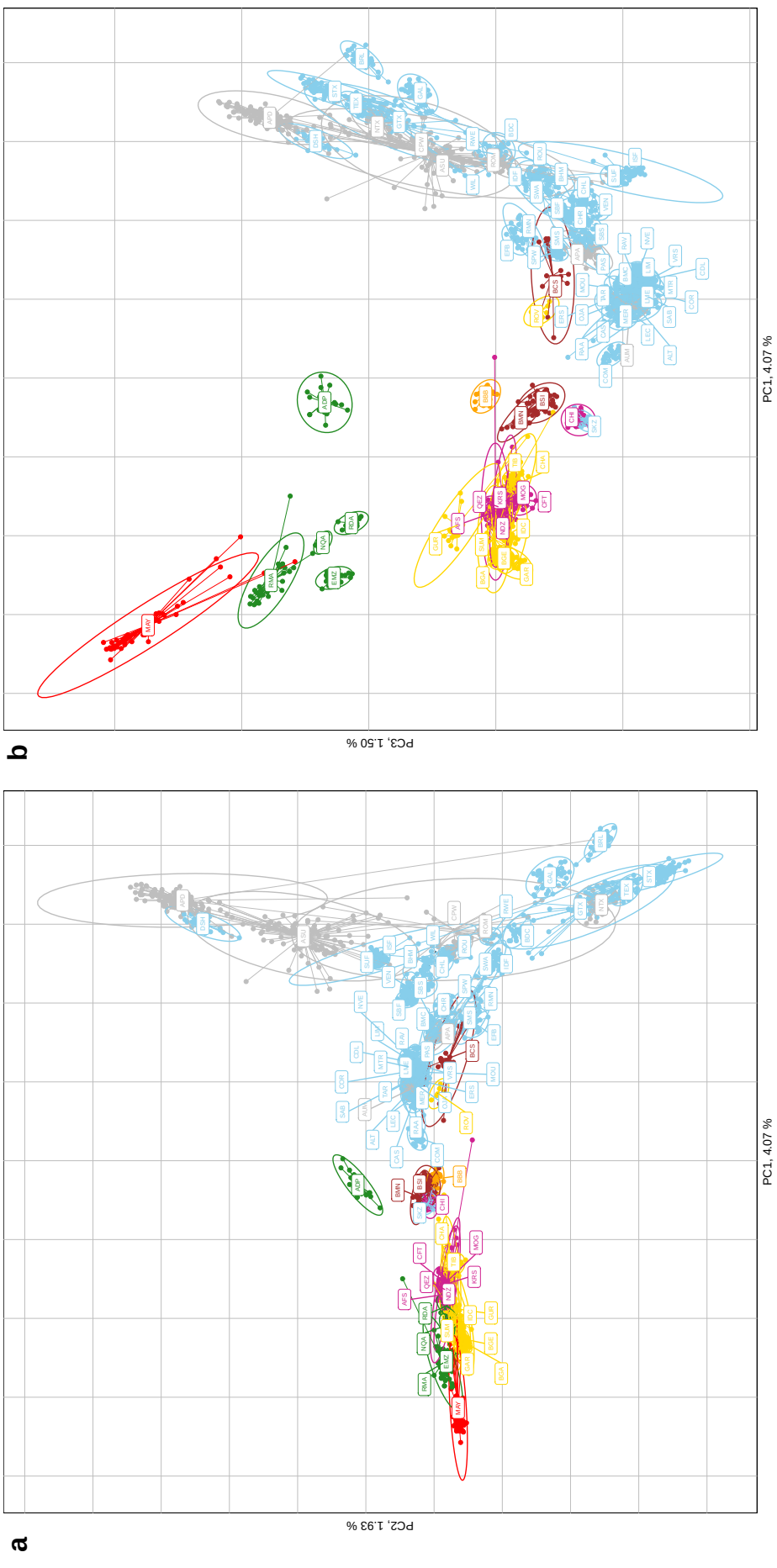


FIGURE 6.4 – Results of the principal component analysis (PCA) including 50K genotyping data (2340 animals from 81 breeds genotyped for 36 902 SNPs). The individuals according to their coordinates are plotted on the three first principal components (**a** PC1-PC2 , **b** PC1-PC3). Ellipses characterize the dispersion of each breed around its centre of gravity. The individuals are colored according to their geographic locations : Africa (green), Asia (yellow), Brazil (brown), Caraïbes (orange), Europe (blue), Mayotte (red), Middle East (purple) and Oceania (grey).

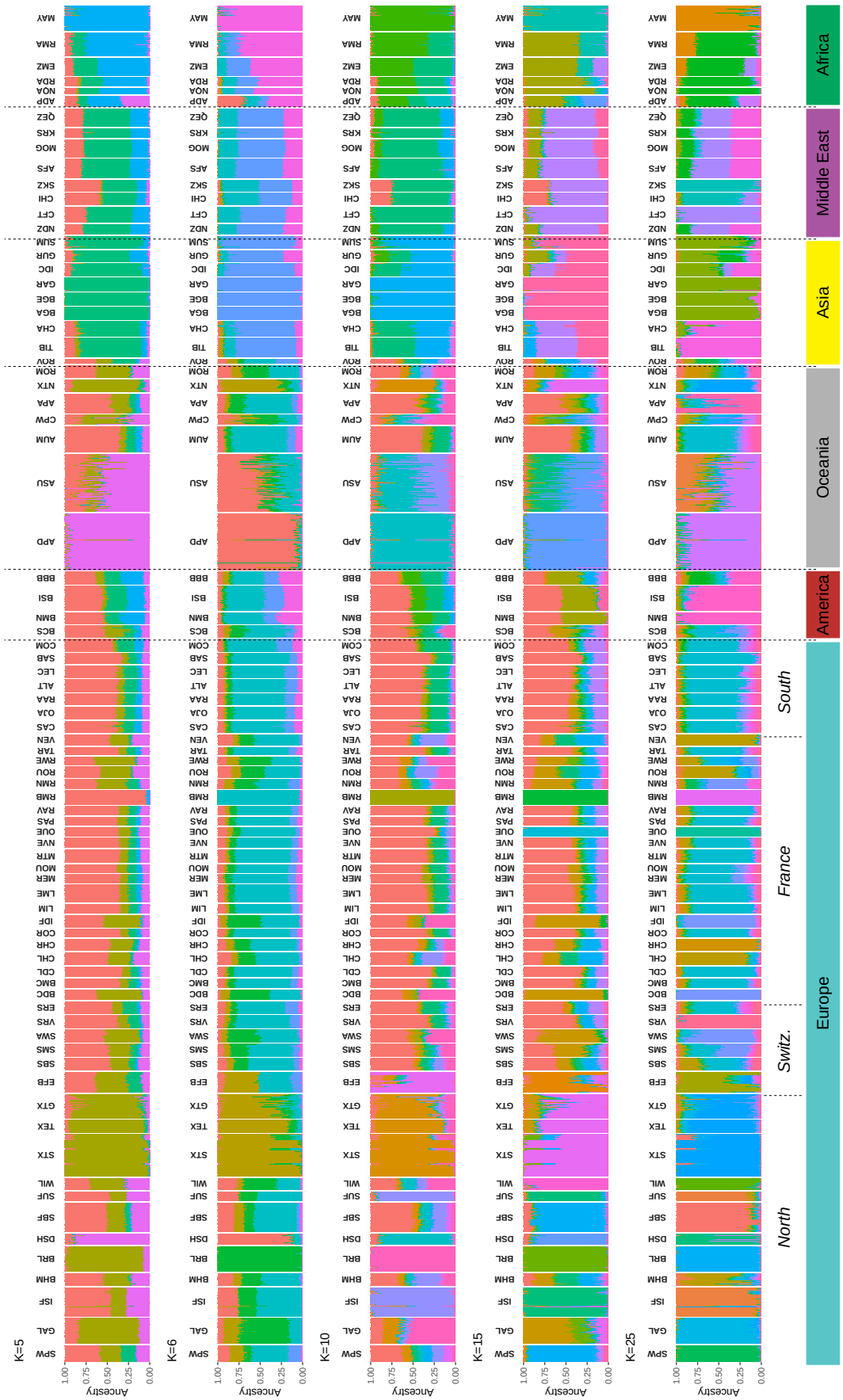


FIGURE 6.5 – Results of unsupervised hierarchical clustering results including 2340 animals from 81 breeds genotyped for 36,902 SNPs, with an inferred number of clusters $K = 5$, *6, 10, 15 and 25* calculated with Admixture 1.04 software. For each individual, proportions of each cluster (y axis) are plotted.

Characterization of individual levels of inbreeding

In order to characterized the individual levels of recent inbreeding at both a global (genome-wide) and local (locus-specific) scale, we estimated the individual inbreeding coefficients in OMAY population using 50K genotyping data and applying the method implemented in RZoo-RoH software [Druet and Gautier, 2017b, Bertrand et al., 2019]. The distribution of these coefficients is detailed in Figure 6.6. The average individual inbreeding levels in OMAY population was found equal to 0.309 (SD= 0.133). However, the estimated individual levels of inbreeding revealed that three OMAY individuals had 50% of their genome classified has HBD. One individual (i.e.#17) presented more than 60 % of its genome in HBD segments. Most HBD segments were concentrated in classes with Rk equal to 32 and 64 (i.e. 16 and 32 generations before sampling, corresponding to 64 and 128 years ago assuming a 4 years generation time). However some OMAY individuals have long HBD segments that belonged to HBD classes with lower Rk (equal to 4 and 8, corresponding to 2 or 4 generations, 8 to 16 years before sampling) although the average inbreeding levels were low for these HBD classes.

Identification of mutation responsible for microtia

We performed four genome-wide association tests and detected the most significantly associated SNPs (OAR6-115757345.1 ;rs421339551) on OAR6 at position 114 296 273 bp (a $p_{BONF}val = 0.003$ for the Fisher exact allelic test, a $p_{BONF}val = 0.0006$ for the χ^2 allelic and genotypic tests and for the Cochran-Armitage genotypic trend test and $p_{BONF}val = 0.026$ for the MLMA test ; Figure 6.7). This SNP is located within 22.5 kb of the closest gene, H6 family homeobox 1 (*HMX1*). Another significant SNP (i.e. s19659.1 ;rs416186858) located on OAR6 within *HMX1* was found with the χ^2 allelic test ($P_{BONF} < 0.01$), the Cochran-Armitage allelic test ($P_{BONF} < 0.01$), and the Fisher exact allelics test ($P_{BONF} < 0.05$)

.

6.3.5 Discussion

This study provides for the first time a fine-scale genetic characterization of the sheep breed from Mayotte by comparing OMAY SNP genotyping data with those previously obtained for several other breeds representative of the major breed groups of sheeps globally distributed [Kijas et al., 2012]. The OMAY individuals, sampled in different farms of Grande Terre, are representative of the whole population. No clear sub-structure are observed among OMAY individuals, suggesting an homogeneous population. Based on PCA and clustering results, OMAY is closed to some African breeds, i.e. the Red Masai from Kenya, the Ethiopian Menz and the Namaqua Afrikaner, a local breed of South Africa. These breeds are also fat-tailed sheep [Porter et al., 2016, Pourlis, 2011], a trait previously characterized in Mayotte population. PCA, clustering and Fst results revealed that OMAY was also genetically nearest from Asian and Middle East populations than European, American and Oceanian sheep breeds. These results suggested an African origin of OMAY breed which is consistent with historical data reporting an ancient presence on the island of Mayotte [Wright et al., 1984, Pauly and Ferrandis, 2018] and fat-tailed sheep in Madagascar [Anonymous, 1950]. It is the first ovine population genotyped in islands of the West Indian Ocean. Characterizing other ovine populations found in Madagascar and other Comoros islands could give more information on the structure of the sheep breeds genetic diversity in this area. The relative high amount of F_{is} compared to the other populations of the dataset indicates an heterozygote deficit that could be due to subpopulation structure, admixture or inbreeding and reveals a lower genetic diversity compared to some other breeds. In Mayotte, sheep are generally raised jointly with goat in weak proportion in small paddock and transhumance is not conducted due to landed

issues. Few exchange of male individuals, in order to mating, occured between the different farms, explaining the diversity level of the OMAY population. The characterization of recent inbreeding of OMAY population revealed a global level of inbreeding of 30 % with a low very recent inbreeding median level (1 generation ago) and a higher proportion of the genome in HBD classes 2, 4, 16 and 32 generations ago (corresponding approximately to 8, 16, 64 and 128 years ago), the oldest inbreeding observed tracing back to the end of the 19th century and the middle of the 20th century (Figure 6.6). Unfortunatly, no clear historical information (related to a decrease of population size) is available to explain these inbreeding increases. But we could suppose that they could derive from infectious episodes or sheep introduction on the islands. Inbreeding in this population could be related to the presence of microtia in some individuals.

By grouping OMAY individuals presenting short external ears or complete atrophia, we detected an association between two SNPs close to the *HMX1* gene and microtia. Mutations in the *HMX1* gene that is involved in the development of craniofacial structures [Munroe et al., 2009] were reported to cause malformations of the external ear in cattle, sheep, mice and humans. Indeed a non-coding 76 bp genomic duplication located in the conserved region near this gene was found associated with congenital crop ears in Highland cattle breed with a monogenic autosomal dominant mode of inheritance and variable expressivity [Koch et al., 2013]. A duplication of 76 bp, which shares an interval 53 bp sequence with the 76 bp duplication identified in cattle was associated with microtia in Altay sheep breed from the county in Xinjiang province in China, suggesting a dominant inheritance mode [He et al., 2020]. In mice, the recessive causative mutation of the ear abnormality in dumbo rats is a 5777 bp deletion located 79 kb downstream of *HMX1* and that is associated with a loss of *HMX1* expression in the craniofacial mesenchyme during development [Munroe et al.,

2009]. In human, recessive loss of function mutations in *HMX1* have been associated with oculoauricular syndrome (OAS, OMIM 142992) characterized by malformation of the external ear and eyes [Schorderet et al., 2008] and, recently a 94.6 Kb duplication in an intergenic region between *HMX1* gene and *CPZ* gene was detected in Chinese families with isolated bilateral concha-type microtia and involves the $\sim$ 600 bp evolutionarily conserved region downstream of *HMX1* (HMX1-ECR) [Si et al., 2020]. *HMX1* gene is then a positional and functional candidate gene involved in microtia in sheeps of Mayotte.

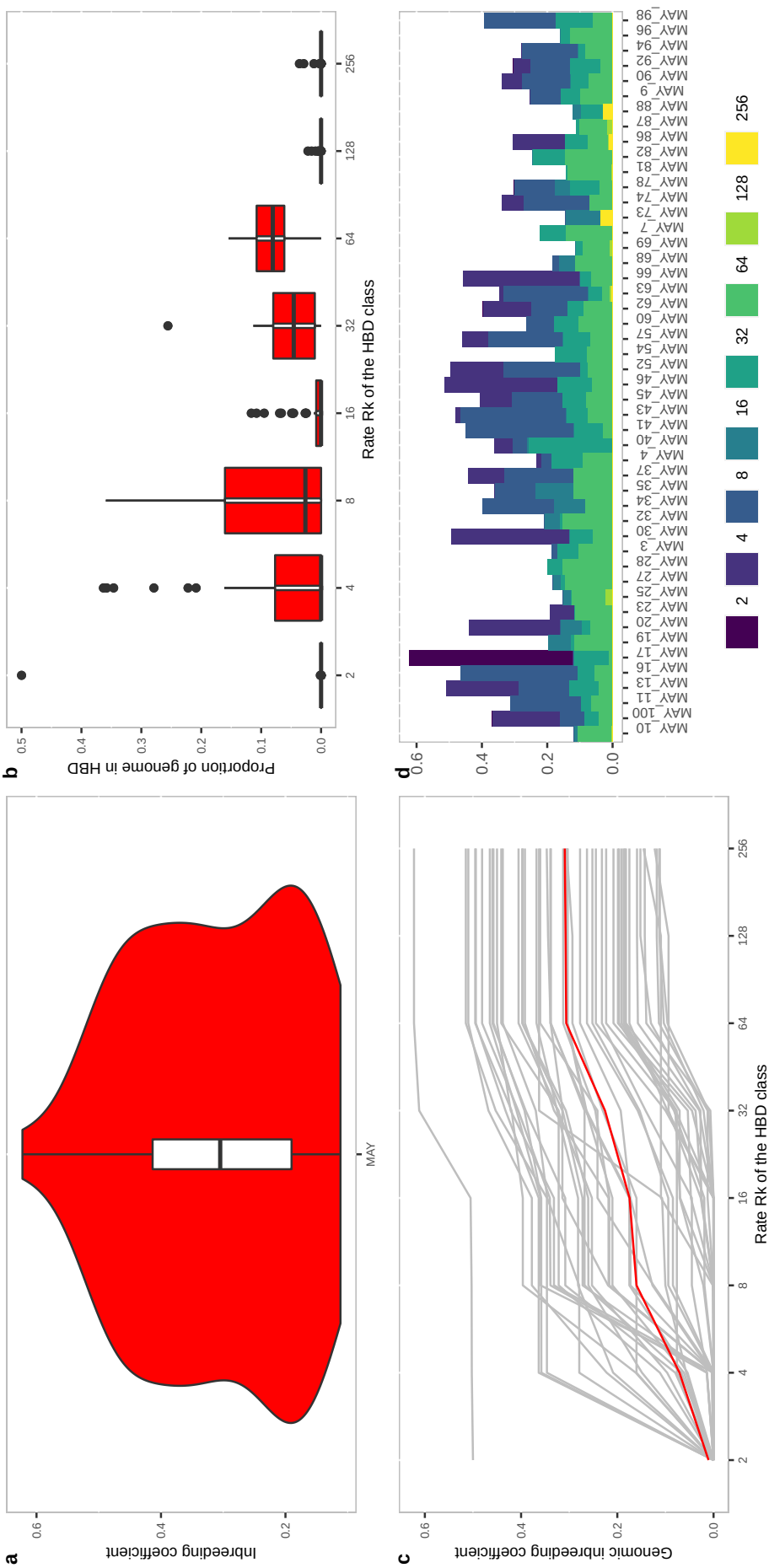


FIGURE 6.6 – Characterization of individual inbreeding levels in ovine population from Mayotte island (n=48). **a** Violin plot represents the distribution of inbreeding coefficients. **b** Average proportion of individual genomes associated with different HBD classes. **c** Cumulated average proportion of individual genomes associated with different HBD classes. Each individual is represented by a grey line. The average value of the population is added in red. **d** Partitioning of individual genomes in different HBD classes. Each bar represents an individual and its total height the total level of inbreeding. The height of different stacks, which appear in different colors, represents the proportion of the genome associated with each HBD class, defined by their rate R_k.

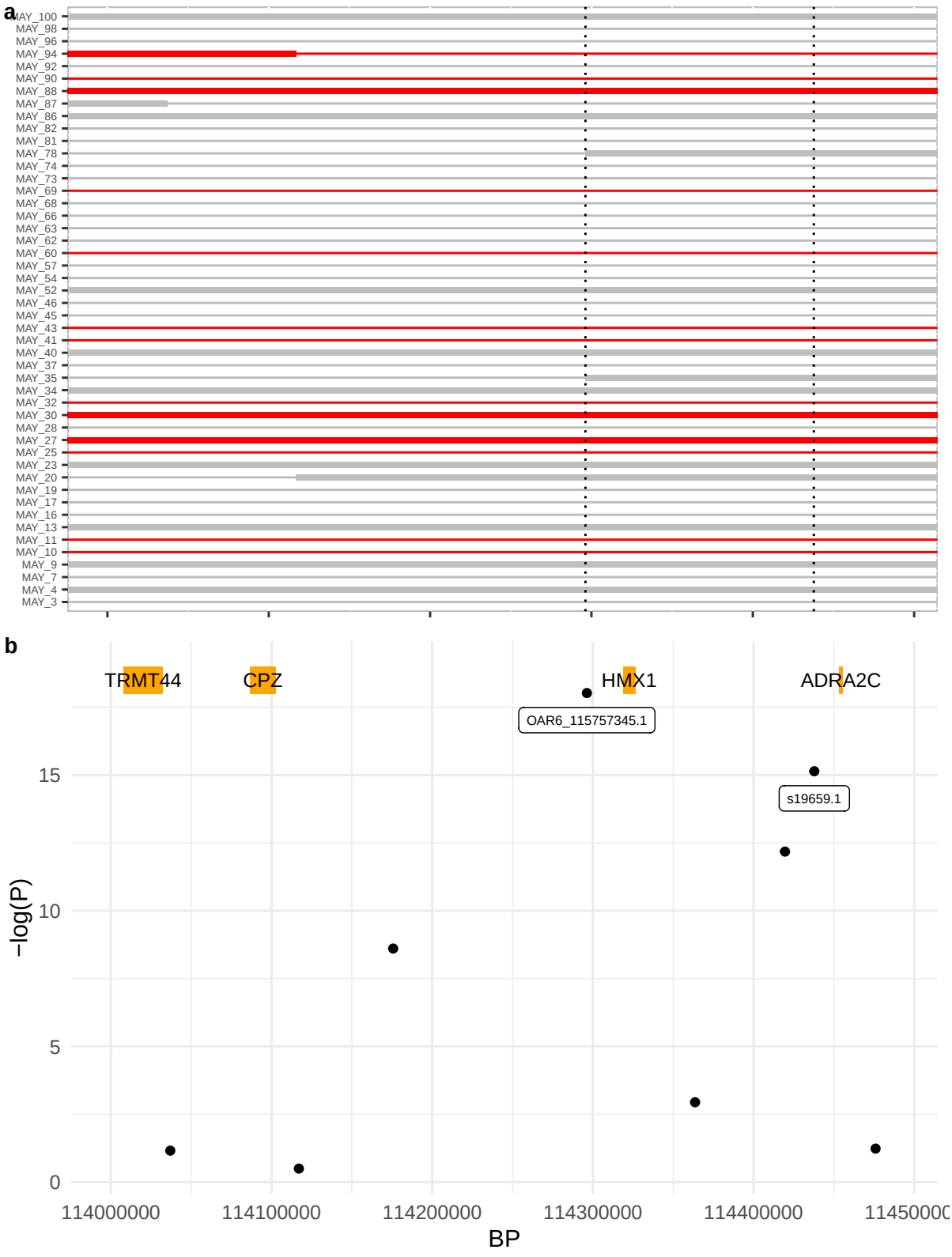


FIGURE 6.7 – Focus on the OAR6 region concerned by microtia mutation. **a** HBD segments calculated with RZooRoh package on the 48 ovine ind. genotyped on 50K ovine beadchip. The larger lines correspond to HBD segments. The color corresponds to the observed phenotype : complete atrophia and small ears (red) and normal ears (grey). The dotted lines correspond to significative SNPS positions. **b** Manhattan plot focused on the region of OAR6 associated with the microtia phenotype with χ^2 genotypic test in the sheep breed of Mayotte. The labeled SNPs present a value of -log(Pval) $>$ 15.

6.3.6 Supplementary Tables and Figures

TABLE S6.2 – The 48 individuals genotyped on Illumina ovine 50K SNP bead chip in sampled population of sheep from Mayotte.

ID	Sex	Brith_date	Age	Visit_date	Farm	Town	Class	Meas.
DAOV003	Female	17/07/2013		09/10/2017	E179	Tsingoni	normal	10
DAOV004	Female			09/10/2017	E179	Tsingoni	normal	10
DAOV007	Female	23/10/2009		09/10/2017	E179	Tsingoni	normal	11
DAOV009	Female	05/03/2014		09/10/2017	E179	Tsingoni	normal	10
DAOV010	Male		1	09/10/2017	E179	Tsingoni	small	5.5
DAOV011	Female		6	10/10/2017	E180	Kani keli	small	7
DAOV013	Female			10/10/2017	E180	Kani keli	normal	9.5
DAOV016	Female		3	10/10/2017	E181	Boueni	normal	9.5
DAOV017	Female		4	10/10/2017	E181	Boueni	normal	10
DAOV019	Female	01/01/2019		10/10/2017	E182	Bandrele	normal	11
DAOV020	Female		3	10/10/2017	E182	Bandrele	normal	12
DAOV023	Female		4	10/10/2017	E182	Bandrele	normal	10
DAOV025	Female		3	10/10/2017	E182	Bandrele	small	5.5
DAOV027	Female		4	10/10/2017	E182	Bandrele	atrophied	2
DAOV028	Male		4	10/10/2017	E182	Bandrele	normal	10
DAOV030	Female		4	11/10/2017	E183	Mtsamboro	atrophied	2
DAOV032	Female	08/09/2009		11/10/2017	E183	Mtsamboro	small	5
DAOV034	Male		1	11/10/2017	E183	Mtsamboro	normal	9
DAOV035	Female			11/10/2017	E184	Bandraboua	normal	10
DAOV037	Male			11/10/2017	E184	Bandraboua	normal	8
DAOV040	Female	23/12/2014		11/10/2017	E185	Acoua	normal	12
DAOV041	Female		5	11/10/2017	E185	Acoua	atrophied	4
DAOV043	Male		2	11/10/2017	E185	Acoua	atrophied	5
DAOV045	Female		3	12/10/2017	E186	Moinatrindri	normal	10
DAOV046	Female		7	12/10/2017	E186	Moinatrindri	normal	10.5
DAOV052	Female		7	12/10/2017	E186	Moinatrindri	normal	11
DAOV054	Female		2	12/10/2017	E186	Moinatrindri	normal	13.5
DAOV057	Female		2	12/10/2017	E187	Bandrele	normal	11
DAOV060	Female		5	13/10/2017	E188	Ouangani	atrophied	6
DAOV062	Female		1	13/10/2017	E188	Ouangani	normal	10.5
DAOV063	Female		4	13/10/2017	E192	Chiconi	normal	10
DAOV066	Female	11/11/2013		17/10/2017	E189	Tsararano	normal	11.5
DAOV068	Male		2	17/10/2017	E190	Tsoundzou 1	normal	11
DAOV069	Female		3	17/10/2017	E190	Tsoundzou 1	small	7
DAOV073	Female		3	17/10/2017	E191	Vahibé	normal	13
DAOV074	Female		3	17/10/2017	E192	Chiconi	normal	10
DAOV078	Female		5	17/10/2017	E193	Ouangani	normal	10
DAOV081	Male		5	17/10/2017	E193	Ouangani	normal	8
DAOV082	Female			18/10/2017	E194	Chiconi	normal	12
DAOV086	Female	18/07/2014		18/10/2017	E195	Sada	normal	12
DAOV087	Female	30/01/2013		18/10/2017	E195	Sada	normal	12
DAOV088	Female		4	18/10/2017	E195	Sada	atrophied	2
DAOV090	Female		5	18/10/2017	E196	Iloni	small	7
DAOV092	Female		4	18/10/2017	E196	Iloni	normal	10
DAOV094	Male		3	18/10/2017	E196	Iloni	small	7
DAOV096	Female		3	18/10/2017	E197	Mronabeja	normal	14
DAOV098	Male	20/01/2010		18/10/2017	E197	Mronabeja	normal	11.5
DAOV100	Female		6	18/10/2017	E198	Mronabeja	normal	12

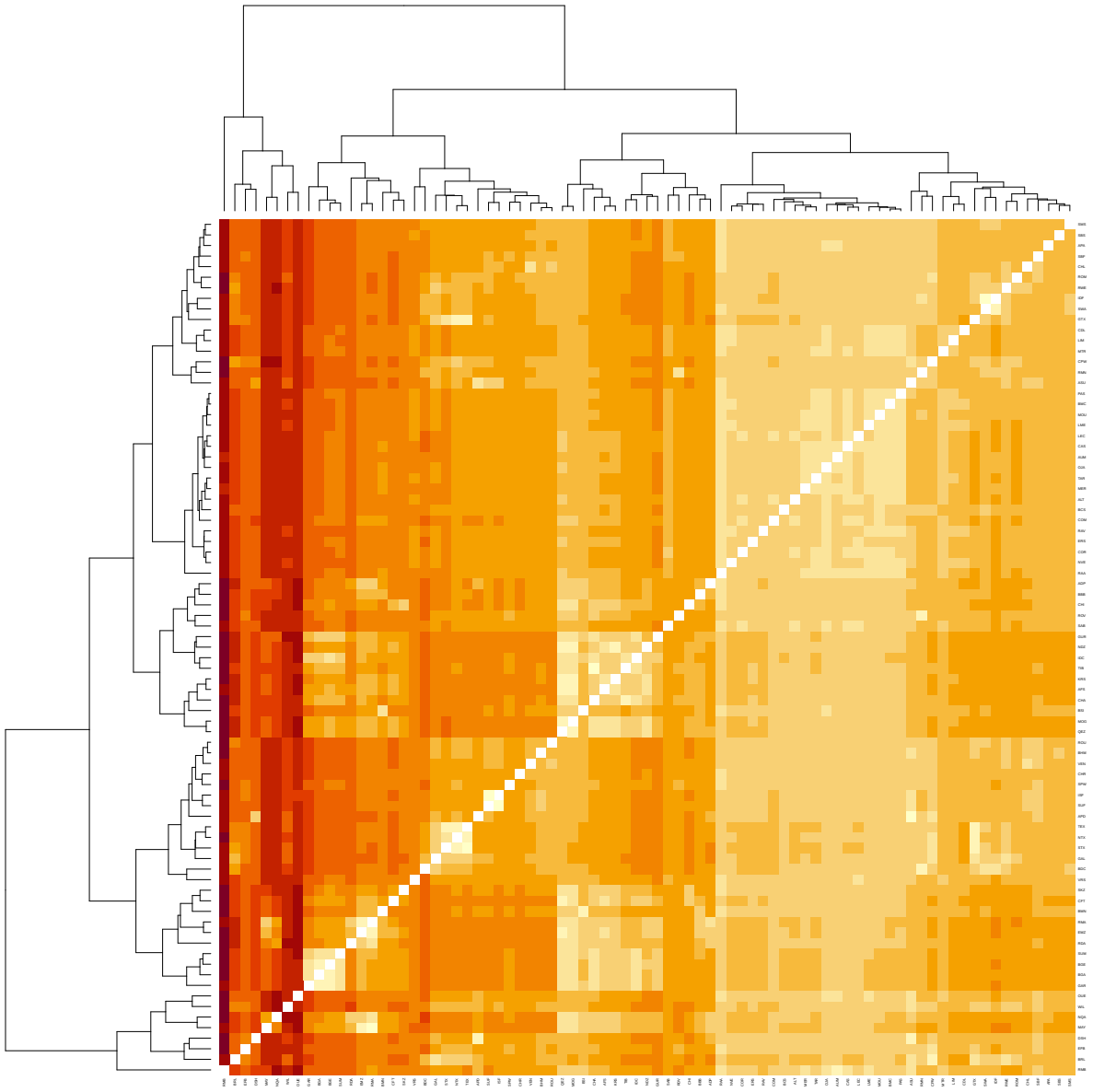


FIGURE S6.8 – Heatmap representation of Fst values between the 81 ovine populations each other.

TABLE S6.3 – Description of the 81 breeds used in exploratory analysis.

Pop	Name	Country	Zone
ADP	African Dorper	South Africa	Africa
AFS	Afshari	Iran	Middle East
ALT	Altamurana	Italy	Europe
APA	Arapawa	New_Zealand	Oceania
APD	Australian Poll Dorset	Australia	Oceania
ASU	Australian Suffolk	Australia	Oceania
AUM	Australian Merino	Australia	Oceania
BBB	Barbados Black Belly	Caraïbes	America
BCS	Brazilian Creole	Brazil	America
BDC	Berrichon du Cher	France	Europe
BGA	Bangladeshi Garole	Bangladesh	Asia
BGE	Bangladeshi East BGE	Bangladesh	Asia
BHM	Black Headed Mountain	UK	Europe
BMC	Blanc du Massif Central	France	Europe
BMN	Morada Nova	Brazil	America
BRL	Border Leicester	UK	Europe
BSI	Santa Ines	Brazil	America
CAS	Castellana	Espagne	Europe
CDL	Causses du Lot	France	Europe
CFT	Cyprus Fat tail	Chypre	Middle East
CHA	Changthangi	Jammu-et-cachemire	Asia
CHI	Chios	Greece	Middle East
CHL	Mouton Charollais	France	Europe
CHR	Charmoise	France	Europe
COM	Comisana	Sicile	Europe
COR	Corse	France	Europe
CPW	Australian Coopworth	Australia	Oceania
DSH	Dorset Horn	UK	Europe
EFB	East Friesian Brown	Germany	Europe
EMZ	Ethiopian Menz	Ethiopia	Africa
ERS	Engadine Red Sheep	Sundarban	Europe
GAL	Galway	Irland	Europe
GAR	Indian Garole	India	Asia
GTX	German Texel	Germany	Europe
GUR	Garut	Indonesia (West Java)	Asia
IDC	Deccani	India	Asia
IDF	Île de France	France	Europe
ISF	Irish Suffolk	Irland	Europe
KRS	Karakas	Turquia	Middle East

Pop	Name	Country	Zone
LEC	Leccese	Italy	Europe
LIM	Limousine	France	Europe
LME	Lacaune (meat)	France	Europe
MER	Mérinos d'Arles	France	Europe
MOG	Moghani	Iran	Middle East
MOU	Mourerous	France	Europe
MTR	Manech Tête Rouge	France	Europe
NDZ	Norduz	East Anatolia	Middle East
NQA	Namaqua Afrikaner	South Africa	Africa
NTX	NewZealand Texel	New_Zealand	Oceania
NVE	Noire du Velay	France	Europe
OJA	Ojalada	Espagne	Europe
OMAY	Mahorais	Mayotte	Mayotte
OUE	Ouessant	France	Europe
PAS	Préalpes du Sud	France	Europe
QEZ	Qezel	Turquia	Middle East
RAA	Rasaaragonesa	Espagne	Europe
RAV	Rava	France	Europe
RDA	Ronderib Afrikaner	South Africa	Africa
RMA	Red Maasai	Kenya, Tanzania, Uganda	Africa
RMB	Mérinos de Rambouillet	France	Europe
RMN	Romane	France	Europe
ROM	NewZealand Romney	New_Zealand	Oceania
ROU	Roussin de la Hague	France	Europe
ROV	Romanov	Russia	Asia
RWE	Rouge de l'Ouest	France	Europe
SAB	Sardinian Ancestral Black	Sardaigne	Europe
SBF	Scottish Blackface	UK	Europe
SBS	Switzerland Black-Brown Mountain	Switzerland	Europe
SKZ	Sakiz	Turquia	Europe
SMS	Switzerland Mirror	Switzerland	Europe
SPW	Spael-white	Norway	Europe
STX	Scottish Texel	UK	Europe
SUF	Suffolk	UK	Europe
SUM	Sumatra	Indonesia (Sumatra)	Asia
SWA	Switzerland White Alpine	Switzerland	Europe
TAR	Tarasconnaise	France	Europe
TEX	Texel	Netherlands	Europe
TIB	Tibetan	Tibet	Asia
VEN	Mouton Vendéen	France	Europe
VRS	Valais Red	Switzerland	Europe
WIL	Wiltshire	UK	Europe

6.4 Résultats supplémentaires

Les résultats présentés dans cette section nécessitent une analyse plus approfondie ultérieure, notamment pour la partie ci-après concernant la recherche des variants causaux de la mutation auriculaire chez les moutons mahorais. Nous présentons dans un premier temps les résultats obtenus pour des séquences mappées sur l'assemblage de référence OAR3.1. Ensuite, nous avons conduit une analyse de signature de sélection chez le mouton mahorais et des gènes associés. Cette liste de gène sera par la suite précisée au regard des résultats ultérieurement obtenus de l'inférence de l'histoire démographique de la population de moutons mahorais.

6.4.1 Identification de la mutation candidate

High-throughput sequencing of sheeps with normal ear and microtia

The whole genome of eight sheeps including three earless, three small ear and two normal ears animals was sequenced on the MGX plateform (Biocampus Montpellier, France). The Illumina TruSeq DNA sample preparation kit (FC-121-2001, Illumina Inc., San Diego, CA) was used according to the manufacturer's protocol. Libraries were then validated on a DNA1000 chip on a Bioanalyzer (Agilent) to determine size and quantified by qPCR using the Kapa library quantification kit to determine concentration. The cluster generation process was performed on cBot (Illumina Inc.) using the Illumina Paired-End DNA sample preparation kit (FC-102-1001, Illumina Inc.). Individuals were further paired-end sequenced on the HiSeq 2000 (Illumina Inc.) using the Sequence by Synthesis technique. Base calling was achieved by the RTA software (Illumina Inc.). Reads of 100 bp from both sides of the fragments were thus obtained after this step. Quality control of the sequences was checked using the FastQC software.

Location of the mutation responsible for microtia

In order to identify the causative mutation located near HMX1 in the sheep breed of Mayotte, we searched for mutations carried by the 6 animals with microtia in the vicinity of this gene and of the SNP OAR-115757345.1 associated with microtia (by comparison with the sequences of the two animals with normal ears and reference sequences). Using the LUMPY software to discover structural variant, we identified a potential candidate duplication of 173 bp located within the second intron of the predicted gene ENSOARG00000013314.1. Individuals with earless or short ears had homozygous genotypes for this variant unlike animals with normal ears, which had heterozygous genotypes, suggesting a dominant mode of inheritance with probably a variable expressivity. This variant was not detected in the reference assembly (OAR3.1). Since the annotation of this region on this assembly is only partial, it appears necessary to use oviAri4 in the assembly to confirm these initial results and better localize this duplication with respect to the HMX1 gene.

We could then check the presence of this variant on larger number of animals suffering from microtia by comparison with animals with normal external ears by performing a genotyping test based on PCR amplification and its association with microtia.

6.4.2 Détection de signatures de sélection dans le génome des ovins mahorais et annotation fonctionnelle des gènes candidats identifiés

Afin de compléter la caractérisation génétique de la population ovine de Mayotte et d'identifier les loci impliqués dans l'adaptation aux conditions spécifiques de l'île de Mayotte, nous avons entrepris de rechercher des signatures de sélection dans le génome de ces animaux, de proposer une liste de gènes candidats et de réaliser une annotation fonctionnelle de ces gènes en suivant l'approche déjà appliquée au Zébu mahorais.

Material and methods

Detection of selection footprints Genotyping data from nine ovine populations including sheep breed of Mayotte (OMAY), three West-Asian (WAS) breeds (i.e. Moghani (MOG), Norduz (NDZ) and Qezel (QEZ) breeds), three East-Asian breeds (EAS) (i.e. Bangladeshi Garole (BGA), Bangladeshi East (BGE) and Garole (GAR) breeds) and two African (AFO) breeds (i.e. Red Masaai (RMA) and Ethiopian Menz (EMZ) breeds) were extracted from the previous dataset used to genetically characterized the OMAY population. This new dataset included 290 individuals genotyped for 21'305 SNPs. SNP positions on oviAri4 assembly positions were considered. Haplotypes of OMAY and each EAS, AFO and WAS ovine populations were reconstructed with FASTPHASE 1.4 [Scheet and Stephens, 2006]. We used population label information to estimate phased haplotype background and the following options : *-Ku50 -Kl9 -Ki9*. Using haplotype information, we further computed *iHS* and *Rsb* scores with REHH3.2.0 package [Voight et al., 2006, Tang et al., 2007, Gautier and Vitalis, 2012, Gautier et al., 2017]. These statistics are designed to detect regions with high level of haplotype homozygosity over an unexpected distance (relative to neutral expectations), within populations (*iHS*) or across populations (*Rsb*). For convenience, *iHS* scores were transformed into $p_{iHS} = -log[1 - 2|\Phi(iHS) - 0.5]$ where $\Phi(x)$ represents the Gaussian cumulative function. Assuming *iHS* scores are normally distributed under neutrality, p_{iHS} might thus be interpreted as $log_{10}(1/P)$ where *P* is the two-sided *P*-value associated with the neutral hypothesis. *Rsb* was computed for the comparisons of OMAY with its candidate parental breeds by standardizing the ratio of the corresponding ancestral cluster (WAS, EAS, AFO) *iES* and the OMAY *iES* [Tang et al., 2007]. Scores were transformed into $p_{Rsb} = -log[\Phi(Rsb)]$. As for *iHS*, assuming *Rsb* are normally distributed under neutrality assumption, p_{Rsb} might be interpreted as $log_{10}(1/P)$ where *P* is the one-sided *P*-value associated with the neutral hypothesis.

TABLE S6.4 – Populations used for the identification of footprints of selection based on EHH-derived tests

ID	Population	Tail	Geographic group	Nb	Reference
BGA	Bangladeshi Garole	Thin	East Asia	24	Kijas et al., 2012
BGE	Bangladeshi	Thin	East Asia	24	Kijas et al., 2012
GAR	Garole	Thin	East Asia	26	Kijas et al., 2012
RMA	Red Masaai	Fat	Africa	45	Kijas et al., 2012
EMZ	Ethiopian Menz	Fat	Africa	34	Kijas et al., 2012
MOG	Moghani	Fat	West Asia	34	Kijas et al., 2012
NDZ	Norduz	Fat	West Asia	20	Kijas et al., 2012
QEZ	Qezel	Fat	West Asia	35	Kijas et al., 2012
OMAY	Sheep from Mayotte	Fat	Mayotte	48	This study

Identification and annotation of candidate genes under selection In order to identify candidate genes associated with selection footprints, following approach proposed by [Flori et al., 2019], all the genotyped SNPs were first annotated using as a gene set reference a list of 217,848 otherRefSeqGenes mapped on oviAri4 assembly that is known protein-coding and non-protein-coding genes for organisms other than cow or sheep, taken from the NCBI RNA reference sequences collection (RefSeq), and that was downloaded from the UCSC xenoRefGene database (https ://hgdownload.soe.ucsc.edu/goldenPath/oviAri4/database/)

A SNP was then considered as representative of one of these otherRefSeq genes if localized within its boundary positions extended by 15 kb up- and downstream to account for persistence of Linkage Desequilibrium [Gautier et al., 2007]. Among the 36 902 SNPs from the dataset included in the analysis, 21 305 SNPs mapped to 83'970 different orthologs RefSeqGenes (from 1 to 38, median = 1) corresponding to 58'363 Gene Symbols while each RefSeqGene was represented by 1.9 SNP (from 1 to 47, median = 1). Genes which included at least **one** SNP displaying p_{iHS} or $p_{Rsb} > \mathbf{3}$ were considered as candidate genes.

Results and discussion

The Manhattan plots given in Figure S6.9 correspond to the four differents tests, *iHS* within OMAY and *Rsb* for OMAYvsWAS, OMAYvsEAS and OMAYvsAFO. According to our criteria, 1, 2, 3 and 10 candidate genes (Table S6.5) displayed significant footprints of selection for iHS_{MAY}, $Rsb_{MAYvsWAS}$, $Rsb_{MAYvsEAS}$ and $Rsb_{MAYvsAFO}$ respectively. Hence, the candidates gene are involved in nervous system function, immune response, metabolism, reproduction system. A particular gene was involved in pigmentation *(ZBBX)* which could be relied to skin properties.

Nervous system

Five genes (i.e. *ZBBX, BIN1, MFSD6, DPP6, ERC2*) were involved in nervous system. Indeed, the gene *ZBBX* (Zinc Finger B-Box Domain-Containing Protein 1) was shown associated with response to paliperidone in schizophrenia [Li et al., 2017] and iris colour [Jonnalagadda et al., 2019]. The gene *BIN1* has been demonstrated to contribute to Alzheimer's disease

particularly in East Asian and Chinese populations [Liu et al., 2013, Tan et al., 2014, Salcedo-Tacuma et al., 2019]. The gene *MFSD6* (major facilitator superfamily domain containing 6) which is associated with the three *Rsb* scores has a role in regulating neural circuit activity in synapses [McCulloch et al., 2017]. Mutations in *MFSD6* gene are associated with locomotion defects or sensitive to insect repellent in wild-type animals [McCulloch et al., 2017]. The *ERC2* gene (ELKS/RAB6-interacting/CAST family member 2) which functions as regulator of neurotransmitter release at the nerve terminals active zone (CAZ) [Takao-Rikitsu et al., 2004, Rafiepour et al., 2021] is related with behavior trait [Dong et al., 2015] in goat and under selection with in Sardinian sheep populations [Mastrangelo et al., 2019]. Interestingly, the *DPP6* (Dipeptidyl peptidase like 6) gene was already associated with a footprint of selection in zebu from Madagascar (Article in Section 5.2, page 143). This gene encodes a subunit of voltage-gated potassium channel subfamily D member 2 that modulates neuronal excitability in the brain and acts in proteolysis and in the physiological processes of the central nervous system [Sun et al., 2011].

Reproduction system *DPP6* is also a candidate gene in GWAS associated with birth and weaning weight in Canchim beef cattle [Buzanskas et al., 2014] Among candidate genes involved in reproduction, *ACTRT2* (actin -related protein T2) plays an important role in the heat-induced reversible reduction in sperm concentration and motility [Wu et al., 2020] but is also associated with severe forms of tick-borne encephalitis in human [Orlov et al., 2019].

Response to diseases

The gene *HPSE2* is homologous to the gene *HPSE* which encodes for heparanase, an enzyme degrading polymeric heparan sulfate molecules into shorter chain length oligosaccharides. Heparan sulfate plays an essential role in inflammation and viral infections and was identified as a primary receptor for the porcine reproductive and respiratory syndrome virus (PRRSV) [Tao et al., 2013]. In pig, a synonymous mutation located in exon5 of the *HPSE* gene was detected, and was significantly associated with the white blood cell count, red blood cell count, hemoglobin, and hematocrit in the peripheral blood [Tao et al., 2013]. Moreover, mutations in *HPSE2* cause urofacial syndrome in human [Daly et al., 2010] and are involved in breast cancer [Wu et al., 2020]. The gene *CACNA2D3* (auxiliary member of the alpha-2/delta subunit three family of the voltage-dependent calcium channel complex) plays a critical role in tumor suppression [Jin et al., 2017, Nie et al., 2019].

The gene *GALNTL6*, in addition to being involved in resistance to internal parasites in thin-tailed Australian sheep [Al Kalaldeh et al., 2019], is associated with physical performance and have a role in gut microbiome regarding short-chain fatty acid regulation and their anti-inflammatory and resynthesis functions [Díaz Ramírez et al., 2020]. It has been shown associated with feed efficiency and growth traits in cattle [Doran et al., 2014, Seabury et al., 2017] and chicken [Moroudi et al., 2021].

Growth and metabolism

One another gene associated with growth traits is the *PPARGC1A* gene (Peroxisome proliferator-activated receptor gamma coactivator 1-alpha) is a member of transcriptional

coactivator of the peroxisome proliferator-activated receptor and involved in lipid metabolism, energy metabolism, adipocyte differentiation and regulation of mitochondrial biogenesis. A 17 bp insertion polymorphism within the *PPARGC1A* gene was shown to influence growth traits of sheep [Chen et al., 2020] and associated with growth, carcass and meat quality traits in grazing Texel sheep [Armstrong et al., 2018].

On chromosome 2, between the positions 119'046 kb and 119'282 kb, 4 genes (*COL5A2, LOC100212904, COL3A1, MIR9182*) overlapped each other. Among them, *COL5A2* and *COL3A1*, encoding for structural components of blood vessels, uterus and bowel, are involved in Classic Ehlers-Danlos syndrome which is a rare autosomal dominant connective tissue disorder that is primarily characterized by skin hyperextensibility, abnormal wound healing/atrophic scars, and joint hypermobility [Ritelli et al., 2013, Frank et al., 2015], althought other pathologies are associated with theses genes (bladder cancer [Zeng et al., 2018], heart disease [Azuaje et al., 2013] or fibrotic diseases [Kuivaniemi and Tromp, 2019]).

All of these functions are the results environmental selection pressure, with a exception for nervous system function which could be a result of human selection.

TABLE S6.5 – Candidate genes associated with iHS_{MAY}, $Rsb_{MAYvsAFO}$, $Rsb_{MAYvsEAS}$ and $Rsb_{MAYvsWAS}$. For each gene, the position (in bp) on the oviAri4 assembly +/- 15 kb is given and express in bp. For *iHS* and *Rsb* tests, the *log(1/P)* value where *P* represents the *P*-value of the corresponding score and the number of SNPs that reach *log(1/P)* > 3 were reported.

Gene symbol	OAR	START	END	iHS_{MAY}	$Rsb_{MAYvsAFO}$	$Rsb_{MAYvsEAS}$	$Rsb_{MAYvsWAS}$	Function
ZBBX	1	217'942'380	218'047'581				3.50 ; 2	Nervous system
GALNTL6	2	106'507'464	108'022'733	3.50 ; 5				Metabolism, Immunitary system
BIN1	2	117'244'041	117'321'301		4.40 ; 2			Nervous system
MFSD6	2	117'659'301	117'748'885		3.90 ; 2	4.39 ; 2	3.68 ; 2	Nervous system
COL5A2	2	119'046'764	119'225'564		3.97 ; 2			Tissue disorders
LOC100212904	2	119'052'265	119'267'660		3.97 ; 3			
COL3A1	2	119'213'238	119'282'722		3.97 ; 3			Tissue disorders
MIR9182	2	119'252'625	119'282'764		3.74 ; 2			
DPP6	4	115'441'566	116'032'195		3.56 ; 3			Nervous system
PPARGC1A	6	43'142'397	43'890'969			3.42 ; 2		Metabolism
ACTRT2	12	48'006'015	48'037'170		3.55 ; 2			Reproduction
ERC2	19	44'538'933	45'567'041			3.59 ; 2		Nervous system
CACNA2D3	19	45'937'847	46'839'841		3.30 ; 2			Immunitary system
HPSE2	22	18'715'479	19'435'480		4.65 ; 3			Immunitary system

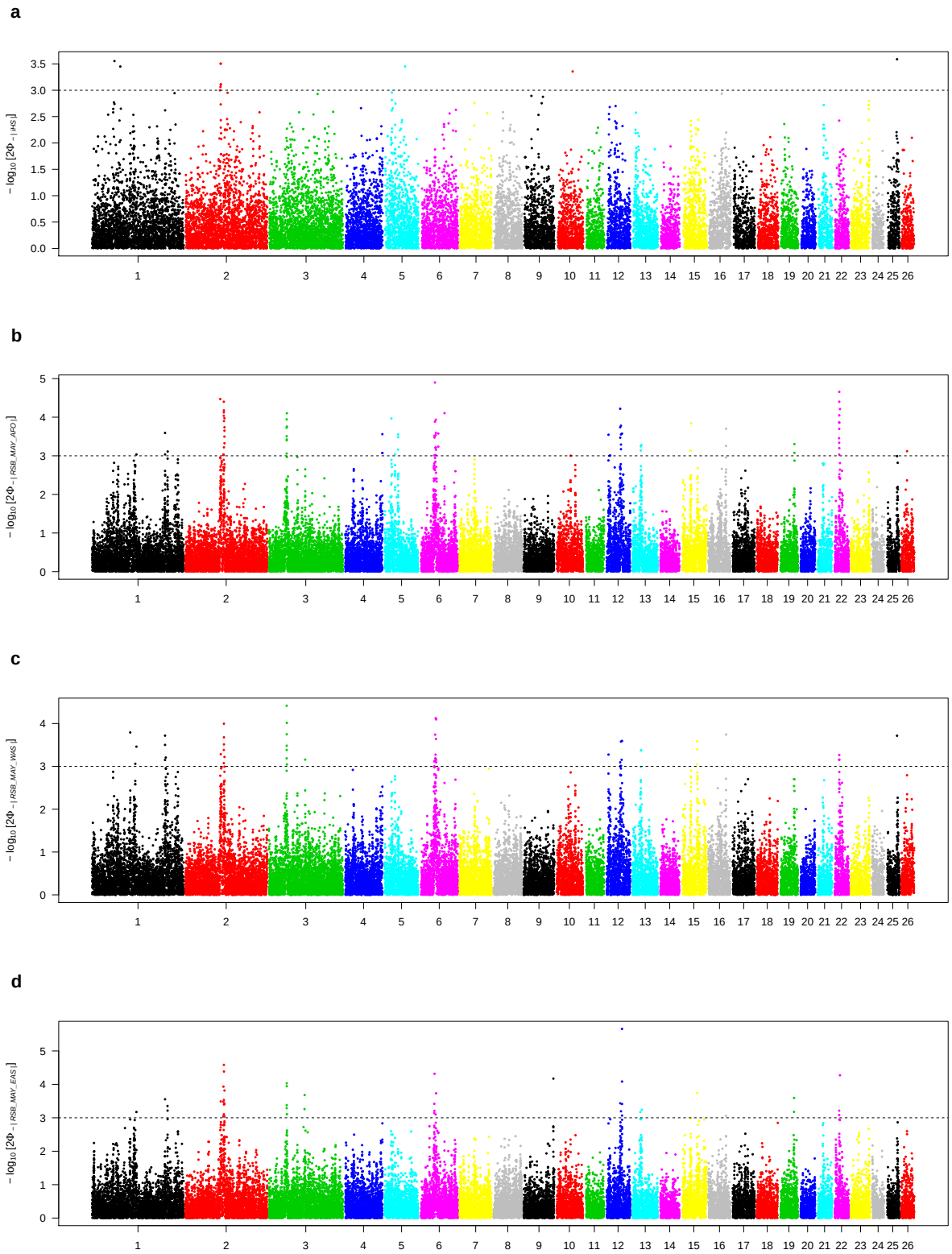


FIGURE S6.9 – Manhattan plot of pvalues for **a** iHS_{MAY}, **b** $Rsb_{MAYvsAFO}$, **c** $Rsb_{MAYvsWAS}$ and **d** $Rsb_{MAYvsEAS}$ analyses.

6.5 Discussion du Chapitre

Les analyses exploratoires ont permis de déterminer la proximité génétique de la population ovine locale de Mayotte avec les moutons africains. Le calcul des segments ROH avec le package RZooroh a apporté quelques premières informations sur l'évolution du statut de consanguinité de la population sur un temps récent : la population présente deux pics de consanguinité l'un daté d'il y a une dizaine d'année et un autre plus ancient datant de la période entre la fin du 19^{éme} siècle et le milieu du 20^{éme} siècle. L'absence de données historiques et de mémoire collective sur l'élevage ovin à Mayotte ne nous permet pas de conclure fermement sur les raisons de ces variations de consanguinité. Un calcul de l'évolution de la taille efficace *Ne* à partir de l'estimation du déséquilibre de liaison permettrait de déterminer si ces variations de consanguinité sont dû à des variations d'effectifs (goulot d'étranglement). L'intégration de nouvelles données de génotypage de populations ovines locales du continent africain pourrait apporter des informations supplémentaires sur les histoires démographiques de la population ovine de Mayotte. Des jeux de données issus de génotypage sur moyenne densité ont été produits récemments (Afrique du Sud [Dzomba et al., 2020], populations locales éthiopiennes [Ahbara et al., 2019] et égyptiennes [Mwacharo et al., 2017]).

Le déterminisme génétique de la mutation auriculaire observée chez 20 % des individus constituant notre échantillon, est constitué d'un polymorphisme candidat situé à la position 114 296 273 bp, situé à proximité du gène *HMX1*, déjà candidat pour le même type de mutation phénotypique chez différentes races locales ovines [He et al., 2020], bovines [Koch et al., 2013], humaines [Si et al., 2020] et chez la souris [Schorderet et al., 2008]. Mais il reste cependant à déterminer le variant causal, puisque la plupart des SNPs montrant une association avec un phénotype donné sont bien souvent non-fonctionnels mais cette association reflète un déséquilibre de liaison avec un ou plusieurs variants causaux qui a pu ne pas être génotypé. De plus, les variants causaux pourraient être autre chose qu'un polymorphisme nucléotidique : insertion, deletion ou une inversion. Il est donc nécessaire de compléter cette GWAS par l'étude des donnnées de séquence. Une première analyse a été effectuée sur des données de séquences mappées sur l'assemblage de référence ISCG Oar_v3.1/oviAri3 et a permis d'identifier une duplication de 73 bp dans le second intron du gène HMX1 à l'état homozygote chez les animaux porteurs de la mutation auriculaire (petites oreilles et microtie) et hétérozygote chez les animaux "sains". Néanmoins, ces résultats restent à confirmer compte tenu de l'absence d'annotation du gène sur l'assemblage oviAri3 alors qu'il est présent sur l'assemblage oviAri4 que nous venons d'utiliser. Une fois la région génomique de la mutation candidate identifiée, il est possible d'effectuer une validation par PCR sur les échantillons d'ADN et en faisant migrer sur gel d'agarose [Jawasreh and Al-Omari, 2020]. L'identification des variants causaux à l'origine de ce phénotype nous permettra d'établir une stratégie de gestion de la mutation à l'échelle de la population. La présence de la mutation auriculaire ne semble pas être accompagnée de désagréments physiologiques ou de baisse de niveau de production à dire d'éleveurs, mais cela doit être confirmé. La forme extrême de la mutation,

dite "moignon", peut représenter un problème pour l'identification des individus par boucle auriculaire, d'autant plus que la réglementation impose la pose de matériel électronique.

L'analyse de l'étendue du déséquilibre de liaison (EHH) de la population de mouton mahorais, comparée avec celles des populations supposés ancestrales (Africaines, Moyen-Orient, Asie de l'Est) a permis de déterminer les régions génomiques et les gènes candidats sous sélection, présentées dans la section 6.4.2. L'annotation fonctionnelle sur les gènes candidats a permis de déterminer les fonctions biologiques relatives au système nerveux, reproductif, la réponse aux agents pathogènes, le métabolisme, les fonctions dermatologiques. Plus particulièrement, les gènes candidats *MFSD6, COL5A2, COL3A1, (LOC 100112904, MIR9162)* chevauchent une région candidate sous sélection en cours de fixation (iHS) chez les moutons égyptiens Barki connus pour être adaptés aux conditions arides du désert occidental [Kim et al., 2016].

Le territoire de Mayotte n'offre que peu de moyens pour la structuration d'une filière ovine qui est à ce jour inexistante. L'identification des ovins demeure la prérogative officieuse de la Chambre d'Agriculture : la base de données d'identification, constituée d'un fichier excel n'a pas été intégrée dans le Système d'information Général et la pose des supports d'identification se fait au grès des arrivages peu fréquents des boucles. L'absence de rôle des pouvoirs publics quant à la structuration de la filière impose aux producteurs l'importation d'ovins (à dire d'éleveurs) mais aucune donnée n'est disponible concernant l'introduction de reproducteurs. Le processus de reconnaissance de la race locale est primordial afin de protéger cette ressource génétique dont l'effectif a été estimé à 1'000 individus en 2011‘ [France, 2011]. Toutefois l'absence de structure porteuse et candidate pour la gestion de la race représente un frein. Une mission d'appui menée par J. Gueux et M. Naves en 2013 proposait différents schémas d'utilisation des ressources génétiques pour répondre aux objectifs de préservation de la race locale, son amélioration et l'augmentation globale des effectifs ovins par le croisement avec des animaux exotiques.

Bibliographie

[Ahbara et al., 2019] Ahbara, A., Bahbahani, H., Almathen, F., Al Abri, M., Agoub, M. O., Abeba, A., Kebede, A., Musa, H. H., Mastrangelo, S., Pilla, F., Ciani, E., Hanotte, O., and Mwacharo, J. M. (2019). Genome-Wide Variation, Candidate Regions and Genes Associated With Fat Deposition and Tail Morphology in Ethiopian Indigenous Sheep. *Frontiers in Genetics*, 9.

[Al Kalaldeh et al., 2019] Al Kalaldeh, M., Gibson, J., Lee, S. H., Gondro, C., and van der Werf, J. H. J. (2019). Detection of genomic regions underlying resistance to gastrointestinal parasites in Australian sheep. *Genetics Selection Evolution*, 51(1) :37.

[Alexander et al., 2009] Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9) :1655–1664.

[Anonymous, 1950] Anonymous, N. (1950). L'élevage du mouton à Madagascar. *Revue d'élevage et de médecine vétérinaire des pays tropicaux*, 4(4) :204.

[Armstrong et al., 2018] Armstrong, E., Ciappesoni, G., Iriarte, W., Da Silva, C., Macedo, F., Navajas, E. A., Brito, G., San Julián, R., Gimeno, D., and Postiglioni, A. (2018). Novel genetic polymorphisms associated with carcass traits in grazing Texel sheep. *Meat Science*, 145 :202–208.

[Azuaje et al., 2013] Azuaje, F., Zhang, L., Jeanty, C., Puhl, S.-L., Rodius, S., and Wagner, D. R. (2013). Analysis of a gene co-expression network establishes robust association between Col5a2 and ischemic heart disease. *BMC Medical Genomics*, 6(1) :13.

[Benjamini and Hochberg, 1995] Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society : Series B (Methodological)*, 57(1) :289–300.

[Bertrand et al., 2019] Bertrand, A., Kadri, N. K., Flori, L., Gautier, M., and Druet, T. (2019). RZooRoH : An R package to characterize individual genomic autozygosity and identify homozygous-by-descent segments. *Methods in Ecology and Evolution*, 10(6).

[Bush and Moore, 2012] Bush, W. S. and Moore, J. H. (2012). Chapter 11 : Genome-Wide Association Studies. *PLoS Computational Biology*, 8(12) :e1002822.

[Buzanskas et al., 2014] Buzanskas, M. E., Grossi, D. A., Ventura, R. V., Schenkel, F. S., Sargolzaei, M., Meirelles, S. L. C., Mokry, F. B., Higa, R. H., Mudadu, M. A., da Silva, M. V. G. B., Niciura, S. C. M., Júnior, R. A. A. T., Alencar, M. M., Regitano, L. C. A., and Munari, D. P. (2014). Genome-Wide Association for Growth Traits in Canchim Beef Cattle. *PLoS ONE*, 9(4) :e94802.

[Chen et al., 2020] Chen, P., Zhao, H., Wu, M., He, S., Yuan, T., Yi, X., Liu, S., Pan, Y., Li, Q., Wang, S., and Sun, X. (2020). A novel 17 bp InDel polymorphism within the PPARGC1A gene is significantly associated with growth traits in sheep. *Animal Biotechnology*, 0(0) :1–9.

[Daly et al., 2010] Daly, S. B., Urquhart, J. E., Hilton, E., McKenzie, E. A., Kammerer, R. A., Lewis, M., Kerr, B., Stuart, H., Donnai, D., Long, D. A., Burgu, B., Aydogdu, O., Derbent, M., Garcia-Minaur, S., Reardon, W., Gener, B., Shalev, S., Smith, R., Woolf, A. S., Black, G. C., and Newman, W. G. (2010). Mutations in HPSE2 Cause Urofacial Syndrome. *The American Journal of Human Genetics*, 86(6) :963–969.

[Dong et al., 2015] Dong, Y., Zhang, X., Xie, M., Arefnezhad, B., Wang, Z., Wang, W., Feng, S., Huang, G., Guan, R., Shen, W., Bunch, R., McCulloch, R., Li, Q., Li, B., Zhang, G., Xu, X., Kijas, J. W., Salekdeh, G. H., Wang, W., and Jiang, Y. (2015). Reference genome of wild goat (capra aegagrus) and sequencing of goat breeds provide insight into genic basis of goat domestication. *BMC Genomics*, 16(1) :431.

[Doran et al., 2014] Doran, A. G., Berry, D. P., and Creevey, C. J. (2014). Whole genome association study identifies regions of the bovine genome and biological pathways involved in carcass trait performance in Holstein-Friesian cattle. *BMC Genomics*, 15(1) :837.

[Druet and Gautier, 2017a] Druet, T. and Gautier, M. (2017a). A model-based approach to characterize individual inbreeding at both global and local genomic scales. *Molecular Ecology*, 26(20) :5820–5841.

[Druet and Gautier, 2017b] Druet, T. and Gautier, M. (2017b). A model-based approach to characterize individual inbreeding at both global and local genomic scales. *Molecular Ecology*, 26(20) :5820–5841.

[Dzomba et al., 2020] Dzomba, E. F., Chimonyo, M., Snyman, M. A., and Muchadeyi, F. C. (2020). The genomic architecture of South African mutton, pelt, dual-purpose and nondescript sheep breeds relative to global sheep populations. *Animal Genetics*, 51(6) :910–923.

[Díaz Ramírez et al., 2020] Díaz Ramírez, J., Álvarez Herms, J., Castañeda-Babarro, A., Larruskain, J., Ramírez de la Piscina, X., Borisov, O. V., Semenova, E. A., Kostryukova, E. S., Kulemin, N. A., Andryushchenko, O. N., Larin, A. K., Andryushchenko, L. B., Generozov, E. V., Ahmetov, I. I., and Odriozola, A. (2020). The GALNTL6 Gene rs558129 Polymorphism Is Associated With Power Performance. *Journal of Strength and Conditioning Research*, 34(11) :3031–3036.

[Flori et al., 2019] Flori, L., Moazami-Goudarzi, K., Alary, V., Araba, A., Boujenane, I., Boushaba, N., Casabianca, F., Casu, S., Ciampolini, R., D'Acier, A. C., Coquelle, C., Delgado, J.-V., El-Beltagi, A., Hadjipavlou, G., Jousselin, E., Landi, V., Lauvie, A., Lecomte, P., Ligda, C., Marinthe, C., Martinez, A., Mastrangelo, S., Menni, D., Moulin, C.-H., Osman, M.-A., Pineau, O., Portolano, B., Rodellar, C., Saïdi-Mehtar, N., Sechi, T., Sempéré, G., Thévenon, S., Tsiokos, D., Laloë, D., and Gautier, M. (2019). A genomic map of climate adaptation in Mediterranean cattle breeds. *Molecular Ecology*, 28(5) :1009–1029.

[France, 2011] France, M. (2011). Synthèse illustrée du recensement agricole 2010. Technical report, Ministère de l'Agriculture, de l'Alimentation, de la Pêche, de la Ruralité et de l'Aménagement du Territoire, Mayotte.

[Frank et al., 2015] Frank, M., Albuisson, J., Ranque, B., Golmard, L., Mazzella, J.-M., Bal-Theoleyre, L., Fauret, A.-L., Mirault, T., Denarié, N., Mousseaux, E., Boutouyrie, P., Fiessinger, J.-N., Emmerich, J., Messas, E., and Jeunemaitre, X. (2015). The type of variants at the COL3A1 gene associates with the phenotype and severity of vascular Ehlers–Danlos syndrome. *European Journal of Human Genetics*, 23(12) :1657–1664.

[Gautier et al., 2007] Gautier, M., Faraut, T., Moazami-Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., Boland, A., Garnier, J.-G., Boichard, D., Lathrop, G. M., Gut, I. G., and Eggen, A. (2007). Genetic and Haplotypic Structure in 14 European and African Cattle Breeds. *Genetics*, 177(2) :1059–1070.

[Gautier et al., 2017] Gautier, M., Klassmann, A., and Vitalis, R. (2017). rehh 2.0 : a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular Ecology Resources*, 17(1) :78–90.

[Gautier and Vitalis, 2012] Gautier, M. and Vitalis, R. (2012). rehh : an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics*, 28(8) :1176–1177.

[Gendron et al., 2016] Gendron, C., Schwentker, A., and van Aalst, J. A. (2016). Genetic Advances in the Understanding of Microtia. *Journal of Pediatric Genetics*, 5(4) :189–197.

[He et al., 2020] He, S., Zhang, Z., Sun, Y., Ren, T., Li, W., Zhou, X., Michal, J. J., Jiang, Z., and Liu, M. (2020). Genome-wide association study shows that microtia in Altay sheep is caused by a 76 bp duplication of *HMX1*. *Animal Genetics*, 51(1) :132–136.

[Hochberg and Benjamini, 1990] Hochberg, Y. and Benjamini, Y. (1990). More powerful procedures for multiple significance testing. *Statistics in Medicine*, 9(7) :811–818.

[Jawasreh et al., 2016] Jawasreh, K., Boettcher, P. J., and Stella, A. (2016). Genome-wide association scan suggests basis for microtia in Awassi sheep. *Animal Genetics*, 47(4) :504–506.

[Jawasreh and Al-Omari, 2020] Jawasreh, K. I. Z. and Al-Omari, H. D.-A. (2020). Duplication in ECR near HMX1 and a SNP in GATA6 Genes Regulate Microtia in Awassi Sheep. *Genes*, 11(6) :597.

[Jiang et al., 2014] Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J. F., Faraut, T., Wu, C., Muzny, D. M., Li, Y., Zhang, W., Stanton, J.-A., Brauning, R., Barris, W. C., Hourlier, T., Aken, B. L., Searle, S. M. J., Adelson, D. L., Bian, C., Cam, G. R., Chen, Y., Cheng, S., DeSilva, U., Dixen, K., Dong, Y., Fan, G., Franklin, I. R., Fu, S., Fuentes-Utrilla, P., Guan, R., Highland, M. A., Holder, M. E., Huang, G., Ingham, A. B., Jhangiani, S. N., Kalra, D., Kovar, C. L., Lee, S. L., Liu, W., Liu, X., Lu, C., Lv, T., Mathew, T., McWilliam, S., Menzies, M., Pan, S., Robelin, D., Servin, B., Townley, D., Wang, W., Wei, B., White, S. N., Yang, X., Ye, C., Yue, Y., Zeng, P., Zhou, Q., Hansen, J. B., Kristiansen, K., Gibbs, R. A., Flicek, P., Warkup, C. C., Jones, H. E., Oddy, V. H., Nicholas, F. W., McEwan, J. C., Kijas, J. W., Wang, J., Worley, K. C., Archibald, A. L., Cockett, N., Xu, X., Wang, W., and Dalrymple, B. P. (2014). The sheep genome illuminates biology of the rumen and lipid metabolism. *Science*, 344(6188) :1168–1173.

[Jin et al., 2017] Jin, Y., Cui, D., Ren, J., Wang, K., Zeng, T., and Gao, L. (2017). CACNA2D3 is downregulated in gliomas and functions as a tumor suppressor. *Molecular Carcinogenesis*, 56(3) :945–959.

[Jonnalagadda et al., 2019] Jonnalagadda, M., Faizan, M. A., Ozarkar, S., Ashma, R., Kulkarni, S., Norton, H. L., and Parra, E. (2019). A Genome-Wide Association Study of Skin and Iris Pigmentation among Individuals of South Asian Ancestry. *Genome Biology and Evolution*, 11(4) :1066–1076.

[Kijas et al., 2012] Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., and Dalrymple, B. (2012). Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biology*, 10(2).

[Kim et al., 2016] Kim, E.-S., Elbeltagy, A. R., Aboul-Naga, A. M., Rischkowsky, B., Sayre, B., Mwacharo, J. M., and Rothschild, M. F. (2016). Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity*, 116(3) :255–264.

[Koch et al., 2013] Koch, C. T., Bruggmann, R., Tetens, J., and Drögemüller, C. (2013). A Non-Coding Genomic Duplication at the HMX1 Locus Is Associated with Crop Ears in Highland Cattle. *PLoS ONE*, 8(10).

[Kuivaniemi and Tromp, 2019] Kuivaniemi, H. and Tromp, G. (2019). Type III collagen (COL3A1) : Gene and protein structure, tissue distribution, and associated diseases. *Gene*, 707 :151–171.

[Li et al., 2017] Li, Q., Wineinger, N. E., Fu, D.-J., Libiger, O., Alphs, L., Savitz, A., Gopal, S., Cohen, N., and Schork, N. J. (2017). Genome-wide association study of paliperidone efficacy. *Pharmacogenetics and Genomics*, 27(1) :7–18.

[Liu et al., 2013] Liu, G., Zhang, S., Cai, Z., Li, Y., Cui, L., Ma, G., Jiang, Y., Zhang, L., Feng, R., Liao, M., Chen, Z., Zhao, B., and Li, K. (2013). BIN1 gene rs744373 polymorphism contributes to Alzheimer's disease in East Asian population. *Neuroscience Letters*, 544 :47–51.

[Mastrangelo et al., 2019] Mastrangelo, S., Jemaa, S. B., Sottile, G., Casu, S., Portolano, B., Ciani, E., and Pilla, F. (2019). Combined approaches to identify genomic regions involved in phenotypic differentiation between low divergent breeds : Application in Sardinian sheep populations. *Journal of Animal Breeding and Genetics*, 136(6) :526–534.

[Mastrangelo et al., 2018] Mastrangelo, S., Sottile, G., Sutera, A. M., Di Gerlando, R., Tolone, M., Moscarelli, A., Sardina, M. T., and Portolano, B. (2018). Genome-wide association study reveals the locus responsible for microtia in Valle del Belice sheep breed. *Animal Genetics*, 49(6) :636–640.

[McCulloch et al., 2017] McCulloch, K. A., Qi, Y. B., Takayanagi-Kiya, S., Jin, Y., and Cherra, III, S. J. (2017). Novel Mutations in Synaptic Transmission Genes Suppress Neuronal Hyperexcitation in Caenorhabditis elegans. *G3 Genes|Genomes|Genetics*, 7(7) :2055–2063.

[Moroudi et al., 2021] Moroudi, R. S., Mahyari, S. A., Torshizi, R. V., Lanjanian, H., and Masoudi-Nejad, A. (2021). Identification of new genes and quantitative trait locis associated with growth curve parameters in F2 chicken population using genome-wide association study. *Animal Genetics*, 52(2) :171–184.

[Munroe et al., 2009] Munroe, R. J., Prabhu, V., Acland, G. M., Johnson, K. R., Harris, B. S., O'Brien, T. P., Welsh, I. C., Noden, D. M., and Schimenti, J. C. (2009). Mouse H6 Homeobox 1 (Hmx1) mutations cause cranial abnormalities and reduced body mass. *BMC Developmental Biology*, 9(1) :27.

[Mwacharo et al., 2017] Mwacharo, J. M., Kim, E.-S., Elbeltagy, A. R., Aboul-Naga, A. M., Rischkowsky, B. A., and Rothschild, M. F. (2017). Genomic footprints of dryland stress adaptation in Egyptian fat-tail sheep and their divergence from East African and western Asia cohorts. *Scientific Reports*, 7(1) :17647.

[Nie et al., 2019] Nie, C., Qin, X., Li, X., Tian, B., Zhao, Y., Jin, Y., Li, Y., Wang, Q., Zeng, D., Hong, A., and Chen, X. (2019). CACNA2D3 Enhances the Chemosensitivity of Esophageal Squamous Cell Carcinoma to Cisplatin via Inducing Ca2+-Mediated Apoptosis and Suppressing PI3K/Akt Pathways. *Frontiers in Oncology*, 9 :185.

[Orlov et al., 2019] Orlov, Y. L., Galieva, E. R., and Melerzanov, A. V. (2019). Computer genomics research at the bioinformatics conference series in Novosibirsk. *BMC Genomics*, 20(7) :537.

[Paradis et al., 2006] Paradis, E., Strimmer, K., Claude, J., Jobb, G., Opgen-Rhein, R., Dutheil, J., Noel, Y., Bolker, B., and Lemon, J. (2006). ape : Analyses of Phylogenetics and Evolution. *R package version*, 1(4).

[Patterson et al., 2006] Patterson, N., Price, A. L., and Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genetics*, 2(12).

[Pauly and Ferrandis, 2018] Pauly, M. and Ferrandis, M. (2018). Le site funéraire d'Antsiraka Boira (Acoua, Grande Terre) : Islamisation et syncrétisme culturel à Mayotte au XIIe siècle. *Afriques. Débats, méthodes et terrains d'histoire*.

[Philipp et al., 2011] Philipp, U., Lupp, B., Mömke, S., Stein, V., Tipold, A., Eule, J. C., Rehage, J., and Distl, O. (2011). A MITF Mutation Associated with a Dominant White Phenotype and Bilateral Deafness in German Fleckvieh Cattle. *PLOS ONE*, 6(12) :e28857.

[Porter et al., 2016] Porter, V., Alderson, L., Hall, S. J. G., and Sponenberg, D. P. (2016). *Mason's World Encyclopedia of Livestock Breeds and Breeding, 2 Volume Pack*. CABI. Google-Books-ID : 2UEJDAAAQBAJ.

[Pourlis, 2011] Pourlis, A. F. (2011). A review of morphological characteristics relating to the production and reproduction of fat-tailed sheep breeds. *Tropical Animal Health and Production*, 43(7) :1267–1287.

[Purcell et al., 2007] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., and Sham, P. C. (2007). PLINK : A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81(3) :559–575.

[Rafiepour et al., 2021] Rafiepour, M., Ebrahimie, E., Vahidi, M. F., Salekdeh, G. H., Niazi, A., Dadpasand, M., Liang, D., Si, J., Ding, X., Han, J., Zhang, Y., and Qanbari, S. (2021). Whole-Genome Resequencing Reveals Adaptation Prior to the Divergence of Buffalo Subspecies. *Genome Biology and Evolution*, 13(1).

[Ritelli et al., 2013] Ritelli, M., Dordoni, C., Venturini, M., Chiarelli, N., Quinzani, S., Traversa, M., Zoppi, N., Vascellaro, A., Wischmeijer, A., Manfredini, E., Garavelli, L., Calzavara-Pinton, P., and Colombi, M. (2013). Clinical and molecular characterization of 40 patients with classic Ehlers–Danlos syndrome : identification of 18 COL5A1 and 2 COL5A2 novel mutations. *Orphanet Journal of Rare Diseases*, 8(1) :58.

[Rochus et al., 2018] Rochus, C. M., Tortereau, F., Plisson-Petit, F., Restoux, G., Moreno-Romieux, C., Tosser-Klopp, G., and Servin, B. (2018). Revealing the selection history of adaptive loci using genome-wide scans for selection : an example from domestic sheep. *BMC Genomics*, 19(1) :71.

[Saitou and Nei, 1987] Saitou, N. and Nei, M. (1987). The neighbor-joining method : a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, 4(4) :406–425.

[Salcedo-Tacuma et al., 2019] Salcedo-Tacuma, D., Melgarejo, J. D., Mahecha, M. F., Ortega-Rojas, J., Arboleda-Bustos, C. E., Pardo-Turriago, R., and Arboleda, H. (2019). Differential Methylation Levels in CpGs of the BIN1 Gene in Individuals With Alzheimer Disease. *Alzheimer Disease & Associated Disorders*, 33(4) :321–326.

[Scheet and Stephens, 2006] Scheet, P. and Stephens, M. (2006). A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data : Applications to Inferring Missing Genotypes and Haplotypic Phase. *American Journal of Human Genetics*, 78(4) :629–644.

[Schorderet et al., 2008] Schorderet, D. F., Nichini, O., Boisset, G., Polok, B., Tiab, L., Mayeur, H., Raji, B., de la Houssaye, G., Abitbol, M. M., and Munier, F. L. (2008). Mutation in the Human Homeobox Gene NKX5-3 Causes an Oculo-Auricular Syndrome. *The American Journal of Human Genetics*, 82(5) :1178–1184.

[Seabury et al., 2017] Seabury, C. M., Oldeschulte, D. L., Saatchi, M., Beever, J. E., Decker, J. E., Halley, Y. A., Bhattarai, E. K., Molaei, M., Freetly, H. C., Hansen, S. L., Yampara-Iquise, H., Johnson, K. A., Kerley, M. S., Kim, J., Loy, D. D., Marques, E., Neibergs, H. L., Schnabel, R. D., Shike, D. W., Spangler, M. L., Weaber, R. L., Garrick, D. J., and Taylor, J. F. (2017). Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics*, 18(1) :1–25.

[Sempéré et al., 2015] Sempéré, G., Moazami-Goudarzi, K., Eggen, A., Laloë, D., Gautier, M., and Flori, L. (2015). WIDDE : a Web-Interfaced next generation database for genetic diversity exploration, with a first application in cattle. *BMC Genomics*, 16(1).

[Sharma et al., 2015] Sharma, A., Lee, J. S., Dang, C. G., Sudrajad, P., Kim, H. C., Yeon, S. H., Kang, H. S., and Lee, S.-H. (2015). Stories and Challenges of Genome Wide Association Studies in Livestock — A Review. *Asian-Australasian Journal of Animal Sciences*, 28(10) :1371–1379.

[Si et al., 2020] Si, N., Meng, X., Lu, X., Liu, Z., Qi, Z., Wang, L., Li, C., Yang, M., Zhang, Y., Wang, C., Guo, P., Zhu, L., Liu, L., Li, Z., Zhang, Z., Cai, Z., Pan, B., Jiang, H., and Zhang, X. (2020). Duplications involving the long range HMX1 enhancer are associated with human isolated bilateral concha-type microtia. *Journal of Translational Medicine*, 18(1) :244.

[Sun et al., 2011] Sun, W., Maffie, J. K., Lin, L., Petralia, R. S., Rudy, B., and Hoffman, D. A. (2011). DPP6 Establishes the A-Type K+ Current Gradient Critical for the Regulation of Dendritic Excitability in CA1 Hippocampal Neurons. *Neuron*, 71(6) :1102–1115.

[Takao-Rikitsu et al., 2004] Takao-Rikitsu, E., Mochida, S., Inoue, E., Deguchi-Tawarada, M., Inoue, M., Ohtsuka, T., and Takai, Y. (2004). Physical and functional interaction of the active zone proteins, CAST, RIM1, and Bassoon, in neurotransmitter release. *Journal of Cell Biology*, 164(2) :301–311.

[Tan et al., 2014] Tan, M.-S., Yu, J.-T., Jiang, T., Zhu, X.-C., Guan, H.-S., and Tan, L. (2014). Genetic variation in BIN1 gene and Alzheimer's disease risk in Han Chinese individuals. *Neurobiology of Aging*, 35(7) :1781.e1–1781.e8.

[Tang et al., 2007] Tang, K., Thornton, K. R., and Stoneking, M. (2007). A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *PLoS Biology*, 5(7) :e171.

[Tao et al., 2013] Tao, C., Wang, W., Zhou, P., Xia, T., Zhou, X., Zeng, C., Zhang, Q., and Liu, B. (2013). Molecular characterization, expression profiles, and association analysis with hematologic parameters of the porcine HPSE and HPSE2 genes. *Journal of Applied Genetics*, 54(1) :71–78.

[Voight et al., 2006] Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A Map of Recent Positive Selection in the Human Genome. *PLoS Biology*, 4(3) :e72.

[Weir and Cockerham, 1984] Weir, B. S. and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, (19) :395–420.

[Wickham, 2016] Wickham, H. (2016). ggplot2 : Elegant Graphics for Data Analysis. *Springer-Verlag New York*.

[Wright et al., 1984] Wright, H. T., Sinopoli, C., Wojnaroski, L., Hoffman, E. S., Scott, S. L., Redding, R. W., and Goodman, S. M. (1984). Early Seafarers of the Comoro Islands : the Dembeni Phase of the IXth-Xth Centuries AD. *Azania : Archaeological Research in Africa*, 19(1) :13–59.

[Wu et al., 2020] Wu, Y.-Q., Rao, M., Hu, S.-F., Ke, D.-D., Zhu, C.-H., and Xia, W. (2020). Effect of transient scrotal hyperthermia on human sperm : an iTRAQ-based proteomic analysis. *Reproductive Biology and Endocrinology*, 18(1) :83.

[Young et al., 2011] Young, E., Kijas, J., McCulloch, R., Scobie, D., McRae, K., Pickering, N., Dodds, K., and McEwan, J. (2011). Arapawa : a novel New Zealand sheep breed of distinct origin. *Proceedings of the New Zealand Society of Animal Production*, 71 :248–250.

[Yu et al., 2017] Yu, G., Smith, D., Zhu, H., Guan, Y., and Lam, T. T.-Y. (2017). ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8 :28–36.

[Zeng et al., 2018] Zeng, X.-T., Liu, X.-P., Liu, T.-Z., and Wang, X.-H. (2018). The clinical significance of COL5A2 in patients with bladder cancer. *Medicine*, 97(10) :e0091.

Chapitre 7

Identification des régions génomiques et des gènes associés avec le climat chez les populations bovines et ovines de l'Océan Indien

La recherche de signatures de sélection dans le génome des populations de zébus de Mayotte et Madagascar et d'ovins de Mayotte a permis d'identifier une liste de gènes candidats. Une revue bibliographique de chaque gène et leur annotation fonctionnelle avec des données d'ontologie de gènes réalisée avec le logiciel IPA a permis de déterminer les principales fonctions biologiques associées à ces signatures de sélection et d'en déduire les pressions de sélection qui ont pu s'exercer sur ces populations, parmi lesquelles le climat a joué un rôle prépondérant.

Afin d'identifier à une échelle géographique plus large, les régions génomiques impliquées dans l'adaptation chez les populations bovines et ovines de l'Océan indien, nous avons sélectionné un panel de races représentatives de la diversité génétique dans la zone et vivant sous des climats contrastés sur le pourtour du littoral de l'Ocean Indien. Nous avons couplé la recherche de signatures de sélection dans leur génome (en analysant la différentiation des fréquences alléliques entre les populations) avec celle de gènes associés avec le climat (à l'aide d'une approche GEA - Genome-Environment Association), en intégrant des variables climatiques populationnelles. Les tests d'associations entre SNPs et variables climatiques permettent en effet d'établir un lien direct entre certaines régions génomiques sous sélection et les contraintes climatiques et d'identifier des régions supplémentaires associées avec le climat dans la zone.

7.1 Synthèse des résultats

Basés sur l'analyse d'un jeu de données de génotypage comprenant pour les bovins, 369 individus de 17 races différentes, génotypés pour 28 556 SNPs et pour les ovins, 271 individus répartis dans 9 races et génotypés sur 44 880 SNPs, les tests de différentiation ont mis en évidence 38 (11) régions candidates associés à 15 (15) gènes candidats pour le jeu de données bovin (ovin). Les tests d'association avec 5 covariables climatiques (température moyenne annuelle, écart moyen de température annuel, précipitations annuelles moyennes, radiation moyenne annuelle, indice moyen d'humidité annuelle, ainsi que le Thermal Humidity Index (THI), combinaison de la température moyenne annuelle et l'indice moyen d'humidité annuelle) ont permis d'identifier respectivement 45 et 23 gènes candidats chez les populations bovines et ovines. Dans les résultats relatifs au jeu de donnée bovin, seules 2 régions sous sélection contiennent des gènes *LEF1, EXOC6B* associés tous deux à la variable climatique Bio12 correspondant aux précipitations moyennes annuelles, ce qui n'est pas le cas pour les ovins. Aucun gène n'est en commun entre les deux espèces. L'annotation fonctionnelle des gènes candidats associés à des covariables climatiques par le logiciel IPA a permis de déterminer que ces gènes étaient impliqués pour les deux espèces dans les propriétés de la peau, le système nerveux, la structure musculo-squelettique, l'embryogénèse et le cancer, notamment à travers la voie de signalisation des protéines Wnt. Ces fonctions biologiques seraient associées au climat de la zone.

7.2 Article 5 : Climate association test and comparison between bovine and ovine populations

(en préparation)

Jessica Magnier[1], Emmanuel Tillard[2], Mélissa Ouvrard[3], Solène Raoul[3], Jérôme Janelle[2], Michel Naves[4], Mathieu Gautier[5], Laurence Flori[1],

[1] UMR SELMET, INRAE, CIRAD, Montpellier SupAgro, Univ. Montpellier, Montpellier, France

[2] UMR SELMET, INRAE, CIRAD, Montpellier SupAgro Univ., Saint Pierre, La Réunion, France

[3] Coopadem, Coconi, 97660 Ouangani, Mayotte

[4] URZ, INRAE, Petit-Bourg, France

[5] UMR CBGP, INRAE, CIRAD, IRD, Montpellier Supagro, Univ. Montpellier, Montferrier-sur-Lez, France

7.2.1 Abstract

The Indian Ocean is a particular geographic area with a rich human history. The tropical climate in coasts of this area is characterized by a high level of temperature and a relative humidity, that represents favorable conditions for pathogens development and heat stress. Due to the availability of genomic resources such as SNP genotyping assays, Genome-Environment Association (GEA) analyses have facilitated identification of genetic variants associated with population-specific environmental covariables. In this study, we searched for footprints of selection and regions/genes associated with 6 climate variables discriminating the different climate subtypes in the area by performing a Genome-Environment analysis, using BayPass Software on 17 bovine and 9 ovine populations. We identified by differentiation method 38 and 11 candidate regions under selection for bovine and ovine datasets respectively. Among these regions, 15 (cattle) and 15 (sheep) candidate genes were identified and none were associated with one climate covariable. We identified 6, 32, 2 and 5 genes associated with temperature annual range, annual precipitation , annual mean radiation and THI climate

variables, respectively for bovine dataset, and 5, 5 and 13 genes associated with annual precipitation, annual mean radiation and THI respectively for ovine dataset. These genes are involved in skin and pigmentation properties, nervous system, skeletal and muscle structure, embryogenese and cancer pathway.

7.2.2 Introduction

Local bovine and ovine populations are relevant biological models to characterize the genetic basis of short-term evolutionary response to abiotic selective pressure such as climate.

The Indian Ocean is a particular geographic area with a rich history. Indeed, long before the arrival of European settlers, this area was a major place of migration and trade by land and sea [Beaujard, 2019b, Beaujard, 2019a]. Along with human migrations from domestication centers, cattle and sheep populations have settled a wide range of new environments on the Indian Ocean coast and differentiated into a variety of populations shaped by the combined action of demographic factors and natural and artificial selections.

The tropical climate in the Indian Ocean coasts, characterized by a high level of temperature and a relative humidity, that represents favorable conditions for pathogens development and heat stress [Bohmanova et al., 2007, Cardoso et al., 2015, Rahman et al., 2018]. It is commonly admitted that there are three types (equatorial, savanah, monsoon) of tropical subclimates according to rainfall pattern [Koppen, 1936] but some variations are possible among the geographic position and island context or not [Peel et al., 2007]. As already observed for cattle populations of Mayotte and Madagascar Islands (Magnier et al ; chapitre 5, section 5.2), these specific climatic conditions certainly exerted a high pressure of selection on animals genome in this area and might have had a strong impact on animal performances [Bohmanova et al., 2007, Cardoso et al., 2015, Rahman et al., 2018]. Hence, in several tropical breeds,

signatures of positive selection in response to heat, solar radiation, poor level of nutrinent and water resources in African [Gautier et al., 2009, Flori et al., 2014, Kim et al., 2017, Kim et al., 2020], Guadeloupean Creole [Gautier and Naves, 2011] and Chinese [Gao et al., 2017] cattle and Ethiopian sheep [Wiener et al., 2021] have been identified.

Due to the availability of genomic resources such as SNP genotyping assays, detecting regions in the whole genome involved in adaptation to local environment is now facilitated. For this purpose, Genome-Environment Association (GEA) analyses have facilitated identification of genetic variants associated with population-specific environmental covariables [Joost et al., 2007, Coop et al., 2010, Frichot et al., 2013, Gautier, 2015]. Combining with functional annotation of candidate genes or variants, these approaches have proved to be efficient at highlighting candidate genes or biological pathways affected by climate variation in bovine species [Gautier et al., 2009, Gao et al., 2017, Flori et al., 2019] and less frequent in ovine species [Lv et al., 2014, Wiener et al., 2021].

We presented here new insights into the adaptative history of bovine and ovine populations around the Indian Ocean, with a focus on climate adaptation. More precisely, we searched for footprints of selection and regions/genes associated with 6 climate variables discriminating the different climate subtypes in the area by performing a Genome-Environment analysis as already applied in [Flori et al., 2019]. We performed a functional annotation of the identified candidate genes using system biology tools in order to uncovered the main physiological pathways mobilized during cattle and sheep adaptation to the Indian Ocean climate. Finally we compared regions under selection and associated with climate data between cattle and sheep.

TABLE 7.1 – List of the 17 cattle and 9 sheep populations integrated into the genotyping dataset after quality filtering.

Species	Population name	Code	n	Reference
Bovine	Aceh	ACE	12	[Decker et al., 2014]
Bovine	Boran	BAN	20	[Decker et al., 2014]
Bovine	Brebes	BRE	9	[Decker et al., 2014]
Bovine	East African Shorthorn Zebu	EAZ	20	[Decker et al., 2014]
Bovine	Gir	GIR	24	[Matukumalli et al., 2009]
Bovine	Kankraj	KAN	10	[Decker et al., 2014]
Bovine	Madura	MAD	7	[Decker et al., 2014]
Bovine	Mahorais	MAY	150	Magnier et al 2021
Bovine	Moka	MOK	31	SNPDOM
Bovine	Ongole Grade	ONG	20	[Decker et al., 2014]
Bovine	Pesisir	PES	6	[Decker et al., 2014]
Bovine	Red Sindhi	RSI	10	[Decker et al., 2014]
Bovine	Sheko	SHK	24	[Matukumalli et al., 2009]
Bovine	Tharparkar	THA	12	[Decker et al., 2014]
Bovine	Thai (Center)	THC	5	[Wangkumhang et al., 2015]
Bovine	Thai (South)	THS	5	[Wangkumhang et al., 2015]
Bovine	Zebu from Madagascar	ZMA	35	[Gautier et al., 2009]
Ovine	Bangladeshi Garole	BGA	24	[Kijas et al., 2012]
Ovine	Bangladeshi East	BGE	24	[Kijas et al., 2012]
Ovine	Ethiopian Menz	EMZ	34	[Kijas et al., 2012]
Ovine	Indian Garole	GAR	26	[Kijas et al., 2012]
Ovine	Garut	GUR	22	[Kijas et al., 2012]
Ovine	Deccani	IDC	24	[Kijas et al., 2012]
Ovine	Mahorais	MAY	48	
Ovine	Red Maasai	RMA	45	[Kijas et al., 2012]
Ovine	Sumatra	SUM	24	[Kijas et al., 2012]

7.2.3 Material and methods

Dataset constitution

A total of 396 individus from 17 different bovine breeds were sampled for DNA extraction in 11 countries and French territories (Figure 7.1 and Table 7.1). For ovine species, 271 individus from 9 different ovine breeds were sampled for DNA extraction in 6 countries and French territories (Figure 7.1 and Table 7.1). Using WIDDE utilities, the detailed constitution of the dataset was :

— 17 bovine breeds representative of Indian indicine (GIR, RSI, THA, KAN, ONG), East-asian indicine (THC, THS, ACE, PES, BRE, MAD) and African indicine (MOK, MAY, ZMA, EAZ, BAN, SHK) ancestry, and

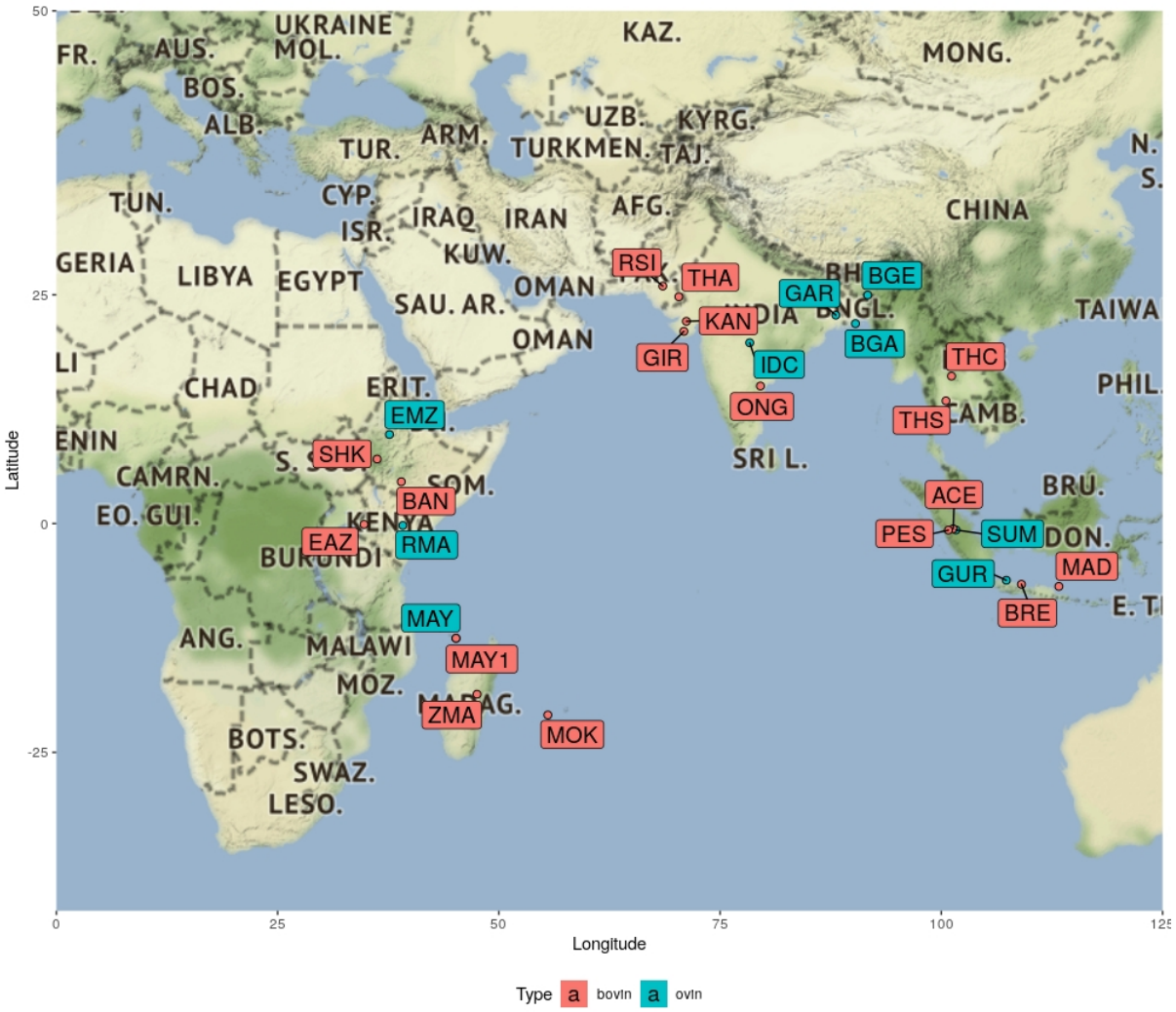


FIGURE 7.1 – Localization of the 17 cattle and 9 sheep populations used in differentiation and climatic association tests.

— 9 ovine breeds representative of African ovine (MAY, RMA, EMZ), Indian ovine (IDC, GAR, BGA, BGE), East Asian ovine (SUM, GUR),

chosen regarding their geographic position (Figure 7.1) .

For all the individuals, a blood sample was collected on each animals and stored on EDTA Vacutainer tubes, following recommendations of the directive 2010/63/EU, standard procedures and relevant guidelines of animal care. Genomic DNA were extracted using the Wizard Genomic DNA purification kit (Promega, France) and stored at -20 C. DNA samples were

genotyped on the Illumina BovineSNP50 beadchip (v1 and v2) at Labogena plateform (INRA, Jouy-en-Josas) using standard procedures (www.illumina.com). For bovine dataset, only 129 MAY individuals and 14 ONG individuals were kept, regarding their European Taurine proportion estimated by Admixture software [Alexander et al., 2009] (See exploratoray analysis in Supplemental Results Section 7.2.7). Hence, the final bovine dataset consisted of 369 individuals.

The minimal SNP genotyping call rate was set to 80% for bovine dataset or 95% for ovine dataset and the minimal individual genotyping call rate to 90% over all populations and 75% within each population (i.e. SNPs genotyped for less than 75% of the animals from at least one population were discarded). SNPs with a MAF < 0.01 or departing from Hardy-Weinberg equilibrium expectation (exact test p <0.001) were also discarded. A total of 28'556 and 44'880 SNPs distributed throughout the 29 and 26 autosomes passed all quality criteria in the bovine and ovine datasets respectively. The SNP genomic positions were actualised with Oar_v4.0/oviAri4 [Jiang et al., 2014] for ovine species and ARS-UCD1.2/Btau9 [Rosen et al., 2020] for bovine species.

Whole-genome scan for adaptative differentiation on bovine and ovine datasets

Whole genome-scans for adaptive differentiation and association with population-specific covariables were performed with BAYPASS 2.1 [Gautier, 2015]. The underlying models explicitly account for the covariance structure among the population allele frequencies, summarized by a scaled covariance matrix Ω, which makes the approach particularly robust to complex demographic histories [Gautier, 2015]. The model also included hyperparameters (a and b) for the Beta distribution of across population allele frequencies which allows to account for the SNP chip ascertainment bias [Gautier, 2015]. Identification of overly differentiated SNPs was

based on the XtX statistics [Gautier, 2015, Günther and Coop, 2013] estimated under the core model of BayPass together with the matrix Ω. To calibrate the XtXs, a pseudo-observed data set (POD) containing 100 000 SNPs simulated under the inference model with hyperparameters equal to those estimated on the real data set was generated and further analyzed under the same conditions following the procedure described in [Gautier, 2015]. In particular, we ensured that the posterior estimate of Ω obtained with the POD was similar to that obtained on the real data since the FMD distance proposed by [Förstner and Moonen, 2003] between the two matrices was found equal to 0.20 for bovine and 0.05 for ovine dataset. Similarly, the posterior means of the two hyperparameters a and b for the Beta distribution of across population allele frequencies obtained on the POD (for bovine $\widehat{a_\pi} = 6.09; \widehat{b_\pi} = 1.50$ and for ovine $\widehat{a_\pi} = 6.93; \widehat{b_\pi} = 2.68$) were almost equal to the ones obtained in the original data (for bovine $\widehat{a_\pi} = 6.02; \widehat{b_\pi} = 1.49$ and for ovine $\widehat{a_\pi} = 6.92; \widehat{b_\pi} = 2.68$). Taken together, these sanity checks indicated that the POD faithfully mimics the real data set, allowing us to define a 0.1% significance threshold on the XtX statistics for bovine species (XtX = 30.4) and a 0.5% significance threshold for ovine species (XtX = 22.99) used to identify genomic regions harbouring footprints of selection.

To that end, the ARS-UCD1.2/Btau9 bovine [Rosen et al., 2020] and Oar_v4.0/oviAri4 ovine [Nicolazzi et al., 2015] genome assemblies were split into 3'570 and 5'610 consecutive 1-Mb windows with 500 kb overlap respectively. Windows which included at least two significant SNPs (at the 0.1% and 0.5% XtX POD threshold for bovine and ovine populations respectively) were deemed significant. After merging overlapping significant windows, we identified for each resulting region the strongest candidate gene as the one located less than 15 kb from the XtX peak. In some instances, some functional candidate genes contiguous to this positional candidate gene were also identified based on their known biological role.

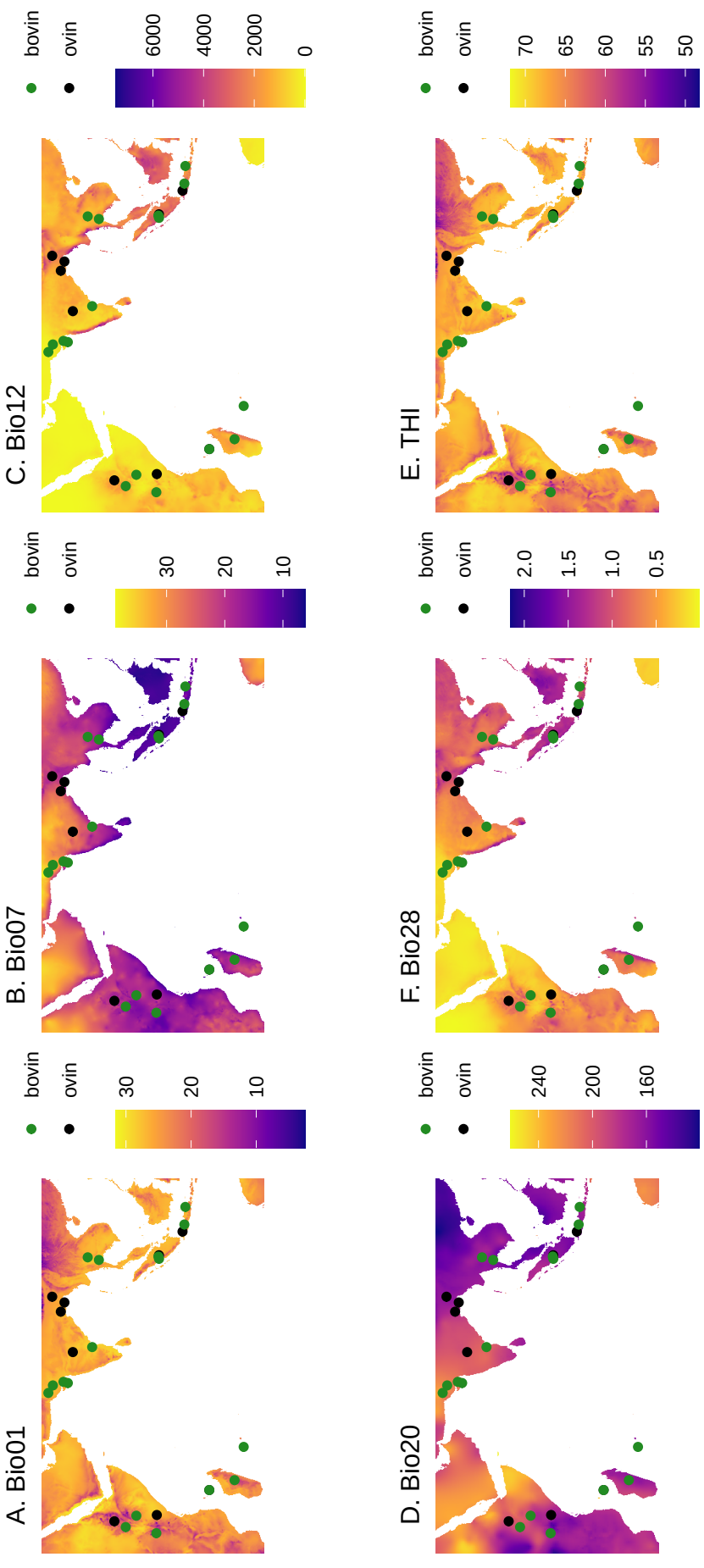


FIGURE 7.2 – Geographic pattern of the annual mean values of the five main climatic variables. Bio01 : annual mean temperature (°C) ; Bio07 : temperature annual range (°C) ; Bio12 : annual precipitation (mm) ; Bio20 : annual mean radiation ($W.m^{-2}$) ; Bio28 : annual mean moisture index ; THI : Thermal Humidity Index.

Genome-environment association analyses and identification of candidate genes associated with climatic covariables

Using GPS coordinates associated to each breed of dataset, 40 climatic covariables (from Bio01 to Bio40) were extracted from the Climond database (`https://climond.org/` ; [Kriticos et al., 2012] ; Supporting information Table S7.7). This database contains bioclimatic variables with original measurable information on annual, weekly, and seasonal temperatures, soil moisture, radiation, and precipitation, that we used a proxy to describe the climate [Ficetola et al., 2017]. These variables are a summary of climatic conditions between 1961 and 1990 in the form of rasters at about 19 km spatial resolution.

Among the 40 climatic covariables, the 5 principal covariables were retained for GEA studies corresponding to :

— Bio01 (annual mean temperature),

— Bio07 (temperature annual range),

— Bio12 (annual precipitation),

— Bio20 (annual mean radiation),

— Bio28 (annual mean moisture index) original climate covariables and

— THI : the annual average of the Temperature Humidity Index (THI), that may be viewed as a measure of discomfort experienced by an animal in warm weather, was calculated for each breed using formula 5 described in [Bohmanova et al., 2007].

The geographic patterns of Bio01, Bio07, Bio12, Bio20, Bio28 and THI were represented in Figure 7.2 using RASTER v3.3.-13 [Hijmans, 2020] and ggspatial v1.1.4 packages [Dunnington, 2020] with bovine populations in pink and ovine populations in blue.

Genome-wide GEA were carried out with the BAYPASS software [Gautier, 2015] using the default options of the AUX model parameterized with the scaled covariance matrix estimated

above (Supporting information Figure S7.3.a and S7.4.a). This model explicitly accounts for multiple tests by integrating over (and estimating) the unknown proportion of SNPs actually associated with a given covariable. The support of the association of each SNP with each covariable was evaluated by a Bayes Factor (BF).

For bovine populations, in order to identify candidate genes associated with climatic covariables, all the genotyped SNPs were first annotated using as a gene set reference a list of 14'562 RefGenes anchored on the ARS-UCD1.2/Btau9 bovine assembly [Rosen et al., 2020] that was uploaded from the UCSC RefGenes database (`http://hgdownload.cse.ucsc.edu/goldenPath/bosTau9/database/`). For ovine dataset, the identification of candidate genes associated with climatic covariables is not possible with the oviAri4 assembly due to a lack of annotated genes for this species. Then, all the genotyped SNPs were first annotated using as a gene set reference a list of 217'848 otherRefSeqGenes that is known protein-coding and non-protein-coding genes for organisms other than sheep, taken from the NCBI reference sequences collection (RefSeq), and that was uploaded from the UCSC RefGene database (`http://hgdownload.cse.ucsc.edu/goldenPath/oviAri4/database/`).

A SNP was then considered as representative of one of these RefGenes genes if localized within its boundary positions extended by 15 kb upstream and downstream to account for persistence of Linkage Disequilibrium across populations [Gautier et al., 2007, Arias et al., 2009]. For bovine populations, among the 28'556 SNPs mapped to 7'507 different RefGenes, which corresponds to 7'433 gene symbols. On average, each SNP mapped within 1.2 RefSeq genes (from 1 to 31, median = 1) while each RefSeq gene was represented by 1.7 SNP (from 1 to 29, median = 1). For ovine populations, among the 44'880 SNPs mapped to 94'094 different otherRefSeqGenes, which corresponds to 31'176 gene symbols. On average, each SNP mapped within 7.9 OtherRefSeq genes (from 1 to 349, median = 6) while each

OtherRefSeq gene was represented by 2.2 SNP (from 1 to 63, median = 1). Genes which included at least one SNP displaying a BF > 20 (decisive evidence for association according to the Jeffrey's rule (Jeffreys, 1961) with a given covariable were considered as candidate genes.

7.2.4 Results

Identification and functional annotation of genes associated with climatic covariables

In order to identify genes involved in adaptation to the Western Indian Ocean climate, we focused on the dataset and performed GEA with BAYPASS [Gautier, 2015] for several climatic covariables representative of the region climate corresponding to 5 original climatic covariables (Bio01, Bio07, Bio12, Bio20 and Bio28) representative of the main climatic covariable categories and related to Annual mean temperature, Temperature annual range, Annual precipitation, Annual mean radiation and Annual mean moisture index (Figure 7.2) and the temperature humidity index, THI. Bayes Factor estimates of the support for association of each SNP with the different covariables are plotted in Figure S7.3 for cattle and Figure S7.4 for sheep. We thus identified 6, 32, 2 and 5 genes associated with Bio07, Bio12 , Bio20 and THI, respectively (Table 7.2) for bovine dataset, and 5, 5 and 13 genes associated with Bio12, Bio20 and THI respectively for ovine dataset (Table 7.3). No genes were associated with two climatic variables and were not in common between the two species.

To obtain a more global view of the main gene functions targeted by the Western Indian Ocean climate, the 45 and 23 genes associated with at least one climatic covariable were then functionally annotated using IPA (Ingenuity (R) Systems). For bovine dataset, the main functional categories, in which 43 genes (excepting MIR12003 and C8H9orf72) that mapped to

IPKB were involved, are listed in Table 7.6.a (see Supporting information Table S7.12 for an exhaustive list of their functional annotation). The top five significant functions belonging to the three IPA categories "Physiological System Development and Function", "Molecular and Cellular Functions", and "Diseases and Disorders" were related to (a) nervous system ("Neurological Disease", "Nervous System Development Disease"), (b) cell death ("Cancer","Cell Death and Survival"), (c) organ and tissue development and morphology ("Organismal Injury and Abnormalities", "Cell Morphology", "Cellular development", "Embryonic Development", "Cellular Function and Maintenance", "Cell Cycle", "Organ morphology", "Organismal development" and "Tissue Morphology"), (d) vascular system ("Cardiovasculaire Disease") and (e) ("Connective Tissue Disorders").

For ovine and following the same workflow, the top five significant functions of IPA (Table 7.6.b) (see Supporting information Table S7.13 for an exhaustive list of their functional annotation), and concerning 20 genes out of the 23 candidate genes, were related to (a) cell death ("Cancer","Cell Death and Survival"), (b) organ and tissue development and morphology ("Organismal Injury and Abnormalities","Cellular Function and Maintenance", "Cell Morphology", "Cellular Assembly and Organisation", "Cellular development", "Organ development" and "Tissue development"), (c) reproductive system ("Reproductive System Disease","Embryonic Development"), (d) nervous system ("Auditory Diseases and Conditions","Visual System Development and Function"), (e) skin properties ("Dermatological Diseases and Conditions") and (f) heart system ("Cardiovascular System Development and Function").

For bovine dataset, 32 and 13 genes participated to 2 significant networks (Supporting information Table S7.10). Overall, the functional annotation of genes associated with climatic variable pointed towards the following main physiological functions involved with (a)

cardiovascular System Development and Function, "Cell Cycle and Organismal Development" (b) "Antigen Presentation", "Cell-mediated Immune Response" and "Cell-To-Cell Signaling and Interaction". For ovine dataset, 20 genes (over 23) participated to 1 significant network (Supporting information Table S7.11) and the functional annotation of these genes associated with climatic variable refered to "Cell Cycle", "Cell Morphology" and "Cellular Assembly and Organization".

Identification of selection footprints in the genome of the local bovine and ovine breeds

If climate was assumed to represent a critical adaptive constraint, other selective pressures may have significantly impacted cattle and sheep populations situated on Indian Ocean littoral. We thus scanned the genome for footprints of adaptive differentiation focusing on the dataset on the XtX statistic [Gautier, 2015]. For bovine dataset, a total of 15 candidate genes under selection could be identified according to their function or to their proximity with the XtX peak (Table 7.4, Supporting information Figure S7.3.c) spread on BTA01, BTA02, BTA04, BTA06, BTA07, BTA09, BTA11, BTA15 and BTA16. Only 13 out of 38 regions contained candidate genes defined by our criteria and no genes were associated with climatic covariables. However, if we considered all the genes contained in regions, 2 regions (#13 and #22) contain genes associated with one climatic covariable : *LEF1* associated with Bio12 in Region #13 and *EXOC6B* associated with Bio12 too in Region #22 (Table 7.2).

For ovine dataset, a total of 15 candidate genes under selection could be identified according to their function or to their proximity with the XtX peak (Table 7.5, Supporting information Figure S7.4.c) spread on OAR03, OAR04, OAR09, OAR15 and OAR19. All the regions harbouring footprints of selection did not contain any gene associated with any of

the considered climatic covariables (Table 7.3) and might thus result from other selective pressures.

TABLE 7.2 – Positions and candidate genes mapped on Btau9 assembly associated with climatic covariable for bovine populations. The THI index was calculated from mean annual temperature (Bio01) and humidity index (Bio28) using formula 5 proposed in [Bohmanova et al., 2007]. The genes marked with ** are found in candidate regions under selection determined in Table 7.4. The gene marked with * is associated with signature of selections identified in Chapitre 5, section 5.2.

Covariable	BTA	Position	RefGene	Gene symbol	BFmc
	1	64,222,505	NM_001076360	TMEM39A	20.01
Temperature	1	108,826,583	NM_001075430	RARRES1	23.03
annual	1	149,676,509	NM_001206014	DYRK1A	23.78
range	3	78,363,470	NM_001076913	MIER1	23.01
	12	66,386,065	NM_001102070	GPC5*	21.20
	24	56,070,207	NM_001206098	WDR7	22.76
	1	145,340,231	NM_001076453	SLC19A1	52.96
	3	98,866,038	NM_001206322	FOXE3	46.94
	6	17,080,889	NM_001192856	LEF1**	31.54
	6	86,710,155	NM_174605	SLC4A4	36.64
	7	25,643,432	NM_001278588	FBN2	30.62
	8	16,699,848	NM_001103088	C8H9orf72	24.39
	8	17,304,615	NM_001105335	PLAA	20.01
	9	36,491,907	NM_001075146	HDAC2	23.82
	10	10,742,098	NM_001281903	CMYA5	20.29
	11	11,642,101	NM_001076416	EXOC6B**	20.06
	12	16,171,675	NM_001045997	CPB2	21.63
	12	25,390,440	NM_001109962	DCLK1	21.43
	12	28,987,000	NM_001205616	FRY	23.57
Annual	13	1,313,761	NM_174817	PLCB1	22.05
precipitation	13	33,074,876	NM_001083670	NSUN6	20.95
	14	65,669,673	NM_001079607	STK3	21.22
	14	69,474,846	NM_001015665	CCNE2	20.30
	14	69,474,846	NM_001102556	INTS8	20.30
	16	39,534,519	NM_001243284	DNM3	27.77
	18	22,223,573	NM_001098142	FTO	27.73
	19	56,948,024	NM_001192456	CD300A	31.68
	23	4,861,885	NM_001191193	HMGCLL1	21.13
	23	43,271,085	NM_001192911	PHACTR1	21.45
	25	109,981	NM_001046617	POLR3K	20.86
	25	109,981	NM_001105400	RHBDF1	20.86
	25	109,981	NM_001166612	SNRNP25	20.86
	25	7,905,066	NR_162331	MIR12003	26.81
	25	18,965,840	NM_001080361	TMEM159	21.38
	25	21,825,112	NM_174587	PRKCB	21.47
	26	19,019,035	NM_001076954	GOLGA7B	24.70
	26	19,019,035	NM_001191289	CRTAC1	24.70
	28	686,6914	NM_001192935	SLC35F3	36.24
Annual mean	2	90,111,726	NM_001078059	C2CD6	21.57
radiation	10	84,699,230	NM_174721	PSEN1	27.77
THI	1	119,126,112	NM_001192286	HLTF	24.08
THI	7	50,631,745	NM_001034636	PAIP2	20.57
THI	7	50,631,745	NM_001102486	MATR3	20.57
THI	15	65,651,813	NM_174013	CD44	21.11
THI	15	65,988,310	NM_001015591	PAMR1	24.77

TABLE 7.3 – Positions and candidate genes associated with climatic covariable for ovine populations. The THI index was calculated from mean annual temperature (Bio01) and humidity index (Bio28) using formula 5 proposed in [Bohmanova et al., 2007]. N corresponds to the number of Gene symbol associated with a candidate SNP.

Covariable	OAR	Position (bp)	Gene symbol	BFmc	N
	7	65,420,670	PELI2	22.21	2
Annual	10	77,161,569	ITGBL1	20.03	3
precipitation	14	38,556,578	AP1G1	21.23	6
	14	38,556,578	AP1G2	21.23	1
	26	29,749,264	UNC5D	28.57	7
	1	68,109,135	TGFBR3	25.01	6
Annual	4	89,121,603	POT1	22.90	4
mean	12	32,887,034	PLD5	21.02	7
radiation	13	62,092,179	BPIFB1	20.64	3
	23	43,556,333	CEP192	24.70	1
THI	3	124,214,371	CG4050	21.39	1
THI	4	92,576,455	PRRT4	22.49	1
THI	4	92,576,455	RBM28	22.49	4
THI	6	61,629,124	GRXCR1	22.74	1
THI	7	76,425,844	CG2065	20.16	1
THI	7	76,425,844	CG2070	20.16	1
THI	7	76,425,844	GSTO1	20.16	3
THI	7	76,425,844	RDH12	20.16	2
THI	7	89,459,525	GTF2A1	25.40	3
THI	14	23,560,155	BBS2	20.97	5
THI	14	23,560,155	OGFOD1	20.97	6
THI	23	43,772,785	FAM210A	26.43	3
THI	23	43,772,785	RNMT	26.43	5

TABLE 7.4 – Regions harbouring footprints of selection based on the XtX measure of differentiation for bovine populations. For each region, the peak XtX value (MAX), its position in Mb (Pos), the putative candidate genes according to region candidate approach (Symb candidate gene), the RefGene localized in the region are indicated. The genes in bold are associated with at least one climatic covariable. The genes in italic were found associated with EHH score in Chapter 5.

Region	BTA	START	END	MAX	N	POSITION	Gene symb	Neighbour RefGene
1	1	4	5	47.24	2	4.66		MIR2284I, MIR2284X, KRTAP11-1
2	1	6	7	43.83	2	6.68		GRIK1
3	1	97	98.5	41.18	2	97.58	MYNN	CLDN11, SKIL, PRKCI, PHC3, SEC62, LRRIQ4, ACTRT3, TERC
4	1	132.5	134	53.63	3	133.81		PCCB
5	1	134.5	135.5	54.96	2	135.01		EPHB1, ANAPC13, SLCO2A1, RAB6B,
6	2	51.5	53	51.63	2	52.21	ZEB2	GTDC1, ARHGAP15
7	2	60.5	61.5	49.22	2	61.01		CXCR4, DARS, MCM6
8	3	81.5	82.5	51.09	2	82.29		PGM1, EFCAB7, ITGB3BP, ALG6
9	4	15	16.5	37.24	2	15.63	MIOS	TAC1, ASNS, C1GALT1, RPA3, ICA1
10	4	101.5	103	46.25	3	102.44	ATP6V0A4	CREB3L2, MIR12013, AKR1D1, SVOPL, TMEM213, KIAA1549, ZC3HAV1L, UBN2, FMC1
11	5	69.5	70.5	33.53	2	69.86		NUAK1, TCP11L2, POLR3B, RFX4, TMEM263, MTERF2, CRY1
12	6	14	15	33.18	2	14.97	ENPEP	MIR2444, PITX2
13	6	16	17.5	34.54	3	16.71		ETNPPL, OSTC, RPL34, **LEF1**, HADH, CYP2U1, SGMS2, PAPSS1
14	7	23.5	25	51.01	2	24.42	ADAMTS19	CHSY3, ISOC1
15	7	85.5	87	31.29	2	86.32		MIR2284D, RASA1
16	8	33	34	54.65	2	33.94		-
17	8	88	89	46.02	2	88.37		GADD45G, SEMA4D, CKS2
18	9	23	24	52.18	2	23.52	SNAP91	PGM3, RWDD2A, ME1, PRSS35, ECHDC1
19	9	56	57.5	53.54	3	56.11		EPHA7
20	9	61.5	63	54.8	2	62.4	RARS2	CNR1, SPACA1, MIR2285DB, AKIRIN2, ORC3, SLC35A1, CFAP206, C9H6orf163, SMIM8, ZNF292, CGA, HTR1E
21	9	71	72.5	50.7	2	71.7		TCF21, TBPL1, SLC2A12, SGK1
22	11	10.5	12	54.24	3	11.09		BOLA3, DGUOK, ACTG2, STAMBP, DUSP11, MIR2295, TPRKB, EGR4, CCT7, PRADC1, SMYD5, RAB11FIP5, MIR2294, *SFXN5*, EMX1, SPR, **EXOC6B**

Region	BTA	START	END	MAX	N	POSITION	Gene symb	Ref gen evoisin
23	11	38	39	47.78	2	38.3	PPP4R3B, PNPT1	PRORSD1, CCDC88A, CFAP36, PNPT1, EFEMP1, MIR217, MIR216A, MIR216B, CCDC85A
24	11	93.5	94.5	49.98	2	94.33	STRBP	PDCL, ZBTB6, RABGAP1, GPR21
25	12	22.5	24	48.84	2	23.07		COG6, LHFPL6, NHLRC3, STOML3, UFM1, TRPC4
26	12	54.5	56	49.18	3	55.5		NDFIP2, SPRY2
27	12	83.5	85	41.82	2	84.42		LIG4, ABHD13, TNFSF13B, LOC107131142, COL4A1
28	14	17.5	18.5	45.43	2	17.61		HAS2
29	15	38	39.5	53.64	2	38.97		PSMA1, COPB1, RRAS2, *SPON1*, MIR2284T, FAR1, MIR2285B-1, PTH, BTBD10, ARNTL, ARNTL
30	15	58	59	47.67	2	58.05		LGR4, LIN7C, BDNF, KIF18A, METTL15
31	15	62.5	64	47.78	2	63.49		ELP4, PAX6, RCN1, EIF3M, CCDC73, PRRG4, QSER1, DEPDC7, TCP11L1, CSTF3, HIPK3
32	15	66	67	52.85	2	66.76	COMMD9	PAMR1, TRIM44, LDLRAD3, PRR5L, TRAF6
33	16	41	42.5	44.79	2	41.81	NPPB	DHRS3, VPS13D, TNFRSF1B, MIIP, MFN2, MFN2, PLOD1, KIAA2013, NPPA, CLCN6, MTHFR, MIR12050, AGTRAP, DRAXIN, MAD2L2, FBXO6, FBXO6, FBXO44, FBXO2, DISP3, ANGPTL7
34	16	65.5	66.5	47.76	2	66.12	RNF2, TRMT1L	C16H1orf21, EDEM3, NIBAN1, MIR12008, IVNS1ABP
35	17	33	34.5	46.14	2	33.66		ANKRD50, SPRY1, MIR12036
36	21	13	14.5	44.85	3	13.59		CHD2, FAM174B
37	26	34	35.5	54.18	2	34.55		HABP2, NRAP, PLEKHS1, NHLRC2, ADRB1, AFAP1L2, MIR2285DG, FAM160B1
38	27	22.5	24	49.67	2	23.23		C27H8orf48, DLC1, TRMT9B

TABLE 7.5 – Regions harbouring footprints of selection based on the XtX measure of differentiation for ovine populations. For each region, the peak XtX value, its position in Mb, the putative candidate genes

Region	OAR	START	END	MAX	N	POSITION	Oth.RefGene	Oth.RefGene in region
1	3	19.00	20.50	24.04	3	19.54	NOL10	
2	4	46.50	48.00	21.23	2	47.19	SYPL1	
3	9	72.50	74.00	22.94	2	73.18	RIMS2, PSME1	
4	10	58.50	60.00	22.29	2	59.36		SLITRK1
5	15	14.50	15.50	22.27	2	15.35	PIWIL4	
6	15	77.00	78.50	20.90	2	77.69	P2RX3, MED19A, MIR466D,LOC781990, PRG3, PRG2	
7	16	33.50	35.00	21.39	3	34.56		DAB2
8	16	42.00	43.50	23.03	2	42.86		CDH6, DROSHA
9	16	64.00	65.50	22.11	2	64.85		MTRR, FASTKD3
10	19	45.00	46.00	24.42	2	45.59	WNT5A, WNT5, WNT5B, LOC10744780	
11	23	8.50	10.00	22.25	2	9.14		DSEL

TABLE 7.6 – Top five significant functions of the putative candidate genes associated with climatic covariables for Zebus (**a**) and Sheep (**b**)

a

Categories	Functions	P-value range	Nb
Diseases and Disorders	Neurological Disease	2.48E-02 – 1.52E-05	16
	Organismal Injury and Abnormalities	2.48E-02 – 1.52E-05	41
	Cancer	2.48E-02 – 1.64E-04	39
	Connective Tissue Disorders	2.48E-02 – 1.64E-04	8
	Cardiovasculaire Disease	2.48E-02 – 2.23E-04	13
Molecular and Cellular Functions	Cell Morphology	2.14E-02 – 1.94E-04	13
	Cell Death and Survival	2.48E-02 – 2.27E-04	11
	Cellular Development	2.48E-02 – 2.27E-04	22
	Cellular Function and Maintenance	2.48E-02 – 2.27E-04	9
	Cell Cycle	2.48E-02 – 3.81E-03	7
Physiological System Development and Function	Nervous System Development and Function	2.48E-02 – 1.52E-05	14
	Organ Morphology	2.42E-02 – 1.52E-05	16
	Organismal Development	2.48E-02 – 1.52E-05	22
	Embryonic Development	2.48E-02 – 6.72E-05	20
	Tissue Morphology	2.44E-02 – 1.94E-04	19

b

Categories	Functions	P-value range	Nb
Diseases and Disorders	Cancer	4.87E-02 – 9.63E-04	20
	Organismal Injury and Abnormalities	4.87E-02 – 9.63E-04	20
	Reproductive System Disease	3.24E-02 – 9.63E-04	4
	Auditory Disease	1.22E-03 – 1.22E-03	1
	Dermatological Diseases and Conditions	4.78E-02 – 1.22E-03	18
Molecular and Cellular Functions	Cellular Function and Maintenance	4.99E-02 – 1.07E-03	5
	Cell Death and Survival	2.64E-02 – 2.43E-03	4
	Cell Morphology	4.99E-02 – 2.43E-03	5
	Cellular Assembly and Organization	4.99E-02 – 2.43E-03	5
	Cellular Development	4.06E-02 – 2.43E-03	4
Physiological System Development and Function	Organ Development	4.87E-02 – 1.07E-03	4
	Tissue Development	4.87E-02 – 1.07E-03	7
	Visual System Development and Function	2.29E-02 – 1.07E-03	2
	Cardiovascular System Development and Function	4.87E-02 – 1.22E-03	1
	Embryonic Development	4.87E-02 – 1.22E-03	6

7.2.5 Discussion

Through climatic association tests, we produced a genomic-map of short-term adaptative response to littoral of Indian Ocean variation that was mainly captured by climatic covariables such as temperature, precipitation, radiation and humidity. The functional analyses carried on the 45 (bovine) and 23 (sheep) genes associated with at least one climatic covariable

brought out several functions or physiological pathways already reported in several Genome Environmental Association studies such as Cancer, Organismal Injury and Abnormalities, Nervous System Development and Function, Tissue morphology, Cell Morphology, Cellular Development, cellular Function and Maintenance, Cellular Assembly and Organization, found in mediterranean [Flori et al., 2019] and chinese [Gao et al., 2017] cattle.

Functions relatives to nervous system ("Neurological Disease", "Nervous System Development and Function" for bovine and "Auditory Diseases and Conditions", "Visual System Development and Function" for ovine) and cell death ("Cancer", "Cell Death and Survival") are in common between the two species. More specifically, several top five biological functions are found for the two species and the three categories : (a) Cancer and Organismal Injury and Abnormalities, (b) Cell Morphology, Cell Death and Survival, Cellular Development, Cellular Function and Maintenance, (c) Embryonic Development.

Skin properties Due to their preponderant importance in heat exchange, skin properties are traits subject to environmental selection pressure. The color gene LEF1 [Porto-Neto et al., 2014, Illa et al., 2021] which was a candidate genes in a region under selection in several local cattle breeds [Gao et al., 2017, Flori et al., 2019] was also found associated with annual precipitation (Bio12) covariable in our study while in [Gao et al., 2017, Flori et al., 2019] this gene was not associated with environment covariable. This gene is localized in bovine region #13. This gene has a central role in hair pigmentation [Guenther et al., 2014] and was found located in QTL for UV-protective eye area pigmentation in cattle [Pausch et al., 2012]. Three genes shown associated with THI index in ovine dataset are involved in pigmentation and skin properties (*RBM28, BBS2, RDH12*) : the gene *RBM28* encodes a nucleolar protein that supports the processing of ribosomal RNAs for inclusion in the 60S large ribosomal subunit [Gerton, 2021] and that regulates hair follicle growth [Warshauer et al., 2015]. The

genes *BBS2* and *RDH12* are envolved in syndrome associated with retinopathy symptoms that are caracterized by a depigmentation of retina [Mackay et al., 2011]. Mutations on *BBS2* gene are responsible of Bardet biedl syndrome, whose symptoms are, in addition to retinopathy, damage to the reproductive system [Innes et al., 2010, Shevach et al., 2015].

Nervous system Another gene involved conjointly in both nervous and reproductive system is *RFX4* gene which is, although not determined as candidate genes under selection, localized in candidate regions for bovine species. The *RFX4* gene, which is a member of Regulatory Factor X (RFX) family of transcriptional regulators that influence MHC class II expression [Aftab et al., 2008], localized in candidate bovine region #11, was found associated with environmental covariables such as annual precipitation [Flori et al., 2019] or environmental composite variables [Gao et al., 2017]. This gene plays a critical role in brain development [Blackshear et al., 2003] and was also found to affect heifer fertility in tropical composition breed Brangus [Fortes et al., 2012]. Other genes involved in nervous system, such as *PRG3* (proteoglycan 3, pro eosinophil major basic protein 2) and *P2RX3* (purinergic receptor P2X 3) [Verma et al., 2017] are candidate genes in ovine region #6 under selection. In the same region, the candidate gene *MED19A* (Mediator subunit 19a) was found part of candidate genes for supernumerary nipple phenotype in sheep [Peng et al., 2017]. The gene *GRXC1* is associated with THI index is implied in hearing loss [Odeh et al., 2010], a function that can take part of in nervous system.

Skeletal and muscle structure Skeletal and muscle structure functions are known to be under selection pressure. These functions are concerned with human selection for the traction capabilities that characterise the uses of local breeds in the tropical context. Hence, the *ZEB2* gene (zinc finger E-box binding homeobox 2), encoding a zinc finger nuclear trans-

cription factor which is involved in numerous processes, is the gene closest to the maximum differentiation signal in bovine region #6. In cattle, this gene is involved in polled Charolais accompanied by congenital malformations [Capitan et al., 2012] and a *de novo* mutation was discovered in Fleckvieh cattle and associated with symptoms including delayed maturity, small body stature and abnormal shape of skull [Gehrke et al., 2020]. In ovine dataset, the gene *RIMS2* (Regulating synaptic membrane exocytosis protein 2), candidate gene in ovine region #3, encodes for a presynaptic active zone protein that regulates synaptic vesicle exocytosis of neurotransmitters and plays an important role in controlling bone remodelling through stimulation, maturation and differentiation of osteoblasts and osteoclasts. Mutations in *RIMS2* have been reported to be significantly associated with scoliosis [Miao et al., 2018]. Moreover this gene was found associated with carcass traits such as bone weight in Simmental beef cattle [Wang et al., 2018].

Tropical climatic conditions also exert selection pressure on skeletal and muscle functions : the gene *EXOC6B*, localized in bovine region #22 although distant from peak position (chr11 : 11 086 534), is also associated with annual precipitation (Bio12) covariable. A mutation of this gene is the cause of spondyloepimetaphyseal dysplasia a genetic condition affecting the bones [Girisha et al., 2016]. The gene *EXOC6B* is part of the exocyst complex, which is critical for cellular trafficking, and mutations in *EXOC6B* have been associated with intellectual disability, language delay, hyperactivity, ear malformations, and craniofacial abnormalities in humans [Wen et al., 2013]. The *FAM210A* (family with sequence similarity 210, member A) genes associated with THI index in ovine dataset is determinant of bone and muscle structure and strenght [Tanaka et al., 2018].

Embryogenes and cancer Several candidate genes *GPC5, WNT5, WNT5A, WNT5B* are involved in Wnt/beta-catenin signaling pathway which is involved in embryogenese and cancer

and plays an important role in skeletal muscle dynamics [Cisternas et al., 2014], intramuscular fat deposition [Jeong et al., 2013] and postnatal skeletal muscle growth and hypertrophy [Steelman et al., 2006]. In pigs, Wnt/beta-catenin signaling pathway mediates fat development and myofiber differentiation and growth [Men et al., 2017] and during embryonic development, Wnt signals control the expressions of myogenic regulatory factors [Men et al., 2017]. More specifically, the *WNT5, WNT5A* and *WNT5B* candidate genes localized in ovine region #10 under selection, were found expressed in adult porcine longissimus dorsi and contribute to differences between porcine genotypes and affected pork quality [Men et al., 2017] while in goat, these genes were higly expressed in longissimus thoraxis muscle in fetal and juvenile domestic goats [Wang et al., 2015b]. Another gene envolved in Wnt/betacatenin pathway is *GPC5*, encoding glypican-5, and associated with temperature annual range covariable (Bio07) was found under selection through EHH scores computing in zebu of Mayotte cattle (Chapitre 5, section 5.2). Several variants of the gene were found associated with acquired nephrotic syndrome [Okamoto et al., 2011]. This gene was downregulated in lung adenocarcinoma tissues compared with adjacent normal lung tissues [Yuan et al., 2016]. It is also a gene candidate found as a genetic locus protective against sudden cardiac arrest [Arking et al., 2010]. This gene is involved in embryonic development and mainly expressed in fetal tissues and is believed to play an important role in growth and differentiation during development through regulation of Wnt, fibroblast growth factor and bone morphogenetic protein pathways in humans [Saunders et al., 1997, Li and Durbin, 2011]. In taurine cattle, this gene is associated with gestation lenght in Charolais breed [Purfield et al., 2019], feed efficiency indicators [Serão et al., 2013] and fat mass [Wang et al., 2015a]. Another gene candidate in ovine region #5 involved in development is *PIWIL4* gene (piwi like RNA-mediated gene silencing 4) [Iliev et al., 2016] is also implied in breast cancer [Wang et al., 2016].

Few regions under selection contained genes associated with any of the considered climatic covariables. This might be the result of other selective pressures. Further association analyses with productions level data or morphology (average size, weight, fertility) or the production of new variables by ecological modelling in order to investigate the response to pathogens [Vajana et al., 2018] could give more informations about the regions under selection.

7.2.6 Supplemental Figures and Tables

TABLE S7.7 – Description of climate covariable from Climond website.

Variable ID	Variable
Bio01	Annual mean temperature (°C)
Bio02	Mean diurnal temperature range (mean(period max-min)) (°C)
Bio03	Isothermality (Bio02 ÷ Bio07)
Bio04	Temperature seasonality (C of V)
Bio05	Max temperature of warmest week (°C)
Bio06	Min temperature of coldest week (°C)
Bio07	Temperature annual range (Bio05-Bio06) (°C)
Bio08	Mean temperature of wettest quarter (°C)
Bio09	Mean temperature of driest quarter (°C)
Bio10	Mean temperature of warmest quarter (°C)
Bio11	Mean temperature of coldest quarter (°C)
Bio12	Annual precipitation (mm)
Bio13	Precipitation of wettest week (mm)
Bio14	Precipitation of driest week (mm)
Bio15	Precipitation seasonality (C of V)
Bio16	Precipitation of wettest quarter (mm)
Bio17	Precipitation of driest quarter (mm)
Bio18	Precipitation of warmest quarter (mm)
Bio19	Precipitation of coldest quarter (mm)
Bio20	Annual mean radiation (W m-2)
Bio21	Highest weekly radiation (W m-2)
Bio22	Lowest weekly radiation (W m-2
Bio23	Radiation seasonality (C of V)
Bio24	Radiation of wettest quarter (W m-2)
Bio25	Radiation of driest quarter (W m-2)
Bio26	Radiation of warmest quarter (W m-2)
Bio27	Radiation of coldest quarter (W m-2)
Bio28	Annual mean moisture index
Bio29	Highest weekly moisture index
Bio30	Lowest weekly moisture index
Bio31	Moisture index seasonality (C of V)
Bio32	Mean moisture index of wettest quarter
Bio33	Mean moisture index of driest quarter
Bio34	Mean moisture index of warmest quarter
Bio35	Mean moisture index of coldest quarter
Bio36	First principal component of the first 35 Bioclim variables
Bio37	Second principal component of the first 35 Bioclim variables
Bio38	Third principal component of the first 35 Bioclim variables
Bio39	Fourth principal component of the first 35 Bioclim variables
Bio40	Fifth principal component of the first 35 Bioclim variables

TABLE S7.8 – Climatic covariable matrix for bovine dataset

Cov	GIR	MOK	SHK	THC	THS	ZMA	ACE	BAN	BRE	KAN	MAD	ONG	PES	RSI	THA	EAZ	MAY
Bio1	24.7	17.4	18.7	26.6	28	18.2	26.4	20.1	27.5	26.2	26.9	28.6	24.4	27.1	26.9	22.5	25.3
Bio7	19.4	12	18	21.6	14.2	16.7	10.2	17.6	10.8	28.5	11.5	20.2	10.8	34.9	33.9	16.5	10.1
Bio12	628	1630	1828	1165	1464	1348	2617	555	2027	600	1659	742	2310	150	290	1368	1265
Bio20	209.3	159.5	192.2	166	152.5	166.8	163.7	185.7	162.9	211.9	170.5	203.8	164.8	227.9	219.6	168.4	237
Bio28	0.5	1.2	0.9	0.7	0.8	0.8	1.3	0.4	1.1	0.4	0.9	0.5	1.2	0.1	0.2	0.7	0.8
THI	66.4	60.4	61.5	67.9	69.1	61	67.9	62.6	68.7	67.6	68.2	69.5	66.2	68.3	68.1	64.6	66.9

TABLE S7.9 – Climatic covariable matrix for ovine dataset

Cov	BGA	BGE	EMZ	GAR	GUR	IDC	OMAY	RMA	SUM
Bio1	25.99	24.86	21.57	26.47	26.88	27.43	25.35	27.27	26.35
Bio7	20.26	19.46	20.47	24.25	11.60	27.75	10.06	17.39	9.98
Bio12	2657.00	3699.00	1073.00	1388.00	2171.00	1134.00	1265.00	354.00	2655.00
Bio20	167.49	162.23	189.54	176.52	158.09	198.23	237.05	168.76	163.10
Bio28	1.11	1.34	0.60	0.74	1.11	0.57	0.84	0.26	1.30
THI	67.48	66.57	63.81	67.83	68.21	68.59	66.92	68.42	67.80

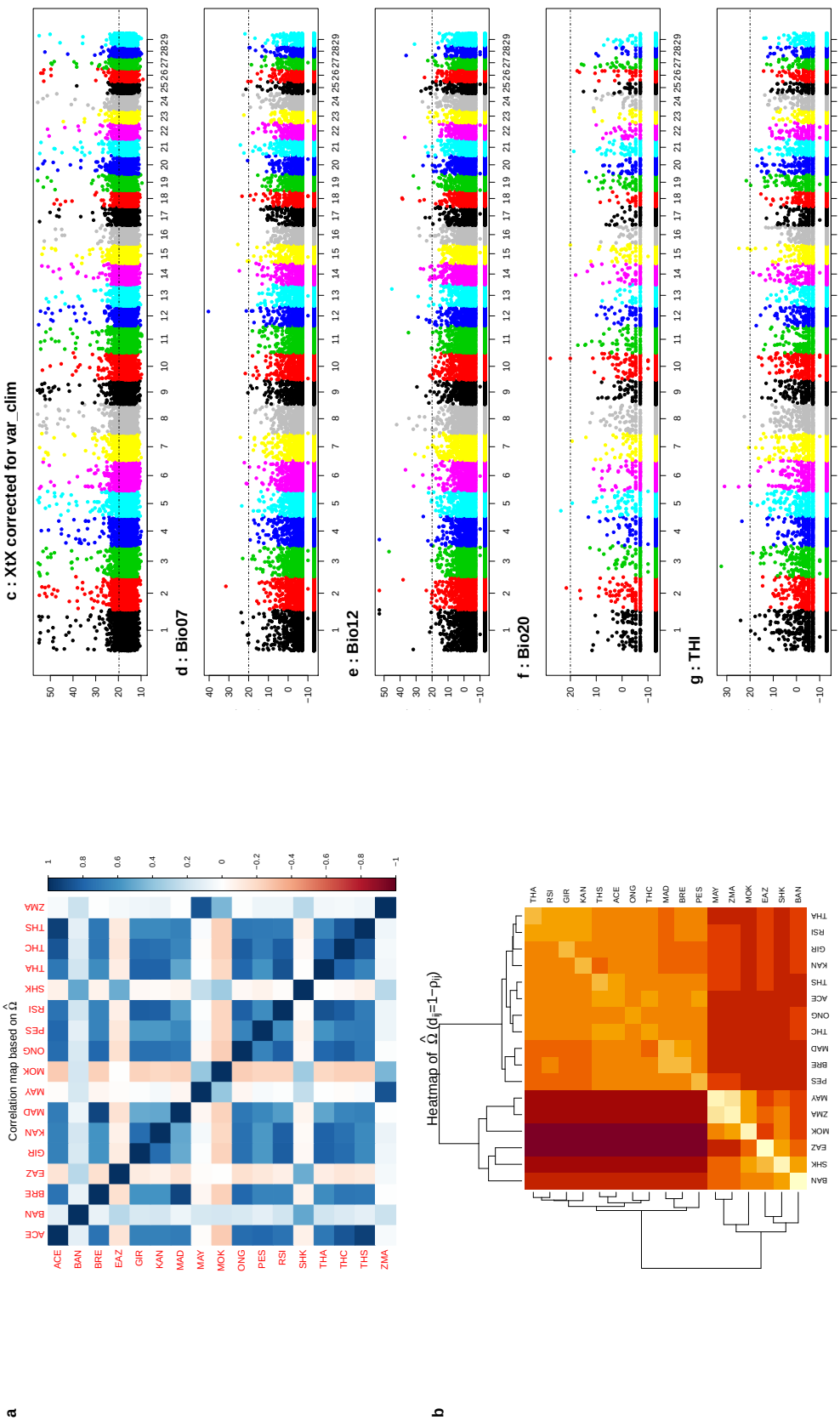


FIGURE S7.3 – Supplemental figures for Genome Environment Analyses with bovine dataset. Correlation matrix (**a**) and Heatmap and clustering tree (**b**) between populations base on estimated Ω matrix. **c** to **g** Manhattan plots for XtX analysis, Bio07, Bio12, Bio20 and THI.

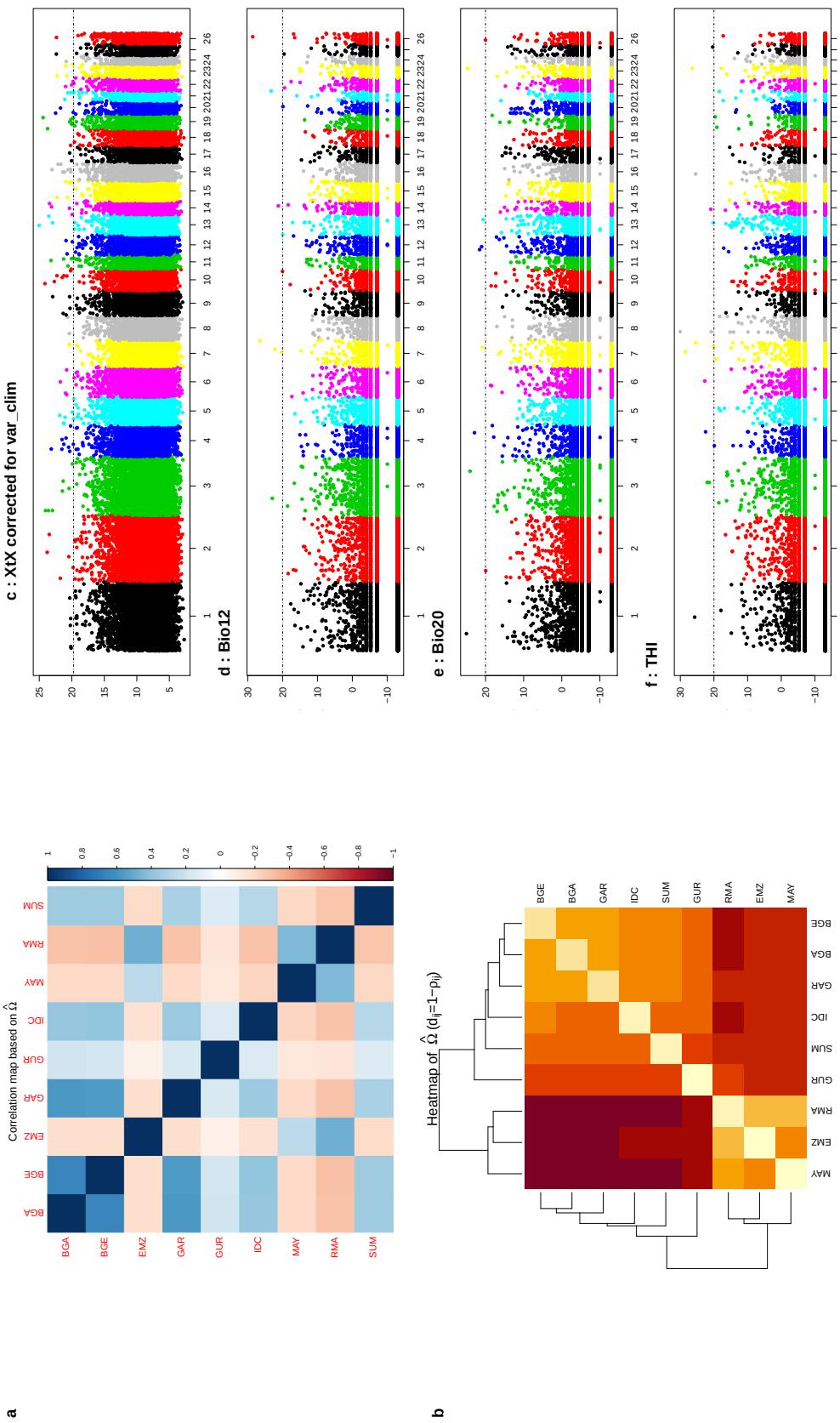


FIGURE S7.4 – Supplemental figures for Genome Environment Analyses with ovine dataset. Correlation matrix (**a**) and Heatmap and clustering tree (**b**) between populations base on estimated Ω matrix. **c** to **f** Manhattan plots for XtX analysis, Bio12, Bio20 and THI.

TABLE S7.10 – Networks bovine BF

ID	Molecules in Network	Score	Focus Molecules	Top Diseases and Functions
1	11,12-epoxyeicosatrienoic acid, 14,15-epoxyeicosatrienoic acid, 15-epi-lipoxin A4, 16,16-dimethylprostaglandin E2, 26s Proteasome, 8,9-epoxyeicosatrienoic acid, Actin, Adaptor protein 1, ADRA1, Akt, Ampa Receptor, ARL11, Atp5e, BCL2, BCR (complex), C17orf97, Calmodulin, calpain, CaMKII, caspase, **CCNE2**, CD164L2, CD3, **CD300A**, CD300LD, **CD44**, CD8B, CDKL2, Ciap, Ck2, **CMYA5**, **CPB2**, **CRTAC1**, CXCR3, Cyb5r3, **DCLK1**, DL glyceraldehyde, **DYRK1A**, EGF, ERK, ERK1/2, ESR1, **FBN2**, Fc receptor, Fibrinogen, Filamin, **FOXE3**, **FRY**, **FTO**, globotriaosylceramide, **GOLGA7B**, **GPC5**, GPR15, GPR55, Groucho, HAS1, HAS3, **HDAC2**, Histone H2b, Histone h3, Histone h4, ***HLTF***, **HMGCLL1**, HTR7, hyaluronidase, IL10, IL12 (family), IL7R, Immunoglobulin, Insulin, **INTS8**, IRS, Jnk, KRAS, Lamin b, **LEF1**, leucovorin, LYN, Mapk, MBNL3, Mi2,***MIER1***, MOB3B, MST1R, MSX2, MUC4, NCAM2, NFkB (complex), Nkx2-2os, NPNT, NRARP, P38 MAPK, **PAIP2**, **PAMR1**, PDPN, **PHACTR1**, PI3K (complex), Pka, PKC , Pkc(s), **PLAA**, **PLCB1**, PLPP3, Pmca, **PRKCB**, PRKX, **PSEN1**, **RARRES1**, **RHBDF1**, Rho gdi, S1PR3, SCN1A, SELL, Sfk, SH3BP2, Shc, SIRT1, **SLC19A1**, SLC35D1, **SLC35F3**, SPZ1, STEAP1, STIM2, **STK3**, Synapsin, Talin, TCR, TEAD, TGFBR, TM4SF5, TMEM158, TNFRSF11A, TRB, TWIST1, TXK, tyrosine kinase, Vegf, Vla-4, ZC3H12D, ZNF260	68	32	[Cardiovascular System Development and Function, Cell Cycle, Organismal Development]
2	ADAM15, ADAR, AKTIP, AMPD2, ANPEP, ANXA11, ATAD1, ATAD3B, ATP1B1, AXL, **C2CD6**, Cd1, CD2, CD300LB, CD8A, CD8B, CDH5, CFP, CHD5, CIRBP, CKMT1A/CKMT1B, COL5A2, CRIM1, CT55, CTNNBIP1, CTNND2, CXADR, CYTIP, DAG1, DAZAP1, DHX33, DMXL2, **DNM3**, DSC2, Dst, EGFR, ELAVL1, EPRS1, ETV6, EXOC6, **EXOC6B**, FAM217B, FAM98A, Fibrin, FKBPL, GAS5, GC, GPM6B, GPR83, GRB2, H2AC1, H2AC7, Hbb-b2, Hd-perinuclear inclusions, ***HLTF***, Hnrnpa3, HNRNPA3, HNRNPH3, HTR7, IARS1, IDH3G, IFNG, IL2, IRAK3, IRF2BP2, LAMC1, LAMC2, Lamin b, LGALS9B, **MATR3**, MERTK, ***MIER1***, MLEC, MST1R, **NSUN6**, NTRK1, PCMTD1, PIGM, PLPP3, PLVAP, POLR3A, POLR3B, POLR3G, POLR3GL, POLR3H, **POLR3K**, PRR3, PTBP2, PTBP3, PXK, RABEP2, RAPH1, RBM15B, RBM6, RO60, RPAP1, RYK, SBF2, SLC11A2, SLC48A1, **SLC4A4**, SLC7A11, SNCB, **SNRNP25**, SOX2-OT, SPATA2, SSR1, ST13, TASOR, TCIRG1, THEMIS, **TMEM159**, TMEM19, **TMEM39A**, TMEM59, TMEM9B, TNFAIP6, TNFRSF11A, TNFRSF21, TNK1, TNK2, TRIM2, TRPC4, TTYH3, TXK, UBC, UMPS, USP9Y, VANGL2, VIM, VSNL1, WARS1, WASL,**WDR7**, YLPM1, ZFYVE21, ZKSCAN1, ZNF474, ZNF609, ZNF668	21	13	[Antigen Presentation, Cell-mediated Immune Response, Cell-To-Cell Signaling and Interaction]

TABLE S7.11 – Networks ovine BF

ID	Molecules in Network	Score	Focus Molecules	Top Diseases and Functions
1	AAGAB, ABCD1, ADAMTS9-AS2, Adaptor, Adaptor protein 1, AIPL1, Ap1 gamma, AP1AR, **AP1G1**, **AP1G2**, AP1M2, AP1S1, AP1S2, AP1S3, AURKA, **BBS2**, BCL3, **BPIFB1**, BRCA1, CALY, CAMK2, CCDC32, CCM2, CCN5, CCND1, CD300C, CDKN1A, CDR1, CELSR1, CENPK, **CEP192**, COL12A1, COL16A1, COL4A6, CRYGS, CYLC2, DCTPP1, DNPH1, DPP3, ENSA, F5, FAM131B, **FAM210A**, FAT10 Cancer Signaling Pathway, FLRT1, FLRT2, GAMT, GFPT2, GNS, GPR183, GPR52, GRHL1, **GRXCR1**, **GSTO1**, **GTF2A1**, GTF2A2, HEATR5B, Histone h3, Hmgb1, HNMT, HTN3, HYAL2, IFI16, IFI6, IFT27, Immunoglobulin, Importin alpha, Inhibin, **ITGBL1**, KLK11, LZTFL1, MAGEA4, MAP4K2, MAP4K5, MED1, miR-223-3p (miRNAs w/seed GUCAGUU), MKX, MTFR2, MYC, NAB1, NECAP1, NFkB (complex), NPTXR, NRXN3, NXNL1, **OGFOD1**, OLIG1, ORC1, PCP4, PDE1B, **PELI2**, Phagosome Formation, PIP5K1C, **PLD5**, POLD2, POLD3, POLD4, **POT1**, **PRRT4**, PTCD2, PTEN, PYCARD, RAMAC, RBKS, **RBM28**, RDH, **RDH12**, **RNMT**, SDCBP, SEC31A, SLC16A3, SLFN5, SP2, SP4, SRGAP1, STK17A, SULT1B1, SULT1C2, SYNRG, TAF4B, TASP1, TBPL1, TFIIA, TGFB1, **TGFBR3**, TKFC, TMEM184B, TMEM219, TMSB10/TMSB4X, TP53, tretinoin, TRIM16, TRPC4AP, UBE2T, UBL7, **UNC5D**, USP42, YBX1, ZFP36L1, ZNF790	48	20	[Cell Cycle, Cell Morphology, Cellular Assembly and Organization]

TABLE S7.12 – 5 first lines of annotation table of candidate genes for bovine dataset sorted by increasing p-value. *The rest of the table is available in the excel file accompanying the manuscript.*

Categories	Functions	Diseases or Functions Annotation	p-value	Molecules	# Molecules
Nervous System Development and Function, Neurological Disease,Organ Morphology, Organismal Development, Organismal Injury and Abnormalities	abnormal morphology	Abnormal morphology of dentate gyrus	1,52E-05	DYRK1A,HDAC2,LEF1,PSEN1	4
Embryonic Development,Organismal Development	morphogenesis	Morphogenesis of embryo	6,72E-05	CD44,FBN2,FTO,HDAC2,LEF1,PSEN1	6
Nervous System Development and Function,Organ Morphology,Organismal Development	morphology	Morphology of hippocampus	1,36E-04	DYRK1A,HDAC2,LEF1,PLCB1,PSEN1	5
Cancer,Connective Tissue Disorders,Organismal Injury and Abnormalities	Ewing sarcoma	Ewing sarcoma	1,64E-04	CD44,FBN2,LEF1,PRKCB	4
Cell Morphology,Nervous System Development and Function,Organ Morphology,Organismal Development,Tissue Morphology	morphology	Morphology of cerebral cortex cells	1,94E-04	DYRK1A,HDAC2,PLCB1,PSEN1	4

TABLE S7.13 – 8 first lines of annotation table of candidate genes for ovine dataset sorted by increasing p-value. *The rest of the table is available in the excel file accompanying the manuscript.*

Categories	Diseases or Functions Annotation	p-value	Molecules	# Molecules
Cancer,Organismal Injury and Abnormalities,Reproductive System Disease	Development of breast carcinoma	9,63E-04	POT1,TGFBR3	2
Cellular Function and Maintenance,Organ Development,Tissue Development,Visual System Development and Function	Maintenance of photoreceptors	1,07E-03	BBS2,RDH12	2
Developmental Disorder,Hereditary Disorder,Ophthalmic Disease,Organismal Injury and Abnormalities	Leber congenital amaurosis type 13	1,22E-03	RDH12	1
Hereditary Disorder,Ophthalmic Disease,Organismal Injury and Abnormalities	Retinitis pigmentosa type 53	1,22E-03	RDH12	1
Dermatological Diseases and Conditions,Hereditary Disorder,Organismal Injury and Abnormalities	Alopecia, neurologic defects, and endocrinopathy syndrome	1,22E-03	RBM28	1
Cardiovascular System Development and Function,Embryonic Development,Organ Development,Organismal Development,Tissue Development	Vasculogenesis of coronary vessel	1,22E-03	TGFBR3	1
Hereditary Disorder,Ophthalmic Disease,Organismal Injury and Abnormalities	Retinitis pigmentosa type 74	1,22E-03	BBS2	1
Auditory Disease,Hereditary Disorder,Neurological Disease,Organismal Injury and Abnormalities	Autosomal recessive deafness type 25	1,22E-03	GRXCR1	1

7.2.7 Supplemental Results : Exploratory analysis

Material and methods An exploratory analysis were conducted on two world-dataset for each species. For bovine populations, the world-datatset, downloaded from Widde plateform [Sempéré et al., 2015] was constitued of 824 individuals grouped in 31 populations and genotyped on 33817 SNPs distributed throughout the 29 autosomes after quality filtering with the following parameters : the minimal SNP genotyping call rate was set to 80% for bovine dataset and the minimal individual genotyping call rate to 80% over all populations and 75% within each population (i.e. SNPs genotyped for less than 75% of the animals from at least one population were discarded). SNPs with a MAF < 0.01 or departing from Hardy-Weinberg equilibrium expectation (exact test p <0.001) were also discarded. Following the same process, for ovine populations the world-dataset was constitued of 983 individuals grouped in 29 populations and genotyped on 32913 variants distributed throughout the 26 autosomes after quality filering with the following parameters : the minimal SNP genotyping call rate was set to 95% and the minimal individual genotyping call rate to 95% over all populations and 75% within each population (i.e. SNPs genotyped for less than 75% of the animals from at least one population were discarded). SNPs with a MAF < 0.01 or departing from Hardy-Weinberg equilibrium expectation (exact test p <0.001) were also discarded.

The two world-dataset were constituted with :

— 31 bovine breeds (Table S7.14) representative of Indian indicine (GIR, NEL, BRM, RSI, THA, KAN, ONG), East-asian indicine (THC, THS, ACE, PES, BRE, MAD), African taurine (LAG, BAO, SOM, NDA), African indicine (MOK, MAY, ZMA, EAZ, BAN, SHK) and European taurine (ANG, HFD, HOL, CHA, MON, SAL, AUB) ancestry, and

— 29 ovine breeds (Table S7.14) representative of African ovine (ADP, AWD, EMZ, OMAY, NQA, RDA, RMA), West Asian ovine (CFT, CHIO, ERS, KRS, MOG, NDZ, QEZ), East Asian ovine (BGA, BGE, GAR, GUR, IDC, SUM), European ovine (BRL, MER, RMB, STX), Oceanian ovine (APD), American ovine (BBB, BCS, BMN, BSI).

Principal Component Analysis (PCA) based on individual SNP genotype data was performed with SMARTPCA [Patterson et al., 2006] and visualized with the R package ggplot2 [Wickham, 2016]. Unsupervised genotype-based hierarchical clustering of the individual animal samples was carried out using the maximum-likelihood method implemented in ADMIXTURE 1.06 [Alexander et al., 2009]. Results were visualized with custom functions in R environment (`http://wwwq.r-project.org`). Within population FIS (Table S7.14) and between population FST were estimated (Table S7.15 and S7.16) with custom R functions implementing the estimator derived by Weir and Cockerham [Weir and Cockerham, 1984].

Results For bovine populations, as previously observed, the first factorial plan of PCA on individual genotyping data (Figure S7.5) recovered the previously described triangle-like two-dimensional global structuring of the worldwide cattle genetic diversity with European taurine (EUT), African taurine (AFT) and indicine (ZEB) populations at the three apexes [Gautier et al., 2009, Gautier et al., 2010, Flori et al., 2014, Decker et al., 2014, Flori et al., 2019].

In accordance with their geographic origin, populations of East Asia clustered together at the right-upper apex that corresponded to indicine origin. We noticed the intermediate position of African indicine populations (BAN, EAZ, MAY, MOK, SHK, ZMA) between AFT and ZEB apex, with an outstanding position of Zebu Moka from Reunion Island. MOK is more positioned at the inside of the triangle, that give some clues of a EUT introgression in this population. Model based hierarchical clustering approaches with varying number K of predefined genetic clusters (Figure S7.5 for K = 3) as implemented in the ADMIXTURE software, confirmed the pattern of the structuring genetic diversity revealed by PCA, in particular when K = 3, if we considered the clusters as overall EUT, AFT and ZEB ancestries. Due to their high level of EUT ancestry estimated by ADMIXTURE software for K = 3 (EUT > 5%), 21 individuals from MAY and 6 individuals from ONG were discarded in principal bovine dataset (Table 7.1) in order to diminished the influence of EUT introgression for analysis of differentiation.

For ovine populations, the first factorial plan of PCA on individuals genotyping data (Figure S7.6) finded a cross-like global structuring the sheep world-wide genetic diversity with the upper left apex grouping local african breeds, the upper right apex for Dorset group, the lower-right apex for Texel and Border type breeds and the lower-keft apex regrouped Indian and South-East Asia sheep breeds. The first component PC1, explaining 5.67 % of the variation, opposed European and Oceanian sheep apart form African, Asian and South-West Asian animals. PC2, explaining 3.04 % of the variation separatated Asian groups from African breeds. We noticed the central position of Merino group and American breeds and the intermediate position of the two African Dorper breeds between local african breeds and Australian Dorset breed. Model based hierarchical clustering approaches with varying number K of predefined genetic clusters (Figure S7.6 for K = 3 to 6) as implemented in the ADMIXTURE software, confirmed the pattern observed on the first factorial plan of PCA.

TABLE S7.14 – Description of 31 cattle and 29 sheep breeds used for exploratory analysis with the fixation index Fis.

		Cattle				Sheep	
Pop	**Name**	**Zone**	**Fis**	**Pop**	**Name**	**Zone**	**Fis**
ACE	Aceh	Asia	0.023	ADP	African Dorper	Africa	-0.011
ANG	Angus	Europe	-0.002	APD	Austr. Poll Dorset	Oceania	-0.003
AUB	Aubrac	Europe	0.010	AWD	African White Dorper	Africa	-0.088
BAN	Boran	Africa	-0.017	BBB	Barbados Black Belly	America	0.069
BAO	Baoule	Africa	-0.013	BCS	Brazilian Creole	America	0.114
BRE	Brebes	Asia	0.030	BGA	Bang. Garole	Asia	0.068
BRM	Brahman	Asia	0.025	BGE	Bang. East	Asia	0.140
CHA	Charolais	Europe	0.001	BMN	Morada Nova	America	0.028
EAZ	E.A Sht.horn Zebu	Africa	0.002	BRL	Border Leicester	Europe	0.002
GIR	Gir	Asia	0.006	BSI	Santa Ines	America	0.023
HFD	Hereford	Europe	0.063	CFT	Cyprus Fat tail	Middle East	-0.036
HOL	Holstein	Europe	-0.001	CHI	Chios	Middle East	0.019
JER	Jersey	Europe	0.000	EMZ	Ethiopian Menz	Africa	0.013
KAN	Kankraj	Asia	-0.052	ERS	Engadine Red Sheep	Europe	-0.009
LAG	Lagune	Africa	0.039	GAR	Garole	Asia	0.024
MAD	Madura	Asia	0.024	GUR	Garut	Asia	-0.004
MAY	Mayotte	Africa	0.043	IDC	Deccani	Asia	0.005
MOK	Moka	Africa	0.019	KRS	Karakas	Middle East	-0.021
MON	Montbeliard	Europe	-0.037	OMAY	Mayotte	Africa	0.159
NDA	N'Dama	Africa	0.008	MER	Merions (Arles)	Europe	-0.002
NEL	Nellore	Asia	-0.045	MOG	Moghani	Middle East	-0.001
ONG	Ongole	Asia	0.007	NDZ	Norduz	Middle East	-0.046
PES	Pesisir	Asia	-0.037	NQA	Namaka Afrik.	Africa	-0.140
RSI	Red Sindhi	Asia	0.068	QEZ	Qezel	Middle East	0.015
SAL	Salers	Europe	0.017	RDA	Ronderib Afrik.	Africa	-0.049
SHK	Sheko	Africa	-0.007	RMA	Red Massai	Africa	-0.003
SOM	Somba	Africa	0.041	RMB	Merinos (Ramb.)	Europe	0.032
THA	Tharparkar	Asia	0.018	STX	Scottish Texel	Europe	-0.053
THC	Thailand (centre)	Asia	0.035	SUM	Sumatra	Asia	0.037
THS	Thailand (south)	Asia	0.007				
ZMA	Zebu of Madagascar	Africa	0.010				

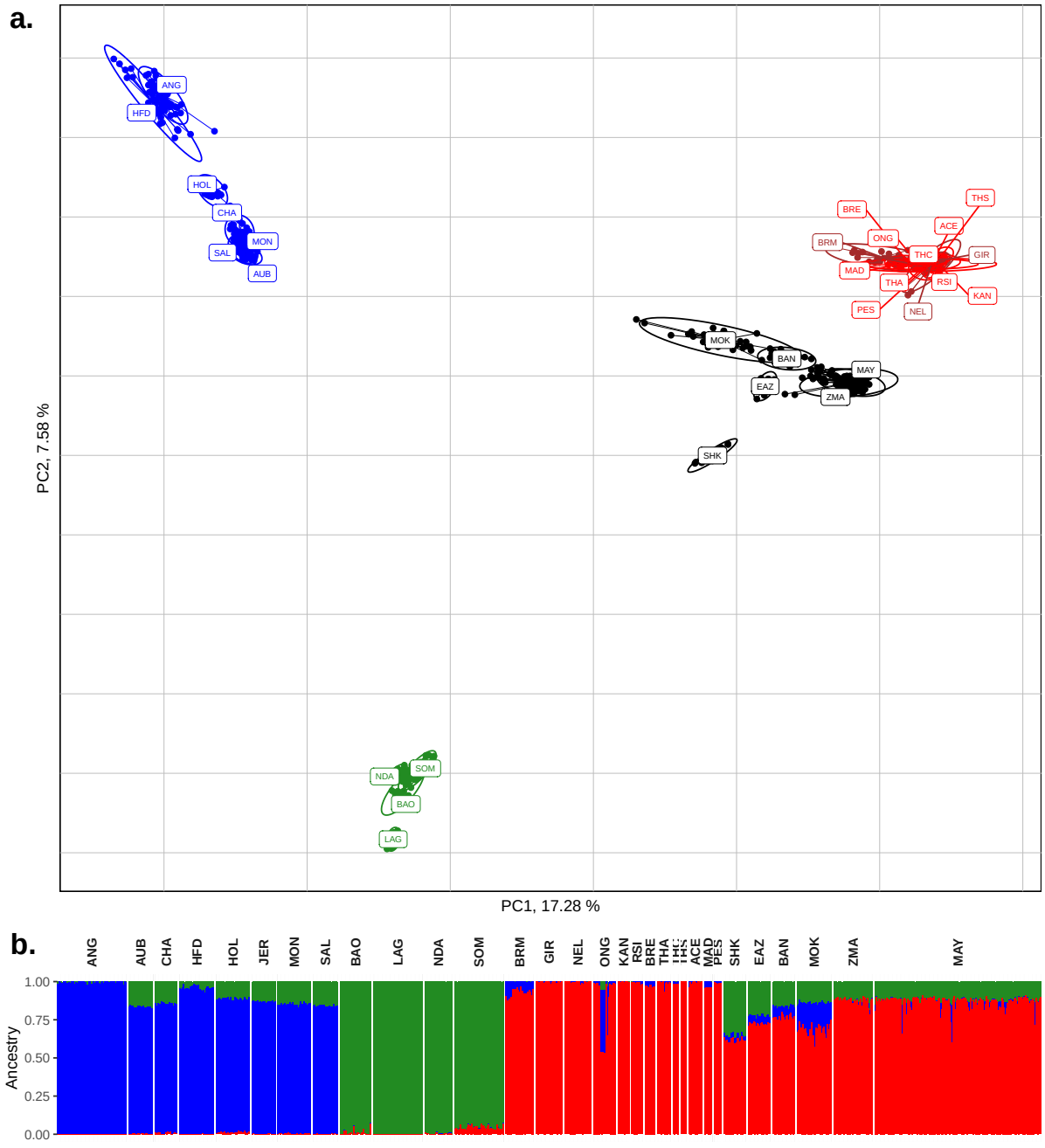


FIGURE S7.5 – Principal Components analysis (PCA) and unsupervised clustering results for bovine populations. **a** PCA results. The individuals according to their coordinates are plotted on the first two principal components, and colored according to their group belonging : EUT in blue, AFT in green, ZEB in red, AFZ in black and Local zebu breeds of Asia in brown. Ellipses characterize the dipersion of each breed around its centre of gravity. **b** Unsupervised hierarchical clustering results of the 824 individuals genotyped for 33'817 SNPs with an inferred number of clusters $K = 3$ obtained with ADMIXTURE software. For each individual, the proportions of each cluster that were interpreted as representative of EUT, AFT and ZEB ancestries are plotted in blue, green and red.

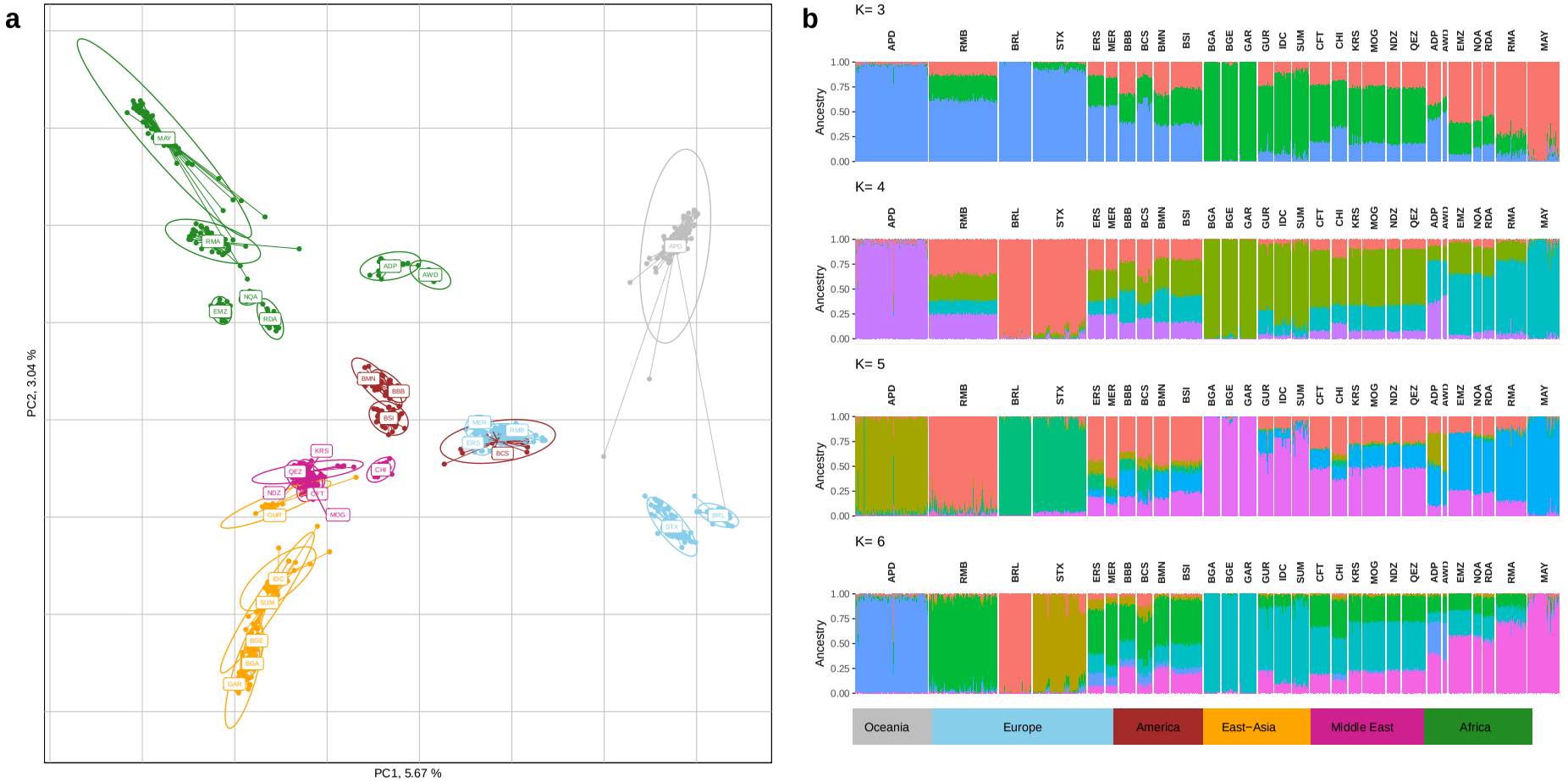


FIGURE S7.6 – Principal Components analysis (PCA) and unsupervised clustering results for ovine populations. **a** PCA results. The individuals according to their coordinates are plotted on the first two principal components, and colored according to their geographic group belonging : Oceania in grey, Europe in blue, America in brown, East Asia in yellow, Middle East in purple and Africa in green. Ellipses characterize the dipersion of each breed around its centre of gravity. **b** Unsupervised clustering results for the 983 individuals genotyped for 32'913 SNPs with an inferred number of clusters from $K = 3$ *to* 6 obtained with ADMIXTURE software.

TABLE S7.15 – Matrix pairwise Fst 31 cattle breeds used for exploratory analysis

	ACE	ANG	AUB	BAN	BAO	BRE	BRM	CHA	EAZ	GIR	HFD	HOL	JER	KAN	LAG	MAD	MAY	MOK	MON	NDA	NEL	ONG	PES	RSI	SAL	SHK	SOM	THA	THC	THS	ZMA
ACE		0.35	0.35	0.13	0.43	0.07	0.07	0.34	0.16	0.09	0.37	0.35	0.39	0.10	0.47	0.10	0.17	0.15	0.36	0.44	0.10	0.05	0.08	0.06	0.37	0.17	0.38	0.07	0.03	0.03	0.18
ANG	0.35		0.12	0.27	0.26	0.33	0.32	0.10	0.27	0.36	0.15	0.13	0.16	0.34	0.30	0.33	0.34	0.23	0.15	0.26	0.36	0.33	0.34	0.33	0.14	0.24	0.25	0.33	0.32	0.34	0.31
AUB	0.35	0.12		0.25	0.23	0.32	0.32	0.05	0.25	0.37	0.13	0.11	0.12	0.34	0.28	0.32	0.32	0.20	0.08	0.23	0.38	0.33	0.34	0.33	0.05	0.21	0.21	0.33	0.30	0.33	0.30
BAN	0.13	0.27	0.25		0.30	0.13	0.10	0.24	0.08	0.13	0.29	0.26	0.30	0.12	0.35	0.15	0.11	0.08	0.27	0.30	0.14	0.10	0.15	0.10	0.27	0.07	0.26	0.10	0.09	0.14	0.11
BAO	0.43	0.26	0.23	0.30		0.41	0.39	0.22	0.28	0.44	0.28	0.25	0.28	0.43	0.12	0.42	0.33	0.25	0.25	0.09	0.44	0.41	0.43	0.41	0.25	0.21	0.04	0.41	0.40	0.43	0.33
BRE	0.07	0.33	0.32	0.13	0.41		0.08	0.31	0.15	0.10	0.34	0.33	0.37	0.11	0.46	0.02	0.17	0.14	0.34	0.42	0.11	0.05	0.12	0.07	0.34	0.16	0.37	0.08	0.05	0.09	0.18
BRM	0.07	0.32	0.32	0.10	0.39	0.08		0.31	0.13	0.05	0.35	0.32	0.36	0.06	0.43	0.11	0.14	0.12	0.33	0.40	0.07	0.04	0.12	0.04	0.34	0.14	0.35	0.04	0.03	0.09	0.15
CHA	0.34	0.10	0.05	0.24	0.22	0.31	0.31		0.24	0.36	0.12	0.09	0.12	0.33	0.27	0.31	0.32	0.19	0.08	0.22	0.37	0.32	0.33	0.32	0.07	0.20	0.21	0.32	0.29	0.32	0.30
EAZ	0.16	0.27	0.25	0.08	0.28	0.15	0.13	0.24		0.16	0.29	0.26	0.30	0.15	0.33	0.17	0.13	0.09	0.27	0.29	0.17	0.14	0.18	0.13	0.27	0.07	0.25	0.14	0.12	0.16	0.13
GIR	0.09	0.36	0.37	0.13	0.44	0.10	0.05	0.36	0.16		0.39	0.37	0.41	0.07	0.47	0.14	0.16	0.16	0.38	0.44	0.08	0.05	0.14	0.04	0.39	0.17	0.39	0.05	0.05	0.12	0.18
HFD	0.37	0.15	0.13	0.29	0.28	0.34	0.35	0.12	0.29	0.39		0.15	0.17	0.36	0.33	0.34	0.37	0.25	0.16	0.28	0.39	0.35	0.35	0.35	0.15	0.26	0.27	0.35	0.32	0.35	0.34
HOL	0.35	0.13	0.11	0.26	0.25	0.33	0.32	0.09	0.26	0.37	0.15		0.15	0.34	0.30	0.33	0.34	0.22	0.13	0.25	0.37	0.33	0.34	0.33	0.13	0.23	0.24	0.33	0.31	0.33	0.31
JER	0.39	0.16	0.12	0.30	0.28	0.37	0.36	0.12	0.30	0.41	0.17	0.15		0.39	0.33	0.37	0.37	0.25	0.15	0.28	0.42	0.38	0.39	0.37	0.15	0.26	0.26	0.38	0.35	0.38	0.35
KAN	0.10	0.34	0.34	0.12	0.43	0.11	0.06	0.33	0.15	0.07	0.36	0.34	0.39		0.47	0.16	0.17	0.15	0.36	0.43	0.09	0.06	0.17	0.05	0.36	0.16	0.38	0.05	0.07	0.14	0.18
LAG	0.47	0.30	0.28	0.35	0.12	0.46	0.43	0.27	0.33	0.47	0.33	0.30	0.33	0.47		0.46	0.37	0.30	0.29	0.15	0.48	0.45	0.47	0.46	0.30	0.27	0.10	0.45	0.45	0.47	0.37
MAD	0.10	0.33	0.32	0.15	0.42	0.02	0.11	0.31	0.17	0.14	0.34	0.33	0.37	0.16	0.46		0.19	0.16	0.34	0.42	0.15	0.08	0.15	0.12	0.34	0.17	0.37	0.12	0.09	0.12	0.20
MAY	0.17	0.34	0.32	0.11	0.33	0.17	0.14	0.32	0.13	0.16	0.37	0.34	0.37	0.17	0.37	0.19		0.07	0.34	0.34	0.17	0.15	0.19	0.14	0.34	0.11	0.30	0.15	0.14	0.18	0.02
MOK	0.15	0.23	0.20	0.08	0.25	0.14	0.12	0.19	0.09	0.16	0.25	0.22	0.25	0.15	0.30	0.16	0.07		0.22	0.26	0.16	0.13	0.17	0.13	0.22	0.07	0.23	0.13	0.11	0.15	0.06
MON	0.36	0.15	0.08	0.27	0.25	0.34	0.33	0.08	0.27	0.38	0.16	0.13	0.15	0.36	0.29	0.34	0.34	0.22		0.25	0.38	0.34	0.35	0.34	0.10	0.23	0.23	0.34	0.32	0.35	0.32
NDA	0.44	0.26	0.23	0.30	0.09	0.42	0.40	0.22	0.29	0.44	0.28	0.25	0.28	0.43	0.15	0.42	0.34	0.26	0.25		0.45	0.42	0.44	0.42	0.25	0.22	0.07	0.42	0.41	0.43	0.34
NEL	0.10	0.36	0.38	0.14	0.44	0.11	0.07	0.37	0.17	0.08	0.39	0.37	0.42	0.09	0.48	0.15	0.17	0.16	0.38	0.45		0.05	0.16	0.06	0.39	0.18	0.40	0.07	0.07	0.13	0.19
ONG	0.05	0.33	0.33	0.10	0.41	0.05	0.04	0.32	0.14	0.05	0.35	0.33	0.38	0.06	0.45	0.08	0.15	0.13	0.34	0.42	0.05		0.11	0.03	0.35	0.14	0.37	0.04	0.02	0.08	0.16
PES	0.08	0.34	0.34	0.15	0.43	0.12	0.12	0.33	0.18	0.14	0.35	0.34	0.39	0.17	0.47	0.15	0.19	0.17	0.35	0.44	0.16	0.11		0.12	0.36	0.18	0.38	0.13	0.10	0.14	0.21
RSI	0.06	0.33	0.33	0.10	0.41	0.07	0.04	0.32	0.13	0.04	0.35	0.33	0.37	0.05	0.46	0.12	0.14	0.13	0.34	0.42	0.06	0.03	0.12		0.35	0.14	0.37	0.02	0.02	0.09	0.15
SAL	0.37	0.14	0.05	0.27	0.25	0.34	0.34	0.07	0.27	0.39	0.15	0.13	0.15	0.36	0.30	0.34	0.34	0.22	0.10	0.25	0.39	0.35	0.36	0.35		0.23	0.23	0.35	0.32	0.35	0.32
SHK	0.17	0.24	0.21	0.07	0.21	0.16	0.14	0.20	0.07	0.17	0.26	0.23	0.26	0.16	0.27	0.17	0.11	0.07	0.23	0.22	0.18	0.14	0.18	0.14	0.23		0.18	0.14	0.13	0.17	0.11
SOM	0.38	0.25	0.21	0.26	0.04	0.37	0.35	0.21	0.25	0.39	0.27	0.24	0.26	0.38	0.10	0.37	0.30	0.23	0.23	0.07	0.40	0.37	0.38	0.37	0.23	0.18		0.37	0.36	0.38	0.30
THA	0.07	0.33	0.33	0.10	0.41	0.08	0.04	0.32	0.14	0.05	0.35	0.33	0.38	0.05	0.45	0.12	0.15	0.13	0.34	0.42	0.07	0.04	0.13	0.02	0.35	0.14	0.37		0.03	0.09	0.16
THC	0.03	0.32	0.30	0.09	0.40	0.05	0.03	0.29	0.12	0.05	0.32	0.31	0.35	0.07	0.45	0.09	0.14	0.11	0.32	0.41	0.07	0.02	0.10	0.02	0.32	0.13	0.36	0.03		0.05	0.15
THS	0.03	0.34	0.33	0.14	0.43	0.09	0.09	0.32	0.16	0.12	0.35	0.33	0.38	0.14	0.47	0.12	0.18	0.15	0.35	0.43	0.13	0.08	0.14	0.09	0.35	0.17	0.38	0.09	0.05		0.19
ZMA	0.18	0.31	0.30	0.11	0.33	0.18	0.15	0.30	0.13	0.18	0.34	0.31	0.35	0.18	0.37	0.20	0.02	0.06	0.32	0.34	0.19	0.16	0.21	0.15	0.32	0.11	0.30	0.16	0.15	0.19	

TABLE S7.16 – Matrix pairwise Fst for 29 sheep breeds used for exploratory analysis

	ADP	APD	AWD	BBB	BCS	BGA	BGE	BMN	BRL	BSI	CFT	CHI	EMZ	ERS	GAR	GUR	IDC	KRS	OMAY	MER	MOG	NDZ	NQA	QEZ	RDA	RMA	RMB	STX	SUM
ADP		0.13	0.11	0.13	0.11	0.19	0.18	0.16	0.23	0.12	0.17	0.16	0.12	0.12	0.20	0.14	0.15	0.13	0.18	0.11	0.11	0.14	0.21	0.10	0.15	0.10	0.13	0.18	0.17
APD	0.13		0.16	0.16	0.12	0.21	0.20	0.18	0.19	0.14	0.19	0.17	0.19	0.12	0.22	0.18	0.18	0.16	0.24	0.12	0.14	0.17	0.26	0.14	0.21	0.19	0.13	0.16	0.20
AWD	0.11	0.16		0.17	0.14	0.23	0.22	0.20	0.27	0.16	0.22	0.20	0.17	0.15	0.25	0.19	0.19	0.17	0.24	0.14	0.14	0.18	0.28	0.14	0.21	0.15	0.16	0.21	0.22
BBB	0.13	0.16	0.17		0.10	0.18	0.17	0.13	0.22	0.10	0.16	0.15	0.13	0.11	0.19	0.14	0.14	0.12	0.19	0.10	0.10	0.13	0.22	0.09	0.17	0.13	0.12	0.17	0.16
BCS	0.11	0.12	0.14	0.10		0.15	0.14	0.13	0.16	0.09	0.14	0.11	0.13	0.07	0.17	0.12	0.11	0.10	0.19	0.05	0.08	0.11	0.21	0.07	0.16	0.13	0.07	0.12	0.14
BGA	0.19	0.21	0.23	0.18	0.15		0.07	0.20	0.28	0.15	0.18	0.18	0.17	0.16	0.10	0.12	0.10	0.14	0.24	0.15	0.11	0.15	0.27	0.10	0.21	0.17	0.17	0.22	0.11
BGE	0.18	0.20	0.22	0.17	0.14	0.07		0.19	0.27	0.14	0.17	0.17	0.16	0.15	0.09	0.12	0.09	0.13	0.23	0.15	0.10	0.14	0.26	0.10	0.20	0.17	0.16	0.21	0.11
BMN	0.16	0.18	0.20	0.13	0.13	0.20	0.19		0.25	0.06	0.19	0.17	0.15	0.13	0.22	0.17	0.16	0.15	0.22	0.12	0.12	0.16	0.25	0.12	0.20	0.15	0.14	0.20	0.19
BRL	0.23	0.19	0.27	0.22	0.16	0.28	0.27	0.25		0.20	0.26	0.24	0.25	0.18	0.29	0.25	0.24	0.23	0.31	0.18	0.20	0.24	0.33	0.20	0.28	0.25	0.18	0.17	0.27
BSI	0.12	0.14	0.16	0.10	0.09	0.15	0.14	0.06	0.20		0.14	0.12	0.12	0.08	0.16	0.12	0.11	0.10	0.18	0.08	0.08	0.11	0.20	0.08	0.15	0.12	0.10	0.15	0.14
CFT	0.17	0.19	0.22	0.16	0.14	0.18	0.17	0.19	0.26	0.14		0.14	0.15	0.14	0.20	0.16	0.14	0.12	0.22	0.14	0.09	0.13	0.25	0.09	0.20	0.16	0.15	0.21	0.18
CHI	0.16	0.17	0.20	0.15	0.11	0.18	0.17	0.17	0.24	0.12	0.14		0.15	0.11	0.19	0.15	0.14	0.12	0.22	0.11	0.09	0.13	0.25	0.09	0.19	0.16	0.13	0.18	0.17
EMZ	0.12	0.19	0.17	0.13	0.13	0.17	0.16	0.15	0.25	0.12	0.15	0.15		0.13	0.18	0.13	0.13	0.10	0.14	0.12	0.08	0.12	0.19	0.08	0.14	0.06	0.14	0.20	0.16
ERS	0.12	0.12	0.15	0.11	0.07	0.16	0.15	0.13	0.18	0.08	0.14	0.11	0.13		0.17	0.13	0.12	0.10	0.20	0.06	0.08	0.11	0.22	0.07	0.16	0.14	0.08	0.13	0.15
GAR	0.20	0.22	0.25	0.19	0.17	0.10	0.09	0.22	0.29	0.16	0.20	0.19	0.18	0.17		0.14	0.11	0.15	0.25	0.17	0.12	0.17	0.28	0.12	0.23	0.19	0.18	0.23	0.13
GUR	0.14	0.18	0.19	0.14	0.12	0.12	0.12	0.17	0.25	0.12	0.16	0.15	0.13	0.13	0.14		0.10	0.11	0.20	0.12	0.09	0.12	0.22	0.08	0.16	0.13	0.14	0.19	0.12
IDC	0.15	0.18	0.19	0.14	0.11	0.10	0.09	0.16	0.24	0.11	0.14	0.14	0.13	0.12	0.11	0.10		0.09	0.20	0.11	0.07	0.11	0.23	0.06	0.17	0.14	0.13	0.19	0.10
KRS	0.13	0.16	0.17	0.12	0.10	0.14	0.13	0.15	0.23	0.10	0.12	0.12	0.10	0.10	0.15	0.11	0.09		0.18	0.09	0.04	0.06	0.21	0.03	0.15	0.11	0.11	0.17	0.13
OMAY	0.18	0.24	0.24	0.19	0.19	0.24	0.23	0.22	0.31	0.18	0.22	0.22	0.14	0.20	0.25	0.20	0.20	0.18		0.19	0.16	0.19	0.25	0.15	0.20	0.11	0.20	0.25	0.23
MER	0.11	0.12	0.14	0.10	0.05	0.15	0.15	0.12	0.18	0.08	0.14	0.11	0.12	0.06	0.17	0.12	0.11	0.09	0.19		0.07	0.11	0.21	0.07	0.16	0.13	0.04	0.13	0.14
MOG	0.11	0.14	0.14	0.10	0.08	0.11	0.10	0.12	0.20	0.08	0.09	0.09	0.08	0.08	0.12	0.09	0.07	0.04	0.16	0.07		0.05	0.18	0.01	0.13	0.10	0.10	0.15	0.11
NDZ	0.14	0.17	0.18	0.13	0.11	0.15	0.14	0.16	0.24	0.11	0.13	0.13	0.12	0.11	0.17	0.12	0.11	0.06	0.19	0.11	0.05		0.23	0.05	0.17	0.13	0.13	0.18	0.14
NQA	0.21	0.26	0.28	0.22	0.21	0.27	0.26	0.25	0.33	0.20	0.25	0.25	0.19	0.22	0.28	0.22	0.23	0.21	0.25	0.21	0.18	0.23		0.18	0.19	0.17	0.21	0.27	0.26
QEZ	0.10	0.14	0.14	0.09	0.07	0.10	0.10	0.12	0.20	0.08	0.09	0.09	0.08	0.07	0.12	0.08	0.06	0.03	0.15	0.07	0.01	0.05	0.18		0.12	0.09	0.09	0.15	0.10
RDA	0.15	0.21	0.21	0.17	0.16	0.21	0.20	0.20	0.28	0.15	0.20	0.19	0.14	0.16	0.23	0.16	0.17	0.15	0.20	0.16	0.13	0.17	0.19	0.12		0.12	0.17	0.22	0.20
RMA	0.10	0.19	0.15	0.13	0.13	0.17	0.17	0.15	0.25	0.12	0.16	0.16	0.06	0.14	0.19	0.13	0.14	0.11	0.11	0.13	0.10	0.13	0.17	0.09	0.12		0.14	0.20	0.16
RMB	0.13	0.13	0.16	0.12	0.07	0.17	0.16	0.14	0.18	0.10	0.15	0.13	0.14	0.08	0.18	0.14	0.13	0.11	0.20	0.04	0.10	0.13	0.21	0.09	0.17	0.14		0.14	0.16
STX	0.18	0.16	0.21	0.17	0.12	0.22	0.21	0.20	0.17	0.15	0.21	0.18	0.20	0.13	0.23	0.19	0.19	0.17	0.25	0.13	0.15	0.18	0.27	0.15	0.22	0.20	0.14		0.21
SUM	0.17	0.20	0.22	0.16	0.14	0.11	0.11	0.19	0.27	0.14	0.18	0.17	0.16	0.15	0.13	0.12	0.10	0.13	0.23	0.14	0.11	0.14	0.26	0.10	0.20	0.16	0.16	0.21	

TABLE S7.17 – Results of clustering for parameter value K = 3 on individuals from bovine MAY and ONG populations.

ID	Cluster 1 (AFT)	Clust 2 (EUT)	Clust 3 (ZEB)	ID	Clust 1 (AFT)	Clust 2 (EUT)	Clust 3 (ZEB)
MAY_2	0.150288	0.070803	0.778908	MAY_224	0.09766	1E-05	0.90233
MAY_3	0.10839	1E-05	0.8916	MAY_226	0.122991	1E-05	0.876999
MAY_6	0.10913	1E-05	0.89086	MAY_228	0.120697	0.011221	0.868082
MAY_11	0.115468	1E-05	0.884522	MAY_236	0.152265	0.129495	0.71824
MAY_12	0.117869	0.004845	0.877286	MAY_237	0.115585	1E-05	0.884405
MAY_14	0.107143	1E-05	0.892847	MAY_239	0.10697	1E-05	0.89302
MAY_15	0.114404	1E-05	0.885586	MAY_243	0.139374	0.100459	0.760167
MAY_21	0.115833	1E-05	0.884157	MAY_246	0.149406	0.0316	0.818995
MAY_22	0.10489	1E-05	0.8951	MAY_248	0.106483	1E-05	0.893507
MAY_25	0.108951	1E-05	0.891039	MAY_252	0.145049	0.040749	0.814202
MAY_28	0.108844	1E-05	0.891146	MAY_253	0.121698	1E-05	0.878292
MAY_35	0.130518	0.007132	0.86235	MAY_255	0.11733	0.009225	0.873444
MAY_36	0.148576	0.107654	0.74377	MAY_263	0.114978	1E-05	0.885012
MAY_37	0.133608	0.034329	0.832064	MAY_267	0.14312	0.01403	0.84285
MAY_40	0.120145	0.01005	0.869805	MAY_268	0.15079	0.118591	0.730619
MAY_43	0.115505	1E-05	0.884485	MAY_269	0.151029	0.250257	0.598714
MAY_48	0.104449	1E-05	0.895541	MAY_273	0.11281	1E-05	0.88718
MAY_49	0.117914	1E-05	0.882076	MAY_275	0.096898	1E-05	0.903092
MAY_51	0.12931	0.019072	0.851617	MAY_276	0.13765	0.0414	0.82095
MAY_52	0.09846	1E-05	0.90153	MAY_279	0.122381	1E-05	0.877609
MAY_56	0.108029	1E-05	0.891961	MAY_280	0.120023	1E-05	0.879967
MAY_63	0.103087	1E-05	0.896903	MAY_285	0.119936	1E-05	0.880054
MAY_66	0.110024	1E-05	0.889966	MAY_286	0.115556	1E-05	0.884434
MAY_70	0.099873	1E-05	0.900117	MAY_290	0.109979	1E-05	0.890011
MAY_71	0.110297	1E-05	0.889693	MAY_293	0.124304	0.091324	0.784372
MAY_75	0.091944	1E-05	0.908046	MAY_296	0.104187	1E-05	0.895803
MAY_78	0.104152	1E-05	0.895838	MAY_300	0.120147	1E-05	0.879843
MAY_80	0.124827	1E-05	0.875163	MAY_303	0.133808	1E-05	0.866182
MAY_82	0.130665	0.010232	0.859102	MAY_304	0.138007	0.034839	0.827154
MAY_83	0.122566	1E-05	0.877424	MAY_305	0.134711	0.059158	0.806131
MAY_85	0.122247	0.026229	0.851524	MAY_306	0.108485	1E-05	0.891505
MAY_86	0.11138	1E-05	0.88861	MAY_311	0.13218	0.049497	0.818322
MAY_87	0.124603	0.003941	0.871455	MAY_312	0.116688	1E-05	0.883302
MAY_89	0.136896	0.148838	0.714266	MAY_314	0.145353	0.089646	0.765001
MAY_91	0.107787	1E-05	0.892203	MAY_316	0.133636	0.004669	0.861695
MAY_94	0.130328	1E-05	0.869662	MAY_319	0.110044	1E-05	0.889946
MAY_95	0.132954	1E-05	0.867036	MAY_321	0.140481	0.01097	0.848549
MAY_96	0.113455	1E-05	0.886535	MAY_324	0.119215	0.002677	0.878108
MAY_98	0.103407	1E-05	0.896583	MAY_326	0.140496	0.023852	0.835651
MAY_101	0.151973	0.039213	0.808814	MAY_327	0.108505	1E-05	0.891485
MAY_106	0.098601	1E-05	0.901389	MAY_328	0.120383	1E-05	0.879607
MAY_108	0.129024	1E-05	0.870966	MAY_331	0.122416	1E-05	0.877574
MAY_113	0.13142	1E-05	0.86857	MAY_334	0.132471	0.054513	0.813016
MAY_115	0.137525	0.024579	0.837896	MAY_336	0.129717	0.014659	0.855624
MAY_119	0.112417	1E-05	0.887573	MAY_342	0.097476	1E-05	0.902514
MAY_125	0.11251	1E-05	0.88748	MAY_345	0.125436	0.001624	0.87294
MAY_127	0.126907	1E-05	0.873083	MAY_346	0.120672	1E-05	0.879318
MAY_129	0.10882	1E-05	0.89117	MAY_348	0.134056	0.038815	0.827128
MAY_131	0.115026	1E-05	0.884964	MAY_352	0.123532	1E-05	0.876458
MAY_132	0.16532	0.174485	0.660195	MAY_355	0.116207	1E-05	0.883783
MAY_135	0.117492	1E-05	0.882498	MAY_357	0.116699	1E-05	0.883291
MAY_136	0.104622	1E-05	0.895368	MAY_361	0.119948	1E-05	0.880042
MAY_139	0.10484	1E-05	0.89515	MAY_364	0.129785	1E-05	0.870205
MAY_141	0.105836	1E-05	0.894154	MAY_367	0.137465	0.048776	0.813759
MAY_143	0.109675	1E-05	0.890315	MAY_369	0.128782	0.034485	0.836733
MAY_151	0.1197	1E-05	0.88029	MAY_371	0.107308	1E-05	0.892682
MAY_153	0.112913	1E-05	0.887077	MAY_372	0.121556	1E-05	0.878434
MAY_155	0.112443	1E-05	0.887547	MAY_373	0.139986	0.02939	0.830624
MAY_157	0.110124	1E-05	0.889866	MAY_376	0.136426	0.079237	0.784337
MAY_159	0.152514	0.186759	0.660727	MAY_379	0.133141	0.042544	0.824314
MAY_161	0.120338	1E-05	0.879652	MAY_382	0.141426	1E-05	0.858564
MAY_163	0.107501	1E-05	0.892489	MAY_386	0.151287	0.119346	0.729367
MAY_165	0.113586	1E-05	0.886404	MAY_388	0.12227	1E-05	0.87772
MAY_166	0.138532	0.079895	0.781573	MAY_391	0.116085	1E-05	0.883905
MAY_170	0.115105	1E-05	0.884885	MAY_395	0.120503	0.101579	0.777918
MAY_173	0.14077	0.072112	0.787118	ONG_614520210	1E-05	0.006958	0.993032
MAY_175	0.123365	1E-05	0.876625	ONG_614520220	1E-05	0.013511	0.986479
MAY_179	0.12827	0.044148	0.827582	ONG_614520240	1E-05	0.007778	0.992212
MAY_180	0.137411	0.049699	0.81289	ONG_614520280	1E-05	0.006996	0.992994
MAY_183	0.10739	1E-05	0.8926	ONG_614520290	1E-05	0.036723	0.963267
MAY_189	0.101796	1E-05	0.898194	ONG_614520300	1E-05	0.025386	0.974604
MAY_192	0.108738	1E-05	0.891252	ONG_614520310	0.053224	0.411506	0.53527
MAY_193	0.104113	1E-05	0.895877	ONG_614520320	0.053619	0.411848	0.534533
MAY_194	0.105488	1E-05	0.894502	ONG_614520330	0.054002	0.409838	0.53616
MAY_196	0.115213	1E-05	0.884777	ONG_614520340	0.05617	0.41205	0.53178
MAY_199	0.140451	1E-05	0.859539	ONG_614520350	0.055902	0.413263	0.530835
MAY_202	0.110058	1E-05	0.889932	ONG_614520370	1E-05	0.016685	0.983305
MAY_205	0.092385	1E-05	0.907605	ONG_614520380	0.034722	0.321137	0.644142
MAY_207	0.106162	1E-05	0.893828	ONG_614520390	0.005421	0.096762	0.897817
MAY_209	0.106811	1E-05	0.893179	ONG_614520560	1E-05	0.015865	0.984125
MAY_210	0.107331	1E-05	0.892659	ONG_614520570	1E-05	0.019016	0.980974
MAY_212	0.107948	1E-05	0.892042	ONG_614520580	1E-05	0.018515	0.981475
MAY_216	0.11201	1E-05	0.88798	ONG_614520590	1E-05	0.00805	0.99194
MAY_218	0.101239	1E-05	0.898751	ONG_614520600	1E-05	0.01226	0.98773
MAY_219	0.106744	1E-05	0.893246	ONG_614520610	1E-05	0.014026	0.985964

7.3 Discussion du chapitre

Dans ce chapitre, nous avons menés deux études d'association génome-environnement à l'échelle de l'Océan Indien pour chaque espèce bovine et ovine. Les études d'association génome-environnement permettent de déterminer si les régions candidates sous sélection contiennent des marqueurs SNPs associés avec des variables climatiques. Si oui, ces régions génomiques particulières aurait subi une pression de sélection environnementale. La variabilité génétique associée à l'adaptation a été auparavant étudiée par des analyses d'association avec les variables climatiques populationnelles qui ont déjà été conduites sur plusieurs zones géographiques comme la Chine [Gao et al., 2017] et le bassin méditerranéen [Flori et al., 2019] chez les bovins, à l'échelle mondiale [Lv et al., 2014, Cao et al., 2021] et plus localement en Ethiopie [Wiener et al., 2021] chez les ovins. Il est a noter que [Lv et al., 2014] avait déjà effectuer une GEA en intégrant des populations ovines africaines (RMA) et asiatiques (SUM, GUR) (données SNP-50K ovines) que nous avons utiliser. Après une conversion des positions sur l'assemblage oviAri4, nous n'avons trouvé aucune des régions associés aux covariables climatiques chez [Lv et al., 2014] en commun avec celles que nous avons déterminé (Table 7.3) ou encore les régions sous sélection (Table 7.5).

Compte-tenu du manque de SNPs associés à une covariable climatique présents dans les régions génomiques candidates sous sélection, il est probable que le seuil de détermination de la significativité des BFMc soit trop élevé : en effet, si nous avons appliqué un seuil égal à celui utilisé par [Flori et al., 2019] (BFMC = 20), d'autres études ont utilisés un seuil plus bas (BFMc = 15 [Gao et al., 2017], BFMc = 10 [Wiener et al., 2021]) qui n'entache pas pour autant la validité des résultats selon la règle de Jeffrey [Jeffreys, 1998]. En outre, d'autres tests d'association peuvent être conduits avec des covariables regroupant des phénotypes populationnels (couleur de la robe, poids vif [Gautier, 2015]), permettant ainsi de définir d'autres pression de sélection à l'origine des signatures de sélection.

Dans ce projet de thèse, la détection des signatures de sélection a été menée avec des méthodes de différentiation (XtX, logiciel BayPass [Gautier, 2015]) ou de méthodes basées sur l'étendue du déséquilibre de liaison dans la population (score iHS [Voight et al., 2006], package rehh [Gautier et al., 2017]) ou entre populations (score Rsb [Tang et al., 2007], package rehh [Gautier et al., 2017]). Peu de régions sous sélection pour ces deux types de statistiques sont en commun : chez les bovins, on retrouve les gènes *SFXN5* (Rsb ZMA vs ZEB) et *SPON1* (Rsb May vs EAZ) dans les régions #22 et #29 (Table 7.4). La région #22 contient également un gène associé avec une covariable climatique. Pour les moutons, on retrouve dans la région #10 (OAR19 : 45-46 Mb) (Table 7.5) les gènes *ERC2* et *CACNA2D3* associés avec des scores Rsb en comparaison avec des populations ovines africaines.

Nos hypothèses de détection des signatures de sélection se basent sur le fait que le motif de ces signaux résulte d'un seul évènement de sélection sur un seul trait et que la fréquence allélique une fois la fixation atteinte est de 100 %. Or, sachant que l'adaptation à des environnements particuliers résulte de l'intéraction de plusieurs traits phénotypiques [Lv et al., 2014]

contrôlés par un réseau complexe de gènes intéragissant entre eux plutôt que par l'action d'un seul gène candidat, la sélection pour les traits complexes laisse peu ou pas de signatures classiques de sélection en raison d'une faible sélection agissant sur plusieurs sites à travers le génome [Kemper et al., 2014] et aboutissant à des motifs de balayage partiel [Barghi et al., 2020].

Bibliographie

[Aftab et al., 2008] Aftab, S., Semenec, L., Chu, J. S.-C., and Chen, N. (2008). Identification and characterization of novel human tissue-specific RFX transcription factors. *BMC Evolutionary Biology*, 8(1) :226.

[Alexander et al., 2009] Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9) :1655–1664.

[Arias et al., 2009] Arias, J. A., Keehan, M., Fisher, P., Coppieters, W., and Spelman, R. (2009). A high density linkage map of the bovine genome. *BMC Genetics*, 10(1) :18.

[Arking et al., 2010] Arking, D. E., Reinier, K., Post, W., Jui, J., Hilton, G., O'Connor, A., Prineas, R. J., Boerwinkle, E., Psaty, B. M., Tomaselli, G. F., Rea, T., Sotoodehnia, N., Siscovick, D. S., Burke, G. L., Marban, E., Spooner, P. M., Chakravarti, A., and Chugh, S. S. (2010). Genome-Wide Association Study Identifies GPC5 as a Novel Genetic Locus Protective against Sudden Cardiac Arrest. *PLOS ONE*, 5(3) :e9879.

[Barghi et al., 2020] Barghi, N., Hermisson, J., and Schlötterer, C. (2020). Polygenic adaptation : a unifying framework to understand positive selection. *Nature Reviews Genetics*, 21(12) :769–781.

[Beaujard, 2019a] Beaujard, P. (2019a). *The Worlds of the Indian Ocean : A Global History*. Cambridge University Press, 1 edition.

[Beaujard, 2019b] Beaujard, P. (2019b). *The worlds of the Indian Ocean : A global history : Volume 2 : From the Seventh Century to the Fifteenth Century CE*, volume 2. Cambridge University Press, Cambridge, United Kingdom ; New York, NY.

[Blackshear et al., 2003] Blackshear, P. J., Graves, J. P., Stumpo, D. J., Cobos, I., Rubenstein, J. L. R., and Zeldin, D. C. (2003). Graded phenotypic response to partial and complete deficiency of a brain-specific transcript variant of the winged helix transcription factor RFX4. *Development*, 130(19) :4539–4552.

[Bohmanova et al., 2007] Bohmanova, J., Misztal, I., and Cole, J. B. (2007). Temperature-Humidity Indices as Indicators of Milk Production Losses due to Heat Stress. *Journal of Dairy Science*, 90(4) :1947–1956.

[Cao et al., 2021] Cao, Y.-H., Xu, S.-S., Shen, M., Chen, Z.-H., Gao, L., Lv, F.-H., Xie, X.-L., Wang, X.-H., Yang, H., Liu, C.-B., Zhou, P., Wan, P.-C., Zhang, Y.-S., Yang, J.-Q., Pi, W.-H., Hehua, E., Berry, D. P., Barbato, M., Esmailizadeh, A., Nosrati, M., Salehian-Dehkordi, H., Dehghani-Qanatqestani, M., Dotsev, A. V., Deniskova, T. E., Zinovieva, N. A., Brem, G., Štěpánek, O., Ciani, E., Weimann, C., Erhardt, G., Mwacharo, J. M., Ahbara, A., Han, J.-L., Hanotte, O., Miller, J. M., Sim, Z., Coltman, D., Kantanen, J., Bruford, M. W., Lenstra, J. A., Kijas, J., and Li, M.-H. (2021). Historical Introgression from Wild Relatives Enhanced Climatic Adaptation and Resistance to Pneumonia in Sheep. *Molecular Biology and Evolution*, 38(3) :838–855.

[Capitan et al., 2012] Capitan, A., Allais-Bonnet, A., Pinton, A., Guienne, B. M.-L., Bourhis, D. L., Grohs, C., Bouet, S., Clément, L., Salas-Cortes, L., Venot, E., Chaffaux, S., Weiss, B., Delpeuch, A., Noé, G., Rossignol, M.-N., Barbey, S., Dozias, D., Cobo, E., Barasc, H., Auguste, A., Pannetier, M., Deloche, M.-C., Lhuilier, E., Bouchez, O., Esquerré, D., Salin, G., Klopp, C., Donnadieu, C., Chantry-Darmon, C., Hayes, H., Gallard, Y., Ponsart, C., Boichard, D., and Pailhoux, E. (2012). A 3.7 Mb Deletion Encompassing ZEB2 Causes a Novel Polled and Multisystemic Syndrome in the Progeny of a Somatic Mosaic Bull. *PLOS ONE*, 7(11) :e49084.

[Cardoso et al., 2015] Cardoso, C. C., Peripolli, V., Amador, S. A., Brandão, E. G., Esteves, G. I. F., Sousa, C. M. Z., França, M. F. M. S., Gonçalves, F. G., Barbosa, F. A., Montalvão, T. C., Martins, C. F., Neto, A. M. F., and McManus, C. (2015). Physiological and thermographic response to heat stress in zebu cattle. *Livestock Science*, 182 :83–92.

[Cisternas et al., 2014] Cisternas, P., Henriquez, J. P., Brandan, E., and Inestrosa, N. C. (2014). Wnt Signaling in Skeletal Muscle Dynamics : Myogenesis, Neuromuscular Synapse and Fibrosis. *Molecular Neurobiology*, 49(1) :574–589.

[Coop et al., 2010] Coop, G., Witonsky, D., Di Rienzo, A., and Pritchard, J. K. (2010). Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics*, 185(4) :1411–1423.

[Decker et al., 2014] Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Molina Alcalá, A., Sonstegard, T. S., Hanotte, O., Götherström, A., Seabury, C. M., Praharani, L., Babar, M. E., Correia de Almeida Regitano, L., Yildiz, M. A., Heaton, M. P., Liu, W.-S., Lei, C.-Z., Reecy, J. M., Saif-Ur-Rehman, M., Schnabel, R. D., and Taylor, J. F. (2014). Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *PLoS Genetics*, 10(3) :e1004254.

[Dunnington, 2020] Dunnington (2020). ggspatial : Spatial Data Framework for ggplot2. *Dewey*.

[Ficetola et al., 2017] Ficetola, G. F., Mazel, F., and Thuiller, W. (2017). Global determinants of zoogeographical boundaries. *Nature Ecology & Evolution*, 1(4) :1–7.

[Flori et al., 2019] Flori, L., Moazami-Goudarzi, K., Alary, V., Araba, A., Boujenane, I., Boushaba, N., Casabianca, F., Casu, S., Ciampolini, R., D'Acier, A. C., Coquelle, C., Delgado, J.-V., El-Beltagi, A., Hadjipavlou, G., Jousselin, E., Landi, V., Lauvie, A., Lecomte, P., Ligda, C., Marinthe, C., Martinez, A., Mastrangelo, S., Menni, D., Moulin, C.-H., Osman, M.-A., Pineau, O., Portolano, B., Rodellar, C., Saïdi-Mehtar, N., Sechi, T., Sempéré, G., Thévenon, S., Tsiokos, D., Laloë, D., and Gautier, M. (2019). A genomic map of climate adaptation in Mediterranean cattle breeds. *Molecular Ecology*, 28(5) :1009–1029.

[Flori et al., 2014] Flori, L., Thevenon, S., Dayo, G.-K., Senou, M., Sylla, S., Berthier, D., Moazami-Goudarzi, K., and Gautier, M. (2014). Adaptive admixture in the West African bovine hybrid zone : insight from the Borgou population. *Molecular Ecology*, 23(13) :3241–3257. WOS :000338014900009.

[Fortes et al., 2012] Fortes, M. R. S., Snelling, W. M., Reverter, A., Nagaraj, S. H., Lehnert, S. A., Hawken, R. J., DeAtley, K. L., Peters, S. O., Silver, G. A., Rincon, G., Medrano, J. F., Islas-Trejo, A., and Thomas, M. G. (2012). Gene network analyses of first service conception in Brangus heifers : Use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors1. *Journal of Animal Science*, 90(9) :2894–2906.

[Frichot et al., 2013] Frichot, E., Schoville, S. D., Bouchard, G., and François, O. (2013). Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. *Molecular Biology and Evolution*, 30(7) :1687–1699.

[Förstner and Moonen, 2003] Förstner, W. and Moonen, B. (2003). A metric for covariance matrices. *Geodesy-the Challenge of the 3rd Millennium*, pages 299–309.

[Gao et al., 2017] Gao, Y., Gautier, M., Ding, X., Zhang, H., Wang, Y., Wang, X., Faruque, M. O., Li, J., Ye, S., Gou, X., Han, J., Lenstra, J. A., and Zhang, Y. (2017). Species composition and environmental adaptation of indigenous Chinese cattle. *Scientific Reports*, 7(1) :1–14.

[Gautier, 2015] Gautier, M. (2015). Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. *Genetics*, 201(4) :1555–1579.

[Gautier et al., 2007] Gautier, M., Faraut, T., Moazami-Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., Boland, A., Garnier, J.-G., Boichard, D., Lathrop, G. M., Gut, I. G., and Eggen, A. (2007). Genetic and Haplotypic Structure in 14 European and African Cattle Breeds. *Genetics*, 177(2) :1059–1070.

[Gautier et al., 2009] Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., Moazami-Goudarzi, K., and Foulley, J.-L. (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC genomics*, 10 :550.

[Gautier et al., 2017] Gautier, M., Klassmann, A., and Vitalis, R. (2017). rehh 2.0 : a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular Ecology Resources*, 17(1) :78–90.

[Gautier et al., 2010] Gautier, M., Laloë, D., and Moazami-Goudarzi, K. (2010). Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PloS One*, 5(9).

[Gautier and Naves, 2011] Gautier, M. and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, 20(15) :3128–3143.

[Gehrke et al., 2020] Gehrke, L. J., Upadhyay, M., Heidrich, K., Kunz, E., Klaus-Halla, D., Weber, F., Zerbe, H., Seichter, D., Graf, A., Krebs, S., Blum, H., Capitan, A., Thaller, G., and Medugorac, I. (2020). A de novo frameshift mutation in ZEB2 causes polledness, abnormal skull shape, small body stature and subfertility in Fleckvieh cattle. *Scientific Reports*, 10(1) :17032.

[Gerton, 2021] Gerton, J. L. (2021). Faulty ribosome biogenesis underlies the ribosomopathy alopecia, neurological defects, endocrinopathy (ANE) syndrome. *Proceedings of the National Academy of Sciences*, 118(24).

[Girisha et al., 2016] Girisha, K. M., Kortüm, F., Shah, H., Alawi, M., Dalal, A., Bhavani, G. S., and Kutsche, K. (2016). A novel multiple joint dislocation syndrome associated with a homozygous nonsense variant in the EXOC6B gene. *European Journal of Human Genetics*, 24(8) :1206–1210.

[Guenther et al., 2014] Guenther, C. A., Tasic, B., Luo, L., Bedell, M. A., and Kingsley, D. M. (2014). A molecular basis for classic blond hair color in Europeans. *Nature Genetics*, 46(7) :748–752.

[Günther and Coop, 2013] Günther, T. and Coop, G. (2013). Robust Identification of Local Adaptation from Allele Frequencies. *Genetics*, 195(1) :205–220.

[Hijmans, 2020] Hijmans, R. J. (2020). *raster : Geographic Data Analysis and Modeling*.

[Iliev et al., 2016] Iliev, R., Stanik, M., Fedorko, M., Poprach, A., Vychytilova-Faltejskova, P., Slaba, K., Svoboda, M., Fabian, P., Pacik, D., Dolezel, J., and Slaby, O. (2016). Decreased expression levels of PIWIL1, PIWIL2, and PIWIL4 are associated with worse survival in renal cell carcinoma patients. *OncoTargets and therapy*, 9 :217–222.

[Illa et al., 2021] Illa, S. K., Mukherjee, S., Nath, S., and Mukherjee, A. (2021). Genome-Wide Scanning for Signatures of Selection Revealed the Putative Genomic Regions and Candidate Genes Controlling Milk Composition and Coat Color Traits in Sahiwal Cattle. *Frontiers in Genetics*, 12 :699422.

[Innes et al., 2010] Innes, A., Boycott, K., Puffenberger, E., Redl, D., MacDonald, I., Chudley, A., Beaulieu, C., Perrier, R., Gillan, T., Wade, A., and Parboosingh, J. (2010). A founder mutation in BBS2 is responsible for Bardet-Biedl syndrome in the Hutterite population : utility of SNP arrays in genetically heterogeneous disorders. *Clinical Genetics*, 78(5) :424–431.

[Jeffreys, 1998] Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.

[Jeong et al., 2013] Jeong, J. Y., Kim, J. S., Nguyen, T. H., Lee, H. J., and Baik, M. (2013). Wnt/B-catenin signaling and adipogenic genes are associated with intramuscular fat content in the longissimus dorsi muscle of Korean cattle. *Animal Genetics*, 44(6) :627–635.

[Jiang et al., 2014] Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J. F., Faraut, T., Wu, C., Muzny, D. M., Li, Y., Zhang, W., Stanton, J.-A., Brauning, R., Barris, W. C., Hourlier, T., Aken, B. L., Searle, S. M. J., Adelson, D. L., Bian, C., Cam, G. R., Chen, Y., Cheng, S., DeSilva, U., Dixen, K., Dong, Y., Fan, G., Franklin, I. R., Fu, S., Fuentes-Utrilla, P., Guan, R., Highland, M. A., Holder, M. E., Huang, G., Ingham, A. B., Jhangiani, S. N., Kalra, D., Kovar, C. L., Lee, S. L., Liu, W., Liu, X., Lu, C., Lv, T., Mathew, T., McWilliam, S., Menzies, M., Pan, S., Robelin, D., Servin, B., Townley, D., Wang, W., Wei, B., White, S. N., Yang, X., Ye, C., Yue, Y., Zeng, P., Zhou, Q., Hansen, J. B., Kristiansen, K., Gibbs, R. A., Flicek, P., Warkup, C. C., Jones, H. E., Oddy, V. H., Nicholas, F. W., McEwan, J. C., Kijas, J. W., Wang, J.,

Worley, K. C., Archibald, A. L., Cockett, N., Xu, X., Wang, W., and Dalrymple, B. P. (2014). The sheep genome illuminates biology of the rumen and lipid metabolism. *Science*, 344(6188) :1168–1173.

[Joost et al., 2007] Joost, S., Bonin, A., Bruford, M. W., Després, L., Conord, C., Erhardt, G., and Taberlet, P. (2007). A spatial analysis method (SAM) to detect candidate loci for selection : towards a landscape genomics approach to adaptation. *Molecular Ecology*, 16(18) :3955–3969.

[Kemper et al., 2014] Kemper, K. E., Saxton, S. J., Bolormaa, S., Hayes, B. J., and Goddard, M. E. (2014). Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics*, 15(1) :246.

[Kijas et al., 2012] Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., and Dalrymple, B. (2012). Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biology*, 10(2).

[Kim et al., 2020] Kim, K., Kwon, T., Dessie, T., Yoo, D., Mwai, O. A., Jang, J., Sung, S., Lee, S., Salim, B., Jung, J., Jeong, H., Tarekegn, G. M., Tijjani, A., Lim, D., Cho, S., Oh, S. J., Lee, H.-K., Kim, J., Jeong, C., Kemp, S., Hanotte, O., and Kim, H. (2020). The mosaic genome of indigenous African cattle as a unique genetic resource for African pastoralism. *Nature Genetics*, 52(10) :1099–1110.

[Kim et al., 2017] Kim, S.-J., Ka, S., Ha, J.-W., Kim, J., Yoo, D., Kim, K., Lee, H.-K., Lim, D., Cho, S., Hanotte, O., Mwai, O. A., Dessie, T., Kemp, S., Oh, S. J., and Kim, H. (2017). Cattle genome-wide analysis reveals genetic signatures in trypanotolerant N'Dama. *BMC Genomics*, 18(1) :371.

[Koppen, 1936] Koppen, W. (1936). Das geographische system der klimat. *Handbuch der klimatologie*, page 46.

[Kriticos et al., 2012] Kriticos, D. J., Webber, B. L., Leriche, A., Ota, N., Macadam, I., Bathols, J., and Scott, J. K. (2012). CliMond : global high-resolution historical and future scenario climate surfaces for bioclimatic modelling : CliMond : climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution*, 3(1) :53–64.

[Li and Durbin, 2011] Li, H. and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357) :493–496.

[Lv et al., 2014] Lv, F.-H., Agha, S., Kantanen, J., Colli, L., Stucki, S., Kijas, J. W., Joost, S., Li, M.-H., and Ajmone Marsan, P. (2014). Adaptations to Climate-Mediated Selective Pressures in Sheep. *Molecular Biology and Evolution*, 31(12) :3324–3343.

[Mackay et al., 2011] Mackay, D. S., Dev Borman, A., Moradi, P., Henderson, R. H., Li, Z., Wright, G. A., Waseem, N., Gandra, M., Thompson, D. A., Bhattacharya, S. S., Holder, G. E., Webster, A. R., and Moore, A. T. (2011). RDH12 retinopathy : novel mutations and phenotypic description. *Molecular Vision*, 17 :2706–2716.

[Matukumalli et al., 2009] Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., O'Connell, J., Moore, S. S., Smith, T. P. L., Sonstegard, T. S., and Van Tassell, C. P. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PloS One*, 4(4) :e5350.

[Men et al., 2017] Men, X.-m., Deng, B., Tao, X., Qi, K.-k., and Xu, Z.-w. (2017). Wnt gene expression in adult porcine longissimus dorsi and its association with muscle fiber type, energy metabolism, and meat quality. *Journal of Integrative Agriculture*, 16(1) :144–150.

[Miao et al., 2018] Miao, J., Wang, X., Bao, J., Jin, S., Chang, T., Xia, J., Yang, L., Zhu, B., Xu, L., Zhang, L., and others (2018). Multimarker and rare variants genomewide association studies for bone weight in Simmental cattle. *Journal of Animal Breeding and Genetics*, 135(3) :159–169.

[Nicolazzi et al., 2015] Nicolazzi, E. L., Caprera, A., Nazzicari, N., Cozzi, P., Strozzi, F., Lawley, C., Pirani, A., Soans, C., Brew, F., Jorjani, H., Evans, G., Simpson, B., Tosser-Klopp, G., Brauning, R., Williams,

J. L., and Stella, A. (2015). SNPchiMp v.3 : integrating and standardizing single nucleotide polymorphism data for livestock species. *BMC Genomics*, 16(1) :283.

[Odeh et al., 2010] Odeh, H., Hunker, K. L., Belyantseva, I. A., Azaiez, H., Avenarius, M. R., Zheng, L., Peters, L. M., Gagnon, L. H., Hagiwara, N., Skynner, M. J., Brilliant, M. H., Allen, N. D., Riazuddin, S., Johnson, K. R., Raphael, Y., Najmabadi, H., Friedman, T. B., Bartles, J. R., Smith, R. J. H., and Kohrman, D. C. (2010). Mutations in Grxcr1 Are The Basis for Inner Ear Dysfunction in the Pirouette Mouse. *The American Journal of Human Genetics*, 86(2) :148–160.

[Okamoto et al., 2011] Okamoto, K., Tokunaga, K., Doi, K., Fujita, T., Suzuki, H., Katoh, T., Watanabe, T., Nishida, N., Mabuchi, A., Takahashi, A., Kubo, M., Maeda, S., Nakamura, Y., and Noiri, E. (2011). Common variation in GPC5 is associated with acquired nephrotic syndrome. *Nature Genetics*, 43(5) :459–463.

[Patterson et al., 2006] Patterson, N., Price, A. L., and Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genetics*, 2(12).

[Pausch et al., 2012] Pausch, H., Wang, X., Jung, S., Krogmeier, D., Edel, C., Emmerling, R., Götz, K.-U., and Fries, R. (2012). Identification of QTL for UV-Protective Eye Area Pigmentation in Cattle by Progeny Phenotyping and Genome-Wide Association Analysis. *PLOS ONE*, 7(5) :e36346.

[Peel et al., 2007] Peel, M. C., Finlayson, B. L., and McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, 11(5) :1633–1644.

[Peng et al., 2017] Peng, W.-F., Xu, S.-S., Ren, X., Lv, F.-H., Xie, X.-L., Zhao, Y.-X., Zhang, M., Shen, Z.-Q., Ren, Y.-L., Gao, L., Shen, M., Kantanen, J., and Li, M.-H. (2017). A genome-wide association study reveals candidate genes for the supernumerary nipple phenotype in sheep (Ovis aries). *Animal Genetics*, 48(5) :570–579.

[Porto-Neto et al., 2014] Porto-Neto, L. R., Reverter, A., Prayaga, K. C., Chan, E. K. F., Johnston, D. J., Hawken, R. J., Fordyce, G., Garcia, J. F., Sonstegard, T. S., Bolormaa, S., Goddard, M. E., Burrow, H. M., Henshall, J. M., Lehnert, S. A., and Barendse, W. (2014). The Genetic Architecture of Climatic Adaptation of Tropical Cattle. *PLoS ONE*, 9(11).

[Purfield et al., 2019] Purfield, D. C., Evans, R. D., Carthy, T. R., and Berry, D. P. (2019). Genomic Regions Associated With Gestation Length Detected Using Whole-Genome Sequence Data Differ Between Dairy and Beef Cattle. *Frontiers in Genetics*, 10 :1068.

[Rahman et al., 2018] Rahman, M. B., Schellander, K., Luceño, N. L., and Van Soom, A. (2018). Heat stress responses in spermatozoa : Mechanisms and consequences for cattle fertility. *Theriogenology*, 113 :102–112.

[Rosen et al., 2020] Rosen, B. D., Bickhart, D. M., Schnabel, R. D., Koren, S., Elsik, C. G., Tseng, E., Rowan, T. N., Low, W. Y., Zimin, A., Couldrey, C., Hall, R., Li, W., Rhie, A., Ghurye, J., McKay, S. D., Thibaud-Nissen, F., Hoffman, J., Murdoch, B. M., Snelling, W. M., McDaneld, T. G., Hammond, J. A., Schwartz, J. C., Nandolo, W., Hagen, D. E., Dreischer, C., Schultheiss, S. J., Schroeder, S. G., Phillippy, A. M., Cole, J. B., Van Tassell, C. P., Liu, G., Smith, T. P. L., and Medrano, J. F. (2020). De novo assembly of the cattle reference genome with single-molecule sequencing. *GigaScience*, 9(3).

[Saunders et al., 1997] Saunders, S., Paine-Saunders, S., and Lander, A. D. (1997). Expression of the Cell Surface Proteoglycan Glypican-5 Is Developmentally Regulated in Kidney, Limb, and Brain. *Developmental Biology*, 190(1) :78–93.

[Sempéré et al., 2015] Sempéré, G., Moazami-Goudarzi, K., Eggen, A., Laloë, D., Gautier, M., and Flori, L. (2015). WIDDE : a Web-Interfaced next generation database for genetic diversity exploration, with a first application in cattle. *BMC Genomics*, 16(1).

[Serão et al., 2013] Serão, N. V., González-Peña, D., Beever, J. E., Faulkner, D. B., Southey, B. R., and Rodriguez-Zas, S. L. (2013). Single nucleotide polymorphisms and haplotypes associated with feed efficiency in beef cattle. *BMC Genetics*, 14(1) :94.

[Shevach et al., 2015] Shevach, E., Ali, M., Mizrahi-Meissonnier, L., McKibbin, M., El-Asrag, M., Watson, C. M., Inglehearn, C. F., Ben-Yosef, T., Blumenfeld, A., Jalas, C., Banin, E., and Sharon, D. (2015). Association Between Missense Mutations in the BBS2 Gene and Nonsyndromic Retinitis Pigmentosa. *JAMA Ophthalmology*, 133(3) :312–318.

[Steelman et al., 2006] Steelman, C. A., Recknor, J. C., Nettleton, D., and Reecy, J. M. (2006). Transcriptional profiling of myostatin-knockout mice implicates Wnt signaling in postnatal skeletal muscle growth and hypertrophy. *The FASEB Journal*, 20(3) :580–582.

[Tanaka et al., 2018] Tanaka, K.-i., Xue, Y., Nguyen-Yamamoto, L., Morris, J. A., Kanazawa, I., Sugimoto, T., Wing, S. S., Richards, J. B., and Goltzman, D. (2018). FAM210A is a novel determinant of bone and muscle structure and strength. *Proceedings of the National Academy of Sciences*, 115(16) :E3759–E3768.

[Tang et al., 2007] Tang, K., Thornton, K. R., and Stoneking, M. (2007). A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *PLoS Biology*, 5(7) :e171.

[Vajana et al., 2018] Vajana, E., Barbato, M., Colli, L., Milanesi, M., Rochat, E., Fabrizi, E., Mukasa, C., Del Corvo, M., Masembe, C., Muwanika, V. B., Kabi, F., Sonstegard, T. S., Huson, H. J., Negrini, R., , T. N. C., Joost, S., and Ajmone-Marsan, P. (2018). Combining Landscape Genomics and Ecological Modelling to Investigate Local Adaptation of Indigenous Ugandan Cattle to East Coast Fever. *Frontiers in Genetics*, 9 :385.

[Verma et al., 2017] Verma, S., Thakur, A., Katoch, S., Shekhar, C., Wani, A. H., Kumar, S., Dohroo, S., Singh, G., and Sharma, M. (2017). Differences in innate and adaptive immune response traits of Pahari (Indian non-descript indigenous breed) and Jersey crossbred cattle. *Veterinary Immunology and Immunopathology*, 192(Supplement C) :20–27.

[Voight et al., 2006] Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A Map of Recent Positive Selection in the Human Genome. *PLoS Biology*, 4(3) :e72.

[Wang et al., 2015a] Wang, M. D., Dzama, K., Hefer, C. A., and Muchadeyi, F. C. (2015a). Genomic population structure and prevalence of copy number variations in South African Nguni cattle. *BMC Genomics*, 16(1) :894.

[Wang et al., 2018] Wang, X., Miao, J., Xia, J., Chang, T., E, G., Bao, J., Jin, S., Xu, L., Zhang, L., Zhu, B., Gao, X., Chen, Y., Li, J., and Gao, H. (2018). Identifying novel genes for carcass traits by testing GxE interaction through genome-wide meta-analysis in Chinese Simmental beef cattle. *Livestock Science*, 212 :75–82.

[Wang et al., 2015b] Wang, Y. H., Zhang, C. L., Plath, M., Fang, X. T., Lan, X. Y., Zhou, Y., and Chen, H. (2015b). Global transcriptional profiling of longissimus thoracis muscle tissue in fetal and juvenile domestic goat using RNA sequencing. *Animal Genetics*, 46(6) :655–665.

[Wang et al., 2016] Wang, Z., Liu, N., Shi, S., Liu, S., and Lin, H. (2016). The Role of PIWIL4, an Argonaute Family Protein, in Breast Cancer *. *Journal of Biological Chemistry*, 291(20) :10646–10658.

[Wangkumhang et al., 2015] Wangkumhang, P., Wilantho, A., Shaw, P. J., Flori, L., Moazami-Goudarzi, K., Gautier, M., Duangjinda, M., Assawamakin, A., and Tongsima, S. (2015). Genetic analysis of Thai cattle reveals a Southeast Asian indicine ancestry. *PeerJ*, 3 :e1318.

[Warshauer et al., 2015] Warshauer, E., Samuelov, L., Sarig, O., Vodo, D., Bindereif, A., Kanaan, M., Gat, U., Fuchs-Telem, D., Shomron, N., Farberov, L., Pasmanik-Chor, M., Nardini, G., Winkler, E., Meilik, B., Petit, I., Aberdam, D., Paus, R., Sprecher, E., and Nousbeck, J. (2015). RBM28, a protein deficient in ANE syndrome, regulates hair follicle growth via miR-203 and p63. *Experimental Dermatology*, 24(8) :618–622.

[Weir and Cockerham, 1984] Weir, B. S. and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, (19) :395–420.

[Wen et al., 2013] Wen, J., Lopes, F., Soares, G., Farrell, S. A., Nelson, C., Qiao, Y., Martell, S., Badukke, C., Bessa, C., Ylstra, B., Lewis, S., Isoherranen, N., Maciel, P., and Rajcan-Separovic, E. (2013). Phenotypic

and functional consequences of haploinsufficiency of genes from exocyst and retinoic acid pathway due to a recurrent microdeletion of 2p13.2. *Orphanet Journal of Rare Diseases*, 8(1) :100.

[Wickham, 2016] Wickham, H. (2016). ggplot2 : Elegant Graphics for Data Analysis. *Springer-Verlag New York*.

[Wiener et al., 2021] Wiener, P., Robert, C., Ahbara, A., Salavati, M., Abebe, A., Kebede, A., Wragg, D., Friedrich, J., Vasoya, D., Hume, D. A., Djikeng, A., Watson, M., Prendergast, J. G. D., Hanotte, O., Mwacharo, J. M., and Clark, E. L. (2021). Whole-Genome Sequence Data Suggest Environmental Adaptation of Ethiopian Sheep Populations. *Genome Biology and Evolution*, 13(3).

[Yuan et al., 2016] Yuan, S., Yu, Z., Liu, Q., Zhang, M., Xiang, Y., Wu, N., Wu, L., Hu, Z., Xu, B., Cai, T., Ma, X., Zhang, Y., Liao, C., Wang, L., Yang, P., Bai, L., and Li, Y. (2016). GPC5, a novel epigenetically silenced tumor suppressor, inhibits tumor growth by suppressing Wnt/beta-catenin signaling in lung adenocarcinoma. *Oncogene*, 35(47) :6120–6131.

Quatrième partie

Discussion et perspectives

Chapitre 8

Discussion

D.1 Résumé des travaux

L'objectif global de ce travail de thèse était d'étudier les histoires démographique et adaptative des bovins et des ovins situés sur le pourtour de l'Océan Indien. La stratégie d'analyse a consisté à caractériser phénotypiquement et génétiquement deux nouvelles populations de bovins et ovins locaux présents sur l'île de Mayotte, permettant ainsi de combler les vides laissés par les données historiques et archéologiques dans les histoires démographiques de ces populations et de compléter la vision des histoires démographiques des populations ovines et bovines Africaines.

La stratégie d'analyse de jeux de données de génotypages à moyenne et haute densité a permis de répondre aux question présentées dans la partie "Axe de recherche" :

Les caractéristiques phénotypiques des ovins et bovins locaux présents sur l'île de Mayotte sont hétérogènes au sein des populations : plusieurs motifs de robe ainsi que les différents types de cornes co-existent. Cette caractérisation a permis de mettre en évidence des caractéristiques phénotypiques liées à l'adaptation des climats tropicaux comme la bosse chez le zébu et la queue grasse chez le mouton présentant un rôle comme organe de réserves corporelles mobilisable en cas de défaut de l'apport nutritif et hydrique. Plus spécifiquement, la caractérisation phénotypique de la population de moutons mahorais a permis de mettre en évidence une mutation phénotypique auriculaire, dont la prévalence est de 20 % dans l'échantillon étudié.

La structure génétique étudiée par l'analyse multidimensionnelle des données de génotypage a permis de déterminer que la population de zébu de Mayotte était issue d'un métissage *Zébu X Taurin Africain* et que les ovins de Mayotte étaient proche génétiquement des moutons africains à queue grasse. Malgré un niveau de consanguinité globale estimé équivalent à celui de la population ovine de Mérinos de Rambouillet (30%) et plus faible que la population locale ovine de Ouessan par le calcul des ROH, la situation n'est pas aussi préoccupante si on prend en compte la variation individuelle de cette estimation. Toutefois il nécessaire de rester vigilant au regard du contexte in-

sulaire, de la faible structuration du processus d'identification des petits ruminants, de l'effectif faible estimé des moutons (environ 1'000 individus) et du risque d'absorption génétique par l'importation en vif et l'insémination articielle de ressources génétiques exotiques européennes. Afin de localiser la mutation déterminant le phénotype auriculaire "oreille courtes/microtie" dans la population de mouton mahorais, plusieurs tests d'association (Fisher, χ^2, Armitage), conduits sur 48 individus génotypés (Ratio cas/contrôles = 35/13) ont détecté un SNP associé avec le phénotype oreille réduite ou atrophiée (SNP OAR-115757345.1) à la position 114 296 273 sur le chromosome 6, à proximité du gène HMX1 qui est associé à la microtie chez les moutons chinois de race Altay [He et al., 2020].

L'origine des zébus locaux de Mayotte et de Madagascar a été déterminée aux moyens de différentes méthodes d'inférence démographique basées sur l'estimation du déséquilibre de liaison, des ROH, des fréquences alléliques. Ces deux populations auraient une origine commune issue d'un métissage ancien entre des populations sources proche des zébus indiens actuels et des populations sources proches des taurins africains actuels datée du 10e siècle, suivi d'une introgression indicine datée du 12e siècle, le lieu de ces métissages restant à déterminer. Après une période d'homogénéisation et à partir du 16e siècle, les trajectoires d'évolution des tailles efficaces des deux populations ont divergé suggérant un isolement des deux populations bovines, concomitant à l'influence croissante des colons Européens dans la zone. Les analyses n'ont pas permis d'écarter ou de confirmer une éventuelle origine de la région indonésienne.

L'identification des signatures de sélection chez les bovins de Mayotte et Madagascar ainsi que chez les ovins de Mayotte a permis de mieux comprendre l'histoire adaptative locale de ces populations dans leur environnement propre. Ainsi l'estimation de l'étendue du déséquilibre de liaison (DL) au sein de nos populations étudiées et en les comparant avec l'étendue du DL chez populations représentatives des ascendances, a permis de déterminer des listes de gènes candidats associés aux signatures de sélection propres à chaque population, avec toutefois des gènes en communs (chez les bovins 5 pour Mayotte vs Madagascar ; 1 gène candidat entre le mouton mahorais et le zébu de Madasgacar). Les fonctions biologiques identifiées concernent la respiration, les propriétés de la peau, le système nerveux, le métabolisme, le structures physiologiques, la respiration, les systèmes hormonal et reproductif, la réponse immunitaire et inflammatoire.

La variabilité génétique associée à l'adaptation a été auparavant étudiée par des analyses d'association avec les variables climatiques populationnelles ont déjà été conduites sur plusieurs zones géographiques comme la Chine [Gao et al., 2017] et le bassin méditerranéen [Flori et al., 2019] chez les bovins, à l'échelle mondiale [Lv et al., 2014, Cao et al., 2021] et plus localement en Ethiopie [Wiener et al., 2021] chez les ovins. Pour la première fois dans la zone de l'Océan Indien, la variabilité génétique associée à l'adaptation a été estimée en déterminant les régions génomiques

associées à des variables climatiques pour des populations bovines et ovines. L'annotation fonctionnelle des gènes candidats chez les deux espèces associés à des covariables climatiques par le logiciel IPA a permis de déterminer les fonctions biologiques sous pression environnementale telles que les propriétés de la peau, le système nerveux, la structure musculo-squelettique, l'embryogénèse et le cancer, notamment à travers la voie de signalisation des protéines Wn.

D.2 Perspectives

D.2.1 Intégration de nouvelles données et recherches de nouveaux gènes candidats sous pression de sélection

Disponibilité des données génétiques Notre étude s'inscrit dans un cadre de croissance du nombre de populations locales bovines et ovines génotypées, s'ajoutant ainsi aux jeux de données de références représentant la diversité génétique mondiale bovine [Decker et al., 2014] et ovine [Kijas et al., 2012]. Néanmoins lors de la récupération de jeux de données SNPs, une attention doit être porté à la qualité et le format de disponibilité des jeux de données qui peuvent être variables d'une publication à l'autre. Il n'est pas rare de se retrouver en présence de données de génotypage déjà filtrées et dont plus du tiers du nombre marqueurs est manquant. Cela implique en fusionnant avec d'autres jeux de données de vérifier les seuils de filtrage de données manquantes.

L'utilisation des données génomiques relève d'un compromis en terme de disponibilité des jeux de données en fonction de la densité des marqueurs et du nombre d'individus par population. Ainsi, le prix de l'analyse (génotypage, séquençage) augmente avec le nombre de variants à typer. De ce fait, à ressources financières égales, plus le nombre de variants augmente, moins il y a d'individus, plus faibles sont les possibilités de déterminer la source de la variabilité génétique observée. Les techniques d'imputation pourraient permettre d'augmenter le nombre d'individus disponibles par population à partir des données de génotypage moyenne densité pour ensuite obtenir des données de génotypage en haute densité.

Affiner les histoires démographiques L'échantillonnage des populations locales bovines et ovines de Mayotte a permis de combler les vides laissés par les données archéologiques, historiques sur l'histoire des bovins et ovins de la zone géographique des îles de l'Ouest de l'Océan Indien. La caractérisation phénotypique et génétique de nouvelles populations locales présentes notamment aux Comores, en Tanzanie Mozambique permettrait de constituer un continnum géographique de données génétiques dans le corridor Swahili et potentiellement répondre aux questions de la localisation des évènements de métissage chez les zébus de Mayotte et Madagascar détectés dans le Chapitre 5. Concernant les ovins, la population de moutons mahorais étant la seule échantillonnée dans la zone, la récolte de données génétiques ovines à Madagascar permettrait d'inférer l'histoire démographique de cette population.

Une autre piste à creuser serait le séquençage d'ADN fossile sur les ossements trouvés dans les différents sites archéologiques de la zone [Wright et al., 1984, Allibert et al., 1989, Pauly, 2013]. Certains défis techniques devront pourtant être relevé du fait de la forte dégradation que présente l'ADN fossile et dont les méthodes actuelles d'extraction ne permettent pas d'obtenir des séquences génétiques complètes. Cette solution pourrait permettre d'avoir une photographie, certes partielle, de la diversité génétique à la période contemporaine des échantillons et ainsi de remonter plus loin dans l'histoire des populations que nous le permet les échantillons modernes [Pereira et al., 2006, Cai et al., 2014].

Association avec d'autres variables La recherche des signatures de sélection candidates déterminées par l'estimation du déséquilibre de liaison chez les bovins (Chapitre 5) et les ovins (Chapitre 6) ne permettent pas de distinguer si la pression de sélection est d'origine humaine ou environnementale. Ainsi, dans le Chapitre 7, une étude d'association Génome-environnement a été conduite sur deux jeux de données regroupant des populations bovines et ovines localisées sur le pourtour du littoral de l'Océan Indien, mais peu de gènes associés aux covariables climatiques ont été retrouvés dans les régions génomiques sous sélection. L'utilisation d'autres types de covariables pourrait affiner les résultats. Cependant la recherche de données populationnelles de production relèvent bien souvent du défi à relever du fait du manque d'informations disponibles. La démocratisation des outils génomiques a entraîné une accélération du génotypage des populations bovines et ovines locales sans une caractérisation phénotypique préalable. La base de données de la FAO (`http://www.fao.org/dad-is/`), seule source d'informations globales sur les races, permet d'obtenir des données déclarées sur la base du volontariat. Ainsi, pour certaines populations, malgré une recherche assidue des bases de données de la littérature, aucune donnée, ne serait-ce que pour la hauteur au garrot et le poids moyen, n'est disponible. Compte-tenu de l'insuffisance des infrastructures de conseil aux éleveurs et de recueil de données de production en routine caractéristiques des pays en développement, les données de productions populationnelles, si elles existent, sont obtenues à partir de données d'expérimentation ou de référence sur de courtes périodes (comme c'est le cas à Mayotte) et ne sont pas forcément représentatives des niveaux de production obtenus dans les systèmes d'élevage particuliers. Ces données sont aussi bien souvent déconnectées des individus génotypés.

Un autre type de données populationnelles peuvent être obtenues à partir de modèles calculant des indicateurs de végétation. Les indicateurs de végétation peuvent être obtenus à partir d'image satellite ou de données Infra-Rouge et pourrait rendre compte de la disponibilité des ressources fourragères. Les modèles écologiques de dispersion spatiale des pathogènes (tiques, parasites) peuvent être considérés comme des covariables phénotypiques dans les Etudes d'Association Genome-Environnement [Vajana et al., 2018] et permettre de détecter les variants associés à la réponse immunitaire.

Méta-analyse des signatures de sélection des bovins et ovins tropicaux La dernière méta-analyse étudiants les signatures de sélection pour différents types de bovins (taurins africains, taurins européens, zébus et hybrides) a été publiée en 2016 [Randhawa et al., 2016]. Mais la majorité des études concernait les races taurines européennes améliorées telle que la Holstein (étudiée dans 33 études sur 56). En outre la plupart des races testées pour la sélection chez les bovins non européens, par exemple le zébu, n'étaient pas représentatives des races locales. En effet, les races étudiées (Brahman, Nellore, Gir, etc.) pourraient avoir connu des goulots d'étranglement parce que les échantillons utilisés dans les analyses de signature de sélection proviennent de troupeaux importés dans des pays (Brésil par exemple) qui ont démarré à partir de quelques centaines d'animaux fondateurs, plutôt que des représentants des populations locales. Ainsi avec le nombre croissant d'études sur les races locales tropicales, ces données pourraient par la suite être intégrées dans une méta-analyse afin de contrebalancer dans une certaine mesure les effets du nombre conséquent d'études menées sur des races améliorées européennes.

Comparer des signatures de sélection entre espèces La comparaison des signatures de sélections entre espèces permet de révéler les régions candidates soumises à une sélection dans un environnement commun aux deux espèces.

Le processus consiste à mapper les régions sous sélection d'un espèce A sur l'assemblage de l'espèce B (et réciproquement) comme suggéré dans [Kim et al., 2016]. D'autres pipelines ont été proposés afin de minimiser la détection de faux positifs : Le test des signatures de sélection commune entre les chèvres et les moutons meditérannéen [Alberto et al., 2018] (20 régions), est constitué de la transformation des p-value des scores hapflk calculés et mappés sur l'autre espèce en q-value en utilisant le processus FDR proposé par [Storey et al., 2004].

Pour vérifier si les régions sous sélection communes entre les bovins et les buffles [Dutta et al., 2020], le processus de permutation a été utilisé afin de déterminer si les régions génomiques d'une espèce chevauchent les pics de score statistique de l'autre espèce plus souvent que prévu par le hasard. Les pics de score statistique du buffle d'eau ont été mappés sur le génome de la vache. Après avoir déterminé le nombre réel de pics se chevauchant, des régions de même nombre et de même taille ont été échantillonnées 100 fois dans le génome de la vache et le nombre de pics se chevauchant a été recalculé. En utilisant ces résultats de permutation, un score Z a été calculé pour chaque comparaison, représentant le nombre d'écarts types du nombre réel de chevauchements au-dessus de la moyenne des résultats de permutation. Ces scores Z ont ensuite été convertis en valeurs P et de nouveaux seuils de significativité statistique ont été déterminés compensant ainsi les taux de faux positifs estimés dans les calculs de scores statistiques.

D.2.2 Caractérisation des mutations candidates impliquées dans l'adaptation

L'étude des signatures de sélection a permis de déterminer des polymorphismes nucléotidiques associées des mutations candidates. Il s'agira par la suite de confirmer leur rôle dans l'adaptation aux conditions d'élevage non-optimal ainsi qu'à l'environnnement tropical caractérisé par l'alternance de périodes de faibles précipitations à l'origine d'un faible apport nutritif et hydrique et de périodes de fortes précipitations permettant la disponibilité en ressources nutritives et hydriques mais favorisant le développement de pathogènes. La caractérisation des mutations candidates se fera à partir des régions candidates obtenues après le calcul des scores EHH et XtX obtenus dans les différentes sections traitées précédemment (Chapitre 5 pour la caractérisation génétique des zébus de Mayotte et Madagascar, chapitre 6 pour les ovins mahorais, chapitre 7 pour les résultats XtX bovins et ovins). L'outil Variant Effect Predictor [McLaren et al., 2016] proposé sur le site d'Ensembl permet une annotation fine des régions codantes. L'effet des changements d'acides aminés engendrés par la mutation candidate est prédit grâce à l'outil SIFT [Kumar et al., 2009], proposant ainsi une annotation délètere ("deleterious") ou tolérée ("tolerated"). D'autres outils sont également disponibles pour annoter les régions régulatrices.

La validation fonctionnelle des mutations candidates pourra à terme être conduite sur des animaux modèles telle la souris afin de relier la biologie adaptative présumée aux mécanismes moléculaires sous-jacents.

D.2.3 Prédiction génomique de l'adaptation

L'effet du changement climatique exerce de fortes pressions de sélection sur le génome des espèces et des populations. Dans le but d'assurer la préservation de certaines races, il est nécessaire de prédire comment les populations vont répondre au changement climatique global [Rellstab et al., 2021] afin d'estimer la capacité d'une population à survivre en cas de changements de l'habitat, ou si la migration ou l'intervention humaine est requise. Les approches d'évaluation de persistance des espèces dans un futur habitat étaient basées auparavant sur l'estimation de la "fitness" dans des dispositifs expérimentaux et des modélisations de niche écologique. Compte tenu des limites de ces deux précédentes méthodes (coût des dispositifs expérimentaux, non prise en compte des caractéristiques démographiques et évolutives des espèces), les résultats des analyses génomiques (structure des populations, diversité génétique neutre, signatures de sélection) donnent des informations sur l'adaptation locale des populations qui peuvent être incorporées dans les modèles de prédiction [Razgour et al., 2019]. Plusieurs approches de mesure du "genomic offset", basées sur la distance génétique entre la composition génomique actuelle et la composition génomique théoriquement requise [1],

1. Si l'écart est petit, les populations ont potentiellement un faible risque d'être maladaptée à leur futur environnement. Si l'écart est grand, les populations sont très éloignées de leur composition génomique théorique requise et peuvent être en situation de haut risque de maladaptation.

sont disponibles et ont été développées [Rellstab et al., 2021] (Figure D.1). Néanmoins ces approches sont assez nouvelles dans leur développement et leur application [Capblancq et al., 2020], surtout pour les espèces animales. Ainsi, les enjeux de ces études seront de tester et valider ces prédictions [Capblancq et al., 2020, Rellstab et al., 2021]. Ces approches permettraient d'identifier les races les plus adaptées dans le contexte du changement climatique.

D.3 Application dans les pays en développement

D.3.1 Utilisation des outils génomiques dans les pays en développement

Une diversité génétique suffisante dans les populations d'animaux d'élevage est nécessaire pour permettre l'adaptation au changement climatique ainsi que pour l'amélioration génétique continue des caractéristiques à visée économique. Contrairement aux pays développés, les programmes d'amélioration des races locales dans les pays en développement n'ont qu'un succès limité : l'utilisation des technologies de reproduction est limitée à des secteurs d'élevage spécifiques dans certains pays, et l'utilisation généralisée des informations génomiques pour améliorer la productivité du bétail est balbutiante [Marshall, 2014, Felius et al., 2015, Marshall et al., 2019].

L'absence de sélection organisée sur les critères de production comme la croissance et les qualités de carcasse (rendement, conformation) explique un désintérêt croissant pour les races locales en même temps que la diffusion des ressources génétiques à haute productivité s'imposait au cours du 20e siècle [Hanotte et al., 2010, Biscarini et al., 2015]. L'effet antinomique de ce manque d'organisation est la disponibilité exploitable de la variabilité génétique des races locales dans les programmes de sélection et de conservation futurs. De plus, la variabilité génétique offre une capacité de résilience des races locales au changement climatique [Hoffmann, 2010, Hoffmann, 2013].

Ainsi, les races locales présentent un intérêt dans les programmes de croisement avec des races exotiques fortement productives afin de répondre à court terme à une demande en produits d'origine animale. Néanmoins, une adaptation au contexte local des méthodes de travail doit être réalisée pour assurer la durabilité des programmes d'amélioration génétique des races locales et de croisement, car les modèles mis en œuvre dans les pays développés ne peuvent pas être appliqués directement [Ducrocq et al., 2018, Marshall et al., 2019]. L'application de la génomique commence tout juste à émerger, et peut servir à :

- identifier le type de race ou de croisement le plus approprié pour différents systèmes d'élevage notamment quand le phénotype ne permet d'assigner un individu à un type racial [Marshall et al., 2019],
- permettre ou améliorer les programmes de sélection en implémentant l'information génétique sans avoir besoin du pedigree, et en augmentant la précision des évaluations génétiques par la connaissance des ascendances dans le cas du croisement,

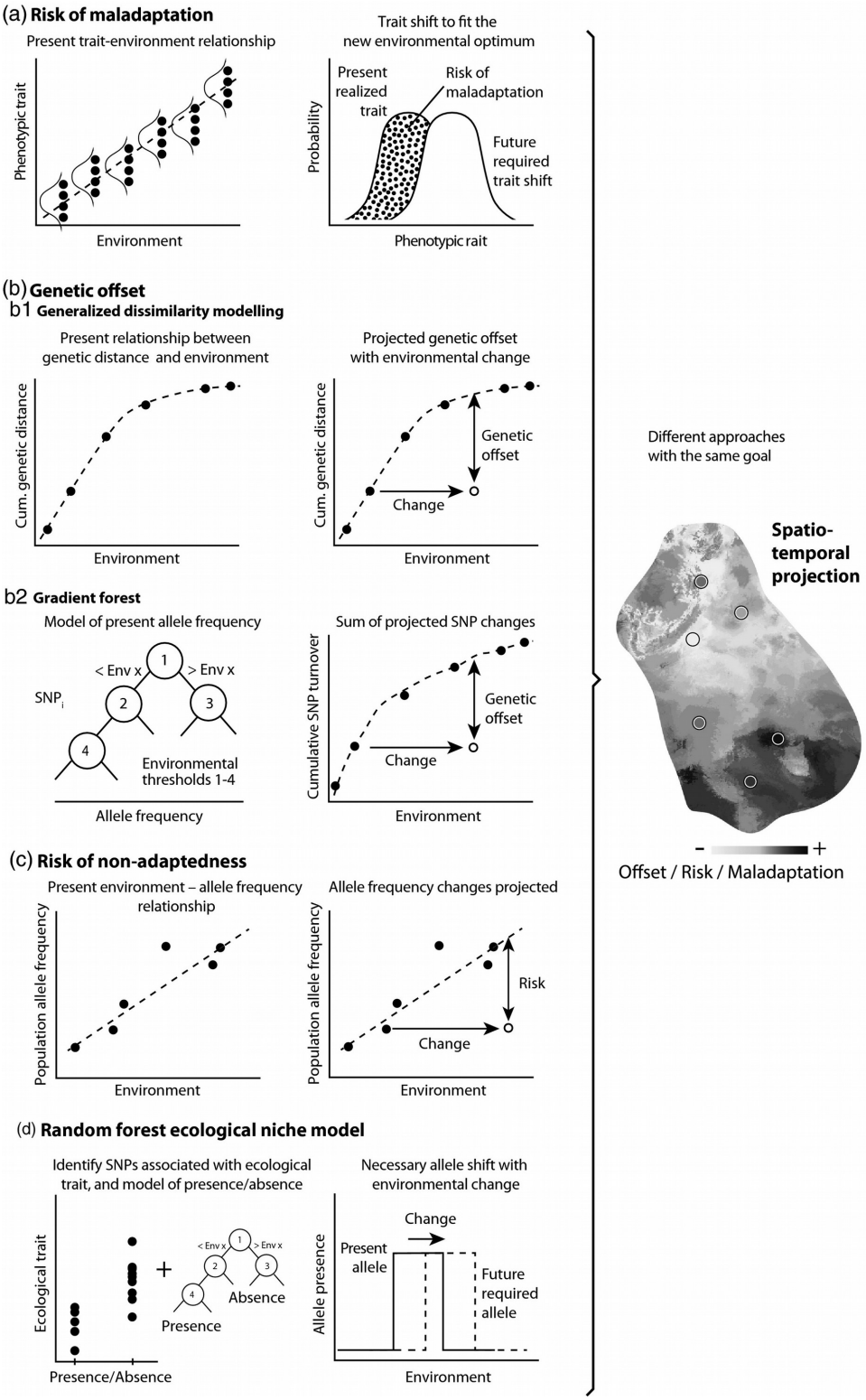


FIGURE D.1 – Principales approches de prédiction génomique basée sur le décalage génomique. **(a)** Risque de maladaptation : approche phénotypes-environnement. **(b)** Approches par gradient : Basé sur des arbres de décisions. **(c)** RONA (Risque de non-adaptation) : différence moyenne entre fréquences alléliques prédites et observées. **(d)** Random forest ecological niche model : modélisation de l'absence/présence des allèles en utilisant des modèles de niche environnementale.

— créer des nouveaux types raciaux présentant une résistance/tolérance aux contraintes environnementales locales,
— découvrir des variants génétiques d'importance économique et environnemental par l'étude des signatures de sélection.

Pour ce dernier enjeu, il demeure nécessaire d'envisager des programmes de gestion des races locales afin de conserver les ressources génétiques. Ici aussi, les outils génomiques ont une utilité : après avoir caractériser génétiquement les populations d'élevage afin de déterminer leur composition (zebus, zébu africains, taurins) [Marshall et al., 2019] et éventuellement d'estimer l'introgression taurine européenne, il est possible d'estimer le niveau de consanguinité, la filiation [Sponenberg et al., 2019, Eusebi et al., 2020] qui sont des indicateurs nécessaires permettant ainsi d'établir une stratégie de conservation des populations [Eusebi et al., 2020] afin de maintenir la variabilité génétique dans les races locales.

Les informations issues des travaux d'identification des signatures de sélection et des mutations causales associées pourront ensuite être intégrées dans des stratégies d'amélioration génétique, soit par le biais de programmes de sélection intégrant l'utilisation de données génomiques/génétiques, soit par la création de nouvelles races par des approches d'introgression ou de modification du génome [Mrode et al., 2018, Marshall et al., 2019].

Cependant, les facteurs limitants pour l'utilisation des outils génomiques de conservation et de sélection dans les pays en développement sont l'insuffisance (voire l'absence) (a) d'infrastructures de services aux éleveurs amenant au manque d'enregistrement de routine de phénotypes fiables empêchant la constitution de la population de référence [Goddard and Hayes, 2009, Biscarini et al., 2015, Ducrocq et al., 2018], (b) d'outils de synthèse des données fournissant un retour d'information aux éleveurs pouvant les aider à améliorer les techniques de gestion d'élevage. De plus, la progression du gain génétique pour les niveaux de production par la sélection génomique s'accompagne nécessairement pour une plus grande efficacité des technologies de reproduction (IA, production d'embryons par FIV, semence sexée), qui sont pour la plupart des pays en développement indisponibles.

D.3.2 Utilisation des résultats pour les populations de Mayotte

Dans le contexte tropical des pays en développement, l'utilisation des races locales de ruminants est orientée principalement vers la production de viande avec toutefois un prélèvement de la production laitière pour l'autoconsommation, mais aussi la traction animale, la valorisation de sous-produits de culture ou de zones difficiles, la fourniture de fumier, l'épargne et en tant qu'emblême patrimonial [Navès et al., 2011]. Les histoires démographiques et adaptatives ont façonné les principales aptitudes reconnues d'adaptation au milieu tropical en général [Hanotte et al., 2010, Hoffmann, 2010, Hoffmann, 2013] : tolérance aux effets directs du climat, valorisation de fourrages tropicaux, capacité de mobilisation des réserves corporelles, résistance aux tiques et aux maladies associées, résistance au parasitisme interne. Leur faible niveau de production comparée aux races améliorées est contrebalancé par les performances de repro-

duction et qualités maternelles : fertilité élevée, longévité, désaisonnement et taille de portée (pour les petits ruminants), croissance sous la mère régulière, faible mortalité [Navès et al., 2011].

Si ces caractéristiques ont été particulièrement bien étudiées chez la race bovine et les petits ruminants Créoles de Guadeloupe par exemple, à Mayotte c'est moins le cas, en raison notamment de l'environnement plus complexe avec une superficie du territoire plus faible (375 km^2), un foncier peu disponible [Barthes, 2003], un contexte géopolitique difficile dû aux relations avec les Comores voisines, une intégration récente du territoire en tant que département qui a freiné les investissements dans l'activité agricole. Le développement des fonds européens FEADER a permis de mettre en place des actions visant à mieux connaître l'organisation et les performances du secteur de l'élevage via le programme PAZEM [Tillard et al., 2013]. Néanmoins bien que ces connaissances acquises constituent des références précieuses et des pistes d'amélioration, les filières ovines et bovines n'ont toujours pas été structurées. L'identification animale n'est effective que pour les bovins et partiellement, la Base de Données Nationale d'Identification (BDNI) n'étant rattachée au système d'information national que depuis 2003. Pour les ovins, cela n'est pas le cas même si des efforts dans la pose des boucles sont faits.

Ainsi, la caractérisation génétique des populations de zébus et moutons locaux a permis de pallier l'absence de données généalogiques et d'établir un état des lieux de la diversité génétique de ces populations, nécessaire pour établir leur risque d'extinction (Figure D.2). De plus, la détermination des histoires démographiques contribuera à renforcer le rôle patrimonial (Figure D.2) du zébu mahorais notamment qui est un levier d'action à développer afin de perenniser la race sur le territoire. En effet, l'arrivée des possibilités de financements par les fonds européens de projets d'amélioration foncière (ouvertures de chemins carrossables, accès à l'eau) pourrait inciter des éleveurs de zébus locaux à se tourner vers des races plus productives du fait des facilités d'élevage nouvelles. Les résultats issus de l'analyse des signatures de sélection appuieront les arguments de l'adaptation à l'environnement spécifique de Mayotte et à un système d'élèvage économe en intrants (Figure D.2). Il est cependant nécessaire de proposer des mesures d'incitation financière à la conservation et à la reproduction d'animaux de races locales comme la Mesure Agro-Environnementale (MAE) "Protection des Races Menacées" (PRM) financée par l'Europe mais qui n'est débloquée que si la race reconnue est inscrite par la Commission Thématique Interfilières "Ressources ZooGénétiques" (CTI - RZG, commission succédant ainsi à la CNAG depuis le 12 août 2020) sur la liste des races animales menacées d'abandon pour l'agriculture, c'est-à-dire que le nombre de femelles reproductrices est en-dessous de 7500 individus (ce chiffre peut être modulé en fonction du contexte de gestion de la race). Un inventaire des zébus mahorais, reconnus officiellement depuis septembre 2018, est en cours afin de fournir des données officielles des effectifs actuels en vue de réviser le statut de la race et d'éventuellement de débloquer les fonds européens. Dans une vision à long terme de l'intégration de la race dans la filière bovine mahoraise future, il s'avère néanmoins nécessaire de mettre en place un service de contrôle de performances, ne serait-ce que pour permettre un suivi des élevages. Concernant les ovins mahorais, la reconnaissance de la race n'est toujours

pas actée du fait de l'absence d'une structure locale porteuse impliquant des éleveurs bien que la situation soit beaucoup plus critique, compte tenu du faible effectif recensé.

A partir des informations issues du processus de caractérisation des races et leur reconnaissance officielle, il pourrait ensuite être possible de proposer des actions de conservation des populations de ruminants de Mayotte, poursuivre ensuite un processus d'amélioration génétique une fois la situation stabilisée (Figure D.2).

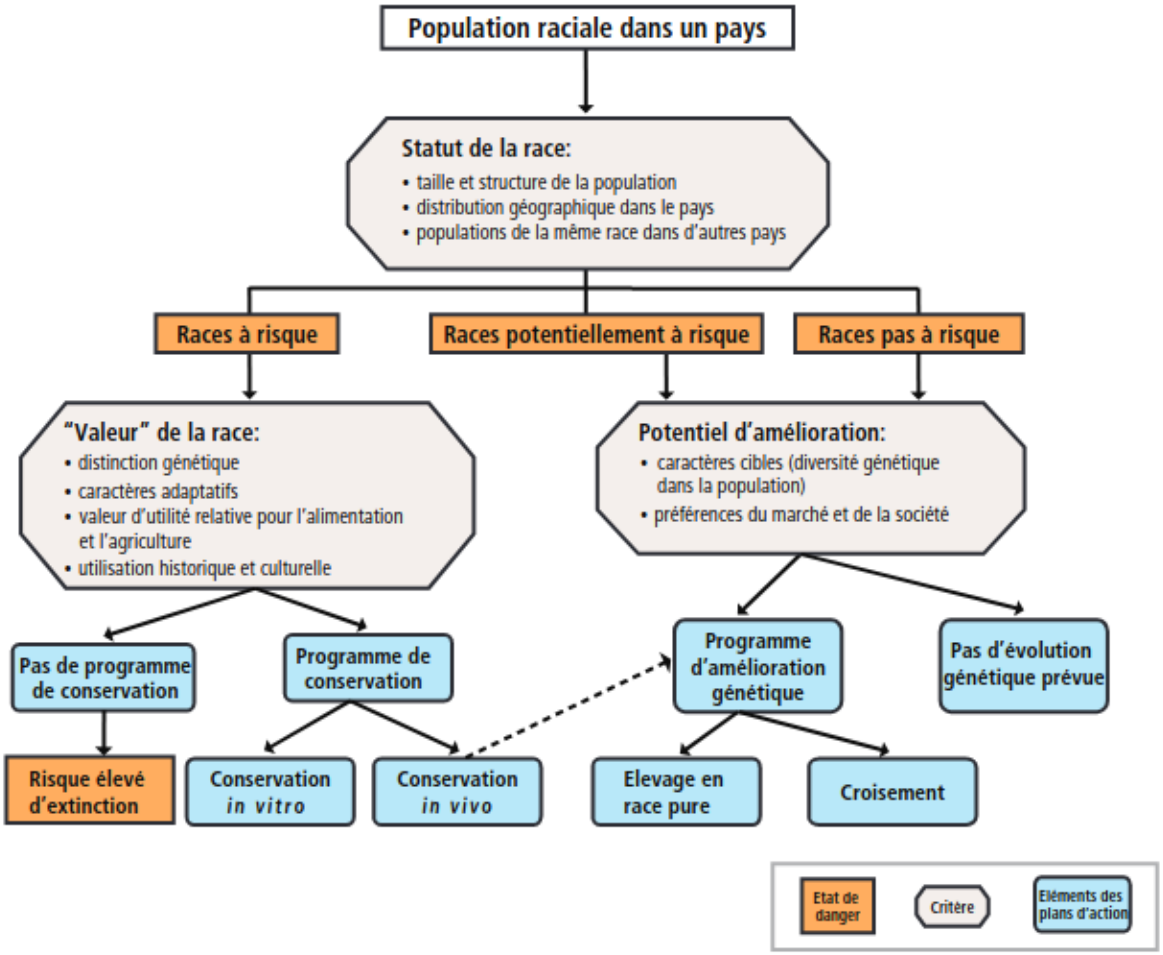


FIGURE D.2 – Schéma des processus de décision pour la gestion des ressources zoogénétiques d'élevage. (Source [Scherf et al., 2015])

D.4 Conclusion

Ainsi, dans le cadre de la recherche fondamentale, les races locales sont des modèles intéressants puisqu'elles sont la résultante de la conjonction entre les quatres forces évolutives [Cavalli-Sforza, 2011] agissant sur le génome des individus, soit de manière globale avec le mécanisme de la dérive génétique et des flux géniques, soit de manière locale avec l'apparition ponctuelle d'une mutation, ou la sélection naturelle ou artificielle par l'homme sur un caractère d'intérêt [Marshall, 2014]. D'un point de vue plus appliquée, bien que leurs niveaux

de performances n'atteignent pas ceux des races améliorées, elles représentent cependant une ressource génétique diverse intéressante pouvant s'adapter aux variations environnementales. Ces races nécessitent donc une attention particulière dans leur gestion, dont les résultats ne seront visibles qu'à long terme, qui vient en contradiction avec l'intérêt économique de court terme.

Bibliographie

[Alberto et al., 2018] Alberto, F. J., Boyer, F., Orozco-terWengel, P., Streeter, I., Servin, B., de Villemereuil, P., Benjelloun, B., Librado, P., Biscarini, F., Colli, L., Barbato, M., Zamani, W., Alberti, A., Engelen, S., Stella, A., Joost, S., Ajmone-Marsan, P., Negrini, R., Orlando, L., Rezaei, H. R., Naderi, S., Clarke, L., Flicek, P., Wincker, P., Coissac, E., Kijas, J., Tosser-Klopp, G., Chikhi, A., Bruford, M. W., Taberlet, P., and Pompanon, F. (2018). Convergent genomic signatures of domestication in sheep and goats. *Nature Communications*, 9(1).

[Allibert et al., 1989] Allibert, C., Argant, A., and Argant, J. (1989). Le site de Dembeni (Mayotte, Archipel des Comores). Mission 1984. *Etudes Océan Indien*, 11 :63–172.

[Barthes, 2003] Barthes, C. (2003). *L'état et le monde rural à Mayotte*. KARTHALA Editions.

[Biscarini et al., 2015] Biscarini, F., Nicolazzi, E. L., Stella, A., Boettcher, P. J., and Gandini, G. (2015). Challenges and opportunities in genetic improvement of local livestock breeds. *Frontiers in Genetics*, 6 :33.

[Cai et al., 2014] Cai, D., Sun, Y., Tang, Z., Hu, S., Li, W., Zhao, X., Xiang, H., and Zhou, H. (2014). The origins of Chinese domestic cattle as revealed by ancient DNA analysis. *Journal of Archaeological Science*, 41 :423–434.

[Cao et al., 2021] Cao, Y.-H., Xu, S.-S., Shen, M., Chen, Z.-H., Gao, L., Lv, F.-H., Xie, X.-L., Wang, X.-H., Yang, H., Liu, C.-B., Zhou, P., Wan, P.-C., Zhang, Y.-S., Yang, J.-Q., Pi, W.-H., Hehua, E., Berry, D. P., Barbato, M., Esmailizadeh, A., Nosrati, M., Salehian-Dehkordi, H., Dehghani-Qanatqestani, M., Dotsev, A. V., Deniskova, T. E., Zinovieva, N. A., Brem, G., Štěpánek, O., Ciani, E., Weimann, C., Erhardt, G., Mwacharo, J. M., Ahbara, A., Han, J.-L., Hanotte, O., Miller, J. M., Sim, Z., Coltman, D., Kantanen, J., Bruford, M. W., Lenstra, J. A., Kijas, J., and Li, M.-H. (2021). Historical Introgression from Wild Relatives Enhanced Climatic Adaptation and Resistance to Pneumonia in Sheep. *Molecular Biology and Evolution*, 38(3) :838–855.

[Capblancq et al., 2020] Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M., and Keller, S. R. (2020). Genomic Prediction of (Mal)Adaptation Across Current and Future Climatic Landscapes. *Annual Review of Ecology, Evolution, and Systematics*, 51(1) :245–269.

[Cavalli-Sforza, 2011] Cavalli-Sforza, L. (2011). *L' Aventure de l'espèce humaine : De la génétique des populations à l'évolution culturelle*. Odile Jacob. Google-Books-ID : YY6dAAAAQBAJ.

[Decker et al., 2014] Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Molina Alcalá, A., Sonstegard, T. S., Hanotte, O., Götherström, A., Seabury, C. M., Praharani, L., Babar, M. E., Correia de Almeida Regitano, L., Yildiz, M. A., Heaton, M. P., Liu, W.-S., Lei, C.-Z., Reecy, J. M., Saif-Ur-Rehman, M., Schnabel, R. D., and Taylor, J. F. (2014). Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *PLoS Genetics*, 10(3) :e1004254.

[Ducrocq et al., 2018] Ducrocq, V., Laloe, D., Swaminathan, M., Rognon, X., Tixier-Boichard, M., and Zerjal, T. (2018). Genomics for Ruminants in Developing Countries : From Principles to Practice. *Frontiers in Genetics*, 9 :251.

[Dutta et al., 2020] Dutta, P., Talenti, A., Young, R., Jayaraman, S., Callaby, R., Jadhav, S. K., Dhanikachalam, V., Manikandan, M., Biswa, B. B., Low, W. Y., Williams, J. L., Cook, E., Toye, P., Wall, E.,

Djikeng, A., Marshall, K., Archibald, A. L., Gokhale, S., Kumar, S., Hume, D. A., and Prendergast, J. G. D. (2020). Whole genome analysis of water buffalo and global cattle breeds highlights convergent signatures of domestication. *Nature Communications*, 11(1) :4739.

[Eusebi et al., 2020] Eusebi, P. G., Martinez, A., and Cortes, O. (2020). Genomic Tools for Effective Conservation of Livestock Breed Diversity. *Diversity*, 12(1) :8.

[Felius et al., 2015] Felius, M., Theunissen, B., and Lenstra, J. A. (2015). Conservation of cattle genetic resources : the role of breeds. *The Journal of Agricultural Science*, 153(1) :152–162.

[Flori et al., 2019] Flori, L., Moazami-Goudarzi, K., Alary, V., Araba, A., Boujenane, I., Boushaba, N., Casabianca, F., Casu, S., Ciampolini, R., D'Acier, A. C., Coquelle, C., Delgado, J.-V., El-Beltagi, A., Hadjipavlou, G., Jousselin, E., Landi, V., Lauvie, A., Lecomte, P., Ligda, C., Marinthe, C., Martinez, A., Mastrangelo, S., Menni, D., Moulin, C.-H., Osman, M.-A., Pineau, O., Portolano, B., Rodellar, C., Saïdi-Mehtar, N., Sechi, T., Sempéré, G., Thévenon, S., Tsiokos, D., Laloë, D., and Gautier, M. (2019). A genomic map of climate adaptation in Mediterranean cattle breeds. *Molecular Ecology*, 28(5) :1009–1029.

[Gao et al., 2017] Gao, Y., Gautier, M., Ding, X., Zhang, H., Wang, Y., Wang, X., Faruque, M. O., Li, J., Ye, S., Gou, X., Han, J., Lenstra, J. A., and Zhang, Y. (2017). Species composition and environmental adaptation of indigenous Chinese cattle. *Scientific Reports*, 7(1) :1–14.

[Goddard and Hayes, 2009] Goddard, M. E. and Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, 10(6) :381–391.

[Hanotte et al., 2010] Hanotte, O., Dessie, T., and Kemp, S. (2010). Time to Tap Africa's Livestock Genomes. *Science*.

[He et al., 2020] He, S., Zhang, Z., Sun, Y., Ren, T., Li, W., Zhou, X., Michal, J. J., Jiang, Z., and Liu, M. (2020). Genome-wide association study shows that microtia in Altay sheep is caused by a 76 bp duplication of *HMX1*. *Animal Genetics*, 51(1) :132–136.

[Hoffmann, 2010] Hoffmann, I. (2010). Climate change and the characterization, breeding and conservation of animal genetic resources. *Animal Genetics*, 41 :32–46.

[Hoffmann, 2013] Hoffmann, I. (2013). Adaptation to climate change – exploring the potential of locally adapted breeds. *animal*, 7(s2) :346–362.

[Kijas et al., 2012] Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., and Dalrymple, B. (2012). Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biology*, 10(2).

[Kim et al., 2016] Kim, E.-S., Elbeltagy, A. R., Aboul-Naga, A. M., Rischkowsky, B., Sayre, B., Mwacharo, J. M., and Rothschild, M. F. (2016). Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity*, 116(3) :255–264.

[Kumar et al., 2009] Kumar, P., Henikoff, S., and Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, 4(7) :1073–1081.

[Lv et al., 2014] Lv, F.-H., Agha, S., Kantanen, J., Colli, L., Stucki, S., Kijas, J. W., Joost, S., Li, M.-H., and Ajmone Marsan, P. (2014). Adaptations to Climate-Mediated Selective Pressures in Sheep. *Molecular Biology and Evolution*, 31(12) :3324–3343.

[Marshall, 2014] Marshall, K. (2014). Optimizing the use of breed types in developing country livestock production systems : a neglected research area. *Journal of Animal Breeding and Genetics*, 131(5) :329–340.

[Marshall et al., 2019] Marshall, K., Gibson, J. P., Mwai, O., Mwacharo, J. M., Haile, A., Getachew, T., Mrode, R., and Kemp, S. J. (2019). Livestock Genomics for Developing Countries – African Examples in Practice. *Frontiers in Genetics*, 10 :297.

[McLaren et al., 2016] McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, 17(1) :122.

[Mrode et al., 2018] Mrode, R., Tarekegn, G. M., Mwacharo, J. M., and Djikeng, A. (2018). Invited review : Genomic selection for small ruminants in developed countries : how applicable for the rest of the world ? *animal*, 12(7) :1333–1340.

[Navès et al., 2011] Navès, M., Alexandre, G., Mahieu, M., Gourdine, J.-L., and Mandonnet, N. (2011). Les races animales locales : bases du développement innovant et durable de l'élevage aux Antilles. In *CIAG 2011 Carrefours de l'innovation agronomique*, volume 16, pages 220–p. INRA.

[Pauly, 2013] Pauly, M. (2013). Acoua-Agnala M'kiri, Mayotte (976), archéologie d'une localité médiévale (XIe-XVe siècles), entre Afrique et Madagascar. *NYAME AKUMA*, (80) :73–90.

[Pereira et al., 2006] Pereira, F., Davis, S. J., Pereira, L., McEvoy, B., Bradley, D. G., and Amorim, A. (2006). Genetic signatures of a Mediterranean influence in Iberian Peninsula sheep husbandry. *Molecular Biology and Evolution*, 23(7) :1420–1426.

[Randhawa et al., 2016] Randhawa, I. A. S., Khatkar, M. S., Thomson, P. C., and Raadsma, H. W. (2016). A Meta-Assembly of Selection Signatures in Cattle. *PLOS ONE*, 11(4) :e0153013.

[Razgour et al., 2019] Razgour, O., Forester, B., Taggart, J. B., Bekaert, M., Juste, J., Ibáñez, C., Puechmaille, S. J., Novella-Fernandez, R., Alberdi, A., and Manel, S. (2019). Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *Proceedings of the National Academy of Sciences*, 116(21) :10418–10423.

[Rellstab et al., 2021] Rellstab, C., Dauphin, B., and Exposito-Alonso, M. (2021). Prospects and limitations of genomic offset in conservation management. *Evolutionary Applications*, 14(5) :1202–1212.

[Scherf et al., 2015] Scherf, B. D., Pilling, D., and others (2015). The second report on the state of the world's animal genetic resources for Food and Agriculture.

[Sponenberg et al., 2019] Sponenberg, D. P., Martin, A., Couch, C., and Beranger, J. (2019). Conservation Strategies for Local Breed Biodiversity. *Diversity*, 11(10) :177.

[Storey et al., 2004] Storey, J. D., Taylor, J. E., and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates : a unified approach. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 66(1) :187–205.

[Tillard et al., 2013] Tillard, E., Moussa, T., Balberini, L., Aubriot, D., and Berre, D. (2013). *Référentiel technico-économique des élevages de bovins à Mayotte*.

[Vajana et al., 2018] Vajana, E., Barbato, M., Colli, L., Milanesi, M., Rochat, E., Fabrizi, E., Mukasa, C., Del Corvo, M., Masembe, C., Muwanika, V. B., Kabi, F., Sonstegard, T. S., Huson, H. J., Negrini, R., , T. N. C., Joost, S., and Ajmone-Marsan, P. (2018). Combining Landscape Genomics and Ecological Modelling to Investigate Local Adaptation of Indigenous Ugandan Cattle to East Coast Fever. *Frontiers in Genetics*, 9 :385.

[Wiener et al., 2021] Wiener, P., Robert, C., Ahbara, A., Salavati, M., Abebe, A., Kebede, A., Wragg, D., Friedrich, J., Vasoya, D., Hume, D. A., Djikeng, A., Watson, M., Prendergast, J. G. D., Hanotte, O., Mwacharo, J. M., and Clark, E. L. (2021). Whole-Genome Sequence Data Suggest Environmental Adaptation of Ethiopian Sheep Populations. *Genome Biology and Evolution*, 13(3).

[Wright et al., 1984] Wright, H. T., Sinopoli, C., Wojnaroski, L., Hoffman, E. S., Scott, S. L., Redding, R. W., and Goodman, S. M. (1984). Early Seafarers of the Comoro Islands : the Dembeni Phase of the IXth-Xth Centuries AD. *Azania : Archaeological Research in Africa*, 19(1) :13–59.

Cinquième partie

Annexes

Annexe A

Monographies

A.1 Phylogénie

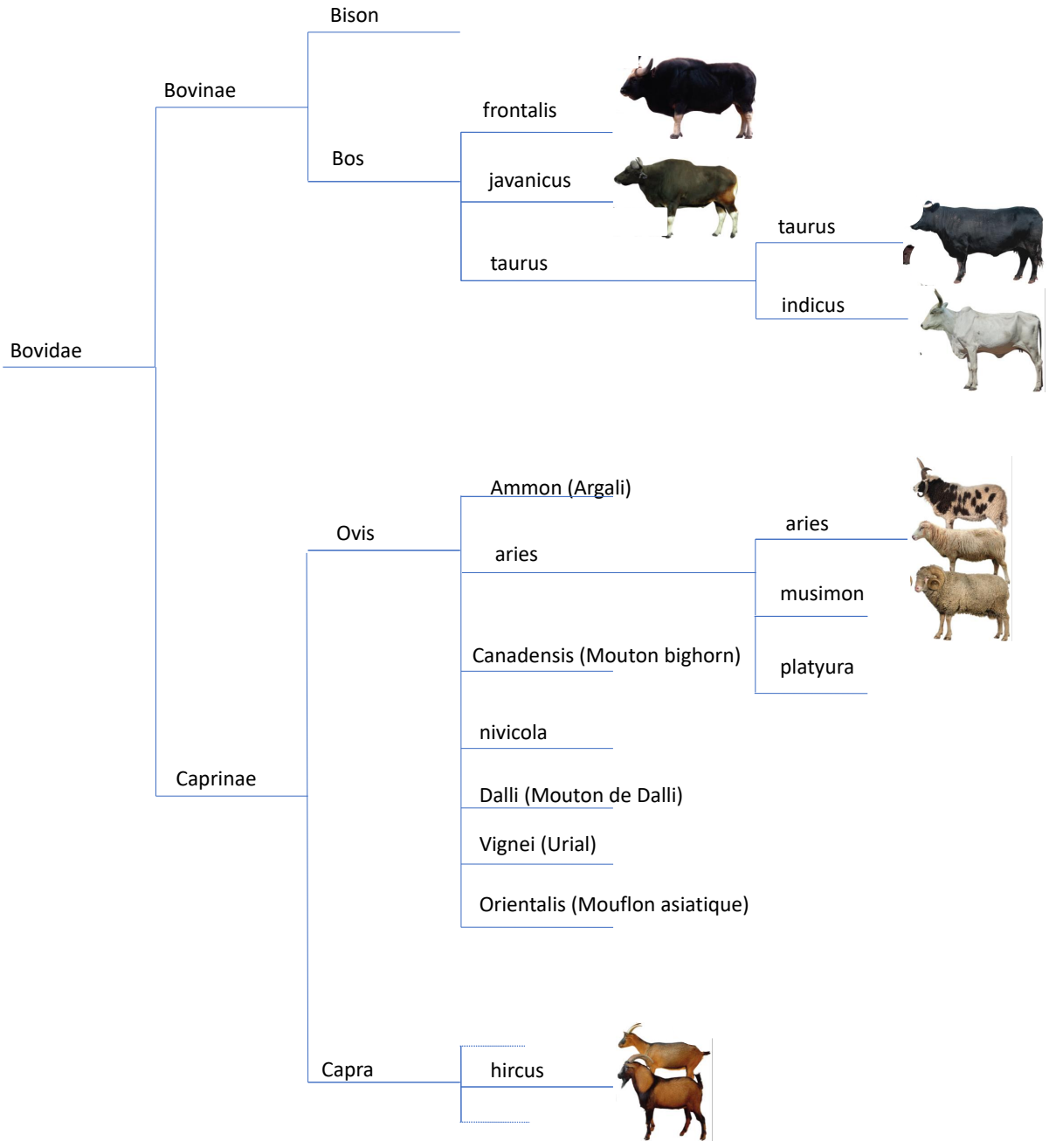
Les bovins et ovins appartiennent à la famille des bovidés (*Bovidae*). Les deux espèces divergent ensuite entre deux sous-familles les bovinés (*bovinae*) et les caprinés (*caprinae*).

Concernant les bovins, les deux rameaux relatifs à l'espèce *Bos taurus*, les plus largement élevées, avec autour de 0.75 milliard d'individus pour chaque effectif (`www.fao.org/faostat/en/`) [Utsunomiya et al., 2019], se rencontrent sur toute la surface du globe : Les bovins de type "taurins" *Bos taurus taurus* se rencontrent partout sauf dans les zones désertiques et les régions polaires. L'élevage de la plupart des races taurines se pratique dans des zones géographiques de moyennes et hautes altitudes, généralement en-dessous de 1500 m, exception faite de la Créole du Pérou qui peut s'élever à plus de 3000 m d'altitude (Figure A.1). Les populations se référant au type "zébu" *Bos taurus indicus* se retrouvent dans les régions tropicales en Inde, Indochine, Indonesie, Chine méridionale (Yunnan, Guangxi, Hainan), Afrique Australe, Brésil, Etats Unis et Australie (`www.fao.org/faostat/en/`, [Felius et al., 2014, Zhang et al., 2020]).

D'autres espèces appartenant au genre *Bos* sont présentes dans la région asiatique : les Banteng, une espèce sauvage *Bos javanicus* [Gardner et al., 2016], les Gayal *Bos frontalis* et son pendant sauvage le gaur *Bos gaurus* [Duckworth et al., 2016], ainsi que le kouprey *Bos sauveli* [Timmins et al., 2016] partagent des aires de distribution communes. Plus au nord, les Yaks *Bos grunniens* et pour le rameau sauvage *Bos mutus* vivent au niveau du plateau du Tibet, et le Nord de l'Inde [Buzzard and Berger, 2016].

Il est à noter en outre que les populations du genre *Bos* côtoient d'autres espèces du groupe "Bovini" et sont quelquefois interfertiles. C'est le cas notamment des espèces *Bison* comme le bison européen *Bison bonasus* présent en totale ou semi-liberté en Europe de l'Est et dans quelques réserves d'Europe Occidentale [Gautier et al., 2016], et le bison Nord-Américain *Bison bison* situé entre le Nord du Mexique, l'Alaska et l'Alberta Central du Canada [Aune et al., 2017].

Cependant, l'interfertilité totale entre les différentes espèces *Bos* n'est possible qu'entre les individus relevant des types *Bos taurus taurus*, les "taurins" et *Bos taurus indicus*, les "zébus" [Zhang et al., 2020]. Cette particularité représente un risque d'absorption de races locales zébuines par des races européennes hautement productives.



Famille	*Sous-Famille*	*Genre*	*Espèce*	*Sous-Espèce*

FIGURE A.1 – Arbre phylogénétique des bovidés

Les ovins se rapportant au genre *Ovis* appartiennent à la sous-famille des *Caprinae*, comprenant également les chèvres. Au sein du genre *Ovis*, plusieurs classifications ont été proposées. L'évolution de la classification des espèces sauvages connues de ce genre est résumée dans [Rezaei et al., 2010]. Néanmoins, nous prendrons comme référence la classification proposée par le NCBI [Schoch et al., 2020]. Les espèces sauvages du genre *Ovis* sont regroupées sous les taxons *O. ammon* (Argali dans les régions montagneuses de l'Asie centrale), *O. canadensis* (Big-Horn se trouvant dans tout le territoire des Montagnes Rocheuses), *O. nivicola* (le mouton des neiges rencontré au Nord Est de l'Asie), *O. dalli* (le mouton de Dalli dans les régions montagneuses Nord Américaines), *O. vignei* (l'Urial présent dans toute l'Asie mineure) et *O. orientalis* (présent en Asie occidentale). Ce dernier taxon correspond au mouflon asiatique dont la domestication aurait abouti au mouton domestique. Le mouton (*Ovis aries*) se rencontre un peu partout dans le monde avec cependant une forte densité dans les régions des îles britanniques, du bassin méditéranéen, la ceinture sahélienne, le Moyen Orient, le Sud de l'Inde, au Nord-Est de la Chine et en Mongolie. L'Afrique du Sud ainsi que la Nouvelle Zélande et les régions tempérées de l'Australie sont des zones connues pour l'élevage d'ovins. L'extrémité Sud du continent Américain et une partie des Andes sont également des territoires d'élevage de moutons [Gilbert et al., 2018].

A.2 Monographie des différentes populations étudiées

Cette monographie présente les différentes races bovines et ovines étudiées, leur histoire et leurs caractéristiques morphologiques et de productions.

A.2.1 Bovines

Nelore

La zone d'élevage du zébu Nelore est circonscrite au Brésil. Cette race descend directement en partie du zébu Ongole d'Inde dans la région administrative Ongole, dont la première importation date de 1895 jusqu'en 1962. Le dernier recensement d'effectif effectué en 2006 porte le nombre d'individus enregistrés à un peu plus de 170 000 (84 % des enregistrements). Les études génétiques ont mis en évidence la présence de 2 sous-populations dont l'origine pourrait être une importation d'autres races Indiennes et pas uniquement d'Ongole. Malgré le métissage avec des races taurines, les éleveurs brésiliens continuent à préserver la "pureté" des lignées issues des premières importations.

Ongole

Les zébus Ongole sont élevés dans la région méridionale d'Andhra Pradesh en Inde et dans des systèmes herbagers. Cette race représente une importante ressource génétique bovine et a été largement diffusée dans diverses régions du monde (Sri Lanka, Fiji, Asie du Sud-Est, aux Etats Unis contribuant ainsi à la Santa Gertrudis, au Brésil). Cette race serait proche génétiquement de populations de la Vallée de Krishna et de Kankrej. Cette race présente des niveaux de productions assez variables.

Gir

Le Gir ou Gyr est l'une des principales races de zébu originaires de l'Inde. Il a été utilisé localement pour l'amélioration d'autres races, notamment le Red Sindhi et le Sahiwal. C'est également l'une des races utilisées pour le développement de la race Brahman en Amérique du Nord. Au Brésil et dans d'autres pays d'Amérique du Sud, le Gir est fréquemment utilisé pour sa résistance aux températures chaudes et aux maladies tropicales. Elle est réputée pour ses qualités laitières et est souvent croisée avec des vaches frisonnes pour former la race Girolando.

Le Gir présente un front arrondi et bombé (c'est la seule race ultraconvexe au monde), de longues oreilles pendantes et des cornes qui s'enroulent en spirale vers l'extérieur et l'arrière. Les gir sont généralement tachetés, leurs couleurs allant du rouge au blanc en passant par le jaune, le noir étant la seule couleur non-reconnue. Originaires du sud-ouest de l'Inde, dans l'État du Gujarat, ils se sont répandus dans les États voisins du Maharashtra et du Rajasthan.

Les femelles pèsent en moyenne 385 kg et mesurent 130 cm ; les taureaux pèsent en moyenne 545 kg et mesurent 140 cm [Felius, 1995, Porter and Tebbit, 2007]. À la naissance,

les veaux pèsent environ 20 kg [Felius, 1995, Porter and Tebbit, 2007]. Le rendement laitier moyen des Gir est de 1590 kg par lactation [Porter and Tebbit, 2007], avec une production record de 3182 kg à 4,5% de matière grasse en Inde.

Les premiers bovins Gir (6300 têtes) ont été importés d'Inde au Brésil en 1906. La majeure partie étant à l'origine des noyaux existant encore aujourd'hui a été introduite en 1920. Il y a eu quelques preuves de perte de valeur génétique par la consanguinité. Dans les années 1960, les éleveurs brésiliens de Gir ont commencé à sélectionner pour la production laitière, en recherchant des caractéristiques morphologiques telles que la conformation de la mamelle et la longueur des pattes, qui ne faisaient pas partie des anciens objectifs liés à la production de viande bovine. Aujourd'hui, les éleveurs incluent des lignées pour la production de lait et de viande. Le niveau de production laitière est en moyenne 3500 kg par lactation [Porter et al., 2016].

En 2003, les Gir étaient environ 915 000, soit 37% des 2,5 millions de bovins de la région de Saurashtra au Gujarat. En 2010, la population au Brésil était estimée à environ cinq millions [Porter et al., 2016].

Brahman

Le Brahman est une race américaine de bovins à viande hybride zébuine-taurine. Elle a été élevée aux États-Unis à partir de 1885 à partir de bovins originaires d'Inde, importés à différentes époques du Royaume-Uni, d'Inde et du Brésil. Il s'agissait principalement des races Gir, Guzerá et Nelore, avec un peu d'Indu-Brasil, Krishna Valley et Ongole. Le Brahman a une grande tolérance à la chaleur, au soleil et à l'humidité, et une bonne résistance aux parasites. Elle a été exportée dans de nombreux pays, notamment sous les tropiques ; en Australie, c'est la race bovine la plus nombreuse. Elle a été utilisée dans la création de nombreux hybrides taurine-indicine, dont certains - comme le Brangus et le Brahmousin - sont établis comme des races distinctes [Felius, 1995, Porter and Tebbit, 2007, Porter et al., 2016].

La race a été introduite en Martinique (ainsi qu'en Guyane) où un organisme de sélection s'est développé et a créé son propre schéma de sélection adapté au contexte Antillais.

Red Sindhi

Les zébus Red Sindhi originaires de la province du Sindh au Pakistan sont largement élevés pour la production laitière au Pakistan, en Inde, au Bangladesh, au Sri Lanka et dans d'autres pays comme le Brésil en moindre proportion. Ils ont été utilisés pour des croisements avec des races laitières d'origine tempérée (européenne) dans de nombreux pays afin de combiner leurs adaptations tropicales (tolérance à la chaleur, résistance aux tiques, résistance aux maladies, fertilité à des températures plus élevées, etc.) avec la production laitière plus élevée que l'on trouve dans les régions tempérées. Elle a été croisée avec des Jerseys dans de nombreux endroits, notamment en Inde, aux États-Unis, en Australie, au Sri Lanka [Porter and Tebbit, 2007].

La couleur des Sindhi rouges varie d'un brun rougeâtre profond à un rouge jaunâtre, mais le plus souvent d'un rouge profond. Ils se distinguent de l'autre race laitière du Sindh, le Tharparkar, à la fois par la couleur et la forme, le Red Sindhi est plus petit, plus rond, avec une forme plus typique des races laitières, et avec des cornes courtes et incurvées. Les taureaux sont généralement d'une couleur plus foncée que les vaches.

Tharparkar

Les zébus Tharparkar (ou Thari, ou White Sindhi au Sri Lanka), élevés au Sud-Ouest du Rajasthan, au Gujarat et au Sud-Est de la province pakistanaise de Kutch, auraient été "influencés" par les zébus Kankrej[1], Red-Sindhi, Gir et Nagori[2]. Le Tharparkar est plus grand (150 - 138 cm, 600/400 kg) que le Red Sindhi avec une forme plus typique des races de trait Zebu, et avec des cornes plus longues en forme de lyre. La couleur de la robe est généralement grise claire avec des zones plus sombres (noires ou rouges).

Boran

La race bovine Boran, classée sous le nom de Bos indicus (humped Large East African Shorthorn Zebu type) est l'une des races bovines locales élevée par les pastoralistes Borana Oromo dans les zones Borana du sud de l'Éthiopie pour la production de viande et de lait. Le Boran est issue d'un métissage originel *Zebu X Taurins Africains* et a été élevée en tant que race pure pendant 1300 ans lui conférant une vigueur hybride beaucoup plus forte que les races composées modernes. Il est cependant sensible aux tiques et leurs maladies associées. Le profil génétique se compose des proportions suivantes : 64 % de Bos indicus, 24 % de Bos taurus européen (ancien et récent) et 12 % de Bos taurus africain.

Depuis 1951, la Boran Cattle Breeders' Society gère et élève stratégiquement des bovins Boran au Kenya. En 2008, on comptait environ 454 ranchs dont l'orientation était la viande dans le pays. Le Boran est de taille moyenne avec une tête courte, de petites oreilles, un fanon lâche et une grande bosse au-dessus des épaules. Ils peuvent être cornus ou polled. Leur hauteur au garrot varie de 114cm à 147cm, et le poids des taureaux est d'environ 500kg à 850kg. Les vaches pèsent environ 380 kg à 450 kg. La robe est de couleur pie noire mais aussi rouge, fauve, ou noire. La peau est lâche, épaisse et extrêmement souple pour mieux repousser les insectes. Elle est pigmentée de couleur foncée et dotée de poils fins et courts pour mieux tolérer la chaleur. Les poils peuvent être de différentes couleurs, à l'exception du bringé ou du noir uni.

1. Le zébu de Kankrej, élevé dans des systèmes pastoralistes, originaire de la zone aride et semi-arid de Rann of Kutch à Gujarat en Inde, sont des animaux de robe grise à noire et présentant des cornes en Lyre. Du fait, des déplacements saisonniers des élevages, il est possible que des métissages avec d'autres zébus à petites cornes comme les Mewati et les Nagori.

2. Le Zébu de Nagori est une race reconnue de travail, notamment de traction, élevée dans le district de Nagaur à Rajasthan en Inde Occidentale. La robe est généralement blanche ou grise, la hauteur au garrot est de 124 cm pour les femelles et de 148 cm pour les mâles.

East African Zebu

Le nom complet est Small East African Short-Horned Zebu (EASZ). Ce type de race regroupe plusieurs populations présentes au Kenya, Tanzanie, une partie de l'Ouganda jusqu'aux frontières de la République Démocratique du Congo et se concentrent dans les régions humides. Au Kenya, cette race est nommée selon le district géographique (Suk, Kamasia, Kavirondo, Teso, Masai, Samburu). La taille est en moyenne autour de 111cm - 118cm. Le niveau de production laitière est faible (2L par jour, 300 L/ lactation).

Sheko

Les Sheko seraient les derniers vestiges du bétail Shorthorn de type Brachyceros en Afrique de l'Est, probablement d'une origine similaire à celle du bétail Shorthorn de Nubie. Ainsi les bovins Sheko sont les derniers vestiges des bovins Hump less shorthorn (Bos taurus) originaires d'Afrique de l'Est. Ils ont été signalés pour la première fois en 1929 dans le sud-ouest de l'Éthiopie, puis en 1982. Cette race fait partie des races bovines d'Éthiopie qui sont traditionnellement élevées par un petit nombre d'agriculteurs locaux dans les districts humides de Sheko et de Bench, dans le cadre de systèmes agricoles mixtes culture-élevage du sud-ouest de l'Éthiopie. Actuellement, certains individus présentent de petites bosses qu'ils ont héritées de l'introgression du zébu. La race est maintenant considérée comme menacée par le métissage progressif avec les zébus locaux. Cette race est connue pour ses meilleures performances et sa capacité à survivre, à produire et à se reproduire dans les zones infestées de tsé-tsé.

Zébu de Madagascar

Ce zébu, nommé également *Malagasy*, serait issu d'un métissage entre des taurins africains présents depuis le 5-7é siècle avec des zébus d'Inde, d'Afrique ou d'Indonésie depuis le 10é siècle. La comparaison des caractéristiques phénotypiques (hauteur au garrot, développement de la bosse, longueur des cornes et des oreilles) laisse à suggérer que l'introduction au sud des bovins serait plus ancienne.

Zébu Moka

Cette race est le résultat de métissage entre des zébus de Madagascar introduits au 17-18^{iéme} sur l'île de la Réunion, des bovins Afrikaners, des Renitelos, ainsi que de bovins originaires du Yémen avec le développement de la culture du café. Reconnue comme race française menacée depuis 2016, les animaux étaient utilisées comme traction dans la culture de la canne à sucre et le pressage du jus de canne puis comme moyen de transports dans les Hauts de l'île. Le bœuf Moka porte une bosse sur le garrot l'apparentant au zébu mais moins visible cependant chez les vaches. Sa robe est généralement fauve, mais peut aller aussi du blanc au noir. Elle peut être uniforme ou tachetée. Les cornes sont en quart de lune.

Bovin créole de Maurice

Cette population bovine locale se rencontrant sur l'île Maurice serait issue d'une absorption de zébu venant de Madagascar arrivés au cours des 16^{iéme} - 17^{iéme} siècles avec des taurins européens introduits plus récemment par des colons Français au 19^{iéme} siècle. Cette population se caractérise par un petit gabarit sans cornes et sans bosse ainsi qu'une robe blanche avec des points noirs ou fauve ou pie rouge. Cette population a été délaissée peu à peu au cours du 20^{iéme} siècle face à la concurrence des races productives européennes pour le lait (Holstein) et zébuines pour la viande (Boran, Brahman).

Races locales Thaïlandaises

Quatre races de zébus autochtones ont été reconnues par le gouvernement thaïlandais. Pour notre étude les races Kho-Lan (THC) et Kho-Chon (THS) seront utilisées. Les THS sont utilisées comme traction animale car ils sont beaucoup plus grands en gabarit par rapport aux autres races. Les dernières études génétiques dont sont issus les jeux de données montrent que le THS seraient une race distincte par rapport aux trois autres [Wangkumhang et al., 2015]. Une autre race locale thailandaise distincte, White Lampun (Khao Lumpoon ou Kow Lumpoon) est élevée dans le nord du pays et présente un pelage blanc et la peau rose. L'effectif représentait 33 000 têtes en 1999 puis 1 000 femelles reproductrices en 2013 aboutissant ainsi à la mise en place du schéma de sauvetage en produisant un veau mâle de cette race par clonage de cellules somatiques.

Aceh

Les zébus Aceh se trouvent dans toute la province portant le même nom au nord de l'île de Sumatra. La couleur de la robe est diverse : gris, rouge clair, rouge foncé, noir, avec de temps en temps des taĉhes blanches sur les pattes et le ventre. La bosse est assez petite pour les 2 sexes. Le zébu Aceh serait la résultante d'une absorbtion de la race locale Madura et de 11% de Banteng [Mohamad et al., 2009]. La lignée mitochondriale est entièrement proche du zébu et génétiquement proche du Peranakan Ongole [3]que du Pesisir. Les données pour le chromosome Y sont incomplètes. Le gabarit aurait évolué au cours des dernières années : les individus seraient devenus plus petits, probablement dû à la systématisation de la castration et de l'abattage des taureaux les plus gros. Ainsi à partir de 1926, la hauteur au garrot serait passée de 116 à 105 cm.

Brebes

Une autre dénomination de ce zébu métis *Peranakan Ongole*[3] *X Madura/ Boeuf de Bali* est "Jabres". Cette population est élevée dans la province du Brebes au nord de l'île centrale de Java. La couleur de la robe est diversifiée : blanche, pie rouge, pie noir, pie brun. La hauteur

au garrot est de 130 cm pour les mâles et de 126 cm pour les femelles. Le poids moyen est de 325 kg pour les mâles et de 310 kg pour les femelles.

Madura

Ce zébu est originaire de l'île de Madura à l'est de l'île de Java et est la résultante d'un métissage *Zébu X Banteng*. C'était une population bien établie au moins à partir du 10^{iéme} siècle avant le développement du Peranakan Ongole[3]. Cette race est élevée pour plusieurs objectifs : la production de viande bovine, les courses de taureaux (karapan) et les concours de beauté de vaches (sonok). Dans le cadre de la campagne de promotion du métissage, le gouvernement local promeut le madrasin, un équivalent du sonok pour le bétail croisé. Les animaux qui participent à ces événements peuvent se vendre trois fois plus cher que leurs congénères non compétitifs et sont de plus grande taille (128 cm vs 116cm pour les bovins n'étant pas orientés vers les compétitions).

Pesisir

Les bovins Pesisir sont présents dans les systèmes d'élevages traditionnels de la côte occidentale de Sumatra et sont issus de métissage de Banteng et de Zébu. Le degré d'introgression du banteng a été estimé à 10-16 %. La couleur de leur robe est très variable et comprend le rouge-brun, le jaune, le brun foncé, le chauve ou le blanc. En 2008, l'effectif était estimé au nombre de 89'995 (une tendance à la baisse due en partie à la politique officielle ayant favorisé le bétail importé) et ils représentent 20% du bétail de l'ouest de Sumatra. Le poids corporel des vaches est de 160 kg, bien inférieur à celui des vaches de Bali ou de Madura (3010 kg ou 248 kg, respectivement), et la taille du corps diminue en raison du manque de pâturages. Les taureaux pesaient 275 kg en 1980, mais 225-250 kg en 2005 ; tandis que les vaches ont diminué de 256 kg à 172 kg au cours de la même période. La hauteur au garrot a également diminué, passant de 1150 cm à 105 cm pour les mâles et de 110 à 98 cm pour les femelles.

A.2.2 Ovines

Deccani

La race Deccani se rencontre du Nord Est du Karnatak, à l'ouest de l'Andhra pradesh, et au Maharashtra central. Cette race est de type laineux à queue courte utilisée pour la production

3. Le Peranakan Ongole, ou *Filial Ongole*, résulte d'un croisement entre des Ongole importés et des bovins autochtones de l'île de Java. Une introgression de Banteng a également été trouvée par l'analyse génétique (Mohamad et al., 2009). Bien que les zébus soient présents en Indonésie depuis 1500 ans, les zébus Ongole n'ont été utilisés en croisement avec des bovins locaux indonésiens, exceptés ceux présents sur l'île de Madura et de Bali, qu'à partir de la fin du XIXe siècle. Les importations ont été effectuées d'abord vers l'île de Sumatra par des compagnies privées mais ont cessé à cause de la peste bovine en Inde à la fin du XIXe siècle. 90% des effectifs de Peranakan Ongole se trouvent à Java, le restant à Bengkulu, à Yogyakarta, au Lampung, au Kalimantan du Sud, au Nusa Tenggara, au Timor. La robe est claire, blanche ou grise. La hauteur au garrot et le poids est estimée à 135 cm / 500kg pour les mâles, à 125 cm / 350 kg pour les femelles.

de viande. Le poids des brebis est de 30 kg à maturité, 1 agneau par mise bas, Le projet "Australian back" a été initié afin d'introduire le gène de la fécondité (FecB) à partir de Garole pour augmenter la prolificité dans les races ovines australiennes. Le pelage est de type laineux et la robe noire, pie, peut être blanche ou marron. Les béliers sont cornus, contrairement aux brebis. Plusieurs déclinaisons locales de cette race se rencontrent : Sangamneri de Ahmedhager est une variété locale dans Maharashtra, le Dhormendi (corné noir) du Maharashtra du Sud est une version plus grande du Deccani.

Garole

La race Garole est élevée dans la partie sud du Bengale occidental en Inde caractérisé par un climat chaud et humide, particulièrement sur la zone côtière des Sundarbans. Il s'agit d'une race très prolifique (nombre d'individus de la portée 2.27 dans la région du Sudarbans et de 1.74 au niveau du plateau du Deccan de Maharashtra) dont le poids adulte est de 10–14 kg. En 1792, ces ovins ont été exportés vers l'Australie et ont été croisés avec des Merinos donnant la race Booroola Merino.

Bangladeshi Garole

Le Bangladeshi Garole correspond au pendant Bangladeshi du Garole indien. Cette population est élevée dans la zone côtière des Sundarbans du Bangladesh.

Bangladeshi East

Cette population est probablement un assemblage de plusieurs populations locales ovines situées dans les régions du Bandarbar (Sud-Est), de Sylhet et Moulovibazar (Nord-Est) et de Noakhali (Sud Côtier) [Khan et al., 2009, Kijas et al., 2012]. Le nom commun pour toutes les populations locales présentes sur le territoire bengali est "Wera" (Mason 1980), néanmoins une tentative de charactérisation des populations locales en fonction des régions géographiques est en cours (Hassan 2011, Deb 2019).

Sumatra

Ce mouton local est élevé dans le Nord de Sumatra et est de type "queue fine" [Sodiq and Tawfik, 2004] avec un petit gabarit [Gatenby et al., 1997] et un potentiel reproductif (Iniguez et al 1991b)

Garut

Aussi nommé *Priangan*, ce mouton est élevé dans la partie occidentale de l'île de Java en Indonésie principalement pour les combats de béliers et la viande. Il s'agit d'une variété de la race Javanese Thin-tailed (à queue fine). Selon certaines indications, des élevages d'Africander et de Mérinos auraient été introduits au 19ème siècle. Ils sont généralement noirs ou pie,

occasionnellement des individus seront gris ou fauve. Les mâles sont cornés et les femelles sont polled (sans cornes). Les individus sont souvent dépourvus d'oreilles externes (microtie).

Red Massai

Les populations de Red Massai sont élevées par des tribus pastoralistes du Nord du Kenya (Samburu, Maasai, Bukusu, Busia, Kipsigis, Luo et Nandi) mais leur distribution spatiale peut s'étendre jusqu'au nord de la Tanzanie et les régions sèches d'Uganda. Ces populations ovines sont reconnues pour être résistantes génétiquement aux parasites intestinaux et ont été utilisées en Australie pour cette caractéristique. Néanmoins la Red Maasai subit le revers de la médaille et est en danger de disparition par l'absorption par des Dorper Australiens. La Red Maasai est du type Africain à queue graisse, à poil relativement long de couleur marron-rouge, rarement pie. De grande taille, la hauteur au garrot est de 70 cm. Quelques béliers portent des cornes mais il s'agit en général d'une population de type "polled".

Ethiopian Menz

Cette population appartient au groupe des "queues grasses" et présente un pelage de type laine grossière et ondulée. Elle représente le principal type de moutons locaux en Ethiopie, Egypte et les hauts plateaux de l'Erythrée. De petite taille, la hauteur au garrot est de 60 cm environ et le poids de 35 kg. Les béliers ont des cornes en spirale et nervurées, les brebis sont sans cornes. La localisation du gras est au niveau de la queue mais aussi sur les côtés de la tête. La robe est de couleur marron ou noir généralement avec des tâches blanches sur la tête et les pattes.

Moghani

Le nom de la race dérive de la plaine du Moghan dans la province d'Ardabil. La race, historiquement circonscrite à la zone du Moghan, a été introduite durant les 40-50 dernières années dans plusieurs régions de l'Iran (Téhéran, Semnan, Khorasan et Golistan) soit par la voie mâle, soit en race pure. Les bêliers étaient croisés avec des races de petites tailles pour produire des agneaux lourds, plus spécifiquement dans les aires de répartition des populations Afshari, Ghezel, et Taleshi. La Moghani est une race à viande dont le pelage est constitué de laine, de grand gabarit (hauteur moyenne du Moghani au garrot de 68.2cm et le périmètre cardiaque des femelles adultes de 95,1 cm), à queue grasse s'étendant jusqu'au-dessus du jarret se terminant en forme de S. Ils sont généralement d'un blanc uni, mais on trouve parfois des marques de couleur pâle sur la tête et les pieds. La tête est relativement petite et fine, avec une ligne de nez droite et un os orbital relativement proéminent. Les oreilles sont longues et de forme trapézoïdale. Le ventre et le cou sont dépourvus de laine. Les mâles et les femelles ont des crêtes.

Norduz

Mouton à queue grasse originaire de la province de Van en Anatolie de l'Est (Turquie), sous-type de la race Akkaraman (Bingol et al., 2006), il est menacé d'extinction. Cette race est similaire à l'Akkaraman et au Karakas en terme de taille, mais la couleur de la face, des oreilles et du cou est très spécifique. La face est gris-bleu et contient des nuances de noir et de brun qui peuvent s'étendre à la poitrine. Le cou peut être coloré ou blanc. La toison est généralement blanche, mais les animaux cendrés sont assez courants. Certains animaux ont des caroncules. En général, les béliers ont des cornes hélicoïdales proéminentes ; près de la moitié des brebis sont cornues. La laine convient à la fabrication de tapis car elle est principalement constituée de fibres longues, ondulées et médullées. La queue grasse est semblable à celle de l'Akkaraman. Le poids au garrot est en moyenne de 70,8 cm et le tour de poitrine de 98,0 cm, pour un poids corporel de 63,5 kg.

Qezel/Ghezel

(Autre dénomination "Red Kamaran") Les moutons Qezel sont originaires du nord-ouest de l'Iran et du nord-est de la Turquie. Cette région d'Iran, connue sous le nom d'Azarbayjan, est caractérisée par un climat sec et froid dans les montagnes. La toison de ce mouton est rouge et comporte des zones rouge noir ou rouge clair. Le Ghezel est utilisé à la fois pour la viande et la laine dans cette région. Les mâles sont généralement sans cornes et les femelles le sont toujours. La queue est de type "queue grasse".

Le poids à la naissance, le poids au sevrage, le poids à six mois, le gain quotidien moyen de la naissance au sevrage et le gain quotidien moyen du sevrage à six mois sont de 4,47, 20,88, 33,14 kg, 0,184 et 0,134 gr.

Mouton Comorien

2 types semblent cohabiter sur l'archipel comorien : type Djallongué et type African long fat tailed.

Mouton Malgache (Malagazy)

Ce mouton présent dans la partie Sud-Ouest de Madagascar est de type à "poil" et à "queue grasse" et a été peu étudié récemment. Les dernières références date du milieu du XX° siècle [Anonymous, 1952]. Le pelage blanc ou brun clair, souvent avec de grandes taches rouges, brunes ou noires et constitué de fibres grossières. La tête et le cou sont souvent noirs, le reste du corps étant blanc selon le même motif que le mouton somali à tête noire. Les sabots sont noirs, le profil de la tête est convexe, les oreilles sont longues et pendantes, les cornes sont petites ou absentes, le fanon est parfois bien développé. La queue longue et grasse bien développée en forme de faucille et le pelage à poils grossiers de cette race suggèrent une

origine commune avec d'autres types de moutons africains à queue grasse. En apparence, ils ressemblent aux moutons à tête noire du Tanganyika et de l'Ouganda.

A.3 Histoire des épizooties principales dans la zone

Peste bovine du XIXe siècle La peste bovine était présente au XIXe siècle en Russie où elle s'est diffusée en Europe et a causé dans les années 1850-1870 la mort de centaines de millions de bovins. Elle était également présente en Afrique mais resta circonscrite à l'Egypte au début du XIXe [Pastoret et al., 2006]. La peste bovine à l'orgine de la Grande Pandémie Africaine s'est introduite en Abyssinie (Erythrée - Ethiopie) au début du XXe siècle au moyen d'un animal par des italiens lors d'une campagne militaire [Spinage, 2003] à partir des Indes [Pastoret et al., 2006, Roeder et al., 2006, Roeder, 2011]. La pandémie s'est ensuite étendue à la globalité du territoire africain jusqu'à la côte Atlantique de l'Afrique et le Cap de bonne espérance, la pointe Sud africaine en décimant une large proportion du cheptel bovin mais également d'espèces *ongulées* de la faune sauvage. Elle ne fut déclarée "éradiquée" qu'en 2011 (OIE) après plusieurs épisodes épidémiques. La peste bovine était aussi présente dans toute l'Asie orientale, indonésienne, indienne, centrale. Elle a également été introduite à plusieurs reprises au Moyen Orient [Roeder, 2011]. Concernant les îles de l'Ouest de l'Océan Indien, Madagascar serait restée exempte, mais la Réunion aurait vu son cheptel décimé entre 1900 et 1902, ainsi qu'une des îles des Comores en 1901 [Barrett et al., 2006].

East Coast Fever Il s'agit d'une maladie dont l'agent étiologique est un parasite protozoaire *Theileria parva* transmis par les tiques (*Rhipicephalus appendiculatus*. Ce parasite a été reconnu pour la première fois en Rhodésie du Sud (aujourd'hui Zimbabwe) en 1901/02. Elle a été introduite dans la région par l'importation de bétail de Tanzanie et du Kenya pour le repeuplement et a dévasté le cheptel bovin d'Afrique australe en trois ans en raison de sa grande sensibilité. Selon une estimation, elle a causé la mortalité de 1,25 million de bovins sur les 4 millions de la zone touchée en 1914. Elle est également apparue en Zambie, au Malawi et au Mozambique (province de Tete) entre 1912 et 1922. L'infection est toujours répandue dans ces pays et inflige des pertes économiques à la communauté des éleveurs (Yusufmiaet al.,2010).

La East Coast Fever a été mentionnée aux Comores (Grande Comores notamment) entre 2003 et 2004 pour la première fois à la suite d'un import massif de bétail (Deken 2007). Du fait des entrées clandestines entre la république des Comores et l'île française de Mayotte, l'entrée de cette pathologie ne peut pas être écartée.

Les bovins taurins de race pure élevés dans des conditions exemptes de tiques sont très sensibles à l'ECF et les races zébus (telles que les Boran améliorés et non améliorés) élevées dans des conditions exemptes de tiques sont modérément sensibles à la maladie. En revanche, les zébus élevés dans des zones endémiques de l'ECF sont peu sensibles à la maladie. Cette sensibilité serait lié aux systèmes de pâturage, car les races exotiques sensibles à l'ECF sont principalement élevées dans des systèmes de petites exploitations laitières où l'exposition aux tiques est limitée, tandis que les races indigènes résistantes à l'ECF sont élevées dans un système de pâturage ouvert qui permet l'exposition aux tiques infectées (Gachochi 2012).

Fièvre de la vallée du Rift La Fièvre de la vallée du Rift (FVR) est une zoonose virale africaine, affectant principalement les animaux domestiques ruminants et pouvant se transmettre aux humains. Elle se manifeste comme une fièvre, souvent hémorragique chez l'animal et parfois (1 à 3 % des cas) chez l'homme. Elle est due à l'espèce de virus *Rift Valley fever phlebovirus* VFVR. Chez l'animal, la maladie se transmet par piqûres de moustiques, principalement du genre Aedes. Une première description probable de la maladie a été faite en 1913 dans la vallée du Rift au Kenya, chez des moutons laineux d'importation. La fièvre de la vallée du Rift (FVR) est décrite en 1931, avec plus de précisions dans la même région, près du lac Naivasha. Le virus est isolé en 1934. La maladie touche les ruminants domestiques, en particulier les petits ruminants, jeunes et femelles gravides. Maladie aiguë et souvent mortelle, elle procède par épizooties périodiques, survenant lors des saisons de pluies intenses et prolongées. Les épizooties débutantes se manifestent habituellement par une vague d'avortement inexpliqués. Le taux d'avortement est pratiquement de 100 % chez les brebis pleines ; relativement moins, jusqu'à 85 %, chez les vaches.

À Mayotte, le virus de la FVR a été détecté pour la première fois en 2007 chez l'homme mais était présente depuis au moins 2004. Le séquençage a permis de rattacher ce virus à la lignée Kenya-1 d'Afrique de l'Est de 2006-2007 et s'étant propagé en Afrique de l'Est et du Sud au cours des mois suivants, notamment en Somalie, en Tanzanie, au Soudan, au Mozambique et dans l'Union des Comores, suggérant une importation récente sur l'île depuis le continent africain. Des analyses sérologiques rétrospectives effectuées sur des sérums de bétail (collectés entre 2004 et 2008) ont montré que le virus de la FVR était présent sur l'île au moins depuis 2004. Depuis 2009, la maladie est surveillée suite à une réémergence observée chez le bétail en 2008-10. Une épidémie beaucoup plus récente est réapparue suite à une épidémie déclarée au Kenya en 2018.

Annexe B

Méthodes et outils

B.1 Récapitulatifs des analyses des jeux de données

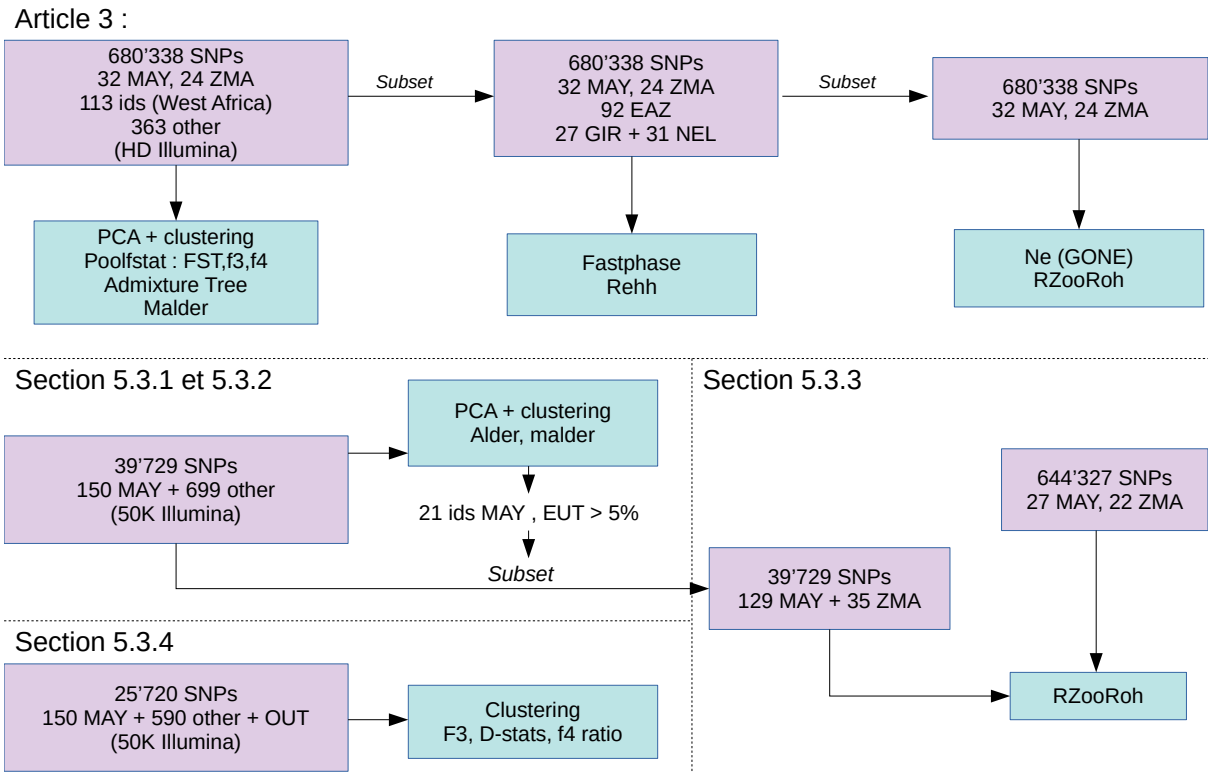


FIGURE B.1 – Schéma récapitulatif des analyses des différents jeux de données utilisés dans les sections du Chapitre 5.

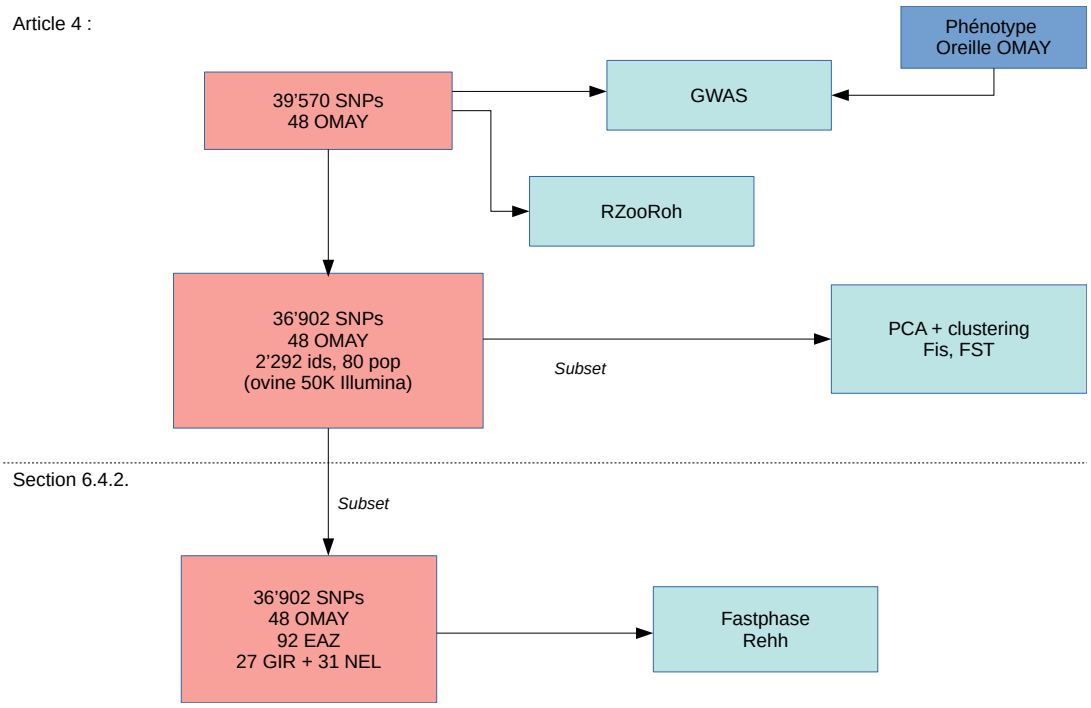


FIGURE B.2 – Schéma récapitulatif des analyses des différents jeux de données étudiés dans le Chapitre 6.

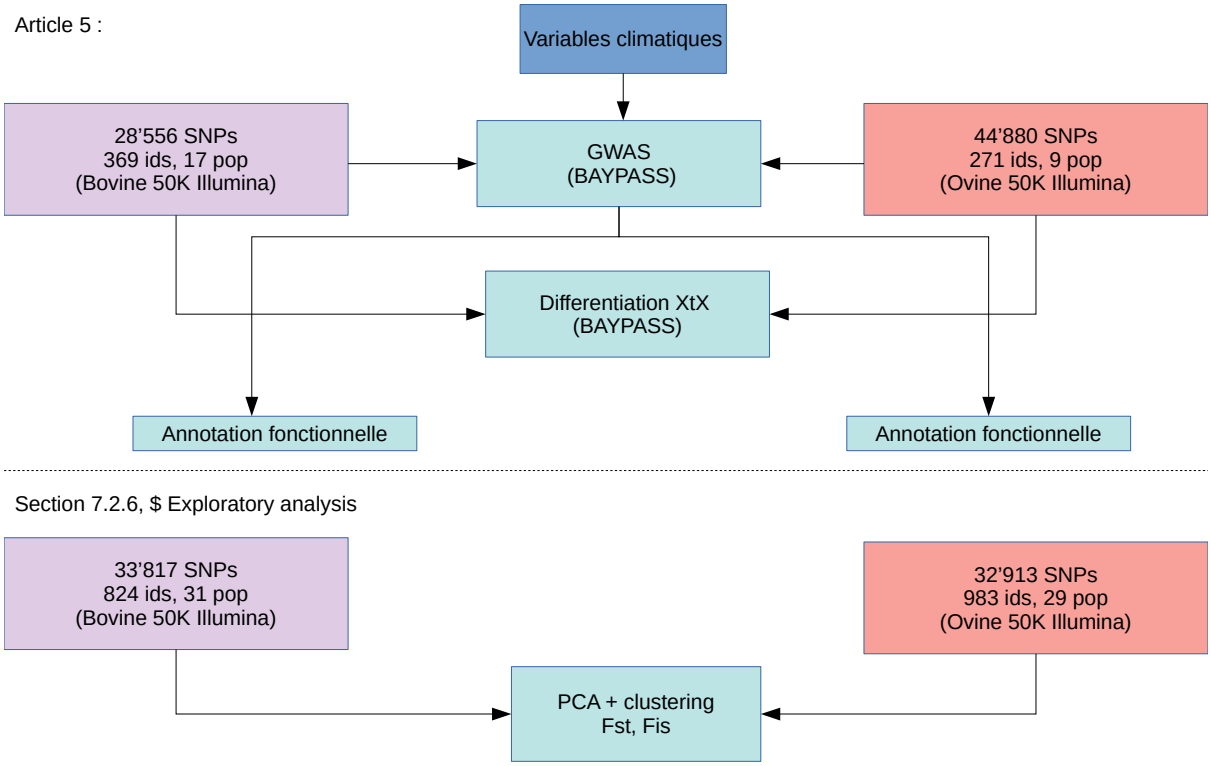


FIGURE B.3 – Schéma récaputilatif des analyses des différents jeux de données étudiés dans le Chapitre 7.

B.2 Pipelines

B.2.1 Conversion des positions des SNPs

```
\$plink --file datav1 --exclude snp_exclude --make-bed --species --noweb
        --out data_tmp
\$plink --bfile data_tmp --species --update-chr assembly.chr --make-bed
        --out chr_data_tmp
\$plink --bfile data_tmp --species --update-map assembly.pos --make-bed
        --out pos_chr_data_tmp
\$plink --bfile pos_chr_data_tmp --species--recode --out data_new
```

B.2.2 Analyse exploratoire

ACP

```
\$ smartpca -i data.geno  -a data.snp -b data.ind  -o output.pca
            -p output.plot -e output.eval  -l output.log -k 33
```

admixture

```
\$ plink --file data --cow --recode12 --noweb --out data_12
\$ admixture data_12.ped 3
```

Statistiques

```
\$ plink --file data --species --recode12 --noweb --out data_12
```

```
>dataset<-read.geno.ped("data_12.ped")
>popdet.obj<-geno2popdet(dataset)
>res<-fstats(popdet.obj,niter=100,c.int=0.95,n.snp.jack=1000)
>mat_fst<-pairwise.fst(popdet.obj)
>dist2njtree(mat_fst)
```

B.2.3 Inférence démographique

Consanguinité récente

```
\$ plink --noweb --species --file data --recode oxford --autosome
--out data_ox
```

```
>library(RZooRoH)
>zoopop<-zoodata(genofile = "data_ox.gen", zformat = "gp", supcol=5,
    samplefile="data_ox.sample")
#Pour puce medium density
>mod9<-zoomodel(K=9)
>res9data <-zoorun(mod9, zoodata)
#Pour puce high density
>mod12<-zoomodel(K=12)
>res12data <-zoorun(mod12, zoodata)
```

F-statistiques

```
>library(poolfstat)
>dataset.fstats<-compute.fstats(dataset,nsnp.per.bjack.block = 1000,
computeDstat = TRUE)
```

B.2.4 Recherche des signatures de sélection -Rehh

Phasage - calcul des scores

```
\$plink --noweb --species --recode fastphase --file data --out data_fast
#Pour chaque chromosome "chr"
#Pour les bovins :
\$fastPHASE -ulist_ind -Ku50 -Kl5 -Ki5 data_fast.{chr}.recode.phase.inp -oRES

#Pour les ovins :
\$fastPHASE -ulist_ind -Ku50 -Kl9 -Ki9 data_fast.{chr}.recode.phase.inp -oRES
```

```
>library(rehh)
#Pour iHS
>for(i in 1:26){
  hap_file=paste("RES_chr-",i,"_hapguess_switch.out",sep="")
  data<-data2haplohh(hap_file = hap_file, map_file = "SNP_file",
    chr.name = i, popsel = 5, allele_coding = "none")
  scan<-scan_hh(data, polarized = FALSE)
  if(i==1){wg.res<-scan}else{wg.res<-rbind(wg.res,scan)}
}

>pop.ihs<-ihh2ihs(wg.res, freqbin = 1)
>data_ihs_pop=pop.ihs\$ihs
>data_ihs_pop_ssNA <- na.omit(data_ihs_pop)

#Pour Rsb

>for(i in 1:26){
  hap_file=paste("RES_chr-",i,"_hapguess_switch.out",sep="")
  data<-data2haplohh(hap_file = hap_file, map_file = "SNP_file",
    chr.name = i, popsel = 5, allele_coding = "none")
  scan<-scan_hh(data, polarized = FALSE)
  if(i==1){wg.res.pop1<-scan}else{wg.res.pop1<-rbind(wg.res.pop1,scan)}
}

>for(i in 1:26){
  hap_file=paste("RES_chr-",i,"_AFO_hapguess_switch.out",sep="")
  data<-data2haplohh(hap_file = hap_file, map_file = "SNP_file",
    chr.name = i, popsel = 10, allele_coding = "none")
  scan<-scan_hh(data, polarized = FALSE)
  if(i==1){wg.res.pop2<-scan}else{wg.res.pop2<-rbind(wg.res.pop2,scan)}
}

>pop1_pop2.rsb<-ines2rsb(scan_pop1 = wg.res.pop1,scan_pop2 = wg.res.pop2,
    popname1 = "MAY", popname2 = "AFO")

>data_rsb_pop1_pop2=pop1_pop2.rsb
>data_rsb_pop1_pop2_ssNA <- na.omit(data_rsb_pop1_pop2)
```

Recherche des gènes candidats

```
#approche region candidates

>reg\_data=calc\_candidate\_regions(pop.rsb,threshold=2,pval=TRUE,
    window\_size=500000,overlap=25000,min\_n\_extr\_mrk=2)
>manhattanplot(pop.ihs, pval = TRUE, threshold = thr, cr = reg\_data)
>data_ihs_pop_ssNA_supthr<-subset(data_ihs_pop_ssNA,
        data_ihs_pop_ssNA[,4]>=thr)
>reg_ihs_pos<-safe_left_join(reg\_data, data_ihs_pop_ssNA_supthr,  ~
                  X("CHR") == Y("CHR") &
                  X("START") <= Y("POSITION") &
                  X("END") >= Y("POSITION") &
                  X("MAX_MRK") == Y("LOGPVALUE")    ,
                conflict = ~.x)

>reg_snp<-cbind(reg_ihs_pos,
    start_pos_25kb = reg_ihs_pos\$POSITION-25000,
    end_pos_25kb=reg_ihs_pos\$POSITION+25000)

>assembly<-read.table("refGene.txt")
>gene_candidat<-safe_left_join( reg_snp,assembly,  ~
                       X("CHR") == Y("CHR") &
                       X("start_pos_25kb") <= Y("START") &
                       X("end_pos_25kb") >= Y("END") ,
                     conflict = ~.x)
```

```
#approche gène candidats

> gene_cand<-safe_inner_join(data_ihs_pop_ssNA_supthr, assembly,  ~
                       X("CHR") == Y("CHR") &
                       X("POSITION") >= Y("START") &
                       X("POSITION") <= Y("END") ,
                     conflict = ~.x)
```

B.2.5 Tests d'association génome-environnement

Préparation des fichiers

```
>library(rgdal)
>library(raster)
>library(plyr)

```

```
#création d'un data frame avec long et lat
>coord2 <- id_gps[,-1]
>coordinates(coord2)<-c("Longitude","Latitude")
>pos_id=row.names(coord2)
##ATTENTION, mettre longitude puis latitude dans le fichier

>folders <- list.files(path="~/CM10_1975H_Bio_V1.2",
        full.names = TRUE, recursive=FALSE)

>bio_df <- c()

#Note: this will cause an error for the last "info" folder
        #which does not have an .adf file
>for (currentfile in folders) {
  #path <- paste0(currentfile)
  bio <- raster(currentfile)
  bioextracted <- extract(bio, coord2)
  bio_df <- cbind(bio_df, bioextracted)
  colnames(bio_df)[colnames(bio_df)=="bio_extracted"]
        <- currentfile
}

>for (currentfile in folders) {
  print(currentfile)
}

#Création du plot pour la variable THI

>Bio01=raster("~/CM10_1975H_Bio01_V1.2.txt")
>Bio28=raster("~/CM10_1975H_Bio28_V1.2.txt")

>thi <- overlay(Bio01, Bio28, fun=function(x,y){
    return((1.8*x+32)-(0.55-0.0055*y)*(1.8*x-26)
        )})

>b <- as(extent(30, 120, -26, 27.5), 'SpatialPolygons')

>crs(b) <- crs(thi)
>bio_crop <- crop(thi, b)

```

```
>bioXX<-ggplot(latlong, aes(Longitude, Latitude)) +
  layer_spatial(bio_crop, aes(fill = stat(band1))) +
  scale_fill_viridis_c(na.value = NA, option = "C")+
  geom_spatial_point() +
  ggtitle("A. BioXX - Nom de la variable (Symb)")+
  theme_void()+
  theme(legend.title=element_blank(),
        plot.title = element_text(size = 12))+
  coord_sf(xlim=c(0,150),ylim=c(-50,75))

#création de la matrice des covariables climatiques

>bio_df <- data.frame(bio_df)
>rownames(bio_df)=row.names(coord2)

#colnames(bio_df)=c(1:40)
>colnames(bio_df)=paste("Bio",1:40,sep="")

#calcul du THI (Formula 5 from National Research Council 1971)
>THI=(1.8*bio_df\$Bio1+32)-(0.55-0.0055*bio_df\$Bio28)*(1.8*bio_df\$Bio1-26)
>bio_df<-cbind.data.frame(bio_df,THI)
>bio_df35 <-bio_df[,-c(36:40)]
>bio_df35 <- cbind.data.frame(bio_df35, id_gps[,c(2:3)])
>bio_df6<- bio_df35[,1:36]
>bio_df6<-bio_df6[,-c(2:6,8:11,13:19,21:27,29:35)]
>transp_bio_df6<-t(bio_df6)

#ACP des covariables climatiques

>library(FactoMineR)
>library(factoextra)
>library(tidyverse)

#ACP Dim 1-2
>res.pca <- PCA(bio_df35, quanti.sup = c(37,38) ,graph=FALSE)

>fviz_eig(res.pca, addlabels = TRUE, ylim = c(0, 50))

>fviz_pca_var(res.pca, col.var = "cos2",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
```

```
            repel = TRUE # Évite le chevauchement de texte
)

>fviz_cos2(res.pca, choice = "var", axes = 1:2)

>fviz_pca_ind (res.pca, col.ind = "cos2",
              gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
              repel = FALSE # Évite le chevauchement de texte
)

#Préparation des données génomiques de populations

>pop<-read.geno.ped("data_12.ped")
>geno2baypass(pop)
```

Running du logiciel baypass- core analysis

```
\$baypass -gfile genobaypass -nthreads 8 -outprefix anacore_data
```

```
###Résultats core model - XtX statistics
#upload the estimated Omega matrix
omega=as.matrix(read.table("anacore_20210729_mat_omega.out"))

pop.names=c("BGA","BGE","EMZ","GAR","GUR","IDC","MAY","RMA","SUM")
#en faisant attention que pop.names = genobaypass.names
dimnames(omega)=list(pop.names,pop.names)

# Visualization of the matrix
# Using SVD decomposition
plot.omega(omega=omega,pop.names=pop.names)
# as a correlation plot
>require(corrplot)
>cor.mat=cov2cor(omega)
>corrplot(cor.mat,method="color",mar=c(2,1,2,2)+0.1,
         main=expression("Correlation map based on"~hat(Omega)))
# as a heatmap and hierarchical clustering tree (using the average
            #agglomeration method)
>hclust.ave <- function(x) hclust(x, method="average")
>heatmap(1-cor.mat,hclustfun = hclust.ave,
        main=expression("Heatmap of
        "~hat(Omega)~"("*d[ij]*"=1-"*rho[ij]*")"))
```

```
>anacore.snp.res=read.table("anacore_data_summary_pi_xtx.out",h=T)
        map=read.table("~/data_12.map", header=F)

#Estimates of the XtX differentiation measures (using the calibrated XtXst
        # estimator)
#check behavior of the p-values associated to the XtXst estimator
>hist(10**(-1*anacore.snp.res\$log10.1.pval.),freq=F,breaks=50)
>abline(h=1)

#layout(matrix(1:2,2,1))
manplot(map, anacore.snp.res\$XtXst, pch = 20,main = "XtXst",ylab="",xlab="",cex.main

>manplot(map, anacore.snp.res\$log10.1.pval, ylim.scan = 3,
    pch = 20,main = "XtX P-value (-log10 scale)",
    ylab="",xlab="",cex.main=1,cex.lab=1,ymin=NA,
    ymax=NA,col.horiz="black",lty.horiz=4)
 #0.001 p--value theshold

#Calibration des XtX statistics

#get estimates (post. mean) of both the a_pi and b_pi parameters of
#the Pi Beta distribution
>pi.beta.coef=read.table("anacore_20210729_summary_beta_params.out",h=T)\$Mean
#upload the original data to obtain total allele count
>IO_sheep.data<-geno2YN("genobaypass")

#Create the POD
>simu.sheep<-simulate.baypass(omega.mat=omega,nsnp=100000,
    sample.size=IO_sheep.data\$NN,
>beta.pi=pi.beta.coef,pi.maf=0,suffix="sheeppods")

\$baypass -gfile G.sheeppods -nthreads 6 -outprefix anapod

>pod.omega=as.matrix(read.table("anapod_mat_omega.out"))
>pod.pi.beta.coef=read.table("anapod_summary_beta_params.out"
        ,h=T)\$Mean
>pod.xtx=read.table("anapod_summary_pi_xtx.out",h=T)\$M_XtX

#get estimate of omega from the POD analysis
>plot(pod.omega,omega) ; abline(a=0,b=1)
```

```

#get estimates (post. mean) of both the a_pi and b_pi parameters of
#the Pi Beta distribution from the POD analysis
>plot(pod.pi.beta.coef,pi.beta.coef) ; abline(a=0,b=1)
>fmd.dist(pod.omega,omega)


##########################################################
#XtX calibration
##########################################################
#get the pod XtX
#pod.xtx
#compute the 0.1\% threshold
>pod.thresh=quantile(pod.xtx,probs=0.999)

>manplot(map, anacore.snp.res\$M_XtX, ylim.scan = pod.tresh)
```

Running du logiciel baypass - AUX analysis

```
\$baypass -gfile genobaypass -efile mat_cov_pop_clim -nthreads 6
    -auxmodel -scalecov -omegafile anacore_data_mat_omega.out
    -outprefix anacovaux
```

```
>covaux.snp.res=read.table("anacovaux_summary_betai.out",h=T)
>covaux.snp.xtx=read.table("anacovaux_summary_pi_xtx.out",h=T)
```

B.2.6 Tests d'association microtie

```
#Préparation des fichiers
\$awk '{for(i=7;i<=79147;i++) printf("\%s\%s",\$i,(i==79147)?"\n":OFS)}'
        may.ped>tt

#création d'un fichier avec les phéno et 2 colonnes supplémentaires
\$awk '{if(FILENAME==ARGV[1]){SEXE[\$1]=\$2;PHEN[\$1]=\$3}
    else
        {print\$1,\$2,\$3,\$4,SEXE[\$2],PHEN[\$2]}}'
            pheno_oreilles.txt may.ped >ttt

\$awk '{if(FILENAME==ARGV[1]){INF[FNR]=\$0}
    else
        {print INF[FNR],\$0}}' ttt tt> may.ped
```

```
14
15 #Test du chi2 + Bonferonni
16 \$plink --file may --sheep --assoc --noweb
17 \$awk '{if(FNR==1){print \$0,"P\_BONF"}
18     else
19         {if(\$9<0.0001){print \$0,\$9*39570}}}'
20             plink.assoc > may_plink_assoc_Pinf0001
21
22 #Test exact de Fisher
23 \$plink --file may --sheep  --fisher  --noweb
24 \$awk '{if(FNR==1){print \$0,"P\_BONF"}
25     else
26         {if(\$8<0.0001){print \$0,\$8*39570}}}'
27         plink.assoc.fisher > may_plink_fisher_Pinf0001
28
29 #Test de Cochran-Armitage
30 \$plink --file may --sheep --model --cell 1 --noweb
31 \$awk '{if(FNR==1){print \$0,"P_BONF"}
32     else
33         {if(\$10<0.0001){print \$0,\$10*39570}}}'
34         plink.model > may_plink_model_cell1_Pinf0001
35
36 #GCTA MLMA : mixed linear model based association analysis
37 \$gcta64 --mlma --bfile may  --autosome-num 26 --pheno pheno
38     --out micro_mlma --thread-num 10
39 \$awk '{if(FNR==1){print \$0,"P_BONF"}
40     else
41         {if(\$9<0.0001){print \$0,\$9*39570}}}'
42         micro_mlma.mlma > micro_mlma_Pinf0001
```

Bibliographie

[Anonymous, 1952] Anonymous, N. (1952). Rapport sur l'élevage à Madagascar. *Revue d'élevage et de médecine vétérinaire des pays tropicaux*, 5 :139.

[Aune et al., 2017] Aune, K., Jørgensen, D., and Gates, C. (2017). Bison bison (errata version published in 2018). *The IUCN Red List of Threatened Species*, pages 2017–3.

[Barrett et al., 2006] Barrett, T., Pastoret, P.-P., and Taylor, W. P., editors (2006). *Rinderpest and peste des petits ruminants : virus plagues of large and small ruminants*. Biology of animal infections. Academic Press, London. OCLC : 249810303.

[Buzzard and Berger, 2016] Buzzard, P. and Berger, J. (2016). Bos mutus. *IUCN Red List of Threatened Species*.

[Duckworth et al., 2016] Duckworth, J., Sankar, K., Williams, A., Samba Kumar, N., and Timmins, R. (2016). *Bos gaurus. The IUCN Red List of Threatened Species 2016 : e. T2891A46363646*.

[Felius, 1995] Felius, M. (1995). Cattle breeds : an encyclopedia. *Cattle breeds : an encyclopedia.*, (Ed. 1).

[Felius et al., 2014] Felius, M., Beerling, M.-L., Buchanan, D. S., Theunissen, B., Koolmees, P. A., and Lenstra, J. A. (2014). On the History of Cattle Genetic Resources. *Diversity*, 6(4) :705–750.

[Gardner et al., 2016] Gardner, P., Hedges, S., Pudyatmoko, S., Gray, T., and Timmins, R. (2016). Bos javanicus. *The IUCN Red List of Threatened Species*, pages 2016–2.

[Gatenby et al., 1997] Gatenby, R. M., Bradford, G. E., Doloksaribu, M., Romjali, E., Pitono, A. D., and Sakul, H. (1997). Comparison of Sumatra sheep and three hair sheep crossbreds. I. Growth, mortality and wool cover of F1 lambs. *Small Ruminant Research*, 25(1) :1–7.

[Gautier et al., 2016] Gautier, M., Moazami-Goudarzi, K., Levéziel, H., Parinello, H., Grohs, C., Rialle, S., Kowalczyk, R., and Flori, L. (2016). Deciphering the Wisent Demographic and Adaptive Histories from Individual Whole-Genome Sequences. *Molecular Biology and Evolution*, 33(11) :2801–2814.

[Gilbert et al., 2018] Gilbert, M., Nicolas, G., Cinardi, G., Van Boeckel, T. P., Vanwambeke, S. O., Wint, G. R. W., and Robinson, T. P. (2018). Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Scientific Data*, 5(1) :180227.

[Khan et al., 2009] Khan, M. Y. A., Husain, S. S., Alam, M. R., Teneva, A., Han, J. L., and Faruque, M. O. (2009). Genetic relationships in different sheep populations of Bangladesh based on microsatellite markers. *Journal of the Bangladesh Agricultural University*, 7(2) :291–294.

[Kijas et al., 2012] Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., and Dalrymple, B. (2012). Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biology*, 10(2).

[Mohamad et al., 2009] Mohamad, K., Olsson, M., Tol, H. T. A. v., Mikko, S., Vlamings, B. H., Andersson, G., Rodríguez-Martínez, H., Purwantara, B., Paling, R. W., Colenbrander, B., and Lenstra, J. A. (2009). On the Origin of Indonesian Cattle. *PLOS ONE*, 4(5) :e5490.

[Pastoret et al., 2006] Pastoret, P.-P., Yamanouchi, K., Mueller-Doblies, U., Rweyemamu, M. M., Horzinek, M., and Barrett, T. (2006). 5 - Rinderpest — an old and worldwide story : history to c. 1902. In Barrett, T., Pastoret, P.-P., and Taylor, W. P., editors, *Rinderpest and Peste des Petits Ruminants*, Biology of Animal Infections, pages 86–VI. Academic Press, Oxford.

[Porter et al., 2016] Porter, V., Alderson, L., Hall, S. J. G., and Sponenberg, D. P. (2016). *Mason's World Encyclopedia of Livestock Breeds and Breeding, 2 Volume Pack*. CABI. Google-Books-ID : 2UEJDAAAQBAJ.

[Porter and Tebbit, 2007] Porter, V. and Tebbit, J. (2007). *Cattle : a handbook to the breeds of the world*. Crowood, Ramsbury, Marlborough, Wiltshire. OCLC : ocm84151515.

[Rezaei et al., 2010] Rezaei, H. R., Naderi, S., Chintauan-Marquier, I. C., Taberlet, P., Virk, A. T., Naghash, H. R., Rioux, D., Kaboli, M., and Pompanon, F. (2010). Evolution and taxonomy of the wild species of the genus Ovis (Mammalia, Artiodactyla, Bovidae). *Molecular Phylogenetics and Evolution*, 54(2) :315–326.

[Roeder, 2011] Roeder, P. L. (2011). Rinderpest : The end of cattle plague. *Preventive Veterinary Medicine*, 102(2) :98–106.

[Roeder et al., 2006] Roeder, P. L., Taylor, W. P., and Rweyemamu, M. M. (2006). 6 - Rinderpest in the twentieth and twenty-first centuries. In Barrett, T., Pastoret, P.-P., and Taylor, W. P., editors, *Rinderpest and Peste des Petits Ruminants*, Biology of Animal Infections, pages 105–VII. Academic Press, Oxford.

[Schoch et al., 2020] Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., Leipe, D., Mcveigh, R., O'Neill, K., Robbertse, B., Sharma, S., Soussov, V., Sullivan, J. P., Sun, L., Turner, S., and Karsch-Mizrachi, I. (2020). NCBI Taxonomy : a comprehensive update on curation, resources and tools. *Database : The Journal of Biological Databases and Curation*, 2020 :baaa062.

[Sodiq and Tawfik, 2004] Sodiq, A. and Tawfik, E. S. (2004). Productivity and breeding strategies of sheep in Indonesia : a review. *Journal of Agriculture and Rural Development in the Tropics and Subtropics (JARTS)*, 105(1) :71–82.

[Spinage, 2003] Spinage, C. A. (2003). *Cattle Plague : A History*. Springer Science & Business Media. Google-Books-ID : t5QUSfS8FfYC.

[Timmins et al., 2016] Timmins, R., Hedges, S., and Duckworth, J. (2016). Bos sauveli. *The IUCN Red List of Threatened Species*, pages e–T2890A46363360.

[Utsunomiya et al., 2019] Utsunomiya, Y. T., Milanesi, M., Fortes, M. R. S., Porto-Neto, L. R., Utsunomiya, A. T. H., Silva, M. V. G. B., Garcia, J. F., and Ajmone-Marsan, P. (2019). Genomic clues of the evolutionary history of Bos indicus cattle. *Animal Genetics*, 50(6) :557–568.

[Wangkumhang et al., 2015] Wangkumhang, P., Wilantho, A., Shaw, P. J., Flori, L., Moazami-Goudarzi, K., Gautier, M., Duangjinda, M., Assawamakin, A., and Tongsima, S. (2015). Genetic analysis of Thai cattle reveals a Southeast Asian indicine ancestry. *PeerJ*, 3 :e1318.

[Zhang et al., 2020] Zhang, K., Lenstra, J. A., Zhang, S., Liu, W., and Liu, J. (2020). Evolution and domestication of the Bovini species. *Animal Genetics*, 51(5) :637–657.