

L. OLLIVIER

INRA Station de Génétique quantitative et
appliquée 78352 Jouy-en-Josas Cedex

Les bases de la génétique quantitative

Le modèle à plusieurs locus

Résumé. Cet article traite de la génétique d'un caractère quantitatif déterminé par plus de deux locus, sur la base du modèle polygénique classique incluant des effets génétiques additifs, de dominance et épistatiques. Par simplifications successives on définit un modèle polygénique additif (sans épistasie) et un modèle polygénique additif strict (sans dominance ni épistasie). Le modèle polygénique additif conduit à définir pour un ensemble de plusieurs caractères les paramètres génétiques d'une population. Ceux-ci conditionnent le comportement de ces caractères dans un programme de sélection. Il s'agit d'un modèle avant tout opérationnel, auquel il est souvent fait le reproche de ne pas approcher d'assez près la réalité biologique. Il existe cependant des caractères quantitatifs qui répondent à un déterminisme génétique additif impliquant un nombre élevé de locus. Des méthodes existent par ailleurs pour évaluer le nombre minimum de locus impliqués dans un caractère et plusieurs expériences sur animaux de laboratoires indiquent que ce nombre est au moins de l'ordre de quelques dizaines pour la plupart des caractères.

L'article précédent traite de la génétique d'un caractère déterminé par un nombre limité de gènes. Les exemples numériques choisis pour illustrer le modèle à un locus montrent clairement que le nombre des valeurs génotypiques possibles est lui aussi limité. On voit donc assez bien comment ce modèle peut s'appliquer à des caractères à variation discontinue, ou, à la rigueur, à des caractères à variation continue pour lesquels les différences entre génotypes sont d'un ordre de grandeur très largement supérieur à celui des erreurs de mesure : par exemple, une quantité de pigment dans le poil, une activité enzymatique... Mais comment concilier ce modèle avec l'observation de caractères à variation strictement continue (caractères quantitatifs), comme le sont la plupart des caractères zootechniques (poids, quantité de lait...) ? Il existe en effet une contradiction apparente entre la variation continue d'un caractère et une transmission mendélienne par essence discontinue, puisqu'elle dépend de la ségrégation des gènes à la méiose. Le modèle à plusieurs locus, ou modèle polygénique, permet de résoudre cette contradiction.

Tableau 1. Nombre de génotypes possibles selon le nombre des locus et des allèles par locus.

Nombre de locus	Nombre d'allèles par locus		
	2	3	5
2	9	36	225
3	27	216	3 375
5	243	7 776	759 375

1 / Un caractère polygénique

1.1 / Valeur génotypique

Le nombre de génotypes possibles, et donc celui des valeurs génotypiques, augmente très vite avec le nombre des locus considérés, comme le montre le tableau 1.

Cependant, comme nous allons le voir, l'extension du modèle à 2 locus au modèle à plus de 2 locus ne fait appel à aucun concept nouveau. Le modèle s'écrit toujours :

$$G = A + D + I \quad (1)$$

où :

- G est la valeur génotypique ;
- A est la somme des effets moyens de tous les gènes, soient 2 fois l termes pour le cas de l locus à 2 allèles ;
- D est la somme des valeurs de dominance, soient l termes pour le cas de l locus à 2 allèles ;
- I est la valeur d'épistasie.

Le terme I est, comme dans le cas de 2 locus, un "résidu" qu'on peut obtenir par différence entre G et A+D. Mais ce résidu est lui-même décomposable en plusieurs termes correspondant à différents types d'interaction. Dans le cas de 2 locus il y a 3 types d'interaction possibles, AA, AD et DD (voir l'article précédent). Rappelons qu'une interaction de type AA existe lorsque la différence entre les effets moyens des 2 allèles à un locus n'est pas la même selon la

valeur génétique additive à un autre locus. On définit de la même façon une interaction AAA dans le cas de 3 locus, lorsque la combinaison des effets additifs à 2 locus varie selon la valeur génétique additive à un troisième locus. Selon le même principe, on définit une interaction de type AAD (la combinaison des effets additifs à 2 locus dépend de la valeur de dominance au troisième locus), et des interactions ADD et DDD.

On a alors la décomposition suivante de la variance génotypique $V(G)$ dans le cas de 3 locus :

$$V(G) = V(A) + V(D) + V(I) \quad (2)$$

où :

- $V(A)$ est la somme de 3 variances génétiques additives : une pour chaque locus

- $V(D)$ est la somme de 3 variances de dominance : une pour chaque locus

- $V(I) = [V(AA) + V(AD) + V(DD)] + [V(AAA) + V(AAD) + V(ADD) + V(DDD)]$: le premier terme entre crochets regroupe les interactions de premier ordre (locus pris 2 à 2) et le deuxième terme regroupe les interactions de deuxième ordre (impliquant les 3 locus).

L'existence d'une variance d'épistasie traduit la non-additivité des effets des différents locus, autrement dit les interactions entre locus. On voit facilement que le nombre des termes de $V(I)$ va augmenter très rapidement avec le nombre des locus impliqués dans le caractère, un peu comme le nombre des génotypes (mais cependant moins vite), alors que le nombre des termes de $V(A)$ reste égal au nombre des locus. Il ne faudrait cependant pas en conclure hâtivement que la variance génétique additive $V(A)$ tend à s'amenuiser au profit de la variance d'épistasie $V(I)$ pour des caractères influencés par un grand nombre de locus. Car l'importance d'une composante d'interaction tend à diminuer à mesure qu'augmente sa complexité. Cela découle du rôle que jouent les fréquences génétiques dans les variances qui composent l'expression (2). On a vu en effet, dans le cas d'un locus, que l'expression générale de $V(A)$ fait intervenir le produit $2pq$, toujours inférieur à 0,5, et celle de $V(D)$ le terme $(2pq)^2$. On montre aussi que les termes correspondants sont $(2pq)^2$ pour $V(AA)$, $(2pq)^3$ pour $V(AD)$ et $V(AAA)$, $(2pq)^4$ pour $V(AAAD)$, $V(AAD)$ et $V(DD)$, etc... Le coefficient multiplicatif des composantes de $V(I)$ tend donc à diminuer quand le degré des interactions augmente. Dans le cas de 5 locus par exemple, avec la valeur maximale de $2pq$ correspondant à $p=q=0,5$ à chaque locus, la variance d'interaction $V(AAAAA)$ est affectée du coefficient $0,5^5 = 0,03$ contre 0,5 pour chacun des termes de $V(A)$. Cet effet des fréquences génétiques est encore accentué si celles-ci diffèrent d'un locus à l'autre. Ainsi, avec $p=q=0,5$ à 4 locus et $p=0,9$ $q=0,1$ au 5^e, le coefficient de $V(AAAAA)$ est réduit à 0,006 contre 0,4 (en moyenne) pour chacun des termes de $V(A)$.

Notons aussi que la décomposition (2) n'inclut aucune covariance. Ainsi, écrire que $V(A)$ et $V(D)$ sont des sommes de variances implique l'hypothèse que les covariances entre les différents locus sont nulles. Le modèle suppose donc l'indépendance des locus entre eux, ce qui correspond à la situation d'une population "idéale" évoquée au chapitre précédent, avec équilibre de liaison. Remarquons au passage que l'interaction entre les composantes A et D des différents locus exprimée par I dans (1) est ici compatible avec l'indépendance des locus entre eux définie par (2). Deux

locus épistatiques peuvent en effet se trouver en situation d'équilibre de liaison.

1.2 / Valeur phénotypique

Comme avec le modèle à 2 locus, en ajoutant l'effet du milieu E, on obtient :

$$P = G + E \quad (3)$$

et, compte tenu de (2) :

$$V(P) = V(A) + V(D) + V(I) + V(E) \quad (4)$$

Ces deux expressions traduisent les deux hypothèses essentielles qui sont faites (voir, par exemple, Ollivier 1981):

- additivité des effets du génotype et du milieu, puisque l'équation (3) ne comporte pas de terme d'interaction GxE (interaction génotype-milieu). Cela veut dire que les effets génétiques ne varient pas quand les conditions de milieu changent. Autrement dit, les différences entre deux génotypes quelconques sont les mêmes quel que soit le milieu.

- indépendance du génotype et du milieu, puisqu'il n'y a pas de covariances dans (4). Cela veut dire que les génotypes et les conditions de milieu sont associés au hasard. Ainsi, par exemple, il est exclu que les individus à haute valeur génétique (à supposer que cette valeur soit connue ou estimée) soient placés dans des conditions de milieu plus favorables que la moyenne.

1.3 / Covariance entre apparentés

La formule donnée dans l'article précédent pour le cas de 2 locus épistatiques est généralisable à un nombre l quelconque de locus et, pour $l=3$ par exemple, la covariance entre 2 génotypes individuels G_x et G_y s'écrit :

$$\text{Cov}(G_x, G_y) = 2f_{xy}V(A) + u_{xy}V(D) + (2f_{xy})^2V(AA) + 2f_{xy}u_{xy}V(AD) + u_{xy}^2V(DD) + (2f_{xy})^3V(AAA) + (2f_{xy})^2u_{xy}V(AAD) + 2f_{xy}(u_{xy})^2V(ADD) + u_{xy}^3V(DDD) \quad (5)$$

Cette covariance est aussi la covariance entre les apparentés x et y, ou leurs phénotypes P_x et P_y , en l'absence de covariance entre les effets de milieu. La généralisation à un nombre quelconque de locus s'obtient en multipliant chaque variance d'interaction par $2f_{xy}$ et u_{xy} , chacun autant de fois que les termes A et D interviennent dans l'interaction. On peut donc prévoir, puisque $2f_{xy}$ et u_{xy} sont inférieurs à 1, que les coefficients multiplicatifs vont rapidement tendre à devenir négligeables à mesure qu'augmente le nombre des locus pris en compte. Ainsi dans le cas de demi-frères, avec $f_{xy} = 1/8$ et $u_{xy} = 0$ (cf article précédent), on a :

$$\text{Cov}(G_x, G_y) = (1/4)V(A) + (1/16)V(AA) + (1/64)V(AAA) + (1/256)V(AAAA) + \dots$$

Nous avons vu que les variances d'interaction elles-mêmes tendent à diminuer quand la complexité des interactions augmente (voir 1.1). On peut donc prévoir que les termes relatifs à l'épistasie dans les covariances entre apparentés seront en général négligeables.

1.4 / Conclusions

Ces considérations justifient que les phénomènes d'épistasie soient le plus souvent ignorés dans l'étude génétique des caractères quantitatifs. En théorie, on peut s'attendre que $V(I)$ dans l'expression (4) soit faible et, en pratique, la mise en évidence expérimentale

tales de cette composante de variance est difficile, du moins intra-population à cause de (5), et les résultats expérimentaux eux-mêmes sont peu convaincants (voir la mise au point de Barker, 1979). En faisant donc les deux hypothèses d'additivité et d'indépendance, à la fois pour les locus entre eux et pour les locus vis-à-vis du milieu, on aboutit au modèle polygénique additif, qui s'écrit :

$$P = A + D + E \quad (6)$$

$$V(P) = V(A) + V(D) + V(E) \quad (7)$$

Le passage de la discontinuité des valeurs génotypiques à la continuité des valeurs phénotypiques, évoqué dans l'introduction, peut maintenant être explicité. On démontre en effet mathématiquement que la somme d'un grand nombre de valeurs (qu'on appelle des variables aléatoires) indépendantes entre elles est elle-même une variable aléatoire dont la distribution tend rapidement à devenir continue, et à suivre en fait une distribution normale. Les propriétés de la valeur phénotypique P impliquées par (6) et (7) répondent exactement à ces conditions, et se trouvent ainsi justifiées par les distributions continues, souvent voisines de la loi normale, observées pour la plupart des caractères quantitatifs. On notera en particulier que la distribution des variables aléatoires élémentaires est quelconque, et que celles-ci peuvent donc suivre des distributions discontinues, par exemple un effet additif à 2 valeurs pour un locus donné.

Notons que le modèle additif (6) et (7) n'implique en fait que l'additivité des effets des locus, et que les effets des gènes du même locus ne sont pas supposés additifs entre eux puisqu'il y a une variance de dominance $V(D)$. Celle-ci est parfois négligée aussi, et on a alors un modèle polygénique additif strict. Ce modèle simplifié est souvent utilisé pour traiter de la sélection. Mais pour traiter des croisements il faut prendre en compte les possibilités de dominance et aussi d'épistasie.

2 / Deux caractères polygéniques

Quand on considère deux caractères, la question se pose de savoir s'ils sont liés (ou corrélés), c'est-à-dire s'ils ont tendance à varier dans le même sens ou en sens contraire, et si cette tendance a des causes génétiques. Cette tendance pour deux caractères à "covarier" se mesure statistiquement par une covariance, qui est la moyenne des produits des écarts de chaque caractère à sa moyenne. Il s'agit donc d'une extension de la notion de variance, qui mesure la tendance d'un caractère à varier autour d'une moyenne par la moyenne des carrés des écarts à cette moyenne. Dans l'article précédent la même notion de covariance a été appliquée à des couples d'apparentés, alors qu'ici elle va être appliquée à des couples de caractères mesurés sur le même individu.

2.1 / Covariance génotypique

Les causes de covariance, comme celles de variance, peuvent être de nature génétique ou non génétique. Il y a covariance de nature génétique entre 2 caractères si au moins un locus exerce une action simultanée sur eux. Nous excluons ici la possibilité d'une covariance due à des locus différents mais liés (cas du déséquilibre de liaison), conformément à l'hypothèse d'une population "idéale" retenue pour la

définition du modèle polygénique additif défini plus haut.

Dans le cas d'un locus, on pourra définir pour chaque caractère une valeur génotypique (soient G et G' pour les deux caractères considérés) et les décomposer comme dans l'article précédent, soit : $G = A + D$ et $G' = A' + D'$

On définira de même une covariance génotypique $Cov(GG')$, une covariance génétique additive $Cov(AA')$ et une covariance de dominance $Cov(DD')$.

Pour illustrer ces notions reprenons l'exemple numérique dans l'article précédent, en associant à chaque génotype un deuxième caractère mesurable (tableau 2). Ce deuxième caractère tend visiblement à varier dans le même sens que le premier, mais la situation de l'hétérozygote par rapport aux 2 homozygotes indique que c'est l'allèle T_2 qui est dominant au lieu de T_1 pour le premier caractère.

Tableau 2. Décomposition de la covariance génotypique entre deux caractères dans le cas d'un locus.

Génotype	T_1T_1	T_1T_2	T_2T_2	
Fréquence	0,16	0,48	0,36	
Mesure ¹⁾	8	6	2	Covariance
	7	4	3	
Valeur génotypique (G) ¹⁾	3,12	1,12	-2,88	2,5344 = $Cov(GG')$
	2,88	-0,12	-1,12	
Valeur génétique additive(A) ¹⁾	3,84	0,64	-2,56	2,7648 = $Cov(AA')$
	2,16	0,36	-1,44	
Valeur de dominance (D) ¹⁾	-0,72	0,48	-0,32	-0,2304 = $Cov(DD')$
	0,72	-0,48	0,32	

¹⁾ première ligne : caractère 1 deuxième ligne : caractère 2

Les valeurs G, A et D pour chaque caractère sont calculées selon les définitions données au chapitre précédent. Les covariances sont les moyennes des produits des valeurs de G, A et D, puisque ces valeurs sont centrées sur zéro : $Cov(GG') = (0,16 \times 3,12 \times 2,38) + (0,48 \times 1,12 \times (-0,12)) + (0,36 \times 2,88 \times 1,12) = 2,5344$, etc... On peut vérifier les relations générales suivantes :

$$Cov(GG') = Cov(AA') + Cov(DD') \quad (8a)$$

$$\text{avec } Cov(AA') = 2pqaa' \quad (8b)$$

$$\text{et } Cov(DD') = (2pq)^2dd' \quad (8c)$$

où a et a' sont les effets moyens de substitution d'un allèle à l'autre pour les deux caractères et d et d' les valeurs génotypiques des hétérozygotes exprimées en écart à la moyenne des deux homozygotes.

Les expressions (8b) et (8c) généralisent celles de $V(A)$ et $V(D)$ de l'article précédent. Il faut noter que a, a', d et d' peuvent être positifs ou négatifs et que les covariances sont donc également affectées d'un signe positif ou négatif selon que les valeurs des deux caractères varient ou non dans le même sens. On note aussi que, puisque $V(A) = 2pq a^2$ et $V(A') = 2pq a'^2$ dans le cas particulier d'un seul locus, on a : $Cov(AA') = \pm [V(A)V(A')]^{1/2}$ d'après (8b)

et que de même, puisque $V(D) = (2pqd)^2$ et $V(D') = (2pqd')^2$,
 $Cov(DD') = \pm [V(D)V(D')]^{1/2}$ d'après (8c)

L'importance pratique de la variance génétique additive découle du fait qu'elle représente la part de variation transmissible à la descendance. De même, la covariance génétique additive est importante parce qu'elle traduit le degré d'association des deux caractères dans leur transmission à la descendance. On peut en effet imaginer deux caractères se transmettant bien à la descendance ($V(A)$ élevée), mais avec des transmissions indépendantes ($Cov(AA')$ nulle) de l'un et de l'autre.

2.2 / Covariance phénotypique

La covariance phénotypique, $Cov(PP')$, prenant en compte les effets de plusieurs locus et ceux du milieu, se décompose comme la variance phénotypique, avec les hypothèses du modèle polygénique additif (additivité des effets et indépendance des locus et des conditions de milieu) :

$$Cov(PP') = Cov(AA') + Cov(DD') + Cov(E E') \quad (9)$$

où $Cov(AA')$ et $Cov(DD')$ représentent les sommes respectives des covariances génétiques additives et des covariances de dominance à chaque locus, et $Cov(E E')$ la covariance due aux effets de milieu. Notons que les covariances élémentaires à chaque locus sont affectées d'un signe et qu'ainsi $Cov(AA')$, par exemple, peut être nulle, alors que de nombreux locus affectent les deux caractères simultanément. Notons aussi que les covariances génétiques et de milieu peuvent être de signe différent.

Tableau 3. Paramètres génétiques relatifs à quatre caractères quantitatifs de l'espèce porcine. Les 4 caractères considérés sont : taille de portée (en nombre de porcelets), vitesse de croissance (en g / j), indice de consommation (aliment consommé / gain de poids) et pourcentage de viande.

Variances-covariances phénotypiques				Variances-covariances génétiques additives			
7	0	0	0	0,7	0	0	0
0	4900	-25	-52	0	1470	-7	-24
0	-25	0,25	-0,5	0	-7	0,08	-0,6
0	-52	-0,5	25	0	-24	-0,6	15

Paramètres génétiques			
0,10	0	0	0
0	0,30	-0,71	-0,15
0	-0,65	0,32	-0,20
0	-0,16	-0,55	0,60

héritabilités : sur la diagonale

corrélations phénotypiques :
au-dessus de la diagonale

corrélations génétiques :
au-dessous de la diagonale

2.3 / Covariance entre apparentés

La formule (5) est applicable au cas où deux caractères différents sont mesurés sur les deux apparentés en remplaçant les variances du second membre de l'équation par les covariances correspondantes. On a donc, en supposant l'absence d'épistasie :

$$Cov(G_x G_y) = Cov(G_y G_x) = 2f_{xy} Cov(AA') + u_{xy} Cov(DD') \quad (10)$$

Cette covariance vaut aussi pour les phénotypes en l'absence de covariance entre les effets de milieu (comme en 1.3 pour le cas d'un caractère).

3 / Les paramètres génétiques d'une population

La variation d'un caractère quantitatif polygénique est, comme on l'a vu, représentée par 3 variances, phénotypique, génétique additive et de dominance, la variance de milieu pouvant être obtenue par différence. Parallèlement la covariation de deux caractères est représentée par 3 covariances de même nature. Au-delà de deux caractères, il faut considérer toutes les covariances entre les caractères pris 2 à 2. Ces variances et covariances peuvent être disposées en tableaux carrés et symétriques (appelés aussi matrices). La génétique de n caractères quantitatifs peut ainsi se résumer à 3 tableaux de dimension $n \times n$, avec les variances dans la diagonale et les covariances hors-diagonale : le tableau des variances-covariances phénotypiques, celui des variances-covariances génétiques additives et celui des variances-covariances de dominance.

Pour avoir des valeurs indépendantes des unités de mesure des différents caractères, on exprime généralement la variance génétique additive relativement à la variance phénotypique et le rapport $V(A)/V(P)=h^2$ est appelé l'héritabilité du caractère. Pour la même raison les covariances phénotypiques et génétiques additives sont généralement présentées sous la forme de coefficients de corrélation : $Cov(PP') / [V(P)V(P')]^{1/2} = r_P$ est la corrélation phénotypique et $Cov(AA') / [V(A)V(A')]^{1/2} = r_A$ est la corrélation génétique (sous-entendu additive). Alors que h^2 varie de 0 à 1, r_A et r_P peuvent varier entre -1 et +1. Notons que, pour un locus, r_A est nécessairement égal à ± 1 (voir équation 8b).

La génétique de n caractères quantitatifs, satisfaisant aux hypothèses du modèle polygénique additif se résume alors à n^2 chiffres (situés entre -1 et +1) appelés paramètres génétiques (en ignorant les paramètres de dominance) : n héritabilités, $n(n-1)/2$ corrélations phénotypiques et $n(n-1)/2$ corrélations génétiques. Le passage des tableaux des variances-covariances à celui des paramètres génétiques est illustré au tableau 3 pour quatre caractères quantitatifs du Porc.

4 / Validité du modèle polygénique additif

Le modèle qui vient d'être décrit, dans cet article et dans le précédent, est le modèle de base de la génétique quantitative. Il permet de passer du monde concret des valeurs phénotypiques au monde concep-

tualisé, idéal pourrait-on dire, des valeurs génétiques additives. Les articles suivants montreront l'usage qui en est fait en amélioration génétique et, notamment, la manière dont les paramètres génétiques conditionnent le comportement des caractères quantitatifs d'une population tout au long des générations successives. C'est donc un modèle dont la principale caractéristique est d'être opérationnel.

Mais nous apprend-il quelque chose sur la génétique ? Le reproche qui lui est souvent fait est d'être trop éloigné de la réalité biologique. La controverse sur l'hérédité de l'intelligence, par exemple, exploite fréquemment ce genre d'argument et illustre bien les réticences de certains biologistes à l'égard des concepts de la génétique quantitative. On a encore pu lire dans un récent article de "La Recherche" que les expériences sur jumeaux (humains), en fait des études de covariances entre apparentés (vrais et faux-jumeaux), "n'apportent certainement rien sur les déterminations de l'intelligence, ni même sur le mode d'action des facteurs génétiques qui pourraient l'influencer".

4.1 / L'additivité est-elle un mythe ?

La première difficulté consiste peut-être à appréhender la notion même de caractère additif, qui associe d'un côté un concept qui n'est pas courant, la variation, et de l'autre une opération arithmétique aussi simple que l'addition. L'idée même de "construire" un caractère de nature biologique en "additionnant" des effets, comme on construit un mur avec des briques, est contraire à tout ce que l'on sait de la complexité du vivant. Et pourtant... Prenons, à titre d'exemple, un caractère biologique comme les empreintes digitales, dont la complexité peut difficilement être niée. On peut le quantifier en comptant le nombre de crêtes entre des repères bien définis et ainsi obtenir un caractère très variable d'un individu à l'autre, variabilité qui, comme on le sait, est exploitée depuis longtemps dans les enquêtes policières. En prenant des couples d'individus apparentés, des covariances $Cov(G_x G_y)$ peuvent être calculées pour ce caractère.

Les résultats du tableau 4 montrent d'abord que les covariances obtenues sont exactement proportionnelles au coefficient de parenté. En se reportant à la formule de la covariance entre apparentés pour le cas de plusieurs locus (équation (5)), on voit que le fait que $Cov(G_x G_y)$ soit proportionnel à f_{xy} implique qu'il n'y ait ni variance de dominance ni variance d'épistasie. On constate de plus que la covariance entre jumeaux monozygotes est très proche de la variance phénotypique (autrement dit, la corrélation est voisine de 1) et $V(A) = 0,98 V(P)$, puisque dans ce cas

Tableau 4. Covariances entre apparentés pour les empreintes digitales de l'Homme.

Relation de parenté	Coefficient de parenté f_{xy}	Covariance entre apparentés $Cov(G_x G_y)$
Jumeaux monozygotes	1/2	0,98 V(P)
Jumeaux dizygotes	1/4	0,49 V(P)
Parent-enfant	1/4	0,49 V(P)

$Cov(G_x G_y) = V(A)$ en l'absence de dominance et d'épistasie. Il s'agit donc d'un caractère dont quasiment toute la variabilité est de nature génétique additive, sans aucun effet de dominance, ni d'épistasie, ni de milieu.

En toute rigueur cependant, il reste 2% de variance non expliquée génétiquement qu'on peut attribuer au "milieu" (les aléas, non génétiques, du développement prénatal, par exemple), et qu'on pourrait, si on disposait d'autres corrélations, répartir d'une manière plus précise entre le milieu, la dominance et l'épistasie. Une étude a d'ailleurs montré l'existence plausible d'une certaine épistasie (AA) pour ce caractère (voir Heath *et al* 1984).

Cet exemple montre, bien qu'il s'agisse d'un cas limite, que le modèle polygénique additif peut, dans certains cas, approcher d'assez près une réalité biologique complexe.

4.2 / Plusieurs gènes c'est combien ?

Le modèle polygénique peut-il nous renseigner sur le nombre des gènes impliqués dans la réalisation d'un caractère quantitatif ? Avant de répondre à cette importante question, il est indispensable de préciser le sens à lui donner (voir, par exemple, Thoday et Thompson 1976). Il ne peut s'agir, comme la description du modèle l'a montré, que de gènes, ou plus précisément de locus, qui interviennent dans la variabilité observée. Leur nombre, quel qu'il soit, sera nécessairement inférieur à celui de tous les locus impliqués dans les processus biologiques relatifs au caractère. Ainsi, par exemple, le gène de l'hormone de croissance intervient biologiquement dans le poids atteint à un âge donné, mais, s'il n'y a pas de variation au locus codant pour cette hormone, il n'est pas impliqué dans les variations de poids observées. Il faut de plus se rappeler que les paramètres génétiques sont relatifs à une population donnée et que la question du nombre de locus est donc également relative à une population. Si celle-ci résulte d'un croisement entre deux races très éloignées génétiquement on peut s'attendre à y trouver plus de variabilité génétique et donc plus de locus.

Le modèle polygénique ne peut pas apporter de réponse très précise à la question du nombre de gènes, car, comme on l'a vu, on appréhende seulement une variation génétique globale, qui dépend à la fois du nombre des gènes et de la taille de leurs effets. Pour passer de la variance totale de A au nombre l de ses composantes dans l'équation (1), il faut supposer, entre autres hypothèses restrictives, que les variances à chaque locus sont égales. Si un caractère est déterminé par l locus, additifs entre eux et chacun possédant 2 allèles de fréquences égales avec les mêmes effets additifs a et $-a$, on peut en effet écrire :

$$V(A) = l(0,5 a^2) \quad (11)$$

Pour une valeur donnée de $V(A)$, la valeur de l ne dépend alors que de a . Si cette inconnue a peut être déterminée par ailleurs, le nombre l en découle. Par sélection divergente sur une population où $V(A)$ répond aux conditions exprimées dans (11), on peut fixer dans une lignée tous les allèles favorables au caractère (d'effet $+a$) et dans une autre tous les allèles défavorables (d'effet $-a$). La différence entre les deux lignées est alors $D=l$ (2a) et en reportant la valeur $a=D/2l$ dans (11) on obtient :

$$V(A) = D^2/8l \quad (12)$$

et une estimation du nombre l de locus :

$$l = D^2/8 V(A) \quad (13)$$

On peut donc, théoriquement, estimer le nombre l de locus, en sachant cependant que ce nombre constitue une limite inférieure, compte tenu de l'hypothèse faite plus haut d'égalité des fréquences géniques et des effets génétiques. La difficulté pratique réside non pas dans l'évaluation de $V(A)$ mais dans la réalisation de la sélection nécessaire pour évaluer D . Cette sélection doit en effet se poursuivre sur de nombreuses générations pour que les lignées arrivent à leur divergence maximum correspondant à des plateaux. De telles expériences ne peuvent être réalisées que sur des espèces de laboratoire. Les résultats obtenus donnent des valeurs de l qui sont rarement inférieures à quelques dizaines et qui peuvent souvent atteindre plusieurs centaines (voir, par exemple : Wright 1968, Falconer 1989). Il est donc raisonnable de penser que, chez les animaux de ferme aussi, plusieurs gènes c'est au moins quelques dizaines et probablement beaucoup plus du fait que les effets des gènes ne sont pas égaux, comme en témoigne l'existence de gènes majeurs.

En guise de conclusion : le modèle polygénique additif n'est-il pas le meilleur ?

On peut spéculer, comme le fait Crow (1985) dans un "éloge de l'additivité" repris ici en conclusion, que

le système génétique idéal du point de vue de l'amélioration par sélection est celui dans lequel interviennent de très nombreux gènes, agissant tous d'une manière additive et indépendante. Si les gènes agissent indépendamment, il est possible de sélectionner sur un caractère sans obtenir des effets indésirables sur d'autres systèmes. De plus, comme on le verra dans les articles suivants, avec des gènes additifs la supériorité des parents sélectionnés se retrouve, au moins partiellement, dans la descendance. Un avantage supplémentaire est la possibilité, s'il y a de multiples facteurs, d'ajuster au mieux chaque caractère, car, plus l'effet individuel de chaque gène est faible, plus les chances sont grandes de trouver la combinaison génétique adéquate. En outre, plus le nombre des facteurs génétiques est élevé, plus la variabilité génétique potentielle est grande relativement à la variabilité génétique exprimée. C'est ce qu'ont montré de nombreuses expériences de sélection de longue durée aboutissant à des lignées dont la moyenne dépasse la valeur de l'individu le plus extrême de la population de départ. Il est probable aussi que dans de telles expériences les mutations jouent un rôle non négligeable et, dans ce contexte, la variabilité génétique à long terme tend à être directement proportionnelle au nombre de locus impliqués dans le caractère (Lande 1975). Pour toutes ces raisons, des gènes additifs en grand nombre constituent le meilleur système possible et il n'est peut-être pas trop surprenant de voir que beaucoup de caractères sont si proches de cet idéal.

Références bibliographiques

Barker J.S.F., 1979. Inter-locus interactions : a review of experimental evidence. *Theoret. Pop. Biol.* 16 323-346.

Chapman A., 1985. *General and Quantitative Genetics*. Elsevier, Amsterdam, 408pp.

Crow J.F., 1985. Transmission genetics : the rules of inheritance. In : *General and Quantitative Genetics* (Chapman A., éd.), Elsevier, Amsterdam, 3-22.

Falconer D.S., 1989. *Introduction to Quantitative Genetics*. 3e édition. Longman, Harlow (Grande-Bretagne), 438pp.

Heath A.C., Martin N.G., Eaves L.J., Loesch D., 1984. Evidence for polygenic epistatic interactions in man. *Genetics* 106 : 719-727.

Lande R., 1975. The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genet. Res.* 26 : 221-235.

Ollivier L., 1981. *Eléments de Génétique Quantitative*. Masson, Paris, 152pp.

Thoday J.M., Thompson J.N. Jr., 1976. The number of segregating genes implied by continuous variation. *Genetica* 46 : 335-344.

Wright S., 1968. *Evolution and the Genetics of Populations*, Vol 1, Genetic and Biometric Foundations, University of Chicago Press, Chicago, 469pp.