

D. BOICHARD, P. LE ROY,  
H. LEVÉZIEL\*, J.-M. ELSEN\*\*

INRA Station de Génétique Quantitative  
et Appliquée, 78352 Jouy-en-Josas Cedex

\* INRA, Laboratoire de Génétique Biochimique  
et de Cytogénétique, 78352 Jouy-en-Josas Cedex

\*\* INRA, Station d'Amélioration Génétique  
des Animaux, BP 27, 31326 Castanet Cedex

## Utilisation des marqueurs moléculaires en génétique animale

Les marqueurs sont des fragments d'ADN qui servent de repères pour suivre la transmission d'un segment de chromosome d'une génération à l'autre. Ils sont notamment utilisés pour rechercher les gènes qui gouvernent les caractères d'intérêt zootechnique, mais ils peuvent aussi servir, par exemple, à maintenir la variabilité génétique des petites populations. L'article rappelle d'abord les notions à la base du concept de marqueur puis présente les différentes utilisations actuelles et potentielles.

Bien que les lois de Mendel décrivant la transmission des gènes entre générations soient connues depuis longtemps, il n'est généralement pas possible de les vérifier sur les caractères économiquement importants. Quelques gènes à effet majeur, comme le gène de sensibilité à l'halothane chez le porc, le gène culard chez le bovin, le gène Booroola chez les ovins, le gène de la caseine  $\alpha_{s1}$  chez la chèvre ou le gène de nanisme de la poule constituent pour l'instant des exceptions. En général, au niveau d'une population comme intra descendance, les caractères d'intérêt présentent souvent une distribution continue, dans laquelle il n'est pas possible de distinguer un nombre fini de génotypes. Ceci s'explique par le fait que ces caractères sont soumis à des effets de milieu, aux effets de

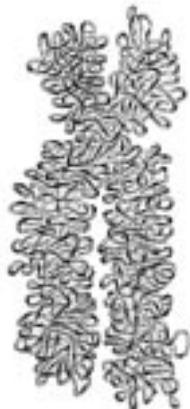
plusieurs gènes, et éventuellement à leurs interactions. Pour analyser ces données et sélectionner les populations, un modèle a été développé initialement par des mathématiciens. Ce modèle, dit infinitésimal, suppose que le caractère est soumis de façon additive à des effets de milieu et aux effets d'un nombre infini de gènes, non identifiables isolément. Ce modèle n'a pas pour but d'expliquer le déterminisme génétique des caractères mais de prédire certains paramètres comme le degré de ressemblance entre apparentés, la supériorité génétique transmissible d'un reproducteur, ou le progrès génétique réalisable par sélection, à partir de deux types d'information, les observations du caractère (ou phénotype) et les généalogies. C'est l'outil de base en amélioration génétique. Le fait que ce modèle infinitésimal, véritable boîte noire mathématique, soit opérationnel et performant ne signifie en aucune façon qu'il est correct mais plutôt qu'il est robuste envers une réalité sans doute très différente. Toutefois, bien que les gènes agissant sur les caractères d'intérêt restent en général méconnus, faute de pouvoir les mettre en évidence individuellement, on peut cependant supposer que leurs effets individuels, pour certains d'entre eux, peuvent être importants.

Grâce aux progrès récents et spectaculaires de la génétique moléculaire qui ont permis, entre autres, la construction de cartes génétiques dans différentes espèces animales, l'identification de ces gènes peut être abordée. Cet article présente les perspectives offertes dans l'étude analytique des caractères et les procédures de sélection et de gestion des populations.

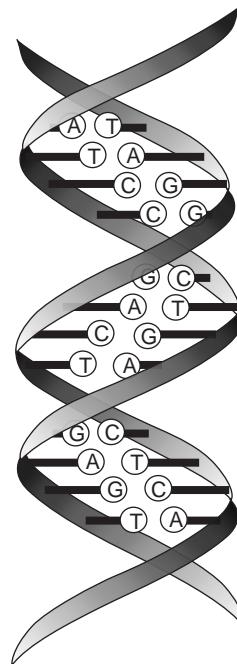
### Résumé

Grâce aux progrès récents de la biologie moléculaire, en particulier la découverte des séquences microsatellites et la technique d'amplification des séquences d'ADN par PCR, de nombreux marqueurs génétiques sont maintenant disponibles et organisés en cartes génétiques dans la plupart des espèces d'élevage. Cet article présente d'abord les principales propriétés des marqueurs génétiques, en particulier le polymorphisme et la liaison, qui permettent d'identifier et de suivre des segments chromosomiques entre générations. Il décrit ensuite les principales applications possibles des marqueurs dans la gestion des populations et l'analyse de leur variabilité génétique, et il met l'accent sur l'application la plus développée à l'heure actuelle, la détection des principales régions chromosomiques (ou QTL) impliquées dans le déterminisme des caractères d'intérêt économique, préliminaire indispensable à leur utilisation par sélection assistée par marqueurs ou introgression. Enfin, cet article présente succinctement les méthodes, en particulier celles basées sur la cartographie comparée, pour passer des marqueurs aux gènes responsables de la variabilité génétique des caractères d'intérêt, ouvrant ainsi la voie à une meilleure compréhension des mécanismes impliqués et à une utilisation plus facile et plus généralisable des résultats.

L'ADN est organisé en chromosomes. L'ensemble des chromosomes constitue le **génome**.



*Chromosome au stade métaphase de la méiose.*

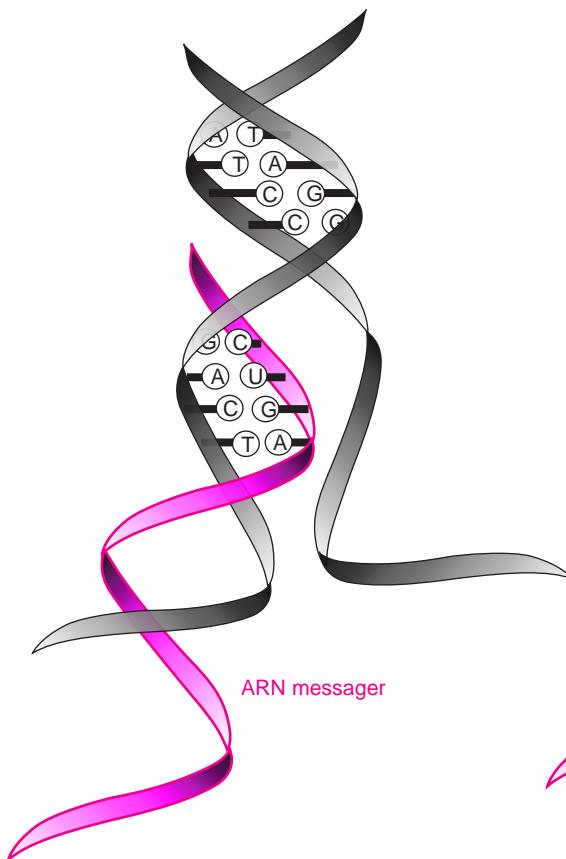


La molécule d'ADN se présente sous forme d'une hélice à deux brins. Chaque brin est constitué de l'enchaînement de 4 éléments de base appelés nucléotides. Ces 4 nucléotides diffèrent par les bases azotées qui les constituent : A, C, G ou T. Les enchaînements des nucléotides sur les deux brins sont complémentaires (bases C avec G, T avec A).

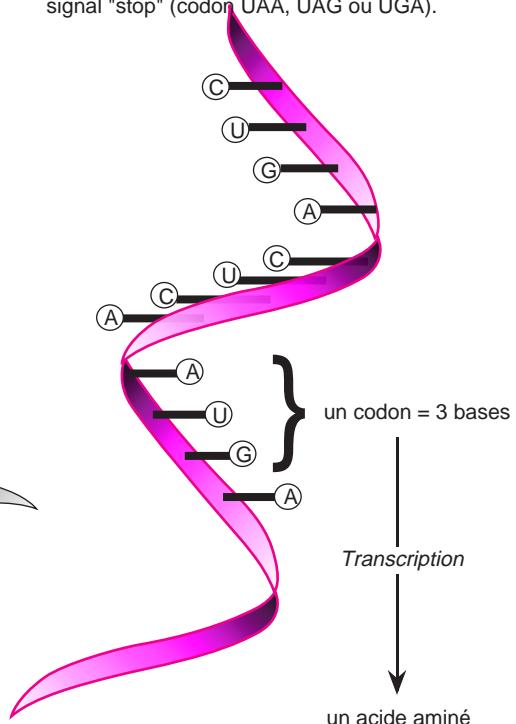
Un **gène** est une portion de la molécule d'ADN. Il se définit par la séquence des nucléotides.

Les gènes sont la partie codante du génome, c'est-à-dire les portions du génome qui seront traduites en protéines. Les gènes ne représentent que 5 à 10 % du génome.

La synthèse des protéines se déroule en deux phases. Au cours de la première, l'ADN est "transcrit" en ARNmessenger (les bases utilisées sont les mêmes que dans l'ADN sauf U à la place de T).



Au cours de la deuxième phase, l'ARN messenger est "traduit" en protéines : à chaque triplet (ou **codon**) de nucléotides correspond un acide aminé. La traduction est initiée par une séquence particulière de nucléotides, puis elle démarre au premier triplet qui code pour le premier acide aminé de la protéine. Elle se poursuit ensuite jusqu'à un signal "stop" (codon UAA, UAG ou UGA).



À un emplacement (= **locus**) donné du génome, c'est-à-dire pour un segment d'ADN donné, que ce soit ou non un gène, la séquence des nucléotides peut varier d'un individu à l'autre. Cette variabilité définit le **polymorphisme** génétique, les différentes séquences au locus étant dites **allèles**.

## 1 / Définition et propriétés des marqueurs

### 1.1 / Rappels

L'information génétique est stockée dans chaque cellule sous la forme de longues molécules d'acide désoxyribonucléique (ADN) constituant le génome. L'ADN est constitué de l'enchaînement précis de quatre éléments de base, les nucléotides, notés A, C, G et T (voir encadré ci-contre). Les gènes correspondent à la partie codante du génome, c'est-à-dire la partie traduite sous forme de protéines, qui globalement ne correspond qu'à une faible part du génome (5 à 10 %). L'ADN est organisé en chromosomes. Chez les espèces diploïdes, et en particulier les espèces d'élevage, l'information génétique est contenue dans  $2n$  chromosomes répartis en  $n$  paires de chromosomes homologues, chacun des chromosomes de chaque paire provenant de chaque parent.

#### a / Polymorphisme

L'analyse de la séquence nucléotidique d'un segment d'ADN (« séquençage ») à un locus donné, c'est-à-dire à un emplacement donné sur le génome (qu'il s'agisse d'un gène ou non) et dans une population donnée, peut montrer certaines variations. Les différentes formes à un locus sont dites allèles. L'existence de différents allèles possibles à un locus définit le polymorphisme génétique. En pratique, le polymorphisme peut être mis en évidence par différentes techniques, généralement plus simples et plus légères que le séquençage. Un individu portant deux fois le même allèle à un locus est dit homozygote, tandis qu'un individu portant deux allèles différents est dit hétérozygote. Seule une petite partie de cette variabilité génétique au niveau de l'ADN se traduit par une variabilité des performances.

#### b / Liaison

Lors de la méiose, le génome d'un gamète est constitué en tirant au hasard un chromosome dans chaque paire du parent. Il en résulte que deux gènes situés sur les chromosomes de deux paires différentes sont toujours transmis indépendamment l'un de l'autre. Pour deux gènes appartenant à une même paire de chromosomes homologues, on constate que s'ils sont situés sur le même chromosome, ils ont tendance à être transmis ensemble, tandis que si ces gènes sont situés sur les deux chromosomes homologues de la paire, ils ne sont généralement pas transmis ensemble. Cette règle est mise en défaut par l'existence d'échanges de matériel entre chromosomes homologues, appelés crossing-over. Ainsi, deux gènes situés initialement sur le même chromosome peuvent se retrouver sur les deux chromosomes homologues en cas de crossing-over. On dit qu'il y a recombinaison. La fréquence de recombinaison est surtout fonction de la distance entre les locus. Deux locus proches recombinent peu et ont donc

tendance à être transmis ensemble, tandis que deux locus très éloignés sur le même chromosome se comportent comme deux locus indépendants. Le taux de recombinaison entre deux locus est d'ailleurs la méthode de mesure la plus courante et la plus ancienne de la distance entre deux gènes, dite distance génétique. En moyenne, chez les mammifères domestiques, 1 % de recombinaison entre deux locus correspond à une distance physique d'un million de bases entre ces deux locus, la taille du génome étant de l'ordre de 3 milliards de paires de bases.

### 1.2 / Concept de marqueur

Polymorphisme et liaison sont à la base de la définition de la notion de marqueur.

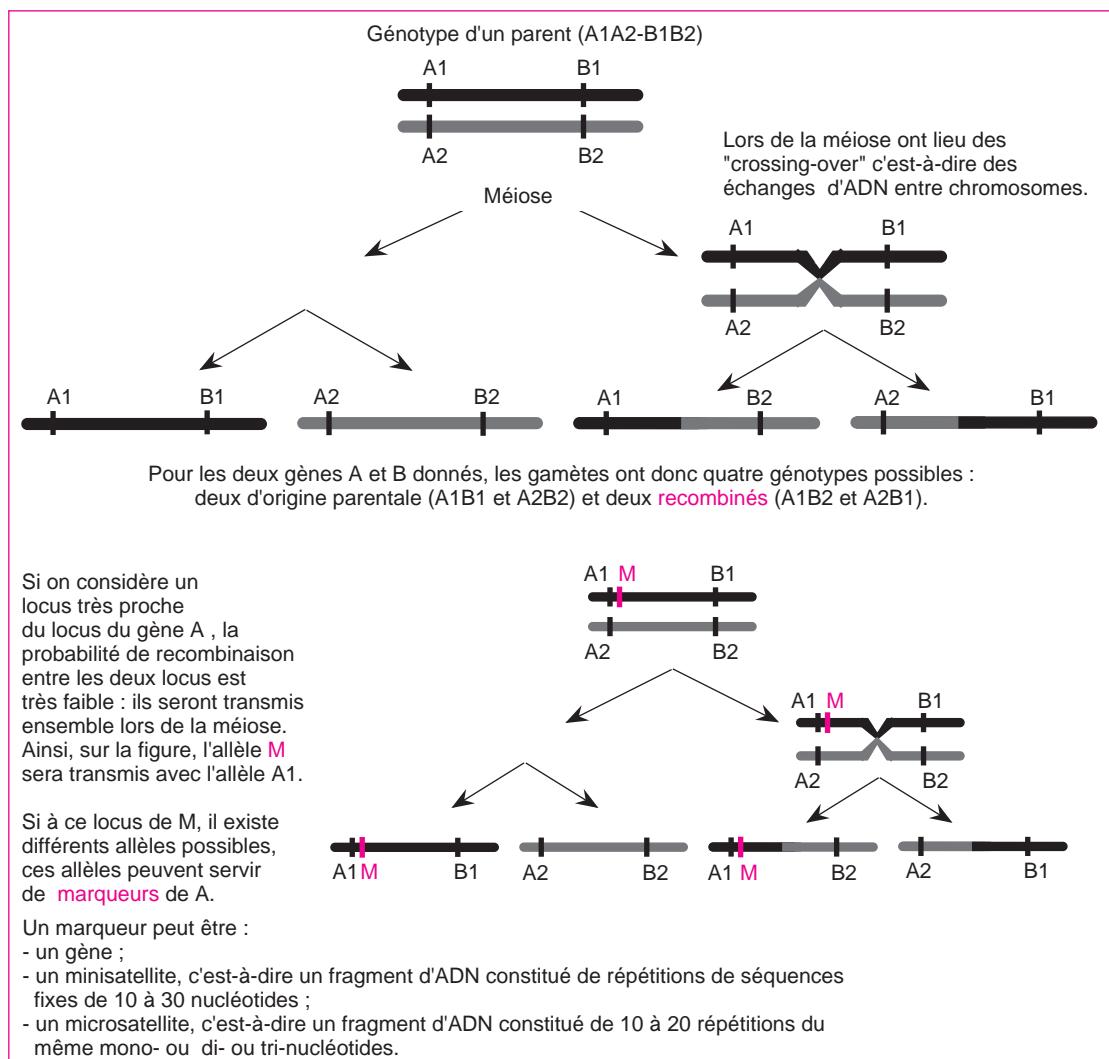
#### a / Polymorphisme

Le polymorphisme permet d'établir l'origine parentale de l'allèle à un locus donné. Ainsi, si un allèle porté par un individu est porté par son père mais pas par sa mère, l'individu l'a reçu de son père. On peut donc distinguer les chromosomes reçus du père et de la mère.

Un allèle marqueur ne permet de suivre un segment chromosomal que s'il peut être distingué des autres allèles. Intra population, un marqueur n'est même utile que s'il est polymorphe intra reproducteur, autrement dit que si le reproducteur est hétérozygote au marqueur. En effet, chez un reproducteur homozygote, le marqueur n'est pas informatif pour distinguer deux types de gamètes et donc deux types de descendants. Cependant, même à l'état hétérozygote, un marqueur n'est pas informatif à 100 %. Ainsi, dans une famille où père, mère et descendant sont M/m, le marqueur n'est pas informatif. D'une façon générale, on mesure l'informativité d'un marqueur par son aptitude à distinguer sans équivoque deux groupes de descendants selon l'allèle marqueur reçu du parent. Le PIC (=polymorphism information content) d'un marqueur, qui varie de 0 à 1, représente la proportion d'individus informatifs dans une population idéale. Un marqueur codominant, c'est-à-dire pour lequel tous les allèles peuvent être déduits simplement de l'observation du phénotype, est plus informatif qu'un marqueur dominant, dont l'allèle récessif n'est observable qu'à l'état homozygote. D'autre part, un marqueur est d'autant plus informatif que le nombre d'allèles est élevé et que leurs fréquences sont équilibrées. Ces considérations expliquent pourquoi les généticiens recherchent des marqueurs codominants très polymorphes. Pour augmenter l'informativité, une alternative consiste à considérer un groupe de marqueurs très liés comme un marqueur unique, appelé haplotype, dont le polymorphisme résulte de toutes les combinaisons alléliques de chaque marqueur élémentaire.

#### b / Liaison (figure 1)

La liaison génétique permet de généraliser les conclusions des observations sur un locus

**Figure 1.** Définition et propriétés d'un marqueur.

particulier à l'ensemble du segment d'ADN entourant ce locus : comme il n'y a pas de recombinaison entre le locus observé et le segment qui l'entoure, ce locus devient un marqueur de ce segment et de tous les gènes qu'il contient.

Le marquage est d'autant plus efficace que le segment considéré est court, ce qui limite le taux de recombinaison entre le marqueur et le gène. En pratique, dans de nombreuses applications, la distance recherchée entre marqueur et gène ne dépasse pas 20 cM. La taille du génome étant d'environ 3000 cM, 150 marqueurs bien placés constituent donc un maillage théoriquement suffisant des 100 000 gènes qui constituent le patrimoine génétique d'une espèce animale. Avant toute utilisation systématique de marqueurs, il convient donc de les localiser les uns par rapport aux autres sur une carte génétique et de vérifier que le génome est totalement couvert.

### c / Déséquilibre de liaison

Deux locus sont en équilibre de liaison si dans l'ensemble des gamètes de la population considérée et pour tout allèle M et tout allèle Q à ces deux locus, la fréquence de l'haplotype

MQ est égale au produit des fréquences des allèles M et Q. Dans ce cas, la connaissance de l'allèle à un locus ne permet pas de prédire l'allèle à l'autre locus. Au contraire, on parle de déséquilibre de liaison s'il existe des associations préférentielles entre les allèles des deux locus. Un déséquilibre de liaison est généralement dû à l'existence d'une liaison. Toutefois, l'existence d'une liaison génétique n'implique pas un déséquilibre de liaison. Dans une population fermée de grande taille, on considère l'équilibre de liaison comme l'hypothèse la plus générale. En effet, tout déséquilibre apparu par migration, mutation, sélection ou dérive génétique, est progressivement détruit par les recombinaisons au cours des générations. Un déséquilibre global au niveau de la population est donc instable et n'existe que si son origine est récente et si les locus sont très liés. Cette situation est assez défavorable car elle rend le marquage plus difficile d'utilisation. En effet, l'association entre deux allèles de deux locus, détectée sur un échantillon de la population, n'est pas généralisable à l'ensemble de la population.

En pratique, deux situations peuvent exister. Première possibilité, on crée un déséquilibre, par exemple par croisement entre deux

populations différentes. Le déséquilibre est maximal au cours des premières générations de croisement, les associations entre locus chez les gamètes ressemblant aux formes présentes dans les chromosomes des lignées parentales. Puis il s'atténue au cours des générations, avec l'accumulation des recombinaisons. Deuxième possibilité, la population est en équilibre de liaison. Dans ce cas, on utilise les déséquilibres qui existent nécessairement dans des sous-populations, en particulier intra famille. Ainsi, par exemple, dans la descendance d'un parent double hétérozygote pour deux locus liés, il y a nécessairement déséquilibre de liaison, car les gamètes parentaux sont plus nombreux que les gamètes recombinants. Toutefois, les allèles préférentiellement associés varient d'une famille à l'autre.

#### **d / Facilité de typage**

Les protocoles impliquant des marqueurs moléculaires sont fréquemment de grande taille et nécessitent des milliers, voire des dizaines ou des centaines de milliers de détermination d'allèles, ou « typages ». Les génotypages doivent donc être rapides, peu coûteux et automatisables. Sans que ce soit une obligation absolue, l'unicité de la technique de marquage est en général recherchée car elle facilite beaucoup la robotisation et diminue donc les coûts. La mise en place de structures spécialisées dans le typage à grande échelle, comme le GIE Labogena, répond à ce souci d'efficacité et de maîtrise des coûts.

Par ailleurs, l'intérêt principal d'un marqueur réside dans la précocité du typage, quand aucune autre information phénotypique n'est encore disponible. La détermination du génotype au marqueur doit être possible chez tout animal, sans limite d'âge, de sexe ou de stade physiologique. Ces considérations expliquent l'intérêt des marqueurs ADN, tandis que les marqueurs visibles ou protéiques requièrent l'expression du gène.

Pour des questions de rapidité d'utilisation et de limitation de la taille des protocoles, on peut souhaiter typer les marqueurs sur des embryons de quelques dizaines de cellules. Cette technique est actuellement utilisée pour sexer des embryons bovins en recherchant des séquences spécifiques du chromosome Y. Les deux problèmes essentiels associés au typage sur embryon sont d'une part l'altération de la qualité de l'embryon (altération de la membrane pellucide, prélèvement de quelques cellules) affectant son aptitude à la congélation et sa survie ultérieure (voir l'article de Colleau *et al.*, ce numéro), mais surtout la quantité très restreinte d'ADN disponible, avec les problèmes de sensibilité de la méthode de typage et les risques de contamination par de l'ADN exogène. A l'extrême, quelques équipes réalisent même des typages moléculaires sur spermatozoïde unique. Cette technique, dont l'intérêt zootechnique est limité, peut s'avérer très utile pour la cartographie fine de locus très liés.

## **2 / Développements récents en génétique moléculaire**

L'utilisation des marqueurs n'a pu se développer rapidement au cours des dernières années que grâce aux progrès importants de la génétique moléculaire et à la mise à disposition d'outils et de techniques, qui ont ouvert la voie à la construction des cartes génétiques dans les différentes espèces d'élevage. Parmi les développements récents, il convient de citer la mise en évidence des microsatellites et la réaction d'amplification par PCR qui ont eu un impact particulièrement marqué sur l'établissement des cartes génétiques.

### **2.1 / Microsatellites**

Jusqu'à la fin des années 1970, la principale limite à l'établissement de cartes génétiques était le manque de marqueurs polymorphes et bien répartis. Au cours des années 1980, de nouvelles sources de polymorphisme ont été découvertes. Très nombreux, les polymorphismes de longueur de fragment de restriction (RFLP), largement utilisés en cartographie végétale, se sont révélés assez limités mais surtout difficiles à analyser de façon rapide et automatisable. Constitués d'un nombre très variable de répétitions de séquences de 10 à 30 bases, les minisatellites permettent l'identification d'empreintes génétiques très spécifiques de chaque individu ; ils sont très polymorphes mais assez mal répartis sur le génome et surtout difficilement interprétables, les bandes révélées correspondant à l'observation simultanée d'un nombre variable et élevé de locus.

En 1989, la description de microsatellites a incontestablement permis l'essor des projets de cartographie. Les microsatellites sont des séquences constituées de répétitions en tandem de mono-, di - ou tri nucléotides répétés de 10 à 20 fois en moyenne. Un motif répété fréquemment rencontré est le dinucléotide TG. Très nombreuses, bien réparties sur le génome, ces séquences se caractérisent par un polymorphisme important dû à la variation du nombre de répétitions selon les allèles. L'utilisation de ce type de marqueur nécessite la mise en œuvre des deux types de techniques suivants :

- comme pour les minisatellites, une telle définition du microsatellite ne caractérise pas un locus particulier, mais toutes les séquences de ce type dans l'ensemble du génome. Le locus doit être défini de façon unique en considérant les séquences flanquantes, c'est-à-dire de part et d'autre, du microsatellite, ce qui implique une phase de séquençage lors de la caractérisation du marqueur ;

- la mise en évidence du polymorphisme est une difficulté majeure, résolue par l'amplification PCR (voir plus loin) de la séquence entourant le microsatellite puis électrophorèse de l'amplifiat sur gel d'acrylamide dont la haute résolution permet de distinguer les allèles dont la taille diffère de deux bases seulement.

Dans toutes les espèces où ils ont été recherchés, les microsatellites ont pu être rapidement identifiés en grand nombre et leurs propriétés (polymorphisme, répartition) ont été confirmées.

## 2.2 / Amplification par PCR

La technique d'amplification par « réaction en chaîne par la polymérase » (ou PCR), décrite il y a quelques années est maintenant utilisée dans de très nombreuses applications de génétique moléculaire. Elle est entre autres indispensable pour mettre en évidence le polymorphisme des microsatellites. Grâce à elle, on peut disposer en quelques heures de millions de copies d'un fragment d'ADN compris entre deux petites séquences connues. Ces deux petites séquences, dites amores, sont fournies comme réactif, avec l'ADN génomique total, la polymérase, des nucléotides et divers facteurs. Dans l'approche classique, les nucléotides fournis pour la réaction sont radioactifs, de sorte que les différents allèles sont distingués par autoradiographie après électrophorèse. Depuis quelques années, la radioactivité est remplacée par l'utilisation d'amores fluorescentes.

**Les progrès importants réalisés en biologie moléculaire ont permis d'établir les cartes génétiques des différentes espèces d'intérêt zootechnique.**

## 2.3 / Stratégie d'établissement des cartes génétiques

L'établissement des cartes génétiques suit la logique suivante :

- constitution d'une collection d'ADN génomique de familles de référence, familles dans lesquelles sont analysées les liaisons entre marqueurs. Les familles généralement choisies comprennent les deux parents et 10 à 30 produits, la structure familiale pouvant être nucléaire ou hiérarchique. Un panel comprend généralement de 150 à 300 individus au total, cet effectif étant un bon compromis entre le travail de typage à réaliser et la précision souhaitée dans l'estimation des distances. Pour favoriser la construction d'une carte unique au niveau international, ce panel est distribué à toutes les équipes impliquées ;

- recherche de microsatellites par criblage de banque d'ADN génomique avec une sonde oligonucléotidique (par exemple (TG)<sub>n</sub>) puis caractérisation par séquençage du motif répété et des séquences flanquantes ;

- typage des individus du panel de référence pour les marqueurs étudiés, centralisation des résultats et analyse de liaison pour ordonner les marqueurs les uns par rapport aux autres et estimer leurs distances génétiques.

## 2.4 / Apports de la carte physique

Des outils de carte physique, sans être indispensables, aident considérablement la construction d'une carte génétique saturée. Ainsi, les lignées cellulaires d'hybrides somatiques, constituées de cellules hôtes d'une espèce différente ayant incorporé un ou quelques chromosomes de l'espèce étudiée, sont un outil simple et rapide de construction des groupes de syténie. Un groupe de syténie

est constitué des marqueurs situé sur un chromosome ou un fragment chromosomal donné et qu'on trouve donc chez les mêmes hybrides cellulaires. D'autre part, la technique d'hybridation *in situ*, employant historiquement la radioactivité et maintenant des sondes fluorescentes (FISH), permet de localiser précisément une séquence sur un chromosome. Elle permet donc d'affecter un groupe de liaison à un chromosome, de l'orienter sur le chromosome, et de vérifier que la carte génétique couvre tout le génome.

## 2.5 / Situation actuelle

Les cartes génétiques sont maintenant très denses chez l'homme et la souris, avec plusieurs milliers de marqueurs microsatellites et de gènes localisés, et elles continuent à progresser très rapidement. Parmi les espèces d'élevage, les cartes bovine et porcine sont les plus avancées (Barendsee *et al* 1997). Fin 1997, la carte bovine comptait 1 200 marqueurs microsatellites et plus de 200 gènes localisés. Tous les groupes de liaison sont assignés à un chromosome et orientés. Un effort important est aussi porté sur le développement des cartes du poulet et des ovins. L'INRA a réalisé la première carte caprine.

Un important travail est en cours pour faciliter l'utilisation en routine des marqueurs. Cette étape consiste à choisir un sous-ensemble de 150 à 250 marqueurs bien répartis sur tout le génome, suffisamment polymorphes, et surtout bien adaptés aux procédures de typage automatisées par séquenceur automatique. La mise au point de ce sous-ensemble est une étape préalable indispensable à l'utilisation à très grande échelle des marqueurs moléculaires.

## 3 / Un exemple d'utilisation des marqueurs : l'analyse du déterminisme génétique d'un caractère complexe

La recherche des gènes régissant un caractère peut se faire selon des méthodes très différentes. L'approche la plus intuitive, qui s'appuie sur l'analyse physiologique des fonctions impliquées, consiste à remonter la chaîne qui va du caractère à la protéine et au gène. Elle consiste donc à vérifier l'effet possible d'un gène « candidat » *a priori* impliqué dans le déterminisme du caractère. Cette approche a connu quelques succès, comme par exemple l'identification de l'effet du gène de la caseïne  $\alpha_{s1}$  sur la composition du lait de chèvre ou la compréhension du mode d'action du gène de nanisme chez la poule. Ces quelques succès ne doivent pas cacher les difficultés de cette méthode pour les caractères complexes, difficultés liées avant tout au nombre très élevé de gènes candidats possibles et au fait que la plupart d'entre eux ne sont pas connus. Dans un premier temps, la recherche à l'aide de marqueurs des régions chromosomiques

impliquées (ou QTL, pour « quantitative trait locus ») est souvent préférable. Bien qu'elle ne permette pas de déterminer la nature du gène responsable ni son mode d'action, elle présente divers avantages décisifs : elle facilite considérablement la recherche ultérieure d'un gène candidat, en comparant sa position sur la carte génétique avec celle du QTL détecté ; même si le gène impliqué reste inconnu, sa localisation par marqueurs permet son utilisation, par exemple en sélection.

### 3.1 / Principe

Le principe découle directement des propriétés des marqueurs. Considérons un reproducteur double hétérozygote MQ/mq pour deux locus liés, le taux de recombinaison entre ces deux locus étant  $r$ . Le premier locus n'est pas observable directement mais il agit sur le caractère d'intérêt, de sorte que la substitution de l'allèle  $q$  par l'allèle  $Q$  chez un descendant entraîne un effet noté  $a$ . Le deuxième locus n'a pas d'effet sur le caractère mais ses deux allèles  $M$  et  $m$  peuvent être observés et distingués l'un de l'autre. Les gamètes produits par ce reproducteur sont de quatre types, deux types parentaux MQ et mq, de fréquence  $(1-r)/2$  et chez lesquels l'haplotype marqueur-QTL est le même que chez le parent, et deux types dits recombinants, Mq et mQ, de fréquence  $r/2$ , et issus d'une recombinaison chez le parent.

Si les locus sont liés, le taux de recombinaison  $r$  est inférieur à 0,5, les gamètes parentaux sont plus fréquents que les gamètes recombinants, et la différence de performance entre les produits de ce reproducteur ayant reçus  $M$  et  $m$  est de  $a(1-2r)$ . Plus le marqueur est proche du QTL, plus l'effet apparent du marqueur est élevé et proche de l'effet vrai du QTL. Au contraire, l'efficacité du marqueur diminue lorsqu'il s'éloigne du QTL, pour devenir nulle lorsque le marqueur et le QTL ne sont pas liés.

Pour qu'un QTL, dont la position est quelconque et inconnue, puisse être détecté, il est nécessaire de disposer de marqueurs nombreux et couvrant tout le génome. La taille des génomes animaux étant proche de 30 Morgans (soit 3 000 cM), un réseau de 150 marqueurs informatifs uniformément répartis tous les 20 cM permet en théorie de couvrir tout le génome et de détecter tous les QTL d'intérêt. Ceci explique la nécessité de constituer une carte génétique précise, avant de rechercher des QTL.

L'utilisation d'une carte précise présente un autre avantage. Lorsque les marqueurs sont analysés un par un, il est possible de détecter un QTL, mais pas de le localiser précisément par rapport au marqueur lié, ni d'en estimer son effet. En effet, avec un seul marqueur, il est difficile de distinguer un QTL proche à petit effet d'un QTL à gros effet mais plus éloigné, les deux situations conférant au marqueur le même effet apparent. L'analyse simultanée de marqueurs adjacents (« interval mapping ») permet d'augmenter un peu la

puissance de détection mais surtout d'estimer à la fois l'effet et la position du QTL (Lander et Botstein 1989). Diverses méthodes statistiques peuvent être utilisées mais le principe général est le suivant. En une position  $x$  supposée du QTL, on calcule, à partir de l'information marqueur des parents et des produits, la probabilité  $p(x)$  que le segment chromosomal reçu du parent par le produit provienne de son grand-père ou de sa grand-mère. L'effet du QTL supposé en  $x$  peut ensuite être estimé par le coefficient de régression linéaire de la performance sur cette probabilité. La présence d'un QTL est testée par le rapport des vraisemblances sous les hypothèses de présence ou d'absence du QTL. Si l'hypothèse d'absence de QTL est rejetée, sa position la plus probable est celle qui maximise le rapport de vraisemblance.

**L'expression d'un caractère quantitatif est gouvernée par des gènes appelés QTL pour Quantitative Trait Loci.**

### 3.2 / Types de protocole (figure 2)

Tous les protocoles reposent sur l'utilisation de reproducteurs doubles hétérozygotes aux marqueurs et au QTL. Dans les protocoles de détection de QTL les plus simples, comme ceux utilisés chez les plantes, le reproducteur est un individu F1 issu du croisement entre deux lignées consanguines, donc totalement homozygotes (Paterson *et al* 1988). La population F1 est totalement homogène et se comporte comme un individu unique. Le déséquilibre de liaison existant en deuxième génération, F2 ou back-cross, est mis à profit pour établir les associations entre marqueurs et QTL. La puissance d'un tel dispositif est maximale, ne dépendant que de l'effectif mesuré, de la part de variance expliquée par le QTL et de la distance entre marqueurs et QTL.

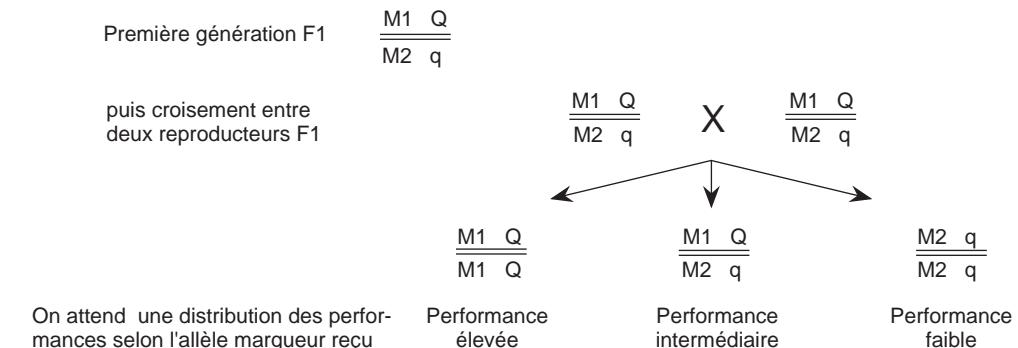
En l'absence de lignées consanguines, très difficiles à obtenir chez les animaux d'élevage, le même principe peut être appliqué en croisant des races différentes, ou des lignées différentes, voire des lignées sélectionnées de façon divergente. Cependant, dans un tel dispositif, les populations parentales ne sont pas totalement homogènes. Les allèles n'étant pas fixés dans ces populations parentales, il subsiste un polymorphisme au QTL, aux marqueurs, ainsi que dans le reste du génome et donc dans la composante polygénique affectant le caractère. La F1 est donc constituée d'un mélange d'individus doubles hétérozygotes au marqueur et au QTL, hétérozygotes au marqueur mais homozygotes au QTL, homozygotes au marqueur et hétérozygotes au QTL, ou doubles homozygotes. L'informativité des marqueurs varie d'un reproducteur F1 à l'autre. Enfin, la composante polygénique varie aussi d'un individu F1 à l'autre. Cette hétérogénéité introduit un bruit de fond qui contribue d'une part à diminuer la puissance de détection du dispositif, d'autre part à rendre beaucoup plus complexes l'élaboration du protocole et l'analyse des résultats. Ce bruit de fond est d'autant plus élevé que les populations parentales sont hétérogènes et proches l'une de l'autre. Le déséquilibre de liaison, dont on cherche à maximiser le niveau par croisement, n'est pas total. La structure familiale de la deuxième

**Figure 2.** Dispositifs de recherche de QTL à l'aide de marqueurs.

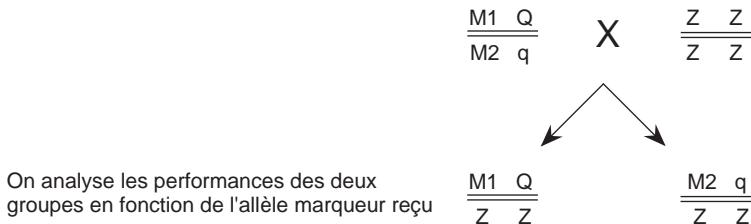
Croisement entre populations (races ou lignées) différant par le marqueur et par la performance

$$\begin{array}{ccc} \text{Performance élevée} & \frac{\text{M1 } Q}{\text{M1 } q} & \times & \frac{\text{M2 } q}{\text{M2 } Q} & \text{Performance faible} \end{array}$$

Si le marqueur M est proche du QTL recherché, marqueur et QTL seront transmis ensemble



Dans une population, on recherche le QTL à l'aide de marqueurs intra-famille



**La recherche de QTL peut se faire à l'aide de marqueurs en utilisant leurs propriétés : polymorphisme et liaison avec le QTL supposé.**

génération, quelconque dans le cas du croisement entre lignées consanguines, doit être optimisée dans le cas du croisement entre races. En effet, un compromis doit être trouvé entre le nombre de familles, suffisamment élevé pour limiter le risque que le dispositif soit peu informatif, et la taille des familles, suffisamment importante pour que l'on distingue l'effet du QTL de l'effet famille. L'informativité variable des marqueurs d'une famille à l'autre nécessite l'analyse simultanée de tous les marqueurs d'un même groupe de liaison, de sorte que la position testée pour le QTL soit le plus souvent possible encadrée par deux marqueurs informatifs.

Les QTL peuvent aussi être recherchés intra population. Dans ce cas, l'hypothèse de base est l'équilibre de liaison entre marqueurs et QTL dans l'ensemble de la population. Le dispositif tire donc profit du déséquilibre de liaison, qui existe nécessairement intra famille. L'effet apparent du marqueur n'est défini qu'au sein d'une famille et il est susceptible de varier d'une famille à l'autre. Les résultats sont combinés entre familles en déterminant la part de variance expliquée par une région chromosomique donnée. Ce type de protocole, souvent bien adapté aux contraintes des populations animales, est cependant moins puissant que les protocoles en croisement. D'une part, le nombre de parents homozygotes au QTL, alourdisant inutilement le protocole, est plus élevé qu'en croisement, et il est généralement incontrôlable. De même, le nombre de marqueurs à considérer est plus élevé, du fait qu'un marqueur n'est pas à l'état hétérozygote chez tous les reproduc-

teurs, et ne permet donc pas de rechercher des QTL dans ces familles. D'autre part, un protocole intra population demande des familles plus grandes qu'un protocole de croisement. Comme il est souvent difficile d'obtenir de grandes familles de pleins-frères soeurs, un seul parent, le père, est complètement informatif, et les différences recherchées entre génotypes sont alors deux fois plus faibles que dans un protocole de croisement.

Deux protocoles intra population sont classiquement proposés et reposent sur l'analyse de 2 ou 3 générations successives. Le protocole sur deux générations (Soller et Genizi 1978) consiste à analyser des familles de produits issus de pères hétérozygotes aux marqueurs et à un QTL supposé, et à comparer intra-père les moyennes de production des deux groupes de descendants ayant reçu l'un ou l'autre allèle marqueur paternel. L'inconvénient de ce dispositif est le nombre très élevé, en général plusieurs milliers, d'individus à élever, génotyper et dont il faut mesurer les performances, pour assurer une bonne puissance de détection.

Pour améliorer cette puissance de détection à nombre de typages fixé, le protocole peut être étendu sur trois générations (Weller *et al* 1990). Le père, supposé double hétérozygote aux marqueurs et au QTL, a des fils dont la valeur génétique est estimée sur descendance à partir des performances de leurs produits. Le QTL supposé est détecté par la recherche d'une association entre la valeur génétique estimée des fils et l'allèle marqueur qu'ils ont reçu de leur père. Dans ce protocole, le caractère est mesuré sur les descendants de troi-

sième génération, mais seuls le père et ses fils sont génotypés aux marqueurs. L'intérêt de ce protocole est de réduire la variabilité résiduelle du caractère, qui s'apparente à une moyenne de performances. En conséquence, la puissance de détection est nettement améliorée. En pratique, le nombre de performances à mesurer est très élevé, de sorte que ce protocole n'est réellement envisageable que si les données existent déjà avec une structure de population adaptée, ce qui est le cas chez les bovins laitiers (Georges *et al* 1995). Dans cette espèce, le testage sur descendance est une généralité et, au moins dans les plus grandes races, il concerne fréquemment des familles de plusieurs dizaines de taureaux demi-frères de pères. Le coût marginal d'un tel programme est alors réduit au développement des marqueurs et au typage des taureaux d'insémination artificielle, dont l'ADN est par ailleurs facilement disponible.

### 3.3 / Méthodes alternatives

L'analyse de liaison classique demande des effectifs importants, entraînant à la fois un coût de typage élevé et un protocole lourd, en particulier lorsque le caractère observé nuit à la santé ou à la bonne reproduction des animaux. Lorsque le caractère est soumis à un déterminisme monogénique et que le phénotype permet de déduire le génotype sans équivoque, une méthode alternative dite « homozygosity mapping » a été proposée, particulièrement bien adaptée aux maladies récessives. Son principe est le suivant. Considérons un ensemble d'individus consanguins, dont l'ancêtre commun au père et à la mère est hétérozygote pour le gène recherché. Si l'anomalie est rare, les descendants présentant l'anomalie, et donc homozygotes, ont reçu vraisemblablement deux copies du même gène porté par l'ancêtre commun. Chez ces descendants, les marqueurs proches du gène recherché sont donc également à l'état homozygote, sous l'hypothèse d'absence de recombinaison. Si un marqueur non lié au gène recherché est très polymorphe, la probabilité qu'il soit par hasard à l'état homozygote chez les individus affectés décroît très vite quand le nombre d'individus considérés et le nombre de maillons dans la chaîne de parenté entre les individus et leur ancêtre augmentent. L'efficacité de la méthode est augmentée en considérant des haplotypes, c'est-à-dire des ensembles de marqueurs adjacents. En pratique, à condition qu'une carte génétique précise et assez dense soit disponible, quelques individus affectés suffisent pour localiser le gène recherché.

Cette méthode a ensuite été étendue à des situations où les hypothèses sont moins restrictives. On recherche chez des individus d'un phénotype donné une concentration plus forte qu'espérée d'un haplotype marqueur « identique par descendance », c'est-à-dire dérivant par transmission mendéienne d'un haplotype ancêtre unique (Houwen *et al* 1994). L'éloignement de l'ancêtre est un paramètre important, qui définit la taille du dispositif. Si l'ancêtre est proche, des concentrations fortuites d'ha-

plotypes sont possibles, demandant alors des effectifs d'individus affectés un peu plus élevés. Cependant, le nombre total d'individus produits et mesurés reste limité. Par contre, les recombinaisons étant peu nombreuses, le gène recherché ne peut pas être localisé précisément. Au contraire, si l'ancêtre commun est éloigné, le nombre d'individus affectés est faible, ce qui nécessite un dispositif de grande taille quant à l'effectif produit et mesuré. Mais la probabilité de concentration fortuite devenant négligeable, le nombre d'individus affectés nécessaire pour conclure est très réduit. La probabilité de recombinaison entre marqueurs et gène recherché étant plus élevée, brisant ainsi l'association définissant l'haplotype, le segment conservé au cours des générations est plus court et demande donc, pour être mis en évidence, une carte de marqueurs plus dense. Mais, le nombre de recombinaisons accumulées étant plus important, la localisation du gène est plus fine. Ces techniques, dites d'identité par descendance (*Identity by descend=IBD*), sont encore très récentes et demandent encore des développements, mais elles sont très prometteuses.

## 4 / Revue des différentes utilisations possibles des marqueurs génétiques

Les propriétés des marqueurs, polymorphisme et liaison, peuvent être mises à profit pour la gestion génétique des populations. Le **polymorphisme** est utile pour établir l'origine génétique du locus concerné : au niveau individuel, puisque pour un individu hétérozygote en un locus il est en général possible de connaître l'origine grand-parentale du segment génomique qu'il transmet à ses descendants ; au niveau des populations, sous l'hypothèse que les formes possibles pour un même segment génomique diffèrent d'autant plus que les populations sont phylogénétiquement plus éloignées. La **liaison génétique** est utile pour généraliser les conclusions des observations sur un locus particulier à l'ensemble de l'ADN proche de ce locus : si, lors de la méiose, il n'y a de recombinaison entre le segment génomique visualisé et les zones avoisinantes, ce segment génomique devient un **marqueur** génétique de ces zones et notamment des gènes qu'elles contiennent. Dans certaines applications, les marqueurs sont utilisés pour suivre l'ensemble du génome, dans d'autres applications, ils nous renseignent sur la transmission de gènes qui leur sont liés.

### 4.1 / Marquage de l'ensemble du génome

L'idée de base est que l'ensemble du génome peut être approché par l'information apportée par un ensemble de marqueurs bien répartis. Ceci peut être mis à profit pour l'identification d'individus ou de groupes, la gestion de la

**La localisation de marqueurs sur l'ensemble du génome permet de connaître plus précisément les différentes populations d'une espèce, contribuant ainsi à améliorer leur gestion.**

diversité génétique entre ou intra populations ou l'organisation d'un plan d'introgression (c'est-à-dire l'introduction par croisements d'un gène précis dans une population).

Le **contrôle de filiation** est une des applications les plus anciennes du marquage génétique. Son principe est la recherche d'incohérences entre les génotypes en différents locus marqueurs d'un individu et de ses parents. Par exemple, un descendant de génotype M1M1 au locus *M* et de mère certaine M1M1 ne saurait avoir un individu de génotype M2M2 pour père. Le contrôle de filiation procède donc par exclusion, celle-ci étant d'autant plus probable que le nombre de locus analysés est important et qu'ils sont plus polymorphes.

En France, le contrôle de filiation est une obligation légale (taureaux d'insémination, embryons) à la charge d'une structure unique (Labogena) qui, en 1994, a vérifié la filiation de 17 000 bovins, 2 000 ovins, 500 caprins et 25 000 équins. Les outils utilisés sont pour l'essentiel des antigènes des globules rouges. Parallèlement à ces techniques immunologiques, des techniques biochimiques permettent de détecter le polymorphisme de protéines sanguines par électrophorèse. Ainsi, pour les bovins, 11 marqueurs sont étudiés qui contiennent de 2 à 800 allèles chacun, conférant une efficacité théorique (ou probabilité *a priori* d'exclusion) variant selon les races, de l'ordre de 0,99 pour la race Charolaise à 0,96 pour la race Holstein. Des efforts portent aujourd'hui sur l'utilisation de marqueurs microsatellites pour réaliser ces contrôles. Ces marqueurs ont l'avantage de leur efficacité due à leur fort polymorphisme, de leur abondance et de leur simplicité d'emploi, puisqu'il n'est plus nécessaire de produire de réactifs immunologiques.

Pour des situations extrêmes dans lesquelles les accouplements ne peuvent pas être contrôlés, c'est l'identification plutôt que le contrôle des paternités qui est recherchée, cette identification se faisant *a posteriori*, à l'aide de marqueurs génétiques. Ce cas est fréquent chez les arbres, les poissons ou les animaux sauvages pour lesquels l'utilisation des marqueurs très polymorphes apparaît comme une solution efficace. Son extension aux mammifères d'élevage pourrait parfois être envisagée.

L'**identification des lignées** en vue d'une protection des droits de leur créateur est une préoccupation importante des sélectionneurs de plantes. Ici encore, les marqueurs de l'ADN sont des candidats de choix pour établir un profil génétique spécifique de chaque lignée. Si l'on se réfère aux réflexions relatives à la brevetabilité du vivant, ce type d'application pourrait se développer dans le futur pour les animaux domestiques.

Les généticiens essaient depuis longtemps de situer génétiquement les différentes sous-populations d'une même espèce (races, lignées, rameaux) les unes par rapport aux autres. Au-delà de son intérêt historique, la connaissance des filiations entre ces sous-populations est nécessaire à l'organisation des

programmes de **conservation de la diversité génétique**. Ces programmes ont pour ambition d'éviter la perte de gènes qui pourraient se révéler utiles dans l'avenir si les conditions économiques changent ou après un épuisement de la variabilité génétique immédiatement disponible. Dans le choix des sous-populations conservées, les redondances doivent être évitées afin d'optimiser l'utilisation de ressources financières limitées.

La distance génétique entre deux populations peut être basée sur les informations apportées par les fréquences alléliques en différents locus. Deux populations sont considérées comme d'autant plus proches que les différences de fréquence d'allèles en différents locus sont faibles, traduisant ainsi l'hypothèse opposée qui veut que, par dérive, sélection et/ou mutation, des populations éloignées possèdent des distributions d'allèles très différentes. L'hypothèse est aussi faite que les marqueurs observés donnent une bonne image de l'ensemble du génome, ce qui suppose qu'ils soient en nombre suffisant et bien répartis. L'utilisation des distances entre populations pose toutefois des problèmes théoriques encore en discussion. La notion de distance ne prend réellement son sens que dans le cas d'une structure arborescente dans laquelle existent de réelles filiations entre les races. Or, pour les animaux d'élevage, il est vraisemblable que les croisements opérés depuis la domestication se traduisent par une structure des relations entre races en réseau plutôt qu'en arbre. Par ailleurs, dans la perspective de recherche de redondances pour les aptitudes d'élevage, il faut s'interroger sur l'intérêt de l'utilisation de marqueurs neutres (n'agissant pas sur les caractères sélectionnés) tels que les microsatellites.

Enfin, à l'image des essais réalisés dans le domaine végétal, les distances génétiques pourraient être un élément de décision dans le choix de lignées devant être croisées pour **exploiter les effets d'hétérosis** (supériorité de l'individu croisé par rapport à la moyenne de ses parents). En effet, l'hétérosis est souvent expliqué par l'hypothèse d'une dominance orientée dans laquelle une majorité de gènes contrôlant un même caractère sont soumis à un effet de dominance (pour un gène *G*, on parle de dominance si les individus G1G2 ont une valeur supérieure à la moyenne des homozygotes G1G1 et G2G2). Sous cette hypothèse, l'hétérosis est donc d'autant plus fort que les lignées soumises au croisement sont génétiquement différentes (et homozygotes), donc que les produits croisés sont hétérozygotes en un plus grand nombre de gènes.

Les marqueurs peuvent être aussi utilisés pour **maintenir la variabilité génétique** des petites populations en conservation. Le principe est de sélectionner les reproducteurs sur leur taux d'hétérozygotie estimé par la proportion de locus hétérozygotes parmi un nombre donné de locus marqueurs. L'efficacité relative de cette technique est d'autant meilleure que la population est de faible effectif.

Les marqueurs sont également un outil très intéressant pour l'**introgression**. Parmi les gènes contrôlant un caractère possédant un intérêt zootechnique, certains ont un effet majeur sur le caractère et peuvent être mis en évidence par la seule observation des phénotypes. Dans la plupart des cas, l'allèle favorable de ces gènes n'est présent que dans certaines races (dites donneuses, *D*), alors que d'autres races (receveuses, *R*), par ailleurs très améliorées ou adaptées à leur milieu, en sont dépourvues. Pour bénéficier de l'effet favorable du gène majeur sans perdre les propriétés générales de la race receiveuse, il faut donc introduire (introgresser) l'allèle favorable par un croisement simple  $F_1 = R \times D$ , un premier croisement en retour ( $BC_1 = F_1 \times R$ ) vers la race receiveuse suivi d'une succession de croisements en retour ( $BC_i = BC_{i-1} \times R$ ), étant entendu qu'à chaque génération  $BC_i$ , seuls les reproducteurs porteurs de l'allèle favorable sont retenus. A l'issue de cette série de croisements, des animaux *R* (ou tout au moins proches de *R*), porteurs de l'allèle favorable, sont donc disponibles. A titre d'exemple, le gène Booroola a été introgressé en Mérinos d'Arles. Ce long processus peut être accéléré en choisissant les reproducteurs  $BC_i$  des différentes générations de croisement en retour sur leurs génotypes en des locus marqueurs, en retenant les porteurs d'un nombre maximum d'allèles marqueurs spécifiques de la race *R*. Comme pour la gestion de la variabilité intra population, les locus marqueurs, s'ils sont nombreux et bien répartis, donnent une image de l'ensemble du génome. L'hypothèse de travail est que le polymorphisme est suffisant en ces locus pour être en mesure de repérer leur origine raciale, et que la sélection sur ces marqueurs, du fait des liaisons génétiques, entraîne une sélection générale du génome receveur. En considérant 2 ou 3 marqueurs bien répartis par chromosome, cette technique fait gagner 2 générations de croisements en retour par rapport à la méthode traditionnelle pour obtenir une même proportion de génome *R* (Hospital *et al* 1992).

## 4.2 / Marquage de zones spécifiques du génome

Les marqueurs, qui sont nécessaires pour repérer les **QTL**, peuvent aussi aider à fixer les allèles favorables dans les populations. En plus du polymorphisme et de la liaison, la troisième propriété des marqueurs, le déséquilibre de liaison présenté plus haut, est aussi nécessaire. En se limitant pour simplifier au cas où il n'existe que deux allèles, deux segments génomiques (les locus *M* et *Q*) sont en équilibre de liaison dans un groupe d'animaux si la proportion de gamètes porteurs de l'allèle *M*1 est la même quel que soit l'allèle au locus *Q*. Dans le cas d'un déséquilibre de liaison entre *M* et *Q*, l'allèle *M*1 sera plus souvent associé à l'allèle *Q*1 (par exemple) et l'allèle *M*2 à *Q*2. Dès lors, une sélection de reproducteurs porteurs par exemple du génotype *M*1*M*1 augmentera de façon indirecte la fréquence de l'allèle *Q*1 dans la génération de

leurs descendants. Si *M* est un locus marqueur et *Q* un QTL, on dispose ainsi d'une méthode indirecte de sélection des gènes favorables en ce dernier locus.

Le déséquilibre de liaison global au niveau d'une population ne s'observe que dans un certain nombre de cas, mais il est toujours présent dans la descendance directe de reproducteurs doubles hétérozygotes (en *M* et *Q*). Il peut donc être utilisé pour la gestion d'un gène majeur ou plus généralement pour la sélection d'un ensemble de QTL affectant un ou des caractères zootechniques. On parle alors de sélection assistée par marqueurs.

Pour la gestion d'un gène à effet majeur, la situation diffère selon que l'allèle favorable est présent ou non dans la population. Dans le premier cas, la fréquence de l'allèle favorable augmente de façon automatique si 1) le caractère contrôlé par le gène majeur est un des objectifs de sélection de cette population, 2) l'allèle favorable n'a pas d'effet adverse sur d'autres caractères, et 3) il n'y a pas de surdominance (c'est-à-dire, dans le cas biallétique, si les individus *Q*1*Q*2 ne sont pas meilleurs que les *Q*1*Q*1 et les *Q*2*Q*2). Cette évolution automatique favorable des fréquences alléliques est toutefois d'autant plus lente que l'allèle favorable est initialement rare et son effet faible. Dans un certain nombre de cas, les marqueurs peuvent donc contribuer à accélérer la fixation de l'allèle favorable.

Quand l'allèle favorable est absent de la population, il faut l'introgresser depuis une race donneuse. Mais, en l'absence de sélection, la proportion de reproducteurs porteurs de l'allèle favorable diminue ensuite de moitié à chaque génération. Il faut donc sélectionner, parmi l'ensemble des individus nés de chaque génération, les porteurs de cet allèle. Si le génotype au locus majeur est déterminé dans le jeune âge, dans les deux sexes et sans erreur par la simple observation du phénotype, les marqueurs n'ont pas d'utilité. Dans les autres cas, ils peuvent aider à l'introgression. Ainsi, le gène Booroola (*F*), qui a un effet considérable sur la taille de portée de la brebis, n'est visible que chez les femelles, et après leur puberté. Au cours de l'introgression de ce gène dans les races Mérinos d'Arles, il a donc été nécessaire de tester sur descendance les jeunes bœufs issus de chaque génération de croisement en retour et de génotype inconnu (++) ou (*F*+) en les accouplant à des brebis des races receiveuses (par hypothèse toutes ++) : les génotypes des filles, toutes ++ si le père est ++, ou à 50 % (*F*+, 50 %++) si le père est (*F*+, sont déterminés par observation des niveaux d'ovulation. La découverte de marqueurs de ce gène évite maintenant ce testage sur descendance, long et coûteux.

L'utilisation d'un **déséquilibre de liaison au niveau de la population** n'est possible que dans un nombre limité de cas. L'exemple du gène de sensibilité à l'halothane chez le porc est de ce point de vue caractéristique, avec un très fort déséquilibre entre ce gène et un marqueur sanguin Phi en race Landrace,

chez laquelle l'allèle défavorable a été éliminé rapidement par sélection sur ce marqueur, alors que dans d'autres lignées cette sélection n'a pas été possible du fait du relatif équilibre de liaison entre ces locus. Un déséquilibre est d'autant plus vraisemblable que le gène majeur et son marqueur sont proches, et que la population est génétiquement plus ouverte aux apports de sang extérieur.

Cette démarche est aussi généralisable à un ensemble de QTL marqués. Un déséquilibre de liaison global existe notamment à la suite de la création d'une population par hybridation de deux lignées. On peut alors, au moins partiellement, sélectionner les animaux de cette population hybride d'après le nombre d'allèles marqueurs « favorables » (en réalité associés à des allèles quantitatifs favorables) qu'ils portent. On trie les animaux sur leur score moléculaire dont les coefficients sont ceux de la régression multiple des phénotypes d'une première génération sur le nombre d'allèles favorables en chaque locus marqueur (Lande et Thompson 1990). L'intérêt de cette forme de sélection assistée par marqueurs est maximale quand l'héritabilité du caractère est faible et quand les QTL expliquent une grande part de la variabilité génétique. Le progrès génétique pourrait être multiplié par deux ou trois dans les situations les plus optimistes. A terme, l'efficacité diminue, du fait de la disparition progressive, au cours des générations, du déséquilibre de liaison du fait des recombinaisons.

Dans le cas où les locus sont en équilibre de liaison au niveau de la population, il est toujours possible d'utiliser le déséquilibre de liaison intra famille. En effet, aux événements de recombinaison près, un reproducteur M1Q1/M2Q2 transmet systématiquement Q1 avec M1 et Q2 avec M2. Un tri intra famille sur le génotype marqueur est donc efficace. A titre d'exemple, supposons qu'un QTL soit détecté dans une population bovine. A partir des données du testage sur descendance d'un taureau hétérozygote M1M2 au marqueur de ce QTL, le génotype global (la phase) de ce taureau peut être établi (par exemple M1Q1 sur le premier chromosome et M2Q2 sur le second), ce qui permet de ne mettre en testage sur descendance que ses fils porteurs de l'allèle marqueur associé à l'allèle quantitatif favorable. L'efficacité du procédé est réduite par les recombinaisons entre les locus M et Q, par un polymorphisme insuffisant au marqueur ne permettant pas de connaître la transmission de l'allèle marqueur et par l'imprécision de la détermination du génotype en Q avec la seule observation des phénotypes quantitatifs des descendants.

Pour des marqueurs très proches du QTL, la démarche est généralisable aux générations suivantes : ayant établi les phases pour chaque reproducteur d'une génération ancestrale, il est théoriquement envisageable, grâce au suivi continu de la transmission des marqueurs et à l'enregistrement des performances, de suivre l'évolution des phases chez

les reproducteurs des différentes générations. La méthode « BLUP modèle animal » d'évaluation des reproducteurs a été ainsi étendue pour inclure cette information sur la transmission des allèles marqueurs (Fernando et Grossman 1989, Goddard 1992). L'idée de base est que les informations sur les génotypes marqueurs permettent d'affiner les coefficients de la matrice de corrélations entre valeurs génétiques des apparentés : à proximité du locus du marqueur M, deux frères se ressembleront plus qu'en moyenne s'ils ont reçu le même allèle de leur père, moins s'ils ont reçu des allèles différents.

L'utilisation des marqueurs pour cette gestion globale sera sans doute une application beaucoup plus fréquente pour les animaux domestiques que l'introgression assistée par marqueurs d'un gène unique. L'efficacité de la sélection assistée par marqueur, comparée à la situation standard où les animaux sont choisis à l'aide du BLUP modèle animal, est maximale pour les caractères d'héritabilité faible, et/ou les caractères qui ne s'expriment que chez une partie des individus, les femelles, par exemple. L'intérêt des marqueurs se résume principalement par l'apport d'information plus précoce pour estimer le potentiel génétique de l'individu, ce qui peut se traduire par une augmentation de la pression de sélection, une diminution de l'intervalle de génération, une augmentation de la précision de l'évaluation génétique, ou une combinaison de ces trois effets favorables au progrès génétique.

## 5 / Du marqueur au gène

### 5.1 / Intérêt

L'identification des régions chromosomiques impliquées dans le déterminisme du caractère permet de mettre en place une sélection ou une introgression assistées par marqueurs, sans connaître la nature des gènes impliqués. De ce point de vue, l'approche marqueurs n'est qu'un raffinement de la génétique quantitative classique, raffinement qui consiste à distinguer dans la valeur globale du génome une composante caractéristique d'un ou plusieurs segments chromosomiques et une composante polygénique due aux effets de gènes non individualisés.

Toutefois, lorsqu'un segment chromosomal d'intérêt est identifié, la caractérisation du gène impliqué et du polymorphisme à l'origine de la variabilité génétique du caractère, présente de grands avantages. D'une part, suivre le gène impliqué est toujours plus précis que suivre des marqueurs car on s'affranchit définitivement du risque de recombinaison entre marqueur et gène. L'utilisation du test génétique est plus simple puisqu'elle n'est plus dépendante d'un déséquilibre de liaison entre marqueur et QTL. La caractérisation des différents allèles et de leurs effets devient possible. L'extension du test à d'autres populations est immédiate, tandis que la recherche

d'un polymorphisme similaire dans d'autres espèces est grandement facilitée.

D'autre part, l'identification de la mutation causale permet de comprendre le déterminisme génétique du caractère, par l'étude des conséquences sur la cascade métabolique de la modification d'un maillon particulier. Elle ouvre la voie à de nouvelles possibilités d'action, comme la thérapie génique en médecine humaine ou la transgénèse en élevage.

## 5.2 / Identification des gènes impliqués

Lorsqu'un gène majeur est localisé, la première étape consiste d'abord à s'en rapprocher, en trouvant d'autres marqueurs plus proches. Cependant, en dessous de 1 à 2 cM entre le marqueur le plus proche et le gène majeur, cette approche perd de son efficacité. En effet, le nombre de recombinaisons devient très faible, entraînant une forte augmentation de la taille du dispositif nécessaire et une plus grande sensibilité aux différentes sources d'erreur, en particulier les erreurs de génotypage du marqueur ou du gène majeur.

Se rapprocher d'un QTL est souvent plus difficile. En effet, alors que le niveau de performance permet généralement de déterminer sans erreur le génotype au gène majeur, c'est beaucoup plus délicat pour un QTL, dont l'effet sur la performance est plus limité. Plus l'effet du QTL est faible, plus l'intervalle de confiance de sa position est large, souvent supérieur à 10 cM, et ceci indépendamment de la densité des marqueurs utilisés.

Même quand la distance génétique entre le marqueur et le QTL est proche de 0, la distance physique peut être encore très importante, de l'ordre du million de bases. L'approche génétique laisse alors la place à l'approche moléculaire. Les banques de grands fragments, récemment développées dans différentes espèces, sont un outil de choix pour « marcher » sur le chromosome. Il s'agit de fragments d'ADN de plusieurs centaines de kilobases insérés dans une bactérie (BAC) ou une levure (YAC) qui les répliquent comme s'il s'agissait de leurs propres chromosomes. Ces banques permettent de constituer des contigs, c'est-à-dire des séries de grands fragments chevauchants qui constituent un ensemble continu suffisamment long pour contenir le gène mais beaucoup plus court que l'ADN génomique total.

Parmi les techniques utilisables pour accéder au gène recherché (clonage positionnel), deux peuvent être mentionnées succinctement. La première, dite « exon trapping » permet d'identifier et d'exciser spécifiquement les exons, c'est-à-dire les séquences codantes de l'ADN. La deuxième approche, probablement plus facile, consiste à rechercher, parmi les ARN messagers produits, les transcrits des gènes portés par ces grands fragments, le criblage pouvant s'effectuer par hybridation des grands fragments. La recherche est grandement facilitée en sélectionnant seulement les ARN spécifiques du tissu exprimant le gène.

Il est important de mentionner que ces approches, si elles sont possibles, restent encore très difficiles et assez aléatoires pour un gène à effet majeur. Beaucoup de spécialistes considèrent qu'elles sont pour l'instant irréalistes pour des QTL, compte tenu de l'incertitude sur leur position.

## 5.3 / Cartographie comparée

Le clonage positionnel pouvant être long et délicat, l'approche « gène candidat » reprend tout son intérêt grâce aux progrès de la cartographie comparée. Le principe est d'utiliser les connaissances sur les cartes génétiques humaine et murine qui sont beaucoup plus développées que les cartes des espèces d'élevage. Pour cela, il convient d'établir les correspondances de segments chromosomiques entre les différentes espèces. Ces segments sont souvent de grande taille, le nombre de remaniements chromosomiques pour passer d'une espèce à l'autre étant limité à quelques dizaines. Cette carte comparée entre espèces a d'abord été construite à partir d'environ 300 locus d'ancre. Ces locus sont des gènes codants mieux conservés entre espèces que les marqueurs anonymes non codants et ils ont été choisis pour leur bonne couverture du génome. Plus récemment, la carte comparée a été construite ou complétée par peinture chromosomique. Cette élégante méthode repose sur le tri de chromosomes d'une espèce par cytométrie de flux et l'amplification par PCR non spécifique de l'ADN de chacun de ces chromosomes. L'hybridation de ces produits d'amplification sur un caryotype d'une seconde espèce révèle tous les segments chromosomiques correspondant au chromosome analysé de la première espèce. Cette technique a récemment permis d'établir la carte physique comparée de l'homme, du bovin et du porc (Goureau *et al* 1996).

Lorsqu'une zone chromosomique d'intérêt est détectée avec des marqueurs, la zone correspondante chez l'homme ou chez la souris est donc connue. L'énorme connaissance accumulée sur les cartes humaine ou murine peut alors être mise à profit pour identifier un petit nombre de gènes candidats localisés dans cette zone, ainsi que pour exclure d'éventuels gènes candidats *a priori* intéressants mais cartographiés ailleurs. L'efficacité de cette méthode peut être illustrée par l'identification récente du gène culard chez le bovin, à partir d'une mutation similaire observée chez la souris (Grobet *et al* 1997).

**La connaissance de marqueurs d'un QTL peut être utilisée directement en sélection, mais elle peut aussi permettre d'identifier précisément le QTL et son polymorphisme à l'origine de la variabilité du caractère.**

## Conclusion

Dans un certain nombre de cas, les marqueurs génétiques apparaissent dès aujourd'hui comme un outil très utile à la gestion génétique des populations. Ils sont la seule solution pour des applications telles que le contrôle des paternités ou le calcul des distances entre populations. Ils enrichissent nos moyens d'actions pour d'autres volets, comme la sélection classique ou la valorisation des

gènes à effet majeur. L'efficacité et plus encore la rentabilité de ces opérations doivent être confirmées. Dès maintenant, il est clair que les marqueurs génétiques, s'ils peuvent largement contribuer à une meilleure gestion des populations, ne sauraient être considérés comme une alternative aux méthodes actuelles basées sur les informations phénotypiques et généalogiques. Les deux approches sont complémentaires et les marqueurs apportent une troisième source d'information. L'utilité des marqueurs génétiques apparaîtra surtout dans les situations où les données classiques sont peu informatives vis-à-vis des valeurs génétiques, ou très coûteuses à obtenir : caractère d'héritabilité faible, ou exprimé tardivement ou dans un seul sexe, ou dont la mesure demande l'abattage des animaux... La sélection assistée par marqueurs sera d'autant plus efficace que ces marqueurs expliqueront une part plus importante de la variabilité génétique.

En amélioration génétique, la qualité des marqueurs est très directement liée à l'inten-

sité de leur liaison avec les zones particulières (QTL, gènes majeurs) qu'ils marquent. Dans la plupart des cas, la pleine efficacité ne sera atteinte que lorsque ces gènes à effets quantitatifs seront isolés et leurs formes alléliques, responsables des variations de performances, identifiées. La construction de génotypes, ou réunion, chez un même individu, d'un certain nombre d'allèles favorables en différents QTL contrôlant plusieurs caractères peut être un objectif à terme. Cette technique, qui est considérée comme réaliste pour l'amélioration des plantes, demande que les gènes intéressants soient assemblés par des introgressions successives. Ceci n'est pas encore envisageable dans le cas des animaux, pour lesquels les coûts d'élevage sont élevés et les intervalles de génération longs, et ne le sera peut-être que grâce à une artificialisation très importante des méthodes de reproduction (Georges et Massey 1991) pour disposer d'une pression de sélection considérablement accrue.

## Références bibliographiques

- Barendse W., Vaiman D., Kemp S., *et al.*, 1997. A medium density genetic linkage map of the bovine genome. *Mamm. Genome*, 8, 21-28.
- Fernando R.L., Grossman M., 1989. Marker-assisted selection using best linear unbiased prediction. *Genet. Sel. Evol.*, 21, 467-477.
- Georges M., Massey J.M., 1991. Velogenetics, or the synergistic use of marker-assisted selection and germ-line manipulation. *Theriogenology*, 35, 151-159.
- Georges M., Nielsen D., Mackinnon M. *et al.*, 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics*, 139, 907-920.
- Goddard M.E., 1992. A mixed model for analyses of data on multiple genetic markers. *Theor. Appl. Genet.*, 83, 878-886.
- Goureau A., Yerle M., Schmitz A., Riquet J., Milan D., Pinton P., Grelat G., Gellin J., 1996. Human and porcine correspondence of chromosome segments using bidirectional chromosome painting. *Genomics*, 36, 252-262.
- Grobet L., Royo Martin L.J., Poncelet *et al.*, 1997. A deletion in the myostatin gene causes double-muscling in cattle. *Nature Genetics*, 17, 71-74.
- Hospital F., Chevalet C., Mulsant P., 1992. Using markers in gene introgression breeding programs. *Genetics*, 132, 1199-1210.
- Houwen R.H.J., Baharloo S., Blankenship K., Raeymakers P., Juyn J., Sandkuijl L.A., Freimer N.B., 1995. Genome screening by searching for shared segments : mapping a gene for benign recurrent intrahepatic cholestasis. *Nature Genetics*, 8, 380-386.
- Lande R., Thompson R., 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, 124, 743-756.
- Lander E.S., Botstein D., 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 121, 185-199.
- Soller M., Genizi A., 1978. The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. *Biometrics*, 34, 47-55.
- Weller J., Kashi Y., Soller M., 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J. Dairy Sci.*, 73, 2525-2537.

## Abstract

### **Use of genetic markers in animal breeding.**

Following some recent major advances in molecular genetics, particularly the discovery of microsatellite sequences in all animal genomes and of the DNA amplification with the PCR reaction, many genetic markers are now available in most animal domestic species and are located in genetic maps. The properties of genetic markers are presented. Out of them, polymorphism and linkage make it possible to trace chromosomal segments across generations. This paper describes the potential applications of genetic markers for population management and analysis of genetic variability. Emphasis is put on detection of loci

(quantitative trait loci = QTL) involved in the genetic determinism of economically important traits. QTL detection is a first step toward marker-assisted selection or marker-assisted introgression. This paper also briefly presents some methods, often based on comparative mapping, to identify the causal genes detected through genetic markers. This final step paves the road toward a better comprehension of the mechanisms involved and to simpler, more general, and more efficient selection procedures.

Boichard D., Le Roy P., Levéziel H., Elsen J.-M., 1998. Utilisation des marqueurs moléculaires en génétique animale. INRA Prod. Anim., 11, 67-80.