

Utilisation de données de séquence en caprins laitiers français : exploration de régions d'intérêt et amélioration de la précision des évaluations génomiques

Using sequence data in French dairy goats: exploring regions of interest and improving genomic evaluations accuracy

TALOUARN E (1), BARDOU P (2), TEISSIER M (1), CLEMENT V (3), PALHIÈRE I (1), LARROQUE H (1), TOSSER-KLOPP G (1), RUPP R (1), ROBERT-GRANIE C (1)

(1) INRAE, UMR1388 Génétique, Physiologie et Système d'Élevage, F-31326 Castanet-Tolosan Cedex

(2) INRAE, Sigénae, F-31326 Castanet-Tolosan Cedex

(3) Institut de l'Élevage, F-31326 Castanet-Tolosan Cedex

INTRODUCTION

Le projet VarGoats a pour but de fournir un jeu de données de plus de 1000 séquences pour l'espèce *Capra hircus*, dont 44 individus de race Alpine et 37 de race Saanen. Ce nombre limité permet néanmoins de constituer une population de référence pour envisager une imputation depuis des informations de typage partiel (typages 50k) vers la séquence. Cette étape permet d'intégrer de l'information fine provenant de la séquence complète des animaux dans des analyses d'association ou dans les évaluations génomiques, pour l'ensemble des individus génotypés en puce 50K. Notre objectif dans cette étude appliquée aux chèvres françaises a été de détecter de nouvelles régions d'intérêt et d'affiner les régions QTL déjà connues. Les informations ainsi acquises ont ensuite été utilisées dans les évaluations génomiques. Nous ne présenterons les résultats qu'en race Saanen, par souci de concision et car ils ont fait l'objet d'une étude approfondie.

1. MATERIEL ET METHODES

Cette étude s'est appuyée sur 757 boucs Saanen génotypés avec la puce Illumina GoatSNP50 BeadChip. Ces génotypes ont été imputés vers la séquence en intra-race à l'aide des 33 séquences Saanen, retenues après contrôle qualité. Le taux de concordance de cette imputation a été estimé¹ à 0.74. Les analyses d'association ont ensuite été effectuées sur les 23 337 436 variants imputés pour des caractères de production, de morphologie et de production de semence. Pour les évaluations génomiques, nous avons retenu plusieurs stratégies pour inclure de l'information provenant de la séquence et des analyses d'associations réalisées. Ainsi, nous avons testé d'ajouter en plus des SNP de la 50k: (1) l'ensemble des variants d'une région QTL (2) une sélection de variants dans cette région.

2. RESULTATS

2.1. CARTOGRAPHIE FINE D'UN QTL

En race Saanen, une région de 5,2 Mb du chromosome 19 est associée aux caractères de production : lait (MY), MG (FY) et MP (PY), aux cellules somatiques du lait et à des caractères de morphologie: avant-pis (FU), tour de poitrine (CD), position du plancher (UFP) et qualité de l'attache arrière (RUA). Enfin cette même région est associée à des caractères de production de semence : nombre de spermatozoïdes (SN), concentration (SC) et volume de la semence (SV) (Figure 1). L'ajout des données de séquence a permis d'affiner le QTL identifié préalablement avec des données 50K. Cette région reste complexe (largeur du signal, nombre de pics) et il ne nous a pas été possible d'identifier précisément une mutation causale, bien que plusieurs mutations candidates soient à l'étude. Il serait intéressant de génotyper plus d'animaux pour affiner les signaux.

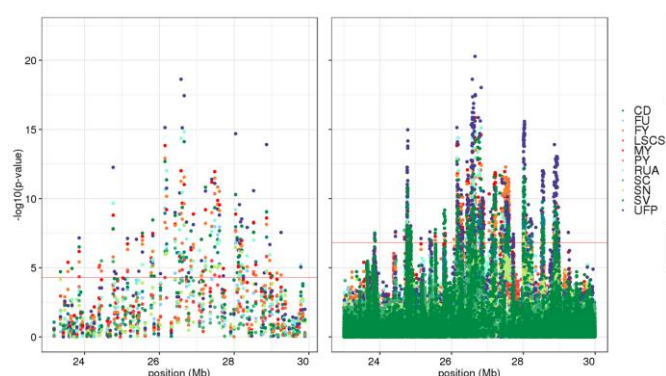


Figure 1: Manhattan plot de la région QTL sur le chromosome 19 (23-30 Mb) en race Saanen : 50k (à gauche) et séquence (à droite)

2.2. DES EVALUATIONS PLUS PRECISES

L'ajout de variants issus de la séquence dans la région du QTL pour les évaluations génomiques a augmenté la précision des prédictions (Tableau 1). Les plus forts gains de précision ont été obtenus en utilisant un ssGBLUP incluant les SNP de la puce 50k et l'ensemble des 22 269 variants imputés de la région du QTL du chromosome 19: +6,2% de précision en moyenne sur les 9 caractères évalués. La sélection de 178 variants dans le QTL constitue un compromis pour certains caractères (cellules par exemple) (entre 3,1 et 6,4% en fonction du scénario) tout en limitant les temps de calculs liés à l'imputation notamment. Ces derniers ont été ajoutés lors de la mise à jour de la puce caprine (en cours).

Stratégie	Lait	Position du plancher	Cellules somatiques
Référence= 50K	0,49	0,63	0,49
Référence +178 SEQ	0,53	0,68	0,49
Référence + 22 269 SEQ	0,55	0,69	0,46

Tableau 1: Précision des évaluations (corrélations DYD/EBV) en ssGBLUP selon l'ajout du nombre de SNP provenant de la séquence (SEQ) du chromosome 19 en race Saanen

CONCLUSION

En conclusion, l'exploitation des données de séquence est prometteuse pour les caprins laitiers français. Elle permet de mieux comprendre l'architecture génétique des caractères et d'améliorer la précision des évaluations génomiques. Il serait opportun d'étendre cette étude à la race Alpine et d'inclure l'information de l'ensemble des QTL identifiés dans les 2 races dans les évaluations génomiques.

Les auteurs remercient Capgènes et la région Occitanie.

¹Talouarn 2020. BMC Genet. doi:10.1186/s12863-020-0826-9.