Habilitation à Diriger les Recherches

Université Paris-Sud

Christine Dillmann
Frédéric Hospital

Document de synthèse commun

«Quels modèles pour la génétique quantitative ?»

Introduction

Un caractère quantitatif est un caractère gouverné par de nombreux gènes (plus de deux) et dont l'expression est influencée par le milieu. En génétique quantitative, on cherche à établir des modèles qui permettent de prédire l'évolution des caractères quantitatifs. On confond souvent génétique quantitative et génétique statistique, pour des raisons qui tiennent tant à l'histoire de la discipline qu'aux outils utilisés. Plus largement, la génétique quantitative au sens où nous l'entendons ici aborde des problèmes liés à l'hérédité des caractères complexes et constitue une extension à plusieurs locus de la génétique des populations.

En génétique des populations, le gène est conçu comme une représentation immatérielle, liée au phénotype, qui permet de faire des calculs sur l'hérédité des caractères. La définition minimaliste que l'on peut en donner est la suivante : *Les gènes sont des entités codant pour une fonction, transmissibles au cours des générations, séparés les uns des autres par la recombinaison, lors de la reproduction sexuée.*

Les limites des modèles d'évolution des caractères quantitatifs En simplifiant, on peut dire que les modèles analytiques disponibles pour prédire l'évolution des caractères quantitatifs sont issus de deux écoles : i) ceux de la Génétique des Populations cherchent à décrire explicitement la fréquence de chaque gène en cause, mais sont en général limités à un ou deux gènes sélectionnés au maximum ; ii) ceux de la Génétique Quantitative font abstraction du déterminisme génétique mais ne sont pas adaptés pour prendre en compte les connaissances biologiques, et d'autre part leurs hypothèses sont questionnables, notamment sur le long terme.

Notre démarche Nos travaux de recherche se sont inspirés d'une démarche commune visant à tester la pertinence biologique des modèles prédictifs de la réponse à la sélection en combinant les approches analytiques, expérimentales et la simulation. Nous nous sommes intéressés pour l'un aux effets de la liaison génétique et pour l'autre aux interactions entre gènes et à la relation entre génotype et phénotype. Une réflexion commune récente au sein de l'équipe de Génétique Quantitative Fondamentale de la Station de Génétique Végétale du Moulon nous a conduits à nous intéresser à l'évolution des systèmes polygéniques dans une population, et à identifier un certain nombre de facteurs qui vont conditionner la réponse, ou l'absence de réponse d'un gène identifié aux pressions évolutives : *i*) la répartition des gènes polymorphes sur le chromosome, *ii*) la relation entre génotype et fitness, *iii*) le nombre de gènes impliqués, à une génération, dans la variation des caractères quantitatifs, et *iv*) la distribution des valeurs des allèles dans une population.

1 La génétique statistique

La formalisation de la description statistique de la composition génétique d'une population date de Fisher (1918). Il propose une décomposition linéaire de la valeur phénotypique P :

$$P = G + E = A + D + E \tag{1}$$

où G représente l'effet du génotype et E l'effet du milieu. A représente la part de la valeur génétique que l'on peut prédire à partir d'un modèle additif, et D une part non additive, qui est un résidu statistique. Formellement, en supposant des gènes connus, l'effet additif A_i d'un allèle i est défini comme le meilleur prédicteur de la valeur génétique des individus qui

possèdent une copie de cet allèle. Il est obtenu par la méthode des moindres carrés en résolvant l'équation :

$$\partial E\left[(G - A_i)^2\right] / \partial A_i = 0 \tag{2}$$

Ces effets statistiques n'ont de valeur que dans la population étudiée et dépendent, entre autres, des fréquences des allèles dans la population. Cependant, il n'existe aucune relation entre l'effet additif d'un allèle et sa valeur au sens biologique (Dillmann 1992). Ainsi, tous les effets d'interaction entre gènes allèles (dominance) ou gènes non allèles (épistasie) ne sont mesurables en génétique statistique que comme des résidus à l'additivité à travers le paramètre D. Quelle que soit l'importance biologique de ces effets, leur valeur statistique mesurable sera toujours plus faible que celle des effets principaux.

Cette décomposition statistique a l'immense avantage de fournir des paramètres mesurables à partir de l'étude de la ressemblance entre apparentés, ou de l'étude de répétitions d'un même génotype, sans faire aucune hypothèse sur le déterminisme génétique des caractères étudiés. Par exemple, dans une population panmictique, si les distributions de P et de G suivent une loi normale, il existe une relation linéaire entre valeur des enfants et valeur des parents, et la pente de la droite de régression parent-enfant est égale à la moitié du rapport entre variance génétique additive et variance phénotypique :

$$b_{P|O} = \frac{1}{2}\frac{\sigma_A^2}{\sigma_P^2} = 1/2\, h^2 \tag{3}$$

La part «transmissible» de la variance phénotypique est quantifiée par l'héritabilité du caractère, h^2. On peut alors prédire la réponse à une génération de sélection :

$$R = ih^2\sigma_P = h^2 S \tag{4}$$

où i représente l'intensité de la sélection, et S la différentielle de sélection, c'est à dire l'écart entre la moyenne des individus sélectionnés et la moyenne de la population.

Les applications de ces modèles statistiques en amélioration des plantes concernent l'optimisation des dispositifs expérimentaux, et l'optimisation de la réponse à la sélection (par exemple, Gallais 1993 ; Goldringer *et al.* 1996 ; Geiger et Tomerius 1997).

La génétique statistique s'est développée essentiellement dans le cadre de l'amélioration génétique et reste valable dans ce cadre uniquement, pour prédire la réponse à court terme à la sélection. En effet, l'équation (4) décrivant la réponse à une génération de sélection n'est extensible sur plusieurs générations de sélection qu'à condition que les variances génétiques et environnementales ne changent pas au cours des générations de sélection, et que la distribution des valeurs génétiques avant et après la sélection suive une loi normale.

2 Modélisation de la réponse à long terme à la sélection

Pour prédire la réponse à long terme à la sélection, un certain nombre de modèles tentent d'étendre à un grand nombre de génération l'équation (4) de la réponse à une génération de sélection, et font pour cela des hypothèses que nous aimerions détailler ici.

2.1 Le modèle infinitésimal

Dans une population infinie, si le caractère sélectionné est gouverné par un nombre infini de gènes à effets individuels infiniment faibles, alors la pression de sélection qui s'exerce à

chaque locus est infiniment faible et les fréquences des allèles ne changent pas sous l'effet de la sélection. La variance génétique reste constante. De même, les valeurs génétiques et phénotypiques résultant de l'addition d'un grand nombre de facteurs d'effet individuel faible, leur distribution tend vers une loi Normale à chaque génération de sélection. C'est le modèle infinitésimal. Ces hypothèses sont suffisantes pour étendre l'équation (4) à plusieurs générations de sélection :

$$R_\tau = \tau \frac{i}{\sigma_P} \sigma_A^2(0) \tag{5}$$

La réponse à la sélection est alors une fonction linéaire du temps (τ), dont la pente dépend de la variance génétique additive de la population initiale (Figure 1). La valeur du caractère augmente indéfiniment. Cependant, il n'y a pas vraiment de raison de penser que la variance génétique reste constante au cours du temps, et les modèles qui suivent ont tenté de prédire son évolution.

2.2 Dérive génétique

En l'absence de sélection, l'échantillonnage aléatoire des gamètes dans une population de taille finie va augmenter la consanguinité (probabilités d'identité par descendance entre gènes) de cette population et conduire à la fixation ultime d'un allèle à chaque locus. Si les effets de la dérive génétique sont extrêmement bien connus dans le modèle neutre (Kimura 1957), ils le sont très mal en sélection pour des caractères quantitatifs. On sait que la sélection réduit l'effectif efficace des populations, mais on ne sait pas dans quelle mesure (Santiago et Caballero 1995).

Dans le cadre du modèle infinitésimal, une solution proposée pour prendre en compte les effets de la dérive génétique a été de «plaquer» artificiellement le modèle neutre sur le modèle sélectif en ne décrivant que l'évolution de la variance génétique en consanguinité (Wright 1931) :

$$\sigma_A^2(\tau) = (1 - exp(-\tau/2N))\, \sigma_A^2(0)$$

La variance génétique additive diminue alors au fil du temps jusqu'à épuisement, et la réponse à la sélection atteint une limite (Figure 1). Il faut noter que dans ce cas, des allèles défavorables peuvent se fixer même en présence de sélection. La limite atteinte ne correspond pas à la valeur maximale du caractère (Robertson 1960).

NB : une extension de ce modèle est le BLUP-modèle animal, qui cherche à estimer la valeur génétique des reproducteurs ayant participé à la sélection en prenant en compte leur relations d'apparentement : $P = \mu + Zu + E$, avec u multinormale de variance $A\sigma_A^2$, et A est la matrice décrivant l'apparentement entre les individus. Ce modèle suppose que la population a été fondée à partir d'individus non apparentés, les estimations obtenues sont bien entendu fortement dépendantes des hypothèses sous-jacentes.

Le modèle infinitésimal n'a plus aucun sens en population finie, puisque les fréquences des allèles évoluent forcément sous l'effet de la dérive génétique. D'autre part, ce modèle simplifié ne prend pas en compte les interactions entre sélection et dérive génétique.

2.3 Déséquilibre de liaison

Le déséquilibre de liaison (D) entre deux locus mesure l'écart à l'association aléatoire des allèles dans les gamètes. Son signe est arbitraire. Par convention, nous donneront un signe positif à un excès de gamètes portant les deux allèles favorables (coupling). La situation

inverse (répulsion) correspond à un déséquilibre négatif. En l'absence de sélection et dans une population panmictique, les déséquilibres de liaison, une fois créés, diminuent d'une génération (τ) à la suivante sous l'effet de la recombinaison.

$$D_{\tau+1} = (1 - r)\, D_\tau \tag{6}$$

où r est le taux de recombinaison entre les deux locus considérés. Ainsi, on s'attend, dans des populations naturelles, à trouver surtout du déséquilibre de liaison entre locus liés, bien que cette règle ne soit évidemment pas absolue. Lorsque deux locus sont liés ($r < 0.5$), on parle de linkage.

2.3.1 Déséquilibre de liaison et variance génétique

La réponse à la sélection à une génération donnée dépend de la variance génétique du caractère sélectionné (4). Le déséquilibre de liaison peut modifier de façon importante cette quantité de variance génétique disponible pour la sélection. Si la valeur génétique d'un individu est égale à la somme des valeurs des allèles portés par cet individu à chaque locus,

$$G = \sum_{i=1}^{n} (g_i + g_i^*) \tag{7}$$

où g_i et g_i^* sont les valeurs des allèles paternels et maternels, alors, la variance génétique dans une population panmictique s'écrit comme la somme des variances géniques à chaque locus (V_g) et des covariances entre paires de locus (C) :

$$\sigma_G^2 \;=\; \sum_{i=1}^{n} \sigma_{g_i}^2 + 2 \sum_{i} \sum_{j \neq i} Cov(g_i, g_j) = V_g + C \tag{8}$$

On montre aisément que la covariance entre deux locus biallèliques d'effets α_i et α_j respectivement vaut, en panmixie :

$$Cov(g_i, g_j) = D\, \alpha_i\, \alpha_j \tag{9}$$

Les covariances entre paires de locus peuvent être positives ou négatives, selon le signe des déséquilibres de liaison. Ainsi, une partie de la variance génétique peut être «cachée» par un excès d'associations en répulsion à des locus différents (C négatif). Ceci aura pour conséquences de ralentir la réponse à la sélection. Au contraire, lorsque C est positif, la variance génétique est plus importante que la variance génique, et la réponse à la sélection est accélérée.

Cas particulier En amélioration des plantes, on utilise pour détecter des QTL des populations issues de l'autofécondation (F_2, lignées recombinantes) ou du rétrocroisement ($BC1$) d'un hybride F_1. Dans ce type de population, on s'attend à ne trouver du déséquilibre de liaison qu'entre locus liés génétiquement. La variance génétique dans ces populations va dépendre, outre des effets des QTL, des associations entre allèles aux locus polymorphes chez les parents (8). Lorsque les parents sont fixés aléatoirement pour l'un ou l'autre allèle à plusieurs locus (phase aléatoire), on s'attend *a priori* à avoir autant de covariances négatives que positives entre paires de locus, et donc à ce que la somme de ces covariances ait une distribution centrée sur zéro. On voit sur la figure 2 que la somme des covariances entre toutes les paires de locus n'est quasiment jamais nulle, et la distribution des valeurs possibles est très disymétrique. Cette distribution est contrainte par le degré d'association des allèles favorables sur un chromosome. Nous aimerions décrire cette contrainte, et ses conséquences sur la quantité de variance génétique exploitable par la sélection.

2.3.2 Déséquilibre de liaison et sélection

Historiquement, l'une des premières limitations apportée au modèle infinitésimal est l'effet Bulmer (1971), qui prédit que la sélection pour un caractère additif crée des déséquilibres de liaison négatifs, et réduit ainsi la variance génétique exploitable pour la réponse à la sélection. La démonstration de Bulmer est purement statistique (modèle infinitésimal multinormal), mais le même résultat peut être démontré génétiquement avec moins d'hypothèses (Hospital et Chevalet 1996).

Si l'on ne connaît pas les gènes qui déterminent un caractère quantitatif, il est difficile de mesurer la part de variance génétique C attribuable aux covariances entre locus. C est un paramètre que l'on ne peut pas estimer expérimentalement. Cependant, on peut avoir une idée de son importance si, partant d'une population sélectionnée, on effectue quelques générations de reproduction en panmixie et sans sélection. Si une partie de la variance génétique était «cachée» par des déséquilibres de liaisons négatifs entre locus, alors on s'attend à ce qu'elle augmente après le relâchement de la sélection sous l'effet des recombinaisons (6). De tels phénomènes ont été obervés expérimentalement dans des expériences de sélection chez la drosophile (Thoday and Boam 1961 ; Mather and Harrison 1949 ; in Lynch and Walsh 2002).

2.3.3 Les effets conjoints de la dérive génétique, de la sélection, et du déséquilibre de liaison

Alors qu'à un locus, la sélection augmente la probabilité de fixation de l'allèle favorable (Kimura 1957), Hill et Robertson (1966) montrent que la sélection à deux locus dans une population de taille finie a pour conséquences de diminuer cette probabilité de fixation, par rapport au résultat à un locus. En particulier, la sélection et le linkage augmentent la probabilité de fixation des gamètes en répulsion. Cette modification des probabilités de fixation entraine une diminution de l'effectif efficace de la population autour de la zone chromosomique sélectionnée. Il n'existe pas d'expression générale des probabilités de fixation dans des systèmes à plus de deux locus.

2.4 Nombre de locus fini

En supposant la multinormalité de la distribution des valeurs génétiques à chaque locus, une tentative a été faite pour prédire l'évolution de la matrice de variance-covariance génétique sous l'effet de la sélection dans une population de taille finie, lorsque le nombre de locus est fini, mais le nombre d'allèles à chaque locus est infini (Chevalet 1994) :

$$G_{kl}^{t+1} = \left(1 - \frac{1}{2N} - r_{kl}\right) G_{kl}^{t} - \left(1 - \frac{1}{N}\right) G_{k.}^{t} G_{l.}^{t} \tag{10}$$

Cependant, ce modèle analytique ne considère la dérive génétique que par ses effets sur la consanguinité, et ne prend pas en compte la fixation des allèles dans une population de taille finie.

Le fait que les caractères quantitatifs soient gouvernés par un nombre fini de locus a une conséquence importante pour la réponse à la sélection. En effet, il existe alors une limite claire au processus de sélection même dans une population infinie, qui correspond à la fixation de tous les allèles favorables, lorsque le nombre d'allèles n'est pas lui-même infini. Avec un nombre fini de locus, la pression de sélection qui s'exerce sur chaque locus ne peut plus être

infiniment faible et la variance génétique évolue au cours du temps, ne serait-ce que sous l'effet de l'évolution des fréquences alléliques.

Evolution de la variance génétique sous l'effet de la sélection Le fait que la variance génétique évolue sous l'effet de la sélection transforme l'équation de la réponse à la sélection (4) en un système ouvert d'équations qui n'a pas de solution (Barton et Turelli 1987). En effet, si la moyenne de la population est fonction de sa variance à la génération précédente, la variance elle même est fonction des moments d'ordre 3 et 4 de la distribution des valeurs génotypiques, et ainsi de suite. On ne peut arriver à des prédictions qu'en faisant des hypothèses sur la normalité des distributions de valeurs génotypiques (Lande 1976). Cependant, ces hypothèses ne sont vraies qu'à court terme et pour peu de locus. Elles ne le sont plus lorsque le nombre de locus à prendre en compte augmente (Turelli et Barton 1990).

2.5 Le rôle de la mutation

La variance génétique $\sigma^2_m = 2U\sigma^2$ dûe à de nouvelles mutations à chaque génération dépend à la fois du taux de mutation U par génome haploide, et de la variance σ^2 des effets des mutations. Il existe deux catégories de modèles pour prendre en compte les effets des mutations, selon que la valeur de l'allèle muté s'ajoute à la valeur de l'allèle avant mutation (Incremental Mutation Model ou IMM) ou que la valeur de l'allèle muté remplace simplement la valeur de l'allèle avant mutation (Random Mutation Model ou RMM). Les résultats qui suivent ont été obtenus avec modèle incrémental.

Sélection directionnelle Dans le cadre du modèle infinitésimal, on peut calculer la réponse à la sélection en combinant la réponse dûe à la variabilité génétique existant dans la population initiale, et celle dûe aux nouvelles mutations (Hill 1982 ; Weber et Diggins 1990 ; in Lynch and Walch 2002) :

$$R_\tau = 2N \frac{i}{\sigma_P} \left(\tau\, \sigma^2_m + 1 - exp(-\tau/2N)\right) \left(\sigma^2_A(0) - 2N\, \sigma^2_m\right) \tag{11}$$

Même lorsque la variance initiale est épuisée, la réponse à la sélection se poursuit grâce aux nouvelles mutations (Figure 1).

Sélection stabilisante dans des populations naturelles D'autres modèles ont été développés en considérant une sélection stabilisante, dans le but d'expliquer par la mutation le maintien de variabilité génétique dans les populations naturelles, malgré la sélection et la dérive génétique. On peut citer trois résultats principaux :

En l'absence de sélection, la dérive génétique conduit à la perte du polymorphisme mais des allèles nouveaux sont introduits à chaque génération par mutation. La variance génétique attendue à l'équilibre est $2N\sigma^2_m$, à condition que $4N\mu << 1$, c'est à dire qu'il n'y ait pas plus de deux allèles par locus (Hill 1982).

Avec une sélection stabilisante, on s'attend également à un état d'équilibre polymorphe, lorsque la valeur phénotypique moyenne de la population est à l'optimum. Il n'existe pas de solution analytique pour prédire la variance génétique à l'équilibre mutation-sélection-dérive génétique, mais des approximations, qui dépendent, entre autres, d'hypothèses sur la distribution des effets des mutations, et surtout, sur leur variance.

Lande (1975) suppose que les effets des mutations sont distribués selon une loi normale de variance faible par rapport à la variance génétique existante à un locus (Gaussian allelic model). Dans ce cas, la variance génétique attendue à l'équilibre, en population infinie, est

$$\sigma_G^2(G) = \sqrt{(2n\sigma_m^2 V_s)} \tag{12}$$

où $V_s = \sigma_E^2 + s^2$ représente l'intensité de la sélection.

Turelli (1984) suppose lui que l'optimum est atteint grâce à de rares mutations d'effet fort, et donc que la variance des effets des mutations est forte, c'est le modèle «House of Cards». Dans ce cas, la variance génétique attendue à l'équilibre, en population infinie, est

$$\sigma_G^2(HC) = 4UV_s \tag{13}$$

Les résultats expérimentaux obtenus à ce jour sur de nombreux organismes modèles ne permettent pas réellement de différencier les deux hypothèses (Barton et Turelli 1989).

Plusieurs auteurs, simultanément, ont tenté de prendre également en compte la dérive génétique (Bürger 1988 ; Keightley et Hill 1988 ; Barton 1989 ; Bürger *et al.* 1989 ; Houle 1989) et aboutissent au modèle «Stochastic House of Cards» :

$$\sigma_G^2(SHC) = \frac{(2N\sigma_m^2)(4UV_s)}{(2N\sigma_m^2) + (4UV_s)} \tag{14}$$

Des résultats de simulation (Bürger et Lande 1994) montrent que l'approximation SHC est correcte, sauf dans les cas suivants, où elle surestime la variance génétique à l'équilibre : locus très liés, variance des effets des mutations très faible, ou distribution fortement leptokurtique des effets des mutations.

3 La relation entre génotype et valeur sélective

Lorsque l'on considère des gènes identifiés, on peut caractériser la relation entre génotype et phénotype à l'aide d'un modèle génétique, qui consiste à attribuer à chaque allèle une valeur et à décrire les interactions entre ces allèles. Les modèles génétiques les plus étudiés sont le modèle additif, le modèle multiplicatif et le modèle métabolique. Ces modèles se distinguent entre eux par l'importance des écarts à l'additivité (linkage, épistasie, dominance) dans les composantes de la variance génétique d'une population, mais aussi par l'existence, ou non, d'interactions entre allèles au sens biologique. On parlera de dominance pour décrire les interactions entre allèles à un même locus lorsque la valeur de l'hétérozygote est plus proche de la valeur de l'un des deux homozygotes. On parlera d'épistasie lorsque la valeur d'un génotype à un locus dépend des allèles présents à d'autres locus. Enfin, lorsque l'on s'intéresse aux effets de la sélection, il faut considérer les relations entre phénotype et valeur sélective.

3.1 Le modèle additif

Un modèle génétique est additif lorsque les valeurs des allèles à chaque locus s'additionnent pour donner le phénotype (Tableau 1). L'avantage du modèle additif est qu'il est simple à manipuler, car on peut utiliser alors tous les raffinements de l'algèbre linéaire. En particulier, on peut calculer assez facilement l'évolution d'un caractère gouverné par deux locus biallélіques (A/a et B/b) et soumis à une sélection directionnelle.

Si A et B sont les allèles favorables, on constate que tous les types de sélection, mis à part la sélection disruptive, ont pour conséquence de créer un déséquilibre de liaison négatif entre les deux locus. Après une génération de sélection, le déséquilibre de liaison vaut :

$$D_{\tau+1} = \frac{1-r}{\bar{w}} D_\tau - \frac{1}{\bar{w}^2} (s\, p_A\, p_a + t\, D_\tau)\, (t\, p_B\, p_b + t\, D_\tau) \tag{15}$$

où $\bar{w}$ est la valeur sélective moyenne de la population, et s et t sont les coefficients de sélection des allèles A et B. On voit bien dans l'équation (15) qu'au cours des générations, le déséquilibre de liaison va évoluer sous l'effet de pressions antagonistes : la dérive génétique peut créer des déséquilibres de liaison de signe arbitraire. La sélection entraine un excès de gamètes en répulsion. La recombinaison fait diminuer le déséquilibre de liaison (6).

3.2 Notion d'additivité statistique

En génétique statistique, les gènes sont des entités plutôt virtuelles dont on ne considère souvent que les effets principaux, au sens de l'analyse de variance, qui s'additionnent pour donner le phénotype. Si l'on considère des gènes particuliers, comme c'est le cas lorsque l'on cherche à détecter des QTL, par exemple, on sort de ce cadre formel.

On fait souvent, à tort, l'amalgame entre un modèle génétique additif et le modèle statistique additif. En génétique statistique, on parle d'un modèle additif lorsque toute la variance génétique est héritable, ou alors lorsque l'on peut expliquer toute la variance phénotypique par les effets principaux des locus et l'effet du milieu. Il est tout à fait possible d'observer, avec un modèle génétique additif, des écarts à l'additivité au sens statistique, par exemple en présence de déséquilibre de liaison (covariances non nulles entre locus).

Cependant, les écarts à l'additivité, au sens statistique, sont difficiles à détecter, du fait de la faible puissance de détection des dispositifs expérimentaux. Ainsi, si la dominance semble être la règle chez les espèces allogames, les interactions épistatiques sont souvent faiblement significatives. Si ce n'est pas un argument en faveur du modèle génétique additif, cela ne veut pas dire non plus que le modèle génétique additif n'a pas de sens biologique. Simplement, on n'en sait rien.

Enfin, il ne faut pas confondre la valeur d'un allèle avec son effet statistique dans une population. A un locus biallélique, si les valeurs des trois génotypes possibles sont $aa : 1-s$, $Aa : 1$, et $AA : 1+s$, alors, la valeur de l'allèle A est $(1+s)/2$, alors que son effet dans la population est $\alpha_A = (1-p)s$, où p est la fréquence de l'allèle A.

3.3 Modèle multiplicatif

Un modèle génétique est multiplicatif lorsque les valeurs des allèles se multiplient pour donner le phénotype (Tableau 2). Ce modèle est souvent proposé lorsque l'on observe une liaison moyenne-variance que l'on peut corriger par une transformation logarithmique du caractère (Enfield 1980). De façon plus générale, les caractères de productivité, que l'on imagine fortement corrélés à la fitness, sont souvent des caractères produits. Par exemple, le rendement en grain chez une plante est le produit du nombre de grains par leur poids (Grafius 1964).

Il s'agit d'un modèle épistatique, puisqu'il suffit d'un seul allèle délétère, de valeur nulle, pour que le caractère ne s'exprime pas. L'épistasie peut être qualifiée dans ce modèle de «plus qu'additive», puisque la valeur d'un caractère croît de façon exponentielle avec le nombre d'allèles favorables. Cependant, ce modèle génère peu d'épistasie au sens statistique dans

une population panmictique (Cockerham 1959). Il s'agit là d'une bonne illustration du fait que le modèle linéaire statistique ne permettent pas de décrire correctement l'épistasie. En particulier, en modèle génétique additif, la variance statistique additive augmente de façon linéaire avec le coefficient de consanguinité ; En modèle multiplicatif, c'est le coefficient de variation additive qui augmente de façon linéaire avec la consanguinité (Dillmann et Foulley 1998). Ces résultats sont en accord avec des observations faites sur la luzerne (Gallais 1989).

Avec un modèle génétique multiplicatif, les fréquences des allèles évoluent plus vite sous l'effet de la sélection qu'avec un modèle additif (Dillmann 1992). De plus, la sélection ne crée pas, dans ce cas, de déséquilibre de liaison entre les locus, s'il n'existe pas au départ :

$$D_{\tau+1} = \frac{1-r}{\bar{w}} (1-s)(1-t) D_\tau \tag{16}$$

Pour cette raison, c'est souvent le modèle de référence en génétique des populations.

Comme l'on très bien montré Wade *et al.* (2001), la perception de l'épistasie dépend complètement du modèle de référence que l'on choisit. Si le modèle de référence est le modèle additif, alors le modèle multiplicatif est épistatique, car $(1+t)(1+s) = 1+t+s+ts$, où le produit ts est un terme d'interaction. Inversement, si le modèle de référence est le modèle multiplicatif, le modèle additif peut être décrit comme un modèle multiplicatif avec épistasie négative : $1+t+s = (1+t)(1+s) - ts$. Par la suite, nous nous en tiendrons aux définitions de la dominance et de l'épistasie données plus haut. Au sens de ces définitions, le modèle génétique multiplicatif est bien épistatique.

Enfin, nous pensons, tout comme Wade *et al.* (2001), que s'il est bien avéré que certains *caractères* se multiplient pour donner la fitness, il est assez improbable que les valeurs des *gènes* se multiplient pour donner le caractère. Il y a plusieurs raisons à celà. 1/ Il est assez difficile d'affecter des valeurs numériques à des allèles agissant de façon multiplicative sans que la gamme de variation des caractères résultants sorte des limites communément admises. 2/ De l'épistasie multiplicative a rarement été mise en évidence dans des expériences de détection de QTL. Nous connaissons un cas dans la littérature, qui concerne la teneur en sucres solubles chez la tomate (Eshed et Zamir 1996). 3/ Le modèle multiplicatif génère de la sous-dominance (Tableau 2), ce qui n'est pas non plus très réaliste. 4/ Même en considérant de l'additivité intra-locus, les fréquences des allèles évoluent extrêmement vite sous l'effet de la sélection, et l'on peut penser que les gènes à effets multiplicatifs éventuels seraient rarement polymorphes dans les populations naturelles. Cependant, ce modèle reste tout à fait intéressant pour étudier les interactions entre les caractères qui se combinent pour donner la valeur sélective.

3.4 Le modèle métabolique

Le modèle métabolique est biologiquement très réaliste puisqu'il décrit les relations entre des enzymes et le flux de produit à travers la chaîne métabolique dans laquelle ces enzymes sont impliquées. Les flux constituent des caractères quantitatifs primaires qui interviennent forcément d'une façon ou d'une autre dans la valeur sélective.

Théorie du contrôle métabolique La théorie du contrôle métabolique a été développée simultanément par Kacser et Burns (1973) et Heinrich et Rappoport (1974), et a permis d'établir des propriétés générales des systèmes biochimiques. Les quantités d'enzymes et leurs constantes cinétiques sont les paramètres du système, soumis à la variation génétique, tandis

que les flux et les concentrations en métabolites sont les caractères observés. On peut décrire le flux comme une fonction hyperbolique de l'activité E_i d'une enzyme :

$$J = \frac{S}{\frac{1}{E_i} + \sum_{j \neq i} \frac{1}{E_j}} \tag{17}$$

Cette relation a plusieurs conséquences importantes concernant les bases génétiques de la variation des caractères quantitatifs :

1/ Tout d'abord, elle fournit une base biochimique à la récessivité des mutations délétères, et à la dominance en général (Kacser et Burns 1981).

2/ Si la sélection naturelle tend à maximiser le flux, elle va sélectionner des enzymes très actives qui, au delà d'une certaine activité, n'auront plus d'effet génétique sur le flux (sélection naturelle de la neutralité sélective, Hartl *et al.* 1985).

3/ Ce système génère des interactions épistatiques fortes entre enzymes. L'écriture du coefficient de contrôle :

$$C_i^J = \frac{\partial J}{\partial E_i}\frac{E_i}{J} = \frac{\frac{1}{E_i}}{\sum_{j=1}^{n} \frac{1}{E_j}} \tag{18}$$

montre que la sensibilité du flux à une variation relative de l'activité d'une enzyme donnée dépend de l'activité de toutes les autres enzymes de la chaîne. Comme $\sum_{i=1}^{n} C_i^J = 1$ pour un génotype donné, il ne peut y avoir qu'une seule enzyme vraiment limitante du flux (C_i^J élevé), mais cette enzyme n'est pas forcément toujours la même si l'on regarde une collection de génotypes.

La nature de ces interactions épistatiques dépend du sens de variation du flux (Keightley 1996). Pour une augmentation du flux, l'épistasie est de type synergique ; Pour une diminution du flux, l'épistasie est de type antagoniste. En d'autres termes, il est plus facile d'augmenter le flux que de le diminuer.

Pourtant, à une génération donnée, on révèle majoritairement, dans les composantes de la variance génétique d'une population, des effets additifs et de dominance (Bost, 1999). Dans ce système, les interactions épistatiques entre enzymes se traduisent par le fait que l'on s'attend «naturellement» à une distribution leptokurtique de la part de variance génétique expliquée par chaque enzyme du flux. Les enzymes expliquant la plus forte part de variance additive sont celles qui ont le plus fort contrôle chez les parents (Bost *et al.* 1999). Le modèle métabolique fournit ainsi des bases biochimiques à la distribution en L des effets des QTL.

Jusqu'à présent, la théorie du contrôle métabolique s'est développée en considérant des enzymes indépendantes. Cependant, si la quantité des enzymes est régulée, alors il suffit d'une seule enzyme régulée négativement avec l'enzyme observée pour que le flux diminue lorsque la quantité de l'enzyme observée augmente (Figure 3). Il existe alors une répartition des quantités de toutes les enzymes de la chaîne qui maximise le flux. Dans le cas particulier de la compétition pour une quantité totale de protéines constante, toutes les régulations sont négatives et le flux acquiert des propriétés particulières (de Vienne *et al.* 2001).

L'étude de l'évolution de ces systèmes sous l'effet de la sélection naturelle montre que, dans le cas où la seule interaction entre les enzymes est la compétition pour les protéines, une enzyme en excès est toujours plus fortement contre-sélectionnée qu'une enzyme en déficit, ce qui introduit un principe d'économie dans la cellule. Par ailleurs, dans un cas plus général, la sélection pour maximiser le flux favorise les systèmes non régulés (Gabriel *et al.* in prep). Ces résultats ont été obtenus de façon analytique, en ne considérant qu'un seul locus variable. Des résultats plus généraux pourront être obtenus par simulation.

Le modèle métabolique constitue pour nous un modèle de référence. D'une part, il va nous permettre de caractériser expérimentalement les relations entre la variabilité génétique quantitative des enzymes et les flux contrôlés par ces enzymes (thèse Julie Fievet). D'autre part, nous l'utiliserons, dans des simulations, pour tester les prédictions sur l'évolution des caractères quantitatifs établies à partir du modèle génétique additif.

3.5 Modèles de sélection

Pour modéliser la réponse à la sélection, il faut également décrire la relation entre le phénotype et la valeur sélective.

La sélection peut être stabilisante lorsqu'il existe un phénotype optimal, ou directionnelle lorsque l'optimum est la valeur maximale du caractère. Une sélection pour un optimum o se modélise le plus souvent de la façon suivante : $w_i = exp\left(-\frac{(P_i-o)^2}{2s^2}\right)$, où s représente l'intensité de la sélection, w la valeur sélective et P le phénotype.

La sélection par troncation est une forme particulière de sélection directionelle utilisée en amélioration génétique. On sélectionne tous les individus dont la valeur phénotypique est supérieure à un certain seuil. La valeur sélective vaut alors 0 ou 1.

A part le cas très particulier de la sélection directionnelle avec valeurs sélectives proportionnelles aux valeurs phénotypiques, tous les modèles sélectifs génèrent de la dominance et de l'épistasie au niveau de la valeur sélective. Une illustration en est donnée dans les tableaux 3 à 5 pour une sélection par troncation.

3.6 Epistasie

Il existe une analogie certaine entre épistasie et déséquilibre de liaison. Le déséquilibre de liaison correspond à des interactions entre les fréquences des allèles à des locus différents. L'épistasie correspond à des interactions entre les valeurs des allèles à des locus différents. Pourtant, on ne dispose pas de bons paramètres permettant de décrire l'importance de ces interactions dans les populations. La raison principale de l'absence de bon descripteur des interactions est très probablement que ces interactions ne sont visibles que dans des dispositifs expérimentaux particuliers (Charcosset *et al.* 2000), lorsque l'on introduit des segments chromosomiques dans des contextes génétiques différents (Eshed et Zamir 1996 ; Monna *et al.* 2002), ou alors par leur conséquence sur l'évolution des caractères quantitatifs.

Ainsi, bien que l'on explique en général une grande part de la variance génétique d'une population par des effets additifs, cela ne veut pas forcément dire que ces effets additifs ont une bonne valeur prédictive de l'évolution du phénotype à moyen ou long terme. Dans l'exemple des tableaux 6 et 7, tiré du modèle métabolique, on voit que c'est toujours l'enzyme B qui explique la plus forte part de la variance génétique du caractère, bien que, au niveau des activités enzymatiques, ce soit l'enzyme A la plus variable. Quelles que soient les fréquences des allèles, plus de 98% de la variance génétique est additive. On s'attend alors à prédire correctement la réponse à la sélection. On voit dans le tableau 7 que le gain génétique réalisé suite à la fixation de l'allèle B peut en réalité être beaucoup moins ou beaucoup plus important qu'espéré, selon les fréquences des allèles dans la population de départ. Les interactions entre locus permettent ainsi d'expliquer, en particulier, pourquoi l'on n'obtient jamais les mêmes estimations de l'héritabilité d'un caractère selon qu'elle soit mesurée dans une population donnée, à partir des covariances entre apparentés, ou à partir de la réponse à la sélection (héritabilité réalisée) (Falconer et Mackay 1989).

Compte tenu des rappels précédents, les paramètres importants des modèles sont bien évidemment le nombre des gènes qui contrôlent la variabilité des caractères complexes, les possibilités de recombinaison entre ces gènes, la distribution de leurs effets et l'impact des mutations.

Dans la suite de ce mémoire, nous proposons notre vision du déterminisme génétique le plus probable pour les caractères quantitatifs et des voies possibles pour le modéliser, en nous appuyant sur les résultats disponibles concernant la détection de QTL, la sélection assistée par marqueurs, les expériences de réponse à long terme à la sélection, et notre propre perception du problème basée sur des résultats de simulations.

4 Ce que nous apprennent les résultats expérimentaux sur l'architecture probable des caractères quantitatifs

4.1 Ce que nous apprend la détection de QTL

La détection de QTL permet d'appréhender une partie du déterminisme génétique des caractères quantitatifs, par la mise en évidence de relations statistiques entre variabilité phénotypique et polymorphisme moléculaire (Kearsey et Pooni 1996 ; Lynch et Walsh 1998 ; Mackay 2001 ; Andersson 2001) Après des années passées à manipuler des caractères complexes sans autre connaissance sur leur déterminisme génétique qu'une «boîte noire» irréaliste, cette méthodologie a évidemment suscité de grands espoirs chez les généticiens quantitatifs. A la grande surprise de ces derniers, les premiers résultats ont mis en évidence une architecture apparente des caractères quantitatifs en complète contradiction avec les hypothèses du modèle infinitésimal : les QTL détectés étaient peu nombreux et d'effets très contrastés, et non pas en grand nombre et d'effets égaux entre eux (pour une synthèse de résultats, voir Kearsey et Farquhar 1998).

Cependant, parallèlement à l'accumulation de résultats expérimentaux de détection de QTL, les réflexions menées sur la méthodologie de détection elle-même, et sur la façon d'interpréter ses résultats permettent d'affirmer que la détection de QTL ne donne pas forcément une image exacte du déterminisme génétique des caractères.

4.1.1 Combien de gènes ?

Dans un croisement donné, on détecte en général de 5 à 10 QTL pour un caractère (Kearsey et Farquhar 1998). Cependant, il a été montré que le nombre des QTL détectables dans un croisement donné dépend de la puissance de détection du dispositif expérimental, et en particulier de la taille de l'échantillon. Or, la limite autorisée par la taille des échantillons couramment utilisés en amélioration des plantes (300) est de l'ordre de 10 (Hyne et Kearsey, 1995). Autrement dit, il est possible que le nombre réel de gènes soit plus élevé, mais on ne peut pas tous les détecter.

Pour un caractère donné, les QTL détectés ne sont pas forcément les mêmes d'un croisement à l'autre, ou d'un milieu à l'autre (Beavis 1994 ; Charcosset *et al.* 2000 ; Goffinet et Gerber 2000). Ceci veut dire que potentiellement, un grand nombre de gènes sont susceptibles de contrôler la variation d'un caractère quantitatif, mais pas dans un croisement donné.

Enfin, les QTL détectés n'expliquent pas forcément toute la variabilité génotypique observée pour le caractère. Dans ce cas, la partie du déterminisme génétique des caractères qui

n'est pas expliquée doit correspondre à de nombreux gènes d'effets faibles, car seuls les QTL d'effets forts sont détectés, et leurs effets sont surestimés (cf ci-dessous).

4.1.2 Quels effets ?

Parmi les QTL détectés, on trouve en général quelques rares QTL d'effet fort, peu d'effet moyen, et beaucoup d'effet faible, la distribution de ces effets étant typiquement «en L» (leptokurtique, Kearsey et Farquhar 1998). Cependant, là aussi, des résultats théoriques ont montré que cette observation pouvait être artefactuelle. Comme seuls les QTL d'effet significatif (supérieur à un certain seuil) sont considérés, les effets de ces QTL ont tendance à être surestimés, sauf si la taille de l'échantillon est très grande (Beavis 1994). De plus, parmi les QTL détectés, l'attribution d'effets trop forts à certains QTL entraîne l'attribution d'effets relativement moins importants aux QTL restants, d'où une distribution des effets apparemment en L, même si les vrais effets des gènes sont tous égaux entre eux. Ce biais sur la distribution observée des QTL est systématique, sauf dans le cas improbable d'une population quasi-infinie, sans déséquilibre de liaison, et pour un caractère non soumis aux effets de l'environnement (Bost *et al.* 2001). On ne peut donc jamais conclure, à partir de la simple observation de la distribution des effets apparents des QTL détectés, si la distribution des «vrais» effets des gènes sous-jacents est en L ou non.

Il convient cependant ici de discuter ce que l'on entend par «vrai» effet d'un gène. On peut le définir comme l'effet d'un gène que l'on pourrait mesurer dans une population idéale au sens statistique du terme : population infinie, sans déséquilibre de liaison, etc. Cependant, cette définition théorique au sens de la génétique statistique n'a qu'une faible valeur opérationnelle, un tel effet n'étant pas possible à mesurer dans une population réelle. Au contraire, on peut penser, étant donnée la puissance prédictive du modèle linéaire additif (voir plus haut) que c'est l'effet estimé du QTL (tous biais inclus) qui sera le plus pertinent dans une population donnée. Simplement, les biais considérés étant versatiles, cet effet estimé n'aura de valeur qu'à très court terme, voire uniquement dans l'échantillon considéré ayant servi à le mesurer, et ne pourra que difficilement être extrapolé pour prédire l'effet du même gène dans un autre échantillon.

4.1.3 Quel linkage ?

La détection de QTL ne permet pas de mettre en évidence directement de vrais gènes ou locus, mais uniquement des segments chromosomiques compris entre deux marqueurs. Même si les méthodes de détection de QTL ont été améliorées pour raffiner la détection (Jansen et Stam 1994 ; Zeng 1994) la précision sur la taille de ces segments est toujours au minimum de plusieurs centimorgans, voire plusieurs dizaines de centimorgans chez les plantes. Un seul de ces segments peut contenir plusieurs «vrais» gènes, qui apparaissent confondus en un seul QTL lors de la phase de détection statistique. Ceci a d'ailleurs pu être mis en évidence expérimentalement dans certains cas, par des techniques permettant de «découper» ces segments (Eshed et Zamir 1995 ; Monna *et al.* 2002 ; Steinmetz *et al.* 2002).

Ainsi, même si la détection de QTL a pu dans certains cas (et après un travail supplémentaire) conduire à la mise en évidence d'un déterminisme réellement simple (peu de QTL d'effets forts) pour certains caractères qu'on aurait pu supposer plus complexes (ex. Doebley et Stec 1991), les expériences de détection de QTL ne donnent pas en général d'information

sur le nombre de gènes impliqués dans la variation d'un caractère quantitatif : il y a des QTL que l'on ne détecte pas, et un QTL donné peut correspondre à plusieurs gènes.

4.2 Ce que nous apprend la sélection assistée par marqueur

La sélection assistée par marqueur (SAM) consiste à utiliser, tant qu'il se maintient dans les populations sélectionnées, un déséquilibre de liaison entre des marqueurs moléculaires et des QTL pour améliorer le phénotype, en sélectionnant sur la présence d'allèles aux marqueurs. Ce principe est utilisé à des degrés divers depuis les années 80 en sélection artificielle car on peut, grâce aux marqueurs moléculaires, s'affranchir des aléas de l'expérimentation agronomique, ou réaliser des sélections précoces. On peut ainsi gagner du temps dans le processus de sélection, et donc réaliser un progrès génétique annuel plus important. Classiquement, à partir d'une population en ségrégation (F2, BC1, multiparentale, ...) un programme de SAM comprend l'estimation des «effets» associés à chaque allèle au marqueur sur le phénotype (ex., par détection de QTL, ou régression multiple), et la sélection sur la base de ces effets, combinés ou non avec la valeur phénotypique. L'étape de détection peut être distincte ou non de la sélection selon des schémas plus ou moins complexes (Dekkers et Hospital, 2002).

A l'heure actuelle, quelques expériences de sélection assistée par marqueurs sont publiées, notamment chez les plantes (Hospital 2002) avec des résultats mitigés. En effet, s'il semble relativement facile d'introgresser un ou deux gènes bien caractérisés dans un nouveau contexte génétique à l'aide de marqueurs moléculaires, et si les programmes visant à augmenter la fréquence d'allèles favorables à des QTL pour des caractères «simples» (contrôlés par peu de gènes à effets forts) ont donné des résultats satisfaisants, en revanche les programmes visant à manipuler des caractères complexes semblent avoir souvent fourni un progrès génétique inférieur, voire opposé, à l'attendu (ex., Bouchez *et al.* 2002). Pour de tels caractères complexes, la sélection assistée par marqueurs apparaît alors moins efficace qu'une sélection phénotypique classique.

Comment peut-on interpréter ce résultat, et quelles conclusions peut-on en tirer sur le déterminisme des caractères ? D'évidence, si tous les caractères étaient contrôlés par peu de gènes d'effets forts et bien identifiés, alors la SAM serait toujours plus efficace que la sélection phénotypique (car la sélection sur marqueurs permet d'augmenter l'héritabilité apparente du caractère et donc de s'affranchir d'une partie de la variation environnementale qui diminue l'efficacité de la sélection classique). Doit-on en conclure que le manque d'efficacité de la SAM pour des caractères plus complexes est nécessairement dû à un déterminisme génétique essentiellement polygénique (beaucoup de gènes d'effets faibles) ? En fait, plusieurs raisons peuvent être invoquées.

4.2.1 Les QTL sont trop nombreux

Le manque d'efficacité de la SAM peut s'expliquer par le fait que les gènes gouvernant le caractère sont en grand nombre. En effet, si les QTL sont trop nombreux, alors :

On ne peut pas les détecter tous L'étape de détection des effets associés aux marqueurs dans un programme de SAM souffre bien évidemment des mêmes limitations que la détection de QTL (problème de puissance du dispositif, voir plus haut). On montre alors que pour des caractères complexes, lorsque le nombre de gènes sous-jacents augmente, la SAM devient moins efficace que la sélection phénotypique classique (Bernardo 2001).

On ne peut pas fixer tous les allèles favorables Le génome étant de taille finie, plus les QTL sont nombreux, plus ils sont liés génétiquement et donc, du fait de l'effet Hill-Robertson (cf section 2.3.3), plus il est difficile de fixer simultanément des allèles favorables à tous les QTL. Ainsi, lorsque le nombre de QTL augmente, la probabilité de fixation d'allèles défavorables augmente également. Même si tous les QTL sont bien marqués et bien identifiés, sélectionner efficacement un grand nombre (50) de QTL liés nécessite alors au moins 10 générations et une optimisation particulière de la méthode de sélection. Même alors, on n'obtient pas une efficacité de 100%, sauf s'il n'y a pas de recombinaison entre marqueurs et QTL (Hospital *et al.* 2000).

On pondère mal leurs effets Lorsque le nombre des gènes impliqués dans la variation d'un caractère quantitatif augmente, le pourcentage de QTL dont l'effet est sous-estimé augmente (Beavis 1994 ; Bost *et al.* 2001). Du fait de ce «biais de sélection des QTL», l'index de sélection accordera alors un poids trop fort aux QTL de fort effet (Moreau *et al.* 1998). Les QTL de faible effet et les QTL non détectés deviennent ainsi quasiment neutres vis à vis de la sélection. La SAM est alors plus efficace que la sélection phénotypique à court terme, car elle fixe plus vite les QTL de fort effet, mais moins efficace à moyen ou long terme terme car elle fixe plus d'allèles défavorables aux QTL de faible effet (Hospital *et al.* 1997), alors que la sélection phénotypique équilibre mieux la sélection entre gros et petits QTL. Dans la pratique, on peut résoudre ce problème en calibrant les poids respectifs des marqueurs et du phénotype dans l'index de sélection en fonction de l'objectif en terme de gain génétique attendu ou de nombre de générations (Dekkers et van Arendonk 1998 ; Dekkers *et al.* 2002).

Ces mécanismes ont été parfaitement illustrés par Bernardo (2001), qui a simulé des expériences de SAM proches de celles pratiquées en sélection végétale, en considérant des nombres croissants de gènes effectivement connus. L'auteur en conclut que la sélection phénotypique classique (sans marqueurs) reste le meilleur moyen de manipuler ces caractères dans les programmes d'amélioration.

4.2.2 Les effets des QTL sont mal estimés dans l'espace et le temps

La SAM appliquant une sélection forte sur certains marqueurs, on peut aussi penser que son manque d'efficacité vient de ce que les marqueurs sont mal choisis, ou que les poids (effets) qui leur sont attribués sont incorrects, autrement dit la SAM ne serait pas efficace parce que les QTL sont mal détectés. Il convient ici de distinguer les problèmes liés à l'estimation des *positions* des QTL, de ceux liés à l'estimation de leurs *effets*.

Estimation des positions des QTL La position d'un QTL n'est connue qu'approximativement, avec une certaine précision qui dépend de la puissance du dispositif. Si la taille de l'échantillon est faible, l'estimation des positions des QTL détectés peut être entachée d'une assez grande erreur, ce qui se traduit par des intervalles de confiance sur la position des QTL (pour autant qu'on sache les calculer, Mangin *et al.* 1994) de plusieurs dizaines de cM, voire le chromosome entier. Même si les gènes en cause sont peu nombreux et peu liés entre eux, les effets d'échantillonnage et les associations entre gènes font qu'on peut parfois détecter un «faux» QTL dans un intervalle vide situé entre deux intervalles, éventuellement éloignés, contenant eux de «vrais» QTL qui ne sont pas détectés. Par exemple, les marqueurs sélectionnés préférentiellement par régression multiple dans le cadre de la SAM ne sont pas forcément

les plus liés génétiquement aux gros QTL (Gimelfarb et Lande 1995). Un QTL de ce genre n'a au fond aucune base biologique.

Cependant, l'incertitude sur la position des QTL ne suffit pas à elle seule à expliquer le manque d'efficacité de la SAM. En effet, il est toujours possible d'améliorer l'estimation des positions des QTL en augmentant la précision du dispositif (ex., taille de l'échantillon), et les méthodes modernes de détection sont désormais capable de séparer plusieurs QTL faiblement liés sur un même chromosome. Surtout, même si les marqueurs sélectionnés n'ont pas de réalité biologique, ils maximisent néanmoins la réponse à la sélection à un instant donné comme on va le voir ci-dessous.

Le paradoxe de la SAM Le problème de la prédiction de la réponse à la sélection, en particulier dans un programme de SAM, se heurte à un paradoxe apparent :

D'une part, si les gènes gouvernant le caractère sont très nombreux, on ne peut pas tous les détecter, mais il suffit d'un très faible nombre de marqueurs pour expliquer efficacement à court terme une part importante de la variabilité génétique du caractère. Un seul marqueur localisé aléatoirement suffit à expliquer environ 40% de la variabilité génétique d'un chromosome (Dekkers et Dentine 1991). Dans Bouchez *et al.* (2002), une méthode sophistiquée de détection de QTL (Composite Interval Mapping) ne donne pas de résultats vraiment meilleurs qu'une méthode plus frustre (ANOVA).

D'autre part, même si on pouvait détecter tous les gènes sous-jacents, on ne pourrait pas tous les contrôler simultanément, mais à court terme la sélection sur seulement quelques marqueurs bien répartis est très efficace. Dans un programme de backcross, il suffit de peu de marqueurs (2 ou 3) pour contrôler efficacement le fond génétique sur tout un chromosome (Servin et Hospital, 2002) ; de même, 3 à 4 marqueurs suffisent à contrôler la présence de l'allèle favorable d'un QTL de position exacte inconnue sur un intervalle de confiance de plusieurs dizaines de centimorgans (Hospital et Charcosset 1997). Inversement, sans sélection, des segments chromosomiques indésirables de grande longueur vont persister longtemps autour des gènes introgressés par backcross, malgré l'apport récurrent de gènes favorables (Naveira et Barbadilla 1992), et la réduction de ces segments par la sélection sur marqueurs nécessite des moyens importants en taille d'échantillon et/ou en temps (Hospital 2001).

A court terme, les marqueurs ne sont donc pas un moins bon résumé du génotype que le phénotype (contrairement à ce qu'on pourrait conclure des résultats de Bernardo 2001). D'ailleurs, Laurence Moreau (comm. pers.) a pu montrer que, à court terme, la variation de fréquence aux marqueurs après sélection purement phénotypique est quasiment identique à celle observée après sélection sur marqueurs seuls. Mais, ce résumé est évanescent (instable dans le temps), pointe éventuellement sur des gènes «virtuels» (cf section 3.2) et doit être ré-évalué à chaque génération.

Ce paradoxe apparent permet en particulier d'expliquer pourquoi on peut résoudre en partie les problèmes de puissance de détection et de sous-estimation des petits effets des QTL, donc améliorer l'efficacité de la SAM, en acceptant un plus fort risque de première espèce (détection de «faux» QTL, Moreau *et al.* 1998), voire en prenant en compte, dans l'index de sélection, l'ensemble des marqueurs dont on dispose, y compris les non-significatifs (Meuwissen *et al.* 2001 ; Lange et Whittaker 2001). En effet, il faut surtout éviter le risque que des régions chromosomiques, voire des chromosomes entiers, ne soient pas marqués, quitte à sur-paramétrer le modèle.

Ainsi, il ne semble pas que le problème principal soit une mauvaise estimation des positions

des QTL. Du reste, il est rare que dans les expériences de SAM publiées à ce jour on ait observé la «disparition» d'un QTL, le segment contrôlé n'intervenant plus sur la variabilité du caractère à la fin du programme (Hospital, 2002). Par contre, il est plus fréquent que le segment considéré conserve un effet sur le caractère, mais que l'amplitude, voire le signe, de cet effet ne soit pas conforme à l'attendu. Il ne semble pas non plus que le problème vienne de l'estimation «instantanée» des effets des QTL, mais plutôt de la stabilité de ces effets au cours du temps, c'est à dire lorsque le QTL va se trouver dans une autre population (échantillon, dérive), un autre fond génétique (recombinaison), et/ou un autre environnement.

Instabilité des effets des QTL Plusieurs causes permettent d'expliquer l'instabilité des effets associés aux marqueurs :

Recombinaison marqueur-QTL Sauf dans des cas favorables rares, le QTL n'est jamais contrôlé directement par un marqueur «interne», mais indirectement par un marqueur «externe», plus ou moins fortement lié génétiquement au QTL *et* en déséquilibre de liaison avec lui. Cette association marqueur-QTL va diminuer sous l'effet des recombinaisons et de la variation des fréquences alléliques, ce qui réduit effectivement l'efficacité de la sélection sur le marqueur. Il est cependant douteux que cet effet soit la cause essentielle du manque d'efficacité de la SAM, d'une part car il est toujours possible de mieux marquer les QTL (avec plus de marqueurs liés plus fortement), et d'autre part car la recombinaison marqueur-QTL est marginale (Hospital *et al.* 2000).

Variations des fréquences et recombinaison inter-QTL En supposant que le QTL soit bien marqué et son effet bien estimé, la sélection sur ce QTL va être efficace, et conduire à l'augmentation de la fréquence de l'allèle favorable et à terme à la fixation du QTL, ce qui fait que son effet dans la population va diminuer. Si le QTL n'explique pas toute la variabilité génétique pour le caractère, le progrès génétique ne pourra être poursuivi qu'en sélectionnant d'autres QTL, d'effets plus faibles. Si la distribution des effets des QTL n'est pas linéaire, la pondération des QTL d'effets faibles restant polymorphes n'est plus correcte.

D'autre part, dans une population réelle, l'effet apparent d'un QTL inclus un biais dû aux effets des autres QTL avec lesquels il est en déséquilibre de liaison (Bost *et al.* 2001). Même si c'est l'effet biaisé qui est pertinent (cf plus haut) ce biais va évoluer en même temps que le déséquilibre de liaison et les fréquences des QTL, et les effets doivent être ré-estimés.

Epistasie entre QTL Une façon plus générale de dire que les effets détectés associés aux marqueurs résultent d'une combinaison d'effets de gènes différents, même peu liés entre eux voire sur des chromosomes différents, est de dire que ces gènes sont en relation épistatique. On sait que l'épistasie au sens biologique existe et pourrait annuler les effets de la sélection en conduisant à de mauvaises prédictions (cf plus haut 3.6). D'ailleurs c'est l'hypothèse que cherche à tester l'équipe d'Avignon travaillant sur la résistance multigénique chez le piment (collaboration, cf Travaux FH).

Recombinaison intra-QTL Un QTL peut correspondre on l'a vu à un segment chromosomique, et non pas à un locus. Un tel segment peut contenir de nombreux vrais gènes. Dans ce cas, l'effet attribué au segment résulte d'une combinaison des effets des gènes qu'il contient, combinaison qui va s'altérer sous l'effet des recombinaisons à l'intérieur du segment.

Le problème est ici encore celui d'un ensemble de gènes dont on estime les effets collectivement, et pas indépendamment, mais sa solution se heurte cette fois à une limite structurelle due à la recombinaison. En effet, il a été montré qu'augmenter le nombre de marqueurs au-delà d'une certaine densité (qui dépend encore une fois de la puissance du dispositif, donc de la taille de l'échantillon) ne fait qu'ajouter du bruit de fond sans augmenter la précision de la détection de QTL, les marqueurs proches étant trop corrélés entre eux. La limite minimale de l'intervalle entre deux marqueurs est de quelques centimorgans pour les plus grandes tailles réalistes d'échantillons, ce qui reste grand par rapport à la densité supposée des gènes.

Il s'agit d'un problème de résolution génétique qui renvoie à la détermination de l'unité génomique de sélection (cf conclusion). Cette cause n'est pas à négliger, en effet nous avons des arguments évolutifs pour dire que c'est un cas probable (cf 6.1) et les résultats expérimentaux s'accumulent montrant que le découpage fin d'un segment QTL fait apparaître plusieurs QTL à l'intérieur du segment (cf 4.1.3).

Interactions génotype×environnement L'autre cause principale d'instabilité apparente des effets associés aux marqueurs dans les programmes de SAM est l'existence d'interactions génotype×environnement, ou plus précisément QTL×environnement, l'environnement étant pris ici au sens large (toute cause non génétique de variation du phénotype : lieu, année, âge, etc.). L'existence d'interactions QTL×environnement est connue (Moreau *et al.* 2002) et souvent invoquée pour expliquer l'echec de la SAM (Hospital 2002). Au cours de la sélection, les populations vont se trouver dans un nouvel environnement, et si les QTL qui s'y expriment ne sont pas les mêmes que ceux détectés dans l'environnement d'origine, alors la SAM sera moins efficace que la sélection phénotypique, qui elle s'ajuste instantanément. Notons cependant que cet argument rejoint celui de l'existence d'un nombre de gènes réellement impliqués dans la variation du caractère plus grand que le sous-ensemble des QTL détectés au départ, puisqu'on considère ici un «pool» de gènes dont seulement une partie s'exprime dans un environnement donné (voir par exemple Bertin et Gallais 2001).

On voit donc que les explications avancées, sauf celles triviales faisant référence à un manque de précision du dispositif, font appel soit à un nombre de gènes en cause nombreux, soit à de l'épistasie. Si les effets de ces gènes peuvent être «résumés» dans une population donnée par quelques facteurs génétiques (locus à effets additifs, éventuellement virtuels) ce résumé doit être reconsidéré dans chaque nouvelle population.

4.3 Ce que nous apprennent les expériences de sélection à long terme

Les expériences de sélection à long terme sont en général un succès. On observe en général un progrès soutenu et linéaire (Yoo 1980 ; Dudley et Lambert 1992 ; Weber 1996). Dans certains cas plus rares, une limite est atteinte, dont on peut penser que c'est une limite physiologique (Dudley et Lambert 1992), ou alors que la sélection artificielle est antagoniste à la valeur sélective (Falconer et King 1953 ; Eklund et Bradford 1977). On voit mal comment une réponse soutenue pourrait être expliquée uniquement par peu de gènes d'effets forts présents au départ, sans apport de variabilité nouvelle. D'ailleurs, dans des expériences de sélection divergente, on estime à plusieurs centaines le nombre de gènes qui ont contribué au progrès génétique sur plusieurs dizaines de générations (Dudley et Lambert 1992).

Simplement, ici encore, si les gènes en cause sont nécessairement nombreux, il n'interviennent pas tous simultanément en permanence dans la variabilité du caractère (puisqu'on

a vu que seul un petit nombre de facteurs génétiques est identifiable dans une population, et qu'il n'est pas possible de sélectionner simultanément sur de nombreux gènes). Il y aurait donc beaucoup de gènes qui interviennent sur la durée totale du processus de réponse à la sélection, mais pas tous en même temps. Comme le soulignent Barton et Keightley (2002), la question est de savoir 1/ si ces gènes étaient tous présents au départ ou s'ils sont apparus au cours du temps par mutation (ou migration) ; 2/ s'ils ont de gros ou petits effets. Nous reviendrons sur les hypothèses possibles dans deux chapitres séparés ci-dessous. En effet, si l'apport de variabilité par mutation est une cause évidente à considérer, il nous semble également nécessaire de bien détailler les autres hypothèses permettant d'expliquer une réponse soutenue à la sélection, notamment l'existence possible de liaisons génétiques fortes entre les gènes en cause.

5 Rôle des mutations dans la réponse à la sélection

5.1 Mesures expérimentales

Deux grands types d'expériences sont envisageables pour mesurer l'apport de variabilité génétique par mutation à partir d'un matériel génétique fixé (Bataillon 2000). Les expériences d'accumulation des mutations consistent à dériver un grand nombre de lignées, en limitant au maximum les effets de la sélection, pour mesurer toute la variabilité génétique créée entre lignées par la mutation. Les expériences de sélection divergente permettent de mesurer, à partir des réponses à la sélection basse et haute, la variance des effets des mutations utilisables pour la sélection.

Nous menons, depuis six années, une expérience de sélection divergente pour la précocité dans des lignées pures de maïs, pour laquelle nous observons déjà des résultats. L'héritabilité mutationelle estimée varie entre 0 et 0.032 selon les populations. On observe également une réponse corrélative sur la hauteur des plantes, qui est un caractère corrélé négativement à la précocité pour des raisons physiologiques. La figure 4 illustre l'apparition de différences génétiques pour la hauteur entre les familles d'une même population à la sixième génération.

Nos estimations de l'héritabilité mutationelle sont cohérentes avec celles de la littérature (Houle *et al.* 1996). Cette expérience confirme que la mutation peut être une pression évolutive importante dans les populations naturelles mais aussi, à plus court terme, dans les populations de sélection artificielle.

5.2 Mutation et adaptation

Nous avons également étudié, par simulation, le rôle de la mutation dans l'adaptation des populations à de nouvelles conditions de milieu (sélection pour un optimum lointain). Contrairement aux études existantes, nous ne cherchions pas à caractériser l'état d'équilibre de la population, mais à décrire la dynamique des fréquences des allèles à partir d'une population initiale, que nous avons choisie entièrement polymorphe.

Les résultats préliminaires montrent que l'adaptation semble se dérouler en trois phases, de durées variables selon la taille de la population, la variance génétique initiale et le taux de mutation (Figure 5). Ces trois phases sont observées quelle que soit la valeur de l'optimum. Durant la première phase, on observe un épuisement très rapide de la variabilité génétique initiale et au delà d'une vingtaine de générations, la variance génétique initiale ne contribue plus au progrès génétique (comme l'observent également Barton et Keightley 2002). Durant la

seconde phase, des mutations favorables apparaissent séquentiellement et se fixent rapidement. Le polymorphisme est à son niveau le plus bas. La troisième phase est une phase d'atteinte de l'équilibre dérive-sélection-mutation. La variabilité génétique ré-augmente alors progressivement, constituée en partie par des allèles faiblement délétères apparus par mutation. En négligeant les locus auxquels se maintiennent des allèles rares, le pourcentage de locus polymorphe à l'équilibre reste faible (10 à 15%). Ce résultat est comparable à ceux obtenus par Bürger et Lande (1994).

A l'équilibre, la distribution des valeurs des allèles ne semble dépendre que de la forme de la fonction de sélection (qui est ici une gaussienne), et non de la distribution des effets des mutations. Ces résultats sont partiellement en accord avec les prédictions des modèles théoriques (2.5). Si l'on s'intéresse au sous-ensemble des locus polymorphes, on observe une distribution leptokurtique de la part de variance expliquée par chaque locus (Figure 6). Le même genre de distribution est observée pour les effets des mutations destinées à se fixer (Orr, 1998).

On peut tirer plusieurs conclusions de ces résultats préliminaires. Tous d'abord, il est frappant de voir à quel point la variance génétique initiale s'épuise vite sous l'effet de la sélection. Dans nos simulations, il est impossible de prédire la capacité d'adaptation d'une population connaissant sa variance initiale. Cependant, nous avons considéré des tailles de population plutôt faibles (effectif démographique entre 300 et 3000). Si ce résultat se confirmait, il pourrait avoir des conséquences importantes sur les programmes de gestion de la variabilité génétique. D'autre part, il semble que l'on puisse, à un instant donné, décrire une population à l'aide d'un petit nombre seulement de gènes polymorphes ayant des effets inégaux. Au cours du temps, ce ne sont jamais les même gènes qui restent polymorphes, car ils se fixent et des mutations apparaissent ailleurs dans le génome.

6 Rôle du linkage dans la réponse à la sélection

On peut conclure des rappels précédents (section 2) que le linkage peut avoir un impact sur la dynamique de la réponse à la sélection, notamment par les interactions sélection-recombinaison-dérive. Même si ces effets sont très difficiles à modéliser, donc encore mal décrits, il nous semble assez évident que les effets du linkage doivent être pris en compte pour modéliser et prédire la réponse à court terme à la sélection, notamment dans une population avec beaucoup de gènes polymorphes. C'est en particulier le cas des programmes de SAM où l'on cherche à tirer le meilleur parti à court terme de tous les gènes disponibles au départ dans un croisement. Ces effets doivent-ils également être pris en compte pour prédire la réponse à long terme ?

La liaison génétique limite la réponse à la sélection, en augmentant la probabilité de fixation des allèles défavorables liés (effet Hill-Robertson) et donc en réduisant la probabilité de fixation des allèles favorables. Par ailleurs, il semble que dans certains cas un équilibre transitoire puisse s'établir entre d'une part la fixation d'allèles défavorables, et d'autre part la régénération de variabilité par recombinaison. Dans de tels cas, le linkage, par ses interactions avec la sélection et la dérive, est capable à lui seul, non pas de maintenir le polymorphisme, mais de retarder sa disparition (dans un système sans mutation) et de modifier sa dynamique par rapport à un système sans liaison (Hospital et Chevalet 1996). Cependant, il semble que les effets du linkage, notamment à moyen ou long terme, n'induisent une différence importante, par rapport à un système sans liaison (recombinaison libre), que pour des taux de recombinaison très faibles.

La question est alors de savoir si une telle situation (l'existence de gènes fortement liés) a une réalité biologique, qualitativement importante dans la dynamique des systèmes, et quantitativement significative sur une échelle évolutive.

Les modèles de Génétique des Populations (cf plus haut) fournissent des prédictions soit sur la dynamique (évolution temporelle) de systèmes liés à deux locus, soit sur l'état d'équilibre de systèmes avec de nombreux locus. Dans le cadre de la thèse d'Emmanuelle Della-Chiesa, nous avons choisi une autre voie d'investigation : comment évolue dans l'espace (génomique) et le temps un système comportant au départ de nombreux gènes liés.

6.1 Structuration «spatiale» du polymorphisme le long des chromosomes

Les premiers résultats indiquent qu'une pression de sélection s'exerçant sur un système polygénique dense et polymorphe au départ conduit à une agrégation apparente du polymorphisme, le polymorphisme étant maintenu plus longtemps dans les régions à forte densité de gènes par unité de recombinaison que dans les régions à faible densité (figure 7). On peut en conclure que, à partir d'une répartition initiale aléatoire des gènes sur les chromosomes, ces phénomènes doivent rendre plus probable l'observation de liaison forte (plutôt que faible) entre les gènes contrôlant la réponse à la sélection.

Il conviendrait alors de modéliser les génomes, non pas comme un magma uniforme de gènes, mais sous une forme structurée comportant des paquets de gènes très liés, paquets séparés les uns des autres par des taux de recombinaison plus importants. La taille (en unités de recombinaison) de ces paquets et le nombre des gènes qu'ils contiennent restent à déterminer (cf conclusion).

Si ces résultats sont généralisés, une telle structuration du polymorphisme devra être prise en compte pour modéliser la réponse à court terme à la sélection. Prendre en compte cette structuration serait par ailleurs essentiel pour faire des inférences moléculaires dans une population (relation entre polymorphisme neutre et polymorphisme sélectionné), ou pour savoir où dans le génome chercher du déséquilibre de liaison (Ardlie *et al.* 2002).

Cependant, l'agrégation mise en évidence semble transitoire (figure 7) en présence de mutation. La nécessité de la prendre en compte sur une échelle évolutive repose alors sur 1/ l'existence de polymorphisme important au départ dans les populations et 2/ la dynamique temporelle des paquets sous l'effet de la sélection et de la mutation.

6.2 Dynamique temporelle, effet des mutations

Sans mutation, la question est de savoir combien de temps les paquets de gènes liés sont capables de maintenir une quantité suffisante de polymorphisme, autrement dit dans quelle mesure la variabilité génétique (C, cf 2.3.1) masquée par les déséquilibres de liaison négatifs dans les paquets peut être libérée par la recombinaison avant la fixation complète des paquets. A ce niveau, la recombinaison entre gènes très liés étant indifférentiable de la mutation, on retrouverait un système avec quelques gènes quasi-indépendants (les paquets) dont la variabilité serait maintenue par mutation (la recombinaison intra-paquets).

L'autre question est de savoir si les «vraies» mutations participent à l'agrégation du polymorphisme (i.e., la renforcent), ou au contraire la réduisent, ce qui est une autre façon de demander si les différents gènes qui interviennent dans la réponse à la sélection le font simultanément ou séquentiellement. Autrement dit, on considère ici les interactions possibles entre les effets mis en évidence aux sections 5.2 et 6.1.

On peut ici tenter un parallèle entre la sélection artificielle et l'adaptation d'une population naturelle. Supposons une population avec un certain polymorphisme au départ. Maximiser la réponse à la sélection dans cette population revient à fixer tous les allèles favorables présents au départ sans en perdre aucun. Un modèle théorique de prédiction, notamment pour optimiser la SAM, doit donc prendre en compte le linkage entre ces gènes, et tenir compte (modéliser) les interactions sélection-recombinaison-dérive. Ceci peut être réalisé en pratiquant une sélection phénotypique «douce» (intensité de sélection faible), c'est le principe de la sélection récurrente. Dans la pratique, un sélectionneur confronté au problème pourrait aussi choisir l'option inverse : fixer rapidement les gènes présents au départ, quitte à perdre des allèles favorables, puis aller rechercher une nouvelle source de variabilité extérieure à la population, c'est le principe de la sélection généalogique.

Par analogie, on peut se demander quelle source de variabilité sera préférentiellement utilisée par une population naturelle pour répondre à une pression évolutive (ex. adaptation à un changement environnemental). Autrement dit, même si le polymorphisme présent au départ dans la population est suffisant, la population sera-t-elle capable d'exploiter au maximum ce polymorphisme (en interne) pour s'adapter, ou sera-t-elle incapable de le faire et devra-t-elle pour survivre faire appel à d'autres sources de polymorphisme ?

Si le polymorphisme initial est faible et/ou s'il est rapidement épuisé, de telle sorte que son impact sur la réponse est faible comparativement à celui des mutations (les mutations réduisent l'agrégation) alors l'essentiel de la réponse adaptative sera dû à des fixations successives de mutations favorables quasi-indépendantes, dans une population peu polymorphe. C'est ce type de dynamique que nous avons observé dans les simulations (cf 5.2). Ces simulations ont été réalisées avec un modèle de mutation aléatoire (Random Mutation Model), où l'effet d'une mutation ne dépend pas de la valeur de l'allèle qu'elle remplace. Une telle mutation pouvant inverser brutalement le déséquilibre de liaison entre l'allèle ancestral et le fond génétique, le choix du modèle de mutation pourrait expliquer que dans les simulations la mutation réduise l'agrégation du polymorphisme comme l'a observé E. Della-Chiesa. Notons que ce modèle de réponse à la sélection, dans lequel l'impact du polymorphisme initial est négligeable, et l'essentiel de la réponse réalisé dans une population peu polymorphe, ne justifie pas à lui seul de chercher à maintenir la variabilité génétique des populations dans les programmes de conservation, si le seul objectif est de maintenir leur potentiel adaptatif (independamment d'autres considérations ex. consanguinité).

Inversement, on peut penser que dans un modèle incrémental (IMM), la mutation ne réduirait que graduellement le déséquilibre de liaison, et donc réduirait moins l'agrégation du polymorphisme. Du reste, il a été montré que la mutation incrémentale crée des conditions plus propices à l'évolution de la reproduction sexuée (Peck *et al.* 1997), donc un avantage à la recombinaison. Ceci pourrait donc augmenter l'impact du polymorphisme initial dans la réponse adaptative.

On peut enfin se demander si, compte tenu du phénomène d'agrégation du polymorphisme, la migration (ou l'hybridation) est capable de générer une variabilité génétique initiale importante.

6.3 Structuration du polymorphisme inter-populations et migrations

On a vu que l'existence d'une liaison forte entre les gènes ralentit la réponse à la sélection. Ceci peut dans certains cas ralentir la disparition du polymorphisme. La disparition complète du polymorphisme dans une population sans mutation est bien évidemment inéluctable, par

fixation complète d'un allèle à chaque locus. Cependant, le taux de recombinaison entre gènes va également affecter l'état dans lequel ces locus vont se fixer, c'est à dire la phase gamétique (enchaînement d'allèles favorables et défavorable le long d'un chromosome).

Une population sélectionnée directionnellement va avoir tendance à fixer majoritairement des allèles favorables pour les locus faiblement liés, mais au contraire des combinaisons alléliques hétérogènes comportant plus d'allèles défavorables dans des groupes de locus fortement liés entre eux. On pourrait alors penser que deux populations identiques soumises à la même pression de sélection (repliquats) vont plutôt diverger pour le second type de locus. Cette prédiction, si elle est confirmée, aurait deux conséquences importantes sur les différences observables entre populations :

6.3.1 Impact de l'histoire évolutive sur l'architecture apparente des caractères quantitatifs

Les gènes présentant du polymorphisme entre deux populations ne seraient pas les mêmes selon la façon dont ces populations ont été préalablement sélectionnées. Pour détecter des QTL chez les espèces agronomiques ou de laboratoire, on part en général de croisements présentant une quantité suffisante de polymorphisme. Deux stratégies sont possibles : soit croiser du matériel déjà sélectionné (ex., deux lignées «élite» de maïs), soit croiser du matériel cultivé et du matériel sauvage. On peut résumer ces deux stratégies en considérant par exemple qu'à partir d'une même population on tire des lignées «hautes» (H, sélection pour augmenter la moyenne du caractère) ou «basses» (L, diminuer la moyenne). On effectue ensuite des croisements H$\times$L ou H$\times$H$'$. C'est la stratégie employée par exemple par Nuzhdin *et al.* (1999) chez la drosophile. Selon la prédiction ci-dessus, les croisements H$\times$H$'$ devraient avoir tendance à mettre en évidence des paquets de gènes liés. C'est ce que semblent confirmer les résultats de la figure 8. Inversement, les croisements H$\times$L devraient mettre en évidence des gènes isolés. Ceci aurait bien évidemment des incidences sur l'image de l'architecture des caractères quantitatifs que nous renvoie la détection de QTL. Cela permettrait en outre d'expliquer les résultats cités plus haut sur le découpage de segments QTL.

6.3.2 Importance du linkage en populations subdivisées

On peut considérer que les mêmes phénomènes se produiraient dans une population subdivisée. Si la sélection agit dans le même sens dans chaque sous-population, alors les génotypes migrants dans une sous-population auraient tendance à apporter du polymorphisme sous la forme de paquets de gènes liés. Un tel résultat, qui reste encore à démontrer, renforcerait l'importance évolutive du linkage fort, et expliquerait comment la migration, ou l'hybridation entre populations est capable de régénérer une source de polymorphisme important et persistant dans les populations naturelles.

7 Synthèse et perspectives

Lorsqu'on veut modéliser la réponse à la sélection, il y a deux objectifs possibles : soit on cherche à expliquer, soit on cherche à prédire. Si l'on cherche à prédire, le modèle n'a absolument pas besoin de coller à la réalité. Il suffit qu'il marche. Si l'on cherche à expliquer, c'est un peu plus compliqué, car on aimerait que l'explication soit réaliste.

Il ressort des rappels ci-dessus que le modèle statistique additif a une forte valeur prédictive à court terme, quelle que soit la complexité réelle du déterminisme génétique des caractères. Du reste, il a été démontré que l'équation (4) est quasiment toujours vraie sur une génération de sélection (Barton et Turelli 1989). Ceci peut s'expliquer par la puissance prédictive du modèle linéaire.

De façon plus surprenante, on pourrait aussi conclure des résultats ci-dessus que, bien que l'équation (4) ne soit valable que sur une génération, son extension à plusieurs générations (5), qui prédit une réponse linéaire à la sélection, acquiert elle-même une bonne valeur prédictive. Bien que cette extension ne soit valable en toute rigueur que sous les hypothèses biologiquement irréalistes du modèle infinitésimal, il semble que des hypothèses plus réalistes fournissent également une réponse linéaire à la sélection sur plusieurs dizaines, voire centaines, de générations, mais pour des raisons différentes : la réponse à la sélection d'un systèmes comportant de nombreux gènes liés en paquets sur un chromosome est linéaire (section 6.1) ; de même la réponse obtenue par fixations successives de mutations à peu gènes est linéaire (section 5.2).

A la lumière de nos travaux respectifs, il nous semble possible d'établir un modèle prédictif simplifié de l'évolution des caractères quantitatifs. C'est cette voie de recherche que nous souhaitons approfondir dans les années à venir.

7.1 Définir l'unité génomique de sélection

Les différents scénarios évoqués ci-dessus contiennent un dénominateur commun : dans une population donnée, même s'il y a beaucoup de gènes réellement impliqués dans la variation d'un caractère, seul un petit nombre d'entités génétiques vont participer effectivement à la sélection. Ces entités peuvent être des locus additifs «virtuels», mais si les gènes gouvernant le caractère sont réellement très nombreux, et à court terme à partir d'un polymorphisme initial important, ces entités sont plus probablement des régions génomiques contenant chacune plusieurs gènes. L'unité de sélection n'est donc sans doute pas le gène (locus individuel) comme le remarquaient déjà Franklin et Lewontin (1970). Cette unité n'est sans doute pas non plus le chromosome entier, mais un segment chromosomique, dont la taille reste à caractériser en fonction des contraintes liées à la recombinaison et des paramètres du système. Prédire la taille de cette «unité génomique de sélection», et son évolution au cours du temps permettrait de modéliser un système polygénique complexe sous forme de blocs qui pourraient être traités de façon indépendante, et ainsi réduire la complexité mathématique du problème.

7.2 Un modèle oligo-génique transitoire

Comme nous l'avons vu, les différentes interactions, entre valeurs des allèles (épistasie), et/ou entre fréquences des allèles (déséquilibre de liaison) font que l'effet statistique associé à un segment chromosomique est un effet évanescent, qui ne dure pas au delà d'une génération. Sans remettre en cause la robustesse du modèle statistique pour prédire la réponse à la sélection d'une génération à la suivante, ceci confirme la difficulté de prédire l'évolution

à long terme. Cependant, les différentes pressions évolutives aboutissent à ce qu'un nombre réduit de ces facteurs polymorphes soit disponible à une génération donnée pour la sélection et que ces facteurs ne soient pas d'effets égaux. Il serait donc possible de résumer la variation des caractères quantitatifs dans une population grâce à un modèle oligogénique relativement simple, sachant que ce résumé est transitoire, c'est à dire qu'il change à chaque génération, en même temps que la distribution apparente de l'effet des gènes.

7.3 Distribution des effets des gènes

S'il est possible de négliger le linkage entre les paquets de gènes liés, et de ne considérer d'une dizaine de facteurs génétiques polymorphes à chaque génération, on pourrait envisager de caractériser une population à une génération donnée par un modèle probabiliste ne considérant que la distribution des effets des gènes et non leur position, et tenter de prédire comment cette distribution se déforme au cours de l'évolution, afin de reproduire l'évolution observée sur les simulations (figure 9). Ceci nous permettrait, entre autres, de quantifier comment la pression de sélection qui s'exerce sur un segment évolue au cours du temps, et inversement, de suggérer des pondérations variables au cours du temps pour les marqueurs dans les programmes de sélection assistée par marqueurs.

Références bibliographiques

Andersson L. (2001) Genetic dissection of phenotypic diversity in farm animals. Nature Rev. Genet. 2 :130-138.

Ardlie K.G., L. Kruglyak and M. Seielstad (2002) Patterns of linkage disequilibrium in the human genome. Nature Rev. Genet. 3 :299-309

Barton N.H., 1989, Divergence of a polygenic system subject to stabilizing selection, mutation and drift. Genet. Res. 54 :59-77.

Barton N.H. and M. Turelli, 1987, Adaptative landscapes, genetic distance and the evolution of quantitative characters. Genet. Res. 49 :157-173.

Barton N.H. and M. Turelli, 1989, Evolutionary quantitative genetics : How little do we know ? Annu. Rev. Genet. 23 :337-370.

Barton N.H. and P.D. Keightley (2002) Understanding quantitative genetic variation. Nature Reviews Genetics 3 :11-21.

Bataillon T. (2000) Estimation of spontaneous genome-wide mutation rate parameters : wither beneficial mutations ? Heredity 84 :497-501.

Beavis W.D. (1994) In *Proceedings of the Corn and Sorghum Industry Research Conference* 250-266. American Seed Trade Association, Washington DC.

Bernardo R. (2001) What if we knew all the genes for a quantitative trait in hybrid crops ? Crop Sci. 41 :1-4.

Bertin P., and Gallais A. (2001). Genetic variation for nitrogen use efficiency in a set of recombinant inbred lines. II QTL detection and coincidences. Maydica 46 : 53-68.

Bouchez A., F. Hospital, M. Causse, A. Gallais and A. Charcosset (2002) Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. Genetics (In Press)

Bost B., 1999, Analyse de la distribution des effets des gènes dans le cadre d'un modèle métabolique de la variation quantitative. Thèse INAP-G, 184pp.

Bost B., Dillmann C., and de Vienne D. (1999). Fluxes and metabolic pools as model traits for quantitative genetics. I. The L-shaped distribution of gene effects. Genetics 153 : 2001-2012.

Bost B., de Vienne D., Moreau L., Hospital F., and Dillmann C. (2001). Genetic and non genetic bases for the L-shaped distribution of QTL effects. Genetics 157 : 1773-1787.

Bulmer M.G., 1971, The effect of selection on genetic variability. Am. Nat. 105 :201-211.

Bürger R., 1988, Mutation-selection balance and contionuum of alleles models. Math. Biosci. 91 :67-83.

Bürger R., J.P. Wagner, and F. Stettinger, 1989, How muche heritable variation can be maintained in finite populations by a mutation-selection balane ? Evolution 43 :1748-1766.

Bürger R. and R. Lande, 1994, On the distribution of the mean and variance of a quantitative trait under mutation-selection-drift balance. Genetics 138 :901-912.

Charcosset A., Mangin B., Moreau L., Combes L., Jourjon M.-F., and Gallais A. (2000). Heterosis in maize investigated using connected RIL populations. In "Quantitative genetics and breeding methods : the way ahead" (A. Gallais, C. Dillmann, and I. Goldringer, Eds.), pp. 90-98, Les Colloques de l'INRA, Paris.

Charlesworth B, 1979, Evidence against Fisher's theory of dominance. Nature(London) 278 :848-849.

Chevalet C. (1994) An approximate theory of selection assuming a finite number of quantitative trait loci. Genet. Sel. Evol. 26 :379-400.

Cockerham C.C. (1959) Partition of hereditary variance for various genetic models. Genetics 44 :1141-1148.

Dekkers J.C.M. and M.R. Dentine (1991) Quantitative genetic variance associated with chromosomal markers in seggregating populations. Theor. Appl. Genet. 81 :212-220.

Dekkers J.C.M. and J.A.M. Van Arendonk (1998) Optimum selection for quantitative traits with information on an identified locus in outbred populations. Genet. Res. 71 :257-275.

Dekkers J.C.M. and F. Hospital (2002) The use of molecular genetics in the improvement of agricultural populations Nature Rev. Genet. 3 :22-32.

Dekkers J.C.M., R. Chakraborty and L. Moreau (2002) Optimal selection on two quantitative trait loci with linkage. Genet. Sel. Evol. (In Press)

de Vienne D., Bost B., Fiévet J., and Dillmann C. (2001) Optimization of enzyme concentrations for unbranched reaction chains : the concept of combined response coefficient. Acta Biotheoretica 49 : 341-350.

Dillmann, C. (1992). Organisation de la variabilité génétique chez les plantes : modélisation des effets d'interaction, conséquences pour la réponse à la sélection, Thèse Institut National Agronomique Paris-Grignon.

Dillmann C., and Foulley J.-L. (1998). Another look at multiplicative models in quantitative genetics. Genetics Selection Evolution 30 : 543-564.

Doebley J. and A. Stec (1991) A genetic analysis of the morphological differences between maize and teosinte. Genetics 129 :285-295.

Dudley J.W. and R.J. Lambert, (1992) Ninety generations of selection for oil and proteins in maize. Maydica 37 :81-87.

Eklund J. and G.E. Bradford (1977) Genetic analysis of a strain of mice plateaued for litter size. Genetics 85 :529-542.

Enfield F.D. (1980) Selection experiments in laboratory and domestics animals. CAB :69-86. Ed Alan Robertson.

Eshed Y. and D. Zamir (1995) An introgression line population of Lycopersicon pennelli in the cultivated tomato enables the identification and fine-mapping of yield-associated QTL. Genetics 141 :1147-1162.

Eshed Y. and D. Zamir (1996) Less-than-additive epistatic interactions of quantitative trait loci in tomato. Genetics 143 :1807-1817

Falconer D.S. and J.W.B. King (1953) A study of selection limits in the mouse. J. Genet. 51 :470-501.

Falconer D.S. and T. Mackay (1989) *Quantitative genetics*. Longman, 4th edition.

Fisher R.A., 1918, The correlation between relatives on the supposition of mendelian inheritance. J. Agric. Research 69 :399-433.

Fisher R.A., (1928) The possible modification of the response of the wild type to recurrent mutations. Am. Nat. 62 :115-126.

Fisher R.A., 1931, The evolution of dominance. Biological Reviews of the Cambridge Philosophical Society 6 :345-368.

Fisher R.A., 1958, The genetical theory of natural selection. 2nd Ed, Dover Publications, New-York.

Franklin I. and R.C. Lewontin (1970) Is the gene the unit of selection ? Genetics 65 :707-734.

Gallais A. (1989) *Théorie de la sélection en amélioration des plantes*. Masson, Paris.

Gallais A., 1993 Efficiency of recurrent selection methods to improve the line value of a population. Plant Breeding 111 :31-41.

Geiger H.H. and A. Tomerius, 1997 Quantitative genetics and optimum breeding plans. In

"The 10th meeting of the Eucarpia section Biometrics in Plant Breeding" 14-16 Mai. P. Krajewski and Z. Kazmarek Eds, pp15-26. Poznan.

Gimelfarb A. and R. Lande (1995) Marker-assisted selection and marker-QTL associations in hybrid populations. Theor. Appl. Genet. 91 :522-528.

Goffinet B. and Gerber (2000) Quantitative Trait Loci : A Meta-analysis. Genetics 155 : 463-473.

Goldringer I., P. Brabant and A. Gallais, 1996 Theoretical comparison of recurrent selection mothods for the imporvement of self-pollinated crops. Crop Sci. 36 :1171-1180.

Grafius J.E., 1964, A geometry for plant breeding. Crop Sci. 4 :241-246.

Hartl D.L., D.E. Dykhuisen and A.M. Dean (1985) Limites of adaptation : the evolution of selective neutrality. Genetics 111 :655-674.

Heinrich R. and T.A. Rapoport (1974) A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. European J. Biochem. 42 :89-95.

Hill W.G., 1982, Rates of change in quantitative traits from fixation of new mutations. Proc. Nat. Acad. Sci. USA 79 :142-145.

Hill W.G. and A. Robertson (1966) The effect of linkage on limits to artificial selection. Genet. Res. 8 :269-294.

Hospital F. (2001). Size of donor chromosome segments around introgressed loci and reduction of linkage drag in marker-assisted backcross programs. Genetics 158 : 1363-1379.

Hospital F. (2002) Marker-assisted selection in plants : lessons for success(?) stories. Molecular Breeding, in prep.

Hospital F., and Chevalet C. (1993). Effect of population size and linkage on optimal selection intensity. Theoretical and Applied Genetics 86 : 775-780.

Hospital F., and Chevalet C. (1996). Interactions of selection, linkage and drift in the dynamics of polygenic characters. Genetical Research 67 : 77-87.

Hospital F., and Charcosset A. (1997). Marker-assisted introgression of quantitative trait loci. Genetics 147 : 1469-1485.

Hospital F., Moreau L., Lacoudre F., Charcosset A., and Gallais A. (1997). More on the efficiency of marker-assisted selection. Theoretical and Applied Genetics 95 : 1181-1189.

Hospital F., Goldringer I., and Openshaw S. (2000). Efficient marker-based recurrent selection for multiple quantitative trait loci. Genetical Research 75 : 357-368.

Houle D., 1989, The maintenance of polygenic variation in finite populations. Evolution 43 :1767-1780.

Houle D., B. Morikawa, and M. Lynch (1996) Comparing mutational variabilities Genetics 1996 143 :1467-1483.

Hyne V. and M.J. Kearsey (1995) QTL analysis : further uses of marker regression. Theor. Appl. Genet. 91 :471-476.

Jansen R.C. and P. Stam (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136 :1447-1455.

Kacser H. and J.A. Burns (1973) The control of flux. Symposium of the Society of Experimental Biology 27 :65-104.

Kacser H. and J.A. Burns (1981) The molecular basis of dominance. Genetics 97 :639-666.

Kearsey M.J. and H.S. Pooni (1996) *The genetical analysis of quantitative traits* Chapman & Hall, London.

Kearsey M.J. and A.G.L. Farquhar (1998) QTL analysis in plants ; where are we now ? Heredity 80 :137-142.

Keightley P.D. and W.G. Hill, 1988, Quantitative genetic variation maintained by mutation-

stabilizing selection balance in finite populations. Genet. Res. 52 :33-48.
Keightley P.D. (1996) Metabolic models of selection response. J. Theor. Biol. 182 :311-316.
Kimura M. (1957) Some problems of stochastic processes in genetics. Ann. Math. Stat. 28 :882-901.
Lande R., 1975, The maintenance of genetic variability by mutation in a polygenic character with linked loci. Genet. Res. 26 :221-235.
Lande R., 1976, Natural seelction and random genetic drift in phenotypic evolution. Evolution 30 :314-334.
Lange C. and J.C. Whittaker (2001) On Prediction of Genetic Values in Marker-Assisted Selection. Genetics 159 : 1375-1381
Lynch M. and B. Walsh (1998) *Genetic and analysis of quantitative traits* (Sinauer Associates, Sunderland, Mass.).
Lynch M. and B. Walsh (1998) *Evolution and selection of quantitative traits* http ://nitro.biosci.arizona.edu/zbook/volume_2/vol2.html.
Mackay T.F.C. (2001) Quantitative trait loci in Drosophila. Nature Rev. Genet. 2 :11-20.
Mangin B., B. Goffinet and A. Rebai (1994) Constructing confidence intervals for QTL location. Genetics 138 :1301-1308.
Meuwissen T.H., B.J. Hayes and M.E. Goddard (2001) Prediction of total genetic value using genome-wide dense marker maps. Geentics 157 :1819-1829.
Monna L., H.X. Lin, S. Kojima and T.Sasaki (2002) Genetic dissection of a genomic region for a quantitative trait locus, Hd3, into two loci, Hd3a and Hd3b, controlling heading date in rice. Theor. Appl. Genet. 104 :772-778.
Moreau L., Charcosset A., Hospital F., and Gallais A. (1998). Efficiency of marker assisted selection in populations of finite size. Genetics 148 : 1353-1365.
Moreau L., Charcosset A. and Gallais A. (2002) Stability of QTL effects investigated in a large range of environmental conditions for grain yield and related traits in Maize. Submitted.
Naveira, H., and A. Barbadilla (1992) The theoretical distribution of lengths of intact chromosome segments around a locus held heterozygous with backcrossing in a diploid species. Genetics 130 : 205-209.
Nuzhdin S.V., Dilda C.L., Mackay T.F. (1999) The genetic architecture of selection response. Inferences from fine-scale mapping of bristle number quantitative trait loci in Drosophila melanogaster. Genetics 153 :1317-1331.
Orr H.A., 1991, A test of Fisher's theory of dominance. Proc. Nat. Acad. Sci. USA 88 :11413-11415.
Orr H.A. (1998) The population genetics of adaptation : the distribution of factors fixed during adaptative evolution. Evolution 52 :935-949.
Peck J., G. Barreau and S.C. Heath (1997) Imperfect genes, Fisherian mutations and the evolution of sex. Genetics 145 :1171-1199.
Robertson A. (1960) A theory of limits in artificial selection. Proc Royal Soc. London B. 153 :234-249.
Santiago E. and A. Caballero (1995) Effective size of populations under selection. Genetics 139 :1013-1030.
Servin B. and F. Hospital (2002) Optimal positioning of markers to control genetic background in marker-assisted backcrossing. J. Hered. In Press.
Steinmetz L.M., H. Sinha, D.R. Richards, J.I. Spiegelman, P.J. Oefner, J.H. McCusker, R.W. Davis (2002) Dissecting the architecture of a quantitative trait locus in yeast. Nature 416 : 326 - 330.

Turelli M., 1984, Heritable variation via mutation-selection balance : Lerch's zeta meets the abdominal bristle. Theor. Pop. Biol. 25 :138-193.

Turelli M. and N.H. Barton (1990) Dynamics of polygenic characters under selection Theor. Pop. Biol. 38 :1-57.

Wade M.J., Winther R.G., Agraval AF, Goodninght CJ (2001) Alternative definitions of epistasis : dependence and interaction. Trends Ecol. Evol., 16 :498-504.

Weber K.E. (1996) Large genetic changes at small fitness cost in large populations of drosophila melanogaster selected for wind tunnel flight : rethinking fitness surfaces. Genetics 144 :205-213.

Wright S., 1931 Evolution in mendelian populations. genetics 16 :97-159.

Wright S., 1934, Molecular and evolutionary theories of dominance. Am. Nat. 68 :24-53.

Yoo B.H. (1980) Long term selection for a quantitative character in large replicate populations of Drosophila melanogaster. I Response to selection. Genet. Res. Camb.

Zeng Z.B. (1994) Precision mapping of quantitative trait loci. Genetics 136 :1457-1468.

TAB. 1 – **Modèle génétique additif** : Valeurs génotypiques à deux locus

		aa	aA	AA
		$1-s$	1	$1+s$
bb	$1-t$	$1-s-t$	$1-t$	$1+s-t$
bB	1	$1-s$	1	$1+s$
BB	$1+t$	$1-s+t$	$1+t$	$1+s+t$

TAB. 2 – **Modèle génétique multiplicatif** : Valeurs génotypiques à deux locus

		aa	aA	AA
		$(1-s)^2$	$(1-s^2)$	$(1+s)^2$
bb	$(1-t)^2$	$(1-s)^2(1-t)^2$	$(1-s^2)(1-t)^2$	$(1+s)^2(1-t)^2$
bB	$(1-t^2)$	$(1-s)^2(1-t^2)$	$(1-s^2)(1-t^2)$	$(1+s)^2(1-t^2)$
BB	$(1+t)^2$	$(1-s)^2(1+t)^2$	$(1-s^2)(1+t)^2$	$(1+s)^2(1+t)^2$

TAB. 3 – **Exemple de valeurs génétiques en modèle additif**

		aa	aA	AA
		0.5	1	1.5
bb	0.5	1	1.5	2
bB	1	1.5	2	2.5
BB	1.5	2	2.5	3

Les valeurs sont calculées d'après la table 1 en prenant $s = t = 1/2$. Dans ce modèle, les écarts entre les génotypes restent constants, et il n'y a pas de dominance.

TAB. 4 – **Exemple de valeurs génétiques en modèle multiplicatif**

		aa	aA	AA
		1/4	3/4	9/4
bb	1/4	1	3	9
bB	3/4	3	9	27
BB	9/4	9	27	81

Les valeurs sont calculées d'après la table 2 en prenant $s = t = 1/2$. Les valeurs génétiques sont exprimées en 1/16èmes. On voit que les écarts entre les valeurs génétiques au locus A dépendent du génotype au locus B. Au niveau du caractère, les allèles a et b sont dominants, ce qui correspond à un cas de sous-dominance.

TAB. 5 – **Valeurs sélectives en sélection par troncation**

	aa	aA	AA
bb	0	0	0
bB	0	0	1
BB	0	1	1

On sélectionne la même fraction des meilleurs génotypes dans les tableaux 3 ou 4. Selon le génotype au locus B, on trouve soit que A n'est pas variable, soit que l'allèle a est dominant, soit que l'allèle A est dominant.

TAB. 6 – **Exemple de valeurs génétiques en modèle métabolique**

		aa	aA	AA
		1	2	3
bb	0.01	000990	0.00995	0.00997
bB	0.5	0.33	0.40	0.43
BB	1	0.50	0.67	0.75

On considère un caractère gouverné par deux enzymes biallélіques A et B dans une population. On appelle g_A et g_B l'activité des enzymes A et B, et la valeur du caractère vaut : $G = \frac{1}{1/g_A+1/g_B}$. Dans cet exemple, l'enzyme A est plus variable que l'enzyme B.

TAB. 7 – **Réponse à la sélection en modèle métabolique**

fréq. alléliques		μ	$\sigma^2_{(A)}$ %	$\sigma^2_{(B)}$%	$\frac{R_B}{R_A}$	
p_A	p_B				prédit	observé
0.3	0.6	0.392	7.1	90.5	3.33	2.35
0.5	0.5	0.359	3.7	94.4	5.05	6.4

En utilisant les valeurs génétiques du tableau 6, on calcule, en fonction des fréquences des allèles aux locus A et B la moyenne μ de la population et les variances expliquées par les effets statistiques additifs à chaque locus, en pourcentage de la variance génétique totale. Presque toute la variance est additive. Dans un modèle additif, si s_i est la valeur de l'allèle au locus i et p_i la fréquence de l'allèle favorable, la variance additive du locus est $\sigma^2_{(i)} = 2p_i(1-p_i)s_i^2$. Le gain attendu par la fixation de l'allèle i est $R_i = s_i$ et peut être estimé à partir de la variance additive au locus i. Le rapport des gains attendus vaut :

$$\frac{R_B}{R_A} = \sqrt{\frac{2p_B(1-p_B)\sigma^2_{(A)}}{2p_A(1-p_A)\sigma^2_{(B)}}}$$

On peut aussi calculer le gain relatif observé en s'aidant des valeurs génétiques du tableau 6.

FIG. 1 – **Réponse à long terme à la sélection prédite par les différents modèles.** Les courbes ont été tracées à partir des paramètres (variance génétique additive, héritabilité, intensité de sélection et effectif génétique) estimés par Yoo (1980) dans une expérience de sélection chez la Drosophile, et des équations de la section 2. La courbe des effets combinés de la dérive génétique et du linkage n'a pas été calculée à partir d'équations analytiques. Elle est donnée à titre indicatif. Les deux points correspondent aux réponses réalisées dans l'expérience de Yoo.

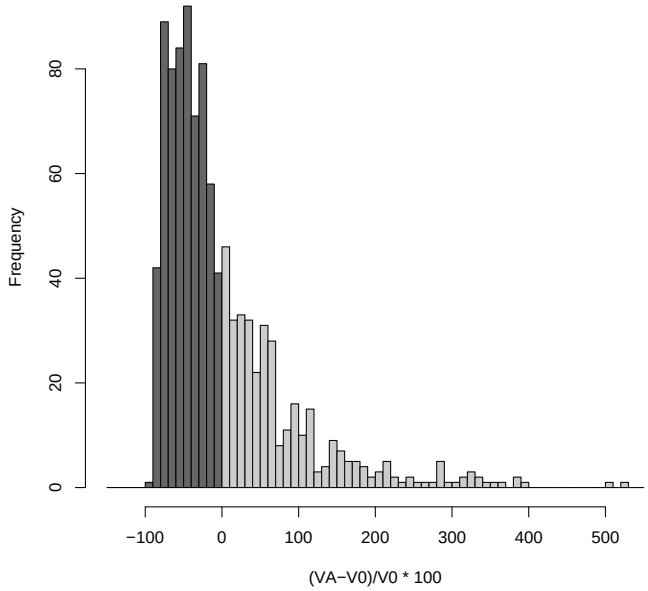


FIG. 2 – **Effet des déséquilibres de liaison sur la variance en F2** En modèle bi-allélique, dans une population F2 issue d'un hybride entre deux parents fixés aléatoirement pour l'un ou l'autre allèle à plusieurs locus (phase aléatoire), on s'attend à avoir autant de covariances négatives que positives entre paires de locus, et donc que la somme de ces covariances ait une distribution centrée sur zéro. Pour obtenir la figure, on a simulé 1000 hybrides F1 de phase aléatoire ($++$, $--$ ou $+-$) à 100 locus distants de 1 cM, et on a calculé la somme des covariances entre tous les locus pris deux à deux. On déduit la variance disponible VA. Sur la figure on donne la distribution pour les 1000 hybrides du différentiel relatif de variance dû aux déséquilibres $\frac{C}{V_g}$. La distribution finale est très dissymétrique, avec la particularité que $2/3$ des valeurs sont négatives, contre $1/3$ positives, mais que la borne inférieure (négative) est beaucoup plus proche de zéro que la borne supérieure (positive). Nous avons également montré que le signe de C ne dépend que de l'alternance des allèles favorables et défavorables chez l'un des parents le long d'un chromosome, et probablement de la longueur des blocs d'allèles favorables. La valeur maximale de C est obtenue lorsque l'un des parents cumule tous les allèles favorables (coupling total). La valeur minimale de C est obtenue lorsque les allèles favorables et défavorables se succèdent en alternance le long des chromosomes (répulsion totale).

FIG. 3 – **Relations possibles entre le flux de produit à travers une chaîne métabolique et la quantité d'une enzyme de cette chaîne pour différents modes de régulation de l'expression des enzymes. A**. En l'absence de régulations et de contraintes, les enzymes sont indépendantes, et la relation entre le flux et l'activité d'une enzyme de la chaîne est hyperbolique. La limite maximale du flux dépend de l'activité des autres enzymes de la chaîne. 1, 2 et 3 sont obtenus pour l'enzyme j dans trois contextes génétiques différents. **B**. En présence d'une contrainte sur la quantité totale de protéines, l'augmentation de la quantité de l'enzyme j pénalise les autres enzymes. Il existe une répartition des quantités d'enzymes qui maximise le flux. **C**. La quantité totale de protéines n'est pas limitée, mais les enzymes de la chaîne sont co-régulées. 4 : la quantité de toutes les enzymes augmente en même temps que celle de l'enzyme j. 5 : Une enzyme de la chaîne est indépendante de l'enzyme j. 6 : Deux des enzymes de la chaîne sont indépendantes de l'enzyme j. 7 : L'une des enzymes de la chaîne est négativement correlée à l'enzyme j. **D**. La quantité totale de protéines est limitée à $Qtot$ et les enzymes de la chaîne sont co-régulées. 8 : Quel que soit le sens de la régulation, il existe une répartition des quantités d'enzymes qui maximise le flux, et qui dépend des paramètres cinétiques des enzymes. Chaque courbe 8 correspond à un mode de régulation différent.

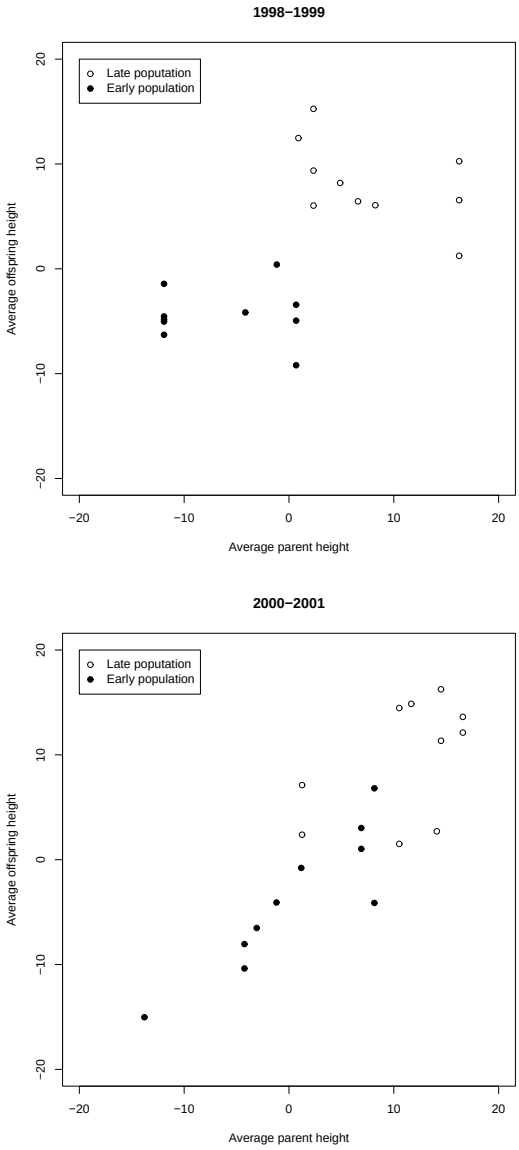


FIG. 4 – **Sélection pour la précocité dans des lignées pures de maïs : relation entre la taille moyenne des parents et celle des enfants** A partir d'une seule lignée de départ, on sélectionne chaque année les 10 plantes les plus précoces dans la population précoce, et les 10 plantes les plus tardives dans la population tardive. Chaque plante sélectionnée est autofécondée et produit 100 descendants. Le dispositif expérimental permet de mesurer avec une bonne précision la taille moyenne des enfants des plantes sélectionnées. Les résultats sont exprimés en écarts (cm) par rapport au témoin, issu du même lot de semences que la lignée de départ.

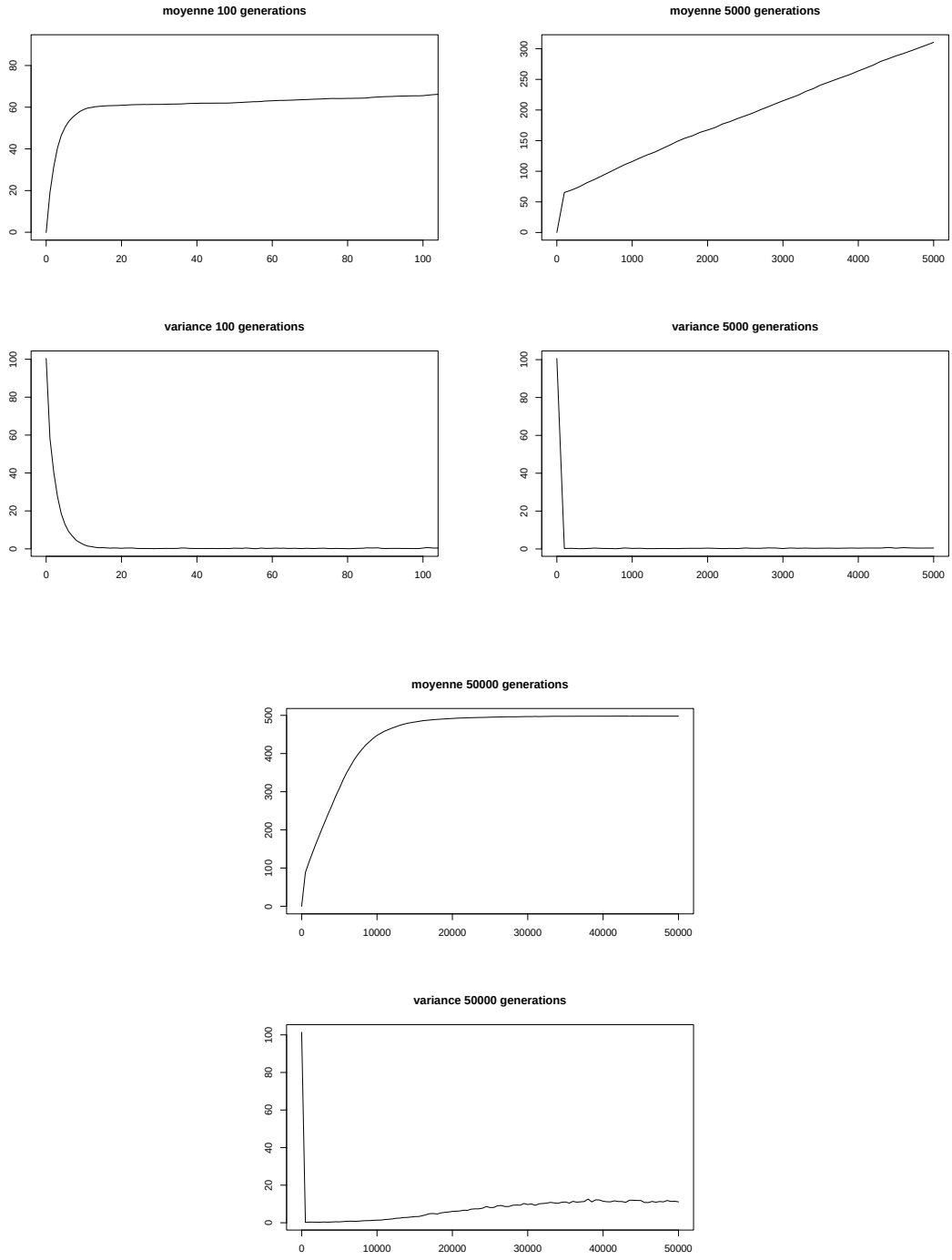


FIG. 5 – **Adaptation à un optimum lointain : Evolution de la moyenne et de la variance génétique des populations au cours du temps**
La moyenne initiale de la population vaut zéro. Les trois séries de graphes montrant l'évolution simultannée de la moyenne et de la variance correspondent à trois échelles de temps différentes. Ces résultats ont été obtenus en simulant une population de taille constante ($N = 500$ individus) soumise à une sélection pour un optimum de valeur $o = 500$, avec une intensité de sélection $s = 30$. Dans la population de départ, tous les locus sont polymorphes et la valeur des $2N$ allèles est tirée dans une loi normale $\mathcal{N}(0, 1)$. La valeur du caractère sélectionné est obtenue en additionnant la valeur de chaque allèle à $n = 50$ locus. L'héritabilité initiale du caractère vaut $h^2 = 0.5$ et sert à déterminer la variance environnementale qui reste constante par la suite. Les effets des mutations ($\mu = 10^{-4}$) sont tirés dans une loi normale $\mathcal{N}(0, 4)$ en utilisant un modèle RMM.

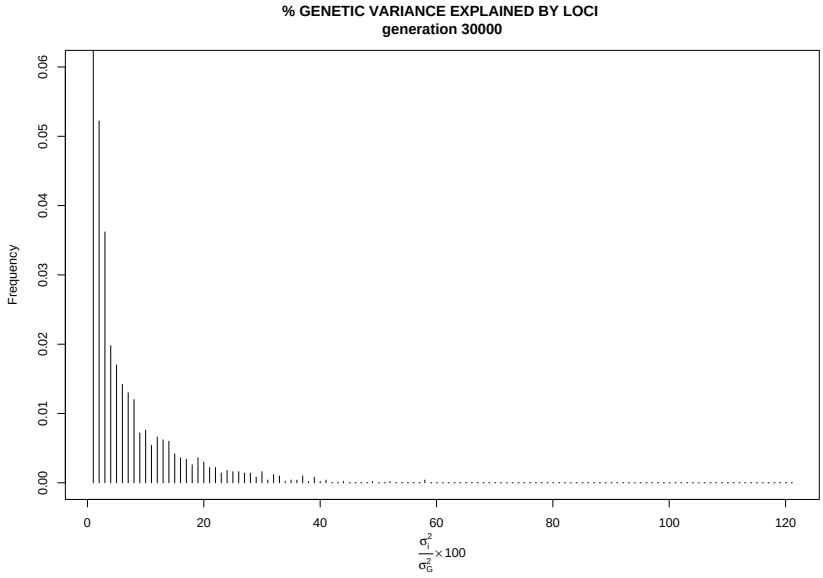


FIG. 6 – **Adaptation à un optimum lointain : Distribution de la part de variance expliquée par chaque locus à l'équilibre mutation/sélection/dérive génétique** Les paramètres des simulations sont les mêmes que pour la figure 5.

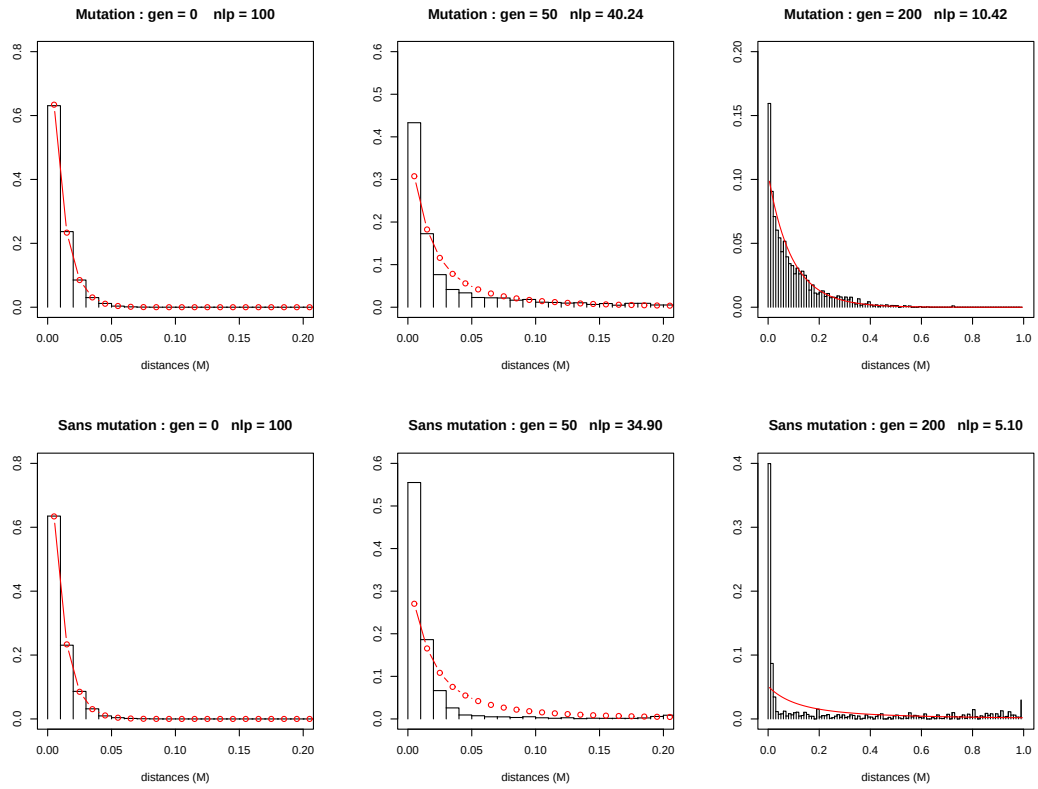


FIG. 7 – **Agrégation du polymorphisme dans le génome au cours des générations de sélection, avec ou sans mutation. Histogrammes** : Distribution observée des distances entre paires de locus polymorphes adjacents, obtenue à partir de 200 répétitions. **Traits** : Distributions attendues pour des locus répartis de façon aléatoire dans le génome (modèle du bâton brisé). Les classes de distances sont en abcisse, et leur fréquence en ordonnée. Ces résultats ont été obtenus en simulant une population de taille constante ($N = 200$ individus) soumise à une sélection pour un optimum de valeur $o = 0$, avec une intensité de sélection $s = 1$. Dans la population de départ, tous les locus sont polymorphes et la valeur des $2N$ allèles est tirée dans une loi normale $\mathcal{N}(0, 4)$. La valeur du caractère sélectionné est obtenue en additionnant la valeur de chaque allèle à $n = 100$ locus répartis de façon aléatoire sur un chromosome de $100cM$. La variance environnementale est nulle. gen indique la génération et nlp indique le nombre de locus polymorphes à la génération considérée. Les effets des mutations ($\mu = 10^{-4}$) sont tirés dans une loi $\Gamma(0, 4)$ en utilisant un modèle RMM.

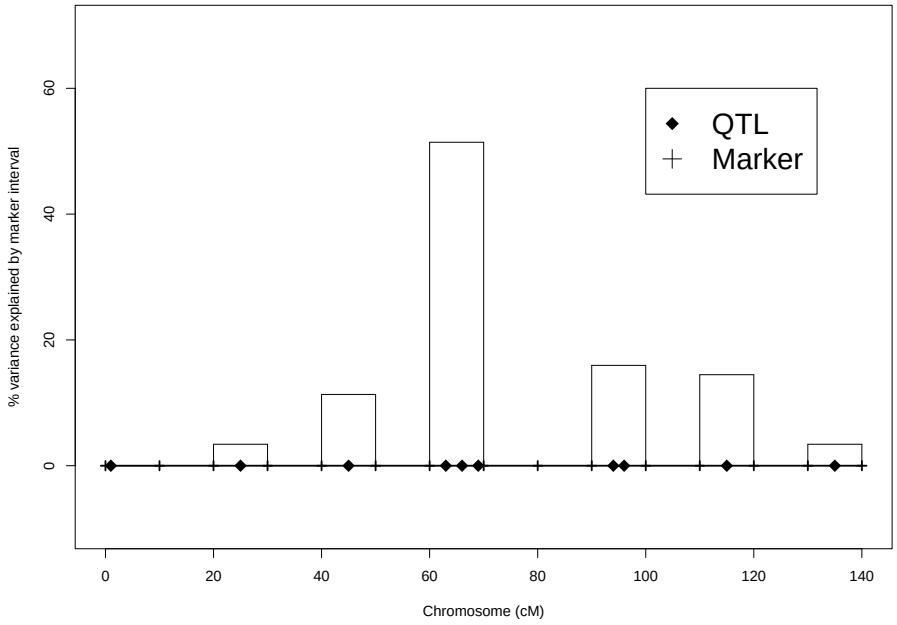
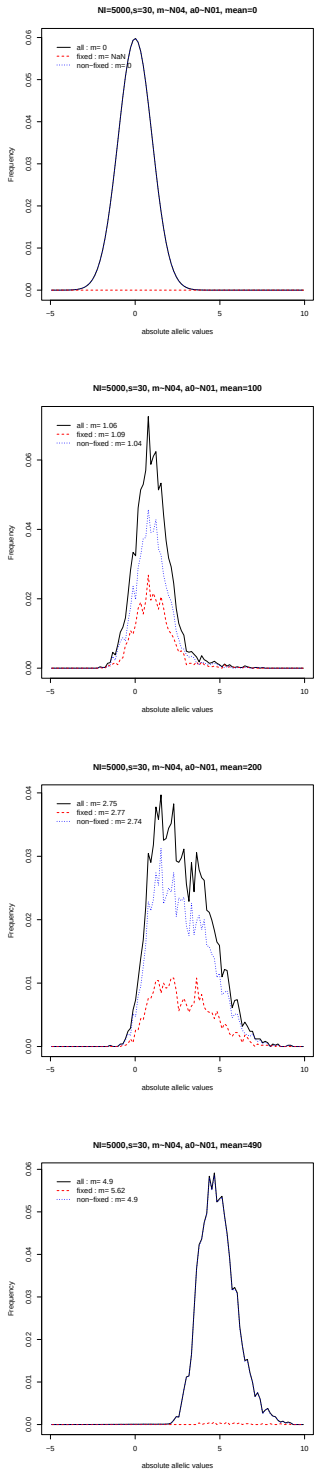


FIG. 8 – **Effet de l'histoire sélective et du linkage sur l'architecture apparente des caractères quantitatifs** Soit un caractère gouverné par les 10 QTL figurant sur la carte génétique ci-dessus. Soit une population panmictique de 200 individus où ces QTL ont des effets identiques sur le caractère et portent des allèles de valeur $+1$ ou -1 en fréquences .1 et .9, respectivement. On a simulé 100 lignées tirées de cette population par sélection directionnelle de 10% des individus à chaque génération jusqu'à fixation complète. On calcule ensuite le pourcentage de variance expliqué par chaque segment compris entre deux marqueurs, sur l'ensemble des hybrides entre les 100 lignées prises 2 à 2. Pour corriger les effets du nombre de QTL par segment, l'effet d'un segment est calculé comme la somme des effets des QTL du segment divisée par le nombre de QTL du segment.



Ces résultats ont été obtenus en simulant une population de taille constante ($N = 5000$ individus) soumise à une sélection pour un optimum de valeur $o = 500$, avec une intensité de sélection $s = 30$. Dans la population de départ, tous les locus sont polymorphes et la valeur des $2N$ allèles est tirée dans une loi normale $\mathcal{N}(0, 1)$. La valeur du caractère sélectionné est obtenue en additionnant la valeur de chaque allèle à $n = 50$ locus répartis de façon aléatoire sur 10 chromosomes de $100cM$. La variance environnementale est constante au cours du temps. Les effets des mutations ($\mu = 10^{-4}$) sont tirés dans une loi $\mathcal{N}(0, 4)$ en utilisant un modèle RMM. Chaque graphe représente la distribution des valeurs des allèles sur l'ensemble des locus et sur 200 répétitions à une génération donnée (générations 0, 100, 600, et 1650, de haut en bas). *mean* est la moyenne de la population à cette génération, et m la valeur moyenne des allèles. Nous avons également distingué, la fréquence des allèles fixés dans une population (courbe rouge, inférieure), celle de ceux qui sont polymorphes (bleue, intermédiaire), et la fréquence totale (noire, supérieure).

FIG. 9 – **Evolution de la distribution des valeurs alléliques au cours de l'adaptation à un optimum lointain**