

# Comment affiner la localisation d'un QTL ?

F. YTOURNEL, H. GILBERT, D. BOICHARD

INRA, UR337 Génétique Quantitative et Appliquée, F-78350 Jouy-en-Josas, France

Courriel : [helene.gilbert@jouy.inra.fr](mailto:helene.gilbert@jouy.inra.fr)

La cartographie fine de QTL est une étape indispensable, tant pour caractériser les gènes responsables de la variabilité génétique des caractères que pour sélectionner efficacement sur la base de l'information moléculaire. Cet article passe en revue les approches et les méthodes permettant de localiser un gène d'intérêt avec la meilleure résolution possible.

La variabilité des caractères quantitatifs mesurés dans les populations animales de rente résulte généralement de l'action combinée de gènes en nombre variable, mais souvent élevé, et d'effets de l'environnement. La variabilité d'origine génétique est due à des polymorphismes génétiques, c'est-à-dire des variations dans la séquence d'ADN modifiant l'effet des gènes. Ce sont ces polymorphismes génétiques, ou du moins les plus importants d'entre eux, que les généticiens cherchent à identifier à la fois pour comprendre le déterminisme des caractères et pour les sélectionner. Les locus impliqués dans la variation des caractères quantitatifs sont appelés *Quantitative Trait Loci* (QTL). Ils sont caractérisés par deux paramètres : leur effet sur le caractère cible et leur position sur le génome. La distinction entre notions de QTL et de gène majeur est arbitraire. On parle de gène majeur lorsque l'effet du locus est suffisamment fort pour que l'analyse des phénotypes seule permette de valider la ségrégation pour ce gène. Au contraire, l'effet individuel de chaque QTL ne peut être mis en évidence qu'à l'aide de marqueurs génétiques car ils expliquent une fraction sensiblement inférieure de la variabilité génétique du caractère. Le génome étant de très grande taille (3 milliards de paires de base, 25 000 gènes et plusieurs millions de polymorphismes dans les espèces de mammifères par exemple), l'identification et la caractérisation d'un QTL (position, effet sur les caractères) impliquent à la fois des appro-

ches de cartographie et des approches fonctionnelles. On considère généralement, sauf argument fonctionnel très fort, que l'approche cartographique est nécessaire dans une première étape pour réduire le nombre de gènes candidats fonctionnels. L'approche cartographique consiste à localiser le QTL dans une région chromosomique aussi réduite que possible, à l'aide de marqueurs génétiques dont les localisations sont connues. Si la cartographie des gènes majeurs est souvent assez précise, de l'ordre du centiMorgan (cM)<sup>1</sup>, celle des QTL l'est beaucoup moins, avec des intervalles de localisation de l'ordre de plusieurs dizaines de centiMorgan. Ces intervalles correspondent à plusieurs dizaines de millions de bases et à plusieurs centaines de gènes, rendant aléatoire la pré-sélection de gènes candidats fonctionnels. La recherche du ou des gènes sous-jacents aux QTL doit donc être précédée d'une étape supplémentaire dite de cartographie fine.

La cartographie fine correspond à une étape de réduction de la taille de l'intervalle de localisation du QTL. Par le passé, cette étape consistait à augmenter à la fois le nombre de marqueurs génétiques et le nombre d'individus pris en compte dans le dispositif. De plus en plus aujourd'hui, des méthodes dites à haut débit, rendues possibles dans un nombre croissant d'espèces par la disponibilité de la séquence de leur génome complet, permettent de saturer le génome en marqueurs dès l'étape de primolocalisation.

Il n'en demeure pas moins que l'étape de cartographie fine par extension du dispositif reste généralement nécessaire pour parvenir à une localisation précise du QTL. Dans le même temps, des méthodes statistiques particulières ont été développées pour exploiter au mieux ces données pour la cartographie fine.

Notre objectif est ici de présenter les principes généraux de la cartographie fine de QTL, puis de nous concentrer sur les facteurs influençant la résolution de la cartographie ainsi que sur les méthodes développées pour affiner la localisation du QTL.

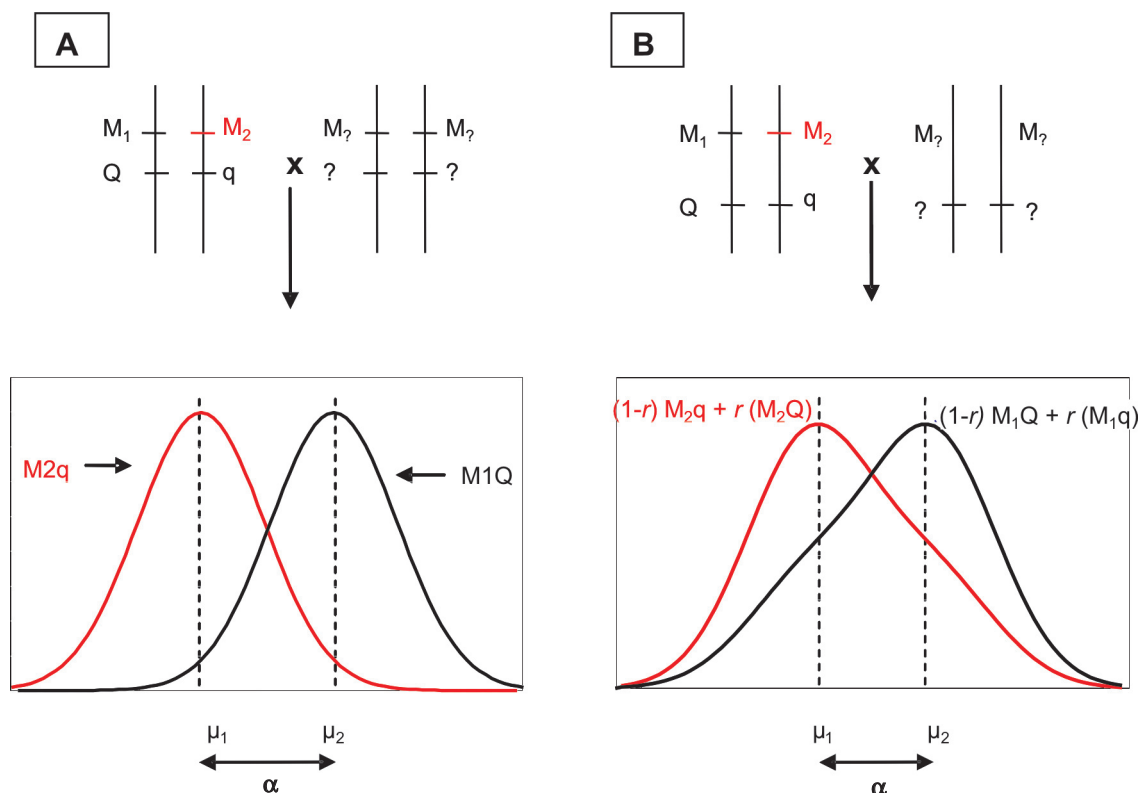
## 1 / Principes et facteurs influençant la résolution de la cartographie

### 1.1 / Principe général

Le principe général de la cartographie de QTL est ancien (Niemann-Sorensen et Robertson 1961) et on pourra se reporter à Boichard *et al* (1998) pour une synthèse en français. Reposant sur les propriétés de la méiose, il utilise le fait que de larges segments chromosomiques sont transmis en blocs de parent à descendant avec conservation chez le descendant de l'association entre locus proches chez le parent. Rechercher des QTL consiste à rechercher chez les descendants de parents supposés hétérozygotes au

<sup>1</sup> La distance entre deux locus sur un chromosome peut être mesurée de différentes manières. Les deux principales sont d'une part la distance physique, c'est-à-dire le nombre de bases séparant les deux locus, d'autre part la distance génétique, c'est-à-dire le pourcentage de recombinaison entre les deux locus lors de la production des gamètes. Pour des distances courtes, 1% de recombinaison correspond à 1 centiMorgan (1 cM). Chez les mammifères, on observe une relation moyenne de l'ordre de 1 million de bases par cM.

**Figure 1.** Ségrégation des allèles d'un QTL ( $Q$  et  $q$ ) dans la descendance d'un individu hétérozygote. A : avec un marqueur  $M$  ( $M_1$  et  $M_2$ ) complètement lié au QTL ( $r=0$ ), B : avec un marqueur  $M$  ( $M_1$  et  $M_2$ ) incomplètement lié au QTL ( $r=0,3$ ).



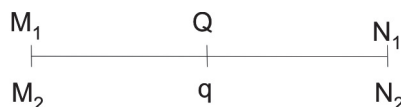
QTL des marqueurs dont le polymorphisme présente une association statistique avec le phénotype (figure 1). S'il est relativement facile de détecter un QTL, il est plus complexe de le localiser précisément. En effet, avec un seul marqueur par exemple, on ne peut pas faire la différence entre un QTL éloigné du marqueur avec un effet fort, et un QTL proche du marqueur avec un effet plus réduit, les deux conférant au marqueur le même effet apparent égal à

$\alpha(1-2r)$ ,  $\alpha$  étant l'effet de substitution d'un allèle du QTL par l'autre, et  $r$  le taux de recombinaison entre marqueur et QTL.

Pour distinguer position et effet des QTL, Lander et Botstein ont proposé en 1989 d'utiliser l'information apportée par deux marqueurs informatifs encadrant la position testée : c'est la cartographie d'intervalle ou «interval mapping». Dans l'intervalle considéré, la

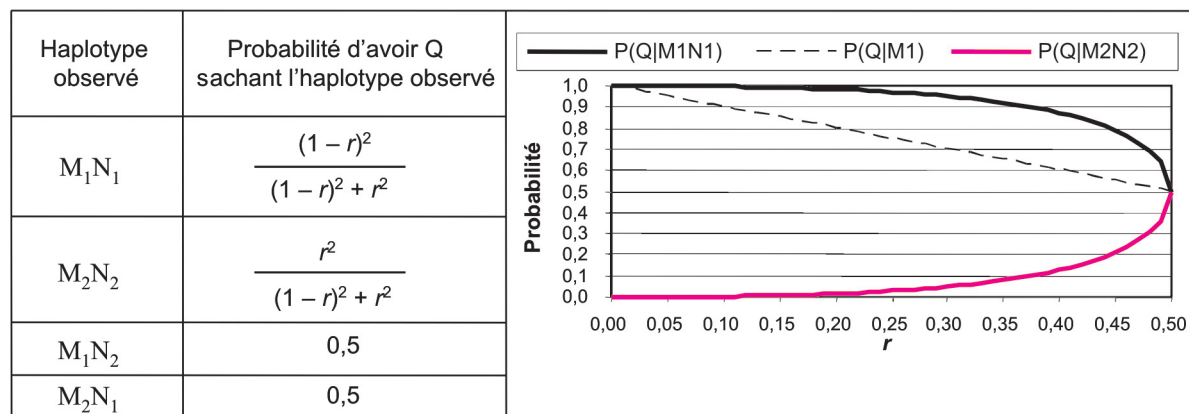
phase chez le parent (c'est-à-dire la succession ordonnée des allèles aux marqueurs sur chacun des deux chromosomes) et les probabilités de recombinaison sont supposées connues. Connaissant les génotypes du parent et du descendant aux marqueurs flanquant la position testée, on estime la probabilité que le descendant ait reçu l'un ou l'autre des allèles parentaux à cette position (figure 2). Cette probabilité est utilisée pour tester par régression ou

**Figure 2.** Probabilités pour l'haplotype  $M-N$  de porter  $Q$ , en fonction de  $r$ .



Taux de recombinaison ( $M-N$ ) =  $2r$

Taux de recombinaison ( $M-Q$ ) = taux de recombinaison ( $Q-N$ ) =  $r$



par test du maximum de vraisemblance la différence de moyennes de performances entre les deux groupes d'individus ayant reçu en probabilité l'un ou l'autre allèle au QTL. Chez des individus non recombinants, la probabilité de transmission est pratiquement constante pour toute localisation comprise entre les deux marqueurs assez proches. En effet, si  $r_1$  est le taux de recombinaison entre le 1<sup>er</sup> marqueur et la position testée,  $r_2$  le taux de recombinaison entre le 2<sup>ème</sup> marqueur et la position testée et  $r$  le taux de recombinaison entre les deux marqueurs, la probabilité  $p$  de transmission de l'allèle QTL parental est  $p = (1-r_1)(1-r_2)/(1-r)$ , proche de 1. On en déduit que seuls les recombinants permettent de localiser le QTL. Prenons l'exemple d'un gène majeur dont le génotype est déduit du phénotype et cartographié à l'aide d'un dispositif *backcross* avec un parent hétérozygote. Si ce gène est situé dans l'intervalle entre deux marqueurs, sa position par rapport à l'un des marqueurs est estimée par la proportion des individus qui n'ont pas recombiné entre le marqueur et le gène, parmi ceux qui ont recombiné entre les deux marqueurs. Pour un QTL (dont le génotype ne peut généralement pas être déduit avec certitude à partir du phénotype), la position montrant l'association la plus significative dans le test de régression ou de maximum de vraisemblance est considérée comme la position la plus probable du QTL.

La résolution atteinte par les méthodes de cartographie relève donc de deux facteurs : l'information apportée par le phénotype et celle apportée par les marqueurs génétiques.

## 1.2 / Facteurs de précision de la cartographie

### a) Informativité du phénotype

Un premier facteur jouant sur la résolution est la capacité à prédire le génotype au QTL d'après le phénotype observé. Lorsque le génotype est déduit du phénotype sans incertitude, la précision est maximale. Au contraire, lorsque les distributions des phénotypes conditionnellement aux génotypes (pénétrance) se superposent, le génotype n'est déduit du phénotype qu'en probabilité, ce qui induit une importante perte de précision dans la localisation du QTL. Idéalement, les distributions des phénotypes en fonction des génotypes doivent donc être peu chevauchantes : pour des phénotypes à distribution gaussienne, le cas idéal correspond à une différence entre les moyennes phénotypiques de chaque groupe d'au moins quatre (2,3% de superposition des distributions), voire six (0,1% de superposition) écarts-types résiduels (figure 3). Cette situation favorable correspond davantage à des gènes majeurs qu'à des QTL. Chez le porc, c'est par exemple le cas du gène RN pour lequel le potentiel glycolytique mesuré *in vivo* est un excellent prédicteur (Milan *et al* 1996), ou du gène Hal, pour lequel le test de réaction au gaz Halothane prédit le génotype (Ollivier 1975). Il est à noter que la variabilité intra-génotype au QTL peut dépendre à la fois de la variabilité génétique du caractère due à d'autres locus, de la variabilité environnementale et de la précision de la mesure. De plus, des phénomènes de dominance entre allèles, tels que ceux observés pour RN sur le rendement technologique Napole pour la transformation du jambon, per-

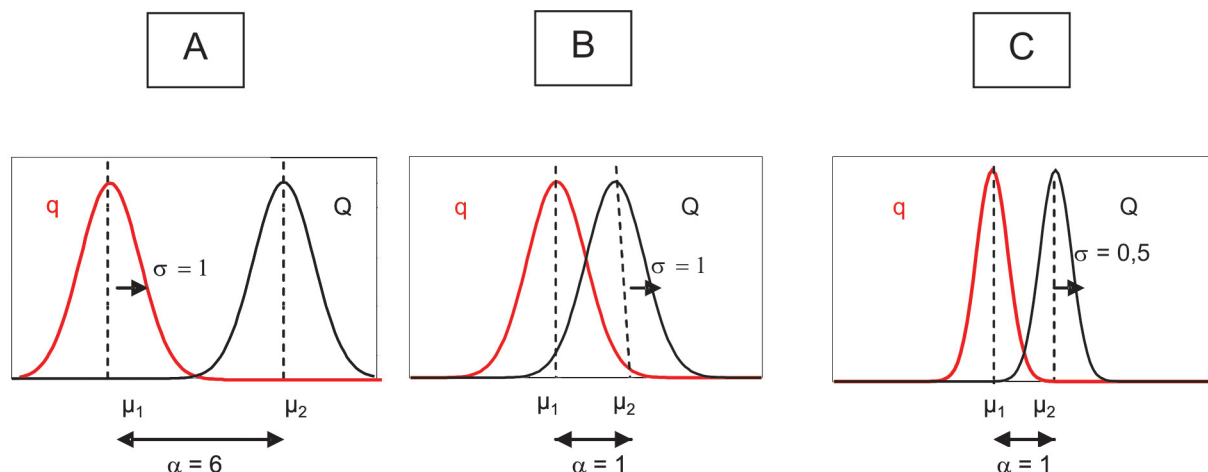
turbe la relation observable entre phénotype et génotype.

Pour optimiser ces critères, plusieurs types de stratégies ont été proposés : ne retenir que des individus de phénotypes extrêmes pour inférer leur génotype avec certitude, diminuer la variabilité résiduelle relativement à l'effet du QTL, ou prendre simultanément en compte plusieurs phénotypes de caractères corrélés.

### b) Génotypage sélectif

Cette méthode, proposée en 1987 par Lebowitz *et al*, et adaptée en 1989 par Lander et Botstein aux caractères quantitatifs, exploite le fait que l'essentiel de l'information sur le lien entre le QTL et le marqueur est apporté par les individus de phénotypes extrêmes : on estime que 81% de l'information quant à la liaison phénotype/génotype est apportée par les 33% des individus s'écartant de plus d'un écart-type de la moyenne de la distribution. Cette propriété vient du fait que la connaissance du génotype est généralement bonne pour les extrêmes et beaucoup plus incertaine sur la zone de chevauchement des distributions. Pour conserver la même puissance de détection que si tous les individus étaient génotypés, il convient, si on ne génotype que les individus de phénotype extrême, d'augmenter le nombre d'individus phénotypés (Lebowitz *et al* 1987) : augmenter d'un quart le nombre de phénotypes enregistrés permet de maintenir le niveau d'information en divisant par 2,5 le nombre de génotypes marqueurs à réaliser. Par ailleurs, comparé au génotypage de l'ensemble de la population, l'intervalle de confiance de la localisation du QTL reste inchangé

**Figure 3.** Distribution des phénotypes pour deux génotypes au QTL selon l'effet du QTL (A vs B), ou la variabilité du phénotype (B vs C).



quand on ne génotype que 40 à 50% des individus, choisis comme les plus extrêmes. Cette solution s'applique donc principalement à des populations où le nombre d'individus à produire et le nombre de phénotypes à mesurer n'est pas limitant. Dans le cas de cartographie fine, la localisation est très dépendante de chaque recombinant pris individuellement. En conséquence, aucune erreur sur le phénotype n'est acceptable et cette stratégie est généralement nécessaire pour inférer le plus justement possible les allèles au QTL. Cependant, le génotypage sélectif introduit un biais dans l'estimation des effets des QTL (Lander et Botstein 1989, Darvasi 1992) et nécessite donc un traitement statistique particulier.

#### c) Réduction de la variabilité résiduelle

Pour un effet de QTL donné, une meilleure discrimination entre distributions peut être obtenue en maîtrisant mieux leurs variances résiduelles. La répétition des mesures phénotypiques d'un même caractère, surtout lorsqu'il est très variable (concentration d'un métabolite, par exemple) pour un individu permet de limiter la variabilité d'origine environnementale et les erreurs de mesure dans l'estimation du phénotype. C'est encore plus net pour

les caractères exprimés plusieurs fois dans le temps, comme les lactations ou les tailles de portées successives.

Une autre façon de réduire la variabilité est d'utiliser comme phénotype la moyenne de nombreux descendants. Ainsi, dans le cas du testage sur descendance utilisé dans les dispositifs de détection de QTL dits «petites filles» ou «granddaughter designs» (Weller *et al* 1990) (figure 4), le phénotype du fils de première génération est la moyenne des performances de ses propres filles (de seconde génération). Par rapport à un dispositif «filles» (figure 4) où le phénotype est mesuré directement chez les descendants de première génération (Soller et Genizi 1978), le phénotype présente une variance résiduelle divisée par le nombre de petites-filles. Cette méthode permet de réduire les effets de l'environnement, des polygènes et autres sources de variabilité qui rendent les mesures directes moins précises. Ainsi, dans un dispositif de type «petites-filles», bien que les effets observés associés au marqueur du grand-père soient divisés par 2 par rapport aux effets estimés chez des filles (l'allèle grand-paternel du fils n'est transmis qu'à une petite-fille sur deux), cette réduction de l'effet est généralement plus que compensée par la réduction de la variance résiduelle. Les dispositifs sur trois générations sont donc généra-

lement plus puissants que ceux sur deux générations (Weller *et al* 1990, Van der Beek *et al* 1995), pour des coûts de génotypage réduits. Au-delà du gain de puissance qui est connu, on obtient également un gain de précision de localisation du QTL. On peut également combiner performances propres et performances de la descendance (Tribout *et al* 2008) et concentrer ainsi le maximum d'information phénotypique sur les individus génotypés.

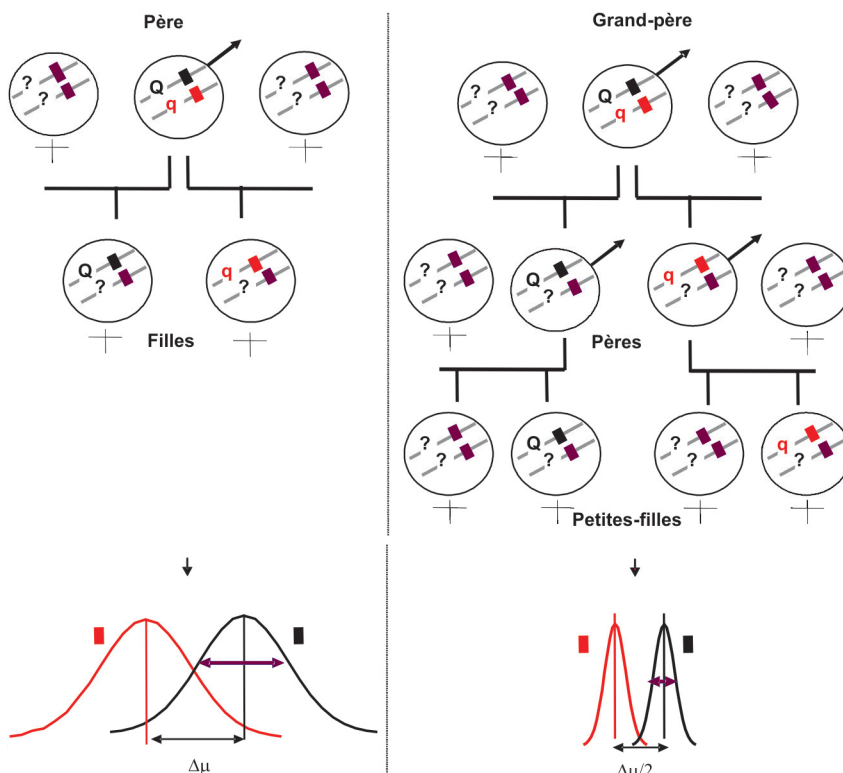
#### d) Utilisation de caractères corrélés

La différence phénotypique entre génotypes peut parfois être accentuée en considérant des caractères corrélés au caractère d'intérêt. Citons l'exemple du gène RN mis en évidence par le phénotype RTN pour lequel une différence de 3 écarts-types est mise en évidence, mais qui est due à une différence de potentiel glycolytique dans le muscle de 7 écarts-types. Ce nouveau phénotype permet une discrimination parfaite entre les génotypes, ce qui est un avantage considérable dans l'étape de cartographie fine.

L'utilisation des distributions conjointes entre caractères pour affiner la connaissance d'un phénotype donné a été proposée par Weller *et al* (1996). Elle permet d'affiner la localisation du QTL, mais aussi de préciser ses effets grâce à la corrélation entre les caractères. Le gain d'information obtenu par ces stratégies s'avère particulièrement important lorsque les corrélations dues au QTL ont des directions opposées aux corrélations phénotypiques entre les caractères (Korol *et al* 1995). L'utilisation de la seule corrélation résiduelle entre les caractères permet également d'augmenter l'information et la puissance du test (Korol *et al* 1995, Gilbert et Le Roy 2003).

Les données d'expression, qui peuvent par exemple refléter dans un tissu l'expression différentielle de gènes entre deux états (animal sain/malade, traité/témoin, nourri/à jeun...), peuvent elles aussi être utilisées comme des phénotypes. On parle alors de détection de QTL d'expression ou de e-QTL. Bien que les résultats soient encore peu nombreux chez les espèces d'élevage, on peut imaginer que l'expression de certains gènes impliqués dans la construction de la différence phénotypique entre génotypes montre des différences beaucoup plus marquées que l'écart macroscopique entre phénotypes classiques. De façon analogue au lien entre le gène RN et le potentiel glycolytique du muscle cité plus haut, ce phénotype

**Figure 4.** Dispositifs «filles» (à gauche) et «petites-filles» (à droite) et distributions des phénotypes selon l'allèle hérité au QTL tracé par le marqueur A/a.





nouveau permet alors en théorie une cartographie beaucoup plus fine.

Un problème fréquemment rencontré en cartographie de QTL est l'hétérogénéité génétique de phénotypes similaires. Les données plus analytiques, et en particulier les données d'expression, peuvent permettre de préciser le phénotype et de concentrer l'étude sur des groupes homogènes, dont la différence est liée uniquement au QTL recherché (Schadt *et al* 2003). En éliminant du bruit de fond, cette méthode peut améliorer la localisation du QTL, au prix cependant d'une perte de données.

### 1.3 / Informativité des marqueurs

Dans un dispositif donné, la résolution est maximale lorsque chaque événement de recombinaison est localisé exactement. En pratique, ceci nécessite pour chaque recombinaison de disposer de deux marqueurs très proches, informatifs et flanquants. Ceci peut être obtenu en saturant le génome, ou tout au moins la région considérée, par des marqueurs génétiques informatifs nombreux.

Jusqu'à une date récente, les dispositifs usuels de cartographie de QTL étaient réalisés avec un maillage de marqueurs peu dense, soit un marqueur tous les 10 à 20 cM. Une région d'intérêt n'était donc couverte que par 2 à 3 marqueurs, pas toujours informatifs. La phase de cartographie fine impliquait alors de rajouter de nombreux marqueurs dans cette région. Avant le séquençage des génomes complets, ces marqueurs n'étaient pas toujours disponibles et devaient souvent être développés spécifiquement. Cette étape était longue et laborieuse.

Chez les espèces dont le génome est séquencé, des polymorphismes génétiques très nombreux ont été mis en évidence. Parmi eux, les polymorphismes de substitution d'une base ou SNP (*Single Nucleotide Polymorphism*) sont des cibles de choix, de par leur nombre (plusieurs millions sur le génome) et leur facilité de typage. En effet, des puces commerciales ont été développées, permettant de génotyper plusieurs milliers, voire plusieurs centaines de milliers de marqueurs SNP simultanément. Par exemple, la puce bovine actuelle la plus utilisée compte 54 000 SNP (en comparaison, la puce humaine de référence en compte près de 400 000). Les méthodologies varient selon les fabricants et reposent sur les

techniques d'hybridation différentielle ou d'extension d'amorce. Pour un coût analogue à celui de quelques dizaines ou centaines de microsatellites, on peut génotyper chaque individu pour des dizaines de milliers de marqueurs couvrant tout le génome. Avec ces nouvelles technologies de génotypage, les phases de primolocalisation et de génotypage dense sont concomitantes. Même si cette densité n'est pas encore totalement optimale (on considère qu'un marqueur tous les 10 kilobases serait en théorie nécessaire pour obtenir la résolution maximale dans les espèces d'élevage, soit une puce de 300 000 SNP), ces puces constituent déjà un progrès considérable en termes de temps, de coût et de travail. Elles ne sont pas encore disponibles pour toutes les espèces de rente mais devraient être progressivement développées en fonction de l'avancée du séquençage des génomes.

Une forte densité en marqueurs informatifs est cependant une condition nécessaire mais pas suffisante pour obtenir une localisation précise des QTL. En effet, Darvasi *et al* (1993) ont démontré que, pour un effet de QTL et une taille de population donnée, la résolution maximale du dispositif expérimental est déterminée par le nombre de recombinaisons informatives présentes dans le dispositif. Par définition, entre deux marqueurs séparés de 1 cM (soit 1 million de paires de bases en moyenne chez les Mammifères), il y a une recombinaison toutes les cent méioses. Les dispositifs expérimentaux classiques pour la cartographie de QTL, composés en général de quelques centaines d'individus répartis en familles de 50 à 100 descendants, ne permettent d'espérer que quelques recombinaisons par cM, ce qui limite la précision de localisation à l'échelle du dispositif de détection à des segments de plusieurs dizaines de cM. Il est généralement admis qu'une forte augmentation de la densité des marqueurs dans ce type de dispositif n'augmente que faiblement la résolution offerte par les dispositifs sur deux générations, le facteur limitant devenant le nombre de recombinaisons. Darvasi et Soller (1997) obtiennent deux estimations de l'intervalle de confiance  $I$  de localisation (en cM) dans des dispositifs expérimentaux F2 ou *backcross* complètement informatifs (les parents F1 sont hétérozygotes au QTL) et une carte saturée en marqueurs. Dans la première estimation, l'intervalle de confiance est exprimé en fonction de l'effet de substitution  $\alpha$  du QTL supposé biallélique. Dans la

seconde, il est exprimé en fonction de la part de variance phénotypique  $p$  expliqué par le QTL :

$$I = 3000 / (N \alpha^2)$$

$$I = 530 / (N p)$$

où  $N$  est le nombre de méioses (le nombre de descendants en *backcross*, 2 fois ce nombre en F2). Cette expression est généralisable à n'importe quel dispositif avec  $N$  le nombre de méioses informatives. Il est à noter que dans certains dispositifs animaux intra-race usuels, seules les méioses paternelles sont utilisées, et uniquement une fraction des pères est hétérozygote au QTL, ce qui limite la valeur de  $N$ .

La disponibilité de nombreuses recombinaisons et la capacité à les identifier sont donc cruciales pour la cartographie fine. La recombinaison étant un phénomène aléatoire que l'on ne peut contrôler, il faut augmenter le nombre de méioses informatives disponibles pour en obtenir plus. Une première solution est de maximiser la proportion de parents hétérozygotes au QTL, d'où la stratégie fréquente de croisement entre lignées consanguines ou entre animaux présentant des phénotypes divergents pour procréer des parents F1. Dans les populations commerciales, le choix des parents présentant les descendances les plus variables est généralement efficace pour augmenter le pourcentage de F1 hétérozygotes au(x) QTL. Dans la mesure du possible, l'utilisation d'un dispositif avec les deux méioses parentales utilisables est à privilégier, ce qui incite à constituer de larges familles non seulement de père mais aussi de mère. L'utilisation d'un très grand dispositif expérimental sur une génération est une option possible mais coûteuse : Darvasi et Soller (1995) ont démontré que, pour cartographier un QTL ayant un effet de substitution  $\alpha = 0,25$  avec un intervalle de confiance à 95% de 5 cM, il faut 4 500 individus F2.

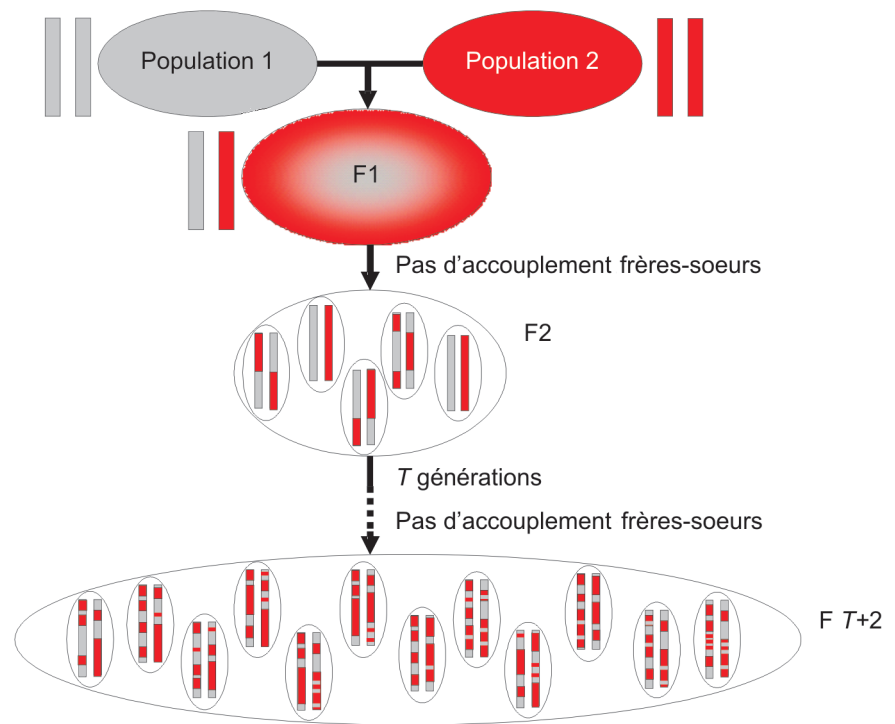
Deux autres stratégies peuvent être privilégiées : la procréation expérimentale de nouveaux individus, et l'utilisation de méioses qui ont eu lieu au cours des générations passées (recombinaisons «historiques»). Des méthodologies différentes sont alors utilisées, basées sur des stratégies dites de dissection chromosomique dans le premier cas, et le développement de nouvelles méthodes statistiques dédiées à la cartographie fine dans le deuxième cas.

## 2 / Dissection chromosomique

La dissection chromosomique résulte de la production de nouvelles méioses sur plusieurs générations, afin d'identifier les recombinaisons produites et de réduire progressivement la longueur de la région chromosomique d'intérêt. Les *lignées d'intercroisement avancé* (ou AIL pour *Advanced Intercross Lines*) sont créées à partir de croisements entre individus de la génération  $F_1$  précédente, en évitant les accouplements frère-sœur (figure 5). Les recombinaisons ainsi obtenues sont accumulées dans une seule population et réduisent la taille des segments parentaux initiaux. Ces croisements sont optimaux en terme de nombre de méioses utiles. Lorsque deux populations consanguines sont utilisées comme parents, Darvasi et Soller (1995) ont ainsi démontré que, sans génotypage supplémentaire et après 10 générations, l'intervalle de confiance de la localisation d'un QTL peut être jusqu'à cinq fois plus petit qu'avec un dispositif BC ou  $F_2$ . La ségrégation des allèles au QTL génère trois génotypes, comme dans les populations  $F_2$ . L'existence d'un effet est donc testée de la même manière. Dans ces conditions, seuls deux allèles par marqueur, d'origine connue, sont en ségrégation dans la population d'AIL, simplifiant grandement l'analyse. Lorsque les populations initiales ne sont pas fixées, l'analyse nécessite un suivi des segments chromosomiques dans la population, rendant beaucoup plus complexe l'analyse, ou se résume à une analyse d'association marqueur par marqueur, de sorte que le gain d'efficacité est à analyser plus précisément. Que les populations parentales soient consanguines ou non, la méthode repose sur une forte densité de marqueurs.

Les *lignées recombinantes consanguines* (ou RIL pour *Recombinant Inbred Lines*) sont rares chez les animaux domestiques car très coûteuses à obtenir. Elles sont par contre fréquentes chez les végétaux ou la souris. Issues d'un certain nombre de générations d'accouplements frère-sœur (figure 6A) ou d'autofécondations après la  $F_1$ , ces lignées sont génétiquement fixées pour la plus grande partie de leur génome. Dès lors qu'elles sont suffisamment nombreuses (ce qui est une contrainte importante), elles peuvent être utilisées en cartographie fine de deux façons. D'une part, comme pour une  $F_2$ , la présence d'un QTL peut être testée par une différence phénotypique

Figure 5. Schéma d'obtention de lignées d'intercroisement avancé (AIL).

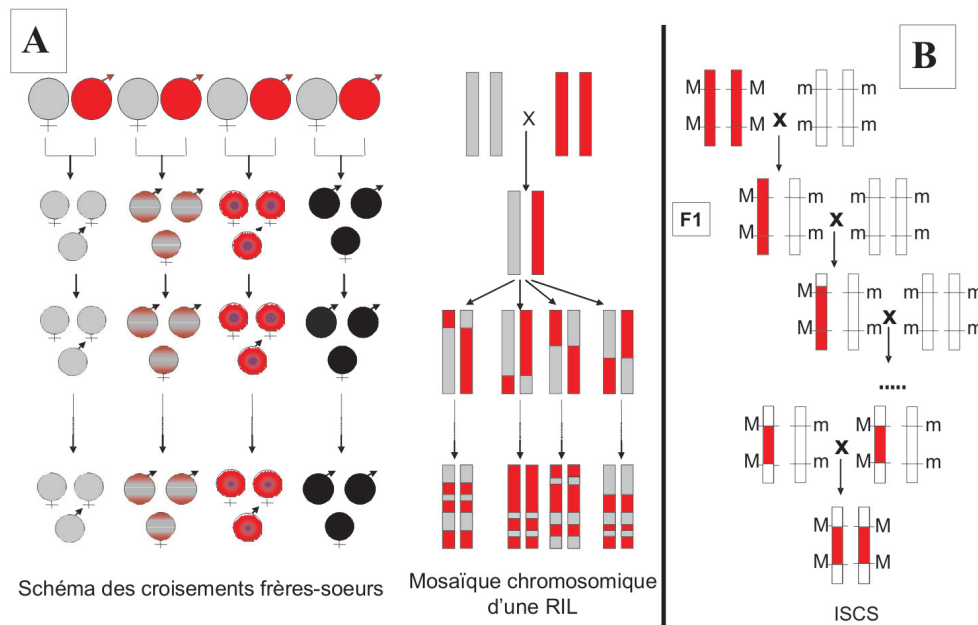


significativement associée à une région différemment fixée entre lignées. On observe un gain par rapport à la  $F_2$  dans la mesure où le nombre de générations réalisées correspond à une accumulation de recombinaisons et donc à une augmentation du taux de recombinaison équivalent sur une génération (ou taux de recombinaison apparent). Ce taux de recombinaison apparent tend vers  $4r(1+6r)$  pour les croisements frères-sœurs et  $2r(1+2r)$  pour les autofécondations,  $r$  étant le taux de recombinaison observé dans la  $F_2$  (Lynch et Walsh 1998). Les segments conservés sont donc de plus petite taille, ce qui améliore la résolution. D'autre part, le gain de résolution passe par une diminution de la variance intra-lignée et par une augmentation de l'information du dispositif, tous les individus étant homozygotes et totalement informatifs (AA vs aa à chaque locus). La puissance des RIL est généralement très élevée, surtout si le caractère est gouverné par peu de QTL et qu'il a une faible héritabilité, du fait de la réduction de la variabilité environnementale et polygénique qu'elles permettent. Les RIL sont aussi un outil facilitant grandement la recherche des épistasies (interactions entre deux ou plusieurs gènes), en individualisant des combinaisons de gènes. Autre atout, comme toute lignée fixée, les RIL, si elles sont conservées, peuvent être phénotypées pour tout nouveau caractère et ne nécessitent pas de génotypage nouveau.

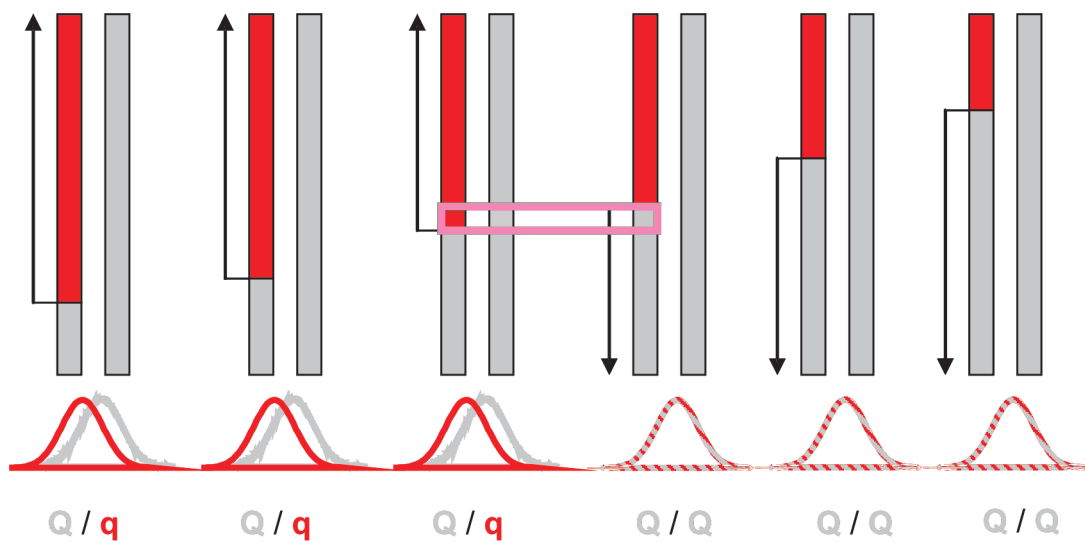
Les RIL présentent un autre avantage moins connu en cartographie fine. Elles gardent statistiquement une petite fraction du génome non fixée. Utilisée comme parent de dispositif de détection de QTL, un RIL peut exprimer une ségrégation du caractère, ce qui permet de localiser finement le QTL à l'intérieur de la région encore hétérozygote, région qui peut ensuite être réduite par des recombinaisons dans une génération suivante. Cette approche a été utilisée avec succès par Loudet *et al* (2007) chez *Arabidopsis*.

Une approche comparable est parfois utilisée chez les espèces d'élevage, le porc par exemple, dans le cadre de *backcross* successifs. On choisit comme parents des individus recombinants à différents points de la région étudiée (figure 7), donc homozygotes ou hétérozygotes au QTL recherché. Ces parents sont ensuite accouplés à des conjoints homozygotes. On déduit le statut du parent en fonction de la présence ou l'absence de ségrégation du QTL chez ses descendants. Ceci permet de localiser le QTL dans une région donnée (le rectangle rose sur la figure 7). Cette méthode est très efficace pour des régions de taille moyenne, mais elle nécessite des effectifs très élevés lorsque la région devient petite et n'est réaliste que lorsque la prolificité est élevée et le coût d'élevage et de phénotypage réduit.

**Figure 6.** Lignées congéniques. A : dispositif pour l'obtention de lignées recombinantes consanguines et chromosomes obtenus. B : Obtention d'une lignée congénique spécifique d'un intervalle.



**Figure 7.** Méthode de dissection chromosomique permettant de localiser le QTL (dans le rectangle rose).



Les *lignées congéniques* (CL) ou *lignées quasi-congéniques* (NIL) sont issues de BC récurrents sur l'une des lignées parentales. Ces lignées sont plus intéressantes si on adjoint au *backcross* récurrent une introgression assistée par marqueurs. On obtient alors des *lignées congéniques spécifiques d'un intervalle* (*Interval-Specific congenic strains*, ISCS) (Darvasi 1996) qui sont identiques à la lignée en retour, à l'exception d'une fraction de chromosome provenant de la lignée donneuse (figure 6B). Toutes les lignées ainsi créées peuvent posséder un fragment introgressé différent. On peut alors tester la présence d'un QTL par

comparaison des moyennes phénotypiques des différentes lignées (Burgio *et al* 2007, L'Hôte *et al* 2007).

### 3 / Méthodes statistiques de cartographie fine

Les méthodes précédemment citées reposent sur des populations expérimentales. De telles populations ne peuvent pas être créées dans toutes les espèces d'élevage, notamment à cause d'intervalles de génération trop longs et/ou du coût de production des animaux. Elles ne sont bien sûr pas envisa-

geables chez l'homme. Une alternative pour la cartographie fine de QTL consiste à utiliser les recombinaisons historiques, en analysant le passé des populations.

#### 3.1 / Illustration de la méthode avec l'«homozygosity mapping»

La méthode de cartographie par recherche d'homozygotie a été développée chez l'homme pour la recherche d'anomalies génétiques récessives (Lander et Botstein 1987). Ces gènes ont en général une fréquence faible et on suppose, au moins dans une popula-



tion donnée, que la maladie est due à un seul évènement de mutation dans un gène donné chez un ancêtre unique, et que cette mutation a provoqué l'apparition de l'allèle «malade». Les porteurs d'un seul allèle malade ont un phénotype normal (qui ne peut pas être distingué de celui des individus n'ayant aucun allèle de la maladie), tandis que les individus atteints sont porteurs de deux allèles malades identiques provenant du même ancêtre. Ces individus malades ont donc reçu deux fois le même allèle muté, de leur père et de leur mère, mais aussi un segment chromosomique entourant ce gène, identique à celui de l'ancêtre. La longueur de ce segment varie d'un individu à l'autre en fonction des recombinaisons survenues au cours des générations depuis l'apparition de la mutation. Ainsi, plus l'ancêtre est éloigné, plus le segment chromosomique commun entre tous les chromosomes des individus atteints est réduit, et plus la cartographie peut être fine. De même, plus le nombre d'individus atteints est élevé, plus le segment commun est court. Le phénomène, illustré en figure 8, montre comment les générations et le nombre d'individus porteurs de la mutation réduisent la longueur des segments conservés par accumulation des recombinaisons. Le principe de cartographie, dans le cas présent, revient à rechercher des zones homozygotes et identiques, héritées de l'ancêtre commun, chez les malades. Cette méthode n'est applicable que pour les gènes récessifs autosomaux

mais elle illustre le principe des méthodes basées sur l'analyse du déséquilibre de liaison : si deux individus partagent le même allèle identique par descendance (dérivant d'un même ancêtre commun), le segment chromosomique entourant ce gène est lui aussi identique par descendance et il est d'autant plus court que le nombre de générations séparant les deux individus est élevé.

### 3.2 / Le déséquilibre de liaison

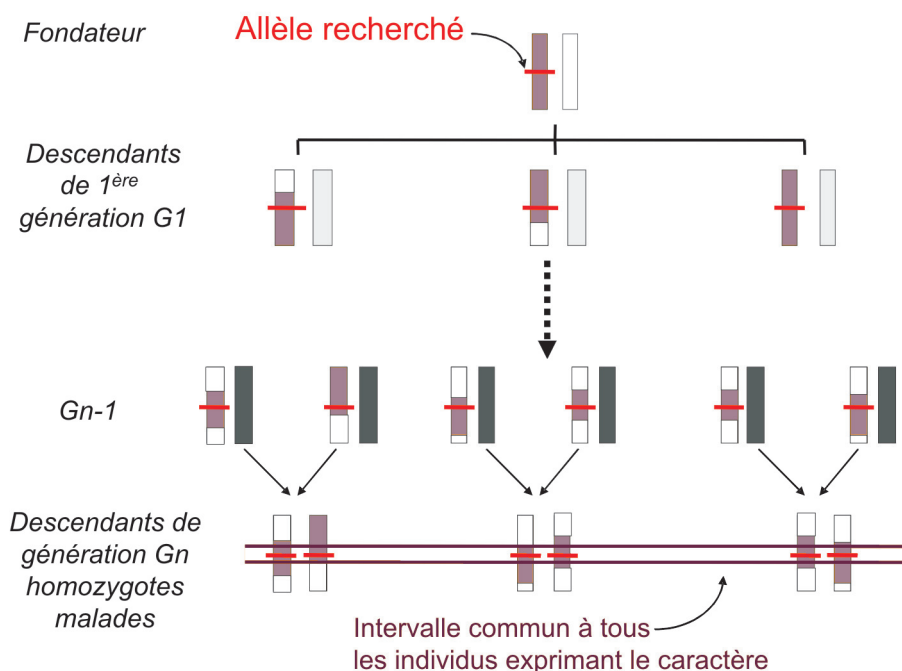
Le Déséquilibre de Liaison (DL) est défini comme l'association non aléatoire d'allèles à des locus différents. Dans le cas de locus bi-alléliques ayant respectivement pour allèles A/a et B/b, il se calcule comme  $D_{AB} = x_{AB} - p_A q_B$ , où  $x_{AB}$  est la fréquence de l'haplotype (ou phase) porteur des allèles A et B, et  $p_A$  ( $q_B$ ) la fréquence de l'allèle A (B) dans la population. La formule précédente se généralise pour les locus multi-alléliques. Cette mesure étant sensible aux fréquences alléliques, diverses fonctions de  $D$  ont été proposées pour éliminer, en général incomplètement, ce défaut.

Le déséquilibre de liaison entre deux locus peut avoir une origine variée et toutes les forces évolutives comme la mutation, la dérive génétique, les régimes d'accouplement ou la sélection créent du déséquilibre. La présence d'un DL entre deux locus n'implique pas de liaison physique entre eux. Cependant, l'évolution de ce déséqui-

bre est fonction de la distance génétique entre les locus et du nombre de générations écoulées entre la création du DL et la génération actuelle. Si  $r$  est le taux de recombinaison entre deux locus et  $D_0$  le déséquilibre initial entre les deux locus dans une population de grande taille, en supposant qu'il n'y a pas de variation des fréquences alléliques entre les générations, le DL entre les locus diminue d'un facteur  $(1-r)$  par génération dans le cas d'accouplements aléatoires. On obtient alors par récurrence la relation  $D_t = (1-r)^t D_0$ , avec  $D_t$  le DL à la génération  $t$ . Le déséquilibre de liaison se maintient donc de manière prolongée entre des locus proches mais il diminue rapidement pour des taux de recombinaison plus élevés.

De ces propriétés on déduit le principe essentiel de la cartographie utilisant le déséquilibre de liaison. Dans une population fermée de taille importante avec des accouplements au hasard, un fort déséquilibre de liaison n'est observé entre un marqueur et un QTL que si le marqueur est très proche du QTL. Le DL se traduit par une association entre un allèle marqueur (ou un haplotype d'allèles marqueurs) avec le phénotype. Cette approche permet en théorie une cartographie très fine, surtout si la population est ancienne et de grande taille. A noter cependant qu'elle repose sur des hypothèses assez fortes et que tout écart à ces hypothèses, en particulier l'existence de migration ou de structuration dans la population, peut se traduire par des résultats faux positifs. Par ailleurs, dans les populations sélectionnées et/ou de taille réduite, des déséquilibres de liaison existent à plus grande distance, réduisant la résolution de la cartographie ; c'est particulièrement le cas de nombreuses populations très sélectionnées, donc d'effectif génétique réduit. Il peut être particulièrement utile de tirer profit de l'histoire des populations d'élevage, marquée par quelques grands événements comme la domestication et la création des races. Intra race, les segments chromosomiques conservés entre individus sont généralement de grande taille, générant un déséquilibre de liaison à grande distance (souvent plusieurs centaines de kilobases) et limitant la précision de la cartographie par déséquilibre de liaison. Au contraire, les segments conservés entre races sont d'autant plus courts que les races ont divergé de longue date. Le déséquilibre de liaison entre races diminue rapidement avec la distance entre marqueurs et offre une excellente résolution potentielle de cartographie (quelques dizaines de kilo-

Figure 8. Principe de l'«homozygosity mapping».





bases), dès lors que le marquage moléculaire est suffisamment dense. Un exemple remarquable chez le chien illustre cette stratégie (Karlsson *et al* 2007).

On peut distinguer les méthodes statistiques de cartographie fine selon plusieurs grands critères :

- en fonction de l'hypothèse faite sur l'origine et la nature du polymorphisme au QTL : création d'un locus biallélique par un polymorphisme génétique ponctuel unique (Terwilliger 1995, Farnir *et al* 2002, Boitard *et al* 2006) ou absence d'hypothèse sur le nombre d'allèles au QTL,

- en fonction de la nature de l'information utilisée pour la cartographie, le DL exclusivement (Spielman 1993, Terwilliger 1995, Boitard *et al* 2006), ou combinant le DL et l'information de liaison apportée par le pedigree (Meuwissen et Goddard 2000, Farnir *et al* 2002).

- en fonction du nombre de marqueurs utilisés pour définir le déséquilibre de liaison.

### 3.3 / Méthodes de cartographie fine basées sur l'utilisation du DL

Sans passer en revue les nombreuses méthodes disponibles, montrons le principe de certaines d'entre elles. Le tableau 1 propose une comparaison des méthodes présentées.

Une première méthode consiste à estimer et tester le coefficient de régression du phénotype sur le nombre de copies (0, 1 ou 2) de chaque allèle d'un marqueur, dans une population d'individus aussi peu apparentés que possible. Bien que très simple, limitée à un seul marqueur à la fois, cette méthode présente un grand intérêt dans les premières analyses à haut débit comptant de très nombreux marqueurs. Plusieurs études visant à comparer

différentes méthodes constatent la bonne efficacité générale de cette approche, souvent compétitive par rapport à des méthodes beaucoup plus complexes.

La méthode de Hästbacka *et al* (1992) permet des études de type cas-témoin et suppose qu'un polymorphisme unique, responsable d'une maladie, est apparu dans un haplotype fondateur,  $t$  générations avant la génération actuelle. Le principe est d'identifier un haplotype très majoritaire dans la population malade, considéré comme l'haplotype initialement porteur du polymorphisme, transmis identiquement au cours des générations. Les autres haplotypes porteurs de cette forme unique du gène sont supposés être issus de recombinaisons. La proportion d'haplotypes IBD dans la population malade est utilisée comme une estimation du DL restant dans la population par rapport au DL initial, égale à  $(1-r)^t$ , qui tend vers  $\exp(-rt)$  lorsque  $r$  tend vers 0 (c'est-à-dire quand le marqueur est très proche de la mutation responsable). En supposant connu l'âge de la mutation ( $t$ ), on en déduit la position du locus par rapport à celle du marqueur. L'estimation de la distance entre le locus de la maladie et le marqueur est cependant très sensible à cette hypothèse, car la date d'apparition du polymorphisme n'est que rarement connue avec précision.

La méthode de Terwilliger (1995) est assez proche dans son principe et repose sur la même hypothèse d'un unique polymorphisme causal unique mais elle formalise davantage le raisonnement. L'haplotype fondateur n'est pas présupposé mais déterminé par maximum de vraisemblance. On analyse la surreprésentation de certains allèles marqueurs chez les individus malades actuels par rapport aux individus sains,  $t$  générations après l'apparition du polymorphisme. La probabilité de présence de

chaque allèle marqueur conditionnellement à la présence de l'allèle muté causal s'écrit en fonction ( $a$ ) des fréquences des allèles au marqueur dans la population avant l'apparition de l'allèle causal (et supposées constantes), ( $b$ ) de la fréquence actuelle de la maladie dans la population, et ( $c$ ) de l'excès d'association ( $\lambda$ ) entre l'allèle marqueur et l'allèle causal. Après sommation sur tous les allèles, l'hypothèse d'une liaison entre le marqueur et le gène recherché est testée par rapport de vraisemblance contre l'hypothèse qu'il n'y a pas d'excès d'association ( $\lambda = 0$ ).

L'information sur plusieurs marqueurs peut être intégrée en multipliant les vraisemblances des différents marqueurs en supposant qu'ils sont indépendants. Cette hypothèse est évidemment fautive pour les marqueurs physiquement liés de la région chromosomique d'intérêt, rendant la distribution de la statistique de test inconnue sous l'hypothèse nulle. Cette distribution peut être construite par simulation. Cette extension de la méthode permet d'obtenir une estimation de la distance entre les marqueurs et le gène : on suppose que la décroissance du DL est fonction du temps et du taux de recombinaison entre le QTL et le marqueur sous l'hypothèse d'équilibre d'Hardy-Weinberg. Cet équilibre correspond à une constance des fréquences des haplotypes au cours des générations en l'absence de dérive génétique, de mutation, de migration et de sélection, et avec des accouplements aléatoires.

Cette méthode a été étendue à la détection de QTL par Abdallah *et al* (2004). Les probabilités de porter un allèle marqueur donné conditionnellement au génotype au QTL sont alors écrites en fonction des distributions des phénotypes selon le génotype. Boitard *et al* (2006) ont développé la méthode HapIM, qui est une extension exacte de la précédente à des haplotypes de deux marqueurs.

**Tableau 1.** Comparaison des hypothèses faites par les méthodes statistiques de cartographie fine.

	Méthode			
	Hästbacka <i>et al</i> (1992)	Abdallah <i>et al</i> (2004)	TDT (Spielman <i>et al</i> 1993)	Meuwissen et Goddard (2000, 2001)
DL initial	Complet	Complet	Complet	Aucun
Type de population	Cas - témoin	Individus supposés non apparentés	Cas - témoin en structures parents-enfants	Structures familiales de type populations animales
Age du polymorphisme causal	Supposé connu	Paramètre estimé	-	Supposé connu
Effets estimés	-	Additif et dominance	-	Effets des haplotypes
Estimation de la position	Distance au marqueur en plus fort DL	Au marqueur ou dans un haplotype	Au marqueur	Milieu de l'haplotype

L'inconvénient majeur des tests d'association est leur sensibilité à un écart aux hypothèses. Une méthode très utilisée en génétique humaine vise à augmenter la robustesse de l'association en l'analysant intra-famille. Le *Transmission Disequilibrium Test*, ou TDT, (Spielman *et al* 1993) repose sur l'étude de la transmission ou la non-transmission des allèles d'un marqueur bi-allélique entre des parents hétérozygotes et des enfants malades. Il utilise donc des données issues de structures familiales de type cas (individus malades) – contrôle (frères-soeurs sains). Si la ségrégation des locus marqueurs est indépendante de celle du gène de la maladie, alors les individus malades doivent hériter des deux allèles parentaux dans les mêmes fréquences que les individus sains. Le test de différence de fréquences est alors un  $\chi^2$  à un degré de liberté. Cette statistique dépend à la fois de la liaison physique et du DL entre les deux locus. Cette méthode présente deux inconvénients : a) elle est peu puissante du fait de l'utilisation exclusive des parents hétérozygotes, b) elle permet de détecter une association avec un marqueur mais pas de localiser le gène par rapport au marqueur.

Une autre méthode pour augmenter la robustesse de l'analyse est de combiner l'information familiale (l'analyse de liaison donnant des résultats robustes mais peu précis en terme de localisation) et l'information du déséquilibre de liaison. Une approche de ce type a été proposée par Farnir *et al* (2002) qui ont combiné la méthode de Terwilliger et l'analyse familiale. Appliquée avec plusieurs marqueurs simultanément, cette méthode a permis la cartographie fine d'un gène majeur affectant la quantité de matière grasse du lait chez les bovins dans un intervalle de confiance de 3 cM, identifié ensuite comme le gène DGAT1.

Meuwissen et Goddard (2000, 2001) ont proposé la méthode LDLA combinant également liaison et déséquilibre de liaison. Contrairement aux autres, cette méthode ne suppose pas un polymorphisme causal unique mais un nombre quelconque d'allèles au QTL. Elle utilise les ingrédients bien connus des généticiens quantitatifs, le modèle linéaire mixte, le BLUP et l'estimation des composantes de variance des effets aléatoires par Maximum de Vraisemblance Restreinte (REML). La présence d'un QTL se traduit par une estimation de variance du QTL significativement non nulle. Pour chaque

QTL supposé additif, le modèle inclut deux effets gamétiques par individu, l'un correspondant à la copie reçue du père, l'autre de la mère. Les effets gamétiques dans la population sont supposés aléatoires et corrélés entre eux, pour prendre en compte le fait qu'ils dérivent les uns des autres entre parents et descendants. La corrélation entre effets est supposée égale à leur probabilité d'identité par descendance (IBD). Les probabilités IBD des QTL sont estimées à partir de l'information moléculaire disponible, d'une part à l'intérieur du pedigree connu, d'autre part entre les fondateurs (individus sans parents connus). A l'intérieur du pedigree, les probabilités sont déduites des probabilités de transmission conditionnellement à l'information moléculaire (Wang *et al* 1995, Pong-Wong *et al* 2001). L'apport essentiel de Meuwissen et Goddard (2000, 2001) est le développement de deux approches, l'une numérique, l'autre déterministe, pour prédire la probabilité IBD entre QTL des fondateurs. Ces approches permettent d'estimer la probabilité pour deux haplotypes d'être IBD, conditionnellement à l'identité par état des différents marqueurs qui définissent l'haplotype. La probabilité d'IBD estimée est celle d'un locus supposé être au milieu de l'haplotype ; elle dépend de la taille et de l'âge de la population (supposés connus), des fréquences alléliques et du taux de recombinaison entre les marqueurs. Les probabilités sont donc tout d'abord calculées pour les deux marqueurs flankant cette position, puis les marqueurs adjacents sont intégrés de proche en proche en tenant compte des taux de recombinaison entre eux.

Ces probabilités d'IBD sont introduites dans un modèle linéaire mixte *via* la matrice de variance-covariance des effets des haplotypes, qui est ici la matrice des probabilités d'IBD des haplotypes multipliée par la variance due au QTL, à estimer. La vraisemblance de la présence d'un QTL est calculée pour chaque position centrale d'haplotype, et elle est comparée à celle d'absence de QTL à la position testée (donc une variance due au QTL nulle), par un test de rapport de vraisemblances. La position maximisant ce test le long d'un chromosome est considérée comme position la plus probable du QTL. La première application de cette méthode concerne la localisation d'un QTL de gémellité dans des populations bovines, avec un intervalle de localisation inférieur à 1 cM (Meuwissen *et al* 2002). Cette méthode, souple et extensible à

plusieurs QTL, qui prend en compte l'information de plusieurs marqueurs simultanément de façon rigoureuse, tend à devenir la référence dans de nombreuses études animales. Elle présente cependant la limitation de l'hypothèse d'un QTL à effet additif.

## Conclusion

La cartographie fine est un enjeu majeur pour les années qui viennent. De son efficacité dépend la réussite des nombreux programmes de caractérisation de QTL, tant chez les animaux que les végétaux ou l'homme. La disponibilité de marqueurs nombreux et peu coûteux est une avancée considérable et on peut s'attendre à une multiplication rapide du nombre de polymorphismes causaux découverts. Dans tous les cas, cependant, le nombre de recombinaisons est l'élément essentiel de réussite, de sorte que des dispositifs de grande taille sont nécessaires, surtout quand les QTL n'expliquent qu'une fraction modérée de la variabilité du caractère. Des dispositifs variés ont été envisagés, en particulier dans les espèces qui se prêtent à des effectifs importants et des croisements faciles. D'autres sont sans doute encore à définir, cherchant à tirer parti du déséquilibre de liaison qui est un concept essentiel pour la cartographie fine de QTL mais aussi l'analyse de l'histoire des populations ou la recherche de signatures de sélection. De nombreuses méthodes sont aujourd'hui disponibles. La plupart d'entre elles, en particulier les plus efficaces, reposent sur des hypothèses plus ou moins fortes sur l'histoire de la population et la nature du QTL, dont on sous-estime probablement l'influence. Des logiciels sont fréquemment disponibles, rendant l'utilisation de ces méthodes plus abordable, bien que certaines méthodes soient difficiles à étendre à de grands pedigrees complexes.

La cartographie fine est un outil et n'est pas une fin en soi. Elle donne accès à un intervalle réduit contenant le polymorphisme causal. Le déséquilibre de liaison existant entre le polymorphisme causal et les marqueurs les plus proches autorise des applications en sélection, à la fois simples et efficaces, sans forcément connaître ce polymorphisme causal. La recherche de ce polymorphisme est cependant un objectif essentiel, pour comprendre le mécanisme et les réseaux géniques impliqués, pour mettre en œuvre une sélection plus facile, mais aussi pour

une meilleure diffusion et protection des résultats. L'identification de ce polymorphisme causal et surtout sa validation reposent alors sur des outils autres comme l'identification du gène

impliqué à partir d'arguments fonctionnels, la recherche exhaustive des polymorphismes de la région, l'identification des polymorphismes en association statistique complète avec le

génotype, le génotypage à grande échelle du polymorphisme supposé pour vérifier sa généralité, ou la recherche d'arguments fonctionnels validant son effet.

## Références

- Abdallah J.M., Mangin B., Goffinet B., Cierco-Ayrolles C., Perez-Enciso M., 2004. A comparison between methods for linkage disequilibrium fine mapping of quantitative trait loci. *Genet. Res.*, 83, 41-47.
- Boichard D., Le Roy P., Levéziel H., Elsen J.M., 1998. Utilisation des marqueurs moléculaires en génétique animale. *INRA Prod. Anim.*, 11, 67-80.
- Boitard S., Abdallah J., Rochambeau H. de, Cierco-Ayrolles C., Mangin B., 2006. Linkage disequilibrium interval mapping of quantitative trait loci. *BMC Genomics*, 7, 54-67.
- Burgio G., Szatanik M., Guenet J.L., Arnau M.R., Panthier J.J., Montagutelli X., 2007. Interspecific recombinant congenic strains between C57BL/6 and mice of the *Mus spretus* species: a powerful tool to dissect genetic control of complex traits. *Genetics*, 177, 2321-2333.
- Butte A., 2002. The use and analysis of microarray data. *Nature reviews*, 1, 951-960.
- Darvasi A., 1996. Interval-specific congenic strains (ISCS): an experimental design for mapping a QTL into a 1-centimorgan interval. *Mamm. Genet.*, 8, 163-167.
- Darvasi A., 1998. Experimental strategies for the genetic dissection of complex traits in animal models. *Nature Genet.*, 18, 19-24.
- Darvasi A., Soller M., 1992. Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theor. Appl. Genet.*, 85, 353-359.
- Darvasi A., Soller M., 1995. Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics*, 141, 1199-1207.
- Darvasi A., Soller M., 1997. A simple method to calculate resolving power and confidence interval of QTL map location. *Behav. Genet.*, 27, 125-132.
- Darvasi A., Weinreb A., Minke V., Weller J.I., Soller M., 1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics*, 134, 943-951.
- Falconer D.S., Mackay T.F.C., 1996. *Introduction to quantitative genetics*. 4th Ed. Longman Sci. and Tech., Harlow, UK, 464p.
- Farnir F., Grisart B., Coppieters W., Riquet J., Berzi P., Cambisano N., Karim L., Mni M., Moisisio S., Simon P., Wagenaar D., Vilkkki J., Georges M., 2002. Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. *Genetics*, 161, 275-287.
- Fredriksson R., Schiöth H.B., 2005. The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Mol. Pharmacol.*, 67, 1414-1425.
- Gilbert H., Le Roy P., 2003. Comparison of three multitrait methods. *Genet. Sel. Evol.*, 35, 281-304.
- Hästbacka J., de la Chapelle A., Kaitila I., Sistonen P., Weaver A., Lander E., 1992. Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genet.*, 2, 204-211.
- Karlsson E.K., Baranowska I., Wade C.M., Salmon Hillbertz N.H., Zody M.C., Anderson N., Biagi T.M., Patterson N., Pielberg G.R., Kulbokas E.J., Comstock K.E., Keller E.T., Mesirov J.P., von Euler H., Kämpe O., Hedhammar A., Lander E.S., Andersson G., Andersson L., Lindblad-Toh K., 2007. Efficient mapping of mendelian traits in dogs through genome-wide association. *Nature Genet.*, 39, 1321-1328.
- Korol A.B., Ronin Y.I., Kirzhner V.M., 1995. Interval mapping of quantitative trait loci employing correlated complex traits. *Genetics*, 140, 1137-1147.
- L'Hôte D., Serres C., Laissue P., Oulmouden A., Rogel-Gaillard C., Montagutelli X., Vaiman D., 2007. Centimorgan-range one-step mapping of fertility traits using interspecific recombinant congenic mice. *Genetics*, 176, 1907-1921.
- Lander E.S., Botstein D., 1987. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science*, 236, 1567-1570.
- Lander E.S., Botstein D., 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 121, 185-199.
- Lebowitz R.J., Soller M., Beckmann J.S., 1987. Trait-based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. *Theor. Appl. Genet.*, 73, 556-562.
- Loudet O., Saliba-Colombani V., Camilleri C., Calenge F., Gaudon V., Koprivova A., North K.A., Kopriva S., Daniel-Vedele F., 2007. Natural variation for sulfate content in *Arabidopsis* is highly controlled by APR2. *Nature Genet.*, 39, 896-900.
- Lynch M., Walsh B., 1998. *Genetics and analysis of quantitative genetics*. Sinauer Associates, Inc. Publishers. Sunderland, Massachusetts, USA, 980p.
- Meuwissen T.H.E., Goddard M.E., 2000. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics*, 155, 421-430.
- Meuwissen T.H.E., Goddard M.E., 2001. Prediction of identity by descent probabilities from marker-haplotypes. *Genet. Sel. Evol.*, 33, 605-634.
- Meuwissen T.H.E., Karlsen A., Lien S., Olsaker I., Goddard M.E., 2002. Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics*, 161, 373-379.
- Milan D., Woloszyn N., Yerle M., Le Roy P., Bonnet M., Riquet J., Lahbib-Mansais Y., Caritez J.C., Robic A., Sellier P., Elsen J.M., Gellin J., 1996. Accurate mapping of the «acid meat» RN gene on genetic and physical maps of pig chromosome 15. *Mamm. Genet.*, 7, 47-51.
- Niemann-Sorensen A., Robertson A., 1961. The association between blood groups and several production characteristics in three Danish cattle breeds. *Acta Agric. Scand.*, 11, 163-196.
- Ollivier L., Sellier P., Monin G., 1975. Déterminisme génétique du syndrome d'hyperthermie maligne chez le porc Piétrain. *Ann. Génét. Sél. Anim.*, 7, 159-166.
- Pong-Wong R., George A.W., Woolliams J.A., Haley C.S., 2001. A simple and rapid method for calculating identity-by-descent matrices using multiple markers. *Genet. Sel. Evol.*, 33, 453-471.
- Schadt E.E., Lum P.Y., 2006. Reverse engineering gene networks to identify key drivers of complex disease phenotypes. *J. Lipid Res.*, 47, 2601-2613.
- Schadt E.E., Monks S.A., Drake T.A., Lusis A.J., Che N., Colinayo V., Ruff T.G., Milligan S.B., Lamb J.R., Cavet G., Linsley P.S., Mao M., Stoughton R.B., Friend S.H., 2003. Genetics of gene expression surveyed in maize, mouse and man. *Nature*, 422, 297-302.
- 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.
- Spielman R.S., McGinnis R.E., Ewens W.J., 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.*, 52, 506-516.
- Terwilliger J.D., 1995. A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *Am. J. Hum. Genet.*, 56, 777-787.
- Tribout T., Iannuccelli N., Druet T., Gilbert H., Riquet J., Gueblez R., Mercat M.J., Bidanel J.P., Milan D., Le Roy P., 2008. Detection of quantitative trait loci for reproduction and production traits in Large White and French Landrace pig populations. *Genet. Sel. Evol.*, 40, 61-78.
- Van der Beek S., van Arendonk J.A.M., Groen A.F., 1995. Power of two- and three-generation QTL mapping experiments in an outbred population containing full-sib or half-sib families. *Theor. Appl. Genet.*, 91, 1115-1124.
- Wang T., Fernando R.L., van der Beek S., Grossman M., van Arendonk J.A.M., 1995. Covariance between relatives for a marked quantitative trait locus. *Genet. Sel. Evol.*, 27, 251-274.



Weller J.I., Wiggans G.R., VanRaden P.M., Ron M., 1996. Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multi-trait experiment. *Theor. Appl. Genet.* 92, 998-1002.

Weller J.I., 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.

Zöllner S., Pritchard J.K., 2004. Coalescent-based association mapping and fine-mapping of complex trait loci. *Genetics*, 169, 1071-1092.

---

## Résumé

De nombreuses méthodes ont été proposées pour cartographier finement des locus impliqués dans la variabilité des caractères quantitatifs (*Quantitative Trait Loci* ou QTL). La diversité de ces méthodes s'explique notamment par les différences qui existent entre les espèces, par exemple du fait des intervalles entre générations, des coûts de production, ou de la variabilité des coûts de phénotypage des caractères. Nous proposons ici une revue des méthodes permettant d'accéder à une cartographie fine de ces locus, décrivant au préalable les conditions d'une réduction de l'intervalle de localisation de ces QTL. Le paramètre clé permettant de réduire cet intervalle est le nombre de recombinaisons utiles. C'est pourquoi nous décrivons, dans les parties suivantes, les méthodes de cartographie fine en distinguant celles qui nécessitent la procréation d'individus ou de générations supplémentaires pour augmenter le nombre de recombinaisons, de celles qui ont recours à des modèles statistiques pour inférer des recombinaisons ayant eu lieu au cours des générations passées.

---

## Abstract

### *How to improve the localisation of a QTL?*

Many methods are offered to finely map loci implicated in the variability of quantitative traits (*Quantitative trait loci* or QTL). The diversity of these methods can be explained notably by the differences that exist between species, for example by generation intervals, production cost, or the variable costs for phenotyping traits. Here, we offer a review of the methods that exist for the fine mapping of these loci, but first we describe the conditions for the reduction of the localisation interval of these QTL. The key parameter that allows reducing the interval is the number of useful recombinations. This is why the fine mapping methods we describe are separated into two groups: those that require reproduction of individuals or additional generations to increase the number of recombinations and those that use statistical models to infer the recombinations having occurred during previous generations.

YTOURNEL F., GILBERT H., BOICHARD D., 2008. Comment affiner la localisation d'un QTL ? *INRA Prod. Anim.*, 21, 147-158.