



**HAL**  
open science

## Methodes mathematiques pour l'etude des genes controlant des caracteres quantitatifs

Bruno B. Goffinet, Jacques Beckmann, Didier Boichard, Mathilde M. Causse,  
Alain A. Charcosset, Claude C. Chevalet, Catherine Christophe, J Jacques J.  
J. Colleau, F. Demenais, Charles Eric Durel, et al.

### ► To cite this version:

Bruno B. Goffinet, Jacques Beckmann, Didier Boichard, Mathilde M. Causse, Alain A. Charcosset, et al.. Methodes mathematiques pour l'etude des genes controlant des caracteres quantitatifs. Colloque BRG/INRA. Ressources génétiques animales et végétales. Méthodologies d'étude et de gestion, Sep 1993, Montpellier, France. hal-02848407

**HAL Id: hal-02848407**

**<https://hal.inrae.fr/hal-02848407>**

Submitted on 7 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Article original

## Méthodes mathématiques pour l'étude des gènes contrôlant des caractères quantitatifs

B Goffinet<sup>1</sup>, J Beckmann, D Boichard, M Causse, A Charcosset,  
C Chevalet, C Christophe, JJ Colleau, F Demenais, CE Durel,  
JM Elsen, JL Foulley, A Gallais, KU Götz, F Hospital,  
A Kremer, M Lorieux, M Lefort-Buson, P Le Roy, P Loisel,  
B Mangin, A Maurice, X Perrier, O Pons, A Rebaï,  
F Rodolphe, M San Cristobal, J Vu Tien Khang  
(groupe MMM)

<sup>1</sup> *Institut national de la recherche agronomique,  
station de biométrie et intelligence artificielle,*

*Chemin de Borde-Rouge, Auzeville, BP 27, 31326 Castanet-Tolosan Cedex, France*

**Résumé** – Nous présentons ici les travaux s'inscrivant dans le cadre du projet «Mathématiques et marqueurs moléculaires», en abrégé «MMM». Ces travaux ont pour objet l'identification et l'utilisation de gènes contrôlant des caractères quantitatifs – ou QTL (*quantitative trait loci*) – à partir de l'information issue des phénotypes, de la généalogie et de marqueurs moléculaires. Une première partie concerne la mise en évidence de gènes à effet majeur et l'utilisation de ces gènes en sélection, à partir d'informations phénotypiques (mesures sur le caractère dont le contrôle génétique est recherché) et généalogiques (relations de parenté entre individus de la population mesurée). La deuxième partie concerne l'utilisation conjointe de ces informations phénotypiques et généalogiques avec celles issues du marquage moléculaire dans la détection, la localisation de QTL, et en amélioration génétique.

**cartographie / gène majeur / marqueur moléculaire / maximum de vraisemblance / locus de caractère quantitatif**

**Summary** – **Mathematical methods for studying genes controlling quantitative traits.** *We present in this paper the results of the project "Mathematics and Molecular Markers" (MMM). These results concern the identification and use of genes that control quantitative characters or QTL (quantitative trait loci), using phenotypic, genealogical and molecular marker information. The first section concerns identification of major genes using information of a phenotypic (measures of the character) and genealogical (relationships between*

*individuals) type and use of these genes in selection. The second section concerns the use of this information jointly with molecular marker information for detection and localisation of QTL, and for genetic improvement.*

**mapping / major gene / molecular marker / maximum likelihood / quantitative trait locus**

## INTRODUCTION

L'ensemble des travaux s'inscrivant dans le cadre du projet «MMM» a pour objet l'identification et l'utilisation de gènes (ou segments du génome) contrôlant des caractères quantitatifs. Ces gènes seront appelés par la suite QTL, pour *quantitative trait loci*, ou plus spécifiquement gène majeur, lorsque leur effet est suffisamment important pour être détecté à l'aide des seules informations phénotypiques. Les méthodes développées pour identifier et utiliser ces gènes font appel à des outils mathématiques complexes. C'est donc en grande partie sur ces aspects mathématiques que nous avons fait porter notre effort.

Ces travaux peuvent être structurés en 2 parties. La première concerne la mise en évidence de gènes à effet majeur et l'utilisation de ces gènes en sélection, à partir d'informations phénotypiques (mesures sur le caractère dont le contrôle génétique est recherché) et généalogiques (relations de parenté entre individus de la population mesurée). La deuxième partie concerne l'utilisation conjointe de ces informations phénotypiques et généalogiques avec celles issues du marquage moléculaire dans la détection, la localisation et l'utilisation de QTL.

## INFORMATION PHÉNOTYPIQUE ET GÉNÉALOGIQUE POUR LA MISE EN ÉVIDENCE DE GÈNES MAJEURS

### *Détection de gènes majeurs*

La méthode de référence pour le test de l'existence d'un gène majeur dans une population est la méthode du maximum de vraisemblance (appelée aussi analyse de ségrégation ou analyse de pedigree) dans laquelle on compare la vraisemblance des observations (ou probabilité d'obtenir ces observations) sous les hypothèses de présence ou d'absence d'un gène majeur. En effet, il s'agit d'une méthode puissante, et qui peut être définie pour toutes les situations expérimentales. Cependant, son emploi pose au moins 3 types de problèmes :

- la détermination des distributions des statistiques de test, élément essentiel pour maîtriser l'erreur de première espèce et connaître la puissance des tests, ne peut pas se faire à partir des théories asymptotiques classiques;
- la complexité des calculs qu'elle met en œuvre et l'absence de logiciel facilement disponible, rendent son utilisation délicate;
- enfin, elle est peu robuste à la non-normalité des distributions, et peut conduire à des conclusions erronées, en particulier pour les caractères à distribution dissymétrique.

Classiquement, la détermination de la distribution des statistiques de test du rapport de vraisemblance se fait en utilisant les résultats généraux de la statistique asymptotique qui permettent par exemple d'affirmer la convergence du rapport de vraisemblance vers un  $\chi^2$  sous l'hypothèse nulle. Ici, les modèles s'inscrivent dans le cadre statistique des mélanges de distribution où les théories asymptotiques classiques ne s'appliquent pas. En effet les conditions de régularité ne sont pas satisfaites. On est confronté en particulier à la non-estimabilité de certains paramètres sous l'hypothèse nulle d'absence de gène, et au caractère non défini-positif des matrices d'information. La résolution de ces problèmes est nécessaire pour déterminer la distribution des statistiques de test afin de maîtriser l'erreur de première espèce et de connaître la puissance de ces tests.

Pons (1993) et Mangin *et al* (1993a) ont recherché le seuil du test du rapport de vraisemblance dans le cas de familles de demi-frères. Mangin *et al* (1993a) montrent que l'utilisation de la méthode de Davies (1977) permet d'avoir un niveau de test correct, alors que l'utilisation du niveau issu de la théorie asymptotique classique conduit à une erreur de première espèce trop importante, et donc à détecter trop de faux gènes majeurs.

Dans le cas de situations expérimentales simples comme le *back-cross* ou la  $F_2$ , les probabilités d'apparition des différents types d'individus, homozygotes et hétérozygotes, sont connues *a priori*. On est alors confronté au caractère non défini-positif de la matrice d'information. Pour obtenir la distribution des statistiques de test, Goffinet *et al* (1992) et Loisel *et al* (1993) ont été amenés à développer ces statistiques à des ordres très élevés (jusqu'à l'ordre 8 dans certains cas). Par exemple, la distribution du test de rapport de vraisemblance sous l'hypothèse nulle dans le cas de la  $F_2$  est le mélange d'un  $\chi^2$  à un degré de liberté avec une probabilité 1/2 et d'un  $\chi^2$  à 2 degrés de liberté avec une probabilité 1/2.

Vis-à-vis du problème lié à la quantité de calculs de la méthode du maximum de vraisemblance (Le Roy, 1992), un besoin existait de création et d'évaluation de tests statistiques plus simples et de mise en œuvre rapide, qui pourraient être utilisés pour une première approche, avec un rôle d'«indicateurs» justifiant ou non le passage à des analyses plus sophistiquées.

La comparaison de plusieurs critères simples, proposés dans la littérature ou à l'occasion de ce travail, a été réalisée par simulation par Le Roy et Elsen (1992a), pour diverses hypothèses sur les structures de populations et les types d'hérédité du caractère. Les conditions d'utilisation (seuils de rejet), la puissance et la robustesse de ces critères ont été mesurées. Deux statistiques sont apparues généralement plus efficaces que les autres : le test d'hétérogénéité des variances intrafamiliales de Bartlett et la méthode de régression moyenne-variance de Fain (1978). Ces critères simples, comme les plus complexes, ont été utilisés dans des analyses de cas (Duchet-Suchaux *et al*, 1992 ; Mandonnet *et al*, 1992 ; Le Roy *et al*, 1990) en introduction d'études plus approfondies. Dans le même esprit, des méthodes d'analyse de ségrégation approchée, qui avaient été développées antérieurement au projet MMM, ont été appliquées à de très grandes populations (Boichard *et al*, 1990).

Ces méthodes sont cependant très peu robustes à la non-normalité des distributions. Cette sensibilité à la non-normalité conduit à travailler sur les techniques non paramétriques (Lemdani, 1991) et sur la robustesse utilisant la différentiabilité

au sens de Fréchet. Notons cependant que ces méthodes ne sont adaptées qu'à de grands effectifs, et nécessitent des calculs très complexes.

### **Utilisation des gènes majeurs**

Notons tout d'abord que les travaux effectués dans ce cadre concernent le monde animal. Un gène majeur ayant été détecté, son utilisation nécessite de connaître le génotype de tout futur reproducteur, ce qui est rarement directement possible d'après son phénotype, parce que le gène peut avoir une expression limitée à un seul sexe ou nécessitant l'abattage de l'animal, ou parce qu'il y a recouvrement des distributions du caractère dans les différents génotypes au locus majeur.

Une situation fréquente lors de la fixation d'un gène majeur depuis une race où il a été détecté vers une race d'accueil est le testage sur descendance de mâles de génotype inconnu par accouplements de ces mâles avec des femelles homozygotes récessives et observation phénotypique de leurs descendance. Ce dispositif permet d'augmenter la précision du choix des reproducteurs (multiplication des mesures) mais aussi de trier des mâles sur un génotype ne s'exprimant que chez les femelles. Il est notamment employé pour l'introggression du gène Booroola (hyperovulation des brebis) dans des races ovines françaises.

En l'absence de marqueur, le problème se ramène donc essentiellement au test d'un mélange de distributions (la descendance d'un mâle AB accouplé à des femelles BB est composée pour moitié de AB et pour moitié de BB) contre une distribution unique (de génotype AA, le même mâle n'aurait eu que des descendants AB). Traité de façon plus exacte, le problème aboutit à la proposition d'un test d'hypothèses non emboîtées. Goffinet *et al* (1990) ont étudié algébriquement ce test afin de déterminer la distribution asymptotique de la statistique correspondante, et évalué la vitesse de convergence vers cette asymptote par simulation.

Des solutions approchées sont toutefois concevables et une étude numérique a montré que certaines d'entre elles sont tout à fait performantes (Elsen *et al*, 1991). En particulier, il est possible, à l'image du modèle unifié de Lalouel *et al* (1983), d'utiliser l'artifice des probabilités de transmission pour créer un test d'hypothèses emboîtées dont la distribution asymptotique est un  $\chi^2$ .

Au-delà du tri des animaux, les stratégies d'utilisation des gènes majeurs identifiés peuvent faire l'objet de réflexions et de modélisation en vue de leur optimisation. Les situations diffèrent selon les caractéristiques du gène en question (fréquence initiale, expression additive ou dominante...) (Mandonnet *et al*, 1992; Elsen, 1992). L'introggression d'un gène «exotique» peut ainsi prendre diverses formes différant par leur rapidité d'exécution et leur robustesse.

## **UTILISATION DE L'INFORMATION ISSUE DU MARQUAGE MOLÉCULAIRE**

La disponibilité croissante des marqueurs moléculaires permet d'envisager de rechercher des gènes dont l'effet est sensiblement moins important et de les positionner sur le génome. Il s'agit tout d'abord de cartographier ces marqueurs sur le génome. Ces marqueurs peuvent alors être utilisés de multiples façons : mise en évidence de QTL,

sélection assistée par marqueurs, aide à l'introgession de gènes identifiés statistiquement ou par l'analyse de liaison génétique, évaluation des distances génétiques entre races, gestion des ressources génétiques.

### **Cartographie génétique**

Dans le cadre végétal, méthodes et logiciels existent pour cartographier les marqueurs dans les schémas simples, rétrocroisement,  $F_2$  ou lignées recombinantes. Cependant, 2 types de situations n'ont pas encore reçu de réponse tout à fait satisfaisante, au moins au niveau des logiciels. Il s'agit des croisements concernant des arbres fruitiers ou forestiers car il ne s'agit pas de croisement entre lignées pures, et du problème des cartes consensus, dont l'objectif est d'utiliser simultanément l'information issue de plusieurs croisements. Mangin et Goffinet (1993) font le point sur le logiciel Join-Map, destiné à répondre à ces 2 questions, et concluent à son efficacité.

Plusieurs grands programmes de cartographie génétique ont débuté ou débutent pour les animaux domestiques. Des projets européens (PIGMAP et BOVMAP) ont été lancés pour les 2 espèces économiquement majeures que sont le porc et le bovin. En parallèle, des contributions à l'établissement des cartes génétiques des ovins, caprins, poules et chevaux sont également mises en place. Ces protocoles utilisent des «populations de référence» dont la taille et la structure généalogiques peuvent être optimisées en vue d'une efficacité maximum du travail de cartographie.

À l'occasion de la mise en place du programme Bovmap, une réflexion a été menée pour cette optimisation (Elsen *et al*, 1992 et 1994). Le critère de choix peut reposer sur un des objectifs importants assignés à ces cartes génétiques, qui est l'aide à la détection de QTL. Pour une même densité de marqueurs, la puissance de détection de QTL sera plus importante si ces marqueurs sont régulièrement répartis le long du génome, et donc choisis à partir d'une connaissance précise de leurs positions. Ces auteurs proposent ainsi de dimensionner la population de référence de telle sorte que la carte de marqueurs résultant de ce choix permette, pour une puissance donnée, de mettre en évidence des QTL avec le minimum d'effort (nombre d'individus à mesurer dans un protocole de détection de QTL utilisant la carte génétique en question).

La structure de la population de référence est déterminée par des choix portant sur le nombre de générations et de descendants par reproducteur et sur les animaux dont il faut déterminer les génotypes. En se basant sur un critère de variance minimale, Elsen *et al* (1993) montrent qu'il n'est pas indispensable de typer la génération grand-parentale, que les familles de pleins-frères sont de loin préférables aux familles de demi-frères et, qu'à nombre total d'observations donné, le typage des conjointes n'est pas plus efficace que le doublement du nombre de descendants par mâle.

### **Mise en évidence de QTL**

La référence à Neimann-Sørensen et Robertson (1961) et Sax (1923) relativise le caractère novateur de la «génétique» moderne. Ce qui est changé, ce n'est pas l'idée de trouver des zones du génome contrôlant des caractères quantitatifs, mais

la découverte de techniques de mise en évidence de polymorphismes de l'ADN qui rendent cette recherche pratique et probablement garantissent son succès.

De la même façon que dans les tests de détection de gènes majeurs, les méthodes de référence sont basées sur le maximum de vraisemblance, comme la cartographie par intervalle (*interval mapping*) de Lander et Botstein (1989) qui est développée pour des situations expérimentales simples (lignées consanguines, rétrocroisement,  $F_2$ ). On rencontre des difficultés du même type que pour la détection de gènes majeurs pour déterminer la distribution des statistiques de tests, avec en particulier l'apparition d'un paramètre de position sur le génome qui pose des problèmes très délicats.

De nombreuses méthodes alternatives sont développées dans les cadres animaux et végétaux de façon à pouvoir traiter des modèles plus complexes (plusieurs QTL par exemple), à permettre des calculs plus simples pour traiter des situations expérimentales complexes, et à offrir plus de robustesse aux hypothèses.

### Les problèmes statistiques généraux

Le paramètre de position du QTL sur le génome disparaît sous l'hypothèse nulle d'absence de QTL. Il n'est donc pas possible d'utiliser les résultats asymptotiques classiques. Rebaï *et al* (1993) proposent d'utiliser la méthode de Davies (1977) pour déterminer le seuil du test du rapport de vraisemblance et sa puissance. Les simulations montrent un bon accord avec les résultats théoriques. Ce travail permet de plus de montrer l'équivalence asymptotique du test de rapport de vraisemblance avec les tests linéarisés de Knapp *et al* (1990), qui sont plus simples à calculer et à généraliser à des situations complexes : multi-QTL, croisement diallele, etc.

D'autre part, on est confronté au caractère non défini-positif de la matrice d'information, comme dans les tests de détection de gène majeur, lorsque l'on introduit des paramètres de dominance du QTL avec des marqueurs dominants. Mangin *et al* (1993b) obtiennent la distribution des statistiques sous l'hypothèse nulle et sous des alternatives locales, en utilisant des développements de la vraisemblance à des ordres très élevés.

### Les situations animales

Les méthodes statistiques les plus développées pour mettre en évidence ces QTL sont basées sur le test, dans un cadre linéaire, des effets des allèles marqueurs intrapère ou grand-père (Soller et Genizi, 1978; Weller *et al*, 1990). Foulley (1991) propose une formule alternative de calcul direct de la puissance de détection d'un QTL associé à un marqueur, pour les dispositifs de contrôles de descendance de mâles des première (Soller et Genizi, 1978), ou de deuxième génération (Weller *et al*, 1990). Cette formule repose sur une approximation asymptotique (grand nombre d'allèles au QTL) et suppose, comme le font Fernando et Grossman (1989), que l'effet du gène au QTL transmis par le père chez un de ses descendants quelconque soit une variable aléatoire.

En se limitant au test d'une simple hérédité monogénique liée au marqueur disponible, Le Roy et Elsen (1992, 1993) évaluent le gain de puissance apporté par l'utilisation d'un test de maximum de vraisemblance qui exploite globalement les

différences de forme de distribution du caractère quantitatif chez les descendants d'un père doublement hétérozygote (au locus marqueur et au locus quantitatif) selon l'allèle marqueur reçu du père, et non seulement les différences dans l'espérance de ces distributions. Ce gain peut être important dans le cas de QTL à faible effet, faiblement liés au locus marqueur et pour des tailles limitées de population (gain de 50% pour un QTL dominant dont l'effet est de 0,2 écart type, et qui est situé à 10% de recombinaison du marqueur, en utilisant une population de 5 grand-pères, 100 pères/grand-père et 100 descendants/père). Cette étude montre également que, dans ces situations où un seul QTL contrôle le caractère, le gain apporté par les marqueurs par rapport à l'analyse de ségrégation classique est faible. Cette étude est maintenant prolongée par un travail équivalent dans le cas d'une hérédité mixte combinant un QTL lié au marqueur étudié et un ensemble d'autres gènes indépendants.

En pratique, la recherche de QTL se fera à partir d'une carte de marqueurs. Or, pour un même individu, un locus marqueur peut ne pas être informatif quand des locus voisins le sont. Les informations apportées par tous les marqueurs présents dans une zone du génome doivent donc être simultanément utilisées. Nous nous sommes placés dans le cas particulier du croisement  $F_1$  et  $F_2$  entre 2 lignées extrêmes du point de vue quantitatif (QTL fixés à l'état homozygote dans les races pures), mais où les allèles aux locus marqueurs sont en ségrégation. Il s'agit vraisemblablement de la situation qui prévaudra pour la détection de QTL chez le porc à partir de croisements entre une race européenne (Large White) et une race chinoise (Meishan). Dans ce contexte, Haley *et al* (1994) ont développé un algorithme exact dans le cas de marqueurs codominants, sans données manquantes, pour le traitement simultané de tous les marqueurs d'un même groupe de liaison. Son évaluation par simulation indique que cette technique est plus puissante et robuste que l'analyse uni marqueur.

Il est intéressant par ailleurs de rechercher des méthodes plus simples à calculer et plus robustes aux hypothèses de normalité des distributions. La méthode de Haseman et Elston (1972) HE basée sur les couples de germains, offre ces avantages (Blackwelder et Elston, 1982), mais est en contrepartie peu puissante. Avec des marqueurs hautement polymorphes, l'utilisation exclusive d'accouplements totalement informatifs peut être envisagée. Götz et Ollivier (1992) ont étudié, à l'aide de simulations, l'effet d'un système génétique hautement polymorphe (5 allèles également fréquents), d'une grande taille de fratrie (6 germains) et d'une structure d'élevage polygynique (25 femelles accouplées à chaque mâle) sur la puissance du test HE. Les résultats montrent que la puissance est considérablement accrue quand on dispose à la fois de marqueurs hautement polymorphes et de grandes fratries. Avec un nombre total d'individus typés de 5000, et pour un locus expliquant 8% de la variance phénotypique du caractère, la puissance du test au niveau 5% est de 0,74 au lieu de 0,14 quand on passe de 2 à 5 allèles au locus marqueur, et de 2 à 6 frères par fratrie.

### Les situations végétales

La plupart des méthodes de recherche de QTL (Lander et Botstein, 1989; Knapp *et al*, 1990) sont fondées sur l'hypothèse que ceux-ci sont bien séparés les uns des autres, et sont donc rares. Nous avons considéré, au contraire, le génome



dans son ensemble et recherché des régions chromosomiques impliquées dans la variation de caractères quantitatifs entre 2 lignées pures et leurs produits de croisement. Rodolphe et Lefort (1993) ont développé une méthode adaptée aux  $F_2$ , rétrocroisements et haploïdes doublés. Les marqueurs doivent être codominants et les données manquantes sont admises. Cette méthode, qui ne repose pas sur des hypothèses génétiques fortes, doit être utilisée avec des densités de marquage raisonnables, mais présente l'avantage de prendre en compte simultanément l'effet de tous les QTL. On doit donc en attendre une réduction de la variance résiduelle dans le cas où coexistent plusieurs QTL. Une perspective intéressante serait d'analyser les données avec cette méthode, en conservant les marqueurs sur tous les groupes de liaison sauf un, puis analyser les résidus avec une méthode séquentielle en conservant cette fois tous les marqueurs disponibles sur le dernier groupe de liaison seulement.

Ces méthodes sont adaptées aux situations expérimentales les plus classiques. Si l'on cherche à étudier la stabilité d'un QTL dans différents fonds génétiques, il est intéressant d'utiliser un croisement diallèle avec plusieurs lignées. Rebaï et Goffinet (1993) ont étudié la puissance de différents tests de détection de QTL dans ce cadre, pour des observations faites sur des topcross ou  $F_3$ . Il apparaît que les tests qui utilisent l'information conjointe des différents croisements du diallèle sont sensiblement plus puissants que les tests classiques, et qu'il est préférable de ne pas intégrer de terme de dominance dans le modèle. Ce travail se poursuit par la mise au point d'une méthode de cartographie par intervalle dans ce cadre.

### ***Utilisation des marqueurs pour l'amélioration génétique***

La sélection assistée par marqueurs peut être basée sur l'utilisation d'un déséquilibre de liaison entre ces marqueurs et des QTL (Lande, Thompson, 1990), ou sur le suivi généalogique des génotypes des candidats à la reproduction (Le Roy, 1992, Elsen, 1992, Le Roy et Elsen, 1992, Boichard *et al*, 1993). Dans le premier cas il est à craindre que le déséquilibre s'estompe au cours des générations, après son apparition à la suite d'un croisement ou d'une mutation, et ceci d'autant plus que la liaison marqueur-QTL est faible. Dans le second cas il convient de mettre en place un suivi précis des performances et des génotypes marqueurs des individus de la population afin d'estimer les génotypes au(x) QTL(s) et de piéger le cas échéant les événements de recombinaison.

L'apport des marqueurs pour l'introgession des gènes est beaucoup plus immédiat. D'une part, il permet d'identifier rapidement, dès la naissance, sur n'importe quel individu, mâle ou femelle, son génotype au locus faisant l'objet de l'introgession (Hospital et Elsen, 1992). D'autre part, il permet d'accélérer le processus de restauration du génome de la race ou de la variété receveuse en utilisant l'information apportée par les génotypes en plusieurs marqueurs disposés le long du génome, les individus retenus étant ceux qui ont la proportion maximum d'allèles spécifiques de la race ou de la variété receveuse. Hospital *et al* (1992) ont étudié les effets de la durée et de l'intensité de la sélection, de la taille de la population, du nombre et de la position des marqueurs sélectionnés sur les chromosomes portant ou ne portant pas le gène introgressé. Ils montrent que l'utilisation de marqueurs permet un gain d'à peu près 2 générations, ce qui est plus faible que les valeurs

données par Hillel *et al* (1990). Les marqueurs sont plus utiles lorsque leur position est connue. Ils montrent que dans les premières générations, il ne sert à rien d'utiliser plus de 3 marqueurs pour chaque chromosome non porteur. Concernant le chromosome porteur, il apparaît qu'il est préférable de typer relativement peu d'individus pour de nombreux marqueurs plutôt que de nombreux individus pour quelques marqueurs proches du gène. Ce travail fournit les bases de raisonnement nécessaires à la définition d'une stratégie optimale de sélection adaptée à chaque cas particulier.

Enfin, les marqueurs permettent de définir des indices de distance génétique entre lignées. Charcosset *et al* (1991) ont étudié la corrélation  $\rho$  entre un index de ce type défini pour 2 lignées parentales et l'hétérosis observé dans la  $F_1$  issue de ces lignées, pour un caractère quantitatif. Les calculs théoriques de cette corrélation sont faits en utilisant un modèle bi-allélique. Lorsque les allèles au locus marqueur et au QTL sont également distribués à l'intérieur de la population des lignées considérées, le coefficient  $\rho$  est une fonction des carrés des déséquilibres de liaison entre les allèles au locus marqueur et les allèles au QTL. Les QTL qui ne sont pas marqués par un locus marqueur, et les marqueurs qui ne sont pas liés à un QTL, jouent un rôle symétrique et peuvent faire décroître  $\rho$  de manière importante. En conclusion, la prédiction de l'hétérosis d'un hybride  $F_1$  à partir de marqueurs serait plus efficace si ces marqueurs étaient sélectionnés sur leurs liaisons avec les QTL. Ce même type de modèle est utilisé par Charcosset et Essioux (1993) pour étudier l'effet sur le coefficient  $\rho$  de la structuration des lignées en groupes hétérotiques. Il apparaît que l'on trouvera des valeurs significatives de  $\rho$  si les lignées appartiennent à un même groupe hétérotique, mais que l'on ne trouvera aucune relation pour des hybrides entre lignées appartenant à des groupes différents, car le déséquilibre de liaison varie de manière aléatoire d'un groupe à l'autre. Ces résultats sont cohérents avec les résultats expérimentaux et soulignent la nécessité de mettre au point de nouveaux modèles de prédiction pour mieux valoriser l'information.

## *Perspectives*

La disponibilité croissante de marqueurs moléculaires ne fait pas disparaître l'intérêt de détecter et d'utiliser des gènes majeurs à partir des seules informations phénotypiques et généalogiques. Il existe en effet un grand nombre de données phénotypiques non encore exploitées et le marquage reste relativement coûteux. Il reste donc intéressant de travailler sur les aspects méthodologiques du traitement de telles données, qui en plus de l'intérêt propre des méthodes, sont souvent un passage obligé avant de traiter de problèmes plus compliqués impliquant des marqueurs. Les aspects prioritaires semblent être le traitement de pedigrees complexes et la recherche de robustesse.

Les problèmes méthodologiques concernant l'utilisation des marqueurs restent très importants et de nouveaux problèmes voient le jour au fur et à mesure que se développent des nouvelles pistes d'utilisation des marqueurs en sélection :

- cartographie : il faut développer des algorithmes de calcul efficaces pour les pedigrees animaux et proposer des algorithmes et des tests pour déterminer l'ordre le plus vraisemblable entre les marqueurs;

- recherche de QTL : il reste encore de nombreux problèmes statistiques non parfaitement résolus, comme l'intervalle de confiance de la position des QTL ou le test de l'existence de plusieurs QTL ;
- par ailleurs il faut développer des algorithmes et des logiciels pour traiter des situations animales; la participation à un réseau européen en cours de constitution devrait permettre d'accélérer les travaux dans cette direction ;
- utilisation en sélection : beaucoup de choses restent à faire dans la mise au point de méthodes et critères de sélection assistée par marqueurs et de construction de génotypes associant un ensemble de QTL favorables.

## REMERCIEMENTS

Nous tenons à remercier l'AIP INRA Agrotech-Prodige, qui est à l'origine de la création du groupe MMM et qui en a assuré le financement de mai 1990 à mai 1993.

## RÉFÉRENCES

- Blackwelder WC, Elston RC (1982) Power and robustness of sib-pair linkage tests and extension to larger sibships. *Commun Statist-Theor Methods* 11, 449-484
- Boichard D, Chevalet C, Le Roy P, Elsen JM (1993) Les perspectives d'utilisation des marqueurs génétiques chez les animaux domestiques. *Réunion de l'AFZ*. Nantes, 25-26 mai 1993
- Boichard D, Elsen JM, Le Roy P, Bonaiti B (1990) Segregation analysis of fat content data in Holstein x European Friesian crossbred cattle. *In: Proc Fourth World Congr Genet Appl Livest Prod* (Hill, Thompson, Wooliams, eds), XIV, 167-170
- Charcosset A, Essioux L (1994) The effect of germplasm structuration on the relationship between heterosis and heterozygosity at marker loci. *Theor Appl Genet* (à paraître)
- Charcosset A, Lefort M, Gallais A (1991) Relationship between heterosis and heterozygosity at marker loci: a theoretical computation. *Theor Appl Genet* 81, 571-575
- Davies RB (1977) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247-254
- Duchet-Suchaux M, Menateau P, Le Roux H, Elsen JM, Lechopier P (1992) Genetic control of resistance to enterotoxigenic *Escherichia coli* in infant mice. *Microb Pathogen* 13, 157-160
- Elsen JM (1992) What should be expected from major genes for reproduction in sheep? *Symposium "Perspektiven der Nutztierwissenschaften"*, Anlässlich des 65. Geburtstages von Prof Dr R Wassmuth, Giessen, 9 März 1993
- Elsen JM, Le Roy P, Goffinet B (1991) Comparison of four statistical tests for genotype determination at a major locus of progeny tested sires. *J Anim Breed Genet* 108, 167-173
- Elsen JM, Mangin B, Goffinet B, Chevalet C (1992) Protocol designs for building genetic linkage map in livestock. 43<sup>e</sup> Réunion annuelle de la FEZ, *Commission de génétique animale*, 14-18 septembre 1992, Madrid, Abstract : vol 1, 68-69 (GII.10)

- Elsen JM, Mangin B, Goffinet B, Chevalet C (1994) Optimal structure of protocol designs for building genetic linkage map in livestock. *Theor Appl Genet* (À paraître)
- Fain PR (1978) Characteristics of simple sibship variance tests for the detection of major loci and application to height, weight and spatial performance. *Ann Hum Genet* 42, 109-120
- Fernando R, Grossman M (1989) Best linear unbiased prediction and marker assisted selection. *Genet Sel Evol* 21, 467-477
- Foulley JL (1991) Exposé critique des dispositifs «filles» et «petites filles» proposés par Weller *et al*, 1990, pour la mise en évidence des QTL. *Rapport interne INRA*, Domaine de Vilvert, 78352 Jouy-en-Josas Cedex
- Goffinet B, Elsen JM, Le Roy P (1990) Statistical tests for identification of the genotype at a major locus of progeny tested sires. *Biometrics* 46, 583-594
- Goffinet B, Loisel P, Laurent B (1992) Testing in normal mixture models when the proportion are known. *Biometrika* 79, 842-846
- Götz KU, Ollivier L (1992) Theoretical aspects of applying sib-pair linkage tests to livestock species. *Genet Sel Evol* 24, 29-42
- Haley CS, Knott SA, Elsen JM (1994) Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* (à paraître)
- Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 1, 3-19
- Hillel J, Schaap T, Haberfeld A *et al* (1990) DNA fingerprint applied to gene introgression breeding programs. *Genetics* 124, 783-789
- Hospital F, Chevalet C, Mulsant P (1992) Using markers in gene introgression breeding programs. *Genetics* 132, 1199-1210
- Hospital F, Elsen JM (1992) Introgression génique assistée par marqueurs. *INRA Prod Anim Hors série «Éléments de génétique quantitative et application aux populations animales»*, 299-302
- Knapp SJ, Bridges WC, Birkes D (1990) Mapping quantitative trait loci using molecular marker linkage maps. *Theor Appl Genet* 79, 583-592
- Lalouel JM, Rao DC, Morton NE, Elston RC (1983) A unified model for complex segregation analysis. *Am J Hum Genet* 35, 816-826
- Lande R, Thompson R (1990) Efficiency of marker assisted selection in the improvement of quantitative traits. *Genetics* 124, 743-756
- Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185-199
- Lemdani M (1991) Estimation du mode d'une densité unimodale par un critère de minimum de distance. DEA de statistiques et modèles aléatoires, Orsay
- Le Roy P (1992) Les méthodes de mise en évidence des gènes majeurs. *INRA Prod Anim Hors série «Éléments de génétique quantitative et application aux populations animales»*, 93-99
- Le Roy P, Elsen JM (1992a) Simple test statistics for major gene detection: a numerical comparison. *Theor Appl Genet* 83, 635-644
- Le Roy P, Elsen JM (1992b) Numerical comparison between maximum likelihood and analysis of variance methods for QTL detection in progeny test designs. 43<sup>e</sup> Réunion annuelle de la FEZ, 14-17 septembre 1992, Madrid

- Le Roy P, Elsen JM (1993) Numerical comparison between maximum likelihood and analysis of variance methods for QTL detection in progeny test designs. Monogenic inheritance. *Theor Appl Genet* (à paraître)
- Le Roy P, Naveau J, Elsen JM, Sellier P (1990) Evidence for a new major gene influencing meat quality in pigs. *Genet Res Camb* 55, 33-40
- Loisel P, Goffinet B, Monod H, Montes de Occa G (1993) Detecting major gene in an F2 population. *Biometrics* (à paraître)
- Mandonnet N, Le Roy P, Caritez JC, Elsen JM, Legault C, Bidanel JP (1992) Existe-t-il un gène à effet majeur expliquant la prolificité exceptionnelle de la race Meishan? 43<sup>e</sup> réunion annuelle de la FEZ, 14-18 septembre 1992, Madrid, *Commission de génétique animale*, Abstract : vol 1, 126-127
- Mangin B, Goffinet B (1993) Comparaison de logiciels pour l'obtention de cartes consensus. Rapport interne INRA, Auzeville, 31326 Castanet-Tolosan
- Mangin B, Goffinet B, Elsen JM (1993a) Testing in normal mixture models with some information on the parameter. *Biometrical J* (à paraître)
- Mangin B, Goffinet B, Rebaï A (1993b) QTL detection and dominant markers. *Rapport interne*, INRA, Auzeville, 31326 Castanet-Tolosan
- Neimann-Sorensen A, Robertson A (1961) The association between blood groups and several production characteristics in three Danish cattle breeds. *Acta Agr Scand* 11, 163-196
- Pons O (1993) Homogeneity test in distribution mixtures and application to major genes detection. *Statistics* 24, 149-160
- Rebaï A, Goffinet B (1993) Power of different tests for QTL detection using progenies derived from a diallel cross. *Theor Appl Genet* 86, 1014-1022
- Rebaï A, Goffinet B, Mangin B (1993) Comparing power of different methods for QTL detection. *Biometrics* (à paraître)
- Rodolphe F, Lefort M (1993) A multi-marker model for detecting chromosomal segments displaying QTL activity. *Genetics* 134, 1277-1288
- Sax K (1923) Association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8, 552-560
- 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, 7-55
- Weller JL, 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