

La sélection porcine assistée par marqueurs

L. OLLIVIER

*Institut National de la Recherche Agronomique
Station de Génétique Quantitative et Appliquée - 78352 Jouy-en-Josas Cedex*

La sélection porcine assistée par marqueurs

Cet article passe en revue les principes de base de la sélection assistée par marqueurs (SAM). Ce type de sélection dépend, pour sa mise en œuvre et son efficacité, des liaisons génétiques (linkages) qui existent entre les locus marqueurs du génome et les locus qui gouvernent le caractère sélectionné, ainsi que de l'éventualité d'associations gamétiques (ou déséquilibres de liaison) entre ces locus. La SAM est en général beaucoup plus efficace quand elle peut exploiter des associations gamétiques. En exploitant au mieux ces associations, le gain de précision atteint par exemple 90 % en sélection combinée pour un caractère faiblement héritable ($h^2 = 0,10$), contre un gain de seulement 32 % en situation d'équilibre, pour une même structure familiale de 8 mères/père et 8 descendants/mère. Une évaluation précise des effets de gènes identifiés sur les caractères sélectionnés, au sein de la population soumise à la SAM, est requise au préalable. Les techniques utilisables sont très coûteuses à cause de la taille des échantillons nécessaires. Le génotypage sélectif, surtout lorsqu'il peut être appliqué à des mélanges d'ADN d'individus sélectionnés, permet de réduire considérablement ces coûts, mais ne peut s'appliquer qu'à des marqueurs en association gamétique avec les locus responsables du caractère sélectionné.

Marker-assisted selection in pigs

This paper reviews the basic principles of marker-assisted selection (MAS). Such a type of selection is operationally very dependent on the physical linkages between markers and quantitative trait loci (QTL) and on potential linkage disequilibrium (D) between those loci. MAS is generally much more efficient when it can exploit D situations. The maximum gain in selection accuracy reaches 90 % in marker-assisted combined selection for a lowly heritable trait ($h^2 = 0.10$) in a D situation, as against only 32 % in a linkage equilibrium situation, assuming the same family structure of 8 dams/sire and 8 progeny/dam. A precise evaluation of gene effects within the population intended for SAM is required. Such evaluations are costly because of the large sample sizes needed. Selective genotyping, especially when it can be associated with DNA pooling, can considerably reduce the costs of gene effects evaluation. This however only applies to markers being in linkage disequilibrium with QTL.

INTRODUCTION

Les cartes génétiques produites ces dernières années chez les animaux d'élevage ont donné un regain de vigueur aux recherches lancées par NEIMAN-SORENSEN et ROBERTSON (1961) sur le rôle que peut jouer l'utilisation en sélection de gènes ou de marqueurs identifiés, comparativement aux méthodes classiques de la génétique quantitative qui n'utilisent que des mesures effectuées sur les caractères concernés. En intégrant l'information fournie par des marqueurs dans la sélection artificielle pour des caractères polygéniques, ce qu'on appelle une Sélection Assistée par Marqueurs (SAM), l'efficacité de la sélection peut être augmentée d'une manière substantielle (voir par exemple LANDE et THOMPSON, 1990). Pour le porc, rappelons que les progrès de la cartographie du génome et les perspectives d'application ont déjà été présentés à ces Journées (OLLIVIER et al., 1995).

Les principes de bases de la SAM seront passés en revue dans cet article. La SAM dépend très étroitement, pour sa mise en œuvre, des liaisons physiques (linkages) qui existent entre les locus marqueurs et les locus qui gouvernent les caractères quantitatifs (QTL pour l'anglais Quantitative Trait Locus), ainsi que de l'éventualité d'associations gamétiques (ou déséquilibres de liaison) entre ces locus. La présentation qui sera faite traitera donc d'abord des situations de déséquilibre de liaison (D), qui se trouvent être aussi celles qui ont été les premières étudiées par les généticiens (NEIMAN-SORENSEN et ROBERTSON, 1961 ; SMITH, 1967 ; LANDE et THOMPSON, 1990). La deuxième partie sera consacrée à la SAM en situation d'équilibre de liaison (E), plus récemment étudiée (CHEVALET et al., 1984 ; FERNANDO et GROSSMAN, 1989 ; OLLIVIER, 1998). L'information nécessaire pour mettre en œuvre efficacement une SAM sera enfin brièvement présentée dans une troisième partie.

1. La sélection assistée par marqueurs en situation de déséquilibre de liaison (D)

Rappelons d'abord qu'un déséquilibre de liaison entre deux locus traduit le fait que les allèles présents à l'un des locus ne sont pas associés au hasard aux allèles présents à l'autre locus. Cette association s'évalue au niveau des gamètes (d'où le terme d'association gamétique) et elle est relative à une population. C'est donc une notion statistique différente de celle de liaison génétique (ou linkage), qui a trait à la position relative de deux locus sur le génome. Deux locus très proches sur le même chromosome peuvent être en équilibre de liaison, et, inversement, deux locus situés sur des chromosomes différents, donc non liés physiquement, peuvent être en déséquilibre. Le déséquilibre de liaison se mesure par un coefficient D qui est la corrélation entre les deux locus. Comme toute corrélation, D peut varier de zéro, qui correspond à l'équilibre, à 1 (ou -1), qui traduit un déséquilibre complet.

Une situation de déséquilibre entre un locus marqueur (que l'on identifie) et un locus de caractère quantitatif (QTL putatif) crée une dépendance statistique entre le caractère et le locus marqueur. Dans ce cas, la connaissance du génotype

d'un individu au locus marqueur fournit une information sur la valeur génétique de cet individu. D'un point de vue génétique, l'information sur le marqueur est équivalente à la mesure d'un caractère additionnel en corrélation avec le caractère que l'on cherche à améliorer. Cette information peut donc être utilisée comme un critère indirect de sélection. Considérons la situation simple d'un caractère quantitatif mesuré sur les candidats à la sélection, et soit P la valeur de cette mesure pour un individu donné. L'information apportée par les marqueurs prendra la forme d'une note marqueur globale (LANDE et THOMPSON, 1990), définie comme la somme des effets sur le caractère considéré des différents gènes marqueurs identifiés sur le candidat. Soit M^D cette note marqueur, exprimée dans la même unité que P. Par exemple, dans le cas de deux locus bialléliques ayant des effets additifs de 0,4 et 0,1 pour les allèles favorables respectifs, et -0,4 et -0,1 pour les allèles défavorables, M^D pourra prendre 9 valeurs différentes selon les génotypes aux deux locus marqueurs, soient 1, 0,8, 0,6, 0,2, 0, -0,2, -0,6, -0,8 et -1. L'application de la théorie classique des indices de sélection montre qu'un critère optimal de sélection individuelle combinant P et M^D peut être calculé pour maximiser la réponse à cette sélection, qui est une Sélection Individuelle Assistée par Marqueurs (SIAM), tel que :

$$I_{SIAM} = h^2 (1-p)(1-ph^2)^{-1}P + (1-h^2)(1-ph^2)^{-1}M^D \quad (1)$$

pour un caractère d'héritabilité h^2 et une proportion p de la variance génétique additive associée aux locus marqueurs (NEIMAN-SORENSEN et ROBERTSON, 1961).

Comme le montre l'équation (1), le poids relatif accordé aux marqueurs dans l'indice de sélection I_{SIAM} augmente avec p. Le gain de précision de la SIAM par rapport à la sélection individuelle stricte suit un comportement analogue, comme le montre le tableau 1. D'autre part la SIAM voit sa supériorité décroître au fur et à mesure que h^2 augmente.

Quand une information familiale est disponible pour la sélection, les marqueurs identifiés chez les apparentés apportent aussi une information supplémentaire et améliorent la précision de l'évaluation génétique (LANDE et THOMPSON, 1990). Chez le porc, l'information familiale comprendra habituellement les moyennes de familles de germains (g) et de demi-germains (d), soient les phénotypes moyens (P_{gm} et P_{dm}) et les notes marqueurs moyennes (M^D_{gm} et M^D_{dm}). L'indice global de Sélection Combinée Assistée par Marqueurs (SCAM) est alors constitué de 6 éléments, tel que :

$$P_{SCAM} = b_1P + b_2M^D + b_3P_{gm} + b_4M^D_{gm} + b_5P_{dm} + b_6M^D_{dm} \quad (2)$$

où les coefficients $b_1 \dots b_6$ sont fonction de h^2 , p et de la structure familiale définie par le nombre de mères par père et la taille des familles de germains. Le tableau 2 montre que les gains de précision évoluent selon les mêmes tendances qu'au tableau 1, mais que les gains sont considérablement réduits, pour la raison évidente que toute information nouvelle a d'autant moins de valeur que l'on dispose déjà d'une information plus importante.

Tableau 1 - Sélection assistée par marqueurs pour un caractère en situation de déséquilibre de liaison, pour des valeurs variables de l'héritabilité (h^2) et de la part (p) de variance génétique additive associée aux marqueurs

| $h^2 \backslash P$ | 0,10 | 0,50 | 1,00 |
|--------------------|----------------------|----------------------|----------------------------------|
| 0,10 | 10 1,35 1,68 | 18 2,29 3,20 | ∞ 3,16 4,47 |
| 0,50 | 1,10 1,03 1,12 | 2,00 1,15 1,53 | ∞ 1,41 2,00 |
| 1,00 | 0 1,00 1,05 | 0 1,00 1,22 | 0 (ou ∞) 1,00 1,41 |

1ère ligne : pondération de la note marqueur relative à celle de la valeur individuelle du caractère dans l'indice de sélection individuelle assistée par marqueurs.

2ème ligne : précision de la sélection individuelle assistée par marqueurs relative à une sélection individuelle stricte : valeur maximale $1/h$ pour $p = 1$.

3ème ligne : précision d'une sélection sur pedigree assistée par marqueurs (phénotype de la mère + note marqueur du père et de la mère) relativement à une sélection strictement sur pedigree (phénotype de la mère) : valeur maximale $\sqrt{2}/h$ pour $p = 1$.

Tableau 2 - Précision de la sélection combinée assistée par marqueurs en situation de déséquilibre de liaison, exprimée relativement à une sélection combinée incluant les mesures sur l'individu et sur des membres de sa famille (h^2 et p définis au tableau 1).

Structure familiale constituée de 8 mères non apparentées par père et 8 descendants par mère

| $h^2 \backslash P$ | 0,10 | 0,50 | 1,00 | m grand ⁽¹⁾ |
|--------------------|------|------|------|------------------------|
| 0,10 | 1,10 | 1,47 | 1,90 | 1,71 |
| 0,50 | 1,02 | 1,10 | 1,27 | 1,25 |
| 1,00 | 1,00 | 1,00 | 1,00 | 1,00 |

(1) Avec $p = 1$

Les considérations précédentes s'appliquent à des caractères qui peuvent être mesurés sur les individus avant qu'ils n'atteignent leur maturité sexuelle. La sélection sur les caractères de reproduction est habituellement basée sur les performances des ascendants, comme par exemple les tailles de portée des mères. Dans ce cas, le caractère n'est mesuré que sur les parents d'un seul sexe, alors que l'information sur les marqueurs peut être disponible dans les deux sexes. La Sélection sur Pedigree Assistée par Marqueurs (SPAM) pour la taille de portée par exemple peut alors profiter à la fois d'une évaluation de la valeur génétique maternelle améliorée par la prise en compte des marqueurs et d'une évaluation de la valeur génétique paternelle qui serait impossible sans marqueurs (en supposant l'absence d'information sur les collatéraux paternels). Les gains de précision relatifs de la

SPAM pour différentes valeurs de h^2 et p sont donnés au tableau 1. On voit que ces gains sont très supérieurs à ceux obtenus en sélection individuelle, ce qui s'explique par la faible valeur informative du pedigree.

En plus de la précision de l'évaluation génétique, un schéma de sélection efficace requiert aussi un bon équilibre entre l'intensité de sélection (i) et l'intervalle de génération (t). En fait, la réponse génétique annuelle attendue dans un schéma de sélection est directement proportionnelle à l'intensité de sélection annuelle, qui est le rapport (i/t) de ces deux paramètres. L'intensité de sélection peut être accrue grâce à la SAM pour des caractères qui ne s'expriment que dans un seul sexe, ou pour des caractères qui ne peuvent être mesurés que sur la carcasse, puisque dans les deux cas le nombre des candidats disponibles pour la sélection peut être accru. Une sélection plus précoce et par conséquent des intervalles de génération plus courts peuvent aussi être obtenus par l'utilisation des marqueurs. Cependant, comme des intervalles de génération courts sont habituellement déjà obtenus dans les schémas porcins, une sélection précoce sur des marqueurs risque de ne guère améliorer la situation de ce point de vue. Une augmentation supplémentaire du rapport i/t peut seulement s'envisager avec la mise en œuvre de nouvelles techniques de reproduction, telles que la production d'embryons in vitro à partir de collectes d'ovules, qui pourrait autoriser une fourniture illimitée de descendants par mère. L'intervalle de génération pourrait alors atteindre une valeur très inférieure à l'âge normal de un an au premier descendant, et une forte pression de sélection pourrait être exercée sur les embryons. Une sélection en deux étapes, d'abord sur les embryons génotypés pour les marqueurs avec une intensité i_e et ensuite sur les adultes avec une intensité i_a sur l'information phénotypique classique, conduirait à multiplier la réponse attendue par le rapport i_e/i_a , par rapport à la SIAM considérée au tableau 1 pour $p = 1$ (LANDE et THOMPSON, 1990).

2. La sélection assistée par marqueurs en situation d'équilibre de liaison (E)

En situation d'équilibre de liaison, aucune liaison statistique n'existe entre les marqueurs et les caractères au niveau de la population considérée dans son ensemble. L'information sur les génotypes individuels aux locus marqueurs n'est donc d'aucune utilité en sélection, toutes les notes marqueurs ayant la valeur zéro. La sélection individuelle assistée par marqueurs ne peut donc pas s'appliquer dans ce cas. Ce qui va pouvoir être utilisé sera la transmission des gènes marqueurs au sein des familles. En l'absence d'information sur des gènes identifiés, les covariances entre apparentés nécessaires au calcul des indices de sélection sont basées sur des coefficients de parenté standards (par exemple 0,25 pour les germains ou 0,125 pour les demi-germains), qui sont relatifs au génome dans son ensemble. Si des gènes marqueurs proche d'un locus de caractère quantitatif peuvent être suivis dans leur transmission des parents aux descendants, des coefficients de parenté plus précis pour ce locus (ou pour un segment du génome incluant ce locus) peuvent être obtenus et utilisés dans la prédiction des valeurs génétiques. On fait appel alors au concept de parenté conditionnelle pour un

locus marqué (ou probabilité d'identité conditionnelle aux marqueurs observés chez les deux individus, définie par CHEVALET et al., 1984), d'où des covariances conditionnelles entre apparentés sont obtenues. Cette information supplémentaire fournie par la transmission des marqueurs peut alors être prise en compte dans les évaluations de valeurs génétiques, par exemple avec une méthodologie BLUP utilisant les phénotypes du caractère et les génotypes des marqueurs (FERNANDO et GROSSMAN, 1989).

Pour simplifier, supposons que les allèles présents à un QTL chez les deux parents d'une famille de germains soient identifiés par des gènes marqueurs, et que le locus marqueur soit situé si près du QTL qu'aucune recombinaison ne se produise entre les deux locus. Si on désigne alors par (ab) et (cd) les quatre gènes marqueurs présents chez les parents, quatre génotypes de descendants peuvent être obtenus : (ac), (ad), (bc) et (bd). Les parentés entre ces quatre classes de descendants, conditionnées par l'observation de leur génotype au locus marqueur, sont obtenues facilement sachant le nombre de gènes identiques entre deux génotypes marqueurs, et elles peuvent se disposer comme indiqué au tableau 3. On voit que les parentés entre les membres de cette famille de germains, au lieu d'être uniformément 0,25 (parenté standard entre germains), peuvent prendre trois valeurs, 0, 0,25 ou 0,5 selon les marqueurs qu'ils portent.

Tableau 3 - Parentés entre les membres d'une famille de germains en fonction des génotypes observés à un locus marqueur.

| Génotype observé | ac | ad | bc | bd |
|------------------|------|------|------|------|
| ac | 0,5 | 0,25 | 0,25 | 0 |
| ad | 0,25 | 0,5 | 0 | 0,25 |
| bc | 0,25 | 0 | 0,5 | 0,25 |
| bd | 0 | 0,25 | 0,25 | 0,5 |

Les quatre gènes présents au locus marqueur chez les parents sont respectivement appelés (ab) et (cd).

La valeur génétique de chaque individu au locus quantitatif concerné peut alors être prédite, connaissant ces parentés et les valeurs moyennes du caractère pour chacun des génotypes observés, en fonction de la proportion (p) de variance génétique additive exprimée à ce locus et du nombre d'individus par génotype. Cette valeur génétique prédite constitue la note marqueur (M^E) utilisable en sélection dans une telle situation. Cette sélection, applicable seulement intra-famille de germains, peut venir compléter une sélection combinée classique.

Le raisonnement ci-dessus se généralise facilement à un nombre quelconque de mères (m) accouplées au même père, et le nombre de génotypes marqueurs est alors égal à 4m dans le cas général. Une note marqueur correspondant à cette situation combinera alors les valeurs moyennes du caractère pour les 4m génotypes observés. Si on désigne par la valeur moyenne du génotype marqueur i, la note marqueur M^E s'écrit :

$$M^E = \sum_i b_i M_i \quad (i \text{ allant de } 1 \text{ à } 4m) \quad (3)$$

On voit que la sélection sur M^E est équivalente dans son principe à une sélection familiale, les familles étant ici remplacées par des groupes définis par leur génotype à un locus marqueur. La note marqueur en situation d'équilibre de liaison (M^E) est visiblement différente de la note (M^D) définie précédemment pour une situation de déséquilibre. M^E est fonction des mesures réalisées sur les candidats testés et prend en compte la taille de famille dans cet ensemble, alors que M^D est un critère strictement individuel indépendant à la fois de la structure familiale et des mesures réalisées sur la famille à laquelle il appartient. Dans les deux cas cependant, une "information extérieure" est nécessaire, soit les effets des gènes pour M^D soit la valeur de p pour M^E (comme on le verra plus loin).

La note marqueur M^E définie en (3) constitue un critère de sélection applicable intra-famille de demi-germains et qui peut venir compléter une sélection combinée classique, selon une procédure dont les détails sont donnés par OLLIVIER (1998). La note M^E est alors intégrée dans un indice de sélection combinée assistée par marqueurs I_{SCAM}^E , avec les mêmes informations phénotypiques individuelles et familiales que dans l'équation (2), tel que :

$$I_{SCAM}^E = b_1 P + b_2 P_{gm} + b_3 P_{dm} + b_4 M^E \quad (4)$$

La précision relative de ce type de sélection est indiquée au tableau 4 pour différentes valeurs de p et h^2 .

Tableau 4 - Précision de la sélection combinée assistée par marqueurs en situation d'équilibre de liaison, exprimée relativement à la sélection combinée (h^2 et p définis au tableau 1, structure familiale définie au tableau 2)

| $h^2 \backslash P$ | 0,10 | 0,50 | 1,00 | m grand ⁽¹⁾ |
|--------------------|------|------|------|------------------------|
| 0,10 | 1,01 | 1,10 | 1,32 | 1,44 |
| 0,50 | 1,00 | 1,05 | 1,23 | 1,27 |
| 1,00 | 1,00 | 1,00 | 1,00 | 1,00 |

(1) Avec $p = 1$

En comparant les tableaux 2 et 4 on peut tirer des conclusions sur les mérites respectifs de la SCAM dans les deux situations. A l'évidence, la SCAM est moins efficace en situation E qu'en situation D. Cette faible supériorité de la SCAM comparée à une sélection BLUP classique en situation E a été confirmée par des simulations (RUANE et COLLEAU, 1995). L'effet de l'héritabilité sur les efficacités relatives est plus important en situation D qu'en situation E. La définition de M^E donnée en (3) montre que sa précision doit augmenter avec la taille de famille puisque les M_i deviennent plus informatifs, alors que M^D a une précision constante. Par voie de conséquence, la SCAM incluant M^D présente un avantage décroissant quand la taille de famille augmente (voir les deux dernières colonnes du tableau 2), alors que la SCAM incluant M^E accroît sa supériorité (voir les deux dernières

colonnes du tableau 4). Il faut cependant noter que les calculs du tableau 4 supposent que toutes les familles sont complètement informatives, ce qui requiert un ensemble de marqueurs très hautement polymorphes. Cette hypothèse aura d'autant moins de chances de pouvoir être vérifiée que le nombre de mères par père augmente. Les précisions relatives de la dernière colonne du tableau 4 sont donc à considérer comme des valeurs théoriques maximales, qu'on ne peut pas s'attendre à obtenir en pratique. Les variations d'efficacité relative de la SCAM en situation D et E sont résumées au tableau 5.

La comparaison des situations D et E doit également couvrir les possibilités offertes pour augmenter l'intensité de sélection annuelle (i/t). De ce point de vue, en situation d'équilibre de liaison, des possibilités d'augmenter l'intensité de sélection sont exploitables pour des caractères exprimés dans un seul sexe ou nécessitant l'abattage, comme l'ont montré MEUWISSEN et GODDARD (1996). Mais on notera qu'une sélection précoce des embryons sur les marqueurs qu'ils portent, évoquée plus haut pour augmenter le rapport i/t , n'est pas possible en situation d'équilibre de liaison, puisque cette procédure suppose qu'une SIAM soit réalisable, ce qui, on l'a vu, n'est pas le cas dans cette situation.

3. L'information nécessaire pour la mise en œuvre de la sélection assistée par marqueurs : la cartographie de QTL

Comme on l'a vu, la SAM requiert une information extérieure pour sa mise en œuvre et pour évaluer son efficacité. Cette information comprend, outre l'héritabilité du caractère, les effets des gènes, qui sont nécessaires au calcul de la note M^D , et la part (p) de variance génétique additive associée aux QTL (ou aux locus marqueurs qui leur sont associés) pour calculer la note M^E et pour, dans les deux situations, évaluer l'efficacité de la SAM (tableaux 1, 2, 4). En fait, quand l'effet d'un gène est connu la valeur de p peut être calculée simplement si l'on connaît la fréquence du gène, en appliquant la définition de la valeur génétique additive à un locus (FALCONER et MACKAY, 1996). En situation E il faut de plus connaître la distance qui sépare le locus marqueur du locus quantitatif, les précisions du tableau 4 étant basées, comme on l'a vu, sur l'hypothèse d'une distance nulle entre ces locus.

La manière la plus simple et la plus directe d'estimer un effet de gène consiste à calculer la régression du caractère étudié en fonction du génotype, puisque ceci correspond à la défini-

tion classique de l'effet génétique additif (FALCONER et MACKAY, 1996). Lorsqu'un QTL est lié à un locus marqueur, la régression en fonction du génotype marqueur estime aussi l'effet du QTL pourvu que les deux locus soient en déséquilibre de liaison complet et situés assez près l'un de l'autre pour qu'aucune recombinaison ne se produise. Si le déséquilibre n'est pas complet et les deux locus distincts, la régression estime un effet "apparent" du QTL, inférieur à l'effet réel. Notons au passage que la valeur maximale de p ne pourra alors atteindre 1. Quand plusieurs locus marqueurs sont connus, une régression multiple peut être calculée pour chaque groupe de liaison (voir LANDE et THOMPSON, 1990). Le statut actuel des cartes génétiques du porc permet ainsi d'envisager un balayage complet du génome. Des échantillons de grande taille sont nécessaires. LANDE et THOMPSON (1990) ont montré qu'un effet ne peut être détecté à un locus au seuil de 1 % sur un échantillon de N individus que s'il explique au moins une part $p = 20/Nh^2$ de la variance génétique additive, ou $p = 10 / Nh^2$ au seuil de 5 %. Ainsi un échantillon de 1000 individus ne pourra détecter au seuil de 1 % que des locus déterminant au moins 20 % de la variance génétique d'un caractère d'héritabilité 0,10. Des valeurs élevées de p , cumulant des effets à plusieurs locus, seront donc difficiles à obtenir pour des valeurs faibles de h^2 , là où précisément la SAM est théoriquement la plus efficace.

Une autre manière d'estimer l'effet d'un gène sur un caractère consiste à observer s'il change de fréquence après sélection sur ce caractère, puisque l'effet d'un gène peut être estimé connaissant ce changement de fréquence (Δq) et l'intensité (i) de la sélection pratiquée sur le caractère (OLLIVIER et al., 1997). Cette estimation ne demande que le typage des individus extrêmes sélectionnés, c'est-à-dire un *génotypage sélectif*, contrairement à l'estimation par régression qui requiert le typage de l'échantillon complet. L'efficacité de l'estimateur par sélection, définie comme l'inverse de sa variance d'échantillonnage, est de $0,5Ni^2q(1-q)$, pour N individus typés (c'est-à-dire $N/2$ individus extrêmes dans chaque direction), contre $2Nq(1-q)$ pour l'estimateur classique de régression (SMITH, 1967). Le premier est donc $0,25i^2$ fois plus efficace que ce dernier. On peut montrer aussi que la procédure optimale de sélection dans une distribution normale consiste à choisir les 27 % supérieurs et inférieurs de l'échantillon dont on dispose, et l'estimateur de sélection est alors 1,5 fois plus efficace (puisque $i = 2,45$) que l'estimateur de régression pour un même nombre d'individus typés. Cela veut dire que l'effet minimum détectable à un locus est les deux-tiers du minimum détectable par régres-

Tableau 5 - Variation de précision de la sélection combinée assistée par marqueurs relativement à une sélection combinée classique.

| En fonction de | En situation de déséquilibre de liaison | En situation d'équilibre de liaison |
|----------------------|--|---|
| m et n | diminue quand m et n augmentent | augmente avec m et n |
| p | augmente jusqu'à au moins 180 % ⁽¹⁾ | augmente jusqu'à au plus 152 % ⁽¹⁾ |
| h² | augmente fortement quand h ² diminue | augmente faiblement quand h ² diminue |

(1) pour $h^2 = 0,10$, m grand et $n = 1$

sion, à coût égal de typage. Mais l'estimateur de régression a l'avantage de s'appliquer à tous les caractères qui peuvent se mesurer sur l'échantillon, alors que l'estimateur de sélection ne s'applique qu'au caractère sélectionné. On peut cependant envisager une sélection sur un indice combinant plusieurs caractères et ainsi mesurer l'effet du gène sur cet indice.

Un avantage supplémentaire du génotypage sélectif pour évaluer les effets d'un locus sur un caractère quantitatif peut être attendu de la possibilité de déterminer directement les fréquences géniques sur des mélanges d'ADN, grâce à des mesures densitométriques de l'intensité des bandes correspondant aux différents allèles d'un locus marqueur (voir, par exemple, MOSIG et al., 1998). Cette technique de *mélange sélectif d'ADN* peut par exemple être appliquée aux deux échantillons constitués en regroupant séparément les individus extrêmes de la distribution du caractère étudié. Le coût de collecte et de typage des deux échantillons ainsi constitués peut s'écrire $Nc + 2t$, si on suppose que c représente le coût individuel de la collecte d'échantillon et de l'extraction de l'ADN et t le coût de typage d'un mélange d'ADN de $N/2$ individus. En supposant le même coût t pour un typage individuel, le coût total des opérations de collecte, extraction de l'ADN et des N typages nécessaires pour l'estimation par régression est $N(c + t)$. Le rapport des coûts est donc $N(c + t) / (Nc + 2t)$, soit approximativement $1 + t/c$, puisque $2t / N$ tend à devenir négligeable par rapport à c quand N est grand. Avec les techniques actuelles de génotypage des microsatellites, le coût individuel t pour un balayage complet du génome (environ 120 microsatellites) représente environ $6c$, et l'estimateur par régression est alors environ 7 fois plus coûteux à mettre en œuvre que le génotypage sélectif sur mélange d'ADN. Si on multiplie le rapport d'efficacité statistique des deux techniques (1,5) par le rapport des coûts (7), on voit que l'efficacité globale de la technique de mélange sélectif d'ADN est environ 10 fois supérieure à la technique de régression. Dans ce cas, si des effets génétiques pour plusieurs caractères sont à estimer, l'échantillon pourra être soumis à une sélection pour chacun des caractères et la procédure restera économiquement avantageuse pour un nombre de caractères étudiés pas trop élevé. Cette procédure d'évaluation des marqueurs par sélection constitue donc une possibilité intéressante par rapport à la procédure de régression préconisée par exemple par LANDE et THOMPSON (1990) en préalable à la SAM.

En situation d'équilibre de liaison, les précédentes techniques ne peuvent pas s'appliquer, puisqu'il n'y a pas d'association entre les marqueurs et les caractères. Des plans d'expérience spécifiques sont alors nécessaires pour localiser les gènes responsables des caractères et pour estimer leurs effets. Chez le porc, le système le plus répandu consiste à réaliser une F_2 à partir d'un croisement entre des races génétiquement très différentes et qu'on s'attend donc à voir exprimer de grandes différences du point de vue de leurs gènes, qu'il s'agisse de gènes marqueurs ou de gènes responsables de caractères quantitatifs. Ces derniers sont très efficacement détectés dans des croisements de ce type, comme le montrent les résultats récents obtenus dans des croisements impliquant le sanglier ou des races chinoises (voir les premiers résultats

français de BIDANEL et al., 1998, et la mise au point de ROTHSCHILD, 1998).

L'utilisation en SAM des gènes détectés dans ces expériences n'est cependant pas très simple. Car une variation génétique détectée à un locus dans de tels croisements peut ne pas exister au sein de la race ou lignée particulière dans laquelle les marqueurs correspondants pourraient être utilisés en sélection, ce qui rendrait la SAM totalement inefficace. Les résultats obtenus dans ces croisements ont d'abord l'intérêt de mettre en évidence les régions du génome susceptibles d'exprimer une variation génétique exploitable par SAM. Il reste à faire un examen plus détaillé de la région correspondante dans la population soumise à la sélection, ce qui est alors possible à un moindre coût par un balayage génomique plus restreint. Dans cette recherche, il y aura intérêt à déceler les situations de déséquilibre de liaison, qui permettront une SAM à la fois plus efficace et plus simple à mettre en œuvre en pratique, comme l'ont indiqué SMITH et SMITH (1993). La cartographie comparée pourra par ailleurs aider à mettre en évidence les gènes impliqués et susceptibles d'être directement sélectionnés (gènes dits candidats).

Il faut également souligner, en conclusion, que les prédictions d'efficacité de la SAM ne peuvent concerner que le court terme. S'il existe des associations entre marqueurs et gènes de caractères quantitatifs (situation D), celles-ci tendent à se dénouer progressivement au cours des générations successives par suite des recombinaisons entre les locus concernés. Si l'on se trouve en situation d'équilibre (E), les changements de fréquences géniques induits par la sélection aux locus responsables du caractère sélectionné vont affecter les variances génétiques à ces locus et par conséquent influencer la valeur du paramètre p . On peut donc s'attendre à voir dans tous les cas l'efficacité de la SAM décroître au cours des générations successives. Un réexamen périodique des relations entre les marqueurs et les caractères sera donc nécessaire.

CONCLUSION

Malgré les nombreuses incertitudes qui demeurent sur la valeur réelle de la SAM, et les difficultés d'obtenir les informations requises pour sa mise en œuvre, il est possible de prédire sans prendre trop de risque qu'une SAM de plus en plus efficace pourra être appliquée chez le porc dans un avenir proche. Ce sera le résultat d'une amélioration constante de notre connaissance du génome porcin et de la prise de conscience de son polymorphisme quasi illimité au niveau de l'ADN. Cette évolution est déjà reflétée dans les avis contrastés émis dans le passé par des généticiens sur une période de temps de plus de 20 ans. En écho au pessimisme exprimé par SMITH en 1967, "il semble n'y avoir aucun locus susceptible d'être utilisé avec confiance dans l'amélioration des caractères économiques chez les animaux d'élevage", on peut en effet opposer le point de vue plus récent de LANDE et THOMPSON (1990), "les polymorphismes génétiques moléculaires peuvent être utilisés pour réaliser des accroissements substantiels d'efficacité de la sélection artificielle". Des espoirs

sont également fondés sur les améliorations à attendre dans les technologies de marquage du génome, qui devraient réduire les coûts d'obtention de l'information nécessaire pour une SAM efficace. Après obtention de

cette information, à cause du nombre restreint de marqueurs utilisés simultanément dans de tels programmes, le coût additionnel de leur mise en application dans la pratique devrait rester modéré.

RÉFÉRENCES BIBLIOGRAPHIQUES

- BIDANEL J.P., MILAN D., CHEVALET C., WOLOSZYN N., BOURGEOIS F., et al., 1998. Journées Rech. Porcine en France, 30, 117-125.
- CHEVALET C., GILLOIS M., VU TIEN KHANG J., 1984. Génét. Sél. Evol., 16, 431- 444.
- FALCONER D.S., MACKAY T.F.C., 1996. Introduction to Quantitative Genetics. Longman, Harlow, UK.
- FERNANDO R.L., GROSSMAN M., 1989. Genet. Sel. Evol., 21, 467-477.
- LANDE R., THOMPSON R., 1990. Genetics, 124, 743-756.
- MEUWISSEN T.H.E., GODDARD M.E., 1996. Genet. Sel. Evol., 28, 161-176.
- MOSIG M.D., LIPKIN E., DARVASI A., EZRA E., FRIEDMANN A., SOLLER M., 1998. In : Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, 26, 253-256.
- NEIMANN-SORENSEN A., ROBERTSON A., 1961. Acta Agric. Scand., 11, 163-196.
- OLLIVIER L., 1998. Genetics, 148, 1367-1372
- OLLIVIER L., GELLIN J., MILAN D., POPESCU P., VAIMAN M., YERLE M., 1995. Journées Rech. Porcine en France, 27, 127-134.
- OLLIVIER L., MESSER L.A., ROTHSCCHILD M.F., LEGAULT C., 1997. Genet. Res., 69, 227-232.
- ROTHSCCHILD M.F., 1998. In: Proceedings Symposium Series 1. The 8th World Conference on Animal Production. Séoul, Corée, 186-192.
- RUANE J., COLLEAU J.J., 1995. Genet. Res., 66, 71-83.
- SMITH C., 1967. Anim. Prod., 9, 349-358.
- SMITH C., SMITH D.B., 1993. Animal Breeding Abstracts, 61, 197-204.