

Journée d'information Agenae
« La génétique quantitative à l'heure du génotypage à haut débit »

Cette journée a été organisée par Vincent Ducrocq et Pascale Hubert à la Résidence Internationale de Paris (44, rue Louis Lumière, 75020 Paris). Elle a regroupé environ 70 personnes dont un peu plus de la moitié issues de lu secteur APA de l'INRA, le reste se répartissant à égalité entre professionnels et INRA hors secteur APA.

Six séquences de 45 minutes (discussion comprise) ont illustré le thème de la journée dans les secteurs animaux (essentiellement en bovins laitiers), végétaux et humains.

1 – Didier Boichard (Chef du Département GA) a présenté la situation de la sélection assistée par marqueurs (SAM) en bovins laitiers, telle qu'elle est appliquée en France. Par rapport à la SAM en place depuis 2001 reposant sur quelques dizaines de microsatellites statistiquement assez éloignés des QTL importants, le recours à la puce Illumina avec 54000 SNP permet dorénavant une cartographie beaucoup plus fine de QTL plus nombreux, ainsi qu'un bien meilleur suivi de ces QTL en SAM. Ainsi, le programme ANR Genanimal Cartofine a conduit durant l'été 2008 et pour chacun des 25 caractères analysés à la détection de plusieurs dizaines de QTL, expliquant plus de la moitié de la variabilité génétique totale. Pour cela, une population de référence de 3200 taureaux des trois races laitières majeures a été génotypée. Les « phénotypes » étudiés étaient très précis car ils résultaient de l'évaluation génétique de chacun de ces taureau sur au moins une centaine de filles. La méthodologie utilisée (LDLA) combine analyse d'association et prise en compte du déséquilibre de liaison entre QTL et marqueurs, alliant les qualités de robustesse de la première à la localisation très précise des QTL permise par la seconde. Avec l'appui des professionnels qui se sont engagés à génotyper au moins 10000 animaux par an à LABOGENA, une SAM de deuxième génération (SAM2) a débuté en Octobre 2008, après validation a posteriori sur les taureaux mis en testage en 2004 pour lesquels on connaît déjà les résultats. Les répercussions sur la sélection sont considérables : la connaissance avec une précision suffisante de la valeur génétique des candidats à la sélection dès la naissance, y compris sur des caractères peu héritable, permet d'envisager à terme l'abandon du testage sur descendance extrêmement coûteux et un fort raccourcissement de l'intervalle de génération.

2- Vincent Ducrocq (SGQA, Jouy-en-Josas) a poursuivi sur le développement de la sélection génomique (SG) toujours en bovins laitiers, probablement en complément (et pas nécessairement à la place) de la SAM2. Comme pour la SAM, une population de taureaux de référence à la fois génotypés et phénotypés est utilisée pour établir une équation de prédiction de la valeur génétique des animaux à partir de tous les marqueurs SNP, sans se soucier de la localisation effective ni de la taille des QTL. Il s'agit donc d'une approche purement statistique permettant comme pour la SAM une évaluation génétique des animaux dès la naissance. Les recombinaisons à chaque génération rendent nécessaire des remises à jour périodiques. La principale difficulté résulte du nombre de prédicteurs très supérieur au nombre de phénotypes à prédire. On assiste actuellement à une explosion des méthodes de prédiction proposées, différant principalement par les hypothèses initiales sur l'effet des QTL et la complexité informatique. Les résultats étrangers montrent des précisions de la prédiction au moins du même ordre que la SAM2, y compris pour les caractères peu héritable, et qui s'accroissent avec la taille de la population de référence, d'où une tendance actuelle à aborder la SG au sein de consortiums internationaux. Le nouveau programme ANR-Genanimal AMASGEN se propose d'étudier tous ces aspects ainsi que la manière de combiner SAM et SG, tout cela à partir de la même population de référence (déjà génotypée) que pour la SAM2. Avec la SAM2/SG, le phénotypage fin de nouveaux caractères va créer de nouvelles

opportunités pour la sélection mais le phénotypage aussi étendu que possible de la population ne doit pas être négligé.

L'application de la SG dans d'autres espèces domestiques, y compris dans des populations en croisement, est tout aussi prometteuse mais est plus difficile car elle ne bénéficie pas de la situation très particulière des bovins laitiers (coût élevé de l'animal et du testage sur descendance, intervalle de génération très élevé, phénotype des taureaux (index) connus très précisément, fort investissement des professionnels).

3- Alain Charcosset (GAP, SGV Le Moulon) a ensuite présenté les perspectives d'application de la SG chez les végétaux. Les programmes de sélection diffèrent fortement du secteur animal, avec pour chaque programme de nombreuses populations de petite taille, qui doivent être évalués en multi-sites sur plusieurs années pour bien prendre en compte les interactions génotype x milieu (G x E). Les succès obtenus n'enlèvent pas l'intérêt pour une accélération du progrès génétique à l'aide de marqueurs moléculaires. Cela passe par exemple par le tri précoce d'individus sur des locus individuels contrôlant des caractères indispensables. Mais compte tenu de la lourdeur de l'évaluation génotypique, l'approche la plus efficace est une SAM combinant phénotypes et marqueurs sélectionnés sur la base d'un test - surtout si on la combine avec une augmentation du nombre de cycles par unité de temps par utilisation de générations « dérobées » (sans phénotypage). Les étapes de détection des QTL et celle d'estimation des effets de ceux-ci sont simultanées. Des résultats convaincants ont ainsi été obtenus chez le maïs, même si ils restent lourds. Avec l'évolution des technologies comme des procédures d'extraction à haut débit, des approches de cartographie LDLA pangénomique sur de grandes populations de base deviennent envisageables. Dans ce contexte, il est possible d'élaborer des scores moléculaires pour la SG impliquant tous les marqueurs utilisés et non plus seulement ceux sélectionnés sur la base d'une détection de QTL. Les simulations estiment alors le gain à 20 à 40%, surtout lorsque l'héritabilité du caractère est faible et le nombre de QTL élevé. La SG est donc promise à un bel avenir, surtout après une première sélection phénotypique pour ne génotyper que les populations les plus intéressantes.

4- Florence Demenais (INSERM U704, CEPH, Paris) a ensuite fait le point sur la situation en génétique humaine. On assiste à une véritable explosion des études d'association pangénomique (GWAS). Le site <http://www.genome.gov/gwastudies/> répertorie 175 études GWAS, la très grande majorité postérieure à mi 2007, concernant une grande variété de maladies, troubles de santé ou caractères quantitatifs. Ces études sont basées sur des effectifs importants (jusqu'à 20.000) et une haute résolution - mais l'intérêt d'aller au-delà de 300.000 SNP est assez limité pour la population européenne - ce qui n'est pas sans poser des problèmes (maintenance des bases de données, contrôle qualité, correction pour la stratification de la population). Malgré ces effectifs, une confirmation indépendante est la plupart du temps nécessaire, en prenant garde aux risques d'interaction G x E ou G x fond génétique. Ces études ont révélé de nouveaux gènes mais aussi l'influence de régions chromosomiques importantes bien que vides de gènes, et surtout de nombreux cas de pléiotropie (marqueurs identiques pour plusieurs cancers ou diverses pathologies des muqueuses, par exemple). Florence Demenais a terminé son exposé par diverses illustrations des défis restant à relever. Ceux-ci concernent l'identification des mutations causales bien sûr mais aussi une meilleure utilisation des phénotypes (par exemple, par classes d'âge), la réalisation d'analyses multi-marqueurs, l'inclusion de nouvelles sources de variation (CVN, variants rares) et de l'information apportée par les autres « omiques ». Mais la complexité de l'appréciation des risques rend illusoire pour encore longtemps une médecine personnalisée en fonction des caractéristiques génétiques personnelles.

5- Claude Chevalet (LGC, INRA Toulouse) a abordé le problème de la détection d'interactions épistatiques entre QTL à partir des données de génotypage à haut débit. La SG se développe en effet actuellement sur la base de modèles additifs assez simplistes. L'amélioration de la qualité des prédictions peut passer par la prise en compte des interactions GxE et l'épistasie. L'analyse de la bibliographie récente met en avant la résurgence de modèles anciens (comme le modèle de Kempthorne) s'intéressant à une ou plusieurs paires de locus de marqueurs, voire toutes les paires de locus. Certaines modélisations simplifient légèrement le problème en estimant l'interaction entre un QTL et le fond génétique représentant l'effet global de toutes les autres régions chromosomiques. La difficulté centrale, comme pour tout ce qui concerne la SG est le nombre d'effets à estimer qui ici dépasse encore plus considérablement le nombre d'observations. Diverses approches ont été proposées (avec présélection, « shrinkage », a priori Bayésien,...). Le défi statistique mais aussi informatique est énorme. Aussi peut-on se poser la question de l'importance de l'épistasie. Elle apparaît difficilement contournable pour la compréhension de la base génétique des fonctions. Mais elle n'est peut-être pas aussi essentielle pour la SG. C'est l'argument de Hill, Goddard et Visscher (2008) qui montrent que la variabilité génétique est essentiellement additive dans la majorité des cas.

6- Enfin, Mathieu Gautier (LGBC, INRA Jouy-en-Josas) a terminé la journée en illustrant l'intérêt des recherches de signatures de sélection par des exemples d'application dans l'espèce bovine. L'idée de base est que la sélection modifie la fréquence des allèles ayant un effet dans la population, mais aussi celle des allèles qui leur sont voisins, avec une baisse de variabilité « locale ». L'analyse repose généralement sur le calcul d'estimateurs de la diversité génétique intra et inter-population à chaque locus, comme par exemple le coefficient F_{ST} de Wright, ou sur des haplotypes (EHH). Par exemple, les loci pour lesquels F_{ST} s'éloigne significativement de la distribution attendue sous l'hypothèse de locus neutre trahissent une sélection passée. A partir du programme Cartofine déjà évoqué, il a ainsi été possible de démontrer que les F_{ST} extrêmes ne sont pas répartis aléatoirement sur le génome, mettant en évidence des signatures de sélection positive sur des gènes organisés en 9 réseaux fonctionnels interconnectés intervenant dans voies métaboliques liées à la production laitière et à la reproduction. La présence de gènes connus dans les régions sélectionnées valide l'approche utilisée, qui peut en fait aussi servir à la cartographie fine de QTL, même sans phénotypes. Les travaux théoriques portent actuellement sur le recours à des approches Bayésiennes pour améliorer les estimateurs de diversité génétique et les critères de décision (statistique de test). Une application récente à la recherche de signature de sélection dans les populations bovines d'Afrique de l'Ouest (taurins trypanotolérants, zébus et hybrides) a donné des résultats particulièrement prometteurs, révélant par exemple plusieurs gènes sous sélection positive impliqués dans la réponse aux maladies hématologiques. Les signatures de sélection permettent ainsi une identification directe de zones d'intérêt à coût modéré (peu d'animaux génotypés, pas de phénotypes) et gagnerait à l'avenir à être combiné avec des analyses d'association.

Cette journée a illustré concrètement à quel point la rupture technologique que constitue l'apparition de puces denses en SNP à un coût abordable est en train de révolutionner la génétique quantitative ainsi que ses prolongement en amélioration génétique, étude des fonctions, génétique des populations et génétique historique. Quel que soit l'espèce ou le règne considéré, la communauté scientifique va se trouver en face d'une masse considérable de données nouvelles et précises. Le management et l'analyse de ces données constituent un défi tout à fait passionnant et prometteur.