

III. EFFETS DES GENES SUR LES CARACTERES

Jusqu'à présent, nous avons en fait peu parlé de gènes et d'allèles. Nous avons défini un concept statistique, la valeur génétique, qui traduit l'effet moyen d'un génotype sur le caractère étudié. Il est indispensable d'affiner notre analyse, c'est-à-dire de considérer les différents allèles d'un même locus, et ce pour au moins deux raisons. Tout d'abord, un géniteur ne transmet à sa descendance que la moitié de son génome : en un locus donné, il ne transmet pas la combinaison de gènes qu'il possède mais un gène tiré au hasard parmi les deux présents ; il est donc nécessaire de définir un paramètre traduisant l'effet des gènes transmis individuellement à la descendance. Ensuite, il est nécessaire de pouvoir faire le lien avec les résultats établis par la génétique des populations ; notamment, un point important concerne le lien entre l'évolution de la moyenne d'un caractère quantitatif (ce qui intéresse le sélectionneur) et l'évolution des fréquences alléliques sous l'effet de la sélection (ce qui constitue le phénomène biologique sous-jacent). Dans le présent chapitre, nous présentons la décomposition de la valeur génétique d'un individu et de la variance génétique au sein d'une population. Nous nous appuierons essentiellement sur un modèle à un locus avec un nombre quelconque d'allèles. Nous verrons ensuite comment ce modèle se généralise à des situations plus compliquées.

A. Le modèle à un locus

1. Hypothèses, expression de la valeur génétique et exemple

Considérons un locus avec un nombre d'allèles (n) quelconque ($n > 1$; dans le cas le plus simple, $n = 2$). Chaque allèle est désigné par un numéro, i , qui varie de 1 à n . L'allèle \mathcal{A}_i a une fréquence p_i dans la population observée ($\sum p_i = 1$). Nous supposons que cette population est de grande taille, et qu'elle est panmictique. Les fréquences des génotypes se déduisent simplement des fréquences alléliques (loi de Hardy-Weinberg ; voir GP) : soit f_{ij} la fréquence du génotype non ordonné $\mathcal{A}_i \mathcal{A}_j$, nous avons :

$$\begin{aligned} f_{ii} &= p_i^2, & \text{pour tout } i \\ f_{ij} &= 2p_i p_j, & \text{pour tout } i < j \end{aligned}$$

Tableau 3. Fréquences des génotypes et moyennes phénotypiques (a) dans le cas d'un locus à plusieurs allèles et (b) dans le cas d'un gène de nanisme affectant le poids à 6 semaines (en grammes) chez la souris (moyenne des deux sexes ; source : King, 1950).

Génotype	cas général	gène de nanisme		
	$\mathcal{A}_i \mathcal{A}_j$	++	+n	nn
Fréquence	f_{ij}	p_1^2	$2p_1p_2$	p_2^2
$p_2 = 0,2$		0,64	0,32	0,04
$p_2 = 0,5$		0,25	0,50	0,25
Moyenne phénotypique	μ_{ij}	14	12	6

On effectue des mesures chez un grand nombre d'individus. On suppose que l'on peut identifier le génotype de chaque individu au locus considéré et donc calculer une moyenne pour chacun des génotypes. On désigne par μ_{ij} la moyenne phénotypique des individus de génotype $\mathcal{A}_i \mathcal{A}_j$. Le tableau 3 (ci-dessus) résume la situation et donne l'exemple d'un gène de nanisme chez la souris (cf. figure 6, Ch. II) : l'allèle « normal », de fréquence p_1 , est désigné par +, le second allèle, de fréquence p_2 , est désigné par n comme « nain »⁴. Deux valeurs de la fréquence (p_2) de l'allèle n sont comparées : 0,2 et 0,5.

La moyenne de la population pour le caractère mesuré (μ) est la moyenne des moyennes phénotypiques par génotype pondérées par les fréquences correspondantes :

$$\mu = \sum_{i \leq j} f_{ij} \mu_{ij}$$

⁴ Cette notation n'est pas en accord avec la nomenclature internationale, qui emploie le symbole py pour « pygmée » ; nous n'avons pas retenu ici cette notation afin d'éviter toute confusion avec les notations employées pour les valeurs phénotypiques ou génétiques ou avec celles relatives aux fréquences

La moyenne de la population dépend donc des fréquences alléliques. Ainsi, dans le cas du gène de nanisme, la moyenne de poids à 6 semaines diminue quand p_2 augmente :

$$p_2 = 0,2 : \mu = 0,64 \times 14 + 0,32 \times 12 + 0,04 \times 6 = 13,04 \text{ g}$$

$$p_2 = 0,5 : \mu = 0,25 \times 14 + 0,50 \times 12 + 0,25 \times 6 = 11,00 \text{ g}$$

La valeur phénotypique d'un individu s'exprime comme vu précédemment :

$$P = \mu + G + E$$

En fait, nous n'étudions qu'un seul locus : cette situation correspond au cas où l'on a mis en évidence un seul gène majeur (ou un seul QTL) pour le caractère mesuré, les polygènes étant, par essence, non détectables. Ainsi, dans l'expression ci-dessus, la valeur génétique (G) est restreinte au génotype au locus étudié :

$$G = E(P \mid \text{génotype au locus étudié}) - \mu$$

Les valeurs génétiques de poids à 6 semaines chez la souris sont reportées au tableau 4. On note que la valeur génétique est une notion relative : elle dépend de la moyenne de la population dans laquelle on se trouve. Le cas de l'hétérozygote est particulièrement illustratif à cet égard : ce génotype a une valeur négative pour $p_2 = 0,2$ et une valeur positive pour $p_2 = 0,5$. Cette remarque est importante dans la pratique de l'amélioration génétique : selon les fréquences alléliques dans la population, un génotype peut être améliorateur ou détériorateur.

Les variations individuelles autour de la valeur génétique (cf., par exemple, figure 6, Ch.II), sont dues aux effets conjugués du micro-milieu et des polygènes : exceptionnellement, dans ce qui suit, nous considérons que la valeur résiduelle (E) rassemble ces deux familles d'effets.

Tableau 4. Valeur génétique (G , en grammes) des 3 génotypes au locus de nanisme selon la fréquence (p_2) de l'allèle « nain » dans la population (voir texte).

Génotype	$p_2 = 0,2$			$p_2 = 0,5$		
	++	+n	nn	++	+n	nn
G	+ 0,96	- 1,04	- 7,04	+ 3	+ 1	- 5

2. Notion d'effet moyen d'un gène

Nous avons défini la valeur génétique pour caractériser l'avantage/désavantage, que confère en moyenne un génotype donné par rapport à l'ensemble de la population où se trouve le porteur de ce génotype. En ce qui concerne les gènes, transmis individuellement à la descendance, on peut se poser une question analogue : quel avantage, ou désavantage, moyen procure un allèle donné aux individus qui en ont hérité ? Nous allons ainsi affecter à chaque allèle une valeur qui caractérisera cet avantage (ce désavantage). Cette valeur dépendra bien entendu de l'effet biologique de l'allèle, de la fréquence de ce dernier dans la population et du régime de reproduction (comme déjà signalé, nous nous restreignons ici à la panmixie).

Représentons-nous l'expérience suivante : un homozygote, de génotype connu $\mathcal{A}_i \mathcal{A}_i$, est utilisé comme géniteur d'un grand nombre (supposé infini) de descendants. Les descendants de ce parent auront tous reçu un exemplaire de l'allèle \mathcal{A}_i . Du fait de la panmixie, le second allèle est tiré au hasard dans la population. L'effet moyen de l'allèle \mathcal{A}_i est alors défini comme l'écart entre la valeur moyenne de la descendance décrite ci-dessus et la moyenne de la population.



L'effet moyen d'un allèle est l'espérance centrée de la valeur phénotypique des individus ayant reçu cet allèle d'un de leurs parents, l'autre allèle étant tiré au hasard. Cet effet moyen est également désigné « effet additif » d'un allèle.

En termes mathématiques, l'effet moyen (α_i) de l'allèle \mathcal{A}_i est une espérance conditionnelle, qui peut s'écrire de la façon suivante :

$$\alpha_i = E(P \mid \text{allèle } \mathcal{A}_i, \text{ allèle ?}) - E(P)$$

Elle se calcule en pondérant les moyennes des différents génotypes par les probabilités conditionnelles d'obtention. Sachant que l'allèle \mathcal{A}_i a obligatoirement été reçu du parent homozygote $\mathcal{A}_i \mathcal{A}_i$ et compte tenu de la panmixie, les probabilités recherchées sont égales aux fréquences alléliques dans la population (tableau 5). Ainsi, nous avons :

$$\alpha_i = \left[\sum_{j=1}^n p_j \mu_{ij} \right] - \mu$$

Tableau 5. Prédiction de la valeur de la descendance d'un parent homozygote $\mathcal{A}_i \mathcal{A}_i$: probabilités des apports gamétiques et moyennes phénotypiques correspondantes (voir texte).

	1er allèle	2° allèle					
	\mathcal{A}_i	\mathcal{A}_1	\mathcal{A}_2	...	\mathcal{A}_j	...	\mathcal{A}_n
Probabilité	1	p_1	p_2	...	p_j	...	p_n
moy. phéno.		μ_{i1}	μ_{i2}	...	μ_{ij}	...	μ_{in}

En notant que la somme des probabilités (p_j) est égale à 1, nous pouvons dire aussi :

$$\alpha_i = \sum_{j=1}^n p_j (\mu_{ij} - \mu)$$

Le terme entre parenthèses n'est autre que la valeur génétique (G_{ij}) du génotype $\mathcal{A}_i \mathcal{A}_j$, telle que nous l'avons définie précédemment. En définitive, l'effet moyen de l'allèle \mathcal{A}_i peut s'écrire également comme suit :

$$\alpha_i = \sum_{j=1}^n p_j G_{ij} = E(G \mid \text{allèle } \mathcal{A}_i, \text{ allèle ?})$$

En guise d'illustration, calculons les effets moyens des deux allèles du gène de nanisme chez la souris. Selon des notations analogues à celles du tableau 5, nous avons :

$$\alpha_+ = p_1 G_{++} + p_2 G_{+n} \quad \text{et} \quad \alpha_n = p_1 G_{+n} + p_2 G_{nn}$$

Soit (cf. tableau 4), pour une fréquence (p_2) de l'allèle n égale à 0,2,

$$\alpha_+ = 0,8 \times 0,96 + 0,2 \times (-1,04) = +0,56 \text{ g} ; \quad \alpha_n = 0,8 \times (-1,04) + 0,2 \times (-7,04) = -2,24 \text{ g}$$

et pour une fréquence de 0,5,

$$\alpha_+ = 0,5 \times 3 + 0,5 \times 1 = +2 \text{ g} ; \quad \alpha_n = 0,5 \times 1 + 0,5 \times (-5) = -2 \text{ g}$$

On note là encore la relativité de la notion d'effet moyen d'un gène : un allèle "favorable" l'est d'autant plus que la moyenne de la population est faible.

3. Décomposition de la valeur génétique

La valeur génétique en un locus est l'effet de la combinaison de deux gènes. Il est ainsi naturel de rapprocher la valeur génétique et la somme des effets moyens des deux gènes qui composent ce génotype (on ne fait pas autre chose en statistique quand on construit un modèle linéaire). Si la valeur génétique à la seule somme des effets moyens des deux gènes, la valeur génétique des hétérozygotes serait égale à la moyenne des valeurs génétiques des deux homozygotes correspondants. En effet, en un locus donné et pour deux allèles \mathcal{A}_i et \mathcal{A}_j ($i \neq j$), nous aurions les valeurs génétiques suivantes :

$\mathcal{A}_i \mathcal{A}_i$	$\mathcal{A}_i \mathcal{A}_j$	$\mathcal{A}_j \mathcal{A}_j$
$2 \alpha_i$	$\alpha_i + \alpha_j$	$2 \alpha_j$

En fait, ce n'est généralement pas le cas. Un examen des valeurs génétiques calculées pour le locus pygmée chez la souris en est une illustration (cf. tableau 4) : les hétérozygotes sont plus proches des homozygotes $++$ que des homozygotes nn . Il faut donc, en plus des effets moyens des gènes, introduire un terme traduisant l'interaction entre les deux gènes. D'une manière générale, nous écrivons :

$$G = \text{effet moyen du gène paternel} + \text{effet moyen du gène maternel} \\ + \text{effet d'interaction entre les 2 gènes}$$

Soit un individu de génotype $\mathcal{A}_i \mathcal{A}_j$ (i et j quelconques), G_{ij} désigne la valeur génétique de cet individu, α_i et α_j désignent les effets moyens des deux allèles et δ_{ij} désigne l'effet de leur interaction. Nous avons donc :

$$G_{ij} = \alpha_i + \alpha_j + \delta_{ij}$$

Par définition, on appelle **valeur génétique additive** la somme des effets moyens des gènes paternel et maternel ($\alpha_i + \alpha_j$). L'effet d'interaction (δ_{ij}) est appelé **valeur ou résidu de dominance** (résidu car il représente l'écart entre la valeur génétique et la valeur génétique additive). Soit A la valeur génétique additive d'un individu et D sa valeur de dominance, nous avons en définitive :

$$G = A + D$$

4. Récapitulation

La figure 10 indique, dans le cas du gène de nanisme chez la souris, comment les moyennes des trois génotypes se répartissent sur une échelle de valeur numériques. Ces moyennes traduisent l'effet biologique du locus dans un milieu donné et ne dépendent évidemment pas des fréquences alléliques dans la population. Par contre, la moyenne générale de la population en dépend, puisque elle est liée aux fréquences des différents génotypes (cf. § A.1). Cette figure permet également d'illustrer le fait que l'écart (d) entre la moyenne des hétérozygotes, d'une part, et la moyenne des deux homozygotes correspondants, d'autre part, traduit les relations de dominance, au sens de la génétique formelle, entre les deux allèles. Ici, on est dans une situation de dominance incomplète de l'allèle + par rapport à l'allèle n .

La figure 11 et le tableau 6 présentent la décomposition de la valeur génétique dans le cas du nanisme chez la souris (cf. les valeurs calculées plus haut). Plusieurs points, qui découlent de la construction même des variables introduites, sont à souligner :

- L'évolution des valeurs génétiques (G) selon le génotype est identique, à un facteur de centrage près, à celle des moyennes phénotypiques.
- Les valeurs génétiques additives (A) dépendent linéairement du nombre d'exemplaires d'un allèle donné dans le génotype.
- Les résidus de dominance (D) sont non nuls, y compris pour les génotypes homozygotes.
- Pour un génotype donné, toutes les valeurs (G , A ou D), dépendent des fréquences alléliques.

Cette décomposition de la valeur génétique, introduite par Fisher en 1918, revient à analyser une variable (G), qui dépend de deux facteurs (les deux gènes paternel et maternel), comme la somme des effets moyens de ces deux facteurs et d'un résidu. Pour définir les effets des gènes, Fisher a utilisé la méthode des moindres carrés, de nature strictement statistique : il a recherché les valeurs (α) qui minimisent sur l'ensemble d'une population l'espérance du carré du résidu de dominance. Dans l'encadré 4, on montre que, pour un locus biallélique, les valeurs obtenues par cette méthode sont bien les effets moyens définis au § A.2.

Figure 10. Moyennes phénotypiques des 3 génotypes au locus de nanisme chez la souris.

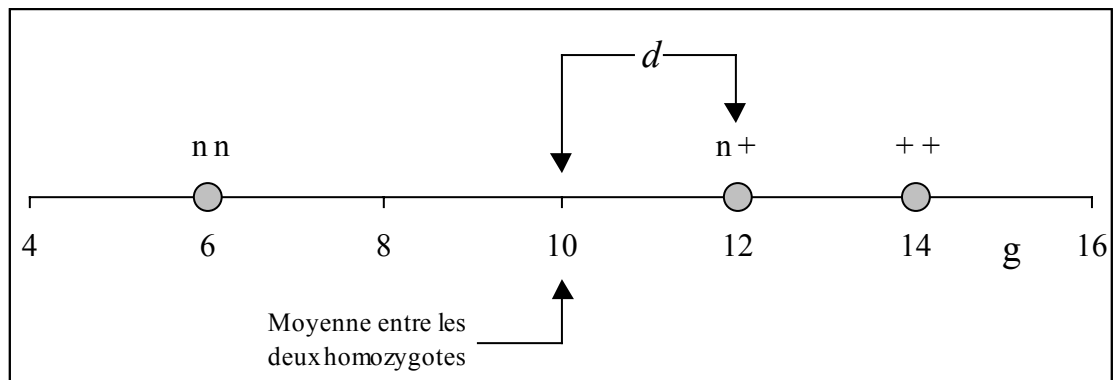


Figure 11. Valeurs génétiques (G), génétiques additives (A) et de dominance (D) dans le cas du gène de nanisme, la fréquence de l'allèle « nain » étant de 0,2 (voir texte). NB. L'axe des ordonnées (en grammes) est dans un repère centré sur la moyenne générale de la population.

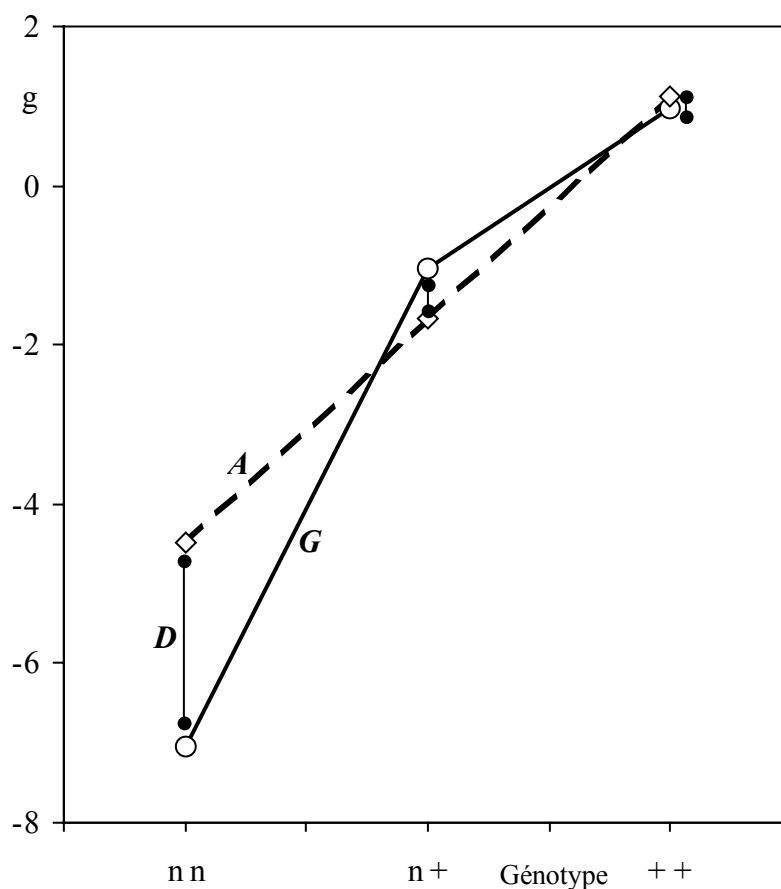


Tableau 6. Valeurs génétiques (G), génétiques additives (A) et de dominance (D) selon la fréquence (p_2) de l'allèle « nain » dans la population (voir texte).

Génotype	$p_2 = 0,2$			$p_2 = 0,5$		
	++	+n	nn	++	+n	nn
G	+ 0,96	- 1,04	- 7,04	+ 3	+ 1	- 5
A	+ 1,12	- 1,68	- 4,48	+ 4	0	- 4
D	- 0,16	+ 0,64	- 2,56	- 1	+ 1	- 1



La valeur génétique additive d'un individu est la somme des effets moyens des gènes qu'il possède. L'effet moyen d'un allèle particulier est l'espérance de la valeur génétique conditionnée par la présence de cet allèle dans le génotype. Cette valeur minimise l'espérance du carré du résidu de dominance.



En génétique animale, le terme « valeur génétique additive » est très souvent abrégé en « valeur génétique ». Dans le présent document, comme il est d'usage en génétique végétale, nous précisons toujours la composante concernée et « valeur génétique » désigne l'effet moyen du génotype, toutes composantes confondues, tel que défini au chapitre II.



Le terme « résidu de dominance » désigne un concept statistique : « dominance » a ici une autre signification que dans le cadre de la génétique formelle (comparer les figures 10 et 11). Toutefois, s'il y a additivité stricte au sens de la génétique formelle, c'est-à-dire si la valeur moyenne des hétérozygotes est strictement égale à la moyenne des deux homozygotes, alors les résidus de dominance sont tous nuls et on dit qu'il y a additivité stricte au sens de la génétique quantitative.

Encadré 4

Calcul de l'effet des gènes par la méthode des moindres carrés dans le cas bi-allélique

NB. Afin d'alléger les écritures, nous employons les notations suivantes :

- Deux allèles \mathcal{A}_1 et \mathcal{A}_2 , de fréquences respectives p et q et d'effets respectifs s et t .
- Les valeurs génétiques des génotypes $\mathcal{A}_1 \mathcal{A}_1$, $\mathcal{A}_1 \mathcal{A}_2$ et $\mathcal{A}_2 \mathcal{A}_2$ sont u , v et w , respectivement.

La valeur génétique se décompose ainsi :

$$G = A + D$$

Nous souhaitons minimiser l'espérance du carré du résidu de dominance (D). Sous l'hypothèse d'un équilibre de Hardy-Weinberg, cette espérance s'écrit :

$$E(D^2) = E[(G - A)^2] = p^2(u - 2s)^2 + 2pq[v - (s + t)]^2 + q^2(w - 2t)^2$$

Nous avons un minimum si les dérivées partielles en s et t sont nulles, soit :

$$\frac{\partial E(D^2)}{\partial s} = -4p^2(u - 2s) - 4pq[v - (s + t)] = 0$$

$$\frac{\partial E(D^2)}{\partial t} = -4pq[v - (s + t)] - 4q^2(w - 2t) = 0$$

Soit, en simplifiant par -4 et en séparant ce qui dépend de u , v et w , d'une part, et de s et t , d'autre part :

$$\begin{cases} p^2 u + pqv = 2p^2 s + pq(s + t) \\ pqv + q^2 w = pq(s + t) + 2q^2 t \end{cases}$$

Sommons ces deux équations membre à membre ; en utilisant la relation $p+q=1$, nous obtenons :

$$p^2 u + 2pqv + q^2 w = 2[ps(p + q) + qt(p + q)] = 2(ps + qt)$$

Le terme de gauche représente la moyenne des valeurs génétiques, qui sont des valeurs centrées ; il est par conséquent nul. Ainsi, la moyenne des effets des gènes pondérés par leur fréquence est nulle :

$$ps + qt = 0$$

Reprenons le système posé plus haut. Admettons que ni p ni q ne sont nuls (si p ou q était nul, il n'y aurait pas de polymorphisme) et simplifions la première équation par p et la seconde par q :

$$\begin{cases} pu + qv = 2ps + qs + qt = s + ps + qt \\ pv + qw = ps + pt + 2qt = t + ps + qt \end{cases}$$

Compte tenu de la liaison entre s et t établie plus haut, nous obtenons directement les expressions des effets des deux allèles :

$$\begin{cases} s = pu + qv \\ t = pv + qw \end{cases}$$

B. Généralisation du modèle

1. Généralisation des notions d'effet moyen

Les notions de valeur génétique et d'effet moyen d'un allèle, en un locus, se généralisent lorsque l'on considère plusieurs locus régissant un caractère quantitatif. La valeur génétique ($G_{l_{ij}}$) pour le génotype ij au locus l se définit comme l'espérance de la valeur phénotypique centrée des individus porteurs des allèles i et j à ce locus l :

$$G_{l_{ij}} = E(P \mid \text{génotype } ij \text{ au locus } l) - \mu$$

De la même façon, on définit l'effet moyen (α_{l_i}) d'un allèle i au locus l comme l'espérance de la valeur génétique conditionnée par la présence de cet allèle, les autres allèles (au locus considéré et aux autres locus) étant choisis au hasard :

$$\alpha_{l_i} = E(P \mid \text{allèle } i \text{ au locus } l) - \mu = E(G \mid \text{allèle } i \text{ au locus } l)$$

On désigne par $\delta_{l_{ij}}$ (cf. les notations du § A.3) le résidu de dominance pour le génotype ij au locus l : c'est l'écart entre l'effet moyen du génotype ($G_{l_{ij}}$) et la somme des effets moyens des gènes ($\alpha_{l_i} + \alpha_{l_j}$). La décomposition de $G_{l_{ij}}$ s'écrit alors ainsi :

$$G_{l_{ij}} = \alpha_{l_i} + \alpha_{l_j} + \delta_{l_{ij}}$$

La définition de l'effet moyen d'un génotype à un locus donné et celle de l'effet moyen d'un allèle à un locus donné sont valables quelle que soit la manière dont les différents locus combinent leurs effets (voir plus loin).

2. Principaux modèles avec plusieurs locus

a. Modèle avec additivité d'un locus à l'autre

Lorsque plusieurs locus interviennent dans la variation d'un caractère quantitatif (ce qui est le cas général), la manière la plus simple de se représenter la valeur génétique d'un individu

consiste à l'écrire comme la somme des valeurs génétiques à chacun des locus impliqués. Ce modèle est qualifié d'additif entre locus car il ne prend pas en compte d'éventuelles interactions entre locus (épistasie). Selon un tel modèle, la valeur génétique (G) d'un individu, en prenant en compte tous les locus, s'écrit donc :

$$G = \sum_l G_{l_{ij}}$$

En définissant la valeur génétique additive (A) et la valeur de dominance (D) d'un individu :

$$A = \sum_l (\alpha_{l_i} + \alpha_{l_j}) \quad ; \quad D = \sum_l \delta_{l_{ij}}$$

on généralise à plusieurs locus la décomposition vue pour un locus : $G = A + D$

b. Modèle avec épistasie

Le modèle précédent a le mérite de la simplicité. Cependant, on peut chercher à prendre en compte les interactions entre gènes situés à des locus différents. La valeur génétique additive (A) et la valeur de dominance (D) étant définies comme précédemment, on calcule plusieurs termes résiduels, dont la somme représente l'écart qui existe entre la valeur génétique multilocus et la somme $A + D$: on parle de résidus d'épistasie. Ces effets d'épistasie sont des résidus au sens de la statistique : par construction, les variations qui leur sont imputables sont minimisées par rapport aux effets définis précédemment, effets additifs et effets de dominance. Ainsi, dans la pratique, les résidus d'épistasie sont le plus souvent négligés.

3. Modèles et réalité

a. Plusieurs locus, c'est combien ?

Nous avons dit au précédent chapitre que l'on était incapable de déterminer le nombre de locus régissant un caractère quantitatif. Au mieux, peut-on mettre en évidence quelques QTL induisant des variations plus ou moins importantes. Ainsi, dans la pratique, on n'est en général pas amené à réaliser les calculs qui viennent d'être présentés. On retient simplement la décomposition de la valeur génétique en une fraction additive et une fraction résiduelle. L'approche des caractères quantitatifs est alors essentiellement statistique. Pour utiliser les

variables introduites plus haut, il est nécessaire de connaître leur distribution. Aussi, devant les limites de nos outils d'analyse, est-on amené à émettre des hypothèses quant aux locus responsables des variations d'un caractère quantitatif. En dehors des gènes que l'on a pu détecter (gènes majeurs, QTL), on suppose qu'un caractère quantitatif est régi par un grand nombre de locus (d'où le terme « polygènes ») chacun induisant de faibles variations. Cette hypothèse permet d'admettre la normalité des distributions des valeurs génétiques (cf. figure 7, Ch. II) et sert de cadre à un modèle couramment utilisé, le modèle polygénique. Il est possible, sur la base du modèle introduit, de réaliser des expériences visant à estimer le nombre de locus induisant des variations sur un caractère quantitatif. Les quelques expériences de ce type qui ont été réalisées montrent que, le plus souvent, plusieurs locus c'est au moins quelques dizaines. Cet ordre de grandeur peut être considéré comme faible mais les développements que l'on peut faire du modèle à plusieurs locus montrent, cependant, que les variables que nous avons introduites (G, A, D) ont des propriétés qui, lorsque l'on fait croître le nombre de locus considérés, tendent très vite vers celles de variables normales.

b. Modèle et complexité biologique

Les concepts introduits sont des concepts fondamentaux en génétique quantitative. Les analyses de génétique quantitative et la pratique de l'amélioration génétique sont largement fondées sur le modèle polygénique, souvent même sur un modèle polygénique strictement additif (c'est-à-dire ignorant la dominance). On peut se demander si cette approche statistique relativement simple est à même de rendre compte de la complexité du fonctionnement biologique des organismes, depuis l'expression des gènes jusqu'à un phénotype en passant par les synthèses protéiques, le métabolisme et la physiologie. De nombreux reproches ont d'ailleurs été formulés à l'égard de ce modèle dans les dernières décennies. Il ne faut cependant pas perdre de vue que cette approche n'a aucune prétention explicative, et que son but est essentiellement opérationnel ou prédictif de l'évolution à plus ou moins court terme de la valeur de caractères quantitatifs au sein de populations soumises à la sélection naturelle ou artificielle. La décomposition suivant un modèle additif est fréquente en statistique et permet souvent des approximations de bonne qualité. Ainsi, les insuffisances du modèle à l'échelon du fonctionnement biologique n'excluent pas un bon pouvoir prédictif à l'échelon macroscopique d'un phénotype quantitatif et de ses lois de transmission.

C. Moyenne des valeurs au sein d'une population

1. Moyenne des valeurs au sein d'une population

La moyenne phénotypique (μ) dépend des fréquences des génotypes et donc des fréquences alléliques (cf. §A.1, pour le cas d'un seul locus).

Les valeurs génétiques sont centrées sur la moyenne phénotypique de la population. Ainsi, par construction, l'espérance de la valeur génétique (G) d'un individu pris au hasard dans la population est nulle :

$$E(G) = E[E(P - \mu)] = E(P) - \mu = 0$$

De la même façon, l'effet moyen d'un allèle, en un locus donné, représente un écart moyen conditionné par la présence de cet allèle. Ainsi, sur l'ensemble de la population (l'ensemble des allèles), les effets moyens des allèles sont des valeurs centrées : leur moyenne pondérée par les fréquences est nulle, soit avec les notations introduites plus haut et pour un locus :

$$\sum_{i=1}^n p_i \alpha_i = 0$$

La valeur génétique additive étant la somme des effets moyens des gènes, la moyenne de ces valeurs est donc, par construction, nulle également, quel que soit le nombre de locus :

$$E(A) = 0$$

Des moyennes génétique (totale) et génétique additive établies ci-dessus, il découle naturellement que la moyenne des résidus de dominance est nulle sur l'ensemble de la population :

$$E(D) = E(G - A) = E(G) - E(A) = 0$$

En définitive, la moyenne phénotypique d'une population est susceptible d'évolution lorsque les fréquences alléliques évoluent. Les autres variables que nous avons définies sont, pour une population donnée et à un instant donné, des variables centrées.

D. Intérêt prédictif de la notion de valeur génétique additive

1. Prédiction de la valeur d'un descendant connaissant un seul parent

Considérons un parent de génotype connu, dont nous connaissons la valeur génétique (G) et la valeur génétique additive (A). Supposons que ce parent se reproduise avec des individus choisis au hasard dans la population. Ce cas de figure correspond à de nombreuses situations concrètes où un géniteur, fournissant un grand nombre de gamètes, est le parent commun d'un grand nombre de descendants ; chez les animaux domestiques, le parent commun est le plus souvent le mâle.

a. Cas d'un locus biallélique

Ramenons nous au cas simple d'un locus à deux allèles et considérons un parent de génotype hétérozygote $A_1 A_2$ (le raisonnement serait identique avec un autre génotype). Nous pouvons établir un tableau de croisement décrivant la situation (tableau 7). Les gamètes s'associant au hasard, les probabilités d'obtention des différents génotypes sont égales aux produits des probabilités marginales correspondantes.

Tableau 7. Tableau de croisement entre un premier parent hétérozygote $A_1 A_2$ et un second parent choisi au hasard dans la population.

p_1, p_2 = fréquences alléliques ; G_{11}, G_{12}, G_{22} = valeurs génétiques (voir texte).

	1er parent	$\frac{1}{2}$ A_1	$\frac{1}{2}$ A_2
2ème parent	p_1 A_1	$\frac{1}{2} p_1$ $A_1 A_1$ G_{11}	$\frac{1}{2} p_1$ $A_1 A_2$ G_{12}
	p_2 A_2	$\frac{1}{2} p_2$ $A_1 A_2$ G_{12}	$\frac{1}{2} p_2$ $A_2 A_2$ G_{22}

A partir de ce tableau de croisement, nous pouvons calculer l'espérance de la valeur génétique (G) d'un descendant conditionnée par le génotype $\mathcal{A}_1 \mathcal{A}_2$ du parent fixé (dénommé "1er parent" dans le tableau) :

$$E(G \mid \text{parent } \mathcal{A}_1 \mathcal{A}_2) = \frac{1}{2} p_1 G_{11} + \frac{1}{2} p_2 G_{12} + \frac{1}{2} p_1 G_{12} + \frac{1}{2} p_2 G_{22}$$

Par définition, la somme des deux premiers termes du membre de droite est égale à l'effet moyen de l'allèle \mathcal{A}_1 , la somme des deux derniers à l'effet de l'allèle \mathcal{A}_2 (cf. § A.2). Ainsi,

$$E(G \mid \text{parent } \mathcal{A}_1 \mathcal{A}_2) = \frac{1}{2} \alpha_1 + \frac{1}{2} \alpha_2$$

Le même raisonnement s'applique à tous les génotypes possibles du parent connu, le second parent étant choisi au hasard dans la population (tableau 8). On voit que, dans tous les cas, la prédiction de la valeur génétique des descendants, sachant un des parents, n'est affectée d'aucun résidu de dominance. Ce résultat découle de la définition même des effets moyens des gènes (cf. le raisonnement qui avait conduit à cette définition, § A.2). Ce résultat rend bien compte du fait qu'en un locus donné, un parent transmet un seul gène et non pas la combinaison des deux gènes qu'il possède.

Tableau 8. Valeur génétique attendue de la descendance d'un parent connu selon le génotype de ce dernier en un locus biallélique.

Parent connu		Valeur attendue de la descendance
Génotype	Valeur génétique	
$\mathcal{A}_1 \mathcal{A}_1$	$2\alpha_1 + \delta_{11}$	α_1
$\mathcal{A}_1 \mathcal{A}_2$	$\alpha_1 + \alpha_2 + \delta_{12}$	$\frac{1}{2} \alpha_1 + \frac{1}{2} \alpha_2$
$\mathcal{A}_2 \mathcal{A}_2$	$2\alpha_2 + \delta_{22}$	α_2

b. Généralisation à un nombre quelconque de locus

Ce résultat se généralise à un nombre quelconque de locus. Soit P le parent connu et I un descendant de ce parent, l'espérance de la valeur génétique de ce dernier (G_I) vaut :

$$E(G_I | G_P) = \frac{1}{2} A_P$$

La prédiction de la valeur phénotypique (P_I) du descendant s'obtient simplement en ajoutant la moyenne de la population (μ) :

$$E(P_I | G_P) = \mu + \frac{1}{2} A_P$$

Si l'on effectue le même raisonnement pour prédire la valeur génétique additive (A_I) et la valeur de dominance (D_I) du descendant (on rappelle que $G_I = A_I + D_I$), nous avons :

$$E(A_I | G_P) = \frac{1}{2} A_P$$

$$E(D_I | G_P) = E(D) = 0$$

Par construction, l'espérance des résidus de dominance est nulle dans toute descendance, quand un des parents est connu et que le second parent est choisi au hasard.



La valeur génétique additive (A) représente la fraction de la valeur génétique (G) dont on peut facilement prédire la transmission de parent à descendant : un parent transmet, en espérance, la moitié de sa valeur génétique additive.

La valeur génétique additive d'un parent représente ainsi, à un facteur $\frac{1}{2}$ près, la valeur attendue des descendants. Cela découle de la définition de l'effet moyen d'un allèle, qui représente la valeur attendue des descendants d'un parent homozygote pour l'allèle en question. On constate donc que ***les concepts d'effet moyen d'un allèle et de valeur génétique***

additive sont d'une grande utilité dans le cas de l'utilisation non spécifique et à grande échelle d'un reproducteur, puisqu'ils *permettent de prédire la valeur de la descendance* de ce dernier. Ce caractère opérationnel de la notion de valeur génétique additive explique sa très large utilisation en sélection, en tant que critère de choix des reproducteurs.

2. Prédiction connaissant les deux parents

Ce que nous venons de décrire quand nous connaissons un seul parent ne se généralise que partiellement quand nous connaissons les deux parents. En effet, dans ce cas, l'espérance de la valeur génétique additive (A) d'un descendant (I) connaissant celles de ses parents (P et M) est égale à la moyenne des valeurs génétique additives parentales :

$$E(A_I | A_P, A_M) = \frac{1}{2} A_P + \frac{1}{2} A_M$$

Par contre, l'espérance de la valeur génétique (G) d'un descendant ne s'exprime plus simplement en fonction des seules valeurs génétiques additives parentales : lorsque nous connaissons le génotype des deux parents, l'espérance du résidu de dominance n'est pas nulle dans toutes les descendance. Il suffit de considérer le cas d'un locus à deux allèles et du croisement entre un homozygote A_1 et un homozygote A_2 : tous les descendants sont hétérozygotes et leur valeur génétique peut être différente de la moyenne des valeurs génétiques additives parentales ($G_{12} = \alpha_1 + \alpha_2 + \delta_{12} \neq \alpha_1 + \alpha_2$).

3. Conséquences pour la pratique de l'amélioration génétique

La décomposition de la valeur génétique permet de dégager deux voies d'amélioration :

- On peut rechercher *les individus* ayant la meilleure *valeur génétique additive* : c'est ce que l'on fait en *sélection*, où l'on choisit les reproducteurs pour ce qu'ils sont susceptibles de transmettre à leur descendance, indépendamment des individus auxquels ils s'uniront.
- On peut rechercher *les couples* dont la descendance présentera des *effets d'interaction* maximaux : c'est le cas lorsque l'on fait se reproduire entre eux des géniteurs provenant de populations différentes (voir Ch. V).

Dans la pratique, ces deux voies d'amélioration ne sont pas antagonistes mais se complètent. Par exemple, quand les différences génétiques entre individus sont principalement de nature additive, la recherche des meilleurs couples peut se faire en deux étapes : (1) sélection d'un nombre restreint de géniteurs à forte valeur additive et (2) recherche des meilleurs couples parmi ceux qu'il est possible de constituer à partir des géniteurs retenus à la première étape.

E. Décomposition de la variance et paramètres génétiques

Au sein d'une population, les variables que nous avons introduites sont considérées comme des variables aléatoires. Dans cette section, nous voyons comment la variance d'un caractère dépend de la structure génétique de la population.

Soit une population de grande taille, en équilibre de Hardy-Weinberg à chaque locus. Nous développons les formules pour le modèle à un locus (cf. les notations au § A.2) et envisageons leur généralisation à plusieurs locus. Pour cela, d'une part, les résidus d'épistasie sont négligés et, d'autre part, nous supposons qu'il y a indépendance statistique entre les différents locus (équilibre d'association des gènes). Par ailleurs, les seuls effets de milieu considérés ici sont des effets de micro-milieu : il est particulièrement important de bien retenir que *les paramètres génétiques qui sont définis ici correspondent à des valeurs phénotypiques dépouillées de tout effet de milieu systématique induit par un facteur contrôlé*. Le modèle de base est donc le suivant :

$$P = G + E \quad \text{et} \quad G = A + D$$

1. Expression des variances

On rappelle que la variance phénotypique (V_P) est la variance des valeurs mesurées sur tous les individus d'une population et qu'elle est égale (cf. Ch.II, § C.3) à la somme des variances génétique (V_G) et environnementale (V_E).

a. Variance génétique

G étant une variable centrée, sa variance est égale à l'espérance de son carré. Soit, en un locus donné, G_{ij} la valeur génétique du génotype non ordonné $\mathcal{A}_i \mathcal{A}_j$ et f_{ij} sa fréquence,

$$V_G = E(G^2) = \sum_{i \leq j} f_{ij} G_{ij}^2$$

Sous les hypothèses d'absence d'épistasie et d'équilibre d'association des gènes, la variance génétique est la somme sur tous les locus des variances génétiques calculées à chaque locus comme ci-dessus.

b. Variance génétique additive

Soit un locus à plusieurs allèles, l'allèle i ayant une fréquence p_i et un effet moyen α_i . Nous rappelons que ces effets moyens sont des variables centrées. La variance des effets moyens des allèles dans la population vaut donc, par définition :

$$\text{Var}(\alpha) = E(\alpha^2) = \sum_i p_i \alpha_i^2$$

La valeur génétique additive d'un individu est la somme des effets moyens des 2 allèles qu'il a reçus au hasard et indépendamment l'un de l'autre (panmixie). Elle s'écrit donc comme la somme de deux variables aléatoires indépendantes de même variance $[\text{Var}(\alpha)]$. La variance génétique additive (V_A) vaut alors 2 fois la variance de la variable élémentaire :

$$V_A = E(A^2) = 2\text{Var}(\alpha) = 2 \sum_i p_i \alpha_i^2$$

Sous l'hypothèse d'équilibre panmictique intra-locus et d'équilibre d'association des gènes entre locus, la variance génétique additive est la somme sur tous les locus des variances génétiques additives calculées à chaque locus.

d. Variance de dominance

Soit δ_{ij} la valeur de dominance pour le génotype ij en un locus donné, la variance de dominance (V_D) s'écrit :

$$V_D = E(D^2) = \sum_{i \leq j} f_{ij} \delta_{ij}^2$$

Pour plusieurs locus en équilibre d'association, V_D est la somme sur tous les locus des variances de dominance calculées à chaque locus.

L'analyse du cas simple d'un locus à deux allèles permet de préciser comment la variance de dominance varie en fonction du mode d'action (au sens de la génétique formelle) des gènes. Désignons par d l'écart entre la moyenne des deux homozygotes et la valeur de l'hétérozygote (cf. figure 10) :

$$d = \mu_{12} - \frac{\mu_{11} + \mu_{22}}{2}$$

On montre alors que V_D s'écrit comme suit :

$$V_D = 4p_1^2 p_2^2 d^2 = (2p_1 p_2 d)^2$$

Lorsque d est nul, la valeur de l'hétérozygote est égale à la moyenne des valeurs des deux homozygotes : nous sommes en situation de stricte additivité des effets des gènes. Les résidus de dominance (D , au sens de la génétique quantitative) sont alors nuls pour les trois génotypes et la variance de dominance est nulle. A l'opposé, V_D est relativement importante en cas de dominance au sens mendélien du terme ($\mu_{11} = \mu_{12} > \mu_{22}$) et plus encore dans le cas de surdominance ($\mu_{12} > \mu_{11} > \mu_{22}$).

d. Evolution des variances en fonction des fréquences alléliques

Les expressions développées ci-dessus indiquent que les variances génétiques sont étroitement liées aux fréquences alléliques. L'analyse de quelques cas simples impliquant un seul locus biallélique va nous permettre de préciser cette liaison (figure 12) :

- D'une manière générale, toutes les variances génétiques sont faibles dès lors qu'un des allèles a une fréquence proche de 1. A l'extrême, lorsqu'il n'y a plus qu'un seul allèle dans la population, toutes les variances sont nulles (pas de polymorphisme).
- Dans les cas allant de l'additivité stricte (fig. 12.a) à la dominance complète (fig 12.b), la variance génétique (V_G) est maximale pour des fréquences alléliques intermédiaires, strictement égales à $\frac{1}{2}$ - $\frac{1}{2}$ dans le premier cas, autour dans le second.
- En cas d'additivité stricte, comme déjà signalé, la variance génétique additive (V_A) est égale à la variance génétique totale (V_G).
- En cas de dominance complète, la variance génétique additive (V_A) est maximale pour une fréquence faible de l'allèle dominant.
- La variance génétique de dominance (V_D) est généralement inférieure à la variance génétique additive. C'est seulement en cas de surdominance qu'elle peut représenter une part majeure de la variance génétique totale. Elle est toujours maximale pour deux allèle équiprobables (cf. aussi l'expression développée plus haut).



Les variances génétique (V_G), génétique additive (V_A) et de dominance (V_D) dépendent (1) des fréquences alléliques et (2) de la nature du caractère.

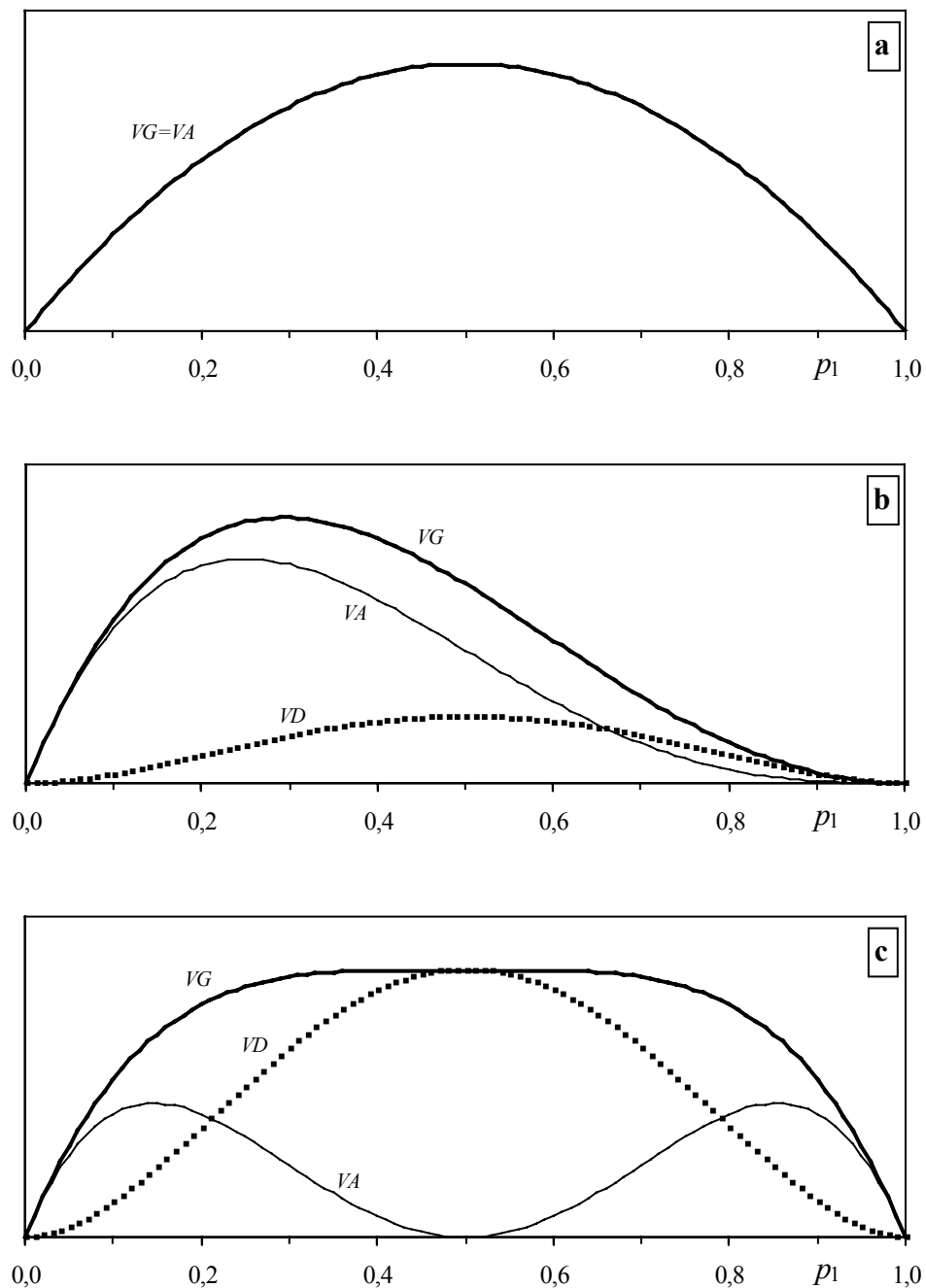
Ainsi, toute facteur d'évolution des fréquences alléliques (sélection, dérive, etc.) fait également évoluer la variance génétique d'un caractère et chacune de ses composantes. De même, lorsque l'on considère plusieurs locus, tout facteur susceptible d'induire un déséquilibre d'association des gènes (sélection, homogamie, croisement, etc.) est susceptible de modifier la variance génétique et ses différentes composantes.

Figure 12. Evolution des variances selon les fréquences alléliques.

Ligne continue épaisse = variance génétique (V_G), ligne continue fine = variance génétique additive (V_A), ligne pointillée = variance de dominance (V_D).

Cas d'un locus biallélique A_1 / A_2 , avec p_1 = fréquence de l'allèle A_1 . Soit μ_{ij} la moyenne phénotypique du génotype $A_i A_j$, 3 situations sont considérées :

- a) stricte additivité : $\mu_{12} = \frac{1}{2}(\mu_{11} + \mu_{22})$ - les deux lignes continues (V_G et V_A) sont confondues -
- b) dominance complète : $\mu_{11} = \mu_{12} > \mu_{22}$
- c) surdominance avec égalité des 2 homozygotes : $\mu_{12} > \mu_{11} = \mu_{22}$



2. Notion d'héritabilité au sens étroit

a. Définition

Nous avons décomposé la valeur phénotypique d'un individu en définissant, de façon statistique, plusieurs variables élémentaires, toutes considérées comme des variables aléatoires, dont on peut calculer une moyenne et une variance. Aussi, pouvons nous exprimer la variance phénotypique, la seule que l'on puisse observer concrètement, en fonction des variances des variables élémentaires. Nous savons déjà que :

$$V_P = V_G + V_E$$

Par ailleurs, on montre (ce qu'ici on admettra) qu'au sein d'une population, les valeurs génétiques additives (A) et de dominance (D) sont non corrélées :

$$\text{Cov}(A, D) = 0$$

Il s'agit là d'une propriété générale de la décomposition par les moindres carrés (voir Stat), qui nous permet d'écrire simplement :

$$V_G = V_A + V_D$$


En définitive, la variance phénotypique s'écrit :

$$V_P = V_A + V_D + V_E$$

La valeur génétique additive étant la fraction de la valeur génétique dont on peut facilement prédire la transmission, il est intéressant d'apprécier l'importance de sa variabilité comparée aux autres sources de variation. Pour ce faire, ***on définit l'héritabilité au sens étroit d'un caractère (h^2) comme le rapport entre la variance génétique additive et la variance phénotypique :***

$$h^2 = \frac{V_A}{V_P}$$

Nous verrons par la suite que l'héritabilité au sens étroit est un concept fondamental en génétique quantitative, qui a une très grande utilité dans la pratique de l'amélioration des plantes et des animaux (voir les enseignements correspondants).

 a) Ne pas confondre $h^2 = V_A/V_P$, hérabilité au sens étroit (parfois notée h_{ss}^2 , pour sensus stricto) et $h_{sl}^2 = V_G/V_P$, hérabilité au sens large (cf. § C.3, Ch. II).

b) En génétique animale, l'usage a fait que quand on parle d'hérabilité sans autre précision, implicitement, il s'agit toujours du sens étroit.


c) Au sens large ou au sens étroit, h_{sl}^2 ou h^2 correspondent à l'hérabilité elle-même et non à son carré. Ce symbolisme provient de la terminologie introduite par Wright (1921) qui utilisait h pour désigner le rapport des écarts-types.

b. Facteurs de variation de l'hérabilité au sens étroit

L'hérabilité au sens étroit est un paramètre génétique que l'on doit estimer, par une analyse statistique de données phénotypiques (voir Ch.IV). Partant de la définition de l'hérabilité au sens étroit, nous allons nous intéresser ici aux facteurs de variation de ce paramètre, sans nous préoccuper de son estimation. En remplaçant V_P par son expression, on obtient :

$$h^2 = \frac{V_A}{V_A + V_D + V_E}$$

L'hérabilité au sens étroit est un rapport compris entre 0 et 1 (ou entre 0 et 100%). La formule ci-dessus nous montre que l'hérabilité dépend (a) des fréquences alléliques dans la population, puisque V_A et V_D en dépendent, (b) de l'importance des variations dues aux effets de dominance (c) de l'importance des variations dues au micro-milieu (on rappelle que l'on ne considère ici que le milieu non contrôlé).

 ***Comme toute composante de la variance, l'hérabilité au sens étroit est un paramètre spécifique du caractère étudié et de la population observée. Elle dépend également du milieu dans lequel se trouve la population***

Toute évolution des fréquences alléliques fait évoluer les variances génétique additive et de dominance. Ainsi, l'hérabilité au sens étroit d'un caractère est susceptible d'évoluer sous l'effet de la sélection ou de la dérive génétique. Très schématiquement, toutes choses étant égales par ailleurs, l'hérabilité au sens étroit est plus élevée pour des fréquences intermédiaires que pour des fréquences extrêmes.

L'héritabilité au sens étroit d'un caractère est d'autant plus grande que les conditions de micro-milieu sont homogènes (V_E faible). On peut mettre à profit cette propriété en rassemblant les individus dans le même lieu où l'on cherche à uniformiser les conditions de milieu. En effet, chez les espèces domestiques, l'héritabilité au sens étroit d'un caractère, estimée à partir d'individus de la même population, est plus grande pour des individus rassemblés en centre expérimental ou en station, plutôt que pour des individus en exploitation agricole, et cela bien que l'on tienne compte des effets de macro-milieu engendrés par les différentes exploitations : ce dont il est question est la variation environnementale d'un individu à un autre (micro-milieu), plus faible en station qu'en exploitation, alors que les effets de macro-milieu sont communs à tous les individus d'une même exploitation.

Enfin, bien que l'héritabilité puisse varier d'une population à une autre, on trouve de grandes tendances qui permettent de distinguer les caractères entre eux. C'est ce qui est analysé dans les deux paragraphes qui suivent ; on ne s'intéressera qu'à des populations de plantes cultivées ou d'animaux d'élevage, compte tenu de l'étude approfondie à laquelle celles-ci sont généralement soumises et de l'utilité en sélection du concept d'héritabilité au sens étroit.

c. L'héritabilité au sens étroit chez les animaux

En ce qui concerne les animaux, on peut donner des ordres de grandeur de l'héritabilité au sens étroit d'un grand nombre de caractères (tableau 9). Ces chiffres représentent une synthèse de résultats obtenus à partir de données de terrain. Schématiquement, les caractères quantitatifs chez les animaux peuvent être regroupés en trois principales catégories :

- Caractères peu héritable ($h^2 < 0,2$) : essentiellement des caractères liés aux aptitudes de reproduction et de viabilité des jeunes. D'une part, ces caractères sont très sensibles aux conditions de milieu (V_E très élevée) et, d'autre part, des phénomènes génétiques non additifs (notamment des interactions de dominance) induisent probablement d'importantes variations sur ces caractères (V_D élevée ; voir chapitre V).
- Caractères moyennement héritable ($0,2 < h^2 < 0,4$) : essentiellement des caractères liés à l'intensité d'une production.
- Caractères fortement héritable ($h^2 > 0,4$) : essentiellement des caractères liés aux caractéristiques qualitatives des produits, notamment leur composition, qui sont beaucoup moins sensibles aux variations liées au milieu que les précédents.

Tableau 9. Ordre de grandeur de l'héritabilité au sens étroit (h^2) de certains caractères mesurés chez des animaux d'élevage.

Sources : Lerner, 1964 ; INRA-Génétique cheval, 1982, 1992 ; Gjedrem, 1983 ; Sellier, 1988 ; Ricordeau, 1992

	h^2
Bovins	
Taux de réussite en 1ère insémination artificielle	0,00 - 0,05
Vitesse de croissance	0,25 - 0,35
Production laitière (traite)	0,25 - 0,35
Teneur en protéines du lait	0,45 - 0,55
Ovins	
Taille de portée	0,05 - 0,10
Taux de mortalité avant sevrage	0,00 - 0,05
Finesse de la laine	0,60 - 0,65
Production laitière (allaitement)	0,20 - 0,30
Caprins	
Production laitière (traite)	0,25 - 0,35
Teneur en protéines du lait	0,50 - 0,55
Porc	
Taille de portée	0,05 - 0,10
Indice de consommation	0,25 - 0,35
Epaisseur de lard dorsal	0,30 - 0,60
Taux de muscle dans la carcasse	0,50 - 0,60
Cheval (sport et course)	
Record sur 1 km entre 3 et 4 ans (trot)	0,30 - 0,40
Gain annuel en concours de saut d'obstacles	0,25 - 0,35
Lapin	
Taille de portée	0,00 - 0,20
Poids de la portée à l'abattage	0,10 - 0,30
Poule	
Taux d'éclosion	0,05 - 0,20
Age à la maturité sexuelle	0,20 - 0,40
Production d'œufs	0,15 - 0,30
Poids moyen de l'œuf	0,40 - 0,60
Truite arc-en-ciel	
Taux de mortalité des alevins	0,05 - 0,20
Poids à 2 ans	0,25 - 0,35
Teneur en gras de la carcasse	0,40 - 0,60

d. L'héritabilité au sens étroit chez les plantes

Chez les animaux, la notion d'héritabilité calculée sur la base des valeurs individuelles a un sens et les régimes de reproduction sont en général considérés comme stables. Chez les plantes, il en va tout autrement. D'une part, l'entité mesurée est souvent un peuplement constitué de nombreux individus. D'autre part, les régimes de reproduction sont facilement modifiés par le généticien : clonage, consanguinité, hybridation forcée (voir Ch. V), etc. Il en découle que selon les conditions expérimentales et selon l'espèce observée, les héritabilités sont calculées très différemment et ne peuvent pas être comparées. On constate toutefois la tendance suivante : plus un caractère est complexe, plus il fait intervenir d'étapes dans le développement de la plante et moins, toutes choses étant égales par ailleurs, il est héritable, au sens étroit comme au sens large.

3. Notions de corrélation phénotypique et de corrélation génétique

La liaison entre caractères est un phénomène courant et dont l'observation est aisée (voir chapitre I). L'importance de la liaison se mesure statistiquement au travers de la corrélation. Soit une population dans laquelle on mesure les individus pour deux caractères différents, avec P_1 la valeur phénotypique pour le premier caractère et P_2 celle pour le second. La corrélation calculée sur un grand nombre d'individus mesurés pour les deux caractères représente, par définition, la **corrélation phénotypique** entre les deux caractères (r_p):

$$r_p = \frac{\text{Cov}(P_1, P_2)}{\sigma_{P_1} \sigma_{P_2}}$$

On montre que la covariance phénotypique peut avoir une origine ou environnementale :

$$\text{Cov}(P_1, P_2) = \text{Cov}(G_1, G_2) + \text{Cov}(E_1, E_2)$$

On appelle **corrélation génétique** (r_g) la corrélation entre les valeurs génétiques du même individu pour deux caractères différents :


$$r_g = \frac{\text{Cov}(G_1, G_2)}{\sigma_{G_1} \sigma_{G_2}}$$

En décomposant encore, on fait apparaître une composante additive et une composante de dominance :

$$\text{Cov}(P_1, P_2) = \text{Cov}(A_1, A_2) + \text{Cov}(D_1, D_2) + \text{Cov}(E_1, E_2)$$

Ainsi, on appelle **corrélation génétique additive** (r_a) la corrélation entre les valeurs génétiques additives du même individu pour deux caractères différents :

$$r_a = \frac{\text{Cov}(A_1, A_2)}{\sigma_{A_1} \sigma_{A_2}}$$

 *En génétique végétale, r_g désigne effectivement la corrélation génétique totale, telle que définie ici. En génétique animale, l'usage a fait que quand on parle de corrélation génétique sans autre précision, implicitement, il s'agit toujours de la corrélation génétique additive. Il convient alors de désigner ce paramètre par r_a et non par r_g , comme on le voit encore souvent, afin d'éviter toute ambiguïté.*

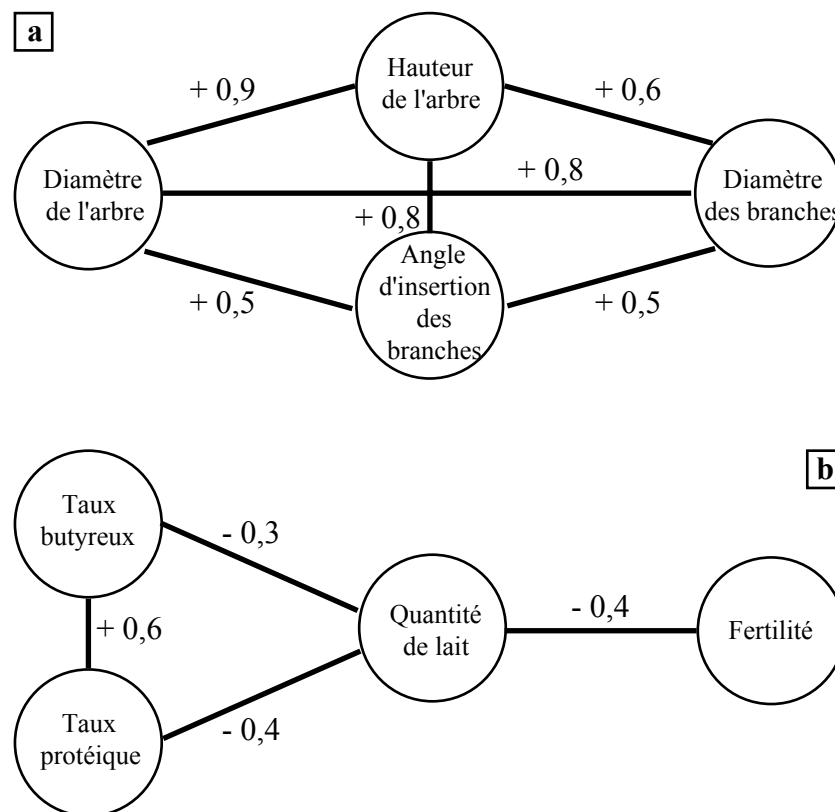
Pour interpréter les notions de corrélation génétique et de corrélation génétique additive, qui sont des concepts purement statistiques, on invoque généralement deux phénomènes génétiques sous-jacents. Premièrement, certains gènes sont pléiotropes, c'est-à-dire qu'ils exercent une influence sur plusieurs caractères à la fois. Deux caractères gouvernés en partie par des locus communs montrent alors une corrélation génétique, dont le sens dépend du sens dans lequel, majoritairement, les gènes jouent simultanément sur les deux caractères. Deuxièmement, l'existence d'un déséquilibre d'association entre certains locus gouvernant le premier caractère et certains locus gouvernant le second peut entraîner une corrélation génétique. En cas de pléiotropie des gènes, la corrélation génétique est permanente et biologiquement difficilement contournable. En cas de déséquilibre d'association des gènes, la corrélation génétique peut n'être que temporaire : en effet, le déséquilibre d'association lui-même peut évoluer dans le temps et, notamment, s'annuler grâce à la recombinaison, après de nombreuses générations de panmixie (voir GP).

La figure 13 donne deux exemples de corrélations génétiques additives liant plusieurs caractères, l'un chez les plantes cultivées et l'autre chez les animaux domestiques. L'existence de corrélations génétiques, totales ou additives, est extrêmement importante à prendre en considération en amélioration génétique. En effet, si deux caractères sont génétiquement liés, toute modification de la moyenne du premier caractère, sous l'effet d'une sélection par exemple, se répercute sur la moyenne du second : si la corrélation est positive, les caractères évoluent dans le même sens, si elle est négative, les caractères évoluent en sens opposé. De même, si l'on souhaite, par sélection, faire évoluer deux caractères dans le même sens, une corrélation génétique positive est un atout alors qu'une corrélation négative représente un frein qui nécessite l'emploi de méthodes de sélection appropriées.

Figure 13. Exemples de corrélations entre caractères quantitatifs.

a : Corrélations génétiques chez le cèdre (Source : Bariteau et Ferrandes, 1992)

b : Corrélations génétiques additives chez la vache laitière (Source : Boichard, 1987)



EXERCICES

Exercice n°1

Reprendre les données numériques obtenues dans le cas du locus de nanisme chez la souris, pour une population supposée en équilibre de Hardy-Weinberg, avec une fréquence (p_2) de l'allèle n de 0,2 ou de 0,5 (cf. tableau 6). En supposant que le reste du génome n'a pas d'effet sur le poids à 6 semaines, calculer pour les deux valeurs de p_2 les variances génétique, génétique additive et de dominance. Calculer l'héritabilité au sens large et l'héritabilité au sens étroit de ce caractère, sachant que la variance environnementale est de 3 g². Quelles conclusions en tirez-vous ?

Exercice n°2

En 1982 en Australie, l'existence d'un locus à effet majeur sur la prolificité a été démontrée dans une souche de brebis Mérinos (gène Booroola, à 2 allèles notés + et F). Un protocole expérimental a permis d'observer les résultats moyens suivants sur les performances de reproduction des brebis Mérinos :

Génotype	Taux d'ovulation	Taille de portée
++	1,40	1,1
$F+$	3,05	2,0
FF	4,70	3,0

- 1) Calculer le taux de mortalité embryonnaire observé pour chaque génotype. Qu'en concluez vous quant à l'effet biologique du gène Booroola ?
- 2) Si l'on suppose l'absence d'effet du reste du génome sur le taux d'ovulation, ainsi que l'absence d'effet de milieu, calculer l'héritabilité au sens étroit de ce caractère dans une population où l'allèle F a une fréquence de 0.25 (la réponse à cette question nécessite un peu d'attention et une seule opération, au demeurant fort simple).

Exercice n°3

Soit un caractère quantitatif gouverné par un locus à deux allèles, avec surdominance :

Génotype	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$
Moyenne phénotypique	11	15	9

Pour quelles fréquences alléliques une population en équilibre de Hardy-Weinberg atteint-elle la plus forte moyenne ? La solution peut être obtenue graphiquement, de façon approximative, ou précisément par calcul algébrique.

