

VI. L'APPORT DES MARQUEURS MOLECULAIRES

Comme nous l'avons vu précédemment, l'étude des caractères quantitatifs est fondée sur l'analyse statistique des performances mesurées des individus. Il en est de même pour les méthodes de sélection dans les différentes espèces domestiques, végétales ou animales (voir les enseignements correspondants). Cependant, l'accès au génome des organismes permettrait sans doute une meilleure appréhension de l'hérédité des caractères quantitatifs et pourrait contribuer à améliorer l'efficacité des programmes de sélection. Dans ce chapitre, nous présentons succinctement les apports des marqueurs moléculaires dans ce domaine.

A. Définitions

La construction de *cartes génétiques* permet de localiser des locus impliqués dans la variation de caractères quantitatifs (*QTL = Quantitative Trait Locus*) liés à des locus plus facilement identifiables, *les marqueurs*. Les premiers marqueurs utilisés ont été des locus affectant le *polymorphisme visible* (sous-entendu visible à l'œil nu) : gènes de coloration, de nanisme, etc. Depuis le milieu des années 60, une nouvelle famille de marqueurs a été fournie par l'analyse du *polymorphisme biochimique* : électrophorèse des protéines, immunologie. Ces deux types de polymorphisme ont pu être exploités à des fins d'amélioration génétique (voir les enseignements correspondants). Cependant, leur utilisation est limitée : d'une part, beaucoup de gènes à effets visibles ont des effets pléiotropes défavorables et, d'autre part, le nombre restreint de locus analysés au travers du polymorphisme visible ou du polymorphisme biochimique ne permet pas une couverture complète du génome.

L'essor des techniques de biologie moléculaire depuis le début des années 80 a permis d'étudier le polymorphisme directement au niveau de l'ADN. Deux principaux outils ont une application possible pour l'analyse des caractères quantitatifs : le polymorphisme de longueur des fragments de restriction (Restriction Fragment Length Polymorphism, *RFLP*) et les *microsatellites*. L'emploi de ces outils s'est trouvé grandement facilité par le développement de la technique de *PCR* (Polymerase Chain Reaction) qui permet de dupliquer en très grande quantité la(les) zone(s) de l'ADN que l'on souhaite étudier. L'usage des polymorphismes sur une seule base (*Single Nucleotide Polymorphism*, SNP) tend aujourd'hui à se développer.

Les *enzymes de restriction* « coupent » la molécule d'ADN en des sites de reconnaissance spécifiques, qui correspondent à de courtes séquences palindromiques (un palindrome est un mot qui peut se lire dans les deux sens ; exemple : radar). S'il y a un changement de base dans la séquence, il n'y a pas de reconnaissance, et donc pas de scission de l'ADN. La « digestion » d'une même région de l'ADN, par une ou plusieurs enzymes de restriction, fournit des brins d'ADN de taille variable selon les individus, c'est-à-dire selon que les sites de reconnaissance correspondants sont altérés ou pas (d'où le terme RFLP). Chaque site de reconnaissance se comporte comme un locus, l'absence ou la reconnaissance d'un site correspondant à deux allèles possibles.

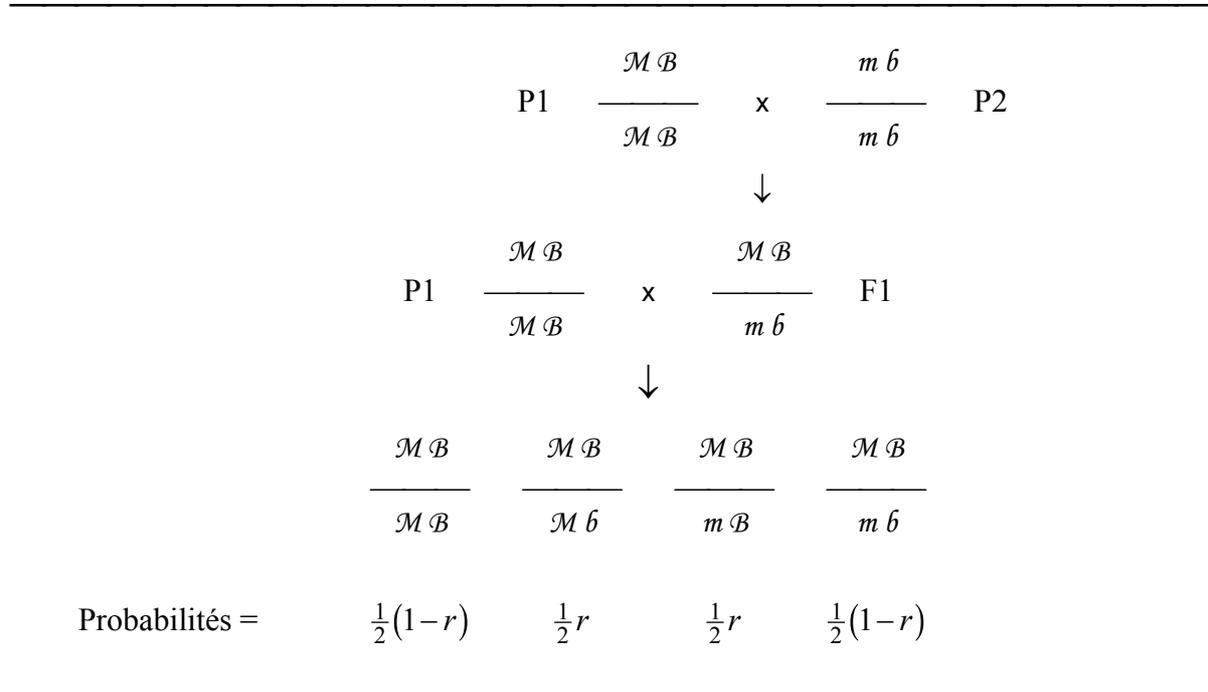
Les *microsatellites* sont des séquences très courtes (2 ou 3 bases), hautement répétées. A un site donné, le nombre de répétitions de la séquence peut varier d'un individu à l'autre. Des enzymes capables de « couper » l'ADN aux deux extrémités permettent de révéler le polymorphisme du nombre de répétitions. Les systèmes de microsatellites sont extrêmement polymorphes.

B. Mise en évidence d'un QTL : un exemple simple

La mise en évidence de QTL au moyen des marqueurs moléculaires a été développée initialement chez des plantes : en effet, chez les plantes, où l'on manipule aisément des populations en ségrégation avec des effectifs importants, il a été facile d'établir des cartes génétiques. La méthode la plus simple pour mettre en évidence la liaison entre un marqueur moléculaire et un QTL consiste à croiser entre elles des lignées pures (au sens homozygotes à tous les locus) et à analyser les ségrégations à la génération suivante ou dans les générations ultérieures. C'est cet exemple que nous allons développer ici.

Soit un marqueur dont un allèle M est fixé dans une première lignée pure (P1) et un allèle m est fixé dans la seconde lignée pure (P2). Soit un QTL dont on présume l'existence, avec un allèle B fixé dans la première lignée et un allèle b dans la seconde. Les individus de la F1 ont tous le même génotype (doubles hétérozygotes) : il faut une deuxième génération de croisement pour voir apparaître un polymorphisme. En cas de back-cross sur une des deux lignées, nous avons 4 génotypes possibles, avec des probabilités dépendant du taux de recombinaison (r) entre le marqueur et le QTL supposé (figure 22).

Figure 22. Croisement de deux lignées pures (P1 et P2) suivi d'un back-cross sur P1, afin de détecter un QTL (voir texte). \mathcal{M} , m = allèles au marqueur. \mathcal{B} , b = allèles au QTL. r = taux de recombinaison entre le marqueur et le QTL.



Dans ce protocole, pour chaque individu, on peut observer le génotype au marqueur, d'une part, et la valeur phénotypique, d'autre part. On peut donc calculer la valeur phénotypique moyenne des homozygotes $\mathcal{M}\mathcal{M}$ et celle des hétérozygotes $\mathcal{M}m$. Soit D la différence entre ces deux moyennes ; un test statistique simple permet de dire si cette différence est significative ou pas. Si elle ne l'est pas, il faut conclure que l'analyse effectuée ne permet pas de mettre en évidence un QTL associé au marqueur ; cela ne signifie pas qu'il n'y en pas, mais que l'on n'a pas les moyens de le voir (par exemple, les effectifs mesurés sont insuffisants). Si D est significative, on peut alors mettre en relation cette différence observée avec la différence que l'on attendrait si un QTL était associé au marqueur. Pour ce faire, nous employons les notations qui sont rassemblées au tableau 19.

Tableau 19. Valeurs phénotypiques moyennes dans le cas du croisement de deux lignées pures suivi d'un back-cross (voir texte). Les moyennes relatives au marqueur (\mathcal{M} / m) correspondent à des valeurs observées, celles relatives au QTL (\mathcal{B} / b) sont des inconnues.

Génotype	Moyenne phénotypique	Différence
$\mathcal{M}\mathcal{M}$	\bar{X}	D
$\mathcal{M}m$	\bar{Y}	
$\mathcal{B}\mathcal{B}$	μ_{11}	Δ
$\mathcal{B}b$	μ_{12}	

D'après la figure 22, on établit la structure génotypique attendue au QTL pour les individus homozygotes au marqueur et les individus hétérozygotes au marqueur :

- Les individus $\mathcal{M}\mathcal{M}$ sont de génotype $\mathcal{B}\mathcal{B}$ en proportion $1-r$ et $\mathcal{B}b$ en proportion r
- Les individus $\mathcal{M}m$ sont de génotype $\mathcal{B}\mathcal{B}$ en proportion r et $\mathcal{B}b$ en proportion $1-r$

On en déduit l'expression de l'espérance des moyennes observées en fonction des paramètres inconnus :

$$E(\bar{X}) = (1-r)\mu_{11} + r\mu_{12}$$

$$E(\bar{Y}) = r\mu_{11} + (1-r)\mu_{12}$$

L'espérance de la différence observée entre génotypes au marqueur vaut donc :

$$E(D) = \bar{X} - \bar{Y} = \mu_{11}(1-2r) + \mu_{12}(2r-1) = (\mu_{11} - \mu_{12})(1-2r) = \Delta(1-2r)$$

Si les deux locus (marqueur et QTL) sont physiquement indépendants ($r = 1/2$), quel que soit l'effet du QTL, la différence phénotypique attendue entre les deux génotypes au marqueur est nulle (il y a autant de $\mathcal{B}\mathcal{B}$ et de $\mathcal{B}b$ parmi les $\mathcal{M}\mathcal{M}$ et parmi les $\mathcal{M}m$). De même, si les deux lignées parentales possédaient le même allèle au QTL, la différence attendue est nulle, car tous les individus ont le même génotype. Dans ces deux situations, on s'attend donc à ce que la différence observée (D) soit non significative.

En cas contraire, la différence attendue est d'autant plus grande que Δ est grand et que r est faible. A ce stade, on constate donc que l'on ne peut pas faire la différence entre un QTL induisant de grosses différences phénotypiques et physiquement loin du marqueur, d'une part, et un QTL proche du marqueur et induisant de faibles différences, d'autre part [on ne peut pas résoudre une équation à deux inconnues (Δ et r)]. Il faut donc soit réaliser d'autres croisements, comme, par exemple, les deux types de back-cross ou une F2 obtenue de façon panmictique à partir de la F1, soit analyser les données en fonction du génotype à deux marqueurs encadrant sur le chromosome le QTL supposé. Si, de la sorte, on peut estimer le taux de recombinaison (r), on déduit un estimateur de l'écart (Δ) entre la valeur moyenne des l'homozygotes $B\bar{B}$ et celle des l'hétérozygotes $B\bar{b}$:

$$\hat{\Delta} = D/(1-2r)$$

On raisonne de la même manière pour déterminer l'écart entre la valeur de l'hétérozygote $B\bar{b}$ et celle de l'homozygote $b\bar{b}$.

C. Autres situations et perspectives

1. Méthodes utilisées

Le croisement de deux lignées pures suivi d'un back-cross a été présenté pour des raisons de simplicité. En fait, ce protocole n'est pas celui qui est le plus couramment utilisé pour détecter des QTL responsables de variations des caractères d'intérêt agronomique ou zootechnique.

Chez les plantes, on part généralement du croisement de deux lignées pures car ce type de matériel est en général facile à obtenir. Les descendances en ségrégation qui, aujourd'hui, sont le plus utilisées sont soit de type F2, soit de type lignées recombinantes obtenues après quelques générations d'autofécondation des individus de la F2 ; dans certains cas, on pratique une haplodiploïdisation à partir des individus de la F1. L'intérêt de ces deux dernières techniques réside dans l'obtention de nombreuses réplifications d'individus ayant exactement le même génotype à tous les locus, ce qui permet de réduire l'impact des facteurs aléatoires de milieu. L'utilisation d'une F1 obtenue à partir de deux lignées pures est une stratégie très puissante car elle assure un déséquilibre d'association maximal entre le marqueur et le QTL

recherché (déséquilibre statistique ; voir GP). En effet, les gamètes produits par une plante F1 sont en majorité de type parentaux, MB et mb , alors qu'une faible proportion est de type recombiné, Mb ou mB (cf. figure 22).

En ce qui concerne les animaux, les choses sont plus compliquées. Il est en effet impossible, sauf chez des espèces de laboratoire, de disposer de lignées pures au sens où on l'entend en génétique végétale. Il est donc impossible de constituer une F1 à partir de laquelle on puisse observer des descendance en ségrégation. Il est donc nécessaire de tirer parti de situations particulières pour isoler, au sein d'une grande population, des sous-populations où existe un déséquilibre d'association des gènes. Le croisement entre deux races peut représenter une situation qui se rapproche du croisement de deux lignées pures. Ce protocole est cependant peu puissant car, d'une part, chacune des deux races présente en général une importante variabilité génétique (ce qui n'est pas le cas de deux lignées pures) et, d'autre part, les différences entre les deux races correspondent le plus souvent à des différences de fréquences géniques et non de nature d'allèles.

A l'heure actuelle, la stratégie vers laquelle s'orientent la plupart des recherches chez les animaux domestiques consiste à étudier les descendance intra-famille, car c'est la seule analyse pertinente en cas d'équilibre d'association des gènes. Cette approche consiste à comparer, au sein de la descendance d'un reproducteur, mâle en général, qui est hétérozygote au marqueur (Mm), les valeurs phénotypiques des descendants qui ont reçu l'allèle M ou l'allèle m . Dans cette approche, on ne peut pas attribuer dans l'absolu un effet à un allèle du marqueur : les effets ne sont définis qu'intra-famille, car un même allèle au marqueur peut être associé à un allèle favorable dans une famille et à un allèle défavorable dans l'autre. Dans ce protocole, un certain nombre de familles sont non informatives : celles des pères homozygotes au marqueur et celles des pères homozygotes au QTL associé au marqueur. Dans le premier cas, on ne peut pas classer les descendants en fonction de l'allèle reçu au marqueur ; dans le second cas, les descendants ont tous reçu de leur père le même allèle au QTL. Il résulte de tout cela que les protocoles de détection de QTL chez les animaux domestiques sont, par nature, moins puissants que ceux que l'on peut appliquer chez les plantes, et nécessitent d'être appliqués sur de grands effectifs.

2. Résultats et perspectives

Les exemples de QTL détectés sont déjà nombreux. Les espèces « pionnières » en la matière sont la tomate et le maïs pour les plantes, les bovins laitiers et le porc pour les animaux. Les résultats connus aujourd'hui montrent que la distribution des effets des QTL suit une courbe « en L » (figure 23) : un très petit nombre de QTL induit de fortes variations alors que l'essentiel des QTL détectés correspond à des locus qui ne sont chacun responsables que d'une faible part de variance. De plus, il subsiste encore des locus non détectés, en nombre indéterminé : la figure 23 indique bien qu'aucun locus responsable d'une part de variance inférieure à 4 % n'a été détecté, du fait de la limite de puissance du protocole. Les locus non détectés correspondent à ce que nous avons appelé des polygènes (cf. § B.2.b, chapitre II).

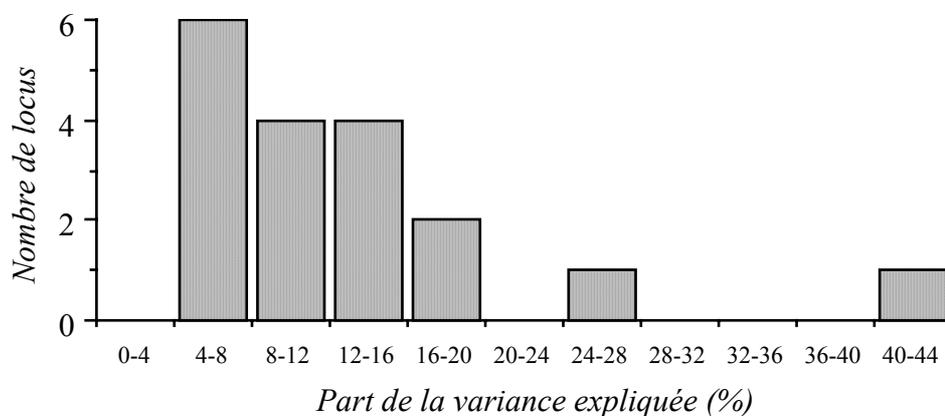
Figure 23. Distribution des QTL détectés pour des caractères mesurés chez la tomate selon la part de variance phénotypique (%) de l'un ou l'autre caractère qu'ils expliquent. Certains QTL jouent sur plusieurs caractères à la fois. Les caractères étudiés sont les suivants :

Poids moyen des fruits : 72 % de la variance phénotypique expliquée par des QTL

Teneur en sucres : 44 % de la variance phénotypique expliquée par des QTL

pH du fruit : 32 % de la variance phénotypique expliquée par des QTL

Source : Paterson et al., 1991.



Les perspectives d'application des marqueurs moléculaires sont nombreuses. On peut citer :

- ❑ La Sélection Assistée par Marqueur (SAM), qui consiste à utiliser le génotype des candidats aux marqueurs comme critère de sélection, en fonction des associations connues avec des QTL. L'information sur les marqueurs présente l'avantage de pouvoir être connue très tôt dans la vie de l'individu et peut donc représenter un critère de sélection précoce. Cette information peut se révéler très utile pour pallier un manque d'information phénotypique, par exemple, lorsque la mesure d'un caractère est très coûteuse ou lorsqu'elle ne peut pas s'effectuer sur certaines catégories d'individus (caractères ne s'exprimant que dans un seul sexe, etc.). On peut enfin combiner l'information sur les marqueurs avec les performances mesurées sur les candidats ou sur leurs apparentés afin d'améliorer la précision de l'évaluation génétique (voir les enseignements d'amélioration des plantes et des animaux).
- ❑ Le contournement de corrélations génétiques négatives, si les QTL détectés permettent de « disséquer » finement certains caractères complexes et de faire la part entre les gènes pléiotropes et ceux qui ne le sont pas.
- ❑ L'introgession de gènes assistée par marqueurs, permettant un gain de temps par rapport aux méthodes de rétrocroisement.
- ❑ La gestion de la variabilité génétique des populations sélectionnées ou des petites populations en conservation, en fournissant des outils fins d'analyse de la variabilité et en procurant des moyens d'optimiser le choix des reproducteurs en fonction du maintien de la variabilité génétique.

L'utilisation de ces nouveaux outils se raisonne en fonction du surcroît d'efficacité attendue et des surcoûts occasionnés.