



PCAA

Programme canadien d'adaptation agricole

Rapport final

**Utilisation de la génomique pour améliorer
la productivité et la santé des troupeaux caprins**

Numéro de projet : 6641

Demandeur: Société des éleveurs de chèvres laitières de race du Québec

Co-demandeur: Ontario Goat

Période: février 2012 à décembre 2013

Préparé par: Sylvie Vermette, Chargée de projet

En collaboration avec Mohsen Jafarikia, CCAP

Laurence Maignel, CCAP

Stefanie Wyss, CCAP

Brian Sullivan, CCAP

Luiz Brito, Université de Guelph

Flavio Schenkel, Université de Guelph

Kevin Weaver, Ontario Goat

Sophie Girouard, SECLRQ

Date de soumission: 30 décembre 2013

« Une partie du financement de ce projet a été fournie par l'entremise des conseils sectoriels du Québec, de l'Ontario et de la Colombie-Britannique qui exécutent le Programme canadien d'adaptation agricole (PCAA) pour le compte d'Agriculture et Agroalimentaire Canada. »

Table des matières

1. DESCRIPTION DU PROJET	3
1.1. <i>Objectif global</i>	3
1.2. <i>Objectifs spécifiques</i>	3
2. RÉSULTATS ET ANALYSES	4
2.1. <i>Échantillonnage des animaux</i>	4
2.2. <i>Génotypage des animaux</i>	5
2.3. <i>Chargement et gestion des données génomiques</i>	6
2.4. <i>Analyses génomiques</i>	7
2.4.1. <i>SNPs analysé</i>	7
2.4.2. <i>Variabilité des SNPs</i>	8
2.4.3. <i>Déséquilibre de liaison et phase de liaison</i>	9
2.4.4. <i>Analyses d'associations (GWAS)</i>	10
2.4.5. <i>Évaluations génomiques</i>	11
3. CONCLUSIONS	13
4. SOMMAIRE DES ACCOMPLISSEMENTS DU PROJET	14
5. ACRONYMES ET DESCRIPTIONS	15
6. ANNEXES	16

1. DESCRIPTION DU PROJET

1.1. Objectif global

Évaluer dans le contexte caprin le potentiel de la sélection génomique pour l'amélioration de la productivité des troupeaux via l'amélioration des performances et de la santé.

1.2. Objectifs spécifiques

Les objectifs spécifiques du projet étaient:

- D'utiliser de nouvelles technologies génomiques (puces à SNP) pour génotyper des chèvres laitières canadiennes.
- D'estimer l'effet des SNPs (Polymorphisme à nucléotide unique) en relation avec les performances des races de chèvres laitières utilisées au Canada.
- De développer les méthodes pour inclure les données génomiques dans l'estimation des valeurs génétiques.
- D'effectuer des évaluations génomiques pilotes pour différents caractères actuellement mesurés dans les exploitations de chèvres laitières.
- De fournir des recommandations concernant l'utilisation de la génomique pour la sélection de chèvres au Canada.

2. RÉSULTATS ET ANALYSES

2.1. Échantillonnage des animaux

Deux conférences téléphoniques se sont tenues en mars et mai 2012 avec le comité consultatif du projet afin de discuter des participants potentiels pour le projet sur la génomique caprine (Annexes 01 & 02). Les étapes afin d'identifier les participants potentiels du projet incluaient l'envoi d'une lettre aux exploitations candidates potentielles en Ontario et de fréquentes communications entre le soumetteur du projet et les éleveurs de chèvres membres de la SECLRQ au Québec. La plupart des participants ont été trouvés assez rapidement. Certaines exploitations ne répondaient pas complètement aux critères ci-dessous et n'ont donc pas été choisies comme troupeaux participants. Les centres d'insémination artificielle en Ontario et en Alberta ont également été contactés et ont fourni de la semence provenant de boucs de race pure. Le Programme canadien pour les ressources génétiques animales d'AAC a également fourni des échantillons d'ADN sur des chèvres spécifiques. Les chèvres ont été choisies selon les informations disponibles dans la base de données GénétiqueCaprine.Ca. Les critères suivants ont été utilisés afin d'identifier les candidats à l'intérieur de chaque troupeau:

- Enregistré auprès de la CGS
- Race pure
- Nombre de données de contrôle laitier
- Nombre de descendants enregistrés
- Nombre de descendants sur le contrôle laitier
- Nombre de descendants ayant des données de conformation
- La priorité a été donnée aux chèvres dans le Top50 de leur race

Liste de participants qui ont fourni des échantillons de tissus:

- AAC à Saskatoon (Programme canadien pour les ressources génétiques animales)
- 16 exploitations au Québec
- 9 exploitations en Ontario
- 2 centres d'insémination artificielle (EastGen et OC Flock)

Entre octobre 2012 et janvier 2013, un total de 1041 échantillons caprins ont été prélevés dans 25 fermes participantes, 2 centres d'IA et au centre AAC pour les ressources génétiques animales à Saskatoon. Les échantillons étaient pour la plupart des fragments de tissus prélevés au niveau de l'oreille (76 %), mais incluaient également des échantillons d'ADN (13 %), des tubes de sang (9 %) et des paillettes de semence (2 %). Tous les échantillons collectés ont été soumis au laboratoire Delta Genomics en Alberta, où ont été réalisés l'extraction d'ADN et le génotypage avec la puce à 50 000 marqueurs SNP.

Le Tableau 1 ci-dessous montre le nombre de sujets prélevés dans les troupeaux participants et centres d'I.A. par race et sexe dans le cadre du projet. L'objectif initial était d'échantillonner 400 chèvres de race Alpine, 400 Saanen, 100 chèvres d'autres races laitières ainsi que 100 chèvres de boucherie, pour un total de 1000 animaux. Le nombre de chèvres prélevées a dépassé l'objectif initial.

Tableau 1 – Nombre total de sujets prélevés par race et sexe

<i>Race</i>	<i>Mâles</i>	<i>Femelles</i>	<i>Total</i>
Alpine	59	367	426
Saanen	55	284	339
LaMancha	12	71	83
Nubienne	24	38	62
Toggenburg	7	50	57
Boer	17	57	74
Total	159	882	1041

2.2. *Génotypage des animaux*



La puce 50K caprine est disponible pour l'industrie caprine depuis la fin de l'année 2011 grâce aux efforts du Consortium international sur le séquençage du génome caprin (IGGC), un organisme international a été créé en 2010 pour regrouper les ressources et les efforts afin de construire un génome de référence pour les chèvres. Ces efforts et la mise en commun des ressources a permis le développement d'une puce SNP 50K caprine, un outil que l'industrie pourrait utiliser comme moyen pour obtenir des informations supplémentaires sur les animaux génotypés.

puce SNP 50K caprine

Source: <http://goatgenome.org>

Dans le cadre du projet, un total de 1,008 puces SNP ont été achetées. Certains échantillons n'ont pas été génotypés en raison d'une quantité insuffisante d'ADN extraite du tissu prélevé ou des problèmes de qualité de l'échantillon à son arrivée au laboratoire. Certains échantillons ont été génotypés mais ensuite exclus des analyses pour cause d'échec à l'étape du génotypage. Au total, 979 chèvres et boucs ont été génotypés avec succès. Le tableau 2 montre le nombre de chèvres génotypées par race et sexe. L'objectif initial était de génotyper 1000 chèvres. Approximativement 98% de cet objectif a été atteint, un très bon résultat si l'on considère que ce fut la première application de cette technologie sur des échantillons caprins au Canada.

Tableau 2 - Chèvres génotypées – Nombre par race et sexe

<i>Race</i>	<i>Mâles</i>	<i>Femelles</i>	<i>Total</i>
Alpine	51	352	403
Saanen	51	267	318
LaMancha	11	70	81
Nubienne	21	33	54
Toggenburg	7	46	53
Boer	17	50	67
Croisés	0	3	3
Total	158	821	979

Les chèvres prélevées dans le cadre de ce projet ont été choisies selon la quantité d'informations (généalogies, données de contrôle laitier et de conformation) disponibles dans la base de données GénétiqueCaprine.ca. Les tableaux 3 et 4 montrent un sommaire des descendants enregistrés et des données de contrôle laitier disponibles sur les chèvres et boucs génotypés avec succès dans le cadre de ce projet.

Tableau 3 – Chèvres laitières génotypées – Bilan de la descendance (nombre de descendants enregistrés)

Race	Mâles					Femelles			
	Descendants enregistrés					Descendants enregistrés			
	0	1-10	11-20	21-50	>50	0	1-5	6-10	>10
Alpine	13	10	6	19	3	120	204	28	0
Saanen	8	14	10	11	8	100	125	30	12
LaMancha	3	1	5	2	0	39	26	4	1
Nubienne	0	4	3	11	3	19	14	0	0
Toggenburg	0	2	1	1	3	6	30	10	0

Tableau 4 - Chèvres laitières génotypées – Bilan des données de contrôle laitier

Race	Nombre de lactations							
	0	1	2	3	4	5	6	7+
Alpine	26	28	98	67	54	36	29	14
Saanen	34	70	57	40	16	23	13	14
LaMancha	7	7	17	37	1	0	1	0
Nubienne	14	9	4	4	2	0	0	0
Toggenburg	23	17	6	0	0	0	0	0

2.3. Chargement et gestion des données génomiques

Les résultats de génotypage ont été fournis sous la forme de fichiers par Delta Genomics et téléchargés à partir d'un site sécurisé. Ces résultats ont été par la suite chargés dans la base de données GénétiqueCaprine.ca. Avec plus de 50,000 résultats SNPs pour chacun des 979 animaux génotypés, plus de 52 millions de données (52,226,713 pour être exact) ont été chargées dans des tables spécifiques de la base de données. Les résultats de génotypage ont été liés à l'identifiant de l'animal dans les fichiers de Delta Genomics, selon les numéros d'enregistrement fournis par les éleveurs participants pour la plupart des animaux enregistrés. Cela a permis que les données génomiques soient liées au bon animal dans la base de données nationale, qui inclut également les généalogies, les données de performance ainsi que les valeurs génétiques.

Le site GénétiqueCaprine.ca comprend des outils de recherche qui permettent de chercher des animaux spécifiques en utilisant les 'Chèvres sur le Web'. Pour les animaux génotypés dans le cadre du projet, une icône spécifique a été créée afin qu'ils puissent être facilement identifiés lors de recherches dans 'Les Chèvres sur le Web' (Figure 2).

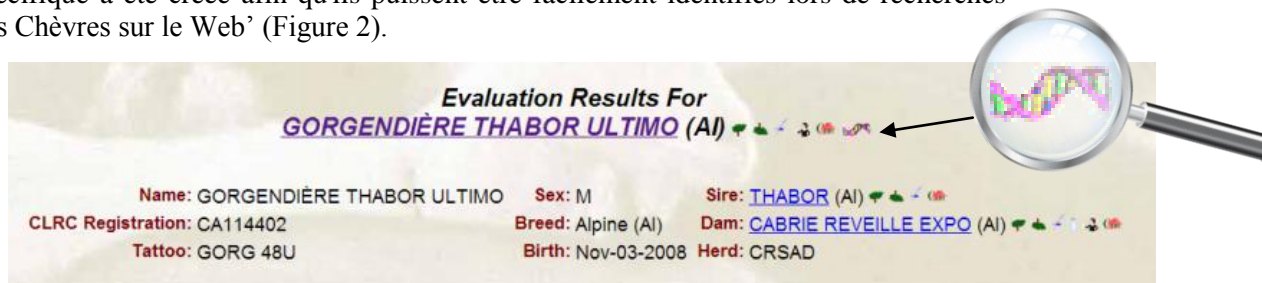


Figure 2. Screen capture of an animal with SNP genotyping data (represented by DNA helix) on Goats on the Web

2.4. Analyses génomiques

Considérons que le génome d'une espèce est un livre possédant différentes versions et que chaque livre possède des chapitres (chromosomes). Le génome est composé de 4 lettres (À,C,G,T) qui se combinent pour créer une séquence de la même façon que les lettres de l'alphabet sont utilisées pour créer différents mots. Les différences de formulation entre les versions sont comparables aux polymorphismes à nucléotide unique ou SNPs et sont tout simplement les différences d'une seule lettre au même endroit dans l'ADN entre deux animaux de la même espèce. Ces différences peuvent avoir un effet sur les caractères mesurables tels que la production de lait. Un locus (pluriel: loci) est l'adresse d'un gène et est semblable à un sous-chapitre d'un livre faisant partie d'un chapitre (chromosome). Un allèle est une variante d'un gène. Les évaluations génomiques utilisent les données de génotypage afin d'ajouter à la précision des évaluations génétiques classiques.

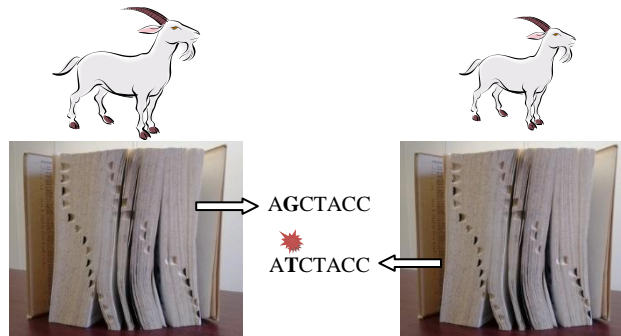


Figure 3: Exemple d'un SNP. Bien que les génomes des deux chèvres ci-dessus sont semblables, des petites différences dans une séquence pourraient être responsables d'une différence qui pourrait être observée chez un animal, par exemple sur la production laitière.

2.4.1. SNPs analysé

Le génome caprin (*Capra hircus*) est composé de 30 paires de chromosomes et a une longueur approximative de 2,636 Mb. La puce 50K caprine a été mise à la disposition de l'industrie à la fin de l'année 2011 grâce à une initiative du Consortium international sur le séquençage du génome caprin (IGGC). Il y a 53 347 SNPs sur la puce caprine à haute densité. Le nombre (pourcentage) des SNPs par chromosomes se trouvant sur la puce varie de 855 SNPs (1,6 %) pour le chromosome 25 à 3258 SNPs (6,1 %) pour le chromosome 1 (Figure 4).

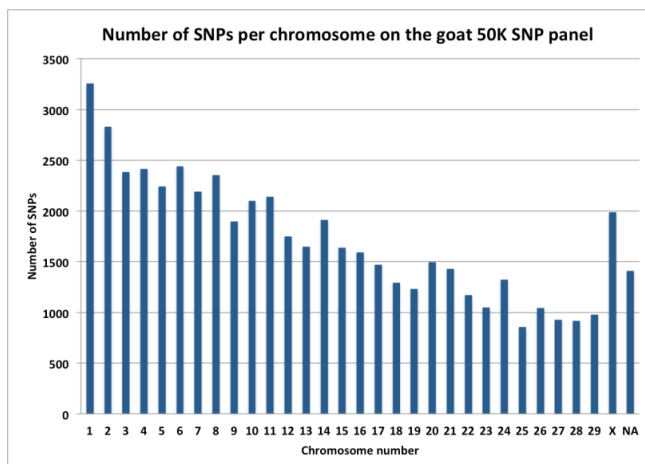


Figure 4. Nombre de SNPs par chromosome sur la puce 50K caprine (NA: SNPs non-cartographiés).

Approximativement 9 % des SNPs sont séparés par une distance de moins de 30kb et 8 % se trouvent à une distance de plus de 80kb. Environ 35 % des SNPs sont séparés par une distance de 30 à 40kb et plus de 80 % se trouvent à l'intérieur d'une distance de moins de 60kb l'un de l'autre. La distribution relative des SNPs à travers le génome fournit des opportunités d'utiliser la puce 50K pour différentes études telles que les études de déséquilibre de liaison, études d'associations à l'échelle du génome entier (GWAS) ou le calcul des valeurs génomiques (GEBV).

Plus d'informations sur les marqueurs SNPs disponibles sur la puce 50K caprine sont fournies à l'**Annexe 05**.

2.4.2. Variabilité des SNPs

Les études telles que les études d'association et le calcul de valeurs génomiques nécessitent de la variabilité dans les fréquences alléliques afin d'identifier les SNPs se trouvant sur la puce 50K caprine qui sont liés avec les différents caractères à l'étude. La fréquence de l'allèle mineur (FAM) est un critère représentant le taux de variabilité des marqueurs bi-alléliques tels que les SNPs. Lorsque la FAM s'approche de 0, cela signifie que l'allèle majeur est presque fixé, indiquant que le SNP en question n'est pas très pertinent pour la sélection génomique. À l'inverse, les valeurs de FAM se rapprochant de 50 % indiquent un taux élevé de polymorphisme, qui est nécessaire pour la sélection. Les fréquences intra-races des allèles SNP ont été calculées pour les 968 animaux de race pure génotypés en utilisant la puce 50K caprine. Les chèvres Alpines, Saanen et LaMancha génotypées avaient un taux de variabilité plus élevé que les chèvres Boer, Nubiennes et Toggenburg (Tableau 5).

Tableau 5. FAM moyenne de tous les SNPs génotypés par race

Race	Fréquence de l'allèle mineur moyenne tous marqueurs confondus (%)
Alpine	28
Saanen	27
LaMancha	28
Nubienne	24
Toggenburg	23
Boer	25

Une proportion plus élevée de SNPs ayant une FAM de moins de 10 % a été observée dans les races Boer, Nubienne et Toggenburg comparé aux chèvres Alpines, Saanen et LaMancha. Inversement, les races Alpine, Saanen et LaMancha avaient une proportion plus élevée de SNPs avec des FAM se situant entre 30 à 40 % ou plus de 40 %. Ces différences entre races peuvent être dues à un effet d'échantillonnage, en raison du faible nombre d'animaux génotypés dans certaines races. Les niveaux de FAM observés intra-races devraient fournir assez de variabilité pour les études génomiques telles que les études d'associations à l'échelle du génome entier (GWAS) et les évaluations génomiques (GEBV).

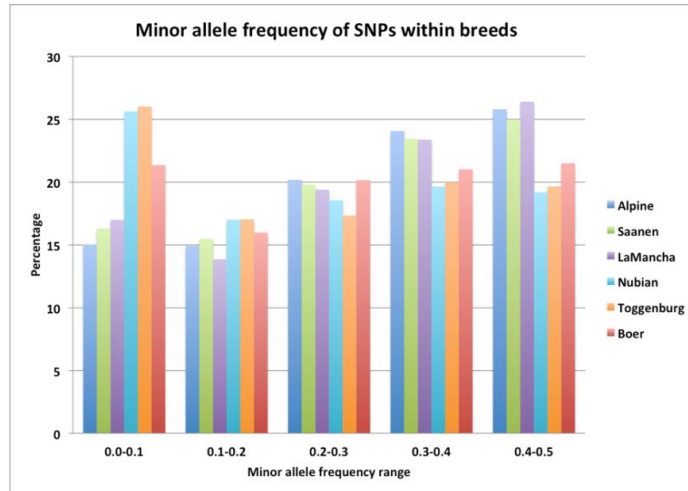


Figure 5: Fréquence de la FAM de différentes races pour les SNPs sur la puce SNP 50K caprine

Plus d'informations sur les FAM des SNPs sont fournies dans l'[Annexe 06](#).

2.4.3. Déséquilibre de liaison et phase de liaison

Une analyse d'association à l'échelle du génome entier réussie ainsi qu'une sélection génomique précise dépendent d'une densité adéquate des marqueurs, qui est déterminée par la mesure de déséquilibre de liaison (DL) à travers le génome. Le déséquilibre de liaison est l'association non aléatoire entre marqueurs à différents loci et est influencé par l'histoire de la population, les stratégies de sélection et les subdivisions géographiques. Le DL exerce un effet important sur la précision des valeurs génomiques estimées. Plus le DL est élevé, plus les marqueurs prédictifs des effets des loci quantitatifs (QTL) seront précis, car il y a plus de certitude que les bons allèles seront prédits au QTL étant donné les allèles présents au marqueur. Les loci quantitatifs sont des séquences d'ADN qui contiennent ou sont liées à des gènes ayant un effet sur un caractère donné. Des phases de liaison élevées entre les marqueurs et les QTL entre races sont également nécessaires pour estimer les effets des marqueurs toutes races confondues.

Certains SNPs ont été exclus des analyses pour les raisons suivantes:

- SNPs ayant des fréquences d'allèle mineur (FAM) de moins de 5% (pour les races Alpine et Saanen) ou 15% (pour les autres races, qui avaient beaucoup moins d'animaux génotypés).
- SNPs avec plus de 10% de génotypes manquants
- SNPs se trouvant sur les chromosomes sexuels ainsi que;
- Les SNPs non-cartographiés (aucun chromosome ou position pour le SNP).

L'étendue du DL entre marqueurs a été mesurée en utilisant r^2 , qui est le carré de la corrélation entre allèles à 2 loci. Cette valeur a été calculée pour chaque paire de loci sur chaque chromosome afin de déterminer le DL entre SNPs adjacents et le niveau de décomposition du DL sur différentes distances. Le r^2 a été calculé en utilisant le logiciel SNPPLD (Dr. Mehdi Sargolzaei). Les données ont été classées dans des groupes selon la distance par paire. Le DL moyen pour chaque groupe a été calculé pour évaluer le motif du DL entre chromosomes. La cohérence des phases de marqueur (r) a été calculée en prenant la racine carrée de la valeur de r . La corrélation entre les valeurs r a ensuite été calculée avec la procédure PORC CORR du logiciel SAS (SAS Institute Inc., Cary, USA). La cohérence de phase au marqueur est utilisée pour identifier le degré de similarité des phases de marqueurs SNP entre races qui, s'il est élevé, rend les évaluations génomiques multiraciales possibles.

La puce caprine 50K SNP, après filtration des données pour le contrôle de qualité, montre une bonne couverture du génome avec des intervalles moyens variant de 0,05 à 0,07 Mb. On observe des valeurs élevées de DL seulement dans le cas de courtes distances entre les paires de SNP. Pour toutes les races, le r^2 moyen diminue avec la distance entre les marqueurs. Les niveaux de DL sont bas pour des distances supérieures à 1.2 Mb chez toutes les races. La tendance du DL selon la distance est similaire dans toutes les races.

À une distance moyenne entre SNPs adjacents sur la puce 50K SNP (~0,06 Mb), la plupart des races approchent ou dépassent la valeur de DL requise pour les évaluations génomiques, sauf dans les races Alpine et Saanen. Ceci indique que la sélection génomique intra-race serait possible avec une population de référence assez grande, mais que les races Alpine et Saanen pourraient bénéficier d'une puce plus dense.

La cohérence de phase était la plus élevée entre les races Alpine et Saanen, indiquant une plus grande similitude entre ces deux races. Cependant, même pour ces deux races, la cohérence de phase entre des marqueurs adjacents n'est pas assez élevée (~0,50) pour permettre le mélange des deux races dans un groupe d'apprentissage pour la sélection génomique. En général les corrélations estimées sont faibles pour toutes les races même pour de courtes distances. En conséquence, plus d'animaux devraient être génotypés dans chaque race pour permettre le développement d'évaluations génomiques.

Des informations complémentaires sur le déséquilibre de liaison et la phase de liaison sont fournies à l'**Annexe 07**.

2.4.4. *Analyses d'associations (GWAS)*

Une autre application possible des données génomiques est l'analyse des liens potentiels entre les marqueurs SNP et les caractères ou les valeurs génétiques disponibles pour les animaux génotypés. Dans ce projet, une part importante des animaux génotypés ont été choisis sur la base de la quantité d'informations relatives aux performances propres ou aux performances de descendants. De cette manière, on crée une bonne population pour "l'apprentissage" des marqueurs en vue de développer des évaluations génomiques, mais cela s'avère également utile pour détecter des associations spécifiques entre certaines régions du génome et certains caractères d'intérêt pour les éleveurs et les producteurs de chèvres. Dans cette étude, différentes méthodes statistiques ont été employées pour détecter des liens importants entre les marqueurs SNP et les caractères évalués dans le cadre du Programme canadien d'amélioration génétique des chèvres laitières (www.GenetiqueCaprine.Ca), en particulier les principaux caractères de production laitière (lait, matière grasse et protéine par lactation), huit caractères de conformation et les comptages de cellules somatiques. Parmi les chèvres génotypées, au total 856 animaux avaient des valeurs génétiques pour les caractères de production, 759 pour les caractères de conformation, et 639 pour les cellules somatiques, avec des niveaux variables de précision.

Dans la race Alpine, les analyses ont permis de détecter 5 SNP associés avec la production de lait (sur les chromosomes 17 et 24), les cellules somatiques (chromosome 24) et l'attache avant de la mamelle (sur les chromosomes 1 et 4). Dans la race Saanen, un plus grand nombre d'associations a été trouvé, avec des SNPs sur les chromosomes 1, 4 et 13 associés avec l'attache avant de la mamelle, des SNPs sur les chromosomes 3 et 5 associés avec les cellules somatiques, et quatre SNPs sur les chromosomes 4, 8 et 9 associés avec la qualité des trayons.

Ces résultats sont encourageants et devraient être validés en génotypant davantage de chèvres laitières. Certains des SNPs détectés dans cette étude sont peut-être localisés dans des gènes qui pourraient être identifiés par de la cartographie fine.

Des informations complémentaires sur les analyses d'association sont fournies à l'**Annexe 08**.

2.4.5. Évaluations génomiques

D'après les résultats de l'analyse du déséquilibre de liaison, la sélection génomique pourrait être mise en place avec une grande population de référence, à l'intérieur de chaque race, mais il semble que les races Alpine et Saanen pourraient nécessiter une puce à SNP plus dense. La cohérence de phase entre marqueurs adjacents n'est pas assez élevée, et pour développer des évaluations génomiques multi-races, davantage de chèvres des différentes races devraient être génotypées. Le déséquilibre de liaison et la cohérence de phase de liaison sont des conditions de base pour développer les évaluations génomiques et la sélection génomique.

L'évaluation génomique est devenue une procédure de routine dans la sélection de différentes espèces animales. Elle consiste à sélectionner sur la base des valeurs génomiques, calculées à partir des effets des marqueurs couvrant l'ensemble du génome. Les données disponibles sur 367 chèvres Alpine ont été utilisées pour étudier l'impact de l'inclusion des données génomiques dans l'évaluation génomique. La race Alpine a été choisie pour cette étude, car elle est la race avec le plus grand nombre d'animaux génotypés. Pour estimer les effets des SNPs de façon précise et valider les prédictions, beaucoup d'animaux génotypés sont nécessaires.

Les caractères évalués sont: la production de lait, matière grasse et protéine, les cellules somatiques (comme indicateur de mammite), les notes de conformation pour l'avant et l'arrière de la mamelle, le ligament suspenseur, les trayons, le caractère laitier, les aplombs, la capacité corporelle et l'apparence générale. Après un contrôle de qualité des données génotypiques (identiques à celui de l'analyse du déséquilibre de liaison) 45 338 SNPs parmi les 53 347 SNPs initiaux ont été inclus dans l'analyse. Les animaux ont été séparés en deux groupes: un groupe d'apprentissage (animaux plus vieux avec plus de données de performances) ont été utilisés pour estimer les effets de marqueurs, qui sont ensuite utilisés pour calculer les valeurs génomiques des animaux conservés pour le groupe de validation (animaux plus jeunes). Les nombres d'animaux inclus dans les groupes d'apprentissage et de validation sont fournis au tableau 6.

Tableau 6. Animaux inclus dans les groupes d'apprentissage et de validation pour les évaluations génomiques.

Caractères	Groupe d'apprentissage	Groupe de validation
Production	317 chèvres	50 chèvres
Conformation	224 chèvres	20 chèvres

Les indices de potentiel génétique (IPGs) officiels utilisés provenaient de l'évaluation génétique nationale du mois d'octobre 2013 publiée par le Centre canadien pour l'amélioration des porcs. Le logiciel GEBV a été utilisé pour estimer les valeurs génomiques directes (VGDs). Le carré de la corrélation entre les VGDs et les IPGs a été calculé pour évaluer la précision réalisée des VGDs et a été comparé à la précision des IPGs officiels pour les animaux du groupe de validation.

Les résultats préliminaires pour la race Alpine indiquent que l'utilisation des données génomiques permet d'augmenter la précision des valeurs génétiques (0,04, 0,10 et 0,10 points de précision pour les caractères de production, les cellules somatiques et les caractères de conformation, respectivement). Ces résultats doivent être interprétés avec précaution du fait du petit nombre d'animaux dans les deux groupes

d'animaux (apprentissage et surtout validation). Ceci est d'autant plus vrai pour les caractères de conformation.

Des informations complémentaires sur les évaluations génomiques préliminaires sont fournies à ***l'Annexe 07.***

3. CONCLUSIONS

Ce projet constitue la première application des puces à SNP dans les populations caprines canadiennes. Depuis que les puces sont disponibles en décembre 2011, plusieurs pays ont lancé des programmes permettant de tester cette nouvelle technologie, mais ce projet sera parmi l'un des premiers permettant de partager des résultats sur un grand nombre (~1000) de chèvres dans six races différentes.

La génomique a été mise en application pour les bovins laitiers dans plusieurs pays à travers le monde au cours de la dernière décennie. Les gains attendus en matière d'intervalles de génération et de précision de la sélection font entrevoir des progrès potentiels importants sur les caractères classiques de production et de conformation, mais aussi sur de nouveaux caractères pas encore sélectionnés. L'industrie toute entière pourrait bénéficier de l'utilisation d'une telle technologie dans une future proche, car elle rend possible la sélection sur des caractères difficiles à mesurer tels que la résistance aux maladies.

Afin d'optimiser la valeur des puces à SNP utilisées dans ce projet, les animaux génotypés ont été choisis principalement sur la quantité d'information disponible (IPGs, données de contrôle laitier et de conformation, descendance, etc). Le but était de génotyper des animaux très informatifs, pour que les liens entre génotypes et phénotypes puissent être étudiés dans une population avec des valeurs génétiques ayant une bonne précision. Les chèvres ont été choisies dans 25 fermes du Québec et de l'Ontario, et deux centres d'insémination, afin d'assurer un bon échantillonnage de chaque race.

Ce projet a fourni des résultats originaux et très intéressants. Cependant, le niveau de précision de certains résultats est assez bas en raison du nombre limité d'animaux génotypés dans chaque race. Des études similaires réalisées en Europe sur un plus grand nombre d'animaux (>3000) ont conclu qu'un nombre de chèvres encore plus grand était nécessaire pour pouvoir mettre en place des évaluations génomiques. Dans la présente étude, où 979 chèvres issues de 6 races ont été génotypées, avec 53 à 403 chèvres génotypées par race, on aboutit évidemment à la même conclusion. Tous les résultats obtenus sont préliminaires et devraient être confirmés avec au moins 2000 animaux génotypés par race. Il est également important de noter que la puce à 50K SNP a été développée sur la base de marqueurs SNP identifiés dans les races suivantes : Alpine, Saanen, Boer, Créole, Kacang et Savannah. Dans notre étude, des chèvres de race Nubienne, Toggenburg et LaMancha ont également été testées en plus des chèvres de race Alpine, Saanen et Boer. Ce projet a permis de commencer à évaluer si la puce 50K SNP est adaptée à ces autres races. Dans les races Nubienne et Toggenburg, plus de 25 % des SNPs sur la puce à 50K SNP sont fixés ou très peu polymorphes, ce qui semble indiquer qu'une puce différente devrait être développée, mais on devrait génotyper davantage d'animaux pour confirmer ces résultats. À l'inverse, pour la race LaMancha on observe moins de 17 % de SNPs fixés ou très peu polymorphes (FAM de 0 à 10 %).

La quantité d'information génomique collectée durant ce projet est phénoménale. Des analyses complémentaires plus poussées seront nécessaires pour exploiter ces données de façon optimale. Ceci inclura le développement d'évaluations génomiques pour la race Saanen, et peut-être le partage de données avec d'autres organisations telles que l'INRA pour améliorer la précision des évaluations génomiques dans les races Alpine et Saanen.

Au cours de ce projet, de nouveaux outils ont été développés pour centraliser, valider, analyser les données génomiques et les inclure dans les évaluations génomiques pour les chèvres laitières. À l'avenir, tout nouveau génotype pourra être traité par ces mêmes programmes. Les éleveurs de chèvres souhaitant adopter cette technologie auront accès à des valeurs génomiques pour leurs animaux à un très jeune âge, ce qui leur fournira une bonne prédiction de leur potentiel génétique pour un large choix de caractères.

La puce à SNP caprine actuellement disponible permet l'identification de plus de 50 000 marqueurs différents sur le génome d'un animal génotypé. À mesure qu'on en apprend de plus en plus sur ces puces, il serait judicieux de développer des puces de plus faible densité et moins dispendieuses, qui seraient plus abordables pour les éleveurs de chèvres souhaitant génotyper l'ensemble ou une partie de leurs animaux. Ces puces à SNP devraient être développées collectivement dans le cadre du Consortium international sur la génomique caprine. Enfin, des collaborations internationales seront développées sur les méthodes d'évaluation génomique, en particulier avec la France et les États-Unis avec qui le Canada échange de la génétique caprine.

4. SOMMAIRE DES ACCOMPLISSEMENTS DU PROJET

La puce caprine 50K SNP apporte des opportunités d'améliorer notre connaissance du génome caprin, d'accélérer les progrès génétiques sur des caractères d'intérêt et de commencer à sélectionner sur de nouveaux caractères. La technologie est basée sur les polymorphismes de nucléotide unique (*Single Nucleotide Polymorphisms* - SNP), des marqueurs présents sur l'ensemble du génome. La puce caprine 50K SNP, qui inclut plus de 50 000 marqueurs SNP, est disponible depuis décembre 2011. Un projet a été conçu pour évaluer le potentiel de la sélection génomique pour améliorer la productivité et la santé des troupeaux caprins. Des échantillons de tissus ont été collectés sur un total de 1041 chèvres et boucs dans 25 fermes au Québec et en Ontario et 2 centres d'insémination. Les animaux génotypés ont été choisis sur la base de l'information disponible afin d'inclure des animaux ayant une grande influence sur la population. Un total de 403 animaux de race Alpine, 318 Saanen, 81 LaMancha, 54 Nubiennes, 53 Toggenburg et 67 Boer ont été génotypés. Le génotypage SNP a généré une énorme quantité d'information pour chaque animal. Au total, plus de 50 millions de résultats SNP ont été centralisés dans la base de données nationale. Les génotypes ont été analysés pour évaluer la variabilité des SNP dans chaque race et plusieurs critères, tels que les fréquences d'allèle mineur (FAM), le déséquilibre de liaison (DL) et la phase de liaison. Ces critères sont importants pour connaître le potentiel et les conditions d'application des évaluations génomiques. Les FAMs sont des indicateurs de variabilité des SNPs. Elles varient de 23 à 28 % dans les six races testées, avec 74 à 85 % des marqueurs SNPs ayant une FAM > 10 %, ce qui indique un bon niveau de polymorphisme. Les analyses de phase de liaison montrent que la plupart des races avoisinent ou dépassent les valeurs de DL requises pour l'évaluation génomique. Les analyses de cohérence de phase ne concluent pas en faveur d'évaluations génomiques multiraciales. Des analyses d'association préliminaires ont permis de détecter des marqueurs SNP associés avec différents caractères de production, de conformation et de santé chez les races Alpine et Saanen. Des valeurs génomiques pilotes ont été calculées en race Alpine pour 12 caractères de production et conformation et comparées aux valeurs génétiques officielles pour les chèvres et boucs génotypés. Ces résultats préliminaires montrent que l'inclusion des génotypes SNP dans les évaluations génétiques améliorerait considérablement la précision de la sélection, en particulier pour les caractères de conformation et les cellules somatiques. Il est nécessaire de centraliser un plus grand nombre de génotypes caprins pour confirmer ces résultats et mettre en place des évaluations génomiques en routine pour les chèvres laitières canadiennes.

5. ACRONYMES ET DESCRIPTIONS

Organisations

CCAP	Centre canadien pour l'amélioration des porcs
CDAQ	Conseil pour le développement de l'agriculture du Québec
IGS	International Goat Symposium
OG	Ontario Goat
SECLRQ	Société des éleveurs de chèvres laitières de race du Québec
SPCQ	Syndicat des producteurs de chèvres du Québec

Genomic glossary

SNP	Single Nucleotide Polymorphism (Polymorphisme de nucléotide unique)
FAM	Fréquence d'allèle mineur
GWAS	Genome-wide Association Studies
GEBV	Genomic Estimated Breeding Values (Valeurs génomiques estimées)
DL	Déséquilibre de liaison


6. ANNEXES

Annexe 05: Marqueurs SNP sur la puce caprine à 50K SNP

Annexe 06: Rapport sur la fréquence d'allèle mineur dans les races caprines étudiées

Annexe 07: Rapport sur les évaluations génomiques

Annexe 08: Rapport sur les analyses d'association (GWAS)



Utilisation de la génomique pour améliorer la productivité et la santé des troupeaux caprins

SNPs sur la puce 50K caprine à haute densité

« Une partie du financement de ce projet a été fournie par les conseils sectoriels du Québec, de l'Ontario et de la Colombie-Britannique qui gèrent le Programme canadien d'adaptation agricole (PCAA) pour Agriculture et agroalimentaire Canada. »



SOCIÉTÉ DES ÉLEVEURS
DE CHÈVRES LAITIÈRES
DE RACE DU QUÉBEC



GoatGenetics.Ca
GénétiQueCaprine.Ca



Investment
Agriculture
Foundation
of British Columbia



Conseil de
l'adaptation
agricole



Agriculture et
Agroalimentaire Canada

Agriculture and
Agri-Food Canada



CONSEIL POUR
LE DÉVELOPPEMENT DE
L'AGRICULTURE DU QUÉBEC

SNPs sur la puce 50K caprine à haute densité

Le génome caprin est composé de 30 paires de chromosomes et a une longueur approximative de 2636 Mb. Le plus petit chromosome est le chromosome 25 et le plus grand est le chromosome 1, avec une longueur de 42 et 155 Mb, respectivement (www.ncbi.nlm.nih.gov). La puce 50K caprine est disponible pour l'industrie depuis la fin de l'année 2011 grâce à une initiative du Consortium international sur le séquençage du génome caprin (IGGC). Le nombre (pourcentage) de SNPs par chromosome localisés sur la puce varie de 855 SNPs (1,6%) pour le chromosome 25 à 3 258 SNPs (6,11%) pour le chromosome 1 (Figure 1). Il existe 1 407 SNPs sur la puce pour lesquels la localisation chromosomique est inconnue. Les cartes SNPs ont été gracieusement fournies par Gwenola Tosser-Klopp de l'IGGC.

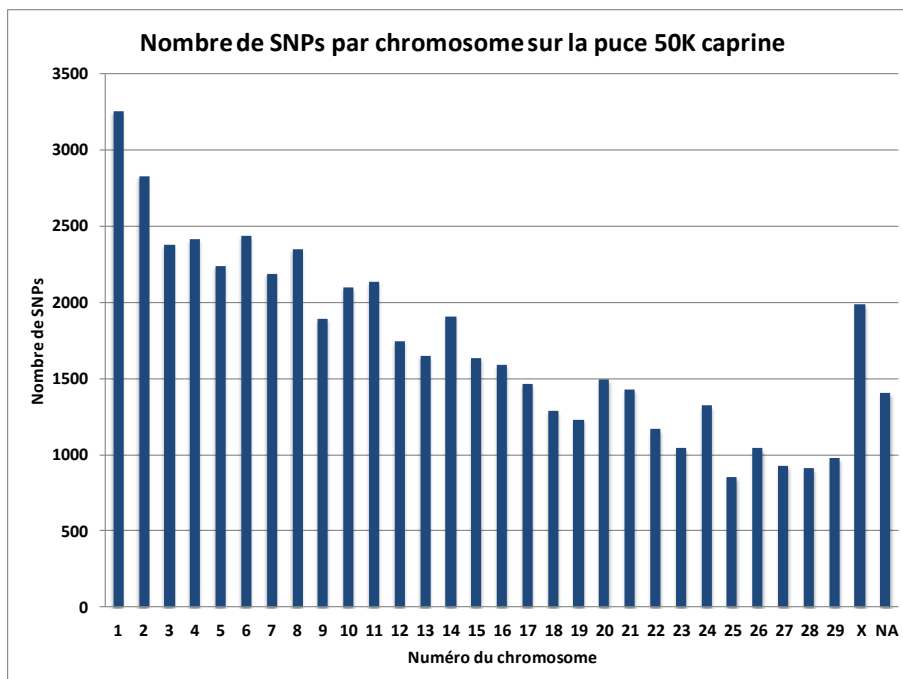


Figure 1. Nombre de SNPs par chromosome sur la puce 50K caprine (NA: SNPs non-cartographiés).

La puce caprine à haute densité contient très exactement 53 347 marqueurs SNP. Les SNPs sont séparés les uns des autres par une distance moyenne de 49 kb. La Figure 2 montre la distribution des SNPs adjacents selon leurs distances deux à deux sur les chromosomes. Approximativement 9% des SNPs sont séparés par une distance de moins de 30 kb et 8% se trouvent à une distance de plus de 80 kb. Environ 35% des SNPs sont séparés par une distance de 30 à 40 kb et plus de 80% se trouvent à une distance de moins de 60 kb l'un de l'autre. La distribution relative des SNPs sur l'ensemble du génome fournit des opportunités d'utiliser la puce 50K pour différentes études telles que les études de déséquilibre de liaison, études d'associations à l'échelle du génome entier (GWAS) ou le calcul des valeurs génomiques (GEBV).

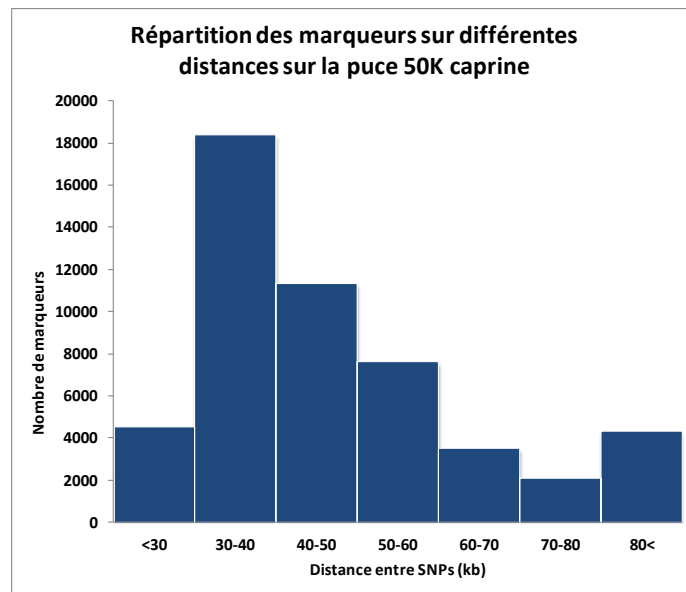



Figure 2: Distribution relative des distances entre SNPs sur la puce SNP 50K caprine.

La validation de l'équilibre de Hardy-Weinberg (HWE) a été réalisée pour chaque race à l'exception des chèvres croisées (seulement 3 animaux génotypés). Toutes les races avaient de faibles niveaux de déséquilibre avec un seuil statistique (p-value) de moins de 10^{-6} . La race Toggenburg connaît le déséquilibre HWE le plus important avec 0,57% et les chèvres de race Boer le niveau le plus faible avec 0,14%. Puisqu'une des hypothèses de départ pour l'équilibre de Hardy-Weinberg est l'absence de sélection, le faible nombre de SNPs qui ne sont pas en équilibre pourrait suggérer que les SNPs sur la puce à SNP 50K ne sont pas soumis à la sélection. Le taux de succès de génotypage des SNPs pour toutes les races était d'environ 93%.



Utilisation de la génomique pour améliorer la productivité et la santé des troupeaux caprins

Fréquences alléliques des marqueurs SNP

« Une partie du financement de ce projet a été fournie par les conseils sectoriels du Québec, de l'Ontario et de la Colombie-Britannique qui gèrent le Programme canadien d'adaptation agricole (PCAA) pour Agriculture et agroalimentaire Canada. »



SOCIÉTÉ DES ÉLEVEURS
DE CHÈVRES LAITIÈRES
DE RACE DU QUÉBEC



GoatGenetics.Ca
GénétiQueCaprine.Ca



Investment
Agriculture
Foundation
of British Columbia



Conseil de
l'adaptation
agricole



Agriculture et
Agroalimentaire Canada

Agriculture and
Agri-Food Canada



CONSEIL POUR
LE DÉVELOPPEMENT DE
L'AGRICULTURE DU QUÉBEC

Fréquences alléliques des marqueurs SNP

Les études telles que les études d'association (GWAS) et le calcul de valeurs génomiques (GEBV) nécessitent de la variabilité dans les fréquences alléliques afin d'identifier les marqueurs SNP de la puce 50K caprine qui sont liés avec les différents allèles de locus quantitatifs (QTL) ayant un effet sur les caractères à l'étude. La fréquence de l'allèle mineur (FAM) est un critère représentant le taux de variabilité des marqueurs bi-alléliques tels que les SNPs. Les fréquences intra-races des allèles SNP ont été calculées pour les 968 animaux de race pure génotypés en utilisant la puce SNP 50K caprine. Les chèvres de race Alpine, Saanen et LaMancha génotypées avaient un taux de variabilité des allèles SNP plus élevé que les chèvres de race Boer, Nubienne et Toggenburg. La fréquence de l'allèle mineur moyenne pour tous les marqueurs confondus était de 28, 28, 27, 25, 24 et 23 % pour les races Alpine, LaMancha, Saanen, Boer, Toggenburg et Nubienne, respectivement (Figure 3).

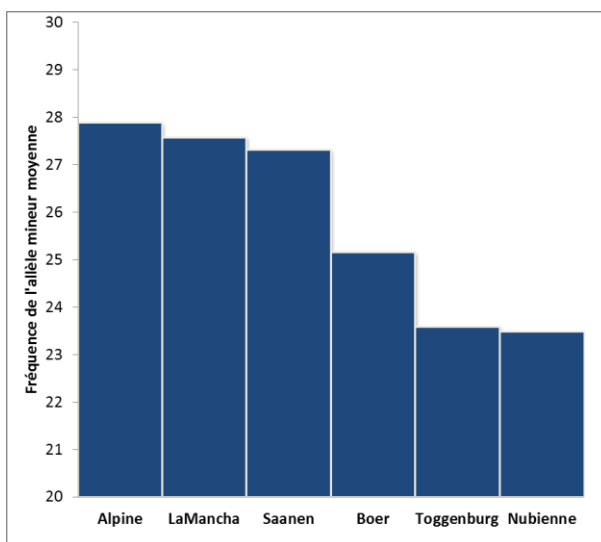


Figure 3: Moyenne de la fréquence d'allèle mineur intra-race, pour tous les SNPs confondus se trouvant sur la puce SNP 50K caprine

La proportion de SNPs avec une FAM inférieure à 0,1 s'élevait à 15, 16 et 17 % pour les races Alpine, Saanen et LaMancha, respectivement. Les races Boer, Nubienne et Toggenburg avaient 21, 26 et 26% des SNPs présentant une FAM inférieure à 0,1, respectivement (Figure 4).

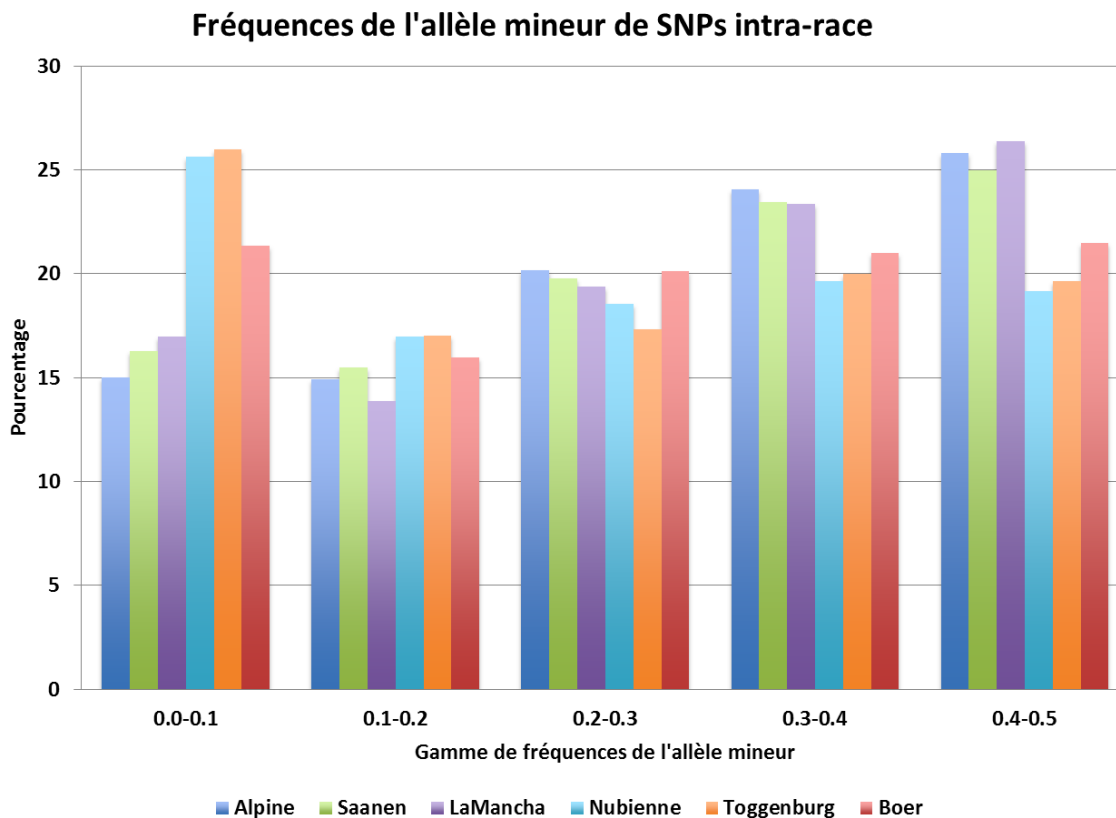


Figure 4: Distribution de différentes classes de FAM pour les marqueurs SNPs sur la puce SNP 50K caprine.

Les races Alpine, Saanen et LaMancha ont une proportion plus élevée de SNPs avec des FAM supérieures à 40%. Le pourcentage de SNPs avec des FAM supérieures à 0,4 est de 26, 25 et 26% pour les races Alpine, Saanen et LaMancha, respectivement. Pour les races Boer, Nubienne et Toggenburg, 22, 19 et 20% des marqueurs SNP ont une FAM supérieure à 0.40, respectivement (Figure 4). Les niveaux de FAM observés intra-race indiquent que la variabilité existante est suffisante pour les études génomiques telles que les études d'associations à l'échelle du génome entier (GWAS) et les évaluations génomiques (GEBV).



**Centre canadien
pour l'amélioration
des porcs Inc.**



**Centre for Genetic
Improvement of
Livestock**

Bref Rapport des Activités de Recherche pour le Projet:

**“Utilisation de la génomique pour améliorer
la productivité et la santé des troupeaux caprins”**

Luiz Fernando Brito*

Superviseur: Dr. Flávio Schenkel

COURRIEL: lbrito@uoguelph.ca

*** Ph.D. étudiant, Centre for Genetic Improvement of Livestock,
Université de Guelph.**

Le 16 décembre 2013

Guelph – ON – Canada

Ce projet a été rendu possible grâce à la contribution financière du Programme canadien d'adaptation agricole d'Agriculture et agroalimentaire Canada offert par le Conseil pour le Développement de l'agriculture du Québec (CDAQ).



SOCIÉTÉ DES ÉLEVEURS
DE CHÈVRES LAITIÈRES
DE RACE DU QUÉBEC



Conseil de
l'adaptation
agricole



Agriculture et
Agroalimentaire Canada



GoatGenetics.Ca
GénétiqueCaprine.Ca

Agriculture and
Agri-Food Canada



Utilisation de la génomique pour améliorer la productivité et la santé des troupeaux caprins

Introduction

La consommation de lait de chèvre est en croissance dans plusieurs régions du monde au cours des dernières années. Cette hausse a mené à un besoin accru de recherche sur la production de chèvres laitières afin d'améliorer la productivité.

Durant de nombreuses années, la sélection des chèvres laitières a été basée sur les mesures de production laitière, matières sèches du lait (gras, protéines, etc.) et l'évaluation de la conformation (qualité du système mammaire, structure des pieds et membres, etc.) sur des chèvres individuelles. Ces mesures phénotypiques sont combinées avec les informations généalogiques utilisant la méthode meilleure prédiction linéaire non-biaisée (BLUP) pour estimer les valeurs génétiques. Au cours des dernières années, d'autres caractères tels que la facilité de traite, la reproduction, la résistance aux maladies ont également été considérés afin d'améliorer la qualité des produits, la santé et le bien-être des animaux et pour réduire les coûts de production pour l'industrie de chèvres laitières et le système d'enregistrement de données dans son ensemble.

La science du génome et les outils génomiques ont été très importants pour améliorer les productions animales au cours des dernières années, particulièrement chez les bovins. Sur une courte période de temps, les coûts ont chuté et la technologie s'est améliorée pour permettre d'assembler les génomes complets de plusieurs espèces.

En utilisant la génomique, des progrès génétiques rapides peuvent être réalisés sur des caractères qui sont déjà mesurés en routine par les éleveurs. Par exemple, les boucs pourraient être contrôlés à un très jeune âge, en utilisant l'ADN de l'animal pour le génotypage, au lieu d'attendre qu'ils soient évalués au moyen du contrôle par descendance de leurs filles, qui nécessite plusieurs années.

L'industrie des bovins laitiers au Canada et ailleurs utilise déjà cette technologie pour améliorer les évaluations génétiques. Cela a été fait de telle sorte que les coûts sont plus faibles que dans les schémas de sélection classiques et les progrès génétiques devraient augmenter considérablement grâce à la précision accrue des valeurs génétiques (IPGs) des jeunes sujets et à un intervalle de génération plus court. Cette technologie offre également l'opportunité de fournir des évaluations génétiques pour des caractères qui sont soit très difficiles ou dispendieux à mesurer, tels que les caractères liés à la résistance aux maladies et la longévité.

Actuellement, les études dans le domaine de la génomique caprine sont rares. Sachant cela, l'objectif de ce projet était de réaliser des études appliquées de base, fondamentales pour la mise en œuvre de la sélection génomique sur l'évaluation génétique des chèvres laitières.

A. Caractérisation du déséquilibre de liaison et phase gamétique chez les races caprines canadiennes

Introduction

Une étude d'association à l'échelle du génome réussie et une sélection génomique précise dépendent de la densité des marqueurs, qui est déterminée par l'ampleur du déséquilibre de liaison (DL) à l'échelle du génome (Khatkar *et al.* 2008). Le DL est défini comme l'association non-aléatoire entre allèles situés à différent loci. Le DL est influencé par l'histoire de la population, les stratégies de sélection et les subdivisions géographiques (Slatkin, 2008).

Ces associations alléliques sont principalement dues à la proximité physique, mais des paires de SNPs éloignées peuvent être en déséquilibre de liaison complet. De plus, le taux de DL peut varier énormément d'une région à une autre. Malgré cette complexité apparente, le DL est organisé en blocs d'haplotypes qui montrent un DL élevé, séparés par des points chauds probables de recombinaison (Daly *et al.* 2001; Ardlie *et al.* 2002; Jeffreys *et al.* 2001). Cette caractéristique est capitale car elle ouvre la

possibilité de choisir un ensemble de SNPs de manière rationnelle, de telle sorte que les SNPs faisant partie d'un bloc d'haplotype puissent être choisis pour la cartographie fine (Johnson *et al.* 2001). Le DL peut être créé par des phénomènes tels que la migration (croisement), les mutations, la sélection et la dérive génétique (consanguinité) dans les populations ayant un petit effectif génétique.

Le déséquilibre de liaison exerce un effet important sur la précision des valeurs génomiques estimées. Plus le DL est élevé, plus les marqueurs prédictifs des effets des loci quantitatifs (QTL) seront précis, car il y a plus de certitude que les bons allèles seront prédits au QTL étant donné les allèles présents au marqueur (Larmer, 2012). De plus, afin d'estimer l'effet de chaque marqueur toutes races confondues, non seulement est-il nécessaire d'avoir des niveaux de DL élevé dans chaque race, mais également des phases de liaison élevées entre marqueurs et QTL entre races.

L'objectif de cette étude était d'estimer les niveaux de déséquilibre de liaison (DL) à l'échelle du génome entier ainsi que la persistance de la phase gamétique pour six races caprines canadiennes.

MATÉRIEL ET MÉTHODES

Animaux et génotypes

Un total de 979 animaux provenant de six races ont été génotypés en utilisant la puce ADN d'Illumina (Illumina goat SNP50 BeadChip) contenant 53 347 SNP. Le nombre d'animaux génotypés par race est présenté au Tableau 1. Les 3 animaux croisés listés dans le tableau ci-dessous ont été exclus des analyses subséquentes.

Tableau 1. Description des animaux génotypés par race et sexe.

Race	Mâles	Femelles	Total
Alpine	51	352	403
Saanen	51	267	318
LaMancha	11	70	81
Nubienne	21	33	54
Toggenburg	7	46	53
Boer	17	50	67
Croisés	0	3	3
Total	158	821	979

Les polymorphismes à nucléotide unique (SNPs) ayant des fréquences d'allèle mineur (FAM) de moins de 5% (pour les races Alpine et Saanen) ou 15% (pour les autres races, qui avaient beaucoup moins d'animaux génotypés) ont été exclus afin d'étudier seulement les SNPs en ségrégation dans un grand nombre d'animaux de l'échantillon. Les SNPs avec plus de 10% de génotypes manquants ont également été exclus. Tous les animaux avaient moins de 10% de génotypes manquants, ils n'ont donc pas été exclus par ce filtre. Seuls les SNPs provenant de chromosomes autosomiques ont été inclus dans l'analyse et les SNPs non-cartographiés (sans numéro de chromosome et/ou position sur le chromosome) ont été exclus.

Détermination de l'étendue du déséquilibre de liaison

L'étendue du DL entre marqueurs a été mesurée en utilisant le critère r^2 , tel que proposé dans l'étude de Hill et Robertson (1968), défini comme le carré de la corrélation entre allèles à 2 loci. Cette valeur a été calculée pour chaque paire de loci sur chaque chromosome afin de déterminer le DL entre SNPs adjacents et le niveau de décomposition du DL sur différentes distances. Le r^2 a été calculé comme suit en utilisant le logiciel SNPPLD (Dr. Mehdi Sargolzaei):

$$r^2 = \frac{D^2}{f(A)f(a)f(B)f(b)}$$

Où $f(A)$ est la fréquence de l'allèle A dans la population et ainsi de suite pour les autres allèles dans la population; et $D=f(AB)-f(A)f(B)$ où $f(AB)$ est la fréquence de l'haplotype AB dans la population. Les valeurs r^2 varient entre 0, pour une paire de loci avec pas de DL entre eux, et 1, pour une paire de loci en déséquilibre de liaison complet.

Pour évaluer le motif du DL le long des chromosomes, les données ont été triées en groupes selon la distance par paire. Le DL moyen a été calculé pour chaque groupe.

Cohérence des phases

Pour chaque paire de marqueurs ayant une mesure r^2 , la valeur positive ou négative r a été calculée en prenant la racine carrée de la valeur de r . Les données ont été triées en différents groupes en fonction de la distance par paire afin de déterminer la composition de la cohérence de la phase gamétique (corrélation des valeurs r entre deux races) sur différentes distances et pour évaluer la cohérence des phases sur les plus petites distances possibles, étant donné le nombre de SNPs génotypés. La corrélation entre les valeurs r a ensuite été calculée entre les six races en utilisant la procédure PROC CORR du logiciel SAS (SAS Institute Inc., Cary, USA).

RÉSULTATS

Le nombre de SNPs exclus selon les critères lors du contrôle de la qualité ainsi que le nombre de SNPs inclus dans les analyses sont présentés dans le Tableau 2.

Tableau 2. Nombre de SNPs exclus lors du contrôle de qualité des données dans les six races et distance moyenne (Mb) entre SNPs adjacents

SNPs exclus				
Race	FAM < 0.05	SNP [†] CR < 0.90	Autres*	SNPs restant
Alpine	1 618	3 338	3 053	45 338
Saanen	1 980	3 338	3 019	45 010
	FAM < 0.15	SNP [†] CR < 0.90	Autres*	SNPs restant
Toggenburg	13 726	3 342	2 518	33 761
LaMancha	8 354	3 338	2 819	38 836
Nubienne	14 360	3 357	2 505	33 125
Boer	11 573	3 352	2 564	35 858

*SNPs sans numéro de chromosome et/ou emplacement chromosomique ou SNPs se trouvant sur les chromosomes sexuels.

[†]CR: Call rate = taux de réussite du génotypage

Le DL moyen et la distance observée entre SNPs adjacents dans les six races étudiées sont présentés dans le Tableau 3. La puce à SNP, suite à l'étape du contrôle de la qualité, a montré une bonne couverture du génome avec un écart moyen entre SNPs allant de 0,05 à 0,07Mb. Les races avec un plus petit nombre d'animaux génotypés avaient le plus grand écart moyen entre SNPs en raison de l'exclusion des SNPs avec des fréquences de l'allèle mineur (FAM) inférieur à 15%, tandis que pour les races Alpine et Saanen, le seuil de FAM utilisé était de 5%. Les valeurs r^2 entre SNPs adjacents ont été calculées comme la moyenne pondérée des valeurs estimées de r^2 et le nombre de paires de SNPs adjacentes sur chaque chromosome.

Les valeurs moyennes de r^2 pour différents intervalles de distances sont présentées dans le Tableau 4 pour toutes les races. La tendance globale pour chaque race est présentée à la Figure 1. Les tendances par chromosome et par race sont présentées à la Figure 2 (Annexe 1). Les valeurs de DL élevées (r^2) ont été observées seulement pour de courtes distances entre paires de SNPs. Pour toutes les races, le r^2 moyen diminuait avec une distance croissante entre marqueurs. Les estimations pour les races Alpine et Saanen donnent les valeurs les plus faibles et elles sont semblables pour les deux races. Cela peut être expliqué par leur ascendance commune. Les races Alpine et Saanen ont le plus grand nombre d'animaux échantillonnés et génotypés. Les valeurs de DL élevées pour les autres races pourraient être dues à l'échantillonnage ou

pourraient être réelles en raison de la plus petite taille génétique de la population. Il serait intéressant de génotyper plus d'animaux dans ces races afin de confirmer les résultats de DL obtenus dans cette étude.

Les animaux de races Boer et Nubienne avaient les plus hauts niveaux de DL pour l'ensemble des distances étudiées. Les niveaux de DL étaient faibles pour toutes les races pour les distances supérieures à 1.2Mb. Les tendances pour toutes races et distances confondues étaient très similaires (Figure 1). L'étendue du DL diminue considérablement entre les intervalles de distances inférieures à 0,02 Mb et ceux de 0,02 Mb à 0,03 Mb. Le nombre de paires de SNPs se trouvant à des distances inférieures à 0,02 Mb est cependant très faible.

Une meilleure précision des valeurs génomiques estimées est liée à un niveau de DL plus élevé. Certaines études (ex. Calus, 2008; Meuwissen *et al.* 2001) ont conclu qu'une valeur de r^2 supérieure à 0,2 est suffisant pour la sélection génomique. À la distance moyenne entre SNPs adjacents sur la puce 50K (~0,06 Mb), la plupart des races dépassent ou s'approchent de cette valeur, à l'exception des races Alpine et Saanen ($r^2 \sim 0,12-0,13$). Cela indique qu'avec une population de référence assez grande, la sélection génomique intra-race pourrait potentiellement être mise en place avec la puce 50K actuelle, mais les races Alpine et Saanen pourraient nécessiter une puce plus dense.

Carillier *et al.* (2013) ont trouvé une valeur de r^2 de 0,17 entre SNPs adjacents dans les races Alpine et Saanen. Ces auteurs ont également évalué l'apport de l'information génomique à la précision des évaluations génétiques et ils ont rapporté certains gains. Cependant, ces gains étaient légèrement plus faibles comparé aux gains obtenus dans les autres espèces. Selon eux, les gains plus faibles seraient dus à la structure et la taille de la population de référence.

Tableau 3. Déséquilibre de liaison moyen (r^2) entre SNPs adjacents par race et distance moyenne entre SNPs adjacents (Mb) dans les six races.

Race	r^2	Distance moyenne entre SNPs adjacents (Mb)
Alpine	0,1445	0,05296
Saanen	0,1534	0,05336
Toggenburg	0,2431	0,07132
LaMancha	0,1934	0,06181
Nubienne	0,2721	0,07247
Boer	0,2860	0,06701

Tableau 4: Valeurs r^2 moyennes et nombre de paires de marqueurs (entre parenthèses) pour les six races pour différents intervalles de distances

Distance (Mb)	Alpine	Boer	LaMancha
<0,02	0,3207 (630)	0,3819 (341)	0,4103 (476)
0,02 - 0,03	0,1692 (11 752)	0,3051 (7 485)	0,2203 (8 656)
0,03 - 0,04	0,1545 (11 424)	0,2954 (7 442)	0,2028 (8 469)
0,04 - 0,05	0,1393 (7 785)	0,2801 (5 073)	0,1906 (5 748)
0,05 - 0,06	0,1220 (7 472)	0,2736 (4 857)	0,1786 (5 519)
0,06 - 0,07	0,1125 (8 958)	0,2683 (5 759)	0,1673 (6 659)
0,07 - 0,08	0,1094 (9 007)	0,2606 (5 734)	0,1689 (6 555)
0,08 - 0,09	0,1028 (8 641)	0,2493 (5 515)	0,1611 (6 377)
0,09 - 0,1	0,1009 (8 523)	0,2495 (5 466)	0,1602 (6 326)
0,1 - 0,2	0,0897 (86 190)	0,2325 (55 127)	0,1472 (63 513)
0,2 - 0,3	0,0799 (86 059)	0,2143 (54 948)	0,1361 (63 413)
0,3 - 0,4	0,0756 (85 839)	0,2019 (54 631)	0,1306 (63 399)
0,4 - 0,5	0,0731 (85 706)	0,1900 (54 627)	0,1272 (63 032)
0,5 - 0,6	0,0703 (85 621)	0,1785 (54 259)	0,1236 (63 058)
0,6 - 0,7	0,0675 (85 504)	0,1709 (54 530)	0,1201 (63 082)
0,7 - 0,8	0,0661 (85 265)	0,1622 (54 223)	0,1160 (62 639)
0,8 - 0,9	0,0642 (85 234)	0,1543 (54 316)	0,1137 (62 808)
0,9 - 1	0,0620 (85 169)	0,1479 (53 981)	0,1109 (62 668)
1 - 1,2	0,0595 (169 874)	0,1388 (107 760)	0,1065 (124 992)
> 1,2	0,0117 (38 841 175)	0,0288 (24 339 482)	0,0328 (28 533 190)

Distance (Mb)	Nubienne	Saanen	Toggenburg
<0,02	0,4460 (369)	0,3470 (601)	0,3626 (295)
0,02 - 0,03	0,2934 (6 446)	0,1787 (11 530)	0,2759 (6 709)
0,03 - 0,04	0,2823 (6 179)	0,1611 (11 298)	0,2584 (6 587)
0,04 - 0,05	0,2621 (4 338)	0,1471 (7 703)	0,2472 (4 518)
0,05 - 0,06	0,2620 (4 091)	0,1321 (7 339)	0,2232 (4 239)
0,06 - 0,07	0,2597 (4 879)	0,1216 (8 838)	0,2201 (5 105)
0,07 - 0,08	0,2529 (4 973)	0,1205 (8 859)	0,2157 (5 103)
0,08 - 0,09	0,2532 (4 834)	0,1130 (8 542)	0,2073 (4 886)
0,09 - 0,1	0,2400 (4 653)	0,1101 (8 414)	0,2039 (4 811)
0,1 - 0,2	0,2312 (47 078)	0,0993 (84 981)	0,1941 (48 788)
0,2 - 0,3	0,2145 (47 066)	0,0895 (84 776)	0,1798 (48 691)
0,3 - 0,4	0,2055 (46 801)	0,0854 (84 692)	0,1738 (48 405)
0,4 - 0,5	0,1962 (46 607)	0,0815 (84 477)	0,1687 (48 073)
0,5 - 0,6	0,1866 (46 400)	0,0786 (84 448)	0,1636 (48 102)
0,6 - 0,7	0,1796 (46 453)	0,0762 (84 337)	0,1609 (48 094)
0,7 - 0,8	0,1723 (46 140)	0,0744 (84 033)	0,1559 (47 785)
0,8 - 0,9	0,1670 (46 365)	0,07169 (84 111)	0,1530 (47 959)
0,9 - 1	0,1609 (46 185)	0,0703 (83 961)	0,1489 (47 940)
1 - 1,2	0,1518 (92 205)	0,0675 (167 458)	0,1443 (95 393)
> 1,2	0,0363 (20 694 120)	0,0146 (38 304 556)	0,0462 (21 633 868)

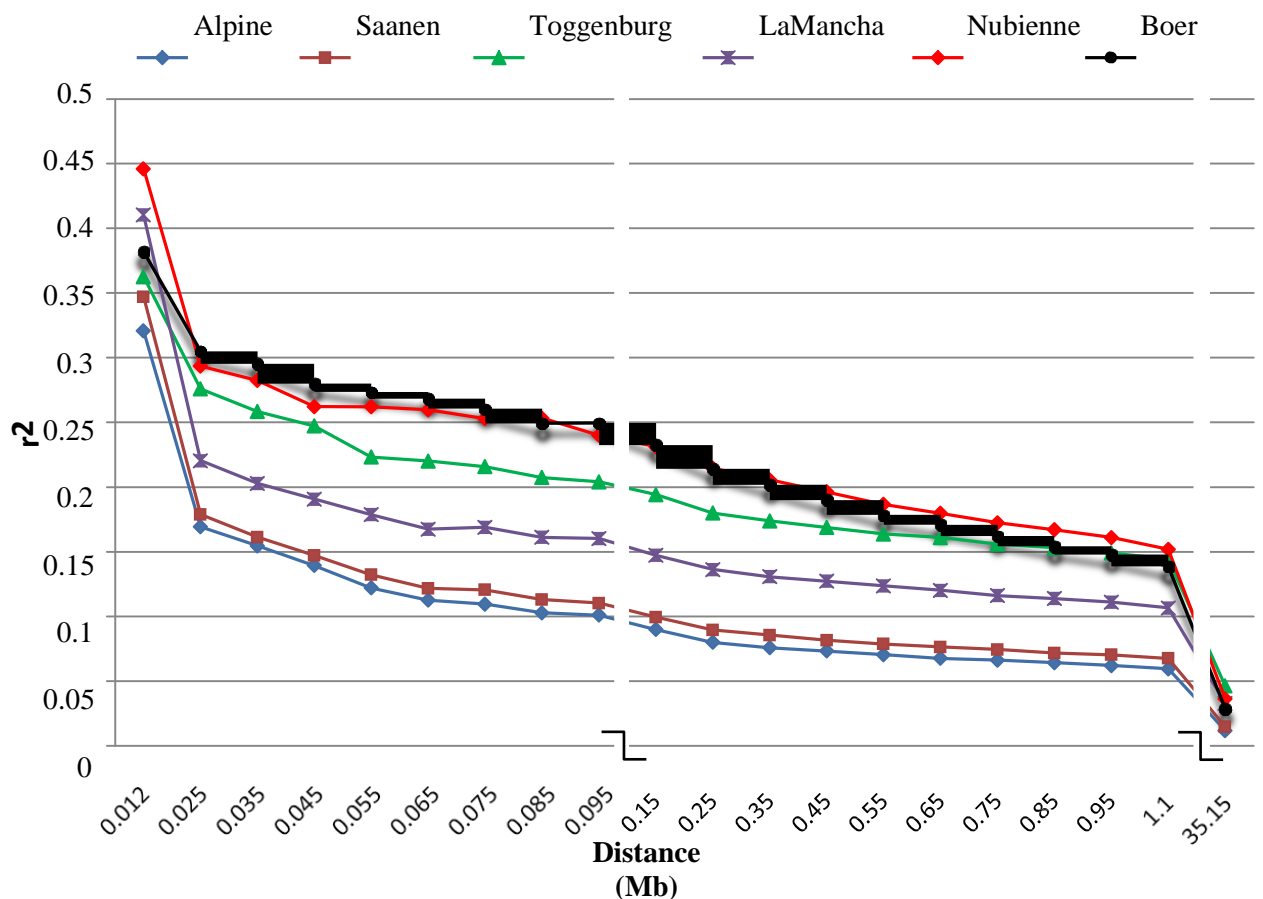


Figure 1. Valeurs r^2 moyennes à des distances données pour six races caprines canadiennes

La cohérence de phase entre races indique si les différentes races pourraient être regroupées en un groupe d'apprentissage pour mieux estimer les effets des SNP. Pour l'évaluation génomique caprine, cet aspect est très important en raison du petit nombre d'animaux génotypés dans les populations de petite taille. Si la cohérence des phases était élevée, il serait possible d'avoir plus animaux dans le groupe d'apprentissage en regroupant les races. Dans les tableaux 5a, 5b, et 5c, la cohérence des phases gamétiques (corrélations entre valeurs r) entre toutes les races est présentée. La plus grande cohérence de phase a été trouvée entre les races Alpine et Saanen, indiquant un niveau de parenté plus élevé entre ces deux races. Cependant, même pour ces deux races, la cohérence des phases entre marqueurs adjacents n'était pas assez élevée ($\sim 0,50$) pour permettre la mise en commun des races dans un groupe d'apprentissage pour la sélection génomique. En général, les corrélations estimées étaient faibles pour toutes les races même pour de courtes distances. Par conséquent, plus d'animaux de chaque race devraient être génotypés pour permettre la mise en œuvre de la sélection génomique intra-race.

Tableau 5a. Corrélations de Pearson entre les phases gamétiques de 15 paires de races ¹ .					
Distance (Mb)	AL - SA	AL - LN	AL - NU	AL - TO	AL - BO
<0,02	0,88	0,87	0,73	0,85	0,65
0,02 - 0,03	0,68	0,61	0,39	0,54	0,34
0,03 - 0,04	0,65	0,61	0,37	0,51	0,28
0,04 - 0,05	0,55	0,52	0,29	0,41	0,18
0,05 - 0,06	0,51	0,49	0,29	0,40	0,19
0,06 - 0,07	0,48	0,42	0,17	0,30	0,13
0,07 - 0,08	0,43	0,41	0,16	0,29	0,12
0,08 - 0,09	0,44	0,42	0,20	0,31	0,08
0,09 - 0,1	0,37	0,37	0,15	0,28	0,05
0,1 - 0,2	0,33	0,32	0,12	0,22	0,05
0,2 - 0,3	0,27	0,26	0,07	0,17	0,02
0,3 - 0,4	0,24	0,23	0,06	0,13	0,01
0,4 - 0,5	0,22	0,24	0,05	0,13	0,006
0,5 - 0,6	0,21	0,23	0,06	0,14	0,002
0,6 - 0,7	0,21	0,23	0,06	0,13	-0,007
0,7 - 0,8	0,21	0,20	0,04	0,12	-0,0001
0,8 - 0,9	0,20	0,19	0,05	0,11	0,01
0,9 - 1	0,20	0,19	0,04	0,11	0,003
1 - 1,2	0,20	0,19	0,04	0,11	0,006
> 1,2	0,07	0,03	0,008	0,03	0,001

¹ AL: Alpine, SA: Saanen, LN: LaMancha, NU: Nubienne, TO: Toggenburg et BO:

Tableau 5b. Corrélations de Pearson entre les phases gamétiques de 15 paires de races¹.

Distance (Mb)	SA - LN	SA - NU	SA - TO	SA - BO
<0,02	0,86	0,71	0,83	0,70
0,02 - 0,03	0,57	0,37	0,54	0,32
0,03 - 0,04	0,55	0,35	0,52	0,28
0,04 - 0,05	0,47	0,29	0,47	0,21
0,05 - 0,06	0,40	0,24	0,39	0,18
0,06 - 0,07	0,39	0,21	0,36	0,11
0,07 - 0,08	0,31	0,13	0,32	0,08
0,08 - 0,09	0,36	0,17	0,29	0,11
0,09 - 0,1	0,27	0,14	0,28	0,06
0,1 - 0,2	0,25	0,10	0,22	0,05
0,2 - 0,3	0,20	0,06	0,17	0,03
0,3 - 0,4	0,15	0,06	0,14	0,008
0,4 - 0,5	0,16	0,05	0,11	0,01
0,5 - 0,6	0,15	0,05	0,14	0,0002
0,6 - 0,7	0,15	0,06	0,11	0,002
0,7 - 0,8	0,13	0,03	0,11	0,0004
0,8 - 0,9	0,14	0,03	0,11	0,003
0,9 - 1	0,13	0,02	0,10	0,008
1 - 1,2	0,12	0,04	0,10	0,008
> 1,2	0,02	0,007	0,02	0,0002

¹SA: Saanen, LN: LaMancha, NU: Nubienne, TO: Toggenburg et BO: Boer.

Tableau 5c. Corrélations de Pearson entre les phases gamétiques de 15 paires de races ¹ .						
Distance (Mb)	LN - NU	LN -	LN -	NU -TO	NU -BO	TO - BO
<0,02	0,80	0,82	0,64	0,67	0,58	0,65
0,02 - 0,03	0,48	0,46	0,36	0,32	0,32	0,28
0,03 - 0,04	0,44	0,47	0,30	0,27	0,29	0,22
0,04 - 0,05	0,40	0,33	0,20	0,19	0,22	0,15
0,05 - 0,06	0,35	0,35	0,20	0,18	0,18	0,14
0,06 - 0,07	0,35	0,29	0,11	0,17	0,14	0,10
0,07 - 0,08	0,28	0,25	0,08	0,10	0,12	0,06
0,08 - 0,09	0,33	0,25	0,07	0,14	0,12	0,06
0,09 - 0,1	0,30	0,26	0,11	0,16	0,14	0,06
0,1 - 0,2	0,26	0,19	0,06	0,07	0,08	0,03
0,2 - 0,3	0,23	0,14	0,02	0,03	0,05	0,002
0,3 - 0,4	0,22	0,10	0,01	0,03	0,05	0,008
0,4 - 0,5	0,20	0,12	-0,001	0,04	0,04	0,006
0,5 - 0,6	0,21	0,11	0,01	0,02	0,02	0,01
0,6 - 0,7	0,21	0,10	0,007	0,04	0,04	0,02
0,7 - 0,8	0,17	0,08	0,01	0,02	0,04	0,001
0,8 - 0,9	0,19	0,08	0,02	0,02	0,02	0,004
0,9 - 1	0,18	0,10	0,002	0,02	0,03	-0,002
1 - 1,2	0,17	0,09	0,01	0,02	0,03	0,005
> 1,2	0,02	0,01	0,001	0,002	0,01	0,0002

¹ LN: LaMancha, NU: Nubienne, TO: Toggenburg et BO: Boer.

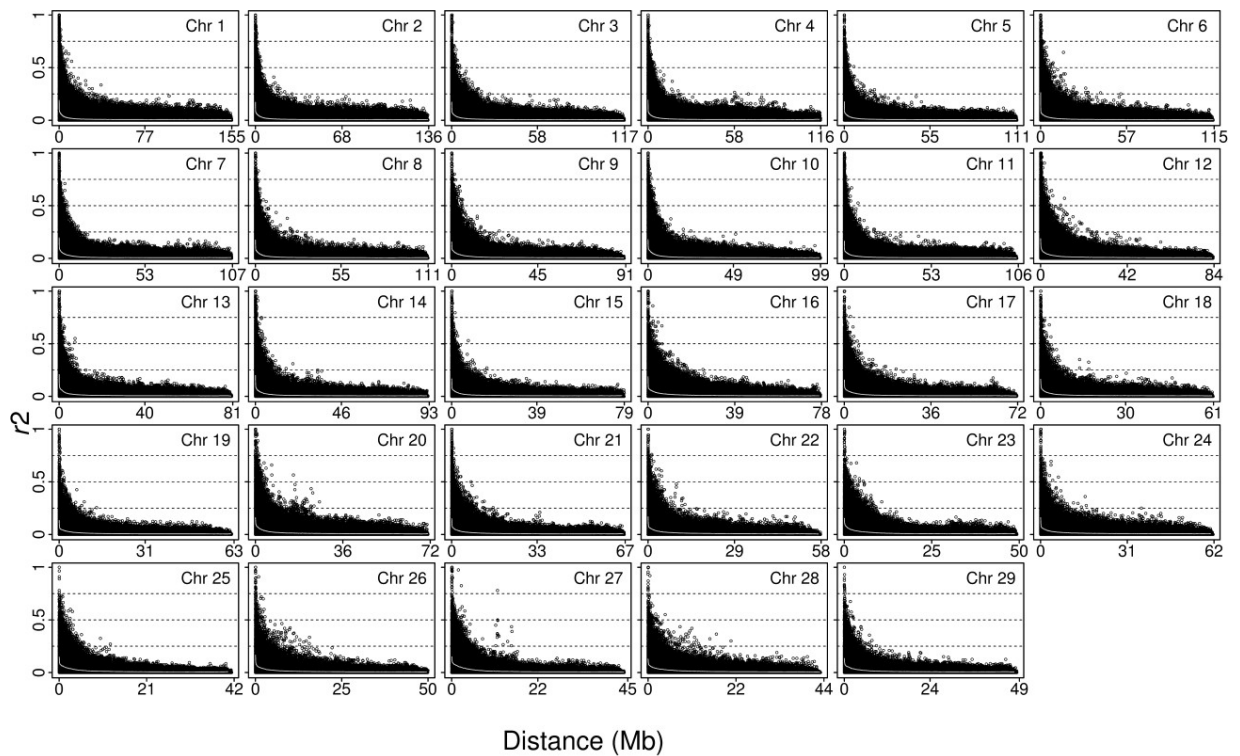
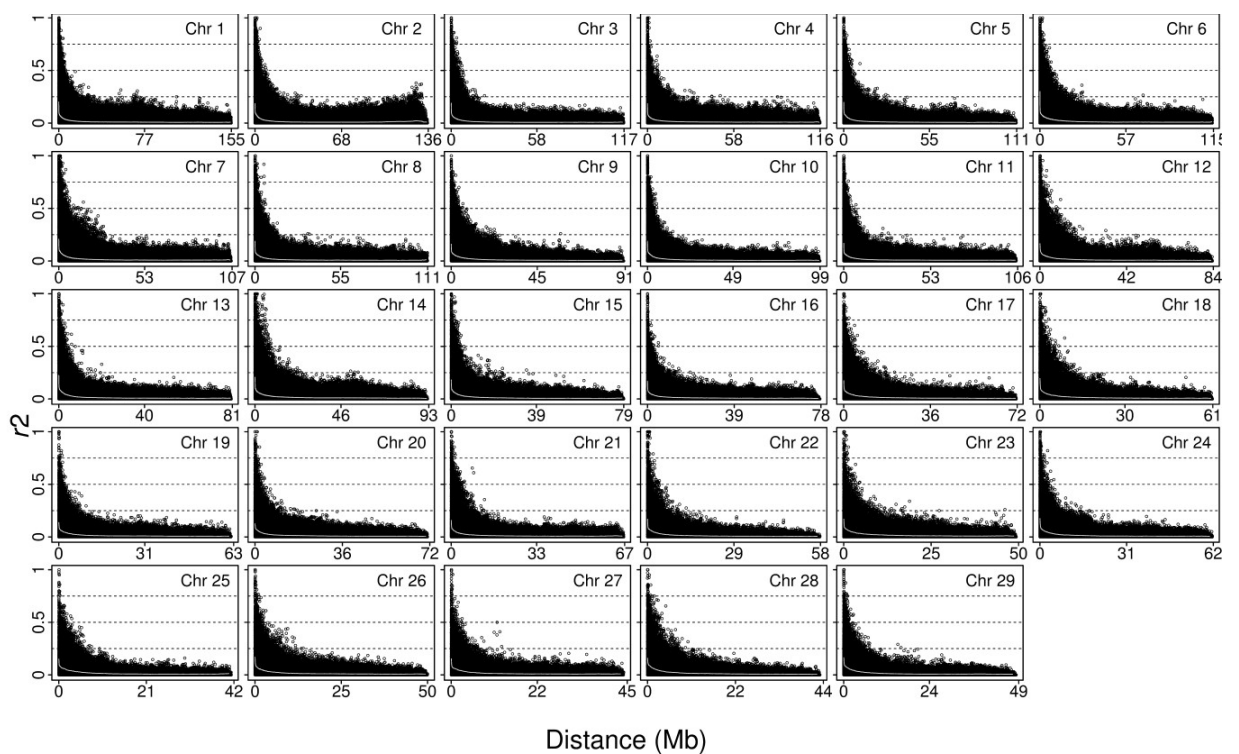
CONCLUSIONS

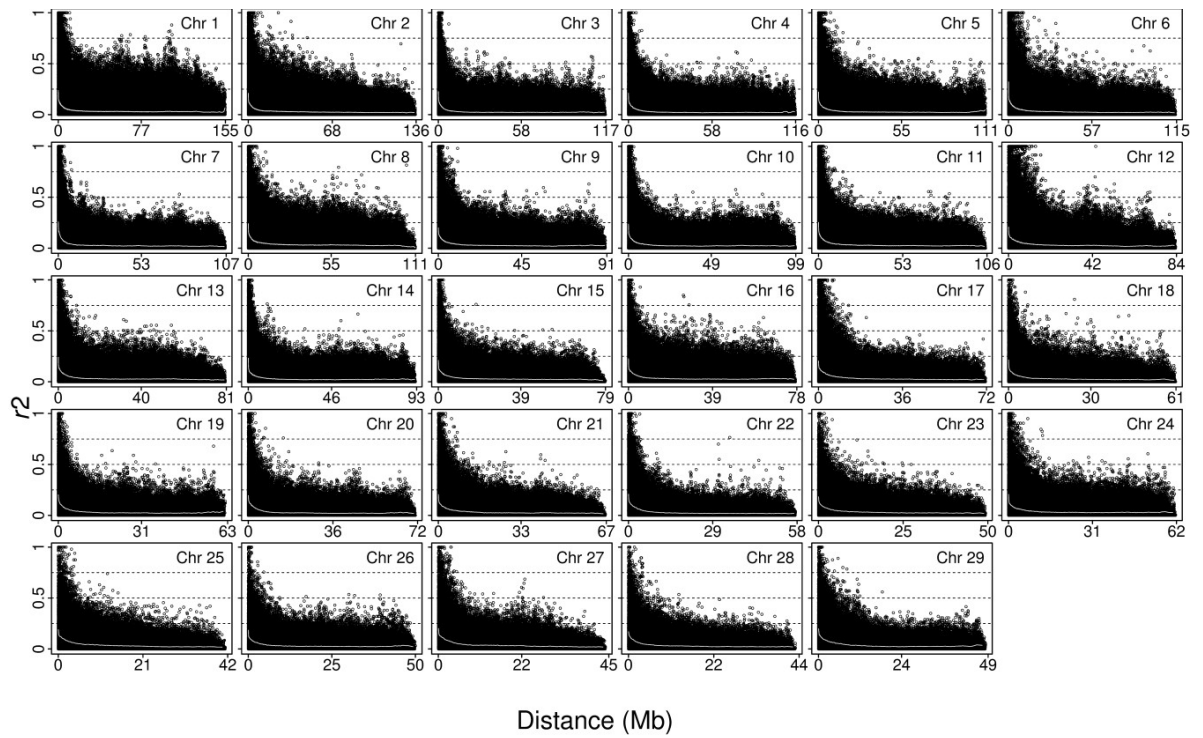
À la distance moyenne entre SNPs adjacents sur la puce 50K actuelle (~0,06 Mb), la plupart des races dépassent ou s'approchent de la valeur de déséquilibre de liaison nécessaire pour la prédiction des valeurs génomiques, à l'exception des races Alpine et Saanen. Cela indique qu'avec une population de référence assez grande, la sélection génomique intra-race pourrait potentiellement être mise en place en utilisant la puce 50K actuelle, mais les races Alpine et Saanen bénéficieraient d'une puce à SNP plus dense.

La plus grande cohérence de phase a été trouvée entre les races Alpine et Saanen, indiquant un plus haut niveau de parenté entre ces races. Cependant, même pour ces deux races, la cohérence de phase entre marqueurs adjacents n'était pas assez grande pour soutenir la mise en commun des races dans un groupe d'apprentissage pour la sélection génomique. Pour l'évaluation génomique multi-race, une puce à SNP plus dense est nécessaire. Par conséquent, d'autres moyens d'augmenter le nombre d'animaux dans le groupe d'apprentissage devraient être trouvés, tels que génotyper plus animaux dans chaque race ou collaborer avec des chercheurs d'autres pays et partager les génotypes et phénotypes (IPGs).

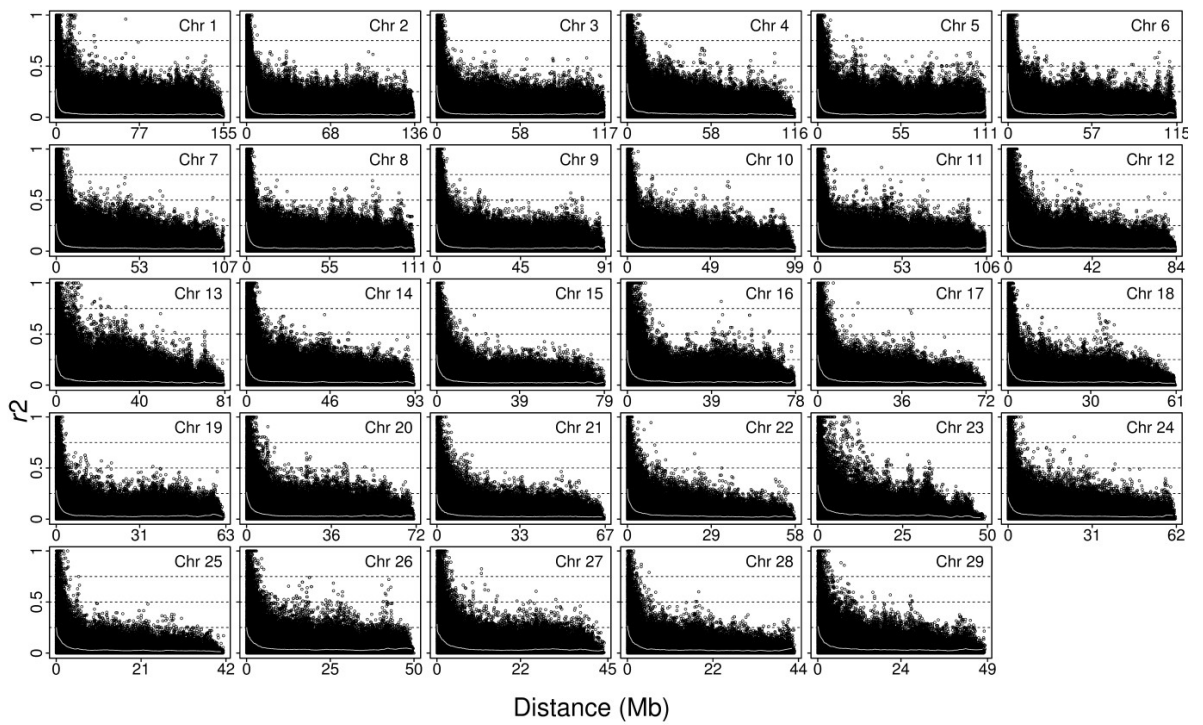
Annexe 1

Figure 2. Distribution de r^2 entre paires de polymorphismes à nucléotide unique (SNP) synténiques en fonction de la distance physique pour chaque race

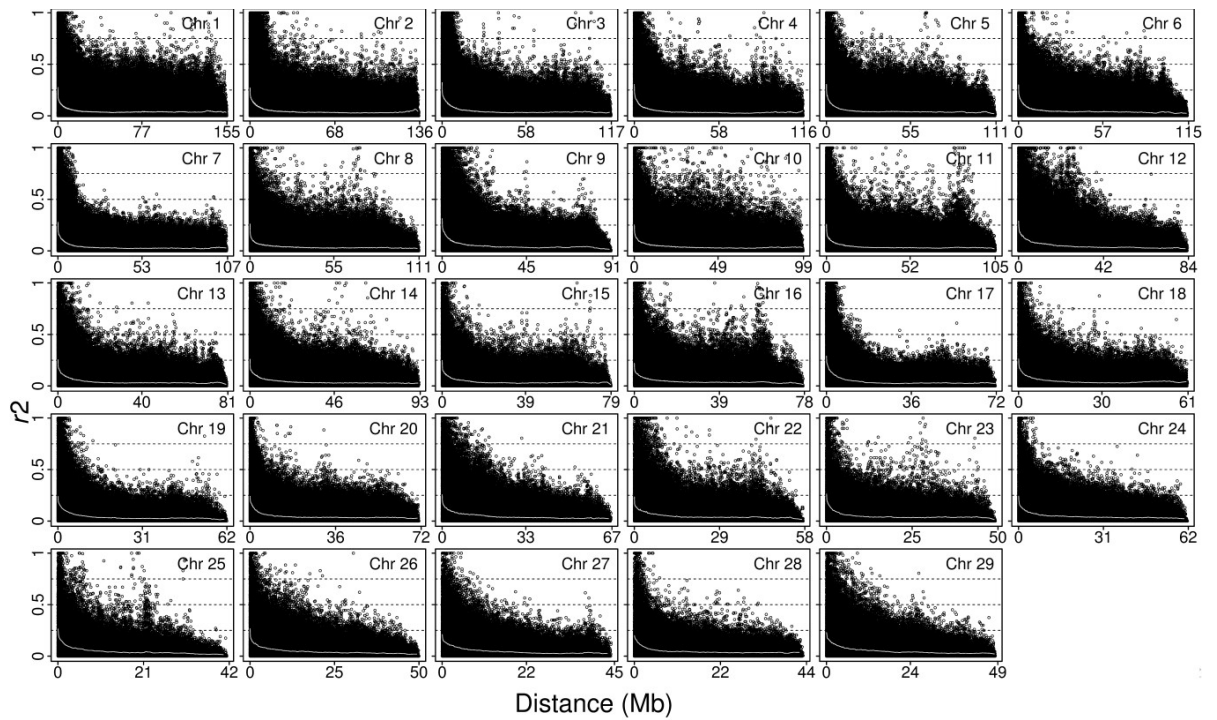
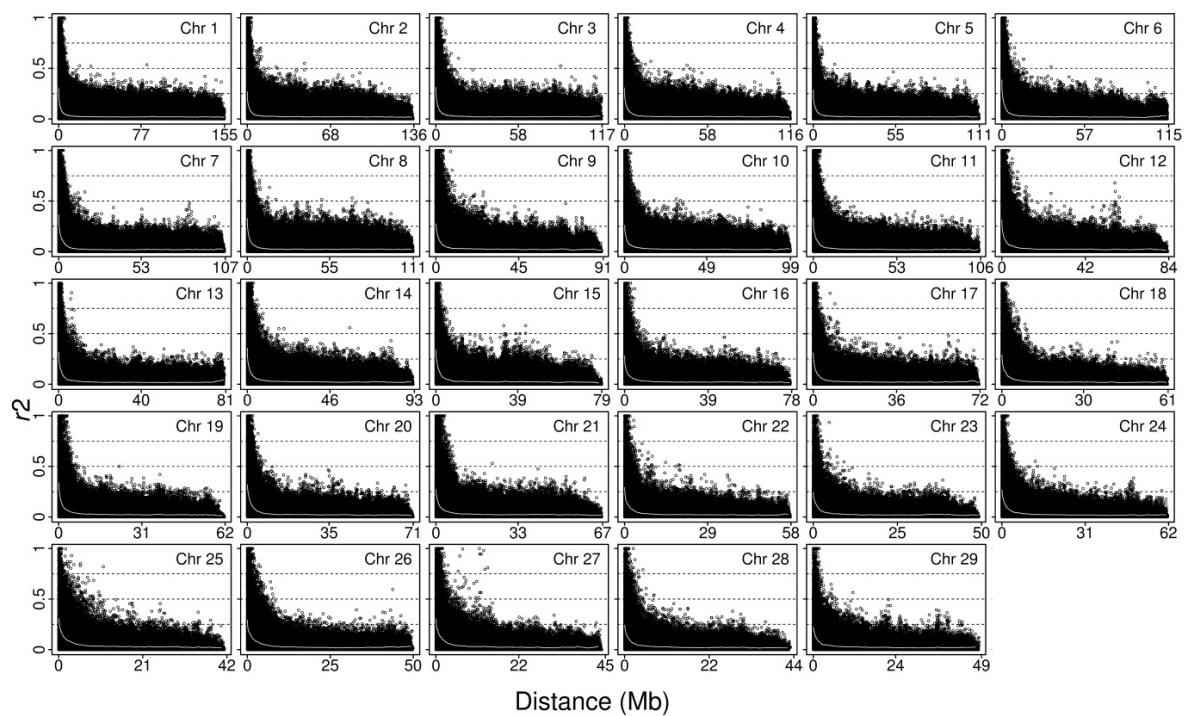
AlpineSaanen

LaMancha

Distance (Mb)

Nubienne

Distance (Mb)

Toggenburg**Boer**

B. Validation de la prédiction des valeurs génomiques chez les chèvres canadiennes de race Alpine (Résultats préliminaires)

L'évaluation génomique est devenue une procédure de routine dans la sélection de différentes espèces animales. Elle consiste à sélectionner sur la base des valeurs génomiques, calculées à partir des effets des marqueurs couvrant l'ensemble du génome. Les données disponibles sur 367 chèvres Alpine ont été utilisées pour étudier l'impact de l'inclusion des données génomiques dans les évaluations génétiques. Dans ces analyses préliminaires, la race Alpine a été choisie car elle est la race avec le plus grand nombre d'animaux génotypés. Il n'est pas possible de réaliser les mêmes analyses pour toutes les races avec le nombre actuel d'animaux génotypés. Par contre, la même analyse pourra être réalisée pour la race Saanen. Pour certaines races, il y avait seulement 60 animaux génotypés. Un plus grand nombre d'animaux génotypés est nécessaire afin d'estimer adéquatement les effets des SNPs et valider les prédictions.

Les caractères évalués sont: la production de lait, matière grasse et protéine, les cellules somatiques (comme indicateur de résistance à la mammites), les notes de conformation pour l'avant et l'arrière de la mamelle, le ligament suspenseur, les trayons, le caractère laitier, les aplombs, la capacité corporelle et l'apparence générale. Après un contrôle de qualité des données génotypiques (identiques à celui de l'analyse du déséquilibre de liaison) 45 338 SNPs sur les 53 347 SNPs initiaux ont été inclus dans l'analyse. Les animaux ont été séparés en deux groupes: un groupe d'apprentissage qui comprenait les mâles et femelles plus âgés avec plus de données de performances pour deux lactations ou plus (en fonction de leur âge) et un groupe de validation qui incluait un groupe de femelles plus jeunes ayant des performances pour deux lactations. Le nombre d'animaux inclus dans le groupe d'apprentissage était de 317 pour les caractères de production de lait, gras et protéine et 224 pour les autres caractères. Le groupe de validation incluait 50 animaux pour les caractères de production (rendement en lait, gras, protéine) et 20 animaux pour les autres caractères. Les indices de potentiel génétique (IPGs) officiels utilisés provenaient de l'évaluation génétique nationale du mois d'octobre 2013 publiée par le Centre canadien pour l'amélioration des porcs. Le logiciel GEBV a été utilisé pour estimer les valeurs génomiques directes (VGDs). Le carré de la corrélation entre les VGDs et les IPGs a été calculé pour évaluer la précision réalisée des VGDs et a été comparé à la précision des IPGs officiels pour les animaux du groupe de validation. Le tableau 1 montre les différences observées entre la précision réalisée des VGDs et la précision des IPGs officiels pour tous les caractères.

Tableau 1. Différences entre la précision moyenne observée des VGDs et la précision des IPGs classiques par caractère

Caractère	Points de précision
Production de lait	0,05
Production de matières grasses	0,04
Production de matières protéiques	0,09
Total - Production	0,04
Nombre de cellules somatiques	0,10
Caractère laitier	-0,12
Ligament suspenseur	0,18
Avant mamelle	0,15
Arrière mamelle	-0,01
Trayons	0,06
Pieds et membres	-0,01
Capacité corporelle	0,32
Apparence générale	0,24
Total - Conformation	0,10
Total	0,09

Ces résultats préliminaires pour la race Alpine indiquent qu'une certaine augmentation dans la précision des valeurs génétiques pourrait être réalisée en utilisant les informations génomiques (0,04, 0,10 et 0,10 points de précision pour la production, le nombre de cellules somatiques et les caractères de conformation, respectivement). Ces résultats doivent être interprétés avec précaution du fait du petit nombre d'animaux dans les deux groupes d'animaux (apprentissage et surtout validation). Ceci est d'autant plus vrai pour les caractères de conformation. En plus, pour les animaux dans le groupe de validation, les VGD devraient être estimées en réalisant l'apprentissage des effets de SNPs en utilisant les IPGs disponibles quand ces animaux étaient jeunes, avec aucune performance individuelle pour imiter les décisions de sélection prises tôt dans la vie de l'animal et pour prévenir l'utilisation des informations dans le calcul des IPGs qui n'étaient pas disponibles à ce moment-là. Par conséquent, les résultats présentés dans ce rapport pourraient être surestimés en termes de gain potentiel de précision.

Les valeurs génomiques seraient plus utiles pour prédire les indices de potentiel génétiques de jeunes sujets et d'animaux qui n'ont pas de performances, ce qui n'était pas le cas dans cette étude de validation.

Pour les caractères de conformation, une plus grande variabilité a été observée en ce qui concerne les résultats. Ceci est une conséquence à la fois du plus petit groupe d'apprentissage et un très petit groupe de validation. La différence globale dans la précision des caractères de conformation pourrait être un meilleur paramètre pour évaluer l'augmentation potentielle de la précision.

La capacité prédictive des VGD peut être améliorée en réalisant l'apprentissage des effets des SNP sur une plus grande population ayant des précisions d'IPGs plus élevées. Les options pour augmenter le nombre d'animaux dans le groupe d'apprentissage comprennent la collaboration avec d'autres groupes, tels que l'INRA (France) pour mettre en commun les informations génomiques disponibles.

Enfin, l'étude de validation sera peaufinée en utilisant les valeurs génétiques estimées lorsque les animaux dans le groupe de validation n'avaient pas de performances individuelles et en fusionnant les VGD avec les indices classiques. De plus, les évaluations génomiques seront également testées pour la race Saanen.

Bibliographie

- Ardlie K. G., L. Kruglyak, M. Seielstad. 2002. Patterns of linkage disequilibrium in the human genome. **Nat. Rev. Genet.** 3:299-309.
- Calus, M. (2008). Accuracy of genomic selection using different methods to define haplotypes. *Genetics*, 178, 553.
- Carillier, C., Larroque, H. Palhière, I., *et al.* (2013) A first step toward genomic selection in the multi-breed French dairy goat population. *Journal of Dairy Science*, 96: issue 11.
- Daly M. J., J. D. Rioux, S. F. Schaffner, T. J. Hudson, E. S. Lander 2001. High-resolution haplotype structure in the human genome. **Nature genetics**, 29: 229-233.
- Jeffreys A. J., L. Kauppi, R. Neumann. 2001. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. **Nature Genetics**. 29: 217-222.
- Johnson G. C. L., L. Esposito, B. J. Barratt, *et al.* 2001. Haplotype tagging for the identification of common disease genes. **Nature Genetics**. 29: 233-238.
- Khatkar M. S., F. W. Nicholas, A. R. Collins, *et al.* 2008. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. **BMC Genomics**.


9:187.

Larmer, S. 2012. Extent of Linkage Disequilibrium, Consistency of Gametic Phase and Imputation Accuracy Within and Across Canadian Dairy Breeds. MSc. thesis. University of Guelph, Guelph, Ontario, Canada.

Meuwissen, T., Hayes, B., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, 157(4), 1819-1829

Sargolzaei, M., Schenkel, F.S., and P. M. VanRaden. 2009. gebv: Genomic breeding value estimator for livestock. Technical report presented at the Dairy Cattle Breeding and Genetics Committee Meeting, October 7, 2009, University of Guelph, ON, Canada.

Slaktin, M. 2008. Linkage disequilibrium – understanding the evolutionary past and mapping the medical future. **Nat. Rev. Genet.** 9:477-485.



Utilisation de la génomique pour améliorer la productivité et la santé des troupeaux caprins

Étude d'association à l'échelle du génome

“Une partie du financement de ce projet a été fournie par les conseils sectoriels du Québec, de l’Ontario et de la Colombie-Britannique qui gèrent le Programme canadien d’adaptation agricole (PCAA) pour Agriculture et agroalimentaire Canada.”



SOCIÉTÉ DES ÉLEVEURS
DE CHÈVRES LAITIÈRES
DE RACE DU QUÉBEC



GoatGenetics.Ca
GénétiQueCaprine.Ca



Investment
Agriculture
Foundation
of British Columbia



Conseil de
l'adaptation
agricole



Agriculture et
Agroalimentaire Canada

Agriculture and
Agri-Food Canada



CONSEIL POUR
LE DÉVELOPPEMENT DE
L'AGRICULTURE DU QUÉBEC

Étude d'associations à l'échelle du génome

Comprendre la structure du génome et la distribution des SNPs ayant une influence sur différents caractères de production donne un aperçu de la « boîte noire » du génome. Les études d'associations à l'échelle du génome (GWAS) ont été réalisées dans le cadre de ce projet afin d'identifier les SNPs sur la puce à haute densité liés à des caractères d'intérêt, y compris ceux se rapportant à la production, la conformation et la santé. En sachant quels SNPs ont un effet significatif sur les caractères d'intérêt, il serait possible de réduire la liste de marqueurs spécifiques ayant un large effet et de les utiliser pour la sélection d'animaux sur des caractères spécifiques. Les GWAS aident également à chercher dans les régions voisines des SNPs significatifs, pour mettre en évidence des différences de séquence ADN qui pourraient être à l'origine de différences phénotypiques entre animaux. Cela est particulièrement intéressant pour étudier les SNPs associés avec les caractères de production, de conformation ou de résistance aux maladies chez les chèvres. Afin d'effectuer les études d'association, les indices de potentiel génétique pour les caractères de production, santé et conformation ont été utilisés comme observations pour les chèvres génotypées. Le nombre d'animaux ayant des IPGs utilisés dans les études d'association variait selon la race. Les études d'association ont seulement concerné les chèvres laitières, car il n'y avait pas de valeurs génétiques disponibles pour les chèvres de race Boer. La race Alpine a été la plus représentée dans ces études d'association tandis que la race Nubienne avait le plus petit nombre dans l'étude. Le nombre d'animaux génotypés et ayant des IPGs pour différents caractères est présenté dans le Tableau 1.

Tableau 1: Nombre d'animaux génotypés et ayant des IPGs pour différents caractères étudiés.

Race	Caractères		
	Production ¹	Santé ²	Conformation ³
Alpine	387	370	287
Saanen	295	283	221
LaMancha	74	68	43
Toggenburg	52	13	48
Nubienne	48	25	40

¹Caractères de production: Production de lait, de matières grasses et protéines (kg) par lactation de 305 jours

²Caractère lié à la santé: Comptage de cellules somatiques, lié avec la résistance à la mammites.

³Caractères de conformation: Capacité corporelle, caractère laitier, aplombs, attache avant, attache arrière, apparence générale, ligament suspenseur, trayons

Les données de conformation et de contrôle laitier sont utilisées pour calculer les indices de potentiel génétique dans le cadre du Programme canadien d'amélioration de chèvres laitières. Les valeurs génétiques étudiées sont économiquement importantes et sont utilisées de façon routinière pour le calcul des indices de production et de conformation dans le cadre du Programme canadien d'amélioration des chèvres laitières. Le compte de cellules somatiques a récemment été inclus dans les programmes d'évaluation génétique. Plus d'informations sur le Programme canadien d'amélioration des chèvres sont disponibles en ligne à l'adresse suivante : www.GenetiqueCaprine.ca. Les indices de potentiel génétique (IPGs) officiels utilisés pour les études d'association provenaient de l'évaluation génétique nationale du mois d'octobre 2013. Cette information a été combinée avec les génotypes 50K de 979 chèvres chargés dans la base de données.

Puisque l'étude de déséquilibre de liaison n'a pas mis en évidence de cohérence de phase suffisante pour permettre les études d'association toutes races confondues, les études d'association ont été effectuées dans chaque race séparément. Seules les races Alpine et Saanen avaient assez d'animaux génotypés pour étudier l'association des SNPs avec différents caractères.

Avant les analyses, une étape de contrôle de qualité a été effectuée. Les SNPs avec les fréquences de l'allèle mineur (FAM) inférieures à 0,05, un taux de réussite de génotypage inférieur à 0,90 et qui n'étaient pas en équilibre de Hardy-Weinberg ($p < 10^{-6}$) ont été exclues d'analyses subséquentes. Un total de 47 700 et 47 302 SNPs ont été inclus dans les études d'association pour les races Alpine et Saanen, respectivement. Le taux de réussite de génotypage a été mesuré pour chaque échantillon et tous les animaux génotypés avaient plus de 90% de leurs SNPs génotypés. Aucun échantillon n'a été exclu en raison d'un faible taux de réussite du génotypage.

Les analyses d'association des marqueurs génétiques avec les IPGs pour différents caractères ont été réalisées en utilisant le modèle mixte à locus unique du logiciel Golden Helix SNP & Variation Suite (SVS). Les relations entre marqueurs ont été incluses comme effets aléatoires dans le modèle mixte GWAS en utilisant la procédure EMMAX (Efficient Mixed-Model Association eXpedited) (Kang et al. 2010). La procédure EMMA (Efficient Mixed-Model Association) prend en compte la parenté entre individus afin d'éviter la surestimation des effets ainsi que les associations possiblement fausses. EMMAX utilise une approche basée sur les composantes de variance, qui est plus efficace qu'EMMA au niveau calculatoire. Le taux de fausses découvertes (FDR) à l'échelle du génome entier (Storey, 2002) a également été utilisé pour tenir compte des tests d'associations sur des milliers de SNPs sur la puce à haute densité.

Dans la race Alpine, les analyses ont permis de détecter 5 SNPs associés avec la production de lait, les cellules somatiques et l'attache avant de la mamelle. Les SNPs liés à la production de lait se trouvaient sur les chromosomes 17 et 24 et un autre SNP également situé sur le chromosome 24 était significativement associé avec les cellules somatiques. Deux autres SNPs significatifs se trouvant sur les chromosomes 1 et 4 étaient liés avec l'attache avant de la mamelle dans la race Alpine (Figure 1-3).

Figure 1: Association des SNPs sur la puce 50K avec l'attache avant chez les chèvres de race Alpine.

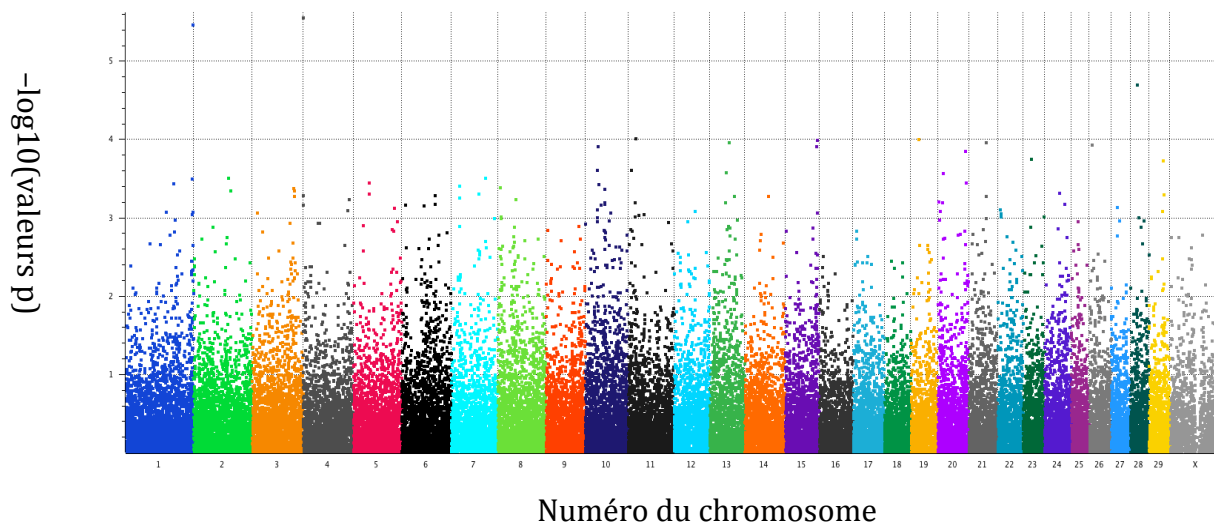


Figure 2: Association des SNPs sur la puce 50K avec la production de lait chez les chèvres de race Alpine.

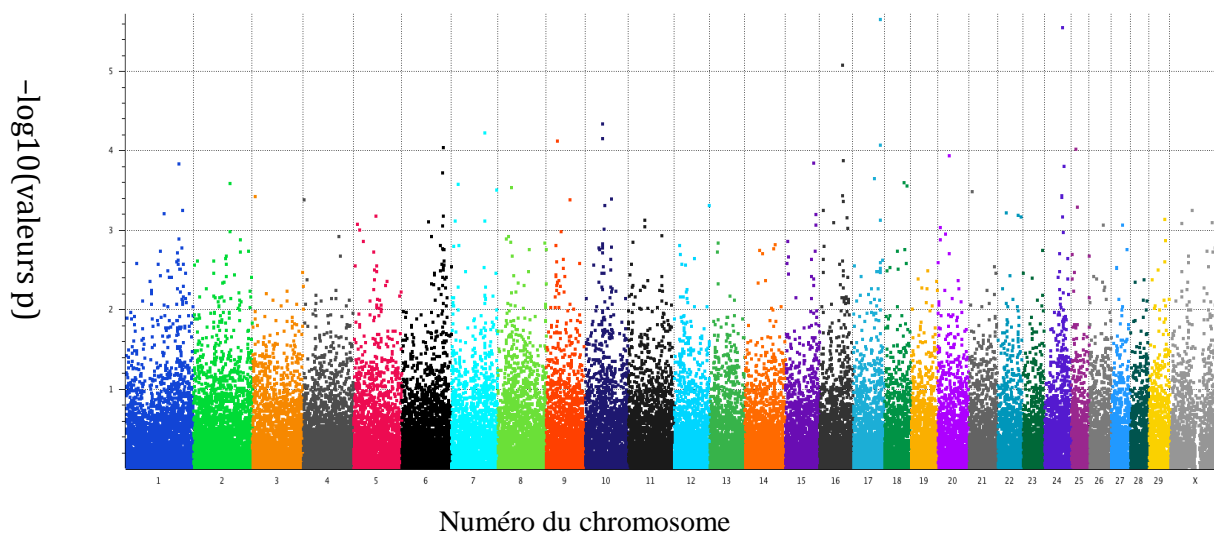
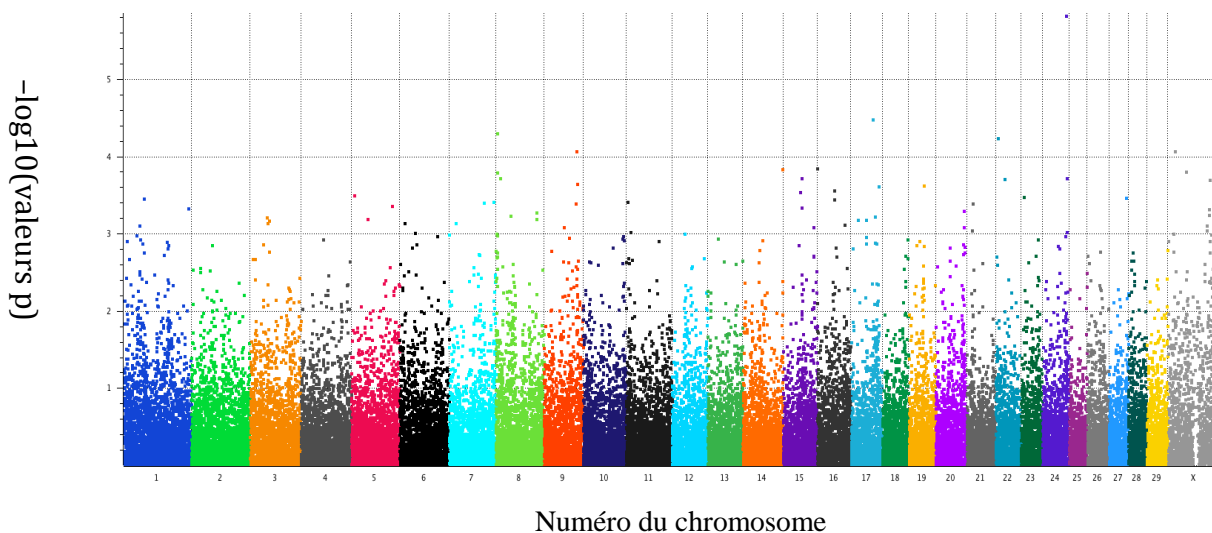


Figure 3: Association des SNPs sur la puce 50K avec le compte de cellules somatiques dans la race Alpine.



Chez la race Saanen, trois SNPs sur les chromosomes 1, 4 et 13 sont liés de façon significative avec l'attache avant de la mamelle, deux SNPs sur les chromosomes 3 et 5 sont liés avec le nombre de cellules somatiques et quatre SNPs sur les chromosomes 4, 8 et 9 sont liés avec l'emplacement des trayons (Figures 4-6). Des SNPs liés significativement avec le nombre de cellules somatiques et l'attache avant de la mamelle ont été détectés chez les deux races (Alpine et Saanen) mais ces SNPs se trouvent dans différentes régions du génome. Maroteau et al. (2013) ont également rapporté des QTL spécifiques aux races dans leur étude d'association des chèvres Alpines et Saanen françaises utilisant la puce 50K caprine.

Figure 4: Association des SNPs sur la puce 50K avec l'attache avant chez les chèvres de race Saanen.

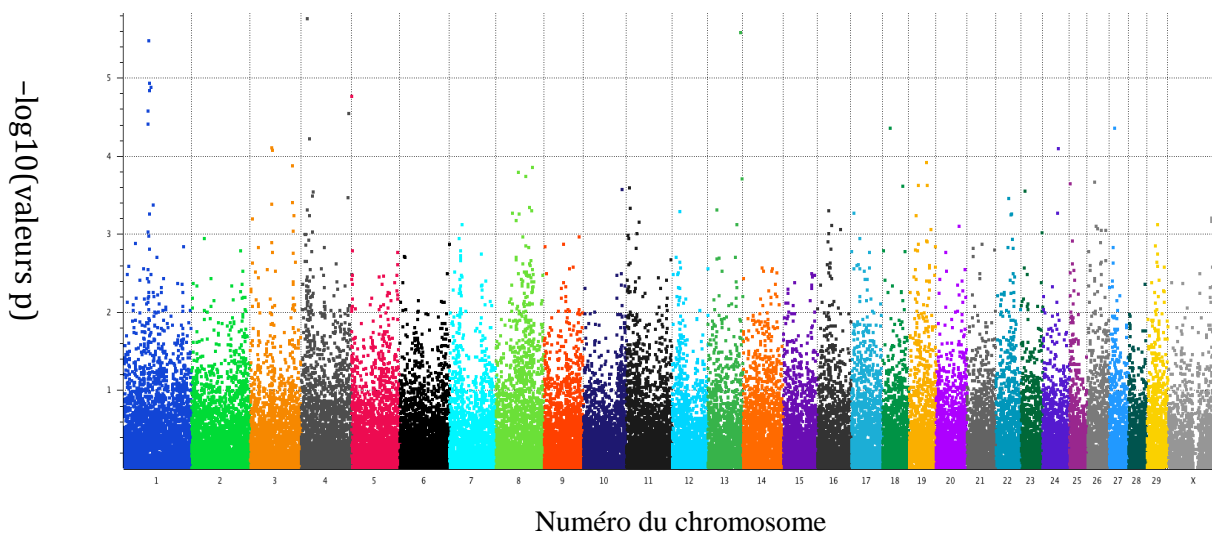


Figure 5: Association des SNPs sur la puce 50K avec le nombre de cellules somatiques chez la race Saanen.

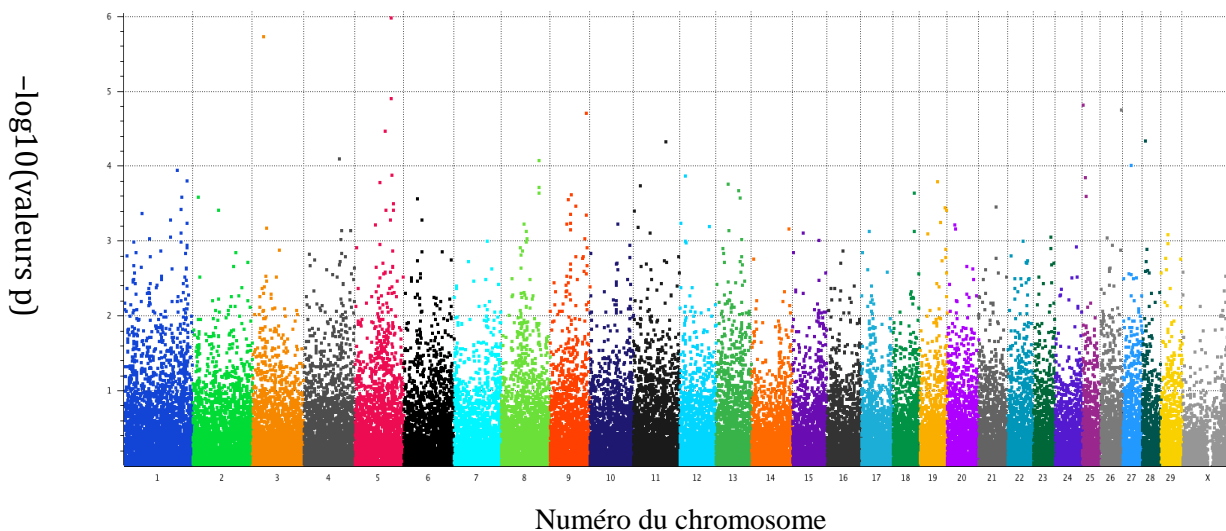
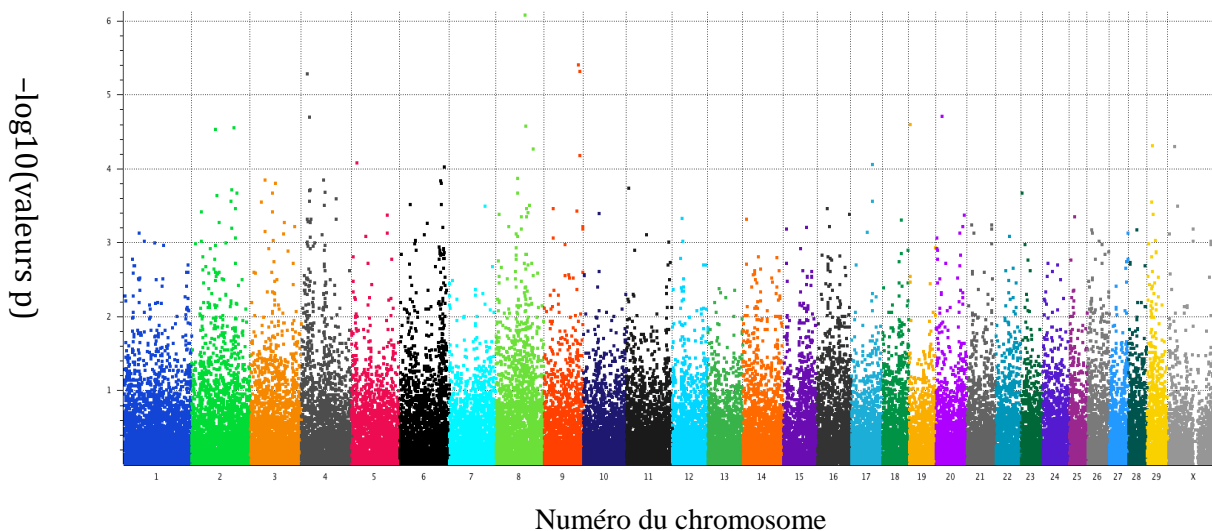


Figure 6: Association des SNPs sur la puce 50K avec l'emplacement des trayons chez les chèvres de race Saanen



Pour toute étude d'association à l'échelle du génome où des milliers de SNPs sont évalués afin d'identifier une liaison avec un caractère d'intérêt, il est nécessaire d'avoir suffisamment d'observations pour être en mesure de détecter une association significative. Dans l'étude présente, une attention particulière a été portée à la qualité des SNP et la méthode EMMAX a été utilisée, car elle est suffisamment puissante pour éviter de trouver de fausses associations. Les animaux retenus pour l'étude d'association avaient des précisions d'IPGs relativement élevées afin d'évaluer les marqueurs contre des valeurs précises. La moyenne (gamme) de précision des IPGs pour la production de lait, gras et protéines était de 0,65(0,13-0,85), 0,67(0,13-0,89) et 0,68(0,13-0,90), respectivement dans la race Alpine et 0,60(0,13-0,85), 0,63(0,13-0,88) et 0,64(0,14-0,90), respectivement dans la race Saanen.

Il est à noter que les niveaux de déséquilibre de liaison (DL) chez les races Alpine et Saanen (r^2 moyen à une distance de 0,09Mb à 0,10Mb de 0,10 et 0,11 pour les races Alpine et Saanen respectivement) étaient plus faibles que ceux trouvés chez le porc (r^2 moyen de 0,32, 0,23 et 0,24 pour les races Duroc, Landrace et Yorkshire respectivement à une distance de 0,10-0,50Mb; Daniela Grossi, résultats non publiés). Les résultats de DL soulignent que plus d'animaux et des puces à plus haute densité pourraient être nécessaires pour effectuer des analyses d'associations à l'échelle du génome chez la chèvre.

Bibliographie:

Kang HM, *et al.* (2008). 'Efficient control of population structure in model organism association mapping', *Genetics*, 178, 1709–1723.

Kang HM, *et al.* (2010). 'Variance component model to account for sample structure in genome-wide association studies', *Nature Genetics* 42, 348–354.

Maroteau, C, *et al.* (2013). 'QTL detection for traits of interest for the dairy goat industry', 64th EAAP meeting, Nantes, France.

Storey JD, (2002), 'A direct approach to false discovery rates'. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64: 479–498.