



HAL
open science

Recherche d'associations entre microARNs, variants génétiques et QTL laitiers chez les bovins, caprins et ovins

Céline Bourdon

► **To cite this version:**

Céline Bourdon. Recherche d'associations entre microARNs, variants génétiques et QTL laitiers chez les bovins, caprins et ovins. Génétique animale. Université Paris-Saclay, 2021. Français. NNT : 2021UPASL020 . tel-03186505

HAL Id: tel-03186505

<https://theses.hal.science/tel-03186505>

Submitted on 31 Mar 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.

Recherche d'associations entre microARNs,
variants génétiques et QTL laitiers chez les
bovins, caprins et ovins

*Association between microRNAs, genetic variants
and dairy QTL in bovine, caprine and ovine
species*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°577,
Structure et Dynamique des Systèmes Vivants (SDSV)
Spécialité de doctorat : Sciences de la vie et de la santé
Unité de recherche : Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France
Réfèrent : Faculté des sciences d'Orsay

Thèse présentée et soutenue à Paris-Saclay, le 15 Mars 2021, par

Céline BOURDON

Composition du Jury

Bernard Mignotte Professeur, Université Paris Saclay	Président
Juan José ARRANZ Professeur, Faculté Vétérinaire-Université de Leon	Rapporteur & Examineur
Christine GASPIN Directrice de recherche, INRAE-MIAT, Castanet-Tolosan	Rapporteur & Examinatrice
Rozenn Dalbies-Tran Chargée de recherche, INRAE-PRC, Nouzilly	Examinatrice
Christine Leroux Directrice de recherche, INRAE-Herbivores, Theix	Examinatrice

Direction de la thèse

Fabienne LE PROVOST Directrice de recherche, INRAE-GABI, Jouy-en-Josas	Directrice de thèse
Gwenola TOSSER-KLOPP Directrice de recherche, INRAE-GenPhySE, Castanet-Tolosan	Co-directrice de thèse

Remerciements

Je tiens tout d'abord à remercier l'ensemble des membres du jury : Juan José Arranz et Christine Gaspin, d'avoir accepté d'être rapporteurs de cette thèse, Rozenn Dalbès-Tran, Christine Leroux et Bernard Mignotte d'avoir accepté d'être examinateurs.

Je remercie également Apis-Gène et le département de génétique animale de INRAE d'avoir financé ma thèse, *via* le projet miRQTLait.

Merci à mes deux directrices de thèse : Fabienne Le Provost et Gwenola Tosser-Klopp pour leur encadrement, leur aide et leur soutien tout au long de ces années de thèse. Elles m'ont fait confiance en me choisissant pour effectuer cette thèse et m'ont accompagné tout au long de ce parcours, merci pour tout !

Merci à l'ensemble de l'équipe GaLac (anciennement GFP-GM) pour m'avoir accueilli au sein de l'équipe et pour leur bienveillance. Je pense aux membres actuels de l'équipe mais aussi aux personnes avec qui j'ai partagé un bureau, ne serait-ce pour quelques mois (Laura, Inès ou encore Thilali). A Antoine et Elitsa en particulier, avec qui j'ai instauré la tradition des goûters de 16h (ou un peu avant). D'ailleurs pour les goûters je peux également remercier Madia, Cathy, Céline, etc...

Je remercie Cathy, pour m'avoir accompagné et formé à la culture cellulaire mais aussi pour les discussions qu'on a pu avoir, professionnelles ou non.

Merci à Sandrine pour son temps et les réflexions scientifiques sur mon sujet de thèse.

Merci à Johann, pour son aide technique mais aussi pour les discussions qu'on a pu avoir et plus globalement pour m'avoir accepté en tant que voisine de bureau (avec mes chocolats et autres bonbons évidemment).

Je remercie également Céline, et désolée encore pour les tracas engendrés par les ddPCR : promis il n'y en a plus !

Nicolas et Etienne, merci pour m'avoir formé à la qPCR et aux Western-Blot respectivement. Nicolas également pour la partie informatique, et l'autre Nicolas pour les solutions de Tris notamment (avec des morceaux ça marche !).

Je n'ai passé que quelques demies-journées à Toulouse, mais je remercie toute l'équipe GenROC de m'avoir accueilli lors de réunions d'équipe, d'avoir pris du temps pour moi et d'avoir réalisé des génotypages pour mon projet de thèse.

De nombreuses personnes ont partagé leurs données et leurs compétences avec moi afin de permettre l'aboutissement de cette thèse, je les remercie toutes : l'équipe G2B et particulièrement Mekki Boussaha, pour son aide pour la partie bioinformatique et pour le partage de ses données, Marie-Pierre Sanchez, Thierry Tribout, Rachel Lefebvre et Didier Boichard ; mais aussi toujours au sein de GABI Denis Laloë pour les analyses statistiques. Je remercie également Rachel Rupp et Hélène Larroque. Merci à Philippe Bardou, pour son aide multiple sur la base de données notamment avec une grande disponibilité, et Cédric Cabau pour les analyses de RNAseq.

Je tiens à remercier l'unité expérimentale du Pin au Haras pour les prélèvements de lait et de sang. Merci également aux différentes personnes qui ont permis le transport de ces échantillons.

Je remercie l'ensemble des membres du comité de pilotage de miRQTLait, pour les réflexions et discussions issues des différentes réunions.

Merci également aux membres de mon comité de thèse : Carine Capel, Christine Gaspin et Sophie Rome.

Merci également à Louise, pour les pauses goûter. Tu as été d'un grand soutien alors même que tu finissais ta thèse. Merci pour tout, pour les moments tard parfois au labo, pour nos petites sorties à V2 ou au zoo (et bonne continuation là-bas loin à la Réunion !). Par la suite c'est Marine et Amandine qui ont pris le pli des pauses goûter, merci également à vous deux.

A Patricia, merci beaucoup pour tout, pour ton aide en tant que gestionnaire mais aussi (et surtout) pour ta bienveillance et ton écoute, pour les moments passés dans ton bureau à parler de tout et rien.

Aux gens de Doc'J, anciens et nouveaux, merci pour tous les bons moments que j'ai pu passer avec vous tous. Audrey, Anna-Charlotte, Jonathan (tous les trois vous avez été de vrais soutiens pendant la thèse, on s'est suivi dans nos galères et nos bons moments), Mathieu pour Doc'Jeux (promis la prochaine fois je ne jouerai pas au Code Names), Marie, Sébastien, Gabriel... Et maintenant Katy (toujours de bonne humeur), Tristan, Alicia... Bon courage aux nouveaux membres de Doc'J, profitez de cette expérience au top !

Dans le cadre de cette thèse j'ai eu la chance de participer à une formation à Paimpont : merci à tous les gens rencontrés là-bas, c'était une super semaine. Et ça s'est poursuivi avec quelques sorties en forêt très sympa également !

Merci à toute la bande "Sacré libanais" : Anaïs, Déborah, Abarna, Bertrand, Mayrone, Emy, Héloïse. Je pense que vous n'imaginez pas à quel point vous avez été et êtes toujours très importants pour moi. La bonne humeur de ce groupe, chacun avec nos spécificité (les chansons d'Anaïs, les gaufres de Débo, les papillons et l'irish coffee d'Abarna, la cave du professeur Bertrand, l'alter-ego chat de Mayrone, la maladresse d'Emy, les chats fous d'Héloïse... et j'en passe !) sans oublier votre bienveillance, gentillesse, grand cœur, humour, etc, etc, etc.... Merci merci merci à vous !

Table des matières

Liste des abbréviations.....	17
Table des illustrations	25
Introduction.....	31
1. Contexte socio-économique	31
a. Agriculture en France	31
b. Production laitière.....	31
2. Composition du lait	35
a. Impact de la génétique.....	39
b. Impact de la physiologie.....	39
c. Impact de l'environnement.....	41
d. Impact de l'alimentation	45
e. Impact de la santé	47
3. Les différentes races présentes en France.....	47
a. Races non majoritaires.....	49
i. Bovins laitiers	49
ii. Caprins laitiers.....	51
iii. Ovins laitiers	51
b. Les races majoritaires.....	53
i. Bovins laitiers	53
1. Holstein	53
2. Normande.....	53
3. Montbéliarde.....	55

ii.	Caprins laitiers	55
1.	Alpine.....	55
2.	Saanen	55
iii.	Ovins laitiers	57
1.	Lacaune.....	57
2.	Manech Tête Noire.....	57
4.	Génétique.....	57
a.	Etat des lieux : connaissances des génomes des ruminants.....	57
b.	Les variants génétiques	63
i.	SNPs.....	65
ii.	Petites insertions et délétions.....	65
iii.	Autres polymorphismes	65
c.	Quantitative Trait Locus	67
i.	Définition.....	67
ii.	Détermination de QTL.....	69
5.	La sélection génomique	75
a.	Principes et but de la sélection génomique	75
b.	Historique de la sélection.....	75
i.	Sélection génétique, évaluation sur descendance.....	75
ii.	Sélection génomique.....	77
1.	Mise en place de la sélection génomique en bovin	79
2.	Mise en place de la sélection génomique en ovin et caprin	81
6.	Les différentes régions du génome.....	83
a.	Les régions codantes	83

b.	Les régions non codantes	83
i.	Les microARNs.....	89
1.	microARNs et espèces animales.....	91
2.	Bases de données existantes	93
3.	Biogenèse de microARNs	95
4.	Mode d'action	95
ii.	Variants de microARNs.....	97
1.	SNPs affectant l'expression du microARN.....	99
2.	SNPs situés dans la région seed	101
iii.	Lien microARNs et production laitière	101
Résultats		109
1.	Création d'une base de données de microARNs de ruminants	109
2.	Détection de variants génétiques d'intérêt en bovin, caprin et ovin	137
3.	Validation de variants de microARNs détectés.....	173
a.	Contexte	173
b.	Matériel et méthodes	173
i.	Validation de variants génétiques via puces à SNPs	173
ii.	Validation de variants génétiques via études de GWAS	177
c.	Résultats.....	177
i.	Ajout de variants sur puces à SNPs	177
ii.	Validation par études de GWAS	177
d.	Conclusion	179
4.	Etudes des potentiels impacts de variants de microARNs validés.....	181
a.	Contexte	181

b.	Etude des impacts potentiels du variant <i>bta-let-7e</i>	183
i.	Matériel et méthodes	183
1.	Culture cellulaire	183
2.	Prédiction d'ARNm cibles	185
3.	Transfection.....	187
4.	Extraction d'ARN	189
5.	Reverse transcription et PCR quantitative	189
6.	Séquençage et traitement des données	191
7.	Analyses statistiques	191
ii.	Résultats	193
1.	Prédiction d'ARNm cibles	193
2.	Expression de cibles prédites	193
3.	Séquençage et traitement des données	195
c.	Test de l'intégrité de microARNs dans le lait à température ambiante et à 4°C... 197	
d.	Etude des impacts potentiels des variants <i>bta-miR-486</i> et <i>bta-miR-92b</i>	201
i.	Matériel et méthodes	201
1.	Échantillons	201
2.	Génotypage des vaches.....	201
3.	Extraction d'ARN	203
4.	Reverse transcription et PCR quantitative	203
5.	Analyses statistiques	207
ii.	Résultats	207
1.	Génotypage des vaches.....	207
2.	Analyse de la présence des microARNs dans le lait	209

Discussion et perspectives	211
1. Détection de variants	213
2. Validation de variants.....	217
3. Tests des potentiels impacts de variants validés	219
a. <i>bta-let-7e</i>	219
i. BME-UV1	219
ii. Effets détectés via l'approche RT-qPCR	221
iii. Effets détectés via l'approche RNAseq	223
iv. Comparaison RT-qPCR et RNAseq	223
v. Le microARN <i>bta-let-7e</i>	225
vi. Perspectives.....	227
b. Conservation des microARNs dans le lait.....	231
c. <i>bta-miR-486</i> et <i>bta-miR-92b</i>	231
i. Échantillons	231
ii. Techniques utilisées pour tester les effets potentiels	233
iii. Perspectives.....	235
d. Perspectives pour le test de potentiels impacts de variants caprins et ovins	237
4. Perspectives générales	239
Bibliographie	243

Liste des abréviations

A2M	Alpha-2-macroglobulin
ACP	Analyse en Composant Principale
ADN	Acide DésoxyriboNucléique
Ago	Argonaute
AKT	AKT serine/threonine kinase 1
AMPKγ1	Adenosine Mono Phosphate protein Kinase gamma 1
AOC	Appellation d'Origine Contrôlée
AOP	Appellation d'Origine Protégée
ARN	Acide Ribo-Nucléique
ARNm	ARN messenger
ARNr	ARN ribosomaux
ARNt	ARN de transfert
B4GALNT2	beta-1,4-N-acetyl-galactosaminyltransferase 2
BME	Bovine Mammary Epithelial cells (cellules épithéliales mammaires bovines)
BMP15	Bone Morphogenetic Protein 15
BRCA1/BRCA2	BRCA1 DNA repair associated/BRCA2 DNA repair associated
Ca	Calcium
CDS	Coding Domain Sequence (région codante)
CEL	Cellules somatiques dans le lait
CHRF	Cardiac Hypertrophy Related Factor
CNV	Copy Number Variation (variation du nombre de copies)
CSN1S1	Casein alpha s1
CSN2	Casein beta
Ct	Cycle threshold (cycles seuil)
ddPCR	Droplet Digital PCR (PCR en gouttelettes)
DGAT1	Diacylglycerol O-Acyltransferase 1

DGCR8	DiGeorge Critical Region 8
DMEM	Dulbecco's Modified Eagle Medium
DNMT3A/DNMT3B	DNA methyltransferase 3 alpha/DNA methyltransferase 3 beta
EEF2K	Eukaryotic Elongation Factor 2 Kinase
EIF4G2	Eukaryotic translation Initiation Factor 4 Gamma 2
ELOVL6	Elongation of long-chain fatty acids family member 6
FAD2	Fatty acid desaturase 2
FAETH	Functional-And-Evolutionary Trait Heritability
FASN	Fatty Acid Synthase
G2B	Génétique et Génomique Bovine
GDD	Grand-Daughter Design (protocole "petites filles")
GWAS	Genome Wide Association Study (études d'associations génome entier)
HK2	Hexokinase 2
IL10	Interleukin 10
InDel	Insertion et Délétion
INEL	Index Economique Laitier
INRAE	Institut National de Recherche pour l'Agriculture, l'alimentation et l'Environnement
ISU	Index de synthèse
K	Potassium
Kb	Kilobase
KO	Knock-Out
lincRNA	Large Intergenic Non-Coding RNA (grand ARN non codant intergénique)
lncRNA	Long NonCoding RNA (ARN long non codant)
MACL	Mammite
MAC-T	bovine Mammary Alveolar Cells

MAF	Minor Allele Frequency (fréquence de l'allèle minoritaire)
Mg	Magnésium
MP	Matière Protéique
MTI	microRNA-Target Interactions (interactions microARNs-cibles)
mTOR	Mechanistic target of rapamycin kinase
Na	Sodium
NCBI	National Center for Biotechnology Information
NCF4	Neutrophil Cytosolic Factor 4
OLAH	Oleoyl-ACP hydrolase
OLR1	Oxidized low density lipoprotein receptor 1
ONU	Organisation des Nations Unies
OSTF1	Osteoclast Stimulating Factor 1
P	Phosphore
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PFA	ParaFormAldehyde
piARN	piwiARN (P-element-induced wimpy testis interacting -RNA)
PPIA	Peptidylprolyl isomerase A (cyclophilin A)
PRLR	Prolactin receptor
PrP	Prion Protein
PTEN	Phosphatase and tensin homolog
QTL	Quantitative Trait Loci
RFLP	Restriction Fragment Length Polymorphism (polymorphisme de longueur des fragments de restriction)
RIN	RNA Integrity Number (facteur d'intégrité de l'ARN)
RISC	RNA-Induced Silencing Complex (complexe d'inhibition induit par ARN)
Rnase III	Ribonucléase 3

<i>RNAseq</i>	Séquençage d'ARN
<i>RT</i>	Reverse Transcription (transcription inverse)
<i>SAM</i>	Sélection Assistée par Marqueurs
<i>SCD</i>	Stearoyl-CoA desaturase
<i>siARN</i>	Small Interfering RNA (petit ARN interférent)
<i>SLC25A24</i>	Solute Carrier Family 25 Member 24
<i>snoRNA</i>	Small nucleolar RNA (petit ARN nucléolaire)
<i>SNP</i>	Single Nucleotide Polymorphism (polymorphisme d'un seul nucléotide)
<i>snRNA</i>	Small nuclear RNA (petit ARN nucléaire)
<i>SO</i>	Sequence Ontology (ontologie de séquence)
<i>spectre MIR</i>	Spectre Moyen InfraRouge
<i>STAT5</i>	Signal Transducer and Activator of Transcription 5
<i>STAT5B</i>	Signal Transducer and Activator of Transcription 5B
<i>SVF</i>	Sérum de Veau Fœtal
<i>TB</i>	Taux butyreux
<i>THI</i>	Temperature and Humidity Index (index de température et humidité)
<i>TP</i>	Taux protéique
<i>TPM</i>	Transcrits Par Millions
<i>TSS</i>	Transcription Start Site (site de démarrage de la transcription)
<i>TWF1</i>	Twinfilin actin binding protein 1
<i>UCSC</i>	University of California Santa Cruz
<i>UE</i>	Unité Expérimentale
<i>URCEO</i>	Union Régionale des Coopératives d'Élevage de l'Ouest
<i>UTR</i>	Untranslated region (région non traduite)
<i>VCF</i>	Variant Call Format
<i>WGS</i>	Whole Genome Sequencing (séquençage du génome entier)

Table des illustrations

Figure 1 : Bovins de race Holstein (crédit photo : Alger MEEKMA [87]), Normande [88], et Montbéliarde [89] ; Caprins de race Alpine [90] et Saanen [91] ; Ovins de race Lacaune [92] et Manech Tête Noire [93]	52
Figure 2 : Proportion de races bovines dans le cheptel de vaches laitières en France	52
Figure 3 : Chromosomes bovins avec contigs assemblés. Chaque changement de couleur correspond à un changement d'un contig à un autre. La partie gauche de chaque chromosome représente les contigs de la version UMD3.1.1 et la partie droite représentent ceux de la version ARS [113]	58
Figure 4 : Localisation des différents types de variants génétiques caractérisés par leur probable impact, d'après Ensembl [130]	62
Figure 5 : Exemple de SNP et de petite insertion.....	64
Figure 6 : Représentation de l'analyse de liaison.....	68
Figure 7 : Proportion des différents caractères de QTL chez les ovins et bovins, présents dans la base de données AnimalQTLdb [152].....	70
Figure 8 : Couverture d'un catalogue URCEO de 1977 et exemple de description d'un animal en tant que père potentiel	74
Figure 9 : Chronologie de la sélection génomique bovine en France [95], [163]	76
Figure 10 : Régulation de la transcription de gènes par répresseur, activateur et promoteur	84
Figure 11 : Différentes classes d'ARN.....	86
Figure 12 : Structure en tige-boucle du précurseur d'un microARN (ici bta-let-7e). Les nucléotides orange composent le microARN mature, la région "seed" est composée des nucléotides 2 à 8 du mature, ici encadré en orange.....	88
Figure 13 : Biogenèse des microARNs	94
Figure 14 : Impacts potentiels de mutations de microARNs dans le précurseur.....	98
Figure 15 : Impacts potentiels de mutations de microARNs dans la région "seed"	100
Figure 16: Précurseur du microARN bta-let-7e, avec le microARN mature en marron. Le variant génétique détecté est encadré en rouge	180

Figure 17 : Résultats relatifs du niveau d'expression de OSTF1 et SLC25A24 détecté en BME-UV1 selon l'allèle présent dans le microARN bta-let-7e transfecté.....	192
Figure 18 : Résultats des analyses statistiques des données de RNAseq	194
Figure 19 : Concentration des microARNs bta-miR-92a et bta-miR-99a selon le temps de conservation des laits à 4°C et à température ambiante. Les diminutions moyennes significatives sont annotées avec un astérisque ($p < 0,05$, tests statistiques Wilcoxon).	198
Figure 20 : Concentration des microARNs selon les temps et la méthode de conservation.	198
Figure 21 : Gel de typage bta-miR-92b	206
Figure 22 : Présence du microARN bta-miR-92b dans le lait de vaches selon leur génotypage pour le variant d'intérêt	208
Figure 23 : Présence du microARN bta-miR-486 dans le lait de vaches selon leur génotypage pour le variant d'intérêt	208
Figure 24 : Visualisation des transcrits gènes étudiés dans le cadre des impacts du variant bta-let-7e	222
Figure 25 : Visualisation de la position des amorces, dessinées pour étudier l'expression d'OSTF1, dans chacun des transcrits décrits du gène (OSTF1-201 en vert et OSTF1-202 en rose).	222

Tableau 1 : Statut des personnes ayant un emploi permanent en exploitations agricoles, en France, d'après [1].....	30
Tableau 2 : Tableau des teneurs en différentes protéines des laits bovins, caprins et ovins	36
Tableau 3 : Caractères de production laitière des trois principales races laitières françaises. Les vaches Holstein ont des productions de lait par an plus importantes que les deux autres races. Les Normande présentent des taux protéiques et butyreux plus élevés	54
Tableau 4 : Liste de caractères de QTL laitiers étudiés par [153]–[156] et Larroque et al. (communication personnelle)	70
Tableau 5 : Synthèse de résultats de liens entre microARNs et production laitière	100
Tableau 6 : Caractéristiques des trois variants génétiques validés en bovin grâce aux données de GWAS, pour des critères de taux butyreux (tb), taux protéique (tp) et de matière protéique (mp).	176
Tableau 7 : Caractéristique du variant génétique situé dans la région "seed" du microARN bta-let-7e.	180
Tableau 8 : Caractéristiques des variants génétiques impactant les microARNs bta-miR-92b et bta-miR-486 et détectés dans des caractères QTL de taux protéique (tp) et taux butyreux (tb) respectivement.	182
Tableau 9 : Séquences des amorces utilisées en qPCR pour les 4 ARNm testés, l'ARNm contrôle positif et le gène de référence	188
Tableau 10 : Taux moyen de diminution de la concentration des microARNs miR-92a et miR-99a entre les temps 0h et 18h de conservation à 4°C ou à température ambiante (TA).....	198
Tableau 11 : Nombre de vaches par génotypage pour le variant génétique impactant bta-miR-92b.....	206
Tableau 12 : Nombre de vaches par génotypage pour le variant génétique impactant bta-miR-486.....	206
Tableau 13 : Séquences nucléotidiques des microARNs de la famille des microARNs bovins let-7.....	224

<i>Statut des agriculteurs/trices</i>	<i>Nombre de personnes (en milliers)</i>
Chefs d'exploitation, coexploitants et associés actifs	564,1
Actifs apparentés au chef d'une exploitation individuelle	104,6
Salariés permanents	142,8
Total des actifs permanents	811,5

Tableau 1 : Statut des personnes ayant un emploi permanent en exploitations agricoles, en France, d'après [1].

Introduction

1. Contexte socio-économique

a. Agriculture en France

En France, l'agriculture et notamment l'élevage tiennent une place prédominante. La France est, en effet, la première puissance agricole européenne, tant sur le plan économique avec 75 milliards d'euros de production agricole (en prenant en compte le prix perçu par le producteur) en 2019, que sur le plan géographique avec sa surface agricole s'étendant sur plus de 28 millions d'hectares de terre. Ces données proviennent des chiffres de Eurostat, la direction générale de la Commission européenne chargée de l'information statistique à l'échelle communautaire [2] et de la fondation de l'ONU (Organisation des Nations Unies) pour l'alimentation et l'agriculture (FAO) [3]. La France est également le pays d'Europe possédant le plus de têtes de bétail, avec plus de 18 millions de bovins en 2019 [2]. L'importance du secteur agricole en France se démontre également *via* l'emploi : plus de 800 000 personnes en France ont un emploi permanent dans une exploitation agricole [4]. Plus de la moitié de ces personnes sont chefs d'exploitation (*Tableau 1*) [1].

b. Production laitière

À l'échelle mondiale, les vaches, chèvres et brebis représentent 86% de la production laitière [5]. Pour les différents cheptels bovins, caprins et ovins, plus de 20% de l'ensemble des animaux sont consacrés à la production laitière : 22,30% des bovins, 20,24% des caprins et 21,31% des ovins [6]–[8].

En France, une large proportion de l'agriculture se consacre à la production laitière : le cheptel des trois espèces bovine, caprine et ovine s'élève à plus de 27 millions d'animaux, dont environ 70% de bovins [6]. Plus de 40% des bovins élevés en France sont des vaches, et près de la moitié d'entre elles sont assignées à la production de lait [9].

La France est le deuxième pays d'Europe en terme de collecte de lait : grâce aux 3,7 millions de vaches laitières, plus de 24 milliards de litres de lait de vache ont été récoltés en 2018, juste derrière l'Allemagne qui collecte 32 milliards de litres de lait [2], [9]. Ces productions de lait bovin sont en majorité orientées vers le lait conditionné [10].

Les productions de lait caprin et ovin sont moins importantes et majoritairement orientées vers la transformation du lait en fromage : les 850 000 chèvres élevées en France produisent 580 millions de litres de lait de chèvre par an et 260 millions de litres de lait ovin sont produits par 1,2 millions de brebis [11]. La France a également une place prédominante dans la production de lait de chèvre et de brebis : il s'agit du premier producteur européen de lait caprin et du cinquième producteur européen de lait ovin. Au total, pour ces trois espèces, près de 25 milliards de litres de lait sont collectés par an en France [11]. La production de lait en France a une importance économique puisqu'environ 298 000 emplois en dépendent, que ce soit en exploitation laitière ou dans l'industrie du lait [12].

En caprin laitier, environ la moitié (49%) des exploitations ont plus de 100 animaux, en ovin laitier la majorité des exploitations ont entre 200 et 499 brebis [13], [14]. En bovin laitier, plus de la moitié des exploitations comptent 50 vaches et plus en 2019. Ces chiffres peuvent être mis en relation avec des chiffres de 1969, quand plus de la moitié des exploitations comptaient moins de 9 vaches en France [9], [15]. L'intensification de l'élevage est également présente en ovin, avec un nombre d'exploitations de brebis laitières ayant diminué d'environ 25% entre 2000 et 2017, alors que le nombre de brebis laitières est lui resté stable entre ces années [14]. L'intensification est encore plus visible en caprin laitier, avec un nombre d'exploitations qui a diminué de plus de 50% alors que le nombre de chèvres est quasiment stable entre 2000 et 2018 [9], [14], [16].

En 2018, des exploitations dites de petite taille subsistent encore, avec 22% des exploitations bovines laitières de plus de 5 vaches comptant moins de 30 vaches. De même pour les caprins et ovins, avec respectivement 28% des exploitations de plus de 10 animaux comptant moins de 50 animaux, et 13% des exploitations ovines laitières ayant moins de 20 brebis [9], [13], [14].

Les bovins, caprins et ovins sont trois espèces laitières avec des caractéristiques différentes en terme de production de lait. Tout d'abord, la durée de leur lactation diffère : en moyenne 334 jours pour les bovins, 319 jours pour les caprins et 167 jours pour les ovins [17].

2. Composition du lait

Le lait est un fluide complexe, avec une composition similaire pour les espèces bovine, caprine et ovine, mais avec des proportions de chacun des éléments qui diffèrent. Le lait est composé pour majeure partie d'eau. Il contient également plus de 100 000 types de molécules différentes. Parmi les grandes catégories de constituants se trouvent des matières grasses, des glucides, principalement du lactose, ainsi que des protéines [18]. Des minéraux sont également présents dans les laits bovin, caprin et ovin et parmi eux le calcium (Ca), potassium (K), magnésium (Mg), phosphore (P) et sodium (Na) [19]–[22].

Les matières grasses du lait se composent de plus de 400 acides gras différents, dont 15 à 16 acides gras principaux [23]. Ceux-ci peuvent être classés en 3 principales catégories : les acides gras saturés (sans double liaison dans la chaîne de carbones), les acides gras monoinsaturés (avec une double liaison, entre deux carbones) et les polyinsaturés (avec plusieurs doubles liaisons). Dans le lait bovin, les acides gras saturés représentent 70% des acides gras présents, les monoinsaturés représentent 25% et les polyinsaturés 5% des acides gras présents dans le lait [24].

Les acides gras à chaîne longue (C16, C18) sont issus de l'alimentation, alors que les acides gras à chaîne courte et moyenne sont synthétisés dans la glande mammaire [25].

En caprin, en ovin, comme en bovin, les acides gras les plus présents sont les C16:0 (acide palmitique) et C18:1 (acide oléique) [26]–[28]. Ils correspondent à environ 22g, pour les C16:0, et 19g, pour les C18:1, pour 100g d'acides gras totaux chez les ovins [27]. Dans le lait bovin, les C16:0 représentent 23% des acides gras et les C18:1 représentent 29% [28], [29]. En caprin ces taux s'élèvent respectivement à 24 et 17% [26], [30].

	Bovin	Caprin	Ovin	Références
Taux Protéiques	3,2 %	3,2 %	6,2 %	[32]–[35]
Taux Butyreux	3,7 %	3,7 %	7,6 %	[32]–[35]
C16:0 (en g/100 g d'acides gras)	23	24	22	[26]–[30]
C18:1 (en g/100 g d'acides gras)	29	17	19	[26]–[30]
α -lactalbumine	2 mg/mL	3 mg/mL	3 mg/mL	[36], [37]
β -lactoglobuline	3 mg/mL	3 mg/mL	7 mg/mL	[36], [37]
Caséines	2,5 %	2,5 %	4-5 %	[33], [34]
Lactose	4-5 %	4-5 %	4-5 %	[32]–[34]

Tableau 2 : Tableau des teneurs en différentes protéines des laits bovins, caprins et ovins

Les protéines majeures du lait sont les différentes caséines : caséines α_1 , α_2 , β et κ , ainsi que la β -lactoglobuline et l' α -lactalbumine [31].

Les taux protéiques et butyreux correspondent aux teneurs en protéines et en matières grasses dans le lait, respectivement. Les laits bovin et caprin sont comparables du point de vue des taux protéiques (3,2% (32 grammes par kilogramme de lait)) et butyreux (environ 3,7%), ce qui les différencie du lait ovin. [17], [32]–[35]. *A contrario*, le lait ovin a des taux protéiques et butyreux plus élevés : respectivement 6,2% et 7,6% environ [33]–[35] (*Tableau 2*).

L' α -lactalbumine et la β -lactoglobuline, deux protéines majeures du lait, sont présentes dans le lait bovin à hauteur de 2 et 3 mg/mL respectivement [36]. Les taux caprins et ovins pour l' α -lactalbumine sont eux estimés à 3 mg/mL, et les taux de β -lactoglobuline sont du même ordre de grandeur entre bovin et caprin, avec 3,16 mg/mL en caprin. Le taux de cette protéine dans le lait ovin est plus élevé : plus de 7 mg/mL [37].

Les taux de caséines sont eux aussi dissemblables entre les espèces, avec des laits bovins et caprins présentant des taux aux alentours de 2,5% et le lait ovin dans lequel ce taux s'élève à 4 à 5% [33], [34].

Une similitude de ces trois laits réside dans le taux de lactose, compris entre 4 et 5% pour ces trois espèces [32]–[34] (*Tableau 2*).

La méthode de spectrométrie moyen infrarouge (ou spectre MIR) est désormais utilisée afin d'analyser les laits et d'en conclure les composants, grâce à des équations permettant de relier les longueurs d'ondes aux différents composants du lait. Des équations de prédiction ont été développées, afin de déterminer les teneurs du lait en caséines (les 4 caséines majeures) et les protéines majeures citées précédemment : l' α -lactalbumine et la β -lactoglobuline [38], [39].

Les éléments tels que la saisonnalité, *via* l'alimentation notamment, ou encore la parité impactent la composition du lait : pour exemple les taux protéiques tendent à être plus importants lors de l'hiver [40].

a. Impact de la génétique

La génétique est un facteur clé des variations de la production laitière. La race des animaux peut influencer la composition du lait. Pour exemple, Stocco *et al.* ont mis en évidence des différences significatives des taux protéiques et de matières grasses dans le lait entre les races Holstein Friesians et Brown Swiss. Le taux protéique est 7% plus élevé en race Brown Swiss qu'en race Holstein Friesian. Le taux de matière grasse est lui plus élevé de 8% en Brown Swiss qu'en Holstein Friesian [41]. Des polymorphismes génétiques présents à des locus de gènes peuvent également moduler la composition du lait. C'est le cas de la substitution K232A dans l'exon VIII de *DGAT1*, qui code une enzyme impliquée dans le métabolisme des acides gras du lait : les vaches avec le génotype KK présentent des teneurs en matières grasses et en protéines plus élevées que les teneurs mesurées chez des vaches présentant le génotype KA ou AA [42].

b. Impact de la physiologie

Un facteur physiologique tel que le stade de lactation module la composition du lait produit. Dès 1924, une publication a décrit l'influence du stade de lactation sur la production de lait chez les vaches, avec des données de 4 races différentes : Ayrshire, Guernsey, Holstein et Jersey [43]. Plus précisément, des diminutions de la production et de la matière grasse du lait sont observées entre le premier mois de lactation et la fin de lactation [43]. Plus récemment, Stoop *et al.* ont eux aussi démontré l'influence du stade de lactation sur la composition du lait en bovin : le taux de matières grasses dans le lait augmente entre les jours 100 et 300 de lactation, passant de 4,24% à 5,02% [44], [45].

Des études en espèce ovine ont également été menées. Une étude de 2008 a montré des pourcentages de matières grasses et de protéines dans le lait qui sont significativement augmentés entre le 33^{ème} et le 191^{ème} jour de lactation, tout comme les caséines. La quantité de lait produit par jour est diminuée entre ces deux temps de lactation [46].

Les résultats en caprins sont similaires à ceux en ovin concernant la quantité de lait produit par jour, qui est là aussi diminuée au fur et à mesure de la lactation d'après l'étude de El-Tarabany *et al.* par exemple. Les taux de protéines et de matières grasses ne sont eux pas significativement diminués [47].

c. Impact de l'environnement

Le changement climatique peut également avoir un impact sur la composition du lait. Une étude a été menée sur 7 ans, en Italie, sur plus de 300 000 vaches de race Holstein. L'index THI a été utilisé pour définir les conditions climatiques : ce "Temperature and Humidity Index" prend en compte les températures minimale et maximale, ainsi que l'humidité relative minimale et maximale de chaque jour. La température maximale et humidité relative minimale sont utilisées pour calculer un index "maxTHI" et la température minimale et humidité relative maximale pour le "minTHI".

Les taux de matières grasses et de protéines dans le lait sont diminués significativement à partir d'une valeur de l'index maxTHI de 50 et 65, respectivement, calculé 2 jours avant l'échantillonnage du lait. Le taux de matières grasses est d'environ 4% dans le lait pour des maxTHI de 35 à 50, après cela il diminue significativement en fonction du maxTHI, jusqu'à atteindre environ 3,7% pour un maxTHI de 85. Le taux protéique est lui d'environ 3,45% d'un maxTHI de 35 à 65, puis diminue significativement jusqu'à atteindre 3,25% pour un maxTHI de 85. Pour le minTHI le point de rupture est situé à 38,0 et 39,4 respectivement pour ces taux de matières grasses et de protéines [48]. Bouraoui *et al.* on également mis en évidence la diminution significative des taux de matières grasses et de protéines lors de l'augmentation du THI : des THI de 68 et de 78 sont comparés, décrits comme correspondant à un climat printannier pour le premier et estival pour le second. Les vaches de race Friesan-Holstein sont élevées en Tunisie. Le taux de matières grasses est de 3,58% pour le THI 68 et de 3,24% pour le THI 78. Pour le taux protéique, de 2,96% pour un THI 68, le taux diminue significativement à 2,88% pour un THI 78 [49]. Une étude sur 142 bovins de race Jersey et 586 de race Holstein a également montré l'influence de la condition climatique sur la composition du lait.

Dans cette étude, un stress thermique est défini par un THI supérieur ou égal à 72. Le stress thermique est alors divisé en stress thermique moyen (THI entre 78 et 79), modéré (THI entre 79 et 90) et enfin sévère (supérieur à 90). Les taux de matières grasses et de protéines dans le lait diminuent significativement lors d'un stress thermique en Holstein, en race Jersey seul le taux de matières grasses est diminué significativement [50].

Des études ont également été menées sur des caprins, là aussi en utilisant cet index THI de température et humidité. Kljajevic *et al.* ont étudié la composition du lait d'environ 250 chèvres Saanen, sur une durée de 4 ans (2012 à 2015). Les teneurs en matières grasses et en protéines du lait sont significativement corrélées négativement avec le THI, tout comme la teneur en lactose du lait. Ces corrélations s'élèvent à -0,93 entre THI et teneur en matières grasses dans le lait, -0,72 entre THI et teneur en protéines du lait et -0,76 entre THI et teneur en lactose du lait [51]. Une étude chez les chèvres laitières de race Murciano-granadina a mis en évidence une diminution significative des teneurs en matières grasses et protéines dans le lait de respectivement 13% et 12% lors de stress thermique, défini dans cette étude par 12h à un THI de 86 suivies de 12h à un THI de 77, par rapport à une neutralité thermique (THI entre 59 et 65) [52].

En espèce ovine également, l'impact des conditions climatiques (toujours estimées par le THI) sur les taux de matières grasses et protéines du lait a été étudié. Dans une étude menée à partir des données de 4 428 ovins de race Valle del Belice entre 1994 et 2003, aucune différence significative du taux de matières grasses ni du taux protéique du lait n'est apparue entre un état contrôle et un stress thermique [53]. Peana *et al.* n'ont également vu aucune différence significative de composition du lait (en terme de taux de matières grasses et protéines) chez des brebis de race Sarda, dans le cas de stress thermique (THI supérieur à 73) [54].

d. Impact de l'alimentation

L'alimentation joue également un rôle important dans la composition du lait. De nombreuses études chez ces trois espèces laitières (bovin, caprin, ovin) ont été menées afin d'évaluer l'impact de l'alimentation, par le biais de restrictions ou des suppléments alimentaires. Une revue recoupe les différents éléments disponibles dans la littérature à ce sujet (Leduc *et al.*, soumis).

En bovin, des effets de suppléments alimentaires sur la composition du lait ont été démontrés. Des études ont, par exemple, montré une baisse de la concentration en matières grasses du lait dans le cas d'un régime faible de fibres alimentaires [55]. D'autres études se sont basées sur une substitution d'une alimentation en son de maïs par des grains de maïs séchés, induisant une réduction des taux protéiques dans le lait [56]. Des suppléments de l'alimentation de vaches Holstein en *Moringa oleifera* (une plante) ont, elles, induit une augmentation de la concentration en matières grasses du lait. Les concentrations protéiques n'ont pas été impactées par ce traitement [57]. Le taux de lactose dans le lait peut être significativement diminué dans le lait en cas de jeûne [58], [59]. Chez des vaches de race Holstein, une restriction alimentaire par substitution de 60% de la ration mixte par de la paille de blé peut entraîner une augmentation de matières grasses dans le lait de 42%, après 48h, par un effet de compensation par mobilisation de graisses [60]. Herve *et al.* ont également mis en avant une augmentation du taux de matières grasses (de 3,42% à 3,63%) dans le lait d'Holstein en cas de restriction alimentaire : cette restriction correspond à 80% de la ration [61].

En caprin, des études se basent sur des suppléments alimentaires en acides linoléiques conjugués, montrant l'induction d'une baisse de taux de matière grasse dans le lait [62]. Des suppléments avec 5% de glycérine brute résultent en une augmentation significative des productions de lait et particulièrement du taux de matières grasses dans le lait [63]. Selon les régions, les suppléments et les régimes alimentaires diffèrent. Ainsi, une étude iranienne se base sur une alimentation en peaux de pistaches, démontrant un impact sur la concentration en matières grasses du lait, sans influencer sur la production [64].

De plus, un régime composé d'un ajout d'huile de thon non encapsulé induit une baisse de la production de lait chez des chèvres [65].

En ovin, une étude basée sur une supplémentation à base de palmier-dattier a montré un lien avec la production de lait, qui est augmentée lors de supplémentation de 20 ou 30%, ainsi qu'avec le taux de calcium dans le lait [66]. Des suppléments en vitamines et minéraux sur des brebis permettent une augmentation des taux de protéines et matières grasses du lait, sans impacter les taux de lactose. Plus globalement l'étude menée montre l'importance de ce type de supplémentation pour le lait destiné à la transformation en fromage (notamment *via* les composants liés à l'aptitude à la coagulation) [67].

e. Impact de la santé

La santé peut avoir un impact sur la production laitière, c'est particulièrement le cas lors de la présence de mammites, une inflammation de la glande mammaire (Bobbo et al., 2017; I Koop et al., 2010; Rupp et al., 2015). En bovin, une diminution de la teneur en lactose dans le lait est observée en cas de mammite dans des études menées sur 80 et 37 vaches en lactation [72], [73]. Une étude réalisée sur des chèvres montre également une diminution significative de la concentration de lactose dans le lait de 47,0 g/L à 41,7 g/L en cas de présence de mammite [74].

3. Les différentes races présentes en France

La France compte 53 races bovines reconnues par un décret du ministère de l'agriculture et de la pêche en 2018 [75], [76]. On compte 11 races caprines en France [77] et 57 races ovines reconnues, dont une majorité de races en conservation, dites "races animales locales menacées d'être perdues pour l'agriculture". Une race est éligible à ce titre si elle respecte des critères d'effectifs insuffisants ou encore de risque de perte de diversité génétique [78], [79].

Plus précisément, deux critères d'effectifs ont été définis : le nombre de femelles reproductrices (ayant déjà eu au minimum un descendant) et l'évolution de ce nombre ; deux critères sont basés sur la génétique : la proportion de femelles qui n'ont que des reproductions en croisement (pas en race pure) et la taille efficace de la population (correspondant aux reproducteurs de la population) ; et enfin deux critères sur l'élevage que sont l'organisation des éleveurs pour la gestion de la race, c'est-à-dire l'existence ou non d'associations d'éleveurs, de programmes de gestion de la race, et le contexte socio-économique global [80].

Parmi les races recensées par le Ministère de l'Agriculture, certaines races sont définies en tant que races locales et d'autres en tant que races à petits effectifs. Les races sont définies comme locales lorsque "30 % des effectifs sont situés dans un seul département ou 70 % dans trois départements limitrophes deux à deux", d'après l'Arrêté du 29 avril 2015 fixant la liste des races des espèces bovine, ovine, caprine et porcine reconnues [81]. Une race à petit effectif correspond à moins de 5 000 femelles reproductrices en bovin et moins de 8 000 en ovin et caprin [78].

a. Races non majoritaires

Les espèces bovine, caprine et ovine les races dites "minoritaires" peuvent donc être des races locales ou en conservation.

i. Bovins laitiers

En France l'élevage de bovins laitiers se compose des trois races principales (Prim'Holstein, Normande et Montbéliarde) ainsi que de races laitières moins importantes en terme d'effectif. Il s'agit de races avec des productions de lait moins importantes.

On peut citer, par ordre de nombres de vaches comptabilisées grâce aux données de contrôles laitiers, les races Abondance, Brune, Simmental, Pie Rouge, Tarentaise et Jersiaise [82]. Certaines races sont plébicitées pour la composition de leur lait. Par exemple, la race Jersiaise est celle présentant les taux protéiques et butyreux les plus élevés parmi les races citées. La Pie Rouge présente une production moyenne de lait et des masses de matières grasses et de

protéines élevées également [82]. La race Brune est prisée notamment pour sa grande résistance face aux infections de la mamelle [83]. L'Abondance ou la Tarentaise ont une forte adaptabilité aux conditions d'élevage, notamment climatiques [84]. La race Salers permet, elle, la production de fromages AOC (Appellation d'Origine Contrôlée) Tradition Salers, et est également valorisée par son adaptabilité climatique et son utilisation pour la filière viande.

Les races Bretonne Pie Noire ou Froment du Léon sont incluses dans des programmes de conservation, comptant respectivement 1 600 et 260 vaches en France [84].

ii. Caprins laitiers

Certaines races caprines laitières ont des effectifs réduits et permettent l'élevage dans des environnements particuliers. Les races Corse, Poitevine ou Des Fossés voient leurs effectifs augmenter à nouveau grâce à des programmes de conservation [84]. Des races minoritaires de caprins laitiers sont parfois privilégiées du fait de la transformation de leur lait en fromage, qui peut être AOC (Appellation d'Origine Contrôlée). C'est le cas par exemple du fromage Brocciu (fromage AOC) ou du fromage Brousse du Rove pour la race du Rove [85]. La race Poitevine est valorisée par les hautes qualités de fromageabilité de son lait [86].

iii. Ovins laitiers

Les races ovines laitières minoritaires sont généralement des races locales, avec une forte adaptabilité aux conditions d'élevage. C'est le cas, par exemple, des races Basco-Béarnaise ou Corse favorisées dans les zones montagneuses car ce sont des races rustiques [84]. En ovin laitier, tout comme en espèce caprine, des races sont également privilégiées du fait de la transformation de leur lait en fromages. On peut citer par exemple le Brocciu, pour la race Corse (fromage AOC), ou encore l'Ossau-Iraty pour les races Manech et Basco-Béarnaise (fromage AOC).



Figure 1 : Bovins de race Holstein (crédit photo : Alger MEEKMA [87]), Normande [88], et Montbéliarde [89] ; Caprins de race Alpine [90] et Saanen [91] ; Ovins de race Lacaune [92] et Manech Tête Noire [93]

Composition du cheptel de vaches laitières françaises

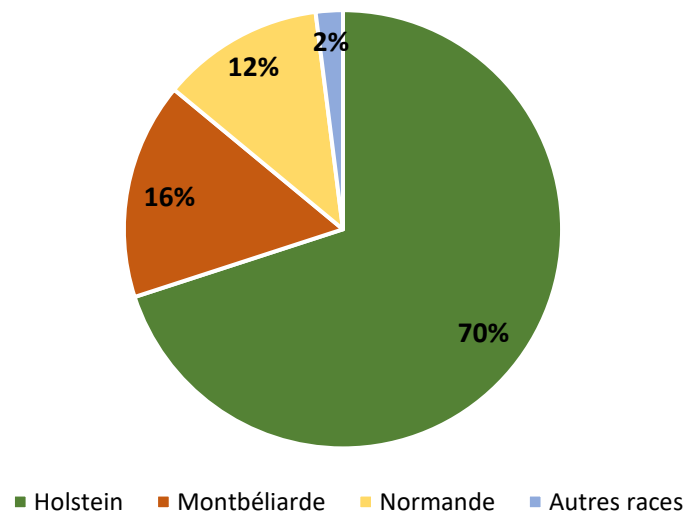


Figure 2 : Proportion de races bovines dans le cheptel de vaches laitières en France

b. Les races majoritaires

i. Bovins laitiers

Les races les plus représentées en France sont la Holstein (avec 2,5 millions de vaches laitières), la Montbéliarde (665 000 vaches) et la Normande (426 000 vaches) (*Figure 1*) [94]. Plus précisément, ces trois races représentent, d'après les chiffres de 2012, environ 98% du cheptel des vaches laitières françaises (70% pour la race Holstein, 16% pour la Montbéliarde et 12% pour la Normande) (*Figure 2*) [95].

Les caractères de production, de taux protéique et butyreux sont les caractères laitiers importants. Le tableau ci-après regroupe les différentes données de ces caractères pour les trois principales races bovines laitières présentées ci-dessus (*Tableau 3*).

1. Holstein

La race Holstein est donc la race bovine laitière la plus répandue en France. Elle a comme caractéristique principale une forte production laitière : il s'agit de la race à la plus forte production de lait au monde, avec jusque 9 tonnes de lait produites par an [96], [97]. Les Holstein sont également prisées de par leur développement rapide [97]. Cette race est plébiscitée du fait de sa grande adaptabilité, notamment du point de vue conditions environnementales [98].

2. Normande

La race Normande est également une des races bovines majoritaires en France. Essentiellement élevée dans l'Ouest de la France et plus particulièrement en Normandie, cette race est particulièrement prisée pour son lait riche en matière protéique et adéquat pour la transformation en fromage. La production de lait de vaches Normande permet la production de Camembert de Normandie ou encore d'autres fromages tels que le Pont-l'évêque [84], [99].

<i>Moyennes</i>	Holstein	Montbéliarde	Normande
Production	9 241 kg	6 933 kg	6 590 kg
TP	32,0 ‰	33,1 ‰	34,7 ‰
TB	39,7 ‰	38,9 ‰	42,4 ‰

Tableau 3 : Caractères de production laitière des trois principales races laitières françaises. Les vaches Holstein ont des productions de lait par an plus importantes que les deux autres races. Les Normande présentent des taux protéiques et butyreux plus élevés

3. Montbéliarde

La race Montbéliarde est une race laitière française ayant des capacités de résistance importante, notamment du point de vue de la longévité, la fertilité mais aussi de la rusticité. Cette race présente une bonne résistance aux mammites [100]. Initialement concentrée en région Franche-Comté, cette race présente une bonne adaptabilité, aux différentes températures par exemple. Tout comme la race Normande, la Montbéliarde est majoritairement liée à la fabrication de fromages. C'est plus particulièrement le cas de la transformation du lait de Montbéliarde en fromage Comté (fromage AOC). Cette race a également le meilleur rapport taux butyreux/taux protéique parmi les principales races laitières et une très bonne capacité de production, avec près de 7,6 tonnes de lait par an [101], [102].

ii. Caprins laitiers

1. Alpine

L'Alpine est la race caprine la plus représentée en France, puisqu'elle représente 55% des effectifs totaux de chèvres [90], [103]. Le lait de cette race est utilisé pour la fabrication de fromages. L'Alpine est caractérisée par une grande adaptabilité, pouvant convenir aux stabulations tout comme aux pâturages. La production laitière de cette race caprine se situe autour de 1 000 kg par lactation (Palhière et al., 2014).

2. Saanen

La Saanen est une race caprine laitière. Il s'agit de la seconde race caprine laitière, avec environ 40% des effectifs totaux de chèvres en France. Dans le monde, il s'agit de la race caprine laitière la plus répandue [91], [105]. Cette race a une très bonne production laitière, aux alentours de 1 000 kg par lactation [104].

iii. Ovins laitiers

1. Lacaune

La Lacaune est la première race ovine française en terme d'effectif avec près de 60% des effectifs totaux [17]. Il s'agit d'une race laitière et allaitante, essentiellement utilisée pour la fabrication de Roquefort [106]. Les Lacaune sont privilégiées pour leur excellente rentabilité, mais également pour leur adaptation à la mono-traite (une seule traite par jour), dans un contexte de production de lait hors production de Roquefort, pour laquelle l'AOP (appellation d'origine protégée) impose 2 traites, avec une perte de lait estimée de moins de 15% [107]. Pour comparaison, la race Préalpes-du-sud a une perte de lait estimée à plus de 34,5% lors de mono-traite en comparaison à une traite deux fois par jour [108]. Son adaptabilité lui permet d'être élevée dans des milieux géographiques différents.

2. Manech Tête Noire

La Manech Tête Noire est une race ovine laitière. Cette race présente le deuxième plus grand nombre d'animaux en race ovine laitière des Pyrénées. Il s'agit d'une race montagnarde, principalement élevée pour la fabrication de fromage Osau-Iraty [109]. Son adaptabilité et sa résistance aux changements thermiques en font une race prisée des éleveurs [110].

4. Génétique

a. Etat des lieux : connaissances des génomes des ruminants

Ces dernières années, l'évolution des connaissances scientifiques a permis le développement de techniques de séquençage des génomes (notamment le WGS : Whole Genome Sequencing ou Séquençage du Génome Entier). Les génomes de 73 espèces de ruminants ont désormais été séquencés [111]. L'obtention de ces séquences de génomes nécessite par la suite l'utilisation d'outils bioinformatiques (accessibles sur les sites tels qu'Ensembl (<https://www.ensembl.org/index.html>), UCSC (<https://genome.ucsc.edu/>), ou encore NCBI (<https://pubmed.ncbi.nlm.nih.gov/>) [112]) afin d'analyser ces séquences, de déterminer la position des gènes par exemple.

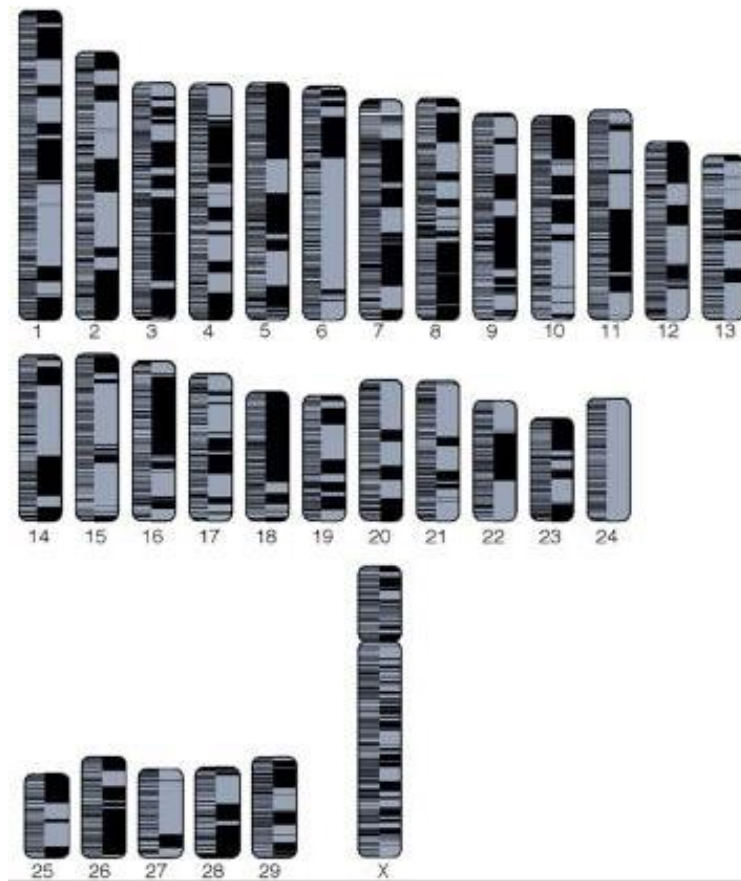


Figure 3 : Chromosomes bovins avec contigs assemblés. Chaque changement de couleur correspond à un changement d'un contig à un autre. La partie gauche de chaque chromosome représente les contigs de la version UMD3.1.1 et la partie droite représentent ceux de la version ARS [113]

Débutée fin 2003 [114], la séquence du génome bovin a été déterminée en 2009, à partir d'une vache de race Hereford, nommée Dominette 01449 [115]. Des données de séquençage moins profond d'individus de races Holstein, Angus, Jersey, Limousin, Norwegian Red et Brahman ont été ajoutées [114]. Les assemblages BCM4 et UMD2 ont été alors réalisés, en se basant sur des chromosomes artificiels bactériens (BACs) et du séquençage génome entier. En résumé, le génome est découpé en plusieurs fragments qui sont alors clonés et insérés dans des BACs. Les extrémités de ces BACs sont chevauchantes et leur ordre connu (fingerprinting). Chacun d'eux est alors séquencé, résultant en un ensemble de contigs, de fragments chevauchants [116]. L'assemblage BCM4 comporte 2,54 gigapaires de bases et plus de 90% des séquences sont assignées à un chromosome spécifique. L'assemblage UMD2 comporte lui 2,86 gigapaires de bases et 91% des séquences sont assignées à un chromosome [114], [117]. La moindre couverture de l'assemblage BCM4 est expliquée du fait de l'utilisation de BACs : des régions du génome se situant en dehors de ces BACs ne sont alors pas prises en compte. Les régions dupliquées sont également plus nombreuses dans le BCM4 du fait du chevauchement de régions : des BACs peuvent se chevaucher, ces régions qui se sont superposées peuvent alors être considérées à tort comme dupliquées dans le génome. L'assemblage UMD2 a, lui, utilisé des données de comparaison entre données de séquences bovines et humaines notamment afin de pouvoir orienter et donc placer les contigs bovins à partir des données humaines [114], [117]. Les couvertures du génome pourront par la suite être améliorées, et les erreurs d'assemblage pourront être réduites grâce aux avancées technologiques [114], [117]. Désormais, la version la plus récente du génome bovin est ARS-UCD1.2 [113]. Ce dernier assemblage comporte 2,7 gigapaires de bases. Ce nouvel assemblage a amélioré de dix fois la précision de l'assemblage par rapport à la précédente version [113]. L'amélioration de l'assemblage est permise par la réduction du nombre de contigs : cet assemblage ARS-UCD1.2 se compose de 345 contigs, l'assemblage précédent UMD3.1.1 comporte 72 264 contigs. La continuité des séquences est donc augmentée de plus de 200 fois (*Figure 3*) [113].

La première séquence du génome caprin a été disponible à partir de fin 2012 : CHIR_1.0 [118]. En ce qui concerne les caprins, un mâle de race San Clemente a été choisi pour le nouvel assemblage ARS1 [119]. Tout comme pour l'espèce bovine, la précision de l'assemblage a été améliorée par l'utilisation de séquençage de lectures longues et donc la réduction du nombre de contigs. La continuité des séquences est donc augmentée d'environ 400 fois en comparaison de l'assemblage précédent [120]. Pour ce qui concerne la séquence du génome ovin, la première version date de 2010 (ISGC Ovis_aries_1.0/oviAri1) [121]. Le nouvel assemblage ovin est désormais Oar_rambouillet_v1.0. Une brebis femelle de race Rambouillet a été choisie pour le réaliser [122]. Dans cette espèce également le séquençage de lectures longues (PacBio long-read sequencing) est utilisé et permet une meilleure continuité des séquences et donc une amélioration de l'assemblage par rapport aux précédents [123], [124].

Les données du site Ensembl, regroupant les informations d'assemblage et d'annotation, sont accessibles pour ces mêmes génomes de référence à l'exception de l'espèce ovine, pour laquelle la version Oar_v3.1 est utilisée [125]. Les derniers assemblages ont été réalisés avec des couvertures de 80X, 50X et 126X pour les bovins, caprins et ovins, respectivement. Leur taille est sensiblement la même pour ces trois espèces : 2,7 gigabases pour les bovins et 2,9 gigabases pour les caprins et ovins [126]. On compte 2 180 séquences non placées dans le génome bovin, 29 875 non placées pour le génome caprin et 2 613 dans le génome ovin [119], [122], [127].

Ces génomes comportent un nombre de paires de bases assez similaire : 2 715 853 792 paires de bases pour les bovins, 2 922 813 246 et 2 534 344 180 paires de bases pour les caprins et ovins. Le nombre de gènes codants est lui aussi très proche chez les trois espèces : 21 880 en bovin, 21 361 en caprin et 20 921 en ovin. Néanmoins, les nombres de variants de petite taille (notés "short variants") sont très dissemblables : 98 844 739 variants pour les bovins, 34 116 678 pour les caprins et 60 323 418 pour les ovins [125], [128], [129]. Pour des tailles de génome similaires, les caprins ont près de trois fois moins de variants de petite taille identifiés que les bovins et près de deux fois moins que les ovins, d'après ces données d'Ensembl. Les ovins ont plus de 1,5 fois moins de variants que les bovins. Un plus grand nombre de recherches réalisées chez les bovins pourrait par exemple expliquer l'effectif de variants plus

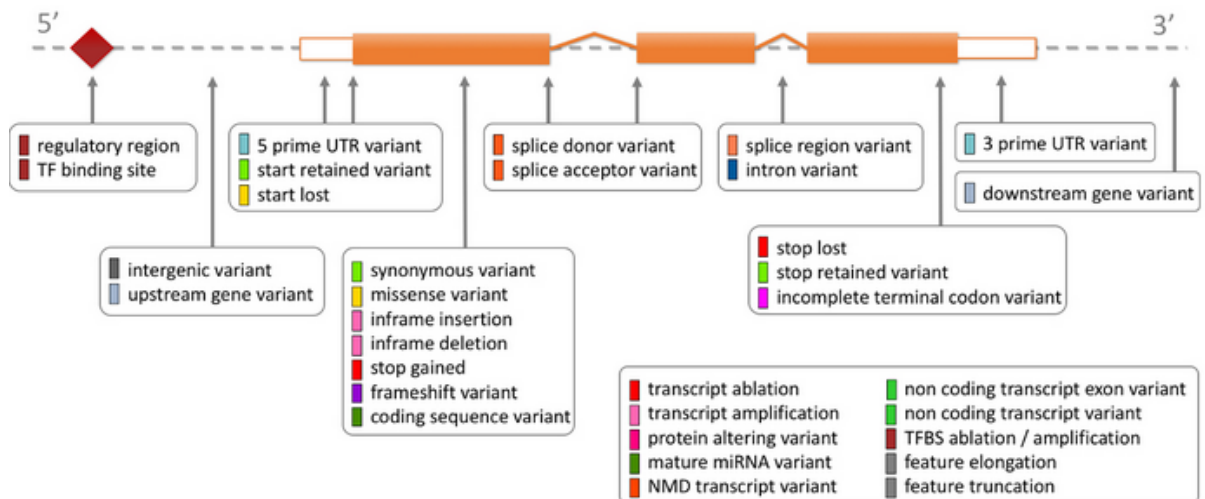


Figure 4 : Localisation des différents types de variants génétiques caractérisés par leur probable impact, d'après Ensembl [130]

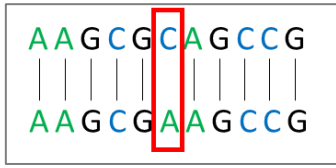
important que chez les caprins ou ovins. Désormais divers outils sont disponibles en ligne, afin d'obtenir des annotations des génomes [115]. L'ensemble de ces outils permettent aux biologistes d'avoir un éventail de données et ainsi de détecter et d'étudier des mécanismes de plus en plus fins.

b. Les variants génétiques

Les variants génétiques sont des modifications dans la séquence d'ADN, qui apparaissent spontanément et peuvent ou non induire un changement de phénotype [131].

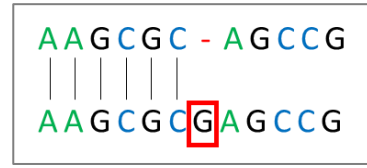
Les variations génétiques peuvent être caractérisées par leur possible conséquence [132]. Généralement, les variations sont classées en trois types : les variations synonymes, ne modifiant pas l'acide aminé qui est traduit, les variations faux-sens, induisant une modification de l'acide aminé traduit, et les variations non-sens, qui induisent un codon stop prématuré et donc une protéine tronquée. Ces potentielles conséquences des variants sont réparties plus précisément selon des "SO term" (sequence ontology term). Plus de 30 ontologies de séquences sont listées à ce jour selon Ensembl (*Figure 4*) [130], [132].

Un exemple d'ontologie est le "synonymous_variant", pour les variations ne modifiant pas le code d'acides aminés de la protéine résultant du gène impacté. D'autres SO caractérisent des variants génétiques impactant les acides aminés, de par la création d'un codon stop prématuré dans le cas du SO "stop_gained", ou en décalant le cadre de lecture *via* une insertion ou délétion d'un nombre de nucléotides non multiple de trois pour le SO "frameshift_variant" par exemple. Des variants d'épissage ou modifiant la protéine de par la modification de la séquence d'acides aminés ("missense_variant") sont également inclus. Chaque sequence ontology est associée à une valeur d'impact : si l'impact du variant est fort, modéré ou faible. D'autres ontologies ont un impact dit "modifier". Ces ontologies sont celles des variants situés dans un intron, dans les régions 5' ou 3' de gènes ou des régions régulatrices (site de fixation de facteurs de transcription par exemple), ou encore dans des régions intergéniques. Des variants de microARNs peuvent être assignés à l'ontologie "mature_miRNA_variant" [130], [132]. Les différentes variations génétiques peuvent être



Un SNP

$C > A$



Une petite insertion

Insertion d'un nucléotide G

Figure 5 : Exemple de SNP et de petite insertion

répertoriées en fonction de leur taille : changement d'un nucléotide, insertion ou délétion de quelques nucléotides.

i. SNPs

Les Single Nucleotide Polymorphisms (SNP) sont des polymorphismes (variations) d'une seule base, d'un seul nucléotide. Les variations sont alors notées de différentes façons selon les bases impactées : il s'agit de transitions lorsque ce sont alors des changements entre bases puriques ou entre bases pyrimidiques et il s'agit de transversion dans le cas d'interchangements entre purines et pyrimidines. En ce qui concerne la nomenclature, la notion de SNP et plus précisément de polymorphisme est utilisée seulement à la condition que la fréquence de l'allèle minoritaire (Minor Allele Frequency (MAF)), soit supérieure à 1%. Dans le cas contraire, il s'agira de mutations et non de polymorphismes [133]. Ces variants génétiques peuvent servir de marqueurs génétiques : des séquences polymorphiques facilement détectables, se situant tout au long du génome et pouvant ainsi servir de balises, de repères, dans le génome.

ii. Petites insertions et délétions

Les InDels sont des insertions et délétions de moins de 1 000 paires de bases [134]. Les petits InDels sont des insertions ou délétions de petites tailles, d'un faible nombre de nucléotides (*Figure 5*). Les petits InDels peuvent être caractérisés par des longueurs inférieures ou égales à 50 paires de bases [135], [136].

iii. Autres polymorphismes

Les microsatellites sont des répétitions de plus petites tailles (1 à 5 bases). Ces microsatellites ont été essentiellement utilisés comme marqueurs génétiques, du fait de leur abondance et de leur répartition tout au long du génome [137].

Un autre type de polymorphisme est la translocation, qui correspond à un déplacement en bloc de nucléotides au sein d'un même chromosome ou sur un autre chromosome. Ces translocations peuvent être robertsoniennes lorsque il y a fusion centrique des chromosomes ou des cassures dans les régions situées autour du centromère, ou réciproques lorsque le point de cassure est en dehors de ces régions. Pour exemple, la translocation robertsonienne 1/29 (donc entre les chromosomes 1 et 29) est connue et recherchée en bovin car cause de baisse de fertilité, particulièrement chez certaines races comme la Blonde d'Aquitaine ou encore la Montbéliarde [138].

Les copy number variations (CNV, variations du nombre de copies) sont des portions d'ADN d'au minimum 1 kilobase dont le nombre de copies diffère d'un individu à l'autre [139]. Ces variations sont distribuées au long du génome. Elles peuvent affecter une portion du génome plus grande que les SNPs et changer la structure de gène [140]. En bovin, des régions de CNV ont été étudiées dans des groupes d'animaux présentant des valeurs d'élevage forte ou faible pour les pourcentages de protéines et de matières grasses du lait. Certaines CNV sont retrouvées comme associées significativement à ces caractères de production laitière [140].

c. Quantitative Trait Locus

i. Définition

Les QTL sont des régions génomiques responsables de caractères quantitatifs (Quantitative Trait Loci), tels que des caractères de production comme les caractères laitiers, de santé ou encore de reproduction. Lorsque 100% de la variance phénotypique est expliquée par le QTL, on est alors en présence d'un gène majeur. C'est le cas en ovin, pour des caractères de prolificité : les gènes majeurs *BMP15*, ou *B4GALNT2* (plus communément appelé gène Lacaune) par exemple [141].

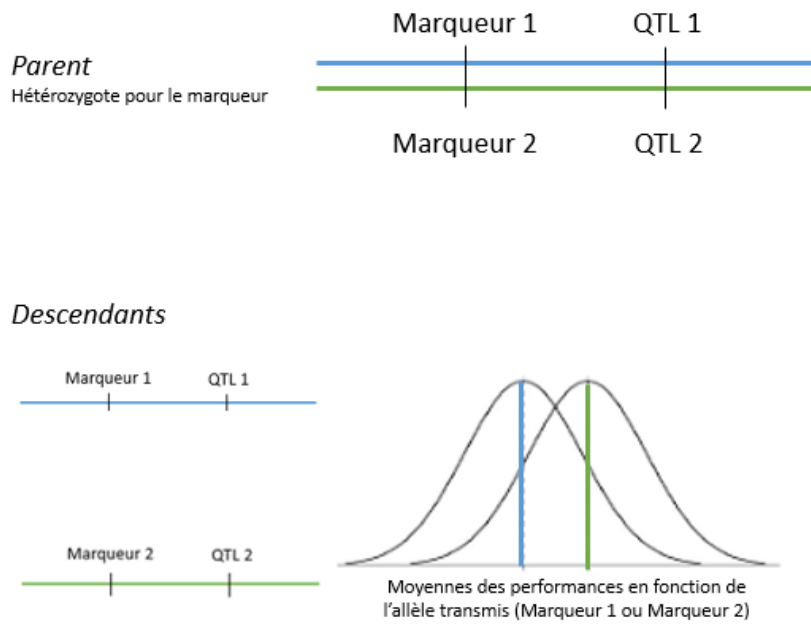


Figure 6 : Représentation de l'analyse de liaison

ii. Détermination de QTL

La détermination de QTL se base sur une méthode statistique qui permet de lier les informations de phénotypes et de génotypes disponibles : des marqueurs moléculaires. Les marqueurs sont neutres pour le caractère quantitatif étudié et sont répartis tout au long du génome, afin de pouvoir associer un ou des marqueurs à toute zone génomique. Des dispositifs QTL ont ainsi pu être mis en place avant que les séquences tout génome aient été disponibles. On utilisait à l'époque d'autres types de marqueurs (microsatellites, gènes, fragments de séquence codante...).

Des analyses de liaison, des analyses d'association ou des analyses combinant ces deux méthodes peuvent être utilisées pour la détection de QTL.

Une analyse de liaison, ou étude de la transmission d'allèles entre apparentés, s'effectue en partant d'un parent hétérozygote pour un marqueur donné. Les données de sa descendance sont alors observées afin de déterminer si les moyennes des phénotypes observés (les performances) sont différentes en fonction de l'allèle qui est transmis (*Figure 6*) [142]. Des dispositifs familiaux peuvent alors être utilisés, comme les protocoles "petites-filles" (ou Grand-Daughter Design, GDD) [143].

Une analyse d'association se base sur le principe du déséquilibre de liaison : la corrélation qui existe entre allèles de plusieurs locus distincts. Plus précisément, il s'agit du fait que des individus qui présentent un même allèle à un locus donné présentent souvent le même allèle pour des SNPs à proximité. Les gènes ne sont alors pas associés de façon aléatoire [144].

Une analyse combinant ces deux types d'analyse est désormais possible grâce à la plus faible distance entre deux marqueurs, permise par les puces hautes densités. L'analyse combinée permet une puissance d'analyse importante, avec les déséquilibres de liaison analysés sur une population entière, couplée à une précision, une robustesse issue de l'analyse d'apparentés [145], [146].

Ces analyses permettent de déterminer des régions génomiques qui comportent de nombreux SNPs et des gènes. Ces régions peuvent atteindre plus de 1 Mb. Le fait de détecter le SNP causal ou un gène responsable du caractère est complexe. En effet, le gène candidat peut être

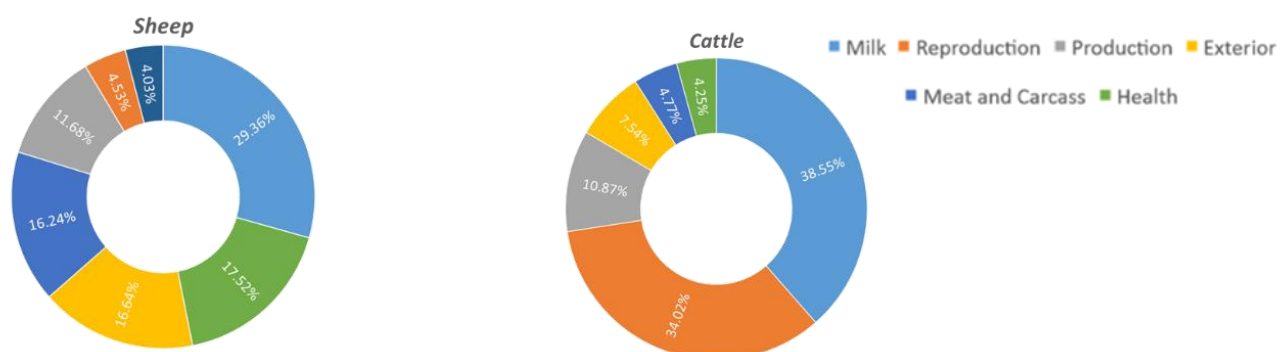


Figure 7 : Proportion des différents caractères de QTL chez les ovins et bovins, présents dans la base de données AnimalQTLdb [152]

Caractères laitiers	Bovins	Caprins	Ovins
α -lactalbumine	X		
β -lactoglobuline	X		X
Caséine α 1	X		X
Caséine α 2	X		X
Caséine β	X		X
Caséine κ	X		X
Total des caséines	X		X
Total α -lactalbumine et β -lactoglobuline	X		X
Acides gras insaturés			X
Acides gras monoinsaturés			X
Acides gras polyinsaturés			X
Acides gras saturés			X
Production de lait	X	X	X
Taux de matières grasses (%)	X	X	X
Rendement en matières grasses (kg)	X	X	X
Taux protéique (%)	X	X	X
Rendement en protéines (kg)	X	X	X
Mammite clinique	X		
Score cellulaire	X		X

Tableau 4 : Liste de caractères de QTL laitiers étudiés par [153]–[156] et Larroque et al. (communication personnelle)

défini comme étant le gène le plus proche du variant génétique qui possède la plus forte association statistique avec le caractère. Néanmoins, le SNP causal peut être éloigné du gène candidat pour des raisons de déséquilibre de liaison par exemple [147]. L'utilisation d'autres données "omics" peut alors permettre d'aider à la priorisation des SNPs détectés. On peut citer pour exemple des outils comme ProGem, ADmiRE ou même la définition d'un score FAETH (Functional-and-Evolutionary Trait Heritability) [148]–[151].

La base de données répertoriant les QTL animaux s'appelle AnimalQTLdb [157]. La dernière mise à jour, d'août 2020, regroupe les données de 2 335 publications. Des régions génomiques sont décrites pour sept espèces : vache, poisson-chat, poulet, cheval, porc, truite arc-en-ciel et mouton. Plus de 2 000 caractères sont ainsi pris en compte, avec plus de 178 000 associations décrites. En ce qui concerne les bovins, 130 407 régions génomiques sont référencées, dont plus de 50 000 concernant des caractères laitiers. Pour les ovins, 3 099 régions génomiques s'y trouvent dont 875 associées à des caractères laitiers. Des données de QTL en espèce caprines sont annoncées pour la prochaine mise à jour, prévue fin décembre 2020 (*Figure 7*) [157].

Des bases de données complémentaires listent des régions génomiques également liées à des caractères quantitatifs, par exemple une base de données regroupant les QTL humains s'est développée, nommée QTLbase. À présent, de nombreux sous-types de QTL sont définis chez l'humain : des QTL reliés aux régulations épigénétiques (caQTL – Chromatin Accessibility – ou encore mQTL – DNA methylation –) ou encore aux régulations de la traduction (riboQTL – Ribosome occupancy – ou pQTL – Protein expression –) [158].

Dans le cas des QTL laitiers, des caractères utilisés sont listés dans le tableau 4 (*Tableau 4*) [153]–[156]. Plusieurs QTL, donc plusieurs régions génomiques, sont détectés pour un même caractère. Et inversement, une même région génomique peut être affectée à plusieurs caractères différents. En effet, il peut y avoir des superpositions partielles ou complètes de régions génomiques QTL associées à un ou plusieurs caractères. Les caractères laitiers reprennent les principaux composants du lait et deux caractères de santé sont également pris en compte : MACL et CEL, caractère de présence de mammites cliniques (MACL) et de

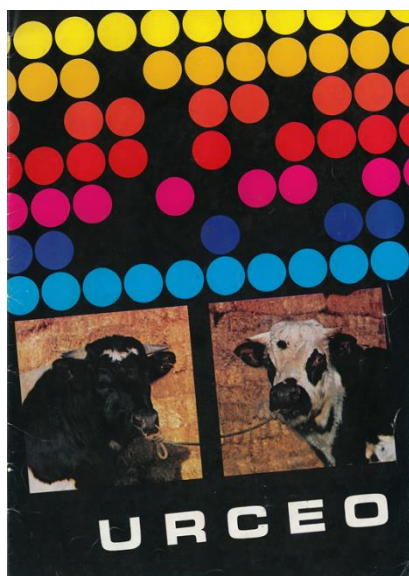
comptage des cellules somatiques dans le lait (CEL), dont l'augmentation peut être un signe d'apparition de cette infection [153].

Des SNPs ont été détectés par Sanchez *et al.* [153] dans des régions codantes du génome pour ces QTL laitiers, avec notamment le gène *DGAT1*, qui est lié à la variabilité des taux de matière grasse et de protéines en bovin [159] et le gène *CSN2* qui est exprimé dans la glande mammaire et code une protéine majeure du lait [160].

Sanchez *et al.* listent au total 22 gènes d'intérêt, dans 11 régions génomiques distinctes. Ces gènes sont des candidats pour expliquer une grande part du caractère associé, ici la composition protéique du lait [153].

Néanmoins il reste toujours des QTL pour lesquels aucun variant dans les régions codantes n'a été détecté. Ces régions du génome n'expliquent pas la totalité de la variance phénotypique pour des caractères, puisque des facteurs autres, tels que des facteurs environnementaux ou le fait de la multi-génécité de caractères, entrent en considération [161].

De plus, les régions codantes ne représentent qu'une faible proportion du génome entier (2% pour les gènes humains), et il a été estimé que 90% des SNPs associés à des pathologies humaines sont situés dans les régions ne codant pas des protéines, donc seuls 10% sont dans le codant [162].



SUCCES MEINTJE

H.B N° 29 70 046 183 - Né le 31 octobre 1970
 Poids : 980 kgs à 3 ans 6 mois
 Hauteur au garrot : 1 m 45
 Naït.-Vend. : M. COTTEN J., Fouesnant (29)

SUCCES HINKE } WATTINCOURT SUCCES MARTHA
 INKE KOLLUMER

HEECHHIEM MEINTJE 85 } HAUBOIS ADEMA
 HEECHHIEM MEINTJE 68

ACCOUPLEMENT

Maintient un développement correct.

A UTILISER : sur des animaux dont on désire améliorer les mamelles et la rapidité de traite.

A ÉVITER : sur la descendance de SUCCES HINKE et de ses fils

Production laitière de la descendance

INDEX 4 PREMIERS CONTROLES	Nbre de filles	Lait Kg/jour	M.G. Kg/jour	M. Az. Kg/jour
	54	+ 1,020	+ 0,030	+ 0,010

Lactations terminées au :	Nbre de filles	Age au vêlage	Lait	Taux butyreux	Matière grasse	Taux azoté	Matière azotée	Durée
31.12.76	45	2 a 4 m	3919	38,1	149	120	30,7	283

Conformation de la descendance

Dévelop.	Av. Main	Dessus	Ar. Main	Aplombs	Mamelles					
					Av.	Ar.	Équil.	Écart.	Impl.	
=	=	=	=	=	+	+	+	+	=	+

Figure 8 : Couverture d'un catalogue URCEO de 1977 et exemple de description d'un animal en tant que père potentiel

5. La sélection génomique

a. Principes et but de la sélection génomique

La sélection a pour objectif la prédiction de la valeur génétique d'animaux et donc la sélection d'animaux présentant des caractères d'intérêt. Cette sélection est possible grâce aux génotypages et phénotypages d'une population de référence, contenant un grand nombre d'animaux, ce qui permet ainsi d'établir un lien statistique entre génotype et phénotype. Ainsi, le génotypage d'un individu pour des marqueurs génétiques amène à la prédiction de son phénotype pour les caractères voulus. Les programmes de sélection génomique utilisent donc des marqueurs génétiques, et le génotypage se déroule grâce à des puces à SNPs regroupant ces marqueurs génétiques. Une amélioration possible de ces programmes pourrait être amenée par l'ajout de mutations causales détectées [155].

b. Historique de la sélection

i. Sélection génétique, évaluation sur descendance

L'apparition et le développement de la génomique avec les possibilités de génotypages ont permis la mise en place de cette sélection précise d'animaux selon leurs performances, mais pour autant la sélection est utilisée depuis bien plus longtemps. En effet, les éleveurs de tout temps réalisaient des choix de reproducteurs selon leurs performances. Dès le XIX^{ème} siècle, les premiers livres généalogiques apparaissent, accompagnés des premières notions de sélection. Les performances des vaches et des filles des taureaux étaient alors scrutées, de façon à planifier les croisements dont il devait résulter des vaches présentant les performances voulues (testage sur descendance). Les premiers organismes de sélection en France, tels que Urceo (Union Régionale des Coopératives d'Elevage de l'Ouest), ont vu le jour dès la fin des années 1940, d'abord sous l'appellation de la Coopérative d'Insémination Artificielle. Ces organismes mettaient à disposition des éleveurs des catalogues des reproducteurs disponibles listant chacune des caractéristiques de ces pères potentiels, ainsi que de brèves données de filiation afin d'éviter la consanguinité (*Figure 8*).

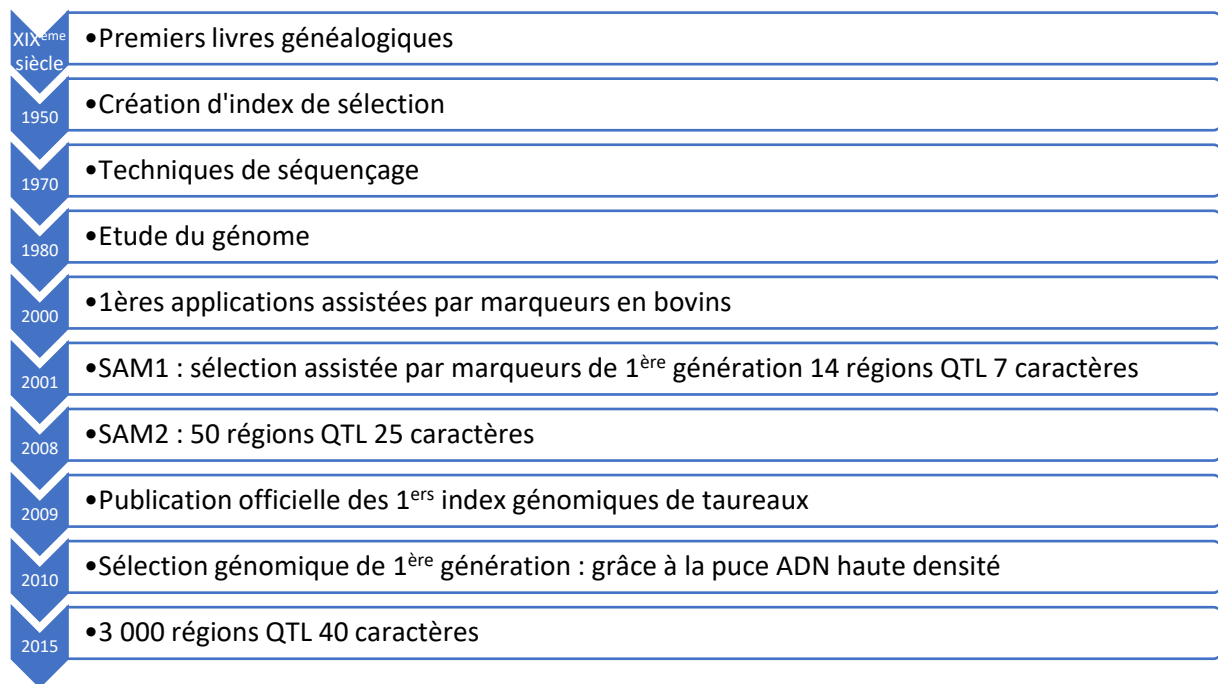


Figure 9 : Chronologie de la sélection génomique bovine en France [95], [163]

Des commentaires sur les performances attendues mais aussi sur les croisements les plus avantageux (en prenant en compte les caractéristiques des mères) étaient d'ores et déjà pris en compte et précisés pour chaque individu.

ii. Sélection génomique

Chez les bovins, des index de sélection ont été définis en 1950 : il s'agit pour les reproducteurs de leurs valeurs génétiques estimées. Par la suite, les techniques de séquençage et l'étude des génomes ont toutes deux permis une avancée technologique dans le domaine de la sélection, respectivement dans les années 1970 et 1980 (*Figure 9*). Par la suite, en 2000, les premières sélections assistées par marqueurs sont mises en place en bovins. Le principe de la sélection assistée par marqueurs se base sur l'identification d'allèles situés dans des QTL d'intérêt et leur suivi au long des générations. Pour cela, des marqueurs sont utilisés, puisque les polymorphismes causaux sont rarement identifiés. Ce sont des marqueurs proches, qui reflètent les impacts des gènes : on pourra donc identifier des marqueurs qui sont associés à des SNPs, notamment grâce à des puces à SNPs dont la puce 50K qui regroupe 54 000 SNPs. Les marqueurs recherchés sont associés à des allèles influant des caractères de production et permettent ainsi la prédiction de la valeur d'individu [164], [165]. Suivies en 2001, de SAM1 : la sélection assistée par marqueurs de première génération, pour 7 caractères ; SAM2, la deuxième génération, apparaît en 2008 et compte alors 25 caractères au total. SAM2 compte également un plus grand nombre de marqueurs et permet ainsi un meilleur suivi des régions génomiques d'importance (les QTL d'intérêt) : un suivi sur de plus nombreuses générations et même sur une population et non plus seulement sur une famille, mais aussi une définition de régions d'intérêt plus petites. L'année suivante, en 2009, les premiers index génomiques (soit des valeurs génétiques estimées [166]) de taureaux sont officiellement publiés, suite à quoi la sélection génomique de première génération est mise en place en grâce à la puce à ADN haute densité. Enfin, en 2015, 3 000 régions QTL pour 40 caractères différents sont comprises dans le programme de sélection génomique [95], [167].

Des index plus complexes sont maintenant développés, tels que l'INEL (INdex Économique Laitier) qui correspond à une formule mathématique : $INEL = 0,98 (MP + 0,2 MG + TP + 0,5 TB)$ avec MP = matière protéique, MG = matière grasse et TB = taux butyreux [168]. Ce nouvel index INEL permet de regrouper plusieurs caractères d'importance en un seul index, qui permet d'avoir un index spécifique à la qualité du lait. Plus globalement, des index peuvent être regroupés en un seul index, l'index de synthèse (ISU) qui diffère selon les races. Il permet de placer un animal par rapport aux performances voulues de la race. Cet ISU regroupe des critères basés sur l'apport de l'animal : la production, la vitesse de traite, la morphologie, mais aussi des index de fertilité, et de santé en particulier de santé de la mamelle et longévité [169].

La France a été un des pays pionniers dans l'évaluation génomique puisqu'il s'agit du premier pays européen à avoir publié ses index génomiques bovins [164]. Les diversités des programmes de sélection dans le monde reposent essentiellement sur les populations de référence utilisées pour définir les liens génotypes/phénotypes : plusieurs milliers d'animaux composent cette population dans des pays tels que les États-Unis, le Canada ou encore le Danemark [170].

En espèce bovine, Interbull (International Bull Evaluation Service) a été créé dans les années 1970, et à présent regroupe et compare les différentes évaluations génétiques et génomiques bovines. Il s'agit donc d'un service d'évaluation internationale ayant, entre autres buts, celui d'assurer la continuité de la comparabilité des différentes valeurs génétiques et génomiques entre pays [171].

1. Mise en place de la sélection génomique en bovin

La sélection génomique est en place en France chez les bovins depuis 2009, pour les bovins laitiers grâce à une population de référence de plus de 76 000 animaux au total, de races Prim'Holstein, Pie-Rouge, Brune, Normande, Montbéliarde, Abondance, Tarentaise, Simmental et Vosgienne [95], [172]. Les trois races bovines laitières majoritaires sont représentées en plus de 6 autres races. La race Pie-Rouge a été choisie de par sa proximité généalogique avec la race Prim'Holstein. La définition d'index précis en race Brune a été

permise par le biais d'une collaboration mondiale. Les races Abondance, Tarentaise, Simmental et Vosgiennes ont elles aussi été choisies, avec des tailles de population suffisantes pour avoir une population de référence permettant d'avoir des index précis [173]. La population de référence bovine se compose notamment de 30 000 taureaux en race Holstein, 2 650 taureaux et 22 000 vaches de race Montbéliarde, 2 300 taureaux et 11 800 vaches Normande et plus de 6 000 taureaux de race Brune [174]. La population de référence en race Abondance compte 345 taureaux et 1 885 vaches, on dénombre également 301 taureaux et 1 273 vaches Tarentaise, 339 taureaux et 215 vaches Simmental ainsi que 56 taureaux et 1 076 vaches Vosgienne [173]. Tous les éleveurs bovins ont eu accès à la sélection génomique pour les femelles dès Janvier 2011. Plus de 230 000 bovins ont ainsi pu être indexés [173].

2. Mise en place de la sélection génomique en ovin et caprin

En ovin, le génotypage de grands nombres d'animaux a été mené dès 2002, afin d'améliorer la résistance à une pathologie ovine répandue : la Tremblante. Afin de contrer l'apparition de cette pathologie, l'état français a mis en place un programme de lutte en se basant sur cet outil qu'est la sélection génomique. En effet, le génotypage d'individus a été mené pour le gène PrP, qui est responsable de la maladie, dans ce cadre.

La sélection génomique est en place depuis 2015 pour la race Lacaune lait, ce qui a été permis par la disponibilité de la puce à SNPs ovine SNP50 sur laquelle sont déposées des sondes pour détecter environ 50 000 SNPs, et depuis 2017 en races Manech tête rousse, Manech tête noire, Basco-Béarnaise et Corse [175], [176]. La population de référence en ovin laitier est composée de plus de 9 000 animaux : 5 368 béliers Lacaune, 2 238 Manech tête rousse , 667 de race Basco-Béarnaise, 510 Manech tête noire et 243 Corse [176].

En caprin, la sélection génomique est mise en place [177]. Une puce à SNPs caprine permettant de détecter plus de 50 000 SNPs a été mise en production et va donc permettre cette sélection génomique, d'autant qu'une population de référence de 825 boucs a également été mise en place [178], [179]. 471 animaux de race Alpine et 354 animaux de race Saanen composent cette population de référence [179].

6. Les différentes régions du génome

Le génome est l'ensemble du matériel, et donc de l'information génétique d'un organisme. Ce matériel génétique est contenu dans les cellules, sous forme d'ADN qui constituent les chromosomes. Des régions codantes (qui codent des protéines) mais également des parties non-codantes composent le génome. Chez les mammifères, les gènes codant des protéines ne représentent qu'environ 2% du génome [180]. Des données en bovin et ovin estiment la part des régions non-codantes à 99% [181]. Néanmoins, près de 80% du génome humain a une activité biochimique, même si cela n'aboutit pas directement à des protéines [182]. Les régions non-codantes sont d'autant plus importantes qu'on considère qu'environ 90% des SNPs identifiés comme étant associés à des pathologies humaines sont situés dans ce non-codant [162]. Elles ne codent pas directement des protéines mais de ces régions peuvent découler des ARNs fonctionnels.

a. Les régions codantes

Les régions codantes sont généralement conservées entre espèces [183]. La conservation de régions génomiques entre espèces peut démontrer leur importance, le fait qu'elles aient un rôle nécessaire. Les régions codantes vont être traduites en protéines. Ces protéines sont considérées comme étant l'élément principal de la cellule, tant au niveau fonctionnel que structurel [184].

b. Les régions non codantes

L'essentiel du génome humain est composé d'ADN ne codant pas des protéines [185]. Ces régions non-codantes du génome avaient été nommées "Junk DNA" ou "ADN poubelle" lors de leur découverte car leur utilité n'avait pas encore été décelée [186]. En réalité, une grande partie des régions ne codant pas des protéines sont fonctionnelles [187]. De plus, ces séquences non-codantes sont parfois retrouvées dans différents organismes et avec un grand degré de conservation, ce qui permet de déduire qu'elles ont certainement une fonction [188].



Figure 10 : Régulation de la transcription de gènes par répresseur, activateur et promoteur

Le non-codant a des rôles de régulation : ce sont des régions pouvant permettre de réguler la transcription de gènes. Elles peuvent définir "où et quand" des gènes sont actifs [189]. Ces régions régulatrices sont recherchées en amont du site de démarrage de la transcription (Transcription Start Site, TSS) [190]. On peut citer les rôles des activateurs (enhancers) : ces portions du génomes se fixent à des facteurs de transcription (des protéines pouvant réguler l'expression de gènes) et ainsi peuvent accroître l'expression de gènes associés [191]. On pourrait les qualifier d'"amplificateurs" de l'expression. Ces activateurs ont des relations fortes avec d'autres régions que sont les promoteurs [192], [193]. Les promoteurs sont des régions situées en amont et proches de gènes, qui jouent également un rôle de régulation de l'expression en interagissant avec des facteurs de transcription pour initier la transcription de gènes *via* leur faculté à être des sites de fixation pour l'ARN polymérase responsable de cette transcription. Une séquence de ces promoteurs est nommée TATA box, il s'agit d'une séquence très conservée sur laquelle se fixent des facteurs de transcription. Cela déclenche alors l'initiation du complexe de transcription de l'ARN polymérase et donc la transcription (*Figure 10*) [194]–[196]. Des polymorphismes dans ces régions non codantes peuvent induire des modifications de niveau d'expression de gènes. Par exemple, des SNPs ont été découverts dans le promoteur du gène *IL10* chez l'humain, induisant un plus faible niveau d'expression de *IL10* et parfois associés à des prédispositions à des pathologies telles que l'hépatite C [197], [198]. Plus globalement, des outils de prédiction de variants génétiques situés dans ces TATA box et associés à des pathologies sont développés. Ainsi, des SNPs sont prédits dans ces régions avec des liens avec des pathologies telles que l'obésité, l'athérosclérose, Alzheimer ou encore des maladies auto-immunes chez humaines [199]. Des polymorphismes peuvent influencer sur des promoteurs de gènes et ainsi avoir un effet indirect sur des gènes. Chez la chèvre, un SNP affecte négativement le promoteur du gène de caséine *CSN2* et pourrait donc avoir un effet sur la présence de caséine β dans le lait de chèvre [200]. D'autres régions non-codantes peuvent être retrouvées dans des gènes ou être intergéniques. Des éléments ont la faculté de moduler l'expression de gènes, tels que les répresseurs (silencers), qui peuvent abaisser le niveau de transcription, ou les ARNs.

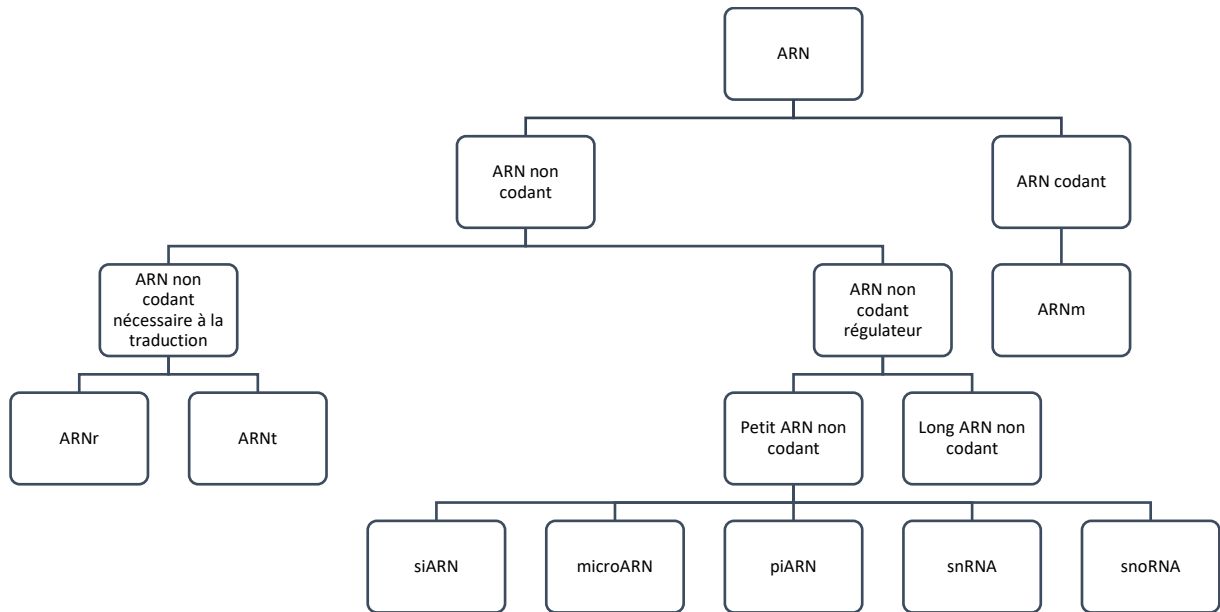


Figure 11 : Différentes classes d'ARN

On décompte plusieurs familles d'ARN ne codant pas des protéines. Ces ARNs n'amènent pas directement à la production de protéines, mais de nombreux systèmes biologiques sont impactés par les ARNs (*Figure 11*) [201].

Les ARNs de transfert (ARNt) et ARN ribosomiaux (ARNr) permettent l'assemblage d'acides aminés et sont ainsi impliqués dans la création de protéines [189]. Les ARNs de transfert sont considérés comme acteurs centraux de la machinerie de traduction permettant la synthèse de protéines en transportant les acides aminés (186).

Les petits ARNs nucléaires (small nuclear RNA, snRNA) sont une classe d'ARNs non codants impliqués dans des événements d'épissage [203].

Les petits ARNs nucléolaires (small nucleolar RNA, snoRNA) sont longs de 60 à 300 paires de bases environ et interagissent avec d'autres ARNs, les ARNr, en les modifiant [203], [204]. Ils sont nécessaires pour la maturation du pre-ARNr [205]. Les snoRNA peuvent également cibler des snRNAs ou des ARNm [206].

D'autres petits ARNs non codants tels que les "P-element-induced wimpy testis interacting RNAs", appelés piwiARNs (piARN), ou les small interfering RNAs appelés siARNs ont également été découverts chez les mammifères [207]. Les piARNs, longs de 24 à 30 nucléotides, ont des rôles de méthylation de l'ADN et de répression de transposons [204]. Un complexe est formé entre piARNs et protéines PIWI. Ce complexe peut alors supprimer l'expression d'éléments transposables *via* un clivage des transcrits, qui s'effectue grâce à l'appariement du piARN sur le transcrit, ou extinction de gène par l'hétérochromatine [204].

Les longs ARNs non codants (long noncoding RNAs, lncRNAs) sont eux aussi des ARNs non codants, de plus de 200 nucléotides [203]. Peu d'entre eux ont une fonction bien identifiée, des rôles de régulation de gènes ont toutefois été mis en évidence [208]. Cette régulation peut s'effectuer *via* les promoteurs ou la stabilité de l'ARNm [203]. Les lncRNAs peuvent être circulaires ou linéaires. Ils peuvent être des médiateurs de modifications épigénétiques, dont l'accessibilité de la chromatine, ou également exercer des rôles de contrôle transcriptionnel, en induisant l'action de facteurs de transcription d'amplificateurs (enhancers) ou en interagissant avec des complexes d'initiation par exemple. Ils peuvent également exercer des



Séquence :

5'-CCC GGCUGAGGUAGGAGGUUGUAUAGUUGAGGAGGACACCCAAGGAGAUACUAUACGGCCUCCUAGCUUCCCCAGG-3'

Figure 12 : Structure en tige-boucle du précurseur d'un microARN (ici bta-let-7e). Les nucléotides orange composent le microARN mature, la région "seed" est composée des nucléotides 2 à 8 du mature, ici encadré en orange.

rôles de contrôle post-transcriptionnel, en interagissant dans des processus d'épissage ou de traduction d'ARNm [209]. Ces ARNs longs non codants sont impliqués dans des processus tels que l'inactivation de l'X chez les mammifères [204]. Leur impact de régulation peut s'effectuer de façon localisée : ils peuvent affecter l'expression de gènes codants situés proches d'eux (appelé phénomène de transvection) [209]. Ces ARNs non codants sont étudiés dans le cadre de pathologie humaines, d'immunité [210].

Les lncRNAs ont également un rôle dans la régulation de microARNs : ils peuvent avoir la capacité de réduire la régulation des microARNs sur les ARNm [211]. Pour exemple, le lncRNA nommé CHRF a été étudié dans le cadavre d'une pathologie cardiaque humaine. CHRF peut se fixer sur le microARN *miR-489* et ainsi réduire son expression [212].

On peut citer, parmi les lncRNAs, les grands ARNs non codants intergéniques (large intergenic non-coding RNAs, lincRNAs) [204]. Ces ARNs sont conservés dans les génomes de mammifères. Un lincRNA régulé par p53 aurait un rôle dans la réponse transcriptionnelle à un dommage de l'ADN et donc dans le déclenchement de l'apoptose [213], [214].

Les microARNs sont eux aussi des ARNs, ayant un rôle de régulation post-transcriptionnelle de l'expression de gènes.

i. Les microARNs

Les microARNs sont des petits ARNs non codants (*Figure 5*) [215]. Ces microARNs sont constitués de 17 à 22 nucléotides [216], [217]. Ils sont très conservés entre espèces et participent à la régulation post-transcriptionnelle de gènes *via* le ciblage d'ARN messagers (ARNm) en s'y appariant. Cet appariement s'effectue grâce à la région "seed" du microARN, la région de reconnaissance entre microARN et ARNm, longue de 7 nucléotides : les nucléotides 2 à 8 du microARN. Par complémentarité de séquences, le ciblage d'ARNm cibles peut s'effectuer. Il existe également une région supplémentaire d'appariement entre microARN et ARNm, constituée des nucléotides 13 à 16 du microARN mature [218] (*Figure 12*).

Le premier microARN a été découvert en 1993, chez l'espèce modèle *Caenorhabditis elegans*. Il s'agit du microARN nommé *lin-4* [219]. Par la suite, en 2000, le microARN *let-7* a été identifié, encore une fois chez ce nématode *C. elegans*. Son rôle dans la transition entre le stade larvaire et le stade adulte, et donc son rôle dans un mécanisme complexe et essentiel a alors été détecté [220].

Les différents éléments impliqués dans la biogenèse de microARNs ont été identifiés en 2001, pour DICER, et en 2003 pour DROSHA et le complexe d'inhibition induit par ARN (RNA-Induced Silencing Complex) : RISC [221], [222].

Les microARNs ont un rôle important dans de nombreux processus biologiques, dont le développement d'organes : ce sont des régulateurs du développement, notamment de la glande mammaire [223].

1. microARNs et espèces animales

Le nombre de microARNs listés chez dans l'espèce humaine s'élève à 2654 microARNs. En murin, 1978 microARNs sont listés, et 437 chez l'espèce modèle *Caenorhabditis elegans*. Dans les espèces de ruminants laitiers, les nombres de microARNs diffèrent de 1025 microARNs bovins, 436 microARNs caprins à 153 microARNs ovins. Ces chiffres sont issus de la dernière version de miRBase (version 22.1, Octobre 2018), base de données de microARNs la plus utilisée [224].

Des données de miRNomes, qui regroupent l'ensembles de microARNs détectés dans un tissu donné, permettent de comparer la présence ou absence de microARNs et leur abondance entre des espèces ou des races. Des miRNomes mammaires, listes de microARNs exprimés dans la glande mammaire, ont été réalisés par Le Guillou *et al.* chez les espèces murines et bovines [225]. Cette étude a permis de comparer la présence de microARNs dans la glande mammaire en lactation dans deux espèces distinctes. Un total de 164 microARNs sont identifiés en murin et 167 en bovin. Les profils d'abondance de ces microARNs sont très conservés entre bovin et murin. Parmi les 30 microARNs les plus exprimés dans la glande mammaire de chacune des espèce, 24 sont communs [225].

Des études se sont basées sur l'impact de la race sur le miRNome de glande mammaire ou de lait. Des différences d'abondance de microARNs ont été mises en évidence dans les miRNomes de lait entier entre races Holstein et Normande [226]. De même, des différences significatives ont été retrouvées dans les miRNomes de glandes mammaires de vaches en lactation entre races Holstein et Montbéliarde [227].

2. Bases de données existantes

Plusieurs bases de données ont été développées ces dernières années afin de lister les microARNs. Elles sont pour la plupart spécifiques : spécifiques d'espèces ou parfois même spécifiques d'une pathologie. On peut citer pour exemple des bases de données de microARNs associés à l'épilepsie ou au cancer chez l'Homme : respectivement EpimiRBase et miRCancer [228], [229], ou plus largement les pathologies humaines pour la base de données MSDD – MicroRNA SNP Disease Database [230].

La base de données la plus généralement utilisée dans le cas des microARNs est miRBase, regroupant les microARNs de 271 espèces différentes [224]. Plus de 2 500 microARNs matures sont désormais listés chez pour l'espèce humaine, ciblant plus de la moitié des ARNm du génome. D'autres bases de données regroupent des microARNs d'animaux, plantes et virus, comme miRNEST 2.0 (pour plus de 400 espèces différentes) [231] ou encore miROrtho pour 46 animaux [232]. Plus récemment, la base de données miRCarta a été créée et liste les microARNs de 148 espèces, en se basant seulement sur la prédiction de nouveaux microARNs [233]. Un outil nommé miRTools-Gallery contient plus de 1 000 outils pour l'étude des microARNs, comme l'identification ou la prédiction de cibles de microARNs [234]. Une nouvelle base de données regroupant les microARNs publiés de ruminants laitiers a été créée lors de cette thèse : RumimiR. Il s'agit d'une base de données regroupant les données de la littérature, listant les microARNs décrits et ajoutant des informations sur ceux-ci (races des animaux, échantillons, contexte d'étude, etc...) lorsqu'elles sont décrites. La création de RumimiR a donné lieu à une publication, et est décrite dans ce manuscrit.

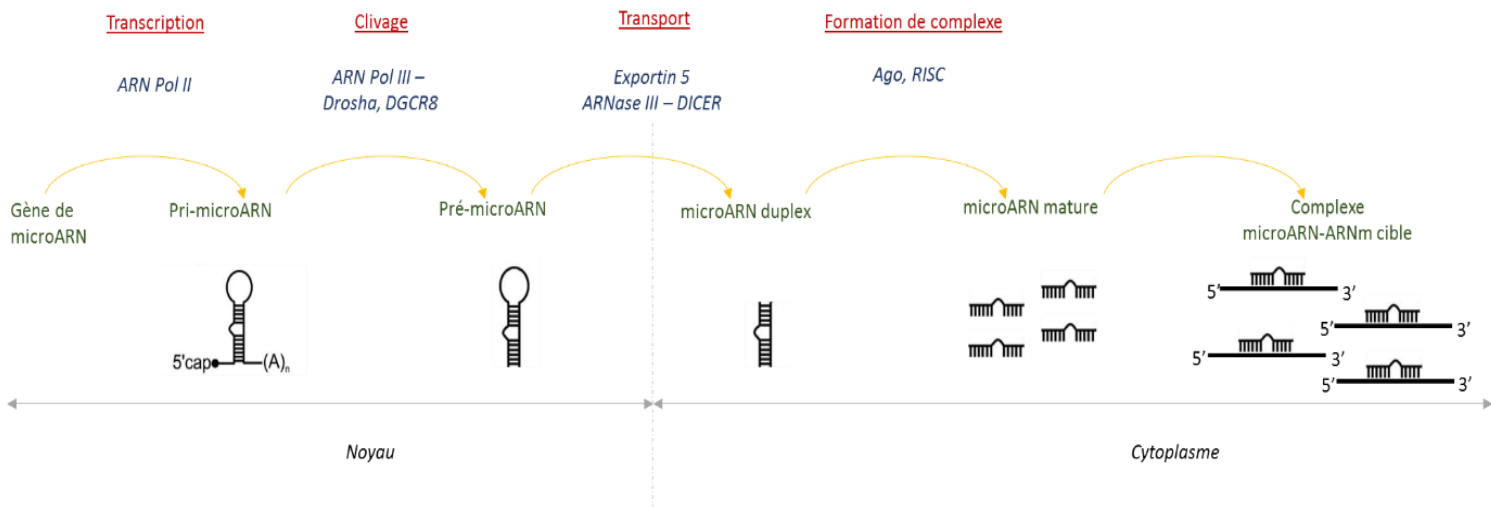


Figure 13 : Biogenèse des microARNs

3. Biogenèse de microARNs

La biogenèse du microARN se déroule en 5 grandes étapes : du gène de microARN au microARN mature, dans le noyau de la cellule puis dans le cytoplasme. La première de ces étapes se déroule dans le noyau il s'agit de la transcription, par l'ARN polymérase Pol II, d'un gène de microARN en microARN primaire. Celui-ci va ensuite être clivé lors de la 2^{ème} étape, par une ARN polymérase Pol III, une enzyme Rnase III –ou ribonucléase 3– en l'occurrence Drosha et *DGCR8* ("DiGeorge Critical Region 8") en un précurseur de microARN.

Drosha et *DGCR8* composent ensemble un complexe spécifique, permettant cette transformation d'un pri-microARN en un pré-microARN en excisant le pri-microARN. Ce pré-microARN –ou précurseur de microARN– est facilement reconnaissable de par sa structure en tige-boucle (stem-loop) caractéristique. Par la suite, ce pré-microARN va être exporté du noyau vers le cytoplasme par l'Exportin-5, et est alors pris en charge et clivée par l'enzyme Rnase III Dicer. Le microARN duplex ainsi formé devient un microARN mature par l'intervention d'une protéine Argonaute (Ago), qui va alors prendre en charge le microARN mature, devenir le complexe RISC, comprenant Ago et le microARN, et donc permettre la présentation du microARN mature à des ARNm cibles (*Figure 13*).

4. Mode d'action

Le rôle des microARNs est la régulation de l'expression de gènes, de façon post-transcriptionnelle [235]. Les microARNs s'apparient donc à des ARNm cibles, et une répression de la traduction ou une dégradation de l'ARNm ciblé peuvent être observées. Ces mécanismes sont mis en place selon l'appariement du microARN à cet ARNm au niveau du site de reconnaissance du microARN, appelé "seed". Une complémentarité plus élevée induira la dégradation, plus faible ce sera une répression de traduction [236]–[238]. La régulation des microARNs a lieu dans 48% des cas par une répression de la traduction, dans 29% des cas il s'agit d'une dégradation de l'ARNm. Pour 23% des cas le microARN agit *via* une combinaison de ces deux actions, dans le cas d'études chez la souris [239].

Une fixation de microARNs dans une autre région d'ARNm (hors 3'UTR) est également possible: dans la région séquence codante ou CDS (Coding Domain Sequence) [240]. Dans le cadre d'une étude sur le cerveau de souris, la fixation de microARN en 3'UTR et la fixation en CDS ont des fréquences respectives de 40% et 25%. D'autres sites de fixation sont présents mais moins fréquents : 12% des fixations se font au niveau d'introns, ou encore 1% au niveau de la région 5'UTR [241].

Un microARN peut cibler plusieurs ARNm, tout comme un ARNm peut être ciblé par plusieurs microARNs. On compte parfois entre 100 et 200 cibles pour un seul microARN [242], [243]. Ceci est notamment dû à l'imperfection possible de la complémentarité entre microARN et ARNm cible.

ii. Variants de microARNs

La première notion de SNPs dans les microARNs est apparue en 2005, dans une étude sur l'Homme [244]. Par la suite, la distinction a été faite entre les SNPs affectant l'expression du microARN et ceux affectant l'expression d'ARNm cibles.

Les variants génétiques de microARNs sont essentiellement étudiés chez l'humain, surtout dans le cadre de pathologies. Dans ce cadre, les études se basent sur les SNPs associés à la susceptibilité (la hausse ou la diminution du risque d'apparition de la pathologie) ou associés au pronostic, à l'espérance de vie (lorsque la pathologie est présente).

De très nombreuses pathologies sont étudiées dans ce cadre de variants génétiques de microARNs : des cancers, asthmes ou autre maladies auto-immunes [245]–[247]. Des bases de données telles que miRSNPs ou MSDD regroupent les microARNs et leurs polymorphismes associés à des pathologies [230], [248].

Une étude de 2007 a défini qu'environ 10% des pré-microARNs humains contient des SNPs, et moins de 1% des microARNs ont des SNPs dans leur région seed [249]. Les polymorphismes des microARNs ont des impacts différents selon leur localisation.

Mutation affectant le niveau d'expression

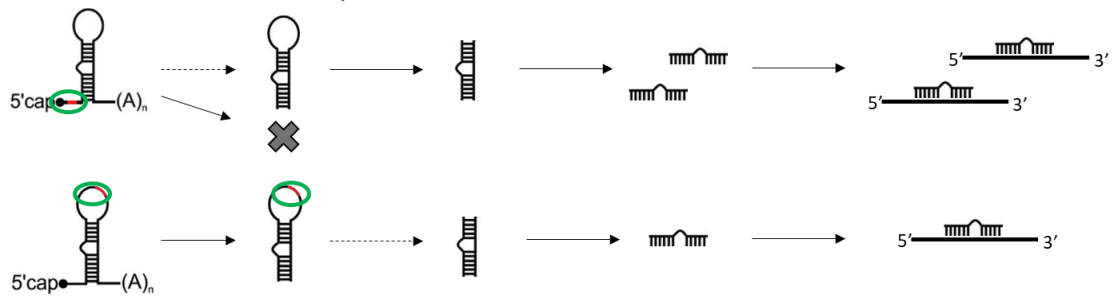


Figure 14 : Impacts potentiels de mutations de microARNs dans le précurseur

1. SNPs affectant l'expression du microARN

Les variants génétiques situés dans le précurseur d'un microARN peuvent moduler le niveau d'expression du microARN. En effet, ils peuvent modifier la biogenèse du microARN : les différents éléments permettant d'aboutir à un microARN mature ciblent des régions du pri-microARN ou du pré-microARN. De ce fait, un polymorphisme dans ces régions peut induire une non prise en charge et donc une biogenèse qui ne peut se faire correctement. Une baisse d'expression de microARN peut alors être observée, et peut induire une diminution du nombre de microARNs matures. L'action de régulation de gènes ciblés par ce microARN est alors elle aussi moins importante : ces gènes peuvent donc s'exprimer davantage. Des modifications de la structure du microARN peuvent impliquer une altération du processus de maturation du microARN (*Figure 14*).

Une étude sur le cancer du sein a montré l'importance de variants de microARNs. Dans cette étude, un SNP est détecté dans le précurseur du microARN *miR-146a*. Ce microARN cible *BRCA1/BRCA2* qui sont des gènes connus comme étant fortement liés à la survenue du cancer du sein. Ce variant génétique implique une baisse de l'expression du microARN mature, et ce *miR-146a* est alors associé au risque de survenue du cancer du sein. Également dans cette publication, un second SNP est retrouvé. Il est situé dans le précurseur du microARN *miR-27a* qui est, lui, de type oncogène. Ce variant implique une baisse de l'expression de ce microARN mature. Du fait du caractère oncogène-like de ce gène, ce variant baisse l'expression de ce gène et *a contrario* du variant du *miR-146a* a un effet protecteur du cancer du sein [250].

Mutation affectant la reconnaissance de cible

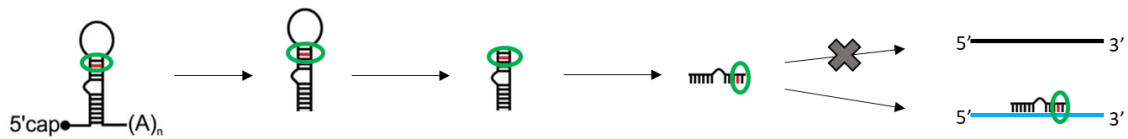


Figure 15 : Impacts potentiels de mutations de microARNs dans la région "seed"

microARNs	Impacts	Espèce	Références
miR-21	Exprimé abondamment dans la glande mammaire	Bovin	[251]
miR-23a	Exprimé abondamment dans la glande mammaire	Bovin	[251]
miR-24	Exprimé abondamment dans la glande mammaire	Bovin	[251]
miR-143	Exprimé abondamment dans la glande mammaire	Bovin	[251]
miR-30b	Gouttelettes lipidiques Croissance des souriceaux	Murin	[252]
miR-145	Synthèse acides gras dans le lait	Caprin	[253]
miR-24	Métabolisme des acides gras	Caprin	[254]
miR-103	Synthèse matière grasse dans le lait	Caprin	[255]
miR-30e	Métabolisme des acides gras	Caprin	[256]
miR-15a	Métabolisme des acides gras	Caprin	[256]
miR-15a	Expression des caséines	Bovin	[257]
miR-141	Ciblage de gènes majeurs pour la lactation	Bovin	[258]
miR-484	Ciblage de gènes majeurs pour la lactation	Bovin	[258]
miR-500	Ciblage de gènes majeurs pour la lactation	Bovin	[258]
miR-486	Ciblage de gènes majeurs pour la lactation	Bovin	[259]
miR-29a	Ciblage de gènes majeurs pour la lactation	Bovin	[260]
miR-29b	Ciblage de gènes majeurs pour la lactation	Bovin	[260]
miR-29c	Ciblage de gènes majeurs pour la lactation	Bovin	[260]
miR-181	Synthèse matière grasse dans le lait	Bovin	[261]
miR-33a	Métabolisme lipidique	Bovin	[262]
miR-152	Métabolisme lipidique	Bovin	[262]
miR-224	Métabolisme lipidique	Bovin	[262]
miR-183	Métabolisme lipidique	Bovin	[263]
miR-424(322)/503	Involution	Murin	[264]

Tableau 5 : Synthèse de résultats de liens entre microARNs et production laitière

2. SNPs situés dans la région seed

Dans le cas de variants génétiques situés dans la "seed" du microARN, on peut observer une modification d'ARNm cibles. Plus précisément, un variant génétique situé dans la région de reconnaissance du microARN avec ses ARNm cibles peut induire, du fait alors d'un appariement modifié, d'une non reconnaissance d'un ARNm cible. De ce fait, un ARNm et donc un gène qui devrait être modulé par ce microARN ne l'est pas : il n'y a pas de régulation post-transcriptionnelle. Parfois, un variant peut engendrer la reconnaissance d'une nouvelle cible : un ARNm qui n'est pas une cible du microARN, mais la présence de celui-ci induit la possibilité d'un appariement entre ce microARN et l'ARNm. Dans ce cas, on peut observer une régulation post-transcriptionnelle de cette nouvelle cible, et donc la modulation de l'expression d'un nouveau gène (*Figure 15*).

iii. Lien microARNs et production laitière

Différents microARNs ont d'ores et déjà été identifiés, notamment chez les ruminants, comme étant liés et donc impliqués dans des caractères laitiers (*Tableau 5*).

Tout d'abord, la présence de microARNs est effective dans le lait (dans le colostrum et le lait mature) et plus précisément dans différentes fractions du lait [265], [266]. Le colostrum présente une plus forte quantité de microARNs que le lait mature, essentiellement des microARNs liés à l'immunité et au développement [267]. Dans la glande mammaire l'expression des microARNs présents est modulée par le stade de lactation, probablement dû à leurs divers rôles (prolifération cellulaire, lipogenèse, immunité, apoptose, par exemple) [268]. Chez la vache, environ 10% des microARNs sont différentiellement exprimés entre les stades de lactation et non-lactation [258]. Des microARNs peuvent aussi être marqueurs de l'état pathologique de la glande mammaire, dans le cas de présence ou non de mammites [269], [270].

Parmi les microARNs liés à des caractères laitiers, certains d'entre eux, comme les *miR-21*, *miR-23a*, *miR-24* ou encore *miR-143* n'ont pas de rôle précis défini : ils sont exprimés abondamment dans la glande mammaire chez des vaches Holstein, suggérant un impact sur la production du lait, *via* le développement de ce tissu par exemple [251].

D'autres microARNs ont des rôles dans la production et/ou la composition du lait décrits plus précisément. L'impact du microARN *miR-30b* sur la lactation a été étudié par Le Guillou *et al.* grâce à des souris transgéniques surexprimant le microARN dans les cellules épithéliales mammaires. Un défaut de formation de gouttelettes lipidiques a été observé : les gouttelettes lipidiques sont moins nombreuses mais d'une aire plus grande en moyenne dans la glande mammaire de souris transgéniques que dans celle de souris contrôle. De plus, la croissance des souriceaux nourris par un lait issu de ces souris transgéniques est en moyenne 37% moins importante à 15 jours après naissance que celle de souriceaux nourris par un lait de souris non-transgéniques [252].

Le microARN *miR-126-3p* est en lien avec le métabolisme lipidique, dans la glande mammaire, de par sa régulation de *FASN* (fatty acid synthase) qui est un gène pouvant avoir un impact sur le pourcentage de matière grasse dans le lait [271], [272]. Le microARN *miR-150* a lui été étudié pour son action sur des gènes tels que *FASN*, comme dans l'étude précédente, mais aussi *OLAH* ou *STAT5B*, essentiels pour lactation et un intermédiaire du récepteur de la prolactine respectivement. Cela entraîne une modification de la lactation et plus précisément de la synthèse lipidique [273]. L'expression constitutive de ce microARN a pour conséquence une réduction de la synthèse d'acides gras, et chez les souriceaux étudiés Heinz *et al.* ont observé une hausse de la mortalité [273].

Des études se concentrant sur le lien entre microARN et lactation ont également été réalisées chez des ruminants, essentiellement bovins et caprins. Le rôle du microARN *miR-145* dans la synthèse d'acides gras dans le lait a été mis en avant en culture cellulaire par Wang *et al.*, en 2017, dans des cellules épithéliales mammaires caprines [253]. Précédemment, le microARN *miR-24* a également été identifié comme impliqué dans le métabolisme des acides gras en cellules épithéliales mammaires caprines [254].

Toujours chez les caprins, le microARN *miR-103* est un régulateur de la synthèse de matière grasse dans le lait, qui est par ailleurs plus exprimé lors de la lactation (en milieu de lactation) qu'en dehors de cette période. Une surexpression de ce microARN a permis de mettre en évidence son rôle tant sur le plan de la synthèse de matière grasse (avec pour conséquence une accumulation des gouttelettes adipeuses) que son rôle de régulation de gènes tels que *FASN* et *SCD*, qui sont inclus dans le processus de synthèse et d'insaturation des acides gras [274].

D'autres publications ont également mis en évidence l'impact de divers microARNs sur des caractères laitiers, tels que la régulation du métabolisme des acides gras du lait par *miR-30e* et *miR-15a* en culture cellulaire chez des cellules épithéliales mammaires caprines [256]. Le microARN *miR-99a-3p* a lui une incidence sur le taux de calcium dans le lait chez les caprins [275].

En bovin également, des études se sont basées sur le lien microARN/production et composition du lait. Le microARN *miR-15a* a été étudié mettant en avant son rôle dans la régulation de lactation et son impact sur les caséines dont l'expression diminue lors de la transfection de ce microARN sur des cellules bovines [257]. Des microARNs tels que *miR-141*, *miR-484*, ou encore *miR-500* ont été décelés comme potentiellement essentiel à la lactation chez les bovins, particulièrement de par leur ciblage de gènes majeurs pour ce phénomène : *STAT5* et *HK2* [258]. Ces gènes sont impliqués dans la lactation car il s'agit respectivement d'un intermédiaire de la prolactine essentiel au développement des alvéoles dans la glande mammaire et *HK2* est un gène impliqué dans le transport de glucose et exprimé seulement lors de la lactation [276]–[279]. Une étude sur des vaches de race Holstein a mis en évidence l'importance du microARN *miR-486* sur la lactation par l'intermédiaire de sa régulation du gène *PTEN* ainsi que des gènes *AKT* ou *mTOR*. Sa surexpression *in vitro* dans des cellules épithéliales mammaires provoque une augmentation des concentrations de composants tels que la β -caséine ou le lactose [259]. Ces gènes cibles *AKT* et *mTOR* sont eux connus comme partie intégrante d'une voie de signalisation impliquée dans la synthèse protéique dans la glande mammaire bovine [280]. En race Holstein, les microARNs *miR-29* affectent la lactation : les *miR-29a*, *miR-29b* et *miR-29c* ciblent les gènes *DNMT3A* et *DNMT3B* mais aussi peuvent

réguler l'expression du gène de la caséine *CSN1S1* en modifiant la méthylation de son promoteur. Des conséquences sur la synthèse de protéines du lait, le taux de lactose et de lipides sont alors possibles [260]. On peut également citer l'étude de Lian *et al.* qui ont mis en évidence le lien entre le microARN *bta-miR-181* et la synthèse de matière grasse du lait ou encore Shen *et al.* liant les microARNs *bta-miR-33a*, *bta-miR-152* et *bta-miR-224* et le métabolisme lipidique en cellules épithéliales mammaires primaires dérivées de deux vaches Holstein produisant des taux de matières grasses très différents [261], [262].

Le rôle du microARN *bta-miR-183* dans ce même métabolisme, avec l'implication du gène de la prolactine, a été mis en évidence récemment en cellules épithéliales mammaires bovines [281].

Une étude utilisant la souris comme animal modèle cite le cluster de microARN *miR-424(322)/503* (comprenant les *miR-424(322)* et *miR-503*) comme étant un élément régulant l'involution de la glande mammaire [264]. L'expression de ce cluster de microARNs étant plus élevée lors de l'involution que lors des autres étapes (puberté, gestation, lactation notamment) et les souris KO pour *miR-424(322)/503* montrant un défaut d'involution, le rôle biologique important de ce cluster lors de l'involution de la glande mammaire est alors décrit [264].

Plusieurs études mettent en évidence le rôle de polymorphismes situés dans les régions 3'-UTR, impactant le ciblage de microARNs, dans la modification de caractères laitiers. Le polymorphisme du gène de la caséine *CSN1S1*, chez la chèvre avec l'impact sur le ciblage par *miR-101* a été étudié [282]. En bovin, les gènes *NCF4* ou *A2M* en lien avec la susceptibilité aux mammites [283], [284], ou encore *OLR1* ou *FAD2* en lien avec la synthèse de matière grasse [285], [286] et leurs polymorphismes dans les régions de fixation des microARNs sont mis en évidence. De même pour le gène *AMPK γ 1* et l'impact de son SNP dans la région 3'UTR sur la production de lait en race Holstein ou encore le gène du récepteur de la prolactine *PRLR* [200], [287].

Résultats

1. Création d'une base de données de microARNs de ruminants

Étudiant les microARNs des espèces bovines, caprines et ovines dans un contexte de production laitière, il est important d'évaluer la quantité de microARNs présents dans ces espèces de ruminants. Non seulement les microARNs officiellement répertoriés et nommés, donc recensés dans miRBase, mais également tous ceux détectés et décrits dans les publications.

Comme décrit dans l'introduction de cette thèse, la base de données de microARNs existante et la plus usitée à ce jour est miRBase [288]. Il s'agit d'une base de données, créée en 2006 listant les microARNs connus et publiés pour des espèces animales et végétales [288].

La dernière version de cette base de données miRBase est en date d'Octobre 2018. À ce jour, un total de 1 025 microARNs matures bovins, 436 microARNs matures caprins et 153 microARNs matures ovins sont présents dans miRBase.

Néanmoins, cette base de données n'est pas complète, puisque des données publiées issues de séquençage haut débit ne sont pas prises en compte pour le moment : seules 5% de ces données y sont présentes [289]. La base de données n'est donc pas exhaustive et, de plus, des informations importantes dans le cadre d'études d'animaux tels que les ruminants, comme la race ou le tissu dans lesquels ont été décelés le microARN, ne sont pas précisées pour chaque microARN. Dans un contexte de production laitière chez les ruminants, ces critères et particulièrement la race de l'animal est une information d'importance.

Afin d'établir une liste exhaustive et détaillée des microARNs de ces espèces de ruminants laitiers, il a été décidé de créer une base de données détaillée regroupant l'ensemble des microARNs décrits dans la littérature en bovin, caprin et ovin. Cette base liste l'ensemble des données tout en y apposant des détails décrits dans la publication associée tels que le tissu d'origine des microARNs décrits, la race, le stade de lactation, l'âge des animaux, le contexte de l'étude : quelles conditions ont été observées (nutrition, pathologie, production...), notamment.

Cette base de données, nommée RumimiR, est donc une base de données de microARNs de ruminants publiés exhaustive et détaillée, offrant des possibilités de tri des données selon les différentes informations ajoutées. Pour cela elle est mise à jour environ deux fois par an afin d'y ajouter les microARNs nouvellement décrits. La version des génomes utilisés est elle aussi mise à jour lorsque c'est nécessaire. Elle est donc adaptée pour des recherches sur les ruminants. Elle est accessible gratuitement en ligne à l'adresse <http://rumimir.siginae.org/>.

Des microARNs issus de 89 publications différentes sont détaillés à ce jour dans cette base de données RumimiR, pour un total de 13 182 microARNs. Pour chacun de ces microARNs, les informations collectées sont rassemblées dans 29 champs différents. Ces informations concernent le microARN en tant que tel, avec sa séquence, sa position, la publication dans laquelle est décrit le microARN avec sa référence et le lien associé. De plus, des informations sont précisées au sujet de l'étude menée et des échantillons : la race, l'âge ou encore le stade de lactation des animaux sont mentionnés lorsque c'est le cas dans la publication associée. Les logiciels utilisés pour la détection de ces microARNs ainsi que les éventuels critères utilisés pour déterminer les microARNs sont notés.

Un code de faux-positivité est élaboré à partir de données ajoutées pour chaque microARN. Ce code prend en compte 4 critères, qui peuvent être des indices sur la réelle existence des microARNs : i) sa présence ou non dans RumimiR au préalable, en effet un microARN déjà décrit et donc détecté dans une autre publication a plus de chance d'être un vrai microARN ; ii) son positionnement multiple ou non, puisque un microARN dont la position sur le génome est multiple ne peut être définie précisément et est possiblement un faux-positif ; iii) son identité de séquence avec un microARN humain et/ou murin, car un microARN présentant une identité de séquence avec un microARN d'une autre espèce a plus de chance d'être réel ; iv) s'il s'agit d'une partie d'un autre petit ARN (snoRNA, tRNA, snRNA ou rRNA), dans ce cas la faux-positivité est probable puisqu'il s'agirait en réalité d'un autre petit ARN tronqué. Ce code permet d'évaluer rapidement selon ces quatre critères si le microARN semble être un "vrai" microARN ou s'il pourrait s'agir d'un faux-positif.

Parmi les 4 301, 2 750 et 6 131 microARNs bovins, caprins et ovins, respectivement, que compte la base de données RumimiR, les taux de microARNs avec des codes de faux-positivité comportant entre 0 et 2 unités "1", donc remplissant *au maximum* deux des quatre critères de faux-positivité définis ci-dessus s'élèvent à plus de 89%.

Ce travail a fait l'objet d'un article publié dans le journal Database (Oxford). Les données mentionnées dans l'article représentent un nombre moins élevé de microARNs décrits ci-dessus car l'article se base sur la version de Juin 2019, une nouvelle version mise à jour est maintenant en ligne (10 715 microARNs issus de 78 publications dans la version de Juin 2019 décrite dans l'article).

La création de cette base de données RumimiR a permis de lister l'ensemble des microARNs publiés dans les trois espèces de ruminants laitiers qui nous intéressent : bovins, caprins et ovins.

Une perspective pour cette base de données est l'ajout d'autres espèces, pour lesquelles les informations consignées dans RumimiR seraient aussi utiles. L'ajout de variants génétiques décrits de microARNs (aussi nommés "miRSNPs") est également une perspective.

Ce travail a fait l'objet d'une publication, citée cinq fois pour le moment, ajoutée ci-dessous :
"RumimiR: a detailed microRNA database focused on ruminant species" BOURDON C., BARDOU P., AUJEAN E., LE GUILLOU S., TOSSER-KLOPP G., LE PROVOST F. ; *Oxford (Database)* doi: 10.1093/database/baz099 (2019) [290].

1 **RumimiR: a detailed microRNA database focused on ruminant** 2 **species**

3 Bourdon Céline¹, Bardou Philippe^{2,3}, Aujean Etienne¹, Le Guillou Sandrine¹, Tosser-Klopp Gwenola²¶,
4 Le Provost Fabienne^{1,*}¶

5 ¹ GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

6 ² GenPhySE, Université de Toulouse, INRA, ENVT, Castanet-Tolosan, France

7 ³ Sigenae, INRA, Castanet-Tolosan, France

8

9 * To whom correspondence should be addressed. Email: fabienne.le-provost@inra.fr

10 ¶ These authors contributed equally to this work.

11

12 **ABSTRACT**

13 The ever-increasing use of Next Generation Sequencing technologies to explore the genome has
14 generated large quantities of data in recent years. Numerous publications have described several
15 thousand sequences of microRNAs, all species included. A new database (RumimiR) has been created
16 from the literature to provide a detailed description of microRNAs for three ruminant species cattle, goats
17 and sheep. To date, 2,887, 2,733 and 5,095 unique microRNAs from bovine, caprine and ovine species,
18 respectively, are included. In addition to the most recent reference genomic position and sequence of
19 each microRNA, this database contains details about the animals, tissue origins and experimental
20 conditions mentioned in the publications. Identity to human or mouse microRNA is also indicated. The
21 RumimiR database allows data filtering by selecting microRNAs on the basis of defined criteria such as
22 animal status or tissue origin. For ruminant studies, RumimiR supplements the widely used miRBase
23 database, by using complementary criteria to allow browsing and filtering, and integrates all newly-
24 described published sequences. The principal goal of this database is to provide easy access to all the
25 ruminant microRNAs described in the literature.

26

27 **INTRODUCTION**

28 MicroRNAs are small, highly conserved, non-coding RNAs approximately 22 nucleotides in length (1)
29 that participate in the post-transcriptional regulation of genes through their impact on targeted
30 messenger RNAs (mRNAs). This phenomenon can lead to translation repression or degradation of the
31 targeted mRNAs and depends on the base-pair binding of the microRNAs to their target *via* a recognition
32 site, the "seed" sequence (2, 3).

33 Many studies currently use Next Generation Sequencing (NGS) technology to explore the
34 transcriptome, notably in a context of microRNA discovery. A significant amount of data is thus being
35 generated, which can attain more than 1,000 microRNA sequences in a single publication (4, 5).

36 Regularly updated tools for data exploration are therefore required and essential. Animal, plant and virus
37 microRNAs are already listed in several databases, such as miRNEST 2.0 (covering more than 400
38 different species) (6) or miROrtho which contains 46 animal genomes (7). Some databases are
39 restricted to human microRNAs in a disease context, such as EpimiRBase (microRNAs associated with
40 epilepsy) (8), miRCancer (microRNAs and cancer) (9) or MSDD, the MicroRNA SNP Disease Database
41 (genetic variants affecting microRNAs in a disease context) (10). The recent miRCarta database (11),
42 which lists microRNAs for 148 species, is solely based on the prediction of novel microRNAs. A
43 database portal, miRToolsGallery, which contains more than 1,000 tools for studying, identifying or
44 predicting the targets of microRNAs, has recently been set up (12). The database most widely used at
45 present is miRBase, which was created in 2006 by Griffiths-Jones and collaborators at the University of
46 Manchester (13), and lists microRNAs for 271 species (14).

47 Studies in cattle, goats and sheep have often focused on production traits such as dairy and
48 meat products (15–17), health (mastitis resistance (18, 19) or reproduction (fertility and fecundity (20–
49 22)). Due to genome conservation between the three species, the miRNomes (all microRNAs expressed
50 in a tissue or cell type) are relatively similar. However there are some differences, as well as those
51 between breeds (23). This explains the importance of generating a database that includes all three
52 species, together with specific information on the breed and physiological status of the animals, as this
53 is frequently lacking in the most commonly-employed databases.

54 RumimiR provides a single access portal to an integrated database containing all the information
55 currently published on ruminant microRNAs. RumimiR is freely available online at the following URL:
56 <http://rumimir.sigenae.org/>.

57 This database offers an exhaustive list of bovine, caprine and ovine microRNAs, collected from
58 the literature. Pertinent information, notably in the context of dairy production, has also been added to
59 the description of each mature microRNA and a filtering option for all these data is included. The entire
60 database (and data filtered by applying one or more filters) can be downloaded in different formats.
61 Moreover, past database versions are traceable and downloadable.

62

63 **MATERIALS AND METHODS**

64 **Data collection and processing**

65 RumimiR contains data collected from all publications describing ruminant microRNAs, which
66 corresponds to the 78 publications cited in PubMed (<https://www.ncbi.nlm.nih.gov/pubmed>). The
67 references for all these publications are listed in Supplementary Data (**Supplementary Figure 1**). Titles
68 and authors, with a hyperlink to the publication, are provided. RumimiR includes all known microRNA
69 sequences, as well as those described as "novel" in the publications, which are not always found in the
70 miRBase database. All microRNA sequences were obtained from publications (usually from the text,
71 figures or supplementary data). The microRNA sequences included in RumimiR database are

72 processed and filtered by the authors of each publication. The RumimiR data were standardized by
73 aligning the sequences to a unique reference genome for each species. For this purpose, Blast was
74 implemented using the NCBI tool (24), based on the latest available versions of the reference genome
75 (UMD3.1.1 for bovine, ARS1 for caprine and Oar v4.0 for ovine). The parameters used for the blast
76 analyses were the NCBI Blastn tool ("Somewhat similar sequences") and general default parameters
77 (short queries, expected threshold: 1,000, word size: 7) (24). In the event of multiple alignments, with a
78 query cover and 100% identity on the genome, the position of the microRNA sequence could be
79 identified from a Blast of the precursor sequence. Multiple positions were obtained for 344 sequences,
80 even after applying the precursor sequence, when this was available in the publication. These
81 microRNAs have therefore been listed in the RumimiR database without their position, but the number
82 of positions on the genome is shown in the "multi-mapping" column (the chromosome, start and end
83 columns are therefore blank). Characterization of a new microRNA was refined by comparing its
84 sequence with small nucleolar RNA (snoRNA), transfer RNA (tRNA), ribosomal RNA (rRNA) and small
85 nuclear RNA (snRNA) sequences. Indeed, part of the snoRNA structure, the stem loop, is almost
86 identical to that of microRNA (25) and tRNAs with three hairpin loops might also be confused with
87 microRNA. To prevent this, the sequences were therefore compared with all species snoRNA and
88 bovine tRNA present in snoRNABase v3 (26) and GtRNAdb 2.0 (27), respectively, and with the bovine,
89 caprine and ovine rRNA and snRNA sequences present in BioMart, an Ensembl tool (28). When a
90 microRNA displayed 100% identity with a snoRNA, tRNA, rRNA or snRNA over its entire length, it was
91 retained in the RumimiR database but this information is indicated in the "small RNA" column. A home-
92 made Python script was applied which compared the microRNA sequences with those extracted from
93 the above-mentioned database (**Supplementary Figure 2**). Finally, each microRNA was assigned a
94 RumimiR identification number: "Rum-species ID-XXXXX". The species IDs for cattle, goats and sheep
95 are BTA, CHI or OAR, respectively, and XXXXX is an incremental five-digit number. To add information
96 about a microRNA listed in the RumimiR database, a search was done to determine its identity with
97 human or mouse microRNA. If it had a 100% sequence match with a known microRNA over the full
98 length of the shorter sequence, using the miRBase database (release 22), it was considered identical
99 (14). These identities were verified by applying a Python script to compare the sequences with other
100 microRNA sequences and those extracted from miRBase (**Supplementary Figure 2**), and to indicate
101 any "strict" homology with human or mouse microRNA. This step was included because of the large
102 number of publications, and hence details, available for human and mouse microRNAs. All these
103 processes for adding microRNAs to the RumimiR database (**Figure 1**, blue box) were used to create a
104 four-digit false positive code representing four positions, each with a value of "0" or "1" (0000 to 1111).
105 A "1" in first position means that the microRNA has multiple genomic locations, a "1" in second position
106 means that the microRNA is not already listed in the RumimiR database (a microRNA occurring in
107 different publications is more likely to be a true microRNA). A "1" in third position means that the

108 microRNA has homology with small RNAs and a “1” in fourth position means that the microRNA has no
109 identity with either human or mouse microRNA present in miRBase. One or more values of this 4-digit
110 code can be filtered out. A microRNA with a code of 1111 is most likely to be a false positive. This
111 indication is to enable RumimiR users to select those microRNAs that they consider as having an
112 acceptable false positive score.

113 All microRNAs, and most of the features described in the literature, have therefore been documented in
114 the RumimiR database. All the microRNA sequences are noted, as is the presence (or not) of isomiRs
115 (microRNA sequences at the same genomic position, with almost identical sequences). These almost
116 identical sequences are highlighted in a specific column ("isomiRs") in RumimiR and also if the
117 microRNA sequence belongs to a known microRNA family. The database also contains details of the
118 number of animals studied, their breeds, ages and lactation stage, as well as the tissue of origin, if this
119 was mentioned in the publication.

120 The RumimiR website was built using HTML5 technology (<https://dev.w3.org/html5>), the bootstrap front-
121 end framework (<https://getbootstrap.com> - v4.0.0) with additional jQuery user interface elements
122 (<http://jquery.com> - v3.2.1), DataTables jQuery plug-in for the data table (<https://datatables.net> -
123 v1.10.16) and the plots were implemented by using the Highcharts JavaScript library
124 (<http://www.highcharts.com> - v6.2.0). RumimiR has been successfully tested on Chrome (version 49
125 and later) and Firefox (version 57 and later). The data, in JSON format, were provided to the DataTables
126 jQuery plug-in by setting the ajax option to the address of the JSON data source. All statistics, charts
127 and drop-down lists have been built into the data-based fly to make any updates as easy as possible.
128 The Blast tool was developed using Perl-CGI.

129 The “Traceability and Archives” box, on the website shows the history of the various RumimiR database
130 versions and allows the user to download entire datasets from current and past versions.

131

132 **RESULTS AND DISCUSSION**

133 **Content of the RumimiR database**

134 The RumimiR database currently contains 10,715 different microRNAs for three ruminant species: 2,887
135 for cattle, 2,733 for goats and 5,095 for sheep. The average numbers of microRNAs described per
136 publication for bovine, caprine and ovine species are 66, 136 and 340, respectively. The microRNA data
137 were collected from 44 publications for bovines, 19 for ovines and 15 for caprines (**Figure 2**). The
138 difference in numbers of microRNAs reported for cattle and goats, as compared to sheep, is due to the
139 wider use of NGS technologies in ovine species. More precisely, 16,041 sequences, corresponding to
140 10,715 different mature microRNAs, are present in the RumimiR database, the difference being due to
141 the presence of isomiRs. 344 microRNA sequences are indicated as having multiple locations (multi-
142 mapping). The microRNAs listed in RumimiR have identities with 889 human or mouse microRNAs:
143 11.78% of the described microRNAs show sequence identities with both human and mouse microRNAs
144 and 4.66% present sequence identity with only one human or one mouse microRNA. Comparisons with

145 snoRNA, tRNA, rRNA and snRNA sequences revealed that 48 microRNAs displayed sequence
146 identities with part of a tRNA, 47 with part of a snoRNA, 34 with part of a rRNA and 9 with part of a
147 snRNA.

148 The importance of each microRNA is evaluated from the false positive score. Around 50% of the
149 microRNAs (56.1% in bovine, 51.2% in caprine and 49.8% in ovine species) have a score with a low
150 risk of being false positive (code with 0 or 1 “1”). Moreover, almost every microRNA has a code with 0
151 to 2 “1” (98.3% in bovine, 92.8% in caprine and 99.4% in ovine species) (**Figure 3 and Supplementary**
152 **Figure 3**). Each scores are representing with the corresponding encoding: for example, the scores of 1
153 are due to the absence of identity with a mouse and human microRNA, in the three species. In fact,
154 when other elements appeared (*i.e.* small RNA or multi-mapping) there are often associated with
155 another false positivity element, as seen in the Figure 3.

156 Most of the microRNAs listed in RumimiR have the expected length (20-24 nucleotides) (1), with 99.27%
157 having 17-25 nucleotides. The microRNA length distributions in bovine and caprine species are similar
158 (**Figure 4A**) with 91% of bovines and 84% in caprines having 20-24 nt. The length distribution in ovine
159 species is unusual (55% having 20-24 nt). The numbers of microRNAs with lengths of 17-18 nt and
160 those with 21-22 nt are similar. The proportion differs if only those microRNAs with weak false positive
161 code are considered, but the percentage of microRNAs with lengths of 17 - 18 nt is still high (around
162 25%) (**Figure 4B**). Most of the ovine sequences with lengths of 17 and 18 nt were obtained from a single
163 paper (4). The authors explained, in their discussion, that they had discovered a large number of new
164 microRNAs because they had not applied the commonly-used restricting pre-screening. All sequences
165 detected at a low count level or with a low sample frequency were considered. Most of the ovine
166 sequences with lengths of 17 and 18 nt correspond to new microRNAs.

167 The majority of microRNAs in the database were taken from studies based on differences
168 between breeds, developmental stages (mainly in a context of meat production) or immune response
169 (comparison of healthy animals with those suffering from mastitis, a widely-studied pathology in
170 ruminants) (**Figure 5**). The different issues addressed in the publications were implemented as filtering
171 options in the RumimiR database. Thus 40 different breeds (17 bovine, 14 caprine and 9 ovine breeds,
172 both meat and dairy), are represented along with about 30 tissues and body fluids such as milk (29,
173 30), adipose tissue (4), mammary gland (31–33) and ovaries (22, 34, 35). Around half of the microRNAs
174 described in the RumimiR database are associated with a single tissue (these “specific” microRNAs are
175 highlighted by the symbol “⊙”). While the probability of a microRNA being a false positive decreases if
176 it is described in several studies, some of the microRNAs described only once can be specific to a breed,
177 tissue, physiological stage or particular condition.

178 Age was mentioned in 38% of the publications and more than 30 different ages have been listed.
179 Some studies described the differential expression of microRNAs throughout development, and at
180 different ages, which is why several ages might be considered in a single publication. The same was
181 true of lactation stages. All these figures will increase as the database is updated. The numbers of

182 microRNAs common to all three ruminant species, or sequences common to these species and to
183 human and mouse sequences, or specific to each species, are presented in **Figure 6**. Between 80.24%
184 and 88.82 % are species-specific microRNAs.

185 RumimiR is limited to three ruminant species, unlike miRBase which is the most widely used
186 and complete database in terms of the number of species covered (plants, animals and viruses).
187 However, this restriction enabled us to generate a more detailed database containing all the microRNAs
188 described in different publications as well as numerous features. Six times more microRNAs are listed
189 in the RumimiR database for bovine, caprine and ovine species than in miRBase: 10,715 in RumimiR
190 *versus* 1,614 in miRBase. Although the latest miRBase release (v22) is quite recent (October 2018), it
191 included only 5% of the data available on deep sequencing of small RNA (14). The risk of assembling
192 all microRNAs mentioned in the literature is still to include false positive microRNAs which is why this
193 risk has been evaluated for each sequence. Thus RumimiR offers a complete and more precise
194 microRNA database for cattle, goats and sheep and should therefore be of value to scientists working
195 on livestock species.

196 The RumimiR database is extensive because it lists all the features mentioned in the
197 publications (unlike miRBase and other databases), the data are homogenized (with blast on the same
198 reference genomes for example), and because information is included about the identity, small RNAs,
199 etc. In addition, the RumimiR database combines the results of a multitude of studies in a single location.

200

201 **User interface**

202 The web interface is user-friendly, allows visualization of all the microRNAs described in bovine, caprine
203 and ovine species, and provides at minimum the genomic position (chromosome, start and end) of each
204 microRNA. Some, or all the columns can be selected to obtain the sequence, name, tissue or other
205 features of the study and of the microRNAs, by clicking on the appropriate name in the "show/hide
206 columns" box (**Figure 7**).

207 The features available for the various microRNAs, such as the sequence, related publications,
208 name, isomiRs, family, tissue origin, breed, condition studied, and others are listed in the 29 columns
209 shown in **Table 1**.

210 Users can also visualize those microRNAs of solely personal interest by selecting the
211 corresponding column (sequence and/or tissue and/or breed, etc.), and then using the "search" box to
212 define different options such as "by position" (chromosome, start, end), "by source" (species, breed,
213 tissue) or "by feature" (conditions, method, software) (**Figure 7**). The above-mentioned criteria can also
214 be applied to filter all the mature microRNAs listed. A sequence or key word can also be entered in the
215 "Search" box to extract a list of the relevant data in the results table. The name of each microRNA in the
216 "name" column is the one given in the original publication. If specific requests are selected, the
217 appropriate filtered data appear and can be downloaded in several formats (Excel, CSV, Fasta or GFF).

218 Users can thus obtain all the data that they require, appropriately filtered, and rapidly. They can also
219 download the entire database without applying any filters.

220 For clarity, some statistics and graphs which summarize the data presented in the database
221 (**Supplementary Figure 4**), are provided in the online version. These include, amongst others, the date
222 of the latest update, the number of sequences listed and the number of publications involved. The
223 distribution of microRNAs by species and by breed is presented graphically, as is the distribution by
224 tissue origin.

225 RumimiR also includes an alignment tool so that sequences can be submitted for analysis and
226 the corresponding microRNAs sorted according to the resulting hits.

227

228 **Future extensions**

229 The NGS approach is generating extensive data and the number of microRNAs will continue to rise in
230 the near future. This new data will be integrated and the database regularly updated (at least twice a
231 year) so that it remains exhaustive. Identity with human or mouse microRNAs will be checked each time
232 that a new release of miRBase becomes available. The three ruminant species probably contain a
233 comparable number of microRNAs as the difference between the numbers of bovine (2,887), caprine
234 (2,733) and ovine (5,095) microRNAs could be due to the more intense efforts assigned to ovine
235 sequencing. However the numbers of human or mice microRNAs in the latest release of miRBase are
236 2,654 and 1,978 respectively. MicroRNA detection in ruminant species, as compared to humans and
237 mice, may have reached saturation. In addition, the numbers will probably evolve as the genome
238 assemblies are updated and the same quality of contiguity and annotation is attained as for human and
239 mouse species.

240 One potential extension for RumimiR might be to add genetic variants of ruminant microRNAs
241 ("miRSNPs" for microRNA and Single Nucleotide Polymorphisms). First those linked with dairy QTL,
242 and then those associated with health or meat QTL. Publications on these genetic microRNA variations
243 are indeed increasingly numerous. For example, in humans, the MSDD database has been created
244 exclusively for miRSNPs linked to human diseases (10). In the same way, the RumimiR database could
245 be completed by including ruminant miRSNPs. Other features and filters will be added depending on
246 the features described in the literature and the needs of the scientific community. Evolution of the
247 reference genomes will also be considered, to take into account the genomic positions of microRNAs in
248 the latest versions of the reference genomes. The RumimiR database could also be extended to include
249 other livestock species.

250

251 **Conclusion**

252 The RumimiR database contains an exhaustive list of all microRNAs described in publications pertaining
253 to three livestock species: cattle, goats and sheep. This database supplements miRbase, one of the
254 most widely used microRNA databases, by including various features mentioned in the literature which

255 are important in the context of animal production and dairy traits, notably the breeds of animals studied
256 or the tissues in which the microRNAs were described. RumimiR can be used to retrieve specific
257 microRNAs and thus provide additional information about the livestock species and a clearer
258 understanding of the context in which these microRNAs were discovered. The database will be regularly
259 updated and will continue to be exhaustive. RumimiR, by standardizing and centralizing information from
260 a large number of publications, constitutes a unique tool that presents and describes all known
261 microRNAs in ruminants.

262

263 **FUNDING**

264 This work was supported by APIS-GENE through the miRQTLait project. CB's grant was financially
265 supported by INRA and APIS-GENE.

266

267 **ACKNOWLEDGEMENTS**

268 The authors are grateful to C. Gaspin, C. Leroux and S. Marthey for their critical reading of the
269 manuscript and to M. Boussaha for his helpful discussions.

270

271

272 **REFERENCES**

- 273 1. Ambros, V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.
- 274 2. Lewis, B.P., Shih, I., Jones-Rhoades, M.W., Bartel, D.P. and Burge, C.B. (2003) Prediction of
275 Mammalian MicroRNA Targets. *Cell*, **115**, 787–798.
- 276 3. Mohr, A. and Mott, J. (2015) Overview of MicroRNA Biology. *Semin. Liver Dis.*, **35**, 003–
277 011.
- 278 4. Miao, X., Luo, Q., Qin, X. and Guo, Y. (2016) Genome-wide analysis of microRNAs
279 identifies the lipid metabolism pathway to be a defining factor in adipose tissue from
280 different sheep. *Sci. Rep.*, **5**.
- 281 5. Zhao, Q., Kang, Y., Wang, H.-Y., Guan, W.-J., Li, X.-C., Jiang, L., He, X.-H., Pu, Y.-B.,
282 Han, J.-L., Ma, Y.-H., *et al.* (2016) Expression profiling and functional characterization
283 of miR-192 throughout sheep skeletal muscle development. *Sci. Rep.*, **6**.
- 284 6. Szczeńniak, M.W. and Makałowska, I. (2014) miRNEST 2.0: a database of plant and animal
285 microRNAs. *Nucleic Acids Res.*, **42**, D74–D77.
- 286 7. Gerlach, D., Kriventseva, E.V., Rahman, N., Vejnar, C.E. and Zdobnov, E.M. (2009)
287 miROrtho: computational survey of microRNA genes. *Nucleic Acids Res.*, **37**, D111–
288 D117.

- 289 8. Mooney,C., Becker,B.A., Raof,R. and Henshall,D.C. (2016) EpimiRBase: a
290 comprehensive database of microRNA-epilepsy associations. *Bioinformatics*, **32**,
291 1436–1438.
- 292 9. Xie,B., Ding,Q., Han,H. and Wu,D. (2013) miRCancer: a microRNA-cancer association
293 database constructed by text mining on literature. *Bioinformatics*, **29**, 638–644.
- 294 10. Yue,M., Zhou,D., Zhi,H., Wang,P., Zhang,Y., Gao,Y., Guo,M., Li,X., Wang,Y.,
295 Zhang,Y., *et al.* (2018) MSDD: a manually curated database of experimentally
296 supported associations among miRNAs, SNPs and human diseases. *Nucleic Acids Res.*,
297 **46**, D181–D185.
- 298 11. Backes,C., Fehlmann,T., Kern,F., Kehl,T., Lenhof,H.-P., Meese,E. and Keller,A. (2018)
299 miRCarta: a central repository for collecting miRNA candidates. *Nucleic Acids Res.*,
300 **46**, D160–D167.
- 301 12. Chen,L., Heikkinen,L., Wang,C., Yang,Y., Knott,K.E. and Wong,G. (2018)
302 miRToolsGallery: a tag-based and rankable microRNA bioinformatics resources
303 database portal. *Database*, **2018**.
- 304 13. Griffiths-Jones,S. (2006) miRBase: microRNA sequences, targets and gene nomenclature.
305 *Nucleic Acids Res.*, **34**, D140–D144.
- 306 14. Kozomara,A., Birgaoanu,M. and Griffiths-Jones,S. (2018) miRBase: from microRNA
307 sequences to function. *Nucleic Acids Res.*, 10.1093/nar/gky1141.
- 308 15. Shen,B., Zhang,L., Lian,C., Lu,C., Zhang,Y., Pan,Q., Yang,R. and Zhao,Z. (2016) Deep
309 Sequencing and Screening of Differentially Expressed MicroRNAs Related to Milk Fat
310 Metabolism in Bovine Primary Mammary Epithelial Cells. *Int. J. Mol. Sci.*, **17**, 200.
- 311 16. Hou,J., An,X., Song,Y., Cao,B., Yang,H., Zhang,Z., Shen,W. and Li,Y. (2017) Detection
312 and comparison of microRNAs in the caprine mammary gland tissues of colostrum and
313 common milk stages. *BMC Genet.*, **18**.
- 314 17. Miao,X., Luo,Q. and Qin,X. (2015) Genome-wide analysis reveals the differential
315 regulations of mRNAs and miRNAs in Dorset and Small Tail Han sheep muscles.
316 *Gene*, **562**, 188–196.
- 317 18. Ju,Z., Jiang,Q., Liu,G., Wang,X., Luo,G., Zhang,Y., Zhang,J., Zhong,J. and Huang,J.
318 (2018) Solexa sequencing and custom microRNA chip reveal repertoire of microRNAs

- 319 in mammary gland of bovine suffering from natural infectious mastitis. *Anim. Genet.*,
320 **49**, 3–18.
- 321 19. Du,J., Gao,S., Tian,Z., Xing,S., Huang,D., Zhang,G., Zheng,Y., Liu,G., Luo,J., Chang,H.,
322 *et al.* (2017) MicroRNA expression profiling of primary sheep testicular cells in
323 response to bluetongue virus infection. *Infect. Genet. Evol.*, **49**, 256–267.
- 324 20. Huang,J., Ju,Z., Li,Q., Hou,Q., Wang,C., Li,J., Li,R., Wang,L., Sun,T., Hang,S., *et al.*
325 (2011) Solexa sequencing of novel and differentially expressed microRNAs in
326 testicular and ovarian tissues in Holstein cattle. *Int. J. Biol. Sci.*, **7**, 1016–1026.
- 327 21. Zi,X., Lu,J. and Ma,L. (2017) Identification and comparative analysis of the ovarian
328 microRNAs of prolific and non-prolific goats during the follicular phase using high-
329 throughput sequencing. *Sci. Rep.*, **7**.
- 330 22. Miao,X. and Qin,Q.L.X. (2015) Genome-wide transcriptome analysis of mRNAs and
331 microRNAs in Dorset and Small Tail Han sheep to explore the regulation of fecundity.
332 *Mol. Cell. Endocrinol.*, **402**, 32–42.
- 333 23. Wicik,Z., Gajewska,M., Majewska,A., Walkiewicz,D., Osińska,E. and Motyl,T. (2016)
334 Characterization of microRNA profile in mammary tissue of dairy and beef breed
335 heifers. *J. Anim. Breed. Genet.*, **133**, 31–42.
- 336 24. Johnson,M., Zaretskaya,I., Raytselis,Y., Merezuk,Y., McGinnis,S. and Madden,T.L.
337 (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res.*, **36**, W5–W9.
- 338 25. Scott,M.S. and Ono,M. (2011) From snoRNA to miRNA: Dual function regulatory non-
339 coding RNAs. *Biochimie*, **93**, 1987–1992.
- 340 26. Lestrade,L. (2006) snoRNA-LBME-db, a comprehensive database of human H/ACA and
341 C/D box snoRNAs. *Nucleic Acids Res.*, **34**, D158–D162.
- 342 27. Chan,P.P. and Lowe,T.M. (2016) GtRNAdb 2.0: an expanded database of transfer RNA
343 genes identified in complete and draft genomes. *Nucleic Acids Res.*, **44**, D184–D189.
- 344 28. Smedley,D., Haider,S., Durinck,S., Pandini,L., Provero,P., Allen,J., Arnaiz,O.,
345 Awedh,M.H., Baldock,R., Barbiera,G., *et al.* (2015) The BioMart community portal: an
346 innovative alternative to large, centralized data repositories. *Nucleic Acids Res.*, **43**,
347 W589–W598.

- 348 29. Sun,J., Aswath,K., Schroeder,S.G., Lippolis,J.D., Reinhardt,T.A. and Sonstegard,T.S.
349 (2015) MicroRNA expression profiles of bovine milk exosomes in response to
350 *Staphylococcus aureus* infection. *BMC Genomics*, **16**.
- 351 30. Do,D.N., Li,R., Dudemaine,P.-L. and Ibeagha-Awemu,E.M. (2017) MicroRNA roles in
352 signalling during lactation: an insight from differential expression, time course and
353 pathway analyses of deep sequence data. *Sci. Rep.*, **7**, 44605.
- 354 31. Mobuchon,L., Marthey,S., Boussaha,M., Le Guillou,S., Leroux,C. and Le Provost,F.
355 (2015) Annotation of the goat genome using next generation sequencing of microRNA
356 expressed by the lactating mammary gland: comparison of three approaches. *BMC*
357 *Genomics*, **16**.
- 358 32. Li,R., Beaudoin,F., Ammah,A.A., Bissonnette,N., Benchaar,C., Zhao,X., Lei,C. and
359 Ibeagha-Awemu,E.M. (2015) Deep sequencing shows microRNA involvement in
360 bovine mammary gland adaptation to diets supplemented with linseed oil or safflower
361 oil. *BMC Genomics*, **16**.
- 362 33. Le Guillou,S., Marthey,S., Laloë,D., Laubier,J., Mobuchon,L., Leroux,C. and Le
363 Provost,F. (2014) Characterisation and Comparison of Lactating Mouse and Bovine
364 Mammary Gland miRNomes. *PLoS ONE*, **9**, e91938.
- 365 34. Zhu,L., Chen,T., Sui,M., Han,C., Fang,F., Ma,Y., Chu,M., Zhang,X., Liu,C. and Ling,Y.
366 (2016) Comparative profiling of differentially expressed microRNAs between the
367 follicular and luteal phases ovaries of goats. *SpringerPlus*, **5**.
- 368 35. Hu,X., Pokharel,K., Peippo,J., Ghanem,N., Zhaboyev,I., Kantanen,J. and Li,M.-H. (2016)
369 Identification and characterization of miRNAs in the ovaries of a highly prolific sheep
370 breed. *Anim. Genet.*, **47**, 234–239.

371

372

373 **TABLE AND FIGURE LEGENDS**

374

375 **Figure 1.** Procedure followed to include microRNAs and their associated features in the RumimiR
376 database.

377 From each publication (black rectangle), the microRNA sequences and their features (orange
378 rectangles) were extracted and included in the RumimiR database. Several analyses were performed
379 to expand the description of the microRNAs collected (blue rectangles): determination of chromosomal
380 position in the latest reference genome of each species (using the NCBI Blastn tool), search for
381 redundancy with microRNAs already present in the RumimiR database, search for identity with other
382 small microRNAs to compare the sequences, search for identity with human or mouse microRNAs
383 present in miRBase. *: using a home-made Python script.

384

385 **Figure 2.** Number of microRNAs per publication.

386 A- Plain bars: average number of microRNAs described per publication for each species. Hatched
387 bars: the number of publications on each species.

388 B- Cumulative number of microRNAs described per year in each species.

389

390 **Figure 3.** Proportion of microRNAs according to the false positive score in bovine.

391 Sequences without a “1-value are more likely to be true microRNAs and those with 4 1-values are less
392 likely to be microRNAs.

393 **Figure 4.** microRNA length distributions across species.

394 The red, blue and orange lines correspond to the percentages of bovine, caprine and ovine microRNAs,
395 respectively, based on their length (in nucleotides (nt)). The grey dotted line represents all the
396 microRNAs listed in RumimiR.

397 A- Percentage of all sequences collected in RumimiR

398 B- Percentage of sequences with a low false positive score (0 and 1).

399

400 **Figure 5**Number of microRNAs according to the conditions studied.

401 The different topics presented here are in line with the issues addressed in the publications. 78
402 publications were considered and each article is mentioned only once. Each microRNA was counted
403 several times, if they were described in several publications corresponding to different topics.

404

405 **Figure 6.** Number of microRNAs common to different species.

406 A- Venn diagram showing bovine, caprine and ovine unique microRNAs listed in the RumimiR
407 database.

408 B- Venn diagram showing bovine, caprine and ovine microRNAs listed in the RumimiR database
409 together with the human and mouse microRNAs listed in miRBase (release 22).

410

411 **Figure 7.** User interface for “search and browse”.

412 Columns can be selected so that only specific features will be visible. The search box allows the
413 detection of microRNAs present in the RumimiR database as a function of the selected choices. The
414 results can be exported in different formats, such as Excel or Fasta.

415

416 **Table 1.** List of columns presenting the microRNAs and features detailed in the RumimiR database.
417 Grey lines indicate a feature that is present in RumimiR but not in miRBase.

418

419 **SUPPLEMENTARY DATA**

420 **Supplementary Figure 1.** List of publications used to collect microRNA data.

421 MicroRNAs from 78 publications were collected and included in this first version of the RumimiR
422 database. Publications 1 to 44 correspond to bovine studies, 45 to 63 to caprine studies and 64 to 78
423 to ovine studies.

424

425 **Supplementary Figure 2.** Python script descriptions.

426

427 **Supplementary Figure 3.**

428

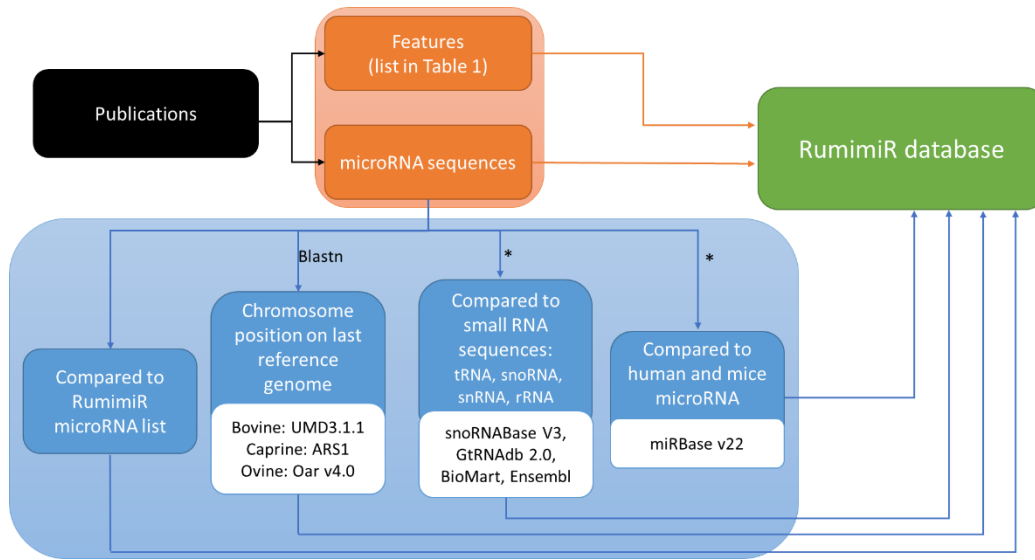
429 **Supplementary Figure 4.** Graphs presented in the online version.

430 Detailed data are available online for each figure. (<http://rumimir.sigenae.org/>).

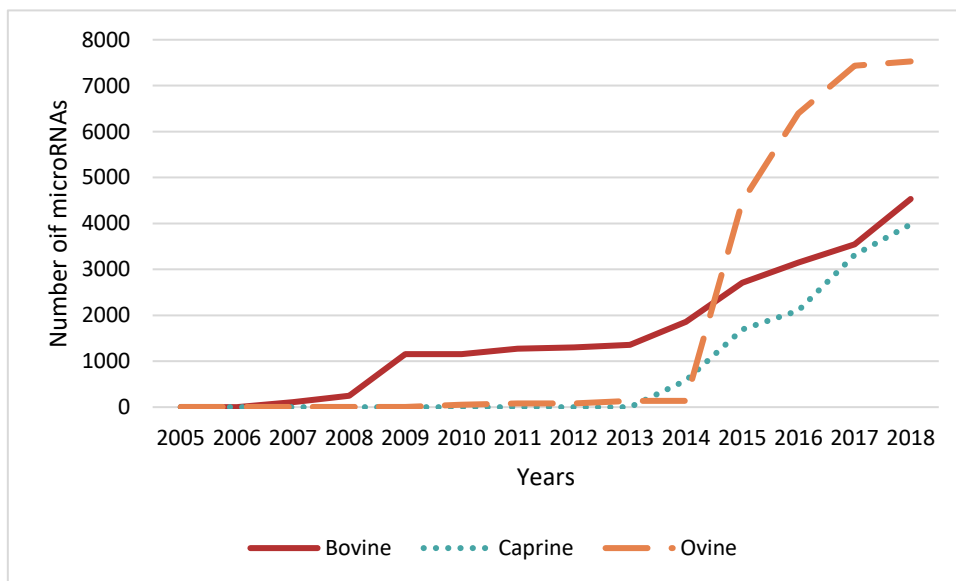
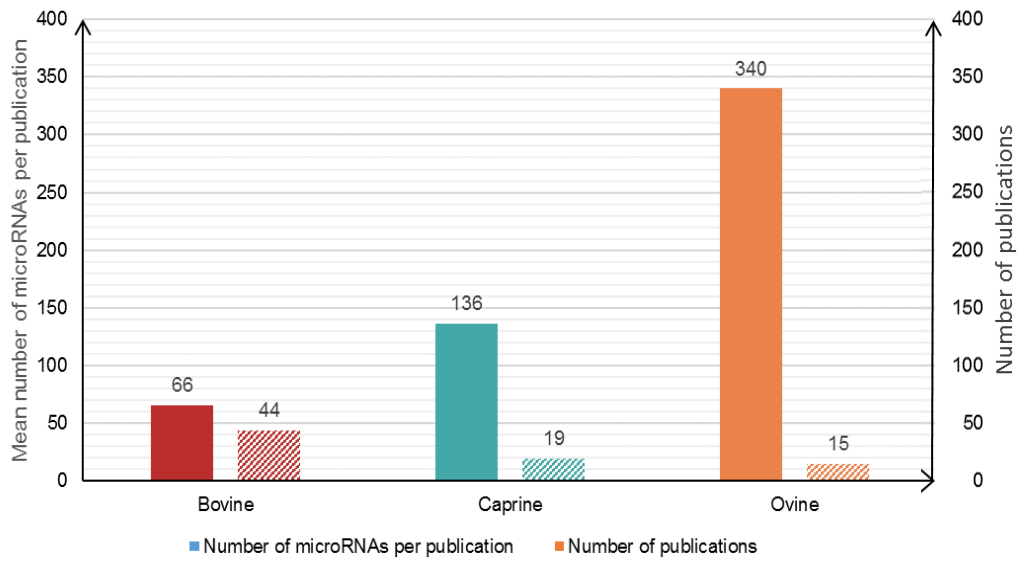
431

432

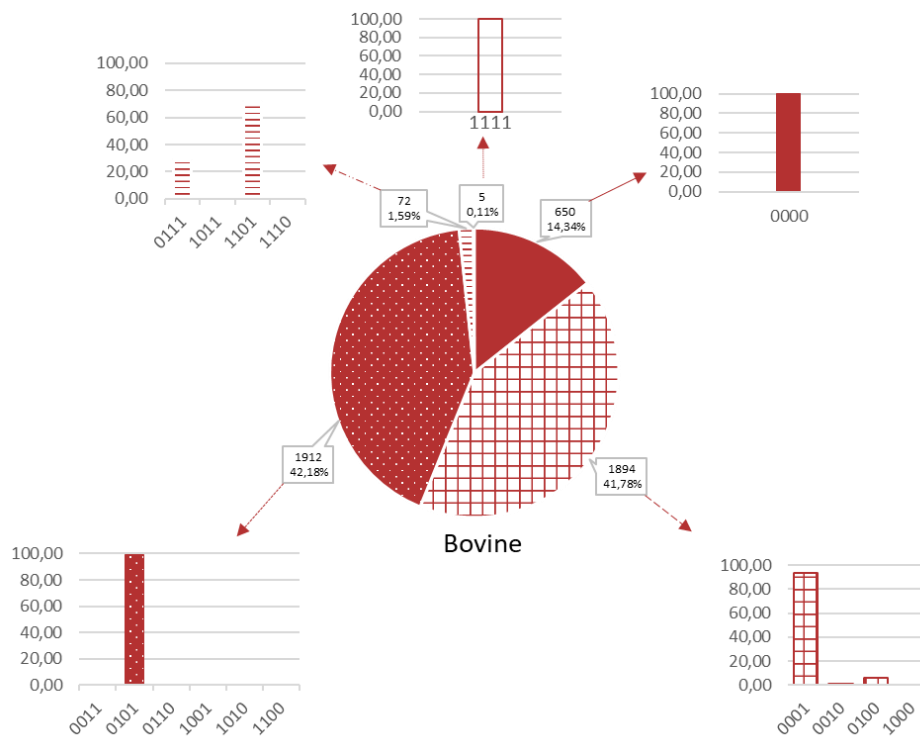
433 Figure 1



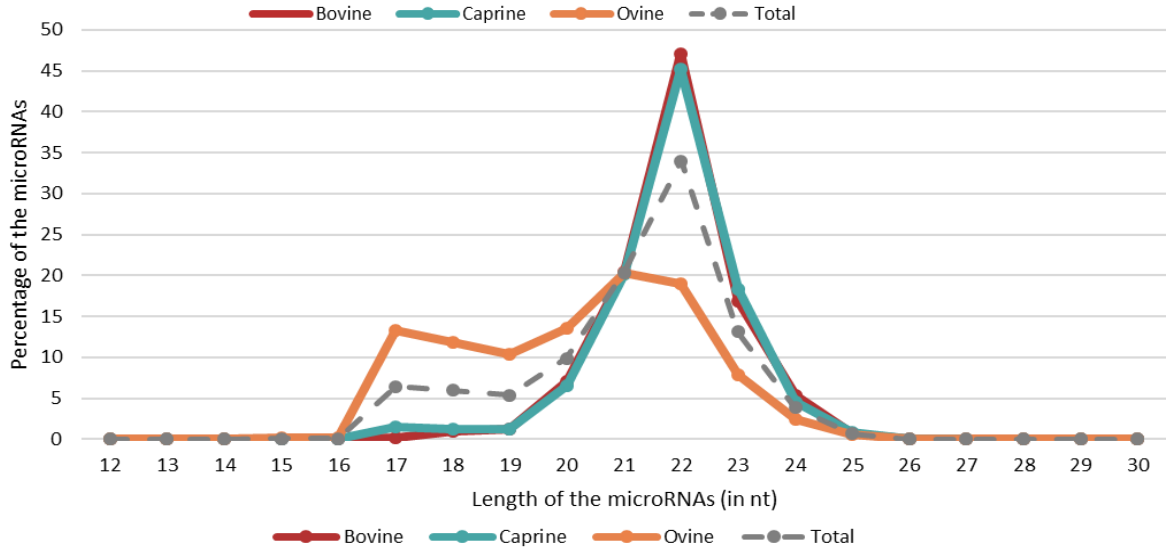
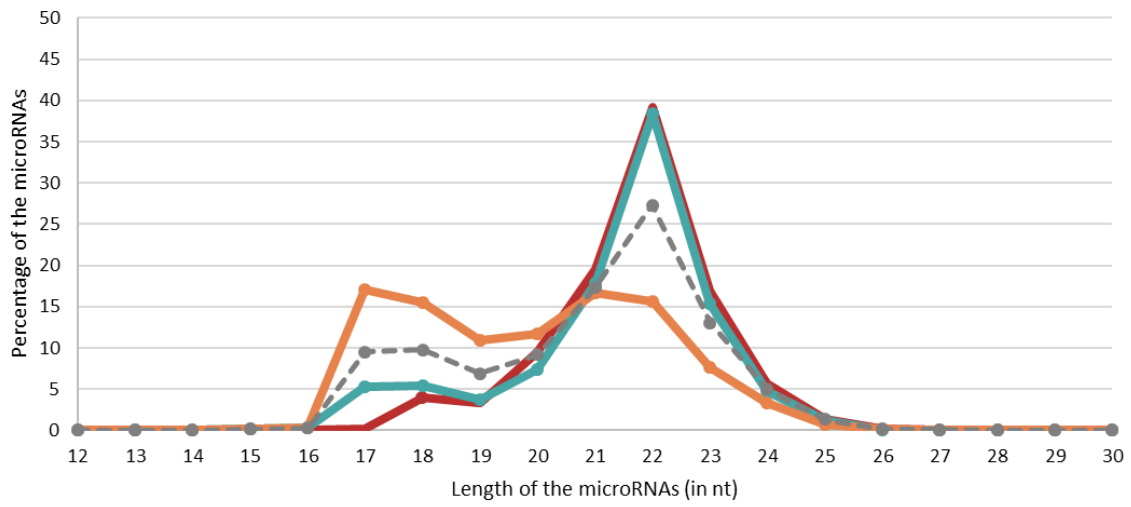
434 Figure 2



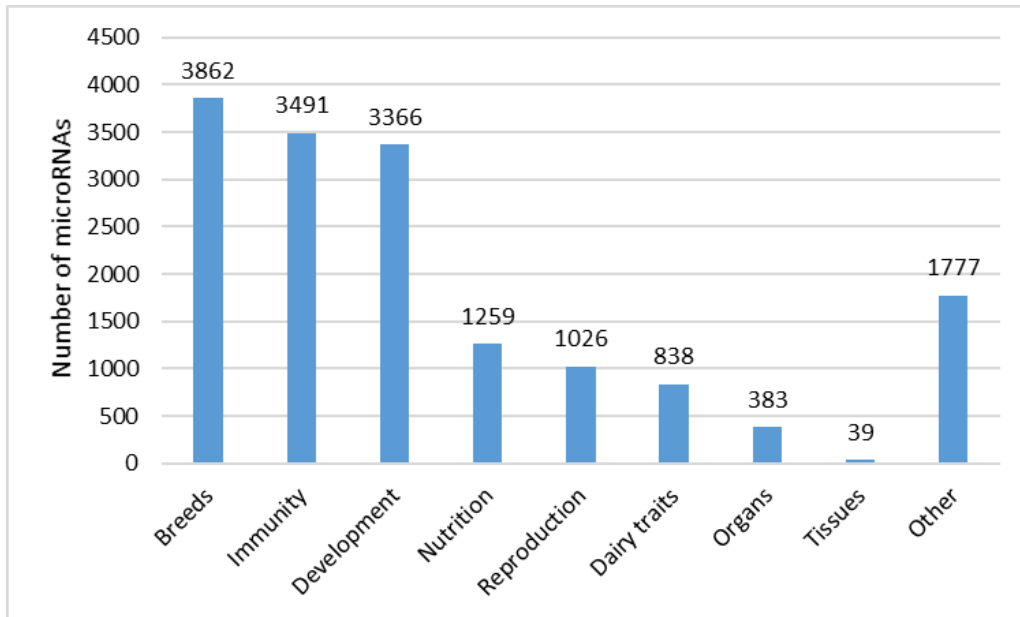
435 Figure 3



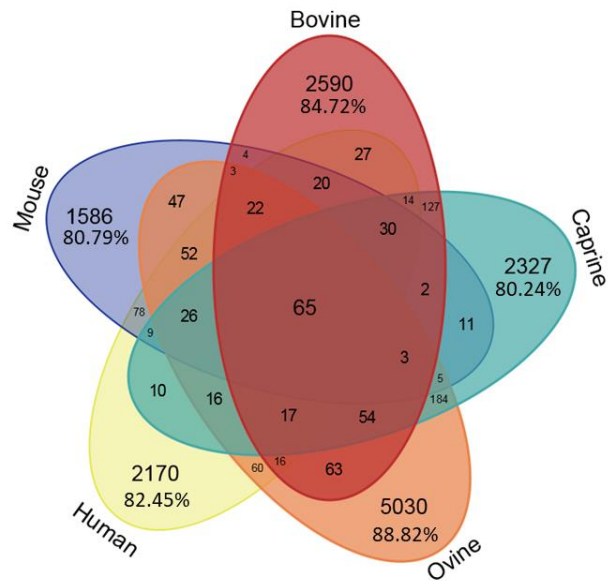
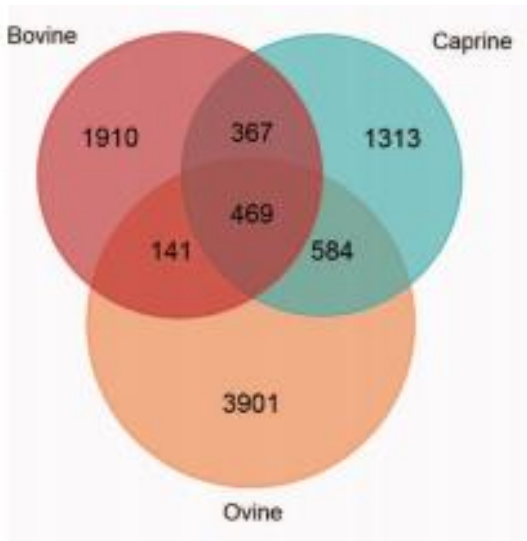
436 Figure 4



437 Figure 5



438 Figure 6



439 Figure 7

SHOW/HIDE COLUMNS

Configure the table below by selecting column name(s)

5 items selected

Number of visible columns: 8/28

All Reset

⚙️ columns visible or "Reset to default"

"Remove all filters" in the search box.

SEARCH

By position | **By source** | By feature | By blast

Species: Nothing selected

Select one or more species.

Breed: 27 items selected

Select one or more breed(s).

Tissue: 46 items selected

Select one or more tissue(s).

Select tissue-specific.

EXPORT

For the filtered row(s) export...

The data:

A fasta file of the miRNA sequence(s).

A GFF file of the miRNA.

Show 10 entries

Chr.	Start	End	Sequence	Name	Tissue	Breed	Species
10	14009887	14009908	ACCUCUCAGCUCAUUGACUGG	chr10_3067	Milk exosome	Holstein	Bovine
10	15225957	15225976	AAAAGGGGGGAGAAAGCCCA	hsa-miR-4484	Left ventricle tissue	Coopworth	Ovine
10	15596370	15596391	AACAGUCUGUUCAGGUUUUCU	chr10_14918	Mammary tissue	Alpine	Caprine
10	15596370	15596391	AACAGUCUGUUCAGGUUUUCU	Novel_022302.1_389730	Rumen	Jiayang Da'er	Caprine

440 Table 1

Chr	Genomic position of the microRNA
Start	
End	
Multi-mapping	Number of localisations in case of multi-mapping (with 100% query cover and 100% identity)
Sequence	Sequence of the mature microRNA
Publication	Publication in which the microRNA is described
Name	Name assigned to the microRNA
IsomiRs	Name of isomiRs of the microRNA (sequence and genomic position almost identical)
RumimiR ID	Name of the microRNA with the RumimiR nomenclature
Family	microRNA family to which the microRNA is affiliated
Tissue	Tissue in which the microRNA was discovered
Breed	Breed used during the study
Study conditions	Conditions prevailing during the study
Lactation stages	Lactation stages of the animals studied
Age	Age of the animals studied
Species	Species of the animals studied (bovine, caprine or ovine)
Number	Number of novel microRNAs described in the publication
Cut-off	Cut-off point used in the publication to detect novel microRNAs
Method	Method used to detect novel microRNAs
Number and breed	Number of animals studied in each breed
Reference genome	UMD3.1.1, ARS1 or Oar v4.0
Bioinformatics tools used	Software used in the study for the microRNA detection
Condition details	Details of study conditions (i.e. number of animals per condition)
5p/3p	If the microRNA is 5p or 3p
Strand (+/-)	Strand in which the microRNA is situated
Star sequence	Sequence of the star microRNA
Matching seed	Known microRNA with the same seed
small RNAs	If the microRNA described is in fact part of a snoRNA, tRNA, snRNA or rRNA
hsa homology	Similarity with a human microRNA
mmu homology	Similarity with a mouse microRNA
False-positive code	Four-digit false positive code

2. Détection de variants génétiques d'intérêt en bovin, caprin et ovin

La première partie de cette thèse consiste en la détection de variants génétiques de microARNs exprimés dans la glande mammaire ou présents dans le lait et situés dans des "QTL laitiers ou mammites". Les trois espèces de ruminants laitiers que sont les bovins, caprins et ovins ont été analysées. Cette détection de variants s'effectue grâce à un script bio-informatique développé en langage Python, permettant d'établir une liste des variants d'intérêt dans notre cas à partir de données initiales. Ces données initiales sont l'ensemble des variants génétiques, des localisations des régions de QTL laitiers/mammites et des microARNs exprimés dans la glande mammaire ou présents dans le lait. Un même script bio-informatique a été créé et modulé pour pouvoir être utilisé dans le cadre de détection de variants génétiques d'intérêt dans les trois espèces. Les filtres sont par la suite appliqués successivement sur ces données de départ.

Un total de 59 124 variants bovins d'intérêt sont alors détectés. En caprin, ce nombre s'élève à 13 427 et 4 761 variants en ovins. Parmi eux, 44 variants bovins sont situés dans la région "seed" du microARN, et 1 des variants détectés est situé dans cette région spécifique en caprin, de même en ovin. Un total de 511 microARNs bovins sont impactés par ces variants génétiques détectés. Ce nombre s'élève à 100 microARNs caprins et 106 microARNs ovins. Un total de 20 microARNs sont impactés par des variants génétiques d'intérêt dans les trois espèces étudiées : à la fois en bovin, caprin et ovin. Par la suite, les régions QTL laitiers et mammites sont ajoutées comme filtre dans ce script. En bovin, 4 679 variants sont alors détectés, parmi lesquels 26 sont situés dans un précurseur de microARN, 10 d'entre eux sont situés dans un microARN mature dont 3 dans la région "seed". En caprin, 127 variants génétiques sont alors détectés, tous situés en-dehors du précurseur de microARN. En ovin, il n'y a pas de variant génétique détecté en ajoutant ce filtre de région QTL.

Ce script bioinformatique, utilisé et adapté pour les trois différentes espèces de ruminants laitiers (bovins, caprins et ovins), a permis de détecter des variants génétiques d'intérêt dans un contexte de production laitière. Ce script pourra donc par la suite être adapté pour la recherche de variants dans d'autres espèces ou de variants d'autres ARNs non codants, sous réserve que les données initiales nécessaires soient disponibles.

Ce travail a fait l'objet d'une publication en cours de soumission à Scientific Reports, ajoutée ci-après.

16 Abstract

17 **Background:**

18 microRNAs are small noncoding RNAs that have important roles in the lactation process and
19 milk biosynthesis. Some polymorphisms have been studied in various livestock species from
20 the perspective of pathology or production traits. To target variants that could be the causal
21 variants of dairy traits, genetic variants of microRNAs expressed in the mammary gland or
22 present in milk and localized in dairy quantitative trait loci (QTLs) were investigated in bovine,
23 caprine, and ovine species.

24 **Results:**

25 In this study, a total of 59,124, 13,427, and 4,761 genetic variants in microRNAs expressed in
26 the mammary gland or present in milk were identified in bovine, caprine, and ovine species,
27 respectively. A total of 4,679 of these detected bovine genetic variants are located in dairy
28 QTLs. In caprine species, 127 genetic variants are localized in dairy QTLs. In ovine species, no
29 genetic variant was identified in dairy QTLs.

30

31 **Conclusions:**

32 This study leads to the detection of microRNA genetic variants of interest in the context of
33 dairy production, taking advantage of whole genome data to identify microRNA genetic
34 variants expressed in the mammary gland and localized in dairy QTLs.

35

36 **Keywords:**

37 microRNA, genetic variants, QTL, dairy traits, miRNomes, bovine, caprine, ovine

38

39 Background

40

41 MicroRNAs are small noncoding RNAs of approximately 22 nucleotides in length that are
42 highly conserved between species ¹. They are involved in posttranscriptional gene regulation
43 through their impact on messenger RNAs (mRNAs). This interaction will lead to the repression
44 of the translation process or to the degradation of the targeted mRNAs, depending on the base-
45 pair binding between the microRNA and the mRNA in the recognition site, the seed sequence
46 ². MicroRNAs have important roles in mammary gland development, lactation and milk
47 biosynthesis ³⁻⁵. Mammary miRNome data, listing the microRNAs expressed in the mammary
48 gland or present in milk, are available in bovine and caprine species ^{6,7}. In the Holstein breed,
49 the mammary gland miRNomes were described in cows with mastitis ⁸ and healthy cows ⁶.
50 The miRNomes of whole milk in Holstein and Normande breeds ⁹, milk fat in Holstein breeds
51 ¹⁰ and milk fat, whey and somatic cells in Holstein breeds ¹¹ are also available. In caprine
52 species, the mammary gland miRNome is available from lactating Alpine goats ⁷. In ovine
53 species, however, no mammary gland data were available until recently, with the publication
54 of a miRNome by Wang et al. ¹².

55 Either single nucleotide polymorphism (SNP) or InDel (small insertion or deletion) variations
56 may alter microRNAs via their biogenesis or impact on mRNA. Indeed, if the genetic variant is
57 located within the seed region, it may affect the binding between the microRNA and targeted
58 mRNAs and thus lead to the nonrecognition of targeted mRNAs or, in contrast, to the
59 recognition of novel mRNAs or novel targets ¹³. If the genetic variant is located in another part
60 of the microRNA or in its flanking regions, the microRNA expression level may be impacted ¹⁴.

61 MicroRNA polymorphisms have been studied in various livestock species, such as chicken,
62 porcine, and bovine, from the perspective of pathology or production traits. Wu et al. (2019)
63 found a genetic variant in *pri-miR-26a-5p*, the primary transcript of *miR-26a-5p*, inducing a
64 modification of the secondary structure of the microRNA gene ¹⁵. The abundance of the
65 mature microRNA is decreased when the mutation is present, impacting chicken egg

66 production traits¹⁵. Similarly, a SNP in porcine *miR-208b* was predicted to affect the secondary
67 structure of *pri-miR-208b* and therefore the expression of the microRNA precursor *pre-miR-*
68 *208b*, the mature microRNA, and the expression of the targets *SOX-6* and *MYH7* genes.
69 Differences in mRNA levels affect the density of muscular fibers and thus muscle and meat
70 quality traits¹⁶.

71 In bovine species, a genetic variant in the seed region of *bta-miR-2899* was detected (chr18:
72 42,198,087 G > A), localized in a QTL affecting the somatic cell score (SCS), and associated with
73 the presence of mastitis. Cows with the AA genotype had a significantly lower SCS than cows
74 with the GG or GA genotype. An impact on the targeted mRNA *SPI1*, which is a potential
75 candidate factor for inflammation in the bovine, was also investigated. The *SPI1* relative
76 expression in mammary glands of cows with *bta-miR-2899* with the AA genotype was
77 significantly higher than those with the GA or GG genotype¹⁷.

78 Bovine, caprine and ovine ruminants provide 86% of the world milk production¹⁸. In these three
79 species, the improvement of dairy traits has always been a core objective of breeding programs
80 and remains in the era of genomic selection. In genomic evaluation, breeding values are usually
81 predicted using SNP beadchips. The markers used for the genomic evaluation do not necessarily
82 have biological function in the genetic determinism of the trait considered, but, being spread
83 across the genome, some of them are in linkage disequilibrium with the causal polymorphisms
84 and capture part of their effect. Although using sets of neutral markers for genomic evaluation
85 is efficient, adding causal variants could improve the reliability of the estimated genomic
86 breeding values, as mentioned by Oget et al., who compared different methods¹⁹.

87 Accounting for causal variants in genomic prediction models could increase the genomic
88 selection accuracy. Numerous studies conducted in all three species described genomic
89 regions or quantitative trait loci (QTLs) associated with dairy traits. Some of these QTLs are
90 currently available in the public AnimalQTLdb database²⁰, and additional studies have listed
91 QTLs for milk production, milk composition, and mastitis resistance traits (see, for example,
92 ^{19,21–23}). However, the identification of causal mutations is still challenging.

93 Missense mutations located in the *DGAT1* and *PAEP* genes were known to be causative for
94 milk production and composition ^{22,24–26}. *DGAT1* encodes an enzyme involved in fatty acid
95 metabolism in milk, while *PAEP* encodes β -lactoglobulin, which is the most abundant whey
96 protein in bovine milk. In addition to these two examples, causal variants located in coding
97 regions of genes are scarce, and most of the candidate variants that are proposed as causative
98 for quantitative traits, and in particular dairy traits, are located in noncoding regions. Indeed,
99 noncoding regions represent approximately 98% of mammalian genomes and have functional
100 impacts on biological processes ^{27–29}. In the GWAS (genome-wide association studies),
101 approximately 90% of the candidate variants are localized within these noncoding regions
102 which can have regulatory roles ^{30,31}. Transcription start sites, enhancers or promoters of
103 genes are noncoding regions leading to the activation of genes in different tissues ^{32–35}.
104 Noncoding RNAs, which are not translated into proteins, could modulate gene expression and
105 impact biological systems, such as milk production or development of the mammary gland
106 ^{36,37}.

107 Some microRNA variants still remain unexplored. The aim of this study was to detect putative
108 causal microRNA variants that could influence dairy traits in three ruminant species. To target
109 microRNA variants that could potentially be the causal variants for these traits, only genetic
110 variants of microRNAs expressed in the mammary gland or present in milk and localized in
111 dairy QTLs were investigated. To this end, we developed a pipeline to filter microRNA genetic
112 variants from whole-genome variants in bovine, caprine and ovine species.

113

114

115

116

117

118

119 Material and Methods

120 1. *In silico* detection of genetic variants

121 a. VCF files

122 Similar processes were used for the three bovine, caprine and ovine species. The genomic
123 variation SNPs and small InDels, defined as insertions or deletions shorter than 60 bp located
124 in microRNA genes, were identified using three INRAE databases, as described in Table 1^{38,39}.

125 Small genomic variations located in bovine microRNA genes were identified using a bovine
126 database containing approximately 28 million SNPs and InDels using whole-genome sequence
127 data from 351 animals corresponding to 16 different breeds³⁸. Annotation of this database
128 was performed using Variant Effect Predictor v81 (VEP)⁴⁰.

129 For caprine species, the microRNA small genomic variations were identified using VCF files
130 containing approximately 87 million SNPs and InDels produced from whole genome
131 sequences for 1,159 animals, corresponding to 102 *Capra hircus* breeds as well as 8 other
132 capra species for 35 of the animals: *Capra aegagrus*, *Capra caucasica*, *Capra pyrenaica*, *Capra*
133 *falconeri*, *Capra ibex*, *Capra nubiana*, and *Capra siberica*. These data are accessible thanks to
134 the VarGoats project (last update: 1,159 goats, November 7th, 2019), corresponding to a
135 resequencing program of 1,000 goat genomes^{39,41}.

136 For ovine species, the VCF files used contained approximately 38 million SNPs from 87 animals
137 representing 12 different breeds and 1 crossbreed.

138 Briefly, sequence alignments against the UMD3.1 bovine reference genome⁴² were carried
139 out using the Burrows-Wheeler Alignment tool (BWA-v0.6.1-r104)⁴³ with the "aln" option.
140 The NCBI *Capra hircus* ARS1 assembly release 102 was used for the caprine species, and
141 Oar_v4.0 version was used for the ovine species. The removal of potential PCR duplicates was
142 performed using the MarkDuplicates tools from the Picard package version 1.4.0 (Broad
143 Institute, <http://picard.sourceforge.net>). Only properly paired reads with a mapping quality of
144 at least 30 ($-q = 30$) were retained. The resulting BAM files were subsequently used to the
145 variant calling, using the Genome Analysis Tool Kit 2.4–9 (GATK) version and GATK-

146 UnifiedGenotyper⁴⁴. All the identified variants were then annotated using the Ensembl
147 Variant Effect Predictor (VEP) v81 tool⁴⁰ based on the Ensembl version 81 transcript set and
148 using dbSNP build 143. The SnpEff v4.3 tool permits variant annotation and prediction of their
149 effects in caprine and ovine species⁴⁵. Finally, the effect of the amino acid changes in bovine
150 was predicted using SIFT v5.2.2^{46,47}.

151 In caprine species, the process differed in the calling variants step: the SAMtools tool version
152 1.6 was used at this step as well as the GATK version 3.6 tool⁴⁴. BCFtools version 1.6 and
153 Freebayes version 1.1.0 were also used⁴⁸.

154

155 b. BED files

156 i. QTL

157 In the three species, the QTLs detected for each trait in French breeds are given in *Table 2*. For
158 each QTL, the genomic location, the nature of trait affected, and the breed(s) in which it was
159 found were specified in the BED file.

160 In cattle, dairy QTLs are those reported in the studies of Sanchez et al. and Tribout et al.^{23,49}.
161 From daughter yield deviations, 84 QTLs were identified for milk production (milk, protein and
162 fat yields), milk composition (protein and fat content) and mastitis resistance (clinical mastitis
163 and somatic cell counts) in 6,321 Holstein, 2,515 Montbéliarde and 2,203 Normande bulls²³.
164 In addition, 34 QTLs were identified for milk protein composition (α 1, α 2, β and κ -caseins,
165 β -lactoglobulin and α -lactalbumin) predicted from the mid-infrared spectra in 2,306 Holstein,
166 2,967 Montbéliarde, and 2,737 Normande cows⁴⁹. A total of 1,019 QTL data were listed for
167 the three breeds and the different traits, corresponding to 118 distinct QTLs. In each study,
168 confidence intervals of the QTLs were defined by the positions of the SNPs included in the
169 upper third of the GWAS peak. As a result, the regions spanned a total length of almost 401
170 megabases, with a mean length of 815 kb (Table 2).

171 In caprine studies, 82 QTLs for five different traits (fat content, protein content, fat yield,
172 protein yield, and milk yield) were studied in two French alpine and Saanen dairy breeds²².

173 Due to different reference genomes (Assembly CHIR_1.0 versus ARS1 (GenBank accession
174 GCA_001704415.1)), the positions of the 82 QTLs were transposed to the ARS1 goat assembly.
175 This step was carried out by selecting the 100 first and the 100 last nucleotides in each
176 confidence interval of the QTLs using the Genome Data Viewer tool on NCBI ⁵⁰. Then, the
177 flanking regions of each QTL were blasted *via* the online NCBI tool
178 (<https://blast.ncbi.nlm.nih.gov/> - ⁵¹). The NCBI megablast tool ("Highly similar sequences") and
179 the general parameters applied were those by default (short queries, expected threshold:
180 0.05) ⁵¹. The QTL coordinates on the ARS1 genome version were deduced from the BLAST
181 results. As a result, the 82 regions corresponded to a total length of almost 18 megabases,
182 with a mean length of 226 kb (Table 2).

183 In ovine species, 107 dairy QTLs were available for Manech Tête Rouse and Lacaune breeds
184 in the publication of Oget et al. ¹⁹ and INRAE unpublished data (Hélène Larroque, personal
185 communication). The QTL regions were transposed to the Oar_v4.0 genome version through
186 the UCSC LiftOver tool ⁵². Seventeen traits are represented in these dairy QTLs, listed in Table
187 2. These 107 regions corresponded to a total length of almost 77 megabases, with a mean
188 length of 1,197 kb.

189

190 *ii. miRNomes*

191 To characterize the microRNAs expressed in the mammary gland and/or present in milk, dairy
192 miRNomes were used. The miRNomes of bovine mammary glands with or without mastitis
193 and the miRNomes of milk or different fractions of milk were used. A total of 1,063 microRNAs
194 are listed in these data sets ^{6,8-11}. For each listed microRNA, its genomic locations in the
195 UMD3.1 bovine genome and its name were indicated.

196 In caprine species, the mammary gland miRNome data were available from Mobuchon et al.
197 ⁷. A total of 239 microRNAs are listed, and their genomic locations in the ARS1 genome version
198 were used.

199 In ovine species, no mammary gland or milk miRNome data were available when the script
200 was developed. Therefore, to prioritize ovine microRNAs detection, bovine and caprine
201 miRNome data were used for the ovine species.

202

203 c. Data filtering

204 Files containing genomic variants, QTLs and miRNomes were combined and further analyzed
205 using an in-house Python script. This script consisted of a cascade of several filtering steps
206 allowing the selection of genetic variants located in dairy and/or mastitis QTLs, within
207 microRNA gene regions expressed in the mammary gland and/or present in milk. The script
208 used the microRNA, miRNome and frequency data as filters added in this script. First, only
209 biallelic variants affecting microRNAs were further filtered from the annotated VCF files.
210 Second, we selected genetic variants that impact microRNAs present in the dairy miRNome.
211 Finally, the variants with a frequency higher than 0.01 were selected and corresponded to our
212 miRNA genetic variant panel (Figure 1a).

213 Then, the QTL data filter was added to this whole script to be more precise and more stringent.
214 Only genetic variants located in dairy or mastitis QTLs were selected. Then, the microRNA,
215 miRNome and frequency filters were processed (Figure 1b). This panel also contained
216 different features from the original three files, such as their genomic position, nature of
217 reference, alternative alleles and their frequencies, the impacted microRNA, the breed in
218 which the variant is polymorphic, and the QTL traits in which the variant was located.
219 Additional filters were applied to the variants to classify them as follows: i) variants located in
220 microRNA precursors (in the seed, the mature microRNA out of the seed or in the precursor
221 microRNA out of the mature) or ii) in flanking regions of different sizes (50, 100, 500 or 1,000
222 base pairs upstream or downstream of the microRNA precursor), which were predefined
223 according to the literature data ⁵³⁻⁵⁸, and iii) in bovine species, CpG island or transcription
224 factor binding site (TFBS) elements that could have an impact on gene expression were added
225 to the scripts. These data were based on UCSC (<https://genome.ucsc.edu>) for the CpG islands
226 and on Bickhart et al. for the TFBS data in bovine ⁵⁹.

227 2. Prediction of mRNA targeted by microRNAs

228

229 The TargetScan release 7.2 bioinformatic tool was used to determine putative mRNA targets
230 of microRNAs of interest (<http://www.targetscan/>) in bovine, caprine and ovine species based
231 on the bovine database. This tool predicts and lists mRNAs that are potential targets of each
232 microRNA. The target prediction is based on the complementarity between the microRNA and
233 mRNAs in the seed sequence corresponding to nucleotides 2 to 7⁶⁰, which is the recognition
234 site in the microRNA. Additionally, the transcripts from *in silico* prediction were compared to
235 the transcriptome data of the bovine mammary gland during lactation and the dry period⁶¹.

236

237 Results

238 1. Detection of dairy microRNA genetic variants

239 Out of a total of 28 million genetic variants in bovine, 87 million in caprine, and 38 million
240 variants in ovine species, 138,442, 26,777 and 4,769 microRNA variants were identified,
241 respectively.

242 Of them, only the variants located in the microRNA genes of the dairy miRNomes were
243 retained, *i.e.*, 895 bovine, 239 caprine and 906 ovine microRNAs. After this filter, 59,843,
244 13,427, and 4,763 variants were kept in bovine, caprine, and ovine species, respectively.

245 Finally, in bovine, the selection of genetic variants with a frequency higher than 0.01 led to a
246 total of 59,124 genetic variants (SNPs and small InDels) of microRNAs expressed in the
247 mammary gland or present in milk. Among them, 360 genetic variants were located in a
248 microRNA precursor, with 213 microRNAs presenting at least one variant. Of them, 101
249 variants were located in 75 mature microRNAs, including 44 in a seed region of 37 different
250 microRNAs (Table 3). The 58,764 additional genetic variants were located in flanking regions
251 of microRNAs. In caprine species, 13,427 microRNA genetic variants were filtered, with 100
252 microRNAs presenting at least one variant. Of them, 33 variants were located in a precursor

253 of 24 different microRNAs, including four variants in three mature microRNAs and one genetic
254 variant in a seed region of *chi-miR-425*. The script with ovine data resulted in 4,763 microRNA
255 genetic variants in 106 different microRNAs. Of them, 33 variants were located in the
256 precursor of 19 microRNAs, with six of them in five mature microRNAs and one variant in the
257 seed region of *oar-miR-539* (Table 3). In addition, the microRNA *miR-539* presented a genetic
258 variant in bovine species in a 100 bp-flanking region of this microRNA.

259 In these analyses, five microRNAs (*miR-93*, *miR-215*, *miR-671*, *miR-874*, and *miR-1307*) had
260 variants in their precursor in both bovine and caprine species (Figure 2). These five microRNAs
261 are very similar in the two species, with the same seeds and, therefore, the same predicted
262 mRNA targets. They differ by only one nucleotide at the 3' end of the mature sequence (Figure
263 3a).

264 Four microRNAs, *miR-30b*, *miR-148a*, *miR-150*, and *miR-654*, had variants in both bovine and
265 ovine species (Figure 2). The microRNAs had very similar sequences with identical seeds and
266 therefore the same predicted mRNA targets (Figure 3b).

267 Out of the 28 million bovine genetic variants (SNPs and InDels), 59,124 were variants of
268 microRNAs expressed in the mammary gland or present in milk, of which 4,679 were located
269 in dairy QTLs. Of these, 1,044, 573, 140, and 69 variants were located within 1 kbp, 500 bp,
270 100 bp, and 50 bp flanking regions of 80, 77, 56, and 38 different microRNAs, respectively
271 (Table 4). Twenty-six variants, located in 21 different microRNA precursors, were found in QTL
272 regions associated with six different dairy traits: protein content, protein yield, fat content,
273 fat yield, somatic cell count, and milk yield. These 26 genetic variants corresponded to 22
274 SNPs, two insertions (one and two nucleotides) and two single nucleotide deletions (Table 5).
275 Eight genetic variants (6 SNPs, 1 insertion and 1 deletion) were identified in seven different
276 mature microRNAs in QTL associated with five traits: protein content, protein yield, fat
277 content, fat yield, and somatic cell count. Finally, three SNPs were present in the seed regions
278 of three different microRNAs, *bta-let-7e*, *bta-miR-2888* and *bta-miR-449b*, in QTL regions
279 associated with protein content, protein yield, and fat yield. Out of the 26 genetic variants
280 located in microRNA precursors, 20 were found in Holstein, four in Normande, and one in

281 Montbéliarde breed, and one genetic variant was found in both the Holstein and
282 Montbéliarde breeds. The frequency of the variants as well as their positions within CpG
283 islands were also considered in this study (Table 5). A total of 74 microRNA genetic variants,
284 including four microRNA precursors, are located in CpG islands. None of the microRNA
285 variants located in QTL regions were located in TFBS.

286 The same approach led to a total of 127 genetic variants in caprine species. These variants
287 impacted only one microRNA, *chi-miR-22*. All of those variants were located upstream and
288 downstream of the precursor microRNA, and the nearest was located 14 nucleotides apart
289 from the precursor. These genetic variants were located in a dairy QTL associated with fat
290 yield. They were identified in 73 caprine breeds, including the Alpine and Saanen dairy breeds.
291 No variant was found in the microRNA *miR-22* in bovine species.

292 In ovine species, no genetic variant was identified using this bioinformatic script.

293 In the three dairy species, the analysis led to 511, 100 and 106 microRNAs showing at least
294 one variant in bovine, caprine, and ovine species, respectively, without considering
295 colocalization with dairy and mastitis QTLs. More specifically, five microRNAs had variants only
296 in caprine species, while 23 microRNAs had variants only in ovine species. Finally, 20
297 microRNAs had genetic variants in all three ruminant species (Figure 4).

298

299 2. Prediction of putative mRNA targets

300 To find a functional link between microRNAs impacted by the genetic variants detected (Table
301 3) and dairy traits, the putative mRNA targets of microRNAs were predicted using *in silico*
302 analyses.

303 *miR-539*, showing a variant in its seed sequence in the ovine, has 5,300 putative target mRNAs,
304 including 63 and 31 mRNAs that are up- and down-regulated, respectively, in the mammary
305 gland during lactation *versus* non-lactation periods. The *miR-425* with a variant in its seed
306 sequence in caprine species is predicted to have 1,621 putative target mRNAs, 25 of which

307 were found to be dysregulated in the mammary gland during lactation *versus* non-lactation
308 periods (5 were down-regulated and 20 were upregulated).

309 Among the 305 transcripts that are predicted as putative targets of the microRNA *miR-874*,
310 with genetic variants in both bovine and caprine microRNA precursors, four putative targets
311 of *miR-874* are shown to be differentially expressed between lactation and non-lactation
312 periods ⁶¹. *ESRRA* (estrogen-related receptor alpha) and *ATF3* (activating transcription factor
313 3) are upregulated during the lactation period compared to the non-lactation period, while
314 *GREB1* (growth regulation by estrogen in breast cancer 1) and *CD248* (CD248 molecule) are
315 down-regulated ⁶¹.

316 For the microRNA *miR-1307*, which is polymorphic in both bovine and caprine species, 11
317 putative targets were found in the bovine mammary gland transcriptome data. Ten of them
318 were reported to be upregulated (*ADAM12*, *AGPAT6*, *FASN*, *MT1A*, *PMM2*, *PTPRT*, *RND1*,
319 *SLC2A4*, *SULF2*, and *THBS2*), and one of them was found to be downregulated (*PRR15L*), during
320 lactation *versus* non-lactation periods ⁶¹.

321

322 3. QTL information depending on species

323 On average, 21.8, 4.6, and 1.7 variants per kbp were detected in bovine, caprine, and ovine
324 species, respectively.

325 The addition of the "dairy and mastitis QTL" filter led to a large decrease in the number of
326 genetic variants. To compare the data, differences in QTL information available in the three
327 species were studied.

328 The number of QTLs differed depending on the species. In the bovine, caprine and ovine
329 species, the QTLs cover 401.5, 18.0, and 76.9 Mbp of the genome, respectively. The mean size
330 of the QTLs, between the bounds of the confidence interval, varies between species: 815 kb
331 in bovine, with QTL sizes from 190 to 4,156,369 bp, approximately 226 kb in caprine, with a
332 range of 174,986 to 403,132 bp, and approximately 1,234 kb in ovine species, from 766,403
333 to 1,997,460 bp. Therefore, the number of QTLs and the length of the genome covered by

334 these regions were much higher in bovine than in the two other species. The total size of the
335 QTLs in bovine is more than fivefold longer than that observed in ovine and more than twenty-
336 two fold longer than that in caprine species (Figure 5). In bovine QTLs, the number of variants
337 in microRNA genes reached 11.65 per megabase, and 7.06 caprine QTLs were identified.

338

339

340 Discussion

341

342 MicroRNA genetic variants are well studied in humans to diagnose or provide a prognosis for
343 diseases such as cancer or cardiovascular disease ⁶²⁻⁶⁵. MicroRNA databases, in particular the
344 MSDD (MicroRNA SNP Disease Database) ⁶⁶ and the miRSNP database ⁶⁷, have been developed
345 in this context. Other databases, such as EpimiRBase or miRCancer, provide access to the
346 microRNAs linked with human databases ^{68,69}.

347 Approximately 10% of human microRNA precursors contain SNPs, with less than 1% located
348 in seed regions ⁷⁰.

349 In the present study, microRNA genetic variants were studied in the context of dairy traits and
350 in three different dairy ruminant species: bovine, caprine, and ovine species.

351 A total of 59,124, 13,427, and 4,763 genetic variants in microRNAs expressed in the mammary
352 gland or present in the milk were identified in bovine, caprine, and ovine species, respectively.

353 Although the number of whole-genome variants investigated in this study was higher in
354 caprine (87 million) and ovine (38 million) species than in bovine (28 million) species, a much
355 higher number of variants in microRNA expressed in the mammary gland or present in milk
356 was observed in bovine species than in the two other species. These differences in genetic
357 variant amounts might be due to the amount of data existing in the different criteria.
358 Knowledge of the different genomes, the number of genetic variants in the different species,

359 and the number of microRNAs described in dairy miRNomes are different depending on the
360 species.

361 In this study, the microRNAs listed in mammary miRNomes were used. A total of 895 bovine
362 microRNAs, including 701 microRNAs detected in mammary gland^{6,8}, 239 caprine microRNAs
363 and 106 ovine microRNAs were then retained. These differences likely explain the larger
364 number of microRNA genetic variants in bovine than in the two other species.

365 As an example, a total of 59,124 microRNA genetic variants were detected in bovine species.
366 Research on genetic variants of interest in bovine species using caprine miRNome data instead
367 of bovine data led to 18,184 microRNA genetic variants. This is in the same order as the caprine
368 results, which leads to 13,427 caprine microRNA genetic variants. The mammary miRNomes
369 seem therefore to be an important factor explaining the difference in the number of genetic
370 variants of interest in the three studied species. The lack of data regarding the caprine and
371 ovine mammary data may explain the difference in results between species.

372 Of the variants found in microRNAs expressed in the mammary gland and present in milk,
373 4,679; 127; and 0 genetic variants were identified in dairy QTLs in bovine, caprine and ovine
374 species, respectively.

375 In addition, the total size of the dairy QTL was longer in bovine species than in ovine (five-fold
376 shorter) and caprine (twenty-two-fold shorter) species. Thus, the total genome length in
377 which the microRNA genetic variants were searched was longer in bovine species, which could
378 explain a part of the differences we observed in the three species: the number of variants
379 detected per QTL megabase is 1.65 higher in bovine than in caprine species. This difference
380 in the size of dairy QTLs may also be induced by the difference in criteria used to define the
381 QTLs.

382 More precisely, the difference in the number of variants in the three dairy ruminant species is
383 highlighted by the mean number of genetic variants of interest (genetic variants of microRNAs
384 expressed in the mammary gland or present in milk) per kilobase: this number is more than

385 four-fold higher in bovine species (21.8 variants per kb) than in caprine species (4.6 variants
386 per kb) and more than 12-fold higher than in ovine species (1.8 variants per kb).

387 Some of the microRNAs found to contain at least one genetic variant of interest appear to
388 have links with dairy traits. These links can appear through some of their putative mRNA
389 targets, which are genes involved in dairy trait determinism, or through the literature,
390 describing some of them as actors in milk traits. *AGPAT6*, impacting milk composition ⁷¹, or
391 the leptin *LEP* genes, impacting milk production ^{72,73}, are putative targets of microRNAs such
392 as *bta-miR-2888*, *miR-1307* and *miR-425*. Another putative target, *PRLR*, has an effect on milk
393 fatty acid composition ⁷⁴.

394 Some microRNAs with genetic variants identified in this analysis, such as *miR-143* and *miR-25*
395 in ovine or *miR-150* in both bovine and ovine species, were found to be involved in milk fatty
396 acid and protein metabolism. For example, the microRNAs *miR-143* and *miR-25* were found
397 to be involved in the regulation of milk fat synthesis ^{75,76}, while *miR-150* was associated with
398 the lactation process and milk protein production ⁷⁷.

399 Among the five microRNAs presenting a genetic variant in their precursor, some are involved
400 in immune responses or dairy traits. For instance, *miR-93* expression is higher in bovine
401 colostrum than in mature milk ⁷⁸, and *miR-874* is upregulated between the galactopoiesis and
402 involution lactation stages in Canadian Holstein cows ^{10,79}.

403 The microRNAs *miR-30b* and *miR-150* have variants in both bovine and ovine species. Both
404 were previously associated with lactation: overexpression of *miR-30b* caused a lactation
405 defect in mice ⁴, and the expression of *miR-150* resulted in the suppression of mRNAs involved
406 in lipid synthesis during secretion ⁷⁷.

407 The *miR-103* and *miR-107* microRNAs contain genetic variants in all three species studied.
408 These two microRNAs belong to the same microRNA family and thus show identical seed
409 sequences in bovine, caprine and ovine species ⁸⁰.

410 The causal role of these variants on dairy traits remains to be demonstrated, but they
411 constitute potential candidate variants for some of the dairy QTLs identified in ruminants.

412 Variants we highlighted in this study in microRNAs could therefore be added to genomic
413 prediction models and participate in the improvement of genomic selection.

414 Conclusions

415

416 This study led to the detection of microRNA genetic variants of interest in the context of dairy
417 production, taking advantage of whole genome sequence data to identify microRNA genetic
418 variants expressed in the mammary gland and localized in some dairy QTLs. A perspective of
419 this study is to choose some genetic variants and characterize them to test their potential
420 functional impacts.

421

422

423

424 Authors' contributions

425 F.LP, G.T-K., D.B., and M.B. conceived the study, P.B., M-P.S., S.LG., T.T., H.L., and R.R. provided
426 data, C.B. and M.B. performed the experimentations, C.B., F.LP. and G.T-K. wrote the paper,
427 and all authors reviewed the manuscript.

428 Competing interests

429 The authors declare no competing interests.

430

431 Funding

432 The financial support of the APIS-GENE, through the miRQTLait project, is acknowledged.
433 Céline Bourdon's grant was financially supported by INRAE Animal Genetics department and
434 APIS-GENE.

- 436 1. Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
- 437 2. Lewis, B. P., Shih, I., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of
438 Mammalian MicroRNA Targets. *Cell* **115**, 787–798 (2003).
- 439 3. Tanaka, T., Haneda, S., Imakawa, K., Sakai, S. & Nagaoka, K. A microRNA, miR-
440 101a, controls mammary gland development by regulating cyclooxygenase-2
441 expression. *Differentiation* **77**, 181–187 (2009).
- 442 4. Le Guillou *et al.* Overexpression of miR-30b in the Developing Mouse Mammary
443 Gland Causes a Lactation Defect and Delays Involution. *PLoS ONE* **7**, (2012).
- 444 5. Yoo, K. H. *et al.* The STAT5-regulated miR-193b locus restrains mammary stem and
445 progenitor cell activity and alveolar differentiation. *Dev. Biol.* **395**, 245–254 (2014).
- 446 6. Le Guillou *et al.* Characterisation and Comparison of Lactating Mouse and Bovine
447 Mammary Gland miRNomes. *PLoS ONE* **9**, e91938 (2014).
- 448 7. Mobuchon, L. *et al.* Food Deprivation Affects the miRNome in the Lactating Goat
449 Mammary Gland. *PLOS ONE* **10**, e0140111 (2015).
- 450 8. Ju, Z. *et al.* Solexa sequencing and custom microRNA chip reveal repertoire of
451 microRNAs in mammary gland of bovine suffering from natural infectious mastitis.
452 *Anim. Genet.* **49**, 3–18 (2018).
- 453 9. Le Guillou *et al.* Characterization of Holstein and Normande whole milk miRNomes
454 highlights breed specificities. *Sci. Rep.* **9**, (2019).
- 455 10. Do, D. N., Li, R., Dudemaine, P.-L. & Ibeagha-Awemu, E. M. MicroRNA roles in
456 signalling during lactation: an insight from differential expression, time course and
457 pathway analyses of deep sequence data. *Sci. Rep.* **7**, 44605 (2017).
- 458 11. Li, R., Dudemaine, P.-L., Zhao, X., Lei, C. & Ibeagha-Awemu, E. M. Comparative
459 Analysis of the miRNome of Bovine Milk Fat, Whey and Cells. *PLOS ONE* **11**,
460 e0154129 (2016).
- 461 12. Wang, J. *et al.* Small RNA deep sequencing reveals the expressions of microRNAs in
462 ovine mammary gland development at peak-lactation and during the non-lactating
463 period. *Genomics* (2020) doi:10.1016/j.ygeno.2020.09.060.
- 464 13. Hill, C. G., Jabbari, N., Matyunina, L. V. & McDonald, J. F. Functional and
465 Evolutionary Significance of Human MicroRNA Seed Region Mutations. *PLoS ONE* **9**,
466 (2014).
- 467 14. Króliczewski, J., Sobolewska, A., Lejnowski, D., Collawn, J. F. & Bartoszewski, R.
468 microRNA single polynucleotide polymorphism influences on microRNA biogenesis
469 and mRNA target specificity. *Gene* **640**, 66–72 (2018).
- 470 15. Wu, H. *et al.* Variants of pri-miR-26a-5p polymorphisms are associated with values for
471 chicken egg production variables and affects abundance of mature miRNA. *Anim.*
472 *Reprod. Sci.* **201**, 93–101 (2019).
- 473 16. Kim, J. M. *et al.* A polymorphism in the porcine miR-208b is associated with
474 microRNA biogenesis and expressions of SOX-6 and MYH7 with effects on muscle
475 fibre characteristics and meat quality. *Anim. Genet.* **46**, 73–77 (2015).
- 476 17. Jiang, Q. *et al.* In silico genome-wide miRNA-QTL-SNPs analyses identify a
477 functional SNP associated with mastitis in Holsteins. *BMC Genet.* **20**, 46 (2019).
- 478 18. FAO, Food and Agriculture Organization of the United Nations. FAOSTAT.
479 <http://www.fao.org/faostat/en/#data/QL> (2019).

- 480 19. Oget, C., Teissier, M., Astruc, J.-M., Tosser-Klopp, G. & Rupp, R. Alternative methods
481 improve the accuracy of genomic prediction using information from a causal point
482 mutation in a dairy sheep model. *BMC Genomics* **20**, 719 (2019).
- 483 20. Hu, Z.-L., Fritz, E. R. & Reecy, J. M. AnimalQTLdb: a livestock QTL database tool set
484 for positional QTL information mining and beyond. *Nucleic Acids Res.* **35**, D604–D609
485 (2007).
- 486 21. Sanchez, M.-P. *et al.* Within-breed and multi-breed GWAS on imputed whole-genome
487 sequence variants reveal candidate mutations affecting milk protein composition in
488 dairy cattle. *Genet. Sel. Evol.* **49**, 68 (2017).
- 489 22. Martin, P. *et al.* A genome scan for milk production traits in dairy goats reveals two
490 new mutations in *Dgat1* reducing milk fat content. *Sci. Rep.* **7**, (2017).
- 491 23. Tribout, T. *et al.* Confirmed effects of candidate variants for milk production, udder
492 health, and udder morphology in dairy cattle. *Genet. Sel. Evol.* **52**, 55 (2020).
- 493 24. Grisart, B. *et al.* Positional Candidate Cloning of a QTL in Dairy Cattle: Identification
494 of a Missense Mutation in the Bovine *DGAT1* Gene with Major Effect on Milk Yield
495 and Composition. *Genome Res.* **12**, 222–231 (2002).
- 496 25. Ganai, N. A., Bovenhuis, H., Arendonk, J. A. M. V. & Visker, M. H. P. W. Novel
497 polymorphisms in the bovine β -lactoglobulin gene and their effects on β -lactoglobulin
498 protein concentration in milk. *Anim. Genet.* **40**, 127–133 (2009).
- 499 26. Scatà, M. C. *et al.* Ovine acyl CoA:diacylglycerol acyltransferase 1– molecular
500 characterization, polymorphisms and association with milk traits. *Anim. Genet.* **40**,
501 737–742 (2009).
- 502 27. Moraes, F. & Góes, A. A decade of human genome project conclusion: Scientific
503 diffusion about our genome knowledge. *Biochem. Mol. Biol. Educ.* **44**, 215–223
504 (2016).
- 505 28. Sun, X. *et al.* The developmental transcriptome sequencing of bovine skeletal muscle
506 reveals a long noncoding RNA, *lncMD*, promotes muscle differentiation by sponging
507 *miR-125b*. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1863**, 2835–2845 (2016).
- 508 29. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in
509 the human genome. *Nature* **489**, 57–74 (2012).
- 510 30. Edwards, S. L., Beesley, J., French, J. D. & Dunning, A. M. Beyond GWASs:
511 Illuminating the Dark Road from Association to Function. *Am. J. Hum. Genet.* **93**, 779–
512 797 (2013).
- 513 31. Giral, H., Landmesser, U. & Kratzer, A. Into the Wild: GWAS Exploration of Non-
514 coding RNAs. *Front. Cardiovasc. Med.* **5**, (2018).
- 515 32. Bansal, M., Kumar, A. & Yella, V. R. Role of DNA sequence based structural features
516 of promoters in transcription initiation and gene expression. *Curr. Opin. Struct. Biol.*
517 **25**, 77–85 (2014).
- 518 33. Chadaeva, I. V. *et al.* Candidate SNP markers of reproductive potential are predicted by
519 a significant change in the affinity of TATA-binding protein for human gene
520 promoters. *BMC Genomics* **19**, (2018).
- 521 34. Gasperini, M., Tome, J. M. & Shendure, J. Towards a comprehensive catalogue of
522 validated and target-linked human enhancers. *Nat. Rev. Genet.* 1–19 (2020)
523 doi:10.1038/s41576-019-0209-0.
- 524 35. Schoenfelder, S. & Fraser, P. Long-range enhancer–promoter contacts in gene
525 expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).

- 526 36. Kawaji, H. & Hayashizaki, Y. Exploration of Small RNAs. *PLoS Genet.* **4**, e22 (2008).
- 527 37. Do, D. N. & Ibeagha-Awemu, E. M. Non-Coding RNA Roles in Ruminant Mammary
528 Gland Development and Lactation. in *Current Topics in Lactation* (ed. Gigli, I.)
529 (InTech, 2017). doi:10.5772/67194.
- 530 38. Boussaha, M. *et al.* Construction of a large collection of small genome variations in
531 French dairy and beef breeds using whole-genome sequences. *Genet. Sel. Evol.* **48**, 87
532 (2016).
- 533 39. VarGoats. VarGoats: data access.
534 http://www.goatgenome.org/vargoats_data_access.html (2019).
- 535 40. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl
536 API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
- 537 41. VarGoats Project. VarGoats. (2019).
- 538 42. Zimin, A. V. *et al.* A whole-genome assembly of the domestic cow, *Bos taurus*.
539 *Genome Biol.* **10**, R42 (2009).
- 540 43. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler
541 transform. *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).
- 542 44. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for
543 analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
- 544 45. Cingolani, P. *et al.* A program for annotating and predicting the effects of single
545 nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster*
546 strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
- 547 46. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous
548 variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081
549 (2009).
- 550 47. Ng, P. C. & Henikoff, S. Predicting Deleterious Amino Acid Substitutions. *Genome*
551 *Res.* **11**, 863–874 (2001).
- 552 48. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read
553 sequencing. *ArXiv12073907 Q-Bio* (2012).
- 554 49. Sanchez, M.-P. *et al.* Within-breed and multi-breed GWAS on imputed whole-genome
555 sequence variants reveal candidate mutations affecting milk protein composition in
556 dairy cattle. *Genet. Sel. Evol.* **49**, 68 (2017).
- 557 50. NCBI. Genome Data Viewer.
558 https://www.ncbi.nlm.nih.gov/genome/gdv/browser/genome/?id=GCF_001704415.1.
- 559 51. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–
560 W9 (2008).
- 561 52. Lee, C. M. *et al.* UCSC Genome Browser enters 20th year. *Nucleic Acids Res.* **48**,
562 D756–D761 (2020).
- 563 53. Chen, J. *et al.* Genetic Variations in the Flanking Regions of miR-101-2 Are
564 Associated with Increased Risk of Breast Cancer. *PLoS ONE* **9**, e86319 (2014).
- 565 54. Duan, R., Pak, C. & Jin, P. Single nucleotide polymorphism associated with mature
566 miR-125a alters the processing of pri-miRNA. *Hum. Mol. Genet.* **16**, 1124–1131
567 (2007).
- 568 55. Gong, J. *et al.* Genome-wide identification of SNPs in microRNA genes and the SNP
569 effects on microRNA target binding and biogenesis. *Hum. Mutat.* **33**, 254–263 (2012).
- 570 56. Liu, C. *et al.* Effects of genetic variations on microRNA: target interactions. *Nucleic*
571 *Acids Res.* **42**, 9543–9552 (2014).

- 572 57. Muiños-Gimeno, M., Montfort, M., Bayés, M., Estivill, X. & Espinosa-Parrilla, Y.
573 Design and evaluation of a panel of single-nucleotide polymorphisms in microRNA
574 genomic regions for association studies in human disease. *Eur. J. Hum. Genet.* **18**, 218–
575 226 (2010).
- 576 58. Thomas, L. F. & Sætrom, P. Circular RNAs are depleted of polymorphisms at
577 microRNA binding sites. *Bioinformatics* **30**, 2243–2246 (2014).
- 578 59. Bickhart, D. M. & Liu, G. E. Identification of Candidate Transcription Factor Binding
579 Sites in the Cattle Genome. *Genomics Proteomics Bioinformatics* **11**, 195–198 (2013).
- 580 60. Agarwal, V., Bell, G. W., Nam, J.-W. & Bartel, D. P. Predicting effective microRNA
581 target sites in mammalian mRNAs. *eLife* **4**, e05005 (2015).
- 582 61. Dai, W., Zou, Y., White, R. R., Liu, J. & Liu, H. Transcriptomic profiles of the bovine
583 mammary gland during lactation and the dry period. *Funct. Integr. Genomics* **18**, 125–
584 140 (2018).
- 585 62. Bastami, M. *et al.* miRNA Polymorphisms and Risk of Cardio-Cerebrovascular
586 Diseases: A Systematic Review and Meta-Analysis. *Int. J. Mol. Sci.* **20**, (2019).
- 587 63. Gao, Z. *et al.* Effects of mutations in porcine miRNA-215 precursor sequences on
588 miRNA-215 regulatory function. *Gene* **701**, 131–138 (2019).
- 589 64. Sung, J.-H. *et al.* miRNA polymorphisms (miR-146a, miR-149, miR-196a2 and
590 miR-499) are associated with the risk of coronary artery disease. *Mol. Med. Rep.* **14**,
591 2328–2342 (2016).
- 592 65. Zeng, J. *et al.* Polymorphisms in four microRNAs and risk of oral squamous cell
593 cancer: a meta-analysis. *Oncotarget* **9**, 8695–8705 (2018).
- 594 66. Yue, M. *et al.* MSDD: a manually curated database of experimentally supported
595 associations among miRNAs, SNPs and human diseases. *Nucleic Acids Res.* **46**, D181–
596 D185 (2018).
- 597 67. Yuan, Y. & Weidhaas, J. B. Functional microRNA binding site variants. *Mol. Oncol.*
598 **13**, 4–8 (2019).
- 599 68. Backes, C. *et al.* miRCarta: a central repository for collecting miRNA candidates.
600 *Nucleic Acids Res.* **46**, D160–D167 (2018).
- 601 69. Xie, B., Ding, Q., Han, H. & Wu, D. miRCancer: a microRNA-cancer association
602 database constructed by text mining on literature. *Bioinformatics* **29**, 638–644 (2013).
- 603 70. Saunders, M. A., Liang, H. & Li, W.-H. Human polymorphism at microRNAs and
604 microRNA target sites. *Proc. Natl. Acad. Sci.* **104**, 3300–3305 (2007).
- 605 71. Littlejohn, M. D. *et al.* Expression Variants of the Lipogenic AGPAT6 Gene Affect
606 Diverse Milk Composition Phenotypes in *Bos taurus*. *PLoS ONE* **9**, (2014).
- 607 72. Balakirev, N. A., Saphina, N. Yu., Yulmeteva, Yu. R., Shakirov, Sh. K. & Zinnatova,
608 F. F. Association of Leptin Gene (LEP) Polymorphism with Growth Rates and Milk
609 Production in Holstein First-Calf Heifers. *Russ. Agric. Sci.* **44**, 460–464 (2018).
- 610 73. Kulig, H. Associations between leptin gene polymorphism and some milk performance
611 traits of cattle. *J. Anim. Feed Sci.* **14**, 235–243 (2005).
- 612 74. Shi, L. *et al.* Polymorphisms and genetic effects of PRLR, MOGAT1, MINPP1 and
613 CHUK genes on milk fatty acid traits in Chinese Holstein. *BMC Genet.* **20**, (2019).
- 614 75. Zhang, L., Wu, Z.-Q., Wang, Y.-J., Wang, M. & Yang, W.-C. MiR-143 Regulates Milk
615 Fat Synthesis by Targeting Smad3 in Bovine Mammary Epithelial Cells. *Animals* **10**,
616 1453 (2020).

- 617 76. Ma, L. *et al.* miR-25 modulates triacylglycerol and lipid accumulation in goat
618 mammary epithelial cells by repressing PGC-1beta. *J. Anim. Sci. Biotechnol.* **9**, 48
619 (2018).
- 620 77. Heinz, R. E. *et al.* Constitutive expression of microRNA-150 in mammary epithelium
621 suppresses secretory activation and impairs de novo lipogenesis. *Development* **143**,
622 4236–4248 (2016).
- 623 78. Sun, Q. *et al.* Immune modulatory function of abundant immune-related microRNAs in
624 microvesicles from bovine colostrum. *Protein Cell* **4**, 197–210 (2013).
- 625 79. Putz, E. J. *et al.* MicroRNA profiles of dry secretions through the first three weeks of
626 the dry period from Holstein cows. *Sci. Rep.* **9**, 19658 (2019).
- 627 80. Chen, H.-Y. *et al.* miR-103/107 Promote Metastasis of Colorectal Cancer by Targeting
628 the Metastasis Suppressors DAPK and KLF4. *Cancer Res.* **72**, 3631–3641 (2012).
629

630 Figure legends

631 Figure 1: Data filtering process without **(a)** and with **(b)** the dairy and mastitis QTL filter.
632 From the genetic variants in the VCF file, different filters were added to kept the
633 variants of interest.

634 Table 1: Information on bovine, caprine and ovine genetic variants files.

635 Table 2: Dairy traits affected by the Quantitative Trait Loci, with the breed for bovine,
636 caprine, and ovine species. The abbreviations of the trait annotations are specified in
637 the table.

638 Table 3: Number of microRNA genetic variants detected, with and without the dairy QTL
639 filter, for the three different species.

640 Figure 2: Number of microRNAs presenting at least one variant in a precursor in bovine,
641 caprine and ovine species. Five microRNAs present variants both in bovine and caprine
642 species. Four microRNAs present variants both in bovine and ovine species.

643 Figure 3: **a)** Nucleotidic sequence of the five microRNAs presenting genetic variants in the
644 precursor of both bovine and caprine species. The differences are located at the end of each
645 microRNA. **b)** Nucleotidic sequence of the three microRNAs presenting genetic variants in the
646 precursor of both bovine and ovine species. The seed regions of the microRNAs are indicated
647 between the black traits.

648 Table 4: Number of detected bovine microRNA genetic variants of interest, according to
649 their locations. The number of microRNAs presenting at least 1 variant was mentioned,
650 as well as the number of variants with a frequency ≥ 0.10 and those located in CpG
651 islands.

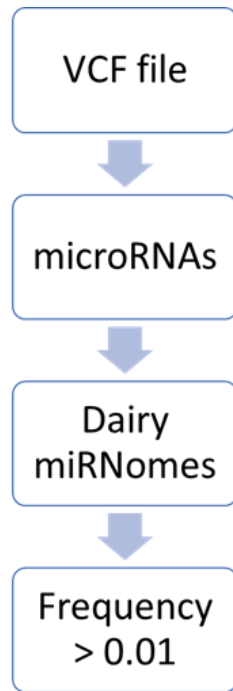
652 Table 5: Description of the 26 genetic variants detected in the precursor of microRNAs in
653 bovine QTLs.

654 Figure 4: Number of microRNAs presenting at least one variant in bovine, caprine or
655 ovine species and common variants.

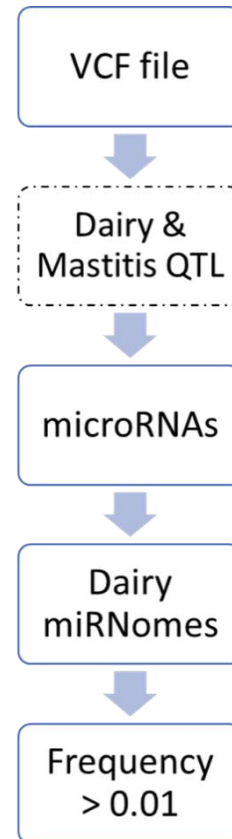
656 Figure 5: Total size of the whole QTLs depending on the species, in megabases (Mb).

677 Figure 1: Data filtering process without (a) and with (b) the dairy and mastitis QTL filter.
678 From the genetic variants in the VCF file, different filters were added to kept the
679 variants of interest.

680 a)



b)



681 Table 1: Information on bovine, caprine and ovine genetic variants files.

	Number of variants	Number of animals	Number of breeds
Bovine	28 million	351	16
Caprine	87 million	1,159 (+ 35)	102(+ 8 other capra species)
Ovine	38 million	87	12 + 1 crossbreed

682

682 Table 2: Dairy traits affected by the Quantitative Trait Loci, with the breed for bovine,
 683 caprine, and ovine species. The abbreviations of the trait annotations are specified in
 684 the table.

Trait		Bovine			Caprine		Ovine	
		<i>Holstein</i>	<i>Normande</i>	<i>Montbéliarde</i>	<i>Alpine</i>	<i>Saanen</i>	<i>Lacaune</i>	<i>Manech</i>
<i>α-lactalbumin content</i>	a-LA	X	X	X				
<i>β-lactoglobulin content</i>	b-LG	X	X	X			X	
<i>αs1- casein content</i>	as1-CN	X	X	X			X	X
<i>αs2- casein content</i>	as2-CN	X	X	X			X	X
<i>β-casein content</i>	b-CN	X	X	X			X	X
<i>κ-casein content</i>	k-CN	X	X	X			X	
<i>Total caseins</i>	s-CN	X	X	X			X	X
<i>Milk yield</i>	MILK	X	X	X	X	X	X	
<i>Fat yield</i>	FAT	X	X	X	X	X	X	
<i>Protein yield</i>	PROT	X	X	X	X	X	X	
<i>Fat content</i>	FC	X	X	X	X	X	X	
<i>Protein content</i>	PC	X	X	X	X	X	X	
<i>Somatic cell score</i>	SCS	X	X	X	X	X	X	
<i>Clinical mastitis</i>	CM	X	X	X				
<i>Monounsaturated fatty acids</i>	MUFA						X	X
<i>Polyunsaturated fatty acids</i>	PUFA						X	
<i>Unsaturated fatty acids</i>	UFA						X	X
<i>Saturated fatty acids</i>	SFA						X	X
<i>Solid proteins</i>	SP							X

685 Table 3: Number of microRNA genetic variants detected, with and without the dairy QTL
 686 filter, for the three different species.

	WITHOUT QTL DATA				DAIRY QTL			
	Total	Precursor	Mature	Seed	Total	Precursor	Mature	Seed
BOVINE	59,124	360	101	44	4,679	26	10	3
CAPRINE	13,427	33	4	1	127	0	0	0
OVINE	4,761	33	6	1	0	0	0	0

686 Figure 2: Number of microRNAs presenting at least one variant in a precursor in bovine,
687 caprine and ovine species. Five microRNAs present variants both in bovine and caprine
688 species. Four microRNAs present variants both in bovine and ovine species.

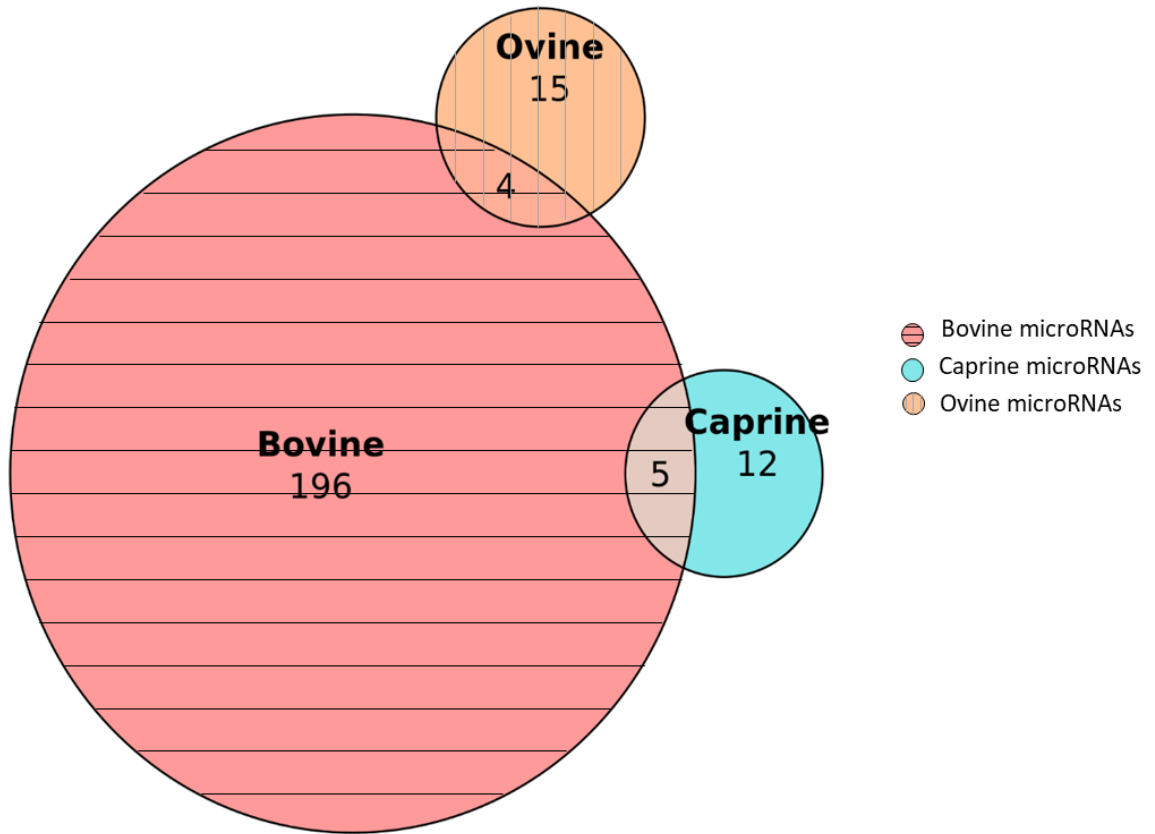


Figure 3: **a)** Nucleotidic sequence of the five microRNAs presenting genetic variants in the precursor of both bovine and caprine species. The differences are located at the end of each microRNA. **b)** Nucleotidic sequence of the three microRNAs presenting genetic variants in the precursor of both bovine and ovine species.

The seed regions of the microRNAs are indicated between the black traits.

a)

miR-93	CAAAGUGCUGUUCGUGCAGGUA- CAAAGUGCUGUUCGUGCAGGUAG	Bovine Caprine
miR-215	AUGACCUAUGAAUUGACAGACA AUGACCUAUGAAUUGACAGAC-	Bovine Caprine
miR-671	AGGAAGCCUGGAGGGGCUGGAG- AGGAAGCCUGGAGGGGCUGGAGG	Bovine Caprine
miR-874	CUGCCCUGGCCCGAGGGACCGA- CUGCCCUGGCCCGAGGGACCGAC	Bovine Caprine
miR-1307	ACUCGGCGUGGCGUCGGUCGUG- ACUCGGCGUGGCGUCGGUCGUGG	Bovine Caprine

b)

miR-30b	UGUAAACAUCCUACACUCAGCU UGUAAACAUCCUACACUCAGC-	Bovine Ovine
miR-150	UCUCCCAACCCUUGUACCAGUGU UCUCCCAACCCUUGUACCAGUG-	Bovine Ovine
miR-654	UAUGUCUGCUGACCAUCACCUU- - UAUGUCUGCUGACCAUCACCUUUA	Bovine Ovine
miR-148a	UCAGUGCACUACAGAACUUUGU UCAGUGCACUACAGAACUUUGU	Bovine Ovine

Table 4: Number of detected bovine microRNA genetic variants of interest, according to their locations. The number of microRNAs presenting at least 1 variant was mentioned, as well as the number of variants with a frequency ≥ 0.10 and those located in CpG islands.

Variant location	Number of variants / <i>Number of different microRNAs</i>	Variants with frequency $\geq 10\%$ / <i>Number of different microRNAs</i>	Number of variants in CpG islands / <i>Number of different microRNAs</i>
Seed region	3 / 3	0	0
Mature microRNA	8 / 7	1 / 1	0
microRNA precursor	26 / 21	8 / 8	4 / 4
± 50 bp	69 / 38	33 / 20	17 / 8
± 100 bp	140 / 56	66 / 33	27 / 10
± 500 bp	573 / 77	291 / 61	50 / 14
± 1 000 bp	1,044 / 80	597 / 78	74 / 24

Table 5: Description of the 26 genetic variants detected in the precursor of microRNAs in bovine QTL regions.

<i>Chr</i>	<i>Position</i>	<i>Reference allele</i>	<i>Alternative allele</i>	<i>Breed</i>	<i>Frequency alternative allele</i>	<i>microRNA</i>	<i>QTL trait</i>	<i>Localisation</i>
3	102923794	G	A	Holstein	0.136	<i>bta-mir-2415</i>	Milk yield & Quantity of proteins	Precursor
4	114614954	T	G	Holstein	0.047	<i>bta-mir-671</i>	Milk yield & Fat content	Precursor
5	30251690	G	C	Holstein	0.673	<i>bta-mir-2425</i>	Protein yield	Precursor
5	30952192	T	G	Holstein	0.038	<i>bta-mir-2426</i>	Protein yield	Precursor
5	62117780	G	A	Holstein	0.019	<i>bta-mir-135a-2</i>	Fat content	Precursor
5	84244517	G	A	Holstein & Montbéliarde	0.071 & 0.017	<i>bta-mir-2436</i>	Fat content & Fat yield	Precursor
5	101542291	C	T	Holstein	0.019	<i>bta-mir-2284r</i>	Fat yield	Precursor
5	118347364	C	CT	Holstein	0.012	<i>bta-mir-2284h</i>	Protein yield	Precursor
6	99976608	T	C	Normande	0.048	<i>bta-mir-2446</i>	Protein yield	Mature
6	99976613	G	A	Holstein	0.013	<i>bta-mir-2446</i>	Protein yield	Precursor
7	50687460	T	C	Holstein	0.210	<i>bta-mir-874</i>	Fat content	Precursor
7	63886927	C	CCA	Holstein	0.026	<i>bta-mir-2461</i>	Quantity of proteins & Fat content	Mature
15	34628957	C	T	Holstein	0.296	<i>bta-mir-2313</i>	Fat content	Precursor
18	56407853	TC	T	Holstein	0.056	<i>bta-mir-150</i>	Quantity of proteins	Precursor
18	56407916	T	G	Holstein	0.013	<i>bta-mir-150</i>	Quantity of proteins	Mature
18	58015050	G	A	Holstein	0.051	<i>bta-let-7e</i>	Quantity of proteins	Seed
18	61145895	G	T	Normande	0.012	<i>bta-mir-371</i>	Protein yield	Precursor
19	38542897	CG	C	Holstein	0.304	<i>bta-mir-2886</i>	Protein yield & Somatic cell count	Mature
19	39081170	T	C	Holstein	0.903	<i>bta-mir-152</i>	Somatic cell count	Precursor
20	23967291	G	T	Normande	0.083	<i>bta-mir-449b</i>	Fat yield	Seed
20	23967292	T	A	Normande	0.083	<i>bta-mir-449b</i>	Fat yield	Mature
21	36134549	C	T	Montbéliarde	0.121	<i>bta-mir-2888-1</i>	Protein yield	Precursor
21	36134560	T	G	Montbéliarde	0.044	<i>bta-mir-2888-1</i>	Protein yield	Mature
25	35300154	G	C	Holstein	0.051	<i>bta-mir-2388</i>	Protein yield	Precursor
25	35300168	G	A	Holstein	0.756	<i>bta-mir-2388</i>	Protein yield	Precursor
29	45520815	C	T	Holstein	0.294	<i>bta-mir-2408</i>	Protein yield	Precursor

Figure 4: Number of microRNAs presenting at least one variant in bovine, caprine or ovine species and common variants.

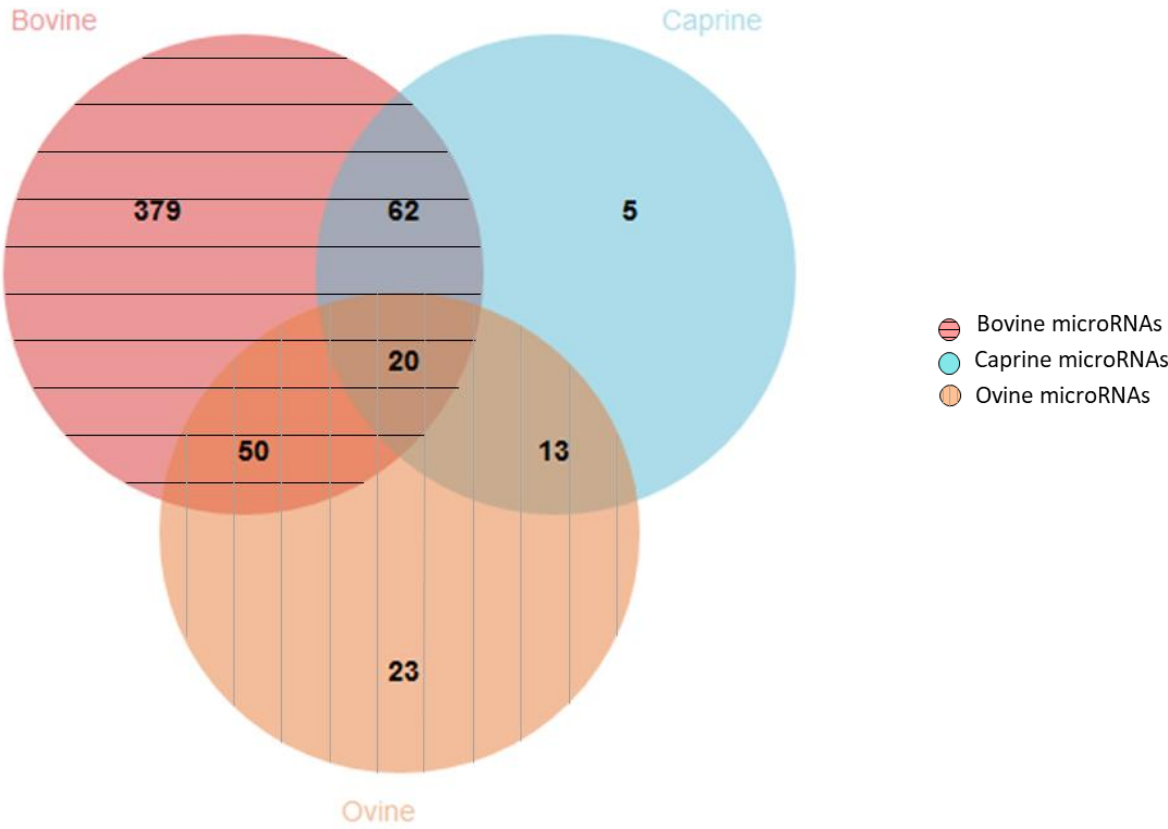
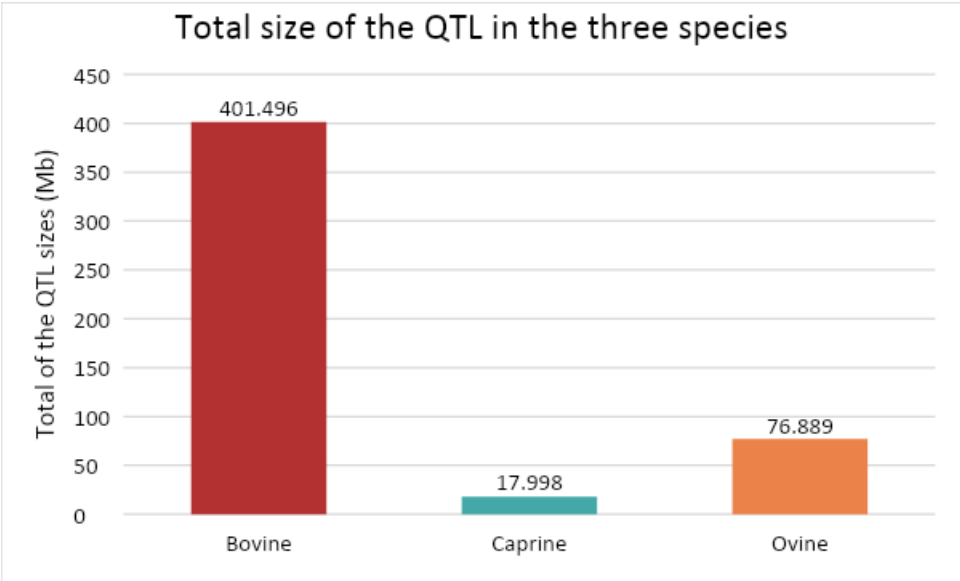


Figure 5: Total size of the whole QTL regions depending on the species, in megabases (Mb).



3. Validation de variants de microARNs détectés

a. Contexte

Cette partie de la thèse a pour objectif de mieux comprendre les mécanismes moléculaires et donc les potentiels impacts fonctionnels de variants de microARNs.

Suite à la détection de variants génétiques de microARNs exprimés dans la glande mammaire ou présents dans le lait et localisés dans des QTL laitiers, des variants génétiques peuvent être ajoutés sur des puces à ADN (puces à SNPs) utilisés dans le génotypage pour la sélection génomique de ces espèces. Ces variants ainsi ajoutés pourront alors être validés : cette technique permet de réaliser un nombre important de génotypages des variants de microARNs, aboutissant ainsi à la possibilité de les imputer par la suite. En effet, les présences de ces variants pourront être statistiquement déterminées à partir des génotypages disponibles d'animaux pour lesquels on dispose de données de performances. Ces données permettront de réaliser de nouvelles analyses d'associations (de type GWAS), et ce sur des animaux différents de ceux utilisés lors des étapes de détection de variants et d'études d'associations préalables. Cette partie de validation permet de confirmer des liens génotype-phénotype et donc de valider des variants détectés.

b. Matériel et méthodes

i. Validation de variants génétiques via puces à SNPs

Des variants génétiques détectés sont sélectionnés selon des critères définis ci-après, afin d'être apposés sur les puces à SNPs bovines et caprines. La puce à SNPs bovine est nommée EuroG10K (Illumina), comportant environ 10 000 SNPs. La puce caprine 50K est utilisée pour des variants génétiques détectés en espèce caprine. Cette puce à SNPs regroupe et permet un génotypage pour un total de plus de 50 000 SNPs. Les variants choisis le sont selon des critères d'importance potentielle du variant génétique : i) La localisation du variant génétique est prise en compte : les variants situés dans une région "seed" ont été sélectionnés prioritairement car ils ont un potentiel impact fonctionnel sur l'expression d'ARNm ;

ii) la localisation du variant dans un îlot CpG est également pris en compte : les variants situés dans ces régions peuvent avoir un potentiel impact sur la transcription *via* notamment des phénomènes de méthylation ; iii) l'importance d'après la littérature du microARN associé est un troisième facteur de sélection ; et iv) la présence dans le lait du microARN est également prise en compte (en privilégiant les variants génétiques impactant des microARNs ayant une plus forte présence dans le lait). De manière générale, la bibliographie et les taux de présence des microARNs comportant des variants dans le lait sont des critères permettant de prioriser des variants pour leur ajout sur la puce bovine. La localisation intronique de variants génétiques est également un critère.

En caprin, la détection de variants génétiques a dans un premier temps été effectuée à partir d'un fichier de variants de 594 animaux, par la suite une mise à jour regroupant 1 159 animaux a été utilisée. L'ensemble des 34 variants génétiques d'intérêt détectés à partir des données des 594 animaux en espèce caprine, lors de l'analyse avec le critère de localisation dans une région de QTL laitier, ont pu être apposés sur la puce à SNP caprine 50K. En ce qui concerne l'espèce bovine, 44 variants génétiques sont sélectionnés selon les critères définis, parmi les 4 679 variants détectés, et ajoutés à la puce bovine LD bovine EuroG10K (Illumina) v2. Au préalable, 50 variants génétiques ont été sélectionnés : deux d'entre eux sont ajoutés du fait de leur localisation dans une région "seed", 26 pour leur localisation dans une région CpG associée à une importance du microARN dans la littérature pour des critères laitiers ou une présence dans le lait, 10 variants génétiques ont été choisis car détaillés dans la littérature pour leur impact sur des caractères laitiers, 9 variants pour leur présence forte dans le lait et finalement trois variants génétiques sont sélectionnés pour leur localisation intronique. Par la suite, six variants génétiques n'ont pas passé les critères de qualité afin de pouvoir être ajoutés sur la puce, c'est pourquoi 44 variants génétiques sont apposés sur la puce à SNPs bovine. Pour le moment, nous sommes dans l'attente de la mise en circulation de ces puces à SNPs et du génotypage d'un nombre d'animaux suffisants.

Chr	Position	Allèle de Référence (0)	Allèle Alternatif (1)	Race	Freq_0	Freq_1	microARN	Caractère	-LOG(p-value)
27	36262931	T	G	Montbéliarde	0.557	0.443	<i>bta-mir-486</i>	tb	9.4
				Normande	0.488	0.512			7.4
3	15480606	A	G	Normande	0.314	0.686	<i>bta-mir-92b</i>	tp	7.6
				Holstein	0.519	0.481			
18	580150150	G	A	Holstein	0.948	0.051	<i>bta-let-7e</i>	mp	8.9

Tableau 6 : Caractéristiques des trois variants génétiques validés en bovin grâce aux données de GWAS, pour des critères de taux butyreux (tb), taux protéique (tp) et de matière protéique (mp).

ii. Validation de variants génétiques via études de GWAS

Cette étape de validation de variants génétiques est réalisée en espèce bovine grâce à des données de GWAS. Ces données de GWAS comptent 27 millions de variants génétiques imputés, à partir des données disponibles pour une population de référence [153]. Trois études GWAS sont utilisées, basées sur des animaux de race Holstein, Normande et Montbéliarde [153], [154], [291]. Ces études sont décrites dans l'introduction de ce manuscrit.

c. Résultats

i. Ajout de variants sur puces à SNPs

L'ajout de ces variants génétiques sur les puces à SNPs a été effectué mais pour le moment le nombre de génotypages réalisés n'est pas suffisant pour permettre la validation de variants par cette méthode. Il faudra attendre un nombre d'animaux génotypés sur ces puces à SNPs suffisant pour pouvoir par la suite en déduire d'éventuels liens phénotypes-génotypes et donc valider des variants génétiques par cette méthode.

ii. Validation par études de GWAS

En utilisant ces données de GWAS, trois variants génétiques de microARNs sont validés : il s'agit de *bta-let-7e* chr18:58015050 G>A, *bta-miR-486* chr27:36262931 T>G, et *bta-miR-92b* chr3:15480606 A>G. Ces variants sont associés à une valeur de $-\log(p\text{-value})$ supérieure à 7 (donc une p-value inférieure à $1e^{-7}$), et donc considérés comme significatifs [292]. Cette valeur de $-\log(p\text{-value})$ correspond à une indication de la significativité statistique de l'association entre génotype et phénotype pour le variant sélectionné. Les variants situés dans les régions flanquantes du précurseur de *bta-miR-486* et *bta-miR-92b* sont associés à des caractères de matières grasses et de taux de protéines dans le lait, respectivement. Le variant génétique situé dans la région "seed" du microARN *bta-let-7e* est, lui, associé au caractère de quantité de protéines dans le lait (Tableau 6). Le variant génétique impactant *bta-miR-486* est détecté dans les races Montbéliarde et Normande, le variant *bta-miR-92b* est lui détecté en races Holstein et Normande. En ce qui concerne le variant génétique situé dans la région "seed" du *bta-let-7e*, celui-ci est détecté seulement en race Holstein.

d. Conclusion

Cette étape de validation de variants génétiques, qui ont été détectés grâce au script bio-informatique, est réalisée de deux manières distinctes. Tout d'abord, des variants génétiques bovins et caprins détectés sont ajoutés sur les puces à SNPs bovine et caprine respectivement. Un grand nombre d'animaux peut donc être génotypés pour ces 44 variants génétiques bovins et 34 variants caprins. Des liens génotypes-phénotypes pourront alors être établis et permettre une éventuelle validation de certains de ces variants génétiques. Cette méthode a des implications à plus long terme, puisqu'il faut attendre la mise en circulation de ces puces à SNPs et le génotypage d'un nombre d'animaux suffisants pour analyser les résultats et faire des études d'association.

L'autre méthode de validation est l'utilisation en bovin de données de GWAS. Cela a permis la validation de lien génotype-phénotype pour trois variants génétiques, impliquant trois microARNs distincts, ainsi que trois caractères laitiers différents. Cette validation a été effectuée sur des données existantes, donc avec des résultats immédiats.

Les impacts fonctionnels potentiels de ces trois variants génétiques bovins validés sont alors testés.

4. Etudes des potentiels impacts de variants de microARNs validés

a. Contexte

En espèce bovine, trois variants génétiques de microARNs ont été détectés, puis validés par des études d'associations génotype-phénotype. Les impacts fonctionnels potentiels de ces variants génétiques sont alors étudiés.

Ceux-ci peuvent différer selon la localisation du variant, dans les régions flanquantes du précurseur de microARN ou dans la séquence du microARN mature.

Si le variant génétique est situé dans la région "seed" du microARN, donc dans le site de reconnaissance entre le microARN et des ARN messagers cibles, ce changement de nucléotide peut compromettre la reconnaissance d'un ARN messager qui est considéré comme étant une cible du microARN et/ou même provoquer la reconnaissance d'un ARN messager comme nouvelle cible.

L'un des variants validés précédemment dans cette étude est situé dans la région "seed" du microARN *bta-let-7e* : il s'agit du nucléotide 8 du microARN mature (*Tableau 7 & Figure 16*).

A contrario, si le variant génétique se situe en-dehors de cette région spécifique du microARN, l'impact ne se situera pas au niveau de la reconnaissance directe de l'ARN messager cible, mais au niveau de l'expression du microARN lui-même. En effet, cela pourra impacter la biogenèse du microARN à diverses étapes, *via* les prises en charge par Drosha ou DICER par exemple. Dans ces cas, la maturation du microARN primaire en précurseur de microARN ou la maturation en microARN mature pourront être compromises, respectivement, et ainsi aboutir à une modification de l'expression du microARN. Dans notre cas, son expression dans la glande mammaire pourrait être impactée. Une étude a mis en évidence une corrélation entre l'expression d'un microARN dans la glande mammaire et sa présence dans le lait [293]. C'est pourquoi l'étude de variants génétiques situés dans ces régions se déroule en utilisant des échantillons de lait et non des biopsies de glande mammaire.

Chr	Position	Allèle de référence	Allèle alternatif	Race	Fréquence référence	Fréquence alternatif	microARN	Caractère	Distance du précurseur
3	15480606	A	G	Normande	0.314	0.686	<i>bta-mir-92b</i>	tp	119 pb
27	36262931	T	G	Montbéliarde Normande	0.557 0.488	0.443 0.512	<i>bta-mir-486</i>	tb	990 pb

Tableau 8 : Caractéristiques des variants génétiques impactant les microARNs *bta-miR-92b* et *bta-miR-486* et détectés dans des caractères QTL de taux protéique (tp) et taux butyreux (tb) respectivement.

Les variants génétiques validés impactant les microARNs *bta-miR-92b* et *bta-miR-486* sont situés dans les régions situées en amont ou en aval de la séquence du microARN respectif et donc pourrait influencer sur cette expression.

Le variant génétique impactant le *bta-miR-92b* est situé à une distance de 119 paires de bases en amont du précurseur du microARN. Le variant impactant le microARN *bta-miR-486* est, lui, situé à 990 paires de bases du précurseur (*Tableau 8*).

Les études réalisées ont pour but, dans le cas de *bta-let-7e*, de déterminer des cibles potentielles de ces variants génétiques et ensuite d'évaluer l'impact potentiel de l'allèle G et de l'allèle A sur l'expression de ces ARNm dans les cellules épithéliales mammaires bovines. Dans le cas de *bta-miR-486* et *bta-miR-92b*, le but est d'analyser la présence de ces microARNs dans le lait de vaches selon le génotype des animaux pour le variant détecté et voir si leur présence est différente en fonction du variant étudié.

Afin de s'assurer des bonnes méthodes de prélèvement et conservation de lait destiné à l'analyse de la présence de microARNs dans le lait, une expérimentation préalable a été nécessaire. Cela a permis de vérifier l'intégrité de microARNs. En effet, la conservation de microARNs dans le lait a été évaluée selon le temps passé par le lait à température ambiante ou à 4°C, afin de définir un délai maximal entre la traite et la congélation des laits.

b. Etude des impacts potentiels du variant *bta-let-7e*

i. Matériel et méthodes

1. Culture cellulaire

La lignée de cellules épithéliales mammaires bovines BME-UV1 est utilisée pour les expérimentations de transfection [294]. Cette lignée a l'avantage d'être une lignée bovine pour laquelle nous disposons des données de RNAseq suite aux travaux de thèse de Lenha Mobuchon [295]. Environ un million de cellules sont cultivées à 37°C, avec 5% de CO₂, en monocouches dans des flasques de culture en plastique. Le milieu de culture est composé comme suit : le DMEM-F12 (Sigma), le RPMI 1640 (Sigma) et le NCTC-135 (Sigma) sont ajoutés

à des proportions de 5:3:2 par volume. Ils sont associés à de l' α -lactose (0,10%) (Sigma), source d'énergie. De la lactalbumine (0,10%) (Sigma), qui est source d'acides aminés. De la glutamine (2,5 $\mu\text{g}/\text{mL}$) (Eurobio) qui est un acide aminé permettant la croissance cellulaire, par la synthèse de protéines. De l'holo-transferrine (5 $\mu\text{g}/\text{mL}$) (Sigma) qui est utilisée dans le milieu de culture pour le transport de fer et sa faculté à détoxifier d'autres métaux. Du glutathion est ajouté (1,2 mM) (Sigma), pour son rôle antioxydant. De l'acide ascorbique (10 $\mu\text{g}/\text{mL}$) (Sigma), qui correspond à la vitamine C, qui est un facteur de croissance et qui permet le maintien du potentiel d'oxydo-réduction. Des hormones comme l'insuline bovine (1 $\mu\text{g}/\text{mL}$) (Sigma), de l'hydrocortisone (1 $\mu\text{g}/\text{mL}$) (Sigma) et de la progestérone (0,5 $\mu\text{g}/\text{mL}$) (Sigma) sont ajoutées dans le milieu de culture. Des antibiotiques sont ajoutés : de la pénicilline-streptomycine (50 UI/mL). Enfin, un sérum permettant l'adhérence et la prolifération cellulaire est utilisé : du SVF (sérum de veau fœtal) décomplémenté (10% vol/vol) (Gibco).

Ce milieu est changé trois fois par semaine, les cellules arrivant à confluence tous les trois à quatre jours.

2. Prédiction d'ARNm cibles

La prédiction *in silico* de cibles putatives du microARN *bta-let-7e* comportant l'allèle de référence puis comportant l'allèle alternatif est effectuée avec le logiciel TargetScan custom release 5.2 (http://www.targetscan.org/vert_50/seedmatch.html) en sélectionnant l'espèce bovine [237]. Parmi les cibles prédites, nous avons sélectionné les ARNm ciblés par le microARN *bta-let-7e* seulement lorsque l'allèle de référence est présent, d'une part, et seulement lorsque l'allèle alternatif est présent, d'autre part. De plus, ces deux listes de cibles ont été réduites aux ARN messagers suffisamment exprimés dans les cellules BME-UV1, selon nos données de RNAseq, pour garantir des niveaux d'expression détectables lors des analyses.

3. Transfection

Afin de déceler les potentiels impacts de la mutation détectée de *bta-let-7e* sur des ARNm cibles, les cellules BME-UV1 sont transfectées avec le microARN présentant l'allèle G ou l'allèle A. Afin de préparer la transfection par lipofectamine, les cellules sont transférées en plaques 24 puits, à raison de 50 000 cellules par puit. Après 24 heures, selon les conditions, le microARN avec l'allèle de G (*bta-let-7e*, assay ID MC10438 - ThermoFisher) ou l'allèle A (avec dans ce cas utilisation d'un mimic réalisé sur-mesure "custom", à partir de la séquence du microARN mature présentant l'allèle alternatif – ThermoFisher) est transfecté. Pour ce faire, 10 µM de mirVana® miRNA mimic sont transfectés en utilisant de la Lipofectamine® RNAiMAX (ThermoFisher), selon les instructions du fabricant, associé à de l'OptiMEM (ThermoFisher) qui est un milieu de culture recommandé lors de l'utilisation de réactifs de transfection de lipides cationiques tels que la Lipofectamine®, puisqu'il permet d'éviter d'utiliser le SVF qui a pour impact une diminution du taux de transfection [296], [297]. Ces mimics sont transfectés dans les cellules BME-UV1 durant 48h.

Afin de valider la transfection, un témoin positif de transfection est également utilisé sur la même plaque : le BLOCK-iT™ Alexa Fluor® Red Fluorescent Control (Ambion™), qui permet de visualiser la transfection par sa couleur rouge apparente. Pour cela, des cellules sont cultivées sur des lamelles de verre et transfectées avec ce BLOCK-iT™ Alexa Fluor®, puis la fixation des cellules est réalisée grâce à une solution de 4% PFA (ParaFormAldehyde) en PBS (Phosphate Buffered Saline) – pendant 20 minutes à 4°C. Après trois lavages, les lamelles sont montées sur des lames, en utilisant ensuite un milieu de montage contenant du DAPI. L'observation de la transfection s'effectue ensuite au microscope.

Un contrôle positif afin de valider l'impact d'un microARN transfecté sur l'expression d'un ARNm est également utilisé : le mirVana® mimic *miR-1* (référence 4464062), dont on connaît l'impact sur un ARNm cible (*TWF1*). Un contrôle a été ajouté : dans ce cas les cellules restent 48h dans un milieu comprenant l'optiMEM et la lipofectamine, sans "mimic".

Ces expérimentations de transfections sont répétées trois fois.

Gène	Sens (Forward)	Antisens (Reverse)
<i>OSTF1</i>	5'-GCAGATATTGTACAGTTGCTTCTGG-3'	5'-CAGGGCCAGCTTCTTCTCAT-3'
<i>SLC25A24</i>	5'-CCCCTCTGGATCGTCTGAAAG-3'	5'-TCTCAAAGCGAGCGGATAC-3'
<i>EIF4G2</i>	5'-GGACAGTCGAAGGATATGCCA-3'	5'-GCTGAAGCTTTGGCACTTGAT-3'
<i>EEF2K</i>	5'-CAAGCTCTGGGGAGAGGAGT-3'	5'-TCTCCTTCAGCTCTATGATGCAC-3'
<i>TWF1</i>	5'-GGCATCCAAGCAAGTGAAGA-3'	5'-GCTTCCTACACGACCCAATCA-3'
<i>PPIA</i>	5'-GGTGACTTCACACGCCATAATGGT-3'	5'-GGACAAGATGCCAGGACCTGTATG-3'

Tableau 9 : Séquences des amorces utilisées en qPCR pour les 4 ARNm testés, l'ARNm contrôle positif et le gène de référence

4. Extraction d'ARN

Des ARNs totaux sont utilisés pour l'analyse de ce variant génétique. Ils sont extraits soit à partir des cellules BME-UV1 après transfection. L'extraction d'ARNs totaux à partir de cellules s'effectue après rinçage des cellules avec du PBS.

L'extraction est réalisée grâce au RNA NOW kit (Ozyme), en suivant les instructions d'utilisation du fournisseur, en y ajoutant une étape de précipitation sur la nuit afin d'optimiser la précipitation des ARNs de petite taille (dont les microARNs). Les ARNs sont repris dans 30 µL d'eau final. La concentration et la pureté des ARNs extraits sont mesurées par spectrophotométrie, grâce à un Nanodrop™, ND-1000.

5. Reverse transcription et PCR quantitative

L'étape de transcription inverse (reverse transcription, RT) est réalisée en utilisant la SuperScript Vilo (11754-050, Invitrogen) en suivant les instructions d'utilisation du fournisseur, à 42°C pendant 60 minutes puis à 85°C pendant cinq minutes.

L'expression des ARNm est analysée par PCR quantitatives (qPCR), qui sont réalisées après choix des amorces grâce aux outils Primer-blast (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) pour les ARN messagers choisis ainsi que pour *TWF1*, cible du *miR-1* utilisé comme contrôle positif. La normalisation des mesures d'expression est réalisée en fonction de celle du gène de référence de la Peptidylprolyl isomerase A (cyclophilin A) : *PPIA*, utilisé comme gène de référence dans des analyses d'expression dans des cellules épithéliales mammaires bovines [298]. Les séquences des amorces choisis pour ces différents gènes sont présentées ci-après (*Tableau 9*).

Les qPCR sont réalisées en utilisant le Absolute Blue QPCR Mix, SYBER Green® selon les instructions d'utilisation du fournisseur. Les cycles de seuil (threshold cycles, Ct) obtenus sont normalisés, et les résultats sont ensuite exprimés en tant que niveau d'expression relatif par l'application de la méthode des $2^{-\Delta\Delta Ct}$ [299].

6. Séquençage et traitement des données

Les ARNs extraits à partir des cellules transfectées sont séquencés (Genewiz) en RNA-seq, avec une sélection polyA grâce au NovaSeq Illumina, avec un RNA-seq brin-spécifique (Illumina NovaSeq 2x150 bp sequencing). Les ARNs de départ ont des RIN (RNA integrity number) de 7,7 à 10,0. L'indexage et l'alignement sont réalisés grâce à l'outil STAR (STAR-2.5.1b) [300], avec pour génome de référence la version UMD3.1, Ensembl release 81 [301] du génome bovin, version utilisée lors de la détection de variants. Ensuite, une étape de nettoyage est réalisée avec le logiciel cutadapt, version 1.14 [302]. L'alignement peut alors être effectué sur le génome bovin, suivi de l'étape de quantification des transcrits grâce à RSEM, version 1.3.0 [303].

7. Analyses statistiques

Les présences des ARN messagers dans les cellules BME-UV1 selon la transfection du microARN *bta-let-7e* avec allèle G ou allèle A et la comparaison avec les transfections contrôles, sont analysées statistiquement avec un test de Wilcoxon, en utilisant le logiciel R, version 3.5.1 (R Development Core Team, 2013, <http://www.R-project.org/>). Le seuil de significativité est placé à une p-value $p < 0.05$.

Les données de séquençage haut débit sont analysées sous R, avec les packages ade4 et vegan pour les analyses multivariées [305], [306, p. 4], par analyse en composant principale (ACP), suivie d'une ANOVA afin de déterminer les axes significatifs [307]. Les différentiels de niveaux d'expression sont analysés par limma et plus précisément *via* une transformation voom des données [308], [309]. Les analyses différentielles sont réalisées en utilisant la procédure de Benjamini-Hochberg afin de calculer une p-value ajustée [310].

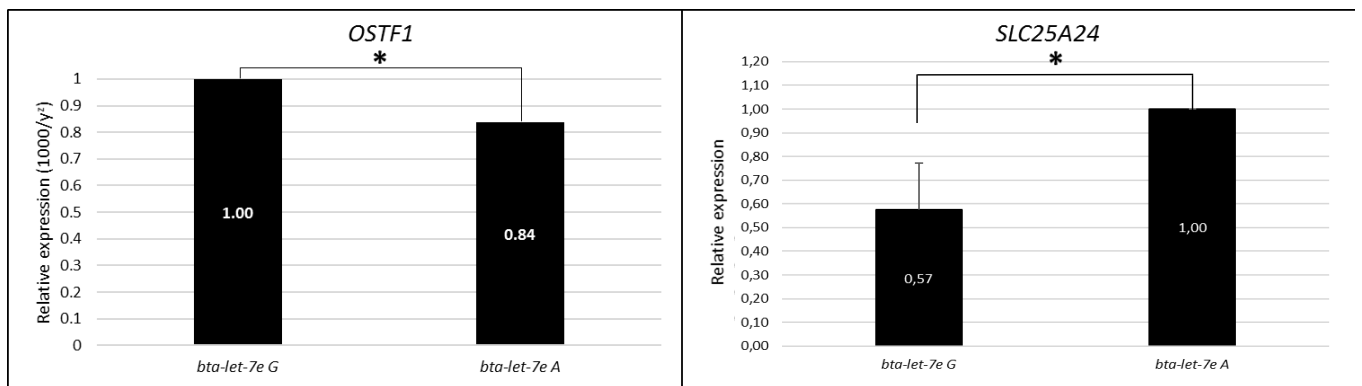


Figure 17 : Résultats relatifs du niveau d'expression de OSTF1 et SLC25A24 détecté en BME-UV1 selon l'allèle présent dans le microARN bta-let-7e transfecté

ii. Résultats

1. Prédiction d'ARNm cibles

Un total de 716 ARNs messagers sont prédits comme étant de potentielles cibles du microARN *bta-let-7e* avec l'allèle G et des 593 cibles potentielles lorsque l'allèle A est présent. Quatre ARNm faisant partie des 1000 transcrits les plus exprimés dans les cellules BME-UV1 sont choisis parmi eux : un ciblé uniquement par le *bta-let-7e* lorsque l'allèle A est présent : *OSTF1* (Osteoclast Stimulating Factor 1), et trois ciblés uniquement par le *bta-let-7e* lorsque l'allèle G est présent : *SLC25A24* (Solute Carrier Family 25 Member 24), *EEF2K* (Eukaryotic Elongation Factor 2 Kinase) et *EIF4G2* (Eukaryotic translation Initiation Factor 4 Gamma 2).

2. Expression de cibles prédites

L'étude de l'effet potentiel du variant génétique situé dans la région de reconnaissance des ARNm cibles du microARN *bta-let-7e* sur l'expression d'ARNm a été réalisée. Les niveaux d'expression de quatre ARNm cibles putatives du microARN sont analysées par RT-qPCR. Concernant les gènes *EIF4G2* et *EEF2K*, on ne détecte pas de différence significative du niveau d'expression relative à celle du gène de référence *PPIA* entre les échantillons issus des cellules transfectées avec le microARN comportant l'allèle G et ceux issus des cellules transfectées avec le microARN comportant l'allèle A. L'allèle A induit une réduction d'expression des transcrits *OSTF1* comparativement à l'allèle G (expression réduite de 14%, avec une p-value inférieure à 0,05) (*Figure 17*).

En ce qui concerne *SLC25A24*, une réduction d'expression est induite par l'allèle G. Son expression relative est significativement plus faible (-57%, $p < 0.05$) en présence de l'allèle G (*Figure 2*).

L'expression de 2 ARNm étudiés a donc été significativement modulée par la présence de *bta-let-7e* comportant l'allèle G ou l'allèle A.

Nombre de transcrits	Conditions	Allèle G vs Allèle A	Contrôle vs Allèle G	Contrôle vs Allèle A
Sous-exprimés		0	22	5
Sur-exprimés		0	11 872	11 899
Sans différentiel significatif d'expression		16 395	23	13

Figure 18 : Résultats des analyses statistiques des données de RNAseq

3. Séquençage et traitement des données

Afin d'avoir une vision plus globale de l'impact potentiel du variant situé dans la région "seed" du microARN, un séquençage d'échantillons issus des expériences en culture cellulaire est réalisé. Le séquençage des 9 échantillons issus de l'expérimentation de culture cellulaire avec transfections est effectué : 3 échantillons Contrôle, 3 échantillons avec transfection du microARN *bta-let-7e* présentant l'allèle G et 3 échantillons avec transfection du microARN *bta-let-7e* présentant l'allèle A. À la suite de l'étape d'indexage et d'alignement des données de séquences brutes, une étape de nettoyage est réalisée conduisant à l'élimination de 0,5% des séquences. Environ 38 millions de lectures par échantillon sont obtenus. Le mapping (la cartographie) peut alors être effectué sur le génome bovin de référence UMD3.1 qui a été choisi afin d'être en concordance avec l'étape de détection de variants d'intérêt qui a été réalisée à partir de cette version. Par la suite, le logiciel RSEM permet un alignement contre le transcriptome et l'étape de quantification des transcrits exprimés est alors effectuée. Les données de comptages et les Transcrits par Millions (TPM) sont obtenus. Un total de 26 740 transcrits uniques sont alors obtenus, après élimination des transcrits sans variation : possédant le même comptage pour les neuf échantillons, ce nombre s'élève à 16 395 transcrits au total.

Un total de 45 transcrits est trouvé différentiellement exprimé entre la condition Contrôle pour une p-value de 0,05, sans transfection du *bta-let-7e* et la condition *bta-let-7e* G ; 18 transcrits le sont toujours pour une p-value de 0,05 entre Contrôle et *bta-let-7e* A, indiquant une influence du microARN *bta-let-7e*, quel que soit l'allèle présent, sur l'expression de transcrits. Certaines p-valeurs seraient donc à la limite de la significativité entre Contrôle et *bta-let-7e* A et significatives entre Contrôle et *bta-let-7e* G, tout en ayant des comptages proches, de fait, entre *bta-let-7e* G et *bta-let-7e* A.

Néanmoins, aucun transcrit ne présente de différence d'expression significative en RNAseq entre la condition de transfection du microARN présentant l'allèle G et celle du microARN présentant l'allèle A. De fait, il ne semble pas y avoir eu d'impact sur le transcriptome des cellules transfectées, incluant les gènes *SLC25A24* et *OSTF1*, qui présentaient des différentiels d'expression entre ces conditions en RT-qPCR (*Figure 18*).

c. Test de l'intégrité de microARNs dans le lait à température ambiante et à 4°C

Afin de s'assurer de la conservation des microARNs dans les échantillons de lait que nous avons utilisés pour étudier les potentiels impacts de variants, cette étude préalable de l'intégrité des microARNs dans le lait a été réalisée au préalable.

Les laits de trois animaux différents sont utilisés pour effectuer des tests de conservation de microARNs : deux animaux de race Holstein et un animal de race Normande. Deux conditions de conservation sont testées, correspondantes aux deux possibilités au domaine expérimental où les laits ont été collectés : une conservation à température ambiante (23°C) et une conservation à 4°C. Les laits étaient préalablement conservés à -80°C en tube de 12 mL. Chacun a alors été décongelé à 4°C, pendant 5 heures, puis divisé en 14 aliquots de 500 µL, qui ont été laissés à 23°C ou 4°C, pendant sept temps différents. Ces temps de conservation des laits permettent de créer une cinétique, pour suivre l'évolution des microARNs en fonction du temps. A chaque point de cinétique, les échantillons correspondants sont recongelés afin d'arrêter les temps étudiés. Les temps étudiés sont les suivants : 0 (immédiatement après aliquotage de l'échantillon), 30 minutes, 1h, 2h, 4h, 6h et 18h. L'impact sur le niveau de présence de deux microARNs dans le lait sera alors évalué : le microARN *bta-miR-92a* et *bta-miR-99a*. Un total de 42 échantillons est donc analysé : trois laits, deux conditions de conservation, sept temps par condition.

Ensuite, l'évaluation de la dégradation éventuelle des microARNs est analysée par la méthode de droplet digital PCR (ddPCR), à partir d'un même volume de RT (1 µL) et non pas à partir d'une même concentration. Ce volume de 1 µL de RT a été choisi selon les résultats de concentrations d'ARNs totaux obtenus au Nanodrop, afin de s'assurer d'avoir des concentrations d'ARNs se situant dans la gamme de fonctionnement optimal de la ddPCR pour chacun des échantillons, ce en se basant sur les concentrations en ARNs les plus faibles et les plus fortes détectées parmi les échantillons. De ce fait, il ne doit pas y avoir de phénomène de saturation de la technique par une trop grande quantité du microARN détecté.

La quantification des microARNs *bta-miR-92a* et *bta-miR-99a* est effectuée par ddPCR, avec les amorces décrites ci-avant.

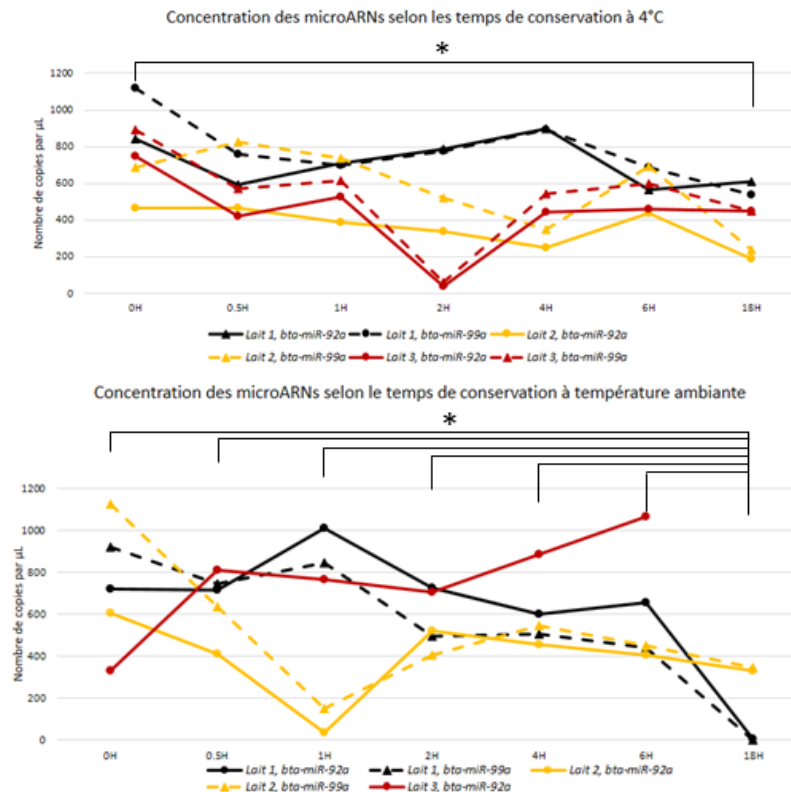


Figure 19 : Concentration des microARNs bta-miR-92a et bta-miR-99a selon le temps de conservation des laits à 4°C et à température ambiante. Les diminutions moyennes significatives sont annotées avec un astérisque ($p < 0,05$, tests statistiques Wilcoxon).

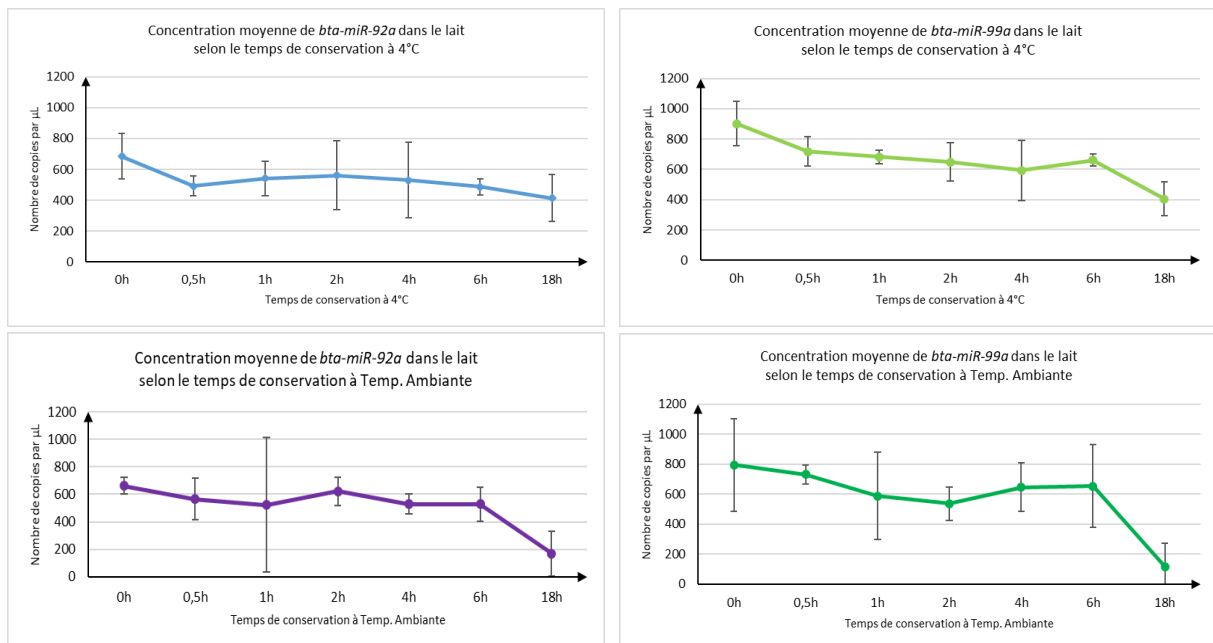


Figure 20 : Concentration des microARNs selon les temps et la méthode de conservation

miR-92a - 4°C	miR-99a - 4°C	miR-92a - TA	miR-99a - TA
42,61%	56,00%	71,90%	84,34%

Tableau 10 : Taux moyen de diminution de la concentration des microARNs miR-92a et miR-99a entre les temps 0h et 18h de conservation à 4°C ou à température ambiante (TA).

La présence relative des microARNs *bta-miR-92a* et *bta-miR-99a* dans des échantillons de lait de 3 animaux différents maintenus à température ambiante, d'une part, et à 4°C, d'autre part, est évaluée au cours du temps pendant 18h (*Figure 4*). Ces microARNs ont été choisis car ils sont souvent utilisés comme gène de référence dans la glande mammaire ou dans le lait et donc ils sont estimés comme ayant une expression stable [298], [311].

Les résultats de chacun des microARNs étudiés selon les temps et mode de conservation sont présentés dans la figure ci-contre (*Figures 19 et 20*).

On observe alors une diminution de la teneur en microARNs lors de la conservation à 4°C au long des 18h et ce dès 30 minutes, mais avec une significativité seulement entre les temps 0h et 18h ($p < 0,05$, tests statistiques Wilcoxon). De même dans le cas d'une conservation à température ambiante, avec dans ce cas une diminution de la concentration qui est significative entre les temps 6h et 18h (et donc entre chacun des temps entre 0h et 6h en comparaison au temps 18h).

Les plus fortes diminutions sont donc visibles entre les temps 0h et 18h, avec des diminutions d'expression allant de 27,50% à plus de 99% (*Tableau 10*).

Au vu de ces résultats, la proposition de l'unité expérimentale d'une conservation des prélèvements des laits pendant 2 heures maximum avant congélation des prélèvements a été validée, et permet l'étude de la présence de microARNs dans le lait.

Les laits des 128 vaches sélectionnées ont été conservés à température ambiante pendant un maximum de 1h41 après prélèvement. D'après les résultats précédents, les bonnes conditions sont réunies pour permettre l'analyse de la présence des microARNs sans craindre un biais lié à leur dégradation.

d. Etude des impacts potentiels des variants *bta-miR-486* et *bta-miR-92b*

i. Matériel et méthodes

1. Échantillons

Des échantillons de lait et de sang ont été prélevés sur un total de 99 vaches de races Holstein et 29 vaches de race Normande, à l'unité expérimentale (UE) INRAE du Pin au Haras. Les échantillons de lait entier sont conservés en tubes de 12 mL et en aliquots de 500 µL à -80°C.

2. Génotypage des vaches

Le génotypage de chacune des 128 vaches pour les deux variants génétiques (les variants impactant les microARNs *bta-miR-92b* et *bta-miR-486*) est réalisé afin de valider les données de génotypages prédites par l'équipe G2B à l'INRAE de Jouy-en-Josas.

Ce génotypage est réalisé à partir du sang par la méthode RFLP. L'étape de d'amplification d'ARN sur le sang est réalisée grâce au kit Terra™ PCR Direct Polymerase Mix. Les amorces 5'-CTGGGCTTCCAGGATAACAA-3' (Sens) et 5'-TGAGGAGGTGATTTCCAAGG-3' (antisens) sont utilisées pour le génotypage du variant associé au microARN *bta-miR-486*. Les fragments amplifiés, de 788 paires de base, sont alors coupés par l'enzyme de restriction Tse I. Les fragments les plus longs sont donc ceux non coupés, présentant l'allèle T. Les amorces 5'-CCCCAGGAAAGATAACAGCA-3' (sens) et 5'-CTTCTCCCTCCCCTTACCAC-3' (antisens) sont utilisées pour le génotypage du variant associé au microARN *bta-miR-92b*. Les fragments amplifiés ont une taille de 562 paires de bases. L'enzyme de restriction Acu I est choisie pour couper ces fragments. Les résultats sur gels permettent de différencier les différents génotypes. Pour le génotypage du variant *bta-miR-92b*, l'enzyme de restriction coupe lors de la présence de l'allèle G. Les fragments les plus longs sont donc ceux non coupés, présentant l'allèle A.

3. Extraction d'ARN

Des ARNs totaux sont utilisés pour l'analyse de chacun des deux variants génétiques. Ils sont extraits à partir de lait entier congelé à -80°C. L'extraction est réalisée grâce au RNA NOW kit (Ozyme), en suivant les instructions d'utilisation du fournisseur, en y ajoutant une étape de précipitation sur la nuit afin d'optimiser la précipitation des ARNs de petite taille (dont les microARNs). Pour 34 échantillons de lait, le kit TRI REAGENT (Ozyme) fut utilisé à la place du RNA NOW du fait de l'arrêt de la commercialisation du précédent, en suivant les instructions du fournisseur et là aussi en y ajoutant une étape de précipitation sur la nuit des petits ARNs. Les ARNs sont repris dans 30 µL d'eau final. La concentration et la pureté des ARNs extraits sont mesurées par spectrophotométrie, avec les données des ratios de longueurs d'onde A260/A280 et A260/A230 grâce à un Nanodrop™, ND-1000.

4. Reverse transcription et PCR quantitative

La présence des deux microARNs dans le lait est analysée par la méthode de droplet digital PCR (ddPCR), pour chaque échantillon d'ARN issu du lait de vaches présentant divers génotypes pour les allèles détectés associés aux microARNs *bta-miR-486* et *bta-miR-92b*. Cette technologie permet une meilleure détection des microARNs en faible présence dans le lait que la qPCR. Leur détection en RT-qPCR a été évaluée au préalable, résultant en des valeurs de Ct trop élevées, ce qui ne permettait pas de faire des analyses d'expression des échantillons. Grâce à des données de miRNomes de lait réalisés au sein de l'équipe, la forme présentant le plus de comptage, donc la plus présente, du microARN *bta-miR-92b* et *bta-miR-486* est identifiée. Il s'agit d'isoformes, de séquences du microARN différant d'un petit nombre de nucléotides, souvent aux extrémités du microARN mature, avec une séquence seed identique. L'isoforme majoritaire du microARN *bta-miR-92b* dans le lait a la séquence suivante : 5'-UAAUUGCACUCGUCCCGCCUCC-3'. Un primer a été défini pour l'amplifier spécifiquement (TaqMan® ID Custom Assay). Dans le cas du microARN *bta-miR-486*, l'isoforme majoritaire dans le lait a pour séquence 5'-UCCUGUACUGAGCUGCCCCGAG-3', ce qui correspond à un primer déjà existant, le TaqMan® ID 001278, Applied Biosystems.

La ddPCR est réalisée selon les instructions des fournisseurs Applied et BioRad. Les étapes de transcription inverse pour la PCR quantitative sont effectuées avec le kit TaqMan® microRNA Reverse Transcription. La solution finale de RT (15 µL) est composée de 7 µL de mix RT, 3 µL des primers spécifiques (5X) et 5 µL d'une solution d'ARN (correspondant à 100 ng d'ARN). La RT est ensuite effectuée selon les recommandations suivantes : 16°C pendant 30 minutes, 42°C pendant 30 minutes et 85°C pendant 5 minutes. La solution finale de ddPCR est composée de 17,67 µL de master mix, 1 µL de TaqMan mRNA assay (20X) et 1,33µL de la RT précédemment effectuée. Les conditions de thermocycleur sont alors les suivantes : 95°C pendant 10 minutes, puis 40 cycles de 95°C pendant 15 secondes et 60°C pendant 1 minute, avant un dernier cycle (95°C pendant 15 secondes et 60°C pendant 1 minute). Les microARNs utilisés pour la normalisation sont *bta-miR-92a* (TaqMan® ID 000431, Applied Biosystems) et *bta-miR-99a* (TaqMan® ID 006254, Applied Biosystems), définis comme gène de référence dans le lait bovin dans les études de Lai *et al.* et Bougarn *et al.* [298], [311].

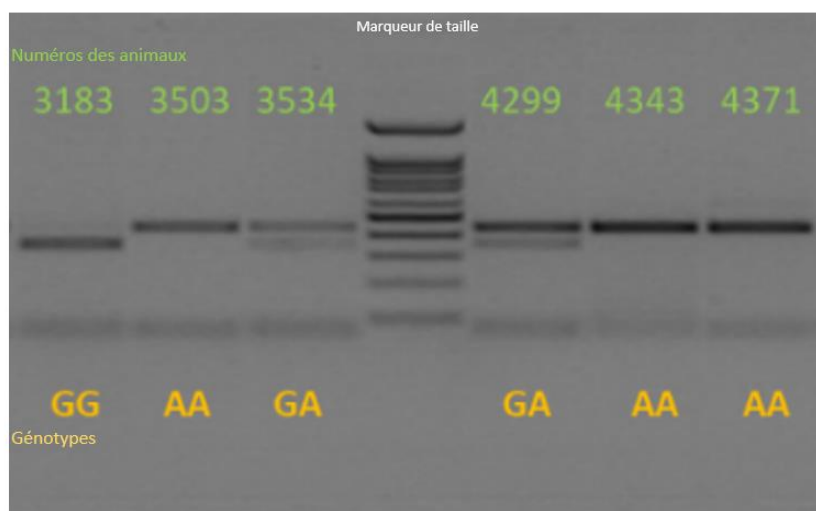


Figure 21 : Gel de typage *bta-miR-92b*

Génotype <i>bta-mir-92b</i>	Nombre d'individus de race Holstein	Nombre d'individus de race Normande	Nombre d'individus total
A/A	16 (16,16%)	9 (31,03%)	25
A/G	57 (57,58%)	12 (41,38%)	69
G/G	26 (26,26%)	8 (27,59%)	34

Tableau 11 : Nombre de vaches par génotypage pour le variant génétique impactant *bta-miR-92b*.

Génotype <i>bta-miR-486</i>	Nombre d'individus de race Holstein	Nombre d'individus de race Normande	Nombre d'individus total
G/G	28 (28,28%)	9 (31,03%)	37
G/T	53 (53,54%)	19 (65,52%)	72
T/T	18 (18,18%)	1 (3,45%)	19

Tableau 12 : Nombre de vaches par génotypage pour le variant génétique impactant *bta-miR-486*.

5. Analyses statistiques

Pour les analyses de présence de microARNs dans le lait selon le génotype, pour les variants des microARNs *bta-miR-486* et *bta-miR-92b*, des analyses de variance ANOVA sont réalisées sous R afin d'observer le potentiel effet d'autres facteurs que le génotype, tels que l'effet de la race ou de la période de prélèvement (et donc ainsi de l'alimentation des vaches) ou encore le nombre de lactations.

ii. Résultats

1. Génotypage des vaches

128 vaches sont génotypées par RFLP pour les deux variants génétiques. En ce qui concerne le variant génétique impactant le microARN *bta-miR-92b*, 19,5% des vaches sont homozygotes A/A, environ 53,9% des vaches sont hétérozygotes et 26,6% sont homozygotes G/G (*Figure 21, Tableau 11*).

Le génotypage des 128 vaches pour le variant impactant le microARN *bta-miR-486* conduit à 28,9% des vaches étant homozygotes G/G, 56,3% sont hétérozygotes G/T et 14,8 % sont homozygotes T/T (*Tableau 12*).

Ces résultats de génotypage peuvent ensuite être comparés aux données de prédiction de génotypes par imputation effectuée par l'équipe G2B. En ce qui concerne le variant impactant *bta-miR-92b*, les génotypes diffèrent pour 15 d'entre eux avec les prédictions effectuées, soit 11,7% des animaux. Deux hétérozygotes AG étaient prédits AA, onze hétérozygotes étaient prédits GG et deux homozygotes AA étaient prédits hétérozygotes. En ce qui concerne le variant impactant *bta-miR-486*, 5 individus, soit 3,9% des animaux présentent un génotypage différent entre l'analyse RFLP et la prédiction : trois hétérozygotes étaient prédits homozygotes TT alors que deux homozygotes TT étaient prédits hétérozygotes.

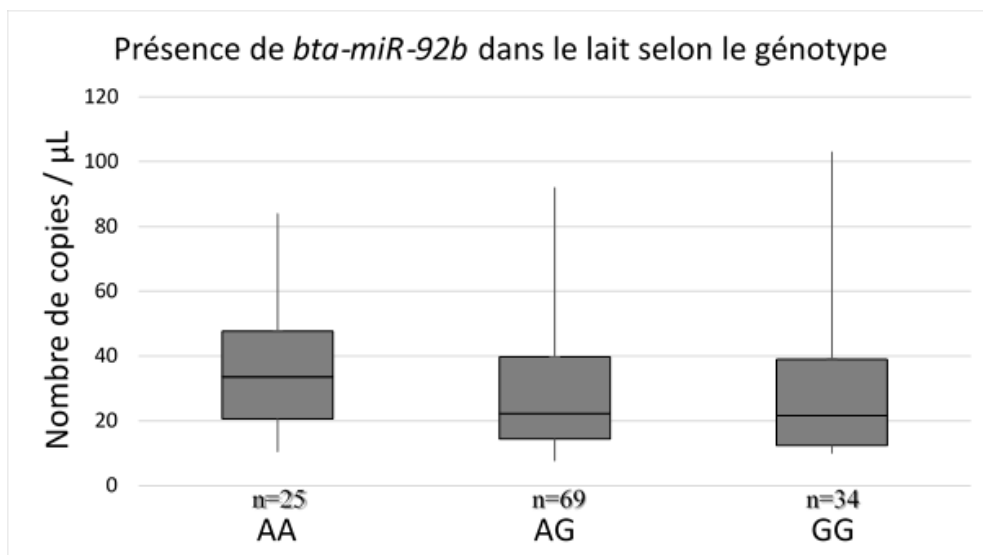


Figure 22 : Présence du microARN *bta-miR-92b* dans le lait de vaches selon leur génotypage pour le variant d'intérêt

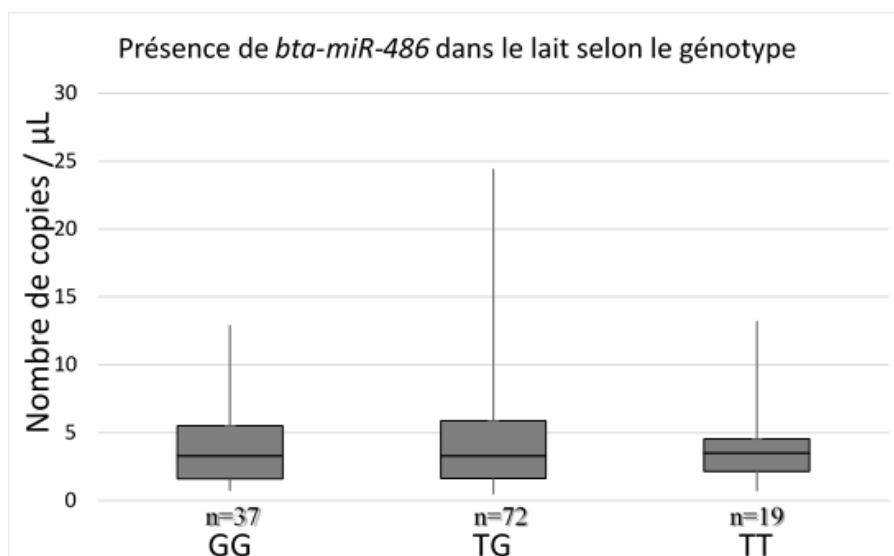


Figure 23 : Présence du microARN *bta-miR-486* dans le lait de vaches selon leur génotypage pour le variant d'intérêt

2. Analyse de la présence des microARNs dans le lait

L'abondance des microARNs *bta-miR-486* et *bta-miR-92b* est évaluée dans le lait de vaches de races Holstein et Normande, présentant les différents génotypes (homozygotes opposés et hétérozygotes, et ce pour chacun des deux variants génétiques).

La présence des deux microARNs étant trop faible dans le lait pour être évaluée par qPCR, avec des Ct supérieurs à 30 pour les microARNs *bta-miR-92b* et *bta-miR-486*, la ddPCR est utilisée.

L'analyse de la présence du microARN *bta-miR-92b* dans le lait ne montre pas de différence significative selon les génotypes au variant (*Figure 22*).

En ce qui concerne le variant génétique situé dans une région flanquante du microARN *bta-miR-486*, comme dans le cas de *bta-miR-92b*, l'analyse de la présence du microARN dans le lait ne montre pas de différence significative selon les génotypes au variant (*Figure 23*).

Des analyses ANOVA avec modèle d'expression permettant de prendre en compte l'effet non seulement des génotypes mais aussi de la race, de l'alimentation et du nombre de lactation des individus ont été réalisées, sans que des différences significatives soient observées.

Discussion et perspectives

L'objectif de cette thèse est de détecter, valider puis évaluer les potentiels impacts fonctionnels de variants génétiques de microARNs exprimés dans la glande mammaire ou présents dans le lait et situés dans des QTL laitiers et mammites. De fait, cela a permis d'effectuer une analyse globale de ces variants de microARNs : de leur détection à leurs effets potentiels. Dans la littérature, les études réalisées se basent principalement soit sur la détection de variants génétiques de microARNs, soit sur l'analyse de leurs impacts.

Dans l'espèce humaine, l'étude de variants génétiques de microARNs s'effectue particulièrement dans le cadre d'études de pathologies d'un point de vue de susceptibilité, soit de hausse ou de baisse du risque d'apparition de la pathologie, ou de pronostic d'espérance de vie (pour exemple : [245], [312]–[314]). Dans ce cadre pathologique, quatre variants génétiques sont régulièrement étudiés : *miR-146a* C>G rs2910164, *miR-149* T>C rs2292832, *miR-196a2* T>C rs11614913 et *miR-499* A>G rs3746444 [315]–[319]. La base de données regroupant les variants génétiques associés aux microARNs (comprenant les variants de microARNs et de régions 3'UTR de gène cible) associés à des pathologies chez l'Humain MSDD a été créée en 2017 et regroupe les données de plus de 2000 publications [230]. Cette base de données liste un total de 525 associations pour 197 SNPs et 164 pathologies au total. Au moins l'un des quatre variants cités ci-dessus est retrouvé dans plus de 38% des associations dans cette base de données (201 sur les 525 recensées). Pour exemple, une pathologie telle que le cancer du sein est associée à trois d'entre eux : les variants de *miR-146a*, *miR-196a2* et *miR-499* [230]. Des publications basées sur le développement de carcinome hépatocellulaire ou sur la susceptibilité au cancer du sein étudient les potentiels impacts de ces quatre variants génétiques [316], [320]. C'est également le cas lors d'études menées sur d'autres cancers, sur la perte récurrente de grossesse idiopathique ou encore sur les risques de maladie coronarienne [245], [315], [317].

Dans les résultats obtenus lors de cette thèse, des variants génétiques bovins ont été retrouvés dans des régions flanquantes du *pré-miR-196a* associés à des caractères liés aux mammites.

1. Détection de variants

En termes de détection de variants de microARNs, Zorc *et al.* ont réalisé en 2012 un catalogue pour six espèces (humain, souris, poulet, chimpanzé, rat et poisson zèbre) de polymorphismes situés dans la région "seed" de microARNs [321]. Par la suite, en 2015, les polymorphismes situés dans les pré-microARNs, les microARNs matures et les régions "seed" ont été ajoutés dans ce catalogue [322]. Les vaches et moutons font partie des 15 espèces étudiées par Zorc *et al.* dans cette nouvelle version du catalogue. Dans ces données, en bovin, 741 des 808 gènes de microARNs étudiés montrent des polymorphismes dans leur pré-microARNs, soit plus de 91%. En ovin, ce pourcentage s'élève à un peu moins de 17% (18 des 106 gènes de microARNs étudiés) [322]. L'outil développé par Zorc *et al.*, miRNA SNIper, permet de lister les variants génétiques de microARNs, mais cet outil n'est plus disponible à ce jour. De plus, il permettait de lister les variants seulement situés dans les pré-microARNs, microARNs matures et les régions "seed". Les variants situés dans les régions flanquantes des précurseurs ne sont pas prises en compte. Les pré-microARNs présentant le plus grand nombre de variants sont *bta-miR-212*, *bta-miR-763* et *bta-miR-877* comprenant 44, 43 et 43 polymorphismes [322]. Durant cette thèse, des variants d'intérêt (variants génétiques de microARNs exprimés dans la glande mammaire ou présents dans le lait) sont détectés en bovin dans ces deux derniers pré-microARNs, avec des variants à des distances de près de 6000 kb d'une région QTL pour *bta-miR-763*, et de 2009 pb d'une région QTL, pour les variants les plus proches, pour *bta-miR-877*, mais n'ont pas été validés par les études de GWAS.

Dans notre cas, un même pipeline bio-informatique a été réalisé pour les trois espèces étudiées : bovins, caprins et ovins. Les résultats sont donc comparables du point de vue de la méthode. Les différences proviennent des données des fichiers de départ.

Une hypothèse a été émise par Siepel *et al.*, selon laquelle la présence d'éléments conservés pourrait être due au fait qu'il s'agit de nœuds dans un réseau de régulation global : en effet, dans notre cas du fait du grand nombre d'interactions des microARNs avec des ARN messagers, une modification d'un de ces microARNs, d'un de ces nœuds, entraînerait des modifications nombreuses et sur différents ARNm donc potentiellement différents processus biologiques [323].

De fait, les microARNs sont des éléments fortement conservés entre espèces. Plus généralement, la régulation de gène est un système dans lequel de nombreuses régions conservées sont impliquées [324]. Une étude comparée de variants de microARNs dans les trois espèces bovine, caprine et ovine est donc possible.

Les résultats obtenus montrent que, pour quelques microARNs, un nombre important de variants est associé à un même microARN. Ce nombre de polymorphismes peut être un signe de moins forte conservation de la région du génome impliquée. Les variants détectés sont parfois situés dans les régions flanquantes, jusqu'à un kilobase en amont ou en aval du précurseur de microARN dans notre cas, et peuvent donc ne pas avoir d'impact crucial. De fait, la conservation de certaines de ces régions flanquantes peut être moins importante.

Le fait d'avoir étudié en parallèle trois espèces de ruminants laitiers a permis de mettre en avant 20 microARNs qui sont impactés par des variants détectés dans les trois espèces étudiées : à la fois en bovin, caprin et ovin. Dans l'espèce bovine, 379 des 511 microARNs impactés par au moins un variant génétique détecté par le script global, sans la prise en compte des données de QTL, sont retrouvés seulement dans cette espèce. Dans l'espèce caprine, seuls 5 microARNs sur les 100 impactés par des variants génétiques détectés ne sont pas retrouvés dans une autre espèce (bovine ou ovine). Ce nombre de microARNs impactés seulement en espèce ovine s'élève à 23, sur un total de 106 microARNs. La présence de microARNs présentant des variants d'intérêt dans plusieurs espèces peut être une indication sur leur conservation et donc leur importance.

Une perspective possible de cette étude est d'étendre ces analyses à la détection de variants d'intérêt d'autres ARNs non codants, à la condition qu'ils soient correctement annotés dans les fichiers VCF de départ, ou en tous cas que leurs positions sur le génome soient connues.

2. Validation de variants

L'étape de validation de variants génétiques détectés s'effectue grâce à des données de GWAS et de puces à ADN, obtenues sur un grand nombre d'animaux. De plus, les études de GWAS sont réalisées sur une population indépendante de celle utilisée pour les études de détection de QTL. Cette étape de validation grâce aux études de GWAS a permis de valider trois SNPs en bovin pour lesquels les liens phénotype-génotype est alors établi.

En amont de cette étape de validation, des arbres de décisions avaient été réfléchis en se basant sur des critères de localisation du variant, bibliographie ou encore de présence dans le lait, dans l'hypothèse qu'une priorisation des variants à étudier ait été nécessaire. Dans le cas de validation de variants, il faut prendre en considération le fait que les effets de certains variants préalablement détectés en bovin peuvent induire un biais dans le potentiel impact de variants génétiques de microARNs situés dans des régions proches, ou plus généralement étant en déséquilibre de liaison avec ces variants [348]. L'exemple du chromosome 14 bovin peut être cité : il est en effet probable que des variants génétiques détectés sur ce chromosome soient en réalité en déséquilibre de liaison avec le gène *DGAT1*, dont on connaît l'impact fort sur des caractères laitiers tels que la teneur en matière grasse du lait [349], [350]. Il s'agit de variants détectés situés à une distance de 228kb à 268kb du gène *DGAT1*, avec des positions, fréquences de l'allèle minoritaire et $-\log P$ proches et qui pourraient être donc en déséquilibre de liaison entre eux. Une étude de 2004 fait également l'hypothèse d'un effet fort du gène *DGAT1* dans cette région, pouvant éventuellement expliquer une non-détection de nouveaux QTL bovins sur ce chromosome pour des caractères protéiques [351]. Le nombre de trois variants validés a permis d'étudier les effets potentiels de chacun d'eux, sans avoir à ajouter une étape de choix de variants. Les données de puces à SNPs sur lesquelles des variants détectés bovins et caprins ont été ajoutés seront disponibles par la suite, lorsqu'un grand nombre d'animaux seront génotypés à l'aide de ces puces.

Ces données pourront éventuellement permettre de valider d'autres variants détectés. Pour le moment, 1200 caprins ont été génotypés à l'aide de la puce à SNPs depuis sa disponibilité en septembre 2020, alors que la version de la puce bovine EURO10K, actualisée tous les 6 à 12 mois, a été mise en production en septembre 2020.

3. Tests des potentiels impacts de variants validés

a. *bta-let-7e*

Après la détection et validation de variants génétiques, leurs potentiels impacts sont étudiés. Les effets du SNP situé dans la région "seed" du *bta-let-7e* sont analysés en utilisant un modèle de lignée de cellules épithéliales bovines mammaires : BME-UV1.

i. BME-UV1

Les lignées cellulaires, ici les cellules épithéliales mammaires bovines BME-UV1, sont utilisées pour les analyses *in vitro*. Elles sont utilisées comme modèle d'épithélium mammaire bovin [325]. Ce modèle BME-UV1 est validé pour l'étude *in vitro* de ciblage d'ARNm par des microARNs. Pour exemple, une surexpression du microARN *miR-1* a abouti à une réduction significative de l'expression d'un de ses gènes cibles, *ELOVL6* en BME-UV1 [326]. Ce matériel permet une représentation d'un composant de la glande mammaire que sont les cellules épithéliales mammaires, responsables dans la glande mammaire de la synthèse et de la sécrétion des constituants du lait. Cette lignée BME-UV1 est une lignée qui permet également d'étudier la prolifération [327]. Le transcriptome de BME-UV1 a été réalisé dans le laboratoire ce qui permet de s'assurer de l'expression par cette lignée d'ARNm ciblés par le microARN *bta-let-7e* présentant l'allèle G ou l'allèle A, et ainsi l'observation des impacts de microARNs sur l'expression de ces ARNs messagers sera possible. Des hypothèses sur les potentiels impacts de variants génétiques de microARNs sur les ARNm pourront donc être émises ; et à plus grande échelle sur la synthèse et composition du lait, au vue des transcrits visiblement impactés en culture cellulaire. L'utilisation de ce modèle cellulaire *in vitro*, au contraire des études *in vivo*, permet de s'affranchir des variations potentielles dues à l'animal et à son environnement.

ii. Effets détectés via l'approche RT-qPCR

Pour les analyses fonctionnelles réalisées par RT-qPCR ou ddPCR, les gènes de référence utilisés ont été systématiquement choisis selon leur utilisation dans la littérature ou dans des données issues du laboratoire. Dans le cas de l'analyse de l'impact potentiel du variant de la région "seed" de *bta-let-7e*, *PPIA*, qui est un gène pouvant être impliqué dans le phénomène d'apoptose, est étudié comme gène de référence, car il est utilisé dans des analyses d'expression dans des cellules épithéliales mammaires bovines [298]. Les effets potentiels attendus en culture cellulaire sont un différentiel d'expression de gènes ciblés par *bta-let-7e* selon la transfection des cellules par le microARN présentant l'un ou l'autre des allèles. Les gènes *OSTF1* et *SLC25A24* sont des cibles potentielles de *bta-let-7e* qui ont été étudiés. Le choix de ces ARNm ne s'est pas effectué selon leur importance dans un contexte laitier, mais sur des critères plus techniques comme leurs niveaux d'expression dans les BME-UV1. Néanmoins, quelques éléments de la littérature pourraient permettre de lier ces gènes à des caractères laitiers et de matières grasses (qui est le caractère de QTL dans lequel le variant du microARN est situé). Le gène *OSTF1* est défini par Schwerin *et al.* comme un gène potentiellement impliqué dans la résistance aux mammites chez les ruminants [328], [329]. *SLC25A24* est associé à l'indice de masse grasse, avec des mutations étudiées dans ce gène : les pourcentages de matières grasses du corps sont plus faibles chez des femelles présentant les allèles majeurs [330]. De plus, des souris KO pour ce gène voient leur poids corporel diminué, ainsi que leur poids de tissus adipeux [330]. *SLC25A24* serait donc lié à un métabolisme des matières grasses moins important.

De ce fait, des variants génétiques permettant une plus forte expression de *OSTF1* et une moindre expression de *SLC25A24* pourraient être à privilégier : dans notre cas, des vaches présentant l'allèle G seraient alors à prioriser.

L'impact direct d'un variant génétique dans la région "seed" d'un microARN sur l'expression d'ARNm cibles dans des cellules épithéliales mammaires est donc montré en RT-qPCR, validant les prédictions bioinformatiques. Le changement d'un seul nucléotide pourrait donc engendrer un changement de cible de microARN et donc une modification de niveaux d'expression d'ARN messager, et donc éventuellement un changement physiologique.

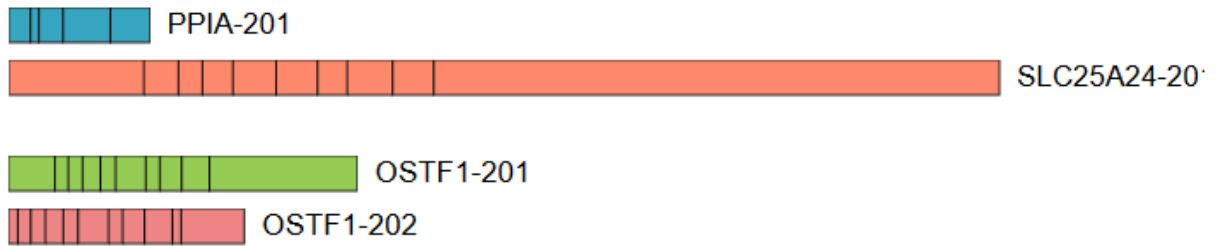


Figure 24 : Visualisation des transcrits gènes étudiés dans le cadre des impacts du variant bta-let-7e

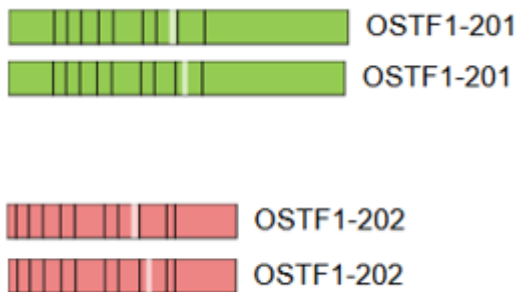


Figure 25 : Visualisation de la position des amorces, dessinées pour étudier l'expression d'OSTF1, dans chacun des transcrits décrits du gène (OSTF1-201 en vert et OSTF1-202 en rose).

Les rectangles blancs représentent les amorces dessinées.

iii. Effets détectés via l'approche RNAseq

L'analyse des données obtenues par RNAseq a été effectuée par la suite afin d'analyser de façon plus globale l'expression des transcrits selon l'allèle présent dans le microARN transfecté en BME-UV1. Aucun transcrit ne montre d'expression différentielle significative entre les conditions. Les résultats obtenus sont donc non concordants avec les analyses en RT-qPCR, et donc avec les prédictions *in silico* pour les gènes *OSTF1* et *SLC25A24*.

iv. Comparaison RT-qPCR et RNAseq

La RT-qPCR est une technique avec une forte sensibilité et spécificité permettant une approche ciblée de gènes, alors que le RNAseq est une approche globale, du transcriptome total [331], [332]. Une perte de précision en RNAseq est possible car il s'agit d'une approche globale, moins précise que la RT-qPCR. Une étude a comparé des données de RT-qPCR et de RNAseq, avec pour résultat un différentiel d'expression concordant entre les deux méthodes pour 19 des 34 gènes sélectionnés [333]. La première hypothèse formulée pour expliquer la non concordance des résultats se base sur le phénomène d'épissage alternatif et donc de transcrits. En effet, les résultats de qPCR obtenus peuvent être discutés du point de vue de la technique en elle-même. Tout d'abord il s'agit d'une technique reposant sur des amorces dessinées pour cibler une région précise du génome. Néanmoins, des amorces différentes pour un même gène peuvent être choisies et ne pas donner de résultats semblables. En effet, ces amorces sont dessinées afin de s'affranchir de l'amplification d'ADN résiduel. Mais, selon leur position sur le gène, cela peut entraîner des expressions détectées différentes en fonction de phénomène d'épissage alternatif. En effet, un transcrit peut être plus exprimé qu'un autre pour un même gène dans nos cellules. Et selon la position des amorces, seul l'un ou l'autre des transcrits peut être ciblé, et donc son expression être mesurée. Dans le cas de notre étude, les différents gènes sont ajoutés dans TraC: Transcript Consensus, un outil permettant de visualiser les différents transcrits d'un gène [334]. Le gène de référence *PPIA* ainsi que le gène étudié *SLC25A24* ne possèdent qu'un seul transcrit recensé (par Ensembl) (*Figure 24*). De fait, les amorces dessinées permettent d'étudier la totalité de l'expression du gène. En ce qui concerne le gène étudié *OSTF1*, deux transcrits sont recensés dans Ensembl (*Figure 24*).

		Région seed																				
<i>bta-let-7a</i>	U	G	A	G	G	U	A	G	U	A	G	G	U	U	G	U	A	U	A	G	U	U
<i>bta-let-7b</i>	U	G	A	G	G	U	A	G	U	A	G	G	U	U	G	U	G	U	G	G	U	U
<i>bta-let-7c</i>	U	G	A	G	G	U	A	G	U	A	G	G	U	U	G	U	A	U	G	G	U	U
<i>bta-let-7d</i>	A	G	A	G	G	U	A	G	U	A	G	G	U	U	G	C	A	U	A	G	U	U
<i>bta-let-7e</i>	U	G	A	G	G	U	A	G	G	A	G	G	U	U	G	U	A	U	A	G	U	_
<i>bta-let-7f</i>	U	G	A	G	G	U	A	G	U	A	G	A	U	U	G	U	A	U	A	G	U	U
<i>bta-let-7g</i>	U	G	A	G	G	U	A	G	U	A	G	U	U	U	G	U	A	C	A	G	U	U
<i>bta-let-7i</i>	U	G	A	G	G	U	A	G	U	A	G	U	U	U	G	U	G	C	U	G	U	U
<i>bta-miR-98</i>	U	G	A	G	G	U	A	G	U	A	A	G	U	U	G	U	A	U	U	G	U	U

Tableau 13 : Séquences nucléotidiques des microARNs de la famille des microARNs bovins let-7.

Les amorces dessinées sont situées dans une région commune à ces deux transcrits, et permettent donc là aussi d'étudier la totalité de l'expression du gène (*Figure 25*). Le séquençage RNAseq prend en compte l'ensemble des transcrits pour un même gène. L'amplification d'une forme non majoritairement exprimée dans les cellules en RT-qPCR aurait donc expliqué cette non concordance entre résultats de RT-qPCR et résultats de RNAseq, mais ce n'est pas le cas ici. Le nombre d'échantillons analysés en RNAseq pourrait également expliquer ces discordances. En effet, un nombre de 12 échantillons par condition est préconisé par Schurch *et al.* dans le cas d'étude visant à identifier les gènes différentiellement exprimés selon les conditions [335]. Dans notre cas, 3 échantillons par condition ont été analysés. Les résultats de RNAseq obtenus montrent que le simple fait de transfecter les cellules avec le microARN *bta-let-7e* entraîne un changement de transcriptome, une modification d'expression de transcrits, mais que cette approche globale ne permet pas de mettre en évidence des différentiels d'expression de transcrits selon la présence de l'allèle G ou de l'allèle A du variant de *bta-let-7e*.

v. *Le microARN bta-let-7e*

Les problématiques de ciblage d'ARNs messagers sont particulièrement présentes pour les microARNs faisant partie de la famille des *let-7*, dont le microARN étudié ici : *bta-let-7e*. En effet, cette famille de microARNs présentant des régions "seed" identiques comporte neuf membres : *bta-let-7a*, *bta-let-7b*, *bta-let-7c*, *bta-let-7d*, *bta-let-7e*, *bta-let-7f*, *bta-let-7g*, *bta-let-7i* et *bta-miR-98* (*Tableau 13*).

Par ailleurs, des publications portent parfois sur l'impact de cette famille de microARNs dans son entièreté et non pas sur un microARN en particulier [336].

Plus précisément, du fait de la courte séquence de cette région d'appariement, peut se poser le problème des séquences de "seed" proches et des systèmes de compensation : un microARN peut cibler plusieurs ARNm et un ARNm peut être ciblé par plusieurs microARNs.

En effet, un variant génétique situé dans la région "seed" peut engendrer le ciblage de nouveaux ARNm, mais qui étaient possiblement déjà ciblés par d'autres microARNs. Son action peut donc être masquée par l'effet des autres microARNs ayant ces mêmes cibles. De même pour le cas de non ciblage d'un ARNm : si d'autres microARNs ciblent cet ARNm, la non-régulation de ce microARN en particulier n'aura qu'un effet limité.

Des phénomènes de compensation sont donc possibles entre chacun de ces microARNs : des relais peuvent être pris lorsque l'un des microARNs de la famille a un variant et ne peut donc plus cibler certains ARN messagers, une hypothèse notamment émise par Gurtan *et al.* [337].

De plus, on peut supposer qu'un variant génétique, un changement d'un seul nucléotide, n'entraînera pas un changement de cible : l'appariement ne sera simplement plus total, ce qui engendrera non pas une dégradation de l'ARN messager cible comme initialement mais une répression de son expression [236]–[238]. De plus Bartel *et al.* ont montré qu'un appariement de la partie 3' du microARN avec l'ARNm cible peut compenser un SNP dans la région "seed" [338].

Du point de vue technique, le mimic du *bta-let-7e* avec l'allèle A (en position 8 du microARN mature, le dernier nucléotide de la région "seed") a été réalisé à façon afin d'obtenir une différence entre les deux mimics.

vi. Perspectives

Les effets potentiels du variant situé dans la région "seed" de *bta-let-7e* ont été étudiés en culture cellulaire, validant les prédictions bio-informatiques de ciblage d'ARNm. En BME-UV1, des microARNs *bta-let-7e* sont présents. L'ajout de mimics de microARNs permet donc de compléter la présence de ces microARNs initialement présents. L'utilisation de mimics permet une expression transitoire, dont l'effet peut être dilué lors des divisions cellulaires. Afin de mieux analyser les effets fonctionnels du variant de *bta-let-7e* dans un contexte spécifique, l'expression des autres microARNs appartenant à la même famille de microARNs pourrait être significativement réduite.

En pratique, l'expression des microARNs de la famille *let-7* à l'exception du *bta-let-7e* serait réduite. En effet, Chang *et al.* ont montré une diminution de l'expression de microARNs matures (les microARNs *miR-200c*, *miR-7* et *miR-141*) ciblés par une technique d'édition de génome de 96% dans des cellules, avec également des incidences sur l'expression d'ARNm cibles de ces microARNs [339]. Une étude publiée en 2020 a utilisé une technique d'édition du génome afin de réduire l'expression du *miR-29b*. Cela a permis non seulement une réduction significative de l'expression de *miR-29b*, mais aussi une absence d'incidence sur les *miR-29a* et *miR-29c* qui font partie de la même famille *miR-29* [340]. Dans notre cas, cela permettrait de tester d'un point de vue fondamental les impacts fonctionnels du variant, dans un contexte écartant toute possibilité de compensations.

Cette hypothèse se limite au fait que les cellules utilisées doivent survivre sans la présence de ces microARNs de la famille *let-7*. Un modèle cellulaire ne présentant pas de *bta-let-7e* pourrait également être utilisé. Pour valider le ciblage des ARNm cibles *OSTF1* et *SLC25A24* par le microARN *bta-let-7e*, il serait possible de transfecter dans les BME-UV1 la séquence 3'-UTR de l'ARNm à tester associée à un gène rapporteur luciférase. L'activité de la luciférase pourrait alors être évaluée selon les concentrations du *bta-let-7e* co-transfectées avec le vecteur rapporteur luciférase. De même, la séquence 3'-UTR mutée associée au gène rapporteur luciférase peut également être transfectée. Une répression de l'activité luciférase de façon dose-dépendante, dans le premier cas, ou une répression de l'activité seulement lorsque la région 3'-UTR n'est pas mutée, dans le second cas, permettrait de valider le ciblage de l'ARNm par le microARN. C'est une approche qui a été utilisée notamment par Jiang *et al.* en utilisant des cellules MAC-T (bovine mammary alveolar cells), pour étudier les effets d'un variant de *bta-miR-2899* qui est associé aux mammites en bovin [341].

b. Conservation des microARNs dans le lait

La conservation des microARNs dans le lait a été étudiée dans le cadre de cette thèse, afin de s'assurer d'une bonne méthode d'échantillonnage des laits. Howard *et al.* ont étudié l'impact de la pasteurisation et du temps (en jours) de stockage à 4°C du lait bovin pasteurisé sur la perte de microARNs [342]. Les concentrations du *miR-29b* et du *miR-200c* dans le lait entier pasteurisé ne sont pas affectées par la conservation à 4°C [342]. Dans notre cas, les concentrations des microARNs dans le lait entier sont significativement diminuées entre 6h et 18h. Les prélèvements effectués à l'Unité Expérimentale du Pin ont permis une conservation à température ambiante pendant un temps maximal de 1h41 avant une congélation, les présences des microARNs dans le lait pourront donc être étudiées en s'affranchissant d'un possible biais dû à la conservation.

c. *bta-miR-486* et *bta-miR-92b*

i. Échantillons

Lors de l'étude des effets potentiels de variants de *bta-miR-486* et *bta-miR-92b*, des laits de 128 vaches de deux races différentes sont utilisés. Le choix d'utiliser le lait se base sur l'observation qu'il existe une forte corrélation entre les microARNs exprimés dans la glande mammaire et leur présence dans le lait [293]. Cela permet d'éviter des procédures invasives que sont les biopsies mammaires sur des vaches. De plus, un nombre important d'échantillons peut alors être obtenu : ici 128 échantillons. Les effets du facteur race (vaches de races Holstein et Normande) et de l'alimentation (selon la période de prélèvement dans l'année) sont également pris en compte dans les analyses statistiques des données. Tous ces échantillons proviennent d'une seule et même unité expérimentale (Unité Expérimentale du Pin), permettant une homogénéité des prélèvements et des conditions d'élevage. Une perspective de ce travail pourrait être l'extension de ces tests d'impacts sur des échantillons de lait de vaches de race Montbéliarde, une autre race bovine laitière dans laquelle le variant génétique du *bta-miR-486* a également été détecté et donc dans laquelle une potentielle modification de l'expression du microARN selon le génotypage de l'animal pourrait être retrouvée.

ii. Techniques utilisées pour tester les effets potentiels

Les effets des variants de *bta-miR-486* et *bta-miR-92b* sur leur expression sont étudiés par ddPCR dans le lait, permettant un comptage brut. Dans la littérature, les polymorphismes de microARNs peuvent être étudiés en culture cellulaire avec transfection, puis analyse de l'expression du microARN ou d'ARNm cibles par qPCR. C'est le cas par exemple de l'analyse du *pre-miR-15b*, ou du *miR-215* en porcin ou encore du *pri-miR-26a* chez les poulets [343]–[345]. Dans notre cas, nous nous basons sur le même principe d'analyse de l'expression, mais en utilisant des laits d'animaux dont nous connaissons les génotypes.

La technique de ddPCR est utilisée. Ce système de PCR quantitative est privilégié dans le cas de détection faible du microARN en RT-qPCR, avec des Ct trop élevés pour pouvoir conclure sur des différences de concentrations.

Des biais peuvent apparaître en amont de l'étape de quantification de ces microARNs, dans l'étape d'extraction des ARNs. Des tests préalables ont été effectués afin d'observer la conservation des microARNs dans le lait selon différents temps et modes de décongélation (à 4°C ou à température ambiante). Les résultats nous ont permis d'indiquer un prélèvement de lait se faisant au maximum deux heures après la traite des vaches et ainsi limiter des dégradations potentielles de microARNs. Toutefois, l'étape d'extraction d'ARN totaux du lait peut aboutir à des résultats variables : dans notre cas, les concentrations d'ARN obtenu s'étendent de 30 à 300 ng/μL pour un même fluide, ici le lait selon les animaux. De telles variations s'observent avec le RNA Now tout comme avec le Tri REAGENT.

Cette technique de ddPCR a l'avantage d'aboutir à des résultats absolus, des concentrations, et non relatifs, ce qui peut ôter le biais amené par les gènes de référence. Toutefois, des microARNs de référence sont utilisés afin de vérifier la stabilité de leur présence selon les échantillons, pour différents calculs pouvant être faits par la suite, à partir des données brutes.

Les microARNs *bta-miR-92a* et *bta-miR-99a* sont choisis comme microARN de référence. En effet, ils sont décrits dans la bibliographie comme microARN de référence et montrent peu de variations individuelles dans des données de RT-qPCR (Leduc *et al.* données non publiées, [311], [346]). Les microARNs *miR-92a*, utilisé comme gène de référence, et *miR-92b*, dont

l'expression est étudiée dans cette thèse, font partie de la même famille de microARNs, mais ont des séquences légèrement différentes : dans le microARN mature, trois nucléotides diffèrent, dont les deux nucléotides situés à l'extrémité 3' du microARN. La brochure de ThermoFisher décrivant les kits utilisés indique que les TaqMan microRNA assays sont spécifiques des microARNs matures, et qu'ils peuvent permettre de distinguer des cibles très homologues, dont des membres d'une même famille de microARN [347]. Le microARN *miR-92a* a donc pu être utilisé en tant que microARN de référence dans notre étude.

Les variants des régions flanquantes de *bta-miR-486* et *bta-miR-92b* n'ont pas d'effet significatif sur l'expression de microARNs matures dans le lait. Du point de vue appliqué, la présence de l'un ou l'autre des allèles à ces positions n'impacte pas la présence du microARN dans le lait et donc n'a pas d'impact fonctionnel dans le lait *via* le microARN mature.

Le fait que les variants associés aux microARNs *bta-miR-486* et *bta-miR-92b* n'aient pas d'impact significatif sur l'expression des microARNs associés peut s'expliquer par le fait que ces variants ne sont peut-être pas situés dans une région régulatrice, et n'ont donc pas d'impact sur la biogénèse des microARNs et donc *a fortiori* sur leur expression. De plus, un changement d'un seul nucléotide pourrait ne pas entraîner de modification dans la prise en charge du pri- ou du pré-microARN par les différents éléments tels que DICER ou Drosha, qui permettent la biogénèse du microARN, et donc pas de modification de l'expression du microARN.

iii. Perspectives

Afin d'évaluer si les variants ont des impacts plus en amont dans le processus de biogénèse de microARNs, il serait possible d'évaluer l'expression non pas des microARNs matures mais des précurseurs. Même si l'intérêt physiologique serait limité, puisque les précurseurs seraient impactés mais pas les matures, donc pas d'impact sur les ARNm cibles, cela permettrait d'étudier de façon fondamentale l'impact de variants génétiques situés en amont ou en aval de précurseurs de microARNs.

Des biopsies mammaires pourraient également être utilisées afin de tester les effets de ces deux variants dans un contexte *in vivo* : sur des échantillons de glandes mammaires bovines. Mais cela impliquerait un protocole lourd et invasif, qui nécessite de plus la rédaction et validation d'une saisine, et donc un nombre d'échantillons potentiellement moins important que pour le lait.

d. Perspectives pour le test de potentiels impacts de variants caprins et ovins

En caprin, le script bioinformatique a résulté en la détection de 127 variants génétiques. Chacun d'eux est situé dans le microARN *chi-miR-22*. Ce microARN est décrit par Yu *et al.* comme étant surexprimé au jour de lactation 30 en comparaison au jour de lactation 5, dans des tissus de glandes mammaires de chèvres de race Laoshan [352]. Il pourrait donc être possible de tester les impacts potentiels de certains de ces variants, par étude de la présence du microARN dans le lait selon le génotype des chèvres. Une réflexion devra alors être menée afin de sélectionner des variants parmi les 127 détectés, pour les analyser. De même en ovin, une étape de validation et de choix de variants pourra permettre d'étudier l'impact potentiel de certains d'entre eux.

4. Perspectives générales

L'étude de variants génétiques de microARNs dans un contexte de production laitière a été réalisée sur les 3 espèces bovine, caprine et ovine. Le fait de travailler sur ces 3 espèces a permis de pointer la différence d'abondance de données disponibles selon l'espèce. Les données de miRNomes bovins et caprins disponibles ont pu être utilisés afin de pallier l'absence de ces données en ovine et donc d'affiner les résultats obtenus en ovins. La similitude des résultats obtenus en bovins, caprins et ovins est toutefois limitée. Pour exemple, les variants caprins détectés lors de l'étude prenant en compte les données de QTL laitiers impactent tous le microARN *chi-miR-22*. Ce microARN *miR-22* n'est impacté par aucun variant génétique détecté en espèce bovine. Aucun variant génétique n'est détecté en espèce ovine lors de la prise en compte des données QTL, ce qui empêche là aussi la comparaison entre espèce. La détection de variants génétiques de microARNs d'intérêt dans un contexte de production laitière peut donc être réalisée de façon individuelle pour chaque espèce, néanmoins l'apport des données d'une espèce (notamment bovines) peut permettre d'affiner les résultats d'une autre.

La base de données RumimiR a permis de regrouper dans une seule et même base l'ensemble des microARNs bovins, caprins et ovins décrits dans la littérature. L'homogénéisation des données est réalisée afin de pouvoir avoir des informations similaires pour chacune des espèces. L'ensembles des données de la bibliographie sont donc ajoutées, sans distinction de critères utilisés pour la détection de microARNs. C'est pourquoi un code de faux-positivité a été ajouté à titre informatif dans RumimiR, afin d'indiquer pour chaque microARN si sa présence est effective ou s'il pourrait s'agir en réalité d'un faux positif. La mise à disposition de davantage de données sur ces trois espèces, en les homogénéisant, permettrait d'obtenir des données similaires en bovin, caprin et ovine. Les différences, notamment en termes de volume de données obtenues, dépendraient alors principalement des études et des connaissances de chacune de ces espèces.

La détection de variants génétiques d'intérêt dans un contexte de production laitière a permis, dans un contexte de sélection génomique, l'ajout de 44 variants bovins sur la puce à SNPs bovine et 34 variants sur la puce caprine. L'ajout de ces informations sur ces outils existants en sélection génomique permettrait alors d'affiner la sélection d'animaux en prenant en compte cette information supplémentaire : leur génotype pour des variants d'intérêt. Même si l'impact de ces variants détectés sur l'expression du microARN associé ou sur celle d'ARNm ciblés par celui-ci n'est pas caractérisé, leur lien avec des caractères laitiers peut être établi si les liens phénotypes/génotypes sont validés. Cette validation pourrait s'effectuer grâce aux données de génotypages obtenues par les puces à SNPs, en les mettant en relation avec les phénotypes des animaux pour les caractères laitiers voulus. Des animaux seront génotypés pour ces variants et ces génotypes pourraient alors être pris en compte pour la sélection d'animaux. En perspective, de nouveaux variants génétiques d'intérêt (variants génétiques de microARNs exprimés dans la glande mammaire ou présents dans le lait, éventuellement localisés dans des régions de QTL laitiers ou mammites) pourraient également être validés et leur potentiel impact pourrait alors être testé. L'impact du variant de la région "seed" de *bta-let-7e* sur d'autres ARNm cibles pourrait également être testé, en culture cellulaire.

Les interactions entre des microARNs et leurs cibles sont appelées MTIs (pour microRNA-Target Interactions). Pour exemple, Pasquini *et al.* se sont basés sur les MTIs dans cinq espèces : bovin, porc, ovin, canin et aviaire, en utilisant les données de plus de 70 articles [353]. Ils ont alors détecté en bovin un total de 265 MTIs, pour 34 microARNs et 123 gènes cibles. En ovin, 3 MTIs ont été retrouvés [353]. L'étude de variants génétiques situés dans ces régulomes a été effectuée dans le cadre de cette thèse en se focalisant sur les variants impactant les microARNs, une perspective de ce travail pourrait être d'étudier les variants de régulomes situés dans les régions d'ARNm ciblées par des microARNs. Cela aurait pour but d'élargir l'étude de variants de microARNs à l'étude plus complète de variants de régulome.

Des études ont mis en évidence de potentiels rôles d'importance d'ARNs long non codants (lncRNAs) dans les processus de lactation et de composition du lait [354], [355]. La perspective d'étendre les analyses effectuées à la détection de variants d'intérêt à ces lncRNAs ou à d'autres ARNs non codants est également possible.

Bibliographie

- [1] Agreste, « Graph'agri 2019 ». 2019, Consulté le: mars 16, 2020. [En ligne]. Disponible sur: <https://agreste.agriculture.gouv.fr/agreste-web/download/publication/publie/GraFraChap4.1/Graf1913%20-%20Population%20active.pdf>.
- [2] Eurostat, « Eurostat - Profil », 2019. <https://ec.europa.eu/eurostat/guip/introAction.do?> (consulté le mars 17, 2020).
- [3] FAO, Food and Agriculture Organization of the United Nations, « FAO Country Profiles:France », *Food and Agriculture Organization of the United Nations*. <http://www.fao.org/countryprofiles/index/en/?iso3=FRA> (consulté le mars 17, 2020).
- [4] Insee, « Exploitations agricoles – Tableaux de l'économie française | Insee ». 2019, Consulté le: mars 16, 2020. [En ligne]. Disponible sur: <https://www.insee.fr/fr/statistiques/3676823?sommaire=3696937>.
- [5] FAO, Food and Agriculture Organization of the United Nations, « La production laitière et les produits laitiers: Les animaux laitiers ». <http://www.fao.org/dairy-production-products/production/dairy-animals/fr/> (consulté le mars 24, 2020).
- [6] G. Pulina *et al.*, « Invited review: Current production trends, farm structures, and economics of the dairy sheep and goat sectors », *J. Dairy Sci.*, vol. 101, n° 8, p. 6715-6729, août 2018, doi: 10.3168/jds.2017-14015.
- [7] « Texas & Southwestern Cattle Raisers | Cow-Calf Corner: Dairy influence on beef markets; Follow BQA guidelines when treating and selling cows ». <https://tscra.org/cow-calf-corner-dairy-influence-on-beef-markets-follow-bqa-guidelines-when-treating-and-selling-cows/> (consulté le déc. 03, 2020).
- [8] « Impacts of the Dairy Industry on Beef Markets | Ohio BEEF Cattle Letter ». <https://u.osu.edu/beef/2016/07/20/impacts-of-the-dairy-industry-on-beef-markets/> (consulté le déc. 03, 2020).
- [9] Idele, (Institut de l'Élevage) et CNE, Confédération Nationale de l'Élevage), « Chiffres-clés Bovins 2019 ». 2019, Consulté le: mars 20, 2020. [En ligne]. Disponible sur: <http://idele.fr/filieres/bovin-viande/publication/idelesolr/recommends/chiffres-cles-bovins-2019.html>.
- [10] CNIEL, « Economie laitière en chiffres - Édition 2019 ». 2019, Consulté le: mars 26, 2020. [En ligne]. Disponible sur: <https://www.calameo.com/read/0022300516dd96abad95b>.
- [11] FranceAgriMer - établissement national des produits de l'agriculture et de la mer, « La filière Lait ». 2019, Consulté le: mars 18, 2020. [En ligne]. Disponible sur: <https://www.franceagrimer.fr/filiere-lait/La-filiere-Lait>.
- [12] CNIEL, « France Terre de lait - Plan de la filière laitière ». déc. 2017.
- [13] Idele, (Institut de l'Élevage) et CNE, Confédération Nationale de l'Élevage), « Chiffres clés Caprins 2019 ». 2019, Consulté le: mars 20, 2020. [En ligne]. Disponible sur: http://idele.fr/no_cache/recherche/publication/idelesolr/recommends/chiffres-cles-caprins-2019.html.
- [14] Idele, (Institut de l'Élevage) et CNE, Confédération Nationale de l'Élevage), « Chiffres clés Ovins 2019 ». 2019, Consulté le: mars 20, 2020. [En ligne]. Disponible sur:

- <http://idele.fr/reseaux-et-partenariats/eurodairy/publication/idelesolr/recommends/chiffres-cles-ovins-2019.html>.
- [15] J. Boichard, « Le lait et les problèmes de l'élevage laitier en France », *Rev. Géographie Lyon*, vol. 47, n° 2, p. 99-135, 1972, doi: 10.3406/geoca.1972.1603.
- [16] Agreste, « Produits agroalimentaires. Ovins, caprins, collecte de lait | Agreste, la statistique agricole ». 2019, Consulté le: mars 26, 2020. [En ligne]. Disponible sur: <https://agreste.agriculture.gouv.fr/agreste-web/disaron/GraFraChap12.7/detail/>.
- [17] Idele, (Institut de l'Élevage), « Résultats de Contrôle Laitier - France 2018 ». 2019, Consulté le: mars 25, 2020. [En ligne]. Disponible sur: http://idele.fr/no_cache/recherche/publication/idelesolr/recommends/resultats-de-controle-laitier-france-2018.html.
- [18] D. Pourchet et N. Gaudilliere, « La composition fine du Lait », p. 23.
- [19] C. L. Manuelian, M. Penasa, G. Visentin, A. Zidi, M. Cassandro, et M. De Marchi, « Mineral composition of cow milk from multibreed herds », *Anim. Sci. J. Nihon Chikusan Gakkaiho*, vol. 89, n° 11, p. 1622-1627, nov. 2018, doi: 10.1111/asj.13095.
- [20] Zwierzchowski et Ametaj, « Mineral Elements in the Raw Milk of Several Dairy Farms in the Province of Alberta », *Foods*, vol. 8, n° 8, p. 345, août 2019, doi: 10.3390/foods8080345.
- [21] S. Currò *et al.*, « Differences in the Detailed Milk Mineral Composition of Italian Local and Saanen Goat Breeds », *Animals*, vol. 9, n° 7, p. 412, juill. 2019, doi: 10.3390/ani9070412.
- [22] A. Polychroniadou et A. Vafopoulou, « Variations of Major Mineral Constituents of Ewe Milk During Lactation », *J. Dairy Sci.*, vol. 68, n° 1, p. 147-150, janv. 1985, doi: 10.3168/jds.S0022-0302(85)80808-0.
- [23] M. W. Taylor et A. K. H. MacGibbon, « Milk Lipids | Fatty Acids », in *Encyclopedia of Dairy Sciences (Second Edition)*, J. W. Fuquay, Éd. San Diego: Academic Press, 2011, p. 655-659.
- [24] M. H. T. Maurice-Van Eijndhoven, H. Bovenhuis, H. Soyeurt, et M. P. L. Calus, « Differences in milk fat composition predicted by mid-infrared spectrometry among dairy cattle breeds in the Netherlands », *J. Dairy Sci.*, vol. 96, n° 4, p. 2570-2582, avr. 2013, doi: 10.3168/jds.2012-5793.
- [25] P. Caboni, C. Manis, I. Ibba, M. Contu, V. Coroneo, et P. Scano, « Compositional profile of ovine milk with a high somatic cell count: A metabolomics approach », *Int. Dairy J.*, vol. 69, p. 33-39, juin 2017, doi: 10.1016/j.idairyj.2017.02.001.
- [26] L. Alonso, J. Fontecha, L. Lozada, M. J. Fraga, et M. Juárez, « Fatty Acid Composition of Caprine Milk: Major, Branched-Chain, and Trans Fatty Acids », *J. Dairy Sci.*, vol. 82, n° 5, p. 878-884, mai 1999, doi: 10.3168/jds.S0022-0302(99)75306-3.
- [27] L. F. De La Fuente *et al.*, « Factors influencing variation of fatty acid content in ovine milk », *J. Dairy Sci.*, vol. 92, n° 8, p. 3791-3799, août 2009, doi: 10.3168/jds.2009-2151.
- [28] D. A. Tzompa-Sosa, G. A. van Aken, A. C. M. van Hooijdonk, et H. J. F. van Valenberg, « Influence of C16:0 and long-chain saturated fatty acids on normal variation of bovine milk fat triacylglycerol structure », *J. Dairy Sci.*, vol. 97, n° 7, p. 4542-4551, juill. 2014, doi: 10.3168/jds.2014-7937.

- [29] E. Samková, J. Špička, M. Pešek, T. Pelikánová, et O. Hanuš, « Animal factors affecting fatty acid composition of cow milk fat: A review », *South Afr. J. Anim. Sci.*, vol. 42, n° 2, Art. n° 2, 2012, doi: 10.4314/sajas.v42i2.1.
- [30] J. Djordjevic, T. Ledina, M. Z. Baltic, D. Trbovic, M. Babic, et S. Bulajic, « Fatty acid profile of milk », *IOP Conf. Ser. Earth Environ. Sci.*, vol. 333, p. 012057, oct. 2019, doi: 10.1088/1755-1315/333/1/012057.
- [31] J. Léonil, M. C. Michalski, et P. Martin, « Les structures supramoléculaires du lait : structure et impact nutritionnel de la micelle de caséine et du globule gras », *INRAE Prod. Anim.*, vol. 26, n° 2, p. 129-144, avr. 2013, doi: 10.20870/productions-animales.2013.26.2.3142.
- [32] R. Arora, N. Bhojak, et R. Joshi, « Comparative Aspects of Goat and Cow Milk », *Int. J. Eng. Sci. Invent.*, vol. 2, n° 1, p. 07-10, janv. 2013.
- [33] J. M. Jandal, « Comparative aspects of goat and sheep milk », *Small Rumin. Res.*, vol. 22, n° 2, p. 177-185, sept. 1996, doi: 10.1016/S0921-4488(96)00880-2.
- [34] Y. W. Park, M. Juárez, M. Ramos, et G. F. W. Haenlein, « Physico-chemical characteristics of goat and sheep milk », *Small Rumin. Res.*, vol. 68, n° 1, p. 88-113, mars 2007, doi: 10.1016/j.smallrumres.2006.09.013.
- [35] K. Raynal-Ljutovac, G. Lagriffoul, P. Paccard, I. Guillet, et Y. Chilliard, « Composition of goat and sheep milk products: An update », *Small Rumin. Res.*, vol. 79, n° 1, p. 57-72, sept. 2008, doi: 10.1016/j.smallrumres.2008.07.009.
- [36] C. Castillo, A. Abuelo, et J. Hernández, « Ruminant (Bovine, Caprine, and Ovine) Milk and Meat Production: The Challenge of Food Quality and Sustainability Through the Use of Plant Extracts », in *Encyclopedia of Food Security and Sustainability*, Elsevier, 2019, p. 25-42.
- [37] M. E. Pintado et F. X. Malcata, « Effect of thermal treatment on the protein profile of whey from ovine and caprine milk throughout lactation », *Int. Dairy J.*, vol. 6, n° 5, p. 497-518, mai 1996, doi: 10.1016/0958-6946(95)00049-6.
- [38] M. P. Sanchez *et al.*, « Short communication: Genetic parameters for milk protein composition predicted using mid-infrared spectroscopy in the French Montbéliarde, Normande, and Holstein dairy cattle breeds », *J. Dairy Sci.*, vol. 100, n° 8, p. 6371-6375, août 2017, doi: 10.3168/jds.2017-12663.
- [39] C. Maroteau *et al.*, « Genetic parameter estimation for major milk fatty acids in Alpine and Saanen primiparous goats », *J. Dairy Sci.*, vol. 97, n° 5, p. 3142-3155, mai 2014, doi: 10.3168/jds.2013-7328.
- [40] B. A. Law et A. Y. Tamime, *Technology of cheesemaking*, 2nd ed. Malden, MA: Blackwell, 2010.
- [41] G. Stocco, C. Cipolat-Gotet, T. Bobbo, A. Cecchinato, et G. Bittante, « Breed of cow and herd productivity affect milk composition and modeling of coagulation, curd firming, and syneresis », *J. Dairy Sci.*, vol. 100, n° 1, p. 129-145, janv. 2017, doi: 10.3168/jds.2016-11662.
- [42] E. Vanbergue *et al.*, « Effects of DGAT1 K232A polymorphism and milking frequency on milk composition and spontaneous lipolysis in dairy cows », *J. Dairy Sci.*, vol. 99, n° 7, p. 5739-5749, juill. 2016, doi: 10.3168/jds.2015-10731.

- [43] A. C. McCandlish, « The Influence of the Stage of Lactation on the Production of Dairy Cows », *J. Dairy Sci.*, vol. 7, n° 3, p. 255-261, mai 1924, doi: 10.3168/jds.S0022-0302(24)94020-3.
- [44] J. M. Morton, M. J. Auldish, M. L. Douglas, et K. L. Macmillan, « Associations between milk protein concentration at various stages of lactation and reproductive performance in dairy cows », *J. Dairy Sci.*, vol. 99, n° 12, p. 10044-10056, déc. 2016, doi: 10.3168/jds.2016-11276.
- [45] W. M. Stoop, H. Bovenhuis, J. M. L. Heck, et J. A. M. van Arendonk, « Effect of lactation stage and energy status on milk fat composition of Holstein-Friesian cows », *J. Dairy Sci.*, vol. 92, n° 4, p. 1469-1478, avr. 2009, doi: 10.3168/jds.2008-1468.
- [46] J. Kuchtík, K. Šustová, T. Urban, et D. Zapletal, « Effect of the stage of lactation on milk composition, its properties and the quality of rennet curdling in East Friesian ewes », *Czech J. Anim. Sci.*, vol. 53, n° No. 2, p. 55-63, févr. 2008, doi: 10.17221/333-CJAS.
- [47] M. S. El-Tarabany, A. A. El-Tarabany, et E. M. Roushdy, « Impact of lactation stage on milk composition and blood biochemical and hematological parameters of dairy Baladi goats », *Saudi J. Biol. Sci.*, vol. 25, n° 8, p. 1632-1638, déc. 2018, doi: 10.1016/j.sjbs.2016.08.003.
- [48] L. Bertocchi, A. Vitali, N. Lacetera, A. Nardone, G. Varisco, et U. Bernabucci, « Seasonal variations in the composition of Holstein cow's milk and temperature–humidity index relationship », *Animal*, vol. 8, n° 4, p. 667-674, avr. 2014, doi: 10.1017/S1751731114000032.
- [49] R. Bouraoui, M. Lahmar, A. Majdoub, M. Djemali, et R. Belyea, « The relationship of temperature-humidity index with milk production of dairy cows in a Mediterranean climate », *Anim. Res.*, vol. 51, n° 6, p. 479-491, nov. 2002, doi: 10.1051/animres:2002036.
- [50] D. L. Smith, T. Smith, B. J. Rude, et S. H. Ward, « Short communication: Comparison of the effects of heat stress on milk and component yields and somatic cell score in Holstein and Jersey cows », *J. Dairy Sci.*, vol. 96, n° 5, p. 3028-3033, mai 2013, doi: 10.3168/jds.2012-5737.
- [51] N. V. Kljajevic, I. B. Tomasevic, Z. N. Miloradovic, A. Nedeljkovic, J. B. Miocinovic, et S. T. Jovanovic, « Seasonal variations of Saanen goat milk composition and the impact of climatic conditions », *J. Food Sci. Technol.*, vol. 55, n° 1, p. 299-303, janv. 2018, doi: 10.1007/s13197-017-2938-4.
- [52] A. Contreras-Jodar, A. A. Salama, S. Hamzaoui, M. Vailati-Riboni, G. Caja, et J. J. Loor, « Effects of chronic heat stress on lactational performance and the transcriptomic profile of blood cells in lactating dairy goats », *J. Dairy Res.*, vol. 85, n° 4, p. 423-430, nov. 2018, doi: 10.1017/S0022029918000705.
- [53] R. Finocchiaro, J. B. C. H. M. van Kaam, B. Portolano, et I. Misztal, « Effect of Heat Stress on Production of Mediterranean Dairy Sheep », *J. Dairy Sci.*, vol. 88, n° 5, p. 1855-1864, mai 2005, doi: 10.3168/jds.S0022-0302(05)72860-5.
- [54] I. Peana, G. Fois, et A. Cannas, « Effects of heat stress and diet on milk production and feed and energy intake of Sarda ewes », *Ital. J. Anim. Sci.*, vol. 6, n° sup1, p. 577-579, janv. 2007, doi: 10.4081/ijas.2007.1s.577.

- [55] M. Zhao *et al.*, « Milk production and composition responds to dietary neutral detergent fiber and starch ratio in dairy cows », *Anim. Sci. J.*, vol. 87, n° 6, p. 756-766, 2016, doi: 10.1111/asj.12482.
- [56] C. Arndt, L. E. Armentano, et M. B. Hall, « Corn bran versus corn grain at 2 levels of forage: Intake, apparent digestibility, and production responses by lactating dairy cows », *J. Dairy Sci.*, vol. 97, n° 9, p. 5676-5687, sept. 2014, doi: 10.3168/jds.2013-7815.
- [57] T. Zhang, B. Si, K. Deng, Y. Tu, C. Zhou, et Q. Diao, « Effects of feeding a *Moringa oleifera* rachis and twig preparation to dairy cows on their milk production and fatty acid composition, and plasma antioxidants », *J. Sci. Food Agric.*, vol. 98, n° 2, p. 661-666, 2018, doi: 10.1002/jsfa.8511.
- [58] P. K. Chelikani, J. D. Ambrose, D. H. Keisler, et J. J. Kennelly, « Effect of short-term fasting on plasma concentrations of leptin and other hormones and metabolites in dairy cattle », *Domest. Anim. Endocrinol.*, vol. 26, n° 1, p. 33-48, janv. 2004, doi: 10.1016/j.domaniend.2003.08.003.
- [59] S. Agenäs, K. Dahlborn, et K. Holtenius, « Changes in metabolism and milk production during and after feed deprivation in primiparous cows selected for different milk fat content », *Livest. Prod. Sci.*, vol. 83, n° 2-3, p. 153-164, oct. 2003, doi: 10.1016/S0301-6226(03)00096-4.
- [60] V. Bjerre-Harpøth *et al.*, « Metabolic and production profiles of dairy cows in response to decreased nutrient density to increase physiological imbalance at different stages of lactation », *J. Dairy Sci.*, vol. 95, n° 5, p. 2362-2380, mai 2012, doi: 10.3168/jds.2011-4419.
- [61] L. Herve *et al.*, « Milk yield loss in response to feed restriction is associated with mammary epithelial cell exfoliation in dairy cows », *J. Dairy Sci.*, vol. 102, n° 3, p. 2670-2685, mars 2019, doi: 10.3168/jds.2018-15398.
- [62] S. Ghazal, V. Berthelot, N. C. Friggens, et P. Schmidely, « Effects of conjugated linoleic acid supplementation and feeding level on dairy performance, milk fatty acid composition, and body fat changes in mid-lactation goats », *J. Dairy Sci.*, vol. 97, n° 11, p. 7162-7174, nov. 2014, doi: 10.3168/jds.2014-8364.
- [63] D. Thoh, P. Pakdeechnuan, et P. Chanjula, « Effect of supplementary glycerin on milk composition and heat stability in dairy goats », *Asian-Australas. J. Anim. Sci.*, vol. 30, n° 12, p. 1711-1717, déc. 2017, doi: 10.5713/ajas.17.0066.
- [64] A. A. Naserian, C. R. Staples, et M. H. Ghaffari, « Effects of replacing wheat bran by pistachio skins on feed intake, nutrient digestibility, milk yield, milk composition and blood metabolites of dairy Saanen goats », *J. Anim. Physiol. Anim. Nutr.*, vol. 100, n° 2, p. 256-263, 2016, doi: 10.1111/jpn.12362.
- [65] S. M. Kitessa, S. K. Gulati, J. R. Ashes, E. Fleck, T. W. Scott, et P. D. Nichols, « Utilisation of fish oil in ruminants: II. Transfer of fish oil fatty acids into goats' milk », *Anim. Feed Sci. Technol.*, vol. 89, n° 3, p. 201-208, févr. 2001, doi: 10.1016/S0377-8401(00)00232-7.
- [66] N. Iqbal *et al.*, « Effect of feeding different levels of discarded date palm (*Phoenix dactylifera*) on digestibility, milk yield, and composition in Damani sheep », *Trop.*

- Anim. Health Prod.*, vol. 51, n° 8, p. 2181-2186, nov. 2019, doi: 10.1007/s11250-019-01919-w.
- [67] V. Tufarelli, R. U. Khan, et V. Laudadio, « Vitamin and trace element supplementation in grazing dairy ewe during the dry season: effect on milk yield, composition, and clotting aptitude », *Trop. Anim. Health Prod.*, vol. 43, n° 5, p. 955-960, juin 2011, doi: 10.1007/s11250-011-9789-1.
- [68] T. Bobbo *et al.*, « Associations between pathogen-specific cases of subclinical mastitis and milk yield, quality, protein composition, and cheese-making traits in dairy cows », *J. Dairy Sci.*, vol. 100, n° 6, p. 4868-4883, juin 2017, doi: 10.3168/jds.2016-12353.
- [69] Idele, (Institut de l'Élevage), « Evaluation de l'incidence des infections et inflammations mammaires sur la production laitière chez la chèvre ». 2003, Consulté le: mars 18, 2020. [En ligne]. Disponible sur: http://idele.fr/no_cache/recherche/publication/idelesolr/recommends/evaluation-de-lincidence-des-infections-et-inflammations-mammaires-sur-la-production-laitiere-chez.html.
- [70] G. Koop, T. van Werven, H. J. Schuiling, et M. Nielen, « The effect of subclinical mastitis on milk yield in dairy goats », *J. Dairy Sci.*, vol. 93, n° 12, p. 5809-5817, déc. 2010, doi: 10.3168/jds.2010-3544.
- [71] R. Rupp *et al.*, « A Point Mutation in Suppressor of Cytokine Signalling 2 (Socs2) Increases the Susceptibility to Inflammation of the Mammary Gland while Associated with Higher Body Weight and Size and Higher Milk Production in a Sheep Model », *PLoS Genet.*, vol. 11, n° 12, déc. 2015, doi: 10.1371/journal.pgen.1005629.
- [72] D. Bagri, R. Pandey, G. Bagri, R. Kumari, et D. Bagdi, « Effect of subclinical mastitis on milk composition in lactating cows », *J. Entomol. Zool. Stud.*, p. 6.
- [73] B. K. Bansal, J. Hamann, N. T. Grabowski, et K. B. Singh, « Variation in the composition of selected milk fraction samples from healthy and mastitic quarters, and its significance for mastitis diagnosis », *J. Dairy Res.*, vol. 72, n° 2, p. 144-152, mai 2005, doi: 10.1017/s0022029905000798.
- [74] G. Leitner, U. Merin, et N. Silanikove, « Changes in Milk Composition as Affected by Subclinical Mastitis in Goats », *J. Dairy Sci.*, vol. 87, n° 6, p. 1719-1726, juin 2004, doi: 10.3168/jds.S0022-0302(04)73325-1.
- [75] Ministère de l'Agriculture et de l'Alimentation, *Arrêté du 17 décembre 2018 modifiant l'arrêté du 29 avril 2015 fixant la liste des races des espèces bovine, ovine, caprine et porcine reconnues et précisant les ressources zoogénétiques présentant un intérêt pour la conservation du patrimoine génétique du cheptel et l'aménagement du territoire*. 2018.
- [76] Races de France, « Bovins - Races de France ». <https://www.racesdefrance.fr/bovins> (consulté le mars 18, 2020).
- [77] Races de France, « Chevres - Races de France ». <https://www.racesdefrance.fr/chevres> (consulté le mars 18, 2020).
- [78] Ministère de l'Agriculture et de l'Alimentation, *Arrêté du 26 juillet 2007 fixant la liste des races des espèces bovine, ovine, caprine et porcine reconnues et précisant les ressources zoogénétiques présentant un intérêt pour la conservation du patrimoine génétique du cheptel et l'aménagement du territoire*. 2007.

- [79] Races de France, « Moutons - Races de France ». <https://www.racesdefrance.fr/moutons> (consulté le mars 18, 2020).
- [80] INRA, « Races animales françaises menacées d'abandon pour l'agriculture », Commission Nationale d'Amélioration Génétique, Ministère de l'Agriculture, de l'Agroalimentaire et de la Forêt, Rapport méthodologique Etude n°SSP-2014-003, nov. 2014. Consulté le: avr. 06, 2020. [En ligne]. Disponible sur: https://agriculture.gouv.fr/sites/minagri/files/races_menacees_rapport_methodologique.pdf.
- [81] Ministère de l'Agriculture et de l'Alimentation, *Arrêté du 29 avril 2015 fixant la liste des races des espèces bovine, ovine, caprine et porcine reconnues et précisant les ressources zoogénétiques présentant un intérêt pour la conservation du patrimoine génétique du cheptel et l'aménagement du territoire - Légifrance*. 2015.
- [82] Web-agri, « Résultats du contrôle laitier 2015 Le détail comparé des races laitières », *Web-agri*, 2016. <http://www.web-agri.fr/conduite-elevage/genetique-race/article/le-detail-compare-des-races-laitieres-1175-118529.html> (consulté le avr. 25, 2019).
- [83] Evolution, « Croisement ». 2019, Consulté le: mars 26, 2020. [En ligne]. Disponible sur: http://www.evolution-int.com/sites/default/files/catalogs/croisement_fr_2019_evolution_international.pdf.
- [84] France Génétique Elevage, « La génétique française pour les filières bovines, ovines et caprines ». 2011, Consulté le: mars 26, 2020. [En ligne]. Disponible sur: http://fr.france-genetique-elevage.org/IMG/pdf/fge_brochure_genetique_francaise_bd.pdf.
- [85] I. INAO (Institut National de l'Origine et de la Qualité, « Aires et produits AOC/AOP et IGP ». data.gouv.fr, nov. 05, 2020, Consulté le: nov. 19, 2020. [En ligne]. Disponible sur: [/fr/datasets/aires-et-produits-aoc-aop-et-igp/](http://fr/datasets/aires-et-produits-aoc-aop-et-igp/).
- [86] F. Grosclaude, P. Martin, G. Ricordeau, F. Remeuf, L. Vassal, et J. Bouillon, « Du gène au fromage : le polymorphisme de la caséine alphas1 caprine, ses effets, son évolution », *INRAE Prod. Anim.*, vol. 7, n° 1, p. 3-19, févr. 1994, doi: 10.20870/productions-animales.1994.7.1.4153.
- [87] AgroParisTech, « Race bovine Holstein ». <http://www2.agroparistech.fr/svs/genere/especes/bovins/primhols.htm> (consulté le nov. 19, 2020).
- [88] AgroParisTech, « Race bovine Normande ». <http://www2.agroparistech.fr/svs/genere/especes/bovins/normand.htm> (consulté le nov. 19, 2020).
- [89] AgroParisTech, « Race bovine Montbéliarde ». <http://www2.agroparistech.fr/svs/genere/especes/bovins/montbel.htm> (consulté le nov. 19, 2020).
- [90] Capgènes, « La race Alpine française », *Capgenes*. <http://www.capgenes.com/les-races-caprines/race-alpine-francaise/> (consulté le oct. 18, 2019).
- [91] Capgènes, « La race Saanen française », *Capgenes*. <http://www.capgenes.com/les-races-caprines/race-saanen-francaise/> (consulté le oct. 18, 2019).

- [92] AgroParisTech, « Race ovine Lacaune lait ». <http://www2.agroparistech.fr/svs/genere/especes/ovins/lacaulai.htm> (consulté le nov. 19, 2020).
- [93] M. C. INRAE, « Photothèque INRAE ». <https://mediatheque.inrae.fr/INRAE/media/537008327> (consulté le nov. 19, 2020).
- [94] France Génétique Elevage, « France génétique Elevage - Sélection des races bovines laitières ». <http://fr.france-genetique-elevage.org/Selection-des-races-bovines.html> (consulté le avr. 25, 2019).
- [95] D. Boichard *et al.*, « Genomic selection in French dairy cattle », *Anim. Prod. Sci.*, vol. 52, n° 3, p. 115, 2012, doi: 10.1071/AN11119.
- [96] The Cattle Site, « Breeds - Holstein », *The Cattle Site*. <http://www.thecattlesite.com/breeds/dairy/22/holstein/> (consulté le avr. 18, 2019).
- [97] Prim'Holstein France, « Présentation de la race | Prim'Holstein France », 2019. <https://primholstein.com/la-prim-holstein/presentation-primholstein/> (consulté le févr. 27, 2019).
- [98] Holstein USA, « History of the Holstein Breed ». http://www.holsteinusa.com/holstein_breed/breedhistory.html (consulté le avr. 18, 2019).
- [99] The Cattle Site, « Breeds - Normande », *The Cattle Site*. <http://www.thecattlesite.com/breeds/dairy/68/normande/> (consulté le avr. 18, 2019).
- [100] Races de France, « Race Bovine Montbéliarde - Races de France ». <https://www.racesdefrance.fr/bovins?id=108> (consulté le avr. 24, 2019).
- [101] France Génétique Elevage, « France génétique Elevage - Montbéliarde ». <http://fr.france-genetique-elevage.org/Montbeliarde.html> (consulté le avr. 24, 2019).
- [102] Les produits laitiers, « Portrait de vache : la Montbéliarde », *Les produits laitiers*, janv. 17, 2011. <https://www.produits-laitiers.com/article/portrait-de-vache-la-montbeliarde> (consulté le avr. 24, 2019).
- [103] France Génétique Elevage, « France génétique Elevage - Alpine ». <http://fr.france-genetique-elevage.org/Alpine.html> (consulté le oct. 18, 2019).
- [104] I. Palhiere, H. Larroque, V. Clement, Tosser-Klopp, Gwenola, et Rupp, Rachel, « Genetic parameters and QTL detection for milking speed in dairy Alpine and Saanen goats. Proceedings, 10th world congress of genetics applied to livestock production. », *Present. 10 World Congr. Genet. Appl. Livest. Prod. WCGALP*, p. 892-892, 2014.
- [105] Races de France, « Race Caprine Saanen - Races de France ». <https://www.racesdefrance.fr/chevres/races-caprines?id=149> (consulté le avr. 25, 2019).
- [106] Races de massif, « Ovin Lacaune ». <https://www.races-montagnes.com/fr/races/lacaune.php> (consulté le avr. 25, 2019).
- [107] P. Hassoun *et al.*, « La monotraite quotidienne appliquée en brebis laitières de race Lacaune : Synthèse de cinq années de recherche », p. 15, 2016.
- [108] J. Labussière, J.-F. Combaud, P. Petrequin, et P. Chesne, « INFLUENCE DE LA FRÉQUENCE DES TRAITES ET DES TÉTÉES SUR LA PRODUCTION LAITIÈRE DES BREBIS

- PRÉALPES DU SUD », *Ann. Zootech.*, vol. 23, n° 4, p. 445-457, 1974, doi: 10.1051/animres:19740403.
- [109] Races de massif, « Races ovines laitières des Pyrénées : Manech tête noire ». <https://www.races-montagnes.com/fr/races/pyrenees-atlantiques/manech-tete-noire.php> (consulté le avr. 25, 2019).
- [110] France Génétique Elevage, « France génétique Elevage - Laitières des Pyrénées ». <http://fr.france-genetique-elevage.org/Laitieres-des-Pyrenees.html> (consulté le oct. 18, 2019).
- [111] B. Wang, L. Chen, et W. Wang, « Genomic insights into ruminant evolution: from past to future prospects », *Zool. Res.*, vol. 40, n° 6, p. 476-487, nov. 2019, doi: 10.24272/j.issn.2095-8137.2019.061.
- [112] A. Bucheton, « LES GÉNOMES ET SES ENJEUX », p. 48, 2004.
- [113] B. D. Rosen *et al.*, « De novo assembly of the cattle reference genome with single-molecule sequencing », *GigaScience*, vol. 9, n° 3, mars 2020, doi: 10.1093/gigascience/giaa021.
- [114] D. W. Burt, « The cattle genome reveals its secrets », *J. Biol.*, vol. 8, n° 4, p. 36, 2009, doi: 10.1186/jbiol137.
- [115] D. E. Hagen, D. R. Unni, A. Tayal, G. W. Burns, et C. G. Elsik, « Bovine Genome Database: Tools for Mining the Bos taurus Genome », in *Eukaryotic Genomic Databases: Methods and Protocols*, M. Kollmar, Éd. New York, NY: Springer New York, 2018, p. 211-249.
- [116] J. Weissenbach et M. Salanoubat, « Séquence des génomes : le feu d'artifice. », *médecine/sciences*, vol. 16, n° 1, p. 10, 2000, doi: 10.4267/10608/1495.
- [117] A. V. Zimin *et al.*, « A whole-genome assembly of the domestic cow, Bos taurus », *Genome Biol.*, vol. 10, n° 4, p. R42, avr. 2009, doi: 10.1186/gb-2009-10-4-r42.
- [118] Y. Dong *et al.*, « Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus) », *Nat. Biotechnol.*, vol. 31, n° 2, Art. n° 2, févr. 2013, doi: 10.1038/nbt.2478.
- [119] NCBI - C. hircus, « capra hircus - Assembly - NCBI », *Assembly - C. hircus*. <https://www.ncbi.nlm.nih.gov/assembly/?term=capra+hircus> (consulté le avr. 06, 2020).
- [120] D. M. Bickhart *et al.*, « Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome », *Nat. Genet.*, vol. 49, n° 4, Art. n° 4, avr. 2017, doi: 10.1038/ng.3802.
- [121] The International Sheep Genomics Consortium *et al.*, « The sheep genome reference sequence: a work in progress », *Anim. Genet.*, vol. 41, n° 5, p. 449-453, 2010, doi: 10.1111/j.1365-2052.2010.02100.x.
- [122] NCBI - O. aries, « ovis aries - Assembly - NCBI », *Assembly - O. aries*. <https://www.ncbi.nlm.nih.gov/assembly/?term=ovis+aries> (consulté le avr. 06, 2020).
- [123] A. C. English *et al.*, « Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology », *PLOS ONE*, vol. 7, n° 11, p. e47768, nov. 2012, doi: 10.1371/journal.pone.0047768.
- [124] B.-H. Baylor College of Medicine (BCM), « Sheep Genome Project », 2017, [En ligne]. Disponible sur: <https://www.hgsc.bcm.edu/other-mammals/sheep-genome-project>.

- [125] Ensembl genome browser 99, « Ovis aries - Ensembl genome browser 99 », janv. 2020. https://www.ensembl.org/Ovis_aries/Info/Annotation (consulté le mars 23, 2020).
- [126] M. Georges, C. Charlier, et B. Hayes, « Harnessing genomic information for livestock improvement », *Nat. Rev. Genet.*, vol. 20, n° 3, p. 135-156, mars 2019, doi: 10.1038/s41576-018-0082-2.
- [127] NCBI, « ARS-UCD1.2 - bosTau9 - Genome - Assembly - NCBI ». https://www.ncbi.nlm.nih.gov/assembly/GCF_002263795.1 (consulté le avr. 06, 2020).
- [128] Ensembl genome browser 99, « Bos taurus - Ensembl genome browser 99 », janv. 2020. https://www.ensembl.org/Bos_taurus/Info/Annotation (consulté le mars 20, 2020).
- [129] Ensembl genome browser 99, « Capra hircus - Ensembl genome browser 99 », janv. 2020. https://www.ensembl.org/Capra_hircus/Info/Annotation (consulté le mars 20, 2020).
- [130] Ensembl, « Calculated consequences ». https://www.ensembl.org/info/genome/variation/prediction/predicted_data.html (consulté le nov. 18, 2020).
- [131] V. Le Morvan, J.-L. Formento, G. Milano, J. Bonnet, et J. Robert, « Techniques de recherche des polymorphismes génétiques », *Oncologie*, vol. 7, n° 1, p. 7-16, mars 2005, doi: 10.1007/s10269-005-0146-8.
- [132] W. McLaren *et al.*, « The Ensembl Variant Effect Predictor », *Genome Biol.*, vol. 17, n° 1, p. 122, déc. 2016, doi: 10.1186/s13059-016-0974-4.
- [133] R. Karki, D. Pandya, R. C. Elston, et C. Ferlini, « Defining “mutation” and “polymorphism” in the era of personal genomics », *BMC Med. Genomics*, vol. 8, n° 1, p. 37, juill. 2015, doi: 10.1186/s12920-015-0115-z.
- [134] J. K. Sehn, « Chapter 9 - Insertions and Deletions (Indels) », in *Clinical Genomics*, S. Kulkarni et J. Pfeifer, Éd. Boston: Academic Press, 2015, p. 129-150.
- [135] M. Boussaha *et al.*, « Genome-Wide Study of Structural Variants in Bovine Holstein, Montbéliarde and Normande Dairy Breeds », *PLOS ONE*, vol. 10, n° 8, p. e0135931, août 2015, doi: 10.1371/journal.pone.0135931.
- [136] M. Lin, S. Whitmire, J. Chen, A. Farrel, X. Shi, et J. Guo, « Effects of short indels on protein structure and function in human genomes », *Sci. Rep.*, vol. 7, août 2017, doi: 10.1038/s41598-017-09287-x.
- [137] P. Bennett, « Demystified ... », *Mol. Pathol.*, vol. 53, n° 4, p. 177-183, août 2000.
- [138] D. Boichard, C. Grohs, C. Danchin-Burge, A. Capitan, et U. Paris-Saclay, « Les anomalies génétiques : définition, origine, transmission et évolution, mode d'action », p. 9, 2016.
- [139] R. Redon *et al.*, « Global variation in copy number in the human genome », *Nature*, vol. 444, n° 7118, p. 444-454, nov. 2006, doi: 10.1038/nature05329.
- [140] Y. Gao *et al.*, « CNV discovery for milk composition traits in dairy cattle using whole genome resequencing », *BMC Genomics*, vol. 18, mars 2017, doi: 10.1186/s12864-017-3636-3.
- [141] G. H. Davis, « Fecundity genes in sheep », *Anim. Reprod. Sci.*, vol. 82-83, p. 247-253, juill. 2004, doi: 10.1016/j.anireprosci.2004.04.001.

- [142] H. Gilbert, P. L. Roy, D. Milan, et J.-P. Bidanel, « De la détection de QTL à l'analyse des grandes fonctions : l'exemple du chromosome 7 porcine et de la composition corporelle », p. 6, 2004.
- [143] J. I. Weller, Y. Kashi, et M. Soller, « Power of Daughter and Granddaughter Designs for Determining Linkage Between Marker Loci and Quantitative Trait Loci in Dairy Cattle », *J. Dairy Sci.*, vol. 73, n° 9, p. 2525-2537, sept. 1990, doi: 10.3168/jds.S0022-0302(90)78938-2.
- [144] B. Hayes, « Linkage disequilibrium mapping (LD) mapping and Bioinformatics Approaches to identify QTL », p. 28, 2020.
- [145] P. Pikkukhokana et M. J. Sillanpää, « Combined linkage disequilibrium and linkage mapping: Bayesian multilocus approach », *Heredity*, vol. 112, n° 3, p. 351-360, mars 2014, doi: 10.1038/hdy.2013.111.
- [146] A. Legarra et R. L. Fernando, « Linear models for joint association and linkage QTL mapping », *Genet. Sel. Evol.*, vol. 41, n° 1, p. 43, déc. 2009, doi: 10.1186/1297-9686-41-43.
- [147] Z. Cai, B. Guldbandsen, M. S. Lund, et G. Sahana, « Dissecting closely linked association signals in combination with the mammalian phenotype database can identify candidate genes in dairy cattle », *BMC Genet.*, vol. 20, n° 1, p. 15, déc. 2019, doi: 10.1186/s12863-019-0717-0.
- [148] B. N. Keel, W. M. Snelling, A. K. Lindholm-Perry, W. T. Oliver, L. A. Kuehn, et G. A. Rohrer, « Using SNP Weights Derived From Gene Expression Modules to Improve GWAS Power for Feed Efficiency in Pigs », *Front. Genet.*, vol. 10, janv. 2020, doi: 10.3389/fgene.2019.01339.
- [149] N. Oak, R. Ghosh, K. Huang, D. A. Wheeler, L. Ding, et S. E. Plon, « Framework for microRNA variant annotation and prioritization using human population and disease datasets », *Hum. Mutat.*, vol. 40, n° 1, p. 73-89, janv. 2019, doi: 10.1002/humu.23668.
- [150] D. Stacey *et al.*, « ProGeM: a framework for the prioritization of candidate causal genes at molecular quantitative trait loci », *Nucleic Acids Res.*, vol. 47, n° 1, p. e3-e3, janv. 2019, doi: 10.1093/nar/gky837.
- [151] R. Xiang *et al.*, « Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits », *Proc. Natl. Acad. Sci.*, vol. 116, n° 39, p. 19398-19408, sept. 2019, doi: 10.1073/pnas.1904159116.
- [152] Z.-L. Hu, E. R. Fritz, et J. M. Reecy, « AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond », *Nucleic Acids Res.*, vol. 35, n° Database, p. D604-D609, janv. 2007, doi: 10.1093/nar/gkl946.
- [153] M.-P. Sanchez *et al.*, « Within-breed and multi-breed GWAS on imputed whole-genome sequence variants reveal candidate mutations affecting milk protein composition in dairy cattle », *Genet. Sel. Evol.*, vol. 49, n° 1, p. 68, déc. 2017, doi: 10.1186/s12711-017-0344-z.
- [154] T. Tribout *et al.*, « Confirmed effects of candidate variants for milk production, udder health, and udder morphology in dairy cattle », *Genet. Sel. Evol.*, vol. 52, n° 1, p. 55, déc. 2020, doi: 10.1186/s12711-020-00575-1.
- [155] C. Oget, M. Teissier, J.-M. Astruc, G. Tosser-Klopp, et R. Rupp, « Alternative methods improve the accuracy of genomic prediction using information from a causal point

- [168] France Génétique Elevage, « France génétique Elevage - Dairy cattle breeds indexes ». <http://en.france-genetique-elevage.org/Dairy-cattle-breeds-indexes.html> (consulté le oct. 07, 2019).
- [169] Alice, « Index : carte des potentiels génétiques - alice ». Consulté le: avr. 08, 2020. [En ligne]. Disponible sur: <https://www.alice.fr/genetique-selection/selection-genetique-des-ruminants/index--carte-des-potentiels-genetiques.html>.
- [170] I. Boujenane, « La sélection génomique : Une révolution dans le monde de la génétique animale », *Elev. Bov.*, vol. 16, p. 9-12, janv. 2010.
- [171] L. Chavinskaia, « Interbull: Constructing International Commensurability for Dairy Cattle Selection », *Interbull Bull.*, n° 51, Art. n° 51, oct. 2017, Consulté le: avr. 08, 2020. [En ligne]. Disponible sur: <https://journal.interbull.org/index.php/ib/article/view/1428>.
- [172] France Génétique Elevage, « La sélection génomique accessible à tous les ruminants ». févr. 2016, Consulté le: avr. 26, 2019. [En ligne]. Disponible sur: http://fr.france-genetique-elevage.org/IMG/pdf/selection_genomique_web.pdf.
- [173] France Génétique Elevage, « Bovins lait - La sélection génomique en France ». févr. 2016, Consulté le: avr. 26, 2019. [En ligne]. Disponible sur: http://fr.france-genetique-elevage.org/IMG/pdf/bovins_lait_web.pdf.
- [174] France Génétique Elevage, « France génétique Elevage - La sélection génomique ». <http://fr.france-genetique-elevage.org/La-selection-genomique.html> (consulté le avr. 26, 2019).
- [175] France Génétique Elevage, « Ovins lait et viande - La sélection génomique en France ». 2016, Consulté le: avr. 26, 2019. [En ligne]. Disponible sur: http://fr.france-genetique-elevage.org/IMG/pdf/ovins_web.pdf.
- [176] Idele, (Institut de l'Elevage), « Les ovins laitiers résolument tournés vers la sélection génomique ». janv. 17, 2019, Consulté le: avr. 08, 2020. [En ligne]. Disponible sur: <http://idele.fr/fr/domaines-techniques/ameliorer-le-troupeau/genomique/publication/idelesolr/recommends/selection-genomique-ovin-lait.html>.
- [177] Capgènes, « La sélection caprine, à l'ère de la génomique ». -, Consulté le: mars 17, 2020. [En ligne]. Disponible sur: https://www.capgenes.com/wp-content/uploads/2019/06/Plaqueette_Genomique_web_V02.pdf.
- [178] France Génétique Elevage, « Caprins - La sélection génomique en France ». févr. 2016, Consulté le: avr. 26, 2019. [En ligne]. Disponible sur: http://fr.france-genetique-elevage.org/IMG/pdf/caprin_web.pdf.
- [179] C. Carillier-Jacquín, H. Larroque, et C. Robert-Granié, « Vers une sélection génomique chez les caprins laitiers », *INRA Prod. Anim.*, vol. 30, n° 1, p. 19-30, juin 2018, doi: 10.20870/productions-animales.2017.30.1.2228.
- [180] The ENCODE Project Consortium, « An integrated encyclopedia of DNA elements in the human genome », *Nature*, vol. 489, n° 7414, p. 57-74, sept. 2012, doi: 10.1038/nature11247.
- [181] J.-N. Volff, *Vertebrate Genomes*. Karger Medical and Scientific Publishers, 2006.

- [182] Encode, « ENCODE: Deciphering Function in the Human Genome », *Genome.gov*. <https://www.genome.gov/27551473/genome-advance-of-the-month-encode-deciphering-function-in-the-human-genome> (consulté le mars 30, 2020).
- [183] N. Jareborg, E. Birney, et R. Durbin, « Comparative Analysis of Noncoding Regions of 77 Orthologous Mouse and Human Gene Pairs », *Genome Res.*, vol. 9, n° 9, p. 815-824, janv. 1999, doi: 10.1101/gr.9.9.815.
- [184] J. Harrow *et al.*, « Identifying protein-coding genes in genomic sequences », *Genome Biol.*, vol. 10, n° 1, p. 201, janv. 2009, doi: 10.1186/gb-2009-10-1-201.
- [185] R. P. Alexander, G. Fang, J. Rozowsky, M. Snyder, et M. B. Gerstein, « Annotating non-coding regions of the genome », *Nat. Rev. Genet.*, vol. 11, n° 8, Art. n° 8, août 2010, doi: 10.1038/nrg2814.
- [186] S. Ohno, « So much “junk” DNA in our genome », *Brookhaven Symp. Biol.*, vol. 23, p. 366-370, 1972.
- [187] F. Moraes et A. Góes, « A decade of human genome project conclusion: Scientific diffusion about our genome knowledge », *Biochem. Mol. Biol. Educ.*, vol. 44, n° 3, p. 215-223, 2016, doi: 10.1002/bmb.20952.
- [188] Study, « Non-Coding Regions of DNA: Sequences & Explanation - Video & Lesson Transcript », *Study.com*. <https://study.com/academy/lesson/non-coding-regions-of-dna-sequences-lesson-quiz.html> (consulté le mars 30, 2020).
- [189] G. H. Genetics Home Reference, « What is noncoding DNA? » mars 31, 2020, Consulté le: avr. 10, 2020. [En ligne]. Disponible sur: <https://ghr.nlm.nih.gov/primer/basics/noncodingdna>.
- [190] I. V. Chadaeva *et al.*, « Candidate SNP markers of reproductive potential are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters », *BMC Genomics*, vol. 19, n° Suppl 3, févr. 2018, doi: 10.1186/s12864-018-4478-3.
- [191] M. Gasperini, J. M. Tome, et J. Shendure, « Towards a comprehensive catalogue of validated and target-linked human enhancers », *Nat. Rev. Genet.*, p. 1-19, janv. 2020, doi: 10.1038/s41576-019-0209-0.
- [192] A. L. Roy, R. Sen, et R. G. Roeder, « Enhancer-promoter communication and transcriptional regulation of Igh », *Trends Immunol.*, vol. 32, n° 11, p. 532-539, nov. 2011, doi: 10.1016/j.it.2011.06.012.
- [193] S. Schoenfelder et P. Fraser, « Long-range enhancer–promoter contacts in gene expression control », *Nat. Rev. Genet.*, vol. 20, n° 8, Art. n° 8, août 2019, doi: 10.1038/s41576-019-0128-0.
- [194] M. Bansal, A. Kumar, et V. R. Yella, « Role of DNA sequence based structural features of promoters in transcription initiation and gene expression », *Curr. Opin. Struct. Biol.*, vol. 25, p. 77-85, avr. 2014, doi: 10.1016/j.sbi.2014.01.007.
- [195] Nature, « promoter | Learn Science at Scitable ». <https://www.nature.com/scitable/definition/promoter-259/> (consulté le avr. 10, 2020).
- [196] Genome.gov, « Promoter », *Genome.gov*. <https://www.genome.gov/genetics-glossary/Promoter> (consulté le avr. 10, 2020).

- [197] A. V. Barkhash, G. V. Kochneva, E. V. Chub, et A. G. Romaschenko, « Single nucleotide polymorphism rs1800872 in the promoter region of the IL10 gene is associated with predisposition to chronic hepatitis C in Russian population », *Microbes Infect.*, vol. 20, n° 3, p. 212-216, mars 2018, doi: 10.1016/j.micinf.2017.11.012.
- [198] B. J. Świątek, « Is interleukin-10 gene polymorphism a predictive marker in HCV infection? », *Cytokine Growth Factor Rev.*, vol. 23, n° 1-2, p. 47-59, févr. 2012, doi: 10.1016/j.cytogfr.2012.01.005.
- [199] M. Ponomarenko *et al.*, « How to Use SNP_TATA_Comparator to Find a Significant Change in Gene Expression Caused by the Regulatory SNP of This Gene's Promoter via a Change in Affinity of the TATA-Binding Protein for This Promoter », *BioMed Res. Int.*, vol. 2015, 2015, doi: 10.1155/2015/359835.
- [200] G. Cosenza, M. Iannaccone, B. A. Pico, L. Ramunno, et R. Capparelli, « The SNP g.1311T>C associated with the absence of β -casein in goat milk influences CSN2 promoter activity », *Anim. Genet.*, vol. 47, n° 5, p. 615-617, 2016, doi: 10.1111/age.12443.
- [201] H. Kawaji et Y. Hayashizaki, « Exploration of Small RNAs », *PLoS Genet.*, vol. 4, n° 1, p. e22, 2008, doi: 10.1371/journal.pgen.0040022.
- [202] A. Katz, S. Elgamal, A. Rajkovic, et M. Ibba, « Non-canonical roles of tRNAs and tRNA mimics in bacterial cell biology », *Mol. Microbiol.*, vol. 101, n° 4, p. 545-558, août 2016, doi: 10.1111/mmi.13419.
- [203] S. Hombach et M. Kretz, « Non-coding RNAs: Classification, Biology and Functioning », *Non-coding RNAs in Colorectal Cancer*, Cham, p. 3-17, 2016.
- [204] M. Esteller, « Non-coding RNAs in human disease », *Nat. Rev. Genet.*, vol. 12, n° 12, Art. n° 12, déc. 2011, doi: 10.1038/nrg3074.
- [205] T. R. Cech et J. A. Steitz, « The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones », *Cell*, vol. 157, n° 1, p. 77-94, mars 2014, doi: 10.1016/j.cell.2014.03.008.
- [206] J. S. Mattick et I. V. Makunin, « Non-coding RNA », *Hum. Mol. Genet.*, vol. 15, n° suppl_1, p. R17-R29, avr. 2006, doi: 10.1093/hmg/ddl046.
- [207] L. Sriyothi, S. Ponne, T. Prathama, C. Ashok, et S. Baluchamy, « Roles of Non-Coding RNAs in Transcriptional Regulation », *Transcr. Post-Transcr. Regul.*, oct. 2018, doi: 10.5772/intechopen.76125.
- [208] F. Ferrè, A. Colantoni, et M. Helmer-Citterich, « Revealing protein–lncRNA interaction », *Brief. Bioinform.*, vol. 17, n° 1, p. 106-116, janv. 2016, doi: 10.1093/bib/bbv031.
- [209] T. R. Mercer, M. E. Dinger, et J. S. Mattick, « Long non-coding RNAs: insights into functions », *Nat. Rev. Genet.*, vol. 10, n° 3, Art. n° 3, mars 2009, doi: 10.1038/nrg2521.
- [210] E. K. Robinson, S. Covarrubias, et S. Carpenter, « The how and why of lncRNA function: An innate immune perspective », *Biochim. Biophys. Acta Gene Regul. Mech.*, vol. 1863, n° 4, p. 194419, avr. 2020, doi: 10.1016/j.bbagr.2019.194419.
- [211] M. D. Paraskevopoulou et A. G. Hatzigeorgiou, « Analyzing MiRNA–lncRNA Interactions », in *Long Non-Coding RNAs: Methods and Protocols*, Y. Feng et L. Zhang, Éd. New York, NY: Springer, 2016, p. 271-286.

- [212] Y. Huang, « The novel regulatory role of lncRNA-miRNA-mRNA axis in cardiovascular diseases », *J. Cell. Mol. Med.*, vol. 22, n° 12, p. 5768-5775, déc. 2018, doi: 10.1111/jcmm.13866.
- [213] M. Guttman *et al.*, « Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals », *Nature*, vol. 458, n° 7235, Art. n° 7235, mars 2009, doi: 10.1038/nature07672.
- [214] M. Huarte *et al.*, « A Large Intergenic Noncoding RNA Induced by p53 Mediates Global Gene Repression in the p53 Response », *Cell*, vol. 142, n° 3, p. 409-419, août 2010, doi: 10.1016/j.cell.2010.06.040.
- [215] D. Gebert, C. Hewel, et D. Rosenkranz, « unitas: the universal tool for annotation of small RNAs », *BMC Genomics*, vol. 18, n° 1, déc. 2017, doi: 10.1186/s12864-017-4031-9.
- [216] V. Ambros, « A uniform system for microRNA annotation », *RNA*, vol. 9, n° 3, p. 277-279, mars 2003, doi: 10.1261/rna.2183803.
- [217] K. V. Bandara, M. Z. Michael, et J. M. Gleadle, « MicroRNA Biogenesis in Hypoxia », *MicroRNA*, vol. 6, n° 2, août 2017, doi: 10.2174/2211536606666170313114821.
- [218] A. Grimson, K. K.-H. Farh, W. K. Johnston, P. Garrett-Engele, L. P. Lim, et D. P. Bartel, « MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing », *Mol. Cell*, vol. 27, n° 1, p. 91-105, juill. 2007, doi: 10.1016/j.molcel.2007.06.017.
- [219] R. C. Lee, R. L. Feinbaum, et V. Ambros, « The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14* », *Cell*, vol. 75, n° 5, p. 843-854, déc. 1993, doi: 10.1016/0092-8674(93)90529-Y.
- [220] B. J. Reinhart *et al.*, « The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans* », *Nature*, vol. 403, n° 6772, p. 901-906, févr. 2000, doi: 10.1038/35002607.
- [221] Y. Lee *et al.*, « The nuclear RNase III Drosha initiates microRNA processing », *Nature*, vol. 425, n° 6956, Art. n° 6956, sept. 2003, doi: 10.1038/nature01957.
- [222] G. Hutvagner, J. McLachlan, A. E. Pasquinelli, É. Bálint, T. Tuschl, et P. D. Zamore, « A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the *let-7* Small Temporal RNA », *Science*, vol. 293, n° 5531, p. 834-838, août 2001, doi: 10.1126/science.1062961.
- [223] M. K. Jena, « MicroRNAs in the development and neoplasia of the mammary gland », *F1000Research*, vol. 6, oct. 2017, doi: 10.12688/f1000research.12005.2.
- [224] A. Kozomara, M. Birgaoanu, et S. Griffiths-Jones, « miRBase: from microRNA sequences to function », *Nucleic Acids Res.*, nov. 2018, doi: 10.1093/nar/gky1141.
- [225] Le Guillou *et al.*, « Characterisation and Comparison of Lactating Mouse and Bovine Mammary Gland miRNomes », *PLoS ONE*, vol. 9, n° 3, p. e91938, mars 2014, doi: 10.1371/journal.pone.0091938.
- [226] Le Guillou *et al.*, « Characterization of Holstein and Normande whole milk miRNomes highlights breed specificities », *Sci. Rep.*, vol. 9, déc. 2019, doi: 10.1038/s41598-019-56690-7.
- [227] P. A. Billa *et al.*, « Deep RNA-Seq reveals miRNome differences in mammary tissue of lactating Holstein and Montbéliarde cows », *BMC Genomics*, vol. 20, n° 1, p. 621, juill. 2019, doi: 10.1186/s12864-019-5987-4.

- [228] C. Mooney, B. A. Becker, R. Raoof, et D. C. Henshall, « EpimiRBase: a comprehensive database of microRNA-epilepsy associations », *Bioinformatics*, vol. 32, n° 9, p. 1436-1438, mai 2016, doi: 10.1093/bioinformatics/btw008.
- [229] B. Xie, Q. Ding, H. Han, et D. Wu, « miRCancer: a microRNA-cancer association database constructed by text mining on literature », *Bioinformatics*, vol. 29, n° 5, p. 638-644, mars 2013, doi: 10.1093/bioinformatics/btt014.
- [230] M. Yue *et al.*, « MSDD: a manually curated database of experimentally supported associations among miRNAs, SNPs and human diseases », *Nucleic Acids Res.*, vol. 46, n° D1, p. D181-D185, janv. 2018, doi: 10.1093/nar/gkx1035.
- [231] M. W. Szcześniak et I. Makołowska, « miRNEST 2.0: a database of plant and animal microRNAs », *Nucleic Acids Res.*, vol. 42, n° D1, p. D74-D77, janv. 2014, doi: 10.1093/nar/gkt1156.
- [232] D. Gerlach, E. V. Kriventseva, N. Rahman, C. E. Vejnar, et E. M. Zdobnov, « miROrtho: computational survey of microRNA genes », *Nucleic Acids Res.*, vol. 37, n° Database, p. D111-D117, janv. 2009, doi: 10.1093/nar/gkn707.
- [233] C. Backes *et al.*, « miRCarta: a central repository for collecting miRNA candidates », *Nucleic Acids Res.*, vol. 46, n° D1, p. D160-D167, janv. 2018, doi: 10.1093/nar/gkx851.
- [234] L. Chen, L. Heikkinen, C. Wang, Y. Yang, K. E. Knott, et G. Wong, « miRToolsGallery: a tag-based and rankable microRNA bioinformatics resources database portal », *Database*, vol. 2018, janv. 2018, doi: 10.1093/database/bay004.
- [235] L. P. Lim *et al.*, « Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs », *Nature*, vol. 433, n° 7027, p. 769-773, févr. 2005, doi: 10.1038/nature03315.
- [236] V. N. Kim, J. Han, et M. C. Siomi, « Biogenesis of small RNAs in animals », *Nat. Rev. Mol. Cell Biol.*, vol. 10, n° 2, p. 126-139, févr. 2009, doi: 10.1038/nrm2632.
- [237] B. P. Lewis, I. Shih, M. W. Jones-Rhoades, D. P. Bartel, et C. B. Burge, « Prediction of Mammalian MicroRNA Targets », *Cell*, vol. 115, n° 7, p. 787-798, déc. 2003, doi: 10.1016/S0092-8674(03)01018-3.
- [238] A. Mohr et J. Mott, « Overview of MicroRNA Biology », *Semin. Liver Dis.*, vol. 35, n° 01, p. 003-011, janv. 2015, doi: 10.1055/s-0034-1397344.
- [239] H. Y. Jin et C. Xiao, « MicroRNA Mechanisms of Action: What have We Learned from Mice? », *Front. Genet.*, vol. 6, nov. 2015, doi: 10.3389/fgene.2015.00328.
- [240] A. Brümmer et J. Hausser, « MicroRNA binding sites in the coding region of mRNAs: Extending the repertoire of post-transcriptional gene regulation », *BioEssays*, vol. 36, n° 6, p. 617-626, 2014, doi: 10.1002/bies.201300104.
- [241] S. W. Chi, J. B. Zang, A. Mele, et R. B. Darnell, « Ago HITS-CLIP decodes miRNA-mRNA interaction maps », *Nature*, vol. 460, n° 7254, p. 479-486, juill. 2009, doi: 10.1038/nature08170.
- [242] D. Jevsinek Skok *et al.*, « Genome-wide *in silico* screening for microRNA genetic variability in livestock species », *Anim. Genet.*, vol. 44, n° 6, p. 669-677, déc. 2013, doi: 10.1111/age.12072.
- [243] P. Landgraf *et al.*, « A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing », *Cell*, vol. 129, n° 7, p. 1401-1414, juin 2007, doi: 10.1016/j.cell.2007.04.040.

- [244] N. Iwai et H. Naraba, « Polymorphisms in human pre-miRNAs », *Biochem. Biophys. Res. Commun.*, vol. 331, n° 4, p. 1439-1444, juin 2005, doi: 10.1016/j.bbrc.2005.04.051.
- [245] H. Ding, Z. Lv, Y. Yuan, et Q. Xu, « MiRNA Polymorphisms and Cancer Prognosis: A Systematic Review and Meta-Analysis », *Front. Oncol.*, vol. 8, déc. 2018, doi: 10.3389/fonc.2018.00596.
- [246] J. Ji, E. Cha, et W. Lee, « Association of miR-146a polymorphisms with systemic lupus erythematosus: a meta-analysis », *Lupus*, vol. 23, n° 10, p. 1023-1030, sept. 2014, doi: 10.1177/0961203314534512.
- [247] H. K. T. Trinh, D. L. Pham, S.-C. Kim, R.-Y. Kim, H.-S. Park, et S.-H. Kim, « Association of the miR-196a2, miR-146a, and miR-499 Polymorphisms with Asthma Phenotypes in a Korean Population », *Mol. Diagn. Ther.*, vol. 21, n° 5, p. 547-554, oct. 2017, doi: 10.1007/s40291-017-0280-1.
- [248] Y. Yuan et J. B. Weidhaas, « Functional microRNA binding site variants », *Mol. Oncol.*, vol. 13, n° 1, p. 4-8, janv. 2019, doi: 10.1002/1878-0261.12421.
- [249] M. A. Saunders, H. Liang, et W.-H. Li, « Human polymorphism at microRNAs and microRNA target sites », *Proc. Natl. Acad. Sci.*, vol. 104, n° 9, p. 3300-3305, févr. 2007, doi: 10.1073/pnas.0611347104.
- [250] S. Parchami Barjui, S. Reisi, S. O. Ebrahimi, et B. Shekari, « Study of correlation between genetic variants in three microRNA genes (hsa-miR-146a, hsa-miR-502 binding site, hsa-miR-27a) and breast cancer risk », *Curr. Res. Transl. Med.*, vol. 65, n° 4, p. 141-147, nov. 2017, doi: 10.1016/j.retram.2017.10.001.
- [251] Z. Gu, S. Eleswarapu, et H. Jiang, « Identification and characterization of microRNAs from the bovine adipose tissue and mammary gland », *FEBS Lett.*, vol. 581, n° 5, p. 981-988, 2007, doi: 10.1016/j.febslet.2007.01.081.
- [252] Le Guillou *et al.*, « Overexpression of miR-30b in the Developing Mouse Mammary Gland Causes a Lactation Defect and Delays Involution », *PLoS ONE*, vol. 7, n° 9, sept. 2012, doi: 10.1371/journal.pone.0045727.
- [253] H. Wang *et al.*, « MiR-145 Regulates Lipogenesis in Goat Mammary Cells Via Targeting INSIG1 and Epigenetic Regulation of Lipid-Related Genes », *J. Cell. Physiol.*, vol. 232, n° 5, p. 1030-1040, 2017, doi: 10.1002/jcp.25499.
- [254] H. Wang *et al.*, « MicroRNA-24 can control triacylglycerol synthesis in goat mammary epithelial cells by targeting the fatty acid synthase gene », *J. Dairy Sci.*, vol. 98, n° 12, p. 9001-9014, déc. 2015, doi: 10.3168/jds.2015-9418.
- [255] X. Lin, J. Luo, L. Zhang, W. Wang, et D. Gou, « MiR-103 Controls Milk Fat Accumulation in Goat (*Capra hircus*) Mammary Gland during Lactation », *PLOS ONE*, vol. 8, n° 11, p. e79258, nov. 2013, doi: 10.1371/journal.pone.0079258.
- [256] Z. Chen *et al.*, « miR-30e-5p and miR-15a Synergistically Regulate Fatty Acid Metabolism in Goat Mammary Epithelial Cells via LRP6 and YAP1 », *Int. J. Mol. Sci.*, vol. 17, n° 11, Art. n° 11, nov. 2016, doi: 10.3390/ijms17111909.
- [257] H.-M. Li, C.-M. Wang, Q.-Z. Li, et X.-J. Gao, « MiR-15a Decreases Bovine Mammary Epithelial Cell Viability and Lactation and Regulates Growth Hormone Receptor Expression », *Molecules*, vol. 17, n° 10, p. 12037-12048, oct. 2012, doi: 10.3390/molecules171012037.

- [258] Z. Li, H. Liu, X. Jin, L. Lo, et J. Liu, « Expression profiles of microRNAs from lactating and non-lactating bovine mammary glands and identification of miRNA related to lactation », *BMC Genomics*, vol. 13, p. 731, déc. 2012, doi: 10.1186/1471-2164-13-731.
- [259] D. Li *et al.*, « MiR-486 Regulates Lactation and Targets the PTEN Gene in Cow Mammary Glands », *PLoS ONE*, vol. 10, n° 3, mars 2015, doi: 10.1371/journal.pone.0118284.
- [260] Y. Bian *et al.*, « Epigenetic Regulation of miR-29s Affects the Lactation Activity of Dairy Cow Mammary Epithelial Cells », *J. Cell. Physiol.*, vol. 230, n° 9, p. 2152-2163, 2015, doi: 10.1002/jcp.24944.
- [261] S. Lian, J. R. Guo, X. M. Nan, L. Ma, J. J. Loor, et D. P. Bu, « MicroRNA Bta-miR-181a regulates the biosynthesis of bovine milk fat by targeting ACSL1 », *J. Dairy Sci.*, vol. 99, n° 5, p. 3916-3924, mai 2016, doi: 10.3168/jds.2015-10484.
- [262] B. Shen *et al.*, « Deep Sequencing and Screening of Differentially Expressed MicroRNAs Related to Milk Fat Metabolism in Bovine Primary Mammary Epithelial Cells », *Int. J. Mol. Sci.*, vol. 17, n° 12, p. 200, févr. 2016, doi: 10.3390/ijms17020200.
- [263] P. Jiao *et al.*, « PRL/microRNA-183/IRS1 Pathway Regulates Milk Fat Metabolism in Cow Mammary Epithelial Cells », *Genes*, vol. 11, n° 2, févr. 2020, doi: 10.3390/genes11020196.
- [264] D. Llobet-Navas *et al.*, « The miR-424(322)/503 cluster orchestrates remodeling of the epithelium in the involuting mammary gland », *Genes Dev.*, vol. 28, n° 7, p. 765-782, avr. 2014, doi: 10.1101/gad.237404.114.
- [265] X. Chen *et al.*, « Identification and characterization of microRNAs in raw milk during different periods of lactation, commercial fluid, and powdered milk products », *Cell Res.*, vol. 20, n° 10, p. 1128-1137, oct. 2010, doi: 10.1038/cr.2010.80.
- [266] H. Izumi, N. Kosaka, T. Shimizu, K. Sekine, T. Ochiya, et M. Takase, « Bovine milk contains microRNA and messenger RNA that are stable under degradative conditions », *J. Dairy Sci.*, vol. 95, n° 9, p. 4831-4841, sept. 2012, doi: 10.3168/jds.2012-5489.
- [267] I. Van Hese, K. Goossens, L. Vandaele, et G. Opsomer, « Invited review: MicroRNAs in bovine colostrum—Focus on their origin and potential health benefits for the calf », *J. Dairy Sci.*, vol. 103, n° 1, p. 1-15, janv. 2020, doi: 10.3168/jds.2019-16959.
- [268] M. Wang, S. Moisés, M. J. Khan, J. Wang, D. Bu, et J. J. Loor, « MicroRNA expression patterns in the bovine mammary gland are affected by stage of lactation1 », *J. Dairy Sci.*, vol. 95, n° 11, p. 6529-6535, nov. 2012, doi: 10.3168/jds.2012-5748.
- [269] J. Ogorevc, T. Kunej, A. Razpet, et P. Dovc, « Database of cattle candidate genes and genetic markers for milk production and mastitis », *Anim. Genet.*, vol. 40, n° 6, p. 832-851, déc. 2009, doi: 10.1111/j.1365-2052.2009.01921.x.
- [270] S. Özdemir, « Expression Profiles of Inflammation-related MicroRNAs in Mycoplasma bovis Infected Milk of Holstein-Friesian and Doğu Anadolu Kırmızısı Cows », *Kahramanmaraş Sütçü İmam Üniversitesi Tarım Ve Doğa Derg.*, vol. 23, n° 3, Art. n° 3, juin 2020, doi: 10.18016/ksutarimdoga.vi.661708.
- [271] M. Chu *et al.*, « MicroRNA-126 participates in lipid metabolism in mammary epithelial cells », *Mol. Cell. Endocrinol.*, vol. 454, p. 77-86, oct. 2017, doi: 10.1016/j.mce.2017.05.039.

- [272] C. A. Morris *et al.*, « Fatty acid synthase effects on bovine adipose fat and milk fat », *Mamm. Genome*, vol. 18, n° 1, p. 64-74, janv. 2007, doi: 10.1007/s00335-006-0102-y.
- [273] R. E. Heinz *et al.*, « Constitutive expression of microRNA-150 in mammary epithelium suppresses secretory activation and impairs de novo lipogenesis », *Dev. Camb. Engl.*, vol. 143, n° 22, p. 4236-4248, nov. 2016, doi: 10.1242/dev.139642.
- [274] X. Lin, J. Luo, L. Zhang, W. Wang, et D. Gou, « MiR-103 Controls Milk Fat Accumulation in Goat (*Capra hircus*) Mammary Gland during Lactation », *PLoS ONE*, vol. 8, n° 11, nov. 2013, doi: 10.1371/journal.pone.0079258.
- [275] S. Chen *et al.*, « 5-Hydroxy-l-tryptophan Promotes the Milk Calcium Level via the miR-99a-3p/ATP2B1 Axis in Goat Mammary Epithelial Cells », *J. Agric. Food Chem.*, vol. 68, n° 10, p. 3277-3285, mars 2020, doi: 10.1021/acs.jafc.9b07869.
- [276] G. L. Kaselonis, E. R. B. McCabe, et S. M. Gray, « Expression of Hexokinase 1 and Hexokinase 2 in Mammary Tissue of Nonlactating and Lactating Rats: Evaluation by RT-PCR », *Mol. Genet. Metab.*, vol. 68, n° 3, p. 371-374, nov. 1999, doi: 10.1006/mgme.1999.2923.
- [277] X. Liu, G. W. Robinson, K. U. Wagner, L. Garrett, A. Wynshaw-Boris, et L. Hennighausen, « Stat5a is mandatory for adult mammary gland development and lactogenesis. », *Genes Dev.*, vol. 11, n° 2, p. 179-186, janv. 1997, doi: 10.1101/gad.11.2.179.
- [278] M. Reichenstein, G. Rauner, et I. Barash, « Conditional repression of STAT5 expression during lactation reveals its exclusive roles in mammary gland morphology, milk-protein gene expression, and neonate growth », *Mol. Reprod. Dev.*, vol. 78, n° 8, p. 585-596, août 2011, doi: 10.1002/mrd.21345.
- [279] S. A. Bhat *et al.*, « Comparative transcriptome analysis of mammary epithelial cells at different stages of lactation reveals wide differences in gene expression and pathways regulating milk synthesis between Jersey and Kashmiri cattle », *PLOS ONE*, vol. 14, n° 2, p. e0211773, févr. 2019, doi: 10.1371/journal.pone.0211773.
- [280] S. A. Burgos, M. Dai, et J. P. Cant, « Nutrient availability and lactogenic hormones regulate mammary protein synthesis through the mammalian target of rapamycin signaling pathway », *J. Dairy Sci.*, vol. 93, n° 1, p. 153-161, janv. 2010, doi: 10.3168/jds.2009-2444.
- [281] P. Jiao *et al.*, « PRL/microRNA-183/IRS1 Pathway Regulates Milk Fat Metabolism in Cow Mammary Epithelial Cells », *Genes*, vol. 11, n° 2, Art. n° 2, févr. 2020, doi: 10.3390/genes11020196.
- [282] A. Zidi *et al.*, « Short communication: Genetic variability in the predicted microRNA target sites of caprine casein genes », *J. Dairy Sci.*, vol. 93, n° 4, p. 1749-1753, avr. 2010, doi: 10.3168/jds.2009-2741.
- [283] Z. Ju *et al.*, « The effect of the SNP g.18475 A>G in the 3'UTR of NCF4 on mastitis susceptibility in dairy cattle », *Cell Stress Chaperones*, vol. 23, n° 3, p. 385-391, mai 2018, doi: 10.1007/s12192-017-0848-z.
- [284] X. G. Wang *et al.*, « Regulatory mutations in the A2M gene are involved in the mastitis susceptibility in dairy cows », *Anim. Genet.*, vol. 45, n° 1, p. 28-37, févr. 2014, doi: 10.1111/age.12099.

- [285] E. M. Ibeagha-Awemu, K. A. Akwanji, F. Beaudoin, et X. Zhao, « Associations between variants of FADS genes and omega-3 and omega-6 milk fatty acids of Canadian Holstein cows », *BMC Genet.*, vol. 15, p. 25, févr. 2014, doi: 10.1186/1471-2156-15-25.
- [286] X. Wang, T. Li, H. B. Zhao, et H. Khatib, « Short communication: A mutation in the 3' untranslated region diminishes microRNA binding and alters expression of the OLR1 gene », *J. Dairy Sci.*, vol. 96, p. 6525-6528, oct. 2013, doi: 10.3168/jds.2013-6873.
- [287] A. Mahmoudi, A. Zargaran, H.-R. Amini, A. Assadi, R. Vajdi Hokmabad, et S. Eghbalsaied, « A SNP in the 3'-untranslated region of AMPK γ 1 may associate with serum ketone body and milk production of Holstein dairy cows », *Gene*, vol. 574, n° 1, p. 48-52, déc. 2015, doi: 10.1016/j.gene.2015.07.077.
- [288] S. Griffiths-Jones, « miRBase: microRNA sequences, targets and gene nomenclature », *Nucleic Acids Res.*, vol. 34, n° 90001, p. D140-D144, janv. 2006, doi: 10.1093/nar/gkj112.
- [289] A. Kozomara, M. Birgaoanu, et S. Griffiths-Jones, « miRBase: from microRNA sequences to function », *Nucleic Acids Res.*, vol. 47, n° D1, p. D155-D162, janv. 2019, doi: 10.1093/nar/gky1141.
- [290] C. Bourdon, P. Bardou, E. Aujean, S. Le Guillou, G. Tosser-Klopp, et F. Le Provost, « RumimiR: a detailed microRNA database focused on ruminant species », *Database*, vol. 2019, janv. 2019, doi: 10.1093/database/baz099.
- [291] M.-P. Sanchez *et al.*, « Sequence-based GWAS, network and pathway analyses reveal genes co-associated with milk cheese-making properties and milk composition in Montbéliarde cows », *Genet. Sel. Evol. GSE*, vol. 51, juill. 2019, doi: 10.1186/s12711-019-0473-7.
- [292] W. S. Bush et J. H. Moore, « Chapter 11: Genome-Wide Association Studies », *PLoS Comput. Biol.*, vol. 8, n° 12, déc. 2012, doi: 10.1371/journal.pcbi.1002822.
- [293] J. Laubier, J. Castille, S. Le Guillou, et F. Le Provost, « No effect of an elevated miR-30b level in mouse milk on its level in pup tissues », *RNA Biol.*, vol. 12, n° 1, p. 26-29, janv. 2015, doi: 10.1080/15476286.2015.1017212.
- [294] B. Zavizion, M. van Duffelen, W. Schaeffer, et I. Politis, « Establishment and characterization of a bovine mammary epithelial cell line with unique properties », *Vitro Cell. Dev. Biol. - Anim.*, vol. 32, n° 3, p. 138-148, mars 1996, doi: 10.1007/BF02723679.
- [295] L. Mobuchon, « Impact des microARNs sur la lactation et la régulation nutritionnelle de leur expression dans la glande mammaire », Université Blaise Pascal - Clermont-Ferrand, 2015.
- [296] ThermoFisher Scientific, « Factors Influencing Transfection Efficiency - FR ». -, Consulté le: oct. 20, 2020. [En ligne]. Disponible sur: [//www.thermofisher.com/fr/fr/home/references/gibco-cell-culture-basics/transfection-basics/factors-influencing-transfection-efficiency.html](http://www.thermofisher.com/fr/fr/home/references/gibco-cell-culture-basics/transfection-basics/factors-influencing-transfection-efficiency.html).
- [297] E. J. Wallenstein, J. Barminko, R. S. Schloss, et M. L. Yarmush, « Serum Starvation Improves Transient Transfection Efficiency in Differentiating Embryonic Stem Cells », *Biotechnol. Prog.*, vol. 26, n° 6, p. 1714-1723, 2010, doi: 10.1002/btpr.472.
- [298] S. Bougarn, P. Cunha, F. B. Gilbert, F. Meurens, et P. Rainard, « Technical note: Validation of candidate reference genes for normalization of quantitative PCR in

- bovine mammary epithelial cells responding to inflammatory stimuli », *J. Dairy Sci.*, vol. 94, n° 5, p. 2425-2430, mai 2011, doi: 10.3168/jds.2010-3859.
- [299] T. D. Schmittgen et K. J. Livak, « Analyzing real-time PCR data by the comparative C(T) method », *Nat. Protoc.*, vol. 3, n° 6, p. 1101-1108, 2008, doi: 10.1038/nprot.2008.73.
- [300] A. Dobin *et al.*, « STAR: ultrafast universal RNA-seq aligner », *Bioinformatics*, vol. 29, n° 1, p. 15-21, janv. 2013, doi: 10.1093/bioinformatics/bts635.
- [301] « Ensembl version 81 release | European Bioinformatics Institute », 2015. <https://www.ebi.ac.uk/about/news/service-news/ensembl-version-81-release> (consulté le oct. 21, 2020).
- [302] M. Martin, « Cutadapt removes adapter sequences from high-throughput sequencing reads », *EMBnet.journal*, vol. 17, n° 1, p. 10, mai 2011, doi: 10.14806/ej.17.1.200.
- [303] B. Li et C. N. Dewey, « RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome », *BMC Bioinformatics*, vol. 12, n° 1, p. 323, déc. 2011, doi: 10.1186/1471-2105-12-323.
- [304] R Development Core Team, « R: The R Project for Statistical Computing », 2013. <https://www.r-project.org/> (consulté le oct. 21, 2020).
- [305] J. Oksanen *et al.*, *vegan: Community Ecology Package version 2.5-6 from CRAN*. 2019.
- [306] S. Dray et A.-B. Dufour, « The **ade4** Package: Implementing the Duality Diagram for Ecologists », *J. Stat. Softw.*, vol. 22, n° 4, 2007, doi: 10.18637/jss.v022.i04.
- [307] C. R. Rao, « The Use and Interpretation of Principal Component Analysis in Applied Research », *Sankhyā Indian J. Stat. Ser. 1961-2002*, vol. 26, n° 4, p. 329-358, 1964.
- [308] G. K. Smyth, « limma: Linear Models for Microarray Data », in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, R. Gentleman, V. J. Carey, W. Huber, R. A. Irizarry, et S. Dudoit, Éd. New York, NY: Springer, 2005, p. 397-420.
- [309] C. W. Law, Y. Chen, W. Shi, et G. K. Smyth, « voom: precision weights unlock linear model analysis tools for RNA-seq read counts », *Genome Biol.*, vol. 15, n° 2, p. R29, févr. 2014, doi: 10.1186/gb-2014-15-2-r29.
- [310] W. Haynes, « Benjamini–Hochberg Method », in *Encyclopedia of Systems Biology*, W. Dubitzky, O. Wolkenhauer, K.-H. Cho, et H. Yokota, Éd. New York, NY: Springer, 2013, p. 78-78.
- [311] Y. C. Lai *et al.*, « Rapid Communication: MiR-92a as a housekeeping gene for analysis of bovine mastitis-related microRNA in milk », *J. Anim. Sci.*, vol. 95, n° 6, p. 2732-2735, juin 2017, doi: 10.2527/jas.2017.1384.
- [312] J. T. Bensen, C. K. Tse, S. J. Nyante, J. S. Barnholtz-Sloan, S. R. Cole, et R. C. Millikan, « Association of germline microRNA SNPs in pre-miRNA flanking region and breast cancer risk and survival: the Carolina Breast Cancer Study », *Cancer Causes Control*, vol. 24, n° 6, p. 1099-1109, juin 2013, doi: 10.1007/s10552-013-0187-z.
- [313] M. Bastami *et al.*, « Evidences from a Systematic Review and Meta-Analysis Unveil the Role of MiRNA Polymorphisms in the Predisposition to Female Neoplasms », *Int. J. Mol. Sci.*, vol. 20, n° 20, oct. 2019, doi: 10.3390/ijms20205088.
- [314] Z. Hu *et al.*, « Genetic variants of miRNA sequences and non-small cell lung cancer survival », *J. Clin. Invest.*, juin 2008, doi: 10.1172/JCI34934.
- [315] M. Alipour, M. Abtin, A. Hosseinzadeh, et M. Maleki, « Association between miR-146a C > G, miR-149 T > C, miR-196a2 T > C, and miR-499 A > G polymorphisms and

- susceptibility to idiopathic recurrent pregnancy loss », *J. Assist. Reprod. Genet.*, vol. 36, n° 11, p. 2237-2244, nov. 2019, doi: 10.1007/s10815-019-01573-z.
- [316] P. Yan *et al.*, « Predictive role of miR-146a rs2910164 (C>G), miR-149 rs2292832 (T>C), miR-196a2 rs11614913 (T>C) and miR-499 rs3746444 (T>C) in the development of hepatocellular carcinoma », *Int. J. Clin. Exp. Pathol.*, vol. 8, n° 11, p. 15177-15183, nov. 2015.
- [317] J.-H. Sung *et al.*, « miRNA polymorphisms (miR-146a, miR-149, miR-196a2 and miR-499) are associated with the risk of coronary artery disease », *Mol. Med. Rep.*, vol. 14, n° 3, p. 2328-2342, sept. 2016, doi: 10.3892/mmr.2016.5495.
- [318] Y. Xiao, H. Liu, L. Chen, Y. Wang, X. Yao, et X. Jiang, « Association of microRNAs genes polymorphisms with arthritis: a systematic review and meta-analysis », *Biosci. Rep.*, vol. 39, n° 7, juill. 2019, doi: 10.1042/BSR20190298.
- [319] M. Bastami *et al.*, « miRNA Polymorphisms and Risk of Cardio-Cerebrovascular Diseases: A Systematic Review and Meta-Analysis », *Int. J. Mol. Sci.*, vol. 20, n° 2, janv. 2019, doi: 10.3390/ijms20020293.
- [320] Z.-J. Dai *et al.*, « Five Common Functional Polymorphisms in microRNAs (rs2910164, rs2292832, rs11614913, rs3746444, rs895819) and the Susceptibility to Breast Cancer: Evidence from 8361 Cancer Cases and 8504 Controls », *Curr. Pharm. Des.*, vol. 21, n° 11, p. 1455-1463, févr. 2015, doi: 10.2174/1381612821666141208143533.
- [321] M. Zorc *et al.*, « Catalog of microRNA seed polymorphisms in vertebrates », *PLoS One*, vol. 7, n° 1, p. e30737, 2012, doi: 10.1371/journal.pone.0030737.
- [322] M. Zorc, J. Obsteter, P. Dovc, et T. Kunej, « Genetic Variability of MicroRNA Genes in 15 Animal Species », *J. Genomics*, vol. 3, p. 51-56, 2015, doi: 10.7150/jgen.11246.
- [323] A. Siepel *et al.*, « Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes », *Genome Res.*, vol. 15, n° 8, p. 1034-1050, janv. 2005, doi: 10.1101/gr.3715005.
- [324] L. A. Pennacchio et E. M. Rubin, « Genomic strategies to identify mammalian regulatory sequences », *Nat. Rev. Genet.*, vol. 2, n° 2, Art. n° 2, févr. 2001, doi: 10.1038/35052548.
- [325] C. Pecorini, D. Sasserà, R. Rebutti, F. Saccone, C. Bandi, et A. Baldi, « Evaluation of the protective effect of bovine lactoferrin against lipopolysaccharides in a bovine mammary epithelial cell line », *Vet. Res. Commun.*, vol. 34, n° 3, p. 267-276, mars 2010, doi: 10.1007/s11259-010-9351-z.
- [326] L. Mobuchon *et al.*, « Sunflower oil supplementation affects the expression of miR-20a-5p and miR-142-5p in the lactating bovine mammary gland », *PLoS ONE*, vol. 12, n° 12, déc. 2017, doi: 10.1371/journal.pone.0185511.
- [327] B. Zavizion, M. van Duffelen, W. Schaeffer, et I. Politis, « Establishment and characterization of a bovine mammary epithelial cell line with unique properties », *Vitro Cell. Dev. Biol. - Anim.*, vol. 32, n° 3, p. 138-148, mars 1996, doi: 10.1007/BF02723679.
- [328] M. Schwerin *et al.*, « Application of disease-associated differentially expressed genes – Mining for functional candidate genes for mastitis resistance in cattle », *Genet. Sel. Evol. GSE*, vol. 35, n° Suppl 1, p. S19-S34, juin 2003, doi: 10.1186/1297-9686-35-S1-S19.

- [329] N. A. and F. U. Lang Xianzheng et N. A. and F. U. Liu Xiaolin, « Polymorphisms of OSTF1 gene and their relationship with somatic cell score in Holstein cows », *J. Huazhong Agric. Univ.*, 2011, Consulté le: août 30, 2019. [En ligne]. Disponible sur: http://agris.fao.org/agris-search/search.do;jsessionid=F46F97C133B7D75D6260D857A04D6C82?request_locale=ru&recordID=CN2012002952&query=&sourceQuery=&sortField=&sortOrder=&agrovocString=&advQuery=¢erString=&enableField=.
- [330] T. Urano, M. Shiraki, N. Sasaki, Y. Ouchi, et S. Inoue, « SLC25A24 as a Novel Susceptibility Gene for Low Fat Mass in Humans and Mice », *J. Clin. Endocrinol. Metab.*, vol. 100, n° 4, p. E655-E663, avr. 2015, doi: 10.1210/jc.2014-2829.
- [331] V. L. T. Hoang *et al.*, « RNA-seq reveals more consistent reference genes for gene expression studies in human non-melanoma skin cancers », *PeerJ*, vol. 5, p. e3631, août 2017, doi: 10.7717/peerj.3631.
- [332] C. Costa, A. Giménez-Capitán, N. Karachaliou, et R. Rosell, « Comprehensive molecular screening: from the RT-PCR to the RNA-seq », *Transl. Lung Cancer Res.*, vol. 2, n° 2, p. 87-91, avr. 2013, doi: 10.3978/j.issn.2218-6751.2013.02.05.
- [333] Y. Shi et M. He, « Differential gene expression identified by RNA-Seq and qPCR in two sizes of pearl oyster (*Pinctada fucata*) », *Gene*, vol. 538, n° 2, p. 313-322, avr. 2014, doi: 10.1016/j.gene.2014.01.031.
- [334] A. Azhir, L.-H. Merino, et D. W. Nauen, « Interactive web-based identification and visualization of transcript shared sequences », *Genomics*, vol. 111, n° 4, p. 860-862, 2019, doi: 10.1016/j.ygeno.2018.05.010.
- [335] N. J. Schurch *et al.*, « How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? », *RNA*, vol. 22, n° 6, p. 839-851, juin 2016, doi: 10.1261/rna.053959.115.
- [336] E. Chirshv, K. C. Oberg, Y. J. Ioffe, et J. J. Unternaehrer, « Let-7 as biomarker, prognostic indicator, and therapy for precision medicine in cancer », *Clin. Transl. Med.*, vol. 8, n° 1, p. 24, août 2019, doi: 10.1186/s40169-019-0240-y.
- [337] A. M. Gurtan et P. A. Sharp, « The role of miRNAs in regulating gene expression networks », *J. Mol. Biol.*, vol. 425, n° 19, p. 3582-3600, oct. 2013, doi: 10.1016/j.jmb.2013.03.007.
- [338] D. P. Bartel, « MicroRNA Target Recognition and Regulatory Functions », *Cell*, vol. 136, n° 2, p. 215-233, janv. 2009, doi: 10.1016/j.cell.2009.01.002.
- [339] H. Chang, B. Yi, R. Ma, X. Zhang, H. Zhao, et Y. Xi, « CRISPR/cas9, a novel genomic tool to knock down microRNA in vitro and in vivo », *Sci. Rep.*, vol. 6, févr. 2016, doi: 10.1038/srep22312.
- [340] J. Li *et al.*, « CRISPR/Cas9-Mediated miR-29b Editing as a Treatment of Different Types of Muscle Atrophy in Mice », *Mol. Ther.*, vol. 28, n° 5, p. 1359-1372, mai 2020, doi: 10.1016/j.ymthe.2020.03.005.
- [341] Q. Jiang *et al.*, « In silico genome-wide miRNA-QTL-SNPs analyses identify a functional SNP associated with mastitis in Holsteins », *BMC Genet.*, vol. 20, n° 1, p. 46, mai 2019, doi: 10.1186/s12863-019-0749-5.

- [342] K. M. Howard *et al.*, « Loss of miRNAs during Processing and Storage of Cow's (Bos taurus)Milk », *J. Agric. Food Chem.*, vol. 63, n° 2, p. 588-592, janv. 2015, doi: 10.1021/jf505526w.
- [343] W. Sun *et al.*, « A mutation in porcine pre-miR-15b alters the biogenesis of MiR-15b\16-1 cluster and strand selection of MiR-15b », *PLOS ONE*, vol. 12, n° 5, p. e0178045, mai 2017, doi: 10.1371/journal.pone.0178045.
- [344] Z. Gao, L. Sun, K. Dai, Y. Du, S. Wu, et W. Bao, « Effects of mutations in porcine miRNA-215 precursor sequences on miRNA-215 regulatory function », *Gene*, vol. 701, p. 131-138, juin 2019, doi: 10.1016/j.gene.2019.03.044.
- [345] H. Wu *et al.*, « Variants of pri-miR-26a-5p polymorphisms are associated with values for chicken egg production variables and affects abundance of mature miRNA », *Anim. Reprod. Sci.*, vol. 201, p. 93-101, févr. 2019, doi: 10.1016/j.anireprosci.2019.01.002.
- [346] P. Rani, V. R. Yenuganti, S. Shandilya, S. K. Onteru, et D. Singh, « miRNAs: The hidden bioactive component of milk », *Trends Food Sci. Technol.*, vol. 65, p. 94-102, juill. 2017, doi: 10.1016/j.tifs.2017.05.007.
- [347] T. S. Appliedbiosystems, « Appliedbiosystems Product bulletin - TaqMan MicroRNA Assays and Arrays ». [En ligne]. Disponible sur: http://tools.thermofisher.com/content/sfs/brochures/cms_042142.pdf.
- [348] C. Dandine-Roulland et H. Perdry, « Where is the causal variant? On the advantage of the family design over the case–control design in genetic association studies », *Eur. J. Hum. Genet.*, vol. 23, n° 10, p. 1357-1363, oct. 2015, doi: 10.1038/ejhg.2014.284.
- [349] J. Lu, S. Boeren, T. van Hooijdonk, J. Vervoort, et K. Hettinga, « Effect of the DGAT1 K232A genotype of dairy cows on the milk metabolome and proteome », *J. Dairy Sci.*, vol. 98, n° 5, p. 3460-3469, mai 2015, doi: 10.3168/jds.2014-8872.
- [350] P. Martin *et al.*, « A genome scan for milk production traits in dairy goats reveals two new mutations in Dgat1 reducing milk fat content », *Sci. Rep.*, vol. 7, n° 1, déc. 2017, doi: 10.1038/s41598-017-02052-0.
- [351] J. Bennewitz *et al.*, « The DGAT1 K232A Mutation Is Not Solely Responsible for the Milk Production Quantitative Trait Locus on the Bovine Chromosome 14 », *J. Dairy Sci.*, vol. 87, n° 2, p. 431-442, févr. 2004, doi: 10.3168/jds.S0022-0302(04)73182-3.
- [352] S. Yu *et al.*, « LncRNA as ceRNAs may be involved in lactation process », *Oncotarget*, vol. 8, n° 58, p. 98014-98028, nov. 2017, doi: 10.18632/oncotarget.20439.
- [353] G. Pasquini et T. Kunej, « A Map of the microRNA Regulatory Networks Identified by Experimentally Validated microRNA-Target Interactions in Five Domestic Animals: Cattle, Pig, Sheep, Dog, and Chicken », *OMICS J. Integr. Biol.*, p. omi.2019.0082, août 2019, doi: 10.1089/omi.2019.0082.
- [354] B. Yang *et al.*, « Transcriptome sequencing to detect the potential role of long non-coding RNAs in bovine mammary gland during the dry and lactation period », *BMC Genomics*, vol. 19, août 2018, doi: 10.1186/s12864-018-4974-5.
- [355] X. Zheng *et al.*, « Integrated analysis of long noncoding RNA and mRNA expression profiles reveals the potential role of long noncoding RNA in different bovine lactation stages », *J. Dairy Sci.*, vol. 101, n° 12, p. 11061-11073, déc. 2018, doi: 10.3168/jds.2018-14900.

Titre : Recherche d'associations entre microARNs, variants génétiques et QTL laitiers chez les bovins, caprins et ovins

Mots clés : microARNs, variants génétiques, QTL

Résumé : La sélection génomique, qui repose sur la prédiction de la valeur génétique des animaux candidats à la sélection à partir de l'information fournie par de très nombreux marqueurs génétiques en utilisant des marqueurs neutres, est un levier pertinent et pérenne. La recherche des mutations causales et leur intégration dans les évaluations génomiques permettraient un gain de précision important. Il est donc essentiel de mieux caractériser les mutations causales responsables de la variabilité des caractères quantitatifs liés à l'efficacité de production et à la qualité des produits tels que le lait. L'objectif de ce projet a été de rechercher des variants génétiques de microARNs exprimés dans la glande mammaire ou présents dans le lait et situés dans des régions génomiques ayant un effet sur des caractères quantitatifs (QTL) laitiers et mammittes, dans trois espèces de ruminants. La détection de 59124 variants de microARNs exprimés dans la glande mammaire ou présents dans le lait en bovin, 13427 variants en caprin et 4761 en ovin a été permise grâce au

développement d'un script bio-informatique. En bovin, 4679 variants génétiques d'intérêt sont situés dans des QTL laitiers et mammittes et 127 en caprin, aucun en ovin. Trois variants bovins détectés ont été validés grâce à des études de GWAS. Les effets biologiques des variants validés ont été étudiés, avec des stratégies différentes selon la localisation et donc l'effet de la mutation. Dans le cas de la mutation située dans la région "seed" du *bta-let-7e*, le niveau d'expression d'ARNm cibles a été testé. Des résultats non concordants ont été obtenus entre les techniques de qRT-PCR et de RNAseq utilisées. Dans le cas de mutations situées dans les régions flanquantes de *bta-miR-92b* et *bta-miR-486*, la présence de ces microARNs a été mesurée dans le lait. Ces analyses n'ont pas révélé de différence d'expression significative des microARNs selon les génotypes. Ce projet a donc permis d'effectuer une analyse globale de variants de microARNs, de leur détection à leurs effets potentiels.

Title : Association between microRNA genetic variants and dairy QTL in bovine, caprine and ovine species

Keywords : microRNAs, genetic variants, QTL

Abstract : Genomic selection, based on the prediction of the genetic value of candidate animals based on the information provided by a large number of genetic markers using neutral markers, is a relevant and perennial lever. The search for causal mutations and their integration into genomic evaluations would allow a significant gain in precision. It is therefore essential to better characterize the causal mutations responsible for the variability of quantitative traits related to production efficiency and the quality of products such as milk. The objective of this project has been to search for genetic variants of microRNAs expressed in the mammary gland or present in milk and located in genomic regions having an effect on dairy and mastitis quantitative traits (QTL), in three ruminant species. The detection of 59,124 microRNA variants expressed in the mammary gland or present in milk in cattle, 13,427 variants in goats and 4,761 in sheep has been allowed through the development of a bioinformatics script.

In cattle, 4,679 genetic variants of interest have been located in dairy and mastitis QTLs and 127 in goats, none in sheep. Three detected bovine variants have been validated through GWAS studies. The biological effects of the validated variants have been studied, with different strategies depending on the location and thus the putative effect of the mutation. In the case of the mutation located in the "seed" region of *bta-let-7e*, the expression level of targeted mRNAs has been tested. Inconsistent results were obtained between the qRT-PCR and RNAseq techniques used. In the case of mutations located in the flanking regions of *bta-miR-92b* and *bta-miR-486*, the presence of these microRNAs has been measured in bovine milk. These analyses did not reveal any significant difference in the expression of microRNAs between genotypes. This project therefore has allowed a global analysis of microRNA variants, from their detections to their potential effects.

