

Outils statistiques et informatiques pour l'analyse du déséquilibre de liaison

Jihad ABDALLAH^{(1)*}, Brigitte MANGIN⁽²⁾,
Christine CIERCO-AYROLLES⁽²⁾, Bruno GOFFINET⁽²⁾,
Hubert de ROCHAMBEAU⁽¹⁾, Miguel PÉREZ-ENCISO^(1, 3)

⁽¹⁾ INRA, Station d'Amélioration Génétique des Animaux, BP 27,
31326 Castanet-Tolosan, France

⁽²⁾ INRA, Unité de Biométrie, BP 27, 31326 Castanet-Tolosan, France

⁽³⁾ Adresse actuelle : Institut Català de Recerca i Estudis Avançats, Facultat de
Veterinària, Universitat Autònoma de Barcelona, 08193 Bellaterra, Espagne

Abstract: Statistical tools for the analysis of Linkage Disequilibrium. Traditionally linkage analysis has been used in gene mapping. The resolution of this approach is limited by the low number of recombination events in the data. Recently linkage disequilibrium (the non random association between alleles at different loci) has been advocated as a fine mapping tool. The advantage of linkage disequilibrium approach over linkage analysis is its ability to incorporate the large number of historical recombination events to provide a fine-scale gene location. Despite the success of this approach in mapping simply inherited disease genes, its usefulness in localizing quantitative trait loci and complex disease genes has not yet been determined. The main difficulty with quantitative traits is the weak phenotype-genotype relationship. As for complex discrete traits, quantitative traits are influenced by environmental factors and, usually, by multiple genes. In this report we cover the results of our research work related to the usefulness of linkage disequilibrium for fine mapping quantitative trait loci under the project 'Statistical tools for quantifying genetic variability using molecular markers'. We first present results from a simulation study to characterise the pattern of linkage disequilibrium and the utility of disequilibrium mapping of quantitative trait loci based on regression analysis. Second we introduce a maximum likelihood linkage disequilibrium based method for fine mapping quantitative trait loci (QTL). We then present a Bayesian framework for combined linkage-linkage disequilibrium mapping approach. Finally we propose some ideas on future research directions related to linkage disequilibrium issue. The main results from these analyses showed that the pattern of linkage disequilibrium depends on allele frequencies. On average, linkage disequilibrium decreased as a function of recombination distance but the variability

* Correspondant et tirés à part : abdallah@toulouse.inra.fr

of disequilibrium measures was generally high. The mapping power by single-marker regression analysis was satisfactory for QTL with large effect and intermediate frequencies but less satisfactory for small QTL with extreme frequencies. The likelihood method based on modelling the decay of linkage disequilibrium using single and multiple markers provided more accurate estimates of QTL location compared to regression analysis particularly when the QTL had dominance effects. A Fortran program was developed to carry out mapping analyses based on the likelihood method using a variable number of markers and that of single-marker regression. The use of information from both linkage and linkage disequilibrium in a unified approach resulted in more precise estimates of QTL location than use of linkage analysis only. This combined linkage-linkage disequilibrium approach provides an appropriate framework for mapping QTL in data consisting of complex pedigrees like animal populations.

Linkage analysis / linkage disequilibrium / quantitative trait locus / fine-scale mapping

Résumé : Le déséquilibre de liaison (c'est-à-dire l'association non aléatoire d'allèles à différents loci) est apparu récemment comme un outil efficace pour cartographier finement les gènes responsables de maladies génétiques à déterminisme simple. Son utilité pour la dissection des caractères quantitatifs est encore à déterminer. Pour ces caractères généralement contrôlés par plusieurs gènes et influencés par des facteurs environnementaux, la principale difficulté des méthodes de cartographie est due à la faible liaison qui existe entre le phénotype et le génotype, ce qui n'est pas le cas des maladies mendéliennes. Dans cet article, nous illustrerons l'utilisation du déséquilibre de liaison pour la cartographie fine. Ces résultats ont été obtenus dans le cadre du projet « Outils statistiques pour quantifier la variabilité génétique à l'aide de marqueurs moléculaires ». Nous chercherons, tout d'abord, à l'aide de simulations, à caractériser le profil du déséquilibre de liaison dans des populations finies. Puis, nous présenterons une méthode basée sur le déséquilibre de liaison pour la cartographie fine des gènes à effet quantitatif. Nous terminerons en proposant une méthode qui permet de combiner analyse de liaison et déséquilibre de liaison, dans un contexte Bayésien. Pour finir, nous envisagerons des directions pour des travaux futurs.

analyse de liaison / déséquilibre de liaison / caractère quantitatif / cartographie fine

1. INTRODUCTION

Les études de cartographie génétique ont pour objectif de localiser les gènes qui expliquent la variabilité entre les individus et, éventuellement

Déséquilibre de liaison et cartographie fine

d'identifier les mutations causales. Il faudra beaucoup de temps pour réaliser un tel travail. Les études de cartographie utilisent classiquement l'analyse de liaison. Cette méthode estime la position des gènes avec un intervalle de confiance trop grand pour réaliser ensuite un clonage positionnel. Le facteur limitant de la résolution des analyses de liaison est le nombre d'événements de recombinaison contenus dans les données (on considère généralement deux ou trois générations).

L'utilisation du déséquilibre de liaison, c'est-à-dire des associations non aléatoires entre les gènes à différents locus, est une alternative pour la cartographie fine. Le déséquilibre de liaison utilise, en principe, tous les événements historiques de recombinaison dans une population d'individus non apparentés. Ceci permet d'estimer avec une meilleure précision la position du gène.

Le déséquilibre de liaison a été utilisé avec succès pour localiser les gènes responsables de maladies déterminées par un seul gène [4], [5]. Par contre, son efficacité pour des maladies à déterminisme complexe et des caractères quantitatifs n'a pas été démontrée. La relation entre le génotype et le phénotype est faible pour les caractères quantitatifs. Les maladies à déterminisme complexe, tout comme les caractères quantitatifs sont le plus souvent soumis à l'influence conjointe du milieu et d'un grand nombre de gènes.

Une mesure idéale du déséquilibre de liaison pour la cartographie génétique est une fonction monotone et décroissante de la distance génétique. Cette mesure doit par ailleurs être robuste aux variations des fréquences alléliques. L'étendue du déséquilibre de liaison dépend de l'histoire démographique de la population, des événements de recombinaison et des mutations. La cartographie fine n'utilise que le déséquilibre de liaison dû aux recombinaisons historiques [9]. Un événement de migration provenant d'une population qui a des fréquences géniques différentes provoque un important déséquilibre de liaison. En conséquence, la région qui présente le déséquilibre de liaison le plus important n'est pas forcément celle qui contient la mutation causale.

Dans cet article, nous illustrerons l'utilisation du déséquilibre de liaison pour la cartographie fine des gènes à effet quantitatif (QTL pour quantitative trait loci). Ces résultats ont été obtenus dans le cadre du projet « Outils statistiques pour quantifier la variabilité génétique à l'aide de marqueurs moléculaires ». Nous chercherons, tout d'abord, à l'aide de simulations, à caractériser le profil du déséquilibre de liaison dans des

populations finies. Puis, nous présenterons une méthode basée sur le déséquilibre de liaison pour la cartographie fine d'un QTL. Nous terminerons en proposant une méthode qui permet de combiner analyse de liaison et déséquilibre de liaison, dans un contexte Bayésien. Pour finir, nous envisagerons des directions pour des travaux futurs.

2. LE PROFIL DU DÉSÉQUILIBRE DE LIAISON

Considérons un locus avec deux allèles Q et q , de fréquences P_Q et $1-P_Q$. Supposons que ce locus est un QTL pour le caractère étudié et que l'allèle Q est apparu par mutation il y a t générations. Nous avons réalisé des simulations dans deux situations très contrastées [1]. Pour le scénario « déséquilibre de liaison », nous supposons que la mutation est apparue dans l'haplotype d'un individu unique. Il y a donc un déséquilibre de liaison complet entre l'allèle Q et les allèles sur les marqueurs du même haplotype. Pour le scénario « équilibre de liaison », nous supposons que l'allèle Q est initialement en équilibre de liaison avec les allèles marqueurs. Ces derniers sont aussi en équilibre de liaison. Le scénario « équilibre de liaison » fait l'hypothèse d'une mutation apparue dans plusieurs haplotypes et conservée par la dérive génique.

Nous avons tout d'abord étudié une région de 18 cM, avant de restreindre nos analyses à un intervalle de 3 cM. Cet intervalle contient le QTL et 30 marqueurs également espacés. La population se compose de 200 ou de 500 individus. Nous utilisons des marqueurs bi- ou multi-alléliques (5 allèles à chaque locus). Les fréquences initiales des marqueurs sont égales. Nous avons simulé 200 générations de panmixie par la méthode dite du « gene dropping » [6]. L'haplotype de chaque descendant est construit en choisissant au hasard deux parents puis en appliquant les lois de Mendel et les probabilités de recombinaison.

Dans le scénario « déséquilibre de liaison », la mutation est apparue après 100 générations de panmixie. Dans le scénario « équilibre de liaison », les allèles de la génération initiale ont été simulés en équilibre de liaison et $P_Q = 0,20$. Nous avons retenu uniquement les simulations dans lesquelles nous n'avons pas observé de fixation pour l'un quelconque des locus. Nous avons réalisé 10 000 simulations pour le scénario « déséquilibre de liaison » et 6 000 pour le scénario « équilibre de liaison ».

Déséquilibre de liaison et cartographie fine

La valeur phénotypique y pour un caractère quantitatif est simulée comme la somme d'une valeur génotypique g et d'une résiduelle e , où g prend les valeurs 1, 0 et -1 respectivement pour les génotypes QQ , Qq et qq . La résiduelle e est tirée dans une loi normale d'espérance 0 et de variance 1. Nous avons réalisé une régression mono-marqueur en analysant la valeur phénotypique y en fonction du nombre de copies des allèles marqueurs que porte chaque individu. Le locus pour lequel le test est le plus significatif sert à estimer la position du QTL.

Nos simulations montrent qu'en moyenne, le déséquilibre de liaison est une fonction décroissante de la distance génétique. Ce résultat est celui qui était attendu. Cependant, les mesures du déséquilibre de liaison ont une variabilité très importante et elles sont influencées par les fréquences alléliques. Du fait de cette grande variabilité, il n'est pas rare d'observer un déséquilibre de liaison élevé pour des marqueurs éloignés. Pour des populations de 500 individus avec une carte dense, ces marqueurs sont rarement à une distance de plus de 3 cM du QTL. La puissance de la régression pour la cartographie est bonne pour des populations de 200 à 500 individus lorsque la fréquence du QTL est intermédiaire et son effet important. Les résultats sont moins bons pour des QTL avec des fréquences plus faibles et/ou des petits effets. Une approche multi-locus est plus puissante et elle sera utile pour cartographier des QTL responsables d'une fraction plus réduite de la variance du caractère.

3. DÉSÉQUILIBRE DE LIAISON ET CARACTÈRES QUANTITATIFS

Les méthodes de cartographie qui utilisent le déséquilibre de liaison ont été développées pour des caractères discrets. Nous avons généralisé une méthode utilisée pour la cartographie des gènes de maladie [10] pour l'adapter à des caractères quantitatifs [2]. Dans la méthode originale, Terwilliger [10] a proposé un test de vraisemblance puissant qui s'adapte à un nombre variable d'allèles aux marqueurs ou de marqueurs considérés simultanément. Le test consiste à maximiser la vraisemblance pour un paramètre λ , défini comme la proportion qui décrit l'excès de l'allèle i sur les chromosomes qui portent Q , par rapport à sa fréquence dans l'ensemble de la population.

Supposons qu'initialement l'allèle i soit en déséquilibre complet avec l'allèle Q . La vraisemblance s'écrit :

$$L = \sum_{i=1}^I P_i \left(\prod_{n=1}^{N_s} \left(\sum_{k=1}^3 (P_{in}(g_k) \phi(y_n; \mu_k, \sigma^2)) \right) \right),$$

où N_s est le nombre d'observations, k est un indice qui prend les valeurs 1, 2 et 3 respectivement pour les génotypes QQ , Qq et qq . $P_{in}(g_k)$ est la probabilité que le n^e individu ait le génotype g_k au QTL, sachant le génotype au marqueur. La fonction $\phi(\cdot)$ est la densité de probabilité de la distribution normale, y_n est la valeur phénotypique, σ^2 est la variance intra-génotype, et μ_k est la moyenne phénotypique du k^e génotype. En utilisant la paramétrisation de Falconer pour les moyennes génotypiques [3], nous avons : $\mu_1 = \mu + a$, $\mu_2 = \mu + d$, et $\mu_3 = \mu - a$, où a et d sont définis comme les écarts additif et de dominance à la moyenne μ , la moyenne des génotypes homozygotes aux QTL. Nous supposons ici que la population est en équilibre de Hardy-Weinberg. Les probabilités $P_{in}(g_k)$ se calculent de la manière suivante :

$$P_{in}(g_1) = P_i(Q | h_n^1) P_i(Q | h_n^2),$$

$$P_{in}(g_2) = P_i(q | h_n^1) P_i(Q | h_n^2) + P_i(Q | h_n^1) P_i(q | h_n^2)$$

$$P_{in}(g_3) = P_i(q | h_n^1) P_i(q | h_n^2)$$

où h_n^1 et h_n^2 sont les allèles au marqueur pour les haplotypes mâle et femelle. Ici, il n'est pas nécessaire de connaître les phases. La probabilité des allèles au QTL sachant les allèles au marqueur ont été calculées en suivant la paramétrisation de Terwilliger [10]. Par exemple, si pour un individu n l'allèle marqueur sur h_n^1 est i et si l'allèle sur h_n^2 est j ($j \neq i$), alors la probabilité s'écrit : $P_{in}(g_j) = [P_Q + \lambda (1 - p_i) P_Q/p_i] [P_Q - \lambda P_Q]$ où p_i est la fréquence de i dans la population et λ est la proportion qui décrit l'excès des allèles i sur les chromosomes qui portent Q . D'une manière plus précise, λ est défini de façon à ce que $P(i | Q) = p_i + \lambda (1 - p_i)$. La vraisemblance est alors maximisée pour les paramètres. λ , P_Q , μ , a , d , et σ^2 .

Pour l'analyse multimarqueurs, Terwilliger a proposé de multiplier les vraisemblances des marqueurs individuels. En procédant ainsi, il suppose que les marqueurs sont indépendants, ce qui n'est généralement pas le cas notamment quand les marqueurs sont proches. Le non-respect de cette hypothèse constitue le principal inconvénient de cette méthode.

Déséquilibre de liaison et cartographie fine

Dans l'analyse multimarqueurs, le paramètre λ s'écrit $\lambda = \alpha(1-r)^t$ où α est la proportion des allèles Q qui sont initialement associés à l'allèle i , r est le taux de recombinaison entre le marqueur et le QTL et t est le nombre de générations depuis l'apparition de l'allèle Q.

Nous avons comparé deux à deux des analyses qui utilisent un marqueur, deux ou six marqueurs. Nous avons aussi comparé ces analyses avec une régression mono-marqueur en simulant différents scénarios. Nous avons apporté quelques modifications à la stratégie de simulation présentée précédemment. Nous considérons un scénario de déséquilibre de liaison avec une mutation apparue à la 20^e génération dans une population de 200 individus. Nous étudions une région de 10 cM avec un intervalle entre marqueurs de 0,25, 1,0 et 2,5 cM. Il y a éventuellement des effets de dominance au QTL.

Notre méthode donne en général de meilleurs résultats que la régression pour un QTL avec un effet de dominance. L'écart-type de l'erreur et l'intervalle de confiance de l'estimation de la position du QTL sont plus petits. Par exemple pour des marqueurs bi-alléliques espacés de 0,25 cM et un QTL avec des fréquences alléliques intermédiaires, l'intervalle de confiance est inférieur à 3 cM (fig. 1).

Notre méthode fournit en outre une estimation de la fréquence du QTL et des effets additifs et de dominance. Notons cependant que ces estimations sont souvent biaisées et qu'elles ont une variance élevée [2].

Nous avons développé un programme - QTLTER - simple et facile à utiliser pour mettre ces méthodes à disposition de nos collègues. Ce programme, écrit en Fortran 90/95, est disponible auprès des auteurs. Le programme possède trois options : une régression, une analyse « Terwilliger » avec un marqueur et une analyse « Terwilliger » avec plusieurs marqueurs. Le programme estime la position du QTL et il fournit en outre une estimation de l'effet et de la fréquence du QTL. L'utilisateur doit fournir les haplotypes et les performances des individus de la population étudiée.

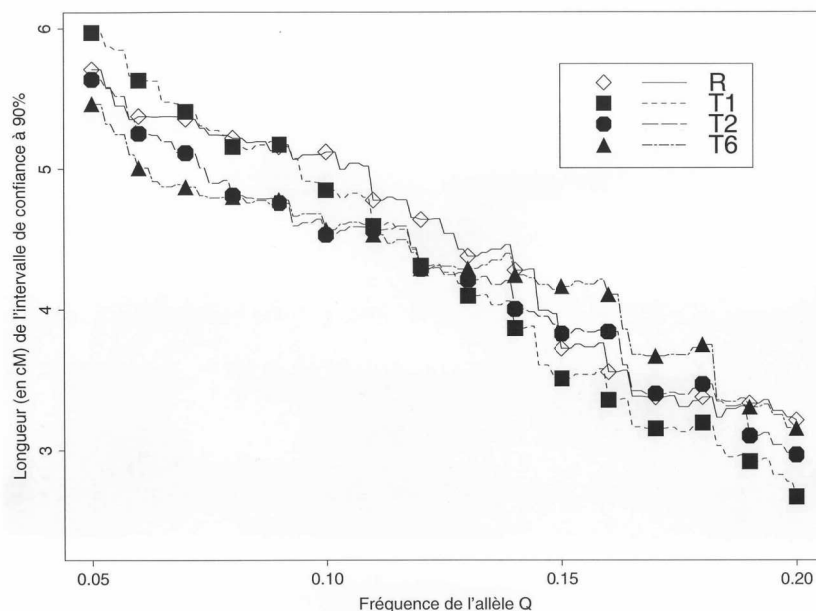


Figure 1 : Taille de l'intervalle de confiance au niveau de 90 % en fonction du seuil de la fréquence de l'allèle Q au QTL. Un seuil de 0,10 rassemble toutes les simulations avec une fréquence de QTL égale ou supérieure à 0,10. Ces simulations ont été faites avec des marqueurs bi-alléliques séparés de 0,25 cM, $a = 1$, $d = 0$ et la fréquence du QTL varie entre 0,05 et 0,50. R = régression, T1 = Terwilliger avec une analyse mono-marqueur, T2 = Terwilliger avec une analyse utilisant 2 marqueurs, T6 = Terwilliger avec une analyse utilisant 6 marqueurs.

4. UNE APPROCHE BAYÉSIENNE UNIFIÉE

Une approche prometteuse pour la cartographie fine est de combiner les méthodes d'analyse de liaison et de déséquilibre de liaison dans un seul cadre théorique qui permette d'additionner leurs avantages respectifs. C'est en particulier, pour améliorer la précision de l'intervalle de confiance du QTL qu'il convient de proposer de telles méthodes robustes et efficaces.

Considérons, par exemple, une expérimentation simple comprenant un ensemble de familles nucléaires dans lesquelles parents et descendants sont typés. Les avantages théoriques d'une analyse utilisant de façon combinée le déséquilibre de liaison et l'analyse de liaison dans un pedigree sont nombreux : i) un parent homozygote à un marqueur

Déséquilibre de liaison et cartographie fine

n'apporte pas d'information dans l'analyse de liaison mais il en apporte dans l'analyse de déséquilibre de liaison ; ii) au contraire, une analyse de déséquilibre de liaison seule peut être trompée par deux parents qui ont le même haplotype mais pas nécessairement les mêmes allèles au QTL. Dans ce cas, l'utilisation du phénotype des descendants et des allèles transmis aide à déterminer le génotype probable au QTL pour les parents ; iii) un individu sans descendance mais observé pour son phénotype peut être inclus dans l'analyse de déséquilibre de liaison ; iv) la comparaison des analyses avec ou sans déséquilibre de liaison peut permettre de tester la validité des postulats du modèle pour le déséquilibre (une mutation unique il y a t générations).

Nous présentons, ici, une méthode Bayésienne qui combine les informations de liaison et de déséquilibre dans un cadre théorique unifié [8]. Notre méthode utilise conjointement toute l'information due aux marqueurs ainsi que toute celle disponible dans le pedigree. Elle n'est pas restreinte à un type de plan d'expérience ou de pedigree particuliers et ne nécessite pas la connaissance des phases des marqueurs. Il est possible d'inclure dans le modèle aussi bien des effets génétiques infinitésimaux que des effets (fixes) environnementaux. Le QTL est supposé bi-allélique, ses effets d'additivité et de dominance peuvent tous les deux être estimés. Nous avons implémenté un échantillonneur combiné de Gibbs et de Metropolis-Hastings afin d'obtenir les distributions marginales *a posteriori* des paramètres d'intérêt.

Nous supposons que le but de l'analyse est de cartographier finement un QTL qui a été détecté auparavant dans une région donnée du génome. Le modèle génétique postule qu'une mutation est apparue il y a t générations sur un gène affectant le caractère étudié. Il y a donc initialement dans la population un seul haplotype ancestral (fondateur) qui possède la mutation. Le nombre d'haplotypes qui possèdent la mutation augmente au cours des générations (si l'allèle mutant ne disparaît pas), cependant la combinaison allélique initiale est perturbée par la recombinaison génétique. La quantité de déséquilibre entre les marqueurs et le QTL décroît proportionnellement à leur distance génétique et au nombre de générations depuis la mutation. Nous utiliserons le modèle de décroissance du déséquilibre de liaison décrit par [7] en y introduisant des modifications.

Soit un QTL dont les effets d'additivité et de dominance sont notés respectivement a et d , c'est-à-dire que la différence entre le phénotype

espéré des individus homozygotes QQ et celui des homozygotes qq est $2a$; tandis que l'espérance du phénotype des individus hétérozygotes Qq ou qQ est d . Supposons, de plus, qu'un nombre N d'individus a été typé pour des marqueurs, typages contenus dans la matrice \mathbf{M} et que des mesures phénotypiques contenues dans le vecteur \mathbf{y} ont été recueillies sur un sous-ensemble de N_s individus. Le modèle explicatif est :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{w}_a a + \mathbf{w}_d d + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

où $\boldsymbol{\beta}$ est un vecteur d'effets fixes (effets environnementaux ou non génétiques), \mathbf{w}_a est un vecteur de variables indicatrices prenant respectivement les valeurs 1 ou -1 si le génotype au QTL est QQ ou qq , et zéro si l'individu est hétérozygote, \mathbf{w}_d contient la valeur 1 si l'individu est de génotype Qq ou qQ au QTL, zéro sinon, \mathbf{u} et \mathbf{e} sont des vecteurs d'effets aléatoires contenant les valeurs de l'effet génétique infinitésimal (effet polygénique) et résiduel, tandis que \mathbf{X} et \mathbf{Z} sont les matrices d'incidence.

Le but de l'analyse est d'estimer l'ensemble des paramètres $\boldsymbol{\theta} = \{\mathbf{S}_0, a, d, \mathbf{u}, \boldsymbol{\beta}, \sigma_u^2, \sigma_e^2, \delta, t, \mathbf{T}, \mathbf{H}\}$, où \mathbf{S}_0 est la matrice qui contient le statut de l'identité par descendance (I.B.D.) des deux allèles au QTL d'un individu par rapport à la mutation causale, σ_u^2 est la variance de l'effet infinitésimal; σ_e^2 , la variance résiduelle, δ est la position du QTL. \mathbf{T} est le vecteur indicateur de la ségrégation au QTL qui contient pour chaque allèle de chaque individu, une variable binaire indiquant le statut I.B.D. de l'allèle au QTL par rapport aux allèles parentaux. Finalement, \mathbf{H} contient les phases (paternelles ou maternelles) pour chaque marqueur. L'inférence Bayésienne est basée sur la distribution *a posteriori* des paramètres :

$$P(\boldsymbol{\theta} | \mathbf{y}, \mathbf{M}) \propto p(\mathbf{y}, \mathbf{M} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) = p(\mathbf{y} | \boldsymbol{\theta}) p(\mathbf{M} | \boldsymbol{\theta}) p(\boldsymbol{\theta})$$

où $p(\mathbf{y}, \mathbf{M} | \boldsymbol{\theta})$ est la vraisemblance (au sens Bayésien) et $p(\boldsymbol{\theta})$ la distribution *a priori* des paramètres.

Des données simulées dans deux types de pedigree : un pedigree simple (familles de pleins frères et sœurs) et un pedigree complexe (quatre générations) nous ont permis d'illustrer cette méthode et de comparer les estimations découlant d'une analyse avec ou sans l'utilisation du déséquilibre de liaison. En général, l'estimation de la position du QTL est meilleure en utilisant le déséquilibre, en particulier lorsque le déséquilibre est complet entre l'allèle mutant au QTL et un

Déséquilibre de liaison et cartographie fine

allèle marqueur. Cependant cet avantage décroît lorsque l'association est seulement partielle. Par contre, et cela dans tous les cas simulés, les paramètres d'additivité et de dominance sont estimés avec précision par les deux analyses.

Nos résultats montrent qu'il est possible d'affiner la zone de localisation d'un QTL en analysant des populations de taille raisonnable, comprenant des familles de taille modeste et un marquage qui ne soit pas extrêmement dense. Cependant, ils mettent en lumière, aussi, que l'intérêt de la prise en compte du déséquilibre dans les analyses classiques de liaison dépend de nombreux facteurs. Premièrement, l'utilité des méthodes combinant l'analyse de liaison et de déséquilibre de liaison par rapport aux méthodes d'analyse de liaison est fortement dépendante de la valeur du déséquilibre entre les allèles au QTL et aux marqueurs. Deuxièmement, la structure de la population a une influence : en effet, pour une cartographie précise, il est extrêmement important de déterminer correctement les phases et les génotypes au QTL. Il semblerait donc préférable de disposer d'un petit nombre d'individus de base ancêtres de grandes familles plutôt que d'un pedigree complexe sur plusieurs générations. De plus, le hasard affecte aussi les résultats : la transmission Mendélienne, la recombinaison, et l'aléa dû à l'environnement sont des processus stochastiques qui produisent des données très différentes même en partant de conditions initiales identiques.

5. CONCLUSIONS ET PERSPECTIVES

La cartographie fine est un secteur de recherche très actif. Le déséquilibre de liaison est un outil prometteur dont l'utilisation n'est pas sans risque. À notre avis, il importe d'améliorer les méthodes de cartographie multi-locus en utilisant l'information apportée par le déséquilibre de liaison, mais aussi par les pedigrees. Il faut aussi étendre ces méthodes au cas de mutations localisées dans la même région à des locus différents (plusieurs QTL) et au cas où il y a plusieurs mutations au même locus (hétérogénéité génétique).

REMERCIEMENTS

Ce travail a été financé par le Bureau des Ressources Génétiques et par une « Action en bioinformatique » du ministère de la Recherche.

RÉFÉRENCES

- [1] Abdallah J.M., Goffinet B., Cierco-Ayrolles C., Pérez-Enciso M., Linkage disequilibrium fine mapping of quantitative trait loci: a simulation study, *Genet. Sel. Evol.* 35 (2003) 513-532.
- [2] Abdallah J.M., Mangin B., Goffinet B., Cierco-Ayrolles C., Pérez-Enciso M., A comparison between methods for linkage disequilibrium fine mapping of Quantitative Trait Loci, *Genet. Res.* (2004) 41-47.
- [3] Falconer D.S., Mackay T.F.C., *Introduction to quantitative genetics*, 4^e ed., Longman, Essex, 1996, 464 p.
- [4] Hästbacka J., Chappelle A. de la, Kaitila I., Sistonen P., Weaver A., Lander E., Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland, *Nat. Genet.* 2 (1992) 204-211.
- [5] Lehesjoki A., Koskiniemi J., Norio R., Tirrito S., Sistonen P., Lander E.S., Chappelle A. (de la), Localization of the EPM1 gene for progressive myoclonus epilepsy on chromosome 21: linkage disequilibrium allows high resolution mapping, *Hum. Mol. Genet.* 2 (1993), 1229-1234.
- [6] MacCluer J.W., VandeBerg J.L., Read B., Ryder O.A., Pedigree analysis by computer simulation, *Zoo Biol.* 5 (1986) 147-160.
- [7] Morris A.P., Whitaker J.C., Balding D.J., Bayesian fine-scale mapping of disease loci, by Hidden Markov Models, *Am. J. Hum. Genet.* 67 (2000) 155-169.
- [8] Perez-Enciso M., Fine mapping for complex trait gene combining pedigree and linkage disequilibrium information: a Bayesian unified framework, *Genetics* 163 (2003) 1497-1510.
- [9] Pritchard J.K., Przeworski M., Linkage disequilibrium in humans: models and data, *Am. J. Hum. Genet.* 69 (2001) 1-14.
- [10] Terwilliger J.D., 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 (1995) 777-787.