

F. HOSPITAL et J.-M. ELSEN*

INRA Laboratoire de Génétique cellulaire
BP 27 31326 Castanet-Tolosan Cedex

*INRA Station d'Amélioration génétique des
animaux BP 27 31326 Castanet-Tolosan Cedex

Apports actuels et futurs des marqueurs génétiques dans l'amélioration des populations animales

Introgression génique assistée par marqueurs

L'introgression génique consiste à inclure dans le génome d'une race sélectionnée A, un et un seul gène favorable, G, d'une race B par ailleurs moins productive. Ceci s'effectue classiquement par un premier croisement AxB, suivi d'une série de croisements en retour (Backcross), des descendants portant le gène G, par la race A. Tous ces descendants sont par construction hétérozygotes Gg (nous supposons pour simplifier qu'il y a un seul allèle de G dans la race A : g). L'introgression s'achève si besoin par un accouplement entre animaux Gg de la race A (intercross), afin de produire des homozygotes GG de "race pure" A.

L'utilisation de marqueurs peut améliorer ce processus de deux façons :

- dans une première étape, pour trier les animaux selon leur génotype au locus majeur, c'est-à-dire choisir des reproducteurs porteurs de G.

- dans une deuxième étape, pour trier, parmi les reproducteurs porteurs de G, ceux qui sont les plus proches du "type génétique" A.

Le processus d'introgression sera amélioré si le résultat final (obtention d'un effectif minimal d'animaux GG possédant un pourcentage maximum de gènes de la race A) est obtenu plus vite et/ou à moindre coût.

1 / Utilisation des marqueurs pour trier sur le génotype majeur

A chaque génération, on choisit des reproducteurs porteurs de G. Si, pour des raisons de coût, le nombre de reproducteurs doit être limité, on choisira en priorité le sexe le plus efficace pour la reproduction, c'est-à-dire les mâles. Dans le cas où le génotype des animaux est visible sans erreur dès la naissance sur les mâles comme les femelles, il est inutile de faire appel à un marqueur. Dans tous les autres cas, l'utilisation d'un marqueur améliore potentiellement l'introgression.

Caractère mesuré sur un seul sexe (exemple : le taux d'ovulation qui n'est mesurable que chez les femelles). En l'absence de marqueur, deux solutions sont envisagées :

- n'utiliser dans les croisements en retour que les femelles porteuses de G, ce qui est peu efficace dans

les espèces où les femelles ont une faible capacité de reproduction.

- tester sur descendance des mâles, en accouplant des candidats à des conjointes gg et en observant leurs filles. Pour le croisement final, et la constitution de la souche homozygote GG, ce testage est obligatoire. Il présente deux inconvénients : il augmente les besoins en animaux puisque des reproducteurs inutiles (se révélant gg) sont utilisés, et peut multiplier par deux la durée du processus puisque, pour chaque génération de mâles, il faut d'abord produire une génération de testage, puis une génération de croisement en retour. Ce délai sera toutefois annulé s'il est possible, par les mêmes accouplements ou en parallèle, de tester les mâles et de produire la génération suivante d'introgression.

Caractère mesuré tardivement dans la vie de l'animal (exemple : la finesse de la laine). En l'absence de marqueur, aucun choix n'est possible avant que le caractère ne soit visible. Il faut donc entretenir, jusqu'à leur mesure, des animaux inutiles. En outre, si cette expression n'apparaît que longtemps après la puberté, ceci allonge les intervalles de génération. Il est malgré tout envisageable de mettre à la reproduction les jeunes, mais seule la moitié des accouplements (le quart pour le croisement final) seront utiles.

Caractère nécessitant la destruction de l'animal (exemple : la qualité de la viande). Si la mesure peut se faire chez l'adulte, il sera envisageable de mettre des candidats à la reproduction puis, a posteriori, d'éliminer la descendance des non porteurs. Cette solution se traduira comme précédemment par la procréation de nombreux animaux inutiles. Dans le cas où la congélation de gamètes, en particulier des spermatozoïdes, est possible, les accouplements pourront attendre les informations sur le génotype au locus majeur et le coût de l'impossibilité de mesurer les génotypes à la naissance sera uniquement l'entretien de futurs reproducteurs inutiles jusqu'à la congélation de leur gamètes. Si la mesure doit se faire chez le jeune (c'est le cas de la plupart des caractères bouchers), seul le testage sur descendance permet actuellement de résoudre le problème - la fabrication de vrais jumeaux étant une solution envisageable dans l'avenir. Ceci se traduira par un éventuel allongement du processus (à moins, comme plus haut, de réaliser en parallèle les accouplements de testage et

d'introggression) et surtout par une augmentation considérable des effectifs impliqués : si le caractère s'exprime sans erreur de classification, il faut 5 descendants par reproducteur pour connaître son génotype propre avec une probabilité d'erreur de 5%.

Gène majeur avec pénétrance incomplète (exemple : la conformation). Le génotype des candidats n'est pas connu de façon certaine. Il est possible d'améliorer la précision du typage en considérant l'ensemble des informations sur les apparentés, mais cette amélioration ne sera substantielle que lorsque les informations sur la descendance seront disponibles, ce qui revient à prévoir un testage sur descendance, donc, encore une fois à allonger les délais et augmenter la taille de la population expérimentale. Dans la plupart des cas, des erreurs de classification seront commises et devront être rattrapées, a posteriori, en éliminant des descendance entières.

En conclusion, les quatre situations qui viennent d'être décrites se traduisent, par rapport au cas idéal, par au moins un des inconvénients suivants :

- entretien d'un effectif accru d'animaux, des individus gg (inutiles) étant conservés jusqu'à la mesure ou la confirmation de leur génotype.

- allongement des intervalles de générations, soit pour obtenir plusieurs reproducteurs par femelle, soit pour tester les reproducteurs sur descendance.

La connaissance d'un marqueur M du locus G remédie à ces inconvénients (voir par exemple Soller et Plotkin-Hazan 1977). Une situation idéale serait celle où les reproducteurs B de départ sont tous porteurs d'allèles marqueurs inexistantes dans la race A. En l'absence de recombinaison, il suffirait de choisir comme reproducteurs à chaque génération les porteurs de ces allèles, le choix final des individus GG se faisant parmi les porteurs de deux allèles de la race B. Par rapport à cette situation idéale, deux phénomènes non indépendants peuvent limiter l'intérêt des marqueurs : un équilibre de liaison partiel entre M et G, et la possibilité de recombinaison entre les deux locus.

Équilibre de liaison partiel. Supposons qu'il n'y ait que deux allèles en M, M1 et M2, que les fondateurs de l'introggression (les animaux GG de la race B) soient tous M1M1 et que les deux allèles soient présents dans la race A. Un reproducteur M1G/M2g, de phase connue, peut donc être accouplé à des conjoints de la race A de génotype M1g/M1g, M1g/M2g ou M2g/M2g. Les accouplements avec les hétérozygotes M1M2 donneront des descendants dont le génotype en G ne pourra se déduire de celui en M dans la moitié des cas. Par ailleurs, les descendants M1G/M1g des accouplements avec les conjoints M1M1 ne sont pas doubles hétérozygotes et le marqueur sera inefficace pour typer leur propre descendance. Il faut donc se restreindre à utiliser des conjoints M2M2, ce qui impose d'une part la mesure de tous les génotypes en M des candidats conjoints, d'autre part une sélection très réduite de ces candidats sur d'autres critères zootechniques que leur génotype en M. L'utilisation de marqueurs très polymorphes réduira cet inconvénient en augmentant la proportion de conjoints utilisables.

Recombinaison possible. Si M et G ne sont pas parfaitement liés, des erreurs seront commises dans l'interprétation des génotypes. Dans ce cas, l'utilisation d'un marqueur doit être conçue comme une pré-sélection des reproducteurs, le choix définitif étant établi après l'observation des phénotypes pour le

caractère quantitatif gouverné par le locus G. Selon la distance entre M et G, cette présélection pourra être plus ou moins intense, l'organisation concrète de l'introggression devant être raisonnée au cas par cas.

A l'avenir, il faudra développer des critères statistiques de tri des animaux incluant l'ensemble des informations sur les performances quantitatives et sur le génotype en M des animaux et de leurs apparentés. Une piste évidente est le prolongement des méthodes de maximum de vraisemblance actuellement utilisées pour typer les reproducteurs dans des introggressions sans marqueurs. Enfin, rappelons que l'utilisation de deux marqueurs encadrant le locus majeur avec des pourcentages de recombinaison r_1 et r_2 , augmente considérablement la puissance de la technique, la probabilité d'erreur n'étant plus que de $r_1 r_2$ dans ce cas.

2 / Utilisation des marqueurs pour augmenter le taux de génome receveur chez les porteurs de g

A chaque génération, une fois les descendants porteurs de G sélectionnés, on peut laisser le restant du génome évoluer au hasard, ou sélectionner les descendants sur des critères phénotypiques caractéristiques de A (méthodes classiques). On peut aussi penser à sélectionner les descendants sur des marqueurs génomiques caractéristiques de A : c'est le point que nous allons développer. D'un point de vue théorique, ce cas particulier de sélection assistée par marqueurs peut être étudié directement en terme de composition génomique des descendants, sans qu'il soit besoin de faire des hypothèses sur le passage du génotype au phénotype qui pose de gros problèmes dans les autres cas. On peut donc penser que les prédictions obtenues dans ce cas seront plus fiables. Les résultats suivants sont tirés de Hospital *et al* (1992).

Il est évident qu'une sélection portant sur des marqueurs moléculaires bien identifiés permettra rapidement d'augmenter la fréquence de ces marqueurs dans la population, encore qu'une fixation complète de nombreux marqueurs en une génération soit impossible (voir plus loin). Mais au-delà, tout le génome ne pouvant être marqué, il s'agit surtout de savoir comment une sélection sur des marqueurs permettra d'entraîner les parties non marquées du génome, en fonction de leur pourcentage de recombinaison avec les marqueurs. Deux cas sont à distinguer :

- les parties du génome non liées au gène introgressé G ("chromosomes non porteurs"), où seul le phénomène précédent sera à considérer (figure 1).

- les parties du génome liées au gène G ("chromosome porteur"), où la sélection de descendants portant tous le gène G crée un entraînement fort de génome de type "B" autour du gène. Techniquement, des résultats analytiques ne peuvent être obtenus que pour une seule génération de sélection. Pour plusieurs générations, il faudra avoir recours aux simulations.

Évolution aléatoire sans sélection. En croisant à chaque génération les descendants par la race "A", on augmente le taux de gènes d'origine "A" dans leur génome, même sans sélection. Les taux attendus en espérance peuvent être calculés (tableau 1).

Figure 1. Réponse locale à la sélection. Sélection appliquée en deux marqueurs, au long d'un chromosome représenté par 21 locus distants de 5 Centimorgans (cM) figurés en abscisse, soit une longueur totale de 100 cM. Moyenne sur 20 chromosomes non porteurs. Ordonnée : taux de génome A après sélection. 20 individus porteurs de G sélectionnés sur 200 à chaque génération. Les générations BC1 à BC6 sont figurées de bas en haut.

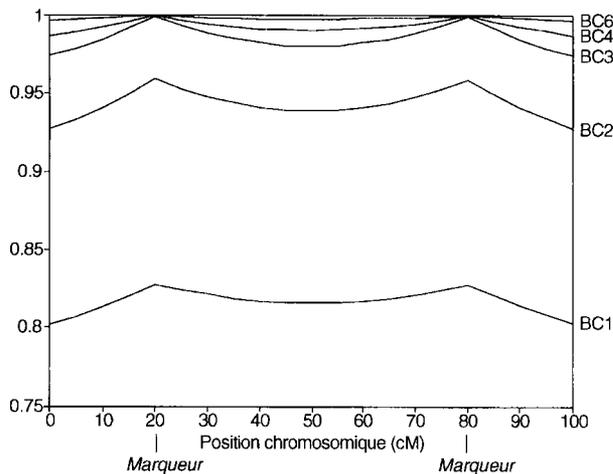


Tableau 1. Pourcentage de génome A sans sélection.

Génération	F1	BC1	BC2	BC3	BC4	BC5	BC6
Chr. non porteur	50,0	75,0	87,5	93,7	96,8	98,4	99,2
Chr. porteur	50,0	60,7	68,4	74,1	78,4	81,7	84,2

Densité des marqueurs. Avec L marqueurs indépendants, la probabilité $(1/2)^L$ que tous les marqueurs se trouvent dans un même gamète devient rapidement négligeable avec L . Si l'on tient compte des recombinaisons dans un génome réaliste, la figure 2 montre qu'à partir d'un certain seuil dépendant de la longueur du génome, augmenter la densité des marqueurs n'augmente pas la réponse à la sélection. Pour des densités moyennes, des marqueurs répartis uniformément sur le génome fournissent de meilleurs résultats que des marqueurs répartis au hasard.

On voit donc que la marge de manoeuvre disponible pour la sélection est limitée à la fois extrinsèquement par l'évolution déjà rapide de la composition du génome sans sélection, et intrinsèquement par l'efficacité même des marqueurs. De plus, les événements de recombinaison s'accumulant au cours du temps, la sélection est d'autant plus efficace qu'elle est pratiquée plus tard, lorsqu'il ne reste que peu à gagner, comme le montre la figure 3 qui présente l'évolution de la réponse à la sélection sur le chromosome porteur. D'un point de vue purement mathématique et abstrait, pour obtenir des descendants contenant 100 % de gènes "A" à part G, il vaudrait mieux sélectionner le plus tard possible, autrement dit pas du tout ! Dans la pratique, on peut se fixer un taux de génome "A" à atteindre et espérer l'obtenir plus rapidement en pratiquant une sélection sur les marqueurs. Le tableau 1 montre que l'effort de sélection doit porter en priorité sur le chromosome porteur, sur lequel le taux de "A" sans sélection est le plus bas.

Figure 2. Effets du nombre et de la répartition des marqueurs. Moyenne sur 20 chromosomes non porteurs. Ordonnée : taux de génome A après sélection. 20 individus porteurs de G sélectionnés sur 200 à chaque génération. Générations BC1 à BC6 de bas en haut. Trait continu : marqueurs répartis uniformément sur chaque chromosome. Tirets : même nombre de marqueurs sur chaque chromosome, positions aléatoires. Pointillés : marqueurs répartis au hasard sur l'ensemble du génome.

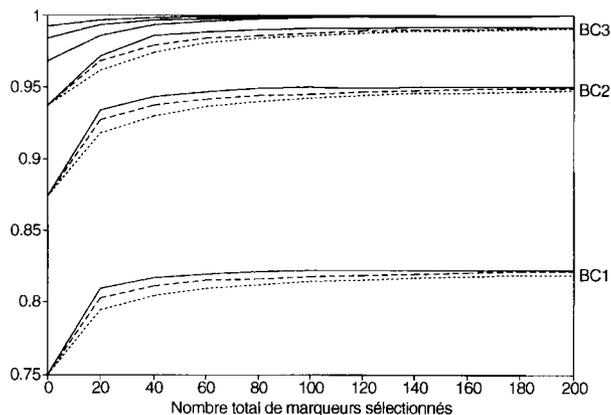
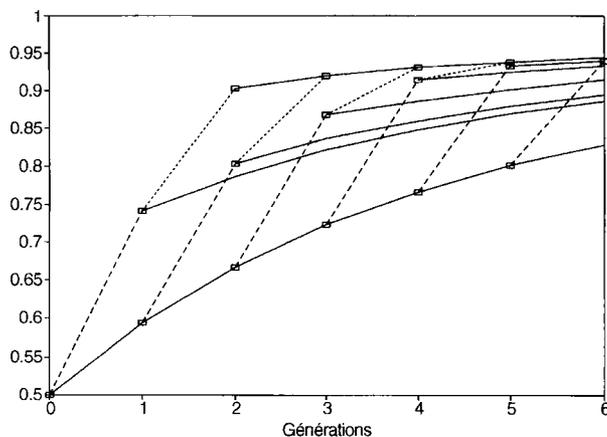


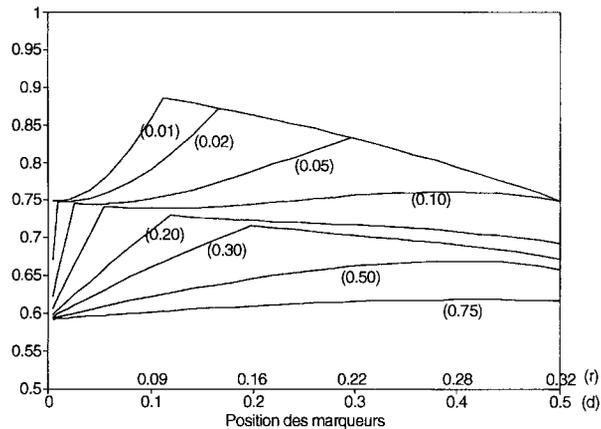
Figure 3. Réponse à la sélection sur le chromosome porteur selon les générations où la sélection est appliquée. Deux marqueurs flanquant G à 10 cM. 20 individus porteurs de G éventuellement sélectionnés sur 200 à chaque génération. Trait continu : pas de sélection. Tirets : première sélection. Pointillés : deuxième sélection. Carrés : choix entre sélection ou pas de sélection.



Sélection sur le chromosome porteur seul. La figure 4 présente la réponse à la sélection sur le chromosome porteur. La sélection retient une proportion de la population (valeurs entre parenthèses) et porte sur deux marqueurs flanquant G à une distance d figurée en abscisse. Cette figure montre qu'il y a une interaction entre le taux de sélection pratiqué, et la distance (ou le pourcentage de recombinaison) entre G et les marqueurs. Autrement dit, il n'est pas toujours plus efficace de sélectionner fortement sur des marqueurs proches du gène comme cela se pratique couramment. Il faut plutôt :

- si une intensité de sélection maximale est fixée par les caractéristiques de l'espèce ou les conditions

Figure 4. Réponse à la sélection sur le chromosome porteur selon la proportion d'individus sélectionnés et la position des marqueurs. Pourcentage de génome A après une génération de sélection. r : taux de recombinaison entre G et chacun des deux marqueurs. d : distance correspondante en Morgans ($d = -1/2 \cdot \ln[1-2r]$).



expérimentales, choisir de sélectionner sur des marqueurs apportant une réponse optimale pour cette intensité donnée.

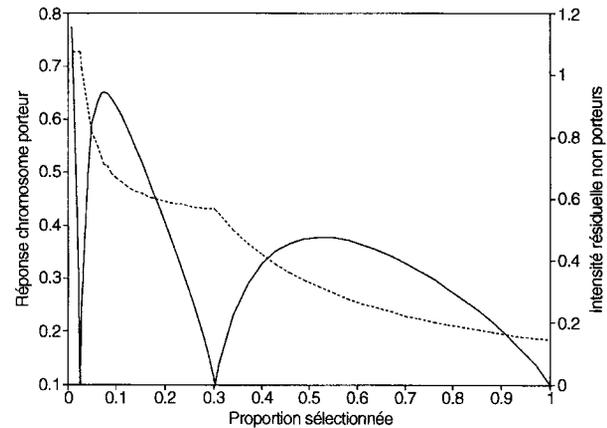
- ou, si le choix du couple de marqueurs est limité par l'avancement de la carte génique de l'espèce, choisir le taux de sélection optimal dans ce cas.

Dans le cas favorable où une carte assez complète est disponible, et où l'intensité de sélection n'est pas limitée, cette figure montre aussi qu'il vaut mieux sélectionner dans les premières générations pour des marqueurs assez distants du gène G, et seulement dans les générations suivantes pour des marqueurs proches. Dans tous les cas, ceci met en évidence l'intérêt des marqueurs distaux, et par là-même d'une carte complète.

Sélection sur chromosomes non porteurs.

A chaque génération, une première sélection ayant été pratiquée sur le chromosome porteur, il est possible d'en pratiquer une deuxième visant à augmenter le taux de gènes d'origine "A" sur les chromosomes non porteurs. Néanmoins, l'intensité de sélection disponible pour cette seconde sélection dépend de la position des marqueurs et du taux de sélection choisi dans la première étape. Ceci est représenté figure 5 : lorsque le taux de sélection lu en abscisse est appliqué sur le chromosome porteur au niveau de 2 marqueurs flanquant G à une distance d , la figure donne à la fois la réponse obtenue sur le chromosome porteur (pointillés) et l'intensité de sélection résiduelle disponible pour une deuxième étape de sélection

Figure 5. Intensité de sélection disponible pour sélectionner sur les chromosomes non porteurs après une première sélection sur le chromosome porteur. Explications dans le texte.



sur les autres chromosomes (trait continu). Cette dernière varie selon que le nombre final d'individus retenus après la deuxième étape de sélection est inférieur, supérieur ou égal au nombre de recombinants entre G et un marqueur sur le chromosome porteur. Par exemple, si le nombre des individus retenus est égal au nombre de recombinants sur le chromosome porteur, les chromosomes non porteurs sont choisis au hasard, et l'intensité de sélection sur ces derniers est nulle.

Conclusion

Les marqueurs sont efficaces pour accélérer les programmes d'introgession génique. D'une part, des marqueurs proches du gène introgressé permettent une économie de temps et de moyens, qui peut être considérable, pour la détermination des individus porteurs du gène. D'autre part, des marqueurs répartis sur l'ensemble du génome permettent de diminuer plus rapidement le taux de génome donneur chez les individus porteurs du gène introgressé. Dans ce dernier cas, l'efficacité de la sélection sur marqueurs est inférieure à celle prédite par Hillel et al (1990), mais néanmoins conséquente : 4 générations de sélection suffisent pour obtenir une réponse correspondant à 6 générations sans sélection (figure 2). Dans tous les cas, une carte génétique complète et des marqueurs bien répartis sont préférables.

Références bibliographiques

Hillel J., Schaap T., Haberfeld A., Jeffreys A. J., Plotzky Y., Cahaner A., Lavi U., 1990 DNA fingerprint applied to gene introgression breeding programs. *Genetics* 124: 783-789.

Hospital F., Chevalet C., Mulsant P., 1992 Using markers in gene introgression breeding programs. *Genetics* : sous presse.

Soller M., Plotkin-hazan J., 1977. The use of marker alleles for the introgression of linked quantitative alleles. *Theor. Appl. Genet.* 51 : 133-137.