

La génomique : quel rôle dans l'évaluation du potentiel génétique du cheval et dans son amélioration ?

Par :

- A. Ricard
- IFCE – INRA
UMR1313 Génétique Animale et Biologie Intégrative
Equipe Biologie Intégrative et Génétique Equine
78350 JOUY EN JOSAS

Résumé

L'amélioration génétique des populations de chevaux repose sur l'évaluation génétique de leurs aptitudes. Cette évaluation permet de sélectionner les meilleurs éléments, de les faire se reproduire et d'améliorer la génération suivante. Jusqu'à aujourd'hui l'évaluation génétique reposait sur la connaissance des caractères mesurés sur les individus et sur les généalogies. La création d'un outil génomique, appelé puce SNP, qui identifie les différences dans les ADN des chevaux en 54 602 points du génome, ouvre la voie d'une part à la recherche des zones du génome ou QTL (pour Quantitative Trait Loci) qui ont un effet sur la variation de la performance et d'autre part à l'évaluation globale du lien entre génome et caractère d'intérêt afin d'augmenter ainsi la précision et la précocité de l'évaluation (sélection génomique). Une meilleure connaissance de la valeur génétique du poulain, dès sa naissance, est attendue. Pour être réalisable, cette avancée technologique nécessite la création d'une base de référence importante génotypée et mesurée pour les caractères d'intérêt. La gestion, le financement et finalement la diffusion des informations relatives à la génomique sont autant de choix à faire dès aujourd'hui.

Mots clés : cheval, QTL, sélection génomique, évaluation génétique, génétique

Summary

Genetic progress of horse population is based on breeding evaluation of production traits. This breeding evaluation is necessary to select horses, mate them and improve the whole population. Until now, breeding evaluation was based on performances and pedigree. A new tool, the Equine SNPchip, identifies 54602 differences in DNA of horses. This tool is the way to 1) find quantitative trait loci responsible for performance differences, 2) perform genomic selection, i.e. global genomic evaluation from all SNP. A more reliable and more precocious estimate of breeding value of the foal is expected. To perform such evaluation, a large reference population, with genotypes and phenotypes is needed. The management, financing, and rules for publication of results must already be debated.

Key-words : horse, QTL, genomic selection, breeding evaluation, genetic

Introduction

L'amélioration génétique des populations de chevaux repose sur l'évaluation génétique de leurs aptitudes. Cette évaluation permet de sélectionner les meilleurs éléments, de les faire se reproduire et d'améliorer la génération suivante.

Jusqu'à aujourd'hui l'évaluation génétique reposait sur la connaissance des caractères mesurés sur les individus et sur les généalogies. Sont déjà publiées les évaluations génétiques des chevaux pour l'aptitude à la compétition de saut d'obstacle, de dressage, de concours complet, leur longévité sportive (bientôt), l'aptitude à la course au trot, le pointage du modèle et des allures ainsi que des outils de gestion de la variabilité génétique.

Les nouvelles technologies de la génétique moléculaire (appelées « puces SNP »). permettent aujourd'hui de connaître la singularité génétique de chaque cheval en 50 000 points de l'ADN, base de notre hérédité. Comment peut-on utiliser cet outil pour améliorer les méthodes d'évaluation du potentiel génétique des chevaux ? Quels sont les choix qui s'offrent aux s éleveurs ?

1. Qu'est-ce qu'une puce SNP ?

A partir d'un prélèvement sanguin, ou du sperme, on extrait l'ADN du cheval. L'ADN ou Acide désoxyribonucléique, est une molécule, présente dans toutes les cellules vivantes, qui renferme l'ensemble des informations nécessaires au développement et au fonctionnement d'un organisme. C'est aussi le support de l'hérédité car il est transmis lors de la reproduction. Il porte donc l'information génétique et constitue le génome des êtres vivants. L'ADN est composé de séquences de nucléotides; on parle de polymère de nucléotides ou encore de polynucléotide. Chaque nucléotide est constitué de trois éléments liés entre eux :

- * un groupe phosphate lié à:
- * un sucre, le désoxyribose, lui-même lié à :
- * une base azotée.

Il existe quatre bases azotées différentes: l'adénine (notée A), la thymine (notée T), la cytosine (notée C) et la guanine (notée G).

Les Single-Nucleotide Polymorphisms (SNP) sont des variations (ou polymorphismes) d'une seule paire de bases du génome, entre individus d'une même espèce. Ces variations sont très fréquentes (e.g. 1/1000 paire de bases dans le génome humain). Cette variation ne prend en général que 2 formes différentes et non pas quatre. Ces formes différentes sont appelées allèles. En un point du génome, des chevaux possèdent la base A et d'autres C, ou bien A ou G, T ou C, T ou G. Une firme spécialisée dans la génomique, Illumina, a réalisé une puce SNP commerciale, à partir des connaissances du génome du cheval et en particulier du séquençage complet d'un génome d'un Pur Sang. Cette puce commerciale permet de connaître les SNP de n'importe quel cheval en 54602 positions définies a priori.

2. La puce SNP permet-elle de connaître directement les différences d'aptitudes des chevaux ?

Les SNP de la puce commerciale sont répartis partout dans le génome, à des positions connues sur les chromosomes. Mais ils ne font pas forcément partis d'une portion du génome qui sert à coder pour une protéine, donc ils ne font pas partis d'un gène particulier. Seul 1,5% du génome codent pour des protéines. De plus, même si ces SNP font effectivement parti d'un gène, leurs fonctions ne sont pas forcément connues et encore moins la modification de la fonction selon l'allèle du SNP. La puce SNP ne permet donc pas, actuellement, de connaître directement les différences de fonctionnement génétique et donc physiologique des chevaux entre eux.

3. A quoi peut elle donc servir ?

Si la puce SNP ne permet pas de connaître de façon mécanique le fonctionnement d'un individu, elle peut permettre de faire une association statistique entre les performances et les différences d'ADN. Ces associations statistiques ont pour objectif :

- de rechercher des zones du génome qui ont un effet sur la variation de la performance, et à partir de là enclencher un travail plus précis de génétique moléculaire (voir exposé L. Schibler) pour identifier les gènes variants (exemple : projet GENEQUIN)
- d'utiliser ces marqueurs pour évaluer globalement le lien entre génome et caractère d'intérêt et augmenter ainsi la précision de l'évaluation apportée par la seule connaissance des performances et des généalogies (exemple : projet JUMPSNP).

4. Recherche de QTL (Quantitative Trait Loci)

4.1. Principe

L'hypothèse est qu'il existe effectivement des variants dans l'ADN qui codent effectivement pour des différences de fonctionnement de ce gène, donc des différences physiologiques et finalement des différences de performance (au sens large, la performance peut être n'importe quel caractère d'intérêt, que ce soit l'aptitude sportive en passant par la conformation ou le tempérament). Cependant, le SNP qui correspond directement à ce variant ne figure pas sur la puce SNP commerciale. Mais comme les SNP de la puce commerciale sont répartis sur tout le génome, il existe forcément des SNP qui sont proches physiquement sur l'ADN du SNP d'intérêt. Du fait de cette proximité physique, l'allèle du SNP d'intérêt va être préférentiellement associé à un allèle particulier du SNP disponible. Ainsi, supposons que les deux allèles du SNP d'intérêt sont T et G et les deux allèles du SNP visible sont A et C. La fréquence des chevaux avec T sur le SNP d'intérêt et par exemple A sur le SNP visible sera plus importante que la fréquence des T-C. Ainsi, connaître l'allèle sur le SNP visible permet de connaître l'allèle au SNP d'intérêt. La relation n'est pas stricte : il existe des T-C et des G-A mais les différences de fréquences permettent de prédire en probabilité l'allèle au SNP d'intérêt à partir de l'allèle au SNP visible.

C'est la proximité physique qui est responsable en parti de cette association. Tous les individus possèdent deux copies de chaque chromosome, et donc chaque gène. Lors de la méiose qui détermine la répartition des gènes dans les gamètes pour produire la génération suivante, les chromosomes se mélangent et de nouvelles associations se créent par crossing-over entre les chromosomes homologues. Ainsi, un morceau de chromosome est coupé et rattaché au morceau de chromosome de l'autre chromosome de la même paire. Si la coupure a lieu entre les 2 SNP qui nous intéressent, l'association est rompue : le T se retrouve avec l'allèle porté par le chromosome homologue. Mais les coupures sont plus fréquentes entre les paires de bases éloignées qu'entre celles qui sont proches. Ces crossing-over diminuent les associations entre portions d'ADN éloignées. Les associations ne demeurent qu'entre SNP proches. Cependant, la proximité physique n'est pas la seule explication à des associations statistiques, cette règle n'est valable qu'en général.

4.2. Tests et modèles

Pour rechercher un QTL, le principe est donc simple : l'effet des SNP de la puce sur la performance est testé, si ce test est positif, la présence d'un gène responsable de la performance à proximité est suspecté.

La détection de QTL se complique car les tests peuvent être positifs pour de mauvaises raisons. En effet, par exemple, des chevaux apparentés ont à la fois plus de chance d'avoir des allèles similaires à un SNP et des performances similaires si le caractère est héritable. L'association entre le SNP et la performance est donc automatique, sans qu'il existe à proximité du SNP étudié un SNP causal. Il faut donc tenir compte de la parenté ambiante ainsi que de l'héritabilité polygénique du caractère pour effectuer la détection. C'est l'utilité du modèle mixte qui inclut l'effet du SNP étudié et un effet aléatoire polygénique simultanément.

Pour plus de précision, il peut aussi être intéressant de ne pas tester l'effet d'un seul SNP mais d'un haplotype, c'est-à-dire un ensemble de SNP associés à proximité sur le même brin d'ADN. Des haplotypes identiques ont plus de chance de renfermer le même variant au SNP d'intérêt car cet haplotype a des chances d'avoir été transmis directement par le premier cheval porteur de la mutation du SNP d'intérêt sans recombinaison. Ce cheval pouvait avoir en commun un SNP proche avec d'autres chevaux, mais beaucoup moins fréquemment tout l'ensemble de SNP proches dont la combinaison est, si ce n'est unique, rare. Tester l'effet de l'haplotype se révèle alors plus discriminant que de ne tester que l'effet d'un seul SNP. Les chevaux qui ont le même haplotype ont plus de chance d'avoir le même SNP causal que ceux qui ont en commun un seul SNP proche.

Un exemple pratique d'application est présenté dans les deux articles du projet GENEQUIN de recherche de marqueurs moléculaires de l'ostéochondrose et de la neuropathie laryngée (Teysseire *et al.*, Dupuis *et al.* ce document)

5. Sélection génomique (ou évaluation génomique)

La sélection génomique repose sur les mêmes principes biologiques que la recherche de QTL : les SNP situés dans les gènes ont un effet sur la performance mais la sélection génomique s'affranchit de la nécessité de localiser ces gènes et suppose qu'ils sont nombreux à intervenir dans la performance. Ainsi tous les SNP sont inclus dans le modèle explicatif, sachant qu'un grand nombre d'entre eux n'auront finalement aucune influence, c'est la somme d'une multitude d'effets qui intervient dans l'estimation de la valeur génétique. Le terme de sélection génomique s'est imposé dans la littérature mais il serait plus juste de parler d'évaluation génomique.

5.1. Sur quoi repose son efficacité ?

Les 50 000 points d'ADN permettent d'estimer correctement, même en l'absence de généalogie, le coefficient de parenté entre deux chevaux. Le coefficient de parenté est la proportion de gènes en commun entre deux individus. La généalogie permet d'estimer le coefficient de parenté en terme de probabilité. Ainsi, entre deux demi-frères, la proportion de gènes en commun est en espérance $\frac{1}{4}$: tous deux héritent la moitié des gènes de leur père mais il n'y a qu'une chance sur deux que le gène hérité soit le même puisque le père a deux copies de chaque gène. Alors que le coefficient de parenté généalogique ne raisonne que par probabilité, la puce SNP permet de mesurer la parenté réalisée. D'où surtout des différences pour les chevaux ou la variabilité autour de l'espérance du coefficient de parenté est plus grande. Entre deux pleins frères il peut y avoir 100% de gènes en commun, comme 0%, et en moyenne bien sur 50%. Entre un parent et son produit, il y a forcément 50% de gènes en commun. Ainsi, les différences d'apparentement entre pleins frères peuvent être mis en évidence.

S'il ne s'agissait que d'estimer au mieux les apparentements, l'efficacité de la sélection génomique serait limitée, notamment pour les relations de parenté bien déterminée (par exemple père/produit). L'avantage de la sélection génomique est de mettre l'accent sur la parenté des SNP qui ont effectivement un effet sur la performance au détriment des autres. L'avantage par rapport à la recherche de QTL est que tous les SNP liés statistiquement pour une raison ou une autre aux SNP causaux sont efficaces, sans que se pose le problème de leur localisation. Le jeu de SNP efficaces peut changer dans le temps, il est basé sur une association statistique qui n'a pas besoin d'être physique.

5.2. Quel rapport avec les indices génétiques actuels ?

Les indices génétiques actuels utilisent la généalogie pour mesurer le degré d'apparentement entre les chevaux et donc le degré de ressemblance de leurs performances. L'indice génétique d'un cheval combine les performances des chevaux qui lui sont apparentés en fonction du coefficient de parenté. Les indices génomiques font de même en remplaçant le coefficient calculé par la généalogie par celui réalisé par les SNP, donc plus précis et plus axés sur les SNP qui influencent réellement la performance. Il est possible de calculer un indice en combinant les deux sources d'information donc de créer une évaluation génétique à partir de chevaux génotypés et de chevaux dont on connaît seulement les généalogies.

5.3. Démonstration de l'efficacité – différents modèles

L'efficacité de la sélection génomique, ou plutôt de l'évaluation génomique à partir des données moléculaires, comme on devrait l'appeler, a été démontrée en production bovine laitière. En France, la précision (corrélation entre valeur estimée et valeur vraie) obtenue sur l'évaluation génétique de la quantité de lait d'un taureau laitier à partir de son seul typage SNP est de 0,56 contre 0,38 obtenu par les indices génétiques des parents (projet AMASGEN). De nombreux modèles statistiques ont été testés, mais aucun n'a montré de supériorité systématique en fonction des caractères étudiés. Le modèle le plus simple, qui explique la performance par la somme des SNP de la puce, permet d'obtenir de très bons résultats.

Chez les chevaux de sport, avec la collaboration de la FNC, de l'ANSF, de l'ANAA et de tous les éleveurs volontaires et le financement de l'IFCE, de l'INRA et du Fonds Eperon, nous avons en cours le projet JUMPSNP. Le projet a pour objectif de tester la validité du modèle d'évaluation génomique pour les chevaux de sport à partir d'un échantillon de 900 étalons de sport génotypés. Les derniers typages ont

eu lieu au moins d'octobre 2010 et seuls les premiers résultats sont mentionnés dans l'article de cette journée (Ricard *et al.*, ce document)

5.4. Avantages de la sélection génomique

Ces nouvelles évaluations apportent deux avantages principaux :

- une évaluation beaucoup plus précoce et précise pour les caractères de performance déjà indexés. La même précision est espérée pour un poulain à sa naissance à partir d'une simple prise de sang que pour un cheval cumulant plusieurs années de concours ou une dizaine de produits performants.

- une évaluation pour des caractères très tardivement disponibles (longévité, fertilité) où très coûteux à mesurer sur une large population (santé) pourrait être mise au point.

6. A quoi ça sert de savoir ? utilisation pratique des résultats

La connaissance d'une mutation dans un gène ayant un effet sur la performance (QTL) et l'estimation par la sélection génomique permettent toutes deux une meilleure évaluation précoce de l'aptitude du cheval. Cette connaissance doit permettre une meilleure valorisation par des soins appropriés et une meilleure orientation du cheval dans sa carrière. Il faut cependant garder à l'esprit que seul un faible pourcentage de la variabilité des performances est dû à la génétique, l'influence du milieu est et restera importante pour la majorité des caractères complexes qui nous intéressent (à la différence de caractères beaucoup moins influençables, comme les robes par exemple). A titre individuel, le pouvoir prédictif de cette évaluation sera limité.

Toutefois une meilleure précision de l'évaluation génétique obtenue plus tôt permet avec certitude une augmentation du progrès génétique sur l'ensemble de la population. Nous avons montré (Dubois *et al.*, 2008), que le progrès génétique, dans les conditions d'élevage des espèces équinées, était conditionné par la précision disponible au moment du taux de sélection le plus réduit, soit avant la mise en testage dans un système type germanique qui limite l'accès aux stations, soit juste après les performances en compétition dans le système Français. L'augmentation précoce de la précision ne peut qu'avoir un effet positif important. En obtenant dès la naissance une précision qui n'était obtenue qu'à 5 ou 6 ans on s'ouvre la possibilité de diminuer l'intervalle de génération de 3 ans en mettant à la reproduction de jeunes mâles et d'augmenter le progrès génétique de 50%. Mais si la sélection sur l'aptitude est aujourd'hui assez efficace quand on respecte les principes de la sélection annoncée dans cette publication, elle est limitée pour tout ce qui concerne la santé et la résistance à l'effort. Un progrès notable peut être attendu sur ces caractères car c'est le manque de précision actuelle et le faible nombre de données disponibles qui limite la sélection.

Enfin, la mise en évidence de processus biologique jusque là ignoré est aussi la voie ouverte à une meilleure prise en charge des pathologies et une optimisation de l'entraînement.

7. Que faut il faire pour bénéficier de ces technologies ?

Mais pour réaliser cette avancée, plusieurs questions sont dans les mains des éleveurs.

7.1. Population de référence

Cet outil nécessite l'existence et la maintenance d'une population de référence importante à la fois génotypée et mesurée pour les caractères d'intérêt. Cette population doit se renouveler constamment pour rester un échantillon génétique en liaison avec l'ensemble de la production. Il faut donc continuer à enregistrer et stocker très largement les performances comme aujourd'hui, ce n'est que leur utilisation qui est décalée dans le temps. Mais qui gère, qui maintient, qui participe à l'élaboration de cette population utile pour tous ? Les entreprises privées qui proposeront des tests, se créeront *de facto* une population de référence avec les chevaux qui donneront leur ADN. Mais ne faut il pas mieux pour les éleveurs français rester maître de leurs choix et décider d'eux même de participer à cette entreprise forcément collective par un acte volontaire ? Faut il créer une banque d'ADN pour tous les produits nés ?

7.2. Caractère « difficiles », phénotypage fin

Les populations chevalines bénéficient depuis longtemps d'un recueil généralisé des performances réalisées en courses ou en concours. C'est une base de données inestimable pour réaliser toutes ces évaluations car elles donnent des renseignements directement sur l'aptitude recherchée mais aussi sur des aptitudes connexes : longévité dans la discipline, reproductibilité des performances. Par contre, nous n'avons pas de recueil systématique de caractères plus difficiles à mesurer comme des mesures biochimiques, des test vétérinaires ou radiographiques, des tests de fertilité etc... Pour obtenir des indices sur les caractères « difficiles » il faut mettre en place le recueil de mesures fines de ces caractères sur un échantillon. Qui réalise, finance, participe à cette œuvre ?

7.3. Diffusion des évaluations génétiques

Le rôle de l'état est d'organiser la libre concurrence entre les détenteurs de matériel génétique (les reproducteurs par excellence) par des informations fiables et transparentes. Les indices génétiques sont réglementés par la loi d'orientation agricole afin de remplir cette mission. Les généalogies sont gérées par l'IFCE, les compétitions équestres par la FFE et les résultats en course par les fédérations de course (France galop, la SECF). Toutes ces structures mettent aujourd'hui en commun leur données pour le bien général. Mais qu'en sera-t-il demain quand certains auront contribué à l'élaboration de la base de référence génomique et d'autres non ? Qui bénéficiera des résultats qui n'auront de valeur que par la création d'une base de référence ?

8. Conclusion

Sans faire de miracles, les outils de la génomique doivent permettre une amélioration de la connaissance des aptitudes génétiques de chaque cheval. Ces connaissances peuvent permettre une meilleure valorisation et orientation dans la carrière de chaque cheval. Cependant, ces informations ne seront disponibles qu'à la condition de la réalisation d'une base de référence importante génotypée et phénotypée sur tous les caractères d'intérêt. L'élevage est donc à un tournant. Les éleveurs ont aujourd'hui le choix de s'orienter dans cette voie, de rester maître des informations disponibles ou de laisser le champs libre à des évaluations non validées ou à des firmes qui élaboreront pour leur propre profit leur propres modèles.