HAL
open science

Détection chez le bovin de polymorphismes génétiques au niveau du génome mitochondrial et des chromosomes sexuels et caractérisation de leurs effets sur les caractères de production, reproduction et santé.

Clémentine Escouflaire

To cite this version:

Clémentine Escouflaire. Détection chez le bovin de polymorphismes génétiques au niveau du génome mitochondrial et des chromosomes sexuels et caractérisation de leurs effets sur les caractères de production, reproduction et santé.. Génétique animale. Université Paris-Saclay, 2020. Français. NNT : 2020UPASA015 . tel-03223090

HAL Id: tel-03223090

https://pastel.hal.science/tel-03223090

Submitted on 10 May 2021

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

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

Détection chez le bovin de polymorphismes génétiques au niveau du génome mitochondrial et des chromosomes sexuels et caractérisation de leurs effets sur les caractères de production, reproduction et santé

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°581, agriculture, alimentation, biologie, environnement et santé (ABIES)
Spécialité de doctorat : Génétique Animale
Unité de recherche : Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France
Référent : AgroParisTech

Thèse présentée et soutenue à Paris-Saclay, le 16 juin 2020, par

Clémentine ESCOUFLAIRE

Composition du Jury

Etienne VERRIER Professeur, AgroParisTech	Président
Juliette RIQUET Directrice de Recherche, INRAE Centre Occitanie-Toulouse	Rapporteur & Examinatrice
Bertrand BED'HOM Professeur, Muséum National d'Histoire Naturelle	Rapporteur & Examinateur
Susana DUNNER Professeure, Universidad Complutense de Madrid	Examinatrice
Laurent TIRET Professeur, EnvA	Examinateur
Helene HAYES Directrice de Recherche, INRAE Centre IdF-Jouy-en-Josas-Antony	Directrice de thèse
Aurélien CAPITAN Ingénieur de Recherche, Allice	Co-Encadrant & Examinateur

Thèse de doctorat

NNT : 2020UPASA015

Remerciements

Au terme de ces trois ans de thèse, je souhaite remercier toutes les personnes qui ont aidé à rendre la route parcourue plus facile et agréable.

Mes premiers remerciements vont à Aurélien. Merci d'avoir accepté de donner une chance à la jeune étudiante qui t'a d'abord demandé un stage, puis qui a vendu son âme pour suivre le cheminement tortueux et tentaculaire de ton esprit et s'engager dans une thèse. Ta passion pour ton travail et les rouages grippés des mécaniques du vivant aboutissant aux anomalies est communicative et a rendu ces trois années très enthousiasmantes. Merci pour ton optimisme toujours prêt à contrer mes doutes. Même si ce sujet se prêtait tout à fait à ta capitanite aigüe, et que nous n'en sommes pas encore au bout, je ne doute pas que les prochaines études seront des défis tout aussi intéressants.

Hélène, merci de t'être embarquée dans cette aventure avec nous, sans même me connaître. Ta confiance et ton soutien sans faille ni doute a permis de maintenir le cap le plus droit possible en nous évitant de nombreux écueils. Merci également de ta patience et de tes relectures bienveillantes.

Je remercie Allice de m'avoir donné l'opportunité de réaliser cette thèse au sein de la structure et plus particulièrement Laurent pour nous avoir conseillé et fait confiance à Aurélien et moi. Je souhaite aussi remercier APIS-GENE et l'ANRT pour le financement de ma thèse. Merci également à l'unité GABI et l'équipe G2B pour leur accueil.

Je remercie également les membres de mon comité de thèse pour leurs conseils, encouragements, et d'avoir aidé à structurer mon sujet de thèse. Je tiens à remercier les membres du jury qui ont relu et/ou évalué mon travail lors de la soutenance pour leur patience et bienveillance malgré les conditions particulières de soutenance de cette année 2020.

Didier, ta bienveillance légendaire a plus d'une fois été salvatrice lors de cette thèse. Ta patience pour expliquer et ré-expliquer encore et toujours les notions qui me paraissaient obscures a été d'une grande aide. Sébastien F merci d'avoir su trouver le dosage idéal entre accompagnement et autonomie tout au long de ces trois années, d'avoir toujours été une oreille attentive. J'ai toujours trouvé ta porte ouverte quand j'en avais besoin (entre deux réunions, évidemment).

Merci à tous ceux qui ont été là pour répondre à mes questions, donner un coup de main, relire du code, mettre à disposition les données dont j'avais besoin. Aurélia et Chris merci d'avoir plus que contribué à faciliter le basculement de ma thèse vers de la quanti, merci pour votre patience et gentillesse. Merci également à la pléthore de personnes qui m'ont dépanné sur du awk, pearl, bcftools, R ou wombat : Mathieu, Maria, Eric, Iola, Marie, Seb T et Julie.

Merci Cécile et Marie-Christine pour vos explications et votre aide très précieuse aux manips ! Merci également à Julie et Marthe à la plateforme d'histologie.

Je remercie également les entreprises de sélection Umotest et EvaJura de leurs contributions aux données de fertilité mâle, et à Eli de m'avoir aidé à bien appréhender les questions liées à ce type de données.

I would also like to thank Ivica, Maulik, Elisabeth and Kim for their kindness, support and patience and for sharing their knowledge on phylogenetic studies.

Un grand merci aux autochtones du bâtiment 211 pour la bonne ambiance qui y règne, et particulièrement aux rares habitants du premier étage.

Un immense merci plein de paillettes à ma co-bureau licorne Iola, merci pour tous les bouleversements que tu as apporté dans ma vie. Merci pour l'organisation et la rigueur que tu t'es efforcée d'insuffler tout au long de ma thèse, pour les discussions sans fin, les échanges culturels et féministes, mais aussi de m'avoir ouvert la porte vers un monde nouveau peuplé de gens aussi fous que créatifs. Merci roxane-toi-et-moi-barbotons-des-provisions-et-ramons-jusqu'en-Espagne pour le partage de deux grandes passions : la nourriture et les voyages. Merci pour le badminton, pour toutes ces expressions de qualitey, et le soutien moral à chaque fois que j'étais marais de Guérande. Merci aux autres compagnonnes de chocolat chaud : Sophie, Sophie et Julie de rajouter du sucre dans ma vie et préparer ma reconversion en fournisseuse de traineau de guépards. Merci Junior pour les milkshakes et le YOLO et merci Sénior pour ta mèche, ta passion pour le rhum et ce zombicide promis.

Merci à Marie et Seb, mes compagnons de galère, c'était très chouette d'avancer de front tous ensemble. Marie, merci pour les discussions jardinage, saucisson, couture et autres conseils de vie, tous ces partages et ce soutien ça aide vraiment à passer les moments difficiles. Merci Seb-le-champion pour Patrick Sébastien, la pétanque les « J'ai une thèse à faire » et « C'est toujours mieux que si c'était moins bien ». Petite pensée également pour ce voyage incroyable en Nouvelle-Zélande et Alexis, son borat et sa manie de la photo en paysage. Merci Gabriel d'avoir été le meilleur accueil de stagiaire possible et un sacré soutien féministe. Et enfin, bon courage aux thésardes bientôt « libres comme Dobby », Audrey, Anna-Charlotte, Laure-Hélène et Fanny. On finit toujours par s'en sortir.

Encore une fois, merci à tous ceux du 211 qui ne sont pas forcément cités ici mais qui ont participé à rendre ce cadre de travail plus chaleureux et agréable.

Merci aux vieux de Charles Peg', Lily pour tous nos débats et discussions autour de dilemmes moraux. Merci de m'avoir apporté de la culture, de la philosophie, des points de vus différents, de m'avoir posé plein de questions et ainsi aider à vulgariser mon travail et prendre du recul.

Merci également à mon kiné de m'avoir porté à bout de bras (littéralement) jusqu'au bout de cette rédaction. De même, merci à Kevin d'avoir développé cette capacité stupéfiante à convertir le chocolat en mots pendant la phase de rédaction.

Merci à ma famille, il n'y a pas vraiment de hasard si j'en suis là. Entre les histoires de Marguerite, ma mère et sa passion pour les bestioles à anomalies et mon père, sa passion pour la génétique bovine et ses explications sur la génomique lors des longs trajets de campagne. Merci également à ma sœur, qui sans comprendre mon travail sera toujours là pour voir le bon côté du verre du FIL à moitié plein.

La légende dit que chacun d'entre nous a une licorne invisible qui sommeille en lui. J'ai la chance d'avoir la mienne matérialisée à mes côtés alors merci à elle de m'avoir accompagnée et soutenue jour après jour, dans la joie comme dans la tourmente.

Table des matières

Liste des figures 6
Liste des tableaux 7
Liste des annexes 8
Liste des abréviations 9
Introduction générale 12
Chapitre 1 : Etat de l'art : Domestication et sélection, la place des chromosomes sexuels et du génome mitochondrial chez les bovins 14
1. De la domestication à la sélection génétique 14
1.1. Domestication du bovin 14
1.2. Histoire de la sélection 16
1.2.1. De la sélection méthodique à la sélection génétique 16
1.2.2. De la sélection génétique à la sélection génomique 18
1.3. Caractères sélectionnés en races laitières 20
1.4. Nécessité de surveiller l'évolution de la variabilité génétique et l'émergence d'anomalies génétiques 22
2. Outils et méthodes de la sélection génomique 24
2.1. Assemblage des génomes 24
2.2. Puces 25
2.3. Imputation et projet « 1000 génomes bovins » 25
2.4. QTL et cartographie génétique 26
2.5. Evaluation génétique et génomique 27
2.5.1. Principe de l'évaluation polygénique 27
2.5.2. Principe de l'évaluation génomique 29
3. Les spécificités des chromosomes sexuels et du génome mitochondrial 33
3.1. Origine et évolution des chromosomes sexuels 33
3.2. Le chromosome X, inactivation et compensation de dose 34
3.3. L'ADN mitochondrial 36
3.3.1. Mécanismes de transmission 36
3.3.2. Rôles physiologiques des mitochondries et pathologies associées à leur dysfonctionnement 38
3.4. Transmission uniparentale du génome mitochondrial et du chromosome Y 39
Conclusion : 40
Chapitre 2 : Diversité du génome mitochondrial et effets sur les caractères de production et reproduction 42

1. Etude 1 : Diversité et origine des haplogroupes mitochondriaux 44
1.1. Etat de l'art : génome mitochondrial et domestication des bovins 44
1.2. Matériels et méthodes 47
1.2.1. Analyse des pedigrees 47
1.2.2. Echantillonnage des individus séquencés 48
1.2.3. Amplification et purification de produits PCR 48
1.2.4. Séquençage du génome mitochondrial 49
1.2.5. Ajouts de séquence de génome mitochondrial dans les séquences complètes produites pour l'équipe G2B 50
1.2.6. Annotation des variants 50
1.2.7. Analyses phylogénétiques 50
1.3. Résultats et discussion 51
1.3.1. Les lignées dans la généalogie 51
1.3.2. Diversité mitochondriale des séquences 54
2. Etude 2 : Recherche de bovins modernes porteurs d'haplogroupes mitochondriaux d'aurochs et de zébu 58
2.1. Méthode 58
2.1.1. Détermination des haplotypes non-T / T et criblage de la population génotypée pour la sélection génomique 58
2.1.2. Amplification et séquençage du génome mitochondrial d'individus déterminés non-T et douteux 59
2.1.3. Analyses phylogénétiques 59
2.2. Résultats 61
2.2.1. Utilisation des génotypages pour identifier des lignées rares 61
2.2.2. Un premier séquençage réalisé sur la région D-loop 62
2.2.3. Exploitation de séquences complètes 64
2.3. Discussion 67
2.3.1. Présence de lignées P, Q et I dans la population de bovins taurins modernes 67
2.3.2. Introgression ou domestication secondaire ? 69
2.3.3. Perspectives 70
3. Etude 3 : Estimation des paramètres génétiques sur les performances laitières 72
3.1. Matériel et méthodes 72
3.1.1. Phénotypes 72
3.1.2. Animaux et lignées 73
3.1.3. Prise en compte d'un effet non-T / T en race Holstein 73
3.1.4. Modèle utilisé pour notre analyse 74
3.2. Résultats et discussion 75

4. Etude 4 : Estimation des paramètres génétiques sur la fertilité mâle 78
4.1. Enjeux liés à la fertilité mâle et rôle de la mitochondrie dans la motilité des spermatozoïdes 78
4.2. Matériel et méthodes 79
4.2.1. Phénotypes 79
4.2.2. Spécificités du jeu de données, filtres et définition des effets du modèle 81
4.2.3. Modèle 84
4.3. Résultats et discussion 85
4.3.1. Héritabilités 85
4.3.2. Part de variance expliquée par la lignée maternelle 86
4.3.3. Des phénotypes peu précis 86
4.3.4. Perspectives 87
5. Etude 5 : Contrôle de la généalogie et étude d'association 89
5.1. Méthode 89
5.1.1. Cohérence des lignées 89
5.1.2. Etude d'association 89
5.2. Résultats et discussion 92
5.2.1. Cohérence des lignées 92
5.2.2. Etude d'association 93
Chapitre 3 : Diversité génétique du chromosome Y et effet sur la fertilité mâle 98
1. Etude 1 : Analyse de la diversité du chromosome Y dans les races françaises 99
1.1. Méthodes 100
1.1.1. Analyse des pedigrees 100
1.1.2. Génotypes et marqueurs 100
1.1.3. Utilisation des données de séquences 101
1.2. Résultats et discussion 102
1.2.1. Analyse des pedigrees 102
1.2.2. Distribution des haplotypes dans la population 103
1.2.3. Exploitations des données de séquences en race Holstein 104
1.3. Discussion 107
2. Etude 2 : Effet du chromosome Y sur la fertilité mâle en race Montbéliarde 109
2.1. Méthodes 109
2.2. Résultats et discussion 110
Conclusion et perspectives 112
Chapitre 4 : Etude du chromosome X et anomalies génétiques 113
1. Etat de l'art sur l'identification de polymorphismes et d'anomalies génétiques 113

1.1. Définitions des anomalies génétiques 113
1.2. Méthodes d'identification des anomalies génétiques 114
1.3. Enjeux et identification 117
2. Polymorphismes délétères et approche de génétique inverse 119
2.1. Préambule 119
2.2. Méthodes 121
2.2.1. Choix des variants 121
1) Exploitation de la base de données du projet 1000 génomes bovin et déterminisme génétique 121
2) Annotation et phénotype délétère 121
3) Choix des variants 122
2.2.2. *Design* des marqueurs et calcul des fréquences 122
Résultats et discussion 123
3. Cas concret d'une anomalie identifiée par génétique classique 127
3.1. Etude : A de novo 3.8-Mb inversion affecting the *EDA* and *XIST* genes in a heterozygous female calf with generalized hypohidrotic ectodermal dysplasia 128
3.2. Discussion 140
Chapitre 5 : Vers une meilleure prise en compte des chromosomes sexuels et du génome mitochondrial dans les applications de génétique 142
1. Limites des études 142
1.1. Phénotypes 142
1.2. Génotypage 143
1.3. Séquençage 144
1.4. Pedigree 145
2. Perspectives et suites 146
2.1. Effets sur les caractères de production 146
2.2. Génétique inverse 147
2.3. Phylogénétique 148
3. Maintien de la diversité génétique du chromosome Y et du génome mitochondrial 148
4. Prise en compte des chromosomes sexuels et du génome mitochondrial dans les évaluations génomiques 149
4.1. Adaptation technique des modèles d'évaluations 149
4.2. Pour aller plus loin, explorer les particularités du chromosome X 152
Conclusion générale 154
Liste des références bibliographiques : 162

Liste des figures

Figure 1 : Grand mâle auroch. Grotte de Lascaux © Unesco, Wikimedia Commons 14
Figure 2 : Estimation de la contribution de chaque allèle à une performance (Croué, 2017)..... 30
Figure 3 : Représentation du fonctionnement d'une évaluation génomique en deux étapes, adaptée de Croué (2017)..... 30
Figure 4 : CD indicatif pour la production laitière (h² élevé) (d'après Labatut et al., 2014; Croue, 2017) 31
Figure 5 *:* Mâle hémizygote (gauche) et femelle hétérozygote (droite) pour une mutation de coloration située sur le chromosome X (© photo MRC Harwell ; Guénet *et al.*, 2015) 35
Figure 6 : Illustration de la répartition aléatoire des mitochondries impliquant des degrés variables d'hétéroplasmie (Beinat, 2013)..... 37
Figure 7 : Représentation de la chaine respiratoire mitochondriale. 38
Figure 8 : Représentation schématique du dispositif petites-filles..... 41
Figure 9 : Chronologie du travail de thèse portant sur le génome mitochondrial..... 42
Figure 10 : Représentation des foyers de domestication et des principales routes migratoires des *Bos taurus* (rouge) et *Bos indicus* (bleu) incluant en vert le foyer d'un potentiel 3ème évènement de domestication (Pitt *et al.* 2019)..... 44
Figure 11 : Distribution géographique des haplogroupes principaux taurins et zébus. 45
Figure 12 : Représentation des différents clades phylogénétiques mitochondriaux et des âges de coalescence. D'après (Achilli et al., 2009 et 2010)..... 46
Figure 13 : Principe de la transmission uniparentale du génome mitochondrial 47
Figure 14 : Répartition du nombre de variants par individu 55
Figure 15 : Représentation en réseau des échantillons. 56
Figure 16: Arbre phylogénétique basé sur la méthode du maximum de vaisemblance..... 65
Figure 17: Réseau phylogénétique comprenant les individus non-T. 66
Figure 18 : Pedigree d'une lignée incohérente 74
Figure 19 : Schéma d'un spermatozoïde (https://www.bio-top.net/Schemas/Spermatozoide.gif) 79
Figure 20 : Description des données (jeu de données A)..... 82
Figure 21 : Description des données de fertilité mâle (jeu de données B). En orange la distribution du PVD et en bleu celle du PV 83
Figure 22 : Distribution géographique des haplogroupes du chromosome Y (Edwards et al., 2011)... 99
Figure 23 : Organisation phylogénétique des différentes lignées..... 105
Figure 24 : Représentation du pedigree de la LP FRUS01322381 et de leurs génotypes pour les variants D et E, présentés dans cet ordre. 106
Figure 25 : Haplotype du chromosome Y issus d'échantillons anciens selon l'étude de a) Götherström et al. (2005) et Bollongino et al. (2008). Figure issue de Bollongino et al. (2008) 108
Figure 26 : Représentation du principe de la cartographie par homozygotie 115
Figure 27 : Représentation d'un gène et de la localisation des variants (Ensembl, 2020) 116
Figure 28 : Représentation de la taille des chromosomes en mégabases (a) et le ratio du nombre de variants identifiés sur le projet 1000 génomes par rapport à la taille du chromosome (b). 119
Figure 29 : Ratio entre le nombre de variants et la taille de chaque chromosome - exemple chez trois espèces sur les 3 premiers autosomes et les chromosomes sexuels..... 120
Figure 30: Arbre décisionnel pour le choix des variants 121
Figure 31 : Représentation schématique des lignes de Blaschko (Jin *et al.*, 2016) d'après Molho-Pessach & Schaffer, 2011)..... 127
Figure 32 : Vache atteinte d'alopécie localisée (Murgiano et al, 2015)..... 128

Figure 33 : Interprétation des génotypes par le logiciel Genome Studio (Michot, P. 2017) (expression des signaux d'hybridation en coordonnées polaires) 143
Figure 34 : Représentation schématique de la matrice des génotypes sans (a) et avec (b) les marqueurs du chromosome X, avec • représentant m_{ij}: génotype à un marqueur j pour un animal i 150
Figure 35 : Fréquence de l'expression de l'allèle paternel chez des animaux sur la région spécifique du chromosome X (A) et le chromosome 5 (B) (Couldrey et al., 2017) 152

Liste des tableaux

Tableau 1 : Index élémentaires en race Montbéliarde (hors morphologie) 21
Tableau 2 : Indicateurs de variabilité génétique sur la population de six races bovines née entre 2014 et 2017 (d'après Institut de l'élevage, 2018)..... 22
Tableau 3 : Amorces pour l'amplification en PCR long range du génome mitochondrial 48
Tableau 4 : Description des structures des races et des lignées sur les animaux nés entre 2015 et 2019 hormis lorsque « toutes années » est spécifié. 52
Tableau 5 : Description des races des échantillons séquencés..... 54
Tableau 6 : Variants portés par individu 55
Tableau 7 : Marqueurs permettant de distinguer les génomes mitochondriaux appartenant au macro-haplogroupe T des autres haplogroupes. *Les positions sont relatives à la séquence AY526085.*..... 58
Tableau 8 : Nombre de génotypages par race sur les puces EuroGMD et EuroG10k v8 en avril 2019 58
Tableau 9 : Echantillons anciens disponibles dans les bases publiques utilisés pour la réalisation de l'arbre phylogénétique 60
Tableau 10 : Etude réalisées sur les animaux génotypés sur la puce LD v8 ou MD v1, dernière mise à jour en novembre 2019 61
Tableau 11 : Tableau répertoriant les mutations sur la région 1-158 et 15782-16338 du génome mitochondrial (sur D-loop). Les haplotypes I, P, Q et T sont représentés respectivement en vert, bleu, gris et orange. Les échantillons au format RACEn° sont ceux qui ont été choisis pour un séquençage mitochondrial complet dans la suite de l'étude. Le détail des groupes et des races des échantillons est donné en annexe 5. 63
Tableau 12 : Correspondance haplotype-échantillon 64
Tableau 13 : Statistiques descriptives des caractères laitiers dans les populations étudiées 72
Tableau 14 : Structure du pédigrée des données utilisées 73
Tableau 15 : Héritabilités estimées dans notre dispositif et écart aux héritabilités estimées par Boichard et Bonaïti (1987)..... 75
Tableau 16 : Estimation de l'effet des lignées T et non-T sur les caractères laitiers 76
Tableau 17 : Part de la variance phénotypique expliquée par la LM 76
Tableau 18 : Echelle de notation de la motilité massale de la semence (Cabannes, 2008)..... 80
Tableau 19 : Description des données et caractères 84
Tableau 20 : Composantes de la variance pour les 6 caractères de fertilité mâle étudiés..... 85
Tableau 21 : Présentation du jeu de données utilisé pour contrôler la cohérence des lignées 89
Tableau 22 : Description par génération du nombre d'individus avec et sans performances affiliés par voie maternelle aux animaux génotypés sur puce EuroGMD. La génération n0 représente les animaux génotypés, n1 leurs mère etc. et d1 les descendants de 1e génération de tous les individus n..... 90
Tableau 23 : Nombre d'individus et de marqueurs utilisés lors de l'étude d'association 91
Tableau 24 : Statistiques effectuées sur les lignées avec au moins 5 descendants typés par lignée ... 92

Tableau 25 : Vérification de parenté des animaux génotypés toutes races comprises (génotypés jusqu'en janvier 2020) 93
Tableau 26 : Estimation des héritabilités obtenues avec GCTA sur les performances corrigées des effets d'environnement 94
Tableau 27 : Présentation des variants filtrés sur leurs p-valeurs et fréquences 95
Tableau 28 : Conservation chez les espèces (*Bos taurus*, *Homo sapiens* (humain), *Mus musculus* (souris), *Equus caballus* (cheval), *Gallus gallus* (poule), *Felis catus* (chat), *Anolis carolinensis* (lézard), *Procavia capensis* (Daman du Cap), *Xenopus tropicalis* (grenouille). Les domaines des protéines ont été identifiées sur (https://www.unip.org) 96
Tableau 29 : Description des marqueurs du chromosome Y de la puce HD utilisés pour cette étude101
Tableau 30 : Description des structures des races et des lignées 102
Tableau 31 : Répartition des haplotypes 103
Tableau 32 : Nombre d'individus porteurs d'allèles alternatifs 104
Tableau 33 : Présentation des variants et la répartition des allèles parmi les 147 taureaux Holstein de notre base de données 105
Tableau 34 : Estimation des effets de chaque LP. 110
Tableau 35 : Estimation de l'effet de l'haplotype Y 111
Tableau 36 : Description des phénotypes de 4 anomalies représentatives 123
Tableau 37 : Répartition des fréquences alléliques des marqueurs chez les femelles et les mâles dans six grandes races. Le choix des variants est codé par « QTL » dans le cas des gènes identifiés par une approche de recherche de QTL de fertilité mâle par De Camargo (2015). Le type de phénotype est renseigné. Les variants « XIST » se situent dans ou autour du gène *XIST* tel qu'identifié chez l'humain. Les fréquences inférieures à 0,01 sont codées « > » tandis que les cases sont vides lorsque la fréquence est nulle. Les variants grisés sont ceux correspondant au tableau 36. 124

Liste des annexes

Annexe 1 : Conditions PCR pour l'amplification en long range du génome mitochondrial 155
Annexe 2 : Liste des séquences mitochondriales disponibles dans Genbank utilisées dans l'étude 1. 156
Annexe 3 : Amorces utilisées pour le séquençage Sanger de l'étude 2. Les amorces F et R proviennent de la table S3 en annexe de (Achilli et al., 2008), tandis que les amorces « t » ont été dessinées lors de cette étude pour combler les trous de séquençage. 157
Annexe 4 : Echantillons modernes disponibles dans les bases publiques utilisés pour la réalisation de l'arbre et du réseau phylogénétique de l'étude 2 158
Annexe 5 : Description des échantillons non-T. 159
Annexe 6 : Synthèse des héritabilités estimées dans la littérature sur des caractères de fertilité mâle. 160
Annexe 7 : Etude de la conservation intra-vertébrés des acides aminés affectés par des polymorphismes non-sens dans les protéines codées par le génome mitochondrial. 161

Liste des abréviations

ABO : race Abondance
ADN : Acide Désoxyribonucléique Nucléique
ADNmt : ADN mitochondrial
AUB : race Aubrac
AZF : *Azoospermia Factor*
BAQ : Blonde d'Aquitaine
BAZ : Bazadaise
BLAD : *Bovine Leukocyte Adhesion Deficiency*
BLUP : *Best Linear Unbiaised Predictor*
BP : *Before Present*
BPN : race Bretonne Pie-Noire
BRU : race Brune
CASA : *Computed-aided sperm analysis*
CD : Coefficient de Détermination
CDH : *Cholesterol Deficiency Haplotype*
CHA : race Charolaise
CIA : Centre d'Insémination Animale
CIFRE : Conventions Industrielles de Formation par la Recherche
CNAG : Commission Nationale d'Amélioration Génétique
CONC : Concentration en spermatozoïdes
CPS *:* Centre de Production de la Semence
CTIG *:* Centre de Traitement de l'Information Génétique
CVM : *Complex Vertebral Malformation*
DAZ : *Deleted in Azoospermia*
DEH : Dysplasie Ectodermique hypohidrotique
D-loop *: Displacement loop*
DYD : *Daughter Yield Deviation*
ECEL : contrôle de performances
EDA *: Ectodysplasin-A*
EDE *:* Etablissements Départementaux d'Elevage
EFFAB : *European Forum of Farm Animal Breeders*
ES *:* Entreprise de Sélection
FGE : France Génétique Elevage
GAS : race Gasconne
GBLUP : *Genomic Best Linear Unbiased Prediction*
GMP : Grand-père maternel
GPP : Grand-père paternel
GWAS : *Genome-Wide Association Study*
h^2 : héritabilité
HOL : race Holstein
HY : effet de l'haplotype du chromosome Y
IA : Insémination Animale
Idèle : Institut de l'Elevage
IFOAM : *International Federation of Organic Agriculture Movements*
INDEL *:* insertion/déletion
INRA ou **INRAE** : Institut National de la Recherche Agronomique et Environnementale
ISU : Index de Synthèse Unique
KIF1C : *Kinesin Family Member 1C*
LAIT : quantités de lait (kg)
LIM : race Limousine
LM : Lignée Maternelle
LP : Lignée Paternelle
Mb : Megabase
MG : Matière Grasse (kg)
MGI : *Mouse Genome Informatics*
MON : race Montbéliarde
MP : Matière Protéique (kg)
MTCP : Mitochondropathie
MRCA : Most Recent Common Ancestor
NCBI *: National Center for Biotechnology Information*
NMID : Note de motilité Individuelle
NMM : Note de motilité massale
NOR : race Normande
OMIA : *Online Mendelian Inheritance in Animals*
OMIM : *Online Mendelian Inheritance in Man*
ONAB : Observatoire National des Anomalies Bovines
OS *:* Organisme de Sélection
PAB : Frontière pseudo-autosomique
PAR : race Parthenaise
PAR : Région pseudo-autosomique
PCR : *Polymerase Chain Reaction*
PV : Pourcentage de spermatozoïdes vivants
QTL : *Quantitative Trait Loci*

RDP : race Rouge des Prés
RepVet : *Reproduction Veterinarians*
RES : Résistance à a décongélation
RFL : race Rouge Flamande
RNA-Seq : *RNA-Sequencing*
RRBS : *Reduced Representation Bisulfite Sequencing*
RZUE : Règlement Zootechnique Européen
SAL : race Salers
SIG *:* Système d'Information Génétique
SIM : race Simmental
SNP : *Single Nucleotide Polymorphism*
SOX3 : SRY-box 3
SRY : sex-determining region of Y chromosome
TAR : race Tarentaise
TB : Taux Butyreux (g/kg)
TEX11 : *Testis expressed 11*
TP : Taux Protéique (g/kg)
Ve!P : *Variant Effet Predictor*
VOL : volume d'éjaculat
VOS : race Vosgienne
XIC : Centre d'inactivation du X
XIST *: X-inactive specific transcript*
YD : *Yield Deviation*

Introduction générale

Afin de sélectionner les meilleurs reproducteurs pour procréer la génération suivante et ainsi augmenter le niveau génétique de son troupeau, un éleveur doit disposer d'une estimation fiable de la valeur génétique de chacun des individus pour les caractères qu'il souhaite améliorer. Les progrès technologiques ont permis de disposer d'outils telles les puces à ADN permettant de baliser l'ensemble du génome avec des marqueurs génétiques et d'étudier l'association entre segments chromosomiques et performances des individus. A partir de ces résultats, il est devenu possible de prédire le potentiel génétique d'un animal directement à partir de son ADN, et ce, dès la naissance, voire dès le stade embryonnaire. Chez le bovin, la mise en place de la sélection génomique s'est faite dès 2009 dans les trois races laitières nationales Holstein, Montbéliarde et Normande (Boichard *et al.*, 2012) et plus récemment dans les races allaitantes, Charolaise, Limousine et Blonde d'Aquitaine (Griffon *et al.*, 2017), puis dans les races régionales à effectifs plus limités Brune (Baur *et al.*, 2014), Abondance, Tarentaise et Vosgienne (Jonas, 2016).

La mise en place de l'évaluation génomique a révolutionné le monde de l'élevage bovin en remplaçant le processus long et coûteux du testage sur descendance qui consiste à estimer la valeur génétique d'un reproducteur à partir des performances de ses descendants. Mais aussi efficace soit-elle, cette méthode ne prend pas en compte l'effet du génome mitochondrial et des chromosomes sexuels. Or, ces derniers comportent de nombreux gènes impliqués dans le déterminisme de multiples caractères de production, reproduction et santé. Les mitochondries jouent par exemple un rôle clé dans le métabolisme énergétique, extrêmement sollicité chez la vache laitière, et la motilité des spermatozoïdes. Le chromosome X est le deuxième plus grand chromosome du génome bovin et comprend 1133 gènes. Il représente donc une source potentielle de variabilité génétique très importante, tous caractères confondus. Enfin, le chromosome Y porte des gènes intervenant dans la différenciation des testicules et la spermatogénèse (Bellott *et al.*, 2014) et les anomalies du chromosome Y sont souvent associées à de la sub-fertilité ou de l'infertilité chez le mâle (Colaco and Modi, 2018).

Ainsi leur non prise en compte contribue sûrement à une part de l'héritabilité manquante. Nous disposons aujourd'hui de tous les éléments nécessaires à l'étude de ces chromosomes dont l'intérêt est renforcé par le rééquilibrage entre les différentes voies de sélection (père-fils, père-fille, mère-fils, mère-fille) permis par l'évaluation génomique.

Avant de mesurer le potentiel des variants génétiques portés par les chromosomes sexuels et le génome mitochondrial, il faut tout d'abord étudier la variabilité, la diversité et la répartition des allèles au sein des races. Plus la variabilité est grande, plus la probabilité que des allèles favorables ou défavorables pour un caractère ségrégent dans la population est élevée. Le génome mitochondrial se transmet par voie maternelle uniquement tandis que le chromosome Y n'est transmis que de père en fils. L'absence de recombinaison et la transmission uniparentale ont fait du génome mitochondrial et de la partie spécifique au chromosome Y des outils intéressants pour retracer l'histoire phylogénétique d'une espèce ou d'un groupe d'espèces.

C'est afin d'étudier la variabilité et le potentiel de sélection des chromosomes sexuels et du génome mitochondrial que ce projet a été initié. Cette thèse vise à détecter chez les bovins, des polymorphismes génétiques au niveau du génome mitochondrial et des chromosomes sexuels mais également à caractériser leurs effets sur les caractères de production, de reproduction et de santé. Elle repose sur l'exploitation de diverses données (séquences de génomes complets, génotypes pour des puces à SNP, bases de données phénotypiques, annotation fonctionnelle des gènes, prédiction de l'effet des variants, examens cliniques *etc.*).

Ce manuscrit est organisé en 5 chapitres. Dans le premier chapitre, je présenterai le contexte de la sélection des bovins à travers l'histoire, les outils et méthodologies de la sélection génomique sur lesquels reposent les études que nous avons menées et enfin les spécificités des chromosomes sexuels et du génome mitochondrial par rapport aux autosomes. Le chapitre 2 rassemblera l'ensemble des études qui ont été menées sur le génome mitochondrial, des analyses phylogénétiques aux études d'association entre lignées maternelles ou variants génétiques et caractères de production. Le chapitre 3 est consacré à l'étude de la variabilité génétique au niveau du chromosome Y et de l'effet de lignées paternelles sur des caractères de fertilité mâle. La chapitre 4 est dédié aux recherches menées sur le chromosome X à travers le prisme des anomalies génétiques. Enfin, dans un cinquième et dernier chapitre, je propose une discussion autour des limites rencontrées au cours des différents projets, des perspectives de valorisation des données non publiées et des moyens nécessaires à l'exploitation des chromosomes sexuels et du génome mitochondrial dans les applications génétiques.

Chapitre 1 : Etat de l'art : Domestication et sélection, la place des chromosomes sexuels et du génome mitochondrial chez les bovins

1. De la domestication à la sélection génétique

1.1. Domestication du bovin

Les espèces du genre *Bos*, comprenant les bovins taurins, les zébus, les gaurs et gayals, les bantengs, les yaks et les bisons, ont connu au moins quatre épisodes de domestication et constituent une source importante de viande, de lait et de force de travail pour de nombreuses cultures humaines.

L'auroch (*Bos primigenius*) appelé aussi us ou urus, est l'ancêtre des taurins (*Bos taurus*) et zébus (*Bos indicus*) domestiques. Il y a deux millions d'années, sa zone géographique s'étendait en Eurasie de l'océan Pacifique à la côte Atlantique et descendait en Afrique du Nord comme l'attestent de nombreux restes archéologiques et les peintures rupestres du Paléolithique (3 millions d'années-12 000 ans) (Gauthier A., 1990). Ces restes archéologiques et peintures rupestres constituent des sources précieuses permettant de reconstituer la morphologie de l'auroch. Ainsi, les ossements retrouvés montrent que sa hauteur au garrot à la fin du Paléolithique était supérieure à 170 cm, contre 120 en moyenne pour les bovins domestiques aujourd'hui (Guintard and Neron de Surgy, 2014; Bouwman *et al.*, 2018).

Figure 1 : Grand mâle auroch. Grotte de Lascaux © Unesco, Wikimedia Commons

La domestication de l'auroch a commencé il y a près de 11 000 ans au Proche-Orient (Troy *et al.*, 2001; Edwards *et al.*, 2007b). L'étude de la diversité génétique des bovins modernes suggère que les taurins actuels descendent d'une population initiale de seulement 80 femelles aurochs sauvages environ (Bollongino *et al.*, 2012; Scheu *et al.*, 2015a). Près de 2000 ans après ce premier évènement, a lieu un deuxième évènement de domestication dans la vallée de l'Indus qui donnera naissance aux zébus à partir d'une autre population d'aurochs (Loftus *et al.*, 1994). Ces événements de domestication pourraient avoir été favorisés par un phénomène climatique ayant provoqué une diminution drastique du nombre d'animaux sauvages à chasser, et ayant poussé les populations humaines vers l'élevage et la gestion des ressources animales. Les bovins représentaient un réservoir alimentaire vivant ne nécessitant pas de conservation et offrant d'autres ressources exploitables comme le cuir, les cornes ou encore les os (Bökönyi, 1976).

Plusieurs millénaires après le début de leur domestication, les formes sauvages et domestiques de la même espèce présentent généralement de nombreuses différences morphologiques. Ainsi, lorsque le naturaliste Louis-Henri Bojanus décrit l'auroch en 1827 pour la première fois, il le distingue immédiatement du bœuf domestique *Bos taurus* (Edel, 2003). Dans les premiers temps de la

domestication, l'humain exerce une sélection dite « inconsciente » et le fait d'être dans un environnement contrôlé et maitrisé par l'humain enclenche un processus de transformation de l'espèce (Darwin, 1859). Les individus les plus massifs et agressifs, donc dangereux pour l'homme, sont généralement éliminés, ainsi que ceux trop sensibles aux maladies ou ne supportant pas le stress lié à la captivité. C'est donc des animaux avec une même typologie qui seront amenés à se reproduire et transmettre ainsi ces caractéristiques.

La domestication animale est un processus cumulatif à long terme qui est caractérisé par des changements progressifs dans le comportement des populations humaines et animales au fil du temps, et qui a eu des avantages sélectifs pour ces deux partenaires biologiques. Deux étapes se succèdent généralement lors de la domestication : une première étape pendant laquelle les humains séparent quelques individus des populations sauvages ; puis une intervention active dans le cycle de vie des populations captives aboutissant à un changement comportemental et physiologique. Il ne s'agit pas d'un événement instantané transformant immédiatement un animal sauvage en domestique mais plutôt d'un processus exigeant un élevage soutenu d'animaux captifs et apprivoisés sur de longues périodes de temps. Le passage des formes sauvages aux formes domestiques se produit après plusieurs générations, ces générations étant sélectionnées indirectement sur des caractères de docilité et de production (Loftus *et al.*, 1999).

Pendant les milliers d'années qui ont suivi leur domestication, les bovins ont été sélectionnés par l'humain sur diverses aptitudes comme la production de lait et de viande, la docilité ou encore la force de travail, à mesure qu'ils migraient hors de leurs zones d'origine. En Europe, l'arrivée de la culture néolithique et donc des bovins domestiques s'est faite il y a 8000 ans à partir du Proche-Orient, et selon deux voies de migrations principales : le long des rives du Danube et des côtes de la Méditerranée (Olalde *et al.*, 2015).

Les déforestations massives ayant pour but d'accroître les surfaces agricoles exploitables et la chasse, ont conduit peu à peu à l'extinction de l'auroch. Il représentait en effet un trophée convoité, chassé notamment pour ses cornes qui servaient d'ornements (Bro-Jørgensen *et al.*, 2018). L'auroch disparut peu à peu des forêts européennes, pour ne subsister plus qu'en Pologne dans une réserve forestière. Le dernier spécimen connu, y mourut en 1627. Cependant, dans les années 1920, Lutz et Heinz Heck, directeurs des zoos de Berlin et Munich initièrent un programme de reconstitution de l'auroch à partir de races peu sélectionnées et présentant certains caractères primitifs comme la Highland, Grise des Steppes, Corse, ou des races de combat. Implanté en France pour la première fois en 1979, l'auroch-reconstitué est aujourd'hui une race officiellement reconnue dans notre pays (code race 30) (Davrou, 2008; Guintard et Neron de Surgy, 2014). Depuis les années 2000, des projets analogues visant à sélectionner certains critères spécifiques comme la rusticité, la taille, la couleur de la robe, la forme et la taille des cornes ont été initiés pour obtenir un animal ressemblant le plus possible à l'auroch mais restant malgré tout proche d'un animal domestique. L'objectif de ces projets va de la production d'animaux rustiques pour l'entretien de zones difficiles au « ré-ensauvagement » de l'Europe (Rewilding Europe; True Nature Fondation, 2013).

Si tous les *Bos taurus* modernes descendent d'une même population d'aurochs, ils ont connu par la suite une diversification régionale en fonction des critères de sélection locaux conduisant à différentes populations initiales qui forment le socle ayant permis la création des races françaises d'aujourd'hui.

1.2. Histoire de la sélection

1.2.1. De la sélection méthodique à la sélection génétique

Dans l'Antiquité, le développement des systèmes agraires à jachère et culture attelée avec l'utilisation de l'araire, outil muni d'un soc permettant de fendre la terre sans la retourner, va s'accompagner d'une utilisation croissante de la traction animale (Mazoyer, 1977). C'est à la même époque que les premières théories sur l'hérédité seront formulées par Aristote et Hippocrate (VI et Vèmes siècles av JC). Hérédité et milieu ne sont pas encore des notions distinctes, mais l'idée que les caractères phénotypiques des parents exprimés lors de l'accouplement sont transmis à la descendance est énoncée. Cette idée va influencer le choix des reproducteurs et conduire les éleveurs à améliorer leur maîtrise de l'alimentation afin que les reproducteurs soient au meilleur de leur état lors de l'accouplement (Gautier, 2003). Le choix des reproducteurs et le suivi de leurs caractères phénotypiques indiquent une transition depuis une sélection inconsciente vers une sélection méthodique.

Au moyen-âge, l'amélioration des outils agricoles et principalement l'attelage vont conduire à la mise en place d'un système dit agraire à jachère et culture attelée lourde (Mazoyer, 1977) car la charrue qui remplace l'araire, nécessite une force de traction plus élevée.

A partir du XVIème siècle, cette vision du bétail considéré principalement comme moyen de traction évolue peu à peu. Commence alors une spécialisation des populations par fonction productrice (lait, viande) qui facilite une différenciation des populations locales.

La révolution agricole initiée en Angleterre qui peut se caractériser par l'ensemble des innovations agricoles des XVIIIème et XIXème siècles, va faire évoluer la conception de l'élevage bovin. En effet, l'utilisation à grande échelle d'engrais, l'amélioration de l'outillage agricole et le drainage des sols humides contribuent à améliorer la qualité et la quantité de l'alimentation animale. Le remplacement des bœufs par les chevaux pour le labour, joue un rôle majeur dans la recherche d'une meilleure valorisation des bovins. Particulièrement marquée en Angleterre, cette situation va engendrer la création d'un système novateur de sélection des troupeaux laitiers, ayant pour objectif à la fois de valoriser le lait mais aussi les taureaux produits en tant que reproducteurs (Bernard et Baudemont, 2016).

Robert Bakewell (1725-1795), éleveur anglais considéré comme le fondateur de la zootechnie moderne, établit le premier ensemble de principes :

- Valoriser au mieux le potentiel de l'animal en subvenant au plus juste à ses besoins, principalement sur le volet alimentaire ;
- Utiliser comme reproducteurs les plus beaux animaux, qu'on destinait préférentiellement à l'alimentation humaine ;
- Accoupler les animaux apparentés afin d'accumuler les qualités recherchées sous forme de lignées.

Deux des élèves de Robert Bakewell, les frères Colling, créent la race Durham à partir d'une race régionale de la vallée de la Tees. Ils sélectionnent les animaux de celle-ci sur la conformation idéale pour l'engraissement et le développement précoce (Bernard et Baudemont, 2016).

En 1840, des éleveurs français souhaitant rattraper leur retard sur les éleveurs anglais, commencent à croiser leurs cheptels avec des animaux de la race Durham. Cette tendance, soutenue d'abord par la presse agricole, décline ensuite dans les années 1860. En effet, les éleveurs commencent à relever le manque de rusticité des produits croisés, leur inaptitude au travail dans le cas des croisements avec des races de type laitière, une production laitière réduite, et une production de gras trop importante qui ne répond pas aux attentes du marché français. Cette étape de croisement permet néanmoins aux éleveurs de réaliser le potentiel de leurs souches locales ainsi que l'importance du rôle de l'environnement et de l'alimentation sur la croissance et la production de leurs animaux. La mise en place des premiers concours de boucherie et du concours universel agricole de Paris en 1856, fut l'élément qui enclencha le processus d'amélioration et de sélection des races locales (Gautier, 2003).

En effet, les théories scientifiques se développent en parallèle et conduisent à une nouvelle conception du choix des reproducteurs ainsi que de nouvelles méthodes d'étude de l'hérédité des caractères.

Publiées en 1866 et redécouvertes trente ans plus tard, les lois de Mendel sur la transmission des caractères héréditaires vont permettre de bâtir les fondations de la génétique moderne. August F.L. Weismann sera le premier en 1883 à distinguer les cellules reproductrices, dites « germinales », des cellules qui composent le reste de notre corps, dites « somatiques », en séparant les caractères acquis au cours de la vie et ceux transmis à la descendance (Gouyon *et al.*, 1997). Puis, les expériences de Johnnsen au début du XXème siècle sur une variété de haricots, comparant des populations sélectionnées et non sélectionnées pour un caractère donné, aboutiront à la distinction du phénotype et du génotype (Johannsen, 2014). Enfin, en 1910, Thomas Morgan décrira chez la drosophile les corrélations qu'il observe entre la transmission des chromosomes sexuels et de certains caractères comme la couleur de l'abdomen.

Toutes ces découvertes permettront de mieux définir la notion de caractères héritables et les mécanismes régulant la transmission des gènes. La notion d'héritabilité devient de plus en plus importante dans la conception de la sélection par les éleveurs et avive leur intérêt pour l'enregistrement des généalogies. Le premier HerdBook français est mis en place en race Durham en 1855. Il est suivi en 1862 par celui de la Charolaise-Nivernaise puis par de nombreux autres HerdBook qui marquent la création officielle des races françaises. Ce sont alors des associations d'éleveurs qui définissent selon un standard l'appartenance des individus à une race et répertorient l'ensemble des filiations. De nombreuses races locales déclarées à cette époque ont par la suite été absorbées par de plus grandes races, comme par exemple la Gâtinaise, la Poitevine et la majeure partie des Maraichines et Nantaises qui ont été intégrées à la Parthenaise dans les années 1960 (Bernard et Baudemont, 2016).

En 1966, la loi sur l'élevage ou loi Poly réorganise la sélection animale et l'amélioration génétique des productions animales. Elle permet la mise en place de contrôles des performances, la création d'une base de données pour identifier les individus et enregistrer leurs pedigrees, régule l'insémination artificielle, structure la profession autour de schémas de sélection et charge l'INRA de la création de la base nationale et du calcul des index (France Agricole, 2011).

1.2.2. De la sélection génétique à la sélection génomique

Le principe de la sélection moderne repose sur la notion de valeur génétique d'un reproducteur. La valeur génétique est, par définition, deux fois ce que le reproducteur transmet en espérance à ses descendants. Elle ne s'observe pas et doit donc être estimée. Son estimation peut être très précise, par exemple lorsque le reproducteur a un grand nombre de descendants. Le modèle liant la valeur génétique entre individus apparentés suppose que le nombre de gènes impliqués dans le déterminisme d'un caractère est très grand et que chaque gène a un petit effet. On parle de modèle polygénique. Dans ces conditions, le modèle est le suivant :

$$u_i = 0{,}5\ u_p + 0{,}5\ u_m + \phi_i$$

Autrement dit, la valeur u_i d'un individu est égale à la valeur moyenne des parents p et m plus un aléa de méiose, d'espérance nulle, indépendant de la valeur des parents, et de variance égale à la moitié de la variance génétique (en l'absence de consanguinité des parents). Ce modèle implique que la covariance entre valeurs génétiques est proportionnelle à l'apparentement entre les individus.

Les phénotypes dépendent à la fois de la génétique et du milieu. Dans sa forme la plus simple, le modèle de description des données s'écrit : $y = M + u + e$, avec y le phénotype, M l'ensemble des effets du milieu identifiés, u la valeur génétique et e un résidu aléatoire. L'intégration du modèle de description des données et du modèle génétique détermine le modèle dit « polygénique ».

Dans ce modèle, la valeur génétique d'un individu est estimée à partir de ses performances propres et de celles de ses apparentés, avec des pondérations en fonction de leur degré d'apparentement avec l'individu évalué et de l'héritabilité du caractère. Sans généalogie, seule la performance propre peut contribuer à l'évaluation de la valeur génétique : on parle d'évaluation massale. Pour tenir compte des apparentés, il faut connaitre les pedigrees. Les pedigrees des animaux de chacune des races élevées en France sont enregistrés dans la base de données nationale.

Cette approche permet de prédire des valeurs génétiques pour chaque animal en distinguant l'effet de l'environnement (élevage, âge, saison, mode de traite, etc.) de l'effet génétique. L'estimation simultanée de ces différents effets repose sur une méthode appelée *Best Linear Unbiased Predictor* ou BLUP décrite dans la partie 2.5.1.

L'évaluation polygénique repose principalement sur le fait de disposer de phénotypes pour des individus apparentés, et en particulier, des descendants. Pour les caractères ne pouvant être mesurés sur l'animal évalué comme la production laitière pour un taureau, l'enregistrement des performances des descendants est primordial. Ce cycle est donc long et coûteux puisqu'il nécessite d'attendre la maturité sexuelle du taureau, la gestation et la maturité sexuelle de ses descendants, avant de pouvoir l'évaluer, éventuellement le sélectionner et le faire reproduire à plus grande échelle. Par conséquent, un nombre limité de taureaux pouvait bénéficier de ce « testage sur descendance » et, de plus, seule une partie de ces animaux se révélaient améliorateurs. Les meilleurs taureaux du classement, les taureaux dits « stars » étaient alors largement diffusés, induisant un fort progrès génétique mais provoquant également une baisse de la diversité génétique de la population (Agriproduction, 2011).

Si les résultats des évaluations polygéniques sont toujours diffusés aujourd'hui, une autre méthode d'évaluation a vu le jour en 2009 : l'évaluation génomique. Les progrès technologiques ont permis de disposer d'outils telles les puces à ADN (cf. partie 2.2.) permettant de baliser l'ensemble du génome avec des marqueurs génétiques et d'étudier l'association entre segments chromosomiques et

performances des individus. A partir de ces résultats, il est devenu possible de prédire le potentiel génétique d'un animal directement à partir de son ADN, et ce, dès la naissance, voire dès le stade embryonnaire. Les animaux sans performances et sans descendant peuvent ainsi être évalués grâce aux informations fournies par la connaissance de leur génome (Meuwissen *et al.*, 2001; Hayes *et al.*, 2009). Les principes des évaluations polygéniques et génomiques seront abordés plus en détail dans la partie 2.5.

La mise en place de la sélection génomique s'est faite dès 2009 dans les trois races laitières nationales Holstein, Montbéliarde et Normande (Boichard *et al.*, 2012) et plus récemment dans les races allaitantes, Charolaise, Limousine et Blonde d'Aquitaine (Griffon *et al.*, 2017) et dans les races régionales à effectifs plus limités Brune (Baur *et al.*, 2014), Abondance, Tarentaise et Vosgienne (Jonas, 2016).

Avant novembre 2018, le cadre des évaluations génétiques était défini par deux lois françaises : la Loi sur l'élevage de 1966 puis la Loi de modernisation agricole de 2006. Dans ce cadre, différents organismes sont agréés pour remplir des missions nécessaires à la sélection :

- Les Organismes de Sélection (OS) qui définissent les choix stratégiques des programmes de sélection dont les objectifs de sélection, tiennent les livres généalogiques (Herdbook), assurent la promotion de la race et l'ingénierie de l'enregistrement de la morphologie. Ils rassemblent et interagissent avec les associations d'éleveurs, les entreprises de sélection (ES), le Centre de Production de Semence (CPS) et les organismes de contrôle de performances
- Les Etablissements Départementaux d'Elevage (EDE) qui sont chargés de l'identification des animaux
- Les organismes de contrôle de performances
- Les centres de mise en place des semences d'insémination
- L'INRA assure la gestion de la base nationale d'information génétique et le calcul des valeurs génétiques. Pour cela, il dispose du Centre de Traitement de l'Information Génétique (CTIG) s'appuyant sur le Système d'Information Génétique (SIG).

Jusqu'en 2006, ces organismes avaient un monopole (sur une zone géographique ou au niveau national). Depuis 2006, l'insémination et, dans une moindre mesure, le contrôle de performances sont entrées dans le monde concurrentiel. Les ES sont, depuis leur origine, dans le monde concurrentiel mais intimement liées aux centres d'insémination animale. Le système est cogéré par le Ministère de l'Agriculture et la filière, avec des concertations organisées au sein de la CNAG (Commission Nationale d'Amélioration Génétique). En 2006, un organisme sous statut d'interprofession, France Génétique Elevage (FGE) est créé pour assurer une coordination et une gouvernance du côté professionnel. L'INRA était alors en charge du calcul des index polygéniques et génomiques.

Depuis 2006, l'INRA accueille une Unité Mixte Technologique de génétique bovine qui réalise le calcul et la diffusion des évaluations génétiques et développe des recherches sur l'amélioration des évaluations génétiques et génomiques. Trois organismes sont représentés au sein de l'UMT eBis : Idele (l'institut de l'élevage), Allice (union de coopératives d'élevage visant à mutualiser les moyens autour de la recherche en génétique et technologies de la reproduction) et INRAE.

Le Règlement Zootechnique Européen (RZUE) du 8 juin 2016 a modifié le cadre structurant la filière de l'amélioration des animaux. Il s'appuie sur l'évolution des OS, qui deviennent les organismes en charge du pilotage des programmes de sélection. Les nouveaux OS ont des missions très larges et sont

responsables des missions de certification des parentés et de la tenue du livre, du contrôle des performances, de la réalisation et de la diffusion des évaluations génétiques. En conséquence, l'INRAE n'est plus responsable des évaluations. Cependant les acteurs de la filière ont choisi de garder un système mutualisé, conduisant à la création par Races de France et Allice d'un nouvel acteur : GenEval. En effet, les OS français, anciens ou nouveaux, n'ont pas les moyens de réaliser eux-mêmes les missions dont ils ont la responsabilité et contractualisent donc avec les opérateurs historiques pour le contrôle de performances (ECEL), la certification de filiation (EDE), et la diffusion (Idele), en plus du nouvel acteur, GenEval, chargé de l'indexation. Depuis novembre 2018, date de mise en application de ce règlement, GenEval est donc en charge du calcul des index. INRAE (et les acteurs de l'UMT eBis) et GenEval continuent de travailler en collaboration sur l'amélioration des méthodes d'évaluation, et la mise en place d'évaluations de nouveaux caractères.

De nombreux acteurs participent donc aux processus de l'évaluation génétique, en remontant de nouveaux besoins d'évaluations, définissant de nouveaux objectifs de sélection, fournissant phénotypes et génotypes ou encore en améliorant les méthodes de calcul des évaluations génétiques et génomiques.

1.3. Caractères sélectionnés en races laitières

Les valeurs génétiques d'une population sont exprimées relativement au niveau génétique moyen d'une population. Ainsi, un animal avec un index pour la production laitière de + 800 kg produira en espérance, 800 kg de lait de plus sur une lactation que la population de référence.

Le tableau ci-dessous présente certains des index élémentaires [1] calculés aujourd'hui en race Montbéliarde. Les caractères évalués n'ont pas tous le même intérêt pour l'éleveur et la filière. C'est pour cela que chaque race définit ses objectifs de sélection, c'est-à-dire le poids relatif à accorder à chacun des caractères. Un index de synthèse unique (ISU) combine les index élémentaires avec leurs pondérations respectives et reflète la valeur génétique globale d'un animal selon les critères propres à la race.

Historiquement, les premiers caractères évalués en races laitières concernaient la production de lait puis la composition du lait. Ensuite, ont été développées des évaluations portant sur la morphologie et d'autres caractères fonctionnels, c'est-à-dire des aptitudes d'élevage (fertilité, résistance aux maladies, longévité, etc.). La plupart des caractères évalués ont pour but de produire des animaux compétitifs et adaptés à un marché donné afin de maximiser le revenu de l'éleveur mais aussi de répondre aux exigences et attentes des clients incluant les consommateurs et aussi les distributeurs. Les objectifs de sélection et les index sont donc amenés à évoluer pour satisfaire les besoins croissants.

Au défi de produire de manière durable des animaux conformes aux besoins de la filière, s'ajoutent désormais de nombreuses questions éthiques autour du bien-être animal, et une attente sur la qualité de la production en lien avec la santé humaine. L'élevage fait l'objet de nombreuses critiques qui reposent sur son impact sur l'environnement, à travers les émissions de gaz à effet de serre (notamment le méthane), l'utilisation de ressources alimentaires pouvant être dédiées à l'alimentation humaine ou encore la surexploitation des terres agricoles (Delanoue et Roguet, 2015).

[1] un index élémentaire correspond à un caractère donné

Si la sélection animale ne peut pas répondre à elle seule à ces problématiques, il est néanmoins nécessaire de tout mettre en œuvre pour contribuer à répondre aux attentes sociétales et aux attentes des éleveurs et de la filière. Ainsi, ces dernières années de nouveaux axes de recherche se focalisent sur la mise en place d'évaluations pour des nouveaux caractères comme la santé de la mamelle, des aplombs, l'efficacité alimentaire ou encore les émissions de méthane (Chesnais *et al.*, 2016; Cottle et van der Werf, 2017; Croué *et al.*, 2017).

	Caractère évalué	**Index**	**Unité de l'index**
Production laitière (par lactation)	Quantité de lait produit (en kg)	LAIT	Unité phénotypique
	Quantité de matières protéiques produites (en kg)	MP	Unité phénotypique
	Quantité de matières grasses produites (en kg)	MG	Unité phénotypique
	Taux protéique du lait (en g/kg)	MP	Unité phénotypique
	Taux butyreux du lait (en g/kg)	MG	Unité phénotypique
Mam-mites	Score cellulaire du lait (leucocytes en unité de score)	CEL	Ecart-type génétique
	Sensibilité aux mammites cliniques	MACL	Ecart-type génétique
Fertilité femelle	Taux de conception des vaches (en %)	FERV	Ecart-type génétique
	Taux de conception des génisses (en %)	FERG	Ecart-type génétique
	Intervalle vêlage - première insémination (en jours)	IVIA1	Ecart-type génétique
	Longévité fonctionnelle (risque de réforme pour une autre cause que la production laitière)	LGF	Ecart-type génétique
Conditions de naissance	Facilité du veau à naître (% de naissances faciles chez les génisses)	NAI	Unité phénotypique
	Facilité de la femelle à vêler (% de vêlages faciles chez les génisses)	VEL	Unité phénotypique
	Vitalité du veau à la naissance (1-mortinatalité (en %) du veau né de l'animal évalué)	VIN	Unité phénotypique
	Vitalité du veau au vêlage (1-mortinatalité (en %) du veau né des filles de l'animal évalué)	VIV	Unité phénotypique
Aptitudes bouchères	Aptitudes bouchères des veaux (abattage entre 70 et 250 jours)	IABvbf	Ecart-type génétique
	Aptitudes bouchères des jeunes bovins (abattage entre 350 et 730 jours)	IABjbf	Ecart-type génétique

Tableau 1 : Index élémentaires en race Montbéliarde (hors morphologie)
(Bérodier, 2019; GenEval, 2019)

Pour conclure, les objectifs de la sélection sont donc multiples : satisfaire les besoins des éleveurs (productivité, résistance et longévité des animaux), qui sont l'un des principaux acteurs de la sélection, et répondre aux demandes du citoyen (environnement, bien-être, respect de la tradition) et du consommateur (qualités organoleptiques et prix) tout en maintenant la diversité génétique.

1.4. Nécessité de surveiller l'évolution de la variabilité génétique et l'émergence d'anomalies génétiques

Grâce à l'amélioration des méthodes pour identifier les meilleurs taureaux en estimant leur valeur génétique, les programmes de sélection sont devenus de plus en plus efficaces. Cependant, avec l'utilisation croissante de l'insémination animale (IA) et l'augmentation de l'intensité de sélection, la diversité génétique des races a fortement diminué. La variabilité génétique se définit comme la diversité des allèles pour un gène donné et de la combinaison de ces allèles à l'échelle d'un chromosome. Le maintien d'un haut niveau de variabilité génétique est le moteur des gains génétiques futurs et une garantie d'adaptabilité en cas d'évolution de l'objectif de sélection. Des indicateurs ont donc été mis en place pour surveiller l'évolution de cette diversité et l'augmentation de la consanguinité à partir des informations généalogiques.

Le tableau ci-dessous (tableau 2) propose un descriptif de la population de six races bovines à travers quelques indicateurs de diversité publiés chaque année par l'observatoire de la variabilité génétique des ruminants et des équidés, VARUME (Danchin-Burge, 2015). Le nombre efficace d'ancêtres (Ae) est un indicateur qui correspond au nombre d'ancêtres non apparentés qui, en ayant tous des contributions égales, engendreraient une population du même niveau de variabilité génétique que la population de référence. Le nombre efficace d'ancêtres reflète le niveau d'apparentement des individus d'une population.

La consanguinité estime la probabilité pour que les deux allèles d'un locus pris au hasard soient identiques par descendance, c'est-à-dire soient la même copie d'un allèle d'un même ancêtre. Un individu est donc consanguin si ses deux parents sont apparentés, c'est à-dire s'ils ont en commun un ou des ancêtres dans leur arbre généalogique (Leroy, 2015). On l'estime généralement à partir du pedigree que l'on suppose connu sur un nombre suffisant de générations. Ce tableau montre que les trois grandes races laitières sont composées d'animaux beaucoup plus apparentés que dans les trois races allaitantes. Malgré une population de très grande taille en termes de nombre de femelles, les races laitières possèdent un nombre efficace d'ancêtres très faible. La différence s'explique par une plus grande utilisation de l'IA dans les races laitières, avec des taureaux « stars » ayant eu un nombre impressionnant de descendants comme Jocko Besné, un taureau Holstein ayant eu entre 300 000 et 400 000 filles (Huet, 2013).

Race	Nombre de femelles	Nombre de mâles d'IA	Ancêtres efficaces	Consanguinité moyenne
Limousine	536 683	422	147	0,9%
Charolaise	784 711	760	171	0,8%
Blonde d'Aquitaine	328 448	353	38	1,5%
Montbéliarde	849 355	1 201	18	4,8%
Normande	411 604	804	21	5,0%
Holstein	3 522 108	5 069	20	4,4%

Tableau 2 : Indicateurs de variabilité génétique sur la population de six races bovines née entre 2014 et 2017 (d'après Institut de l'élevage, 2018)

Une augmentation de consanguinité se traduit par une réduction de l'hétérozygotie. Le contrôle de l'augmentation de la consanguinité permet donc de prévenir l'apparition des anomalies génétiques récessives (qui s'expriment à l'état homozygote). En effet, statistiquement, chaque individu est porteur de plusieurs polymorphismes délétères à l'état hétérozygote. Or les races bovines laitières sont des populations génétiquement petites, c'est à dire avec une taille efficace de 50 individus génétiquement différents (Danchin-Burge *et al.*, 2012). En sélectionnant un nombre restreint de reproducteurs et en les diffusant trop largement, les races bovines laitières ont connu une augmentation de la consanguinité favorisant une augmentation de l'homozygotie et donc l'émergence d'anomalies génétiques récessives.

Selon l'importance d'une anomalie par rapport aux autres caractères, la gestion des reproducteurs du programme de sélection peut permettre d'en maîtriser la fréquence. Jusqu'à récemment, le nombre d'anomalies connues et leur fréquence allélique étant faibles, la pratique était d'éliminer tous les reproducteurs porteurs d'une anomalie afin de l'éradiquer progressivement et d'éviter les accouplements à risque. Cependant, éliminer de façon immédiate l'ensemble des porteurs peut amener à restreindre le choix des reproducteurs et ainsi diminuer la diversité génétique. Il est désormais préconisé de diminuer progressivement et de manière raisonnée le nombre des porteurs au sein du groupe de reproducteurs proposés par le schéma, et de prendre en compte le statut des animaux dans la gestion des accouplements (Boichard *et al.*, 2016b). Le statut des reproducteurs pour certaines anomalies peut également être inclus dans les index de synthèse, en fonction de l'impact économique de l'anomalie, sa fréquence et du type de transmission (Hoze *et al.*, 2018).

De façon générale, les principales races font régulièrement face à des émergences d'anomalies. L'émergence d'anomalies récessives est inévitable car celles-ci résultent de mutations aléatoires dans la population. Cependant, plus elles sont repérées et prises en charge tôt, plus il est possible d'en limiter les conséquences. C'est dans cette optique qu'a été créé l'Observatoire National des Anomalies Bovines (www.onab.fr, Grohs *et al.*, 2016) sous l'instance du Ministère de l'Agriculture, conçu pour collecter et traiter les signalements de cas supposés d'anomalies génétiques, observés dans les élevages bovins français. L'ONAB rassemble différents acteurs : INRAE, Idele, les Groupements Techniques Vétérinaires, les Ecoles Vétérinaires, Allice, les OS, Races de France et le Contrôle de Performance. En pratique, les déclarations de cas d'anomalies sont effectuées par les centres d'IA, vétérinaires, éleveurs, techniciens d'organismes de sélection, et techniciens de contrôle de performance.

En résumé :

- La domestication des animaux de ferme comme le bovin a joué un rôle majeur dans le développement des sociétés humaines.
- Des populations animales dotées de spécificités se sont développées au niveau local, ce qui a constitué une base d'individus plus ou moins homogènes pour constituer les races.
- Les moyens de sélection ont évolué dans le temps et notamment avec la découverte des bases de l'hérédité des caractères, jusqu'à aboutir à une sélection fondée directement sur les informations contenues dans le génome de l'animal.
- Aujourd'hui les animaux sont sélectionnés sur de nombreux caractères et non plus uniquement sur la production.
- Cette sélection doit s'accompagner d'une surveillance continue pour détecter l'apparition des anomalies génétiques et contrôler l'augmentation de la consanguinité.

2. Outils et méthodes de la sélection génomique

Pour pouvoir sélectionner efficacement les meilleurs reproducteurs mais aussi étudier le déterminisme génétique des différents phénotypes, les outils et les méthodes d'évaluation génétique ont connu de nombreuses évolutions ces dernières années.

2.1. Assemblage des génomes

Pour analyser les informations génétiques portées par des individus distincts de races différentes, il est nécessaire de travailler avec une séquence de référence. Cette séquence type permet de définir la position des gènes et des variants sur le génome. Ainsi, chaque génome étudié sera comparé à cette référence, permettant d'uniformiser les positions des allèles et de standardiser les résultats d'analyses génétiques des différents laboratoires de recherche en génétique.

Avoir une séquence de référence permet également d'accélérer les procédés mathématiques permettant l'alignement du génome d'un nouvel individu. En effet, lors du séquençage, ce sont des fragments d'ADN de petite taille qui sont amplifiés, appelés « *read* » ou lectures. Celles-ci sont ensuite alignées informatiquement sur le génome de référence, permettant de recréer la séquence entière de l'individu étudié.

Afin de réaliser l'assemblage des génomes *de novo*, il a fallu développer des outils bioinformatiques en parallèle des technologies de séquençage. Le tout premier assemblage du génome d'un bovin (une femelle de race Hereford) a été réalisé par le *Baylor College of Medicine* en 2006. Plusieurs assemblages du même animal ont été publiés ces dernières années, afin de corriger des erreurs d'assemblage et améliorer la couverture de celui-ci en complétant les régions non couvertes. La dernière version de cet assemblage a été publiée au cours de ma thèse, en avril 2018 (ARS-UCD1.2, GenBank accession NKLS00000000.2). L'assemblage utilisé jusqu'à cette date était l'UMD3.1, la troisième version de l'assemblage publié par Zimin *et al.* en 2009.

Chaque nouvelle mise à jour de l'assemblage du génome bovin a des répercussions sur l'utilisation des données de génotypage et s'accompagne d'une période de transition pour remettre à jour certains outils. En effet, chaque assemblage redéfinit la position des gènes et variants, ce qui change la position des marqueurs génétiques. Un nouvel assemblage permet bien souvent de corriger les erreurs de la version précédente et de réordonner des segments chromosomiques qui étaient jusqu'alors mal placés.

Chaque génome est unique. En effet, à chaque méiose, plusieurs dizaines de neo-mutations se produisent aléatoirement. Plusieurs dizaines de millions de variations sont répertoriées au total et deux animaux de la même race présentent généralement plusieurs millions de différences dans leur génome. En étudiant les séquences ou fragments de séquence génomique, des polymorphismes, ou variants génétiques qui différencient les individus entre eux ont été identifiés ; ceux-ci peuvent être utilisés comme des marqueurs ou balises du génome. L'assemblage du premier génome bovin a permis de déterminer la position physique des marqueurs sur le génome et de placer les séquences partielles déjà disponibles. Ces informations ont contribué à l'élaboration de puces à ADN permettant de baliser le génome d'un animal à moindre coût par rapport au séquençage complet. A titre de comparaison, un génotypage de moyenne densité coûte environ 20 euros par échantillon contre 500 à 1000 euros pour le séquençage du génome entier.

2.2. Puces

Depuis 2008, des puces à ADN avec des marqueurs de type SNP (*single nucleotide polymorphism*) sont utilisées pour les cartographies génétiques et la sélection génomique. Les polymorphismes SNP sont des variations d'un seul nucléotide, généralement bi-alléliques et très fréquentes sur le génome. Les SNP choisis pour être inclus sur une puce sont informatifs, grâce à leur répartition dense sur le génome et la fréquence de chacun des allèles (Jussiau *et al.*, 2013). Ces puces sont utilisées pour réaliser le génotypage d'un individu, ce qui permet de connaitre une partie de la séquence d'un individu à partir d'un échantillon ADN.

En bovin, Il existe des puces avec différentes densités de SNP, BovineSNP50 (54 000 SNP), BovineHD (High density ; 778 000 SNP), BovineLD (Low density, 7 000 SNP) et BovineMD (Medium Density, 45 000 SNP) mises sur le marché par la société Illumina (Illumina Inc, San Diego, USA). D'autres puces existent comme celles proposées par la société Affymetrix mais ne sont pas utilisées en France pour le génotypage bovin.

En France jusqu'à l'année dernière, la puce principalement utilisée était la puce EuroG10k produite par Illumina, une puce LD customisée qui possède en plus de la partie commune de 6909 marqueurs utilisés pour l'indexation, une partie propre à chaque pays dédiée à des marqueurs dits « de recherche », principalement utilisée pour la validation de mutations causales par les laboratoires de recherche (Boichard *et al.*, 2018). Elle a été mise en place par le consortium EuroGenomics composé de huit pays européens : France, Allemagne, Pays-Bas, Danemark, Finlande, Suède, Espagne et Pologne. L'objectif initial d'EuroGenomics a été d'accroître la précision des index génomiques par la mise en commun de données de génotypage 50K et de phénotypes de taureaux indexés sur descendance pour constituer une base de référence pour la sélection génomique (Lund *et al.*, 2011). Cependant, EuroGenomics a également suscité différentes innovations et a joué le rôle de centrale d'achat de puces. Le faible coût de la puce EuroG10k a favorisé l'accessibilité de cette technologie aux éleveurs et donc aux index génomiques tout en permettant la validation des mutations causales et autres variants d'intérêt grâce à la partie dédiée aux marqueurs de « recherche ». Depuis avril 2019, la puce EuroG10k a été remplacée à l'échelle du consortium Eurogenomics par l'EuroGMD, dont la partie commune compte dorénavant 45 000 marqueurs.

Dans le cadre des travaux de cette thèse, nous avons exploité des génotypages provenant de puces EuroG10k, EuroGMD et BovineHD.

2.3. Imputation et projet « 1000 génomes bovins »

Malgré l'informativité des puces utilisées en routine, il est parfois nécessaire de travailler avec une densité plus élevée de marqueurs sur le génome. L'imputation est un processus mathématique qui permet de reconstituer des génotypes manquants et qui repose sur la connaissance des haplotypes[2] existant au sein d'une population. Elle permet de prédire les informations de marqueurs moléculaires manquants en analysant les génotypes obtenus aux SNP voisins. Grâce à la diversité génétique limitée des races bovines, et à la disponibilité des séquences des génomes des principaux fondateurs de

[2] Série ordonnée d'allèles à des marqueurs ou *loci* adjacents portés par un même segment chromosomique

chaque race, il est même possible, par imputation, de reconstruire la séquence entière du génome d'un animal à partir d'un génotypage sur puce LD (Daetwyler *et al.*, 2014).

C'est l'un des objectifs du projet international « 1000 génomes bovins » lancé en 2002 afin de mutualiser les données de séquences complètes générées par différentes équipes de recherche dans le monde (Daetwyler *et al.*, 2014; Bouwman *et al.*, 2018). Cette immense base de données contient aujourd'hui toutes les variations des séquences de plus de 3100 taureaux et vaches, représentant près de 60 races. Le nombre d'échantillons de cette population permet d'améliorer la précision des imputations. Depuis sa création, plusieurs versions ou « runs » se sont succédés, le septième et dernier datant de juin 2019.

Cette base de données d'animaux phénotypiquement sains permet également de filtrer avec une grande efficacité les génomes d'animaux atteints d'anomalies dominantes causées par des mutations *de novo* (Bourneuf *et al.*, 2017).

Avec 103 000 000 variants de type SNP ou indels (insertion/délétion de quelques nucléotides) répertoriés sur le dernier « run », cette liste représente la grande majorité des variants présentes dans la population à une fréquence non négligeable. Cependant, les variants structuraux, tels que les insertions/délétions de grande taille, inversions, remaniements chromosomiques ou duplications ne sont pas renseignés sur ce « run » faute de procédure fiable et standardisée pour les détecter en routine.

2.4. QTL et cartographie génétique

Connaître à la fois le phénotype et le génotype d'une population dite de référence suffisamment large permet d'étudier les segments chromosomiques influençant les caractères quantitatifs. Ces segments sont appelés « *quantitative trait locus* » ou locus à effets quantitatifs (QTL).

L'analyse d'association appelée aussi « *genome-wide association study* » (GWAS) est la méthode la plus communément utilisée pour détecter les QTL. Les cartographies exploitent le déséquilibre de liaison c'est-à-dire l'association préférentielle entre allèles situés à des loci différents sur un chromosome (Robert-Granié *et al.*, 2011). Grâce aux nombreux marqueurs répartis le long du génome, il y a un déséquilibre de liaison suffisant entre ceux-ci et les variants ayant un effet sur le caractère étudié. Il est donc possible d'associer l'allèle d'un marqueur à l'allèle d'un variant d'intérêt. De manière générale, plus deux loci sont proches, plus le déséquilibre de liaison entre les allèles est élevé (Le Roy, 2015).

Grâce aux études de cartographie génétique, les premiers QTL de production laitière ont été identifiés dès 1994 chez les bovins. L'identification de QTL permet d'établir sur quels segments chromosomiques et allèles il faut maximiser les efforts de sélection pour améliorer la génétique de la race étudiée.

2.5. Evaluation génétique et génomique

Afin d'améliorer le niveau génétique d'un troupeau et être efficace dans le choix des reproducteurs pour produire la prochaine génération, un éleveur ou un sélectionneur doit disposer d'une estimation de la valeur génétique de chacun des individus pour chacun des caractères qu'il souhaite améliorer. Cela présuppose cependant qu'il s'agisse de caractères héritables, c'est-à-dire qui puissent être transmis aux descendants.

2.5.1. Principe de l'évaluation polygénique

L'évaluation génétique dite « polygénique » permet d'estimer pour chaque animal une valeur génétique (sa valeur intrinsèque, indépendante de l'environnement, qu'il transmettra en espérance pour moitié à ses descendants) et de le classer parmi l'ensemble de la population. Cette évaluation donne les outils permettant de réaliser une sélection génétique, c'est-à-dire de sélectionner les meilleurs reproducteurs et les accoupler. La maîtrise des accouplements, en favorisant ceux des animaux avec la plus haute valeur génétique pour le ou les caractères d'intérêt, permet d'augmenter le potentiel génétique moyen de la génération suivante.

Pour estimer cette valeur génétique, il faut distinguer le rôle du milieu sur l'expression du potentiel de l'animal, de la part uniquement due à son matériel génétique. Par exemple, une année très chaude a des conséquences sur la production laitière d'un animal et cette baisse de performances n'est pas le fait d'un mauvais potentiel génétique. De même, la conduite du troupeau, le type de bâtiment ou d'élevage, et l'âge sont autant d'effets qui influencent les performances d'un animal. Ces effets doivent être pris en compte pour estimer aussi précisément que possible la valeur génétique d'un animal.

Le modèle utilisé pour l'évaluation polygénique décompose la performance en un effet génétique, des effets de l'environnement et un effet résiduel ou erreur du modèle. Cette estimation simultanée des effets se traduit par l'absence de biais dans les estimations, d'où le nom de BLUP (Henderson, 1984). Pour des caractères dont les effets aléatoires sont normalement distribués, cette méthode est la plus précise, d'où l'attribut « *Best* ». Le fait d'estimer les effets génétiques et de milieu simultanément permet, sous réserve de connexion statistique suffisante du dispositif, de comparer les valeurs génétiques d'animaux dans des conditions différentes (différents élevages, différentes années, différentes régions, etc.).

Comme indiqué précédemment, le modèle polygénique suppose que le déterminisme génétique d'un caractère est soumis à un grand nombre de gènes, chacun ayant un petit effet. Du fait de la loi des grands nombres, la distribution des valeurs génétiques est normale. Avec ce modèle, un parent transmet en moyenne la moitié de sa valeur génétique à ses descendants mais ce qui est réellement transmis résulte d'un tirage aléatoire des chromosomes et des recombinaisons. Cet aléa de méiose, qui explique la différence entre pleins frères-sœurs, est indépendant de la valeur du parent, il a une espérance nulle et une variance σ_ϕ^2 égale à $0{,}5\ \sigma^2\ (1 - 0{,}5\ F_p - 0{,}5\ F_m)$. En l'absence de consanguinité des parents, la variance de l'aléa de méiose représente donc la moitié de la variance génétique. Lorsque les parents sont consanguins, ils sont plus homozygotes qu'attendu en panmixie, et leur aléa de méiose est réduit. La formule [1] montre le lien entre valeurs génétiques du produit et des parents. Cette formule induit que la covariance entre valeurs génétiques d'apparentés est proportionnelle à leur coefficient de parenté.

Avec un modèle de type BLUP, on suppose donc que les performances, les effets de milieu les affectant, et les généalogies sont connus. On suppose également que les effets aléatoires ont une distribution normale de variance connue (variances génétique et résiduelle, notamment).

Le modèle s'exprime sous la forme :

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{e}$$

avec $\mathbf{y}$ le vecteur des performances des individus évalués ; $\boldsymbol{\beta}$ le vecteur des effets d'environnement ; $\mathbf{u}$ le vecteur des valeurs génétiques ; $\mathbf{X}$ et $\mathbf{Z}$ les matrices d'incidence reliant chaque performance aux effets fixes et valeurs génétiques associées : $\mathbf{e}$ le vecteur des erreurs résiduelles, soit l'écart entre la valeur prédite par le modèle et la valeur mesurée.

Les valeurs génétiques étant estimées, elles peuvent avoir un degré de précision variable. La fiabilité de l'estimation génétique est représentée par un indicateur : le coefficient de détermination (CD). Celui-ci est compris entre 0, non fiable, et 1, identique à la valeur vraie. Le CD est la part de variance génétique expliquée par le modèle : $\mathrm{CD} = 1 - \frac{\mathrm{var}(\hat{u}\,|u)}{\mathrm{var}(u)}$. Le CD dépend de trois paramètres : l'héritabilité du caractère étudié, ainsi que la qualité et la quantité des informations utilisées pour calculer la valeur génétique d'un individu.

L'héritabilité (h^2) se définit comme la part du phénotype d'origine génétique, et se calcule comme ci-dessous :

$$h^2 = \frac{\sigma u^2}{\sigma p^2},$$

avec σ_A^2 la variance génétique du caractère considéré et σ_P^2 sa variance phénotypique.
L'héritabilité se rapproche de 1 lorsque la part du phénotype expliquée par le génotype augmente. De manière générale, on considère comme peu héritable un caractère avec une héritabilité inférieure à 0,2 et fortement héritable un caractère avec une héritabilité supérieure 0,4.

La précision de la valeur génétique d'un animal augmente également avec le nombre de ses performances connues (performances propres ou de ses apparentés). Elle dépend aussi du degré d'apparentement entre cet animal et ses apparentés avec performances. Si les seules performances à disposition sont celles de ses parents, alors l'aléa de méiose ne peut pas être estimé et le CD est borné à 0,5, valeur obtenue quand le CD des deux parents est égal à 1. Seule la connaissance d'un grand nombre de performances propres à l'animal ou de ses descendants permet d'estimer avec précision la valeur du matériel génétique qu'un animal a reçu de ses parents. Plus le nombre de ses descendants avec performances est grand et plus il a de performances propres, plus la précision de l'estimation de sa valeur génétique augmente.

L'estimation de la valeur génétique d'un individu est donc amenée à évoluer lorsque de nouvelles performances pour lui-même ou ses apparentés sont connues. C'est pour cela que des évaluations génétiques sont réalisées régulièrement, afin de réactualiser cette estimation en fonction des nouvelles informations disponibles.

Il est important de bien définir les effets d'environnement pouvant influer sur l'expression du génotype. La non prise en compte d'un effet de milieu peut, au mieux, limiter la précision de l'évaluation, ou sinon induire un biais dû à l'association statistique entre valeur génétique et effet de

milieu. D'un autre côté, un effet inutile pourrait réduire la précision des estimations. Un effet d'environnement peut être considéré comme fixe ou aléatoire dans le modèle. Un effet est dit fixe si son estimation ne dépend que des données. Un effet est aléatoire si son estimation combine l'information des données et l'information a priori sur sa distribution. Un effet aléatoire gagne en précision compte tenu de l'information a priori mais est généralement biaisé. Pour éviter les biais, les effets de milieu sont généralement considérés comme fixes. Par construction, les effets génétiques sont aléatoires puisqu'on prend en compte l'information de leur distribution (héritabilité, covariance entre individus). Il y a deux catégories parmi les effets fixes : les covariables dont on peut traduire l'effet sur un caractère par une fonction et les facteurs qui contiennent des classes (Croué, 2017).

Si l'évaluation polygénique permet d'estimer de façon très fiable la valeur génétique d'un individu avec un grand nombre de descendants, elle ne permet pas de prendre en compte l'aléa de méiose pour un individu sans performances. Cela signifie que pour pouvoir estimer précisément quel reproducteur représente le plus d'intérêt pour la filière au sein d'un lot d'individus sur le critère de la production laitière, il faut attendre que tous les taureaux de ce lot aient un nombre suffisant de filles, que ces filles aient un premier veau et terminé leur première lactation. L'utilisation des puces à ADN et l'amélioration de nos connaissances sur le génome bovin ont permis de mettre en place une nouvelle méthode d'évaluation : l'évaluation génomique.

2.5.2. Principe de l'évaluation génomique

Le principe de l'évaluation génomique repose sur l'utilisation de marqueurs moléculaires polymorphes qui permettent de suivre les transmissions alléliques entre parent et produit. Les génotypes des marqueurs sont déterminés grâce à des puces à ADN, qui permettent de connaître précisément les allèles portés par un individu. Du fait du déséquilibre de liaison entre chaque marqueur et le segment chromosomique qui l'entoure, le marqueur a un effet apparent qui capte et résume la valeur génétique du segment. Par ailleurs, pour les QTL bien caractérisés, il est possible d'associer un statut sur marqueurs à un QTL donné (cf. partie 2.4.).

L'évaluation génomique repose sur les mêmes principes que les méthodes d'évaluation polygénique : à partir de mesures phénotypiques, on estime la valeur génétique d'un animal, même s'il n'a pas de performances. L'évaluation génomique utilise directement l'information des marqueurs pour identifier si l'individu possède des allèles favorables pour les caractères d'intérêt.
Pour pouvoir mettre en place la sélection génomique, il est nécessaire dans un premier temps de génotyper et phénotyper précisément une population d'individus (population de référence), afin d'estimer statistiquement la contribution de chaque allèle de chaque marqueur aux performances mesurées, au sein de cette population. Une fois que chaque allèle est associé à une variation de performance, par exemple +2 kg de lait pour l'allèle *A* du marqueur 1, alors l'estimation de la valeur génétique se fait en sommant les effets de l'ensemble des allèles portés par un individu (cf. figure 2).

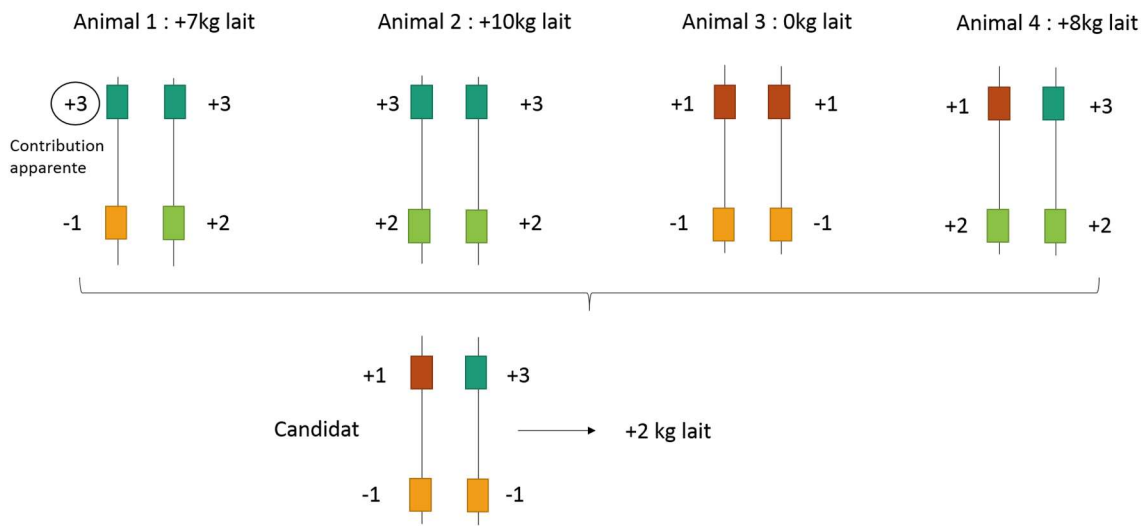


Figure 2 : Estimation de la contribution de chaque allèle à une performance (Croué, 2017)

Les liens entre marqueurs et variants causaux de QTL pouvant être cassés par des recombinaisons, il est essentiel de renouveler régulièrement la population de référence afin qu'elle soit la plus représentative possible des animaux que l'on souhaite évaluer.

Des équations de prédiction permettent ensuite de déduire la valeur génétique de n'importe quel animal, à partir de son génotype, même sans performances. Les performances utilisées pour l'évaluation génomique (sauf dans le cas de l'évaluation Single-Step présentée par la suite) doivent être déjà corrigées des effets d'environnement (Robert-Granié *et al.*, 2011). Une évaluation polygénique est donc réalisée en amont et permet de générer ces performances corrigées, appelées Yield Deviation (YD) pour les individus génotypés ayant des performances propres. Pour les taureaux, des performances sont calculées en faisant une moyenne pondérée des performances corrigées de leurs filles non génotypées, corrigées pour la contribution génétique des mères (*Daughter Yield Deviation* ou DYD).

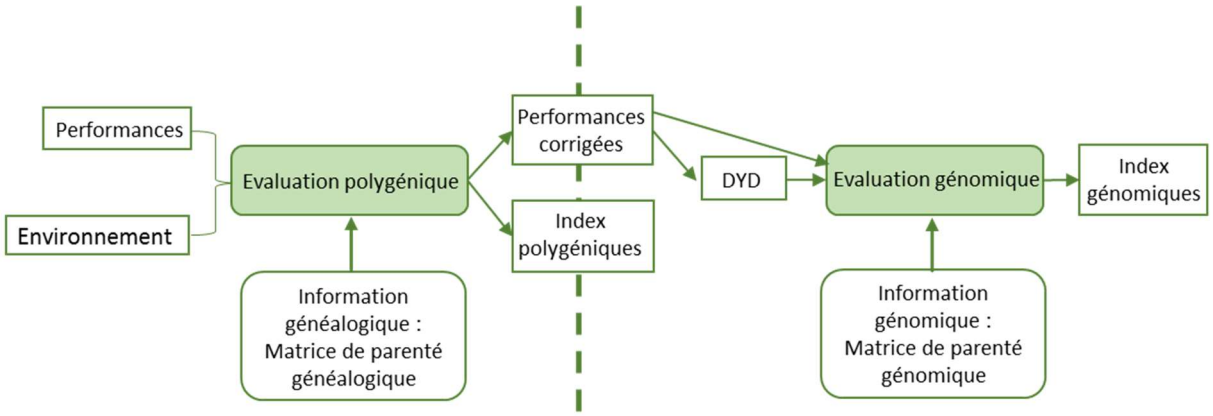


Figure 3 : Représentation du fonctionnement d'une évaluation génomique en deux étapes, adaptée de Croué (2017)

Le GBLUP ou *Genomic Best Linear Unbiased Prediction*, utilise le modèle du BLUP en y apportant des informations génomiques. Ce modèle d'évaluation nécessite une première étape polygénique pour corriger les performances des effets de l'environnement avant la phase génomique (cf. figure 3). L'évaluation génomique ne peut être réalisée que pour les animaux génotypés, les animaux non-génotypés ne descendant pas directement d'animaux génotypés ne sont pas pris en compte. Cela peut entrainer un biais, car les individus sur lesquels les calculs de l'évaluation sont réalisés ne sont pas représentatifs de l'ensemble de la population. En effet, les animaux génotypés ne sont pas choisis aléatoirement dans la population, il s'agit généralement des produits d'accouplement d'animaux de haute valeur génétique (Patry et Ducrocq, 2011).

Une méthode en une étape dérivant du BLUP et appelée single-step GBLUP a été développée pour pouvoir réaliser ces deux étapes en même temps et prendre en compte à la fois les animaux génotypés et non génotypés (Misztal *et al.*, 2009; Christensen et Lund, 2010). Le principe consiste à effectuer une évaluation similaire au BLUP, en intégrant l'information génomique. Une transition vers un modèle single-step est prévue pour les évaluations génomiques françaises à court-terme.

La sélection génomique permet de s'affranchir des multiples contraintes liées aux schémas de sélection sur descendance, tels les longs et coûteux protocoles de contrôle des performances et testage sur descendance pour chacun des reproducteurs mâles entrant dans le schéma de sélection. Les taureaux sont indexés et sélectionnés très tôt et peuvent ainsi être mis à la reproduction dès qu'ils ont atteint la maturité sexuelle, vers 1 an et demi, ce qui réduit l'intervalle de génération. De plus, il a été montré que la précision des index d'individus sans descendant est améliorée en moyenne de 40% en évaluation génomique par rapport à une évaluation polygénique, pour la production laitière (figure 4, Croue, 2017). Alors que la précision de l'index sur ascendance est bornée à 50%, la précision des index génomiques est supérieure car il rend possible la prédiction de l'aléa de méiose à partir des marqueurs.

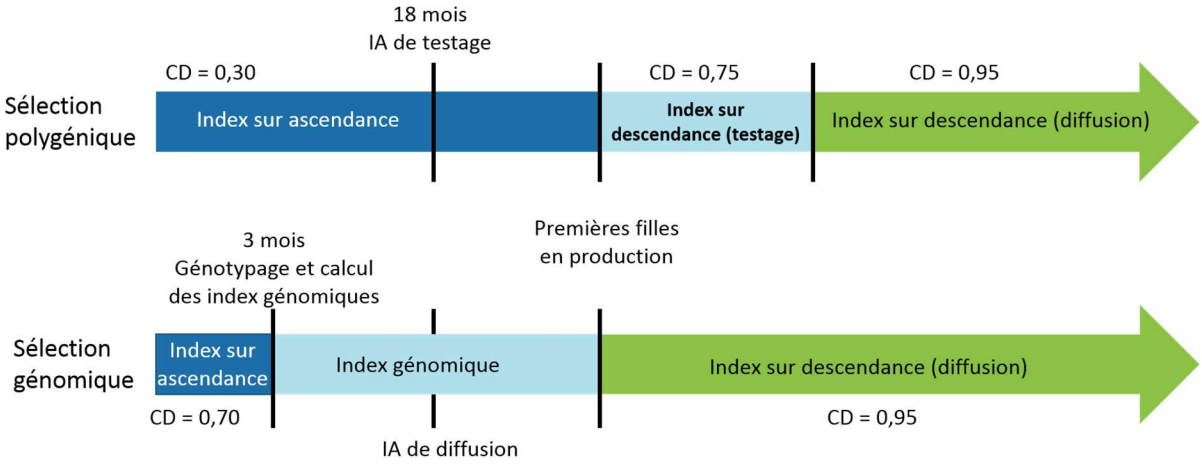


Figure 4 : CD indicatif pour la production laitière (h^2 élevé) (d'après Labatut *et al.*, 2014; Croue, 2017)

En permettant d'identifier des taureaux de haute valeur génétique directement dans les fermes, sans passer par une étape de testage en station, l'évaluation génomique permet de sélectionner des animaux dans l'ensemble de la population et pas uniquement de quelques lignées choisies, et donc d'augmenter la diversité des profils génétiques proposés (France Agricole, 2011). L'augmentation de l'intensité de sélection des reproducteurs et la réduction de l'intervalle de génération permettent une augmentation importante du progrès génétique annuel (Schaeffer, 2006; Pryce *et al.*, 2010; Lillehammer *et al.*, 2011). Ainsi, grâce à l'utilisation de l'évaluation génomique chez les mâles en races Holstein, Montbéliarde et Normande, le progrès génétique annuel a augmenté de 33, 50 et 71%, respectivement chez ces races (Doublet *et al.*, 2019, calcul effectué sur l'ISU).

En résumé :

- Les séquences sont étudiées comparativement à une séquence de référence commune dont l'assemblage est régulièrement actualisé.
- Des puces à SNP sont aujourd'hui utilisées en routine pour identifier des mutations présentant un intérêt pour la sélection.
- La plupart des caractères évalués en bovin dépendent d'un grand nombre de gènes ayant chacun de petits effets.
- La sélection polygénique estime la valeur génétique d'un animal en utilisant les performances de ses apparentés et les liens de parenté qui les relient.
- La sélection génomique utilise les informations des puces à SNP pour estimer la valeur génétique d'un animal en calculant l'effet de chacun des variants.

3. Les spécificités des chromosomes sexuels et du génome mitochondrial

L'amélioration des outils génétiques et de nos connaissances du génome bovin nous permettent désormais d'investiguer des parties du génome peu ou non exploitées en sélection génomique. Du fait de spécificités de transmission, les chromosomes sexuels et le génome mitochondrial ne sont pas pris en compte dans le calcul des évaluations génomiques françaises. L'objectif de ma thèse est d'étudier leur intérêt potentiel pour la sélection.

3.1. Origine et évolution des chromosomes sexuels

Les chromosomes sexuels ont co-évolué ensemble dans la longue histoire des mammifères et dérivent d'une paire ancestrale d'autosomes. La différenciation entre les chromosomes X et Y s'est faite il y environ 300 millions d'années, avec l'acquisition d'un nouveau gène, *SRY* (*sex-determining region of Y chromosome*) qui détermine le sexe, probablement suite à une mutation du gène *Sox3 (SRY-box 3*) (Stevanovlć *et al.*, 1993; Graves *et al.*, 2006). Le chromosome portant ce nouveau gène qui détermine le sexe d'un individu est ainsi devenu le proto-Y.

Les chromosomes (X et proto-Y) sont par la suite devenus suffisamment différents pour qu'une recombinaison entre ceux-ci soit en grande partie impossible. Seul(s) un ou deux petit(s) segment(s) recombinent et sont appelés région(s) pseudo-autosomique(s) (PAR). Il existe plusieurs hypothèses qui pourraient expliquer la sélection au cours de l'évolution contre une recombinaison des chromosomes sexuels :

- De nombreuses inversions identifiées sur le chromosome Y pourraient avoir joué un rôle dans la suppression des recombinaisons (Ross *et al.*, 2005).
- Il est possible que la recombinaison entre chromosomes sexuels portant chacun des gènes de fertilité mâle ou femelle puissent conduire à des phénotypes sexuels désavantageux (Charlesworth, 1996).
- Le lien fort entre *SRY* et d'autres gènes impliqués dans la détermination du sexe mâle pourrait avoir été fortement sélectionné entraînant une suppression progressive de la recombinaison entre les deux chromosomes ancestraux (Ellegren, 2011).

L'efficacité de la sélection naturelle est fortement réduite sur un chromosome ne pouvant pas recombiner (Rice et Chippindale, 2001). Les populations naturelles voient apparaître des mutations, dont certaines augmentent la survie ou le succès de reproduction et d'autres ont des conséquences négatives sur la valeur adaptative. La sélection naturelle favorise la fixation de mutations bénéfiques dans le génome d'une espèce, tandis que les mutations nuisibles sont contre-sélectionnées, ce qui permet et maintient l'adaptation. Sur un autosome, la sélection peut agir indépendamment à plusieurs loci grâce aux recombinaisons qui cassent le déséquilibre de liaison. Lorsque la recombinaison est impossible, la sélection opère sur le chromosome entier qui est sélectionné positivement et fixé dans la population après un certain nombre de générations s'il est porteur d'une mutation bénéfique, ou éliminé s'il porte une mutation délétère.

Du fait de l'absence de recombinaison, il y a eu une différenciation entre les deux chromosomes sexuels. Le chromosome Y est devenu vulnérable à l'accumulation de délétions délétères qui ne peuvent être éliminées ou réparées, on parle de « dégradation » ou « dégénérescence » (Ellegren, 2011; Bachtrog, 2014). Le chromosome X, au contraire du Y, peut encore recombiner lorsqu'il est présent en deux copies chez la femelle, il est donc protégé de la dégénérescence. Le chromosome X

actuel des mammifères conserve plus de 1000 gènes, alors que le chromosome Y a perdu 95% de son contenu génétique original (Skaletsky *et al.*, 2003).

La dégénérescence du chromosome sexuel non recombinant peut être expliquée par trois forces évolutives: un taux de mutation élevé, une sélection inefficace et une dérive génétique (Graves, 2006). Le chromosome Y humain aurait perdu 1393 de ses 1438 gènes originaux au cours de son évolution, et l'extrapolation linéaire de cette perte, résulte en un taux de perte génétique de 4 à 10 gènes par million d'années menant à la dégénerescence totale des gènes portés par le chromosome Y dans les 5 à 10 millions d'années à venir selon les estimations (Graves, 2006; Goto *et al.*, 2009). Cependant, les comparaisons des chromosomes Y de l'homme et du chimpanzé montrent que le chromosome Y de l'homme n'a perdu aucun gène depuis la divergence entre l'homme et le chimpanzé, il y a 6 à 7 millions d'années, Cela suggère que le modèle linéaire n'est pas applicable et que le chromosome Y humain actuel ne rétrécit plus ou rétrécit à un rythme beaucoup plus lent qu'estimé initialement (Hughes *et al.*, 2005). Pendant longtemps, le chromosome Y a été perçu comme un chromosome X profondément dégénéré ne codant que pour très peu de gènes et contenant une forte proportion d'ADN répété (Smith *et al.*, 1987). Les études suivantes ont plutôt démontré une divergence entre le contenu génétique des chromosomes X, Y et des autosomes d'origine. Les deux chromosomes sexuels ont recruté ou perdu de nombreux gènes (Bachtrog *et al.*, 2011). Ainsi, le chromosome X a été enrichi en gènes impliqués dans la détermination du sexe, la reproduction et la cognition (Vallender et Lahn, 2004). Par exemple, chez le bovin, des QTL de fertilité femelle ont été identifiés sur le chromosome X (Fortes *et al.*, 2013).

Sur le chromosome Y, les gènes présents sont principalement impliqués dans la détermination sexuelle et la spermatogenèse (Graves *et al.*, 2006). Cette différence entre les contenus génétiques de ces deux chromosomes est essentiellement due à la sélection liée au sexe (Bachtrog *et al.*, 2011). En effet, la transmission paternelle du chromosome Y rend les gènes spécifiques des mâles plus susceptibles de se fixer. Par exemple, le chromosome Y a acquis des gènes transposés depuis les autosomes tels que *DAZ* (*Deleted in Azoospermia*), un gène impliqué dans la spermatogenèse (Saxena *et al.*, 1996).

3.2. Le chromosome X, inactivation et compensation de dose

Comme nous venons de le voir, le chromosome X comporte un nombre de gènes largement supérieur à celui du chromosome Y. Avec ses 139 mégabases (Mb), il est le deuxième plus grand chromosome du génome bovin et comprend 1133 gènes (version 98 d'Ensembl).

A cause du déséquilibre entre le nombre de X présents chez les mâles et les femelles, le nombre de copies des gènes liés au chromosome X diffère entre les deux sexes qui, en absence de compensation, conduirait à un déséquilibre dans la quantité de produits géniques (ARN et protéines). Cela impliquerait l'existence d'adaptations complexes dans le contrôle métabolique et d'autres processus cellulaires. Des mécanismes de compensation de dose existent pour équilibrer la quantité de produits géniques liés à l'X entre les sexes. Trois types de mécanisme sont connus : (i) une double régulation à la hausse de l'expression des gènes liés à l'X chez les mâles (par exemple chez la drosophile et l'homme); (ii) une double régulation à la baisse des gènes sur chacun des deux chromosomes X chez les femelles (chez le ver *Caenorhabditis elegans*); et (iii) l'inactivation complète de l'un des deux chromosomes X chez les femelles (chez les mammifères) (Brockdorff et Turner, 2015).

Ainsi chez les mammifères femelles, un des deux chromosomes X est inactivé aléatoirement dans chaque cellule de l'embryon. Toutes les cellules filles inactiveront le même chromosome X que celui

inactivé dans la cellule souche embryonnaire de la lignée cellulaire. Chaque cellule somatique d'un individu femelle n'exprime que l'un des deux chromosomes X de son génome. En moyenne, 50% des cellules devraient exprimer chacun des deux chromosomes X. Dans certains cas cependant, des biais d'inactivation causés par une sélection positive ou négative de l'un des deux chromosomes X peuvent se produire (Renault *et al.*, 2007).

En 1961, en étudiant différentes lignées de souris, Mary Lyon observe que des souris femelles avec une monosomie sur le chromosome X sont viables et fertiles, et donc qu'un seul chromosome X suffit à assurer le bon développement de la femelle. Elle observe également que des souris hétérozygotes pour une mutation de coloration située sur le chromosome X sont bicolores, alors que les souris hétérozygotes pour un gène autosomique de pigmentation sont généralement unies (cf. figure 5). Elle formule alors l'hypothèse d'une inactivation irréversible et aléatoire d'un des deux chromosomes X à un stade embryonnaire précoce (Lyon, 1961).

Figure 5 *:* Mâle hémizygote (gauche) et femelle hétérozygote (droite) pour une mutation de coloration située sur le chromosome X (© photo MRC Harwell ; Guénet *et al.*, 2015)

Ce phénomène d'inactivation du chromosome X se traduit également par un effet sur la coloration de la robe chez le chat : le locus responsable de la couleur rousse est situé sur le chromosome X. Chez une femelle hétérozygote selon le chromosome X qui sera inactivé, la couleur des poils sera rousse ou d'une autre couleur (noire, chocolat ou cannelle), formant une robe bringée dite « écailles de tortue ». Les seuls cas de mâles présentant cette coloration concernent des individus XXY (Pedersen *et al.*, 2014).

Les mécanismes de l'inactivation du chromosome X ne sont pas toujours les mêmes selon les espèces. Chez les marsupiaux, le chromosome X inactivé est toujours celui d'origine paternelle. Chez les mammifères placentaires, le chromosome X inactivé est aléatoirement choisi dans chaque cellule embryonnaire et hérité de façon somatique dans les cellules filles. Le processus d'inactivation dépend d'un centre d'inactivation du chromosome X (XIC ou *X inactivation centre*) codant notamment pour un long ARN non codant *XIST* (*X inactivation specific transcript*) qui initie l'inactivation d'un des deux chromosomes X. *XIST* se lie au chromosome qui le produit et l'inactive. Il a été montré sur des cellules souches humaines porteuses de trisomie 21 qu'en implantant le gène *XIST* sur un des trois chromosomes 21, celui-ci était inactivé (Jiang *et al.*, 2014).

Les chromosomes X et Y partagent une région relativement petite d'homologie de séquence connue sous le nom de région pseudo-autosomique (PAR). L'appariement et la recombinaison au sein de la PAR des chromosomes X et Y est une étape importante permettant la ségrégation des chromosomes sexuels lors de la méiose chez les mâles (Ellis et Goodfellow, 1989). En raison de la similitude de séquence et de la recombinaison qui se produit entre les chromosomes X et Y dans la PAR, cette région est héritée selon un modèle autosomique. Le domaine de la PAR est délimité par la frontière pseudo-autosomique (PAB), qui est une transition vers la séquence purement spécifique à chacun des sexes.

Chez la plupart des mammifères, les PAR ont une longueur de quelques centaines de kb à plusieurs Mb (Das *et al.*, 2009). La taille, les gènes et la localisation de la PAR diffèrent selon les espèces. Chez l'humain, il y a même deux PAR, l'une de 2,6 Mb et l'autre de 0,32 Mb (Gusmão *et al.*, 2008). Chez le bovin, la PAR mesure environ 6 Mb et comprend 20 gènes (Van Laere *et al.*, 2008; Das *et al.*, 2009; Johnson *et al.*, 2019). La PAB a été localisée sur le nouvel assemblage ARS-UCD.1 à la position ChrX:133,300,518 et la PAR s'étend jusqu'à la fin du chromosome X à 139,009,144 pb (Johnson *et al.*, 2019). La restriction des possibilités de recombinaison des chromosomes sexuels à la région de la PAR entraîne un taux de recombinaison particulièrement élevé par Mb sur cette région (Charlesworth, 2017).

3.3. L'ADN mitochondrial

Les mitochondries sont des organites présents dans les cellules de tous les organismes eucaryotes. Elles mesurent entre 0.5 à 10 microns. Les mitochondries jouent de nombreux rôles dans le métabolisme cellulaire. Elles interviennent dans la production d'énergie sous forme de molécules d'ATP (adénosine triphosphate) et forment un maillon essentiel de la chaîne respiratoire.

3.3.1. Mécanismes de transmission

Les mitochondries possèdent leur propre ADN, distinct du génome nucléaire, ce qui indique une origine extérieure à la cellule. Elles seraient apparentées à une α-protéobactérie de l'ordre des Rickettsiales, groupe comportant de nombreux parasites intracellulaires obligatoires, qui, Il y a environ 2 milliards d'années, auraient intégré une cellule d'archée (microorganisme unicellulaire procaryote) et formé une relation endosymbiotique avec celle-ci.

Le génome mitochondrial des mammifères prend la forme d'une boucle circulaire à double brin d'ADN dont la longueur varie de 15 000 à 17 000 paires de bases (pb) selon les espèces, la séquence d'ADN mitochondrial (ADNmt) bovin de référence UMD3.1 contenant 16 338 pb. La disposition des gènes sur le génome mitochondrial est compacte, avec notamment deux couples de gènes donc les cadres de lecture sont chevauchants. Une région non codante et hyper-variable, appelée D-loop (*Displacement loop*), permet l'initiation de la réplication et la transcription de l'ADNmt (Beinat, 2013). L'ADNmt étant circulaire, la position 1 est fixée arbitrairement et peut être différente chez les espèces. Chez les bovins, le début de la séquence linéaire a été fixée au milieu des 900pb de la région D-loop.

L'ADNmt ne contient que 37 gènes alors qu'on estime à environ un millier le nombre de gènes nécessaires au métabolisme mitochondrial qui ont été intégrés au génome nucléaire au cours de l'Evolution (Calvo *et al.*, 2016). Les 37 gènes de l'ADNmt codent 13 sous-unités de la chaîne de transport d'électrons qui contribuent à la production d'énergie dans la cellule par phosphorylation oxydative (OXPHOS) ; et 24 ARN, comprenant 22 ARNt et deux ARN ribosomiques (16S ARN et 12S ARN), nécessaires à la transcription et la traduction des protéines mitochondriales. Contrairement au génome nucléaire, le génome mitochondrial des mammifères est hérité de manière uniparentale, seulement par voie maternelle, et ne subit pas de recombinaison (Hagström *et al.*, 2014). L'ADNmt est présent en plusieurs copies, avec un nombre variant de 100 à 10 000 dans chaque cellule et s'adaptant aux besoins des cellules selon le tissu (Chinnery et Hudson, 2013).

Différents mécanismes ont été théorisés pour conserver une transmission maternelle comme la fragmentation du génome mitochondrial à l'intérieur des spermatozoïdes (Wei, 1992; DeLuca et O'Farrell, 2012; Amaral *et al.*, 2013) ou une régulation permettant de diminuer la réplication de

l'ADNmt pendant la spermatogenèse (Moraes et Meyers, 2018). Chez certaines espèces, une grande partie des mitochondries est éliminée lors de la phase d'élongation de la spermiogenèse (DeLuca et O'Farrell, 2012). L'autophagie des mitochondries des spermatozoïdes après la fécondation est un autre mécanisme postulé pour prévenir la transmission de l'ADNmt paternel (Al Rawi *et al.*, 2011). Les mitochondries paternelles qui restent après la fécondation subissent une dégradation dans le cytoplasme de l'ovocyte (Thompson *et al.*, 2003) et si l'ADNmt paternel n'est pas éliminé ou dégradé, il serait fortement dilué dans l'ovocyte, qui contient plus de 200 000 copies d'ADNmt, limitant ainsi son effet global (Wolff et Gemmell, 2008). Toutes ces stratégies de prévention de la transmission de l'ADNmt paternel, contribuent à assurer une transmission mitochondriale maternelle. Il existe cependant des cas répertoriés de transmission biparentale de l'ADNmt (Luo *et al.*, 2018).

Tous les bovins modernes[3] sont supposés descendre d'un même ancêtre commun théorique (appelée most recent common ancestor en anglais ou MRCA) qui leur a transmis ses mitochondries. Au fil des générations, le génome mitochondrial de cet MRCA a ensuite acquis des mutations ponctuelles, entraînant le développement d'un certain nombre d'haplogroupes, chacun défini par un sous-ensemble spécifique de polymorphismes (Achilli *et al.*, 2008). Un haplogroupe se définit comme un ensemble d'haplotypes sensiblement similaires et permet de distinguer des populations génétiquement proches.

Chaque nouvelle mutation mitochondriale existe d'abord sous forme de mutation hétéroplasmique, c'est-à-dire dans un état hétérogène dans lequel les mitochondries de type sauvage coexistent avec celles mutées dans la même cellule, avant de disparaître ou d'être fixées à l'état homoplasmique en fonction de facteurs tels que la sélection ou la dérive liée à la transmission au hasard au cours des division cellulaires successives, et au faible nombre de mitochondries transmises aux ovocytes lors de la gamétogénèse (cf. figure 6).

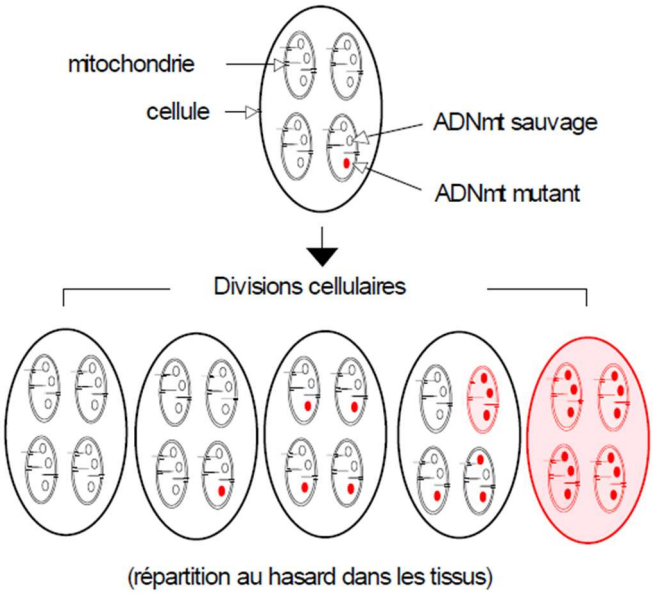


Figure 6 : Illustration de la répartition aléatoire des mitochondries impliquant des degrés variables d'hétéroplasmie (Beinat, 2013)

[3] Tout au long de cette thèse, j'utiliserai l'expression « bovins modernes » pour décrire la population de bovin actuels (en opposition aux animaux et échantillons d'ADN anciens).

Les mutations hétéroplasmiques peuvent être héréditaires ou acquises par mutation somatique *de novo* lors de l'embryogenèse ou du développement postnatal. Enfin, il est important de noter que le degré d'hétéroplasmie varie entre organes, au sein même d'un individu, à la suite d'échantillonnages aléatoires réalisés à chaque division cellulaire.

3.3.2. Rôles physiologiques des mitochondries et pathologies associées à leur dysfonctionnement

Les mitochondries jouent un rôle central au sein du métabolisme énergétique cellulaire en permettant la production d'énergie sous forme d'adénosine triphosphate (ATP) par le processus de phosphorylation oxydative (Ernster et Schatz, 1981).

La principale source de production d'énergie cellulaire est la réaction de phosphorylation oxydative (ou OXPHOS), un processus couplant la réoxydation des coenzymes réduits (NADH et FADH2) générés lors de la glycolyse et du cycle de Krebs, et la synthèse d'ATP à partir de la phosphorylation de l'ADP. Cette réaction se produit au niveau de la chaîne respiratoire mitochondriale constituée de quatre complexes protéiques localisés dans la membrane interne mitochondriale. Cependant, seule une partie des sous-unités de ces complexes est codée par le génome mitochondrial, les autres protéines étant d'origine nucléaire :

- Le complexe I (NADH *ubiquinone oxydoréductase*), constitué d'environ 43 sous-unités protéiques dont 7 d'origine mitochondriale
- Le complexe II (*succinate ubiquinone oxydoréductase*), constitué de 4 sous-unités toutes codées par le génome nucléaire
- Le complexe III (*ubiquinol cytochrome c réductase*), constitué de 11 sous-unités dont 1 codée par l'ADNmt
- Le complexe IV (*cytochrome c oxydase*), constitué de 13 sous-unités dont 3 d'origine mitochondriale
- Le complexe V (*ATP synthétase*) est constitué de 16 sous-unités dont 2 codée par l'ADNmt

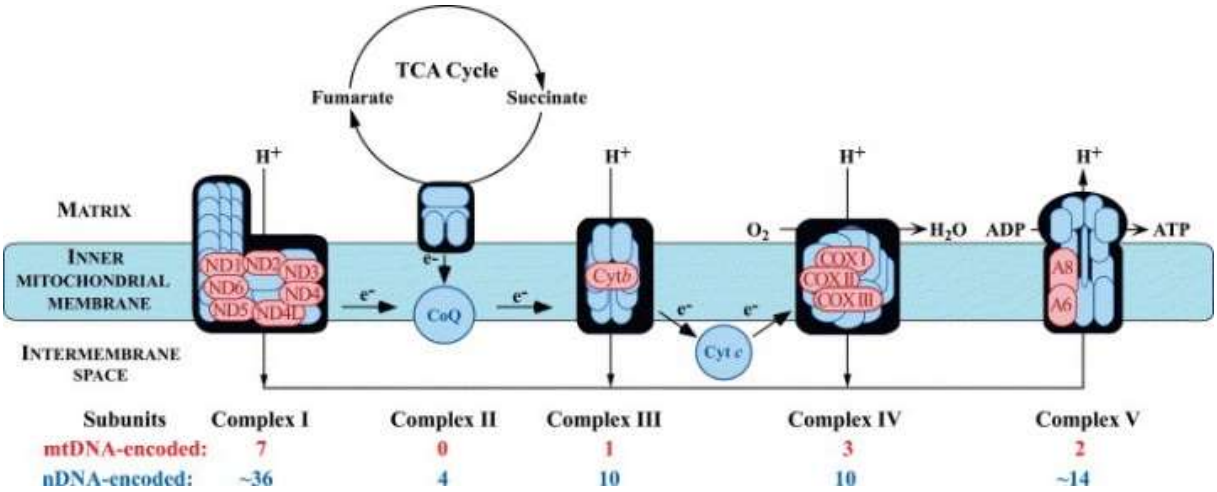


Figure 7 : Représentation de la chaine respiratoire mitochondriale.
En bleu sont représentées les sous-unités codées par l'ADN nucléaire et en rouge celles codées par le génome mitochondrial. (DiMauro, 2004)

Elles jouent également un rôle essentiel dans la biosynthèse des macromolécules, telles que les nucléotides et les lipides. En effet, les mitochondries sont le siège de plusieurs réactions catalytiques : le cycle de Krebs, la ß-oxydation des acides gras et le cycle de l'urée. Les mitochondries interviennent également dans la régulation des signaux biologiques, comme la mort cellulaire programmée ou apoptose, l'homéostasie des ions calciums, du fer et du cuivre, ainsi que dans des aspects spécifiques du processus de vieillissement, comme la sénescence cellulaire et l'inflammation chronique (Galluzzi *et al.*, 2008). Elles protègent aussi la cellule contre le stress oxydatif, dû aux radicaux libres. Par conséquent, les mutations délétères du génome mitochondrial peuvent avoir un effet très néfaste sur l'organisme.

Les troubles mitochondriaux, causés par des mutations du génome mitochondrial, constituent un groupe de maladies hétérogènes chez l'humain. Plus de 250 mutations ponctuelles pathogènes ainsi que des réarrangements à petite et à grande échelle de l'ADNmt ont été identifiés (Greaves et Taylor, 2006). Les troubles mitochondriaux sont considérés comme l'une des formes les plus communes de maladie métabolique (Chinnery *et al.*, 2000). Ils se manifestent habituellement dans les tissus consommateurs d'énergie tels que le système nerveux central, les muscles, le système auditif et le système visuel, mais presque tous les organes, peuvent être impliqués.

Du fait de la double origine des protéines mitochondriales, les anomalies génétiques impliquées dans les maladies mitochondriales peuvent provenir soit des gènes mitochondriaux soit des gènes nucléaires. Le déterminisme génétique de ces maladies est donc variable et peut être autosomique récessif, autosomique dominant ou lié au chromosome X dans le cas de mutation des gènes nucléaires, ou bien à transmission maternelle lorsqu'il s'agit de variants délétères portés par le génome mitochondrial. Seuls 10 à 15% des maladies mitochondriales sont dues à des mutations du génome mitochondrial (Mancuso *et al.*, 2007).

Dans le cas d'anomalies du génome mitochondrial associées à de l'hétéroplasmie, un seuil de proportion de molécules d'ADNmt mutées doit être dépassé pour que le fonctionnement normal des mitochondries dans le tissu concerné soit affecté et que la maladie mitochondriale s'exprime (Baranowska *et al.*, 2009).

3.4. Transmission uniparentale du génome mitochondrial et du chromosome Y

Dans les années 1980 et 1990, le séquençage du génome humain était encore incomplet et les données de séquence n'étaient pas suffisamment fiables pour réaliser des études de génétique des populations. L'absence de recombinaison et la transmission uniparentale ont fait du génome mitochondrial et de la partie spécifique au chromosome Y de très bons outils pour retracer l'histoire phylogénétique d'une espèce ou d'un groupe d'espèces. La faible diversité génétique des espèces domestiquées signifie que pour tout gène nucléaire, il faut séquencer des milliers ou des dizaines de milliers de paires de bases pour avoir une chance de trouver des SNP informatifs permettant de distinguer des populations (Kivisild, 2015). A une période où les technologies de séquençage étaient limitées, le génome mitochondrial a donc été une cible d'étude de choix car le taux élevé de mutation de l'ADNmt rendait plus probable la découverte d'un variant intéressant à l'échelle de la population que pour tout autre locus nucléaire. Chez les vertébrés, le taux de mutation mitochondriale est en effet 25 fois supérieur au taux de mutation de l'ADN nucléaire (Lynch *et al.*, 2006).

Il est important de noter que le génome mitochondrial et la partie spécifique du chromosome Y seuls ne permettent pas de retracer précisément l'origine phylogénétique d'une espèce. Par exemple, l'utilisation de marqueurs sur le génome nucléaire a permis de montrer que nous possédons 2 à 3% d'ADN de l'homme de Néandertal (Green *et al.*, 2010) alors qu'en l'absence de mitochondries d'origine néandertalienne dans les populations modernes, les études basées sur le génome mitochondrial allaient à l'encontre de l'hypothèse d'une hybridation (Serre *et al.*, 2004).

L'ADNmt a permis de créer les premiers arbres phylogénétiques. Des étiquettes alphabétiques ont été attribuées aux branches les plus courantes, et ont, par la suite, défini les haplogroupes d'ADNmt.

Les haplotypes mitochondriaux sont des groupes de génomes mitochondriaux qui définissent collectivement les lignées maternelles de groupes d'individus. Il a également été démontré certains les haplotypes mitochondriaux prédisposent ou protègent les individus contre un certain nombre de maladies et de caractéristiques phénotypiques. Par exemple, ils sont associés à la tolérance à la chaleur et la longévité chez l'humain (Tanaka *et al.*, 1998; Ruiz-Pesini *et al.*, 2004), à la fertilité et à la prolificité chez le porc domestique (Tsai *et al.*, 2016), ou encore à la durée de gestation et la qualité du lait et de la viande chez le bovin (Brown *et al.*, 1989).

En génétique des populations, les analyses réalisées à partir du chromosome Y permettent d'étudier les lignées paternelles. Par exemple, de nombreuses études démontrent également l'existence de lignées très prolifiques chez l'être humain comme par exemple une lignée associée à Genghis Khan et sa descendance qui aurait transmis un même chromosome Y à environ 16 millions de descendants mâles vivants aujourd'hui (Zerjal *et al.*, 2003). Chez les bovins, l'utilisation de l'IA permet la diffusion d'un même mâle de manière intensive, ce qui peut conduire à une dérive très importante. Ainsi, les chromosomes Y de l'ensemble des taureaux Holstein Nord-Américains ne proviennent que de deux ancêtres (Yue *et al.*, 2015).

Aujourd'hui l'accès aux séquences de génome complet pour de nombreux individus a permis d'utiliser les SNP pour connaître en détail les structures génétiques des populations. Cependant, le génome mitochondrial et la part non recombinante du chromosome Y demeurent des marqueurs sources d'information.

En résumé :

- Les chromosomes sexuels dérivent d'une paire d'autosomes. Ils se sont ensuite différenciés et ont acquis des gènes spécifiques.
- Seuls un ou deux petit(s) segment(s) recombinent encore et sont appelés région(s) pseudo-autosomique(s) (PAR).
- Chez les femelles, un des deux chromosomes X est inactivé aléatoirement dans chaque cellule.
- Les mitochondries forment un maillon essentiel de la chaîne respiratoire qui permet la production d'énergie sous forme de molécules d'ATP et possèdent leur propre information génétique sous la forme d'un ADN circulaire.
- Le génome mitochondrial se transmet uniquement par voie maternelle.
- Le génome mitochondrial et la partie non recombinante du chromosome Y permettent de réaliser des études phylogénétiques pour retracer l'histoire d'une espèce sur la voie femelle et mâle, respectivement.

Malgré leurs rôles dans l'expression de caractères de fertilité et dans le métabolisme énergétique, les chromosomes sexuels et le génome mitochondrial ne sont actuellement pas pris en compte dans les évaluations génomiques. Les raisons sont obscures et sans doute historiques. En effet, les travaux initiaux de cartographie génétique ayant permis l'identification de QTL se sont appuyés sur un dispositif de population de type petites-filles ou « *granddaughter design* » (Weller *et al.*, 1990).

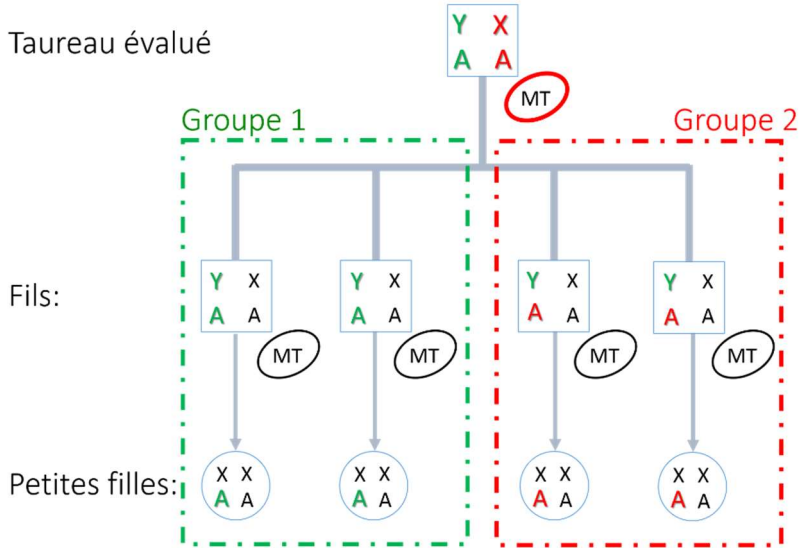


Figure 8 : Représentation schématique du dispositif petites-filles.

Pour la paire de chromosomes sexuels (XY) et une paire d'autosome (AA) du taureau évalué, le chromosome hérité du père de l'animal est représenté en vert et de la mère en rouge. Les fils ayant reçu l'un ou l'autre des autosomes d'une même paire sont regroupés selon la provenance de l'autosome (A).

Dans ce dispositif qui utilise le testage sur descendance réalisé, on analyse la moyenne des performances des filles d'un taureau en l'assimilant à une performance du taureau. On compare ainsi les performances entre groupes de fils ayant reçu l'un ou l'autre segment chromosomique d'une paire d'autosomes, les performances de ces fils étant elles même estimées à partir de leurs propres filles. Comme l'illustre la figure 8, les fils portent tous le chromosome Y du père et les petites filles ne portent ni les chromosomes sexuels, ni le génome mitochondrial du père. Ce type de dispositif ne permet donc pas de réaliser des études de cartographie de QTL sur ces régions spécifiques. Nous disposons aujourd'hui de tous les éléments nécessaires à l'étude de ces chromosomes grâce au rééquilibrage entre les différentes voies de sélection (père-fils, père-fille, mère-fils, mère-fille) permis par l'évaluation génomique.

Cette thèse repose sur l'exploitation de divers types de données (séquences de génomes complets, génotypes de puces à SNP, bases de données phénotypiques) afin de détecter des polymorphismes génétiques au niveau du génome mitochondrial et des chromosomes sexuels et de caractériser leurs effets sur les caractères de production, reproduction et santé. Les trois principaux objectifs de ce projet sont :

- d'identifier et caractériser des variants d'intérêt au niveau du génome mitochondrial et des chromosomes X et Y,
- d'estimer la part de variance phénotypique expliquée par le génome mitochondrial,
- de détecter des variants délétères portées par ces mêmes régions génomiques et pouvant entraîner des anomalies génétiques.

Chapitre 2 : Diversité du génome mitochondrial et effets sur les caractères de production et reproduction

L'objectif des travaux exposés dans cette première partie est d'étudier le lien entre génome mitochondrial et caractères d'intérêt afin d'évaluer l'intérêt de prendre en compte l'ADN mitochondrial dans le cadre des évaluations génomiques. Nous avons tout d'abord étudié la variabilité, la diversité et la répartition des haplotypes au sein des races dans le but de déterminer le potentiel de sélection sur le génome mitochondrial. Plus la variabilité est grande, plus la probabilité que des allèles favorables ou défavorables pour un caractère ségrégent dans la population est élevée. Les deux premières études de ce chapitre sont donc consacrées à l'estimation et à la description de la diversité mitochondriale au sein des races françaises tandis que les études suivantes portent sur l'estimation d'effets mitochondriaux sur des caractères de production et de fertilité à partir d'informations généalogiques (études 3 et 4) et de données de génotypage (étude 5).

Au cours de cette thèse, nous avons étudié le génome mitochondrial selon trois approches différentes : par séquençage, génotypage, et étude généalogique à travers l'identification de lignées maternelles. Afin de clarifier la lecture de ce manuscrit, j'utiliserai « lignée maternelle » ou LM pour décrire une lignée partageant le même plus vieil ancêtre maternel commun identifié à partir des pedigrees renseignés dans les bases de données nationales.

Pour bien comprendre l'agencement des études au cours de ma thèse, il faut replacer chacune des étapes dans son contexte temporel (figure 9).

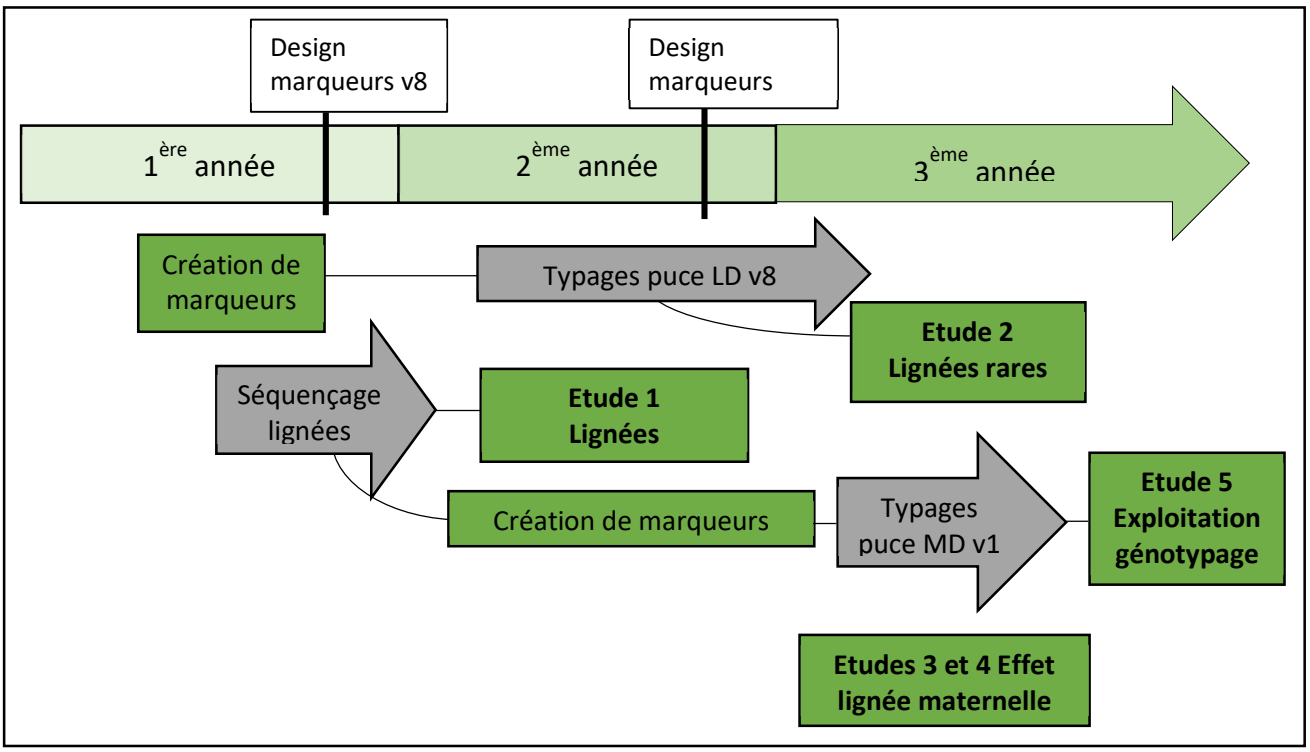


Figure 9 : Chronologie du travail de thèse portant sur le génome mitochondrial

L'étude du génome mitochondrial s'est réalisée en quatre étapes. Tout d'abord, fin 2017, au début de ma 1ère année de thèse, j'ai utilisé 300 séquences présentes dans la base de données « nucleotide » du *National Center for Biotechnology Information* (NCBI ; https://www.ncbi.nlm.nih.gov/nuccore/) pour dessiner les amorces de 29 marqueurs qui ont été ajoutés sur la puce EuroG10k v8, mise en production en juin 2018. Ces marqueurs m'ont permis d'identifier des individus porteurs de lignées maternelles rares au début de la 3ème année, lorsque nous avons disposé de suffisamment d'individus génotypés (étude 2). En parallèle, j'ai sélectionné des animaux à partir d'informations de pedigree et de race dont le génome mitochondrial a été entièrement séquencé en juin 2018. Ces deux études m'ont permis de réaliser un inventaire de la diversité mitochondriale dans la population bovine française. De plus, grâce à l'analyse de ces séquences (étude 1), j'ai pu ajouter 464 nouveaux marqueurs sur la puce EuroGMD qui a été mise en production en mai 2019. Les études 1 et 2 ont été réalisées en partie avec la collaboration de l'équipe de génomique des populations de l'université LMU (Munich) dirigée par le Dr. Ivica Medugorac auprès laquelle j'ai fait un séjour de deux mois à l'automne 2018.

En l'absence de génotypage sur les marqueurs mitochondriaux au début de ma 3ème année de thèse, j'ai estimé la part de variance due au génome mitochondrial sur des caractères de production laitière et de fertilité mâle uniquement à travers l'effet lignée (LM) déterminé par le pedigree, sans information génomique (études 3 et 4).

Finalement, en novembre 2019, j'ai disposé de suffisamment d'individus génotypés pour réaliser une première étude en exploitant les données de la puce EuroGMD et pu faire une première analyse sur les caractères de production laitière basée non plus sur la généalogie seule mais sur les informations de génotypage (étude 5).

En résumé, dans ce chapitre nous avons utilisé plusieurs approches différentes, phylogénétique, séquençage, estimation de paramètres génétiques et étude d'association, afin de caractériser au mieux la variabilité du génome mitochondrial et son effet sur certains caractères d'intérêt en sélection. Ce sont ces différentes approches que je vais vous présenter à travers cinq études distinctes.

1. Etude 1 : Diversité et origine des haplogroupes mitochondriaux

Dans l'optique de faciliter le cheminement du lecteur à travers les approches parfois très différentes adoptées dans ma thèse, j'ai choisi de débuter ce chapitre par une synthèse des analyses phylogénétiques basées sur le génome mitochondrial bovin dans la littérature.

1.1. Etat de l'art : génome mitochondrial et domestication des bovins

Le passage de l'agriculture nomade de subsistance à l'agriculture sédentaire a eu des conséquences culturelles, économiques et démographiques profondes. Grâce à la culture des plantes et à l'élevage, l'homme a pu manipuler et gérer son environnement naturel d'une manière totalement nouvelle (Diamond, 2002). Les bovins taurins modernes (*Bos taurus*) proviennent de la domestication de l'auroch dans le croissant fertile pendant le Néolithique, il y a plus de 10 000 ans (Ajmone-Marsan *et al.*, 2010). Un deuxième événement de domestication d'une autre population d'aurochs a eu lieu dans la vallée de l'Indus 1500 ans après, à l'origine des zébus modernes (Larson *et al.*, 2014). Un potentiel troisième évènement à l'origine des taurins Africains aurait eu lieu au nord-est de l'Afrique il y a 8000-9000 ans, mais il ne fait aujourd'hui pas consensus parmi la communauté scientifique (figure 10 ; Pitt *et al.*, 2019).

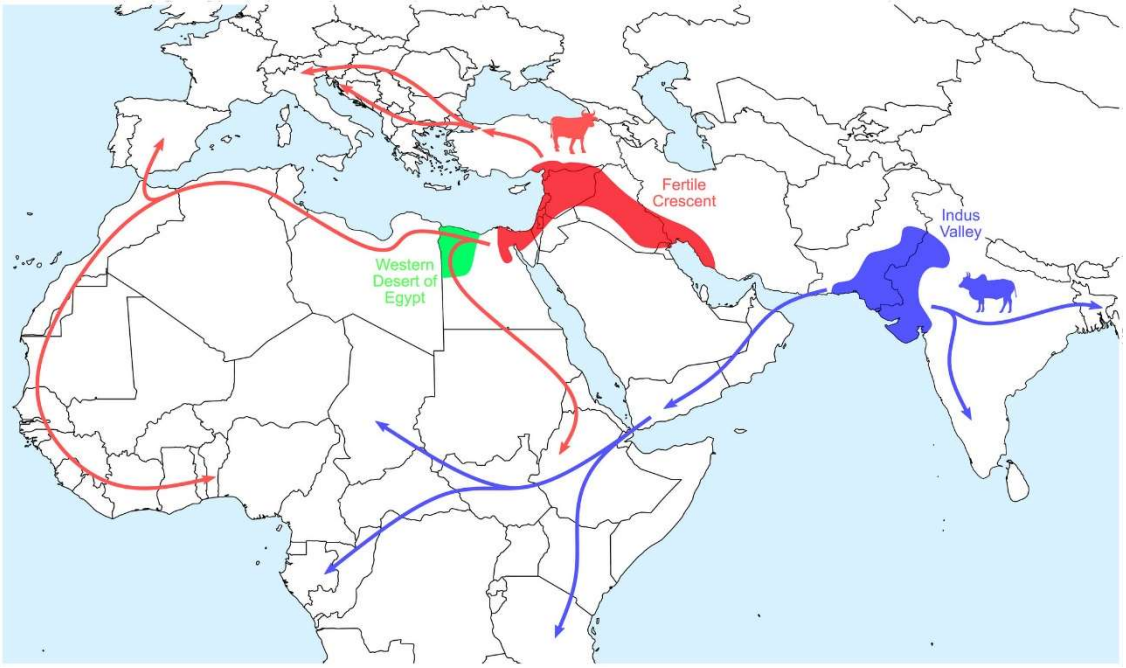


Figure 10 : Représentation des foyers de domestication et des principales routes migratoires des *Bos taurus* (rouge) et *Bos indicus* (bleu) incluant en vert le foyer d'un potentiel 3ème évènement de domestication (Pitt *et al.* 2019)

Les étapes de la domestication animale peuvent être reconstituées grâce à des analyses phylogénétiques et géographiques en utilisant des données génomiques nucléaires (autosomes ou chromosome Y) ou mitochondriales (Groeneveld *et al.*, 2010).

De nombreux goulots d'étranglement ont eu lieu lors de la dissémination des bovins dans le monde, façonnant la structure des populations actuelles (figure 11, Lenstra *et al.*, 2014). La majorité des *Bos taurus* modernes possède des mitochondries appartenant au macro-haplogroupe T (Troy *et al.*, 2001). Il semble qu'il ait été prédominant au Proche-Orient dans les populations d'aurochs à la fin du Pléistocène (Edwards *et al.*, 2007).

Les ancêtres maternels des animaux porteurs de l'haplogroupe T sont vraisemblablement tous nés et ont été domestiqués dans le Croissant fertile. Ils se sont ensuite reproduits et disséminés parmi les troupeaux domestiques en suivant les migrations et échanges commerciaux des êtres humains (Ajmone-Marsan *et al.*, 2010).

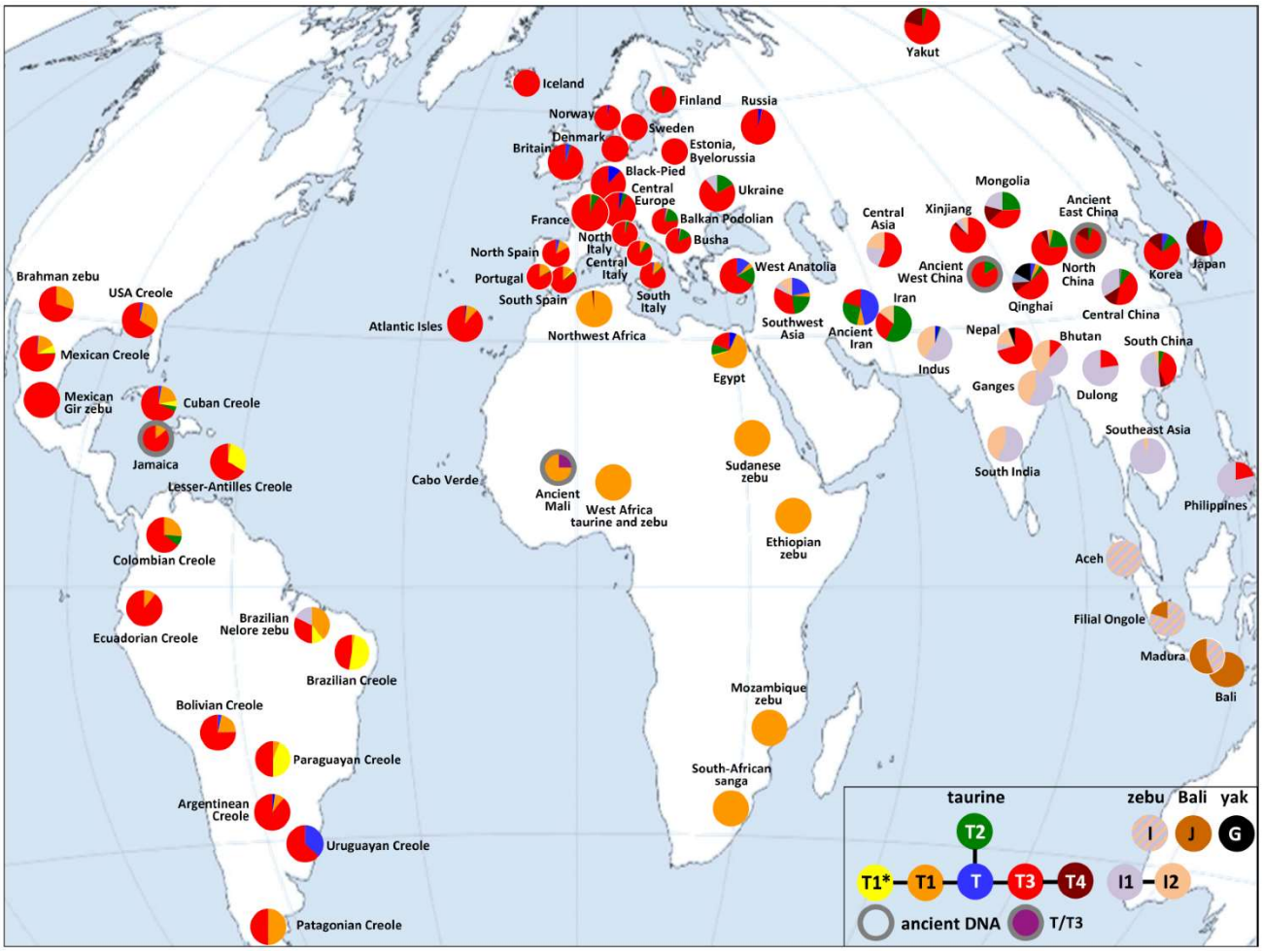


Figure 11 : Distribution géographique des haplogroupes principaux taurins et zébus.
T1 représente l'haplogroupe T1c1a. (Lenstra et al., 2014)*

Le macro-haplogroupe T est composé de deux clades principaux. Le premier regroupe les haplogroupes T1, T2, T3 et le T4 désormais assimilé à l'haplogroupe T3, tandis que le deuxième est formé uniquement du T5, groupe non représenté sur la figure 11 (Achilli *et al.*, 2009). L'haplogroupe T1 est principalement observé chez les bovins Africains ce qui étaye l'hypothèse d'un troisième foyer de domestication (Bradley *et al.*, 1996). L'haplogroupe T4 est une sous-catégorie de T3, observée uniquement en extrême orient : chez un individu de la Chine orientale datant de 4500 BP[4] (Cai *et al.*, 2014), chez une partie des bovins coréens modernes (Achilli *et al.*, 2008), et dans plus de la moitié des bovins japonais (Noda *et al.*, 2018).

Chez les *Bos indicus*, deux haplogroupes principaux ont été différenciés : I1 et I2. La répartition entre les animaux porteurs d'haplotypes I1 et I2 est relativement identique en Inde et en Indonésie. L'haplotype I2 est majoritaire en Chine et dans le reste de l'Asie du Sud-Est possiblement du fait d'une dérive génétique accompagnant la migration des bovins vers le Sud-Est.

[4] *Before present ;* avant le présent

La distribution des haplogroupes mitochondriaux au sein des différentes populations géographiquement distinctes apporte de nombreuses informations sur l'expansion des populations bovines (voir figure 11 ; Lenstra *et al.*, 2014). Par exemple, la fréquence très élevée de l'haplogroupe T1c1a1 dans les races créoles nous renseigne sur le déroulement de la colonisation des Amériques depuis la fin du 15ème siècle. Les colons portugais ont importé les premiers bovins au Brésil en 1534 en provenance du Cap-Vert et parmi eux des animaux porteurs de l'haplotype T1c1a1. Aujourd'hui, cet haplotype est presque absent en Europe mais représente 31% des bovins aux Antilles (d'après Lenstra *et al.*, 2014 par Pitt *et al.*, 2019).

Même si le macro-haplogroupe T est le plus fréquent, il ne représente pas la totalité des génomes mitochondriaux des races taurines modernes. D'autres haplogroupes plus rares, appelés P, Q et R ségrègent dans les races modernes, provenant possiblement d'évènements ponctuels d'introgression d'aurochs dans les cheptels domestiques (Bonfiglio *et al.*, 2010; Lari *et al.*, 2011). Les haplogroupes Q et R ont été retrouvés notamment dans des races italiennes, iraquiennes (Achilli *et al.*, 2009) et égyptiennes (Olivieri *et al.*, 2015). Les haplogroupes Q et T sont probablement d'origine proche-orientale, tandis que les groupes P et R proviendraient de populations européennes d'aurochs sauvages (Achilli *et al.*, 2009). L'haplogroupe R n'a été observé que chez des bovins modernes italiens et est phylogénétiquement distinct des haplogroupes P, Q, et T (Bonfiglio *et al.*, 2010 ; figure 12). L'haplotype P est assimilé aux aurochs européens mais, à ce jour, n'a pas été retrouvé dans des races européennes modernes. Il a cependant été identifié chez des bovins coréens et japonais (Achilli *et al.*, 2008; Noda *et al.*, 2018).

Deux autres haplotypes ont également été identifiés à partir des ossements de deux individus anciens. A ce jour, l'haplogroupe C est caractérisé par un unique génome mitochondrial séquencé seulement partiellement, issu d'un échantillon bovin du nord de la Chine datant de 10 660 ans BP (Zhang *et al.*, 2013). Le seul individu appartenant à l'haplogroupe E est un échantillon partiel datant du néolithique retrouvé sur le site de Eilsleben en Allemagne et distinct à la fois de l'haplogroupe P et du macro haplogroupe T (Edwards *et al.*, 2007b).

La figure 12 représente les principaux haplogroupes mitochondriaux. Les haplogroupes C et E n'ont pas été étudiés dans la suite de cette étude en raison de l'absence de séquences mitochondriales complètes.

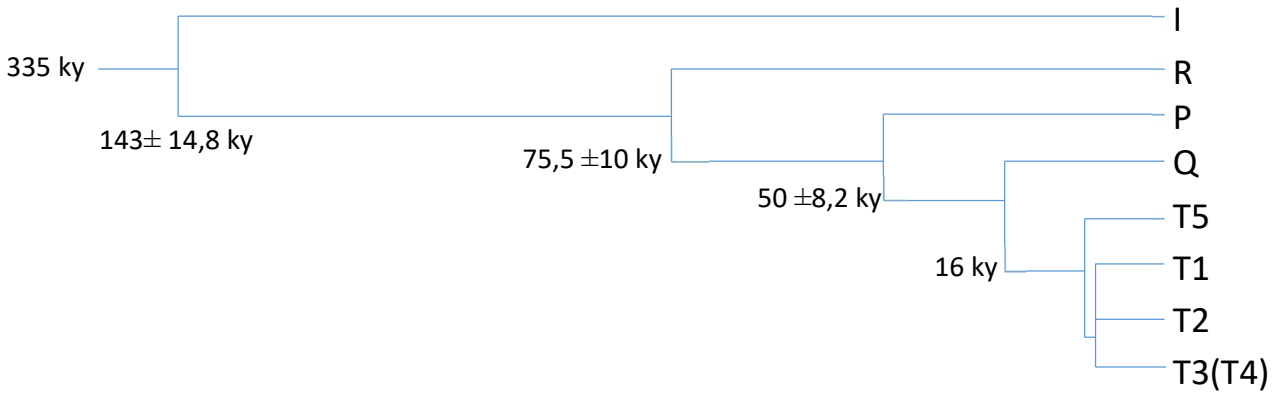


Figure 12 : Représentation des différents clades phylogénétiques mitochondriaux et des âges de coalescence. *D'après (Achilli et al., 2009 et 2010)*

1.2. Matériels et méthodes

Dans la partie précédente, nous avons présenté une image globale de la répartition des haplotypes mitochondriaux bovins. L'objectif de cette étude est de caractériser plus en détail la diversité génétique du génome mitochondrial dans la population bovine française.

1.2.1. Analyse des pedigrees

Dans un premier temps, nous avons étudié le nombre de lignées maternelles (LM) dans les races. L'analyse des pedigrees a été réalisée à partir des informations de généalogie renseignées dans les bases nationales et du programme Lineage, une des dernières évolutions du package pedig (Boichard, 2002). Ce programme permet d'identifier, pour chaque individu, son plus vieil ancêtre connu sur la voie maternelle.

Puisque l'ADNmt se transmet par voie maternelle et sans recombinaison, cela signifie que tous les descendants d'une même LM possèdent théoriquement le même génome mitochondrial. Ainsi, le séquençage d'un seul descendant d'une lignée permet de connaître le génome mitochondrial de chacun des descendants par voie maternelle du même ancêtre maternel (voir figure 13). Cependant, ce principe repose sur trois postulats : (i) la justesse des informations généalogiques ; (ii) l'absence d'hétéroplasmie[5] ; et (iii) l'absence de mutations *de novo* au sein d'une lignée. Ces trois points seront discutés plus loin dans ce chapitre.

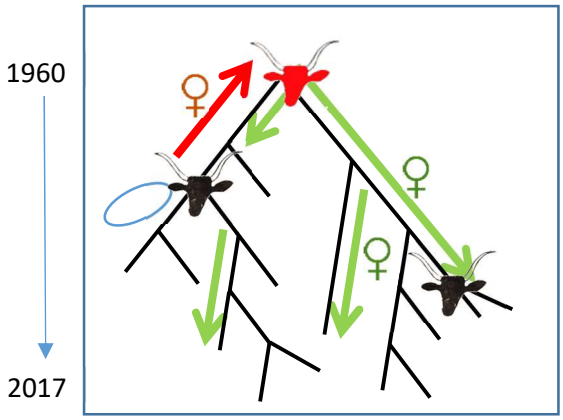


Figure 13 : Principe de la transmission uniparentale du génome mitochondrial
La connaissance de la séquence nucléotidique mitochondriale d'un individu (en bleu) permet d'attribuer la même séquence à son plus vieil ancêtre maternel (en rouge) puis à l'ensemble des descendants par voie maternelle de cette même lignée (en vert)

Pour chaque race, j'ai repéré les animaux nés entre 2015 et 2019 pour lesquels les deux parents étaient renseignés. J'ai ensuite calculé le nombre d'ancêtres maternels et paternels (pour comparaison) de cette cohorte. Est considéré comme ancêtre tout individu n'ayant pas de parent connu. J'ai également calculé le nombre moyen de descendants par mère et père, ainsi que son écart-type (précédé du signe « ± »). Enfin, j'ai étudié le nombre moyen de générations connues lors d'une remontée de généalogie sur la voie mâle et la voie femelle, et indiqué le pourcentage d'animaux avec strictement plus de cinq générations connues sur la voie femelle. Ces chiffres sont présentés dans le tableau 4 dans la partie résultats et discussion.

[5] Hétéroplasmie : présence de mitochondries avec des génomes différents au sein des cellules d'un même individu des suites de mutations *de novo* ou d'une transmission de mitochondries paternelles

1.2.2. Echantillonnage des individus séquencés

L'objectif de cette étude était de représenter un grand nombre de races et que chacune des lignées séquencées représente le plus grand nombre de descendants vivants. Notre budget pour ce projet de recherche nous permettait de séquencer 384 échantillons ADN. Nous avons choisi de prioriser les six grandes races (Holstein, Montbéliarde, Normande, Charolaise, Limousine, Blonde d'Aquitaine) pour lesquelles nous avons séquencé environ 60 individus par race. Nous avons complété cet échantillon par des individus de races à plus faibles effectifs : Abondance, Aubrac, Brune, Salers, Bazadaise, Bretonne, Tarentaise, Simmental, Rouge des Prés, Vosgienne, Rouge Flamande, Parthenaise, Gasconne.

Nous avons choisi les lignées maternelles les plus prolifiques et sélectionné les échantillons après les étapes suivantes :

1) Identification de l'ancêtre maternel le plus ancien connu pour chaque animal avec le programme Lineage ;
2) Comptage du nombre de descendants par lignée et sélection des lignées les plus prolifiques par ordre croissant ;
3) Recherche d'échantillons ADN pour au moins un représentant de ces lignées dans les banques d'ADN accessibles.

La plupart des échantillons ADN proviennent des bases de données d'échantillons de l'ONAB, et des animaux génotypés dans le cadre de la sélection génomique par Labogena, laboratoire national de génotypage animal. Labogena a été un partenaire clé dans la mise en place de la sélection génomique et c'est le principal laboratoire de génotypage bovin en France. Enfin, une partie des échantillons de race à faibles effectifs proviennent de projets de recherche européens sur la biodiversité (RESGEN ; (Cañón *et al.*, 2001; Gautier *et al.*, 2010).

Chaque échantillon ADN qui s'est révélé inexploitable a été remplacé par un échantillon d'un autre descendant de la même lignée si possible et dans le cas contraire, la lignée a été remplacée par une autre lignée en suivant un ordre décroissant de prolificité. L'ADN des échantillons provient de différents projets et a donc été extrait à partir de différents tissus (biopsie d'oreilles, sang, semence).

1.2.3. Amplification et purification de produits PCR

J'ai amplifié le génome mitochondrial en trois segments partiellement chevauchants avec le kit GoTaq® Long PCR Master Mix de Promega et les amorces présentées dans le tableau ci-dessous. Ces amorces ont été dessinées et testées par nos collaborateurs allemands.

PCR	Taille fragment (bp)	F/R	5' np	3' np	Taille amorce (nt)	Séquence (5'→3')	T (°C)
Bt1	6631	F	1271	1290	20	CCGTCACCCTCCTCAAATAG	59,54
		R	7902	7883	20	CTGGGATTGCGTCTGTTTTT	60,11
Bt2	6498	F	7176	7195	20	GATGCCCTCCACCATATCAC	60,17
		R	13674	13655	20	GAACTTGAAGGCGTTTGAGG	59,85
Bt3	6803	F	11818	11837	20	TACACGGGAAAATGCACTCA	60,11
		R	2320	2301	20	GGAATGCTGGAGGTGATGTT	59,93

Tableau 3 : Amorces pour l'amplification en PCR long range du génome mitochondrial

L'amplification PCR a été faite avec un appareil Mastercycler Eppendorff Nexus et le cycle de PCR est détaillé en annexe 1.

J'ai ensuite vérifié la qualité des amplicons en faisant migrer 2 µL de produits PCR par électrophorèse sur gel d'agarose 1% coloré au bromure d'éthidium. Enfin, j'ai groupé les trois produits de PCR pour chaque échantillon et les ai purifiés avec le kit Wizard® SV Gel and PCR Clean-Up System de Promega, avant de procéder à leur envoi à nos collaborateurs allemands, qui ont ensuite réalisé le séquençage de ces échantillons.

1.2.4. Séquençage du génome mitochondrial

Le séquençage des 384 génomes mitochondriaux a été réalisé par le laboratoire Gene center de Munich. Les concentrations en ADN des 384 pools de PCR ont été mesurées avec le kit Qubit dsDNA HS (ThermoFischer Scientific) et une dilution finale de 0,2ng/ul a été préparée. Les bibliothèques de séquençage ont été réalisées avec le kit Illumina NexteraXT selon le protocole du fabricant avec les modifications suivantes : la quantité d'ADN et les volumes de réactifs ont été réduits de 25% par rapport aux valeurs initiales. Les réactifs ont été distribués dans des plaques de 96 puits par un microdistributeur Mantis (Formulatrix, Bedford MA, USA). Après amplification PCR, les bibliothèques à double code barre ont été regroupées en quantités équimolaires selon les mesures réalisées avec un Nanodrop (ThermoFischer Scientific) et les bibliothèques groupées ont été analysées sur un Bioanalyseur 2100 (Agilent, Santa Clara CA, USA). Les bibliothèques groupées ont été séquencées sur un séquenceur Illumina HiSeq1500 (Illumina, San Diego CA, USA) en paired-end avec des lectures de 100 pb.

Lors de mon séjour en Allemagne, j'ai procédé à l'identification des variants génétiques ou *variant calling* avec le programme varscan v2.3.6 (Koboldt *et al.*, 2009) en alignant sur le génome mitochondrial de référence (AY526085, associée à l'UMD3.1. disponible sur https://bovinegenome) les lectures correspondant au séquençage de chacun des 384 individus. La couverture moyenne de chacun des génomes mitochondriaux analysés était supérieure à 100x au sein de de notre panel. Selon les recommandations de l'équipe allemande, j'ai conservé uniquement les variants présents sur au moins 70% des lectures afin de m'affranchir des erreurs de séquençages et d'une hétéroplasmie éventuelle. J'ai ensuite construit une séquence au format « fasta » pour chacun des 384 génomes mitochondriaux en remplaçant dans la séquence de référence (AY526085) les allèles des variants détectés. Ensuite, j'ai réalisé une deuxième session d'alignement des lectures de chaque génome mitochondrial sur sa propre séquence fasta afin d'améliorer le pourcentage d'alignement de certaines lectures qui auraient pu mal s'aligner sur le génome mitochondrial de référence, notamment dans la région hypervariable, et conduire à l'omission de certains variants. A la suite de cette seconde étape, j'ai modifié les séquences fasta en fonction des nouveaux variants détectés et procédé à une troisième session d'alignement selon le même principe. Comme aucun nouveau variant n'a été détecté, j'ai arrêté les itérations. A titre indicatif, l'équipe de LMU qui utilise la même procédure chez le mouton réalise un nombre plus élevé d'itérations car l'animal dont est issu le génome de référence est phylogénétiquement plus éloigné de leur population domestique que le référent bovin de notre population (communication personnelle, Elisabeth Kunz).

1.2.5. Ajouts de séquence de génome mitochondrial dans les séquences complètes produites pour l'équipe G2B

Lors de ma thèse, j'ai également récupéré les séquences mitochondriales des animaux dont le génome complet avait été séquencé dans le cadre de différents projets menés par l'équipe G2B. Au total, j'avais à disposition 344 séquences alignées sur une version de l'assemblage du génome bovin UMD3.1 privée du génome mitochondrial. Pour chacun des génomes à disposition, j'ai donc récupéré toutes les lectures non alignées et les ai réalignées sur le génome mitochondrial de référence (AY526085) avec les outils samtools (Li *et al.*, 2009). Ce faisant j'ai observé que près de 1000 paires de bases n'étaient pas couvertes avec ces seules lectures. J'ai donc réalisé à l'aide du « génome *browser* » de l'Université de Californie, Santa Cruz (http://genome-euro.ucsc.edu) un BLAT du segment manquant sur l'assemblage UMD3.1 et identifié une région du chromosome 19 possédant un pourcentage d'identité de 91.4% avec ce dernier. J'ai supposé que les séquences mitochondriales manquantes avaient été alignées sur cette portion du génome nucléaire et j'ai donc ajouté ces lectures de la région nucléaire Chr19:22814301-22815400 aux lectures non alignées avant de réitérer l'alignement sur le génome mitochondrial de référence. Enfin, j'ai observé une forte variation de couvertures de lectures sur le génome mitochondrial entre individus/génomes qui avaient pourtant des couvertures similaires sur le génome nucléaire. La majorité des génomes produits à partir d'ADN extrait de semence avaient une couverture moyenne de 1X, du fait d'une dégradation du génome mitochondrial au moment de la spermatogénèse (cf. chapitre 1, partie 3.4.). J'ai donc choisi de ne sélectionner que les séquences avec une couverture moyenne supérieure à 100x. Au total, j'ai pu ajouter 94 séquences à notre jeu de données représentant 81 lignées différentes et j'ai procédé à une détection de variants génétiques avec le programme varscan v2.3.6 de façon similaire à celle utilisée pour les 384 nouvelles séquences (Koboldt *et al.*, 2009 ; 1.2.4.).

1.2.6. Annotation des variants

J'ai réalisé l'annotation des variants avec le programme snpEff version 4.3. (Cingolani *et al.*, 2012). J'ai utilisé une option pour spécifier l'utilisation d'un code génétique spécifique, le génome mitochondrial ayant un code génétique légèrement différent du code génétique de l'ADN nucléaire[6]. La séquence mitochondriale de référence n'a pas été modifiée lors du changement d'assemblage UMD3.1 vers ARS-UCD1. J'ai donc utilisé la base d'annotation la plus récente Ensembl 99 faite sur l'ARS-UCD1.2 (accessible sur ftp://ftp.ensembl.org/pub/release-99/gff3/bos_taurus/Bos_taurus.ARS-UCD1.2.99.gff3.gz).

1.2.7. Analyses phylogénétiques

Afin de déterminer l'haplotype de chacun des échantillons, j'ai réalisé un arbre phylogénétique en joignant à mon jeu de données 33 échantillons présents dans la littérature et disponibles dans GenBank (https://www.ncbi.nlm.nih.gov/nuccore/ , annexe 2) et dont l'haplotype a déjà été identifié.

J'ai aligné ces séquences entre elles avec Clustal O v1.2.4. (Madeira *et al.*, 2019), puis j'ai utilisé MegaX (Kumar *et al.*, 2018) pour contrôler « manuellement » l'alignement multiple obtenu. J'ai ensuite utilisé DnaSP version 6.12.1 (Rozas *et al.*, 2017) pour générer un fichier ne comprenant que les positions présentant des polymorphismes parmi mon jeu de données.

[6] Par exemple, AGA et AGG codent pour l'arginine dans l'ADN nucléaire et un codon stop sur le génome mitochondrial des vertébrés.

L'arbre phylogénétique a été construit à partir des polymorphismes de 508 séquences nucléotidiques selon la méthode utilisée par Bro-Jørgensen *et al.* (2018), en utilisant la méthode du maximum de vraisemblance et le modèle de Tamura-Nei (Tamura et Nei, 1993) avec le logiciel MEGAX (Kumar *et al.*, 2018). Un arbre consensus a été déduit à partir d'un bootstrap de 500 répétitions (Felsenstein, 1985). La taille de l'arbre consensus ne permet pas sa présentation dans le manuscrit de thèse, mais peut être mis à disposition sur un espace de stockage sur demande. L'enracinement de l'arbre phylogénétique a été fait sur le groupe d'échantillons porteurs d'un haplotype I. L'arbre initial a été obtenu en appliquant la méthode de *Neighbour-joining*. Les positions avec des trous de séquençage ont été éliminées et l'analyse a porté sur 1270 positions représentant 8% (1270/16 338) du génome mitochondrial.

J'ai également réalisé une carte de réseau génétique (figure 15, 1.2.3.) avec le logiciel Popart (Leigh et Bryant, 2015), avec le modèle *Median Joining Network* afin de contrôler l'attribution des haplogroupes des échantillons. J'ai utilisé l'outil DIVEIN (Deng *et al.*, 2010) pour générer une séquence consensus des génomes mitochondriaux de tous les individus appartenant à l'haplogroupe T3 pour ensuite calculer le nombre moyen de polymorphismes séparant chaque génome de la séquence consensus en prenant en compte les trous de séquence.

1.3. Résultats et discussion

1.3.1. Les lignées dans la généalogie

Dans une première étape, j'ai étudié la distribution des lignées dans la population à partir des données de pédigrée. Le tableau ci-dessous présente quelques indicateurs de structure de population dans les sept races les plus représentées dans notre jeu de données (les six grandes races et une race d'extension régionale, l'Abondance).

Ce tableau illustre tout particulièrement les disparités entre voie mâle et voie femelle et entre races laitières et allaitantes. Même si le développement du transfert embryonnaire a permis d'augmenter le nombre de descendants par femelle, cette technologie n'est utilisée que sur les femelles de haute valeur génétique et le nombre moyen de descendants reste bien plus élevé par père que par mère. L'IA permet d'augmenter considérablement le nombre de descendants d'un père, tandis que les femelles ont souvent une carrière relativement courte, avec un taux de renouvellement du troupeau assez élevé en France (31% en moyenne sur le troupeau laitier contre 20% sur le troupeau allaitant, www.reproscope.fr (Idele)). L'insémination animale permet la diffusion des taureaux de haute valeur génétique, ce qui se traduit par l'utilisation massive de certains taureaux. L'écart-type du nombre de descendants par père est très élevé et révèle la disparité entre les quelques taureaux « stars » massivement utilisés et les autres taureaux. Cependant, avec les améliorations permises par la sélection génomique, maintenir la diversité génétique est devenu un objectif important de la sélection et gestion des accouplements, ce qui se traduit par un plus grand nombre de taureaux proposés depuis une dizaine d'années.

Nous pouvons également observer un nombre moyen de descendants par femelle plus élevé dans les races allaitantes, ce qui s'explique par une plus longue carrière. Dans les races allaitantes, il n'est pas rare que des femelles de plus de 10 ans soit encore en production tandis qu'une grande partie des femelles laitières sont déjà réformées avant l'âge de 5 ans.

	Abondance	**Normande**	**Montbéliarde**	**Holstein**	**Limousine**	**Charolaise**	**Blonde d'Aquitaine**
Nombre total d'animaux	133 125	1 243 834	2 068 463	9 590 353	5 203 511	6 742 863	2 249 809
Animaux donc les deux parents sont connus	99 711	871 873	1 669 357	7 301 931	1 199 543	1 664 163	708 546
Nombre d'ancêtres mâles	113	121	351	1074	1 080	1 401	897
Nombre d'ancêtres femelles	12 262	90 594	157 495	688 655	122 662	183 333	85 899
Nombre moyen de descendants par mère (**toutes années**)	2,69 ±2,02	2,33 ±1,58	2,64 ±1,91	2,29± 1,49	3,86 ±2,95	3,45±2,66	3,37 ±2,67
Nombre moyen de descendants par père (**toutes années**)	65,74 ±439	225,07 ±2097	127,48 ±1908	160,22 ±2347	60,24 ±384	48,64 ±426	68,27 ±555
Nombre moyen de descendants par mère	2,01 ±1,2	1,88 ±1,12	1,91 ±1,17	1,79 ±1,03	2,36 ±1,38	2,28 ±1,37	2,2 ±1,36
Nombre moyen de descendants par père	89,11 ±246	425,72 ±1141	217,25 ±1178	283,96 ±1819	51,46 ±195	52,64 ±374	60,45 ±344
Nombre moyen de générations remontées (voie femelle)	5,23	8,04	7,21	5,16	5,17	5,83	4,57
Nombre moyen de générations remontées (voie mâle)	6,83	10,78	9,79	6,68	11,53	14,44	9,78
Proportion d'animaux avec strictement plus de 5 générations de connues (voie femelle)	40,08	76,8	68,96	43,95	43,53	38,86	32,25

Tableau 4 : Description des structures des races et des lignées sur les animaux nés entre 2015 et 2019 hormis lorsque « toutes années » est spécifié.

De manière générale, nous pouvons observer que le nombre de générations connues est plus important sur la voie mâle que sur la voie femelle, ce qui parait à première vue contre-intuitif puisqu'il semble plus évident d'identifier la mère d'un veau à la naissance, que son père. Cependant, cela s'explique car, au début, les efforts de la sélection se sont principalement concentrés sur les mâles, qui permettaient d'obtenir rapidement plus de descendants et d'augmenter le progrès génétique global de la population plus rapidement pour un nombre d'individus sélectionnés faible. Les parents des mâles d'IA doivent être connus pour que ceux-ci soient autorisés à la reproduction. De plus, les taureaux sélectionnés descendent bien souvent de taureaux eux-mêmes gardés dans le schéma de sélection des entreprises de sélection, avec un meilleur suivi des accouplements et donc une déclaration des parents plus souvent renseignée. L'écart entre le nombre moyen de générations connues moyen sur les voies mâle et femelle est beaucoup plus marqué chez les races allaitantes que chez les races laitières.

Le nombre moyen de générations remontées sur la voie mâle est plus important dans les races allaitantes que laitières. Cela peut s'expliquer par l'utilisation plus importante de taureaux en monte naturelle, qui sont utilisés dès la puberté (vers 2 ans). L'utilisation du testage sur descendance pour la majorité des taureaux laitiers jusqu'à l'avènement de la sélection génomique a eu pour effet de rallonger l'intervalle de générations et donc de diminuer le nombre de générations connues depuis le début de l'enregistrement des généalogies. Chez la Holstein, le nombre moyen de générations remontées est plus faible que dans les autres races laitières, ce qui pourrait s'expliquer par le caractère international de la race et la gestion de bases de données. La généalogie des animaux étrangers utilisés n'est pas forcément entièrement renseignée.

Les données généalogiques peuvent parfois être faussées par des choix de gestion ayant été effectués à la création de la base de données de recensement des animaux. En race Charolaise, nous avons répertorié plusieurs ancêtres femelles fictifs. Par exemple, dans le cas d'une lignée, la femelle FR000S288888 est indiquée comme mère de 154 descendants, tous nés en 1950 alors que la technique de transfert d'embryon n'existait pas encore. Les quatre lignées fictives identifiées n'ont donc pas été prises en compte dans la suite des analyses.

En conclusion, le nombre d'ancêtres femelles est beaucoup plus important que le nombre d'ancêtres mâles, ce qui laisse présupposer une plus grande diversité génétique au niveau du génome mitochondrial que pour le chromosome Y. Par ailleurs, l'étude des voies de transmission uniparentale est limitée par l'ancienneté et l'exhaustivité des informations renseignées dans les bases de données nationales.

1.3.2. Diversité mitochondriale des séquences

Nous avons séquencé 384 individus appartenant à des races différentes et extrait la séquence mitochondriale de 94 séquences entières de génomes déjà présents dans les bases de données de l'équipe G2B. Le tableau ci-dessous présente pour chaque race le nombre d'individus séquencés ainsi que le nombre de descendants minimum et maximum de chaque lignée.

Race	Abréviation	Nombre d'individus	Nombre de descendants par lignée	Nombre total descendants des lignées séquencées
Abondance	ABO	11	[266-1646]	8321
Aubrac	AUB	11	[4-8129]	44 301
Brune	BRU	9	[698-2294]	10 530
Salers	SAL	3	[20-1521]	1541
Bazadaise	BAZ	2	[345-434]	779
Bretonne	BPN	6	[12-879]	2 283
Tarentaise	TAR	13	[4-1192]	10 895
Limousine	LIM	72	[1-6648]	141 132
Simmental	SIM	13	[1-1671]	12 421
Charolaise	CHA	68	[145-15297]	286 098
Rouge des Prés	RDP	2	[296-296]	296
Montbéliarde	MON	61	[91-5459]	100 743
Normande	NOR	48	[973-5426]	77 463
Vosgienne	VOS	9	[210-432]	2 881
Rouge Flamande	RFL	2	[145-207]	352
Holstein	HOL	73	[3-18014]	420 543
Parthenaise	PAR	14	[98-1741]	16993
Gasconne	GAS	4	[210-371]	895
Blonde d'Aquitaine	BAQ	57	[104-5804]	117 594
	Total	**478**		

Tableau 5 : Description des races des échantillons séquencés

L'analyse de ces 478 génomes a permis d'identifier un total de 1056 variants (1021 SNP, 45 Indels) par rapport à la séquence de référence du génome mitochondrial. Parmi ces 1056 variants, j'ai identifié 185 variants faux-sens et 566 variants synonymes. La plus grande partie des polymorphismes n'altèrent donc pas la séquence protéique. L'absence de variants délétères (gain d'un codon stop, décalage du cadre de lecture, etc.) pourrait en partie s'expliquer par le goulot d'étranglement ayant lieu dans les cellules germinales primordiales éliminant une partie des ADN mitochondriaux porteurs de variants délétères (cf. chapitre 1, 3.3.1., Melvin et Ballard, 2017).

Comme attendu, la région de D-loop est la plus variable avec 215 variants, soit environ 14% des variants alors que cette région ne représente que 5% de la taille du génome mitochondrial bovin. Chaque individu possède entre 12 et 30 variants par rapport à la séquence de référence (cf. figure 14) et plus de 50% de ces variants sont individu-spécifique, ce qui est cohérent avec l'histoire des populations, à savoir un fort goulot d'étranglement au moment de la domestication induisant une

perte drastique de variabilité génétique, suivi par une augmentation très importante de la population domestiquée favorisant la conservation des mutations *de novo* qui se sont produites dans ce laps de temps. Le tableau 6 présente le nombre de variants portés par 1 à 478 individus.

Nombre de variants	Nombre d'individus porteurs
607	1
208	2
199	3-9
25	10-34
3	181-284
14	460-478

Tableau 6 : Variants portés par individu

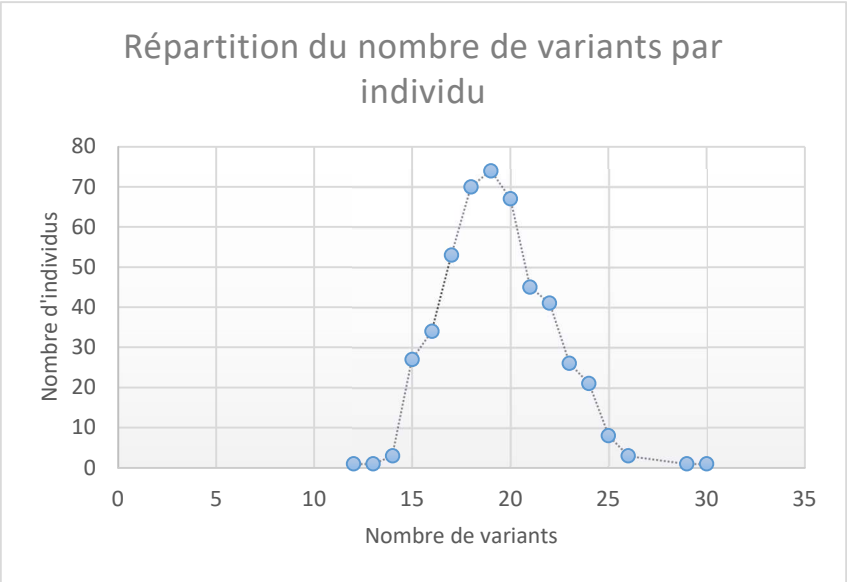


Figure 14 : Répartition du nombre de variants par individu

Conformément aux études antérieures (Achilli *et al.*, 2008; Kantanen *et al.*, 2009; Lenstra *et al.*, 2014), l'analyse phylogénétique que j'ai menée avec le logiciel MEGAX sur l'intégralité du génome mitochondrial de nos 478 échantillons a révélé que la presque totalité d'entre eux appartiennent à l'haplogroupe T3 (n = 469/478). Les autres haplogroupes représentés sont par ordre décroissant le T2 avec huit individus (3 en race Montbéliarde, 1 Brune des Alpes, 1 Blonde d'Aquitaine, 2 Charolaise et 1 Limousine), le T1 avec deux représentants (en races Tarentaise et Bazadaise) et le T5 avec un individu unique de race Montbéliarde.

La séquence de référence utilisée aujourd'hui provient de l'ADN d'un bovin coréen et est légèrement éloignée de celle des T3 que nous avons observés dans notre panel d'individus. J'ai donc créé une séquence consensus à partir de l'ensemble des animaux porteurs de l'haplogroupe T3 séquencés dans notre étude afin d'estimer le nombre de polymorphismes séparant chacun des échantillons T3 de ce consensus. Le nombre de polymorphismes entre chaque individu porteur de T3 et la séquence consensus créée est d'environ 5,3 polymorphismes, avec un maximum de 11 polymorphismes.

L'étude du réseau phylogénétique (figure 15, page suivant) nous montre qu'il n'y a pas de regroupements par race. La différenciation entre les haplogroupes est beaucoup plus ancienne que la création des races. Néanmoins, nous aurions pu nous attendre à trouver certains haplogroupes et sous-haplogroupes rares dans une seule race. Le faible nombre de lignées séquencées explique peut-être que nous n'observions pas de regroupement de lignées proches dans une même race.

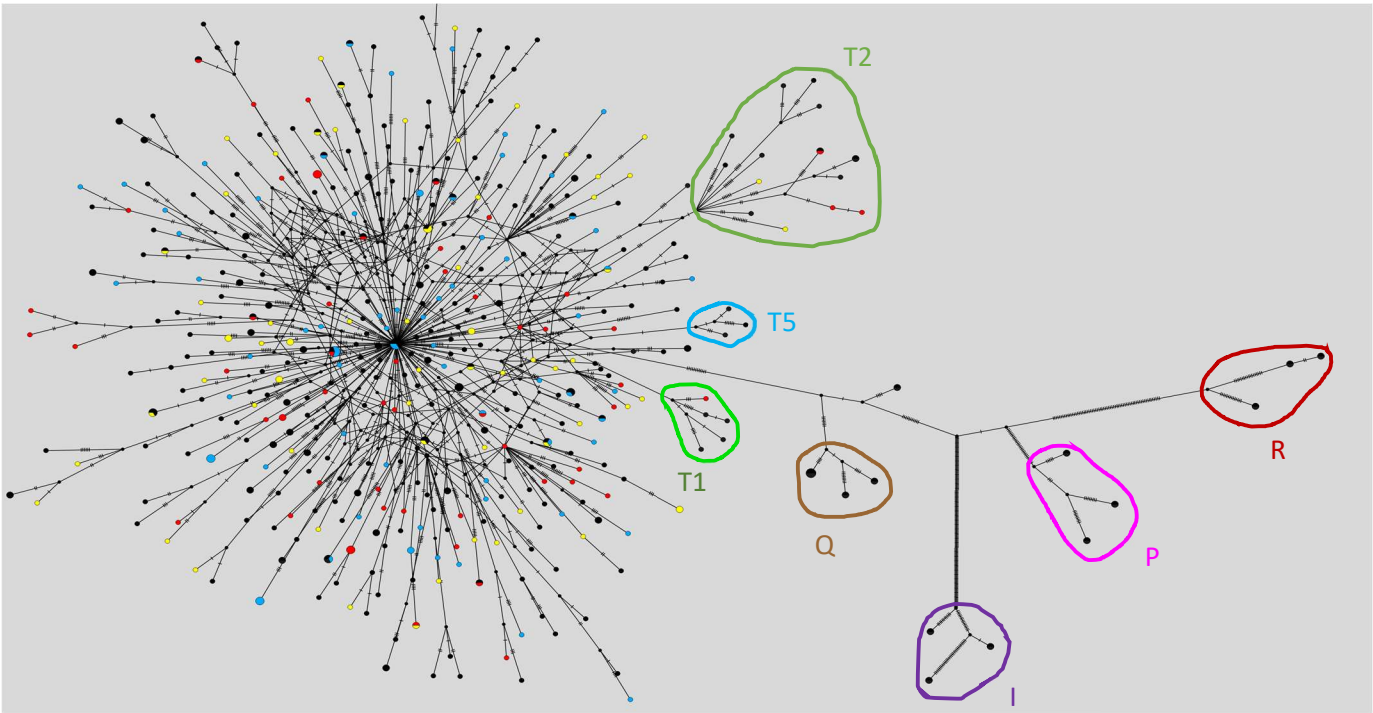


Figure 15 : Représentation en réseau des échantillons.

Les points représentent chacun un individu. En rouge sont représentés les échantillons de race Montbéliarde, en bleu de race Holstein, en jaune de race Charolaise. Les haplogroupes différents du T3 sont entourés et légendés.

Dans leur étude de 2010 réalisée sur 1747 animaux italiens, Bonfiglio *et al.* identifiaient des animaux porteurs des haplotypes T1, T2 et T5 respectivement à une fréquence de 4,2, 5,0 et 0,7%, contre 88,6% d'animaux porteurs de T3. Nous observons également une prédominance de l'haplotype T3 ainsi que quelques animaux porteurs de T1, T2 et T5 à une fréquence plus faible.

Nous avons identifié deux lignées T1 en race Tarentaise et Bazadaise. En Europe, l'haplotype T1 est principalement représenté dans le sud et pourrait provenir des échanges méditerranéens avec les bovins Africains (Lenstra *et al.*, 2014). La race Bazadaise proviendrait d'un croisement d'animaux ibériques et aquitains. Le type ibérique, provenant de l'invasion des Ibères, serait à l'origine des races pyrénéennes (La Race Bazadaise, 2005). La race Tarentaise (appelée aussi Tarine) est originaire de la vallée de la Tarentaise en Savoie, non loin d'Albertville, mais elle doit néanmoins ses origines à des bovins bruns indo-asiatiques. Cette race s'est développée dans toute la Savoie, puis sur l'ensemble des Alpes (Collectif des Races des Massifs). Ainsi, les haplotypes T1 dans les races Bazadaise et Tarentaise pourraient provenir d'échanges avec les populations de bovins hispaniques et italiens. Nous avons également identifié des porteurs de l'haplotype T2 dans plusieurs races : Montbéliarde, Limousine, Charolaise et Brune des Alpes. Cet haplogroupe a déjà été identifié à travers une grande partie de l'Europe, et également en France (cf. figure 11, partie 1.1.). Enfin, seule une lignée T5 a été identifiée, en race Montbéliarde. L'haplotype T5 est le plus rare des sous-clades du macro-haplogroupe T partagé par les races européennes et du Proche-Orient (Achilli *et al.*, 2008). Malheureusement, la lignée de l'échantillon T5 (provenant des séquences de l'équipe G2B) n'a plus de descendants depuis 1999 et il est donc impossible d'étudier son effet sur les caractères de production.

Enfin, il est intéressant de noter que parmi les six races les plus représentées dans notre panel de séquences complètes de génomes mitochondriaux, c'est en Montbéliarde, une race du rameau Pie rouge des montagnes, (Denis, 2010) issue de métissage entre variétés franc-comtoises et suisses durant la période du XVIIIe au XIXe siècles que nous avons observé le plus d'haplogroupes différents : T3, T2 et T5. Il sera intéressant de vérifier dans des études futures si cette variabilité est le fruit de ce métissage, et/ou de sa position à l'Est de la France, à la croisée des voies Danubienne et Méditerranéennes de pénétration de la culture néolithique en Europe.

Afin d'étudier la variabilité des haplotypes et sous-haplotypes mitochondriaux à l'échelle raciale, j'ai ajouté 464 nouveaux marqueurs génétiques à la dernière version de la puce Illumina EuroGMD, qui a été mise en production en mai 2019. Ces derniers correspondent à l'ensemble des variants observés chez au moins deux des 478 génomes mitochondriaux analysés. Une analyse des premiers résultats sera présentée dans la partie 3 de ce chapitre.

En conclusion, l'étude des génomes mitochondriaux de 478 bovins de races françaises a permis de montrer la présence d'un grand nombre de variants individu-spécifiques. Nous avons pu observer qu'il existe peu de variabilité entre les différents haplotypes et que ceux-ci appartiennent majoritairement au groupe T3.

Résultats clés de cette étude :

- Il existe peu de variabilité des haplotypes mitochondriaux dans les races françaises. La plupart des individus séquencés appartiennent à l'haplogroupe T3.
- Nous avons identifié quelques individus porteurs de l'haplotype T1, T2 ou T5.
- Nous n'avons pas identifié de séquences appartenant aux haplotypes plus rares comme Q et R.
- Environ 14% des variants identifiés sur nos séquences sont situés sur la région de D-loop.

2. Etude 2 : Recherche de bovins modernes porteurs d'haplogroupes mitochondriaux d'aurochs et de zébu

En attendant d'obtenir les résultats du séquençage des lignées maternelles présenté dans l'étude 1, au tout début de ma thèse, j'ai dessiné les amorces de marqueurs mitochondriaux à partir des séquences présentes dans les bases de données publiques NCBI. Trois de ces marqueurs m'ont permis de distinguer des individus non porteurs de l'haplotype T (dits « non-T » dans la suite du manuscrit) parmi plus de 160 000 individus de 12 races bovines françaises génotypés pour la sélection génomique. Le séquençage du génome mitochondrial d'un petit nombre d'individus déterminés non-T sur génotypage a révélé l'existence d'haplotypes mitochondriaux des groupes P, Q et I dans la population bovine française dont nous avons étudié la variabilité.

2.1. Méthode

2.1.1. Détermination des haplotypes non-T / T et criblage de la population génotypée pour la sélection génomique

Dans une étude préliminaire, j'ai aligné plus de 300 séquences entières de génome mitochondrial présentes dans la base de données NCBI (https://www.ncbi.nlm.nih.gov/nuccore/; dernier accès le 23/06/2017), et choisi des variants permettant de séparer cet ensemble de séquences en plusieurs groupes (travaux non présentés) que j'ai ajoutés aux puces à SNP Illumina EuroG10k v8 puis EuroGMD. Trois de ces marqueurs, décrits dans le tableau ci-dessous m'ont permis de distinguer le macro-haplogroupe T des autres haplogroupes (P, Q, R et I), appelés « non-T ».

Position	Allèle T	Allèle non-T
5503	C	T
11000	C	T
12749	T	C

Tableau 7 : Marqueurs permettant de distinguer les génomes mitochondriaux appartenant au macro-haplogroupe T des autres haplogroupes. *Les positions sont relatives à la séquence AY526085.*

En avril 2019, j'ai analysé une population de 157 042 individus génotypés provenant de 12 races, et j'ai utilisé les trois marqueurs pour identifier des animaux porteurs d'un haplotype non-T (2 allèles T et un non déterminé ou 3 allèles non-T) et des individus avec un haplotype « douteux » (1 allèle non-T, 2 allèles T). Le nombre de génotypages par race est présenté dans le tableau 8.

	ABO	BRU	TAR	LIM	SIM	CHA	MON	NOR	VOS	HOL	PAR	BAQ
Nombre de génotypages	2134	1487	1074	369	197	4920	54010	10712	410	81048	480	904

Tableau 8 : Nombre de génotypages par race sur les puces EuroGMD et EuroG10k v8 en avril 2019

2.1.2. Amplification et séquençage du génome mitochondrial d'individus déterminés non-T et douteux

A des fins de vérification, j'ai fait amplifier par PCR et séquencer avec la méthode Sanger (Eurofins MWG, Hilden, Allemagne) un segment de 1138 pb de la région D-loop pour ces 69 animaux. Les PCR ont été réalisées avec la GoTaq Flexi de Promega (Charbonnières-les-Bains, France) sur un Mastercycler Eppendorff Nexus (Montesson, France). Les amorces utilisées et les conditions de PCR sont présentées en annexe 3 (couple 1, F et R). L'étude des séquences D-loop a validé le statut non-T des 60 animaux identifiés comme tels par génotypage sur puce, tandis que les neuf animaux douteux se sont révélés appartenir à l'haplogroupe T. J'ai ensuite réalisé un alignement multiple avec Clustal o v1.2.4. (Madeira *et al.*, 2019). J'ai alors sélectionné 28 individus porteurs de séquences représentatives de l'échantillon et j'ai procédé à l'amplification et au séquençage de l'ensemble de leur génome, avec l'aide d'une technicienne Allice, Marie-Christine Deloche, selon la même approche, afin d'identifier l'haplogroupe de chacun des échantillons et d'étudier la variabilité au sein de ces haplogroupes.

Pour ce faire, nous avons amplifié 10 segments supplémentaires, partiellement chevauchants (voir annexe 3, et supplementary table 3 de Achilli *et al.*, 2008) que nous avons fait séquencer à partir des deux extrémités.

Les amorces utilisées pour l'amplification sont celles présentées dans le tableau en annexe 3 (F et R) et ont été dessinées par Achilli *et al.*, 2008. Des amorces intermédiaires (amorces « t » en annexe 3) ont également été utilisées pour le séquençage lorsque la lecture du fragment de PCR à partir des amorces F er R ne permettait pas de couvrir l'ensemble d'un amplicon. Elles ont été dessinées avec Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/ , (Untergasser *et al.*, 2012)) et sont présentées dans le tableau en annexe 3.

2.1.3. Analyses phylogénétiques

En plus des séquences obtenues dans cette étude, j'ai sélectionné un panel de 34 génomes mitochondriaux bovins accessibles dans GenBank (https://www.ncbi.nlm.nih.gov/nuccore/) ou provenant des séquences de l'étude 1 (tableau 9 et annexe 4) représentant les principaux haplogroupes T, I, Q et R et l'ensemble des génomes mitochondriaux complets disponibles appartenant à l'haplogroupe P. Les échantillons issus d'ADN ancien sont présentés dans le tableau 9 tandis que les échantillons modernes sont présentés en annexe 4. J'ai également sélectionné un génome mitochondrial de *Bison bonasus* (identifiant pubmed = KX898013) comme exogroupe pour l'enracinement de l'arbre phylogénétique. J'ai aligné ces séquences entre elles avec Clustal O v1.2.4. (Madeira *et al.*, 2019), puis j'ai utilisé MegaX (Kumar *et al.*, 2018) pour contrôler « manuellement » l'alignement multiple obtenu.

Les haplogroupes ont été assignés en se basant sur les variants discriminants identifiés dans de précédentes études de la littérature (Troy *et al.*, 2001; Achilli *et al.*, 2008, 2009; Bonfiglio *et al.*, 2010)

GenBank ID	Haplogroupe	Nature du prélèvement et origine géographique	Datation	Source
KX898013	*Bison bonasus* 2	Restes osseux (métacarpe) nord du Caucase, Kudaro 3	38500-37000 BP	(Massilani *et al.*, 2016)
MF169213 CMLIX	P	Corne à boire, Allemagne	XVe siècle	(Bro-Jørgensen *et al.*, 2018)
MF169211 10537	P	Corne à boire, Muséum National du Danemark, possible appartenance au roi de Norvège Håkon VI	XIVe siècle	(Bro-Jørgensen *et al.*, 2018)
MF169212 10543	P	Corne à boire, Muséum National du Danemark, origine inconnue	XVe siècle	(Bro-Jørgensen *et al.*, 2018)
Ch22	T	Restes osseux (métatarse), Çatalhöyük, Turquie	7590-7570 BP	(Verdugo *et al.*, 2019)
Gyu2	Q	Restes osseux (crâne), Arménie	7076 BP	(Verdugo *et al.*, 2019)
GU985279	P	Restes osseux (humérus), Derbyshire, Angleterre	6738 BP	(Edwards *et al.*, 2010)
JQ437479	P	Restes osseux (vertèbre) Pologne	1500 BP	(Zeyland *et al.*, 2013)

Tableau 9 : Echantillons anciens disponibles dans les bases publiques utilisés pour la réalisation de l'arbre phylogénétique

L'arbre phylogénétique a été construit à partir de 63 séquences nucléotidiques en utilisant la méthode du maximum de vraisemblance et le modèle de Tamura-Nei (Tamura et Nei, 1993) avec le logiciel MEGAX (Kumar *et al.*, 2018). L'arbre consensus présenté en figure 16 a été déduit à partir d'un bootstrap de 500 répétitions (Felsenstein, 1985). Les branches correspondant aux partitions reproduites dans moins de 50% des répétitions bootstrap sont regroupées. Le pourcentage d'arbres répliqués dans lesquels les taxons associés se sont regroupés dans le test bootstrap (500 réplicats) est indiqué à côté des branches. Les arbres initiaux ont été obtenus en appliquant la méthode *neighbour-joining* à une matrice de corrélations estimées deux à deux. Les positions non couvertes par l'ensemble des séquences ont été éliminées et l'analyse a porté sur 13 065 positions représentant 80% (13 065/16 338 pb) du génome mitochondrial.

La carte du réseau génétique a été réalisée avec le logiciel Popart (Leigh et Bryant, 2015), avec le modèle *Median Joining Network* et contient les mêmes séquences que l'arbre phylogénétique, sauf les séquences d'un échantillon chimérique HOL18 et du *Bison bonasus*.

J'ai utilisé l'outil DIVEIN (Deng *et al.*, 2010) pour générer une séquence consensus des génomes mitochondriaux d'une partie des individus avec un haplotype P pour ensuite calculer le nombre moyen de polymorphismes séparant chaque génome du consensus en prenant les trous de séquence en compte.

2.2. Résultats

2.2.1. Utilisation des génotypages pour identifier des lignées rares

Le tableau ci-dessous présente la proportion de lignées T et non-T dans la population des animaux nés depuis 2015 et génotypés entre juin 2018 (date de mise en service) et novembre 2019. Ne sont considérés pour ces comptages, que les animaux typés pour lequel nous avons un génotype pour au moins deux des trois marqueurs discriminants. Les lignées sans aucun descendant génotypé à la date de l'étude ne peuvent pas être prises en compte.

Race	Nombre d'animaux génotypés	Nb animaux T	Nb animaux non T	Nb animaux douteux	Nb de lignées T	Nb de lignées non T	Nb de lignées douteuses
Normande	10712	10696	13	3	4673	5	2
Montbéliarde	54010	53936	65	9	18643	32	4
Holstein	81048	80775	225	48	34109	87	37
Charolaise	4920	4911	4	5	1951	4	4
Limousine	369	366	3	0	117	2	0
Bonde d'Aquitaine	904	894	2	8	604	2	7
Abondance	2134	2123	11	0	1012	7	0
Brune des Alpes	1487	1484	0	3	603	0	3
Tarentaise	1074	1074	0	0	449	0	0
Simmental	197	197	0	0	117	0	0
Vosgiennes	410	410	0	0	190	0	0
Parthenaise	480	480	0	0	303	0	0

Tableau 10 : Etude réalisées sur les animaux génotypés sur la puce LD v8 ou MD v1, dernière mise à jour en novembre 2019

J'ai ainsi identifié des individus porteurs d'un haplotype non-T dans sept races, dont un nombre important dans les races Normande, Montbéliarde, Holstein et Abondance. Un plus petit nombre d'animaux non-T a été identifié dans les trois grandes races allaitantes, ce qui s'explique en partie par une plus faible utilisation du génotypage dans ces races. Dans le cas de la race Limousine, une grande partie des animaux sont génotypés par un laboratoire n'ayant pas accès à la partie privée de la puce qui inclue nos trois marqueurs. Avec l'accumulation continue de nouveaux génotypages, le nombre de lignées T identifiées devrait augmenter.

Parmi tous les individus identifiés précédemment, nous avons pu récupérer l'ADN restant des échantillons utilisés pour le génotype au laboratoire Labogena, pour un petit nombre d'individus non-T et douteux afin de séquencer dans une première étape une portion de la région D-loop et de valider l'appartenance de ces animaux à des haplogroupes distincts de T. Ce premier séquençage nous a également permis d'étudier le cas de lignées ambigües, possédant deux marqueurs de type T et un marqueur de type non-T.

2.2.2. Un premier séquençage réalisé sur la région D-loop

Sur la séquence de référence AY526085 du génome mitochondrial que j'ai utilisée, la séquence de la région D-loop couvre 910 pb des positions 15 792 à 16 338 pb (fin de la séquence) et 1 à 363 pb. La région que nous avons séquencée couvre 77% de la région D-loop (des positions 15 782 à 16 338 et 1 à 158 pb).

Le séquençage de la région D-loop des neuf génomes mitochondriaux classés comme ambigus a révélé leur appartenance à l'haplotype T. Ce résultat nous démontre la nécessité de considérer comme non-T uniquement les individus avec au moins deux allèles « non-T » pour déterminer avec fiabilité le statut non-T du génome mitochondrial d'un individu.

La qualité d'un échantillon ADN parmi les 60 individus non-T n'était pas suffisante pour permettre une amplification correcte. Ce premier séquençage nous a permis de distinguer trois groupes parmi les 59 génomes mitochondriaux non-T : I, P, Q. Il est important de noter que dans la région D-loop, Q partage des allèles spécifiques avec certains variants de T et aussi avec certains variants de P, et c'est la combinaison sur 13 variants informatifs qui permet de différencier Q des autres haplogroupes.

Nous avons pu identifier deux porteurs d'un haplotype I en race Holstein et 21 porteurs d'un haplotype P dans diverses races : 1, 4, 4 et 12, respectivement, dans les races Blonde d'Aquitaine, Abondance, Montbéliarde et Holstein. Enfin, nous avons mis en évidence l'appartenance à l'haplogroupe Q de 17 individus Holstein, 17 Montbéliardes et 2 Abondances.

Le tableau 11 ci-après détaille les allèles de chaque échantillon pour les positions polymorphiques de notre jeu de données. Nous pouvons y observer quatre haplogroupes : I, P, Q et T.

	15807	15818	15881	15893	15921	15951	15953	15959	15994	16019	16022	16043	16049	16050	16051	16053	16055	16057	16058	16074	16079	16082	16084	16085	16102	16109	16113	16116	16117	16119	16121	16122	16126*	16130	16131	16135	16137	16138	16139	16141	16143	16147	16195	16196	16199*	16200	16229	16231	16247	16248	16250	16255	16264	16300	16301	8	62*	62*	106	119
BRS	G	A	C	T	G	T	C	G	A	T	G	A	C	C	T	T	T	G	C	T	A	G	C	T	G	T	T	T	G	T	G	T	C	T	T	T	T	T	C	T	A	T	A	G	-	G	A	C	C	C	A	T	G	A	C	G	-	-	T	C
BIRS	.	G	.	.	.	C	G	A	G	.	A	.	T	.	.	.	.	A	T	C	.	A	T	.	A	C	C	C	A	C	A	C	-	C	.	.	C	C	.	C	-	C	.	A	G	A	G	.	T	T	.	.	.	G	T	A	-	-	C	.
I_GU256940	.	G	.	.	.	C	G	A	G	.	A	.	T	.	.	.	.	A	T	C	.	A	T	.	A	C	C	C	A	C	A	C	-	C	.	.	C	C	.	C	-	C	.	A	G	A	G	.	T	T	.	.	.	G	T	A	-	-	C	.
HOL8	.	G	.	.	.	C	G	A	G	.	A	.	T	T	.	.	.	A	T	.	.	A	T	.	A	C	C	C	A	C	A	C	-	C	.	.	C	C	.	C	-	C	.	A	G	A	G	.	T	T	.	.	.	G	T	A	-	-	C	.
HOL2	.	G	.	.	.	C	G	A	G	.	A	.	T	.	.	.	.	A	T	C	.	A	T	.	A	C	C	C	A	C	A	C	A	C	.	.	C	C	.	C	-	C	.	A	G	A	G	.	T	T	.	.	.	G	T	A	-	-	C	.
PRI_GU985279	.	.	.	.	.	C	G	.	G	C	.	.	T	.	C	.	.	.	T	C	.	.	.	C	.	.	.	.	.	.	?	C	-	.	.	.	.	.	.	C	.	.	.	.	-	.	.	T	.	.	.	C	A	.	T	.	-	-	.	.
MON9 ABO27	C	.	.	.	.	C	G	.	G	C	.	.	T	.	C	.	.	.	T	C	.	.	.	C	.	.	.	.	.	.	.	C	-	.	.	.	.	.	.	C	.	.	.	.	-	.	.	T	.	.	.	C	A	.	T	.	-	-	.	.
groupe B	.	.	.	.	.	C	G	.	G	C	.	.	T	.	C	C	.	.	T	C	.	.	.	C	.	.	.	.	.	.	.	C	-	.	.	.	.	.	.	C	.	.	.	.	-	.	.	T	.	.	.	C	A	.	T	.	-	-	.	.
HOL19	.	.	T	.	.	C	G	.	G	.	.	.	T	.	C	.	.	.	T	C	.	.	.	C	.	.	.	.	.	.	.	C	-	.	.	.	.	.	.	.	.	.	.	.	-	.	.	T	.	.	.	C	A	.	T	.	-	-	.	T
groupe A	.	.	.	.	.	C	G	.	G	.	.	.	T	.	C	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	C	-	.	C	.	.	.	.	.	.	.	.	.	-	.	.	T	.	.	.	C	A	.	T	.	-	-	C	.
PRI_JQ437479	.	.	.	.	.	C	G	.	G	.	.	.	T	.	.	.	.	.	T	C	.	.	.	C	.	.	.	.	.	.	.	C	-	.	.	.	.	.	.	.	.	.	.	.	-	.	.	T	.	.	.	C	A	.	.	.	-	-	.	.
P2_DQ124389	.	.	.	.	.	C	G	.	G	.	.	.	T	.	C	.	.	.	T	C	.	.	.	C	.	.	.	.	.	.	.	C	-	.	.	.	.	.	.	.	.	.	.	.	-	.	.	T	T	.	.	C	A	.	.	.	-	-	C	.
groupe C	.	.	T	.	.	C	G	.	G	.	.	.	T	.	C	.	.	.	T	C	.	.	.	C	.	.	.	.	.	.	.	C	-	.	.	.	.	.	.	.	.	.	.	.	-	.	.	T	.	.	.	C	A	.	.	.	-	-	.	T
HOL6	.	.	T	.	.	C	G	.	G	.	.	.	T	.	C	.	.	.	T	C	.	.	.	C	.	.	.	.	.	.	.	C	-	.	.	.	.	C	.	.	.	.	.	.	-	.	.	T	.	.	.	C	A	.	.	.	-	-	.	.
Q2_FJ971080	.	.	.	.	A	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	C	.	.	.	.	-	-	.	.
Q1_HQ184039	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	A	-	.	.	.	.	.	.	.	C	.	.	.	.	-	-	.	.
groupe D	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	C	.	.	.	.	-	-	.	.
h20	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	-	.	.	T	.	.	.	C	.	.	.	G	A	.	.
groupe E	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	-	.	.	T	.	.	.	C	.	.	.	.	-	-	.	.
h1 h21	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	C	.	.	.	.	-	-	.	.
h17	.	.	.	.	.	.	G	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	C	.	.	.	.	-	-	.	.
HOL24 h7	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	C	-	.	.	.	.	.	T	.	.	.	.	.	-	.	.	.	.	.	.	C	.	.	.	.	-	-	.	.
h13	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	-	.	.	C	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	C	.	.	.	.	-	-	.	.
m12 m11 h4	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	A	-	.	.	.	.	.	.	.	C	.	.	.	.	-	-	.	.
h8 h9 h18	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.
h5 h19 m15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.
h12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	G	.	-	.	.	.	.	.	.	.	.	.	.	.	-	-	C	.
b1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	-	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.
h10	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	G	.	.	.	.	.	-	-	.	.

Tableau 11 : Tableau répertoriant les mutations sur la région 1-158 et 15782-16338 du génome mitochondrial (sur D-loop).

*. Les « * » représentent des positions avec des insertions et les « . » des allèles identiique à la séquence de référence Bos taurus reference (BRS). BIRS est la séquence de référence en Bos indicus. Les haplotypes I, P, Q et T sont représentés respectivement en vert, bleu, gris et orange. Les échantillons au format RACEn° sont ceux qui ont été choisis pour un séquençage mitochondrial complet dans la suite de l'étude. Le détail des groupes et des races des échantillons est donné en annexe 5.*

2.2.3. Exploitation de séquences complètes

Le délai très court entre l'obtention des échantillons et la fin de la thèse ne m'a pas permis de finaliser le séquençage de certains fragments de séquences. Pour certains échantillons, notamment ceux appartenant à l'haplogroupe I, l'amplification de certains segments s'est révélée difficile et nous n'avions pas assez d'ADN pour finir le séquençage. Rappelons que nous avons récupéré pour chaque échantillon la quantité d'ADN restant après génotypage pour la sélection génomique. Malgré tout, j'ai pu construire un réseau phylogénétique permettant d'identifier l'haplogroupe de chacun des échantillons.

Le tableau ci-dessous retrace l'appartenance de chacun des échantillons.

	P	**Q**	**I**	**Non déterminé**	**Total**
Abondance	4	1			**5**
Montbéliarde	3	2			**5**
Holstein	12	1	2	1	**16**
Blonde d'Aquitaine	1				**1**
Total	**20**	**4**	**2**	**1**	**27**

Tableau 12 : Correspondance haplotype-échantillon

Au total, nous avons observé 340 variants chez ces individus par rapport à la séquence de référence sur 16 338 bases. Une séquence trop incomplète parmi les 28 séquences produites n'a pas été prise en compte dans les analyses.

Sur l'arbre présenté en figure 16, l'échantillon HOL18 semble se situer entre l'haplogroupe T et Q. Il est possible que malgré le soin que nous avons apporté à la traçabilité des échantillons ADN et des produits PCR à séquencer, une erreur d'échantillon se soit produite et qu'il s'agisse d'une chimère entre deux haplotypes. Un deuxième séquençage de l'ADN mitochondrial de cet individu sera donc à effectuer à partir d'un nouveau prélèvement sur le même individu pour vérifier ces résultats.

Hormis l'haplotype identifié uniquement en race Holstein, les haplotypes P et Q ont été retrouvés dans plusieurs races, Holstein, Abondance, Montbéliarde. Parmi les individus séquencés, nous avons identifié 2, 4 et 21 séquences appartenant respectivement aux haplogroupes I, Q, et P.

Le réseau phylogénétique (figure 17) représente l'ensemble des séquences produites pour ce projet, hormis l'échantillon chimérique HOL18. Le nombre de traits sur chaque branche représente le nombre de variants séparant deux individus. Cependant, ce chiffre n'est indiqué qu'à titre indicatif puisque certaines séquences ne sont pas complètes à ce jour (valable également pour certaines séquences P des bases publiques).

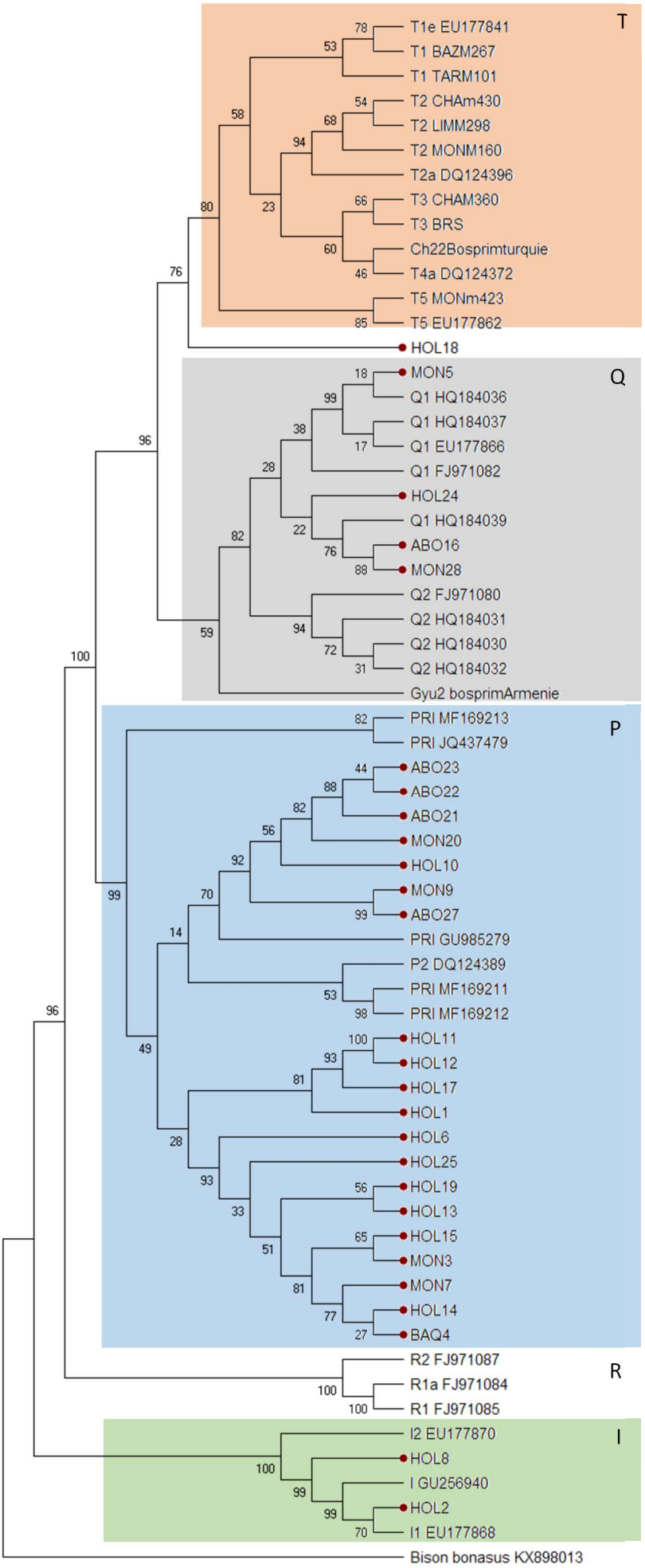


Figure 16: Arbre phylogénétique basé sur la méthode du maximum de vaisemblance.
Chaque groupe monophylétique est affecté de valeurs bootstrap, qui fournissent une estimation de la confiance dans la structure de clade de l'arbre. En rouge, les individus non-T séquencés dans cette étude.

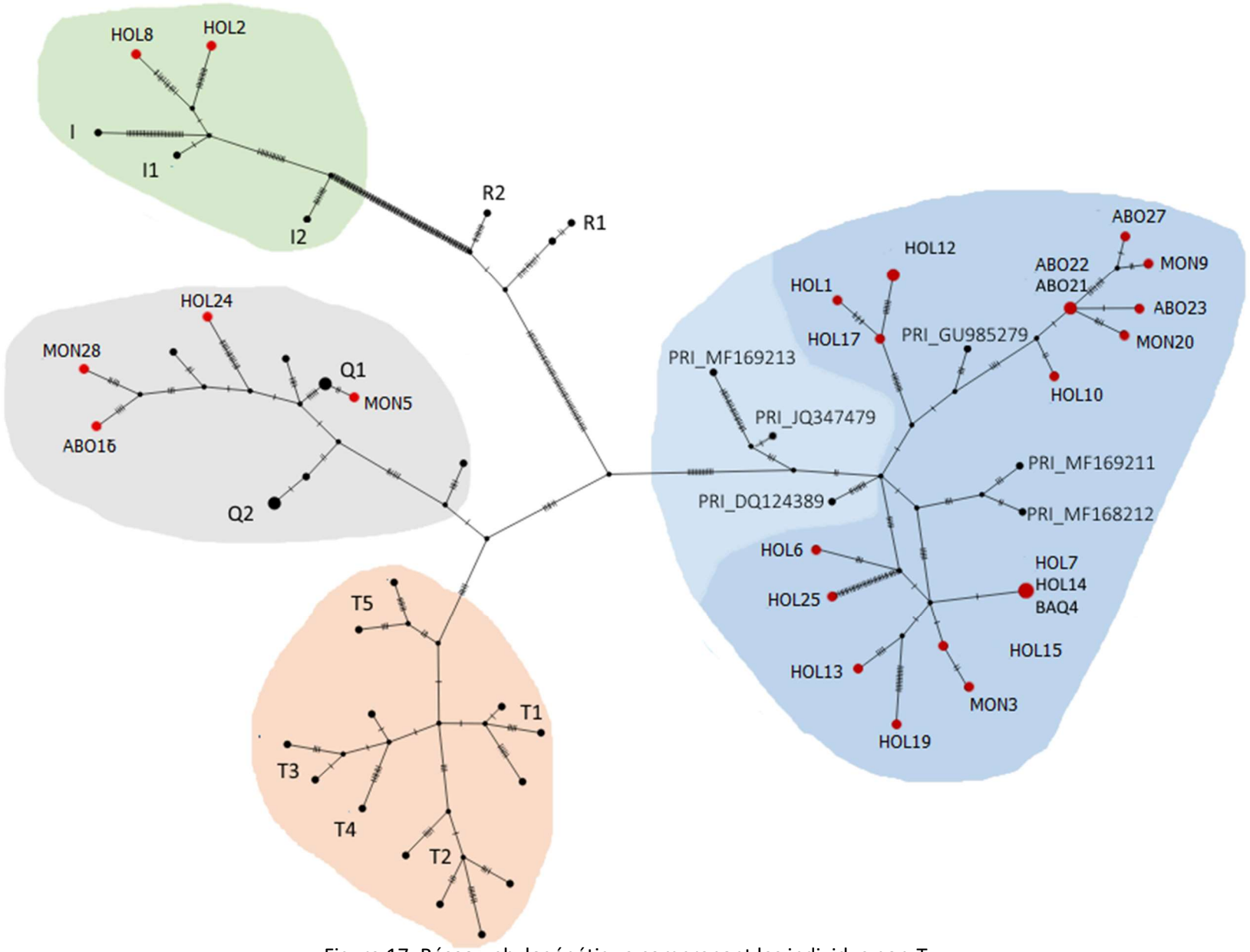


Figure 17: Réseau phylogénétique comprenant les individus non-T.

Le nombre de barres entre deux nœuds symbolisent le nombre de mutations séparant les deux séquences. Les individus non-T séquencés dans cette étude sont symbolisés par les nœuds rouges. L'haplogroupe I est représenté en vert, le P en bleu, le Q en gris et le T en orange. Les séquences appartenant à l'haplogroupe P et délimitées en bleu plus foncées ont été utilisées pour créer un consensus des séquences P.

Il est intéressant de noter que l'échantillon P2_DQ124389, issu d'un de bovin moderne coréen, diverge de l'ensemble des échantillons P identifiés dans notre étude, ce qui indique une divergence très ancienne entre les échantillons européens et asiatiques.

J'ai utilisé l'ensemble des séquences entourées en noir sur la figure 17 pour créer un consensus des séquences d'aurochs avec le programme DIVEIN. Ce consensus a été réalisé sur les positions ne comprenant pas de trous d'assemblage sur ces séquences, soit 15 135 pb. Le nombre de polymorphismes (hors trous de séquence) séparant chacun des échantillons au consensus varie de 8 à 37 (avec une moyenne de 16), indiquant une diversité importante à l'intérieur même de l'haplogroupe P. Pour rappel, le nombre de polymorphismes entre chaque individu T3 et la séquence consensus créée était d'environ 5,3 polymorphismes, avec un maximum de 11 polymorphismes (calculé sur l'entièreté de la séquence).

2.3. Discussion

2.3.1. Présence de lignées P, Q et I dans la population de bovins taurins modernes

Lors de cette étude, nous avons observé trois haplotypes différents du macro-haplogroupe identifié chez la plupart des *Bos taurus* domestiques modernes.

Nous avons mis en évidence la présence de lignées Holstein porteuses d'un haplotype mitochondrial de zébu. La présence de ces lignées dans la race Holstein pourrait provenir des Etats-Unis, carrefour de nombreux croisements au cours des derniers siècles. L'import de la génétique Holstein américaine par l'achat de semence mais aussi d'embryons est une pratique courante depuis les années 1990. La présence de lignées I pourrait également provenir d'Afrique. Le flux de gènes entre les *Bos taurus* et *indicus* a souvent été identifié dans les races africaines et du Proche et Moyen-Orient (Edwards *et al.*, 2007a). Des études d'introgression de gènes ont révélé qu'une partie du génome d'une grande partie des races européennes (hors races britanniques) dont la Holstein, provient des bovins taurins africains (Park *et al.*, 2015). Or il a été montré une contribution des *Bos indicus* très importante chez les populations bovines africaines, principalement par introgression mâle (MacHugh *et al.*, 1997; Bruford *et al.*, 2003). Une autre étude de Decker *et al.* (2014) a également identifié une part du génome nucléaire héritée des bovins africains et des zébus dans trois races italiennes Machininna, Chianinna et Romagnola, sans toutefois identifier de génome mitochondrial appartenant à l'haplotype I. L'introgression de zébu sur des bovins domestiques s'est principalement faite par voie mâle, et l'haplotype I n'a, à ma connaissance, pas été identifié chez des races européennes. Il est cependant possible que les lignées I n'aient à ce jour jamais été identifiés dans les races du bassin méditerranéen à cause d'une puissance insuffisante de détection. Nous avons déterminé la fréquence de non-T à 1,4‰ dans la population Holstein. Or, parmi les 31 lignées non-T Holstein séquencées sur la région D-loop, seule deux sont porteuses de l'haplotype I. Etant donné la fréquence extrêmement faible des lignées I, il est possible que celles-ci soient également présentes dans des populations Holstein étrangères mais aussi dans d'autres races. Nous n'avons pu détecter ces lignées que grâce à l'utilisation massive du génotypage et à l'opportunité que représente la mutualisation des outils pour les projets de recherche en génétique des populations et en sélection génomique au sein de notre équipe.

Dans notre échantillon d'animaux, nous avons observé quatre animaux porteurs d'un haplogroupe Q et appartenant à trois races laitières : Abondance, Montbéliarde et Holstein. En étudiant la fréquence de l'haplogroupe Q dans les populations modernes, l'une des premières hypothèses concernant son

origine l'assimilait à une population d'aurochs présente dans l'extrême sud de l'Europe (Beja-Pereira *et al.*, 2006). Cependant, il est également possible que la fréquence de l'haplotype Q ait été sous-évaluée dans des études n'étudiant qu'une partie de la région D-loop et ne couvrant pas les variants spécifiques de Q (Niemi *et al.*, 2015). En effet, comme indiqué précédemment, seule la combinaison de variants des haplotypes P et T permet d'identifier l'appartenance à l'haplogroupe Q. Une autre hypothèse a également été émise d'une domestication conjointe des haplogroupes Q et T (Achilli *et al.* 2009 ; Bonfiglio *et al.* 2010) au Moyen-Orient. Ces dernières années, de nombreux animaux porteurs de l'haplotype Q ont été identifiés dans des ossements datant du Néolithique dans différentes régions d'Europe y compris le nord de la Baltique (Niemi *et al.*, 2015; Scheu *et al.*, 2015a). Ces résultats mettent en évidence une complexité croissante de la répartition géographique de l'haplogroupe Q, et appuie l'hypothèse d'une domestication conjointe des lignées Q et T.

La découverte d'animaux modernes porteurs de l'haplotype P est tout particulièrement intéressante. A notre connaissance, hormis un individu de race Murbodner identifié par nos collaborateurs allemands (race autrichienne, séquence non publiée,) l'haplotype P n'avait été observé que dans des populations originaires de Corée et du Japon, sur la base de séquences de la région D-loop (Achilli *et al.*, 2008; Noda *et al.*, 2018). Près de la moitié des échantillons parmi 181 individus dans la race Japanese shorthorn appartenait à l'haplotype P (Noda *et al.*, 2018). En l'absence de données sur le reste du génome mitochondrial, les auteurs ont interprété cette découverte comme résultant de l'apport de génétique de bovins de la péninsule coréenne (d'où est originaire la culture japonaise) sans pouvoir vérifier le lien entre ces génomes mitochondriaux et ceux d'échantillons anciens européens appartenant également à l'haplogroupe P.

L'extrême rareté des haplotypes mitochondriaux portés par les aurochs européens contraste avec l'importance des flux de gènes nucléaires entre les aurochs et les bovins domestiqués identifiés dans la population moderne européenne. L'étude de la contribution génétique (à partir du génome nucléaire) du seul auroch européen entièrement séquencé à ce jour (haplotype P) à la population moderne européenne a montré un héritage significatif dans les populations irlandaises et britanniques, et plus particulièrement marqué dans des races locales et primitives comme la Highland ou la Dexter (Park *et al.*, 2015). Une autre étude de flux de gènes nucléaires de l'auroch britannique parmi des races européennes non françaises a montré que la fréquence d'allèles assimilés à l'auroch était plus élevée dans les races britanniques, et plus faible parmi les races de la péninsule ibérique (Anderung *et al.*, 2005; Upadhyay *et al.*, 2017). Ce contraste, et d'autres éléments liés à l'étude des haplogroupes du chromosome Y (cf. chapitre 3, étude 1) étayent l'hypothèse selon laquelle la majorité des flux de gènes entre aurochs et populations domestiquées aurait été réalisé par la voie mâle. De telles pratiques de retrempe pour apporter de la vigueur aux populations domestiques sont attestées chez d'autres espèces de bovidés originaires d'Asie comme le yak et le gayal. La pratique traditionnelle consiste à laisser des femelles domestiques plusieurs jours en lisière de l'aire de répartition des espèces sauvages (yak sauvage et gaur en l'occurrence) en espérant que des accouplements se produisent (Wiener *et al.*, 2003; Porter *et al.*, 2016).

L'identification de 20 génomes mitochondriaux appartenant à l'haplogroupe P dans des populations modernes européennes révèle que des introgressions ont pu, même de façon anecdotique avoir lieu par voie maternelle. De plus, elle permet d'accroître de façon importante le nombre de séquences quasiment complètes de génomes mitochondriaux de l'haplogroupe P qui passent d'un effectif de 6 à

26. L'étude de ces séquences pourrait permettre d'en apprendre plus sur la variabilité de l'haplogroupe P et sur le nombre et l'âge de ces introgressions.

Dans notre étude, nous n'avons pas identifié d'animaux appartenant à l'haplogroupe R dans les animaux séquencés. Il est cependant intéressant de noter que l'haplogroupe R aurait été introgressé à partir de populations d'aurochs du Nord de l'Italie. D'après Bonfiglio *et al.* (2010), Virgile en 29 av. J-C décrit une maladie affectant les bovins domestiques en Italie du Nord ainsi que l'utilisation d'aurochs sauvages au lieu d'animaux domestiques pour tirer un char rituel lors d'une fête religieuse. La domestication ou dressage d'aurochs sauvages dans cette région semble finalement être une pratique presque courante et plus facile que pour les aurochs de très grande taille décrits à la même époque dans le sud de l'Allemagne. Cet exemple d'introgression d'aurochs étaye notre hypothèse d'une introgression similaire d'aurochs appartenant à l'haplogroupe P dans les populations domestiquées.

L'étude de la géolocalisation et de la parenté entre races dans lesquelles les lignées P, Q et I ont été identifiées sont des éléments très intéressants à prendre en compte. L'Abondance a été identifiée parmi les races européennes comme la plus proche parente de la Montbéliarde (Decker *et al.*, 2014), ce qui peut expliquer l'identification de lignées Q dans ces deux races conjointes. Selon l'étude de Park *et al.* 2015, les races Vosgienne, Gelbvieh (race laitière allemande), Tarentaise, Abondance et Simmental appartiennent à un même rameau phylogénétique. Il est donc possible que nous puissions dans la suite de nos travaux, identifier d'autres lignées porteuses de ce même haplogroupe. Nos collaborateurs de LMU ont également observé un animal porteur de l'haplotype Q dans la race Murnau-Werdenfelser et une lignée P en Murbodner autrichienne a été identifié par le projet MitoTAUROmics dont nous reparlerons dans la suite de ce manuscrit (séquence non publiée ; communication personnelle Ivica Medugorac ; Curik et al., 2017). L'haplotype Q a également été identifié en Brune des Alpes (Achilli *et al.*, 2008). Le berceau des races Montbéliarde et Abondance est relativement proche de l'Italie et de l'Allemagne et il est donc tout à fait possible que les différentes lignées Q et P observées dans ces différentes races proviennent d'ancêtres communs à ces races.

2.3.2. Introgression ou domestication secondaire ?

L'origine néolithique unique commune à tous les haplogroupes T dans le Croissant Fertile est communément admise. Des scénarios alternatifs de domestications secondaires ou de flux de gènes du sauvage vers le domestique à partir de femelles aurochs en Europe ont été discutés à plusieurs reprises dans la littérature (Beja-Pereira *et al.*, 2006; Achilli *et al.*, 2008; Bonfiglio *et al.*, 2010; Schibler *et al.*, 2014). Parmi les questions que se posent les chercheurs en phylogénétique, l'une d'elles revient très souvent lorsque des haplotypes d'espèces sauvages sont retrouvés chez leurs homologues domestiques : s'agit-il d'un évènement de domestication secondaire ou d'introgression ponctuelle ? Dans la première hypothèse, cela signifierait une domestication d'un échantillon d'individus d'une population complètement différente des aurochs domestiqués au Proche-Orient ou en Inde. Ces individus domestiqués à une période potentiellement différente de celle des deux (ou trois) premiers événements, possèderaient donc un pool de gènes nucléaires et mitochondriaux différents de la première population. Dans le second cas, l'introgression serait anecdotique et se ferait dans un troupeau d'animaux déjà domestiqués. Cela se produit généralement sur la voie mâle par exemple avec des femelles *Bos taurus* domestiques saillies par des mâles aurochs sauvages.

La présence de l'haplogroupe P chez les bovins domestiques modernes, défini comme prédominant chez les aurochs européens et la présence de lignées proches de T chez les aurochs italiens prénéolithiques (Beja-Pereira *et al.*, 2006; Mona *et al.*, 2010) rend plus difficile la détermination d'un évènement d'introgression ou de domestication secondaire par simple comparaison de la distribution des haplogroupes. Il y a également de nombreux arguments allant à l'encontre de l'hypothèse de domestication secondaire. Premièrement, jusqu'à notre étude et la découverte d'un animal porteur de l'haplotype P en Autriche, l'ensemble des données européennes jusqu'ici suggérait une introgression de gènes locaux très anecdotique et géographiquement limitée (par exemple, l'haplogroupe R en Italie du Nord) dans les populations de bovins domestiques. En effet, la présence d'haplogroupes différents de T était l'exception chez les bovins modernes européens. Deuxièmement, dans le cas d'évènements de domestication secondaires, la population initiale telle qu'estimée aujourd'hui à environ 80 femelles aurochs du Moyen–Orient devrait également être plus importante (Scheu *et al.*, 2015b). Une part plus importante de l'héritage génétique de l'auroch serait également attendu dans les modèles étudiant les flux de gènes entre aurochs et taurins (Upadhyay *et al.*, 2019). Ces éléments amènent à penser que des évènements de domestication indépendants et séparés d'aurochs européens sont hautement improbables et qu'il s'agirait donc plutôt de rares introgressions de lignées maternelles sauvages dans la population domestiquée.

Il est nécessaire de relativiser les enseignements tirés à partir des génomes mitochondriaux modernes, comme le démontre l'exemple de porcs domestiques modernes européens. Les porcs ont été domestiqués indépendamment en Anatolie et dans la vallée du Mékong environ 9000 ans avant aujourd'hui. Les analyses de génome mitochondriaux d'animaux anciens ont révélé que les premiers porcs domestiques en Europe ont été transportés par les premiers éleveurs du Proche-Orient en Europe environ 7500 ans avant aujourd'hui. Cependant, quelques milliers d'années après leur introduction, les porcs domestiques en Europe avaient complètement perdu le génome mitochondrial spécifique du Proche-Orient et avaient acquis à la place les haplotypes mitochondriaux de sangliers européens locaux. Ces résultats suggèrent que les premières populations domestiques ont connu un flux génétique post-domestication provenant de populations de sangliers qui n'étaient pas impliquées dans le processus de domestication anatolien et que le sanglier européen n'a pas été domestiqué de manière indépendante (d'après Frantz *et al.*, 2015).

Le séquençage d'un plus grand nombre de bovins anciens, domestiques ou sauvages pourra améliorer notre compréhension de l'évolution des lignages mitochondriaux en Europe. L'analyse de la diversité des haplotypes P identifiés dans un plus grand nombre de races Européennes, la datation de la divergence entre ces haplogroupes et l'étude de leur répartition géographique pourrait apporter des éléments importants et préciser quand et de quelle manière ont été intégrées des femelles aurochs dans les populations domestiques. C'est un des objets du projet que nous proposons en perspective.

2.3.3. Perspectives

Malgré les faibles quantités d'ADN que nous avons pu récupérer post-génotypage, nous avons identifié l'haplotype de chacun des échantillons. Pour la suite de notre étude et dans le but de publier un article présentant un plus grand nombre d'échantillons, nous prévoyons d'organiser des prélèvements de sang afin d'avoir une quantité d'ADN suffisante pour garantir un séquençage complet du génome mitochondrial.

La première étude de diversité parmi les 400 séquences a été réalisée en collaboration avec l'équipe du LMU. Elle a représenté une opportunité très précieuse d'échanger sur les méthodes en phylogénétique et les questionnements amenés par les résultats de cette seconde étude. Nous avons ainsi pu profiter de l'expérience de cette équipe en matière d'étude phylogénétique.

Pour prolonger cette étude, nous envisageons de monter un projet de plus grande envergure, en collaboration avec le LMU, et le consortium Eurogenomics (http://www.eurogenomics.com/), qui regroupe des acteurs de la sélection bovine issus de huit pays Européens (France, Allemagne, Danemark, Pays-Bas, Suède, Finlande, Pologne et Espagne) et est à l'origine du développement des puces de génotypage Illumina EuroG10K et EuroGMD. Notre objectif est de cribler l'ensemble des animaux génotypés dans ces pays pour identifier des individus porteurs de mitochondries « non-T » et distinguer l'haplogroupe auquel ils appartiennent. A cette fin, j'ai dessiné de nouveaux marqueurs spécifiques, identifiés dans cette étude, que j'ai intégrés à la partie publique de la puce EuroGMD en février 2020 afin que l'ensemble des membres du consortium Eurogenomics y ait accès et puisse discriminer les haplogroupes Q, R et I et les sous haplogroupes de P. Nous souhaiterions ensuite séquencer le génome mitochondrial complet de 384 bovins européens porteurs de mitochondries non-T soit par PCR-long range et séquençage Illumina, ou par une nouvelle approche développée par nos collègues de l'équipe Biologie Intégrative et Génétique Equine de l'UMR GABI, l'objectif étant d'éviter les aléas dus à l'approche PCR et séquençage Sanger. La nouvelle approche, en cours de test sur des échantillons d'ADN de cheval, de truite et de bovin est basée sur un enrichissement de l'ADN total en mitochondries grâce à l'utilisation d'exonucléases qui détruisent les brins d'ADN linéaire et pas l'ADN mitochondrial circulaire, suivi par une amplification du génome mitochondrial. A plus long terme, j'aimerais enrichir cette étude en réalisant des analyses de flux de gènes sur le génome nucléaire comme par exemple celles réalisées par le logiciel ADMIXTURE. Nous pourrions également étudier et comparer les flux de gènes provenant de l'auroch dans les différentes races sur le génome mitochondrial, et observer si les fréquences évoluent de façon similaire à la fréquence de non-T dans chaque race avec des outils comme Treemix. En attendant, nous continuons tous les jours à enregistrer de nouveaux typages, et donc possiblement de nouvelles lignées non-T. De nouveaux marqueurs permettant de discriminer directement les haplotypes P, Q et I ont été également été placés en février 2020 sur la prochaine version de la puce EuroGMD qui va être utilisée en routine.

Pour conclure, le séquençage de génome mitochondriaux nous a permis de mettre en évidence le paradoxe entre une faible diversité globale entre individus avec une grande part de variants lignée-spécifique et la présence à faible fréquence de lignées très anciennes héritées d'ancêtres sauvages des bovins modernes. Il est à noter que nous n'avons séquencé qu'une infime partie des lignées répertoriées dans chaque race et l'existence d'autres lignées porteuses de variants d'intérêt est tout à fait possible.

Résultats clés de cette étude :

- Nous avons identifié des lignées rares dans la population européenne des bovins modernes grâce à l'utilisation trois marqueurs discriminants l'haplogroupe T des autres haplogroupes sur les puces de génotypages utilisées dans le cadre de la sélection génomique.
- Nous avons mis en évidence la présence de l'haplotype Q, haplotype déjà identifié dans des populations de bovins européens modernes mais principalement dans des races italiennes.
- Nous avons observé la présence de l'haplotype I dans deux animaux Holstein.
- Cette étude a permis de révéler la présence de lignées P parmi les bovins européens modernes. Cet haplotype est associé à une population d'aurochs européens mais n'avait jamais été identifié chez des individus modernes.
- Nos résultats ne sont que le début d'une étude que nous souhaitons mener à l'échelle européenne afin de caractériser plus précisément la fréquence de ces lignées rares sur le continent.

La mitochondrie joue un rôle important dans le métabolisme énergétique, il est donc probable que des variations de son génome aient un effet sur des processus biologiques énergivores comme la production de lait ou la fertilité. Dans la suite de ce chapitre, j'ai étudié l'effet de l'ADN mitochondrial sur deux groupes de caractères : des caractères de production laitière évalués en routine (étude 3) et des paramètres de fertilité des mâles (étude 4), à l'aide de deux jeux de données aux caractéristiques distinctes.

3. Etude 3 : Estimation des paramètres génétiques sur les performances laitières

L'influence de la lignée mitochondriale sur les caractères évalués chez le bovin n'a jamais été étudiée dans les populations laitières françaises, mais a fait l'objet dans de plusieurs études dans des populations étrangères (Bell *et al.*, 1985; Boettcher *et al.*, 1996; Boettcher et Gibson, 1997). Elle est généralement étudiée en intégrant l'effet de la LM aux modèles d'évaluation génétique, comme un effet fixe ou aléatoire. L'objectif de notre étude est d'estimer la contribution de l'effet de la LM à la variabilité de cinq caractères laitiers.

3.1. Matériel et méthodes

3.1.1. Phénotypes

Les données phénotypiques utilisées pour l'analyse sont issues des bases de données nationales d'enregistrement des performances laitières. Il s'agit des quantités de lait (LAIT), de matière grasse (MG) et de matière protéique (MP), exprimées en kg, et des taux butyreux (TB) et protéique (TP), exprimés en g/kg. Une performance résume la production d'un animal, corrigée pour la durée de sa lactation, de la manière suivante : (production totale x 385 jours) / (durée de lactation + 80) (Poutous *et al.*, 1981), ce qui revient à traiter des lactations standardisées à 305 jours. A titre indicatif, les moyennes et écarts-types des performances sont présentés dans le tableau ci-dessous:

Caractère (unité)	Race	Moyenne	Ecart-type	Minimum	Maximum
LAIT (kg)	Holstein	10224	2116	328	23285
	Montbéliarde	8043	1793	345	19269
	Normande	7283	1563	368	14659
MG (kg)	Holstein	398	81,9	10,5	1056
	Montbéliarde	315	72,7	11,1	823
	Normande	310	70,2	11,3	659
MP (kg)	Holstein	326	69,0	10,8	744
	Montbéliarde	266	61,9	11,1	694
	Normande	251	56,4	13,3	511
TB (g/kg)	Holstein	39,2	4,52	9,54	97,1
	Montbéliarde	39,2	3,57	12,2	93,0
	Normande	42,6	3,98	15,1	90,8
TP (g/kg)	Holstein	31,8	2,13	17,8	65,0
	Montbéliarde	33,1	2,04	16,6	68,6
	Normande	34,5	2,21	21,0	67,3

Tableau 13 : Statistiques descriptives des caractères laitiers dans les populations étudiées

3.1.2. Animaux et lignées

J'ai utilisé des données provenant de vaches de races Holstein, Montbéliarde et Normande en première lactation, ayant vêlé entre janvier 2002 et décembre 2015. En race Holstein, j'ai restreint le jeu de données aux performances des femelles ayant vêlé après 2005 pour des raisons de calcul. J'ai également réalisé un second filtre consistant à éliminer les lignées « incohérentes » comportant des animaux avec des mitochondries différentes sur la base de données de génotypages, tel que décrit dans la partie 3.2.3. Les performances de 539 918 Holstein, 650 687 Montbéliardes et 370 697 Normandes ont été utilisées.

L'assignation de la LM à chacun des individus a été effectuée avec le programme Lineage, du package Pedig (Boichard, 2002). Afin de maximiser le nombre d'animaux par niveau d'effet, j'ai éliminé les animaux issus des LM présentes dans moins de cinq cheptels et ceux dans des cheptels dans lesquels moins de cinq LM sont présentes. La description des données pour chacune des races est présentée dans le tableau ci-dessous.

	Holstein	**Montbéliarde**	**Normande**
Nombre d'animaux	539 918	650 687	370 697
Nombre de lignées	8048	13 126	9355
Nombre moyen de descendants par lignée maternelle	67,1	49,6	39,6
Nombre moyen de descendants par mère	1.54	1.62	1.54
Nombre moyen de descendants par père	57,9	82	130

Tableau 14 : Structure du pédigrée des données utilisées

Au total, 8048, 13 126 et 9355 lignées maternelles ont été étudiées respectivement en races Holstein, Montbéliarde et Normande avec une moyenne de 67,1, 49,6 et 39,6 descendants phénotypés par lignée.

3.1.3. Prise en compte d'un effet non-T / T en race Holstein

En plus de l'effet LM, j'ai souhaité étudier l'effet des haplotypes T versus non-T identifiés en race Holstein (seule race pour laquelle les effectifs des individus non-T étaient suffisants). Les haplotypes T et non-T diffèrent par un grand nombre de variants génétiques, ce qui augmente la probabilité que l'une de ces polymorphismes soit responsable d'un effet sur les caractères étudiés.

Pour ce faire, j'ai utilisé les données des animaux génotypés sur les puces Illumina EuroGMD et EuroG10Kv8 (puces contenant les 3 marqueurs me permettant de discriminer T et non-T, *cf*. étude 2 de ce chapitre), au sein de la population étudiée.

Dans une analyse préliminaire à cette étude, j'ai réalisé une première étape de vérification de la cohérence des typages des descendants d'une même lignée. Parmi les LM de l'ensemble de la population Holstein française, dont au moins cinq descendants ont été typés sur la puce MD ou LD v8, j'ai mis en évidence des lignées dont une partie des descendants était T et l'autre non-T (lignées dites « incohérentes », les descendants d'une même femelle héritant tous de son génome mitochondrial).

Sur 101 lignées comportant des individus non-T, 13 étaient incohérentes du fait de possibles erreurs de généalogie.

La figure 18 ci-dessus résume le suivi de l'une des lignées Holstein incohérentes dont j'ai remonté le pedigree. J'ai également identifié une année de divergence, 2008, date de naissance du premier individu potentiellement incohérent (en jaune). Malheureusement les femelles n'étant pas génotypées, nous ne pouvons pas vérifier la parenté des individus au moyen de leur génotypage. La grande majorité des animaux représentés sur ce schéma sont du même élevage. Notre hypothèse est que, pour l'un des deux individus signalés par un astérisque sur la figure 18, deux veaux sont nés le même jour dans le cheptel, et que leur parenté a été inversée.

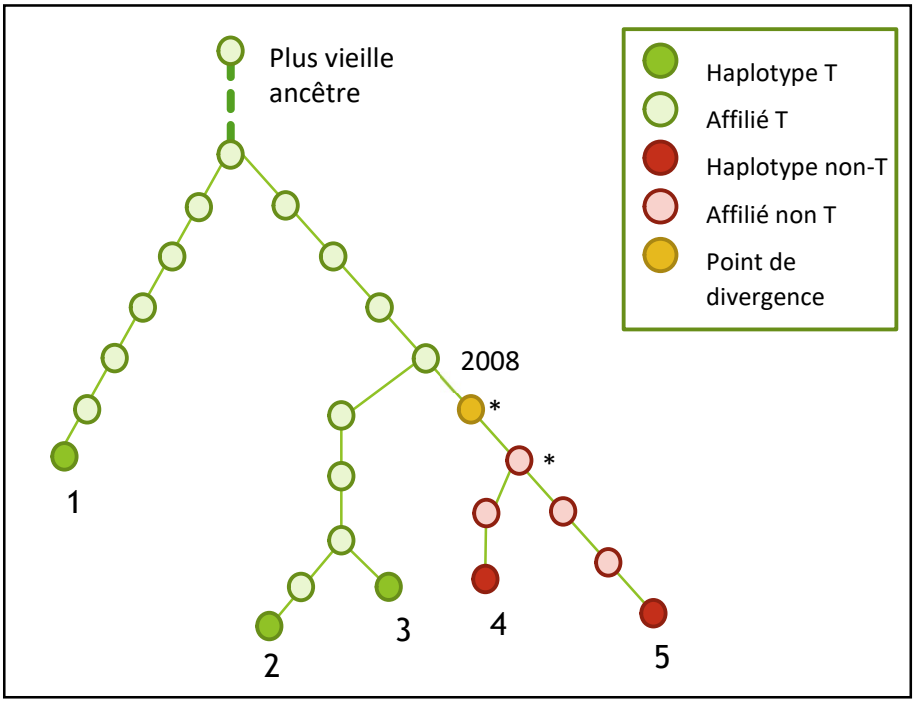


Figure 18 : Pedigree d'une lignée incohérente

Les individus numérotés sont les individus pour lesquels les statuts T et non-T ont été identifiés par génotypage. Les astérisques signalent les deux individus dont la parenté a pu être erronée

Parmi les individus phénotypés sur les caractères laitiers, j'ai éliminé les descendants de lignées identifiées comme incohérentes. Au total, l'analyse en race Holstein a opposé 8027 lignées T représentant 538 598 descendants T contre seulement 21 lignées non-T représentant 1320 descendants.

3.1.4. Modèle utilisé pour notre analyse

L'estimation des paramètres génétiques a été faite à partir d'un modèle mixte linéaire univarié (les caractères sont analysés séparément), en utilisant la méthodologie du maximum de vraisemblance avec le logiciel Wombat (Meyer, 2007).

Le modèle linéaire utilisé pour cette étude est le modèle utilisé aujourd'hui dans les évaluations génétiques de ces caractères laitiers (GenEval, 2019), en rajoutant un effet de la lignée maternelle et en utilisant uniquement les performances de première lactation.

Le modèle utilisé est le suivant :

$$Y_{ijklm(n)} = \mu + gc_j + av_k + mv_l + lm_m + a_i + \varepsilon_{ijklm(n)}$$

où $Y_{ijklm(n)}$ est la performance de l'animal (LAIT, MG, MP, TB ou TP), μ la moyenne générale des performances, gc_j le groupe de contemporains cheptel x campagne laitière, av_k l'effet combiné âge au vêlage x campagne laitière x région, mv_l l'effet combiné mois de vêlage x campagne laitière x région, lm_m la lignée maternelle, suivant une loi normale $\mathcal{N}$ (0, σ^2_{lm}), a_i l'effet génétique de l'animal, le vecteur **a** suivant une loi normale $\mathcal{N}$ (0, σ^2_a**A**), avec **A** la matrice de parenté, et $\varepsilon_{ijklm(n)}$ l'erreur résiduelle, suivant une loi normale $\mathcal{N}$ (0, σ^2_e).

Dans le cas de l'étude de l'haplotype T vs. non-T, le modèle utilisé est le suivant :

$$Y_{ijklm(n)} = \mu + gc_j + av_k + mv_l + t_n + a_i + \varepsilon_{ijklm(n)}$$

où t_n est l'effet lignée T ou non-T (race Holstein seulement), les autres effets ayant la même définition que dans le modèle précédent.

3.2. Résultats et discussion

Dans un premier temps, j'ai estimé l'héritabilité des caractères étudiés (tableau 15). Les héritabilités (h^2) obtenues varient de 0,32 à 0,34 pour la quantité de lait, 0,30 à 0,34 pour la matière grasse, 0,26 à 0,30 pour la matière protéique et enfin de 0,62 à 0,73 et de 0,64 à 0,66 pour les taux butyreux et protéique respectivement.

h²	**Holstein**	**Montbéliarde**	**Normande**
Nombre d'animaux	539 918 (335 581)	650 687 (99 058)	370 697 (91 020)
LAIT	0,34 ↘ (-0,03)	0,32 ↗ (+0,03)	0,34 ↗ (+0,02)
MG	0,31 =	0,30 ↗ (+0,03)	0,30 ↗ (+0,01)
MP	0,26 ↘ (-0,03)	0,26 ↗ (+0,05)	0,30 ↗ (+0,03)
TB	0,73 ↘ (-0,06)	0,62 ↗ (+0,12)	0,64 ↗ (+0,12)
TP	0,66 ↗ (+0,13)	0,64 ↗ (+0,09)	0,61 ↗ (+0,08)

Tableau 15 : Héritabilités estimées dans notre dispositif et écart aux héritabilités estimées par Boichard et Bonaïti (1987)

Les héritabilités estimées sont en majorité plus élevées mais peu différentes que celles estimées par Boichard et Bonaïti en 1987. Plusieurs points peuvent expliquer la différence entre ces deux estimations. L'étude de 1987 portait sur un nombre limité de taureaux ayant un nombre important de filles phénotypées (dispositif père), tandis que mon étude portait sur un effectif d'animaux 1,5 à 6 fois plus important avec des individus génétiquement plus hétérogènes. La variation de l'héritabilité pourrait également être expliquée par l'utilisation d'un modèle différent. Le modèle que nous avons utilisé est un modèle dit « animal » qui estime les valeurs génétiques des individus avec des performances tandis que l'étude réalisée par Boichard et Bonaïti repose sur modèle « père ». Dans ce modèle père, on estime l'effet du père sur la performance de ses filles et la variance génétique est estimée par 4 fois la variance père. De plus, dans Boichard et Bonaïti, les parentés entre pères n'étaient pas prises en compte, ce qui tend à sous-estimer la variance génétique.

En race Holstein, nous avons tout d'abord estimé l'effet T ou non-T. Le tableau ci-dessous présente les moyennes des performances brutes des animaux T et non-T, l'estimation de l'effet d'être non-T sur les caractères laitiers après correction des autres effets du modèle et le pourcentage que représente cet effet par rapport à l'écart-type phénotypique.

	LAIT	MG	MP	TB	TP
Performance moyenne brute animaux T	10225	398,47	326,06	39,248	31,911
Performance moyenne brute animaux non-T	10066	393,96	315,67	39,240	31,370
Effet non-T (par rapport à T)	-36,05	-0,74	-1,00	-0,07	-0,02
En pourcentage par rapport à l'écart-type phénotypique	2%	<1%	1%	2%	<1%

Tableau 16 : Estimation de l'effet des lignées T et non-T sur les caractères laitiers

Ainsi, être porteur d'un haplotype non-T semble avoir un effet faiblement défavorable pour chacun des caractères. J'ai comparé l'estimation de cet effet avec l'écart-type phénotypique de chacun des caractères. L'effet « porteur d'un haplotype non-T » est relativement faible. Il ne représente que 1 ou 2% de l'écart-type de chacun des caractères, soit la magnitude d'un QTL. J'ai également testé la significativité de cet effet à partir d'un test de student. L'effet d'être non-Test non significatif pour l'ensemble des caractères (p.value>0,3). Il est important de noter que dans cette étude, nous n'avons eu la possibilité que de distinguer les lignées T et non-T. Or, nous avons montré dans l'étude 2 de ce chapitre, que des lignées I, P et Q ségrégent en race Holstein. Grâce aux nouveaux marqueurs placés sur la prochaine version de l'EuroGMD, il nous sera possible de distinguer chacun de ces haplotypes et d'étudier leurs effets propres, à condition d'identifier un nombre suffisant de lignées de chaque type.

Dans un deuxième temps, nous avons estimé pour les trois races, la part de la variance phénotypique expliquée par la LM (voir tableau ci-dessous).

Variance lignée	Holstein	Montbéliarde	Normande
LAIT	0,000 ± 0,000	0,001 ± 0,001	0,000 ± 0,000
MG	0,000 ± 0,000	0,002 ± 0,001	0,001 ± 0,001
MP	0,000 ± 0,000	0,001 ± 0,001	0,001 ± 0,001
TB	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,001
TP	0,000 ± 0,000	0,000 ± 0,000	0,000 ± 0,001

Tableau 17 : Part de la variance phénotypique expliquée par la LM

La part de variance expliquée par la LM est très proche ou égale à 0, pour tous les caractères et dans les trois races.

Dans notre dispositif, la part de variance phénotypique expliquée par la LM n'est pas significative chez aucune des trois races. Dans la littérature, certaines analyses ont également considéré la LM comme un effet aléatoire, estimé la part de variance de celle-ci et, contrairement à nous, identifié un effet non nul sur les caractères laitiers. L'étude de Spehar *et al.* (2017) sur 46 696 animaux Holstein croates, selon un protocole sensiblement similaire au nôtre, montre que la LM expliquerait 2% de la variance phénotypique pour le TB et le TP. La proportion de la variance phénotypique expliquée par la LM était de 3% pour les caractères LAIT, MG et MP. Dans une étude de Boettcher *et al.* (1996), la proportion de

la variance phénotypique expliquée par la LM pour les caractères LAIT, MG, MP était respectivement de 0,38%, 0,71% et 2,90%. Une autre étude sur la population Simmental Autrichienne par Schnitzenlehner et Essl (1999) a estimé la composante de variance attribuable à la LM à 2% pour la quantité de lait produite en première lactation.

D'autres études au contraire, considèrent que l'effet de la LM est insuffisant pour qu'elle soit prise en considération dans les modèles d'évaluation génétique. Ainsi, aucune composante significative de la variance attribuable à la LM n'a été trouvée pour le rendement laitier à 305 jours, les caractères de composition et la persistance en race Holstein (Roughsedge *et al.*, 2000).

La méthodologie des modèles mixtes permet une estimation de la part de variance due aux différents effets et l'estimation non biaisée des effets de LM. Toutefois, elle repose sur les informations généalogiques pour la définition de la LM. Or comme nous l'avons montré sur les lignées divergentes T/non-T, il existe parfois des erreurs d'attribution de la LM. Je n'ai pu tester les erreurs d'affiliation que sur les lignées avec des descendants T et non-T. Seules les lignées avec au moins cinq descendants et en race Holstein ont pu être vérifiées de cette façon. Les erreurs d'enregistrements de généalogie et le manque d'exhaustivité des pedigrees qui multiplient le nombre de lignées en diminuant leur effectif, ont une incidence sur la puissance statistique des analyses (Gibson *et al.*, 1997). Plus la généalogie est longue, plus la probabilité qu'elle inclue une erreur de parenté augmente. Roughsedge et ses collaborateurs (2001) ont montré sur données simulées que les erreurs de généalogie pouvaient conduire à une sous-estimation de l'effet réel de la LM. Les erreurs de généalogie et les pedigrees tronqués peuvent expliquer la variabilité des résultats d'estimation d'un effet LM observés dans la littérature.

Il est possible que les erreurs du pédigrée dans nos données nous amènent à sous-estimer la part de variance expliquée par la LM. Il est également possible que des lignées intéressantes ne soient pas suffisamment représentées dans la population (< 5 descendants) et ne soient donc pas inclues dans cette analyse.

En conclusion, la lignée maternelle ne semble pas avoir d'effet significatif sur la variabilité phénotypique des caractères laitiers en race Holstein, Montbéliarde et Normande au sein de notre dispositif expérimental. Nous avons néanmoins observé en race Holstein, qu'être porteur d'une lignée non-T par rapport aux lignées T a un effet faiblement négatif. De manière générale, les études présentées dans la littérature sont mitigées sur l'intérêt réel que représente la prise en compte d'un effet mitochondrial.

Résultats clés de cette étude :

- Nous avons étudié un effet mitochondrial à travers la prise en compte d'un effet lignée maternelle à partir des données généalogiques des bases de données nationales.
- Nous avons mis en évidence l'absence d'effet de la LM sur les caractères de production laitières suivant dans trois races laitières : quantités de lait, de matière grasse et de matière protéique, et des taux butyreux et protéique.
- En race Holstein nous avons également estimé l'effet des lignées mitochondriales non-T versus lignées T, et montré un très faible désavantage pour les individus non-T (environ 1% de l'écart-type phénotypique pour chacun de ces caractères).
- Nous avons observé la présence de lignées incohérentes dans la population Holstein, et mis en évidence l'existence d'erreurs de généalogie.

4. Etude 4 : Estimation des paramètres génétiques sur la fertilité mâle

4.1. Enjeux liés à la fertilité mâle et rôle de la mitochondrie dans la motilité des spermatozoïdes

Chez les bovins, la semence est traditionnellement conservée dans des paillettes de 250 µL avec en moyenne de 15 millions à 18 millions de spermatozoïdes par paillette (GEN'IAtest, 2020). La semence est principalement distribuée après cryopréservation, ce qui peut altérer la survie des spermatozoïdes et diminuer la fertilité de la semence. La qualité est donc importante et elle est vérifiée lors de la production de semence avec certains indicateurs qui seront détaillés par la suite (partie 4.2.) et après l'IA à partir du taux de conception des femelles inséminées.

La fertilité mâle est un enjeu important pour les entreprises de sélection, puisqu'elle impacte fortement la qualité des paillettes de semence qu'elles produisent. De nombreuses études portent sur la prédiction de la fertilité des taureaux en utilisant des modèles fonctionnels (Sellem *et al.*, 2015), moléculaires (Rahman *et al.*, 2017) ou génomiques (Puglisi *et al.*, 2016) des spermatozoïdes. Il n'existe cependant pas de test unique, ou de combinaison de tests, qui puisse prédire de façon fiable la fertilité des taureaux. Dans la plupart des situations commerciales, la fertilité est définie comme la proportion de vaches qui ne reviennent pas en chaleur après insémination (taux de non-retour) ou dont la gestation est confirmée par une échographie, une palpation rectale, un dosage de la progestérone sanguine ou enfin par un événement de vêlage (on parle alors de taux de conception). Toutefois, les indicateurs après insémination ne permettent pas de différencier les raisons d'un échec de la gestation, qui peuvent être multiples : subfertilité de l'un des deux parents, problème sanitaire, insémination réalisée en dehors de la période d'œstrus, mortalité embryonnaire, etc. Il est donc intéressant d'évaluer la fertilité mâle directement à partir de phénotypes fins sur les éjaculats plutôt que sur les données femelles lorsqu'on souhaite étudier spécifiquement la subfertilité des mâles.

Le taux de conception dépend des caractéristiques du sperme fourni par les centres d'IA : les organismes d'IA s'efforcent de collecter un maximum de spermatozoïdes de la meilleure qualité possible au moindre coût. Les spermatozoïdes sont cryopréservés dans des paillettes. Or, la fertilité d'un éjaculat est grandement altérée après congélation de la semence et la conservation de la qualité de la semence représente un enjeu important pour les entreprises de production de semence. Le nombre plus élevé de spermatozoïdes dans la semence cryopréservée compense les dommages causés par le processus de congélation-décongélation par rapport à la semence fraîche et en moyenne, le même niveau de fertilité est atteint avec les deux types de semence (Murphy *et al.*, 2015). Une réduction de la concentration de spermatozoïdes dans une dose d'IA augmenterait le nombre de paillettes produites à partir d'un éjaculat mais pourrait également réduire le taux de conception des vaches inséminées (Shannon et Vishwanath, 1995). Par conséquent, le nombre de spermatozoïdes par paillette est choisi de manière à équilibrer le taux de conception et le nombre de paillettes produites (Mathevon *et al.*, 1998a). Le nombre de paillettes produites par éjaculat dépend alors du volume de l'éjaculat, de la concentration de spermatozoïdes dans l'éjaculat et du pourcentage de spermatozoïdes mobiles.

Les caractéristiques de motilité des spermatozoïdes sont particulièrement étudiées dans des projets sur la fertilité mâle, et la recherche récente en physiologie du sperme se concentre sur la mitochondrie en tant que biomarqueur de la santé et de la fertilité du sperme (Darr *et al.*, 2016; Losano *et al.*, 2017). De nombreuses études chez le cheval ont montré que la mitochondrie est l'un des organites des

spermatozoïdes le plus endommagé pendant la cryopréservation, et pourrait être responsable de la majorité de la perte de motilité et de fertilité après décongélation (revue par Meyers *et al.*, 2019).

Les spermatozoïdes de mammifères possèdent généralement entre 50 et 75 mitochondries, qui assurent la production d'énergie cellulaire nécessaire au battement de leur flagelle (Ankel-Simons et Cummins, 1996). Elles sont localisées dans la pièce intermédiaire, entre le flagelle et la tête qui comprend notamment le matériel génétique et l'acrosome essentiel à la fécondation (voir figure 19). Les mitochondries des spermatozoïdes matures ont une morphologie condensée avec une matrice compactée, ce qui permet une plus grande efficacité de la production d'énergie. Chaque mitochondrie contient entre 0 et 11 copies d'ADN mitochondrial, et certaines études considèrent le nombre de copies comme un marqueur fiable de la qualité et de la fertilité des spermatozoïdes (May-Panloup *et al.*, 2003; Song et Lewis, 2008; Benkhalifa *et al.*, 2014). Le nombre de copies d'ADNmt des spermatozoïdes a été mesuré chez plusieurs espèces, dont l'homme, la souris, le chien et le cheval, et il a été montré que l'augmentation du nombre de copies est associé avec une diminution de qualité des spermatozoïdes (May-Panloup *et al.*, 2003; Song et Lewis, 2008; Orsztynowicz *et al.*, 2016; Hesser *et al.*, 2017). Les mitochondries possédant leur propre génome, l'étude de l'ADNmt pourrait permettre de déterminer des causes inconnues du dysfonctionnement des spermatozoïdes (May-Panloup *et al.*, 2003; Malik et Czajka, 2013).

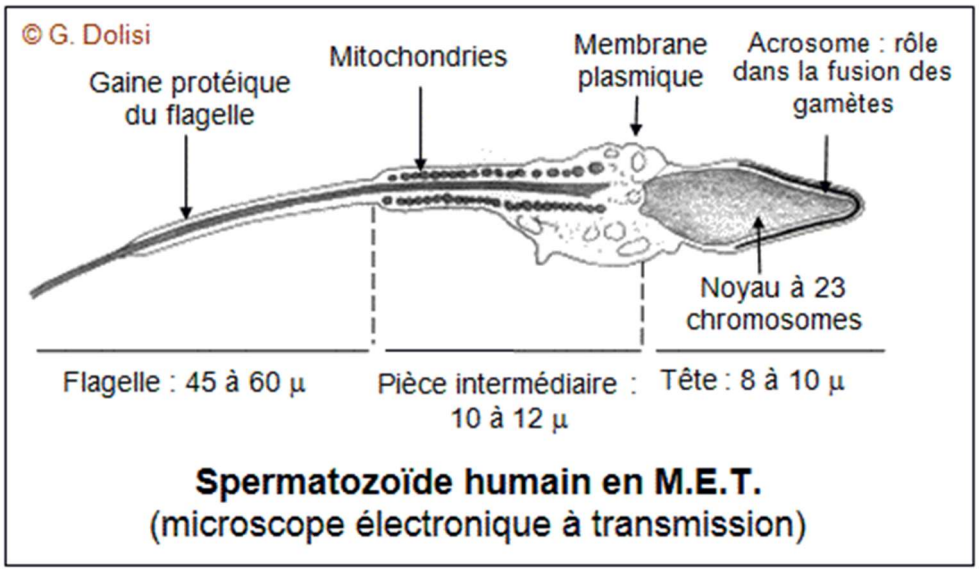


Figure 19 : Schéma d'un spermatozoïde (*https://www.bio-top.net/Schemas/Spermatozoide.gif*)

Dans cette étude, nous avons utilisé des données collectées en routine sur les mâles reproducteurs de deux entreprises de sélection en race Montbéliarde et avons estimé l'héritabilité des caractères ainsi que la part de variance phénotypique expliquée par la LM.

4.2. Matériel et méthodes

4.2.1. Phénotypes

Les données ont été collectées par deux entreprises de sélection en race Montbéliarde : Umotest et EvaJura. Les caractères mesurés par ces deux entreprises n'étaient pas systématiquement identiques et nous avons choisi de ne retenir que les caractères communs afin de maximiser le nombre d'animaux étudiés. Les caractères étudiés sont les suivants : pourcentage de spermatozoïdes motiles ou vivants,

note de motilité avant et après décongélation. Les mesures sont effectuées par collecte et sont répétées pour un même animal. La semence d'un taureau est généralement prélevée tous les 3 à 4 jours sous la forme de deux éjaculats successifs. Chacun des prélèvements du jour est identifié par son rang d'éjaculat (« éjaculat numéro 1 », « 2 » et même « 3 » dans certains cas). Pour une même entreprise de sélection, les données peuvent provenir de différents centres de collecte. Certains caractères n'ont pas été mesurés pour tous les animaux (pourcentage de spermatozoïdes mobiles après décongélation et motilité individuelle après décongélation). Les caractères pour lesquels nous avons des performances et leurs abréviations sont détaillés ci-dessous.

- Volume d'éjaculat (VOL)

Une fois collecté, le sperme est transmis au laboratoire où une mesure du volume est effectuée par lecture directe sur le tube de collecte gradué ou par pesée (Gérard *et al.*, 2008).

- Concentration en spermatozoïdes (CONC)

La concentration peut être mesurée soit par évaluation de la couleur de l'échantillon de manière subjective, soit de manière indirecte par spectrophotométrie, en calculant le rapport de la densité optique finale sur la densité optique émise (Cabannes, 2008).

- Note de motilité massale (NMM)

La motilité massale est évaluée sur un éjaculat dans un environnement et avec un matériel placés à une température de 37°C afin de minimiser le risque de choc thermique. Elle est estimée au microscope à un grossissement x100 : le mouvement global des spermatozoïdes est noté de 0 à 5 dans une microgoutte de sperme selon l'intensité des vagues. Une note de 0 signifie qu'aucun mouvement de vague n'est décelé tandis que 5 traduit la présence de tourbillons rapides. Le tableau 18 présente la définition de chacune des notes.

La motilité massale ne reflète pas la motilité individuelle des spermatozoïdes mais la frénésie dans l'échantillon provoquée par l'ensemble des mouvements de tous les spermatozoïdes présents dans la microgoutte observée.

	Nature et intensité du mouvement
0	Aucun mouvement à la surface de la goutte
1	Léger mouvement à la surface de la goutte
2	Mouvement net mais ne formant pas de vague
3	Début de vagues
4	Vagues très nettes
5	Tourbillons nettement visibles

Tableau 18 : Echelle de notation de la motilité massale de la semence (Cabannes, 2008)

Les trois caractères présentés précédemment sont évalués sur des éjaculats frais en routine, mais d'autres caractères sont mesurés après décongélation (ou à la fois avant et après décongélation).

Par les changements de température qu'elles provoquent, la congélation dans l'azote liquide et la décongélation endommagent les membranes cytoplasmiques, le cytosquelette et les mitochondries. Les capacités fonctionnelles du spermatozoïde sont altérées. La qualité de la semence est donc évaluée

sur un échantillon de paillettes après congélation, sur la motilité massale (et/ou individuelle) et le pourcentage de spermatozoïdes vivants.

De manière générale, il existe une corrélation élevée entre l'aptitude à la congélation et la qualité du sperme frais (Dumont, 1997). Cependant, une semence jugée de bonne qualité en frais (note de motilité massale supérieure à 3 et au moins 60% de spermatozoïdes vivants) peut s'avérer mauvaise après décongélation.

- Note de motilité individuelle (NMID, après décongélation)

La motilité individuelle des spermatozoïdes évalue le pourcentage de spermatozoïdes motiles. La vitesse de déplacement est appréciée par l'opérateur au microscope. Les spermatozoïdes peuvent avoir des mouvements oscillants, c'est-à-dire qu'ils vibrent et n'ont pas de direction définie et des mouvements dits « fléchants » ou « progressifs » en se déplaçant rapidement et en ligne droite (Marc, 2012). L'éjaculat est dilué 10 à 40 fois dans un tampon isotonique tiède et une goutte de cette solution est ensuite observée à grossissement (x100 ou x200). Ne sont pas comptabilisés les spermatozoïdes présentant des mouvements circulaires ou des mouvements d'amplitude très réduite. Aujourd'hui, des appareils automatisés associés à des logiciels d'analyse d'image permettent de réaliser cette mesure.

- Pourcentage de spermatozoïdes vivants (PV : avant-décongélation, PVD : après décongélation)

Ce caractère est évalué à la fois avant et après décongélation de la semence. Le pourcentage de spermatozoïdes vivants est établi par l'opérateur au microscope optique de manière approximative. Une corrélation très forte existe entre cette valeur et la note de motilité. Il existe également parfois une confusion entre les pourcentages de spermatozoïdes vivants et motiles. Les spermatozoïdes vivants sont définis par une structure membranaire intacte. L'examen peut être fait après traitement à l'éosine-nigrosine, qui permet de colorer les spermatozoïdes dont la membrane n'est plus imperméable. Lorsque le seuil de 60% de spermatozoïdes vivants n'est pas atteint, la semence n'est pas conservée.

- Résistance à la décongélation (RES)

Nous avons rajouté un dernier caractère, appelé résistance à la décongélation (RES) qui représente la différence entre le pourcentage de spermatozoïdes vivants avant décongélation et après décongélation.

4.2.2. Spécificités du jeu de données, filtres et définition des effets du modèle

Deux jeux de données ont été considérés à partir de la base initiale qui nous a été transmise. Certains animaux n'avaient pas de mesures pour les caractères après décongélation. Le jeu de données A contient toutes les performances avant décongélation tandis que le jeu de données B est restreint aux animaux possédant des mesures pour tous les caractères.

Les caractères spermatiques sont complexes et influencés par des facteurs génétiques et environnementaux, tels que la saison, l'âge, le taurelier (qui assure la collecte) et l'intervalle entre deux collectes de sperme consécutives en jours (Stälhammar *et al.*, 1989; Fuerst-Waltl *et al.*, 2006).

Nous avons étudié l'effet des facteurs décrits dans la littérature sur notre jeu de données afin d'en étudier la pertinence. Les figures ci-dessous proposent une description de certaines des observations que nous avons pu faire sur la population étudiée.

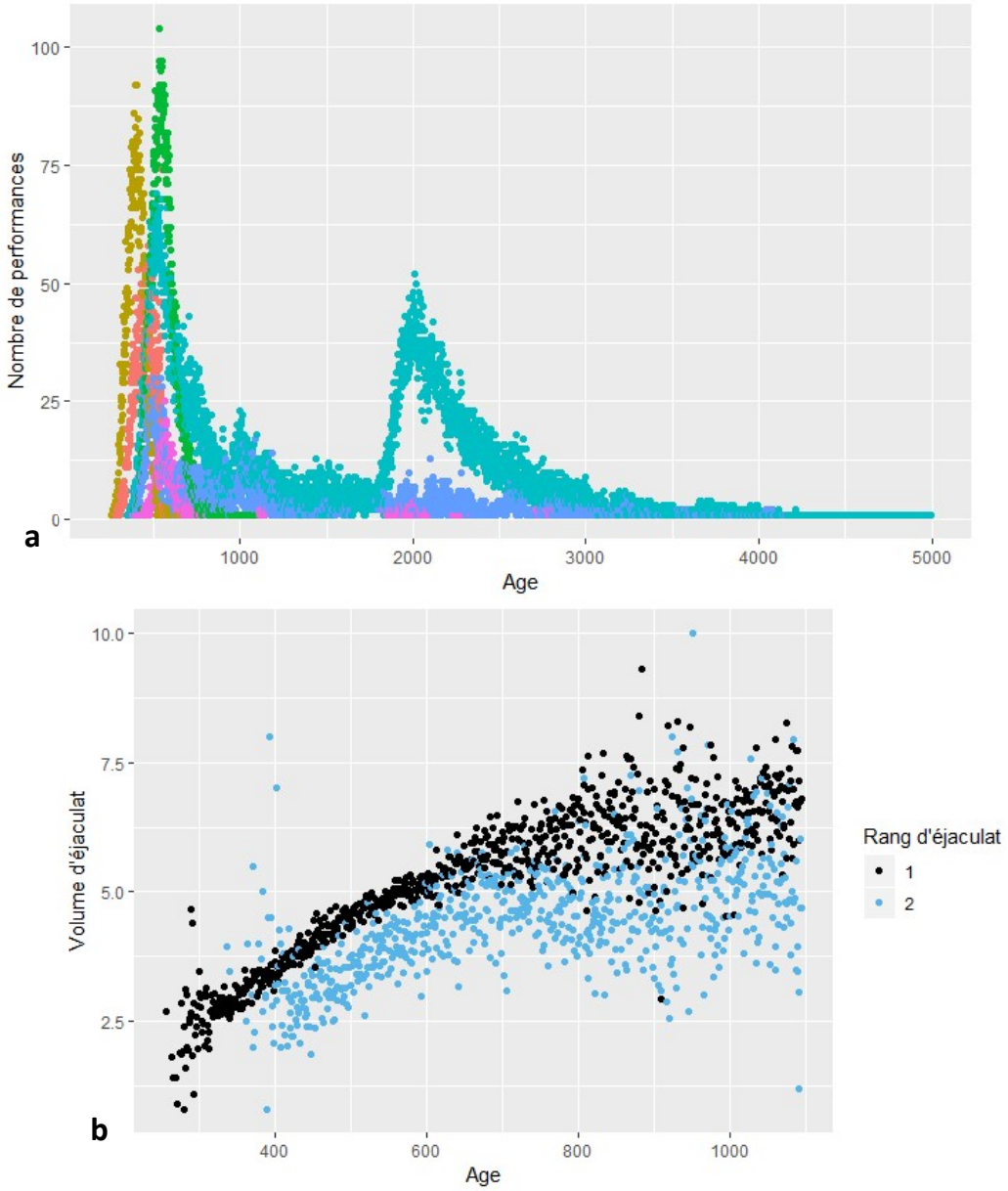


Figure 20 : Description des données (jeu de données A).
L'âge est indiqué en mois. Sur la figure a sont indiquées le nombre de performances par centre de collecte (anonyme) en fonction de l'âge du taureau en mois. La figure b indique le volume d'éjaculat selon l'âge en mois du taureau collecté et le rang d'éjaculat

La figure 20.a représente le nombre de données de notre jeu de données initial (non filtré) par centre de collecte (anonymisé), selon l'âge des taureaux. Les deux pics observés à environ 1 et 4,5 ans illustrent le phénomène de lay-off propre au testage sur descendance. Avec cette méthode, les taureaux sont collectés en taurellerie au tout début de leur carrière pour produire un premier stock de semence, puis durant les quatre ans de la phase de testage (gestation, élevage et première lactation de ses descendantes) sont placés en phase non-productive ou « lay-off ». Les performances de leurs filles une fois en lactation permettent de calculer les index génétiques de l'ensemble des taureaux, dont les 10% meilleurs sont ensuite remis en production pour être diffusés plus largement (Allice, 2017). J'ai choisi de ne garder que les performances mesurées sur les trois premières années de vie de chaque animal afin de ne pas comparer des taureaux ayant eu une phase non-reproductive avec ceux sans phase reproductive. Cela permet de modéliser de façon plus juste l'effet de l'âge sur les caractères, la courbe de volume d'éjaculat selon l'âge formant un plateau chez les taureaux plus âgés.

Cela évite également des différences provenant de concentrations de paillettes différentes, celles-ci étant parfois ajustées chez les taureaux plus âgés comme les taureaux « stars » pour lesquels les paillettes sont parfois moins concentrées pour répondre à la demande en produisant plus de paillettes.

La figure 20.b illustre la corrélation entre l'âge de l'animal et le volume de l'éjaculat, ainsi que l'effet du rang d'éjaculat. Ce sont deux des effets qui ont été pris en compte dans le modèle.

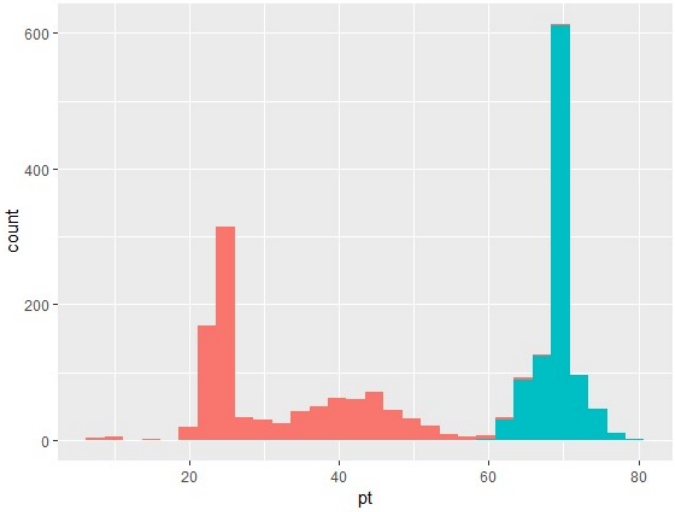


Figure 21 : Description des données de fertilité mâle (jeu de données B). En orange la distribution du PVD et en bleu celle du PV

L'histogramme de la figure 21 représente la distribution du pourcentage de spermatozoïdes vivants dans l'éjaculat avant congélation et après décongélation. Nous pouvons observer une faible variabilité de ce caractère, près des 2 tiers des données étant égales à 70% de spermatozoïdes vivants. Avant congélation, le pourcentage de spermatozoïdes vivants dans l'éjaculat suit une distribution proche de celle d'une loi normale tandis qu'après décongélation ce n'est plus le cas.

Sur les 87 265 données initiales pour 2728 taureaux, je n'ai retenu que les performances de rang d'éjaculat 1 ou 2 et les mesures collectées avant que l'animal ait atteint trois ans. J'ai également gardé les performances dont la durée entre deux prélèvements n'excédait pas 60 jours. Afin de pouvoir estimer correctement l'effet de la lignée maternelle, je n'ai conservé que les descendants de lignées dont les performances d'au moins trois descendants étaient enregistrées.

Le tableau ci-dessous présente les moyennes et écarts-types des performances.

	Nombre collectes	Nombre d'animaux	Nombre de lignées	Caractères	Moyenne ± écart-type
Avant congélation (jeu A)	32743	1454	243	VOL	4,62 ± 1,97
				CONC	13,54 ± 5,54
				PV	64,85 ± 13,92
				NMM	3,54 ± 0,71
Post décongélation (jeu B)	23354	1008	188	PVD	33,91 ± 12,99
				NMID	3,75 ± 0,61
				RES	34,6 ± 0,61

Tableau 19 : Description des données et caractères

4.2.3. Modèle

La définition des effets du modèle linéaire a été réalisée après avoir vérifié leur significativité avec la fonction (lm) de R. Le modèle utilisé est le suivant :

$$y_{ijklmn} = \mu + gc_j + k \times b + d_k + r_l + lm_m + p_n + a_i + \varepsilon_{ijklmn}$$

où y_{ijklmn} est la performance de l'animal, gc_j le groupe de contemporains, k la saison x centre de collecte x année, b le coefficient de régression sur l'âge k, d_k la durée entre deux prélèvements, r_l le rang de l'éjaculat, m_m la lignée maternelle, suivant une loi normale $\mathcal{N}$ (0, σ^2_{lm}), P_n l'environnement permanent, suivant une loi normale $\mathcal{N}$ (0, σ^2_p), a_i l'effet génétique de l'animal. Le vecteur **a** suit une loi normale $\mathcal{N}$ (**0**, σ^2_a**A**), avec **A** la matrice de parenté, et ε_{ijklmn} est la résiduelle, suivant une loi normale $\mathcal{N}$ (0, σ^2_e).

L'effet d'environnement permanent permet de prendre en compte, dans le cas de mesures répétées, un effet propre à l'animal (sans être génétique) et qui va influer sur toutes les performances de celui-ci sur un même caractère.

4.3. Résultats et discussion

4.3.1. Héritabilités

Le tableau ci-dessous présente les composantes de la variance pour les six caractères et la part de variance phénotypique expliquée par la lignée maternelle.

Caractère	Héritabilité	Variance animal	Variance de l'environnement permanent	Variance résiduelle	Variance LM	Part LM
Avant congélation						
VOL	0,166 ± 0,030	0,39	0,35	1,57	0,03	0,011 ± 0,003
PV	0,027 ± 0,020	3,89	41,33	100,36	0,20	0 ± 0
CONC	0,117 ± 0,029	2,35	4,71	12,77	0,26	0,013 ± 0,008
NMM	0,022 ± 0,017	6,99e-3	7,46e-2	0,23	0,01	0 ± 0
Après décongélation						
PV	0,12 ± 0,029	7,04	6,65	45,00	1e-3	0 ± 0,004
NMID	0,021 ± 0,012	5,0e-2	1,4e-2	0,21	1e-3	0,004 ± 0,004
RES	0,096 ± 0,026	5,62	6,20	46,99	1e-3	0 ± 0,004

Tableau 20 : Composantes de la variance pour les 6 caractères de fertilité mâle étudiés

De manière générale, les héritabilités estimées dans cette étude sont plus basses que celles rapportées dans la littérature. L'environnement a une influence importante sur les caractères mesurés. Les héritabilités les plus hautes (0,12 à 0,16) sont obtenues sur les caractères VOL et CON ainsi que PVD. La note de motilité avant et après décongélation a une héritabilité très faible (0,02). Dans la littérature, les estimations de l'héritabilité des caractères de fertilité mâle mesurées sur le sperme varient considérablement selon les études. La variabilité des résultats peut être expliquée par des facteurs tels que les différences de taille de population étudiée, la modélisation statistique, la population ou la race, l'âge et la maturité des taureaux, ainsi que la manière dont les caractères sont enregistrés et définis. Une méta-analyse des performances de reproduction des mâles chez les bovins laitiers et les bovins de boucherie a été réalisée par Berry *et al.* (2014) en synthétisant les résultats de nombreuses études. J'ai repris leur synthèse en ajoutant les résultats d'études plus récentes et en renseignant les races sur lesquelles ces estimations ont été faites. Les résultats individuels de chacune de ces études sont présentés en annexe 6. Les héritabilités estimées varient de 0,04 à 0,65 pour le VE, de 0,13 à 0,56 pour la concentration et de 0,01 à 0,31 pour le PV et donc sont très variables.

4.3.2. Part de variance expliquée par la lignée maternelle

L'effet de la LM sur les différents caractères étudiés est très faible. La part de variance phénotypique la plus importante expliquée par la LM n'est que de 1% environ pour la concentration et le volume d'éjaculat (1.3 et 1.1% respectivement) tandis qu'elle est proche de 0% pour les quatre caractères reflétant la motilité des spermatozoïdes avant et après décongélation. En comparaison, dans une étude réalisée sur 1281 taureaux Angus représentant 100 LM différentes, Garmyn *et al.* (2011) ont estimé une part de variance mitochondriale sur la circonférence scrotale (< 0,0001), le pourcentage de spermatozoïdes motiles (0,013) et les spermatozoïdes anormaux (< 0,001), et conclu un effet marginal sur le PV et non significatif sur les autres caractères. Bien que leurs résultats soient du même ordre de grandeur, l'effet le plus élevé se porte sur un caractère de motilité (pourcentage de spermatozoïdes motiles avant congélation), comme nous l'aurions attendu dans notre étude, étant donné le rôle joué par la mitochondrie pour la production d'ATP nécessaire aux mouvements du flagelle. Les problèmes de fonctionnement de la chaîne respiratoire mitochondriale peuvent affecter la motilité des spermatozoïdes et, par conséquent, la fertilité des mâles. Ruiz-Pesini et ses collaborateurs (1998) ont mené une étude chez l'homme afin de déterminer les relations entre la motilité des spermatozoïdes et les activités des enzymes de la chaîne mitochondriale. Les résultats ont indiqué que des échantillons d'hommes ayant une motilité normale des spermatozoïdes présentaient des activités des complexes I, II et IV de la chaîne respiratoire significativement plus importantes que les sujets asthénozoospermes (présentant une anomalie de motilité du spermatozoïde). Les résultats ont également montré que la motilité dépend largement de la quantité de mitochondries entourant la pièce médiane (Ruiz-Pesini *et al.*, 1998), un indicateur dont nous ne disposons malheureusement pas en routine chez le bovin et qu'il serait intéressant de mesurer. Bien que la LM ne représente pas une part significative de la variance phénotypique, nous espérions observer une corrélation entre la motilité des spermatozoïdes et la LM, comme ce qui a été observé par Garmyn *et al.* (2011). Plusieurs raisons pourraient expliquer la différence entre nos résultats et les leurs, comme par exemple un ensemble de données non exhaustif.

Enfin, si les mitochondries sont importantes pour la motilité du spermatozoïde, il faut rappeler que la majorité des gènes codants pour des protéines de la chaine respiratoire mitochondriale sont dans le génome nucléaire (cf. chapitre 1 : partie 3.3.2.). Dans le but d'étudier ces gènes nucléaires, mon encadrant de thèse Aurélien Capitan a dessiné près de 300 marqueurs placés sur la puce EuroGMD situés sur ces gènes pour estimer leur effet. Des études d'association entre le génome nucléaire et certains caractères de fertilité mâle sont en cours dans notre équipe, ce qui permettra peut-être d'étudier l'impact du génome nucléaire sur la chaine respiratoire mitochondriale.

4.3.3. Des phénotypes peu précis

A la lecture de ces résultats, nous avons recontacté les entreprises de sélection ayant mesuré ces caractères. Il s'avère que les techniques de mesure ont évolué au fil des années sans que les changements de méthode soient consignés. Par exemple, avant 2010, la concentration était mesurée par observation visuelle, puis depuis 2010 elle est mesurée avec un spectrophotomètre. Depuis 2 ans chez une entreprise et 1 an chez l'autre, le pourcentage de spermatozoïdes motiles est mesuré avec un appareil de mesure de la semence, l'IVOS de la société IMV. Les entretiens qui m'ont permis de récolter ces informations ont aussi révélé des différences de traitement avant la congélation des éjaculats. Si un taureau est connu pour avoir un faible pourcentage de spermatozoïdes motiles post-

congélation, le temps d'équilibration de la température à 4°C avant la congélation est passé de 4 à 24h, ce qui permettrait d'augmenter le pourcentage de spermatozoïdes motiles.

Il est également signalé dans une thèse réalisée au sein d'une autre entreprise de sélection « qu'il est possible de convertir cette note (motilité massale) en un pourcentage approximatif de spermatozoïdes mobiles (la note 3 correspondant approximativement à 70% de spermatozoïdes mobiles) » (Cabannes, 2008) ce qui laisse penser que cette approximation a également pu être faite dans notre jeu de données selon les années. La distinction entre le caractère de motilité évalué sur le pourcentage de spermatozoïdes vivants n'a peut-être pas toujours été faite de la même façon. Le manque de standardisation des données est donc aujourd'hui un véritable frein dans l'utilisation des données recueillies par les centres de collecte de semence. La thèse citée précédemment note aussi, concernant le pourcentage de spermatozoïdes vivants, que « cette estimation est subjective et dépend fortement de l'expérience de l'opérateur ».

Enfin, notre jeu de données ne renseignait pas les informations relatives au manipulateur du taureau et la personne chargée de collecter le sperme, celles-ci étant responsables de la préparation avant collecte de l'éjaculat. Mathevon *et al.* (1998a) ont observé un effet significatif de l'équipe de collecte sur le volume de l'éjaculat et le nombre total de spermatozoïdes. Il aurait été intéressant de pouvoir avoir accès à cette information non enregistrée dans les données fournies, et l'absence de prise en compte de cet effet pourrait avoir un impact important sur l'estimation de la variance génétique des caractères étudiés.

4.3.4. Perspectives

Afin d'augmenter la précision de notre analyse, il faudrait pouvoir appliquer des filtres plus stricts, par exemple en ne gardant que les années pour lesquelles la méthodologie suivie est connue. Néanmoins les filtres qu'il faudrait appliquer sur la précision des données risquent en même temps de réduire trop drastiquement le nombre de descendants par lignée, et ainsi de réduire le nombre de lignées pour lesquelles nous pouvons estimer correctement l'effet LM. La prise en compte de données de génotypage sur un grand nombre de variants génétiques mitochondriaux permettrait de regrouper des LM portant les mêmes mitochondries, et donc d'améliorer la puissance des analyses et de remédier à ce problème. Cela permettrait également d'éliminer les biais dus aux erreurs de filiation. Toutefois, cela nécessitera le regénotypage de taureaux anciens puisque seuls les veaux, qui n'ont pas encore de performances sont génotypés sur la puce EuroGMD sur laquelle j'ai intégré ces variants.

Nous avons travaillé avec des données dont une partie était relativement ancienne afin de pouvoir avoir un nombre d'individus par lignée suffisamment grand pour étudier l'effet de la LM. Aujourd'hui des phénotypes plus fins sont souvent enregistrés par les équipes de collecte ou de recherche, notamment grâce à l'automatisation des mesures de la motilité des spermatozoïdes. Les anomalies spermatiques peuvent être mesurées, et classées dans différentes catégories et le type de mouvement individuel grâce aux analyses CASA (*computer-assisted semen analysis*) (Broekhuijse *et al.*, 2011).

Une technique de mesure de l'activité mitochondriale dans les spermatozoïdes a été mise au point permettant de marquer les mitochondries actives et non actives avec une fluorescence différente. La lecture par cytométrie de flux d'un éjaculat traité permet de définir un pourcentage de mitochondries actives (Sellem *et al.*, 2015). Il pourrait être intéressant de comparer l'activité mitochondriale de lignées différentes. Ce dispositif est relativement coûteux et ne pourra pas être mis en place sur l'ensemble des lignées identifiées chez les taureaux reproducteurs mais pourrait par exemple

permettre de comparer les éjaculats de spermatozoïdes d'individus ayant un haplotype mitochondrial I, P Q et T.

Dans la suite de l'étude du génome mitochondrial, nous aimerions tester l'effet de la LM sur d'autres caractères, notamment la fertilité femelle. En 2002, Sutarno *et al.* ont identifié une association entre des polymorphismes du génome mitochondrial et des caractères de fertilité femelle dans deux races de bovins, montrant que le génome mitochondrial pourrait être une source de variation génétique.

Résultats clés de cette étude :

- Dans l'analyse présentée, nous avons estimé l'héritabilité des caractères de fertilité mâle mesurés par deux entreprises de sélection et la part de la variance phénotypique expliquée par le génome mitochondrial. Nous avons mis en évidence un effet de la LM d'environ 1% pour la concentration et le volume de l'éjaculat.
- Contrairement à nos attentes, cette étude ne montre pas d'effet de la LM sur les caractères de motilité avant congélation et après décongélation.
- Nous avons présenté des limites propres à notre jeu de données pouvant expliquer nos résultats non-significatifs.

5. Etude 5 : Contrôle de la généalogie et étude d'association

Dans l'étude 3 portant sur l'estimation de la part de variance phénotypique des caractères laitiers expliquée par la LM, nous avons montré l'existence d'erreurs d'attribution des lignées maternelles et discuté d'une éventuelle sous-estimation de la variance de la LM. Grâce aux marqueurs que j'ai placés sur la puce EuroGMD, j'ai vérifié la cohérence des généalogies en comparant les génotypages des descendants d'une même lignée. Ces marqueurs m'ont ensuite permis de m'affranchir des erreurs de généalogie en réalisant une étude d'association entre les marqueurs du génome mitochondrial et les caractères de production.

5.1. Méthode

5.1.1. Cohérence des lignées

Dans une première étape j'ai étudié la cohérence des génotypes intra-lignée en vérifiant que tous les descendants d'une même lignée maternelle possèdent bien le même génotype. La même approche avait été réalisé en race Holstein sur trois marqueurs distinguant les lignées T et non-T lors de l'étude 3 (cf. 3.2.3.).

J'ai utilisé pour cela tous les typages d'animaux génotypés par Labogena sur la puce EuroGMD entre juin et novembre 2019. Dans chaque race, je n'ai gardé que les marqueurs informatifs (bi-alléliques sur les génotypages de la race) sur les 563 marqueurs mitochondriaux disponibles. Cela m'a permis d'identifier le nombre de lignées non cohérentes parmi les lignées avec au moins cinq descendants génotypés sur la puce EuroGMD. Malgré le nombre important de lignées génotypées, seul un faible nombre d'entre elles avaient suffisamment de descendants génotypés en novembre 2019 (6 mois après la mise en production de l'EuroGMD). Le tableau ci-dessous présente le nombre d'animaux, de marqueurs, de lignées utilisés pour contrôler les LM attribuées grâce au pedigree. A titre indicatif, je présente également le nombre moyen de descendants parmi les lignées sélectionnées (avec 5 descendants et plus).

	Holstein	**Normande**	**Montbéliarde**
Nombre d'animaux génotypés	35 386	4 631	17 398
Nombre de marqueurs bialléliques	468	391	439
Nombre LM	17 806	2 316	8 192
Nombre LM avec au moins 5 descendants	1 136	188	678
Nombre moyen de descendants par LM*	9,3	7,9	8,4

Tableau 21 : Présentation du jeu de données utilisé pour contrôler la cohérence des lignées

**calculé sur les LM avec au moins 5 descendants*

Pour chacune des lignées, j'ai ensuite remonté la généalogie pour identifier l'année de naissance du dernier ancêtre commun aux deux groupes de descendants discordants.

5.1.2. Etude d'association

Dans une deuxième étape, j'ai réalisé une étude d'association entre les caractères de production laitière et les marqueurs mitochondriaux. Les analyses d'association ou GWAS exploitent le déséquilibre de liaison entre les marqueurs et les QTL. Les analyses de type GWAS testent l'effet individuel des allèles de chaque marqueur tout en corrigeant pour la structure familiale de la population (estimée grâce à une matrice de parenté calculée à partir du pedigree ou des génotypes pour l'ensemble des marqueurs du génome nucléaire disponibles). Dans cette analyse, nous avons exploité les génotypages pour construire la matrice de parenté.

a) Animaux, génotypes et performances

Dans cette étude, nous sommes partis des 4631, 17 398 et 35 386 animaux génotypés entre juin et novembre 2019 sur la puce EuroGMD respectivement en races Normande, Montbéliarde et Holstein (tableau 22, ligne « n0 »). Il s'agit pour la plupart de jeunes animaux n'ayant pas encore été mis en production et donc sans performance. Nous avons donc associé le génotype des animaux génotypés à leurs ancêtres femelles aux performances enregistrées.

Afin d'éviter tout biais dû à de potentielles erreurs de pedigree, j'ai attribué les génotypes mitochondriaux d'un individu à sa mère, seulement si cette dernière avait été génotypée et que l'examen des marqueurs nucléaires avait permis de valider leur filiation. Cette vérification de compatibilité génétique est effectuée en routine dans le cadre de la sélection génomique. De proche en proche, j'ai pu attribuer des génotypes mitochondriaux sur plusieurs générations (tableau 22, n1 à n6) lorsque mère, grand-mère, arrière-grand-mère, etc. avaient été génotypées sur puce à SNP (Illumina Bovine SNP50, EuroG10Kv1 à v7, EuroGMDv1) et que leurs filiations étaient validées. Selon le même principe, j'ai également pu propager des génotypes mitochondriaux de manière descendante, par exemple aux cousins maternels d'animaux génotypés sur la puce EuroGMD (tableau 22, d1 à d5). J'ai ensuite croisé cette liste d'animaux disposant de génotypes mitochondriaux inférés avec les bases de données de performances corrigées pour les effets d'environnement pour les cinq principaux caractères laitiers que nous avons analysés dans l'étude 3.

La stratégie utilisée nous a permis de récupérer un grand nombre de performances pour notre étude. Le tableau ci-dessous présente le nombre d'individus dont la parenté a été vérifiée, avec et sans performances pour chacune des générations. Il faut noter que le génotypage femelle est disponible dans le cadre des schémas de sélection depuis seulement 2009, et depuis 2011 est ouvert à tous les éleveurs, ce qui limite le nombre de générations bénéficiant d'un contrôle de parenté sur génotypage.

	Holstein		**Montbéliarde**		**Normande**	
	Sans	**Sans**	**Avec**	**Avec**	**Sans**	**Avec**
n0	35135	17317	81	251	4540	91
n1	3615	1240	6816	9716	367	1719
n2	607	59	2341	4184	57	756
n3	73	10	427	1250	5	210
n4	13	1	50	338	0	46
n5	1	1	4	72	0	6
n6	0	0	0	7	0	0
d1	18952	10877	3708	8288	3598	1481
d2	9025	3869	1106	3710	1346	486
d3	3345	865	142	1119	382	97
d4	1081	80	13	187	56	22
d5	253	14	1	11	9	2

Tableau 22 : Description par génération du nombre d'individus avec et sans performances affiliés par voie maternelle aux animaux génotypés sur puce EuroGMD. La génération n0 représente les animaux génotypés, n1 leurs mère etc. et d1 les descendants de 1e génération de tous les individus n.

Parmi les milliers d'individus génotypés, seuls 91, 81 et 251 bénéficiaient de performances enregistrées respectivement en races Normande, Montbéliarde et Holstein. Notre dispositif nous a permis de récupérer 4825, 14 608 et 28 882 individus supplémentaires avec performances pour notre analyse.

Les performances ajustées (pour les effets d'environnement) proviennent des évaluations officielles de chacune des trois races. Il s'agit des mêmes caractères que ceux analysés lors de l'étude 3 (LAIT, MG, MP, TB, TP). Le nombre d'animaux retenus et le nombre de lignées maternelles (telles que définies uniquement sur information de pedigree sont présentés dans le tableau ci-dessous.

	Holstein	**Montbéliarde**	**Normande**
Nombre d'individus	28 939	14 653	4 799
Nombre de LM (pedigree)	6 005	3 930	1 084
Nombre SNP biallélique	425	406	361
Nombre SNP avec au moins 50 porteurs de l'allèle alternatif	229	178	74

Tableau 23 : Nombre d'individus et de marqueurs utilisés lors de l'étude d'association

Le jeu de données final comprenait 4799 individus de race Normande, 14 653 de race Montbéliarde et 29 133 de race Holstein génotypés pour 563 marqueurs mitochondriaux sur la puce EuroGMDv1. Parmi les marqueurs mitochondriaux de la puce EuroGMD, seuls les variants bialléliques intra-race ont été gardés. Comme je l'ai démontré lors de l'étude 1, une grande partie des variants mitochondriaux sont lignée-spécifique. Afin d'éviter de surestimer l'effet d'un allèle porté par un nombre faible d'animaux aux performances atypiques, je n'ai gardé que les variants dont l'allèle alternatif était présent chez au moins 50 individus. Le filtre que nous avons choisi réduit drastiquement le nombre de variants étudiés (voir tableau 23), mais l'accumulation des génotypages nous permettra par la suite d'étudier un plus grand nombre de variants.

b) Modèle utilisé

Nous avons effectué des analyses d'association univariées entre tous les variants polymorphes et les cinq caractères laitiers: LAIT, MG, MP, TB, TP. Les analyses d'association ont été effectuées à l'aide de l'option mlma du logiciel GCTA (*Genome-wide Complex Trait Analysis*), qui applique un modèle linéaire mixte incluant les variants candidats (Yang *et al.*, 2011) :

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{x}b + \mathbf{u} + \mathbf{e}$$

où **y** est le vecteur des performances pré-ajustées (LAIT, MG, MP, TB ou TP), µ la moyenne générale des performances, b l'effet fixe du variant candidat à tester pour l'association, **x** le vecteur des génotypes imputés codés avec le nombre d'allèles alternatifs (0 ou 1), **u** le vecteur des effets polygéniques aléatoires, $N \sim (\mathbf{0},\mathbf{G}\sigma^2_u)$, avec **G** la matrice de parenté génomique, calculée en utilisant les génotypes de puce BovineHD, et σ^2_u la variance polygénique, estimée sur la base du modèle nul ($\mathbf{y} = \mathbf{1}\mu + u + \mathbf{e}$) et fixée lors du test d'association entre chaque variants et le caractère, **e** le vecteur des effets résiduels aléatoires $N \sim (\mathbf{0}, \mathbf{I}\,\sigma^2_e)$ avec **I** la matrice d'identité et σ^2_e la variance résiduelle.

Le logiciel GCTA permet de prendre en compte une matrice de parentés génomiques afin de corriger d'éventuels biais de stratification dans la population étudiée. J'ai produit cette matrice avec le logiciel PLINK v1.9. (Chang *et al.*, 2015) à partir des données réelles ou imputées de génotypage sur puce BovineHD. L'imputation des puces basse densité vers la puce Illumina BovineHD est réalisée régulièrement dans le cadre de projet réalisant des études d'association.

5.2. Résultats et discussion

5.2.1. Cohérence des lignées

Les typages nous ont permis de contrôler la justesse des informations généalogiques, et de calculer un pourcentage de lignées discordantes.

	Holstein	Montbéliarde	Normande
Nombre total de lignées avec au moins un descendant génotypé MD	17 806	8 192	2 316
Nombre de lignées avec au moins 5 descendants génotypés MD	1136	678	188
Nombre de lignées discordantes	263	100	22
Année moyenne de disruption	1998	1991	1995
Nb moyen de marqueurs distinguant les descendants d'une même lignée discordante	11,7	10,3	8,6
Nombre moyen de liens de filiation maternels connus à l'intérieur des lignées discordantes	47	38	32

Tableau 24 : Statistiques effectuées sur les lignées avec au moins 5 descendants typés par lignée

Le nombre moyen de marqueurs permettant de distinguer des individus porteurs de mitochondries différentes est relativement élevé (8 à 12). Sur l'ensemble du jeu de données, 59, 200 et 306 marqueurs permettent de discriminer les lignées respectivement dans les races Normande, Montbéliarde et Holstein. Cela permet d'affirmer qu'il ne s'agit pas de cas de mutations *de novo* intra-lignée, ou d'erreurs dues à petit nombre de marqueurs peu fiables, mais plutôt d'erreurs de filiations. En effet, la séparation des veaux et de leurs mères dès la naissance en élevage laitier peut être propice à ce type d'erreur si l'éleveur attend d'avoir plusieurs jours pour apposer les boucles d'identification à une série de veaux. Il faut noter également que, même si cela parait peu probable, nous ne pouvons pas exclure la possibilité de transmission d'origine paternelle de mitochondries. Une surveillance régulière de la cohérence des lignées, pourra permettre d'identifier des cas récents de divergence, et d'étudier les haplotypes des deux parents déclarés afin de déterminer si des cas de transmission paternelle se produisent.

Les résultats de ces typages indiquent 15% de lignées divergentes en Montbéliarde, 23% en Holstein et 12% en Normande. La présence de discordances dans une lignée sur 5 signifie que deux haplotypes mitochondriaux différents sont assimilés au sein d'une même lignée dans près de 20% des cas. Il est important de noter que si 20% des lignées présentent des erreurs cela ne reflète pas le taux d'erreur par génération. Nous pouvons estimer « grossièrement » le pourcentage d'erreurs de déclaration par génération à partir (i) de la proportion de lignées discordantes et (ii) du nombre moyen de liens de filiations maternels connus dans les lignées discordantes, en faisant l'hypothèse qu'une seule erreur de filiation a été commise dans un pedigree discordant. Ce calcul donne des taux d'erreurs de filiation sur la voie maternelle de 0,34% en Montbéliarde, 0,40% en Holstein et 0,33% en Normande. A titre indicatif, le calcul réalisé est le suivant avec l'exemple de la race Montbéliarde [1 erreur de déclaration /38 liens de filiation maternelle en moyenne] * [100 lignées discordantes /778 lignées avec plus de 5 descendants génotypés])

Afin d'estimer plus précisément ce taux d'erreur de filiation, je me suis rapprochée de la société Valogène afin d'obtenir des informations sur le nombre de problèmes de déclaration de pédigrée sur les animaux dont les deux parents sont génotypés à partir des contrôles de filiation effectués en routine dans le cadre de la sélection génomique.

	Nombre	Pourcentage
Père et mère typés et compatibles	813 325	99,87
Père compatible, mère incompatible	97	0,01
Père incompatible, mère compatible	575	0,07
Père incompatible, mère incompatible	402	0,05

Tableau 25 : Vérification de parenté des animaux génotypés toutes races comprises (génotypés jusqu'en janvier 2020)

Sur un total de 813 325 animaux génotypés entre 2008 et 2020, toutes races confondues, un faible pourcentage d'individus (0,06%) n'est pas compatible avec la mère déclarée dans le Système d'Information Généalogique. Ce taux est 15 à 18 fois plus faible que celui que nous avons estimé à partir des LM. Cette forte différence pourrait s'expliquer par le fait qu'il s'agisse d'animaux génotypés, et donc bien souvent représentant un intérêt génétique pour l'éleveur, qui assure alors un suivi plus rigoureux à la généalogie de l'animal. Dans la littérature, de nombreuses analyses étudient les erreurs de filiation mais principalement celles concernant la voie paternelle (synthèse par Banos *et al.*, (2001)). En France, une étude réalisée en 2003 par l'Institut de l'Elevage sur 30 taureaux de races Normande, Montbéliarde et Holstein en testage sur descendance identifiait grâce à des analyses de compatibilité génétique (ADN) parmi leur descendance, 7,8% de génisses incompatibles avec les parents déclarés (Ledos *et al.*, 2003).

Le nombre de lignées avec au moins cinq descendants typés est encore faible mais va augmenter au fur et à mesure de l'accumulation de nouveaux génotypages, ce qui permettra d'étudier le pourcentage d'erreurs sur un plus grand nombre de lignées. Notre étude permet de montrer que les données de pedigree actuelles ne sont pas fiables en ce qui concerne les données anciennes des pedigrees et remettent en partie en question l'utilisation d'un effet de la LM comme présentée dans l'étude précédente.

Afin de pouvoir estimer l'effet de variants mitochondriaux, nous avons donc utilisé une deuxième approche, en réalisant une étude d'association par marqueur.

5.2.2. Etude d'association

L'objectif de cette analyse a donc été de tester directement l'effet des variants mitochondriaux sur les caractères de production laitière, en mettant à profit les génotypages réalisés dans le cadre de la sélection génomique à partir de mai 2019 avec la puce EuroGMD v1.

La majorité des animaux génotypés étant trop jeunes pour avoir des performances, j'ai dû tirer profit des informations de pedigree et de contrôle de filiation pour attribuer, sans risque d'erreur, des génotypes à leurs apparentés disposant de performances, de typages sur les autosomes, mais pas de données de génotypage sur marqueurs mitochondriaux. Cette approche, bien que très fiable, induit une perte de charge importante qui m'a contrainte à abandonner l'étude sur les caractères de fertilité mâle, le nombre d'animaux disposant de performances étant déjà réduit (n = 2728).

J'ai réalisé une étude d'association génétique sur les caractères laitiers en estimant un effet SNP par SNP sur les caractères.

Dans un premier temps, j'ai estimé les héritabilités obtenues sur les cinq caractères, présentées dans le tableau ci-dessous, en indiquant la variation par rapport aux héritabilités estimées dans l'étude 3 sur 370 697, 650 687 et 539 918 individus respectivement en races Normande, Montbéliarde et Holstein.

	Holstein	**Montbéliarde**	**Normande**
Nombre d'animaux	28 706	14 589	4 916
LAIT	0,30 ↘ (-0,01)	0,26 ↘ (-0,06)	0,37 ↗ (+0,03)
MG	0,26 ↘ (-0,08)	0,28 ↘ (-0,02)	0,31↗ (+0,01)
MB	0,21 ↘ (-0,05)	0,21 ↘ (-0,05)	0,31↗ (+0,01)
TB	0,71 ↘ (-0,02)	0,64 ↗ (+0,02)	0,68↗ (+0,04)
TP	0,66 =	0,67 ↗ (+0,03)	0,68↗ (+0,07)

Tableau 26 : Estimation des héritabilités obtenues avec GCTA sur les performances corrigées des effets d'environnement

Les héritabilités sont sensiblement similaires à celles obtenues dans l'étude 3, la plus grande variation étant une diminution de 0,08 par rapport à l'étude 3 sur le caractère MG en race Holstein.

Nous avons ensuite étudié l'effet des SNP et n'avons conservé que les valeurs de p inférieures à 0,01. Une correction de Bonferroni pourrait être appliquée sur ces résultats (en multipliant les p-valeurs par le nombre de SNP étudiés, i.e., 74, 178 et 229 pour les races Normande, Montbéliarde et Holstein, respectivement) mais la puissance de notre dispositif n'est pas suffisante pour obtenir un résultat significatif après correction. Le tableau 27 présente les marqueurs filtrés et indique l'impact du variant sur le gène lorsque celui-ci est situé dans une région codante.

Caractère	Race	Position	Allèles	Fréquence	Effet	Ecart type d'erreur	p-valeur	Impact	Gene
LAIT	HOL	16165	G/A	0,002	-267,212	61,66	0,00001		
LAIT	HOL	9480	A/G	0,005	-114,544	36,34	0,00162	faux-sens	COX3
LAIT	HOL	15965	G/A	0,008	-95,941	30,46	0,00163		
LAIT	HOL	9987	A/G	0,002	-172,592	60,88	0,00459	synonyme	ND3
LAIT	HOL	16026	A/G	0,003	-134,864	50,01	0,007		
LAIT	HOL	15492	A/G	0,015	64,223	23,86	0,00711	faux-sens	CYTB
LAIT	HOL	106	G/A	0,057	-31,929	12,00	0,00777		
LAIT	MON	16232	A/G	0,011	84,532	30,99	0,00637		
LAIT	NOR	12711	A/G	0,022	121,919	46,57	0,00884	synonyme	ND5
MG	HOL	16165	G/A	0,002	-9,633	2,55	0,00016		
MG	HOL	11698	A/G	0,003	-7,665	2,51	0,00227	synonyme	ND4
MG	HOL	15965	G/A	0,008	-3,829	1,26	0,0023		
MG	HOL	7976	A/G	0,008	-4,217	1,55	0,00656	synonyme	COX2
MG	HOL	11248	A/G	0,012	-2,912	1,07	0,00675	synonyme	ND4
MG	HOL	15511	A/G	0,004	-5,055	1,92	0,00842		
MG	MON	8247	A/G	0,004	6,705	2,36	0,00446	faux-sens	ATP8
MG	MON	16232	A/G	0,011	3,276	1,26	0,00951		
MG	NOR	16122	G/A	0,034	3,675	1,41	0,00913		
MP	HOL	16165	G/A	0,002	-7,791	1,90	0,00004		
MP	HOL	9480	A/G	0,005	-3,649	1,11	0,00107	faux-sens	COX3
MP	HOL	15965	G/A	0,008	-2,824	0,93	0,00237		
MP	HOL	106	G/A	0,057	-1,107	0,36	0,00245		
MP	HOL	16026	A/G	0,003	-4,516	1,54	0,00329		
MP	HOL	15492	A/G	0,015	1,975	0,72	0,00629	faux-sens	CYTB
MP	HOL	9987	A/G	0,002	-4,821	1,86	0,00966	synonyme	ND3
MP	MON	177	G/A	0,014	2,921	0,94	0,00198		
MP	MON	12189	A/G	0,019	2,457	0,87	0,0049	synonyme	ND5
MP	NOR	12711	A/G	0,022	4,356	1,54	0,00457	synonyme	ND5
TB	HOL	11803	G/A	0,002	0,545	0,16	0,00066	synonyme	ND4
TB	HOL	2558	A/G	0,005	0,284	0,09	0,00271		
TB	HOL	4740	A/G	0,007	0,295	0,10	0,00289	faux-sens	ND2
TB	MON	814	A/G	0,018	0,177	0,06	0,00271		
TB	MON	3871	A/G	0,014	0,198	0,07	0,00393	synonyme	ND1
TB	MON	8925	G/A	0,014	0,190	0,07	0,00564	synonyme	ATP6
TB	MON	7712	G/A	0,007	0,269	0,10	0,00609	synonyme	COX2
TB	NOR	16133	G/A	0,023	0,329	0,12	0,00588		
TP	HOL	4740	A/G	0,007	0,157	0,05	0,00135	faux-sens	ND2
TP	HOL	16197	A/G	0,004	-0,156	0,05	0,00303		
TP	HOL	11440	A/G	0,002	0,234	0,08	0,00408	synonyme	ND4
TP	HOL	16139	A/G	0,018	0,075	0,03	0,00442		
TP	HOL	12876	G/A	0,003	-0,146	0,06	0,00983	synonyme	ND5
TP	NOR	5855	A/G	0,012	0,246	0,09	0,00696	faux-sens	COX1

Tableau 27 : Présentation des variants filtrés sur leurs p-valeurs et fréquences

L'effet de l'allèle alternatif (après le « / ») est indiqué dans l'unité de mesure du caractère (LAIT, MP et MG en kg, TB et TP en g/kg). Les marqueurs associés à un effet sur plusieurs caractères sont surlignés en gris.

Cette étude nous a permis d'identifier 30 SNP différentes avec un effet significatif (p-valeur brute < 0,01) sur au moins un des cinq caractères étudiés. Parmi ceux-ci, on dénombre 9 marqueurs associés à une variation de la quantité de lait produite (variation entre 31 et 267 kg), 9 pour la MG (2 à 9 kg), 10 pour la MP (1 à 7 kg), 8 pour le TB (0,01 à 0,06 g/kg) et 6 pour le TP (0,07 à 0,24 g/kg). Dix de ces variants ont un effet sur plusieurs caractères.

Sur les 30 SNP identifiés, 17 sont situés dans la partie codante du génome mitochondrial dont 16 sont des variants synonymes (ou silencieux, qui ne changent pas la séquence protéique) et 8 sont des variants faux sens (dont 3 affectent plusieurs caractères). Les gènes affectés par des variants faux sens sont :

- *ATP8* (*ATP synthase*) codant une des sous-unités du complexe V de la chaîne respiratoire ;
- *COX1* et *COX3* (*cytochrome c oxydase*), codant des protéines intervenant dans la chaîne respiratoire et catalysant la réduction de l'oxygène en eau ;
- *ND2* codant pour une des sous-unités du complexe I de la chaine respiratoire mitochondriale (NADH-ubiquinone oxidoreductase) ;
- *CYTB* (*cytochrome B*) codant l'unique sous-unité du complexe III de la chaine respiratoire mitochondriale codée par le génome mitochondrial.

Nous avons étudié la conservation des cinq variants faux-sens chez 9 espèces comprenant *Bos taurus*. Les résultats sont présentés en annexe 7 sous forme de de représentation logo et synthétisés dans le tableau 28.

Position	Changement	Gène	Domaine	Acide-aminés de référence dans 9 espèces de vertébrés
4740	p.V159I	ND2	Hélical transmembranaire	I, L, V
5855	p.V57I	COX1	Hélical transmembranaire	I, V
8247	p.T40M	ATP8	Hors domaine	K, Q, T
9480	p.V171I	COX3	Hélical transmembranaire	F,L,V
15492	p.A327T	CYTB	Hélical transmembranaire	A, I, L

Tableau 28 : Conservation chez les espèces (*Bos taurus*, *Homo sapiens* (humain), *Mus musculus* (souris), *Equus caballus* (cheval), *Gallus gallus* (poule), *Felis catus* (chat), *Anolis carolinensis* (lézard), *Procavia capensis* (Daman du Cap), *Xenopus tropicalis* (grenouille). Les domaines des protéines ont été identifiées sur (https://www.unip.org)

De manière générale, ces cinq variants sont associés à des acides aminés peu conservés, ce qui est compatible avec un effet de type quantitatif, puisque des polymorphismes affectant des acides aminés sous fortes contraintes évolutives ont plus de chances d'être à l'origine de phénotypes sévères de type anomalie génétique. Trois de ces variants sont des substitutions d'une Valine par une Isoleucine et que l'allèle alternatif chez le bovin est rencontré en tant qu'allèle de référence chez certaines espèces de vertébrés. Cela démontre que cette substitution est bien tolérée et n'affecte pas la *fitness* des individus porteurs. Bien que ces deux acides aminés soient apolaires et de structure proche, des études ont montré que leur substitution pouvaient avoir des effets phénotypiques observables lorsqu'elles affectent un domaine fonctionnel (comme ici) notamment dans le cas de la protéine du Prion (Ito et al., 2018). Enfin, deux autres variants entraînent des substitutions vers des acides aminés qui ne sont pas observés chez les espèces contrôles que nous avons sélectionnées, et qui n'ont pas les mêmes propriétés physico-chimiques (ATP8 p.T40M et CYTB p.A327T, la thréonine étant polaire alors que l'alanine et la méthionine sont apolaires). Il serait intéressant de vérifier leurs effets grâce à des études fonctionnelles.

Il est important de préciser que les allèles associés à un effet phénotypique ne sont pas forcément les variants causaux. Il peut simplement s'agir d'allèles qui sont liés physiquement aux variants causaux, parce qu'ils appartiennent au même génome mitochondrial. Dans la suite de ces travaux, il sera intéressant de séquencer le génome complet de certains individus choisis et de les comparer aux séquences déjà produites dans cette étude afin d'identifier les variants spécifiques des individus porteurs de l'allèle d'intérêt au marqueur.

La principale limite de notre étude tient au grand nombre de variants situés sur le génome mitochondrial et de leur fréquence souvent faible à l'échelle de la population. Etant donné qu'un faible nombre d'individus est porteur d'un allèle alternatif, nous éliminons beaucoup de variants de l'analyse en appliquant notre filtre à 50 animaux porteurs dans notre jeu de données. Cependant, le nombre des génotypages a massivement augmenté avec la mise en place du génotypage femelle, atteignant sur l'année 2019, 142 490 en race Holstein, 69 848 en race Montbéliarde et 18 957 en Normande (communication personnelle, Valogène). L'étude n'a été réalisée que sur le panel d'individus génotypés entre mai 2019 et novembre 2019 mais avec l'augmentation du nombre d'animaux génotypés, nous devrions pouvoir rapidement augmenter la population étudiée et la puissance de détection. L'augmentation du nombre d'animaux génotypés devrait également nous permettre d'augmenter le nombre d'individus porteurs des allèles alternatifs et ainsi étudier plus de variants.

Une autre limite de notre étude est le nombre de marqueurs étudiés. Il est possible que des différences entre sous-sous-catégories d'haplotypes ne puissent pas être identifiées par nos marqueurs et dépendent de polymorphismes que nous n'avons pas placés sur la puce. Cependant au vue du nombre moyen de marqueurs permettant de distinguer différents haplotypes dans une même lignée, ce risque parait relativement minime.

Dans notre analyse, nous avons également voulu regarder si l'effet des marqueurs distinguant les animaux porteurs et non porteurs de l'haplogroupe T était significatif. Malheureusement, nous n'avions pas assez d'animaux porteurs avec des performances dont la parenté est vérifiée sur génotypage pour pouvoir étudier leurs effets. Avec l'accumulation de nouveaux individus et de nouveaux couples animal-mère, nous espérons pouvoir augmenter suffisamment le nombre d'individus avec performances pour pouvoir étudier l'effet de nos marqueurs distinguant les haplogroupes T et non-T. De même, grâce aux marqueurs placés sur la prochaine version de puce, nous pourrons distinguer directement les animaux porteurs de l'haplotype P, Q, T ou I afin d'étudier l'effet associé aux marqueurs.

Cette approche nous a permis de nous affranchir des erreurs de pedigree en entrainant une perte de puissance qui sera corrigée grâce aux données collectées sur de nouveaux individus génotypés qui s'accumulent rapidement. A moyen terme, nous souhaitons attendre jusqu'à septembre 2020, lorsque le nombre de nouveaux génotypages sera suffisamment augmenté afin de réaliser une nouvelle étude d'association et la valoriser au sein d'un même article avec le séquençage des lignées présenté dans l'étude 1.

Résultats clés de cette étude :

- Nous avons montré grâce aux nombreux marqueurs placés sur la puce EuroG10k / EuroGMD que les données de pedigree actuelles ne sont pas fiables en ce qui concerne la remontée de généalogies bovines anciennes sur la voie femelle.
- Une première étude d'association a été réalisée entre les marqueurs mitochondriaux et les caractères laitiers.
- L'accumulation de nouveaux génotypages nous permettra d'augmenter la puissance de notre analyse et le nombre de SNP étudiés.

Chapitre 3 : Diversité génétique du chromosome Y et effet sur la fertilité mâle

L'objectif des travaux présentés dans ce chapitre est d'étudier la diversité du chromosome Y au sein des races bovines françaises. Le chromosome Y, comme le chromosome X, dérive d'une paire ancestrale d'autosomes. Seuls 3 % des gènes ancestraux existent encore sur le chromosome Y humain (Skaletsky *et al.*, 2003) contre 98 % sur le chromosome X. La nature hautement répétitive et dégénérative du chromosome Y rend difficile l'assemblage de sa séquence et sa caractérisation transcriptomique. Cela explique l'absence de ce chromosome au niveau de l'assemblage du génome bovin UMD3.1 disponible au début de ma thèse. L'animal dont est issu le génome de référence est une femelle mais des banques BAC de son père ont été utilisées lors de la construction de l'assemblage et auraient pu aboutir à l'assemblage du chromosome Y sans ces difficultés techniques. Dans ce contexte, nous avons décidé d'attendre la sortie du nouvel assemblage ARS-UCD1.2 pour étudier les variants présents sur le chromosome Y et avons préféré faire porter notre travail principalement sur les données de génotypage et de pedigree. A la sortie du nouvel assemblage, nous avons constaté que l'assemblage du chromosome Y n'avait pas été inclus. Nous avons donc basé nos travaux sur la séquence du chromosome Y (Bellott *et al.*, 2014), utilisé par le projet « 1000 génomes bovins » et téléchargeable sur https://sites.ualberta.ca/~stothard/1000_bull_genomes/. Cette séquence de 43 Mb contient 4,5 Mb non couverts qui sont localisés principalement au début du chromosome et dont la séquence est inconnue et codée « N ».

Dans ce chapitre, je décris d'abord l'étude de la distribution des différents haplogroupes du chromosome Y dans la population bovine française à partir de données de génotypage, puis, l'analyse de la diversité génétique des chromosomes Y présents en race Holstein à partir de données de séquençage de génomes complets, et enfin, l'estimation de l'effet de deux haplogroupes sur la fertilité mâle en race Montbéliarde.

1. Etude 1 : Analyse de la diversité du chromosome Y dans les races françaises

Dans le chapitre précédent, nous avons étudié la diversité mitochondriale dans les races françaises. L'étude des lignées paternelles (LP) via l'analyse de la variation du chromosome Y permet d'apporter des informations complémentaires à celles apportées par le génome mitochondrial.

L'existence de schémas d'introgression et de croisements sexuellement biaisés a été démontré chez les bovins domestiques. Par exemple, l'introgression de gènes provenant de zébus chez les bovins taurins domestiques africains s'est faite principalement par voie mâle (MacHugh *et al.*, 1997; Edwards *et al.*, 2007a). Cela a produit des distributions asymétriques entre génome mitochondrial et chromosome Y, les autosomes représentant une image intermédiaire entre ces deux héritages uni-parentaux. Les analyses basées sur le chromosome Y peuvent donc apporter des informations complémentaires à celles présentées dans le chapitre 2.

Comme pour le génome mitochondrial, l'absence de recombinaison chromosomique en dehors de la région pseudo-autosomique (PAR) préserve les haplotypes originaux, et l'héritage du chromosome Y peut donc être retracé par l'étude des LP dans la population bovine. Le faible nombre de reproducteurs mâles et l'utilisation massive des taureaux « stars » ont fait du chromosome Y un indicateur sensible aux événements démographiques récents, tels que les goulots d'étranglement, les effets fondateurs et les expansions de la population (Edwards *et al.*, 2011). Par le passé, l'étude de la variabilité du Y a mis en évidence la présence de seulement trois haplogroupes : Y1 et Y2 chez les *Bos taurus* et Y3 chez *Bos indicus* (Götherström *et al.*, 2005; Edwards *et al.*, 2011). Les haplogroupes Y1 et Y2 sont dominants respectivement en Europe du Nord et du Sud, selon le gradient présenté dans la figure 22 *(Edwards* et al.*, 2011)*.

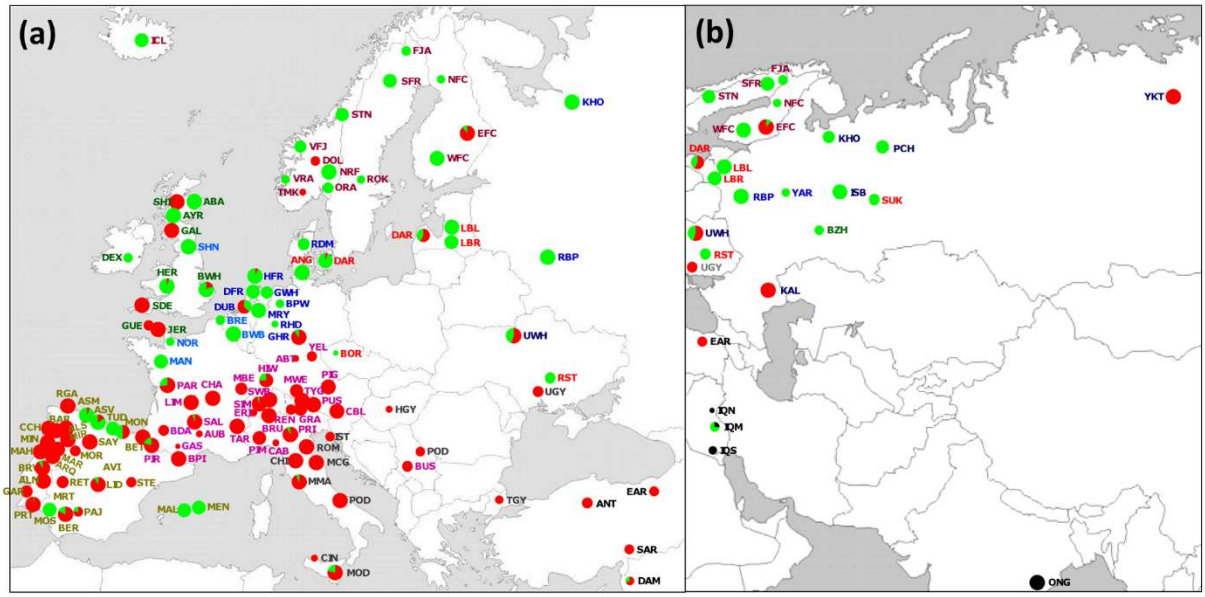


Figure 22 : Distribution géographique des haplogroupes du chromosome Y (Edwards *et al.*, 2011)
Les haplotypes Y1 sont représentés en vert, et Y2 en rouge.

L'étude qui suit a été réalisée en collaboration avec Chris Hozé (Allice, UMTeBis). Une première analyse avait été réalisée lors du montage du projet de thèse en 2016, exploitant des marqueurs de la puce Illumina BovineHD pour identifier des haplotypes du chromosome Y. J'ai ensuite utilisé ses résultats pour réattribuer les lignées des chromosomes Y parmi la population des taureaux autorisés à la reproduction.

1.1. Méthodes

1.1.1. Analyse des pedigrees

L'analyse des pedigrees a été réalisée à partir des informations de généalogie renseignées dans les bases de données nationales. La remontée des généalogies a été conduite avec le programme Lineage (Boichard, 2002). La contribution des plus vieux ancêtres mâles connus aux LP a été calculée sur la population des mâles nés entre 2015 et 2019 dont les deux parents sont connus. J'ai estimé le pourcentage d'animaux descendants de chaque ancêtre paternel et compté le nombre minimum d'ancêtres représentant 95% de la cohorte étudiée.

1.1.2. Génotypes et marqueurs

Les génotypages ont été réalisés dans le cadre du projet ANR GEMBAL (Génomique multiraciale des bovins allaitants et laitiers ; https://anr.fr/Projet-ANR-10-GENM-0014), ayant pour objectif d'étendre la sélection génomique à l'ensemble des races bovines française laitières et allaitantes sélectionnées. Ce projet a permis le génotypage d'un grand nombre d'individus de races diverses sur puce Illumina BovineHD. Parmi les 778 000 marqueurs de cette puce, 1224 sont positionnés sur le chromosome Y.

Nous avons utilisé le programme plink v1.07 (Purcell *et al.*, 2007) pour filtrer les animaux et marqueurs et générer les haplotypes. Sur les 5059 individus génotypés, 240 n'ont pas été pris en compte car leurs génotypages étaient incomplets. J'ai également éliminé 134 SNP dont les génotypes étaient manquants chez plus de 0,1% des individus. Enfin, parmi les marqueurs restants, seuls 13 n'étaient pas monomorphes et ont permis de différencier les haplogroupes. J'ai ensuite utilisé les informations trouvées dans la littérature concernant les aires de répartition des haplotypes Y1 et Y2 pour en déduire les allèles Y1 et Y2 (Götherström *et al.*, 2005; Edwards *et al.*, 2011). Les marqueurs utilisés pour l'identification des haplotypes du chromosome Y sont présentés dans le tableau 29. La position des marqueurs est indiquée par rapport à l'assemblage du chromosome Y utilisé par le projet 1000 génomes bovins (Bellott *et al.*, 2014).

L'objectif initial de ma thèse était d'exploiter la grande quantité de typages sur la puce EuroG10K pour réaliser de nouveaux comptages. Cependant, seuls deux des marqueurs discriminants l'haplotype Y1 de Y2 étaient présents sur la puce EuroG10k. Je suis donc repartie des résultats obtenus sur la puce HD pour réassigner les animaux récents aux différents haplotypes. J'ai donc remonté la généalogie, calculé le nombre de lignées pour chacun des haplotypes, réattribué les LP parmi les taureaux autorisés à la reproduction. La liste des taureaux mise à disposition par l'Institut de l'élevage a été téléchargée le 25/01/2020 depuis la plateforme API-Agro v2 (Idele, 2020). J'ai ensuite vérifié les discordances entre descendants d'une même lignée. L'identification de la LP de chacun des animaux a été effectuée avec le programme Lineage.

Position	Allèles	Séquence nucléotidique de l'amorce
3729035	A/C	AAGTGAGTTCTGAACACATTTACAAGTACTCTTTCTTATATATTACACTTTTCAACAAAA [**A/C**]AGTTACCATAGTAATAAAGTGAATTTGGTCCAGAAAACTCATGATGACTGTACCAATTAA
2915639	A/G	TCTGCCATACCTATTAAAAAGGGAGAAGAAGAATGTTTATCAAAAACAAAAGCAAAAATA [**A/G**]TAAGAATAGCAACTGCTGCTTTAATAAATATGTATACAGCTGCCTGTAATTTTTATGACT
3645679	A/G	AAACACACACCCTCCTCCTCCATCCTCTGCACCTCAAGCTCCAGCACCAGAGCTGGGGAT [**A/G**]GGACCAAAGGTTTTTTGCCCAACTTCCCATCACCTCATGGAAAACACGTCCTGAGGACCA
42636486	A/T	TTCCATGACAAAGGACTGCTTGGTTGGCTTGATGGAGAAATGGCACCTCCTCGCCCTCAA [**A/T**]TTGTCCTCAATAAATGGGTTGGGTCGAAACAATGAGACATCTGGAAACTTGCTGCAAGTA
3626919	A/T	GTCCATCACATCTGAAAACGATCCAGAAGTGATAAATTGCCTCACTGCTTTCAGGGCACC [**A/T**]TCTCCAGCAGCGGAGATGCTAGCTGTCCCGGGGCAGGGAGGTGTTGCTCCCTTGCACCCC
5332935	C/G	GTTCACAGTGTGCTCTATTCTCAACAACCTCAGAAAAAAGTTACCATCTCTAGAAAGATC [**C/G**]TTATACTTTCACAGAAGGTGACTGTGGGCTTCTACCTCAACAGGACATACAAAGGACCTA
3617345	T/C	TTCAGTCTCAAAAAGGCTGCCCGTTCCACGCATCCAGCCTTCCACGCATCCAGCCCTCAG [**T/C**]TATGTCTTCTGTTTCTCCTCCACAGAAGACTCATGTTGCATCTCAGGGGCCAACTGCTCC
4478777	T/C	GTGCATGGTATACCACTTGCTTTTTCTATGATCCAAACATCAATTCATATGCAAATCTTA [**T/C**]GATTAAATCACTAACTACACATTCCTGTGTTTGTTAGGGTTGTTTGTTTGTTTTGTGGTC
42642996	T/C	CTGTATTCAAATCCAGACTTTATTAAGATTAATAAGCTTAACAAACTTGTTTCTTATTGG [**T/C**]TCCATTATTTAGGTTATATGTTTCAATACAGACCTCTACTCCAGAAGAGAGGTGGGTAAT
3633742	T/C	TCCCTGCCAGGAGCATATGAAGAAGTGTCCTTACTGACCAGCAGAGAATGTAAATTTCAC [**T/C**]TTTGGGAGTAAAAATTTCACCACATACGCAATGGTACGAGGTTCTAATTACTTTTCTGCA
42627318	T/C	ATACAGAGTAACCGCTAGATATATTTTTACTGTCTTTAACTTTTTAATTTTCTTTTCTCT [**T/C**]TTTTTTTTTTTTTTTTAAGTTCTTTCATTTCAGTTCAGTTCAGTCGCTCAGTCGTGTTCG
28561229	T/G	AGGAGGCCTTACAAATAGCTGAGAAAAGAAGAGAAGTGAAAGGCAAAGGAGAAAAGAAAA [**T/G**]ATATTCTCATTTGAATGCAGAGTTAGGAGACTAGCAAGGAGTGATGAGAAAGCATTCCTC
42643285	T/G	GAAAAAAAAATCAAGGGAAGCAGAAGAATGGCATACATACCACAAGAAAATGACTAGAGC [**T/G**]AGTCAGTGACTTTATGGAAGTTGCTGGATACATATATGCAATAGATTATTACTCAGTCCA

Tableau 29 : Description des marqueurs du chromosome Y de la puce HD utilisés pour cette étude

1.1.3. Utilisation des données de séquences

J'ai utilisé l'annotation des variants des séquences de génome générées par mon équipe de recherche (G2B) dans le cadre de différents projets (n = 613 génomes de différentes races). J'ai comptabilisé les variants hétérozygotes comme des génotypes manquants. Parmi les 190 taureaux Holstein, j'ai éliminé 43 individus avec plus de 30% de génotypes manquants. J'ai ensuite éliminé les variants avec un taux de génotypage inférieur à 70% parmi les 147 animaux restants, et les variants dont l'allèle alternatif est fixé dans la population. Au total, j'ai étudié la variabilité du chromosome Y en race Holstein sur 147 taureaux et 260 variants.

1.2. Résultats et discussion

1.2.1. Analyse des pedigrees

Dans une première étape, j'ai étudié la distribution des LP dans la population mâle née entre 2015 et 2019 à partir des données de pédigrée. Le tableau ci-dessous présente quelques indicateurs décrivant les LP dans les six grandes races bovines.

	NOR	MON	HOL	BAQ	LIM	CHA
Nombre de mâles 2015-2019	425 477	718 037	3 436 432	355 998	604 519	847 854
Nombre d'ancêtres expliquant 95% des LP	4	6	2	15	18	11
Nombre d'ancêtres avec une contribution > 1%	4	6	3	11	12	9
Plus grande contribution par ancêtre(%)	56	38	49	63	39	23

Tableau 30 : Description des structures des races et des lignées

Calculée sur les mâles de parents connus nés entre 2015 et 2019. (NOR : Normande, MON : Montbéliarde, HOL : Holstein, BAQ : Blonde d'Aquitaine, LIM : Limousine, CHA : Charolaise)

Le nombre de LP est très réduit dans les populations bovines en comparaison des lignées maternelles, et parmi ce faible nombre, certaines LP ont une contribution très élevée comme en race Blonde d'Aquitaine avec un ancêtre représentant 63 % des LP de la population des mâles nés entre 2015 et 2019. Cela signifie qu'un nombre très faible de chromosomes Y génétiquement distincts ségrègent chez les mâles (destinés à la reproduction ou l'engraissement). Nous pouvons observer une nette distinction entre les races laitières et allaitantes. Les races laitières avec un recours plus important à l'IA ont connu une dérive génétique plus forte sur le chromosome Y. Ce phénomène est particulièrement marqué en race Holstein avec l'existence de deux lignées principales issues de FRUS01013415 (IDEAL W B né en 1956) et FRUS01322381 (R PEARL HA, né avant 1950) ayant chacune une contribution respective de 49,6% et 48,6%.

1.2.2. Distribution des haplotypes dans la population

Les résultats de cette première étude comptant le nombre d'animaux par haplotypes distincts sont présentés dans le tableau 31.a. Le tableau 31.b présente le nombre de lignées identifiées pour chacun des haplogroupes, et les résultats sur la population des taureaux autorisés à la reproduction sont indiqués dans le tableau 31.c)

	Haplotype	HOL	RDP	NOR	RFL	BPN	SIM	CHA	ABO	MON	PAR	AUB	BSW	SAL	BAZ	TAR	BLA	LIM	VOS	
Y1	CAATAAAGGGAGA	690		429	18		2		9	4										13
	CAATGAAGGGAGA		123	1	23	26		1												0
	CAATGGAGGAAGA										12									12
Y2	AGGAGGCACGTAA					3	120	619	186	498	66	242	99	237	45	175	324	436	36	2344
	AGGAGGCACGTAC																	1		1
	CGGTGGCACATAA																		16	16
a	NB d'animaux	690	123	430	41	29	122	620	195	502	78	242	99	237	45	175	324	437	52	

	Haplotype	HOL	RDP	NOR	RFL	BPN	SIM	CHA	ABO	MON	PAR	AUB	BSW	SAL	BAZ	TAR	BLA	LIM	VOS	
Y1	CAATAAAGGGAGA	9		7	2		2		3	4										7
	CAATGAAGGGAGA		15	1	10	7		1												0
	CAATGGAGGAAGA										8									8
Y2	AGGAGGCACGTAA					1	16	32	6	6	9	13	9	19	11	9	29	43	9	163
	AGGAGGCACGTAC																	1		1
	CGGTGGCACATAA																		2	2
b	Nb LP divergente										3							1		
	Nombre LP	9	15	8	12	8	18	33	9	10	17	13	9	19	11	9	29	44	11	

	Haplotype	HOL	RDP	NOR	RFL	BPN	SIM	CHA	ABO	MON	PAR	AUB	BSW	SAL	BAZ	TAR	BLA	LIM	VOS	
Y1	CAATAAAGGGAGA	12929		3370	33		52		3	42										45
	CAATGAAGGGAGA		189	62	59	41		1												0
	CAATGGAGGAAGA										104									104
Y2	AGGAGGCACGTAA					5	376	2500	458	3490	40	155	536	179	82	382	794	658	67	6841
	AGGAGGCACGTAC																	322		322
	CGGTGGCACATAA																		40	40
c	NB d'animaux	12929	189	3432	92	46	428	2501	461	3532	144	155	536	179	82	382	794	980	107	

Tableau 31 : Répartition des haplotypes
a : par nombre d'individus typés, b : par nombre de lignées, c : par nombre d'individus dont l'origine est réattribuée. Les abréviations des noms de races sont renseignées dans le chapitre 2, tableau 5 et dans la liste des abréviations.

Une lignée en race Limousine était divergente pour un individu, le seul animal possédant l'haplotype AGGAGGCACGTAC, différant d'un seul marqueur avec l'haplotype présent chez tous les autres individus. S'il ne s'agit pas d'un artefact, ce qu'il conviendrait de vérifier avec un second génotypage, cet animal pourrait représenter une source de variabilité génétique intéressante.

Dans la majorité des races, on retrouve un haplotype très majoritaire appartenant à l'haplogroupe Y1 ou Y2, ce qui démontre la très forte dérive génétique associée au chromosome Y. Dans le cas des races dans lesquelles les deux haplotypes Y1 et Y2 ségrègent, des signes d'introgression ont été confirmés par examen des pedigrees avec notamment des lignées mâles d'origine Holstein en races Montbéliarde et Abondance. On peut également observer des introgressions intra-haplogroupe avec l'utilisation dans les années 1970 de taureaux d'une race Norvégienne appelée Télémark pour ajouter de la

variabilité génétique en race Vosgienne. Comme observé dans la littérature, deux haplogroupes sont présents selon un gradient Nord/Sud et Est/Ouest. L'haplogroupe Y1 est majoritaire dans les races d'origine nordique comme les races Holstein, Normande ou encore Rouge Flamande, tandis que dans les races qui viennent plutôt du sud comme par exemple les races Charolaise et Salers, l'haplotype Y2 est majoritaire.

Sur les 34 534 taureaux d'IA répertoriés (toutes races), 28 838 appartiennent aux lignées dont l'haplotype est connu grâce aux génotypages sur la puce Illumina BovineHD obtenus dans le cadre du Projet GEMBAL.

1.2.3. Exploitations des données de séquences en race Holstein

L'analyse des séquences des chromosomes Y de 147 mâles Holstein a révélé la présence d'un très faible nombre de variants génétiques à l'échelle de la taille de ce chromosome (n = 260 variants sur 38 Mb). Le tableau ci-dessous présente le nombre de variants selon le nombre d'individus porteurs de l'allèle alternatif.

Nombre de variants	Nombre d'individus porteurs de l'allèle alternatif
234	1
19	2
1	4
1	13
1	20
1	51
1	67
1	76
1	103

Tableau 32 : Nombre d'individus porteurs d'allèles alternatifs

On peut également observer que la majorité des variants (234/260) ne sont présents que chez un seul animal. Ces résultats sont compatibles avec le fort goulot d'étranglement observé dans la première étude de ce chapitre, avec l'existence de deux lignées mâles principales en Holstein remontant aux années 1950, et la survenue de mutations *de novo* depuis.

Afin de retracer l'héritage des mutations à travers les lignées paternelles, j'ai utilisé les six variants dont l'allèle alternatif est porté par plus de 5 % des individus sur les 147 taureaux étudiés (soit n > 7/147). Le tableau 33 ci-après présente chacun des polymorphismes et le code lettre qui leur est attribué pour la suite de l'étude. Les individus hétérozygotes proviennent d'erreurs de séquençage et sont indiqués à titre indicatifs.

Code	Position	REF	ALT	Qualité	Nombre de génotypes manquants	0/0	1/1	0/1
A	43128337	G	T	14125,6	2	77	67	1
B	42507916	G	A	17911,5	3	68	76	0
C	42313722	ATG	ATGTG	39320,1	11	77	51	8
D	41064651	AT	A	2513,24	20	103	20	4
E	42436171	C	T	3093,14	4	130	13	0
F	42857798	AAGAG	A	39743,9	24	11	103	9

Tableau 33 : Présentation des variants et la répartition des allèles parmi les 147 taureaux Holstein de notre base de données

Le variant F (chrY : 42857798) n'a pas été pris en compte dans la suite des analyses car il est incohérent avec la définition des lignées (présent et absent chez au moins un individu dans 4 lignées sur 5) et l'arbre reproduit ci-dessous.

Pour pouvoir établir une chronologie entre ces cinq mutations, j'ai supprimé les animaux avec deux génotypes ou plus manquants ou douteux (car hétérozygotes) sur les cinq variants ainsi que deux animaux d'origine néerlandaise dont je n'ai pas pu retracer le pedigree. L'analyse retraçant l'héritage des mutations à travers les LP a porté sur les 127 animaux restants. J'ai retracé la transmission des variants dans le groupe d'individus et dessiné un arbre classant les LP en fonction de leurs allèles.

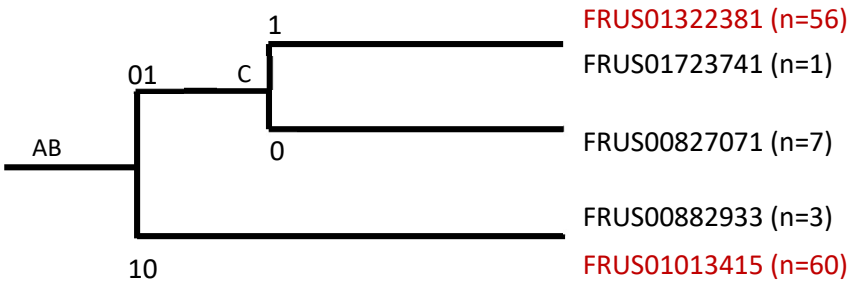


Figure 23 : Organisation phylogénétique des différentes lignées.

Les variants discriminants sont indiqués selon leur code défini dans le tableau 33. Les génotypes codés 0 pour l'allèle de référence et 1 pour l'allèle alternatif sont indiqués sur les branches. Les identifiants nationaux des ancêtres à la base de chaque branche sont présentés ainsi que le nombre « n » de leurs descendants analysés dans l'étude. En rouge, sont indiqués les deux principaux fondateurs nés dans les années 1950 contribuant à 98,2% des chromosomes Y présents dans la race.

Cet arbre phylogénétique, confirme l'existence de deux groupes de chromosomes Y représentés par 63 et 64 taureaux, qui se distinguent par deux marqueurs A et B. Ces deux groupes correspondent aux descendants des deux LP principales (issues des taureaux FRUS01322381 et FRUS01013415) et de quelques LP minoritaires. Nous n'avons observé aucun individu porteur des allèles alternatifs de chacun des deux variants. Un troisième variant (C) permet de distinguer deux sous-groupes au sein de l'une des branches principales, et de séparer la lignée FRUS00827071 des lignées FRUS01322381 et FRUS01723741.

De plus, parmi les 56 descendants de la lignée majoritaire du taureau FRUS01322381, deux variants (D et E) viennent délimiter trois groupes d'individus. Seuls 49 taureaux avaient un génotype pour ces deux

variants discriminants dans notre jeu de données. Je les ai étudiés pour retracer l'apparition de ces deux mutations D et E dans ce pedigree (cf. figure 24).

Au sein de ce pedigree, la grande majorité des individus sont 00, c'est-à-dire que comme leur ancêtre FRUS01322381 ils possèdent l'allèle de référence pour les deux variants D et E. Une première branche se distingue des autres par une mutation du variant D, et dont l'ancêtre commun à l'ensemble des neuf individus porteurs est le taureau VALIANT né en 1973. Une deuxième branche possédant les allèles alternatifs des variants D et E se différencie. Les informations de pedigree relient cette branche au taureau BLACKSTAR dont un autre descendant ne possède aucun allèle alternatif. Il parait peu probable que la mutation apparue chez VALIANT se soit produite une deuxième fois dans le pedigree de cette lignée paternelle. Il est possible que malgré le soin généralement apporté au suivi de la parenté chez les taureaux d'IA, une erreur d'assignation se soit produite entre 1983 et 2003 dans la généalogie du taureau PLANET, ancêtre commun aux 10 taureaux possédant les deux allèles alternatifs des variants D et E.

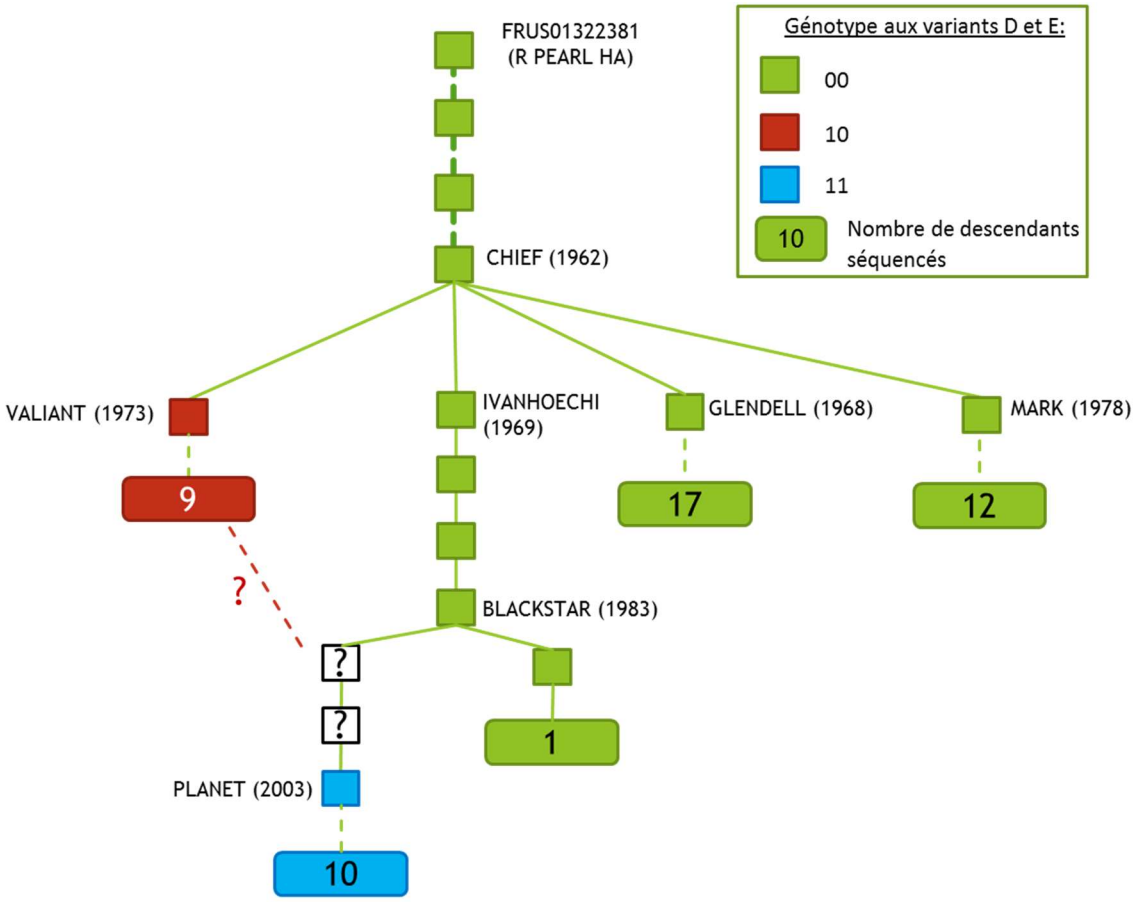


Figure 24 : Représentation du pedigree de la LP FRUS01322381 et de leurs génotypes pour les variants D et E, présentés dans cet ordre.

Les traits pleins indiquent un lien direct de parenté tandis que les pointillés signalent l'existence d'ancêtres intermédiaires entre deux individus. Le nom des taureaux clés est indiqué ainsi que leur année de naissance

1.3. Discussion

Nos résultats montrent une faible diversité génétique intra-race au niveau du chromosome Y. Le petit nombre de lignées Y est probablement une conséquence indirecte de la sélection artificielle intensive et de l'utilisation massive de l'IA. En 2015, une étude a mis en évidence que presque la totalité des taureaux d'IA Holstein descendent de deux ancêtres mâles nés en 1880. Ces deux lignées Y ancestrales ont continué à évoluer en Amérique du Nord grâce à deux fondateurs majoritaires qui ont contribué respectivement à 48,78 % et 51,06 % des taureaux dans la population Holstein de 2013 (Yue *et al.*, 2015). Nous retrouvons également la présence de deux lignées principales en race Holstein à la fois sur pedigree et sur données de séquences. Les deux ancêtres identifiés sont d'origine américaine et nés dans les années 1950. Nous n'avons pas d'information nous permettant de remonter leur pedigree jusqu'aux années 1880, et ainsi de déterminer si ces deux taureaux Holstein descendent des deux fondateurs identifiés par Yue *et al.* (2015). Toutefois, c'est très problable étant donné la forte dépendance de la génétique Holstein française vis-à-vis de la génétique américaine, et la similarité entre les contributions observées pour les deux principaux fondateurs des années 1880 pour la population américaine et des années 1950 pour la population française (48,8 et 50% contre 48,6% et 49,6%, respectivement).

Le faible nombre de variants génétiques identifiés ne nous permet pas d'estimer avec précision la vitesse d'accumulation des mutations sur le chromosome Y. En l'absence d'horloge moléculaire fiable nous pouvons toutefois remarquer que deux mutations sont apparues et se sont transmises au sein du pedigree de la LP de FRUS01322381 (R PEARL HA) sur une période de 70 ans (1950-2020). Sachant que seules deux polymorphismes (variants A et B) séparent les deux grands groupes de LP en race Holstein, dont les plus anciens représentants connus datent des années 1880 pour les données Américaines, nous pouvons estimer que l'ancêtre commun à tous les mâles Holstein le plus récent date de la première moitié du XIXème siècle (1800-1850) et qu'il est donc contemporain du début de la formation des races bovines (dont les premiers herdbook furent fondés au début du XIXe, Bernard et Baudemont (2016). Les premières importations enregistrées de Holstein aux Etats-Unis datent de 1852 (Lush et Holbert, 1936).

Dans cette discussion, je souhaite également aborder la problématique de l'origine, sujette à débat dans la littérature, des deux grands haplogroupes Y1 et Y2 observés chez les taurins. En se basant sur la distribution discontinue de deux haplogroupes distincts (Y1 et Y2) chez les taureaux taurins européens, et sur le génotypage d'échantillons d'aurochs mâles européens, l'hypothèse a été émise que la présence de l'haplogroupe Y1 uniquement chez les bovins d'Europe du Nord indiquerait une hybridation ancienne entre les aurochs mâles européens et les femelles domestiques (Götherström *et al.*, 2005). Cette hypothèse reposait principalement sur la présence de l'haplotype Y1 chez la quasi totalité des individus anciens de cette étude (Götherström *et al.*, 2005). La publication de Bollongino *et al.* trois ans plus tard démontre la présence de l'haplotype Y2 chez un panel d'animaux anciens européens.

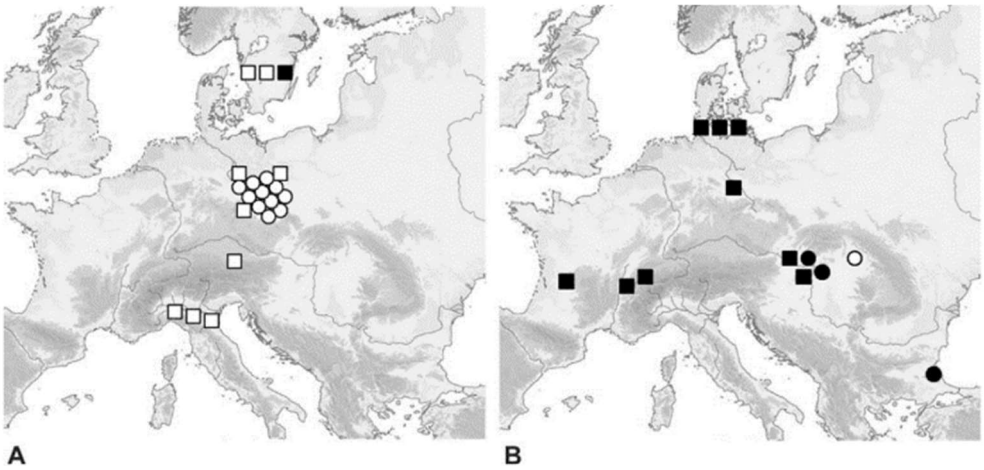


Figure 25 : Haplotype du chromosome Y issus d'échantillons anciens selon l'étude de a) Götherström *et al.* (2005) et Bollongino *et al.* (2008). Figure issue de Bollongino *et al.* (2008)
En blanc sont indiqués les haplotypes Y1 et en noir les Y2. Les carrés représentent les individus avec un haplotype mitochondrial P, et les ronds les T.

D'autres travaux ultérieurs ont également remis en question cette interprétation et il est maintenant généralement admis que les marqueurs génétiques du chromosome Y bovin actuellement disponibles (SNP et microsatellites) ne permettent pas de résoudre la question de savoir si l'un des deux chromosomes Y provient de l'introgression de taureaux aurochs sauvages européens dans les troupeaux domestiques (Svensson et Götherström, 2008; Edwards *et al.*, 2011; Bonfiglio *et al.*, 2012). Dans l'Europe médiévale et post-médiévale, où les possibilités d'échanges commerciaux à longue distance sont plus importantes, les bovins auraient pu facilement être déplacés d'une partie de l'Europe à une autre. La répartition des haplogroupes ne permet pas de déduire l'origine de ces deux haplotypes.

Il est important de noter que l'analyse de flux de gènes sur le génome nucléaire soutient également l'hypothèse d'une contribution génétique importante des aurochs britanniques vers les ancêtres des bovins domestiques d'Angleterre et de l'Irlande (Park *et al.*, 2015), ce qui va dans le sens d'une origine de l'auroch européen pour l'haplotype Y1. L'amélioration de notre connaissance du chromosome Y et de sa séquence de référence pourra permettre d'apporter plus d'éléments pour répondre à ce questionnement. Le séquençage du génome complet d'un plus grand nombre d'échantillons d'ADN anciens, d'animaux domestiqués et sauvages, sur des périodes variées pourra également permettre de retracer plus précisément la distribution des différents chromosomes Y et comprendre l'héritage laissé aux bovins modernes.

Résultats clés de cette étude :

- En étudiant les informations de pedigree, nous avons mis en évidence un nombre très petit de lignées paternelles au sein de chacune des grandes races bovines françaises.
- Nous avons confirmé, grâce à des données de séquençage de génomes complets l'existence d'une très faible variabilité génétique au niveau du chromosome Y en race Holstein.
- Les données de génotypages ont permis de mettre en évidence l'aspect géographique de la localisation des deux haplotypes Y1 et Y2 dans les races bovines françaises.

2. Etude 2 : Effet du chromosome Y sur la fertilité mâle en race Montbéliarde

Le chromosome Y porte des gènes intervenant dans la différenciation des testicules et la spermatogénèse (Bellott *et al.*, 2014). Les anomalies du chromosome Y sont souvent associées à de la sub-fertilité ou infertilité (Colaco et Modi, 2018). Comme nous l'avons détaillé dans le chapitre 2 (4.1.), la fertilité mâle est un enjeu important pour la filière bovine.

Parmi les races étudiées lors de l'analyse précédente sur la diversité des haplotypes du chromosome Y, seule la race Montbéliarde présentait un effectif de taureaux porteurs de chacun des deux haplotypes (Y1, Y2) a priori suffisant pour étudier l'effet des deux haplotypes sur les caractères de fertilité des mâles. Pour cela nous avons utilisé comme phénotype la fertilité moyenne à l'insémination, estimée dans le cadre de l'évaluation génétique de la fertilité des femelles.

2.1. Méthodes

2.1.1. Données

Le caractère utilisé dans cette analyse est un indice de fertilité mâle calculé à partir de l'évaluation génétique de la fertilité femelle. Il s'agit d'un effet aléatoire du taureau sur le taux de réussite à l'IA des femelles. Cet effet est corrigé de l'ensemble des autres effets du modèle utilisé par l'indexation officielle de la fertilité femelle, soit : le cheptel dans lequel est élevée la femelle, le mois et le jour d'insémination, l'inséminateur, le rang du vêlage. Le modèle inclut également l'intervalle vêlage-IA pour les vaches ou l'âge à l'IA pour les génisses. Tous ces effets sont estimés intra année. Le modèle inclut également le taux de consanguinité de la vache et le taux de consanguinité de l'embryon produit. Enfin, le modèle prend en compte la valeur génétique de la vache et son effet d'environnement permanent. Les IA réalisées avec de la semence sexée ne sont pas prises en compte dans notre jeu de données. L'indice de fertilité mâle est exprimé en pourcentage de réussite à l'IA, par taureau, par an, en écart à la moyenne de la population de base (les taureaux de race Montbéliarde nés de 2008 à 2011 pour les évaluations de 2019). Les performances des taureaux sont comprises entre -26 et 9, avec une moyenne de 0,90 et un écart-type de 2,93.

L'attribution des LP a été faite avec le programme Lineage du package Pedig (Boichard, 2002). Sur les deux lignées Y1 identifiées précédemment en Montbéliarde (étude 1), une seule était présente dans notre jeu de données.

2.1.2. Modèle

L'estimation des paramètres génétiques a été réalisée avec le logiciel Wombat (Meyer, 2007). Nous avons réalisé deux estimations pour deux effets différents : tout d'abord l'effet de la LP (effet LP) puis de l'haplotype (Y1 ou Y2) porté (effet HY). Les performances utilisées sont des performances corrigées pour tous les effets du modèle d'évaluation autres que l'effet du taureau.

$$Y_i = \mu + lp_i + a_i + \varepsilon_i$$

où: Y_i^* : Performance de l'animal : indice de fertilité mâle

μ : Moyenne générale des performances corrigées des filles

lp_i : Effet de la lignée paternelle

a_i : Effet génétique de l'animal, suivant une loi normale $\mathcal{N}$ (0, $\sigma^2\mathbf{A}$), avec **A** la matrice de parenté

et ε_i: Erreur résiduelle, suivant une loi normale $\mathcal{N}$ (0, σ^2_e)

L'effet LP a été considéré comme un effet fixe afin d'estimer l'effet de chaque lignée sur le caractère. Une deuxième analyse a été réalisée en considérant l'effet de l'haplotype du Y (HY) :

$$Y_i = \mu + hy_i + a_i + \varepsilon_i$$

où hy_i : Effet de l'haplotype Y1 ou Y2, les autres effets ayant la même définition que dans le modèle précédent.

Considérer les effets LP et HY comme des effets fixes, nous a permis d'estimer les effets de leurs différents niveaux, et ainsi d'estimer l'effet de chaque lignée ou haplotype.

2.2. Résultats et discussion

L'héritabilité du caractère estimée lors de cette analyse est de 0,44, ce qui est plutôt élevé comparé à l'héritabilité des caractères de fertilité femelle bien souvent inférieure à 0,10 (Jamrozik *et al.*, 2005). Cette valeur s'explique par le nombre d'inséminations prises en compte dans chaque phénotype. Ainsi, si l'héritabilité à l'échelle de chaque insémination est de l'ordre de 1%, l'héritabilité de la moyenne de 100 inséminations est de $\frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2/100}$, de l'ordre de 0,5.

Le tableau ci-dessous présente l'estimation des effets des différents niveaux de l'effet LP sur l'indice de fertilité mâle.

LP	HY	Nombre d'animaux	Moyenne brute	Solution
FR0000011189	Y2	897	0,79	0,23
FR0000011228	Y2	7	-0,92	-0,99
FR0000011594	Y2	342	1,38	0,20
FR0000012571	Y2	158	0,34	-0,02
FR0000013705	Y2	461	0,73	-0,04
FR0000014347	Y2	677	0,92	0,28
FR0000016403	Y2	221	1,67	0,76
FRUS01599157	**Y1**	**38**	**-0,72**	**-0,41**

Tableau 34 : Estimation des effets de chaque LP.

La moyenne brute représente la moyenne des performances intra-lignée avant correction des effets lignée paternelle et animal. Les solutions représentent une estimation de l'effet de chacune des LP sur l'indice de fertilité. Nous pouvons constater que si avant correction de l'effet génétique animal la LP du taureau Y1 semble défavorable, l'écart entre l'effet de celle-ci et des autres lignées n'est plus aussi important. Il est important de noter qu'une lignée Y2 (FR0000011228) semble également défavorable mais n'est estimée qu'à partir de sept animaux qui sont tous des fils ou petits-fils de GIBET MS né en 1991. Il est possible que cet effet soit un effet dû au génome nucléaire de leur ancêtre commun qui est relativement proche.

Dans cette analyse, nous pouvons voir que certaines lignées semblent plus favorables que d'autres. Ces résultats sont à relativiser sur certaines lignées dont l'effet est calculé sur un petit nombre de taureaux.

Dans un deuxième temps, nous avons comparé directement l'ensemble des lignées Y2 à la lignée Y1.

HY	Nombre d'animaux	Moyenne brute	Solution
Y1	38	-0,72	-0,30
Y2	2763	0,92	0,30

Tableau 35 : Estimation de l'effet de l'haplotype Y

Le nombre de LP est très faible et le nombre d'individus par lignée très déséquilibré. Considérant le petit nombre de taureaux porteurs de l'haplotype Y1, nous avons imaginé un dispositif pour comparer directement les performances de taureaux porteurs de l'haplotype Y1 de ceux porteurs de Y2. Le principe consiste à comparer deux à deux les performances de taureaux partageant un même grand-père, mais une fois sur la voie paternelle (grand-père paternel G) et une fois sur la voie maternelle (grand-père maternel GPM). En comparant les performances de taureaux porteurs de l'haplotype Y1 ou Y2 chacun possédant 25% du génome nucléaire d'un même grand-père mais sans avoir le même haplotype au chromosome Y, cela nous permettrait de corriger l'effet génétique dû au génome nucléaire du grand-père.

Malheureusement, les 38 individus porteurs de l'haplotype Y1 n'avaient que six grands-pères différents et l'un d'entre eux (FR0186008878) était commun à 30 des 38 descendants. De plus, FR0186008878 n'était GPM que d'un seul taureau Y2. Cela nous a permis de mettre en évidence l'apparentement très proche entre les taureaux porteurs de l'haplotype Y1 (tous descendants de la même LP). Du fait de ce déséquilibre entre les structures familiales sur les voies mâle et femelle, ce dispositif ne nous permettrait pas de différencier un effet de HY1 d'un effet dû à un effet génétique du grand-père commun à une grande partie des taureaux Y1.

Nous avons également étudié l'effet de la LP sur les données de qualité de semence récoltées en centre de collecte (volume de l'éjaculat, pourcentage de vivant, note de motilité, pourcentage de spermatozoïdes vivants), et n'avons mis en évidence aucun effet. Seulement six taureaux sur les 1454 individus étudiés étaient porteurs de l'haplotype Y1, ce qui ne nous a pas permis d'étudier l'effet de l'haplotype sur ces caractères.

Nous avons démontré dans le chapitre précédent que les informations basées sur la généalogie n'étaient pas toujours exactes. Il est donc possible que, dans cette étude, l'attribution de la LP soit parfois faussée. Cependant, dans la mesure où la généalogie des mâles d'IA est généralement mieux suivie sur la voie mâle (cf. chapitre 2 partie 1), le taux d'erreur associé devrait être relativement minime.

Les liens de filiation mâle-mâle sont généralement plus fiables car ils sont contrôlés depuis plus longtemps que le début des contrôles de parenté sur SNP. Ils étaient contrôlés sur des marqueurs de type microsatellite ou bien encore par groupes sanguins et ces contrôles étaient obligatoires depuis leur mise en place technique.

Conclusion et perspectives

L'introgression de l'haplotype Y1 dans la race Montbéliarde est relativement récente, mais à terme l'effet fondateur de FR0186008878 va s'estomper et ses descendants vont petit à petit se répartir de façon plus homogène et équilibrée sur les voies maternelles et paternelles des futures générations de taureaux d'IA, nous permettant de mieux décrire l'effet de ce chromosome Y sur la fertilité mâle. Aucun haplotype Y1 de race Montbéliarde n'étant disponible parmi les séquences de génome complet disponibles au sein de l'équipe G2B au moment de l'étude, nous n'avons pas comparé les variants entre ces deux haplotypes. Nous pourrions faire une estimation de cette variabilité en comparant les haplotypes Y2 avec un haplotype Y1 Holstein mais cette estimation serait biaisée, ne sachant pas lequel des haplotypes Y1 a été transmis dans la race Montbéliarde.

Enfin il est important de rappeler que la qualité de l'assemblage du chromosome Y est très inférieure à celle de l'assemblage des autosomes, avec notamment de nombreux trous de séquence. L'étude des polymorphismes du chromosome Y pour une comparaison fiable des haplotypes ne pourra se faire qu'avec l'amélioration de la séquence de référence. La création d'un nouvel assemblage du chromosome Y en race Holstein a été présentée lors du congrès Plant and Animal Genomics de San Diego début 2020 mais, à ma connaissance, il n'est pas encore disponible dans les bases publiques (Hua *et al.*, 2020).

Résultats clés de cette étude :

- Nous avons étudié l'effet des lignées paternelles sur un indice de fertilité mâle.
- Nous avons mis en évidence une structure familiale forte des taureaux Montbéliards Y1, et le déséquilibre dans le schéma de transmission par voie mâle et femelle de la lignée Y1 ne nous a pas permis de conclure quant à un effet de l'un ou l'autre des chromosomes Y.

Chapitre 4 : Etude du chromosome X et anomalies génétiques

Afin de faciliter la transition des chapitres précédents vers une thématique concernant plus particulièrement les anomalies génétiques, j'ai choisi de présenter plus en détail cette thématique qui me tient à cœur, en décrivant les principales méthodes d'identification utilisées et les enjeux liés à leur maîtrise en élevage et en sélection animale.

1. Etat de l'art sur l'identification de polymorphismes et d'anomalies génétiques

1.1. Définitions des anomalies génétiques

Les anomalies génétiques se caractérisent par une série de symptômes définissant un tableau clinique rare dans une population dont la cause est d'origine génétique et non environnementale. (Boichard *et al.*, 2016a). Les anomalies génétiques ne sont pas, à proprement parler, toutes délétères en élevage bovin. On peut citer par exemple l'absence de cornes (« polled ») qui est recherchée en tant que solution alternative à l'écornage ou l'anomalie de coloration « milka » en Montbéliarde, due à une dilution des pigments dans le poil (ONAB, 2014) et n'affectant pas la carrière productive des animaux. Dans ce chapitre, nous nous intéresserons principalement aux anomalies délétères.

Les anomalies génétiques sont causées par des mutations de l'ADN qui, dans la plupart des cas, sont neutres, c'est-à-dire qu'elles n'ont pas d'effet sur une fonction biologique. Lorsque ces mutations touchent une région codante ou régulatrice du génome, elles peuvent modifier la structure d'une protéine, altérer sa fonction, ou encore la quantité de protéine produite. Chaque individu possède plusieurs millions de variants génétiques qui le distinguent d'un autre individu pris au hasard dans la même espèce, et seulement quelques dizaines de ces variants proviennent d'événements de mutations « *de novo* ».

Différents mécanismes peuvent influer sur la fréquence d'un variant génétique dans une population. Une mutation apportant une amélioration du succès reproducteur d'un individu est amenée à être sélectionnée de manière naturelle ou artificielle (activité humaine) augmentant ainsi sa fréquence. Un autre phénomène pouvant faire évoluer la fréquence d'un polymorphisme est la dérive génétique. Il s'agit de la variation aléatoire des fréquences alléliques au sein d'une population et au cours des générations, liée au nombre fini de reproducteurs. Plus l'effectif des reproducteurs est restreint, plus la fréquence d'un allèle est aléatoire entre deux générations (Boichard *et al.*, 2016a). Or, les populations bovines sont génétiquement petites, avec un nombre d'ancêtres efficaces de seulement quelques dizaines d'individus en races laitières (cf. tableau 2 chapitre 1). La fréquence d'un allèle peut brutalement augmenter comme pour le variant responsable de l'anomalie *Bovine Leukocyte Adhesion Deficiency* (BLAD) passée de 0 à 6% dans la population Holstein française en 4 ans (Boichard *et al.*, 2016a). Si un fondateur porteur d'une anomalie génétique récessive contribue pour 3% au pool génétique d'une race, il sera responsable d'une augmentation de fréquence de cette anomalie de 1,5 % (3 %/2). Le polymorphisme se propagera à l'état hétérozygote dans la population jusqu'à l'apparition des premiers individus homozygotes atteints. Les homozygotes sont issus de deux parents porteurs, eux-mêmes descendant du même ancêtre porteur. C'est donc la consanguinité qui induit cette apparition d'homozygotes. Le « star système » encourageant indirectement l'utilisation massive de quelques taureaux a ainsi favorisé l'émergence d'une quinzaine d'anomalies génétiques récessives en race Holstein depuis les années 1970. Dans les décennies 1990 et 2000 en France, les deux premières anomalies récessives qui ont nécessité un plan d'éradication, ont été BLAD et CVM (*Complex Vertebral Malformation*) respectivement (Shuster *et al.*, 1992; Thomsen *et al.*, 2006; Grohs *et al.*, 2016). En plus

de la dérive, il existe un autre phénomène appelé surdominance, qui va concerner les polymorphismes récessifs délétères à l'état homozygote mais présentant un avantage à l'état hétérozygote. C'est le cas par exemple d'une mutation du gène *KIF1C* (*Kinesin Family Member 1C*) en race Charolaise qui a été sélectionné favorablement pendant de nombreuses années avant l'émergence des premiers cas homozygotes. Cette mutation améliore la conformation musculaire chez l'hétérozygote mais est responsable d'ataxie progressive chez l'homozygote et est létale pour l'animal (mort vers 18 mois ; Duchesne *et al.*, 2018). Dans ces situations de sélection, la fréquence de l'anomalie atteint parfois des niveaux élevés, 13% pour l'anomalie décrite en race Charolaise, ce que l'on n'observe généralement pas en situation de dérive.

Si les anomalies génétiques récessives sont, du fait de la structure des populations bovines, celles le plus fréquemment rencontrées et les plus délétères, il n'est pas rare d'observer également des anomalies avec des déterminismes génétiques différents, affectant la descendance d'un seul taureau. Il peut s'agir d'une mutation dominante associée à un mosaïcisme germinal, c'est-à-dire n'affectant qu'une partie des cellules du géniteur qui sera phénotypiquement sain, contrairement à la faible proportion de ses descendants qui auront reçu un gamète porteur de la mutation dominante et seront donc entièrement constitués de cellules porteuses de ce variant. Il peut également s'agir d'aneuploïdie (c'est-à-dire de monosomies et/ou trisomies partielles ou complètes) causées par des biais dans la ségrégation des chromosomes au moment de la gamétogénèse chez un individu porteur d'une translocation chromosomique équilibrée. Depuis le passage à la sélection génomique et l'arrêt du testage sur descendance des taureaux, qui agissait comme un filtre, l'ONAB recense l'émergence d'environ un à deux cas par an de taureaux porteurs de translocations équilibrées ou porteurs en mosaïque d'une mutation dominante. Les autres déterminismes génétiques (lié au sexe, dû à des mutations de l'ADN mitochondrial, oligogénique, etc.) ne sont presque jamais rencontrés, ou ne peuvent tout simplement pas être mis en évidence du fait du faible nombre de cas induits et de la structure des populations.

1.2. Méthodes d'identification des anomalies génétiques

On regroupe sous le terme génétique « classique » ou « forward », les approches génétiques partant d'un phénotype pour rechercher des variants causaux dans le génome des animaux atteints. La cartographie par homozygotie est une méthode utilisant le génotypage sur puce. Cette méthode repose sur la recherche de régions à l'état homozygote chez les atteints et rarement à l'état homozygote chez les autres (Charlier *et al.*, 2008). Si une mutation « m » se produit chez un individu, il la transmet ensuite avec les régions qui l'encadrent de chaque côté. Au fur et à mesure des générations, des évènements de recombinaison vont « casser » et donc réduire l'haplotype entourant le polymorphisme (figure 26). Après un certain nombre de générations, des descendants consanguins et homozygotes pour ce polymorphisme émergent révélant la présence d'une anomalie génétique. Le suivi des haplotypes recombinés associés au polymorphisme chez les individus atteints et les descendants sains permet de réduire la fenêtre chromosomique dans laquelle se trouve le polymorphisme. Le variant est localisé dans le plus petit haplotype homozygote commun à tous ces descendants atteints et non homozygote chez les descendants sains.

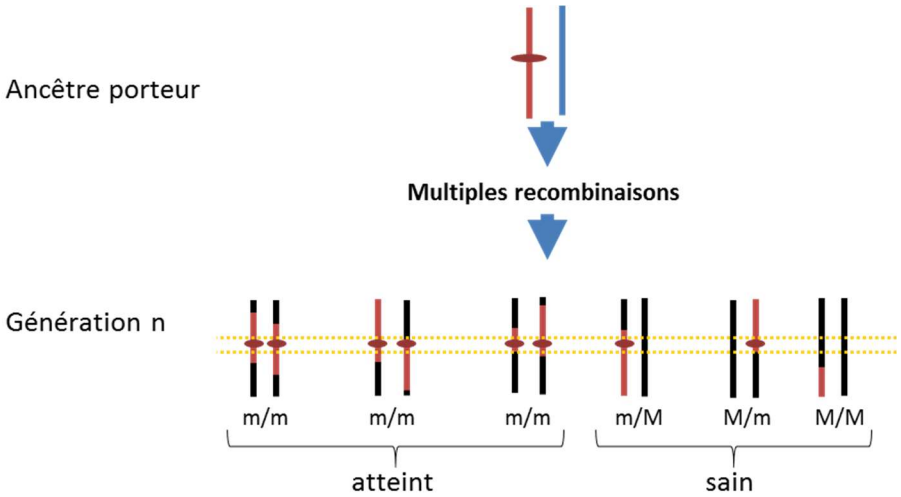


Figure 26 : Représentation du principe de la cartographie par homozygotie

Une mutation délétère symbolisée par l'ovale rouge apparait de novo chez un ancêtre et se transmet en association avec l'haplotype (en rouge) sur laquelle la mutation s'est produite. M représente l'allèle sain et m la version mutée du locus étudié

Cette approche utilisée par Charlier *et al.* en 2008 dans l'article qui fait référence en la matière chez les espèces d'élevage permet de réduire l'intervalle de localisation du polymorphisme. Une fois la région identifiée, des variants candidats peuvent être identifiés en utilisant par exemple les données de séquençage d'un voire deux animaux atteints porteurs du segment commun à tous les atteints. L'accès à un grand nombre de séquences d'animaux sains grâce au projet 1000 génomes a permis de donner plus de puissance à l'identification de variants candidats en opposant les variants observés chez les atteints à ceux de la population contrôle et de réduire ainsi le nombre de variants candidats.

L'accès aux séquences complètes d'un grand nombre de bovins, a également permis de développer d'autres approches de recherche de variants causaux. Depuis 2010, des approches de génétique « inverse » ont vu le jour et ont permis d'identifier des anomalies génétiques passées inaperçues et d'anticiper l'émergence de nouvelles anomalies génétiques. Le principe repose sur l'identification de variants a priori délétères pour le fonctionnement d'une protéine.

Un gène est structuré en régions régulatrices, en exons (parties codantes) et en introns (non-codantes) et l'impact d'une mutation affectant ce gène dépendra directement de sa nature et de sa position (Figure 27). En règle générale, les variants les plus délétères sont ceux qui affectent la transcription ou la traduction. On peut citer par exemple des délétions de tout ou d'une partie d'un gène, des variants affectant les sites d'épissage, le gain d'un codon stop au milieu du gène (mutation non-sens) qui va arrêter la traduction, et les mutations dites « frameshift » qui décalent le cadre de lecture. Les substitutions d'un acide aminé par un autre ou mutations faux-sens ont souvent un effet phénotypique moins marqué que les précédentes sauf si elles changent les propriétés de la protéine dans un domaine fonctionnel important. Enfin, les variants localisées au niveau des régions régulatrices de l'expression de gènes conduisent à une augmentation ou à une diminution de la quantité de protéine produite et ont ainsi le plus souvent un effet phénotypique quantitatif (de type QTL) plutôt que qualitatif.

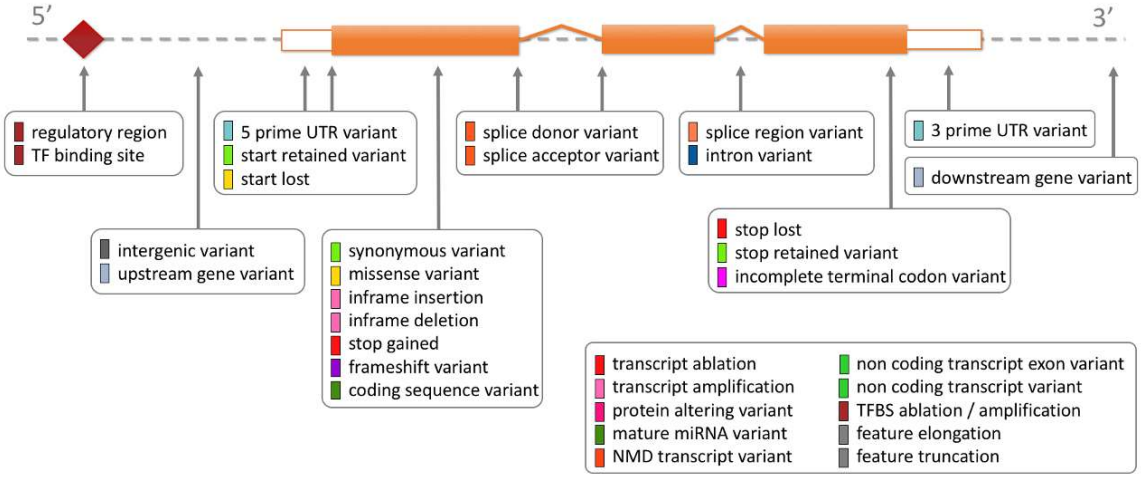


Figure 27 : Représentation d'un gène et de la localisation des variants (Ensembl, 2020)

La prédiction des effets des différents variants sur un phénotype précis repose sur différentes sources : les données phénotypiques recueillies chez l'humain et les espèces modèles ou domestiques, la conservation des polymorphismes entre espèces et l'annotation fonctionnelle de l'effet du variant mais aussi du gène.

Les données de séquences générées au sein de l'équipe mais aussi les variants renseignés dans le projet 1000 génomes bovins constituent une base répertoriant un grand nombre de variants d'intérêt. Des tests sont régulièrement dessinés pour les variants dont le phénotype semble le plus intéressant et sont ensuite placés sur la puce à SNP. Lorsque suffisamment de génotypes ont été récoltés, des analyses phénotypiques permettent de valider ou non le phénotype délétère prédit.

Une autre approche proposée par VanRaden et ses collaborateurs (2011) utilise les données de génotypage produites en masse dans le cadre des évaluations génomiques pour identifier les haplotypes présentant un déficit significatif en animaux homozygotes et ainsi cartographier des variants responsables de létalité embryonnaire ou périnatale récessive. Si un haplotype est observé à une fréquence assez élevée dans une population mais n'est jamais observé à l'état homozygote, cela signifie que les animaux homozygotes n'ont jamais été génotypés. Deux raisons peuvent expliquer ce biais de génotypage : soit l'homozygotie est létale à l'état embryonnaire ou dans les jours suivants la naissance de l'animal, soit les atteints présentent un phénotype anormal et ne sont pas gardés pour la sélection.

Ces dernières années, de nombreuses études reprenant cette stratégie ont été réalisées dans différentes races (Fritz et al., 2013, 2018; Daetwyler et al., 2014; McClure et al., 2014; Pausch et al., 2015; Michot et al., 2016) conduisant à l'identification de dizaines de nouveaux défauts récessifs. En amont puis en parallèle de mes travaux de thèse, j'ai contribué à la caractérisation de deux de ces polymorphismes et présenté ces résultats au *World Congress Applied to Livestock Production* (WCGALP) de 2018 (Escouflaire *et al.*, 2018; Fritz *et al.*, 2018; Hozé *et al.*, 2020).

L'identification des haplotypes létaux à l'état homozygote permet de constituer un test sur haplotype pour les reproducteurs et de tester l'effet d'être porteur ou non sur le taux de conception et le taux de non-retour[7] dans les accouplements à risque (entre parents porteurs ou entre un taureau porteur

[7] Taux de femelles ne revenant en chaleur après une insémination

et la fille d'un taureau porteur). Des variants candidats présents dans la fenêtre définie par l'haplotype sont également recherchés dans les séquences des animaux porteurs ou des principaux contributeurs de la race. Il arrive parfois que la mutation ne soit pas située directement sur l'haplotype détecté. C'est le cas par exemple si l'haplotype incluant le polymorphisme existe aussi à une fréquence élevée sans le polymorphisme. Dans ce cas, l'étude d'haplotypes un peu plus éloignés mais en fort déséquilibre de liaison peut permettre d'identifier l'anomalie. L'utilisation de données de séquences couplée à la recherche de déficit en homozygotes a permis d'identifier plus d'une dizaine de mutations létales à l'état homozygote (Hozé *et al.*, 2020).

Les approches de génétique inverse permettent d'étudier la fonction de gènes méconnus. Néanmoins, ces méthodes ne sont pas infaillibles et nécessite un important travail de validation. En effet, il y a un grand nombre de faux positifs parmi les variants de type perte de fonction (gain de codon stop, décalage du cadre de lecture ou mutation affectant un site d'épissage). Charlier *et al.* (2016) ont estimé que seuls 15% des variants prédits « perte de fonction » et 6% des variants non-synonymes causaient réellement une létalité à l'état embryonnaire. La recherche de déficit en homozygotes produit bien moins de faux-positifs mais nécessite une grande taille de population génotypée pour espérer détecter des déficits pour des allèles avec une fréquence faible. En général, une population de 2000 animaux est suffisante pour détecter les haplotypes responsables de mortalité avec une fréquence de 10%, mais plus de 200 000 animaux sont nécessaires pour obtenir des résultats significatifs pour les haplotypes avec une fréquence de 1% (Hozé *et al.*, 2020). Le déséquilibre de liaison parfois incomplet entre un haplotype et la mutation causale peut également entrainer la présence de faux positifs, des taureaux possédant l'haplotype mais sans la mutation, ce qui peut conduire à une diminution de la puissance de détection.

Pour conclure, chaque individu est porteur de mutations délétères. De par leur effectif génétique réduit qui induit à la fois dérive et consanguinité, les populations bovines voient augmenter le risque d'émergence d'anomalies génétiques récessives. Des observatoires et réseaux de surveillance ont été mis en place pour détecter et répertorier les cas d'anomalies. Les nouvelles technologies de séquençage et le génotypage à grande échelle sont de puissants outils pour identifier les variants causaux. L'amélioration des connaissances du génome bovin et les bases de données existantes répertoriant les phénotypes mutés dans d'autres espèces comme l'être humain, la souris, et le chien permettent d'augmenter la rapidité de résolution des cas.

1.3. Enjeux et identification

Afin de faire face à l'émergence des anomalies génétiques et de pouvoir identifier les variants causaux, des observatoires chargés de répertorier des cas ont été mis en place dans de nombreux pays en Europe. Ces observatoires reposant principalement sur un réseau de vétérinaires ont permis d'identifier de nombreuses anomalies génétiques. Le réseau le plus ancien a été créé au Danemark et a permis d'identifier la mutation responsable du CVM (Thomsen *et al.*, 2006). L'université de Liège en Belgique a également identifié de nombreux variants causaux grâce à sa cellule de surveillance des défauts génétiques dans la race blanc bleu belge comme celles d'un type de nanisme et du syndrome de la queue tordue décrit précédemment (Fasquelle *et al.*, 2009; Sartelet *et al.*, 2012).

En France c'est l'Observatoire National des Anomalies Bovines (ONAB) mis en place en 2002, qui recense les cas d'anomalies en centralisant les déclarations et les échantillons biologiques. Ses autres missions sont la veille scientifique, la sensibilisation des acteurs, la diffusion des résultats, et plus

généralement la coordination du réseau de détection (Grohs *et al.*, 2016). L'ONAB a permis d'identifier les variants causaux de nombreuses anomalies génétiques dont un décalage du cadre de lecture présent à une fréquence de 27 % dans la population Normande, responsable de dégénérescence progressive de la rétine (Michot *et al.*, 2016) ou encore de résoudre des cas de veaux dits « bulldog » en Holstein et dans la descendance croisée d'un taureau Charolais (Bourneuf *et al.*, 2017).

L'augmentation du nombre d'individus génotypés augmente la puissance de détection des anomalies génétiques et facilite l'étude du déterminisme associé (Duchesne *et al.*, 2016; Michot, 2017). Grâce à la détection par déficit en homozygotes, il est même possible désormais d'identifier des anomalies récessives liées à la mortalité embryonnaire ou chez les jeunes animaux (Fritz *et al.*, 2013, 2016; ONAB, 2019). Les tests génétiques sur haplotype ou variant causal sont effectués sur les animaux génotypés pour la sélection génomique, et le statut des animaux, notamment des mâles reproducteurs, est alors retransmis au propriétaire de l'animal. Les programmes d'accouplements peuvent ainsi prendre en compte les statuts des animaux proposés à la reproduction afin de limiter le risque d'apparition d'animaux homozygotes. Puisqu'un accouplement porteur x sain ne risque pas d'aboutir à la naissance d'un animal atteint, la gestion des accouplements et le génotypage des animaux permet d'éviter l'apparition de nouveaux cas (Cole, 2015).

Les stratégies de gestion des anomalies sont différentes en fonction du déterminisme génétique sous-jacent. Dans le cas d'une anomalie dominante très délétère par exemple avec mosaïcisme germinal, le taureau porteur doit être retiré des reproducteurs mais ses descendants sains (non porteurs) peuvent être gardés dans le schéma de sélection. Lorsqu'il s'agit d'une anomalie récessive, une élimination progressive des porteurs de l'anomalie est conseillée, afin de ne pas provoquer une diminution brutale du progrès génétique (Boichard *et al.*, 2016b), l'effort à consentir dépendant du coût par cas et de la fréquence allélique de l'anomalie. Comme le nombre d'anomalies découvertes a augmenté grâce à une amélioration de la puissance de détection, l'élimination systématique de porteurs d'anomalies parmi les reproducteurs est donc difficilement envisageable ; cela risquerait de créer un goulot d'étranglement génétique et donc de provoquer une chute de la diversité génétique , qui augmenterait la consanguinité et favorisait d'autant plus l'apparition de nouvelles anomalies génétiques (Howard *et al.*, 2017).

Dans certains cas d'anomalie, le statut des reproducteurs peut également être inclus dans les index de synthèse, en fonction de l'impact économique de l'anomalie, sa fréquence et du type de transmission (Hoze *et al.*, 2018). Dans les simulations effectuées par Hoze *et al.* (2018), sont prises en compte une anomalie entrainant la mort des veaux à la naissance (Brachyspina), deux anomalies responsables de mortalité chez les jeunes veaux (Mitochondropathie et CDH, Cholesterol Deficiency Haplotype) et plusieurs anomalies létales à l'état embryonnaire (MH1, MH2 en race Montbéliarde, HH1, HH2, HH3, HH4, HH5, HH6, HH7 en race Holstein). Sur le volet économique, dans les conditions françaises, le coût de la perte d'un embryon est inférieur à celui de la perte d'un jeune veau. Le coût d'une anomalie létale à l'état embryonnaire dépend du stade auquel survient l'avortement et du délai avant que sa mère puisse initier une nouvelle gestation. Dans le cas de mortalités juvéniles, le coût inclut l'entretien et les éventuels soins du veau jusqu'à sa mort. L'étude de Hoze et al 2018 estime ainsi le poids économique d'une anomalie entrainant de la mort embryonnaire à entre -70€ et -100€ selon le stade embryonnaire atteint lors de l'avortement, -130€ pour l'anomalie Brachyspina entrainant la mort de l'animal lors de sa naissance, et jusqu'à -400€ pour les anomalies responsables de mortalité juvénile.

2. Polymorphismes délétères et approche de génétique inverse

2.1. Préambule

Au cours d'un contrat de 4 mois précédant ma thèse, j'ai travaillé sur l'étude de variants prédites comme délétères. Ce travail faisait suite à la thèse de Pauline Michot dans l'équipe G2B (Michot, 2017) et entrait dans le cadre du projet Bovano, un projet financé par l'Agence Nationale pour la Recherche et APIS-GENE pour une durée de 5 ans portant sur l'identification et l'analyse fonctionnelle d'anomalies génétiques (Grohs *et al.*, 2016).

J'ai souhaité appliquer la même approche de génétique inverse sur les variants du chromosome X tout en prenant en compte certaines particularités du chromosome X qui rendent l'approche de génétique inverse plus difficile que pour les autosomes.

Le chromosome X est le deuxième plus grand chromosome du génome bovin (voir figure 28.a) mais en proportion de sa taille, il possède un nombre de variants identifiés moindre que les autosomes (figure 28.b).

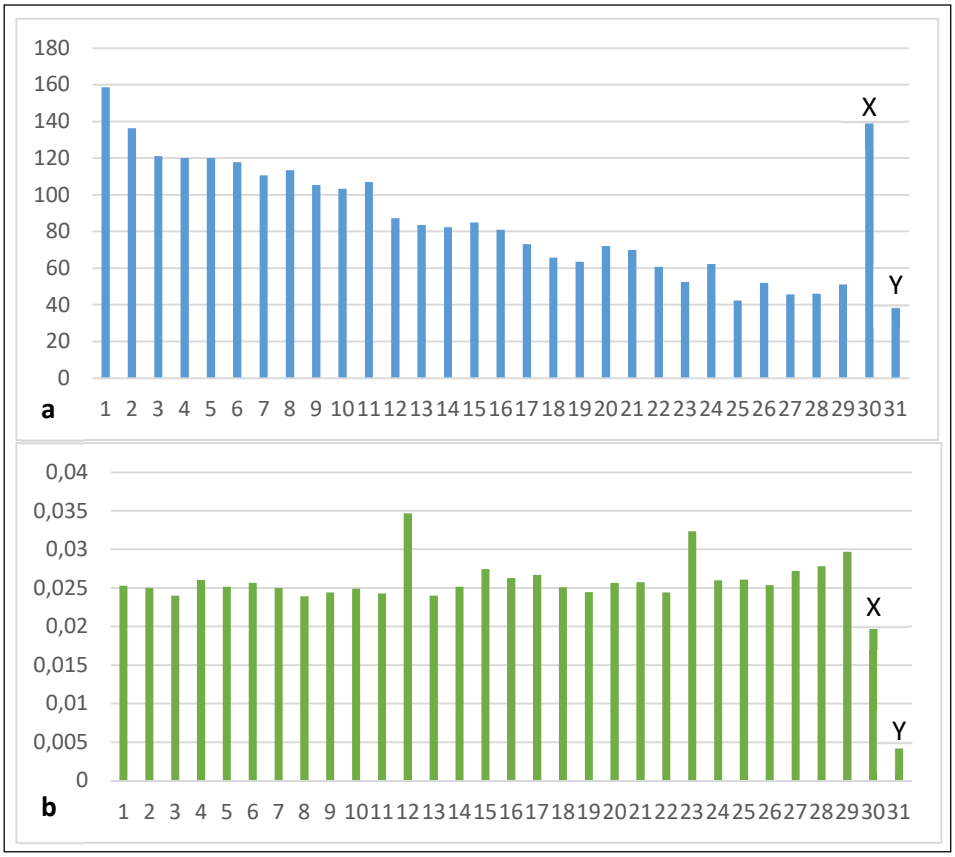


Figure 28 : Représentation de la taille des chromosomes en mégabases (a) et le ratio du nombre de variants identifiés sur le projet 1000 génomes par rapport à la taille du chromosome (b).

J'ai ensuite recherché si ce phénomène était observé chez une espèce sans sélection artificielle (l'humain) et une espèce chez laquelle le sexe hétérogamétique est différent (la poule dont les femelles sont ZW). J'ai utilisé pour cela le nombre de variants et la taille renseignés par Ensembl (par exemple : http://www.ensembl.org/Bos_taurus/Location/Chromosome?r=1). En l'absence de séquence disponible sur Ensembl pour le chromosome Y bovin, les calculs ont été effectués sur la séquence du chromosome Y de l'assemblage ARS-UCD1.2_Btau5.0.1Y sans prendre en compte les 6 Mb inconnus

au début du chromosome. Nous pouvons observer que chez ces trois espèces le nombre de variants sur le chromosome X (ou W) est inférieur à celui sur les autosomes (cf. figure 29 ci-après).

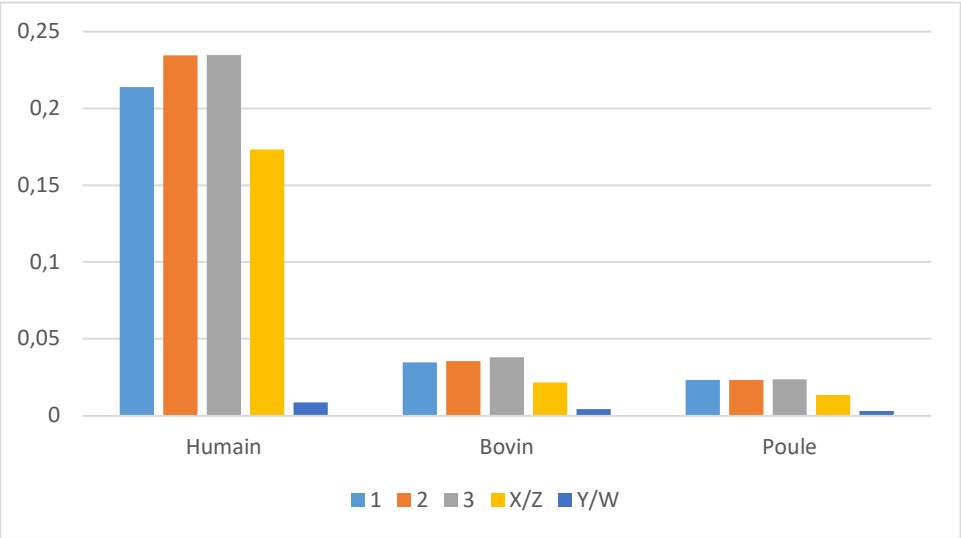


Figure 29 : Ratio entre le nombre de variants et la taille de chaque chromosome - exemple chez trois espèces sur les 3 premiers autosomes et les chromosomes sexuels.

En effet, la sélection favorise un taux de mutation plus faible sur le chromosome X que sur les autosomes, en raison des risques de polymorphismes récessifs délétères à l'état hémizygote (McVean et Hurst, 1997). La dérive est également plus forte sur le chromosome X que sur les autosomes. Le chromosome X du père et l'un des chromosome X de la mère ont donc plus de chance de s'éteindre. Cela augmente les risques d'avoir des faux positifs dans les annotations. Nous avons donc pris ce risque en compte et avons privilégié les phénotypes d'intérêt. Il est intéressant de noter que le nombre de polymorphismes sur le chromosome Y (ou W) rapporté à sa taille est très faible chez les trois espèces.

Deuxièmement, chez l'humain, des variants responsables de déficiences intellectuelles liées au chromosome X ont été identifiés sur 102 des 852 gènes répertoriés sur ce chromosome (Lubs *et al.*, 2012). Or, ce type de phénotype ne représente que peu d'intérêt lorsque l'on s'intéresse à des animaux d'élevage, ce qui réduit le champ des gènes d'intérêt de notre étude.

En novembre 2017, lors du design de la première série de marqueurs mitochondriaux, j'ai également sélectionné des polymorphismes situés sur le chromosome X prédits comme délétères et dessiné des marqueurs correspondants qui ont été placés sur la version 8 de la puce EuroG10k. Nous avons axé notre choix de variants grâce à la base de données de variants du run 6 du projet 1000 génomes bovins. Celle-ci répertorie près de 3 millions de variants présents sur le chromosome X.

Le nombre de marqueurs que nous pouvions placer sur la puce étant limité, nous avons dû faire une sélection. J'ai choisi de présenter en détail les critères habituellement utilisés dans le choix des variants ainsi que les particularités liées au chromosome X et le statut hémizygote des mâles.

2.2. Méthodes

2.2.1. Choix des variants

Le schéma ci-dessous synthétise le processus nous ayant amenés au choix des variants, les numéros 1, 2 et 3 faisant référence aux paragraphes numérotés ci-dessous.

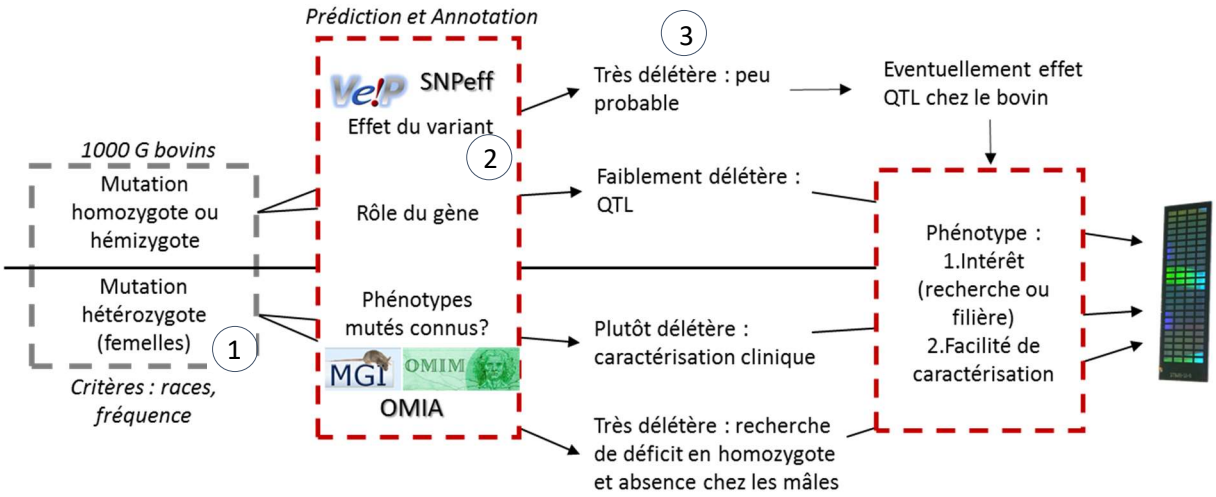


Figure 30: Arbre décisionnel pour le choix des variants

1) Exploitation de la base de données du projet 1000 génomes bovin et déterminisme génétique

La mutualisation d'un grand nombre de séquences de génome entier par le consortium 1000 génomes bovins permet de rechercher des variants délétères grâce aux approches de génétique inverse de façon particulièrement efficace. Les séquences partagées au sein de ce groupement sont majoritairement celles des taureaux les plus influents de chaque race. Cette base de séquence peut donc être à la fois utilisée pour identifier des polymorphismes récessifs délétères présentes à l'état hétérozygote chez certains individus et servir de contrôle pour réduire la liste des variants candidats dans le cas d'anomalies avec un déterminisme dominant. Nous avons utilisé le run6 de cette base en novembre 2017 pour choisir des polymorphismes sur différents critères et dessiner des marqueurs permettant l'identification massive d'animaux porteurs grâce aux puces de génotypage utilisées pour la sélection génomique.

Comparé aux autosomes, le cas des variants sur le chromosome X est très spécifique puisqu'hormis sur la région pseudo-autosomique partagée avec le chromosome Y, les mâles seront hémizygotes. Cela signifie qu'il est hautement improbable que les mâles soient porteurs de polymorphismes très délétères, sauf dans le cas où ces variants n'affectent que des caractères exprimés chez la femelle (fertilité femelle, production laitière). De même, il est peu probable que nous identifiions des anomalies létales à l'état homozygote puisqu'elles seront systématiquement létales chez les mâles et ne pourront être transmises.

2) Annotation et phénotype délétère

La détection des variants de type SNP et petits InDels (insertion-délétion) a été réalisée avec le logiciel Samtools (Li *et al.*, 2009) par le consortium 1000 génomes bovins. L'annotation des variants a été réalisée avec le programme *Variant Effet Predictor* (Ve!P) d'Ensembl (McLaren *et al.*, 2016) et les effets des substitutions d'acides aminés faux-sens ont été prédits par le logiciel SIFT (Kumar *et al.*, 2009). J'ai

considéré comme très délétère et retenu les variants entraînant : soit (i) une modification d'épissage, (ii) le gain ou la perte d'un codon stop, (iii) des modifications du cadre de lecture, et (v) la substitution d'un seul acide aminé prédite comme délétère par SIFT (score <0,05).

Lorsqu'une mutation induit un mauvais fonctionnement d'une protéine, si celle-ci n'intervient pas dans une fonction métabolique majeure, elle n'impactera pas la survie ou les caractères de production de l'animal. Il est donc important de prendre également en compte le rôle du gène en utilisant des bases de données comme GeneCards (https://www.genecards.org/) ou OMIM (https://www.ncbi.nlm.nih.gov/omim). J'ai aussi utilisé plusieurs bases de données recensant les caractéristiques phénotypiques d'animaux mutés. Je me suis appuyée sur la base d'annotation des phénotypes du modèle murin de *Mammalian Phenotype Ontologies* (Mouse Genome Informatics, accessible depuis http://www.informatics.jax.org/), la base *Online Mendelian Inheritance in Man* (OMIM) des syndromes génétiques chez l'être humain, et le répertoire développé par l'université de Sydney des anomalies génétiques répertoriées chez les autres espèces que l'humain, la souris et le rat : *Online Mendelian Inheritance in Animals* (OMIA ; https://omia.org/home/).

3) Choix des variants

La population des animaux répertoriés dans le projet « 1000 génomes bovins » est supposée saine, étant composée principalement de mâles sélectionnés comme reproducteurs. Il est donc a priori peu probable qu'un variant gravement délétère pour la santé de l'animal ou entraînant une stérilité mâle y soit présent à l'état homozygote ou hémizygote. Cependant on ne peut exclure la possibilité d'identifier des variants entraînant une baisse substantielle de la fertilité, de la résistance à des maladies ou enfin ayant un impact chez les femelles. Un variant prédit comme fortement délétère chez les espèces murines peut se révéler avoir un effet plus modéré chez le bovin.

Pour le choix des variants, nous avons privilégié ceux pouvant avoir un effet sur la fertilité (MGI, OMIM). François Vialard, directeur du secteur génétique de l'hôpital Poissy-Saint-Germain et travaillant dans l'unité de Biologie de la Reproduction Environnement, Epigénétique et Développement (BREED) de INRAE nous a fourni une liste de gènes d'intérêt suspectés d'être impliqués dans des anomalies spermatiques.

2.2.2. *Design* des marqueurs et calcul des fréquences

Pour dessiner les marqueurs, j'ai récupéré les régions de 100 pb encadrant chacune des variants choisis sur le navigateur de génomes de l'UCSC (http://genome-euro.ucsc.edu, assemblage UMD3.1.) et inséré entre crochets les deux allèles possibles au variant (P.ex. : …CATGC[A/G]CCTCT…). Dans le cas des InDels, le design des marqueurs a été adapté afin de mimer des SNP tout en imposant une amorce en ajoutant des nucléotides inconnus « NN » de l'autre côté.

Les fréquences génotypiques et alléliques ont été calculées pour chacun des marqueurs par comptage, dans les six grandes races en distinguant mâles et femelles, à partir des génotypages des puces EuroG10k et EuroGMD entre juin 2018 et janvier 2020.

Résultats et discussion

Au total, 31 marqueurs ont été ajoutés sur la puce, leur choix reposant sur les raisons qui suivent. Seize ont été choisis principalement pour le phénotype délétère associé au gène chez la souris dont trois variants prédits comme responsable d'un gain de codon stop. Trois variants situés sur le gène *TEX11* (*Testis expressed 11*), gène d'intérêt pour l'unité BREED, ont été ajoutés ainsi que huit marqueurs choisis pour leur localisation proche ou dans *XIST* (*X-inactive specific transcript*), en lien avec les travaux présentés dans la partie 3 de ce chapitre. Enfin, quatre variants ont été sélectionnés car ils affectent des gènes identifiés dans des QTL de fertilité mâle (de Camargo *et al.*, 2015).

Le recours à la génétique inverse avait pour objectif d'utiliser des données des séquences pour détecter des variants candidats responsables d'anomalies et des polymorphismes pouvant avoir un effet sur la fertilité mâle ou femelle ou entraînant un biais d'inactivation du chromosome X. Les fréquences sont calculées sur des animaux génotypés dans le cadre de la sélection génomique. Ce sont donc, a priori, des animaux en bonne santé lors du prélèvement servant au génotypage. Une première étape consiste donc à rechercher un éventuel déficit en homozygotes chez les femelles et/ou de porteurs chez les mâles. Le tableau 36 ci-dessous recense la fréquence allélique chez les mâles et les femelles. La position des marqueurs est indiquée sur l'assemblage UMD3.1 utilisé lors de l'annotation des variants sur le Run6 du projet 1000 génomes bovins.

Parmi ces marqueurs j'ai choisi de détailler le choix d'un variant de chaque type (tableau 36). Le variant situé sur *XIST* est annoté comme un variant intergénique car ce gène n'est pas placé sur le génome bovin dans la base Ensembl utilisée pour réaliser cette annotation.

Gène	Variant	Position (UMD3.1)	AA	Phénotype (source : MGI)	Choix
TAZ	Faux-sens délétère (0,03)	40386059	N/S	Les souris homozygotes pour un allèle knock-out présentent une infertilité masculine, une réduction de la taille des testicules, une azoospermie et une spermatogenèse anormale.	Phénotype
XIST	Variant intergénique	82277065	-	-	*XIST*
TEX11	Faux-sens toléré (0,1)	85115764	G/D	Défaillance de la spermatogénèse	Fertilité Mâle
TAF7L	Faux-sens délétère (0,03)	55133073	C/R	L'inactivation ciblée de ce gène provoque une sous-fertilité masculine, une oligozoospermie, une tératozoospermie, une athénozoospermie, des tubules séminifères anormaux et une réduction du poids des testicules sur des souches congénitales mais pas sur des souches mixtes. On observe également une diminution de la quantité de tissu adipeux brun et une augmentation de la formation des muscles squelettiques.	QTL

Tableau 36 : Description des phénotypes de 4 anomalies représentatives

Position	Normande		Montbéliarde		Holstein		Limousine		Charolaise		Blonde d'Aquitaine		Gène	Effet attendu
(UMD3.1)	M	F	M	F	M	F	M	F	M	F	M	F		
1957307	<	<	<	<		<	0,10	0,12	0,06	0,09	0,07	0,08	CCDC86	Survie
20327316	0,06	0,06	0,05	0,06	0,05	0,08	0,08	0,13	0,01	0,01	<	0,01	Arghf6	Immunité
22455700	0,03	0,03					<	<	0,14	0,18	<	<	IL13RA2	Immunité
39027899			0,25	0,24		<							MPP1	Immunité
40247641		<	0,12	0,09	<	<	0,01	<	<	<	0,01	0,01	TEX28	Fertilité M
40323817	0,10	0,09		<		<			0,04	0,03		<	FLNA	Survie
40386059		<	**<**	<	**<**	<			**0,08**	**0,03**	**<**	<	**TAZ**	**Fertilité M**
50812374		<				<							RNF168	Fertilité M
53516054	0,05	0,07		<		<				<			NRK	Survie
54971267	0,09	0,10	<	0,01	0,18	0,18	0,03	0,06	0,09	0,06	0,10	0,09	CENPI	QTL
55133073						<			**0,03**	**0,02**			**TAF7L**	**QTL**
55134586	0,06	0,05		<	0,02	0,01			0,04	0,02	<	<	TAF7L	QTL
82277065	**0,47**	**0,45**	**0,38**	**0,35**	**0,19**	**0,23**	**0,28**	**0,31**	**0,19**	**0,18**	**0,22**	**0,22**	**XIST**	**XIST**
82277693			0,08	0,07		<			0,02	0,03			XIST - EDA	XIST
82277847		<		<	<	0,01			0,07	0,06		<	XIST – EDA	XIST
82279663		<		<	<	0,01			0,07	0,06		<	XIST – EDA	XIST
82286297	0,22	0,19	0,02	0,02	0,17	0,20	0,17	0,21	0,05	0,06	0,17	0,19	XIST – EDA	XIST
82287530	0,09	0,07	<	<		<				<	0,03	0,04	XIST – EDA	XIST
82288518	0,22	0,19	<	0,01	0,17	0,20	0,17	0,21	0,04	0,04	0,16	0,18	XIST – EDA	XIST
82288520	0,22	0,19	<	0,01	0,17	0,20	0,17	0,21	0,04	0,04	0,16	0,18	XIST – EDA	XIST
85082301		<		<		<			0,03	0,04			TEX11	Fertilité M
85115764		<		<		<			**0,03**	**0,04**			**TEX11**	**Fertilité M**
85130594		<		<		<			0,03	0,04			TEX11	Fertilité M
90516788						<							RP2	Vision
92801539	0,08	0,07	0,17	0,16	0,15	0,18	0,49	0,43	0,35	0,36	0,33	0,30	SPACA5	QTL
96045540	0,09	0,08				<			<	<			KDM5C	Survie
118534737							<						NR0B1	Fertilité M
118535277				<		0,003		<				<	NR0B1	Fertilité M
128793730	0,06	0,06		<		<			<				SMPX	Surdité
130546627		<		<		<							SH3KBP1	Constitution
139925456		<		<	<	<							Mid1	Survie

Tableau 37 : Répartition des fréquences alléliques des marqueurs chez les femelles et les mâles dans six grandes races. Le choix des variants est codé par « QTL » dans le cas des gènes identifiés par une approche de recherche de QTL de fertilité mâle par De Camargo (2015). Le type de phénotype est renseigné. Les variants « XIST » se situent dans ou autour du gène *XIST* tel qu'identifié chez l'humain. Les fréquences inférieures à 0,01 sont codées « > » tandis que les cases sont vides lorsque la fréquence est nulle. Les variants grisés sont ceux correspondant au tableau 36.

Les variants prédits comme ayant un effet sur la fertilité mâle seront analysés dans le cadre d'études d'association sur le chromosome X. Nous souhaitions initialement utiliser les bases de données de prélèvements de semence fournies par les entreprises de sélection. Cependant, au vu de l'hétérogénéité des méthodes de mesure en race Montbéliarde (cf. chapitre 2, étude 4), nous avons reconsidéré ce choix. Une étude d'association entre ces variants et des phénotypes de fertilité femelle tels les taux de conception et le taux de non-retour des femelles à 90 et 56 jours pourrait être intéressante. Les variants prédits comme responsables d'une chute du taux de survie et d'une baisse de l'immunité pourront être étudiés grâce aux phénotypes recueillis dans le cadre du projet MORPHE. Ce projet financé par FGE avait pour objectif de mesurer la part génétique responsable de la mortalité juvénile chez les bovins dans le but de proposer une évaluation génétique (Leclerc *et al.*, 2016).

Lors du choix des variants, j'ai privilégié l'intérêt du phénotype délétère en rapport avec mes travaux de recherche (*XIST*, fertilité) plus que la fréquence au sein de la race. Ce choix avait pour but de maximiser la probabilité de détecter un variant intéressant pour un QTL de fertilité qui était notre sujet d'intérêt principal sur le chromosome X. Cependant, au regard des fréquences calculées sur la population génotypée entre juin 2018 et janvier 2020, nous pouvons observer que la plupart des variants choisis sont présents dans plusieurs races. Il aurait pu être plus judicieux de faire une sélection accrue sur les variants spécifiques d'une seule race, puisque les variants responsables d'anomalies génétiques sont le plus souvent d'apparition récente et race-spécifiques (Michot, 2017).

L'étude de variants rares nécessite une population d'individus génotypés très importante pour identifier des individus homozygotes ou hémizygotes et réaliser une caractérisation phénotypique. Les effectifs identifiés en janvier 2020 vont nous permettre par la suite d'étudier le phénotype des animaux mutés. Ces variants pourront être étudiés ultérieurement, avec l'accumulation de nouveaux génotypes, le développement de tests sur haplotype pour récupérer des individus supplémentaires et l'imputation des génotypes pour ces polymorphismes sur la population génotypée sur des puces ne possédant pas ces marqueurs.

Les variants autour de *XIST* ont été choisis car ils pouvaient potentiellement entraîner un biais d'inactivation qui pourrait être important à prendre en compte dans les évaluations génomiques future puisqu'impactant l'ensemble des gènes du chromosome X et donc des phénotypes déterminés par ces derniers. Ce biais d'inactivation pourrait être mis en évidence en étudiant dans différents tissus le niveau d'expression de gènes du chromosome X portant des polymorphismes synonymes chez des animaux hétérozygotes pour ces variants autour de *XIST*.

En conclusion, j'ai présenté une méthodologie d'identification des variants d'intérêt par une approche de génétique inverse ainsi que les particularités liées au chromosome X de cette méthode. L'efficacité de cette stratégie s'appuyant sur l'exploitation de données de séquences complètes a été démontrée dans la littérature (Charlier *et al.*, 2016).

Même si nous n'avons pas pu aller jusqu'à la caractérisation phénotypique des polymorphismes, nous avons pu étudier la fréquence de chacune d'entre elles dans la population. Grâce à ce travail, nous avons mis en place en routine de nouveaux outils pour étudier de façon distincte les génotypes des mâles et des femelles, ce qui nous permettra de suivre l'évolution de la fréquence allélique chez les deux sexes avec l'accumulation de nouveaux génotypages. Cette étude nous a également amenée à collaborer avec l'unité BREED et discuter de l'intérêt de la caractérisation sur le bovin, considéré comme espèce modèle, de mutations affectant la fertilité mâle chez l'homme.

L'addition de plus de 1000 génomes supplémentaires dans le dernier run du projet 1000 génomes bovins pourrait permettre de détecter de nouveaux variants d'intérêt avec une fréquence interraciale

plus pertinente. De même, le changement d'assemblage et la correction des trous dans la séquence ont permis d'améliorer la séquence de référence du chromosome X, ce qui va influer sur l'annotation des variants et augmenter sa précision.

Résultats clés de cette étude :

- Nous avons réalisé une approche de génétique inverse en choisissant 31 variants d'intérêt à partir des annotations réalisées dans le cadre du run6 du projet 1000 génomes bovins.
- La caractérisation clinique de ces variants n'a pas pu être faite faute de temps mais l'étude des fréquences nous apporte des premiers renseignements concernant la pertinence du choix de ces variants.

3. Cas concret d'une anomalie identifiée par génétique classique

En 2013, l'ONAB s'est vu confier le cas d'une génisse présentant un pelage fin de type « duvet » et une malformation des dents, faisant penser à une dysplasie ectodermique hypohidrotique (DEH). Cette anomalie génétique est bien connue chez les mammifères et se caractérise par des cheveux ou poils épars (hypotrichose), des dents anormales ou manquantes (oligodontie), et une diminution de la transpiration (hypohidrose) (Clarke *et al.*, 1987). Les caractéristiques cliniques typiques incluent également un défaut de développement d'autres glandes exocrines, qui provoque une sécheresse de la peau et des muqueuses, ainsi que des ongles parfois absents et dystrophiques (Lefebvre et Mikkola, 2014).

Plus de la moitié des cas de DEH rapportés chez l'homme sont des individus mâles présentant une mutation du gène *EDA (ectodysplasin-A)* situé sur le chromosome X. Or dans le cas d'anomalies génétiques liées au chromosome X, les femelles et les mâles ne sont pas affectés de la même manière. Les femelles hétérozygotes, porteuses de la mutation, montrent des symptômes moins sévères que les mâles hémizygotes. En effet, bien qu'un seul chromosome X soit exprimé dans une cellule de femelle du fait du phénomène d'inactivation du chromosome X (voir chapitre 1), les deux chromosomes sont exprimés à l'échelle de l'individu, ce qui permet une certaine compensation. Dans le cas d'anomalies génétiques cutanées liées au chromosome X, les cellules saines et mutées se répartissent le long des lignes de Blaschko.

Les lignes de Blaschko définissent les voies de développement ectodermique. Les affections suivant les lignes de Blaschko sont des manifestations de mosaïcisme cutané, avec un contraste cliniquement apparent entre une peau génétiquement normale et une peau anormale (Molho-Pessach et Schaffer, 2011). Les problèmes cutanés mosaïques ne suivent pas toujours les lignes de Blaschko (Type 1 sur la figure 31) mais peuvent être également en bloc (type 2), phylloïde comme les nervures d'une feuille (type 3), avec de grandes taches sans séparation des lignes médianes (type 4) ou latéralisé (type 5) (Happle, 1993).

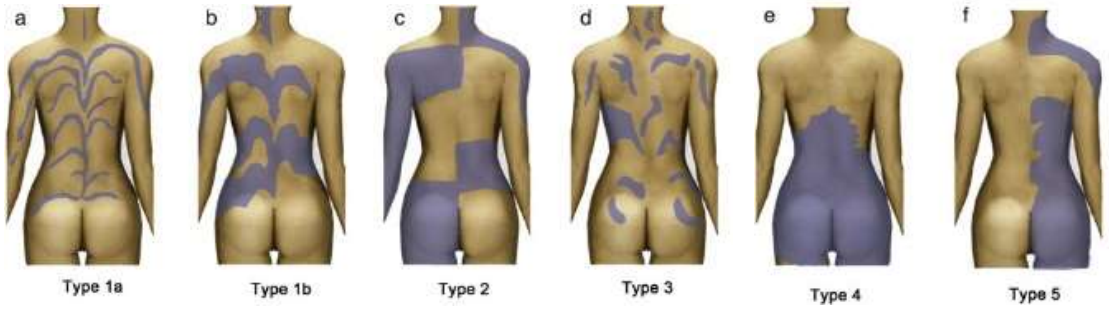


Figure 31 : Représentation schématique des lignes de Blaschko (Jin *et al.*, 2016) d'après Molho-Pessach & Schaffer, 2011)

Chez le bovin, Murgiano *et al.* (2015) ont rapporté le cas de femelles apparentées présentant un motif poils/sans poils le long les lignes de Blaschko. Cette anomalie était dûe à un variant dans un site d'épissage du gène *TSR2 (Ribosome maturation factor)* situé sur le chromosome X.

Figure 32 : Vache atteinte d'alopécie localisée (Murgiano et al, 2015)

Le fait que notre animal atteint soit (i) un cas isolé, (ii) de sexe femelle, et (iii) présentant un phénotype homogène et sévère était particulièrement intéressant puisqu'un phénotype non homogène et moins sévère serait attendu chez une femelle dans le cas d'une mutation du gène *EDA*. C'est ce qui a motivé les recherches présentées ci-dessous qui ont fait l'objet d'un article scientifique publié dans la revue BMC Genomics en décembre 2019.

3.1. Etude : A de novo 3.8-Mb inversion affecting the *EDA* and *XIST* genes in a heterozygous female calf with generalized hypohidrotic ectodermal dysplasia

Escouflaire et al. BMC Genomics (2019) 20:715
https://doi.org/10.1186/s12864-019-6087-1

BMC Genomics

RESEARCH ARTICLE Open Access

A *de novo* 3.8-Mb inversion affecting the *EDA* and *XIST* genes in a heterozygous female calf with generalized hypohidrotic ectodermal dysplasia

Clémentine Escouflaire[1,2*], Emmanuelle Rebours[1], Mathieu Charles[1,3], Sébastien Orellana[1], Margarita Cano[1], Julie Rivière[1,4], Cécile Grohs[1], Hélène Hayes[1] and Aurélien Capitan[1,2*]

Abstract

Background: In mammals, hypohidrotic ectodermal dysplasia (HED) is a genetic disorder that is characterized by sparse hair, tooth abnormalities, and defects in cutaneous glands. Only four genes, *EDA*, *EDAR*, *EDARADD* and *WNT10A* account for more than 90% of HED cases, and *EDA*, on chromosome X, is involved in 50% of the cases. In this study, we explored an isolated case of a female Holstein calf with symptoms similar to HED.

Results: Clinical examination confirmed the diagnosis. The affected female showed homogeneous hypotrichosis and oligodontia as previously observed in bovine *EDAR* homozygous and *EDA* hemizygous mutants. Under light microscopy, the hair follicles were thinner and located higher in the dermis of the frontal skin in the affected animal than in the control. Moreover, the affected animal showed a five-fold increase in the number of hair follicles and a four-fold decrease in the diameter of the pilary canals. Pedigree analysis revealed that the coefficient of inbreeding of the affected calf (4.58%) was not higher than the average population inbreeding coefficient (4.59%). This animal had ten ancestors in its paternal and maternal lineages. By estimating the number of affected cases that would be expected if any of these common ancestors carried a recessive mutation, we concluded that, if they existed, other cases of HED should have been reported in France, which is not the case. Therefore, we assumed that the causal mutation was dominant and de novo. By analyzing whole-genome sequencing data, we identified a large chromosomal inversion with breakpoints located in the first introns of the *EDA* and *XIST* genes. Genotyping by PCR-electrophoresis the case and its parents allowed us to demonstrate the de novo origin of this inversion. Finally, using various sources of information we present a body of evidence that supports the hypothesis that this mutation is responsible for a skewed inactivation of X, and that only the normal X can be inactivated.

Conclusions: In this article, we report a unique case of X-linked HED affected Holstein female calf with an assumed full inactivation of the normal X-chromosome, thus leading to a severe phenotype similar to that of hemizygous males.

Keywords: Hypohidrotic ectodermal dysplasia, EDA, XIST, Whole-genome sequencing, Holstein, Cattle

* Correspondence: clementine.escouflaire@inra.fr; aurelien.capitan@inra.fr
[1]INRA, GABI, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France
Full list of author information is available at the end of the article

© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Background

In Mammals, hypohidrotic ectodermal dysplasia (HED) is a well-described genetic disorder that is characterized by sparse hair (hypotrichosis), abnormal or missing teeth (oligodontia), and reduced ability to sweat (hypohidrosis) [1]. Typical clinical features also include defective development of other exocrine glands, which causes dryness of the skin and mucosa, and occasionally absence of teats and dystrophic nails [2]. This condition is known in humans and represents the most frequent form of human ectodermal dysplasia, but has also been reported in a number of species such as rat [3], dog [4] and cattle [5]. Without adequate medical care, the life span of affected individuals is markedly reduced because of sensitivity to temperature fluctuations, respiratory problems and feeding difficulties [6, 7].

Studies in humans have demonstrated that only four genes (*EDA, EDAR*, *EDARADD*, and *WNT10A*) account for more than 90% of HED cases, and that *EDA* is involved in more than 50% of cases [8]. *WNT10A* encodes the wingless-type mouse mammary tumor virus (MMTV) integration site family member 10A (WNT10A), which is a small signaling protein belonging to the Wnt family. WNT10A is a key modulator of cell–cell interactions in many tissues, and also contributes to the embryonic development of skin, hair, nails and teeth [9–11]. *EDA*, *EDAR* and *EDARADD*, each encode one of the three proteins of the ectodysplasin pathway, namely ectodysplasin-A (EDA), which is a member of the tumor necrosis factor (TNF) superfamily, its receptor (EDAR), and a specific adaptor protein, the EDAR associated death domain (EDARADD) [12]. The ectodysplasin pathway functions downstream of the Wnt/β-catenin signaling and plays an essential role in the morphogenesis of ectodermal organs [13].

Mutations in *EDAR*, *EDARADD* and *WNT10A* are most frequently autosomal recessive, whereas mutations in EDA, which is located on the X chromosome, are inherited in a X-linked recessive pattern and cause the X-linked form of HED (XLHED) [14]. In contrast to males, females that are heterozygous for *EDA* mutations typically show less severe symptoms with a reduced degree of hypodontia, mosaic patterns of hypotrichosis, and defective sweat glands along Blaschko's lines because of random X inactivation [15–17].

In 2014, a female calf that presented typical symptoms of HED was reported to the French National Observatory for Bovine Genetic Defects (ONAB, https://www.onab.fr/). Its parents were both phenotypically normal. ONAB has no other record of such cases in the French Holstein population during the period 2002–2018. This isolated case offered a rare opportunity to identify new genes and/or a new mode of inheritance involved in HED.

In this study, we report the identification of a 3.8-Mb inversion on chromosome X of a heterozygous female calf that causes a dominant and generalized form of HED via skewed X-inactivation and truncation of the EDA protein.

Results

Clinical examination

A Holstein female calf, conceived through artificial insemination and born after a normal gestation period of 275 days, was reported to the ONAB for abnormal hair coat. Hairs were short and thin over the whole body and tended to be slightly longer on the fetlocks and the tail. Whiskers were normal and horn buds appeared normal for its age (Fig. 1 and Additional file 5: Figure S1). Dental examination revealed a complete absence of teeth with the exception of one malformed premolar on each side of the upper jaw (Fig. 1 and Additional file 5: Figure S1 and Additional file 6: Figure S2). This animal died at 2 weeks of age of pulmonary infection. Unfortunately, we did not have the possibility to investigate its internal organs but obtained formalin-fixed skin samples from the forehead and horn bud areas for post-mortem histological analyses. Observation under light microscopy showed that, in the affected calf, the hair follicles were thinner and located higher in the dermis than in the control animal (Fig. 2a-b). Sebaceous

Fig. 1 Pictures of a control animal (1 month old, **a-c**) and the affected animal (2 weeks, **b-d**). Please note the difference in hairiness between the two animals (**a-b**) and the absence of incisors in the affected calf (**d**)

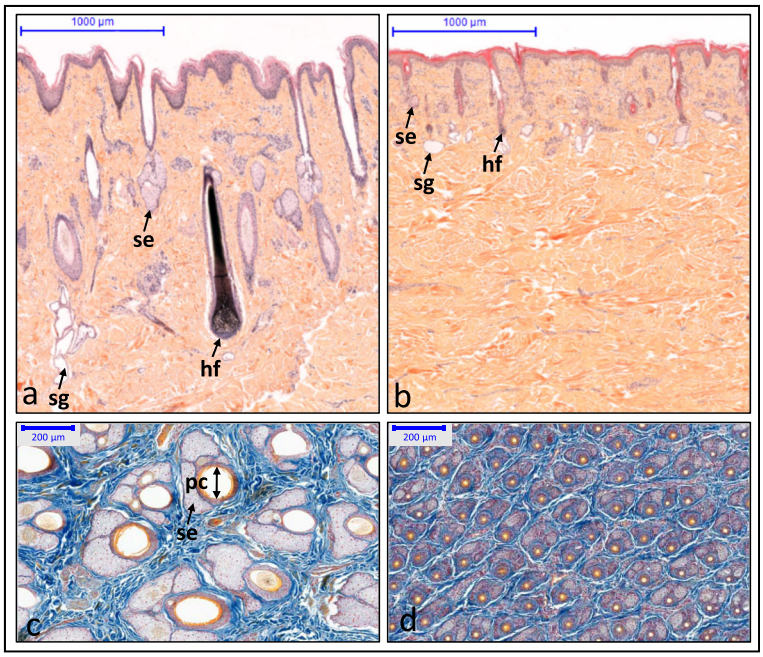


Fig. 2 Sections of frontal skin samples from the affected calf and a control. Transversal sections of the skin of control (**a**) and HED animals (**b**) stained with hematoxylin-eosin-safran. Note the reduced size of sweat glands (sw), sebaceous glands (se), and hair follicles (hf) in the affected (**b**) versus control (**a**) animals. Longitudinal sections of the skin of control (**c**) and HED animals (**d**) stained with a Roan solution (Orange G, aniline blue staining and Kernechtrot). Note the increased density but reduced size of sebaceous glands (se), and pilary canals (pc) in the affected (**d**) versus control (**c**) animals

and sweat glands were hypoplastic. Analysis of longitudinal skin sections demonstrated that there were five times more hair follicles in the frontal skin of the affected calf than the control (mean = 47.7 vs 9.3 hairs in 1-mm^2 quadrats) and that the diameter of their pilary canals was four times smaller (mean = 21.6 vs 87.2 µm, Fig. 2c-d, Table 1). Histological analyses also revealed that the affected calf had normal horn buds (Additional file 7: Figure S3).

Pedigree analysis

To better understand the genetic inheritance of this syndrome in the affected animal, we analyzed pedigree data and showed that with a coefficient of inbreeding of 0.0458, this calf was not more inbred than the Holstein calves born the same year (2014) with an average coefficient of inbreeding of 0.0459. We identified ten loops in the pedigree with common ancestors in the maternal and paternal lineages. Then, we predicted the number of inbred animals that were homozygous for alleles inherited from each of these common ancestors in the French Holstein population born from 2002 to 2017, which represents the putative number of affected animals born since the creation of the ONAB assuming a recessive inheritance. This number varied from 10 to 38,529 individuals depending on the ancestor (Additional file 1: Table S1). Considering that HED is a distinctive syndrome and that no other individual with a similar phenotype was reported in Holstein during this period, it is very unlikely that this case is due to a recessive mutation.

Thus, we investigated the hypothesis of a dominant inheritance associated with somatic or germline mosaicism,

Table 1 Measurement of hair count, diameter of the pilary canal and area covered by pilary canals in three different quadrats of 1 mm^2 of frontal skin samples from the affected animal and a control

		1st quadrat	2nd quadrat	3rd quadrat	Mean
Control	Hair count	9	10	9	9.3
	Mean diameter (µm)	89.9	85.1	86.6	87.2
	Total area (µm^2)	59,689	59,274	56,286	58,416
Affected	Hair count	43	44	56	47.7
	Mean diameter (µm)	22.9	23.8	18.0	21.6
	Total area (µm^2)	18,802	20,348	14,773	17,974

since both parents were clinically normal. The dam had only two other descendants, which were both unaffected. The sire had 9201 descendants and among these, none were reported to be affected by HED.

Finally, we estimated the mortality rate in the progeny of the sire of the affected calf. With 10.22% of calves that died between 0 and 6 months of age, its mortality rate is slightly lower than the average rate for the population (10.93%, +/− 0.21 sd). Thus, there is no evidence of increased mortality and of putative non reported HED calves among its progeny.

Based on these findings, we made the assumption that the mutation responsible for this isolated case of HED is a de novo and dominant mutation.

Analysis of whole-genome sequences

In order to identify putative causal variants, we sequenced the whole genome of the affected calf. Based on the assumption of a de novo and dominant mutation, we retained only the heterozygous SNPs and small indels that were absent from the whole-genome sequencing data from 2331 bovine individuals in run 6 of the 1000 bull genomes project [18, 19]. We identified 5382 private heterozygous mutations among which only nine were predicted to affect the primary structure of a protein (Additional file 2: Table S2).

Six of these polymorphisms consisted of additions of one to three amino-acids in stretches of repeated amino-acids. The three remaining mutations were one putative deleterious missense mutation (SIFT score = 0) in the *solute carrier family 34 member 3 protein* (*SLC34A3* p.Y463D) gene and two frameshift mutations in the *suppressor of Ty 20 homolog* (*SUPT20H* p.Q642PfsX22) and the *phosphatidylinositol-4-phosphate 3-kinase C2 domain-containing beta* (*PIK3C2B* p.L441WfsX21) genes.

Since with this approach, we investigated only the small variants that were located in coding regions [20], we then searched for structural variations in candidate genes using the Integrative Genomics Viewer (IGV) [21]. The study of *EDA*, *EDAR*, *EDARADD* and *WTN10* revealed a 3.8-Mb inversion with breakpoints located in the first intron of *EDA* and the first intron of *XIST* (Fig. 3, Additional file 8: Figure S4). Average sequencing depth in the region that encompasses the two breakpoints revealed no amplification or deletion of the inverted segment. (Additional file 9: Figure S5). PCR amplification of four segments that span these breakpoints in the mutant and wild-type alleles confirmed the existence and the de novo nature of this inversion. Indeed, only the affected calf was heterozygous for the mutation, whereas its parents were homozygous wild-type (Fig. 4). Subsequent Sanger sequencing of the PCR products enabled us to specify the nature of the nucleotide sequence at each breakpoint in the mutant allele. For the first breakpoint (between positions 82,271,052 and 82,271,053 bp on chromosome X, based on the bovine genome assembly UMD3.1), we observed a partial duplication of a 12-bp segment (5′-GTACAAGAAACT-3′), which suggested that microhomology-mediated break-induced replication [22] was involved in DNA repair. In contrast, for the second breakpoint, we observed a deletion of a single thymine (between positions 86,034,441 and 86,034,442 bp), which suggests that DNA repair resulted from non-homologous end-joining at this position. Finally, we analyzed 500-bp segments that encompass the breakpoints using dotplot [23] and identified no identical DNA sequence motif that could have promoted this inversion. Therefore, this mutation most probably results from a random break event.

Immunostaining of EDA protein

To determine if a truncated EDA protein is expressed in the affected calf, we performed immunostaining in frontal

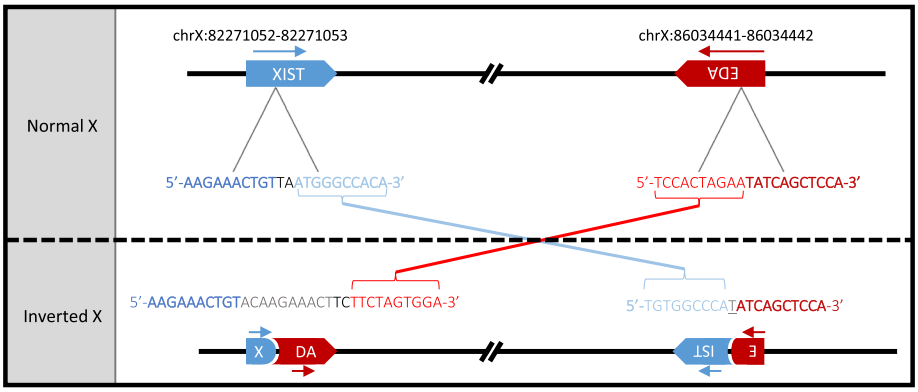


Fig. 3 Scheme representing the de novo inversion between *XIST* and *EDA*. The first line provides information on the position of the breakpoints on the bovine UMD3.1 genome assembly. The nucleotides inserted in the first breakpoint after microhomology DNA repair are represented in grey while the nucleotide shared by both *XIST* and *EDA* segments at the second breakpoint is underlined

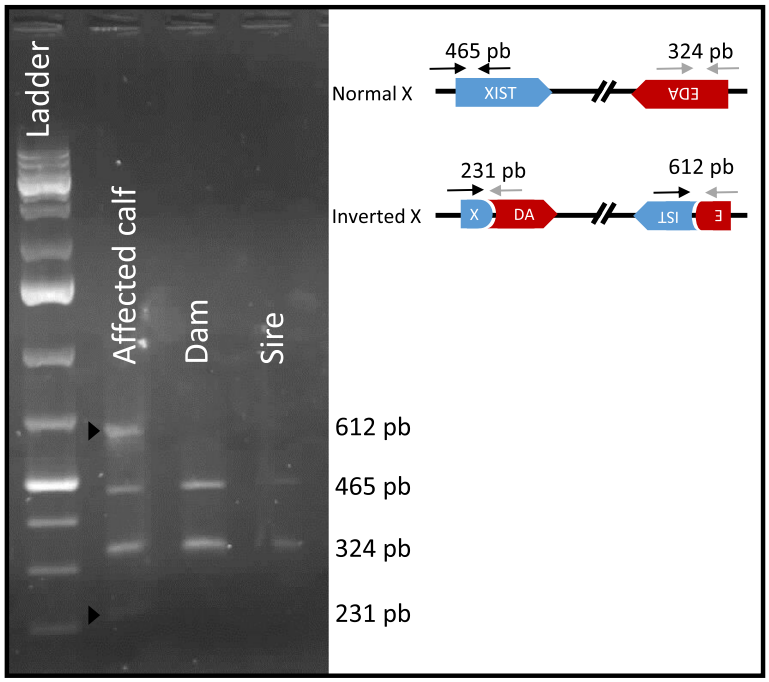


Fig. 4 Genotyping of the de novo inversion using PCR and electrophoresis. Amplicons specific of the mutant allele (231 and 612 bp, in red) were observed only in the affected calf (see arrow-heads) while amplicons corresponding to the wild-type allele (324 and 465 bp) were amplified in the affected calf and both of its parents

skin tissue with antibodies specific for the EDA C-terminal and N-terminal ends. We observed relatively similar intensity and localization of staining in the case and control samples (Fig. 6; Additional file 11: Figure S7). However, the contrast of staining between the wall and the cytoplasm of the sebaceous glands cells was less marked in the affected animal than in the control with the antibody directed against the N-terminal region of the protein.

Discussion

In this study, we explored an isolated case of a female Holstein calf with homogeneous hypotrichosis and oligodontia. Clinical and histopathological examinations revealed symptoms similar to those previously reported in *EDAR* homozygous and *EDA* hemizygous mutant cattle e.g. [5, 16, 17, 24], thus confirming the diagnosis of HED. In the affected animal, hair follicles were more numerous, thinner and located higher in the dermis of the frontal skin; sebaceous and sweat glands were hypoplastic; teeth were absent with the exception of one malformed premolar on each side of the upper jaw; while horn development appeared normal. This animal died at 2 weeks of age of pulmonary infection, probably aggravated by the absence of ciliae and mucous glands in the respiratory tract, as usually observed in HED [6, 14].

Analyzing the pedigree of the affected female, we identified ten ancestors present in both its paternal and maternal lineages. Nevertheless, this animal was not more inbred than the average of Holstein calves born the same year, with inbreeding coefficients of 4.58% versus 4.59%, respectively. By estimating the number of affected cases that would be expected if any of the common ancestors carried a recessive mutation, we concluded that, if they existed, other cases of HED should have been reported in France, which is not the case. Therefore, we assumed that the causal mutation is dominant and de novo.

Based on this hypothesis, we sequenced the whole genome of the affected calf, and conserved the small indels and heterozygous SNPs that were not observed in the population of 2331 bovine individuals of the 1000 bull genomes project [18, 19]. We identified nine variants that were predicted to affect the primary structure of a protein (Additional file 2: Table S2). Among them, six consist of additions of one to three amino acids in stretches of repeated amino acids. Since these stretches showed variation in size and/or poor conservation among orthologous proteins in mammals, we assumed that insertions at these sites are likely tolerated. The three remaining mutations are one putative deleterious missense mutation (SIFT score = 0) in the *SLC34A3* (p.Y463D) gene and two frameshift mutations in *SUPT20H* 5p.Q642PfsX22) and *PIK3C2B* (p.L441WfsX21). Mutations for knock-out alleles in these three genes have been well characterized in mouse. To our knowledge, no particular phenotype has been reported in heterozygous carrier animals, and none

of the symptoms observed in homozygous individuals were consistent with HED. *SLC34A3* −/− mice have a modified calcium absorption and excretion [25, 26]. Some *SUPT20* −/− animals show major defects in the epithelial-to-mesenchymal transition during gastrulation, while others survive post-natally and appear normal [27]. Finally, *PIK3C2B* −/− mice were found to exhibit normal epidermal growth, differentiation and function in a study on the role of phosphoinositide 3-kinases in epidermal differentiation [28]. In conclusion, none of these nine SNPs and small indels appear to be relevant candidate mutations.

In addition, we conducted a search for structural variations in candidate genes (*EDA*, *EDAR*, *EDARADD* and *WTN10*) using IGV [21]. In doing so, we identified a large chromosomal inversion (3.8 Mb) with breakpoints located in the first intron of *EDA* and the first intron of *XIST* (Fig. 3, Additional file 8: Figure S4). Genotyping the case and its parents by PCR-electrophoresis allowed us to demonstrate the de novo origin of this mutation (Fig. 4).

XIST (X inactive specific transcript) encodes a long non-coding RNA that is responsible for the initiation of X-chromosome inactivation (XCI), i.e. the mechanism of dosage compensation that equalizes the expression of X-specific genes between genders in mammals. In female cells, one chromosome X is randomly inactivated, resulting in the expression of only one of the two sex chromosomes. The mechanism that regulates XCI is complex and not yet fully understood. XCI starts with the progressive coating of the future inactive chromosome by XIST RNA in cis, and involves a number of XIST activator [29] and repressor molecules [30] such as TSIX, a long non-coding RNA partially overlapping XIST and transcribed in the antisense orientation. In some cases, skewed-inactivation can occur due to negative or positive selection of one of the X chromosomes [7, 31, 32]. In humans and mouse, several mutations that affect the promoter or the coding regions of *XIST* are known to result in the preferential inactivation of the normal X chromosome in heterozygous females [33–35]. In parallel, analysis of a series of genome-edited human cultured cells demonstrated that partial deletions of *XIST* reactivate the inactive X chromosome [36]. While XCI is a conserved mechanism among mammals, the nature and importance of the *XIST* functional sites differ between man and mouse, the two species in which this mechanism has been most studied [36–38]. In agreement with this observation, analysis of Ensembl GERP conservation scores showed that, among the 91 eutherian mammals studied, the number of constrained elements in the *XIST* gene is small and these are mainly located in the region corresponding to human exon 4 (Additional file 10: Figure S6).

The inversion that we describe in this paper leads to the separation of the *XIST* exon 1 from the rest of the gene. Thus, if transcribed, the mutant XIST RNA would lack the main evolutionary constrained element among mammals, and contain only 48% of the normal transcript. In addition, this mutation does not affect the integrity of TSIX, which encodes the main repressor of XIST.

Although we could not conduct expression studies to validate our hypothesis, these arguments combined with the observation of a generalized HED phenotype in a female heterozygous for a mutation that truncates *EDA*, suggest that XCI is impaired on the X chromosome carrying the mutation and results in the skewed inactivation of the normal X.

In humans, several examples of females with X-linked HED due to complete skewed X-inactivation have been reported in the literature [7, 39, 40]. The patients were heterozygous carriers of a partial X-autosome translocation causing a disruption of the *EDA* gene. In these cases, the integrity of the *XIST* gene was not affected and the mechanism invoked to explain the skewed X inactivation was selective survival of embryonic cells inactivating the normal X. Indeed, inactivation of the derivative chromosome X is assumed to cause partial monosomy due to the extension of the inactivation wave onto the translocated autosome and is incompatible with further development [7]. In parallel, as previously mentioned, a number of mutations in the regulatory or coding regions of *XIST* have been reported to cause skewed inactivation and it is generally accepted that about 1% of females exhibit extremely *skewed* X-inactivation patterns i.e. 95:5 [41–43]. Nevertheless, cases of heterozygous females who carry mutations in *XIST* and in another chromosome X gene and are affected by a generalized syndrome, as observed here, are very uncommon [34, 44].

The second breakpoint is situated at position 86,034,441 bp within the *EDA* intron 1. *EDA* encodes two main transcripts EDA-A1 and EDA-A2 which differ only by two amino-acids (307 V and 308E localized in the TNF domain of the protein; exon 8). Both of these isoforms possess a membrane form and a secreted form after cleavage by furine at position 159–160 (exon 2). The extracellular part includes a collagen-like Gly-X-Y repeat that is essential for the trimerisation of the protein and a TNF-like signaling domain at the C-terminal end. The putative truncated protein would possess only the 132 first AA encoded by exon 1 and thus lack the cleavage site and the collagen-like and TNF-homology domains of the protein (Fig. 5). Many mutations in *EDA* have already been described in humans and mice, which provide insights on the impact of the deletion that we report in this article. Karlskov et al. [14] described a case

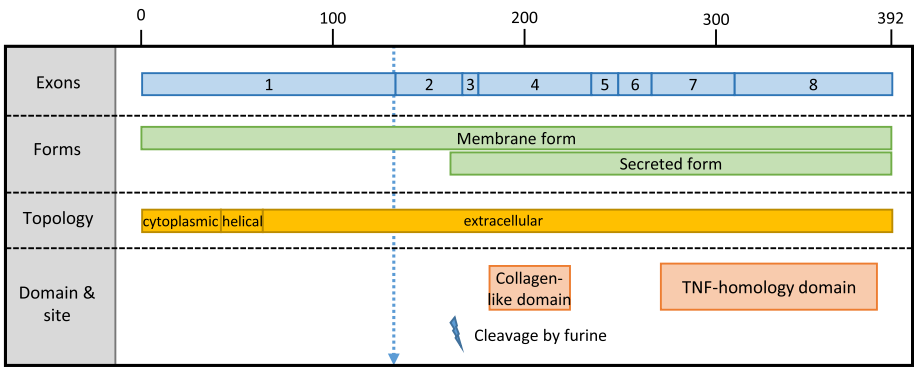


Fig. 5 EDA isoforms and protein domains. Note that EDA-A1 and EDA-A2 are not separated since they only differ by two amino-acids in exon 8. Both of these isoforms possess a membrane form and a secreted form after cleavage by furine at position 159–160 (exon 2)

of HED in cattle that was caused by a transcript variant containing a LINE1-derived pseudo exon between the first and the second exons of *EDA*. Since this pseudo exon results in a frameshift and a stop codon early in exon 2, we conclude that, in this HED case, the truncated protein encoded by only the first *EDA* exon would not be functional. In parallel, a complete deletion of the first exon has been reported to cause HED in humans [45] which supports the hypothesis that no putative alternative transcripts starting after exon 1 can compensate for the lack of the full transcripts. Thus, if expressed, the two parts of EDA that have been truncated by the inversion reported in this article are assumed to be non-functional.

Finally, to determine if a truncated EDA protein is expressed in the affected calf, we performed immunostaining with antibodies specific for the EDA C-terminal and N-terminal ends in frontal skin tissue. Immunostaining of the N-terminal end (Fig. 6a-c) revealed a relatively similar intensity and localization of staining in the case and control samples, thus supporting the existence of peptides containing the proximal part of the EDA protein in the skin of the affected calf. However, in a general manner, the contrast of staining between the wall and the cytoplasm of the sebaceous glands cells was less marked in the affected animal than in the control. While the region of EDA corresponding to exon 1 contains the helical domain required for membrane binding, it can be assumed that the lack of the C-terminal region affects its stability and/or transport to the cell wall. Surprisingly, immunostaining of the C-terminal part of the protein revealed the presence of peptides that correspond to the distal region of the EDA protein in the affected animals (Fig. 6b-d, Additional file 11: Figure S7). Considering the particular nature of the inversion and the transcription sense of the two broken genes, these peptides could be produced by a fusion transcript (e.g. [46]) that would start with the first exon of *XIST* and continue with exons 2 to 8 of *EDA*. Unfortunately, the lack of RNA material precluded further investigation. In any case, these analyses demonstrated that both parts of the EDA protein are expressed in spite of the rearrangement involving *EDA* and leading to a HED phenotype.

Conclusions

In this study, we characterized a new mutation which is a strong candidate causal mutation for the hypohidrotic ectodermal dysplasia syndrome, that we detected in the female Holstein calf reported here. Mutations that occur in *EDA* on the X chromosome usually affect only the male carriers while the female carriers express few or no symptoms. The severe and homogeneous phenotype displayed by this animal could be explained by a preferential inactivation of the wild-type chromosome. Indeed, the affected calf carries an inversion that breaks both *EDA* and *XIST*, the keystone genes involved in X-inactivation. In spite of some limitation due to accessibility and conservation of biological samples, we present a body of evidence supporting the hypothesis that this mutation is responsible for a skewed inactivation of the X chromosomes, and that only the normal X can be inactivated. Based on our results, we present a new mode of inheritance for HED that may highlight the underlying mechanisms of dosage compensation when XIST is not functional.

Methods

Animals and sampling

The affected animal was photographed and reported by its breeder and veterinarian to the ONAB in its first week of life. It died the following week and we were informed of its death only 2 days later. In this context, we could obtain only a limited number of samples from the local veterinarian before disposal of the body. These specimens included formalin fixed samples from the

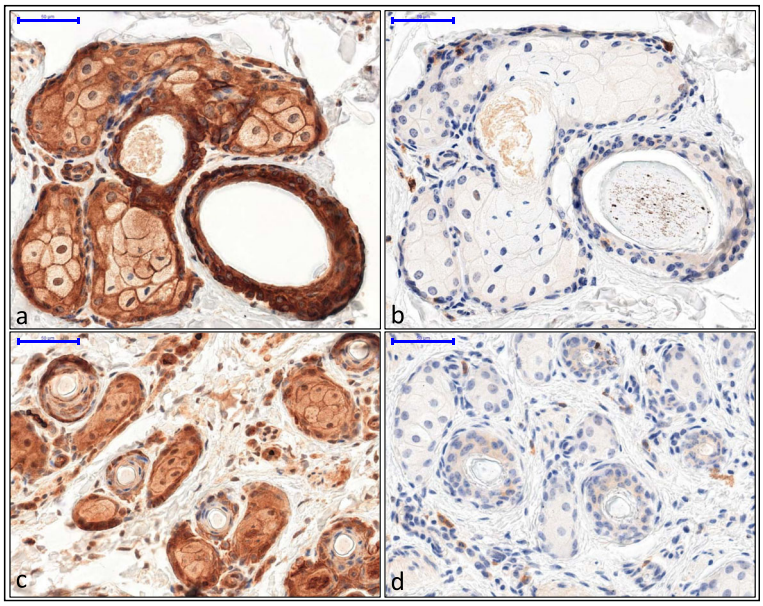


Fig. 6 Immunostaining of the EDA protein in frontal tissue from control (**a, b**) and HED animals (**c, d**). Results of the N-ter immunostaining show a stronger contrast between cell wall and cytoplasm coloration in the control animal (**a**) than in the affected calf (**c**). Immunostaining of C-ter part of the protein appear similar between control (**b**) and affected (**d**) calves. Negative control of the C-ter is available in Additional file 11: Figure S7

frontal skin, the horn buds and the distal extremity of the lower jaw. Sampling for expression studies was not possible since RNA degrades in just a few hours after death. Blood samples for the female calf and its dam, as well as semen straws for the sire, were available for DNA extraction. Finally, we also sampled control tissues 6 h after death on a young Holstein calf that died of diarrhea.

Histological analyses

Skin samples were fixed in 10% neutral buffered formalin for 24 h and embedded in paraffin wax according to standard histological protocols. Four-micrometer thick sections were mounted on adhesive slides (Klinipath-KP-PRINTER ADHESIVES). Paraffin-embedded sections were deparaffinized and stained with HES (haematoxylin, eosin and safran) to observe the skin morphology and with a Roan solution (kernechtrot, orange gelb and aniline blue) to study the hair follicles. Immunostaining for the C-terminal and N-terminal ends of EDA was performed using the Leica Bond RXm IHC staining system with the Bond Polymer Refined kit (DS9800), and antibodies against the EDA C-terminal (ab198022, abcam) and EDA N-terminal (OAAB03423, Aviva biosystem) ends, both diluted at 1:50. Incubation with the primary antibody and antigen retrieval were performed at pH = 9 for 20 min.

All slides were scanned with the Pannoramic Scan 150 (3D Histech) and analyzed with the CaseCenter 2.9 viewer (3D Histech). The number of hair follicles, their diameter and the total area covered by hair were measured in 1 mm^2 quadrats on three different Roan-stained skin sections from the forehead region of the control and affected animals.

Pedigree analysis

The average coefficient of inbreeding in the French population of Holstein females born in 2014 and in the affected animal were calculated with the program meow.f of the pedig package [47], which is based on the direct construction of the Cholesky decomposition of the kinship matrix. We used pedigree information from the French national databases.

The ancestors found in both the paternal and maternal lineages of the affected animal were identified based on pedigree data. Then, for each of these ancestors we calculated the expected numbers of affected animals in the French Holstein population assuming a recessive inheritance for HED. To do this, for any individual i in the population, we computed the contribution c (i) of this ancestor to i by gene dropping and then the probability of i of being homozygous for a mutation inherited from its ancestor as $p_{homo}(i) = c\ (sire) * c\ (dam)\ /4$. Finally, we estimated the number of affected cases per year by summing the risk probabilities for all the calves born during the same period.

In addition, for 3577 bulls (including the sire of the HED female) with more than 100 descendants, we

calculated the percentage of calves that died a natural death between zero and 6 months of age in their progeny.

DNA extraction

Genomic DNA was extracted from blood and semen using the DNeasy Blood and Tissue Kit (Qiagen). DNA quality was controlled by electrophoresis and quantified using a Nanodrop spectrophotometer.

Whole-genome sequencing

A paired-end library with an insert size of 450 bp was generated with the NEXTflex PCR-Free DNA Sequencing Kit (Bioscientific), and sequenced on a HiSeq 2000 sequencer (Illumina) with an average sequence coverage of 13.8x. Reads were mapped to the UMD3.1. bovine genome assembly using the Burrows Wheeler Aligner (BWA) tool [48]. Variants were called by using SAMtools 1.3.1 [49]. Assuming that the causative mutation is a dominant de novo mutation, all heterozygous variants detected for the affected calf were filtered against whole-genome sequence data from 2333 animals from more than 40 breeds that are part of the 1000 Bull Genomes Run 6.0 [19]. Private heterozygous SNPs and small indels with a quality score greater than 30 were annotated using the Ensembl Variant Effect Predictor pipeline (VeP) on the Ensembl v88 transcript set [50]. The SIFT software was used to predict the effect of non-synonymous coding variants [51]. Alignments of orthologous proteins among mammals were conducted using the CLUSTAL W (1.81) multiple sequence alignment tool implemented in Ensembl (www.ensembl.org) to evaluate the conservation of the amino-acids surrounding the sites of inframe insertions. Information on the domains that were missing in the proteins affected by frameshift mutations were obtained from the uniprot database (www.uniprot.org). Finally, the Integrative Genomics Viewer (IGV) [26] was used to search for structural variations within candidate genes (*EDA, EDAR, EDARADD and WTN10*) and 500-kb long regions up and downstream.

Primer design and PCR

Four primers were designed to amplify segments that encompass the two breakpoints of the inversion based on the *Bos_taurus*_UMD_3.1/bosTau6 assembly (UCSC Genome Browser Gateway) using the Primer3 software [52]. Details on the PCR primers are in Additional file 3: Table S3. Four independent PCR reactions were performed to amplify the mutant and wild type alleles using the GO-Taq Flexi polymerase (Promega) according to the manufacturer's instructions on a Mastercycle pro thermocycler (Eppendorf). Aliquots of PCR products were pooled for ethidium bromide-stained 2% agarose gel electrophoresis. The remaining amplicons were purified and sequenced by Eurofins (Eurofins MWG, Ebersberg Germany) using conventional Sanger sequencing.

Alignment and annotation of *XIST*

The boundaries of the *XIST* exons were determined by alignment of the *XIST* sequence NR_001464.2 [38] with the BLAT tool of the UCSC genome browser on the UMD3.1 genome assembly [53], and comparative analysis of the human and cattle sequences of *XIST* made by Yen et al. [37]. The positions of the seven exons of the bovine XIST transcript are presented in Additional file 4: Table S4.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12864-019-6087-1.

Additional file 1: Table S1. Number of expected calves with the same disease in the case of a recessive mutation.

Additional file 2: Table S2. Details on private heterozygous SNPs and small indels carried by the HED animal. QUAL and MQ scores: quality score of the variant and mapping quality score of the read containing the variant as calculated by SAMtools. DP4: number of reads containing the reference allele on the forward strand, the alternative allele on the forward strand, the reference allele on the reverse strand and the alternative allele on the reverse strand.

Additional file 3: Table S3. PCR primers for the amplification of segments spanning the inversion.

Additional file 4: Table S4. Position of *XIST* exons on the UMD3.1 bovine genome assembly.

Additional file 5: Figure S1. Phenotypic comparison of the affected calf (2 weeks, b-d-f) and a control (1 month, a-c-e). Clinical examination reveals abnormal eyelashes (a-b) but normal whiskers (c-d) and horn buds (e-f) in the affected versus control animal.

Additional file 6: Figure S2. Dental examination of the affected calf. X-ray of the proximal region of the lower jaw (a) reveals complete absence of deciduous and permanent incisors.

Additional file 7: Figure S3. Histological section of a horn bud from the affected animal. Note that the horn bud showed no notable morphological specificity. (PNG 472 kb)

Additional file 8: Figure S4. Visualization with IGV of the two breakpoints of the inversion affecting *XIST* (a) and *EDA* (b). Reads from pairs mapped with an aberrant orientation are color-coded to indicate the orientation of the second read of the pair (blue: reverse-reverse; green: forward-forward). Soft clipping parts of the reads are bright-colored.

Additional file 9: Figure S5. Plot of the average sequence depth in a 6-Mb interval encompassing the 3.8-Mb inverted segment. The sequence depth in the genomes of the affected animal (a) and a female control (b) was calculated for non-overlapping windows of 10 kb. Note the similarity of the profiles. A region that consisted of highly repeated centromeric satellites (ChrX:83151156–83,181,156) was removed for the sake of clarity (*). The localization of the inverted fragment is indicated with a double-headed arrow.

Additional file 10: Figure S6. Conservation of *XIST* exons (analyzed with Ensembl).

Additional file 11: Figure S7. Negative control of C-ter immunostaining in the affected calf. Negative control of C-ter immunostaining was performed on the affected animal without the primary antibodies. Note the complete absence of brown staining, which contrasts with the coloration in Fig. 6.

Abbreviations
BWA: Burrows Wheeler Aligner; *EDA*: *Ectodysplasin-A*; *EDAR*: *Ectodysplasin-A receptor*; *EDARRAD*: *EDAR associated death domain*; HED: Hypohidrotic Ectodermal Dysplasia; IGV: Integrative Genomics Viewer; INRA: French National Institute for Agricultural Research; ONAB: French National Observatory for Bovine Genetic Defects; PCR: Polymerase Chain Reaction; *PIK3C2B*: *Phosphatidylinositol-4-phosphate 3-kinase C2 domain-containing beta*; *SLC34A3*: *Solute carrier family 34 member 3 protein*; SNP: Single Nucleotide Polymorphism; SRA: Sequence Read Archive; *SUPT20H*: *suppressor of Ty 20 homolog*; TNF: Tumor necrosis Factor; VEP: Variant Effect Predictor; *WNT10A*: *Wingless-type mouse mammary tumor virus integration site family member 10A*; XCI: X-Chromosome Inactivation; *XIST*: *X inactive specific transcript*; XLHED: X-linked form of HED

Acknowledgements
This study could not have been completed without the generous cooperation of many people. We would like to thank DVM Marco Cartesegna (Clinique des Hortensias, 29290 Saint-Renan, France) for having reported the affected animal to the ONAB and for having accepted to perform sampling for the current study. We are also grateful to Marthe Vilotte (INRA) for technical assistance in preparing histological sections, Sarah Barbey (INRA) for providing pictures of a control animal, Chantal Kang (INRA) for performing X-ray of the jaw of the affected animal, Didier Boichard (INRA) for introducing the use of the pedig package, Hélène Leclerc (Idele) for her help in obtaining information on calf survival for 3577 Holstein bulls, and partners of the 1000 bull project and Mekki Boussaha for providing access to data of the 1000 bull genomes project. Carolyn Brown is specifically acknowledged for her advice and help to understand the role of *XIST* inactivation in our study case. Finally, we are grateful to Christian Beauvallet, Didier Chevret, Hélène Kiefer, Hélène Jammes and Aurélie Chaulot-Talmon for interesting discussions on experimental protocols.

Authors' information (optional)
Sebastien Orellana and Margarita Cano do not work at INRA anymore.

Authors' contributions
AC conceived and supervised the study. CE carried out the clinical, pedigree, WGS and histological analyses. CG provided samples and data information on the animal. MCh assisted CE in the WGS data analysis. SO and ER examined the candidate genes sequences and amplified the segments spanning the inversions. MCa and JR performed the histological analyses. AC and CE drafted the manuscript. HH participated in the discussions and preparation of the manuscript. All authors read and approved the final manuscript.

Funding
This study is part of the BOVANO project (ANR-14- CE19–0011) funded by the French Agence Nationale de la Recherche (Paris, France) and APIS-GENE (Paris, France). Clémentine Escouflaire is a recipient of a CIFRE PhD grant from Allice (Paris, France), with the financial support of the French Association Nationale de la Recherche et de la Technologie (Paris, France) and APIS-GENE. The funding bodies were not involved in the design of the stuy and collection, analysis and interpretation of data and in writing the manuscript.

Availability of data and materials
The genome sequence of the affected calf was deposited in the Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/sra) under the accession number: PRJNA534204. The genome sequence of the control female Holstein used for comparison of average sequencing depth is accessible under the accession number SAMEA3869561 in project PRJEB12703. The whole-genome sequence data of 2331 bovine individuals used for filtering is part of the 1000 Bull Genomes Run 6.0 [19]. More than half of these are already publicly available at accession numbers PRJNA238491, PRJEB9343, PRJNA176557, PRJEB18113, PRNJA343262, PRJNA324822, PRJNA324270, PRJNA277147, PRJEB5462. The complete dataset with all the variants of the 1000 Bull Genomes Run 6.0 used as control is available from the corresponding author on reasonable request.

Ethics approval and consent to participate
Blood samples from the affected calf and its dam were collected by a veterinarian during sampling for viral testing. Semen from the sire was bought from Bovec (Lentilly, France). Any invasive sampling was performed only post-mortem on the affected and control calves, after natural death; thus no ethical evaluation was required. Experiments reported in this work comply with the ethical guidelines of the French National Institute for Agricultural Research (INRA). All the samples and data analyzed were obtained with the permission of their owners.

Consent for publication
Written informed consent to publish was given by the breeder of the affected calf.

Competing interests
The authors declare they have no competing interest.

Author details
[1]INRA, GABI, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France. [2]Allice, 75595 Paris, France. [3]INRA, SIGENAE, GABI, 78350 Jouy-en-Josas, France. [4]INRA, Micalis Institute, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France.

Received: 25 April 2019 Accepted: 9 September 2019
Published online: 18 September 2019

References
1. Clarke A, Phillips DI. Clinical aspects of X-linked hypohidrotic ectodermal dysplasia. Arch Dis Child. 1987;62(10):989. https://doi.org/10.1136/adc.62.10.989.
2. Lefebvre S, Mikkola M. Ectodysplasin research—where to next? Semin Immunol. 2014;26:220–8.
3. Kuramoto T, Yokoe M, Hashimoto R, Hiai H, Serikawa T. A rat model of hypohidrotic ectodermal dysplasia carries a missense mutation in the Edaradd gene. BMC Genet. 2011;12:1–8.
4. Waluk D, Zur G, Kaufmann R, Welle M, Jagannathan V, Drögemüller C, et al. A splice defect in the EDA gene in dogs with an X-linked Hypohidrotic ectodermal dysplasia (XLHED) phenotype. G3: Genes|Genomes|Genetics. 2016;6:2949–54.
5. Drögemüller C, Kuiper H, Peters M, Guionaud S, Distl O, Leeb T. Congenital hypotrichosis with anodontia in cattle: a genetic, clinical and histological analysis. Vet Dermatol. 2002;13:307–13.
6. Seeliger F, Drögemüller C, Tegtmeier P, Baumgärtner W, Distl O, Leeb T. Ectodysplasin-1 deficiency in a German Holstein bull associated with loss of respiratory mucous glands and chronic Rhinotracheitis. J Comp Pathol. 2005;132:346–9.
7. Zankl A, Addor M-C, Cousin P, Gaide A-C, Gudinchet F, Schorderet D. Fatal outcome in a female monozygotic twin with X-linked hypohydrotic ectodermal dysplasia (XLHED) due to a de novo t (X;9) translocation with probable disruption of the EDA gene. Eur J Pediatr. 2001;160:296–9.
8. Cluzeau C, Hadj-Rabia S, Jambou M, Mansour S, Guigue P, Masmoudi S, et al. Only four genes (EDA1, EDAR, EDARADD, and WNT10A) account for 90% of hypohidrotic/anhidrotic ectodermal dysplasia cases. Hum Mutat. 2011;32:70–2.
9. Tziotzios C, Petrof G, Liu L, Verma A, Wedgeworth EK, Mellerio JE, et al. Clinical features and WNT10A mutations in seven unrelated cases of Schöpf–Schulz–Passarge syndrome. Br J Dermatol. 2014;171:1211–4.
10. Gordon M, Nusse R. Wnt signaling: multiple pathways, multiple receptors, and multiple transcription factors. J Biol Chem. 2006;281:22429–33.
11. Zeng B, Xiao X, Li S, Lu H, Lu J, Zhu L, et al. Eight mutations of three genes (EDA, EDAR, and WNT10A) identified in seven Hypohidrotic ectodermal dysplasia patients. Genes. 2016;7:65.
12. Wiśniewski SAA, Kobielak A, Trzeciak WHH, Kobielak K. Recent advances in understanding of the molecular basis of anhidrotic ectodermal dysplasia: discovery of a ligand, ectodysplasin a and its two receptors. J Appl Genet. 2002;43:97–107.
13. Cui C, Schlessinger D. EDA signaling and skin appendage development. Cell Cycle. 2006;5(21):2477.
14. Karlskov-Mortensen P, Cirera S, Nielsen OL, Arnbjerg J, Reibel J, Fredholm M, et al. Exonization of a LINE1 fragment implicated in X-linked hypohidrotic ectodermal dysplasia in cattle. Anim Genet. 2011;42:578–84.
15. Cambiaghi S, Restano L, Pääkkönen K, Caputo R, Kere J. Clinical findings in mosaic carriers of hypohidrotic ectodermal dysplasia. Arch Dermatol. 2000; 136:217–24.

16. Barlund CS, Clark EG, Leeb T, Drögemüller C, Palmer CW. Congenital hypotrichosis and partial anodontia in a crossbred beef calf. Can Vet J. 2007;48:612–4.
17. Ogino A, Kohama N, Ishikawa S, Tomita K, Nonaka S, Shimizu K, et al. A novel mutation of the bovine EDA gene associated with anhidrotic ectodermal dysplasia in Holstein cattle. Hereditas. 2011;148:46–9.
18. Daetwyler H, Capitan A, Pausch H, Stothard P, van BR, Brøndum R, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat Genet. 2014;46:ng.3034.
19. Bouwman A, Daetwyler H, Chamberlain A, Ponce C, Sargolzaei M, Schenkel F, et al. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nat Genet. 2018;50:362–7.
20. Michot P, Fantini O, Braque R, Allais-Bonnet A, Saintilan R, Grohs C, et al. Whole-genome sequencing identifies a homozygous deletion encompassing exons 17 to 23 of the integrin beta 4 gene in a Charolais calf with junctional epidermolysis bullosa. Genet Sel Evol. 2015;47:1–7.
21. Robinson J, Thorvaldsdóttir H, Winckler W, Guttman M, Lander E, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24.
22. Hastings IG, Lupski J. A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genet. 2009;5: e1000327.
23. Sonnhammer E, Durbin R. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene. 1995; 167:GC1–GC10.
24. Bourneuf E, Otz P, Pausch H, Jagannathan V, Michot P, Grohs C, et al. Rapid discovery of De novo deleterious mutations in cattle enhances the value of livestock as model species. Sci Rep. 2017;7:11466.
25. Segawa H, Onitsuka A, Kuwahata M, Hanabusa E, Furutani J, Kaneko I, et al. Type IIc sodium–dependent phosphate transporter regulates calcium metabolism. J Am Soc Nephrol. 2009;20:104–13.
26. Myakala K, Motta S, Murer H, Wagner C, Koesters R, Biber J, et al. Renal-specific and inducible depletion of NaPi-IIc/Slc34a3, the cotransporter mutated in HHRH, does not affect phosphate or calcium homeostasis in mice. Am J Physiol-renal. 2014;306:F833–43.
27. Zohn I, Li Y, Skolnik E, Anderson K, Han J, Niswander L. p38 and a p38-interacting protein are critical for downregulation of E-cadherin during mouse gastrulation. Cell. 2006;125:957–69.
28. Harada K, Truong AB, Cai T, Khavari PA. The class II phosphoinositide 3-kinase C2β is not essential for epidermal differentiation. Mol Cell Biol. 2005; 25(24):11122. https://doi.org/10.1128/MCB.25.24.11122-11130.2005.
29. Barakat T, Jonkers I, Monkhorst K, Gribnau J. X-changing information on X inactivation. Exp Cell Res. 2010;316:679–87.
30. Senner C, Brockdorff N. Xist gene regulation at the onset of X inactivation. Curr Opin Genet Dev. 2009;19:122–6.
31. Renault N, Dyack S, Dobson M, Costa T, Lam W, Greer W. Heritable skewed X-chromosome inactivation leads to haemophilia a expression in heterozygous females. Eur J Hum Genet. 2007;15:5201799.
32. Belmont JW. Genetic control of X inactivation and processes leading to X-inactivation skewing. Am J Hum Genet. 1996;58:1101–8.
33. Clerc P, Avner P. Role of the region 3′ to Xist exon 6 in the counting process of X-chromosome inactivation. Nat Genet. 1998;19:ng0798_249.
34. Plenge RM, Hendrich BD, Schwartz C, Arena JF, Naumova A, Sapienza C, et al. A promoter mutation in the XIST gene in two unrelated families with skewed X-chromosome inactivation. Nat Genet. 1997;17:353–6.
35. Marahrens Y, Panning B, Dausman J, Strauss W, Jaenisch R. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. 1997;11(2):156. https://doi.org/10.1101/gad.11.2.156.
36. Lee HJ, Gopalappa R, Sunwoo H, Choi S-WW, Ramakrishna S, Lee JT, et al. En bloc and segmental deletions of human XIST reveal X chromosome inactivation-involving RNA elements. Nucleic Acids Res. 2019;47(8):3875.
37. Yen Z, Meyer I, Karalic S, Brown C. A cross-species comparison of X-chromosome inactivation in Eutheria. Genomics. 2007;90:453–63.
38. Chureau C, Prissette M, Bourdet A, Barbe V, Cattolico L, Jones L, et al. Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. Genome Res. 2002;12:894–908.
39. Kere J, Grzeschik KH, Limon J, Genomics G-M. Anhidrotic ectodermal dysplasia gene region cloned in yeast artificial chromosomes. Genomics. 1993;16(2):305 https://www.sciencedirect.com/science/article/pii/S0888754383711894.
40. MacDermot KD, genetics H-M. Female with hypohidrotic ectodermal dysplasia and de novo (X; 9) translocation. Hum Genet. 1990;84(6):577 https://link.springer.com/article/10.1007/BF00210814.
41. Gale RE, Wheadon H, Boulos P, Blood L-D. Tissue specificity of X-chromosome inactivation patterns. Blood. 1994;83(10):2899 http://www.bloodjournal.org/content/83/10/2899.short.
42. Naumova AK, Plenge RM, Bird LM, Leppert M, Morga K, Willard HF, Sapienza C. Heritability of X chromosome--inactivation phenotype in a large family. Am J Hum Genet. 1996;58(6):1111 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1915075/.
43. Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. Am J Hum Genet. 1992;51(6):1229 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1682906/.
44. Pereira L, Zatz M. Screening of the C43G mutation in the promoter region of the XIST gene in females with highly skewed X-chromosome inactivation. Am J Med Genet. 1999;87:86–7.
45. Pääkkönen K, Cambiaghi S, Novelli G, Ouzts L, Penttinen M, Kere J, et al. The mutation spectrum of the EDA gene in X-linked anhidrotic ectodermal dysplasia. Hum Mutat. 2001;17:349.
46. Puig M, Castellano D, Pantano L, Giner-Delgado C, Izquierdo D, Gayà-Vidal M, et al. Functional impact and evolution of a novel human polymorphic inversion that disrupts a gene and Creates a fusion transcript. PLoS Genet. 2015;11:e1005495.
47. Boichard D. PEDIG: a fortran package for pedigree analysis suited for large populations. 7th world congress on genetics applied to livestock production; 2002. p. 28–13.
48. Li H, Durbin R. Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics. 2009;25:1754–60.
49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25: 2078–9.
50. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP effect predictor. Bioinformatics. 2010;26:2069–70.
51. Kumar P, Henikoff S, Ng P. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4:nprot.2009.86.
52. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth B, Remm M, et al. Primer3—new capabilities and interfaces. Nucleic Acids Res. 2012;40:e115.
53. Kent W. BLAT—the BLAST-like alignment tool. Genome Res. 2002;12:656–64.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC

3.2. Discussion

L'étude du mécanisme génétique à l'œuvre et la caractérisation de l'anomalie génétique affectant cette génisse sans poils et sans dents est un cas unique qui illustre la complexité des phénomènes liés à l'inactivation aléatoire du chromosome X. Chez certaines femelles, notamment dans le cas d'anomalies chromosomiques de grande taille touchant l'un des deux chromosomes X ou d'une anomalie génétique très délétère, un biais d'inactivation se produit, favorisant l'expression du chromosome X non muté (Payet, 2017). Dans le cas de l'animal étudié ici, le phénomène inverse s'est produit, dû à l'altération du gène *XIST*, qui initie l'inactivation.

Au-delà de l'article présenté précédemment, nous avons réalisé de nombreuses tentatives afin de démontrer de manière plus directe l'inactivation non aléatoire d'un des deux chromosomes X. La principale contrainte était l'absence d'échantillons d'ARN, nous avons donc essayé d'autres alternatives.

Une première étude protéomique par analyse de spectrométrie de masse a été réalisée à partir de protéines extraites d'échantillons de peau fixée avec du formol 10%. L'objectif était de mettre en évidence un biais de représentation allélique pour au moins une des 78 protéines codées par des gènes du chromosome bovin X et présentant un polymorphisme non synonyme porté à l'état hétérozygote chez notre cas. Malheureusement, ces protéines n'étaient pas ou trop peu exprimées dans la peau. Seuls quatre peptides ont été identifiés et le nombre d'observations indépendantes pour chacun d'eux (n = 3 pour le plus représenté) était insuffisant pour réaliser des analyses statistiques.

Nous avons ensuite choisi d'utiliser les propriétés de méthylation du chromosome X inactivé. L'inactivation se traduit par la méthylation d'un chromosome X empêchant la fixation de ribosomes nécessaires pour la transcription. Une première étape d'appariement des chromosomes X permet de compter leur nombre dans la cellule. L'inactivation ne s'initie que si la présence de plusieurs chromosomes X est détectée. Cela explique notamment que les femmes souffrant d'une monosomie du chromosome X (appelée aussi syndrome de Turner) soient viables. Des gènes échappent cependant à l'inactivation, il s'agit principalement des gènes situés sur la région PAR, même si d'autres clusters de gènes ont été identifiés, notamment sur les bras courts du chromosome. Ce sont ces gènes qui expliquent que même si les femmes atteintes d'une monosomie du chromosome X sont viables, elles présentent également des particularités phénotypiques comme une petite taille, une dysgénésie gonadique, voire des malformations rénales et cardiaques.

Le chromosome X inactivé étant méthylé, nous avons essayé une méthode similaire au « HUMARA assay » utilisé chez l'humain, sur les conseils d'un des membres du comité de thèse, François Vialard. Cette méthode consiste à amplifier un segment chromosomique comprenant à la fois un polymorphisme hétérozygote et un site de méthylation, après utilisation d'une enzyme de restriction sensible à la méthylation et un contrôle de son stéréoisomère non sensible, afin de mettre en évidence un biais de méthylation associé à un allèle du polymorphisme étudié. Cependant, nous n'avons pas réussi à obtenir une digestion enzymatique suffisante à l'échelle du génome, et malgré plusieurs tentatives d'amélioration de notre protocole, nous n'avons pas obtenu de résultats utilisables. Toujours dans le but d'étudier la méthylation en lien avec l'inactivation du chromosome X, nous avons alors fait séquencer le génome de l'animal atteint ainsi que celui de sa mère en RRBS (*Reduced Representation Bisulfite Sequencing ;* Integragen SA, France) sur les conseils de collègues de l'unité

BREED. L'objectif était là encore d'utiliser comme marqueurs des polymorphismes proches des sites de méthylation dans le but de démontrer une différence statistique de méthylation entre les deux chromosomes X chez l'atteint, puis d'utiliser la mère en contrôle. Contrairement à notre hypothèse, nous n'avons pas observé une hausse du nombre de régions présentant un biais de méthylation chez le contrôle par rapport à sa mère sur l'ensemble du chromosome (et donc en majorité dans des régions non codantes). Il semblerait que le biais de méthylation associé au biais d'inactivation du chromosome X, tel qu'observé chez l'humain au niveau du promoteur du récepteur aux androgènes, ne concerne que les régions promotrices des gènes. Nous avons réalisé un dernier test en regardant si dans nos données de séquence nous avions des régions méthylées dans les promoteurs de gènes. Malheureusement, aucune des régions promotrices de gènes couvertes par le RRBS sur le chromosome X ne présentait de marqueurs informatifs (hétérozygotes chez notre cas et chez sa mère). Devant le cumul de ces tentatives infructueuses et le délai imposé par ma thèse, nous avons décidé de laisser de côté la démonstration du biais d'inactivation et de publier rapidement, sans cette démonstration.

Plusieurs autres possibilités n'ont pas été explorées, par manque de temps et de moyens. Le pyroséquençage aurait pu permettre d'amplifier des régions dans les promoteurs de gènes contenant des îlots CpG et un polymorphisme présent chez l'individu atteint et sa mère. Cependant rien n'indique que le profil de méthylation aurait été plus indicatif que ce que nous avons obtenu avec le RRBS.

Lors de cette étude, nous avons exploré la possibilité de l'existence d'un transcrit chimérique fusionnant une partie de XIST avec une partie d'EDA. Il aurait été très intéressant de confirmer cette transcription de fusion, mais malheureusement, nous n'avons pas pu y parvenir en raison des conditions dans lesquelles les tissus ont été prélevés, c'est-à-dire échantillonnés et fixés dans le formol 2 jours après la mort. La façon la plus simple de détecter une transcription de fusion aurait été d'analyser la séquence d'ARN par qRT-PCR ou RNA-Seq ou encore par hybridation *in situ* sur des lames histologiques. Cela n'était pas possible en raison du délai entre la mort et le prélèvement des tissus et de la dégradation rapide de l'ARN dans les échantillons biologiques.

Une autre méthode aurait consisté à effectuer une analyse immuno-histochimique avec un anticorps monoclonal personnalisé. Cela impliquerait de : 1- produire un peptide qui résulterait de la traduction de la partie XIST du transcrit chimérique présumé, 2- l'injecter dans une souris pour la synthèse d'anticorps, 3- recueillir les cellules B de sa rate, et 4- fusionner les cellules B avec un myélome et rechercher le clone cellulaire approprié. Il faudrait au moins six mois pour produire cet anticorps monoclonal. De plus, nous ne disposons pas de tissu témoin positif pour vérifier la spécificité de l'anticorps. Nous n'avons testé pas cette méthode car elle est coûteuse, longue et sans garantie de résultats.

Résultats clés de cette étude :

- Nous avons mis en évidence un modèle de transmission lié au chromosome X d'un cas unique de dysplasie ectodermique hypohidrotique chez une femelle Holstein causée par une inversion de 4 Mb entre les gènes *EDA* et *XIST.*
- Un biais d'inactivation dû à l'inactivation du gène *XIST* amène à l'inactivation complète du chromosome X sain.
- Nous avons émis l'hypothèse de la présence d'un transcrit chimérique commençant avec le premier exon de *XIST* et continuant avec les exons 2 à 8 de *EDA*. Malheureusement, l'absence d'échantillon d'ARN n'a pas permis une démonstration plus poussée.
- Dans la discussion nous avons évoqué différentes méthodes permettant de démontrer ce mécanisme et les limites dues au matériel biologique à notre disposition.

Chapitre 5 : Vers une meilleure prise en compte des chromosomes sexuels et du génome mitochondrial dans les applications de génétique

Les objectifs de cette thèse étaient de décrire la diversité génétique des chromosomes sexuels et du génome mitochondrial, d'étudier leur potentiel de sélection et de mesurer l'intérêt de leur prise en compte dans les évaluations génomiques. Avec le recul que m'ont procuré ces trois ans de thèse, je prends conscience que nous aurions dû être moins ambitieux. La diversité des sujets et des méthodes nous a conduit à prospecter en parallèle de nombreux aspects.

L'objectif de ce chapitre est (i) de revenir sur les limites rencontrées au cours de cette thèse et des solutions proposées, (ii) de proposer des perspectives de recherche et de valorisation des résultats obtenus, et enfin (iii) de discuter l'intérêt de - et les moyens nécessaires à - la prise en compte des chromosomes X, Y et du génome mitochondrial dans les évaluations génomiques.

1. Limites des études

Au cours de ma thèse, nous avons fait face à plusieurs défis. Ils concernent les trois types d'informations que nécessitent les études de génétique : les phénotypes, les génotypes ou séquences et les informations de parenté.

1.1. Phénotypes

La qualité et la quantité des échantillons biologiques sont primordiales lorsqu'on traite un cas d'anomalie génétique, d'autant plus si ce cas est unique. L'étude du cas d'un veau Holstein atteint de dysplasie ectodermique hypohidrotique, une anomalie génétique liée au chromosome X, l'illustre bien. L'absence d'ARN, due à la mauvaise qualité de conservation des tissus, nous a conduit à tester différentes méthodes alternatives pour démontrer le biais d'inactivation du X, qui finalement se sont révélées inefficaces. Ces nombreuses tentatives que nous n'avions pas anticipées avant la découverte de la nature de la mutation causale, ont été motivées par l'envie de prouver l'originalité du déterminisme génétique de cette anomalie et d'assurer une première publication dans une revue scientifique de qualité. Malheureusement, elles ont eu comme conséquence de retarder les autres études menées en parallèle.

Le travail de l'ONAB en collaboration avec les vétérinaires et les éleveurs est très important pour essayer de limiter le nombre de situations comme la nôtre. Lorsqu'une anomalie est déclarée, avant la mort de l'animal, il faut pouvoir décrire précisément la nature, la quantité et le type de conservation des prélèvements, tout en se coordonnant avec la personne sur le terrain qui va pouvoir réaliser ces échantillonnages. Ce travail est primordial pour pouvoir assurer le travail de recherche qui va ensuite être réalisé. Afin d'anticiper l'apparition de nouvelles anomalies non décrites auparavant, l'ONAB a mis en place une feuille de description permettant de fournir le plus d'éléments possible sur les symptômes d'un animal atteint (ONAB, 2020). Ces renseignements permettent d'orienter l'équipe sur la méthodologie à suivre pour optimiser les prélèvements.

En dehors des échantillons biologiques, la justesse et la bonne caractérisation des phénotypes sont un problème très souvent rencontré lorsqu'un nouveau caractère est étudié. Au cours de ma thèse, lorsque j'ai exploité les données de collecte de semence pour étudier la fertilité mâle, je me suis rendue compte qu'en fonction des années, les mesures n'avaient pas toutes été faites en suivant la même méthodologie. A l'origine, ces données ne sont pas à destination de projets de recherche, et servent

aux ES à contrôler à court terme les performances des taureaux. Cependant, ces données représentent un réel potentiel pour permettre de prédire la fertilité des taureaux. L'homogénéisation des méthodes est primordiale pour pouvoir traiter efficacement un ensemble de données. Si l'on souhaite étudier le lien statistique entre le génotype et le phénotype pour, par exemple, développer à long terme une évaluation génétique, il faut mettre en place une démarche durable de collecte de données. Cela nécessite de réfléchir, en amont de la création de la base de collecte, à la meilleure manière de standardiser les mesures entre les différents opérateurs. Non seulement, il faut que les données collectées soient mesurées de la même manière, mais aussi que les informations relatives aux effets de l'environnement soient enregistrées. Dans notre cas, la mention de l'identité du taurelier pour chacune des collectes aurait probablement pu améliorer la précision de notre analyse avec la prise en compte d'un effet opérateur.

Dans le cas des deux entreprises de sélection qui m'ont fourni les données, des méthodes de mesure CASA (*computed-aided sperm analysis*) ont été mises en place depuis quelques années. L'utilisation d'un outil comme l'IVOS (IMV technologies), plus précis pour effectuer les mesures devrait permettre de les standardiser. Cependant, les mesures dépendent du paramétrage de l'appareil qui est ajusté selon des critères définis par chaque utilisateur. Les laboratoires paramètrent leur appareil selon une grille de notation qui leur est propre. Au sein de l'EFFAB (*European Forum of Farm Animal Breeders*), le groupe RepVet (*Reproduction Veterinarians*) a créé deux bureaux d'étude Qualivet en bovin et porcin chargés de projets autour de la qualité de semence. Les objectifs sont de faire un état des lieux des pratiques, d'étudier la variation des résultats selon le paramétrage des appareils et d'établir une grille de paramètres ainsi qu'un guide de bonnes pratiques de préparation de la semence (article soumis par le groupe Qualivet fin mars). En fonction des paramétrages, ils ont observé une variation de 15 à 60% du pourcentage de spermatozoïdes progressifs (communication personnelle, Eli Sellem président du groupe Qualivet bovin). Cela démontre l'importance de standardiser les protocoles de collecte et de mesure des critères utilisés pour pouvoir comparer les performances des taureaux entre centres de collecte différents.

Suite à l'analyse des enregistrements de collecte de semence, j'ai sensibilisé les ES aux problématiques rencontrées concernant ce jeu de données, ainsi que mes collègues d'Allice souhaitant exploiter ces données pour des projets futurs comme l'étude de l'impact des effets d'environnement sur la production de semence.

1.2. Génotypage

Le génotypage sur les puces à ADN peut se révéler difficile pour les variants rares ou sans possibilité d'hétérozygotie. A Labogena, les signaux d'hybridation des marqueurs sont captés par un scanner et analysés par le logiciel Genome Studio. Les résultats de chaque marqueur sont ensuite interprétés selon leurs regroupements ou « cluster ». Dans une situation classique, les trois génotypes A1A1, A1A2 et A2A2 sont répartis sur trois nuages de points distincts (cf. figure 33).

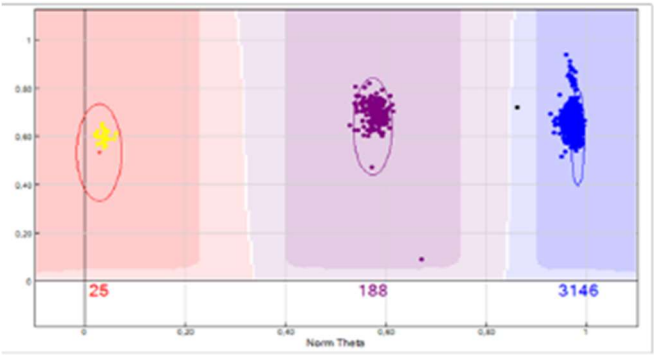


Figure 33 : Interprétation des génotypes par le logiciel Genome Studio (Michot, P. 2017) (expression des signaux d'hybridation en coordonnées polaires)

Les points localisés en dehors des clusters formés automatiquement sont catégorisés comme génotypages indéterminés ou manquants. Lorsque les nuages de points ne sont pas suffisamment distincts, il est impossible d'interpréter les résultats de l'ensemble des génotypages à ce marqueur.

Dans sa thèse réalisée au sein de notre équipe, Pauline Michot (2017) avait placé 785 marqueurs de variants rares sur la puce EuroG10k. Parmi ces marqueurs, 239 marqueurs ont été éliminés à cause de leur manque de fiabilité : 59 marqueurs pour lesquels les clusters n'ont pas pu être définis et 180 marqueurs avec un call rate[8] inférieur à 95%. Le taux de chute des variants rares est donc relativement élevé.

La bonne définition des clusters est importante pour pouvoir définir au mieux le génotype. Dans le cas de nos études sur le chromosome X, nous avons étudié des variants à fréquence rare, pour lesquels une éventuelle mauvaise définition des clusters pourrait artificiellement créer des hétérozygotes. Concernant les marqueurs du chromosome Y et du génome mitochondrial dont le génotype devrait systématiquement être homozygote, une fréquence relativement faible d'hétérozygotes a été parfois observé. Par exemple en race Montbéliarde, sur les 563 marqueurs mitochondriaux placés sur la puce EuroGMD, 210 présentent des hétérozygotes avec cependant une fréquence très faible (en moyenne 0,000322 avec un écart-type de 0,0017). Afin d'améliorer le génotypage de ces individus atypiques, nous pourrions retravailler avec Labogena sur la définition des clusters et mettre en place, par exemple, une séparation entre mâles et femelles pour les marqueurs du chromosome X afin de déterminer plus précisément le génotype des mâles.

1.3. Séquençage

Au cours de cette thèse, nous avons utilisé deux méthodes différentes pour séquencer le génome mitochondrial. Dans le cadre des 384 génomes mitochondriaux séquencés dans l'étude 1 du chapitre 2, nous avons amplifié l'ADNmt en trois fois, avec des PCR long-range et un séquençage NGS a été effectué par nos collaborateurs de Munich. Pour des raisons logistiques, le séquençage des échantillons non-T a été réalisé par amplification de 11 segments (chacun réalisé deux fois) suivi d'un séquençage Sanger. De par le nombre de segments amplifiés, cette seconde méthode augmente considérablement le temps passé à produire une seule séquence complète d'ADNmt. Après l'amplification, il faut encore contrôler la qualité des séquences produites, réaliser les alignements pour fusionner ensuite les séquences partielles et recréer la séquence du génome mitochondrial. En multipliant le nombre d'amplifications, on augmente également le risque d'erreurs de pipetage et d'interversion d'échantillons. Cette seconde méthode nécessite aussi une quantité d'ADN plus élevée.

Le génotypage dans le cadre de la sélection génomique représente pour nous une base d'ADN très précieuse qui nous permet de récupérer les échantillons biologiques grâce à Labogena. Cela représente des gains de temps et d'énergie importants en évitant de demander des prélèvements sanguins ou biopsies à des dizaines de techniciens différents dans les élevages et de réaliser l'extraction nous-mêmes. Cependant, cela limite aussi la quantité d'ADN que nous pouvons récupérer par ce biais puisque nous récupérons le reste d'ADN après génotypage. Cette limite s'est fait particulièrement ressentir avec l'approche de séquençage Sanger.

L'amplification de certains couples échantillon – amorce n'a pas pu aboutir malgré plusieurs essais et tentatives d'amélioration de notre protocole d'amplification, nous contraignant à demander des

[8] nombre de génotypages déterminés / nombre total de génotypages, exprimé en pourcentage

échantillons d'ADN supplémentaires pour compléter le panel d'individus. Ce phénomène a été encore plus marqué lors du séquençage mitochondrial des animaux porteurs d'haplotypes non-T en 22 amplifications différentes. La qualité et la quantité d'ADN que nous avions à disposition ne nous a pas permis d'amplifier l'ensemble des fragments pour tous les individus, ce qui explique pourquoi les analyses ont porté sur 13 065 bp au lieu de l'intégralité des séquences mitochondriales (16 338 bp).

1.4. Pedigree

L'exactitude des informations de pedigree et la prise en compte de l'apparentement entre les animaux sont importantes pour assurer la précision des évaluations génétiques. Dans les évaluations génétiques basées sur le BLUP, une matrice de parenté est construite sur la base des informations généalogiques enregistrées. Cependant, les erreurs de pedigree sont courantes, avec des taux d'erreurs significatifs (Banos *et al.*, 2001). Au Royaume-Uni, le taux global d'erreurs de pedigree a été estimé à 10% et Visscher *et al.* (2002) ont estimé que cela entraînerait une perte de réponse à la sélection de 2 à 3 %. Banos *et al.* (2001) ont montré qu'avec 11 % d'erreurs de pedigree, il y avait une réduction de 11 à 18 % de la moyenne des index (calculé sur la quantité de lait, matières grasses et protéiques).

Même si la perspective d'une généalogie faussée peut être un sujet d'inquiétude concernant la justesse des liens de parentés dans les évaluations génétiques, il faut noter que les erreurs de pedigrees que nous avons identifiées sont relativement anciennes. Or, plus les erreurs de pedigree sont anciennes, moins elles ont d'impact sur les évaluations des animaux récents. En évaluation génétique, les parentés les plus proches ont un effet prépondérant, limitant l'impact des erreurs anciennes de pedigree, puisque la covariance entre apparentés (pour les autosomes) est divisée par deux à chaque génération. Ce risque est donc négligeable dans le contexte des évaluations génétiques à partir de cinq générations ($1/2^n$ ~0,03), et ce d'autant plus que les races sont consanguines (principalement les races laitières) et que l'ancêtre réel et l'ancêtre affiché sont forcément apparentés de près ou de loin. Le problème de la justesse des généalogies anciennes est davantage un frein pour l'étude des transmissions uniparentales, telles que conduites dans cette thèse.

Pour améliorer la précision du pedigree, des méthodes de vérification de la parenté ont été mises en place. Jusqu'à récemment, les marqueurs microsatellites constituaient l'approche standard de la vérification de la filiation. Avec l'introduction des polymorphismes à nucléotide unique (SNP) et de la sélection génomique, les méthodes de filiation basées sur les SNP deviennent désormais l'approche standard. La norme internationale a tout d'abord été d'utiliser l'ensemble de SNP proposés par l'ISAG (ISAG100 ou ISAG200) ; cependant, McClure *et al.* (2015) ont suggéré qu'un panel avec un minimum de 500 SNP est plus approprié pour la vérification et surtout la prédiction (ou assignation) de la filiation mais le gain est relativement réduit. Lorsque le test génétique n'est réalisé que pour le contrôle de filiation, le panel de 200 marqueurs est généralement techniquement suffisant et constitue un bon rapport qualité-prix en conditions bovines (Barbotte *et al.*, 2012)(Barbotte et al., 2012)(Barbotte et al., 2012)(Barbotte et al., 2012)(Barbotte et al., 2012)(Barbotte et al., 2012)(Barbotte et al., 2012)(Barbotte et al., 2012)(Barbotte et al., 2012)(Barbotte et al., 2012)(Barbotte et al., 2012)(Barbotte et al., 2012)(Barbotte et al., 2012). Ces données sont aussi facilement échangeables entre pays. Si le test génétique est réalisé à des fins de sélection génomique, alors le nombre de marqueurs disponibles est important et on peut en utiliser plus que 200 sans surcoût. En France, les tests officiels des vérifications de parenté se basent sur le panel officiel ISAG200 mais la chaine d'évaluation génomique en considère 500. Les SNP complémentaires ont été choisis parmi les

marqueurs déjà présents sur la plupart des puces Illumina selon leur répartition sur les 29 autosomes bovins et leur informativité (communication personnelle, Sébastien Fritz).

Jusqu'à la mise en place de la sélection génomique à la fin des années 2000, la vérification de la parenté concernait uniquement les individus appartenant à des schémas de sélection ou inscrits à des livres généalogiques. Depuis lors, cette vérification est faite en routine pour tous les animaux génotypés dont les parents sont génotypés, ce qui est le cas pour la quasi-totalité des pères et augmente progressivement pour les mères. Grâce aux nombreux marqueurs mitochondriaux placés sur la puce EuroGMD, nous avons pu retracer les erreurs de pedigree. Nous avons également discuté la possibilité d'erreurs de pedigree en nous basant sur l'étude de variants du chromosome Y en race Holstein. L'étude des voies de transmission uni-parentale n'est pas exhaustive mais permettrait de corriger des erreurs relativement anciennes. D'un point de vue pragmatique, nous ne pourrions pas corriger officiellement les pedigrees. Le risque de déclassification d'animaux ou de contentieux avec les éleveurs est trop élevé. Cependant cette correction nous permettrait d'améliorer la fiabilité des pedigrees dans le cadre des études menées sur les voies de transmission uniparentale.

2. Perspectives et suites

2.1. Effets sur les caractères de production

A court-terme, notre objectif est de combiner dans un même article l'étude de la diversité génétique des séquences mitochondriales produites au cours de cette thèse et de nouvelles études d'association après l'accumulation de nouveaux génotypages. Nous ne sommes pas la seule équipe à étudier l'impact du génome mitochondrial sur les caractères de production. En effet, Le projet « MitoTAUROmics » (2014-2018) financé par l'université de Zagreb en Croatie à des objectifs similaires aux nôtres : analyser la diversité du génome mitochondrial chez les bovins modernes et étudier son impact fonctionnel sur les caractères de production bovins (Curik *et al.*, 2017). Le projet est fini depuis 2018 mais aucune étude d'association n'a encore été publiée à ce jour sur ce thème chez le bovin d'après une recherche bibliographique dans la base de données « Entrez pubmed » (https://www.ncbi.nlm.nih.gov/pubmed, dernier accès le 06/04/2020). De plus, la mise en évidence d'un taux non négligeable d'erreurs de filiations sur la voie maternelle remettrait en cause un certain nombre d'études antérieures portant sur l'étude de l'effet de lignées maternelles (à partir d'informations de pedigree) sur des caractères phénotypiques. Cet article serait donc particulièrement intéressant et novateur pour la communauté scientifique de la génétique animale.

Concernant l'étude du chromosome X, j'ai réalisé de premières études d'association sur des caractères de production laitière et de fertilité femelle (résultats non présentés) qui nécessitent d'être poursuivies en collaboration avec mes collègues de l'équipe G2B en charge des détections de QTL. Une première étape consistera à vérifier la qualité des imputations réalisées des génotypes du chromosome X, qui ont été mises en place très récemment suite au changement d'assemblage du génome bovin de référence (cf. partie 4.1). Il serait intéressant de réaliser une étude à la fois sur la fertilité femelle et mâle car des QTL liés à ces deux caractères ont été identifiés sur le chromosome X chez le bovin (Fortes *et al.*, 2013).

Nous avons démontré la faible diversité des LP, avec une fixation de l'un ou l'autre des haplogroupe Y1 et Y2 dans la majorité des races et une forte dérive liée à l'utilisation d'un nombre limité de pères à taureaux génération après génération. Ce manque de variabilité du chromosome Y au sein de chaque race limite grandement son intérêt en sélection. Il serait intéressant de suivre la transmission de la lignée paternel

le Y1 en race Montbéliarde si celle-ci est amenée à se développer, voire de développer des schémas de croisement pour comparer les effets des haplogroupes Y1 et Y2 sur les caractères de fertilité mâle. En fonction des résultats, cela pourrait permettre d'envisager l'introgression de chromosomes Y plus favorables dans des races fixées pour un chromosome Y défavorable.

2.2. Génétique inverse

Les séquences mitochondriales que nous avons produites pourront par la suite nous permettre d'étudier le rôle des variants sur des anomalies putatives. Le projet « MitoTAUROmics » a permis d'identifier chez le bovin deux variants candidats pour une anomalie de la vision, connue chez l'humain comme la neuropathie optique de Leber (Novosel *et al.*, 2019). Au cours de ces trois ans de thèse, nous avons privilégié l'étude d'association avec les caractères de production plutôt que la recherche d'anomalies génétiques. Cependant l'exploitation des séquences produites représente un potentiel intéressant pour une approche de génétique inverse. Les anomalies de la chaine respiratoire mitochondriale ne concernent pas uniquement les gènes portés par le génome mitochondrial. Comme indiqué précédemment, plus de 1000 gènes nucléaires codent pour des protéines mitochondriales qui sont essentielles au métabolisme mitochondrial et interagissent avec les protéines produites par le génome de cet organite. En amont de ma thèse, j'ai participé à la caractérisation d'un polymorphisme nucléaire impactant la chaine respiratoire mitochondriale. L'anomalie baptisée MTCP ou mitochondropathie, est présente en race Montbéliarde et provoque de la mortalité juvénile, majoritairement avant que les veaux n'atteignent 6 mois d'âge. J'ai pu participer aux débuts de la caractérisation de cette anomalie, principalement en contactant les éleveurs pour établir la liste des symptômes. La protéine mutée induit une défaillance de la chaine respiratoire mitochondriale, ce qui se traduit par une maladie systémique. Il serait intéressant de rechercher des polymorphismes potentiellement délétères dans les régions codantes du millier de gènes nucléaires codant pour des protéines mitochondriales identifiées par le projet MitoCarta2.0 (Calvo *et al.*, 2016) , de dessiner des marqueurs pour les puces à SNP et de phénotyper les animaux homozygotes mutés selon une approche de génétique inverse. Par ailleurs, il existe un phénomène appelé incompatibilité mito-nucléaire (IMN) qu'il serait intéressant à prendre en compte à ce niveau et dans l'étude des effets sur les caractères de production. L'IMN est due à une incompatibilité entre des variants affectant des protéines d'origine nucléaire et d'origine mitochondriale, qui n'arrivent plus à interagir de manière efficace dans la mitochondrie et sont à l'origine de pathologies uniquement lorsqu'elles sont combinées. Leur détection nécessite des modèles particuliers et un jeu de données de taille conséquente pour garantir une puissance de détection suffisante, mais cela en vaut la peine puisque cela permettrait de mettre en évidence des mécanismes moléculaires très intéressants très peu décrits actuellement.

Nous avions initié des études de génétique inverse pour des variants génétiques sur le chromosome X, qui sont restées à un stade préliminaire pour des raisons de calendrier. Le suivi de ces polymorphismes et le phénotypage des individus mâles et femelles aux génotypes d'intérêt permettront de déterminer si nos prédictions s'avèrent justes et aboutissent à la description de nouvelles anomalies génétiques. Dans un deuxième temps, il serait intéressant de revoir et de mettre à jour la liste des variants suivis sur le chromosome X en exploitant les nouveaux variants du Run7 du consortium 1000 génomes bovins dans le cas où des nouveaux variants prédits comme délétères nous semblent pertinents.

Concernant le chromosome Y, notre principal sujet d'intérêt est l'impact de ses variations sur la fertilité mâle. Les délétions du chromosome Y sont souvent associées à l'infertilité, et trois régions de facteurs d'azoospermie, AZFa, AZFb et AZFc (*azoospermia factor*), ont été identifiées sur le bras long du

chromosome Y (revue par Anderson *et al.*, (2019). La région AZFc, qui contient la famille de gènes *DAZ* est complètement absente dans 5 à 10 % des cas d'azoospermie/oligospermie grave. Il s'agit de la cause génétique la plus fréquente de stérilité chez l'homme (Ferlin *et al.*, 2005). De plus, l'étude du chromosome Y ne représente pas un intérêt que chez les mâles. Dans un article de Mcdaneld *et al.* (2012) réalisé sur plusieurs populations de bovins allaitants, les auteurs ont révélé la présence de matériel génétique associé au chromosome Y chez 21 à 29 % des femelles ayant un faible taux de reproduction. Je propose donc de réaliser des recherches de variants structuraux pour identifier des gains/pertes de matériel génétique au niveau du chromosome Y et des translocations de matériel du Y vers les autosomes. L'effet de ces variants pourra également être étudié par une approche de génétique inverse.

2.3. Phylogénétique

Comme nous l'avons détaillé dans le chapitre 2, nous souhaitons monter un projet permettant d'étudier la répartition et les séquences complètes de génomes mitochondriaux appartenant à l'haplogroupe P dans plusieurs pays européens, en exploitant les marqueurs spécifiques des différents haplogroupes que j'ai placés sur la dernière version de puce. Nous prévoyons le séquençage de génomes mitochondriaux de 384 bovins répartis dans les différents pays adhérant au consortium Eurogenomics. A court-terme, il faut donc affiner ce projet, monter les partenariats ainsi que commander des prélèvements pour de nouvelles extractions d'ADN de bovins français porteurs des haplotypes d'intérêt.

Les premières analyses que nous avons menées sur le chromosome Y ouvrent également des perspectives intéressantes sur la phylogénétique des races. Grâce à l'exploitation des données du consortium 1000 génomes bovins, nous pourrions élargir notre étude sur la variabilité du chromosome Y. De plus, en étudiant les séquences supplémentaires de taureaux Holstein présentes dans le run7 du projet 1000 génomes, nous pourrons peut-être retracer plus précisément la transmission des variants au sein de la lignée de R PEARL HA pour étoffer notre hypothèse d'une erreur de parenté.

Pour conclure, si les aléas de la thèse et la durée limitée n'ont pas permis de valoriser et de traiter plus en profondeur certains sujets, nous avons posé des jalons pour de nombreuses études qui pourront être valorisées par la suite. Nous souhaitons prioriser la valorisation des séquences mitochondriales et l'effet sur des caractères de production. La soumission d'un article après accumulation de nouveaux génotypages est envisageable d'ici la fin de l'année 2020. En parallèle, nous pourrons concrétiser le projet d'identification à l'échelle européenne de séquences mitochondriales appartenant aux haplotypes P et Q.

3. Maintien de la diversité génétique du chromosome Y et du génome mitochondrial

Etant donné le peu de variabilité génétique du chromosome Y intra-race, ces dernières ne peuvent transmettre qu'un même chromosome Y à la génération suivante, qu'il soit favorable ou défavorable. L'étude de races comme la Montbéliarde chez laquelle on retrouve les deux haplotypes Y1 et Y2, ou encore la Vosgienne avec deux sous-haplotypes du même haplogroupe, représente une opportunité de comparer l'effet des variants portés par le chromosome Y et, le cas échéant, suggérer des programmes d'introgression d'haplotypes favorables du chromosome Y dans des races ayant un chromosome Y défavorable. Dans le cas d'un effet sur la fertilité mâle, un gain même seulement d'1 % de taux de réussite à l'IA par ce biais, aurait des conséquences très importantes. En effet, en estimant

le prix d'une IA à environ 30 euros et sachant que les IA se chiffrent à plusieurs dizaines de milliers à des millions par an selon la taille des races, l'apport de l'introgression d'un chromosome Y favorable sur la production de semence représenterait un intérêt économique majeur pour la filière.

Le génome mitochondrial n'est pas sujet à un risque de perte de diversité à moyen terme et ce même dans l'hypothèse d'une montée en puissance de la superovulation et du transfert embryonnaire. En effet, le nombre de descendants produits par une femelle ne dépasse que rarement une vingtaine d'individus, ce qui est sans commune mesure avec les milliers à centaines de milliers de descendants qu'ont les mâles d'IA. Toutefois, il pourrait être intéressant de mener des plans de préservation d'haplogroupes très rares. Cela pourrait notamment se faire au travers des différents programmes de reconstitution de l'aurochs ou de dédomestication des bovins mentionnés en introduction, qui pourraient introgresser des mitochondries P et R dans leurs populations pour se rapprocher du matériel génétique des ancêtres sauvages.

4. Prise en compte des chromosomes sexuels et du génome mitochondrial dans les évaluations génomiques

Comme nous l'avons abordé à la fin du premier chapitre, l'absence de la prise en compte des chromosomes sexuels et du génome mitochondrial dans les évaluations trouve son origine dans le dispositif petites-filles ou « *granddaughter design* » utilisé pour réaliser les premières détections de QTL. Aujourd'hui, nous bénéficions de suffisamment de données de mâles et femelles pour pouvoir étudier les informations portées par ce matériel génétique, mais leur intégration dans les évaluations génomiques pose des problèmes spécifiques.

4.1. Adaptation technique des modèles d'évaluations

Intégrer l'information obtenue sur le chromosome X dans l'évaluation génomique fait partie des projets prévus à moyen terme au sein de l'équipe G2B. Cependant, pour pouvoir prendre en compte cette information, il faut repenser le modèle de l'évaluation et l'adapter aux particularités relatives du modèle de transmission du chromosome X, qui est différent de celui des autosomes.

Plusieurs points clés sont à contrôler tout au long de ce processus. Tout d'abord, il s'agit de traiter différemment les marqueurs situés sur la PAR (transmission identique à celle des autosomes) et sur la région spécifique du chromosome X (liée au sexe). Au sein de l'équipe G2B, nous n'avons actuellement pas de marqueurs situés sur la PAR, il faudrait donc en dessiner afin d'étudier cette région particulière. Ensuite, lors de l'imputation des marqueurs, il est nécessaire de contrôler que les génotypes imputés des mâles sont bien « homozygotes » pour toute la région spécifique du chromosome X. Il faut également déterminer comment les génotypes des mâles vont être considérés. En effet, pour les autosomes, les génotypes sont exprimés en nombre de copies de l'allèle majoritaire : 0, 1 ou 2. Pour le chromosome X, les mâles sont hémizygotes et deux hypothèses sont envisageables : soit considérer que c'est équivalent à être homozygote, soit considérer qu'ils ne possèdent qu'une seule copie de l'un des allèles et qu'être hémizygote n'est pas équivalent à être homozygote. Enfin, il faut considérer la transmission des marqueurs du chromosome X de manière différente des autosomes et adapter la matrice de parenté.

Si les marqueurs du chromosome X sont déjà pris en compte dans les évaluations génomiques bovines réalisées dans certains pays, peu d'articles expliquent précisément de quelle façon ces marqueurs sont pris en compte. J'ai choisi de présenter ci-dessous l'étude de Su *et al.* (2014), qui détaille comment

adapter une approche GBLUP en modèle animal pour intégrer les marqueurs du chromosome X à l'évaluation. Cette étude a été réalisée sur 5643 taureaux Holstein testés sur descendance.

Su *et al.* ont étudié la manière d'ajuster la matrice de parenté génomique à l'intégration des marqueurs du chromosome X. La matrice de parenté génomique se calcule à partir d'une matrice de génotypage **M** (cf. figure 34).

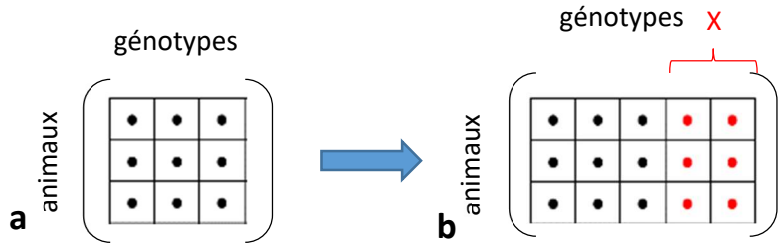


Figure 34 : Représentation schématique de la matrice des génotypes sans (a) et avec (b) les marqueurs du chromosome X, avec • représentant m_{ij}: génotype à un marqueur j pour un animal i

où les éléments m_{ij} de la matrice **M** sont 0, 1 et 2 pour les génotypes A_1A_1,A_1A_2 et A_2A_2 du SNP autosomique. Ces éléments peuvent être centrés selon la fréquence de l'allèle A_2 de chaque SNP de la façon suivante : 0 - $2p_j$,1-$2p_j$ et 2 - $2p_j$, avec p_j la fréquence de l'allèle A_2 au marqueur j.

Cette matrice M permet de calculer la matrice de parenté **G**, mesurant la relation de parenté génomique entre deux individus. Elle est initialement calculée de la façon suivante:

$$\mathbf{G} = \mathbf{MM}'$$

où **M** la matrice des génotypes et **M'** la matrice transposée de **M**

La matrice de parenté, peut être réduite comme présenté ci-après:

$$\text{Soit : } \mathbf{G} = \mathbf{MM}' / \Sigma\, 2p_j(1-p_j)$$

où p_j : Fréquence de l'allèle A_2 au marqueur j

Incorporer les marqueurs présents sur le chromosome X consiste à ajouter des informations de nouveaux marqueurs dans la matrice **M** (cf. figure 34.b). Seulement, si dans le cas des marqueurs autosomiques et des marqueurs sur le chromosome X chez les femelles, qui sont présents en deux copies on observe bien trois génotypes A_1A_1, A_1A_2 et A_2A_2 possibles à chaque marqueur, en ce qui concerne les mâles, seuls deux génotypes sont observables : A_1O ou A_2O. Dans cette situation, soit on considère qu'avoir une seule copie est identique au fait d'être homozygote et donc que les génotypes des mâles, en nombre de copies de l'allèle majoritaire, sont 0 ou 2, soit on considère qu'avoir une seule copie n'est pas identique au fait d'en posséder deux et alors les génotypes sont 0 ou 1.

Dans leur démonstration mathématique, Su *et al.* montrent que les deux hypothèses alternatives considérant les génotypes 0, 1 ou 0, 2 chez les mâles conduisent au même coefficient de relation génomique entre individus. Dans les deux cas, la matrice **G** incluant à la fois les marqueurs du chromosome X et autosomiques peut être calculée en considérant les marqueurs du chromosome X comme les marqueurs autosomiques, mais en divisant l'élément m_{ij} de la matrice **M** par $\sqrt{2}$ si le marqueur j est un marqueur spécifique du chromosome X et que l'individu i est un mâle.

Ensuite, Su *et al.* ont estimé la précision des évaluations génomiques obtenue avec plusieurs modèles GBLUP utilisant des matrices de parenté différentes : (i) calculée sur les autosomes uniquement, (ii) construite en utilisant l'ensemble des marqueurs et en considérant les marqueurs de la région spécifique du chromosome X comme des marqueurs d'autosomes, (iii) construite en utilisant tous les marqueurs et en prenant en compte la transmission liée au sexe des marqueurs du chromosome X, et (iv) en utilisant séparément les marqueurs autosomiques et du chromosome X, associés à des matrices de parenté spécifiques.

Les résultats des modèles génomiques utilisant l'ensemble des marqueurs disponibles ont une fiabilité légèrement supérieure (+0,5 % en moyenne sur l'estimation du coefficient de détermination) à ceux négligeant les marqueurs du chromosome X. Dans une étude similaire réalisée en race Holstein, l'ajout des marqueurs du chromosome X dans les évaluations représentait une amélioration de 1,5 % de la fiabilité (VanRaden *et al.*, 2009). Dans l'article de Su *et al.*, prendre en compte la transmission liée au sexe des marqueurs du chromosome X n'a pas amélioré la prédiction. Cependant, cela peut être dû au fait que les évaluations n'ont été réalisées que sur des taureaux.

Cette étude est très intéressante car elle présente une méthode pour prendre en compte les marqueurs du chromosome X dans la matrice de parenté génomique de façon relativement simple. Cependant, cette méthode est applicable pour une évaluation génomique de type GBLUP suivant un modèle animal. Or, les évaluations génomiques françaises utilisent un modèle SNP qui évoluera prochainement en Single-Step. Ce modèle passe par de nombreuses équations de prédictions permettant d'estimer l'effet d'un SNP, et la matrice de parenté intervient à de nombreuses étapes dans ce système d'équation. La prise en compte de la transmission particulière du chromosome X ne semble donc pas aussi simple à mettre en place dans ce type de modèles et nécessite des études plus approfondies.

Malgré la difficulté pour adapter le modèle, la part de variance expliquée par le chromosome X, estimée respectivement à 1 et 1,7 % par VanRaden *et al.* (2009) et Su *et al.* (2014), est suffisamment importante pour que cela représente un réel intérêt pour l'amélioration des évaluations génomiques.

Concernant le chromosome Y, sa faible variabilité intra-race ne représente pas un intérêt suffisant pour que l'intégration de variants du chromosome Y dans les modèles d'évaluations soit considérée. De plus, prendre en compte les informations portées par le chromosome Y n'aurait de sens et d'intérêt que pour les caractères exprimés chez le mâle puisque les femelles n'ont pas de chromosome Y. Puisque les mâles reçoivent le chromosome Y de leur père, cela reviendrait à inclure un effet LP dans le modèle.

Dans le cas du génome mitochondrial, si les études à venir démontrent l'intérêt de variants mitochondriaux sur les caractères d'intérêt, il sera également nécessaire de réfléchir à faire évoluer le modèle. Contrairement au chromosome X, les individus sont tous homozygotes donc les génotypes seront 0 ou 1, selon l'allèle porté. Au sein de la matrice de parenté, la covariance sera de 1 entre un animal et sa mère et 0 avec son père. La méthode la plus simple consisterait à ajouter un effet lignée mitochondriale dans le modèle d'évolution.

4.2. Pour aller plus loin, explorer les particularités du chromosome X

A ce jour, les études d'association reposent principalement sur des modèles additifs, souvent d'une utilité limitée pour l'analyse du chromosome X, sauf si la population étudiée est entièrement femelle (tous les individus portent 2 copies du chromosome X). Même lorsque tous les individus étudiés sont des femelles, l'utilisation de modèles additifs ne prend pas en compte l'inactivation aléatoire d'un chromosome X dans chaque cellule. Si l'inactivation du chromosome X est aléatoire, les biais d'inactivation sont relativement fréquents (Belmont, 1996; Couldrey *et al.*, 2017), ce qui complexifie l'estimation de l'effet de chaque allèle du variant. Prenons l'exemple théorique d'une femelle avec un biais d'inactivation et porteuse d'un allèle améliorant la production laitière. Si seules 30 % de ses cellules expriment l'allèle favorable alors la prédiction de performance de l'animal sera inférieure à la valeur génétique estimée.

En élevage laitier, un petit nombre de taureaux est utilisé, et ils transmettent leur unique chromosome X à toutes leurs filles. La transmission de ce chromosome X et d'un éventuel biais d'inactivation peuvent donc avoir des conséquences importantes sur la diffusion du progrès génétique. Une étude de Couldrey *et al.* (2017) a étudié l'expression des chromosomes X d'origine paternelle et maternelle au sein des tissus de la glande mammaire à partir de RNA-Seq de 298 vaches en lactation. A l'échelle de la population, les deux chromosomes X sont exprimés de manière égale quelle que soit leur origine parentale dans la glande mammaire en lactation, ce qui correspond à une inactivation aléatoire du chromosome X. Cependant au niveau individuel, l'expression moyenne du chromosome paternel varie de 10 à 90 % selon l'animal (cf. figure 35).

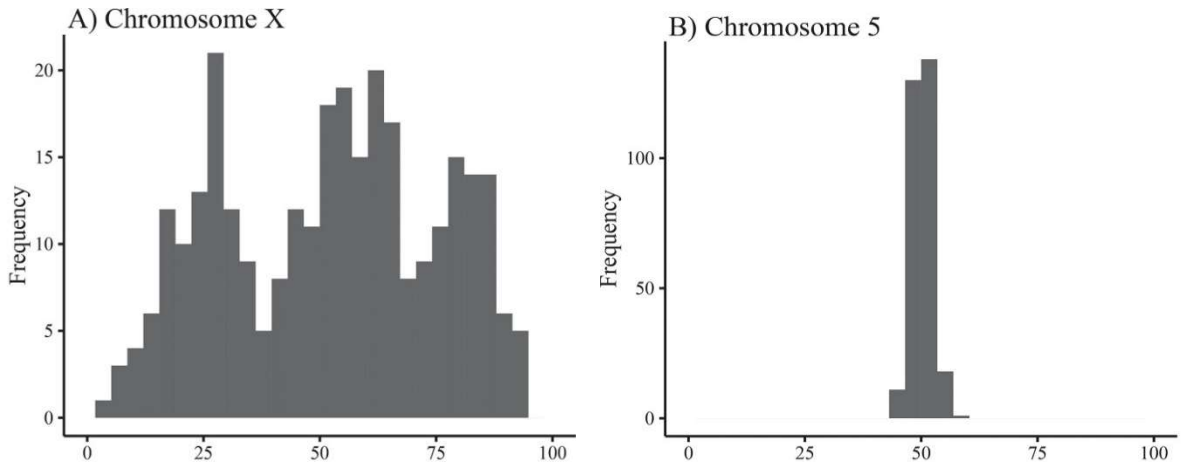


Figure 35 : Fréquence de l'expression de l'allèle paternel chez des animaux sur la région spécifique du chromosome X (A) et le chromosome 5 (B) (Couldrey *et al.*, 2017)

L'étude des mécanismes intervenant dans la régulation de l'expression du chromosome X pourrait se révéler importante pour estimer au mieux l'effet de chaque variant et ainsi améliorer la précision des prédictions de la valeur génétique des femelles.

Il est important de noter que le biais d'expression allélique concerne également les gènes soumis à empreinte parentale sur le reste des chromosomes dont le mécanisme est également lié à la méthylation de l'ADN. Hormis la prise en compte du biais d'inactivation du chromosome X, la gestion de l'expression uniparentale des gènes soumis à empreinte portés par les autosomes pourrait donc également représenter une source d'amélioration des évaluations génomiques. Si à ce jour, chez les bovins, seuls une vingtaine de gènes soumis à empreinte ont été identifiés, on peut présumer que ce

nombre est proche des 200 gènes soumis à empreinte identifiés chez l'humain (100 caractérisés et 100 prédits) (http://geneimprint.com/site/genes-by-species, dernier accès 10/04/2020). Les mécanismes épigénétiques peuvent ainsi représenter une source potentielle de variance phénotypique qui n'est actuellement pas prise en compte par la sélection génomique. Le projet POLYPHEME soumis en octobre 2019 et porté par l'unité INRAE BREED et l'équipe G2B ainsi que Allice, vise à étudier l'épigénome afin de déterminer s'il peut apporter des informations complémentaires à prendre en compte dans les stratégies de sélection. Ce projet pourrait être une occasion unique d'estimer l'effet de biais d'inactivation du chromosome X ainsi que d'identifier des prédicteurs de l'origine parentale de l'allèle principalement exprimé et du ratio d'expression et d'inclure ces informations dans l'évaluation génomique chez le bovin.

En résumé :

- Nous avons présenté et discuter les limites rencontrées pendant ces trois ans.
- Beaucoup d'études prometteuses ont été initiées au cours de cette thèse que nous souhaitons valoriser sous forme d'articles dans un futur proche.
- Dans le cas où des analyses ultérieures mettraient en évidence un réel intérêt à prendre en compte le génome mitochondrial et le chromosome Y, ceux-ci pourraient être pris en compte sous la forme d'un effet fixe intégré au modèle.
- Le modèle français d'évaluation de permet pas actuellement d'intégrer les informations portées par les variants portés par le chromosome X et une évolution du modèle est donc à concevoir.
- L'étude de l'épigénétique pourrait également permettre d'améliorer les estimations des effets des variants.

Conclusion générale

L'objectif principal de cette thèse était de détecter des polymorphismes génétiques au niveau du génome mitochondrial et des chromosomes sexuels et de caractériser leurs effets sur les caractères de production, reproduction et santé. Pour répondre à cette problématique, nous avions trois grands objectifs :

- identifier et caractériser des variants d'intérêt au niveau du génome mitochondrial et des chromosomes X et Y,
- estimer, pour différents caractères d'intérêt, la part de variance génétique expliquée par le génome mitochondrial, et le chromosome Y,
- détecter des polymorphismes délétères portées par ces mêmes régions génomiques et pouvant entraîner des anomalies génétiques.

Au court de ces trois ans de thèse nous avons abordé ces thématiques à travers des approches très variées : séquençage et détection de variants génétiques, analyses phylogénétiques, études généalogiques, estimations de paramètres génétiques, études d'associations entre variants génétiques et phénotypes, génétique inverse, mais aussi analyses épigénétiques, examens cliniques et histologiques dans le cadre de l'étude d'une anomalie génétique. Nous avions été relativement ambitieux sur ce sujet de thèse très vaste. Au terme de ces trois ans, je pense qu'il aurait été avisé de restreindre nos objectifs aux seules études du chromosome Y et du génome mitochondrial, les méthodes utilisées étant relativement similaires et les résultats complémentaires. De surcroît, si la caractérisation de l'anomalie DEH due à une inversion génétique au niveau du chromosome X présentait l'avantage de m'assurer une publication scientifique, le temps que j'ai investi sur ce sujet m'a manqué en fin de thèse. Cela ne m'a pas permis d'aborder d'entreprendre des études d'associations au niveau du chromosome X et d'intégrer des variants génétiques dans les modèles d'évaluations génomiques. Si nous n'avons pas pu traiter ce vaste sujet dans son ensemble, nous avons tout de même pu obtenir une première vue d'ensemble des différents sujets, soulevé des problèmes et identifié les solutions pour y répondre. Plusieurs des études que j'ai initiées pourront, avec l'accumulation à court terme de nouvelles données de génotypage ou séquence, être valorisées par la rédaction d'articles scientifiques intéressants.

Une thèse est à la fois une aventure professionnelle mais aussi personnelle. Malgré l'insatisfaction que j'éprouve de ne pas avoir pu aller plus loin sur certains sujets, la multidisciplinarité des études m'a permis de voir un large panel des approches qu'il est possible de mener en génétique, des approches populationnelles à la caractérisation très précise d'une mutation en passant par des méthodes quantitatives. Cela m'a permis d'acquérir une vision plus large de la complémentarité des approches. Je voudrais d'ailleurs revenir sur l'opportunité précieuse que représente la diversité des sujets traités au sein de l'équipe G2B et l'unité GABI. En mêlant génétique quantitative et moléculaire, génétique des populations et caractérisation des anomalies, il est possible d'exploiter à leur maximum les possibilités permises par chacune des disciplines. Cela s'est principalement fait ressentir à travers l'identification des lignées mitochondriales rares dans la population, qui n'aurait pu se faire sans la mutualisation des outils pour les projets de recherche en génétique des populations et en sélection génomique au sein de notre équipe.

Annexe 1 : Conditions PCR pour l'amplification en long range du génome mitochondrial

Etape	Température (°C)	Durée (minute)
1	94°	2:00
2	94°	0:30
3	60°	0:30
4	72°	6:00
5	94°	0:30
6	59°	0:30
7	72°	6:00
8	94°	0:30
9	58°	0:30
10	72°	6:00
11	72°	10:00
12	15°	END

Annexe 2 : Liste des séquences mitochondriales disponibles dans Genbank utilisées dans l'étude 1.

Identifiant genbank	Haplotype	Localisation	Age (BP)	Référence
GU256940	I			(Wang et al., 2016)
EU177868	I1	Irak		(Achilli *et al.*, 2008)
EU177870	I2	Iran		(Achilli *et al.*, 2008)
DQ124389	P	Corée		(Achilli *et al.*, 2008)
GU985279	P	Derbyshire, Angleterre	6738 BP	(Edwards *et al.*, 2010)
JQ437479	P	Pologne	1500 BP	(Zeyland *et al.*, 2013)
FJ971085	R1	Italie		(Achilli *et al.*, 2009)
FJ971084	R1a	Italie		(Achilli *et al.*, 2009)
FJ971087	R2	Italie		(Achilli *et al.*, 2009)
EU177841	T1	Italie		(Achilli *et al.*, 2008)
DQ124393	T1	Corée		(Shin and Kim, 2005)
EU177842	T1	Italie		(Achilli *et al.*, 2008)
EU177846	T1	Italie		(Achilli *et al.*, 2008)
DQ124396	T2	Corée		(Shin and Kim, 2005)
EU177849	T2	Grèce		(Achilli *et al.*, 2008)
EU177851	T2	Italie		(Achilli *et al.*, 2008)
EU177856	T2	Irak		(Achilli *et al.*, 2008)
EU177861	T2	Italie		(Achilli *et al.*, 2008)
EU177860	T2	Iran		(Achilli *et al.*, 2008)
EU177853	T2	Italie		(Achilli *et al.*, 2008)
AY526085	T3	Corée		(Chung and Ha, 2004)
DQ124372	T4	Corée		Soumission directe
EU177862	T5	Italie		(Achilli *et al.*, 2008)
EU177863	T5	Italie		(Achilli *et al.*, 2008)
EU177864	T5	Irak		(Achilli *et al.*, 2008)
EU177865	T5	Irak		(Achilli *et al.*, 2008)
HQ184039	Q1	Italie		(Bonfiglio *et al.*, 2010)
FJ971082	Q1a	Italie		(Achilli *et al.*, 2009)
HQ184030	Q2	Italie		(Bonfiglio *et al.*, 2010)
HQ184031	Q2	Italie		(Bonfiglio *et al.*, 2010)
Ch22	T	Turquie	7750-7570 BP	(Verdugo *et al.*, 2019)
Gyu2	Q	Arménie	7076 BP	(Verdugo *et al.*, 2019)

Annexe 3 : Amorces utilisées pour le séquençage Sanger de l'étude 2. Les amorces F et R proviennent de la table S3 en annexe de (Achilli et al., 2008), tandis que les amorces « t » ont été dessinées lors de cette étude pour combler les trous de séquençage.

PCR	Nom	Taille de l'amplicon (bp)	5' np	3' np	Taille amorce (nt)	Sequence (5'→3')	T° (°C)
1	1_F	1138	15718	15741	24	CCTAAGACTCAAGGAAGAAACTGC	59,95
	1_R		517	496	22	AACCTAGAGGGCATTCTCACTG	59,77
	1tF		109	128	20	CCCGGAGCATCTATTGTAGC	60,39
2	2_F	1935	312	335	24	CAATTTAGCACTCCAAACAAAGTC	59,27
	2_R		2246	2225	22	TTTCCTTAGATGCACTCCTGTG	59,38
	2tR810		832	810	23	GGTGTAAACTGGATGCTTTGTTT	60,29
	2tR1399		1418	1399	20	AAGCAAGGCACTGAAAATGC	59,47
3	3_F	1843	1807	1828	22	TAGCTGGTTGTCCAGAAAATGA	59,75
	3_R		3649	3630	20	TATTGCTAGAGGCCATGCTG	59,04
	3tF		2516	2536	21	GGGAATGCACAAATAAGACGA	59,95
4	4_F	1861	3339	3360	22	GCCTAGCCTTAACCATGTGAAT	59,53
	4_R		5199	5178	22	TTATGTTGTTTGTGGAGGGAAA	59,36
	4tF		4102	4125	24	AGAGGTTCAAACCCTCTTATTTCT	58,02
5	5_F	1869	4922	4943	22	TATAGCCAATTCCACCACCACT	60,61
	5_R		6790	6768	23	GTGAAGAACAATATCGAGGGAAG	59,15
	5tF		5667	5686	20	TTTGCTCAGCCATTTTACCC	60,07
6	6_F	1901	6525	6546	22	TCGGATTTCTAGGTTTCATCGT	59,97
	6_R		8425	8406	20	GGAGGGTTACAAAGCGATTG	59,57
7	7_F	1849	8109	8132	24	CATATACTCTCCTTGGTGACATGC	59,92
	7_R		9957	9938	20	TCCTATGGGGTCAAATCCAC	59,61
	7tR		9471	9450	22	TGATGGTGATAAATAGGGCTTG	58,97
8	8_F	1881	9599	9620	22	TGGGTCCACCTTCTTAATTGTC	60,22
	8_R		11479	11460	20	AATCATAAGGGCGGTTGCTC	61,33
	8tF		10449	10471	23	GCAGCCCTAGGTCTATCTCTACT	57,45
9	9_F	189	11201	11220	20	ATCGCAGGCTCCATAGTCCT	61,15
	9_R		13090	13069	22	GGAGGAAAGCTAGGTAAGGTTG	58,48
	9tF		12080	12100	21	AAAAATTGGTGCAACTCCAAA	59,48
10	10_F	1876	12750	12771	22	TGGACTAGCATTAGCTGCAACC	61,65
	10_R		14625	14608	18	CCAGGAGGGAACCGAAAT	59,86
	10tF		13489	13508	20	AAGCCTCTTCGCAGGATACA	59,98
11	11_F	1751	14210	14231	22	TTCTGTAGCCATAGCCGTTGTA	59,80
	11_R		15960	15939	22	CCCTTGCGTAGGTAATTCATTC	59,86
	11tF		14937	14956	20	GGAGCAACAGTCATCACCAA	59,68

Annexe 4 : Echantillons modernes disponibles dans les bases publiques utilisés pour la réalisation de l'arbre et du réseau phylogénétique de l'étude 2.

Identifiant genbank	Haplotype	Race ou localisation	Référence
GU256940	I		(Wang et al., 2016)
EU177868	I1	Irak	(Achilli *et al.*, 2008)
EU177870	I2	Iran	(Achilli *et al.*, 2008)
DQ124389	P	Corée	(Achilli *et al.*, 2008)
FJ971087	R2	Italie	(Achilli *et al.*, 2009)
FJ971085	R1	Chianina	(Achilli *et al.*, 2009)
FJ971084	R1a	Italie	(Achilli *et al.*, 2009)
HQ184036	Q1	Brunes des Alpes	(Bonfiglio *et al.*, 2010)
HQ184037	Q1	Brunes des Alpes	(Bonfiglio *et al.*, 2010)
HQ184039	Q1	Chianina	(Bonfiglio *et al.*, 2010)
HQ184030	Q2	Chianina	(Bonfiglio *et al.*, 2010)
HQ184031	Q2	Chianina	(Bonfiglio *et al.*, 2010)
HQ184032	Q2	Chianina	(Bonfiglio *et al.*, 2010)
FJ971080	Q2	Romagnola	(Achilli *et al.*, 2009)
HQ184033	Q2	Romagnola	(Bonfiglio *et al.*, 2010)
FJ971082	Q1a	Pie rouge Italienne	(Achilli *et al.*, 2009)
HQ184034	Q1a	Romagnola	(Bonfiglio *et al.*, 2010)
EU177841	T1	Chianina	(Achilli *et al.*, 2008)
BAZM267	T1	Bazadaise	Cette étude
TARM101	T1	Tarentaise	Cette étude
MONM160	T2	Montbéliarde	Cette étude
LIMM298	T2	Limousine	Cette étude
CHAm430	T2	Charolaise	Cette étude
DQ124396	T2	Beef cattle	DQ124396
CHAM360	T3	Charolaise	Cette étude
AY526085	T3	Corée	(Anderson *et al.*, 1982)
DQ124372	T4	Corée	(Shin and Kim, 2005)
EU177862	T5	Valdostana	(Achilli *et al.*, 2008)
MONm423	T5	Montbéliarde	Cette étude

Annexe 5 : Description des échantillons non-T.

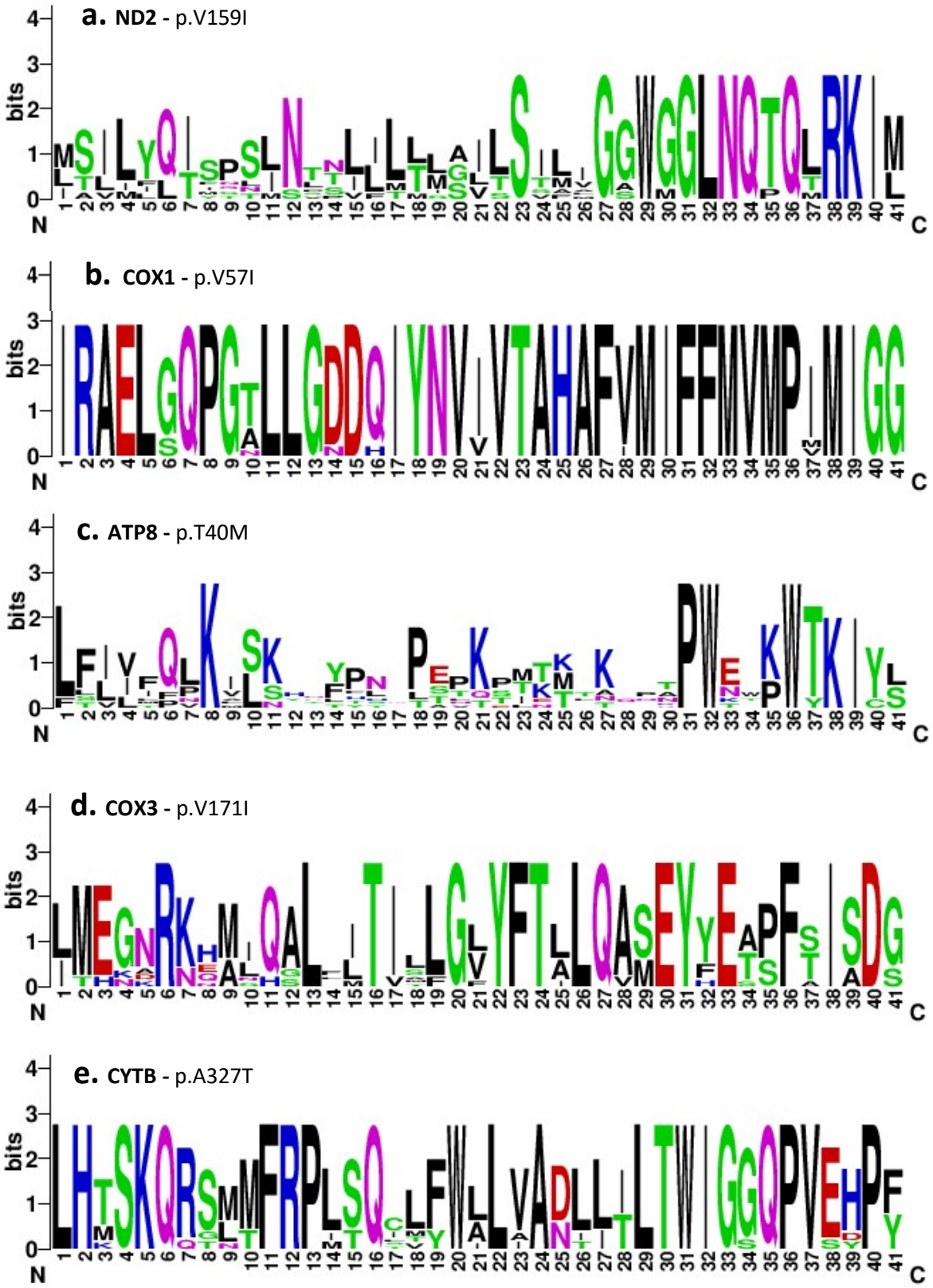
Haplotype	Groupe	Race	Identifiants des échantillons
mauvaise qualité	éliminée	Holstein	h22
I		Holstein	HOL2 HOL8
P	groupe A	Holstein	HOL1 HOL11 HOL12 HOL17 HOL18
P	groupe B	Abondance	ABO21 ABO22 ABO23
P	groupe B	Holstein	HOL10
P	groupe B	Montbéliarde	MON20
P	groupe C	Blonde d'Aquitaine	BAQ4
P	groupe C	Holstein	HOL13 HOL14 HOL15 HOL25 HOL26
P	groupe C	Montbéliarde	MON3 MON7
P		Abondance	ABO27
P		Holstein	HOL19
P		Holstein	HOL6
P		Montbéliarde	MON9
Q	groupe D	Montbéliarde	MON5 m1 m2 m5 m6 m9 m10 m13 m14 m16
Q	groupe E	Montbéliarde	MON28 m3 m4 m7 m8
Q	groupe E	Abondance	a1
Q	groupe E	Holstein	h2 h3 h6 h11 h14 h15 h16 h23 h25
Q		Abondance	ABO16
Q		Holstein	h1
Q		Holstein	h4
Q		Holstein	h7
Q		Holstein	h13
Q		Holstein	h17
Q		Holstein	h20
Q		Holstein	HOL24
Q		Holstein	h21
Q		Montbéliarde	m11
Q		Montbéliarde	m12
T		Blonde d'Aquitaine	b1
T		Holstein	h5
T		Holstein	h8
T		Holstein	h9
T		Holstein	h10
T		Holstein	h12
T		Holstein	h18
T		Holstein	h19
T		Montbéliarde	m15

Annexe 6 : Synthèse des héritabilités estimées dans la littérature sur des caractères de fertilité mâle.

Référence	Pays	Race	Nombre d'animaux	Age moyen (mois)	Volume	Concentration	Pourcentage de motiles	Pourcentage de vivants	Nombre de spermatozoïdes	Anormaux/ Normaux
(Berry *et al.*, 2011)	Irlande	Multiples								
(Corbet *et al.*, 2013)	Australie	« Tropical composite »	2080	18			0,24 (0,07)			0,25 (0,09)
(Corbet *et al.*, 2013)	Australie	Brahman	1398	18			0,13 (0,05)			0,20 (0,06)
(Druet *et al.*, 2009)	France	Holstein	515	12-18	0,22	0,19		0473		
(Ducrocq et Humblot, 1995)	France	Normande	1957	12-15	0,65 (0,09)	0,37 (0,09)	0,23 (0,08)			0,19 (0,07)
(Garmyn *et al.*, 2011)	USA	Angus					0,05 (0,03)			0,25 (0,07)
(Gipson *et al.*, 1985)	USA	Angus, Hereford, Simmental		12		0,2 (0,13)	0,11 (0,12)		0,19 (0,14)	
(Gredler *et al.*, 2007)	Autriche	Simmental		41	0,18 (0,02)	0,14 (0,04)	0,04 (0,01)	0,1 (0,05)	0.22, (0,02)	
(Kealey *et al.*, 2006)	USA	Hereford		37	0,09 (0,08)	0,16 (0,08)	0,22 (0,09)	0,22 (0,09)		0,35 (0,10)
(Knights *et al.*, 1984)	USA	Hereford		12-14		0,13 (0,06)	0,13 (0,06)		0,24 (0,05)	
(Mathevon *et al.*, 1998a)	Canada	Holstein			0,24	0,52	0,31		0,38	
(Mathevon *et al.*, 1998a)	Canada				0,44	0,36	0,01		0,54	
(Mathevon *et al.*, 1998b)	France	Montbéliarde			0,08	0,32			0,15	
(Olsen *et al.*, 2020)	Norvège	Rouge Norvégienne			0,14 (0,02)					
(Suchocki et Szyda, 2015)	Pologne	Holstein			0,26 (0,062)	0,56 (0,024)	0,31 (0,064)	0,26 (0,054)	0,27 (0,060)	
(Yin *et al.*, 2019)	Chine	Holstein			0,15 (0,03)	0,22 (0,04)		0,12 (0,03)	0,16 (0,03)	

Annexe 7 : Etude de la conservation intra-vertébrés des acides aminés affectés par des polymorphismes non-sens dans les protéines codées par le génome mitochondrial.

Les acides aminés d'intérêt sont placés à la position 21 sur les graphiques. La représentation logo a été réalisée à partir des séquences protéiques de neuf espèces représentant les principaux clades des vertébrés et des mammifères sur le site internet : https://weblogo.berkeley.edu/logo.cgi



Liste des références bibliographiques :

Achilli, A., S. Bonfiglio, A. Olivieri, A. Malusà, M. Pala, B.H. Kashani, U.A. Perego, P. Ajmone-Marsan, L. Liotta, O. Semino, H.J. Bandelt, L. Ferretti, and A. Torroni. 2009. The multifaceted origin of taurine cattle reflected by the mitochondrial genome. PLoS One 4:e5753. doi:10.1371/journal.pone.0005753.

Achilli, A., A. Olivieri, M. Pellecchia, C. Uboldi, L. Colli, N. Al-Zahery, M. Accetturo, M. Pala, B.H. Kashani, U.A. Perego, V. Battaglia, S. Fornarino, J. Kalamati, M. Houshmand, R. Negrini, O. Semino, M. Richards, V. Macaulay, L. Ferretti, H.-J. Bandelt, P. Ajmone-marsan, and A. Torroni. 2008. Mitochondrial genomes of extinct aurochs survive in domestic cattle. Curr. Biol. 18:157–158.

Ajmone-Marsan, P., J.A. Lenstra, J. Fernando Garcia, and The Globaldiv Consortium. 2010. On the origin of cattle: how aurochs became domestic and colonized the world. Evol. Anthropol. 19:148–157. doi:10.1002/evan20267.

Allice. 2017. L'insémination Animale. De La Paillette à l'ère Du Génome, 70 Ans d'aventure Humaine.

Amaral, A., B. Lourenço, M. Marques, and J. Ramalho-Santos. 2013. Mitochondria functionality and sperm quality. Reproduction 146. doi:10.1530/REP-13-0178.

Anderson, K., M. Cañadas-Garre, R. Chambers, A.P. Maxwell, and A.J. McKnight. 2019. The Challenges of Chromosome Y Analysis and the Implications for Chronic Kidney Disease. Front. Genet. 10:1–9. doi:10.3389/fgene.2019.00781.

Anderson, S., M.H.L. de Bruijn, A.R. Coulson, I.C. Eperon, F. Sanger, and I.G. Young. 1982. Complete sequence of bovine mitochondrial DNA conserved features of the mammalian mitochondrial genome. J. Mol. Biol. 156:683–717. doi:10.1016/0022-2836(82)90137-1.

Anderung, C., A. Bouwman, P. Persson, J.M. Carretero, A.I. Ortega, R. Elburg, C. Smith, J.L. Arsuaga, H. Ellegren, and A. Götherström. 2005. Prehistoric contacts over the Straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle. Proc. Natl. Acad. Sci. U. S. A. 102:8431–8435. doi:10.1073/pnas.0503396102.

Ankel-Simons, F., and J.M. Cummins. 1996. Misconceptions about mitochondria and mammalian fertilization: Implications for theories on human evolution. Proc. Natl. Acad. Sci. U. S. A. 93:13859–13863. doi:10.1073/pnas.93.24.13859.

Bachtrog, D. 2014. Y chromosome evolution: emerging insights into processes of Y chromosome degeneration. Nat. Rev. Genet. 14:113–124. doi:10.1038/nrg3366.Y.

Bachtrog, D., M. Kirkpatrick, J.E. Mank, S.F. McDaniel, J.C. Pires, W.R. Rice, and N. Valenzuela. 2011. Are all sex chromosomes created equal?. Trends Genet. 27:350–357. doi:10.1016/j.tig.2011.05.005.

Banos, G., G.R. Wiggans, and R.L. Powell. 2001. Impact of paternity errors in cow identification on genetic evaluations and international comparisons. J. Dairy Sci. 84:2523–2529. doi:10.3168/jds.S0022-0302(01)74703-0.

Baranowska, I., K.H. Jäderlund, I. Nennesmo, E. Holmqvist, N. Heidrich, N.G. Larsson, G. Andersson, E.G.H. Wagner, Å. Hedhammar, R. Wibom, and L. Andersson. 2009. Sensory ataxic neuropathy in golden retriever dogs is caused by a deletion in the mitochondrial tRNATyr gene. PLoS Genet. 5:e1000499. doi:10.1371/journal.pgen.1000499.

Barbotte, L., L. Genestout, S. Fritz, C. Chantry-Darmon, and D. Boichard. 2012. Assignation de parenté

par marqueurs SNP chez les bovins français. 19. Rencontres Recheche Ruminants 19:92.

Baur, A., S. Fritz, J. Promp, O. Bulot, D. Boichard, V. Ducrocq, and P. Croiseau. 2014. Implementation of the French Official Genomic Evaluation in Brown Swiss Dairy Cattle. 10th World Congr. Genet. Appl. to Livest. Prod. 4–6.

Beinat, M. 2013. Caractérisation génétique des atteintes hépatiques mitochondriales. Thèse, Université paris 5 René Descartes,.

Beja-Pereira, A., D. Caramelli, C. Lalueza-Fox, C. Vernesi, N. Ferrand, A. Casoli, F. Goyache, L.J. Royo, S. Conti, M. Lari, A. Martini, L. Ouragh, A. Magid, A. Atash, A. Zsolnai, P. Boscato, C. Triantaphylidis, K. Ploumi, L. Sineo, F. Mallegni, P. Taberlet, G. Erhardt, L. Sampietro, J. Bertranpetit, G. Barbujani, G. Luikart, and G. Bertorelle. 2006. The origin of European cattle: Evidence from modern and ancient DNA. Proc. Natl. Acad. Sci. U. S. A. 103:8113–8118. doi:10.1073/pnas.0509210103.

Bell, B.R., B.T. McDaniel, and O.W. Robison. 1985. Effects of Cytoplasmic Inheritance on Production Traits of Dairy Cattle. J. Dairy Sci. 68:2038–2051. doi:10.3168/jds.S0022-0302(85)81066-3.

Bellott, D.W., J.F. Hughes, H. Skaletsky, L.G. Brown, T. Pyntikova, R.S. Fulton, S. Dugan, Y. Ding, D. Morton, Z. Khan, L. Lewis, C. Buhay, W. Quiaoyan, J. Watt, M. Holder, S. Lee, L. Nazareth, J. Alföldi, S. Rozen, D.M. Muzny, W.C. Warren, R.A. Gibbs, R.K. Wilson, and D.C. Page. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature 508:494–499. doi:10.1038/nature13206.Mammalian.

Belmont, J.W. 1996. Genetic control of X inactivation and processes leading to X-inactivation skewing. Am. J. Hum. Genet. 58:1101–1108.

Benkhalifa, M., Y.J. Ferreira, H. Chahine, N. Louanjli, P. Miron, P. Merviel, and H. Copin. 2014. Mitochondria: Participation to infertility as source of energy and cause of senescence. Int. J. Biochem. Cell Biol. 55:60–64. doi:10.1016/j.biocel.2014.08.011.

Bernard, D., and E. Baudemont. 2016. Les Vaches Ont Une Histoire. Naissance Des Races Bovines.

Bérodier, M. 2019. Utilisation des données de génotypage en ferme pour une gestion optimisée et durable de l'élevage laitier. Thèse, Université Paris-Saclay,.

Berry, D.P., R.D. Evans, and S. Mc Parland. 2011. Evaluation of bull fertility in dairy and beef cattle using cow field data. Theriogenology 75:172–181. doi:10.1016/j.theriogenology.2010.08.002.

Berry, D.P., E. Wall, and J.E. Pryce. 2014. Genetics and genomics of reproductive performance in dairy and beef cattle. Animal 8:105–121. doi:10.1017/S1751731114000743.

Boettcher, P.J., and J.P. Gibson. 1997. Estimation of Variance of Maternal Lineage Effects among Canadian Holsteins. J. Dairy Sci. 80:2167–2176. doi:10.3168/jds.S0022-0302(97)76164-2.

Boettcher, P.J., D.W.B. Steverink, D.C. Beitz, A.E. Freeman, and B.T. Mcdaniel. 1996. Multiple Herd Evaluation of the Effects of Maternal Lineage on Yield Traits of Holstein Cattle. J. Dairy Sci. 79:655–662. doi:10.3168/jds.S0022-0302(96)76411-1.

Boichard, D. 2002. PEDIG : A FORTRAN PACKAGE FOR PEDIGREE ANALYSIS SUITED FOR LARGE POPULATIONS. Pages 28–29 in 7th World Congress on Genetics Applied to Livestock Production, 23 août 2002, Montpellier, France.

Boichard, D., and B. Bonaïti. 1987. Genetic parameters for first lactation dairy traits in Friesian, Montbéliarde and Normande breeds. Genet. Sel. Evol. 19:337. doi:10.1186/1297-9686-19-3-337.

Boichard, D., M. Boussaha, A. Capitan, D. Rocha, C. Hozé, M. Sanchez, T. Tribout, R. Letaief, P. Croiseau, W. Li, C. Harland, C. Charlier, M. Lund, G. Sahana, M. Georges, S. Barbier, W. Coppieters, S. Fritz, and B. Guldbrandtsen. 2018. Experience from large scale use of the EuroGenomics custom SNP chip in cattle. Pages 1–6 in 11th World Congress on Genetics Applied to Livestock Production.

Boichard, D., C. Grohs, C. Danchin-Burge, and A. Capitan. 2016a. Les anomalies génétiques: Définition, origine, transmission et évolution, mode d'action. INRA Prod. Anim. 29:297–305.

Boichard, D., C. Grohs, P. Michot, C. Danchin-Burge, A. Capitan, L. Genestout, S. Barbier, and S. Fritz. 2016b. Prise en compte des anomalies génétiques en sélection: Le cas des bovins. INRA Prod. Anim. 29:351–358.

Boichard, D., F. Guillaume, A. Baur, P. Croiseau, M.N. Rossignol, M.Y. Boscher, T. Druet, L. Genestout, J.J. Colleau, L. Journaux, V. Ducrocq, and S. Fritz. 2012. Genomic selection in French dairy cattle. Anim. Prod. Sci. 52:115–120. doi:10.1071/AN11119.

Bökönyi, S. 1976. Development of early stock rearing in the Near East. Nature 264:19–23. doi:10.1038/264019a0.

Bollongino, R., J. Burger, A. Powell, M. Mashkour, J.D. Vigne, and M.G. Thomas. 2012. Modern taurine cattle descended from small number of near-eastern founders. Mol. Biol. Evol. 29:2101–2104. doi:10.1093/molbev/mss092.

Bollongino, R., J. Elsner, J.D. Vigne, and J. Burger. 2008. Y-SNPs do not indicate hybridisation between European Aurochs and domestic cattle. PLoS One 3:e3418. doi:10.1371/journal.pone.0003418.

Bonfiglio, S., A. Achilli, A. Olivieri, R. Negrini, L. Colli, L. Liotta, P. Ajmone-Marsan, A. Torroni, and L. Ferretti. 2010. The enigmatic origin of bovine mtDNA haplogroup R: Sporadic interbreeding or an independent event of Bos primigenius domestication in Italy?. PLoS One 5:e15760. doi:10.1371/journal.pone.0015760.

Bonfiglio, S., A. De Gaetano, K. Tesfaye, V. Grugni, O. Semino, and L. Ferretti. 2012. A novel USP9Y polymorphism allowing a rapid and unambiguous classification of Bos taurus y chromosomes into haplogroups. Anim. Genet. 43:611–613. doi:10.1111/j.1365-2052.2012.02328.x.

Bourneuf, E., P. Otz, H. Pausch, V. Jagannathan, P. Michot, C. Grohs, G. Piton, S. Ammermüller, M.C. Deloche, S. Fritz, H. Leclerc, C. Péchoux, A. Boukadiri, C. Hozé, R. Saintilan, F. Créchet, M. Mosca, D. Segelke, F. Guillaume, S. Bouet, A. Baur, A. Vasilescu, L. Genestout, A. Thomas, A. Allais-Bonnet, D. Rocha, M.A. Colle, C. Klopp, D. Esquerré, C. Wurmser, K. Flisikowski, H. Schwarzenbacher, J. Burgstaller, M. Brügmann, E. Dietschi, N. Rudolph, M. Freick, S. Barbey, G. Fayolle, C. Danchin-Burge, L. Schibler, B. Bed'hom, B.J. Hayes, H.D. Daetwyler, R. Fries, D. Boichard, D. Pin, C. Drögemüller, and A. Capitan. 2017. Rapid Discovery of de Novo Deleterious Mutations in Cattle Enhances the Value of Livestock as Model Species. Sci. Rep. 7:1–19. doi:10.1038/s41598-017-11523-3.

Bouwman, A.C., H.D. Daetwyler, A.J. Chamberlain, C.H. Ponce, M. Sargolzaei, F.S. Schenkel, G. Sahana, A. Govignon-Gion, S. Boitard, M. Dolezal, H. Pausch, R.F. Brøndum, P.J. Bowman, B. Thomsen, B. Guldbrandtsen, M.S. Lund, B. Servin, D.J. Garrick, J. Reecy, J. Vilkki, A. Bagnato, M. Wang, J.L. Hoff, R.D. Schnabel, J.F. Taylor, A.A.E. Vinkhuyzen, F. Panitz, C. Bendixen, L.E. Holm, B. Gredler, C. Hozé, M. Boussaha, M.P. Sanchez, D. Rocha, A. Capitan, T. Tribout, A. Barbat, P. Croiseau, C. Drögemüller, V. Jagannathan, C. Vander Jagt, J.J. Crowley, A. Bieber, D.C. Purfield, D.P. Berry, R. Emmerling, K.U. Götz, M. Frischknecht, I. Russ, J. Sölkner, C.P. Van Tassell, R. Fries, P. Stothard, R.F. Veerkamp, D. Boichard, M.E. Goddard, and B.J. Hayes. 2018. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nat. Genet. 50:362–367. doi:10.1038/s41588-018-0056-5.

Bradley, D.G., D.E. Machugh, P. Cunningham, and R.T. Loftus. 1996. Mitochondrial diversity and the origins of African and European cattle. Proc. Natl. Acad. Sci. U. S. A. 93:5131–5135. doi:10.1073/pnas.93.10.5131.

Bro-Jørgensen, M.H., C. Carøe, F.G. Vieira, S. Nestor, A. Hallström, K.M. Gregersen, V. Etting, M.T.P. Gilbert, and M.H.S. Sinding. 2018. Ancient DNA analysis of Scandinavian medieval drinking horns and the horn of the last aurochs bull. J. Archaeol. Sci. 99:47–54. doi:10.1016/j.jas.2018.09.001.

Brockdorff, N., and B.M. Turner. 2015. Dosage compensation in mammals. Cold Spring Harb. Perspect. Biol. 7. doi:10.1101/cshperspect.a019406.

Broekhuijse, M.L.W.J., E. Šostarić, H. Feitsma, and B.M. Gadella. 2011. Additional value of computer assisted semen analysis (CASA) compared to conventional motility assessments in pig artificial insemination. Theriogenology 76:1473–1486. doi:10.1016/j.theriogenology.2011.05.040.

Brown, D.R., C.M. Koehler, G.L. Lindberg, A.E. Freeman, J.E. Mayfield, A.M. Myers, M.M. Schutz, and D.C. Beitz. 1989. Molecular analysis of cytoplasmic genetic variation in Holstein cows.. J. Anim. Sci. 67:1926–1932. doi:10.2527/jas1989.6781926x.

Bruford, M.W., D.G. Bradley, and G. Luikart. 2003. DNA markers reveal the complexity of livestock domestication. Nat. Rev. Genet. 4:900–910. doi:10.1038/nrg1203.

Cabannes, C. 2008. Comparaison des méthodes d'évaluation de la qualite de la semence dans les espèces bovine, canine et humaine. Thèse vétérinaire, Ecole Nationale Vétérinaire de Toulouse,.

Cai, D., Y. Sun, Z. Tang, S. Hu, W. Li, X. Zhao, H. Xiang, and H. Zhou. 2014. The origins of Chinese domestic cattle as revealed by ancient DNA analysis. J. Archaeol. Sci. 41:423–434. doi:10.1016/j.jas.2013.09.003.

Calvo, S.E., K.R. Clauser, and V.K. Mootha. 2016. MitoCarta2.0: An updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res. 44:D1251–D1257. doi:10.1093/nar/gkv1003.

de Camargo, G.M.F., L.R. Porto-Neto, M.J. Kelly, R.J. Bunch, S.M. McWilliam, H. Tonhati, S.A. Lehnert, M.R.S. Fortes, and S.S. Moore. 2015. Non-synonymous mutations mapped to chromosome X associated with andrological and growth traits in beef cattle. BMC Genomics 16:384. doi:10.1186/s12864-015-1595-0.

Cañón, J., P. Alexandrino, I. Bessa, C. Carleos, Y. Carretero, S. Dunner, N. Ferran, D. Garcia, J. Jordana, D. Laloë, A. Pereira, A. Sanchez, and K. Moazami-Goudarzi. 2001. Genetic diversity measures of local European beef cattle breeds for conservation purposes. Genet. Sel. Evol. 33:311–332. doi:10.1186/1297-9686-33-3-311.

Chang, C.C., C.C. Chow, L.C.A.M. Tellier, S. Vattikuti, S.M. Purcell, and J.J. Lee. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience 4:1–16. doi:10.1186/s13742-015-0047-8.

Charlesworth, B. 1996. The evolution of chromosomal sex determination and dosage compensation. Curr. Biol. 6:149–162. doi:10.1016/S0960-9822(02)00448-7.

Charlesworth, D. 2017. Evolution of recombination rates between sex chromosomes. Philos. Trans. R. Soc. B Biol. Sci. 372. doi:10.1098/rstb.2016.0456.

Charlier, C., W. Coppieters, F. Rollin, D. Desmecht, J.S. Agerholm, N. Cambisano, E. Carta, S. Dardano, M. Dive, C. Fasquelle, J.C. Frennet, R. Hanset, X. Hubin, C. Jorgensen, L. Karim, M. Kent, K. Harvey, B.R. Pearce, P. Simon, N. Tama, H. Nie, S. Vandeputte, S. Lien, M. Longeri, M. Fredholm,

R.J. Harvey, and M. Georges. 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. Nat. Genet. 40:449–454. doi:10.1038/ng.96.

Charlier, C., W. Li, C. Harland, M. Littlejohn, W. Coppieters, F. Creagh, S. Davis, T. Druet, P. Faux, F. Guillaume, L. Karim, M. Keehan, N.K. Kadri, N. Tamma, R. Spelman, and M. Georges. 2016. NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. Genome Res. 26:1333–1341. doi:10.1101/gr.207076.116.

Chesnais, J.P., T.A. Cooper, G.R. Wiggans, M. Sargolzaei, J.E. Pryce, and F. Miglior. 2016. Using genomics to enhance selection of novel traits in North American dairy cattle. J. Dairy Sci. 99:2413–2427. doi:10.3168/jds.2015-9970.

Chinnery, P.F., and G. Hudson. 2013. Mitochondrial genetics. Br. Med. Bull. 106:135–159. doi:10.1093/bmb/ldt017.

Chinnery, P.F., M.A. Johnson, T.M. Wardell, R. Singh-Kler, C. Hayes, D.T. Brown, R.W. Taylor, L.A. Bindoff, and D.M. Turnbull. 2000. The epidemiology of pathogenic mitochondrial DNA mutations. Ann. Neurol. 48:188–193. doi:10.1002/1531-8249(200008)48:2<188::AID-ANA8>3.0.CO;2-P.

Christensen, O.F., and M.S. Lund. 2010. Genomic relationship matrix when some animals are not genotyped Genomic prediction models. Genet. Sel. Evol. 42:1–8.

Chung, H.Y., and J.M. Ha. 2004. Haplotype analysis of mitochondrial DNA in Korean native cattle. GenBank AY526085.

Cingolani, P., A. Platts, L.L. Wang, M. Coon, T. Nguyen, L. Wang, S.J. Land, X. Lu, and D.M. Ruden. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 6:80–92. doi:10.4161/fly.19695.

Clarke, A., D.I.M. Phillips, R. Brown, and P.S. Harper. 1987. Clinical aspects of X-linked hypohidrotic ectodermal dysplasia. Arch. Dis. Child. 62:989–996. doi:10.1136/adc.62.10.989.

Colaco, S., and D. Modi. 2018. Genetics of the human Y chromosome and its association with male infertility. Reprod. Biol. Endocrinol. 16:1–24. doi:10.1186/s12958-018-0330-5.

Cole, J.B. 2015. A simple strategy for managing many recessive disorders in a dairy cattle breeding program. Genet. Sel. Evol. 47:1–13. doi:10.1186/s12711-015-0174-9.

Collectif des Races des Massifs. Bovin Tarentaise. Accessed February 17, 2020. https://www.races-montagnes.com/fr/races/tarentaise.php.

Corbet, N.J., B.M. Burns, D.J. Johnston, M.L. Wolcott, D.H. Corbet, B.K. Venus, Y. Li, M.R. McGowan, and R.G. Holroyd. 2013. Male traits and herd reproductive capability in tropical beef cattle. 2. Genetic parameters of bull traits. Anim. Prod. Sci. 53:101–113. doi:10.1071/AN12163.

Cottle, D.J., and J.H.J. Van Der Werf. 2017. Optimising the proportion of selection candidates measured for feed intake for a beef cattle breeding objective that includes methane emissions. J. Anim. Sci. 95:1030–1041. doi:10.2527/jas2016.1177.

Couldrey, C., T. Johnson, T. Lopdell, I.L. Zhang, M.D. Littlejohn, M. Keehan, R.G. Sherlock, K. Tiplady, A. Scott, S.R. Davis, and R.J. Spelman. 2017. Bovine mammary gland X chromosome inactivation. J. Dairy Sci. 100:5491–5500. doi:10.3168/jds.2016-12490.

Croué, I. 2017. Évaluation génétique et génomique de nouveaux caractères en bovins laitiers. Thèse, Université Paris-Saclay,.

Croué, I., F. Fikse, K. Johansson, E. Carlén, G. Thomas, H. Leclerc, and V. Ducrocq. 2017. Genetic evaluation of claw health traits accounting for potential preselection of cows to be trimmed. J. Dairy Sci. 100:8197–8204. doi:10.3168/jds.2017-13002.

Curik, I., V. Brajkovic, D. Novosel, S. Ristov, M. Spehar, M. Cacic, M. Ferencakovic, D. Salamon, N. Raguz, and V. Cubric-Curik. 2017. Impact and utilisation of the mitogenome in livestock breeding & genetics. Page 23 in 25th International Symposium Animal Science Days, Brandlucken, Austria.

Daetwyler, H.D., A. Capitan, H. Pausch, P. Stothard, R. Van Binsbergen, R.F. Brøndum, X. Liao, A. Djari, S.C. Rodriguez, C. Grohs, D. Esquerré, O. Bouchez, M.N. Rossignol, C. Klopp, D. Rocha, S. Fritz, A. Eggen, P.J. Bowman, D. Coote, A.J. Chamberlain, C. Anderson, C.P. Vantassell, I. Hulsegge, M.E. Goddard, B. Guldbrandtsen, M.S. Lund, R.F. Veerkamp, D.A. Boichard, R. Fries, and B.J. Hayes. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat. Genet. 46:858–865. doi:10.1038/ng.3034.

Danchin-Burge, C. 2015. Bilan de Variabilité Génétique Pour Les Filières Caprines, Ovines et Bovines. Accessed. http://idele.fr/no_cache/recherche/publication/idelesolr/recommends/bilan-de-variabilite-genetique-pour-les-filieres-caprines-ovines-et-bovines.html.

Danchin-Burge, C., G. Leroy, M. Brochard, S. Moureaux, and E. Verrier. 2012. Evolution of the genetic variability of eight French dairy cattle breeds assessed by pedigree analysis. J. Anim. Breed. Genet. 129:206–217. doi:10.1111/j.1439-0388.2011.00967.x.

Darr, C.R., G.A. Cortopassi, S. Datta, D.D. Varner, and S.A. Meyers. 2016. Mitochondrial oxygen consumption is a unique indicator of stallion spermatozoal health and varies with cryopreservation media. Theriogenology 86:1382–1392. doi:10.1016/j.theriogenology.2016.04.082.

Darwin, C. 1859. De l'origine Des Espèces.

Das, P.J., B.P. Chowdhary, and T. Raudsepp. 2009. Characterization of the bovine pseudoautosomal region and comparison with sheep, goat, and other mammalian pseudoautosomal regions. Cytogenet. Genome Res. 126:139–147. doi:10.1159/000245913.

Davrou, S. 2008. Etat des lieux de la race aurochs-reconstitué (code race 30) sur la période 2005-2008 : présentation des caractéristiques de la race et typologie de l'élevage. Ecole nationale vétérinaire de Nantes,.

Decker, J.E., S.D. McKay, M.M. Rolf, J.W. Kim, A. Molina Alcalá, T.S. Sonstegard, O. Hanotte, A. Götherström, C.M. Seabury, L. Praharani, M.E. Babar, L. Correia de Almeida Regitano, M.A. Yildiz, M.P. Heaton, W.S. Liu, C.Z. Lei, J.M. Reecy, M. Saif-Ur-Rehman, R.D. Schnabel, and J.F. Taylor. 2014. Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. PLoS Genet. 10:e1004254. doi:10.1371/journal.pgen.1004254.

Delanoue, E., and C. Roguet. 2015. Acceptabilité sociale de l'élevage en France: Recensement et analyse des principales controverses à partir des regards croisés de différents acteurs. Organ. Sci. 28:39–50.

DeLuca, S.Z., and P.H. O'Farrell. 2012. Barriers to Male Transmission of Mitochondrial DNA in Sperm Development. Dev. Cell 22:660–668. doi:10.1016/j.devcel.2011.12.021.

Deng, W., B.S. Maust, D.C. Nickle, G.H. Learn, Y. Liu, L. Heath, S.L.K. Pond, and J.I. Mullins. 2010. DIVEIN: A web server to analyze phylogenies, sequence divergence, diversity, and informative sites. Biotechniques 48:405–408. doi:10.2144/000113370.

Denis, B. 2010. Races Bovines. C.E. Pollux, ed.

Diamond, J. 2002. Evolution, consequences and future of plant and animal domestication 418.

DiMauro, S. 2004. Mitochondrial diseases. Biochim. Biophys. Acta - Bioenerg. 1658:80–88. doi:10.1016/j.bbabio.2004.03.014.

Doublet, A.-C., P. Croiseau, S. Fritz, A. Michenet, C. Hozé, C. Danchin-Burge, D. Laloë, and G. Restoux. 2019. The impact of genomic selection on genetic diversity and genetic gain in three French dairy cattle breeds. Genet. Sel. Evol. 51:1–13. doi:10.1186/s12711-019-0495-1.

Druet, T., S. Fritz, E. Sellem, B. Basso, O. Gérard, L. Salas-Cortes, P. Humblot, X. Druart, and A. Eggen. 2009. Estimation of genetic parameters and genome scan for 15 semen characteristics traits of Holstein bulls. J. Anim. Breed. Genet. 126:269–277. doi:10.1111/j.1439-0388.2008.00788.x.

Duchesne, A., C. Grohs, P. Michot, M. Bertaud, D. Boichard, S. Floriot, and A. Capitan. 2016. Du phénotype à la mutation causale: Le cas des anomalies récessives bovines. INRA Prod. Anim. 29:319–327.

Duchesne, A., A. Vaiman, M. Frah, S. Floriot, S. Legoueix-Rodriguez, A. Desmazières, S. Fritz, C. Beauvallet, O. Albaric, E. Venot, M. Bertaud, R. Saintilan, R. Guatteo, D. Esquerré, J. Branchu, A. Fleming, A. Brice, F. Darios, J.L. Vilotte, G. Stevanin, D. Boichard, and K.H. El Hachimi. 2018. Progressive ataxia of Charolais cattle highlights a role of KIF1C in sustainable myelination. PLoS Genet. 14:e1007550. doi:10.1371/journal.pgen.1007550.

Ducrocq, V., and P. Humblot. 1995. Genetic characteristics and evolution of semen production of young Normande bulls. Livest. Prod. Sci. 41:1–10. doi:10.1016/0301-6226(94)00029-7.

Edel, P. 2003. Bojanus et l'honneur retrouvé des aurochs. Alsace monde 17:86–89.

Edwards, C.J., J.F. Baird, and D.E. MacHugh. 2007a. Taurine and zebu admixture in Near Eastern cattle: A comparison of mitochondrial, autosomal and Y-chromosomal data. Anim. Genet. 38:520–524. doi:10.1111/j.1365-2052.2007.01638.x.

Edwards, C.J., R. Bollongino, A. Scheu, A. Chamberlain, A. Tresset, J.D. Vigne, J.F. Baird, G. Larson, S.Y.W. Ho, T.H. Heupink, B. Shapiro, A.R. Freeman, M.G. Thomas, R.M. Arbogast, B. Arndt, L. Bartosiewicz, N. Benecke, M. Budja, L. Chaix, A.M. Choyke, E. Coqueugniot, H.J. Döhle, H. Göldner, S. Hartz, D. Helmer, B. Herzig, H. Hongo, M. Mashkour, M. Özdogan, E. Pucher, G. Roth, S. Schade-Lindig, U. Schmölcke, R.J. Schulting, E. Stephan, H.P. Uerpmann, I. Vörös, B. Voytek, D.G. Bradley, and J. Burger. 2007b. Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. Proc. R. Soc. B Biol. Sci. 274:1377–1385. doi:10.1098/rspb.2007.0020.

Edwards, C.J., C. Ginja, J. Kantanen, L. Pérez-Pardal, A. Tresset, F. Stock, L.T. Gama, M.C.T. Penedo, D.G. Bradley, J.A. Lenstra, and I.J. Nijman. 2011. Dual origins of dairy cattle farming - Evidence from a comprehensive survey of european Y-chromosomal variation. PLoS One 6:e15922. doi:10.1371/journal.pone.0015922.

Edwards, C.J., D.A. Magee, S.D.E. Park, P.A. McGettigan, A.J. Lohan, A. Murphy, E.K. Finlay, B. Shapiro, A.T. Chamberlain, M.B. Richards, D.G. Bradley, B.J. Loftus, and D.E. MacHugh. 2010. A complete mitochondrial genome sequence from a mesolithic wild aurochs (Bos primigenius). PLoS One 5:e9255. doi:10.1371/journal.pone.0009255.

Ellegren, H. 2011. Sex-chromosome evolution: Recent progress and the influence of male and female heterogamety. Nat. Rev. Genet. 12:157–166. doi:10.1038/nrg2948.

Ellis, N., and P.N. Goodfellow. 1989. The mammalian pseudoautosomal region. Trends Genet. 5:406–410. doi:10.1016/0168-9525(89)90199-6.

Ensembl. 2020. Ensembl Variation - Calculated Variant Consequences. Accessed. http://www.ensembl.org/info/genome/variation/prediction/predicted_data.html.

Ernster, L., and G. Schatz. 1981. Mitochondria : A Historical Review 91.

Escouflaire, C., S. Fritz, C. Hozé, E. Rebours, A. Barbat, M. Bizard, A. Chamberlain, C. Vander Jagt, M. Boussaha, C. Grohs, A. Allais-Bonnet, M. Philippe, A. Vallée, Y. Amigues, B.J. Hayes, D. Boichard, and A. Capitan. 2018. Identification and characterization of two new recessive embryonic lethal mutations in Holstein cattle. Page 12 in 11th World Congress on Genetics Applied to Livestock Production, Auckland (Nouvelle-Zélande).

Fasquelle, C., A. Sartelet, W. Li, M. Dive, N. Tamma, C. Michaux, T. Druet, I.J. Huijbers, C.M. Isacke, W. Coppieters, M. Georges, and C. Charlier. 2009. Balancing selection of a frame-shift mutation in the MRC2 gene accounts for the outbreak of the crooked tail syndrome in Belgian blue cattle. PLoS Genet. 5:e1000666. doi:10.1371/journal.pgen.1000666.

Felsenstein, J. 1985. Confidence Limits on Phylogenies : An Approach Using the Bootstrap. Evolution (N. Y). 39:783–791.

Ferlin, A., A. Tessari, F. Ganz, E. Marchina, S. Barlati, A. Garolla, B. Engl, and C. Foresta. 2005. Association or partial AZFc region deletions with spermatogenic impairment and male infertility. J. Med. Genet. 42:209–213. doi:10.1136/jmg.2004.025833.

Fortes, M.R.S., K.L. DeAtley, S.A. Lehnert, B.M. Burns, A. Reverter, R.J. Hawken, G. Boe-Hansen, S.S. Moore, and M.G. Thomas. 2013. Genomic regions associated with fertility traits in male and female cattle: Advances from microsatellites to high-density chips and beyond. Anim. Reprod. Sci. 141:1–19. doi:10.1016/j.anireprosci.2013.07.002.

France Agricole ed. . 2011. La Révolution Génomique Animale. Agri Produ.

Frantz, L.A.F., J.G. Schraiber, O. Madsen, H.J. Megens, A. Cagan, M. Bosse, Y. Paudel, R.P.M.A. Crooijmans, G. Larson, and M.A.M. Groenen. 2015. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. Nat. Genet. 47:1141–1148. doi:10.1038/ng.3394.

Fritz, S., A. Capitan, A. Djari, S.C. Rodriguez, A. Barbat, A. Baur, C. Grohs, B. Weiss, M. Boussaha, D. Esquerré, C. Klopp, D. Rocha, and D. Boichard. 2013. Detection of Haplotypes Associated with Prenatal Death in Dairy Cattle and Identification of Deleterious Mutations in GART, SHBG and SLC37A2. PLoS One 8:e65550. doi:10.1371/journal.pone.0065550.

Fritz, S., C. Hoze, E. Rebours, A. Barbat, M. Bizard, A. Chamberlain, C. Escouflaire, C. Vander Jagt, M. Boussaha, C. Grohs, A. Allais-Bonnet, M. Philippe, A. Vallée, Y. Amigues, B.J. Hayes, D. Boichard, and A. Capitan. 2018. An initiator codon mutation in SDE2 causes recessive embryonic lethality in Holstein cattle. J. Dairy Sci. 101:6220–6231. doi:10.3168/jds.2017-14119.

Fritz, S., P. Michot, C. Hoze, C. Grohs, U. Paris-saclay, and M. Nationale. 2016. Anticiper l ’émergence d ’anomalies génétiques grâce aux données génomiques. INRA Prod. Anim. 29:339–349.

Fuerst-Waltl, B., H. Schwarzenbacher, C. Perner, and J. Sölkner. 2006. Effects of age and environmental factors on semen production and semen quality of Austrian Simmental bulls. Anim. Reprod. Sci. 95:27–37. doi:10.1016/j.anireprosci.2005.09.002.

Galluzzi, L., N. Joza, E. Tasdemir, M.C. Maiuri, M. Hengartner, J.M. Abrams, N. Tavernarakis, J. Penninger, F. Madeo, and G. Kroemer. 2008. No death without life: Vital functions of apoptotic effectors. Cell Death Differ. 15:1113–1123. doi:10.1038/cdd.2008.28.

Garmyn, A.J., D.W. Moser, R.A. Christmas, and J. Minick Bormann. 2011. Estimation of genetic

parameters and effects of cytoplasmic line on scrotal circumference and semen quality traits in Angus bulls. J. Anim. Sci. 89:693–698. doi:10.2527/jas.2010-3534.

Gauthier A. 1990. La Domestication - Et l'homme Créa l'animal. ERRANCE.

Gautier, M. 2003. Cartographie fine de QTL chez les bovins laitiers. Thèse, AgroParisTech,.

Gautier, M., D. Laloë, and K. Moazami-Goudarzi. 2010. Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. PLoS One 5:e13038. doi:10.1371/journal.pone.0013038.

GEN'IAtest. 2020. Production Conventionnelle. Accessed March 12, 2020. https://www.geniatest.com/production-conventionnelle.html.

GenEval. 2019. Evaluation génétique des taureaux Montbéliards.

Gérard, O., C. Ponsart, M. Petit, and P. Humblot. 2008. Evolution des techniques de préparation de la semence et d'insémination artificielle chez les bovins. Rencontres Rech. Ruminants 351–354.

Gipson, T.A., D.W. Vogt, J.W. Massey, and M.R. Ellersieck. 1985. Associations of scrotal circumference with semen traits in young beef bulls. Theriogenology 24:217–225. doi:10.1016/0093-691X(85)90186-4.

Götherström, A., C. Anderung, L. Hellborg, R. Elburg, C. Smith, D.G. Bradley, and H. Ellegren. 2005. Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. Proc. R. Soc. B Biol. Sci. 272:2345–2350. doi:10.1098/rspb.2005.3243.

Goto, H., L. Peng, and K.D. Makova. 2009. Evolution of X-degenerate y chromosome genes in greater apes: Conservation of gene content in human and gorilla, but not chimpanzee. J. Mol. Evol. 68:134–144. doi:10.1007/s00239-008-9189-y.

Gouyon, P.-H., J.-P. Henry, and J. Arnould. 1997. Les Avatars Du Gène. Belin.

Graves, J.A.M. 2006. Sex chromosome specialization and degeneration in mammals. Cell 124:901–914. doi:10.1016/j.cell.2006.02.024.

Graves, J.A.M., E. Koina, and N. Sankovic. 2006. How the gene content of human sex chromosomes evolved. Curr. Opin. Genet. Dev. 16:219–224. doi:10.1016/j.gde.2006.04.007.

Greaves, L.C., and R.W. Taylor. 2006. Mitochondrial DNA mutations in human disease. IUBMB Life 58:143–151. doi:10.1080/15216540600686888.

Gredler, B., C. Fuerst, B. Fuerst-Waltl, H. Schwarzenbacher, and J. Sölkner. 2007. Genetic parameters for semen production traits in Austrian dual-purpose simmental bulls. Reprod. Domest. Anim. 42:326–328. doi:10.1111/j.1439-0531.2006.00778.x.

Green, R., J. Krause, A. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. Fritz, N.F. Hansen, E.Y. Durand, A.-S. Malaspinas, J.D. Jensen, T. Marques-Bonet, C. Alkan, K. Prüfer, M. Meyer, H. Burbano, J.M. Good, R. Schultz, M. Rasilla Vives, M. Lachmann, D. Reich, and S. Pääbo. 2010. A draft sequence of the neandertal genome. Science (80-.). 328:710–722. doi:10.1126/science.1188021PubMedID:20448178.

Griffon, L., P. Boulesteix, A. Delpeuch, A. Govignon-Gion, J. Guerrier, O. Leudet, S. Miller, R. Saintilan, E. Venot, and T. Tribout. 2017. La sélection génétique des races bovines allaitantes en France: Un dispositif et des outils innovants au service des filières viande. INRA Prod. Anim. 30:107–124. doi:10.20870/productions-animales.2017.30.2.2237.

Groeneveld, L.F., J.A. Lenstra, H. Eding, M.A. Toro, B. Scherf, D. Pilling, R. Negrini, E.K. Finlay, H. Jianlin, E. Groeneveld, and S. Weigend. 2010. Genetic diversity in farm animals - A review. Anim.

Genet. 41:6–31. doi:10.1111/j.1365-2052.2010.02038.x.

Grohs, C., A. Duchesne, S. Floriot, M.C. Deloche, D. Boichard, A. Ducos, and C. Danchin-Burge. 2016. L'Observatoire National des Anomalies Bovines, son action et ses résultats pour une aide efficace à la gestion des anomalies génétiques. Prod. Anim. 29:307–317.

Guénet, J.L., J.J. Panthier, P. Avner, E. Heard, and X. Montagutelli. 2015. L'héritage de Mary F. Lyon (1925-2014). Medecine/Sciences 31:687–689. doi:10.1051/medsci/20153106024.

Guintard, C., and O. Neron de Surgy. 2014. L'aurochs. De Lascaux Au XXIe Siècle. LE GERFAUT.

Gusmão, L., M. Brión, and I. Gomes. 2008. Chapter 30 The human Y chromosome male-specific polymorphisms and forensic genetics. Handb. Anal. Sep. 6:969–1000. doi:10.1016/S1567-7192(06)06034-7.

Hagström, E., C. Freyer, B.J. Battersby, J.B. Stewart, and N.G. Larsson. 2014. No recombination of mtDNA after heteroplasmy for 50 generations in the mouse maternal germline. Nucleic Acids Res. 42:1111–1116. doi:10.1093/nar/gkt969.

Happle, R. 1993. Mosaicism in Human Skin: Understanding the Patterns and Mechanisms. Arch. Dermatol. 129:1460–1470. doi:10.1001/archderm.1993.01680320094012.

Hayes, B.J., P.J. Bowman, A.J. Chamberlain, and M.E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92:433–443. doi:10.3168/jds.2008-1646.

Henderson, C.R. 1984. Applications of Linear Models in Animal Breeding. U. of Guelph, ed.

Hesser, A., C. Darr, K. Gonzales, H. Power, T. Scanlan, J. Thompson, C. Love, B. Christensen, and S. Meyers. 2017. Semen evaluation and fertility assessment in a purebred dog breeding facility. Theriogenology 87:115–123. doi:10.1016/j.theriogenology.2016.08.012.

Howard, J.T., J.E. Pryce, C. Baes, and C. Maltecca. 2017. Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. J. Dairy Sci. 100:6009–6024. doi:10.3168/jds.2017-12787.

Hozé, C., C. Escouflaire, M. Mesbah-Uddin, A. Barbat, M. Boussaha, M.C. Deloche, D. Boichard, S. Fritz, and A. Capitan. 2020. Short communication: A splice site mutation in CENPU is associated with recessive embryonic lethality in Holstein cattle. J. Dairy Sci. 103:607–612. doi:10.3168/jds.2019-17056.

Hoze, C., S. Fritz, A. Baur, C. Grohs, C. Danchin-Burge, and D. Boichard. 2018. Prise en compte des gènes d'intérêt dans les objectifs de sélection en bovins laitiers. Rencontres autour la Rech. sur les Ruminants 1–4.

Hua, L., G. Rincon, and W. Liu. 2020. Poster : Sequence and assembly of the Holstein Y chromosome. Page in Plant and Animal Genome.

Huet, Y.-A. 2013. L'hommage à Jocko, le taureau aux 400 000 filles. Ouest Fr.

Hughes, J.F., H. Skaletsky, T. Pyntikova, P.J. Minx, T. Graves, S. Rozen, R.K. Wilson, and D.C. Page. 2005. Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. Nature 437:101–104. doi:10.1038/nature04101.

Idele. Reproscope. Accessed January 15, 2020. http://www.reproscope.fr/.

Idele. 2020. Lise Des Taureaux Utilisables En Monte Publique Par Insémination. Accessed. http://idele.fr/no_cache/recherche/publication/idelesolr/recommends/liste-des-taureaux-utilisables-en-monte-publique-par-insemination.html.

Institut de l'élevage. 2018. Indicateurs de variabilité génétique.

Ito, Y., N. Sanjo, M. Hizume, A. Kobayashi, T. Ohgami, K. Satoh, T. Hamaguchi, M. Yamada, T. Kitamoto, H. Mizusawa, and T. Yokota. 2018. Biochemical features of genetic Creutzfeldt-Jakob disease with valine-to-isoleucine substitution at codon 180 on the prion protein gene. Biochem. Biophys. Res. Commun. 496:1055–1061. doi:10.1016/j.bbrc.2018.01.119.

Jamrozik, J., J. Fatehi, G.J. Kistemaker, and L.R. Schaeffer. 2005. Estimates of genetic parameters for Canadian Holstein female reproduction traits. J. Dairy Sci. 88:2199–2208. doi:10.3168/jds.S0022-0302(05)72895-2.

Jiang, J., Y. Jing, G.J. Cost, J. Chiang, H.J. Kolpa, A.M. Cotton, D.M. Carone, B.R. Carone, D.A. Shivak, D.Y. Guschin, J.R. Pearl, E.J. Rebar, M. Byron, P.D. Gregory, C.J. Brown, F.D. Urnov, L.L. Hall, and J.B. Lawrence. 2014. Translating Dosage Compensation to Trisomy 21. Nature 500:296–300. doi:10.1038/nature12394.Translating.

Jin, H., G. Zhang, Y. Zhou, C. Chang, and Q. Lu. 2016. Old lines tell new tales: Blaschko linear lupus erythematosis. Autoimmun. Rev. 15:291–306. doi:10.1016/j.autrev.2015.11.014.

Johannsen, W. 2014. The genotype conception of heredity1. Int. J. Epidemiol. 43:989–1000. doi:10.1093/ije/dyu063.

Johnson, T., M. Keehan, C. Harland, T. Lopdell, R.J. Spelman, S.R. Davis, B.D. Rosen, T.P.L. Smith, and C. Couldrey. 2019. Short communication: Identification of the pseudoautosomal region in the Hereford bovine reference genome assembly ARS-UCD1.2. J. Dairy Sci. 102:3254–3258. doi:10.3168/jds.2018-15638.

Jonas, D. 2016. Evaluation of haplotype-based genomic selection methods with focus on their performances in a multi-breed context in dairy cattle. Thèse, AgroParisTech,.

Jussiau, R., J. Rigal, A. Papet, and E. Zanchi. 2013. Amélioration Génétique Des Animaux D'élevage, Génome, Caractères, Sélection et Croisements. Educagri e.

Kantanen, J., C.J. Edwards, D.G. Bradley, H. Viinalass, S. Thessler, Z. Ivanova, T. Kiselyova, M. Ćinkulov, R. Popov, S. Stojanovic, I. Ammosov, and J. Vilkki. 2009. Maternal and paternal genealogy of Eurasian taurine cattle (Bos taurus). Heredity (Edinb). 103:404–415. doi:10.1038/hdy.2009.68.

Kealey, C.G., M.D. MacNeil, M.W. Tess, T.W. Geary, and R.A. Bellows. 2006. Genetic parameter estimates for scrotal circumference and semen characteristics of Line 1 Hereford bulls. J. Anim. Sci. 84:283–290. doi:10.2527/2006.842283x.

Kivisild, T. 2015. Maternal ancestry and population history from whole mitochondrial genomes. Investig. Genet. 6:1–10. doi:10.1186/s13323-015-0022-2.

Knights, S.A., R.L. Baker, D. Gianola, and J.B. Gibb. 1984. Estimates of heritabilities and of genetic and phenotypic correlations among growth and reproductive traits in yearling Angus bulls.. J. Anim. Sci. 58:887–893. doi:10.2527/jas1984.584887x.

Koboldt, D.C., K. Chen, T. Wylie, D.E. Larson, M.D. McLellan, E.R. Mardis, G.M. Weinstock, R.K. Wilson, and L. Ding. 2009. VarScan: Variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 25:2283–2285. doi:10.1093/bioinformatics/btp373.

Kumar, P., S. Henikoff, and P.C. Ng. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4:1073–1082. doi:10.1038/nprot.2009.86.

Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura. 2018. MEGA X: Molecular evolutionary genetics

analysis across computing platforms. Mol. Biol. Evol. 35:1547–1549. doi:10.1093/molbev/msy096.

Van Laere, A.S., W. Coppieters, and M. Georges. 2008. Characterization of the bovine pseudoautosomal boundary: Documenting the evolutionary history of mammalian sex chromosomes. Genome Res. 18:1884–1895. doi:10.1101/gr.082487.108.

Lari, M., E. Rizzi, S. Mona, G. Corti, G. Catalano, K. Chen, C. Vernesi, G. Larson, P. Boscato, G. De Bellis, A. Cooper, D. Caramelli, and G. Bertorelle. 2011. The complete mitochondrial genome of an 11,450-year-old aurochsen (bos primigenius) from Central Italy. BMC Evol. Biol. 11:32. doi:10.1186/1471-2148-11-32.

Larson, G., D.R. Piperno, R.G. Allaby, M.D. Purugganan, L. Andersson, M. Arroyo-Kalin, L. Barton, C.C. Vigueira, T. Denham, K. Dobney, A.N. Doust, P. Gepts, M.T.P. Gilbert, K.J. Gremillion, L. Lucas, L. Lukens, F.B. Marshall, K.M. Olsen, J.C. Pires, P.J. Richerson, R.R. De Casas, O.I. Sanjur, M.G. Thomas, and D.Q. Fuller. 2014. Current perspectives and the future of domestication studies. Proc. Natl. Acad. Sci. U. S. A. 111:6139–6146. doi:10.1073/pnas.1323964111.

Leclerc, H., R. Lefebvre, M. Douguet, F. Phocas, S. Mattalia, I.D.L. Elevage, D. Génétique, J. Cedex, H. Leclerc, R. Lefebvre, M. Douguet, F. Phocas, and S. Mattalia. 2016. Mortalité des veaux : analyse phénotypique et étude de la composante génétique Calf mortality : phenotypic and genetic analysis 149–152.

Ledos, H., S. Moureaux, and P.L.E. Mezec. 2003. Mesure du taux d'incompatibilité des génisses issues de taureaux de testage et incidence sur le protocole de testage. Rencontres Rech. Ruminants 66.

Lefebvre, S., and M.L. Mikkola. 2014. Ectodysplasin research-Where to next?. Semin. Immunol. 26:220–228. doi:10.1016/j.smim.2014.05.002.

Leigh, J.W., and D. Bryant. 2015. POPART: Full-feature software for haplotype network construction. Methods Ecol. Evol. 6:1110–1116. doi:10.1111/2041-210X.12410.

Lenstra, J.A., P. Ajmone-Marsan, A. Beja-Pereira, R. Bollongino, D.G. Bradley, L. Colli, A. De Gaetano, C.J. Edwards, M. Felius, L. Ferretti, C. Ginja, P. Hristov, J. Kantanen, J.P. Lirón, D.A. Magee, R. Negrini, and G.A. Radoslavov. 2014. Meta-Analysis of mitochondrial DNA reveals several population bottlenecks during worldwide migrations of cattle. Diversity 6:178–187. doi:10.3390/d6010178.

Leroy, G. 2015. Guide de compréhension des indicateurs de variabilité génétique.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. doi:10.1093/bioinformatics/btp352.

Lillehammer, M., T.H.E. Meuwissen, and A.K. Sonesson. 2011. A comparison of dairy cattle breeding designs that use genomic selection. J. Dairy Sci. 94:493–500. doi:10.3168/jds.2010-3518.

Loftus, R.T., O. Ertugrul, A.H. Barbar, M.A.A. El-Barody, D.E. Machugh, S.D.E. Park, and D.G. Bradley. 1999. A microsatellite survey of cattle from a centre of origin: The Near East. Mol. Ecol. 8:2015–2022. doi:10.1046/j.1365-294X.1999.00805.x.

Loftus, R.T., D.E. MacHugh, D.G. Bradley, P.M. Sharp, and P. Cunningham. 1994. Evidence for two independent domestications of cattle. Proc. Natl. Acad. Sci. U. S. A. 91:2757–2761. doi:10.1073/pnas.91.7.2757.

Losano, J.D.A., J.F. Padín, I. Méndez-López, D.S.R. Angrimani, A.G. García, V.H. Barnabe, and M. Nichi.

2017. The stimulated glycolytic pathway is able to maintain ATP levels and kinetic patterns of bovine epididymal sperm subjected to mitochondrial uncoupling. Oxid. Med. Cell. Longev. 2017. doi:10.1155/2017/1682393.

Lubs, H.A., R.E. Stevenson, and C.E. Schwartz. 2012. Fragile X and X-linked intellectual disability: Four decades of discovery. Am. J. Hum. Genet. 90:579–590. doi:10.1016/j.ajhg.2012.02.018.

Lund, M.S., A.P. De Roos, A.G. De Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, R. Reents, C. Schrooten, F. Seefried, and G. Su. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. Genet. Sel. Evol. 43:43. doi:10.1186/1297-9686-43-43.

Luo, S., C.A. Valencia, J. Zhang, N.C. Lee, J. Slone, B. Gui, X. Wang, Z. Li, S. Dell, J. Brown, S.M. Chen, Y.H. Chien, W.L. Hwu, P.C. Fan, L.J. Wong, P.S. Atwal, and T. Huang. 2018. Biparental inheritance of mitochondrial DNA in humans. Proc. Natl. Acad. Sci. U. S. A. 115:13039–13044. doi:10.1073/pnas.1810946115.

Lush, J.L., and Holbert. 1936. Genetic history of the Holstein-Friesian cattle in the United States. J. Hered. 27:61–72. doi:https://doi.org/10.1093/oxfordjournals.jhered.a104174.

Lynch, M., B. Koskella, and S. Schaack. 2006. Mutation pressure and the evolution of organelle genomic architecture. Science (80-.). 311:1727–1730. doi:10.1126/science.1118884.

MacHugh, D.E., and D.G. Bradley. 2001. Livestock genetic origins: Goats buck the trend. Proc. Natl. Acad. Sci. U. S. A. 98:5382–5384. doi:10.1073/pnas.111163198.

MacHugh, D.E., M.D. Shriver, R.T. Loftus, P. Cunningham, and D.G. Bradley. 1997. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (Bos taurus and Bos indicus). Genetics 146:1071–1086.

Madeira, F., Y.M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A.R.N. Tivey, S.C. Potter, R.D. Finn, and R. Lopez. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 47:W636–W641. doi:10.1093/nar/gkz268.

Malik, A.N., and A. Czajka. 2013. Is mitochondrial DNA content a potential biomarker of mitochondrial dysfunction?. Mitochondrion 13:481–492. doi:10.1016/j.mito.2012.10.011.

Mancuso, M., M. Filosto, A. Choub, M. Tentorio, L. Broglio, A. Padovani, and G. Siciliano. 2007. Mitochondrial DNA-related disorders. Biosci. Rep. 27:31–37. doi:10.1007/s10540-007-9035-2.

Marc, S. 2012. Actualités en cryoconservation des semences des principales espèces d'intérêt vétérinaire. Vetagro Sup- Campus vétérinaire de Lyon,.

Massilani, D., S. Guimaraes, J.P. Brugal, E.A. Bennett, M. Tokarska, R.M. Arbogast, G. Baryshnikov, G. Boeskorov, J.C. Castel, S. Davydov, S. Madelaine, O. Putelat, N.N. Spasskaya, H.P. Uerpmann, T. Grange, and E.M. Geigl. 2016. Past climate changes, population dynamics and the origin of Bison in Europe. BMC Biol. 14:93. doi:10.1186/s12915-016-0317-7.

Mathevon, M., M.M. Buhr, and J.C.M. Dekkers. 1998a. Environmental, Management, and Genetic Factors Affecting Semen Production in Holstein Bulls. J. Dairy Sci. 81:3321–3330. doi:10.3168/jds.S0022-0302(98)75898-9.

Mathevon, M., J.C.M. Dekkers, and M.M. Buhr. 1998b. Environmental, management and genetic factors affecting semen production in French Montbéliard bulls. Livest. Prod. Sci. 55:65–77. doi:10.1016/S0301-6226(98)00116-X.

May-Panloup, P., M.F. Chrétien, F. Savagner, C. Vasseur, M. Jean, Y. Malthièry, and P. Reynier. 2003. Increased sperm mitochondrial DNA content in male infertility. Hum. Reprod. 18:550–556.

doi:10.1093/humrep/deg096.

Mazoyer, M. 1977. Evolution et différentiation des systèmes agricoles d'exploitation de la nature. J. d'agriculture Tradit. Bot. appliquée 24:267–275. doi:10.3406/jatba.1977.3289.

McClure, M.C., D. Bickhart, D. Null, P. VanRaden, L. Xu, G. Wiggans, G. Liu, S. Schroeder, J. Glasscock, J. Armstrong, J.B. Cole, C.P. Van Tassell, and T.S. Sonstegard. 2014. Bovine exome sequence analysis and targeted SNP genotyping of recessive fertility defects BH1, HH2, and HH3 reveal a putative causative mutation in SMC2 for HH3. PLoS One 9:e92769. doi:10.1371/journal.pone.0092769.

Mcdaneld, T.G., L.A. Kuehn, M.G. Thomas, W.M. Snelling, T.P.L. Smith, T.S. Sonstegard, L.K. Matukumalli, E.J. Pollak, and J.W. Keele. 2012. Y are you not pregnant : Identification of Y chromosome segments in female cattle. Am. Soc. Anim. Sci. 2142–2151. doi:10.2527/jas2011-4536.

McLaren, W., L. Gil, S.E. Hunt, H.S. Riat, G.R.S. Ritchie, A. Thormann, P. Flicek, and F. Cunningham. 2016. The Ensembl Variant Effect Predictor. Genome Biol. 17:1–14. doi:10.1186/s13059-016-0974-4.

McVean, G.T., and L.D. Hurst. 1997. Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. Nature 386:388–392. doi:10.1038/386388a0.

Melvin, R.G., and J.W.O. Ballard. 2017. Cellular and population level processes influence the rate, accumulation and observed frequency of inherited and somatic mtDNA mutations. Mutagenesis 32:323–334. doi:10.1093/mutage/gex004.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Meyer, K. 2007. WOMBAT: a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML).. J. Zhejiang Univ. Sci. B. 8:815–821. doi:10.1631/jzus.2007.B0815.

Meyers, S., E. Bulkeley, and A. Foutouhi. 2019. Sperm mitochondrial regulation in motility and fertility in horses. Reprod. Domest. Anim. 54:22–28. doi:10.1111/rda.13461.

Michot, P. 2017. Utilisation des séquences de génome complet pour l'identification de mutations délétères responsables d'anomalies génétiques récessives chez le bovin. Université Paris-Saclay,.

Michot, P., S. Chahory, A. Marete, C. Grohs, D. Dagios, E. Donzel, A. Aboukadiri, M.C. Deloche, A. Allais-Bonnet, M. Chambrial, S. Barbey, L. Genestout, M. Boussaha, C. Danchin-Burge, S. Fritz, D. Boichard, and A. Capitan. 2016. A reverse genetic approach identifies an ancestral frameshift mutation in RP1 causing recessive progressive retinal degeneration in European cattle breeds. Genet. Sel. Evol. 48:1–15. doi:10.1186/s12711-016-0232-y.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648–4655. doi:10.3168/jds.2009-2064.

Molho-Pessach, V., and J. V. Schaffer. 2011. Blaschko lines and other patterns of cutaneous mosaicism. Clin. Dermatol. 29:205–225. doi:10.1016/j.clindermatol.2010.09.012.

Mona, S., G. Catalano, M. Lari, G. Larson, P. Boscato, A. Casoli, L. Sineo, C. Di Patti, E. Pecchioli, D. Caramelli, and G. Bertorelle. 2010. Population dynamic of the extinct European aurochs: Genetic evidence of a north-south differentiation pattern and no evidence of post-glacial expansion. BMC Evol. Biol. 10:83. doi:10.1186/1471-2148-10-83.

Moraes, C.R., and S. Meyers. 2018. The sperm mitochondrion: Organelle of many functions. Anim. Reprod. Sci. 194:71–80. doi:10.1016/j.anireprosci.2018.03.024.

Murgiano, L., V. Shirokova, M.M. Welle, V. Jagannathan, P. Plattet, A. Oevermann, A. Pienkowska-Schelling, D. Gallo, A. Gentile, M. Mikkola, and C. Drögemüller. 2015. Hairless Streaks in Cattle Implicate TSR2 in Early Hair Follicle Formation. PLoS Genet. 11:e1005427. doi:10.1371/journal.pgen.1005427.

Murphy, C., S.A. Holden, E.M. Murphy, A.R. Cromie, P. Lonergan, and S. Fair. 2015. The impact of storage temperature and sperm number on the fertility of liquid-stored bull semen. Reprod. Fertil. Dev. 28:1349–1359. doi:10.1071/RD14369.

Niemi, M., A. Bläuer, T. Iso-Touru, J. Harjula, V.N. Edmark, E. Rannamäe, L. Lõugas, A. Sajantila, K. Lidén, and J.P. Taavitsainen. 2015. Temporal fluctuation in North East Baltic Sea region cattle population revealed by mitochondrial and Y-chromosomal DNA analyses. PLoS One 10:e0123821. doi:10.1371/journal.pone.0123821.

Noda, A., R. Yonesaka, S. Sasazaki, and H. Mannen. 2018. The mtDNA haplogroup P of modern Asian cattle: A genetic legacy of Asian aurochs?. PLoS One 13:e0190937. doi:10.1371/journal.pone.0190937.

Novosel, D., V. Cubric-Curik, M. Simcic, P. Dovc, and I. Curik. 2019. Two detrimental mutations in cattle mitogenome indicate the presence of leber's hereditary optic neuropathy. J. Cent. Eur. Agric. 20:19–24. doi:10.5513/JCEA01/20.1.2413.

Olalde, I., H. Schroeder, M. Sandoval-Velasco, L. Vinner, I. Lobón, O. Ramirez, S. Civit, P.G. Borja, D.C. Salazar-García, S. Talamo, J.M. Fullola, F.X. Oms, M. Pedro, P. Martínez, M. Sanz, J. Daura, J. Zilhão, T. Marquès-Bonet, M.T.P. Gilbert, and C. Lalueza-Fox. 2015. A common genetic origin for early farmers from mediterranean cardial and central european LBK cultures. Mol. Biol. Evol. 32:3132–3142. doi:10.1093/molbev/msv181.

Olivieri, A., F. Gandini, A. Achilli, A. Fichera, E. Rizzi, S. Bonfiglio, V. Battaglia, S. Brandini, A. De Gaetano, A. El-Beltagi, H. Lancioni, S. Agha, O. Semino, L. Ferretti, and A. Torroni. 2015. Mitogenomes from Egyptian cattle breeds: New clues on the origin of haplogroup Q and the early spread of Bos taurus from the Near East. PLoS One 10:e0141170. doi:10.1371/journal.pone.0141170.

Olsen, H.B., B. Heringstad, and G. Klemetsdal. 2020. Genetic analysis of semen characteristic traits in young Norwegian Red bulls. J. Dairy Sci. 103:545–555. doi:10.3168/jds.2019-17291.

ONAB. 2014. La Vache Milka. Accessed December 21, 2019. https://www.onab.fr/Anomalies-en-cours-d-etude/Milka.

ONAB. 2019. Mitochondropathie En Race Montbéliarde. Accessed January 18, 2020. https://www.onab.fr/Anomalies-identifiees/MTCP.

ONAB. 2020. Déclarer Une Anomalie, Mode d'emploi. Accessed March 30, 2020. https://www.onab.fr/Declarer-une-anomalie.

Orsztynowicz, M., P. Pawlak, Z. Podstawski, W. Nizanski, A. Partyka, M. Gotowiecka, K. Kosiniak-Kamysz, and D. Lechniak. 2016. Mitochondrial DNA Copy Number in Spermatozoa of Fertile Stallions. Reprod. Domest. Anim. 51:378–385. doi:10.1111/rda.12689.

Park, S.D.E., D.A. Magee, P.A. McGettigan, M.D. Teasdale, C.J. Edwards, A.J. Lohan, A. Murphy, M. Braud, M.T. Donoghue, Y. Liu, A.T. Chamberlain, K. Rue-Albrecht, S. Schroeder, C. Spillane, S. Tai, D.G. Bradley, T.S. Sonstegard, B.J. Loftus, and D.E. MacHugh. 2015. Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and

evolution of cattle. Genome Biol. 16:1–15. doi:10.1186/s13059-015-0790-2.

Patry, C., and V. Ducrocq. 2011. Evidence of biases in genetic evaluations due to genomic preselection in dairy cattle. J. Dairy Sci. 94:1011–1020. doi:10.3168/jds.2010-3804.

Pausch, H., H. Schwarzenbacher, J. Burgstaller, K. Flisikowski, C. Wurmser, S. Jansen, S. Jung, A. Schnieke, T. Wittek, and R. Fries. 2015. Homozygous haplotype deficiency reveals deleterious mutations compromising reproductive and rearing success in cattle. BMC Genomics 16:312. doi:10.1186/s12864-015-1483-7.

Payet, F. 2017. Apport de l'étude de l'inactivation du chromosome X dans l'interprétation des microremaniements détectés par ACPA dans le bilan étiologique des anomalies du développement. Faculté de Médecine Henri Warembourg,.

Pedersen, A.S., L.C. Berg, K. Almstrup, and P.D. Thomsen. 2014. A tortoiseshell male cat: Chromosome analysis and histologic examination of the testis. Cytogenet. Genome Res. 142:107–111. doi:10.1159/000356466.

Pérez-Pardal, L., L.J. Royo, A. Beja-Pereira, S. Chen, R.J.C. Cantet, A. Traoré, I. Curik, J. Sölkner, R. Bozzi, I. Fernández, I. Lvarez, J.P. Gutiérrez, E. Gómez, F.A. Ponce De León, and F. Goyache. 2010. Multiple paternal origins of domestic cattle revealed by Y-specific interspersed multilocus microsatellites. Heredity (Edinb). 105:511–519. doi:10.1038/hdy.2010.30.

Pitt, D., N. Sevane, E.L. Nicolazzi, D.E. MacHugh, S.D.E. Park, L. Colli, R. Martinez, M.W. Bruford, and P. Orozco-terWengel. 2019. Domestication of cattle: Two or three events?. Evol. Appl. 12:123–136. doi:10.1111/eva.12674.

Porter, V., L. Alderson, S.J.G. Hall, and D.P. Sponenberg. 2016. Mason's World Encyclopedia of Livestock Breeds and Breeding.

Poutous, M., M. Briend, S. Calomiti, D. Doan, C. Felgines, and G. Steier. 1981. Méthode de calcul des index laitiers. Bull. Tech. Inf. 361:433–446.

Prado-Martinez, J., P.H. Sudmant, J.M. Kidd, H. Li, J.L. Kelley, B. Lorente-Galdos, K.R. Veeramah, A.E. Woerner, T.D. O'Connor, G. Santpere, A. Cagan, C. Theunert, F. Casals, H. Laayouni, K. Munch, A. Hobolth, A.E. Halager, M. Malig, J. Hernandez-Rodriguez, I. Hernando-Herraez, K. Prüfer, M. Pybus, L. Johnstone, M. Lachmann, C. Alkan, D. Twigg, N. Petit, C. Baker, F. Hormozdiari, M. Fernandez-Callejo, M. Dabad, M.L. Wilson, L. Stevison, C. Camprub, T. Carvalho, A. Ruiz-Herrera, L. Vives, M. Mele, T. Abello, I. Kondova, R.E. Bontrop, A. Pusey, F. Lankester, J.A. Kiyang, R.A. Bergl, E. Lonsdorf, S. Myers, M. Ventura, P. Gagneux, D. Comas, H. Siegismund, J. Blanc, L. Agueda-Calpena, M. Gut, L. Fulton, S.A. Tishkoff, J.C. Mullikin, R.K. Wilson, I.G. Gut, M.K. Gonder, O.A. Ryder, B.H. Hahn, A. Navarro, J.M. Akey, J. Bertranpetit, D. Reich, T. Mailund, M.H. Schierup, C. Hvilsom, A.M. Andrés, J.D. Wall, C.D. Bustamante, M.F. Hammer, E.E. Eichler, and T. Marques-Bonet. 2013. Great ape genetic diversity and population history. Nature 499:471–475. doi:10.1038/nature12228.

Pryce, J.E., M.E. Goddard, H.W. Raadsma, and B.J. Hayes. 2010. Deterministic models of breeding scheme designs that incorporate genomic selection. J. Dairy Sci. 93:5455–5466. doi:10.3168/jds.2010-3256.

Puglisi, R., G. Gaspa, D. Balduzzi, A. Severgnini, R. Vanni, N. Macciotta, and A. Galli. 2016. Genomewide analysis of bull sperm quality and fertility traits. Reprod. Domest. Anim. 51:840–843. doi:10.1111/rda.12747.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. De Bakker, M.J. Daly, and P.C. Sham. 2007. PLINK: A tool set for whole-genome association and

population-based linkage analyses. Am. J. Hum. Genet. 81:559–575. doi:10.1086/519795.

La Race Bazadaise. 2005. Origines. Accessed February 17, 2020. http://www.bazadaise.fr/z/site.php?act=2_1.

Al Rawi, S. Al, Louvet-Valle, A. Djeddi, M. Sachese, E. CUletto, C. Hajar, L. Boyd, R. Legouis, and V. Galy. 2011. Postfertilization autophagy of sperm organelles prevents paternal mitochondrial DNA transmission. Science (80-.). 334:1144–1147. doi:10.1126/science.1211878.

Renault, N.K., S. Dyack, M.J. Dobson, T. Costa, W.L. Lam, and W.L. Greer. 2007. Heritable skewed X-chromosome inactivation leads to haemophilia A expression in heterozygous females. Eur. J. Hum. Genet. 15:628–637. doi:10.1038/sj.ejhg.5201799.

Rewilding Europe. Tauros. Accessed January 3, 2020. https://rewildingeurope.com/rewilding-in-action/wildlife-comeback/tauros/.

Rice, W.R., and A.K. Chippindale. 2001. Sexual recombination and the power of natural selection. Science (80-.). 294:555–559. doi:10.1126/science.1061380.

Robert-Granié, C., A. Legarra, and V. Ducrocq. 2011. Principes de base de la sélection génomique. Prod. Anim. 24:331–340.

Ross, M.T., D. V Grafham, A.J. Coffey, S. Scherer, K. Mclay, D. Muzny, M. Platzer, G.R. Howell, C. Burrows, C. P, A. Frankish, F.L. Lovell, K.L. Howe, J.L. Ashurst, R. S, R. Sudbrak, G. Wen, M.C. Jones, M.E. Hurles, T. Daniel, C.E. Scott, S. Searle, J. Ramser, A. Whittaker, N.P. Carter, S.E. Hunt, R. Chen, A. Cree, P. Havlak, A. Hodgson, M.L. Metzker, S. Richards, G. Scott, D. Steffen, E. Sodergren, D. a Wheeler, K.C. Worley, R. Ainscough, K.D. Ambrose, M.A. Ansari-lari, S. Aradhya, R. I, S. Ashwell, A.K. Babbage, C.L. Bagguley, A. Ballabio, R. Banerjee, G.E. Barker, K.F. Barlow, I.P. Barrett, K.N. Bates, D.M. Beare, O. Beasley, A. Beck, G. Bethel, K. Blechschmidt, S. Bray-allen, A.M. Bridgeman, A.J. Brown, M.J. Brown, E. a Bruford, C. Buhay, P. Burch, D. Burford, W. Burrill, J. Burton, J.M. Bye, C. Carder, L. Carrel, J. Chako, J.C. Chapman, D. Chavez, E. Chen, G. Chen, Z. Chen, C. Chinault, A. Ciccodicola, S.Y. Clark, C.M. Clee, S. Clegg, K. Clerc-blankenburg, K. Clifford, C.G. Cole, J.S. Conquer, N. Corby, R.E. Connor, J. Davies, C. Davis, J. Davis, O. Delgado, D. Deshazo, P. Dhami, Y. Ding, H. Dinh, S. Dodsworth, H. Draper, A. Dunham, M. Dunn, K.J. Durbin, I. Dutta, M. Ellwood, A. Emery-cohen, H. Errington, K. L, L. Faulkner, F. Francis, J. Frankland, A.E. Fraser, J. Gilbert, R. Gill, G. Glöckner, S.G. Gregory, C. Griffiths, R. Grocock, Y. Gu, R. Gwilliam, E. a Hart, A. Hawes, P.D. Heath, K. Heitmann, J. Hernandez, B. Hinzmann, S. Ho, M. Hoffs, P. J, E.J. Huckle, J. Hume, P.J. Hunt, A.R. Hunt, L. Jacob, D. Johnson, S. Jones, P.J. De Jong, S. S, S. Klages, A.J. Knights, A. Kosiura, C. Kovar-smith, G.K. Laird, C. Langford, S. Lawlor, M. Leversha, L. Lewis, W. Liu, C. Lloyd, D.M. Lloyd, H. Loulseged, J.E. Loveland, J.D. Lovell, R. Lozado, J. Lu, R. Lyne, J. Ma, P. Meidl, T. Meitinger, S. Milne, G. Miner, S. L, G. Nordsiek, G. Nyakatura, C.N.O. Dell, D. Patel, A. V Pearce, D.M. Pearson, S.E. Pelan, D. Schlessinger, M.G. Schueler, H.K. Sehra, C. Shaw-, R.M. Swann, D. Swarbreck, P.E. Tabor, S. Taudien, T. Taylor, A.C. Tromans, M. Urso, D. Verduzco, and D. Villasana. 2005. The DNA sequence of the human X chromosome. Genome 434:325–337. doi:10.1038/nature03440.The.

Roughsedge, T., P.M. Visscher, and S. Brotherstone. 2000. Effects of cow families on production traits in dairy cattle. Anim. Sci. 71:49–57. doi:10.1017/S1357729800054886.

Rozas, J., A. Ferrer-Mata, J.C. Sanchez-DelBarrio, S. Guirao-Rico, P. Librado, S.E. Ramos-Onsins, and A. Sanchez-Gracia. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. 34:3299–3302. doi:10.1093/molbev/msx248.

Ruiz-Pesini, E., C. Diez, A.C. Lapena, A. Perez-Martos, J. Montoya, E. Alvarez, J. Arenas, and M.J. Lopez-Perez. 1998. Correlation of sperm motility with mitochondrial enzymatic activities. Clin. Chem. 44:1616–1620. doi:10.1093/clinchem/44.8.1616.

Ruiz-Pesini, E., D. Mishmar, M. Brandon, V. Procaccio, and D.C. Wallace. 2004. Effects of Purifying and Adaptive Selection on Regional Variation in Human mtDNA. Science (80-.). 303:223–226. doi:10.1126/science.1088434.

Sartelet, A., T. Druet, C. Michaux, C. Fasquelle, S. Géron, N. Tamma, Z. Zhang, W. Coppieters, M. Georges, and C. Charlier. 2012. A splice site variant in the bovine rnf11 gene compromises growth and regulation of the inflammatory response. PLoS Genet. 8:e1002581. doi:10.1371/journal.pgen.1002581.

Saxena, R., L.G. Brown, T. Hawkins, R.K. Alagappan, H. Skaletsky, M.P. Reeve, R. Reijo, S. Rozen, M.B. Dinulos, C.M. Disteche, and D.C. Page. 1996. The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. Nat. Genet. 14:292–299. doi:10.1038/ng1196-292.

Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123:218–223. doi:10.1111/j.1439-0388.2006.00595.x.

Scheu, A., A. Powell, R. Bollongino, J.D. Vigne, A. Tresset, C. Çakirlar, N. Benecke, and J. Burger. 2015a. The genetic prehistory of domesticated cattle from their origin to the spread across Europe. BMC Genet. 16:54. doi:10.1186/s12863-015-0203-2.

Scheu, A., A. Powell, R. Bollongino, J.D. Vigne, A. Tresset, C. Çakirlar, N. Benecke, and J. Burger. 2015b. The genetic prehistory of domesticated cattle from their origin to the spread across Europe. BMC Genet. 16:54. doi:10.1186/s12863-015-0203-2.

Schibler, J., J. Elsner, and A. Schlumbaum. 2014. Incorporation of aurochs into a cattle herd in Neolithic Europe: Single event or breeding?. Sci. Rep. 4:5798:8–13. doi:10.1038/srep05798.

Schnitzenlehner, S., and A. Essl. 1999. Field data analysis of cytoplasmic inheritance of dairy and fitness-related traits in cattle. Anim. Sci. 68:459–466. doi:10.1017/S1357729800050463.

Sellem, E., M.L.W.J. Broekhuijse, L. Chevrier, S. Camugli, E. Schmitt, L. Schibler, and E.P.C. Koenen. 2015. Use of combinations of in vitro quality assessments to predict fertility of bovine semen. Theriogenology 84:1447-1454.e5. doi:10.1016/j.theriogenology.2015.07.035.

Serre, D., A. Langaney, M. Chech, M. Teschler-Nicola, M. Paunovic, P. Mennecier, M. Hofreiter, G. Possnert, and S. Pääbo. 2004. No evidence of Neandertal mtDNA contribution to early modern humans. PLoS Biol. 2:313–317. doi:10.1371/journal.pbio.0020057.

Shannon, P., and R. Vishwanath. 1995. The effect of optimal and suboptimal concentrations of sperm on the fertility of fresh and frozen bovine semen and a theoretical model to explain the fertility differences. Anim. Reprod. Sci. 39:1–10. doi:10.1016/0378-4320(95)01376-B.

Shin, H.D., and L.H. Kim. 2005. Bos taurus mitochondrion, complete genome sequence of Korean cattle, beef cattle and Holstein-Friesian. GenBank DQ124371–D.

Shuster, D.E., M.E. Kehrli, M.R. Ackermann, and R.O. Gilbert. 1992. Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. Proc. Natl. Acad. Sci. U. S. A. 89:9225–9229. doi:10.1073/pnas.89.19.9225.

Skaletsky, H., T. Kuroda-Kawaguchl, P.J. Minx, H.S. Cordum, L.D. Hlllier, L.G. Brown, S. Repplng, T. Pyntikova, J. All, T. Blerl, A. Chinwalla, A. Delehaunty, H. Du, G. Fewell, L. Fulton, R. Fulton, T. Graves, S.F. Hou, P. Latrielle, S. Leonard, E. Mardis, R. Maupin, J. McPherson, T. Miner, W. Nash, C. Nguyen, P. Ozersky, K. Pepin, S. Rock, T. Rohlfing, K. Scott, B. Schultz, C. Strong, A. Tin-Wollam, S.P. Yang, R.H. Waterston, R.K. Wllson, S. Rozen, and D.C. Page. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825–837. doi:10.1038/nature01722.

Smith, K.D., K.E. Young, C.C. Talbot, and B.J. Schmeckpeper. 1987. Repeated DNA of the human Y chromosome. Development 101:77–92.

Song, G.J., and V. Lewis. 2008. Mitochondrial DNA integrity and copy number in sperm from infertile men. Fertil. Steril. 90:2238–2244. doi:10.1016/j.fertnstert.2007.10.059.

Spehar, M., M. Ferencakovic, V. Brajkovic, and I. Curik. 2017. Variance estimation of maternal lineage effect on milk traits in Croatian Holstein cattle. Agric. Conspec. Sci. 82:263–266.

Stälhammar, E.M., L. Janson, and J. Philipsson. 1989. Genetic studies on fertility in A.I. bulls. I. Age, season and genetic effects on semen characteristics in young bulls. Anim. Reprod. Sci. 19:1–17. doi:10.1016/0378-4320(89)90042-0.

Stevanovlć, M., R. Lovell-badge, J. Collignon, and P.N. Goodfellow. 1993. SOX3 is an X-linked gene related to SRY. Hum. Mol. Genet. 2:2013–2018. doi:10.1093/hmg/2.12.2013.

Su, G., B. Guldbrandtsen, G.P. Aamand, I. Strandén, and M.S. Lund. 2014. Genomic relationships based on X chromosome markers and accuracy of genomic predictions with and without X chromosome markers. Genet. Sel. Evol. 46:1–12. doi:10.1186/1297-9686-46-47.

Suchocki, T., and J. Szyda. 2015. Genome-wide association study for semen production traits in Holstein-Friesian bulls. J. Dairy Sci. 98:5774–5780. doi:10.3168/jds.2014-8951.

Sutarno, J.M. Cummins, J. Greeff, and A.J. Lymbery. 2002. Mitochondrial DNA polymorphisms and fertility in beef cattle. Theriogenology 57:1603–1610. doi:10.1016/S0093-691X(02)00664-7.

Svensson, E., and A. Götherström. 2008. Temporal fluctuations of Y-chromosomal variation in Bos taurus. Biol. Lett. 4:752–754. doi:10.1098/rsbl.2008.0342.

Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees.. Mol. Biol. Evol. 10. doi:10.1093/oxfordjournals.molbev.a040023.

Tanaka, M., J.S. Gong, J. Zhang, M. Yoneda, and K. Yagi. 1998. Mitochondrial genotype associated with longevity. Lancet 351:185–186. doi:10.1016/S0140-6736(05)78211-8.

Thompson, W.E., J. Ramalho-Santos, and P. Sutovsky. 2003. Ubiquitination of Prohibitin in Mammalian Sperm Mitochondria: Possible Roles in the Regulation of Mitochondrial Inheritance and Sperm Quality Control1. Biol. Reprod. 69:254–260. doi:10.1095/biolreprod.102.010975.

Thomsen, B., P. Horn, F. Panitz, E. Bendixen, A.H. Petersen, L.E. Holm, V.H. Nielsen, J.S. Agerholm, J. Arnbjerg, and C. Bendixen. 2006. A missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation. Genome Res. 16:97–105. doi:10.1101/gr.3690506.

Troy, C.S., D.E. MacHugh, J.F. Bailey, D.A. Magee, R.T. Loftus, P. Cunningham, A.T. Chamberlain, B.C. Sykes, and D.G. Bradley. 2001. Genetic evidence for Near-Eastern origins of European cattle. Nature 410:1088–1091. doi:10.1038/35074088.

True Nature Fondation. 2013. True Nature Foundation Starts International Project to Bring Back the Aurochs. Accessed January 3, 2020. https://truenaturefoundation.org/species-restoration/true-nature-foundation-starts-international-project-to-bring-back-the-aurochs/.

Tsai, T.S., S. Rajasekar, and J.C. St. John. 2016. The relationship between mitochondrial DNA haplotype and the reproductive capacity of domestic pigs (Sus scrofa domesticus). BMC Genet. 17:67. doi:10.1186/s12863-016-0375-4.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, and S.G. Rozen. 2012.

Primer3-new capabilities and interfaces. Nucleic Acids Res. 40. doi:10.1093/nar/gks596.

Upadhyay, M., C. Bortoluzzi, M. Barbato, P. Ajmone-Marsan, L. Colli, C. Ginja, T.S. Sonstegard, M. Bosse, J.A. Lenstra, M.A.M. Groenen, and R.P.M.A. Crooijmans. 2019. Deciphering the patterns of genetic admixture and diversity in southern European cattle using genome-wide SNPs. Evol. Appl. 12:951–963. doi:10.1111/eva.12770.

Upadhyay, M.R., W. Chen, J.A. Lenstra, C.R.J. Goderie, D.E. Machugh, S.D.E. Park, D.A. Magee, D. Matassino, F. Ciani, H.J. Megens, J.A.M. Van Arendonk, M.A.M. Groenen, P.A. Marsan, V. Balteanu, S. Dunner, J.F. Garcia, C. Ginja, and J. Kantanen. 2017. Genetic origin, admixture and population history of aurochs (Bos primigenius) and primitive European cattle. Heredity (Edinb). 118:169–176. doi:10.1038/hdy.2016.79.

Vallender, E.J., and B.T. Lahn. 2004. How mammalian sex chromosomes acquired their peculiar gene content. BioEssays 26:159–169. doi:10.1002/bies.10393.

VanRaden, P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor, and F.S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24. doi:10.3168/jds.2008-1514.

Verdugo, M.P., V.E. Mullin, A. Scheu, V. Mattiangeli, K.G. Daly, P. Maisano Delser, A.J. Hare, J. Burger, M.J. Collins, R. Kehati, P. Hesse, D. Fulton, E.W. Sauer, F.A. Mohaseb, H. Davoudi, R. Khazaeli, J. Lhuillier, C. Rapin, S. Ebrahimi, M. Khasanov, S.M.F. Vahidi, D.E. MacHugh, O. Ertuğrul, C. Koukouli-Chrysanthaki, A. Sampson, G. Kazantzis, I. Kontopoulos, J. Bulatovic, I. Stojanović, A. Mikdad, N. Benecke, J. Linstädter, M. Sablin, R. Bendrey, L. Gourichon, B.S. Arbuckle, M. Mashkour, D. Orton, L.K. Horwitz, M.D. Teasdale, and D.G. Bradley. 2019. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent.. Science (80-.). 365:173–176. doi:10.1126/science.aav1002.

Visscher, P.M., J.A. Woolliams, D. Smith, and J.L. Williams. 2002. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. J. Dairy Sci. 85:2368–2375. doi:10.3168/jds.S0022-0302(02)74317-8.

Wang, Z., X. Shen, B. Liu, J. Su, T. Yonezawa, Y. Yu, S. Guo, S.Y.W. Ho, C. Vila, M. Hasegawa, and J. Liu. 2016. Deep genetic divergences within domesticated yak are the result of quaternary climatic oscillations and orogeny on the Qinghai-Tibetan Plateau. GenBank GU256940.

Wei, Y.H. 1992. Mitochondrial DNA alterations as ageing-associated molecular events. Mutat. Res. DNAging 275:145–155. doi:10.1016/0921-8734(92)90019-L.

Weller, J.I., Y. Kashi, and M. Soller. 1990. Power of Daughter and Granddaughter Designs for Determining Linkage Between Marker Loci and Quantitative Trait Loci in Dairy Cattle. J. Dairy Sci. 73:2525–2537. doi:10.3168/jds.S0022-0302(90)78938-2.

Wiener, G., H. Jianlin, and L. Ruijun. 2003. The Yak Second Edition.

Wolff, J.N., and N.J. Gemmell. 2008. Lost in the zygote: The dilution of paternal mtDNA upon fertilization. Heredity (Edinb). 101:429–434. doi:10.1038/hdy.2008.74.

Yang, J., S.H. Lee, M.E. Goddard, and P.M. Visscher. 2011. GCTA: A tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88:76–82. doi:10.1016/j.ajhg.2010.11.011.

Yin, H., L. Fang, C. Qin, and S. Zhang. 2019. Estimation of the genetic parameters for semen traits in Chinese Holstein bulls. BMC Genet. 20:51. doi:10.1186/s12863-019-0752-x.

Yue, X.P., C. Dechow, and W.S. Liu. 2015. A limited number of Y chromosome lineages is present in North American Holsteins. J. Dairy Sci. 98:2738–2745. doi:10.3168/jds.2014-8601.

Zerjal, T., Y. Xue, G. Bertorelle, R.S. Wells, W. Bao, S. Zhu, R. Qamar, Q. Ayub, A. Mohyuddin, S. Fu, P. Li, N. Yuldasheva, R. Ruzibakiev, J. Xu, Q. Shu, R. Du, H. Yang, M.E. Hurles, E. Robinson, T. Gerelsaikhan, B. Dashnyam, S.Q. Mehdi, and C. Tyler-smith. 2003. The Genetic Legacy of the Mongols 2:717–721.

Zeyland, J., Wolko, J. Bocianowski, M. Szalata, R. Słomski, A.M. Dzieduszycki, M. Ryba, H. Przystałowska, and D. Lipiński. 2013. Complete mitochondrial genome of wild aurochs (Bos primigenius) reconstructed from ancient DNA. Pol. J. Vet. Sci. 16:265–273. doi:10.2478/pjvs-2013-0037.

Zhang, H., J.L.A. Paijmans, F. Chang, X. Wu, G. Chen, C. Lei, X. Yang, Z. Wei, D.G. Bradley, L. Orlando, T. O'Connor, and M. Hofreiter. 2013. Morphological and genetic evidence for early Holocene cattle management in northeastern China. Nat. Commun. 4. doi:10.1038/ncomms3755.

Zimin, A. V., A.L. Delcher, L. Florea, D.R. Kelley, M.C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C.P. Van Tassell, T.S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J.A. Yorke, and S.L. Salzberg. 2009. A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol. 10. doi:10.1186/gb-2009-10-4-r42.

ECOLE DOCTORALE N°581
Agriculture, alimentation, biologie, environnement et santé (ABIES)

Titre : Détection chez le bovin de polymorphismes génétiques au niveau du génome mitochondrial et des chromosomes sexuels et caractérisation de leurs effets sur les caractères de production, reproduction et santé.

Mots clés : génome mitochondrial, chromosome Y, chromosome X,

Résumé : Malgré leurs rôles dans l'expression de caractères de fertilité et dans le métabolisme énergétique, le génome mitochondrial et les chromosomes sexuels ne sont actuellement pas pris en compte dans les évaluations génomiques bovines françaises. Cette thèse a pour but d'étudier la variabilité génétique du génome mitochondrial et des chromosomes X et Y, de détecter des polymorphismes génétiques et de caractériser leurs effets sur les caractères de production, reproduction et santé. L'étude des schémas de transmission uniparentale, a mis en évidence une disparité entre une faible diversité des haplotypes du chromosome Y et un grand nombre de lignées mitochondriales dans de nombreuses races bovines. La présence à très faible fréquence de porteurs d'haplogroupes dont la divergence est antérieure à la domestication des bovins taurins a été identifié par génotypage. Deux études ont été réalisées pour estimer l'effet des différents variants identifiés sur le génome mitochondrial et le chromosome Y sur certains caractères d'intérêt zootechnique. Puis, une approche de génétique inverse a permis d'exploiter des données des séquences pour détecter sur le chromosome X des mutations candidates responsables d'anomalies et des mutations pouvant avoir un effet sur la fertilité mâle ou femelle ou entraînant un biais d'inactivation du chromosome X. En parallèle, le mécanisme génétique d'une anomalie liée au chromosome X responsable d'un cas de dysplasie ectodermique hypohidrotique a été décrit chez une génisse de race Holstein. Enfin, des pistes de réflexions sont proposées afin d'initier une meilleure prise en compte des chromosomes sexuels et du génome mitochondrial dans la sélection des bovins.

Title : Identification of genetic variants on the bovine mitochondrial genome and sex chromosomes and characterization of their effects on production, reproduction and health traits.

Keywords : mitochondrial genome, Y chromosome, X chromosome

Abstract : In spite of the roles of the mitochondrial genome and sex chromosomes in the expression of fertility traits and energy metabolism, currently they are not considered in French bovine genomic evaluations. This work is aiming to study the genetic variability of the mitochondrial genome and the X and Y chromosomes, to detect genetic polymorphisms, and to characterize their effects on production, reproduction and health traits. Analysis of uniparental transmission patterns revealed a discrepancy between the low diversity of Y-chromosome haplotypes and the large number of mitochondrial lines in many cattle breeds. Identification of carriers of haplogroups whose divergence predates the domestication of taurine cattle has been made at a very low frequency by genotyping. Initial studies have been carried out to estimate the effect of the variants identified on the mitochondrial genome and the Y chromosome on certain traits of interest to the breeding sector. Then, a reverse genetics approach has been applied to use sequence data to detect candidate mutations responsible for defects and mutations that may have an effect on male or female fertility or lead to an X chromosome inactivation bias. In parallel, the genetic mechanism of an X-linked defect responsible for a case of hypohidrotic ectodermal dysplasia was investigated in a Holstein heifer. In conclusion, several lines for future research are proposed to initiate a better consideration of sex chromosomes and the mitochondrial genome in the selection of cattle.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France