



Quel futur pour l'amélioration génétique chez les espèces animales domestiques ?

Didier BOICHARD, Pascal CROISEAU, Sébastien FRITZ, Vincent DUCROCQ
INRA, UMR1313 Génétique Animale et Biologie Intégrative, 78350 Jouy en Josas

Manuscrit révisé le 24 mars 2014 - Publié le 16 avril 2014

Résumé : La sélection génomique est une méthode de sélection qui s'appuie sur une évaluation du potentiel des candidats à partir de marqueurs moléculaires couvrant tout le génome. Elle repose sur l'analyse d'une population de référence disposant à la fois de phénotypes et de génotypes, permettant d'établir les équations de prédiction que l'on utilise ensuite sur les candidats génotypés. La sélection génomique peut être précoce, peu coûteuse si le génotypage est bon marché, efficace sur tous les caractères dès lors que la prédiction est précise. Elle déconnecte l'obtention des phénotypes, utilisés pour l'établissement des formules de prédiction, de la sélection des candidats. S'il existe une variabilité génétique, elle offre donc des possibilités nouvelles de sélection sur des caractères classiquement difficiles à sélectionner comme la santé des animaux. Depuis que les outils de génotypage sont disponibles, la sélection génomique se développe très rapidement chez les bovins laitiers chez lesquels elle constitue une innovation de rupture qui a mis fin au testage sur descendance. Elle s'étendra à l'avenir, sans doute dans toutes les espèces mais à des degrés divers en fonction de son intérêt économique. L'intérêt est d'autant plus important que l'intervalle de génération est long, la valeur d'un reproducteur élevée et le phénotypage difficile ou coûteux. Les principaux enjeux dans le futur sont les suivants : réduire les coûts pour favoriser le développement à grande échelle ; élargir le panel de caractères sélectionnés, particulièrement les caractères de santé, pour une production durable ; développer des méthodes présentant une bonne robustesse de prédiction au manque d'apparement entre population de référence et candidats à la sélection ; à terme prendre en compte les interactions entre gènes et les interactions génotype x milieu. Des applications naîtront au-delà de la sélection au sens strict, par exemple dans la gestion du plan d'accouplement ou la minimisation de la perte de variabilité. D'un point de vue organisationnel, une plus grande intégration est attendue entre sélection et production des phénotypes. Compétition et différenciation se développeront entre acteurs mais aussi naîtront des collaborations nouvelles pour gagner en efficacité.

Mots clefs : sélection génomique, marqueurs génétiques, objectif de sélection.

1. Le développement de la génomique et des outils de génotypage

La sélection génomique consiste à sélectionner des reproducteurs sur la base de leur valeur génétique prédite à partir de marqueurs génétiques répartis sur le génome. Cette approche inventée il y a plus de 10 ans (Meuwissen *et al.* 2001) est restée d'abord sans application, faute de disposer des outils génomiques adéquats. Il faut en effet pouvoir génotyper les individus, c'est-à-dire déterminer leur génotype, pour un grand nombre de marqueurs couvrant tout le génome. Des progrès considérables ont été réalisés au cours des années 2000 dans ce domaine

très technologique. D'une part, le génome d'un nombre croissant d'espèces d'élevage a été séquencé, donnant accès à la connaissance de millions, voire de dizaines de millions de marqueurs. Les marqueurs les plus fréquents sur le génome sont dénommés SNP (pour Single Nucleotide Polymorphism). Un SNP est un polymorphisme simple qui consiste en la substitution d'une base d'ADN par une autre. Les SNP n'ont généralement que deux allèles, l'allèle ancestral et l'allèle dérivé issu d'une mutation de l'allèle ancestral.

D'autre part, des dispositifs dénommés « puces à SNP » ont été élaborées et commercialisées par deux compagnies principales, Illumina et Affymetrix. Ces puces permettent de génotyper des individus pour plusieurs dizaines de milliers de marqueurs simultanément. Ce sont des dispositifs miniaturisés comportant des sondes d'ADN correspondant à chacun des deux allèles des milliers de marqueurs testés. Pour réaliser le génotypage, une petite quantité d'ADN est extraite à partir d'un tissu quelconque de l'individu (sang, cartilage, muqueuse buccale ou nasale, sperme, muscle...), amplifiée par PCR (Polymerase Chain Reaction pour amplification en chaîne par la polymérase) et hybridée de manière très spécifique sur la puce. L'hybridation est associée à un signal lumineux qui est lu avec un scanner à très haute résolution. L'hybridation préférentielle sur la sonde de l'allèle A, ou de l'allèle B ou des deux allèles A et B permet de déduire le génotype de l'individu, respectivement AA, BB ou AB.

Des puces à SNP, permettant de typer plusieurs dizaines de milliers de marqueurs simultanément sont disponibles dans la majorité des espèces d'élevage. Ainsi, chez le bovin, la puce de référence, commercialisée depuis 2008 par Illumina, permet de tester 54 000 marqueurs SNP simultanément. Elle a été complétée depuis par une puce à haute densité (777 000 marqueurs) et par une puce basse densité (7000), bon marché.

2. Principe de la sélection génomique

Les caractères que l'on cherche à sélectionner ont un déterminisme généralement complexe : ils sont soumis d'une part à l'effet du milieu, d'autre part à un effet génétique constitué de l'effet de nombreux gènes. La valeur génétique est définie comme l'espérance de ce que transmet un individu à ses descendants, en écart à la moyenne de la population. Cette valeur génétique est la somme des effets de nombreux gènes. Pour sélectionner de façon rationnelle, on estime la valeur génétique des candidats et sont retenus les individus présentant la supériorité souhaitée de valeur génétique. Les gènes en cause étant inconnus, cette valeur génétique est classiquement estimée à partir des phénotypes mesurés sur les candidats eux-mêmes ou sur leurs apparentés proches. La sélection utilise donc des phénotypes et des généalogies, qui sont combinés dans un modèle mathématique dont l'effet est double : d'une part, il corrige les phénotypes pour les effets de milieu, d'autre part, il pondère de façon optimale l'ensemble des phénotypes corrigés pour estimer les valeurs génétiques.

Plus récemment, avec l'émergence des marqueurs génétiques, les approches de prédiction des valeurs génétiques ont cherché à utiliser ce nouveau type d'information. Dans un premier temps, les approches ont cherché à détecter les régions du génome les plus importantes dans le déterminisme génétique d'un caractère pour les utiliser ensuite en sélection. Pour cela, les gènes en cause ne doivent pas forcément être connus. En effet, les marqueurs qui n'ont généralement pas d'effet biologique propre, peuvent être situés à proximité de gènes ayant un effet sur les caractères. Du fait de leur proximité sur le génome et de la relative rareté des recombinaisons génétiques (en moyenne 1% de recombinaison à la méiose par million de bases

d'ADN chez les mammifères), les marqueurs sont co-transmis et ont donc un effet apparent, celui de la région les entourant. Cet effet apparent est utilisé pour détecter les régions portant les gènes influençant les caractères. Par ailleurs, sélectionner sur les allèles des marqueurs présentant un effet apparent favorable conduit à sélectionner les allèles favorables de ces gènes. Lorsqu'on sélectionne sur un petit nombre de régions, on parle de *sélection assistée par marqueurs*.

Cependant, pour la plupart des caractères, chaque gène (ou région) n'explique qu'une faible fraction de la variabilité génétique, de sorte que la sélection assistée par marqueurs de quelques régions n'a souvent qu'une efficacité limitée.

En 2001, Meuwissen et al. ont proposé une nouvelle approche, utilisant des marqueurs couvrant l'ensemble du génome, ou *sélection génomique*. Le principe est relativement simple : une population de référence de quelques milliers d'individus est phénotypée et génotypée. Cette population permet d'établir les relations statistiques entre génotype et phénotype. Autrement dit, cette méthode estime à partir de ces données les effets apparents de tous les marqueurs génotypés. Ces effets apparents sont ensuite appliqués à des candidats génotypés mais sans performance et leur somme est ainsi une prédiction de leur valeur génétique.

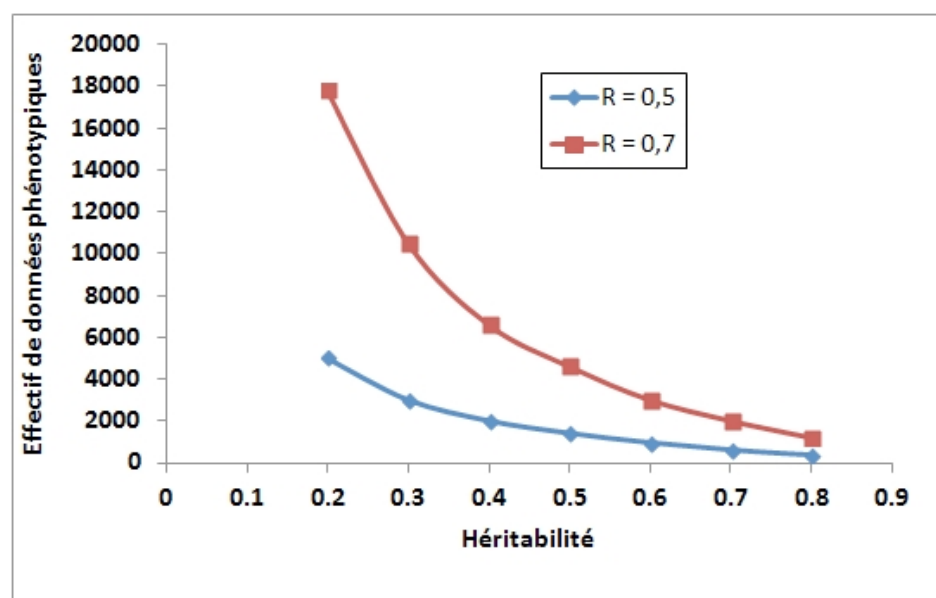
3. Facteurs d'efficacité des prédictions génomiques

Plusieurs facteurs affectent la précision des évaluations génomiques. La précision est généralement mesurée par la corrélation R entre la valeur estimée et la valeur génétique vraie, ou plus fréquemment par le carré de cette corrélation R^2 également appelé coefficient de détermination (CD). La valeur génétique vraie étant inconnue, le CD l'est également mais il peut être estimé indirectement, soit à partir de la quantité d'information disponible, soit *a posteriori* à partir de la qualité des prédictions réalisées.

Deux grands types de facteurs déterminent la précision des prédictions. Le premier facteur est purement statistique, il détermine la précision des estimations des effets des marqueurs. Comme dans toute estimation statistique, la précision est d'autant meilleure que l'effectif de la population de référence est élevé. De façon similaire, la précision est d'autant meilleure que l'héritabilité du caractère est élevée. En effet, plus l'héritabilité est faible, plus l'effet du milieu est important et plus il y a de bruit de fond dans l'estimation des effets génétiques. La *figure 1*, d'après Hayes et al. (2009), fournit un abaque de la précision R obtenue avec un effectif et une héritabilité donnés, dans une population de mammifères d'effectif génétique 100.

Figure 1. Relation entre précision (mesurée ici par la corrélation R) de l'évaluation génomique, effectif de la population de référence, et héritabilité du caractère.

(d'après Hayes et al., 2009).



Mais les effets apparents des marqueurs peuvent être très différents des effets réels des gènes. Pour que ces effets soient similaires, il faut que marqueurs et gènes soient en déséquilibre de liaison, c'est-à-dire qu'il existe une association statistique entre allèles du marqueur et allèles du gène. Cette association existe si marqueur et gène sont à l'intérieur d'un segment statistiquement conservé dans la population. Ces segments sont d'autant plus petits que la population est génétiquement grande. En effet, dans une grande population, les recombinaisons sont plus nombreuses, mais aussi plus efficaces car rarement entre segments identiques. Donc plus l'effectif génétique d'une population est important, plus les segments sont petits, plus les marqueurs doivent être nombreux, et plus la population de référence doit être grande pour estimer précisément leurs effets.

De façon similaire, plus le génome (mesuré en nombre de recombinaisons, ou Morgan) est grand, plus les marqueurs doivent être nombreux, plus la population de référence doit être grande. Ce facteur est cependant assez théorique car chez les mammifères, la longueur du génome est assez comparable d'une espèce à l'autre.

On peut également supposer que la méthode d'estimation des effets des marqueurs influence la précision des estimations de valeur génétique. Comme le nombre d'effets à estimer est très élevé, les méthodes de moindres carrés sont généralement très peu efficaces. Il est essentiel de supposer que les effets des marqueurs sont aléatoires (c'est-à-dire tirés d'une distribution connue) et non fixés. Deux grands types de méthodes sont proposés :

- Des méthodes cherchant à sélectionner ou à privilégier les marqueurs les plus prédictifs. Cette approche est bien adaptée à un déterminisme avec nombre de gènes fini et cherche à se rapprocher de la biologie ;
- Des méthodes cherchant à mesurer l'apparentement entre individus : deux individus ont des prédictions d'autant plus proches que leurs génomes se ressemblent. Cette seconde approche est adaptée au modèle polygénique utilisé en sélection classique, l'apparentement « vrai » mesuré par les marqueurs remplaçant l'espérance d'apparentement déduit du pedigree.

En pratique, les deux types d'approche ont souvent des performances similaires, avec un léger avantage à la méthode la plus proche du déterminisme biologique vrai du caractère évalué. Toutefois, la précision de l'évaluation dépend fortement de l'apparentement entre populations de référence et candidats, ce qui montre que les estimations des effets de marqueurs sont sensibles à la distance génétique ou au nombre de générations. Cela signifie que les estimations reposent sur des relations à grande distance entre marqueurs et gènes, liées à l'apparentement proche et cassées ensuite avec le nombre de générations. Les deux types d'approche, explicitement ou implicitement, utilisent donc largement l'apparentement dans leurs prédictions. En pratique, il est donc essentiel que la population de référence soit représentative de la population des candidats et qu'elle soit renouvelée pour « suivre » l'évolution des candidats : avec les méthodes actuelles, toute augmentation de distance (en termes de nombre de générations, de déterminisme des caractères ou d'évolution du milieu) induit une perte de précision.

4. Propriétés de la sélection génomique

La sélection génomique permet d'estimer le potentiel génétique d'un individu à partir de son génotype à de nombreux marqueurs couvrant tout le génome. Les conséquences de la sélection génomique sont nombreuses. Pour bien mesurer les conséquences, il faut revenir à la formule de prédiction du progrès génétique DG, en supposant une population simple à génération séparée : $\Delta G = \frac{i.R.\sigma_g}{T}$ avec R la corrélation entre valeur génétique prédite et valeur génétique vraie, i l'intensité de sélection (définie par la supériorité des sélectionnés par rapport aux candidats, mesurée en écart type du prédicteur, fonction de la proportion de sélectionnés), T l'intervalle de génération (défini par l'âge moyen des parents à la naissance de leurs descendants), et σ_g l'écart type génétique (égal au produit de l'écart type phénotypique par la racine carrée de l'héritabilité). Si σ_g est une constante du caractère, les paramètres i, R et T peuvent varier en fonction des méthodes de sélection, comme l'illustrent les paragraphes suivants.

La sélection génomique produit des évaluations précoces, la seule contrainte étant la disponibilité d'une quantité suffisante d'ADN de qualité. Une analyse dès la naissance est donc possible, voire même éventuellement sur l'embryon. La sélection devient alors indépendante d'une mesure du phénotype sur les candidats ou ses descendants, le reproducteur peut être utilisé dès qu'il est en âge de reproduire, ce qui peut permettre de réduire l'intervalle de génération T.

La précision R ne dépend plus que de la population de référence. Elle est donc potentiellement élevée si cette population de référence est de taille suffisante, y compris pour des caractères peu héritables. Elle ne dépend plus du temps nécessaire à accumuler l'information phénotypique, comme c'est le cas en sélection classique. L'avantage devient évident lorsque le caractère est non mesurable sur le candidat, coûteux à mesurer, tardif dans la vie, invasif ou incompatible avec le statut de reproducteur, toutes conditions rendant la sélection classique plus difficile et coûteuse. La précision obtenue est équivalente chez les mâles et les femelles, ce qui est un avantage considérable pour les caractères exprimés dans un seul sexe comme par exemple la production laitière.

Enfin, l'intensité de sélection dépend surtout du nombre de candidats que l'on peut génotyper pour un budget donné, et donc du coût du génotypage. Lors de la comparaison économique entre sélection génomique et sélection classique, un paramètre central est le rapport entre le coût du génotype et celui du phénotype.

Pour résumer, les gains permis par la sélection génomique sont de plusieurs types : une précision potentiellement élevée si la population de référence le permet ; un intervalle de génération qui n'est plus limité que par le potentiel de reproduction et non plus par l'obtention d'information ; l'intensité de sélection peut être élevée si le coût du génotypage est suffisamment faible pour permettre un criblage large de la population. La situation est assez variable entre espèces et filières. Elle est généralement beaucoup plus favorable aux grosses espèces dont l'intervalle de génération peut être raccourci et pour lesquelles le coût du génotypage n'est pas trop limitant par rapport à la valeur de l'animal.

La sélection génomique a été très rapidement adoptée en bovins laitiers. Les raisons du succès immédiat de cette innovation sont les suivantes. Tout d'abord, la sélection classique est lourde et coûteuse. La prolificité des femelles est réduite, de sorte que l'intensité de sélection possible des vaches est réduite. L'essentiel de la sélection passe donc par la sélection des mâles diffusables largement en insémination artificielle. Cependant, la majorité des caractères en sélection n'étant pas mesurables chez le mâle, il convient de les évaluer sur descendance, en procréant 50 à 100 filles puis en mesurant leurs phénotypes pour plusieurs dizaines de caractères (production, composition du lait, résistance aux mammites, fertilité, morphologie, aptitude à la mise bas, longévité...). Ce processus est coûteux et dure 5 ans. Une caractérisation dès la naissance est donc une solution particulièrement attractive. Grâce à la diminution sensible de l'intervalle entre générations, dans cette espèce le progrès génétique peut être doublé sans augmenter le coût de la sélection.

Par ailleurs, les conditions de mise en œuvre ont été grandement facilitées par la disponibilité de populations de référence. Ainsi, chez les bovins laitiers, les premières populations de référence ont été constituées des taureaux testés sur descendance dans les années précédentes. Un index sur descendance est assimilable à une performance exprimée par le taureau, d'héritabilité égale à la précision de l'index. Ceci permet de se placer dans la partie droite de la figure 1 et de donc bénéficier d'une bonne précision pour tous les caractères. Ainsi plusieurs milliers de taureaux testés ont été génotypés (dont plus de 20 000 en Holstein dans le cadre du consortium Eurogenomics), permettant de prédire la valeur des jeunes candidats avec des R^2 compris entre 0,5 et 0,7 pour plus de 30 caractères, y compris certains peu héréditaires comme la fertilité. Dans cette espèce, les phénotypes préexistaient de sorte que le coût de la population de référence initiale a été limité au coût du génotypage.

La sélection génomique s'est ainsi développée particulièrement vite chez les bovins laitiers. Elle est officielle en France depuis 2009 (Boichard *et al.* 2012), alors qu'elle reste encore expérimentale dans les autres espèces. La sélection génomique est une alternative au testage qui n'est plus nécessaire et voit sa pratique disparaître. En 2013, 70% des inséminations réalisées dans les trois principales races laitières françaises (Holstein, Normande et Montbéliarde) étaient issues de taureaux évalués sur information génomique. Depuis 2011, le service est ouvert à tout éleveur qui souhaite caractériser ses vaches.

Mais la situation n'est pas aussi simple dans les autres espèces. Un paramètre essentiel est le coût global du processus d'évaluation génétique. En évaluation classique, il comprend le coût du phénotypage des candidats et de leurs apparentés directs (descendants, collatéraux) nécessaire pour obtenir la précision souhaitée. En évaluation génomique, il comprend le phénotypage et le génotypage de la population de référence ainsi que le génotypage (et éventuellement un certain niveau de phénotypage) des candidats. En évaluation classique, le coût de l'évaluation est grossièrement proportionnel au nombre de reproducteurs à sélectionner. En évaluation génomique, le coût comprend une composante variable correspondant au génotypage des candidats mais aussi une composante fixe, celle de la population de référence dont la taille détermine la précision. Pour les populations en sélection de taille réduite, une population de référence de plusieurs milliers d'individus constitue un coût majeur et donc un frein à la mise en place. Au contraire, pour des populations en sélection de très grande taille comme la race bovine Holstein, une population de référence même de plusieurs dizaines de milliers d'individus constitue une solution réaliste et économiquement attractive.

Le facteur limitant de la mise en œuvre est donc le coût de création puis d'entretien de la population de référence. De plus, dans les conditions actuelles de sélection génomique, la précision d'évaluation n'est élevée que lorsque les candidats sont fortement apparentés à la population de référence, ce qui impose de la renouveler constamment au cours des générations. Cette contrainte est bien sûr d'autant plus forte que l'intervalle de génération est court.

Toutefois, même dans les espèces peu favorables à la sélection génomique, il est fréquemment montré que celle-ci est optimale si la sélection porte sur des caractères non mesurables chez le candidat (car incompatibles avec le statut de reproducteur). Ainsi, Tribout et al. (2013) ont montré qu'en sélection porcine qui bénéficie pourtant d'un intervalle de génération court et de performances majoritairement mesurables chez les candidats, la sélection génomique est plus efficace que la sélection classique pour les caractères mesurés sur collatéraux et en conséquence pour l'objectif global.

On peut distinguer trois grandes situations :

- (a) La sélection génomique est au moins aussi efficace et moins chère que la sélection classique, il est probable qu'elle va se développer rapidement, en substitution de la sélection classique. C'est le cas chez le bovin laitier pour lequel la sélection génomique conduit à un progrès génétique potentiel deux fois plus élevé que la sélection classique, principalement grâce à la diminution de l'intervalle de génération.
- (b) elle est plus coûteuse que la sélection classique sans gain d'efficacité clair, les conditions ne sont pas encore remplies pour son développement. C'est par exemple le cas chez le poulet de chair ou le poisson, des espèces où la sélection classique est efficace et peu coûteuse.
- (c) elle induit un gain d'efficacité mais au prix d'un surcoût, son développement est probable mais nécessite des investissements supplémentaires, généralement rentabilisés en quelques années. C'est le cas par exemple du porc (Tribout et al., 2013) ou de la poule pondeuse.

Un point critique est le coût du génotypage. Il est généralement réalisé par des puces permettant le génotypage de nombreux marqueurs simultanément. Ce coût intègre différentes composantes : prélèvement de tissus, extraction de l'ADN, puce de génotypage, travail de laboratoire correspondant, enfin évaluation génétique proprement dite. Pour le réduire, différentes approches sont possibles. Le premier facteur est le temps : le prix du génotypage diminue progressivement du fait du progrès des technologies et de la concurrence entre fabricants et entre laboratoires. Le deuxième facteur est l'augmentation des volumes. Sous les effets combinés du temps et du volume, le coût de la puce bovine a ainsi été divisé par 5 entre 2008 et 2013. De ce point de vue, les approches reposant sur des produits génériques libres de droit permettent une plus grande diffusion et donc un prix plus bas. Les stratégies sont différentes entre espèces, selon que la priorité est mise sur l'exclusivité (en volailles par exemple, chaque sélectionneur a fait développer sa propre puce de génotypage et doit en assumer seul le coût) ou sur le prix (en bovins par exemple, avec un produit générique commercialisé librement). Une troisième approche pour diminuer le coût est l'utilisation de puces de plus basse densité, contenant moins de marqueurs, mais moins chères (Boichard et al., 2012). Cette stratégie demande une organisation et un travail bioinformatique d'imputation, en vue d'estimer statistiquement l'information manquante avec la meilleure précision possible. À titre d'illustration, en 2013, le coût du génotypage variait de quelques dizaines à 150-200 euros par animal, selon l'espèce (et donc le volume d'activité) et le type de puce.

À l'avenir, beaucoup d'espoir est placé sur les technologies de génotypage par séquençage. Dans cette approche, une proportion réduite, quelques pour cent, mais prédéfinie du génome, est séquencée. Par la comparaison des séquences des mêmes régions entre individus, le génotype de chaque individu peut être déterminé pour plusieurs dizaines de milliers de marqueurs simultanément. La réduction du coût par individu est obtenue par multiplexage, c'est-à-dire par le mélange dans la même opération de séquençage, de plusieurs dizaines d'individus, permettant de diviser le coût du séquençage par ce nombre d'individus traités simultanément. L'inconvénient actuel est l'effort bioinformatique nécessaire pour obtenir un niveau équivalent d'information et de précision à celui des puces de génotypage. Cette approche se développera d'abord dans les espèces qui ne disposent pas de puce de génotypage ou pour lesquelles cette puce reste coûteuse.

Dans certaines situations, il est possible de construire des cercles vertueux induisant l'accumulation d'information. Ainsi, chez les bovins, la puce basse densité dite LD (low density), moins chère intrinsèquement, permet une plus forte diffusion et donc une réduction supplémentaire du coût de revient. Sa forte diffusion à l'échelle des élevages en contrôle de performances permettra un renouvellement facilité de la population de référence. La large diffusion de cette puce permet de la remettre à jour régulièrement et donc d'inclure de nouveaux tests issus des travaux de recherche, augmentant l'intérêt de son utilisation.

5. Implémentation de la sélection génomique

L'utilisation de ce nouvel outil entraîne généralement des conséquences importantes dans l'organisation pratique de la sélection. Tout d'abord, il convient de trouver le meilleur compromis entre l'investissement dans la population de référence (qui détermine la précision) et le nombre de candidats génotypés (qui détermine l'intensité de sélection).

Mais dans les populations à générations chevauchantes, le compromis entre précision de l'évaluation et intervalle de générations est profondément modifié. En sélection classique, les jeunes ont une précision d'évaluation faible et il est souvent préférable d'attendre une meilleure précision, au prix d'un âge à l'utilisation plus élevé. En sélection génomique, il convient d'utiliser les reproducteurs le plus tôt possible pour diminuer l'intervalle de génération et profiter pleinement de l'évaluation précoce.

Chez les bovins, il a été ainsi rapidement démontré (Schaeffer, 2006 ; Colleau et al., 2009) que le testage sur descendance n'était plus nécessaire. En effet, il est plus intéressant de réduire l'intervalle de génération en utilisant les taureaux dès qu'ils sont en âge de reproduire, même si l'évaluation génomique est un peu moins précise que l'évaluation sur descendance. Le coût d'un taureau mis sur le marché étant fortement diminué, il est rentable économiquement et intéressant génétiquement d'en utiliser un nombre plus grand qu'auparavant, chaque taureau n'ayant alors qu'une diffusion limitée dans le temps et en nombre de doses. Colleau et al. (2009) ont même montré que cette option était largement préférable, comme illustré en *figure 2*. En effet, le schéma dit « AXMAX » (accélération maximale) reposant sur l'utilisation exclusive de jeunes taureaux engendre le progrès génétique quasi maximal tout en réduisant l'augmentation annuelle du taux de consanguinité. En effet, l'augmentation attendue de la consanguinité du fait du turn-over plus rapide des générations est plus que compensée par

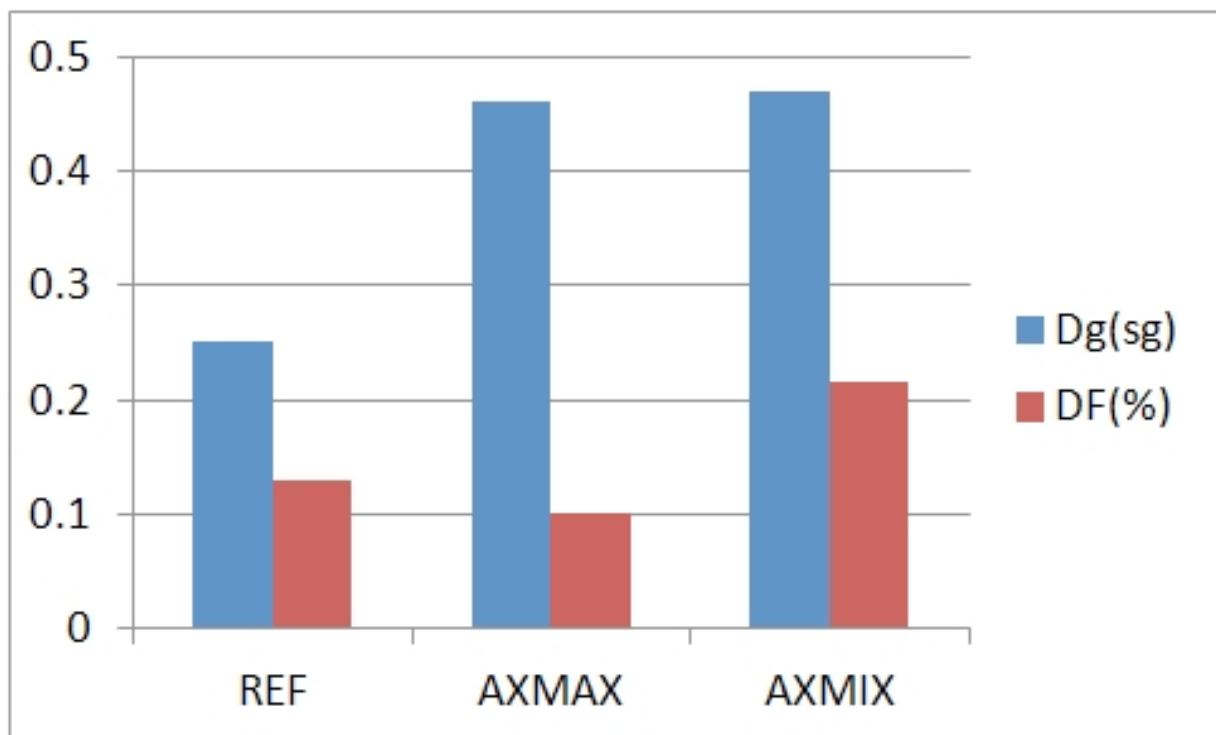


Figure 2. Comparaison de trois schémas de sélection possibles chez les bovins laitiers, d'après Colleau et al. (2009). Δg : progrès génétique annuel attendu, en écart type génétique ; ΔF : évolution annuelle de consanguinité, en %. REF : schéma maintenant un testage sur descendance, n'utilisant l'évaluation génomique que pour la présélection avant testage ; AXMAX : utilisation exclusive de jeunes taureaux évalués sur information génomique ; AXMIX : utilisation pour moitié de jeunes taureaux évalués sur information génomique et des meilleurs taureaux, plus âgés, évalués sur descendance.

l'augmentation du nombre de reproducteurs. Le schéma dit « AXMIX » où 50% du marché de la semence bovine est couvert par des taureaux plus âgés confirmés sur descendance fournit un progrès à peu près équivalent, mais une évolution annuelle de consanguinité deux fois plus élevée que le schéma de référence maintenant le testage sur descendance. Il est donc essentiel pour la durabilité de la sélection de privilégier un schéma de type AXMAX et d'éviter un schéma de type AXMIX, même si ce dernier est sans doute plus attractif et moins en rupture, donc plus acceptable, pour les sélectionneurs comme pour les éleveurs utilisateurs de la sélection.

Cet exemple montre que la sélection génomique peut mieux utiliser la variabilité génétique que la sélection classique, surtout quand le phénotype n'est pas disponible sur le candidat ou que le caractère est peu héritable. D'une façon générale, l'information familiale (en particulier la valeur génétique des parents) joue un rôle très important en sélection classique. La valeur génétique présente la propriété suivante : $G_i = 0,5 G_p + 0,5 G_m + f_i$ avec G_i , G_p et G_m les valeurs génétiques de l'individu i , de son père p et de sa mère m , et f_i l'aléa de méiose ou écart à la valeur des parents. En sélection classique, cet aléa de méiose (qui explique la différence entre pleins-frères) ne peut pas être prédit à partir des parents, mais seulement à partir des performances du candidat ou de ses descendants. La sélection étant séquentielle et coûteuse, elle tend à rechercher les candidats chez les meilleurs parents, ce qui entraîne une concentration vers un nombre limité d'origines. La sélection génomique permet d'estimer l'aléa de méiose et donc de s'affranchir, dans une certaine mesure, de cette sélection sur ascendance. En effet, si l'évaluation génomique n'est pas trop coûteuse, il est possible de cribler l'ensemble de la population et donc d'éroder beaucoup moins vite la variabilité génétique.

6. Sélection pour de nouveaux caractères et évolution des objectifs de sélection

Alors que la sélection classique est liée à la mesure des caractères sur les candidats ou leurs apparentés proches, la sélection génomique sur un caractère est possible dès lors que l'on dispose d'une population de référence adaptée, en taille et en nature, et, bien sûr qu'il existe une variabilité génétique. En déconnectant l'obtention des phénotypes et l'évaluation des candidats, la sélection génomique offre donc une beaucoup plus grande souplesse que la sélection classique, pour les caractères difficiles à mesurer chez les candidats.

Par ailleurs, la sélection génomique engendre généralement plus de progrès génétique, ce qui permet plus facilement d'augmenter le nombre de caractères sélectionnés, sans obérer les gains sur les autres caractères. Cela ouvre des perspectives importantes dans la définition des objectifs de sélection, au moment même où les objectifs doivent prendre en compte des besoins de plus en plus complexes (Boichard, 2013).

Au cours des 40 dernières années, les objectifs de sélection initialement orientés sur la productivité ont progressivement intégré les caractères de qualité des produits, puis d'adaptation des animaux. Aujourd'hui, les demandes se diversifient et deviennent de plus en plus complexes pour satisfaire les trois facettes, économiques, sociétales et environnementales de la durabilité.

Plusieurs stratégies sont envisagées selon les filières et l'organisation de la sélection. Lorsque la sélection est réalisée dans un noyau privé, l'obtention d'un nouveau phénotype correspond généralement à la mise en œuvre d'un protocole de mesure spécifique et entraîne des coûts. L'inclusion d'un caractère nouveau dans l'objectif correspond alors à un choix réfléchi en amont, dont la rentabilité doit être avérée, en particulier en termes d'avantage concurrentiel. Le sélectionneur bénéficie cependant des avancées de l'électronique, des automatismes, de la métrologie pour accéder à de nouveaux phénotypes comme par exemple la production ou le comportement des poules pondeuses dans les systèmes de production sans cages.

Dans toutes les espèces, il est possible de tirer profit des dispositifs croissants de l'élevage de précision. Chez les bovins, les robots de traite constituent des systèmes privilégiés d'observation et de phénotypage des animaux (analyse du lait, débit de traite, poids des animaux...) et peuvent être complétés par des systèmes d'analyse additionnels (imagerie externe de l'état corporel). Des perspectives importantes se dessinent dans l'utilisation de dispositifs sur les animaux (par exemple des accéléromètres pour l'analyse du comportement, la détection des chaleurs ou des boiteries).

Comme montré dans le cadre du projet PhenoFinlait, le recours à la spectrométrie moyen infra rouge (MIR) déjà largement utilisée pour mesurer les taux butyreux et protéique du lait peut être étendu à d'autres composants avec une précision suffisante pour des fins de sélection. Le profil en acides gras du lait ou sa composition en protéines individuelles est ainsi prédit à partir des spectres infra rouge. Avec 25 millions d'analyses MIR réalisées chaque année, un phénotypage à haut débit est possible avec un surcoût marginal. Des études sont en cours pour prédire d'autres composants comme la teneur en minéraux ou l'aptitude fromagère, ou mesurer des signatures d'état physiologique (mobilisation corporelle par exemple) ou de troubles de santé (maladies métaboliques, mammite).

Dans toutes les espèces, des efforts très importants sont réalisés pour mesurer la quantité d'aliment ingéré, l'efficacité alimentaire et/ou l'importance des rejets. Ces caractères sont en effet essentiels d'une part parce que l'aliment représente le poste de charges le plus important de l'élevage et d'autre part parce que des animaux plus efficaces rejettent moins d'effluents. La mesure de l'ingestion passe généralement par des dispositifs de distribution individuelle d'aliment, les animaux étant reconnus par leur identification électronique. La mesure des effluents est souvent complexe. Par exemple, pour l'émission de méthane des bovins, divers dispositifs sont proposés, dosant à la fois le méthane et un gaz de référence (SF₆ émis par un bolus ruminal ou CO₂ respiratoire) dans un échantillon d'air ambiant prélevé autour du museau ou dans un espace confiné comme le robot de traite ou un distributeur de concentré (Greenfeed).

Lorsque la sélection est réalisée dans une population commerciale, des données préexistantes peuvent être valorisées (Boichard et Brochard, 2012). Ainsi, chez les bovins, les données issues de la gestion technique ou de l'activité commerciale peuvent être valorisées. A titre d'exemples qui constituent des enjeux pour le futur, citons les données de troubles des sabots collectées lors du parage, les données du carnet sanitaire dont la tenue est obligatoire dans les élevages, ou les caractéristiques de carcasse collectées lors de l'abattage.

Dans bien des situations, des dispositifs de collecte devront être spécifiquement mis en œuvre. Par exemple, chez le bovin, des travaux sont en cours sur la paratuberculose, la difficulté essentielle étant de disposer de phénotypes fiables permettant de caractériser les individus. Par exemple, il faut éviter de confondre un animal résistant avec un individu non exposé au risque. Il faut également tenir compte des limites des méthodes de diagnostic, tant en termes de sensibilité que de spécificité.

A la fois pour des raisons objectives d'importance des caractères dans l'économie de la production, mais aussi par la volonté de différenciation des sélectionneurs entre eux, il est probable que le nombre de caractères sélectionnés augmente, surtout quand leur mesure peut être réalisée à un coût marginal. Le nombre de caractères sélectionnables augmentant, il sera essentiel de redéfinir périodiquement le poids de chacun dans l'objectif de sélection, pour éviter d'orienter la sélection dans une direction inappropriée.

7. Sélection génomique robuste

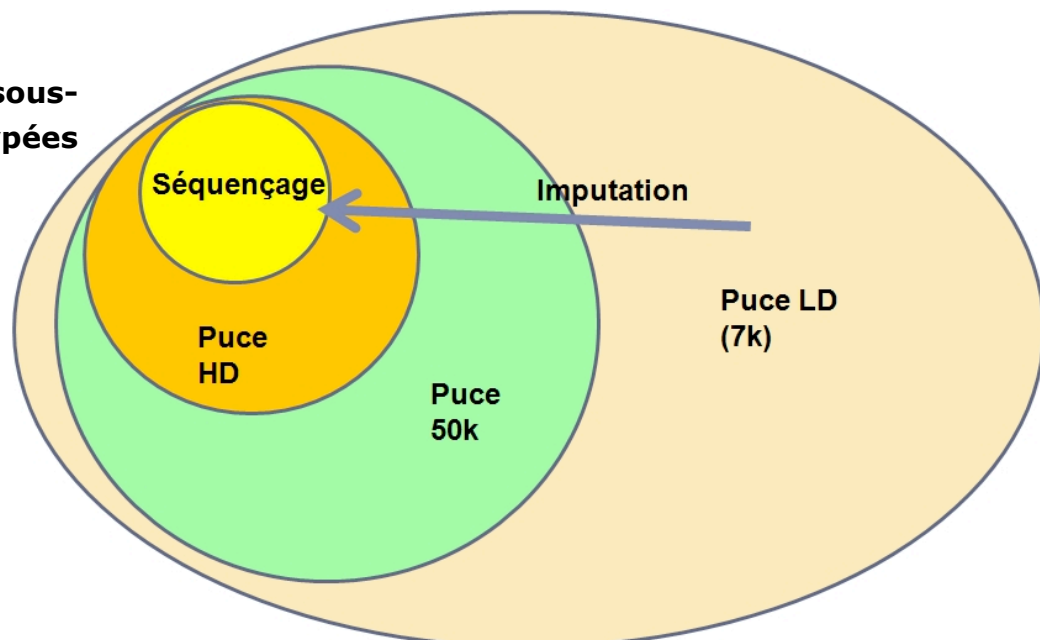
Les méthodes de sélection génomique actuelles ne fonctionnent bien que lorsque la population de référence est fortement apparentée aux candidats, idéalement lorsqu'elle comprend la génération des parents des candidats. En effet, deux types d'information sont utilisés en sélection génomique : d'une part le déséquilibre de liaison à grande distance qui permet de mesurer l'apparentement vrai entre individus ; d'autre part le déséquilibre à courte distance qui permet de suivre les QTL dans la population. Dans les méthodes actuelles, la part de l'apparentement dans la prédiction est forte, rendant la prédiction sensible à cet apparentement.

Ceci induit des limites importantes : il faut reconstruire une population de référence spécifique à chaque population en sélection, avec les coûts que cela induit. La conséquence la plus fâcheuse est que cela rend la sélection génomique peu applicable dans les petites populations ou dans les populations très segmentées. Idéalement, on souhaiterait pouvoir constituer une méta-population de référence unique permettant une prédiction globale.

À l'avenir, les méthodes devront donc être moins sensibles à l'apparement et davantage reposer sur l'utilisation explicite des mutations causales ou de marqueurs très proches en fort déséquilibre de liaison. Ce problème essentiel n'est pas résolu à ce jour, même si des efforts de recherche importants sont consentis à l'échelle nationale et internationale. Les premières approches reposent sur l'hypothèse qu'une fraction des QTL sont conservés entre races et donc que le déséquilibre de liaison à courte distance apporte une certaine capacité de prédiction entre populations. Pour cela, chez les bovins, une puce à haute densité a été produite, apportant une résolution suffisante pour mettre en évidence les segments chromosomiques préexistant à la différenciation entre races et conservés depuis dans les différentes populations. Si les segments chromosomiques sont conservés entre races, leurs effets sont aussi en grande partie conservés, ce qui doit permettre une prédiction commune. Les premières estimations montrent qu'un gain de précision est apporté mais qu'il n'est pas suffisant pour fournir une solution opérationnelle pour les populations les plus petites. Les travaux sont en cours pour raffiner l'approche et améliorer la précision de l'évaluation multi- raciale.

Mais la solution repose sans doute sur l'identification des mutations causales, responsables de la variabilité génétique des caractères. A l'avenir, l'approche précédente sera étendue aux données de séquence de génome complet. La séquence doit être vue comme un génotypage exhaustif, contenant tous les polymorphismes du génome, y compris les polymorphismes responsables de la variabilité des caractères. Même s'ils sont cachés parmi un très grand nombre de variants sans effet sur le caractère, les identifier est théoriquement possible. Le problème statistique est difficile, car il faut identifier quelques milliers de variants causaux parmi des dizaines de millions de polymorphismes de séquence, ce qui demande des dispositifs de très grande taille. Comme le séquençage reste trop cher pour être généralisé, l'approche envisagée combine le séquençage de quelques centaines d'individus et le génotypage du plus grand nombre. L'idée est ensuite d'imputer statistiquement les marqueurs manquants, à partir de l'information des marqueurs génotypés et des associations entre marqueurs observées chez les animaux séquencés. Cette approche est en cours dans le projet « 1000 bull genomes » (Daetwyler et al., 2014). L'imputation étant appliquée à des dizaines de milliers d'individus génotypés et phénotypés (*figure 3*), le résultat est un énorme dispositif d'individus phénotypés et séquencés réellement ou *in silico*, dont il faut extraire l'information sur les mutations causales par analyse d'association. Les premières approches sont très prometteuses.

Figure 3. Emboîtement de sous-populations bovines génotypées à des densités variées.



Une fois les mutations causales identifiées, l'évaluation génétique devient une tâche beaucoup plus simple. Si les effets de ces mutations sont au moins partiellement conservés entre fonds génétiques, l'évaluation devient universelle et robuste au manque d'apparement.

8. Prise en compte du déterminisme génétique complexe

Les méthodes de sélection génomique actuelles les plus courantes ne tirent parti que de l'effet additif des gènes. En effet, compte tenu du nombre de marqueurs pris en compte, il n'est pas possible statistiquement d'estimer toutes les interactions. Notons toutefois que lorsqu'un effet haplotypique est mesuré, il concerne tout un segment chromosomique et donc tient compte des interactions à courte distance.

Des modèles récents incluent également des effets de dominance (Vitezica et al., 2013). Toutefois, les interactions entre gènes ne pourront être efficacement prises en compte que lorsque les mutations causales seront identifiées. Les grands dispositifs de sélection génomique futurs, avec beaucoup d'individus phénotypés sur performances propres et génotypés pour de nombreuses mutations causales, seront favorables à la recherche des interactions entre gènes et l'estimation de leurs effets, le plus souvent négligés jusqu'à présent.

Un autre sujet futur touche les interactions entre génotype et milieu. Ce sujet est important à double titre. D'une part, la sélection génomique peut perdre de l'efficacité si la population de référence est conduite dans un milieu très différent de celui de la production. Ce sujet est bien connu des sélectionneurs de volailles ou de porc, mais était largement négligé chez les ruminants, principalement du fait du testage sur descendance qui limitait beaucoup les risques, l'index étant alors l'effet moyen sur toute une gamme de milieux. Si les futures populations de référence sont conduites dans des milieux spécifiques, les interactions pourraient conduire à de fortes pertes d'efficacité de la sélection. Il est donc essentiel de considérer une gamme large de milieu dans la constitution des populations de référence.

Les porcs et volailles sont des espèces sélectionnées dans des lignées de race pure pour une production en croisement. L'environnement est aussi beaucoup plus maîtrisé en sélection qu'en production. Ceci induit de fortes interactions entre sélection et production et donc une perte d'efficacité de la sélection. Dans le futur, il est envisageable de constituer des populations de référence chez les animaux croisés de production pour sélectionner les lignées pure directement sur la base des caractères mesurés en production, réduisant ainsi les interactions génotype x milieu.

Plus généralement, de larges populations génotypées et phénotypées dans une gamme de milieux variés permettent d'étudier ces interactions et, éventuellement, de les prédire, augmentant les possibilités de personnalisation de l'objectif de sélection et d'adaptation du choix des reproducteurs à un milieu donné.

9. Autres applications du génotypage

Chez les bovins, le génotypage devrait se développer très fortement, et concerner l'ensemble des génisses du troupeau. Associée à l'utilisation de semence sexée, cette pratique permettra de renouveler le troupeau à partir de la moitié supérieure du cheptel choisie sur l'objectif défini par l'éleveur, et d'appliquer à la seconde moitié une politique de reproduction autre, par exemple le croisement industriel ou la production de femelles pour la vente. Grâce à cette

pression de sélection disponible chez les femelles, à la bonne précision des index femelles pour tous les caractères, ainsi que la gamme nettement plus large de reproducteurs mâles disponibles, des objectifs de sélection définis à l'échelle de l'élevage et s'écartant sensiblement de l'objectif général de la population deviennent réalistes. Pour la première fois, il apparaît ainsi une solution opérationnelle à la question récurrente de l'adaptation de la sélection aux souhaits individuels des éleveurs ou aux conditions locales.

Dans les filières avec un fort développement du génotypage (principalement les bovins aujourd'hui, mais sans doute d'autres à l'avenir), de nouvelles applications verront le jour. Si tous les animaux sont génotypés, de nouveaux plans d'accouplements peuvent être construits intégrant des informations génétiques précises sur plusieurs dizaines de caractères. Les tares pourront être éliminées de façon volontariste et rapide et cette information est déjà fournie aux clients actuels de la sélection génomique. L'apparement vrai entre individus sera mesuré, permettant d'orienter les accouplements pour minimiser la consanguinité. La perte de variabilité liée à la sélection pourra être limitée, en surpondérant les allèles rares pour éviter de les perdre par fixation. L'accumulation progressive des allèles favorables des QTL les plus importants pourra être recherchée au cours des générations.

10. Évolution du contexte organisationnel autour de la sélection génomique

La sélection génomique est une innovation de rupture qui peut modifier profondément le schéma organisationnel. Les opérateurs historiques doivent évoluer et intégrer cette technologie, des nouveaux venus sont possibles. L'organisation actuelle chez les ruminants, souvent nationale, est potentiellement remise en cause.

L'enjeu premier est de constituer et maîtriser les populations de référence. Les structures augmentent en taille, intègrent davantage sélection et phénotypage, développent leur capacité de recherche et développement. Les données sont davantage un enjeu stratégique et ne sont échangées que dans un cadre contractuel précis, pour bénéficier de populations plus grandes et réaliser des économies, tout en protégeant son investissement de la concurrence.

Un phénotype nouveau, objectivement important ou simplement ressenti comme tel, est un signe de différenciation vis-à-vis de la concurrence. Les données ont un cycle : d'abord spécifiques, elles ont un statut privé ; puis lorsqu'elles se banalisent, elles sont plus échangées pour mieux partager les coûts. Des efforts importants sont consentis pour inventer des méthodes de mesure, les mettre en place dans les élevages, contractualiser des réseaux de fermes sur une fonction de phénotypage, développer les bases de données et les systèmes d'interprétation correspondants.

Des efforts sont également consentis dans les approches de prédiction statistique de la valeur génétique car il est évident que les méthodes actuelles ne sont qu'une étape. Les méthodes du futur seront plus robustes et plus universelles, elles intégreront les données de séquençage et beaucoup de travaux et de collaborations internationales sont en cours pour produire ces données et les utiliser de la façon la plus efficace.

La sélection génomique n'a pas encore percé dans toutes les populations, loin de là. Pourtant, son émergence semble inexorable. Des collaborations internationales se mettent en place pour assurer les transferts technologiques et également dessiner de nouvelles sphères d'influence.

11. Conclusion

La sélection animale du futur intégrera la sélection génomique. La sélection génomique sera moins coûteuse, donc plus facile à mettre en œuvre dans toutes les populations. Elle reposera davantage sur la connaissance du déterminisme génétique moléculaire des caractères et donc sera plus facilement partagée entre populations. Elle sera étendue à plus de caractères, en particulier de santé et d'adaptation, grâce à la mise en place de populations de référence adaptées. Plus généralement, elle rendra possible la sélection dans des conditions considérées comme difficiles ou très coûteuses en sélection classique. La constitution de populations de référence sera au cœur des stratégies de sélection, avec ses implications sur le phénotypage mais aussi les stratégies d'entreprise.

Références

- Boichard D., Brochard M. 2012. New phenotypes for new breeding goals in dairy cattle. *Animal* 6: 544-550.
- Boichard D., Guillaume F., Baur A., Croiseau P., Rossignol M.N., Boscher M.Y., Druet T., Genestout L., Colleau J.J., Journaux L., Ducrocq V., Fritz S. 2012. Genomic selection in French dairy cattle. *Animal Production Science* 52: 115-120.
- Boichard D., Chung H., Dasonneville R., David X., Eggen A., Fritz S., Gietzen K.J., Hayes B.J., Lawley C.T., Sonstegard T.S., Van Tassell C.P., Vanraden P.M., Viaud K., Wiggans G.R. 2012. Design of a bovine low-density SNP array optimized for imputation. *PLoS ONE* 7: e34130.
- Boichard D., 2013. La sélection génomique : une opportunité pour l'amélioration de la santé des animaux d'élevage. *Bulletin de l'Académie Vétérinaire de France* 166, 25-31.
- Colleau J.J., Fritz S., Guillaume F., Baur A., Dupassieux D., Boscher M.Y., Journaux L., Eggen A., Boichard D. 2009. Simulation des potentialités de la sélection génomique chez les bovins laitiers. 16èmes Rencontres Recherches Ruminants, Paris, 419.
- Daetwyler H.D., Capitan A., Pausch H., Stothard P., Van Binsbergen R., Brøndum R.F., Liao X., Djari A., Rodriguez S., Grohs C., Jung S., Esquerré D., Bouchez O., Gollnick N., Rossignol M.N., Klopp C., Rocha D., Fritz S., Eggen A., Bowman P., Coote D., Chamberlain A., Vantassell C.P., Huggsle I., Goddard M.E., Guldbbrandtsen B., Lund M.S., Veerkamp R., Boichard D., Fries R., Hayes B.J. 2014. The 1000 bull genomes project. *Nature Genetics* (in press).
- Hayes B.J., Bowman P.J., Chamberlain A.J., Goddard M.E. 2009. Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science* 92: 433-443
- Meuwissen T.H.E., Hayes B.J., Goddard M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819-1829.
- Schaeffer L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics* 123: 218-223.
- Tribout T, Larzul C, Phocas F. 2013. Economic aspects of implementing genomic evaluations in a pig sire line breeding scheme. *Genetics Selection Evolution* 45:40.
- Vitezica Z.G., Varona L., Legarra A. 2013. On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics* 195, 1223-1230.