

Recherche de marqueurs moléculaires d'aptitude au Concours de saut d'obstacles

Par :

- S. Brard¹, A. Ricard²
- ¹ INRA, UR 631, 31326 Castanet-Tolosan
- ² INRA, GABI, UMR1313, 78350 Jouy-En-Josas

Résumé

Une détection de QTL a été menée chez des chevaux de sport dans le but de localiser des gènes influençant les performances des animaux en Concours de saut d'obstacle (CSO). Les 1 010 chevaux génotypés sont des Selles Français (70%), des Selles Etrangers (17%) et des Anglo-arabes (13%). Sur les 54 602 SNP disponibles grâce à la puce Illumina EquineSNP50, 44 424 ont été conservés pour l'étude. Les pseudo-performances ont été obtenues après dé-régression de l'indice génétique des animaux pour le CSO afin de prendre en compte les performances propres des animaux et celles de leurs apparentés. Deux modèles ont été utilisés : le modèle mixte uni-SNP teste l'effet des SNP pris individuellement sur la pseudo-performance, le modèle haplotypique fait de même en utilisant des séquences de SNP adjacents se transmettant ensemble. Les tests ont été réalisés sur tous les animaux et sur un échantillon sans Anglo-arabes. Pour l'échantillon total les deux modèles détectent le même QTL sur le chromosome 1. Pour l'échantillon sans Anglo-arabes, seul le modèle haplotypique détecte ce QTL.

Mots clés : cheval, saut d'obstacle, sélection génomique, cartographie de QTL

Summary

QTL mapping was performed for performances of sport horses in jumping. The 1 010 horses included in the study were Selle Français (70%), Foreign sport horses (17%) and Anglo-Arabians (13%). Horses were genotyped using Illumina EquineSNP50 BeadChip. 54 602 SNP were available, and 44 424 SNP were kept in the study. Pseudo-performances were got by de-regressing breeding values for jumping competition in order to use own performances of horses and performances of their relatives. Two models were used to test effects of genotypes on pseudo-performances: a single marker mixed model (testing effects of single SNPs on performances) and a haplotype-based model (testing the effect of groups of closed SNPs). Tests were done on two samples: all animals, and Selle Français and Foreign sport horses only. In the complete sample, the two models detected a QTL on chromosome 1. In the sample without Anglo-Arabians, only the haplotype-based model detected this QTL.

Key-words : horse, jumping, genomic selection, QTL mapping

Introduction

Le développement de l'équitation de loisir et de compétition a conduit à la création et à l'amélioration de races dédiées suivant la discipline. Jusqu'à présent, la sélection des animaux a été réalisée en utilisant leur pédigrée, leurs performances propres en compétition, et celles de leurs descendants quand elles deviennent disponibles. Aujourd'hui, la génomique permet à l'aide de marqueurs moléculaires de tenter d'expliquer les performances d'animaux génotypés, et d'éventuellement localiser les régions du génome influant sur ces performances. La recherche de QTL (Quantitative trait loci) est une méthode possible. Dans le cas où un ou plusieurs QTL seraient localisés, il serait possible par comparaison avec des génomes d'autres espèces d'identifier un ou plusieurs gènes pouvant expliquer la performance, et de rechercher ces QTL chez les jeunes animaux afin d'estimer plus tôt leur aptitude au CSO. Le but de cette étude est de tenter de localiser des QTL pour le CSO en utilisant les génotypes d'un peu plus de 1000 chevaux de sport. Les données utilisées sont celles du projet JUMPSNP. Ce projet a pu être réalisé grâce aux propriétaires des chevaux qui ont participé à l'analyse, grâce à la FNC, à l'ANSF et à l'ACA qui ont soutenu le projet, grâce à LABOGENA qui a réalisé les analyses, et grâce à l'IFCE, l'INRA et le Fond Eperon qui ont financé les recherches.

1. Matériel et méthodes

1.1. Animaux

Sur les 1 101 chevaux génotypés, certains ont dû être éliminés à cause de données manquantes dans les génotypes. 1 010 chevaux ont été conservés pour l'étude. Ils sont de race Selle Français (70%), Selle Etranger (16%) et Anglo-arabe (13%). Plus de la moitié sont nés entre 1993 et 2006. 94% sont des étalons. Des relations de parenté existent dans l'échantillon : 600 chevaux génotypés ont leur père génotypé, et 200 chevaux génotypés ont aussi leur grand-père génotypé.

1.2. Phénotypes

Les animaux génotypés ont un indice génétique pour le CSO (BSO), qui a l'avantage de prendre en compte les performances propres de l'animal, mais aussi celles de tous ses apparentés, et d'être corrigé pour les effets fixes et les choix d'accouplement. Le BSO synthétise lui-même deux indices génétiques : un indice cumul annuel de points et un indice classement (Ricard *et al.*, 2012). Pour pouvoir être utilisé comme un phénotype, le BSO a été dé-régressé : une pseudo-performance, plus informative que la performance propre seule, a été calculée pour chaque cheval à partir de son indice et de ceux de ses apparentés (descendants, collatéraux et ancêtres), sauf si ceux-ci faisaient aussi partie de l'échantillon génotypé (Ricard *et al.*, soumis à *J. Anim. Sci.*).

1.3. Marqueurs

Les 54 602 marqueurs disponibles pour chaque cheval ont été soumis à des tests de qualité afin d'éliminer les erreurs éventuellement survenues au cours du génotypage. Les critères étaient les suivants : la fréquence de l'allèle minimum devait être d'au moins 5%, le taux de données manquantes devait être inférieur à 20%, et la P-value du test de Hardy-Weinberg vérifiant la cohérence des fréquences alléliques dans la population devait être supérieur à 10^{-8} . Finalement, 44 424 SNP ont été retenus pour l'étude. Le chromosome sexuel n'a pas été utilisé.

1.4. Modèles

1.4.1. Equation

Deux modèles ont été utilisés : un modèle mixte uni-SNP, qui teste l'effet des SNP pris individuellement sur les pseudo-performances, et un modèle mixte haplotypique, qui fait de même mais avec des séquences de SNP adjacents se transmettant ensemble d'une génération à l'autre : les haplotypes. Le modèle mixte uni-SNP a la forme suivante :

$$Y = Xb + Za + e$$

Avec Y le vecteur des pseudo-performances, b le coefficient de régression de l'effet d'un allèle de référence au SNP, X le vecteur des génotypes des animaux (codé 0, 1 ou 2 suivant le nombre d'allèle de

référence au SNP), a le vecteur des effets aléatoires génétiques polygéniques (avec $V(a)=A\sigma_a^2$), Z une matrice d'incidence reliant la pseudo-performance i au cheval j , et e le vecteur des résiduelles.

Les haplotypes ont été construits en utilisant le logiciel PHASEBOOK (Druet *et al.*, 2010). L'équation du modèle utilisant les haplotypes est similaire à celle du modèle mixte uni-SNP. Dans le cas du modèle haplotypique, la matrice X contient les différentes versions des haplotypes des animaux : les états cachés. Chaque ligne contient deux 1 pour les deux états cachés portés par l'individu, et des 0 sinon. b contient les coefficients de régression des effets de chaque état caché sur la pseudo-performance. Pour les deux modèles les calculs ont été réalisés avec le logiciel ASReml (Gilmour *et al.*, 2006).

1.4.2. Seuils

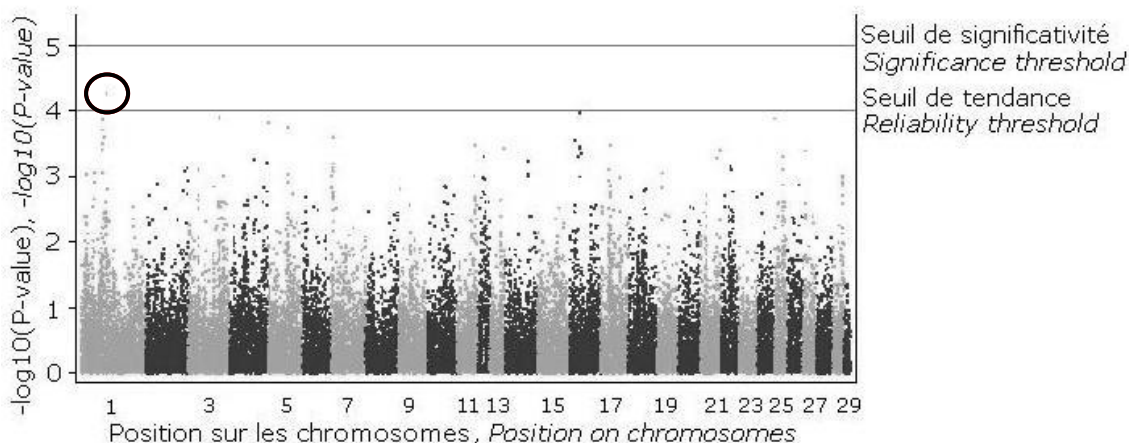
Dans le cadre de notre étude, plus de 44 000 tests indépendants sont réalisés, en conséquence le seuil de 0,05 utilisé usuellement ne convient pas : il conduirait à la détection d'un nombre élevé de faux positifs. Une correction de Bonferroni (Lander *et al.*, 1995) a été appliquée, et le seuil retenu est $P < 10^{-5}$ ($P < 10^{-4}$ pour le seuil de tendance).

2. Résultats

2.1. Résultats du modèle mixte uni-SNP

Avec le modèle mixte uni-SNP, aucun marqueur n'a une P-value atteignant le seuil de significativité. En revanche deux SNP situés sur le chromosome 1 se trouvent au dessus du seuil de tendance (P-value= $5,404 \times 10^{-5}$) (figure I). Avec le même modèle, ce QTL n'est pas détecté dans l'échantillon sans Anglo-arabes (P-value= $3,139 \times 10^{-4}$).

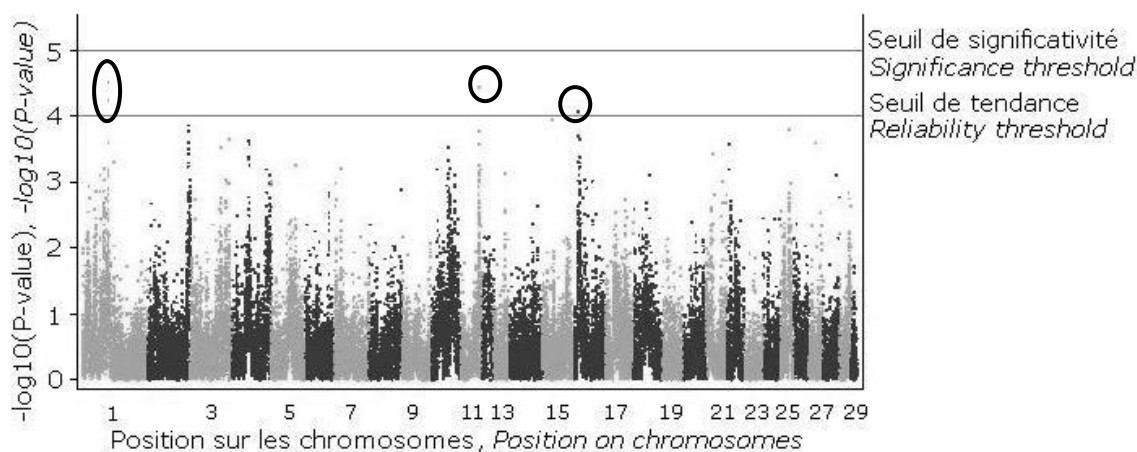
Figure I : Manhattan plot de $-\log_{10}(P\text{-value})$ avec le modèle mixte uni-SNP pour l'échantillon total
Figure I: Manhattan plot of $-\log_{10}(P\text{-value})$ with the single-SNP mixed model for the whole sample



2.2. Résultats du modèle haplotypique

Comme pour le modèle mixte uni-SNP, aucun marqueur n'atteint le seuil de significativité. Cependant trois SNP atteignent le seuil de tendance. Ils sont situés sur les chromosomes 1 (P-value= $3,048 \times 10^{-5}$), 11 (P-value= $3,700 \times 10^{-5}$) et 16 (P-value= $8,403 \times 10^{-5}$) (figure II). Avec l'échantillon sans Anglo-arabes, les QTL des chromosomes 1 et 16 sont détectés également (P-value= $1,719 \times 10^{-5}$ et P-value= $1,344 \times 10^{-5}$ respectivement).

Figure II : Manhattan plot de $-\log_{10}(P\text{-value})$ avec le modèle haplotypique pour l'échantillon total
 Figure II: Manhattan plot of $-\log_{10}(P\text{-value})$ with the haplotypes model for the whole sample



3. Discussion - Conclusion

3.1. Homogénéité des résultats

Parmi les trois QTL détectés dans les différentes associations modèle-échantillon, celui du chromosome 11 n'est détecté que dans un seul des deux échantillons. La comparaison des P-values montre que la différence de détection n'est pas due au choix des seuils. Ce marqueur serait témoin des différences entre races, alors que l'on recherche des QTL expliquant les performances en CSO quelle que soit la race.

Le SNP du chromosome 16 est détecté dans les deux échantillons, mais avec le modèle haplotypique seulement.

En revanche le QTL du chromosome 1 (1-31196) est plus intéressant. Il est détecté par les deux modèles dans l'échantillon total, par le modèle haplotypique dans l'échantillon sans Anglo-arabes, et sa P-value est faible dans l'échantillon sans Anglo-arabes étudié avec le modèle mixte uni-SNP ($P\text{-value}=3,141 \times 10^{-4}$). Sa détection dans les deux échantillons peut signifier que la région du chromosome 1 où il se situe joue un rôle dans la performance en CSO.

3.2. Exploitation de ce résultat

Le QTL du chromosome 1 est le seul détecté dans les deux échantillons. Des précautions ayant été prises dans le choix des seuils afin d'éviter de détecter des faux-positifs, il serait envisageable d'abaisser les seuils de détection. Cependant les distributions des P-values montrent que le nombre de QTL détectés augmenterait, mais sans homogénéité entre les méthodes. La performance en CSO est donc un caractère complexe et polygénique : elle n'est pas expliquée par un gène majeur mais par plusieurs régions du génome dont les effets pris individuellement ne sont pas assez importants pour être détectés.

Toutefois, même si le QTL 1-31196 n'explique pas à lui seul les performances des animaux, il va pouvoir être étudié de façon plus approfondie en recherchant de potentiels gènes d'intérêt à proximité.

Références

- Druet, T., Georges, M. 2010. A Hidden Markov Model Combining Linkage and Linkage Disequilibrium Information for Haplotype Reconstruction and Quantitative Trait Locus Fine Mapping. *Genetics* 184, 789–798.
- Lander, E., Kruglyak, L. 1995. Genetic dissection of complex traits : guideline for interpreting and reporting linkage results. *Nature Genetics* 11, 241-247.
- Ricard, A., Danvy, S., Blouin, C., *et al.* 2010. Les indices de chevaux de sport revisités. *36eme journée de la recherche équine*.
- Ricard, A., Danvy, S., Legarra, A. First results on genomic selection in French show-jumping horses and computation of deregressed proofs when own phenotypes exist. Soumis à *J. Anim. Sci.*
- Gilmour, A.R., Gojel, B.J., Cullis, B.R., Thompson, R. 2006. ASReml User Guide Release 2.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.