

Génétique des caractères quantitatifs

PIERRE DARLU

1	Valeur phénotypique	4	Estimation de la ressemblance génétique
2	Analyse de la valeur phénotypique	4.1	Covariance génétique
2.1	Facteur "temps" et valeur phénotypique	4.2	Covariance environnementale
2.2	Additivité et valeur phénotypique	4.3	Conclusion
2.2.1	<i>Contribution génétique</i>	5	Héritabilité
2.2.2	<i>Contribution environnementale</i>	5.1	Définitions
2.2.3	<i>Contribution d'interaction</i>	5.2	Estimation de l'héritabilité
3	Analyse de la variabilité phénotypique		Bibliographie

LE SUCCÈS de l'expression génétique quantitative vient essentiellement de la traduction française, en 1974, de l'ouvrage classique de Falconer (1960) : *Introduction to quantitative Genetics*. Ces termes méritent cependant quelques éclaircissements. En effet, la problématique n'est pas de rechercher si la génétique est "quantitative" ni de trancher sur la nature qualitative ou quantitative de l'information génétique. Elle est plutôt de déterminer si des caractères quantitatifs, c'est-à-dire mesurables, peuvent avoir des déterminants génétiques, simples ou complexes. La démarche est donc d'analyser ces caractères afin de remonter aux gènes qui en sont éventuellement responsables. Les méthodes reposent sur l'interprétation des variations de ces caractères en fonction de divers facteurs, génétiques ou non, que l'on tente de contrôler. En ce sens, la génétique quantitative est mieux définie comme étant la génétique des variations quantitatives (Wright, 1952), ou comme génétique biométrique (Mather, 1949 ; Kempthorne, 1960).

L'un des facteurs génétiques importants que l'on peut contrôler est celui de la parenté. Il permet d'établir la relation entre le degré d'identité génétique et les variations du caractère étudié. L'analyse de cette relation est le fondement de la génétique mendélienne, qui est de comparer les caractères observés chez les parents et chez les enfants afin d'en déduire la façon dont ils se transmettent.

Dans ce domaine du quantitatif, les travaux de Galton à la fin du dix-neuvième siècle sont exemplaires, à une époque où les lois de Mendel n'avaient pas encore été "redécouvertes". La figure 11.1 donne un exemple ingénieux de la tentative qu'il fit pour effectuer des "prédictions" en s'appuyant sur la ressemblance familiale. Ces travaux sont d'ailleurs à l'origine du concept de régression en biométrie. En 1918, Fisher proposa un modèle rendant compte de cette ressemblance entre apparentés, en se fondant sur les principes mendéliens de la transmission d'entités "qualitatives" : les gènes. La figure 11.2 donne une brève illustration de son approche.

Dans ce chapitre, on pose d'abord la question de la nature d'un caractère quantitatif. La question suivante porte sur la description et l'interprétation causale des variations d'un tel caractère. Quelques figures proposent des exemples de distributions : stature dans une population (fig. 11.3), pression systolique dans deux groupes distincts appartenant à une même population (fig. 11.4), concentration en magnésium dans les érythrocytes chez des couples de jumeaux mono- et dizygotes (fig. 11.5). Ces exemples montrent que les variations d'un caractère quantitatif peuvent se situer à des niveaux différents (chez un même individu, entre individus, entre populations) et que leurs origines peuvent également être très diverses (génétiques, environnementales, etc.).

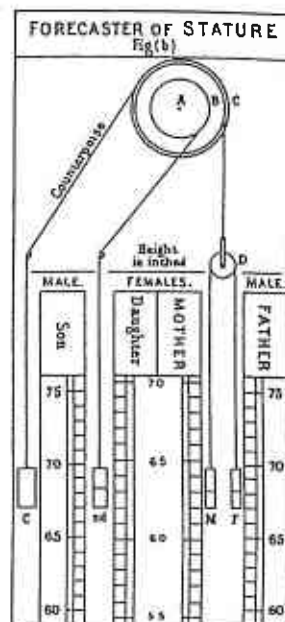
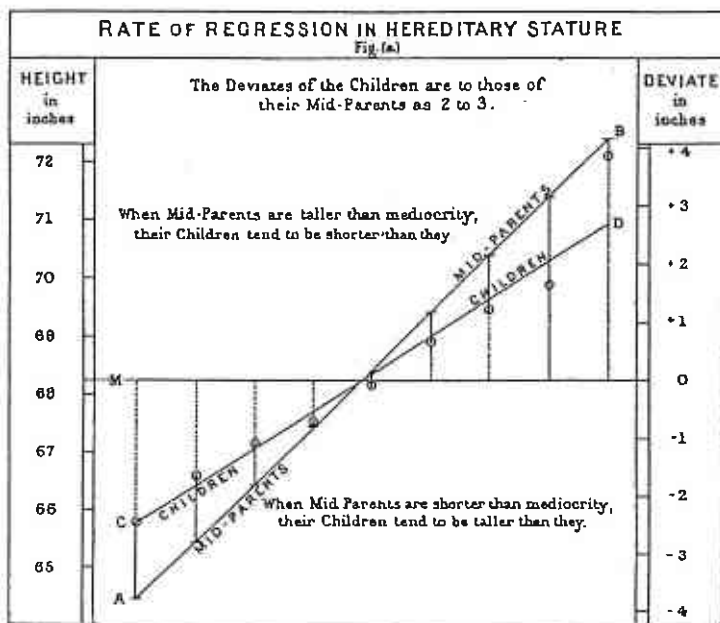


Figure 11.1 : Relation entre la stature des enfants et la stature moyenne des parents établie à partir d'un échantillon de 928 familles (Galton, 1885, in Pearson, 1930).

Dans la figure de gauche (a), Galton montre que lorsque les statures sont exprimées en écarts à la moyenne, la stature des enfants est, environ, les 2/3 de la moyenne des statures des deux parents. L'écart à la moyenne est toujours plus faible chez les enfants que chez les parents, d'où l'expression de *regression* à la moyenne. Galton proposa un système de poulies pour prédire la stature des garçons ou des filles à partir de celle du père et de la mère (b). Dans ce schéma, les diamètres des poulies A et B sont dans un rapport 2/3. Les échelles sont différentes pour les hommes et pour les femmes. Ce travail historique de Galton sur la stature et ses premières observations sur l'hérédité du poids des pois de senteur (1877, in Pearson, 1930) sont à l'origine de l'interprétation biométrique de l'hérédité et du concept statistique de régression.

La science de l'hérédité, qui s'intéresse à la transmission génétique et non génétique des caractères, et la génétique quantitative qui s'intéresse à la transmission génétique des caractères mesurables sont donc les deux concepts étroitement liés qui font l'objet de ce chapitre.

1 Valeur phénotypique

La mesure d'un caractère quelconque à l'instant t chez un individu i constitue ce que l'on appelle sa valeur phénotypique $P_{i,t}$.

Concrètement, le phénotype peut aussi bien être la mesure du poids corporel d'un individu que son taux de cholestérol plasmatique, la concentration en magnésium dans ses globules rouges, l'activité de certaines enzymes dans ses fibres musculaires, la réponse cardiovasculaire à une stimulation particulière, le score attribué lors d'une épreuve comportementale, etc.

Cette valeur phénotypique peut s'expliquer par l'existence d'un certain nombre de variables génétiques, $g_1, g_2, g_3, \dots, g_r$, constituant l'ensemble génétique G_i :

$$G_i = (g_1, g_2, g_3, \dots, g_k, \dots, g_r)$$

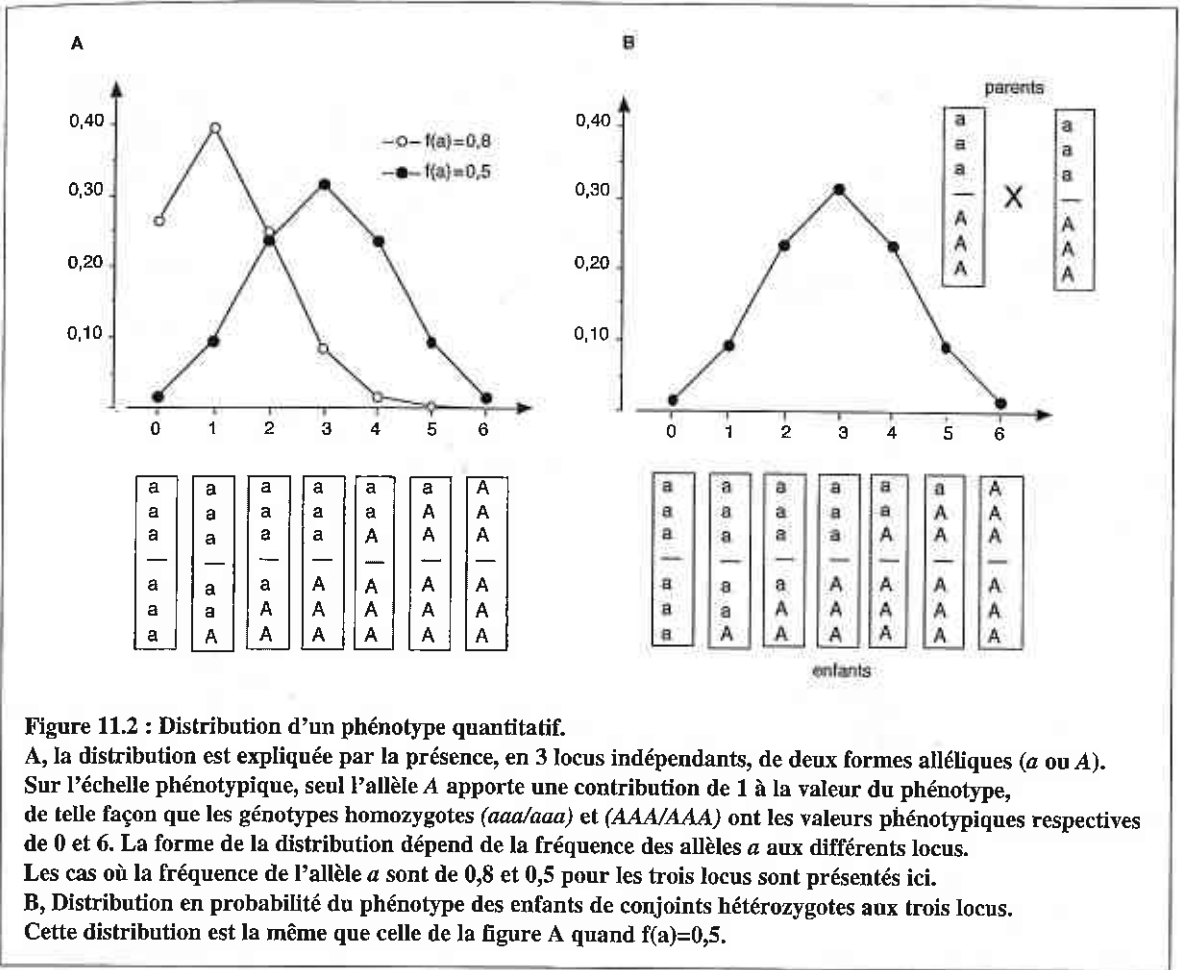


Figure 11.2 : Distribution d'un phénotype quantitatif.

A, la distribution est expliquée par la présence, en 3 locus indépendants, de deux formes alléliques (*a* ou *A*). Sur l'échelle phénotypique, seul l'allèle *A* apporte une contribution de 1 à la valeur du phénotype, de telle façon que les génotypes homozygotes (*aaa/aaa*) et (*AAA/AAA*) ont les valeurs phénotypiques respectives de 0 et 6. La forme de la distribution dépend de la fréquence des allèles *a* aux différents locus. Les cas où la fréquence de l'allèle *a* sont de 0,8 et 0,5 pour les trois locus sont présentés ici.

B, Distribution en probabilité du phénotype des enfants de conjoints hétérozygotes aux trois locus. Cette distribution est la même que celle de la figure A quand $f(a)=0,5$.

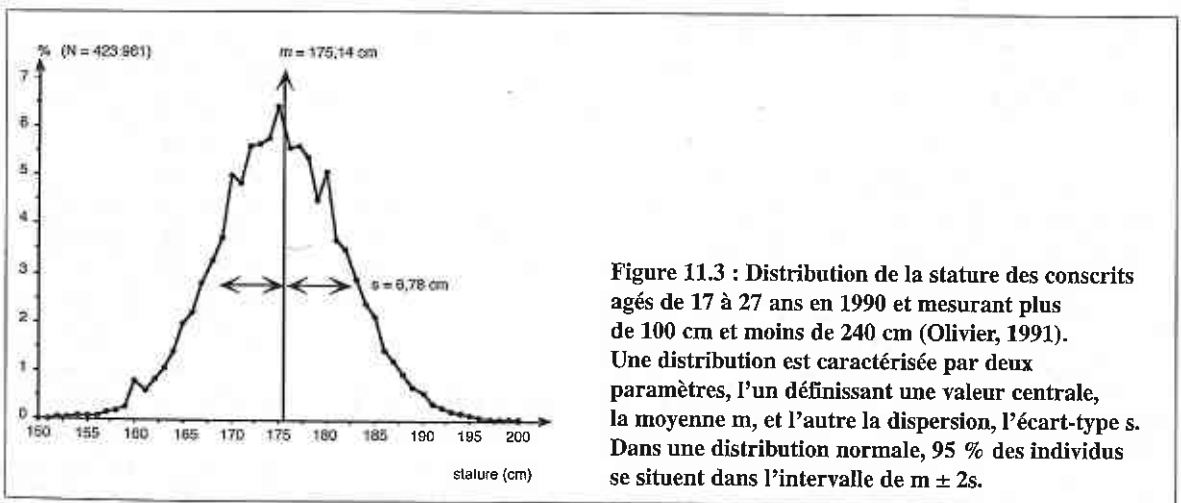


Figure 11.3 : Distribution de la stature des conscrits âgés de 17 à 27 ans en 1990 et mesurant plus de 100 cm et moins de 240 cm (Olivier, 1991). Une distribution est caractérisée par deux paramètres, l'un définissant une valeur centrale, la moyenne m , et l'autre la dispersion, l'écart-type s . Dans une distribution normale, 95 % des individus se situent dans l'intervalle de $m \pm 2s$.

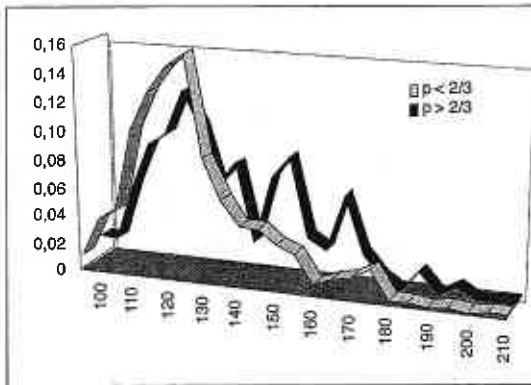


Figure 11.4 : Distribution de la pression artérielle systolique des habitants de la Désirade (Antilles françaises) regroupés selon la proportion p de leurs gènes d'origine africaine ($p < 2/3$ et $p \geq 2/3$) (Darlu *et al.*, 1990).

Des personnes vivant dans un environnement homogène, mais ayant des caractéristiques génétiques différentes, peuvent donc montrer une différence significative ($P \leq 0,05$) au niveau d'une variable quantitative.

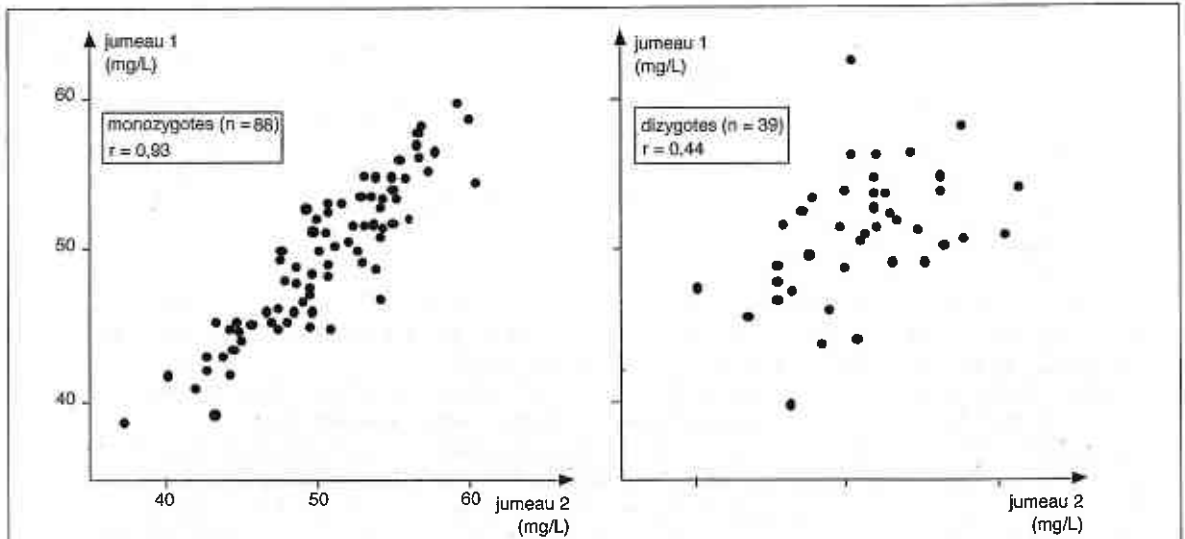


Figure 11.5 : Relation entre la concentration érythrocytaire en magnésium chez des jumeaux mono- (MZ) et dizygotes (DZ).

L'objet de la génétique quantitative est d'interpréter le contraste entre la forte ressemblance observée chez les MZ par rapport à celle des DZ. L'interprétation peut être à la fois génétique (coefficient de parenté des MZ 2 fois plus élevée que celui des DZ) et environnementale (milieu plus homogène chez les MZ que chez les DZ).

Ces variables peuvent traduire la présence ou l'absence de tel ou tel gène ou groupes de gènes, la présence ou l'absence de telle ou telle forme allélique de ces gènes, leurs interactions ainsi que leur degré d'expression ou d'inhibition réciproque.

La valeur phénotypique s'explique également par l'influence que peuvent exercer plusieurs variables de milieu, e_1, e_2, \dots, e_s , qui constituent l'ensemble mésologique (ou environnemental) :

$$E_i = (e_1, e_2, e_3, \dots, e_s)$$

Ces variables représentent, par exemple, des paramètres climatiques (température, pH), la présence d'inhibiteurs, la qualité de certains nutriments ingérés, le niveau socio-économique, l'état sanitaire, etc.

Ces deux ensembles G_i et E_i et leurs éléments respectifs n'ont pas de raison, a priori, d'être indépendants les uns des autres, c'est-à-dire ne pas présenter d'interactions. Ils ne sont pas non plus indépendants du facteur temps. En effet, les gènes ne s'expriment pas tous simultanément et une fois pour toute, mais tout au long d'un processus de dévelop-

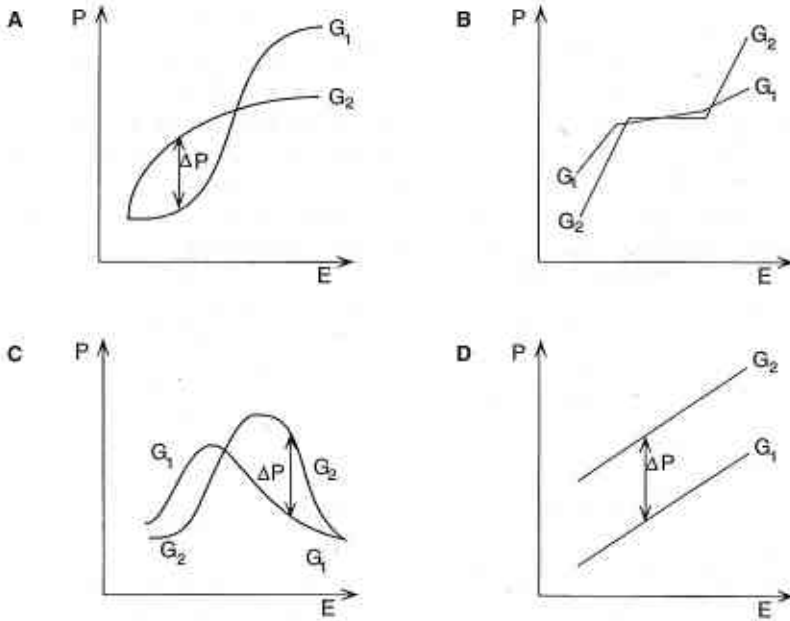


Figure 11.6 : Normes de réaction.

Exemples de variations du phénotype P_i en fonction des modifications de l'ensemble environnemental E_i , pour deux ensembles génétiques distincts G_1 et G_2 . Dans le cas de la figure A, la différence phénotypique ΔP augmente, diminue, s'annule puis s'inverse selon la valeur prise par E .

La figure B illustre le cas où les différences phénotypiques entre les deux ensembles génétiques G_1 et G_2 ne sont perceptibles que dans les environnements extrêmes (petites ou fortes valeurs de E).

Le troisième exemple (C) pourrait être celui de deux formes alléliques (G_1 et G_2) d'une même enzyme dont la variation d'activité (phénotype P) serait différente selon la température (variable E).

La dernière figure correspond à l'hypothèse particulière d'additivité où la variation phénotypique ΔP est indépendante de la valeur prise par l'ensemble environnemental et ne dépend donc que de la différence génétique entre G_1 et G_2 .

(Adapté de Lewontin, 1974.)

pement et de croissance. De leur côté, les différents facteurs définissant le milieu fluctuent également avec le temps et peuvent se modifier progressivement ou brusquement.

La valeur phénotypique $P_{i,t}$ est donc une fonction temporelle des ensembles génotypique et mésologique, que l'on peut écrire :

$$P_{i,t} = f_i(G_i, E_i)$$

Les questions alors pertinentes sont comment préciser le contenu de ces deux ensembles, certainement fort complexes, et quelle formulation donner à la fonction f_i qui puisse être valable pour un ensemble donné d'in-

dividus. Enfin, quelles limitations posent de telles évaluations dans le cas de l'espèce humaine et quelles sont les simplifications que l'on peut admettre sans prendre le risque de fausser les interprétations ?

L'ensemble G_i est généralement ramené au nombre minimal de variables compatible avec les observations, en se fondant sur le principe – omniprésent en génétique – que l'hypothèse interprétative la plus simple est toujours la meilleure. C'est, par exemple, l'hypothèse d'un locus présentant deux allèles sans dominance. Le modèle génétique peut être compliqué progressivement jusqu'à ce qu'il rende compte des observations de manière satisfaisante.

Seuls les modèles relativement simples peuvent être vérifiés et testés dans l'espèce humaine, dans la mesure où des raisons éthiques compréhensibles ne permettent pas d'élaborer des plans expérimentaux complexes comme cela est possible chez l'animal.

Hormis des cas simples, le modèle explicatif le plus commode est souvent le modèle polygénique décrit par Fisher (1918), qui suppose l'action additive de nombreux gènes indépendants (cf. § 2.2.1 & fig. 11.2).

L'ensemble E_i est également délicat à définir et dépend de la nature du caractère abordé. Ainsi, pour un gène dont l'expression est indépendante de la qualité du milieu, un tel ensemble peut être considéré comme vide. En revanche, pour un caractère physiologique comme la pression artérielle, le poids de naissance, la cholestérolémie, ou l'activité de la catéchol-O-méthyl-transférase, l'ensemble E_i recouvre un nombre élevé de variables plus ou moins corrélées : qualité et équilibre de l'alimentation, niveau socio-économique, cadre de vie, médication, etc.

Les difficultés d'observation dans l'espèce humaine contraignent le plus souvent à ne retenir que quelques variables mésologiques, qu'il faut donc choisir aussi pertinentes que possible. Le risque qu'une variable vraiment déterminante échappe à l'investigation n'est pas toujours négligeable.

On insiste également sur le fait que la définition donnée du phénotype est une définition "poly-factorielle", c'est-à-dire due à plusieurs facteurs, sans préciser l'origine de ces facteurs, à ne pas confondre avec une définition qui ne serait que "polygénique", c'est-à-dire due à plusieurs facteurs strictement génétiques. Enfin, l'adéquation à un modèle n'est jamais une preuve absolue de sa réalité biologique, ni qu'il soit le seul possible.

La définition de la fonction f_t est nécessaire pour comprendre la nature causale et le rôle explicatif des deux ensembles G_i et E_i sur la valeur phénotypique. Pour évaluer cette fonction, la méthode consiste à observer les modifications de $P_{i,t}$ lorsque l'on fait varier l'ensemble G_i tout en maintenant constant l'ensemble E_i et, réciproquement, en faisant varier l'ensemble E_i en fixant l'ensemble G_i . Cela permet ainsi de définir les modes de variation respectifs de $P_{i,t}$, G_i et E_i et d'établir ce que l'on appelle des normes de réactions. La figure 11.6 donne quelques

exemples de ces normes de réactions, adaptés de Lewontin (1974). La définition de f_t impose donc l'étude des variations des éléments qui la composent.

Dans le cas où l'on s'intéresse non à un individu et à sa valeur $P_{i,t}$, mais à un ensemble d'individus d'une population donnée, la méthode consiste à comparer des populations aux structures génétiques différentes et vivant dans le même environnement d'une part et à comparer des populations aux structures génétiques identiques vivant dans des milieux contrastés d'autre part. On retrouve ainsi l'une des méthodes les plus couramment utilisées en anthropologie biologique (Harrison & Boyce, 1972).

2 Analyse de la valeur phénotypique

Si l'on s'en tient à la définition du Larousse, l'hérédité représente "les mécanismes de transmission aux enfants des caractères des parents". La science de l'hérédité s'intéresse donc à la relation (a) entre $P_{i,t}$ et $P_{j,t}$ (fig. 11.7), où i et j sont des apparentés. En ce sens, l'hérédité ne peut se réduire à la seule étude des mécanismes de transmission de l'ensemble G_i du sujet i au sujet j (relation b). Elle englobe également tous les modes de transmission non génétiques de l'en-

$$P_{i,t} = f_t(G_i, E_i)$$

$$\begin{array}{ccc} \text{(a)} \downarrow & \text{(b)} \downarrow & \text{(c)} \downarrow \\ P_{j,t} = f_t(G_j, E_j) & & \end{array}$$

Figure 11.7 : Les ensembles génétiques G et non génétiques E sont les éléments d'une fonction dépendante du temps t et définissent le phénotype P d'un individu i .

Les transmissions des ensembles G et E d'une personne à une autre, par exemple d'un parent à son enfant, obéissent à des lois différentes (flèches b et c). La difficulté est de distinguer entre ces deux modes de transmission dans les familles nucléaires. L'un des objets de la génétique quantitative est d'inférer, à partir des différences phénotypiques entre $P_{i,t}$ et $P_{j,t}$, (flèche a) les modalités de ces transmissions.

semble E_i (relation c) qui peut se réaliser selon des modalités sociales, culturelles, économiques qui n'obéissent évidemment pas aux lois de Mendel (Cavalli-Sforza & Feldman, 1981)

Les problèmes sont alors les suivants :

– comment distinguer les différents types d'hérédité, c'est-à-dire les différents modes de transmission des caractères, à partir de la seule observation des ressemblances phénotypiques entre $P_{i,t}$ et $P_{j,t}$?

– cette distinction est-elle aussi efficace quel que soit le caractère P envisagé ? peut-elle se faire sans ambiguïté et dans quelles limites ?

En règle générale, il serait plus correct de ne pas parler de "génétique" mais d'"hérédité" des caractères quantitatifs, car il serait illusoire, sauf cas très particuliers ne posant pas de problème, de traiter du mode de transmission génétique sans prendre en compte les autres modes de transmission pouvant interférer avec lui. En effet, cela reviendrait à faire abstraction de la fonction f_t reliant les deux ensembles G_i et E_i . La connaissance, même imparfaite ou approchée, de cette fonction reste nécessaire pour tenter de comprendre la nature causale et le rôle explicatif que jouent les différentes variables g et e sur la valeur phénotypique observée.

Étant donné la complexité aussi bien des ensembles G_i et E_i que de la fonction f_t qui les relie, l'ambition de vouloir analyser dans le détail une telle complexité doit être limitée, dans la mesure où le nombre d'inconnues intervenant dans le modèle général est certainement bien supérieur au nombre d'équations dont pourrait disposer l'expérimentateur, et à plus forte raison l'observateur dans le cas de l'espèce humaine.

Des hypothèses simplificatrices sont donc nécessaires, même si, dans une première étape, elles doivent réduire le niveau de compréhension du phénotype. Ces simplifications concernent essentiellement deux aspects, qui sont examinés successivement :

– l'indépendance de la fonction f et du temps (hypothèse H_1) ;

– l'additivité des deux ensembles G_i et E_i au niveau de leurs effets sur la valeur phénotypique (hypothèse H_2).

2.1 Facteur "temps" et valeur phénotypique

Les variations ou modifications du phénotype avec le temps sont une évidence. Elles résultent à la fois du

fait qu'un phénotype s'exprime dans un contexte temporel précis : heure, jour, saison, etc., mais aussi que cette expression s'inscrit avec une certaine durée dans un processus ontogénétique. Non seulement les gènes participant au phénotype peuvent différer d'un stade de développement à un autre (Waddington, 1957), mais encore les effets du milieu se modifient ou s'accumulent avec le temps.

La méthode à suivre pour s'affranchir du facteur temps consiste à étudier soigneusement en préalable les variations du caractère avec le temps, ou à les prendre en compte systématiquement. Pour cela, il est nécessaire de constituer un groupe d'individus que l'on aura synchronisé avec une précision d'autant plus grande que le caractère étudié présente de grandes fluctuations avec le temps. Cela revient à distinguer avec soin les sujets selon leur stade de développement (les grouper en classes d'âge, de maturité osseuse ou sexuelle, en générations, etc.) et à réaliser les observations dans les mêmes conditions temporelles (mêmes saisons, heures de la journée, etc.). Quelles que soient les précautions prises, l'analyse d'un phénotype n'a de sens que dans certaines limites temporelles. En effet, l'expression phénotypique d'un certain ensemble génétique peut fluctuer au cours du temps en fonction de facteurs non contrôlés et aléatoires, au point qu'à cet ensemble peuvent correspondre des valeurs phénotypiques différentes. Inversement, des ensembles génétiques différents peuvent, en raison d'aléas de développement, conduire à des convergences phénotypiques. Ces notions sont bien illustrées à partir de ce que Waddington appelait le "paysage épigénétique" (fig. 11.8).

Si l'on accepte les limites imposées par une telle hypothèse H_1 , on peut écrire, en supprimant l'indice t de temps des formules :

$$H_1 \Rightarrow f_t(G_i, E_i) = f(G_i, E_i)$$

L'importance du facteur temps est également illustrée à propos de la discussion sur l'héritabilité (cf. § 5).

2.2 Additivité et valeur phénotypique

Une deuxième hypothèse simplificatrice, celle d'additivité (H_2), revient à définir la fonction f comme la somme de trois fonctions distinctes :

– l'une, G_i , dépendant uniquement des variables génétiques g_k et qui représente donc la fraction du phénotype due aux seuls effets des gènes :

$$G_i = f_1(g_1, g_2, g_3, \dots, g_k, \dots, g_r) ;$$

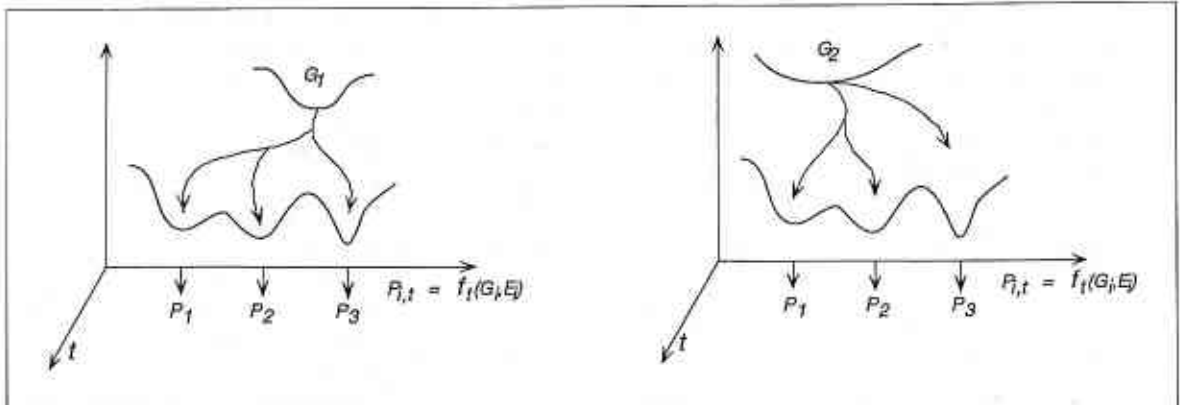


Figure 11.8 : Paysages épigénétiques.

Deux sujets aux ensembles génétiques distincts (G_1 et G_2) se “déplacent” au cours du temps t sur une surface où des alternatives successives conduisent à l’acquisition d’un phénotype particulier parmi plusieurs phénotypes possibles (P_1 , P_2 ou P_3). Au cours du développement, représenté par l’axe vertical, un même génotype peut donc conduire à des phénotypes distincts et deux génotypes différents peuvent aboutir à un même phénotype.

(D’après Waddington, 1957.)

– la deuxième, E_i , est une fonction des seules variables d’environnement et représente la fraction du phénotype due aux seuls effets du milieu ;

$$E_i = f_2(e_1, e_2, e_3, \dots, e_i, \dots, e_r) ;$$

– la troisième enfin $R = GE_i$ représente l’interaction entre ces deux premières fonctions :

$$GE_i = f_3(G_i, E_i)$$

On peut donc écrire :

$$H_i \Rightarrow P_i = G_i + E_i + GE_i$$

Une telle définition a pour conséquence de donner une expression de la fonction génétique G_i et de la fonction d’environnement E_i dans les mêmes unités que la variable phénotypique, et cela bien que les variables génétiques g_k ou mésologiques e_i n’aient pas a priori de raisons d’être considérées sur une même échelle d’observation. Pour prendre l’exemple simple de la stature, dont on sait que le déterminisme est à la fois génétique et environnemental, on ne peut évidemment pas dire que les 110 premiers centimètres de la stature d’un homme de 170 cm sont dus aux gènes et les 60 restants sont dus au milieu.

En fait, cette décomposition de P_i n’apporte aucune information sur les causes diverses qui produisent P_i . Elle ne prend tout son sens que par l’étude des variations de P_i , ce qui soulève alors les difficultés propres à l’analyse des variations phénotypiques (cf. § 3).

2.2.1 CONTRIBUTION GÉNÉTIQUE

La contribution globale du génotype G_i à la valeur du phénotype P_i du sujet i peut, à son tour, être décomposée en diverses contributions que l’on considère comme additives et exprimées dans les mêmes unités, soulevant ainsi les mêmes problèmes que précédemment. Ces contributions sont les suivantes, si le caractère étudié dépend de n locus bialléliques indépendants.

L’une résulte de la somme des effets α_{ik} de chacun des $2n$ allèles impliqués dans l’expression du génotype. C’est l’effet additif :

$$A_i = \sum_{k=1}^{2n} \alpha_{ik}$$

Cela signifie que chacun des $2n$ allèles apporte sa contribution au génotype et que ces contributions s’ajoutent.

Une deuxième traduit l’existence de phénomènes de dominance, quantifiée δ_{ij} , entre les deux allèles situés à un même locus j : c’est l’effet de dominance D_i qui signifie que deux allèles ne produisent pas, lorsqu’ils sont ensemble au même locus, un effet égal à la somme des effets de chacun d’eux. S’il y a n locus, l’effet de dominance global s’écrit :

$$D_i = \sum_{i=1}^n \delta_{ij}$$

Une troisième exprime les effets d'interaction I_{ij} entre les allèles situés cette fois à différents locus : c'est l'effet d'épistasie I_i . Il signifie qu'une combinaison d'allèles situés à des locus différents est porteuse d'une information différente de celle donnée par la somme des informations qu'apporte chacun d'eux séparément :

$$I_i = \sum_{n=1}^{2n(n-1)} I_{im}$$

Cette contribution I_i représente, pour résumer, tous les effets non décrits précédemment, par exemple l'interaction entre les effets de dominance de deux locus différents (effet DD) ou celle entre la dominance à un locus et l'additivité à un autre locus (effet AD), entre les additivités à un locus et l'additivité à un autre locus (effet AA). Dans ces conditions, la fonction $G_i = f_i(g_1, g_2, g_3, \dots, g_r)$ s'écrit donc :

$$G_i = A_i + D_i + I_i$$

Par ailleurs, l'hypothèse H_3 que les effets I_i sont négligeables par rapport aux effets additifs et de dominance est souvent formulée, d'où :

$$H_3 \Rightarrow G_i = A_i + D_i$$

Cette hypothèse H_3 est cohérente avec la façon dont sont estimés les divers effets génétiques. L'estimation de ces effets s'effectue en effet en minimisant, par des méthodes de moindres carrés par exemple, les effets de dominance par rapport aux effets additifs, les effets d'interactions par rapport aux effets additifs et de dominance. C'est là une position a priori qu'il ne faut pas perdre de vue et qui correspond au modèle proposé par Fisher (1918, 1941).

2.2.2 CONTRIBUTION ENVIRONNEMENTALE

La contribution globale de l'environnement E_i à la valeur du phénotype P_i n'est que rarement décomposée, mais le plus souvent considérée en bloc. Cette position est révélatrice de l'"asymétrie" (Lewontin, 1974) avec laquelle sont considérés les différents facteurs affectant la valeur phénotypique d'un caractère et sa variabilité. Certains auteurs tentent cependant – à la faveur d'un dispositif expérimental approprié ou d'observations adéquates dont quelques exemples sont donnés – d'éclater cette composante E_i , en recherchant, par exemple, les effets dus au milieu intra-utérin, à la santé et à l'état nutritionnel

de la mère, à la taille de la famille, à l'ordre de naissance, au niveau socio-économique, etc. Quoiqu'il en soit, les problèmes soulevés précédemment à propos de la décomposition de E_i se retrouvent ici, en ce sens que la plupart des variables considérées sont interactives (non additives), d'une part, et que l'on ne mesure que les effets de leurs variations, d'autre part.

2.2.3 CONTRIBUTION D'INTERACTION

La contribution d'interaction est le plus souvent considérée comme nulle : $GE_i = 0$. Cette hypothèse présente l'avantage de simplifier grandement le modèle proposé, car cela revient à supposer une additivité simple entre effets des gènes et effets du milieu. Si cette position peut sembler confortable pour un esprit cartésien, elle n'est cependant qu'une situation très particulière : elle traduit qu'une différence génétique reste constante quelles que soient les variations du milieu ou, inversement, qu'une variation d'environnement a les mêmes effets quelle que soit la valeur du génotype. C'est le cas de la figure 11.6D des normes de réaction. Les problèmes relatifs à l'interaction ont été abordés par Haldane (1946) et remis en mémoire par Lewontin en 1975 à propos des problèmes relatifs à l'hérédité de l'intelligence.

3 Analyse de la variabilité phénotypique

La définition de la fonction $f(G_i, E_i)$ n'est possible qu'à la condition d'effectuer une analyse des modifications du phénotype lorsque varient les ensembles génétiques G_i et environnementaux E_i .

Il est possible, en utilisant les développements de Taylor et de Mac Laurin d'analyser les variations ΔP de P_i en fonction de celles des g_k et e_l (Lewontin, 1975 ; Jacquard, 1978) en écrivant :

$$\begin{aligned} \Delta P = & a \frac{\partial f(G,P)}{\partial G} \Delta G + b \frac{\partial f(G,P)}{\partial E} \Delta E \\ & + c \frac{\partial^2 f(G,P)}{\partial G^2} \Delta G^2 + d \frac{\partial^2 f(G,P)}{\partial E^2} \Delta E^2 \\ & + e \frac{\partial^2 f(G,P)}{\partial G \partial E} \Delta G \Delta E + \dots \end{aligned}$$

où a, b, c, d, e sont des constantes numériques.

Pour se ramener à un modèle additif, il convient de négliger les variations de degré 2 ou supérieur à 2 [les $\Delta(G)^2$ et $\Delta(E)^2$, etc.] ; il convient également que les dérivés partielles puissent être considérées comme constantes lorsque G et E varient. Cela implique donc que la décomposition des variations du phénotype en variations additives des effets des gènes et de l'environnement n'est valable que localement, c'est-à-dire lorsque les groupes étudiés sont suffisamment homogènes d'un point de vue génétique [$\Delta(G)$ faible] et vivent dans un milieu peu différencié [$\Delta(E)$ faible].

L'approche statistique des variations de P_i par rapport à la moyenne de la population conduit aux mêmes conclusions. Si l'on représente par P_{ige} la valeur phénotypique du sujet i de génotype g et de milieu e, alors on peut écrire :

$$\begin{aligned} (P_{ige} - P_{...}) &= (P_{.g} - P_{...}) \\ &+ (P_{.e} - P_{...}) \\ &+ [(P_{ge} - P_{...}) - (P_{.g} - P_{...}) + (P_{.e} - P_{...})] \\ &+ (P_{ge} - P_{...}) \\ \Delta P &= \Delta G + \Delta E + \Delta R + \varepsilon \end{aligned}$$

où $P_{...}$ est la moyenne de la population et ΔP l'écart entre la valeur phénotypique de l'individu i et la moyenne de la population $P_{...}$. Cet écart se décompose en :

- ΔG , écart entre la valeur moyenne des individus ayant le génotypique g ($P_{.g}$) et la moyenne de la population ;
- ΔE , écart entre la valeur moyenne des individus ayant le même environnement e ($P_{.e}$) et la moyenne de la population ;
- ΔR , écart dû à l'effet d'interaction, celui qui ne s'explique pas par la somme des valeurs génotypiques g et environnementales e ;
- ε , déviations de tous ordres, spécifiques de l'individu i et dues aux erreurs de mesures.

L'application de ce modèle d'analyse du phénotype utilisé dans l'étude des variations du phénotype, implique que, comme le rappelle Lewontin (1974) :

- les effets génétiques et environnementaux sont exprimés dans les mêmes unités que le phénotype (problème d'unité, cf. § 2.2) ;
- l'on ne peut préciser si, par exemple, une petite variation ΔE provient de ce que l'environnement est homogène ou de ce que le génotype en question est insensible aux variations de l'environnement (problème de causalité) ;

- la décomposition fait intervenir à la fois la moyenne générale de la population et les valeurs moyennes des individus de même génotype, d'une part, et de même environnement, d'autre part. S'il advient des modifications quelconques de ces moyennes, par exemple si les structures génétiques et environnementales de la population se modifient, les résultats donnés par la décomposition sont également modifiés. En d'autres termes, l'analyse des variations du phénotype n'est valable que dans des limites spatio-temporelles bien définies. Ce point est repris à propos de la discussion sur le concept d'héritabilité (cf. § 5).

Sans perdre de vue ces limitations inhérentes à toute analyse de variances, on peut ensuite écrire, si les écarts de $P_{.g}$, $P_{.e}$ et P_{ge} à leurs moyennes se distribuent normalement avec des variances respectives, V_g , V_e , V_r et V_ε :

$$\begin{aligned} V_P &= V_g + V_e + V_r + V_\varepsilon \\ &+ 2 \text{Cov}(G,E) + 2 \text{Cov}\{(G+E),R\} \end{aligned}$$

De nouvelles simplifications sont souvent introduites à ce niveau. La première (H_4) revient à supposer que les effets d'interaction, R, sont indépendants des effets génétiques et environnementaux. Dans ce cas, le dernier terme de l'équation précédente est nul. La deuxième (H_5) suppose l'absence de relation entre les effets du génotype et ceux de l'environnement, au niveau du groupe étudié. Lorsque ces effets sont indépendants, on a alors :

$$H_5 \Rightarrow \text{Cov}(G,E) = 0$$

Donc, dans chaque environnement, tous les génotypes sont équiprobables et, réciproquement, un génotype donné se retrouve de manière équiprobable dans tous les milieux possibles. En d'autres termes, l'appartenance d'un individu à un certain milieu n'apporte aucune information sur la valeur de son génotype et inversement.

La validité d'une telle hypothèse est, évidemment, fortement contestable dans le cas de l'espèce humaine. En effet, s'il existe au sein du groupe que l'on étudie une quelconque stratification socio-économique, professionnelle ou géographique par exemple, il se peut qu'il existe également de façon parallèle, une stratification génétique. Certains auteurs l'ont montré, par exemple, à propos de la stratification professionnelle de la stature, de la stra-

tification sociale et géographique de la pigmentation au Japon, de la stratification socio-économique de taux de créatinine au Guatemala, de la relation entre ordre de naissance et poids à la naissance, etc.

Dans ces conditions, et si les caractères présentent un certain déterminisme génétique, alors l'hypothèse $Cov(G, E) = 0$ peut ne plus être vérifiée.

Comme on l'a mentionné, la valeur génotypique a été décomposée en effets additifs, de dominance ou d'épistasie. De la même façon, la variance génotypique V_g peut se décomposer en variance additive, de dominance et d'épistasie :

$$V_g = V_a + V_d + V_i$$

avec $V_i = V_{aa} + V_{ad} + V_{dd} + \dots$

4 Estimation de la ressemblance génétique

La ressemblance génétique entre deux individus doit être évaluée à partir de la ressemblance phénotypique, la seule qui soit observable directement.

On suppose un couple de deux individus i et j dont les valeurs phénotypiques s'écrivent, après avoir retenu l'hypothèse d'additivité sans interaction des effets des gènes et du milieu :

$$P_i = G_i + E_i$$

$$P_j = G_j + E_j$$

L'intensité de la ressemblance entre un ensemble de couples (i, j) peut être évaluée, pour un caractère donné, par la covariance qui s'écrit :

$$Cov(P_i, P_j) = Cov(G_i, G_j) + Cov(E_i, E_j)$$

avec l'hypothèse H_5 d'indépendance des effets G et E : $Cov(G, E) = 0$.

Le problème est donc de différencier ces deux composantes de la ressemblance, afin de préciser dans quelle mesure on peut estimer la covariance génétique à partir de la ressemblance phénotypique. Pour cela, il faut savoir au préalable estimer la covariance génétique, d'une part, et la covariance environnementale, d'autre part.

4.1 Covariance génétique

Lorsque deux individus sont apparentés, la covariance génétique $Cov(G_i, G_j)$ est fonction :

– des variances additives V_a , de dominance V_d , d'épistasie V_i ;

– du coefficient de parenté ϕ entre les deux sujets i et j (cf. chap. 10). Ce coefficient de parenté ϕ est la probabilité pour qu'un allèle pris au hasard chez i soit identique, par ascendance, à un allèle pris au hasard, au même locus, chez j (Malécot, 1948) ;

– de la probabilité δ d'identité, pour un locus donné, des deux allèles de i avec les deux allèles de j .

On peut écrire en effet (Fisher, 1918 ; Kempthorne, 1954, 1955) :

$$Cov(G_i, G_j) = 2\phi V_a + \delta V_d + 2\phi\delta V_{ad} + 4\phi^2 V_{aa} + \delta^2 V_{dd} + \dots$$

Par exemple :

– pour deux germains (frère-frère, frère-sœur ou sœur-sœur) ou deux jumeaux dizygotes : $\phi = 1/4$ (cf. chap. 10) et $\delta = 1/4$, d'où

$$Cov(G_i, G_j) = (1/2) V_a + (1/4) V_d + (1/4) V_{aa} + (1/8) V_{ad} + (1/16) V_{dd} + \dots$$

– pour deux jumeaux monozygotes, $\phi = 1/2$ et $\delta = 1$:

$$Cov(G_i, G_j) = V_a + V_d + V_{aa} + V_{ad} + V_{dd} + \dots$$

– pour un parent et son enfant, $\phi = 1/4$ et $\delta = 0$:

$$Cov(G_i, G_j) = 1/2 V_a + 1/4 V_{aa} + \dots$$

Cependant ces formules simplifiées ne sont valables que si l'on tient compte de l'ensemble des hypothèses suivantes :

– les différents locus qui contribuent à la valeur du caractère sont considérés comme indépendants les uns des autres et la population est supposée en équilibre de liaison (cf. chap. 10) ;

– les deux sujets i et j sont non consanguins. Cette dernière hypothèse limitative peut être levée si l'on retient une formulation plus complexe de la covariance génétique proposée par Kempthorne (1955), Gillois (1969) et Jacquard (1970).

Les conditions pour calculer cette covariance génétique sont la connaissance préalable des coefficients d'identité et de la structure génique de la population, ainsi que l'hypothèse de locus indépendants dans une population en équilibre de liaison.

4.2 Covariance environnementale

En l'absence de toute indication permettant d'estimer $\text{Cov}(E_i, E_j)$, l'attitude la plus fréquente consiste à rechercher une situation expérimentale ou certaines conditions d'observation qui permettent d'annuler une telle covariance. C'est le cas lorsque les caractéristiques du milieu que l'on suppose pouvoir influencer le phénotype sont "randomisées", c'est-à-dire tirées au sort entre chaque membre des couples d'individus apparentés que l'on étudie. Alors, $\text{Cov}(E_i, E_j) = 0$.

Une telle situation est bien concevable en expérimentation animale mais éthiquement bien difficile à réaliser dans le cas de l'espèce humaine. Elle peut s'observer cependant lorsque des frères, des demi-frères ou des jumeaux sont séparés, puis placés de manière non sélective dans des milieux hétérogènes. Cette séparation doit être aussi précoce que possible, sans que l'on puisse espérer toutefois supprimer totalement cette covariance environnementale. En effet, une partie de celle-ci est due à l'identité du milieu intra-utérin, de l'environnement maternel prénatal, de l'environnement familial précédant la séparation, etc. ; elle est donc pratiquement irréductible. Une autre possibilité consiste à étudier la covariance entre demi-germains de même père. Si les différentes mères du même père ne sont pas choisies sur des critères environnementaux communs et s'il n'existe pas de relations entre l'environnement du père et ceux des mères, alors on peut conclure que la corrélation environnementale entre de tels demi-germains est nulle : $\text{Cov}(E_i, E_j) = 0$. Une telle covariance n'a pas de raison d'être nulle dans le cas de demi-germains de même mère, puisque ils partagent, au minimum, un environnement intra-utérin identique.

Une autre méthode pour prendre la mesure des effets de covariabilité dus à des facteurs d'environnement revient à étudier la covariance phénotypique chez des jumeaux vivant ensemble, d'une part, et chez des jumeaux vivant séparés, d'autre part (mono- et/ou dizygotes). Si la variabilité génétique n'est pas différente entre ces deux groupes comparés, alors les modifications de la ressemblance phénotypique entre jumeaux d'un même couple, selon qu'ils vivent ensembles ou séparés et selon la durée de leur séparation éventuelle, donnent alors une estimation de l'impact des facteurs de milieu sur cette ressemblance.

4.3 Conclusion

L'estimation de la covariance génétique à partir de la covariance phénotypique n'est donc possible qu'à la condition de pouvoir prendre en compte la covariance environnementale. Cette condition est impérative si l'on se donne pour projet de distinguer entre hérédité génétique et hérédité culturelle. Lorsque cette condition est remplie, l'estimation des covariances entre apparentés de différents degrés permet donc une estimation de la variance génétique, additive et de dominance. Cette méthode est le fondement de toutes les analyses génétiques, aussi bien des analyses de ségrégation, selon les modèles mixtes ou régressifs, que des analyses de liaison génétique (*linkage*), sur des traits qualitatifs ou quantitatifs.

5 Héritabilité

5.1 Définitions

Si l'on peut accepter les limites imposées par le modèle additif et si l'on distingue entre analyse des causes et analyse des variations, l'héritabilité, qui est un prolongement de l'analyse de variance, peut être définie de deux façons distinctes :

– l'héritabilité au sens large H représente le rapport entre la variance génotypique V_g et la variance phénotypique V_p ;

– l'héritabilité au sens étroit h^2 représente le rapport entre la variance additive V_a et la variance phénotypique V_p .

$$H = \frac{V_g}{V_p} ; \quad h^2 = \frac{V_a}{V_p}$$

De telles définitions appellent un certain nombre de remarques, qui permettent de préciser à la fois les limites et les implications d'un tel concept.

L'héritabilité est donc un rapport de variance qui varie entre 0 et 1, selon que la part du génotype dans le cas de H , ou seulement sa fraction additive dans le cas de h^2 , est faible ou forte dans la variabilité phénotypique du caractère. Ces deux définitions – héritabilité au sens large H , héritabilité au sens strict ou étroit h^2 – se rejoignent quand les effets de dominance et d'interaction sont négligeables devant l'importance supposée des effets additifs.

La comparaison de ces deux héritabilités est d'ailleurs un moyen d'estimer l'ampleur de ces effets

non additifs, à la condition toutefois que le caractère étudié ne soit soumis à aucune pression sélective.

Ces deux définitions de l'héritabilité (H et h^2) reposent sur l'analyse de la diversité d'un groupe humain donné. La valeur de l'héritabilité trouvée est donc étroitement dépendante des facteurs influant sur la composition de ce groupe ; elle est donc dépendante

à la fois de sa structure génétique (la variance additive par exemple est fonction de la fréquence des gènes dans la population ; cf. encadré 11.1), de sa structure démographique et sociale ainsi que de l'influence que peut jouer l'environnement sur ces diverses structures. Il s'ensuit qu'une héritabilité n'a de sens qu'après avoir précisé ces structures.

ENCADRÉ 11.1

Valeurs génotypiques et leurs variances

La moyenne et la variance d'un caractère quantitatif dans une population peuvent s'exprimer à partir d'un modèle génétique simple faisant intervenir un locus et deux allèles $A1$ et $A2$, de fréquences p et q .

Génotypes	Fréquences	Valeurs génotypiques	Décomposition en effets additifs et d'interaction
$A1A1$	p^2	i	$2 a_1 + (aa)_{11}$
$A1A2$	$2 pq$	j	$a_1 + a_2 + (aa)_{12}$
$A2A2$	q^2	k	$2 a_2 + (aa)_{22}$

Les valeurs génotypiques associées aux trois génotypes possibles sont i , j et k . Ce sont les valeurs observées. Elles sont estimées par un modèle régressif où chacun des allèles apporte sa contribution (a_1 et a_2) :

$$i = 2 a_1 + (aa)_{11} ; j = a_1 + a_2 + (aa)_{12} ; k = 2 a_2 + (aa)_{22}$$

L'écart à l'additivité est pris en compte par l'une des trois interactions possibles : $(aa)_{11}$, $(aa)_{12}$ et $(aa)_{22}$. Dans ce modèle statistique, l'interaction n'est autre que la dominance du modèle génétique.

Pour estimer les effets a_1 et a_2 propres à chacun des allèles, et en supposant les données centrées, c'est-à-dire que la moyenne soit égale à 0 :

$$m = p^2 i + 2 pqj + q^2 k = 0$$

il suffit de minimiser les effets d'interactions. Pour cela, on calcule la somme f des carrés des écarts entre la valeur génotypique observée et les effets additifs,

$$f = p^2(i - 2 a_1)^2 + 2 pq(j - a_1 - a_2)^2 + q^2(k - 2 a_2)^2$$

puis on recherche les valeurs de a_1 et a_2 minimisant cette fonction, en annulant les dérivées partielles de f en fonction de a_1 et a_2 . On obtient :

$$a_1 = pi + qj ; a_2 = pj + qk$$

La variance additive est :

$$V_a = p^2(2 a_1)^2 + 2 pq(a_1 + a_2)^2 + q^2(2 a_2)^2$$

$$V_a = 2 pq [p(i - j) + q(j - k)]$$

et la variance de dominance, égale à f , est :

$$V_d = p^2q^2 (i - 2j + k)^2$$

Deux points importants sont à souligner :

- le modèle statistique revient à minimiser la variance de dominance ;
- la décomposition de la variance génotypique en variance additive et de dominance est dépendante des fréquences des allèles $A1$ et $A2$ dans la population comme l'illustre la figure 11.9. Cette décomposition n'est donc pas a priori identique dans deux populations qui auraient des fréquences alléliques différentes.

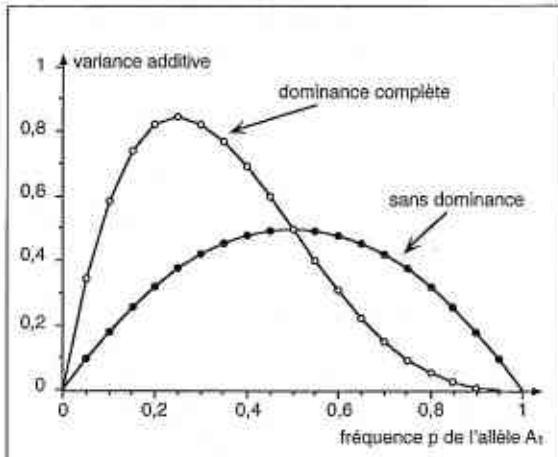


Figure 11.9 : Variations de la variance additive calculée en un locus à deux allèles A_1 et A_2 , en fonction de la fréquence de l'allèle A_1 dans la population et du degré de dominance.

Dans la mesure où, de ce point de vue, deux populations ne peuvent être identiques, la valeur d'une hérabilité n'a donc de sens que dans la population où elle a été étudiée, et au moment où elle a été étudiée, sans qu'aucune généralisation soit possible, sinon avec quelques risques d'erreur. Ce point rejoint les conclusions à propos des limites spatio-temporelles de l'analyse de variance.

Il est nécessaire de remarquer également que l'hérabilité n'apporte aucune information sur le mécanisme de transmission génétique d'un caractère. Celui-ci peut en effet être transmis de façon strictement génétique, en observant cependant une hérabilité nulle : il suffit pour cela que la variabilité génétique dans la population soit nulle ($V_g = 0$), ou que les effets de l'environnement sur ce caractère soient très diversifiés (milieu très hétérogène, V_e très élevée).

Les deux coefficients d'hérabilité, au sens large et étroit, n'ont pas la même signification. En effet, du point de vue du généticien effectuant des sélections expérimentales, seule l'hérabilité au sens étroit présente un intérêt, car les effets de dominance, pris en compte dans l'hérabilité au sens large, ne sont pas transmissibles aux générations suivantes. Les allèles intervenant dans l'expression du caractère ségrègent indépendamment : aussi les quantités atta-

chées à leurs arrangements ne peuvent pas être transmises à la descendance. Quant aux effets d'interaction entre deux gènes, ils ne sont observables que tant qu'il existe un déséquilibre de liaison.

L'hérabilité au sens étroit d'un caractère (h^2) est d'autant plus faible que ce caractère est davantage impliqué dans le maintien de processus vitaux pour l'organisme : cela ne signifie pas l'absence de déterminisme génétique pour ce caractère, mais seulement une homogénéité plus grande des différents génotypes individuels. Cette homogénéité est obtenue vraisemblablement à la suite de l'action des forces de sélection stabilisante, forces qui sont d'autant plus vives que le caractère étudié doit être maintenu dans une marge étroite pour assurer la survie de l'organisme. Tout écart à la valeur acquise par sélection au fil des générations lui est alors néfaste. Cela se traduit par une variabilité individuelle d'origine génétique faible. Cette remarque ne s'applique qu'à l'hérabilité au sens étroit, car la sélection diminue la variance additive, mais non aux variances de dominance ou d'épistasie, si elles existent, puisque celles-ci ne se transmettent pas. L'écart entre hérabilité au sens étroit et large pour un caractère donné est donc un moyen global de mesurer l'importance des effets de dominance et d'épistasie, mais aussi de mesurer l'intensité et la durée de l'action sélective sur ce caractère.

Une hérabilité n'est définie qu'à l'instant t , car sa valeur peut varier au cours du temps. En effet, les gènes ne s'expriment pas tous simultanément, mais au contraire de façon échelonnée dans le temps. Par ailleurs, la fréquence des gènes intervenant dans la variance additive V_a et donc dans h^2 varie avec le temps par pression sélective, ou par le seul hasard. De plus, les effets du milieu peuvent, eux aussi, fluctuer dans le temps : s'ils s'homogénéisent, alors la variance environnementale (V_e) diminue et l'hérabilité h^2 augmente ; s'ils se diversifient, V_e augmente tandis que h^2 diminue.

Ce sont ces deux mécanismes qui, en agissant conjointement, expliquent les variations de l'hérabilité observée au cours de la croissance pour les caractères anthropométriques (Tanner & Israelsohn, 1963 ; Susanne, 1969). Ils expliquent également les différences d'hérabilité selon les classes d'âge pour des caractères tels que le cholestérol (Martin *et al.*, 1973) ou selon que les sujets apparentés vivent en milieu homogène ou contrasté, ensemble ou séparément.

Un autre aspect du facteur temps ne doit pas être négligé. En effet, l'une des particularités des caractères physiologiques tient à l'importance de leurs fluctuations dans le temps, ces fluctuations prenant parfois l'aspect de rythmes, circadiens ou saisonniers par exemple.

Pour atteindre une meilleure définition de l'héritabilité, il est donc nécessaire de tenir compte de ces variations, en prenant soin d'effectuer de multiples mesures, espacées dans le temps, sur chaque sujet. Les facteurs génétiques qui semblaient disparaître – noyés sous l'amplification des facteurs de milieu apparemment exogènes prenant parfois l'aspect des rythmes – resurgissent alors, une fois ceux-ci pris en compte. Ainsi, par exemple, la pression artérielle systolique qui présente une faible héritabilité ($h^2 = 0,15$) lorsqu'une seule mesure est effectuée sur chaque sujet, s'élève à $h^2 = 0,68$ lorsque sont effectuées 6 mesures par jour, sur chaque sujet ; de même, le h^2 de la diastolique passe de 0,0 à 0,60 (Barcal *et al.*, 1969).

Une telle analyse peut être extrapolée. Il est en effet possible que l'amplitude des rythmes, et non plus seulement le niveau moyen autour duquel il s'effectue, aient une composante génétique. C'est ce qui a été observé pour la fréquence cardiaque et, de façon moins probante, pour le taux d'AVM urinaire (Barcal *et al.*, 1968).

Une étude d'héritabilité doit donc prendre en considération le facteur temps sous tous ses aspects : variation entre générations, auxologiques (croissance), chronologiques (rythmes, variabilité individuelle).

5.2 Estimation de l'héritabilité

Lorsque l'ensemble des conditions énumérées au paragraphe 4.1 à propos de l'estimation de la covariance génétique à partir de la ressemblance phénotypique (essentiellement absence de covariance environnementale et de covariance gènes-milieu, $CovE_iE_j = 0$ et $CovGE = 0$) sont satisfaites, alors on peut écrire :

– dans les cas où $\delta = 0$, c'est-à-dire parent-enfant par exemple ($\varphi = 1/2$) :

$$h^2 = \frac{V_a}{V_p} = \frac{Cov(P_i, P_j)}{V_p}$$

– dans le cas contraire où $\delta \neq 0$; seule la limite supérieure de l'héritabilité stricte peut donc être atteinte. En effet, dans le cas de germains par exemple :

$$\varphi = 1/4 \text{ et } Cov(P_i, P_j) = 1/2 V_a + 1/4 V_{AA} + \dots$$

$$h^2 \leq \frac{2 Cov(P_i, P_j)}{V_p} = \frac{V_a}{V_p} + \frac{1/2 V_d + 1/2 V_{aa} + \dots}{V_p}$$

La méthode gémellaire fournit également des estimations d'héritabilité. Toutefois, il ne s'agit pas de l'héritabilité telle que l'on vient de la définir, si bien qu'il est préférable de parler, à propos des études gémellaires, d'indice d'héritabilité.

Un grand nombre de formulations de cet indice ont été rappelées par Kang *et al.* (1978). Elles se fondent toutes sur l'estimation des ressemblances intra- et interaires de jumeaux et sur leurs comparaisons selon que les jumeaux sont mono- ou dizygotes.

Ces indices peuvent être considérés comme de simples rapports de variance : ils comparent ainsi la ressemblance entre jumeaux. Cependant, si l'on veut approfondir l'analyse et attribuer à ces index un sens génétique précis, plusieurs hypothèses doivent être retenues, comme le développent par exemple Christian *et al.* (1974) :

– les effets génétiques et les effets environnementaux doivent être additifs, sans interaction, selon le modèle : $P_i = G_i + E_i$;

– les covariances entre les effets génétiques et les effets d'environnement sont identiques, que l'on prenne l'un ou l'autre membre du couple gémellaire :

$$Cov(G_1, E_1) = Cov(G_2, E_2) =$$

$$Cov(G_1, E_2) = Cov(G_2, E_1)$$

– les covariances entre les effets environnementaux des deux membres du couple sont identiques chez les mono- et chez les dizygotes. Parfois, une hypothèse plus forte exige qu'elles soient nulles :

$$Cov_{MZ}(E_1, E_2) = Cov_{DZ}(E_1, E_2)$$

– Dans ces conditions, l'index d'héritabilité H^* est égal à :

$$H^* = \frac{V_a + 1/2 V_d + \dots}{V_g + V_e}$$

Il ne s'agit donc pas de l'héritabilité au sens étroit ou large, mais d'un rapport entre une combinaison de variances additive et de dominance et la variance totale.

BIBLIOGRAPHIE

- Barcal R, Sova J, Krikanovska M, Levy J & Matousek J (1968) Genetic background of circadian rhythms. *Nature* 220, 1128-1131.
- Barcal R, Simon J & Sova J (1969) Blood pressure in twins. *Lancet* 1, 1321.
- Cavalli-Sforza LL & Feldman M (1981) *Cultural transmission and Evolution. A quantitative approach*. Monographs in population biology, 16. Princeton University Press, Princeton.
- Christian JC, Kang KW & Norton JA Jr (1974) Choice of an estimate of genetic variance from twin data. *Amer J Hum Genet* 26, 154-161.
- Darlu P, Sagnier PP & Bois E (1990) Genealogical and genetical african admixture estimations, blood pressure and hypertension in a Caribbean community. *Ann Hum Biol* 17, 387-397.
- Falconer DS (1960) *Introduction to quantitative genetics*. Oliver & Boyd, Londres.
- Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans Roy Soc Edimb* 52, 399-433.
- Fisher RA (1941) Average excess and average effect of a gene substitution. *Ann Eugen* 11, 53-63.
- Gillois M (1969) *La relation d'identité génétique*. Thèse Faculté des sciences, Paris, 294 p.
- Haldane JBS (1946) The interaction of nature and nurture. *Ann Eugen* 13, 197-205.
- Harrison GA & Boyce AJ (1972) *The structure of human populations*. Clarendon Press, Oxford.
- Jacquard A (1970) *Structures génétiques des populations*. Masson, Paris.
- Jacquard A (1978) Modèle additif, génétique et idéologie. *La Pensée* 201, 19-32.
- Kang KW, Christian JC & Norton JA Jr. (1978) Heritability estimates from twin studies. I : Formulae of heritability estimates. *Acta Genet Med Gemellol* 27, 39-44.
- Kempthorne O (1954) The correlation between relatives in a random mating population. *Proc Roy Soc Lond B* 143, 103-113.
- Kempthorne O (1955) The theoretical values of correlations between relatives in random population. *Genetics* 40, 153-167.
- Kempthorne O (1960) *Biometrical Genetics*. Pergamon Press, New York.
- Lewontin RC (1974) The analysis of variance and the analysis of causes. *Amer J Hum Genet* 26, 400-425.
- Lewontin RC (1975) Genetics aspects of intelligence. *Annu Rev Genet* 9, 387-405.
- Malécot G (1948) *Les mathématiques de l'hérédité*. Masson, Paris.
- Martin AO, Kurczynski TW & Steinberg AG (1973) Familial studies of medical and anthropometric variables in a human isolate. *Amer J Hum Genet* 25, 581-593.
- Mather K (1949) *Biometrical genetics*. Methuen & Co, London.
- Olivier G (1991) La stature des français en 1990. *Bull Mem Soc Anthropol Paris* 3, 145-150.
- Pearson K (1930) The life, letters and labours of Francis Galton. IIIa : *Correlation, personal identification and eugenics*. Cambridge University Press.
- Susanne C (1969) *Recherche sur la transmission des caractères mesurables chez l'homme*. Thèse de doctorat, Université libre de Bruxelles, 167 p.
- Tanner JM & Israelsohn WJ (1963) Parent-child correlations for body measurements of children between the ages one month and 7 years. *Ann Hum Genet* 26, 245-259.
- Waddington CH (1957) *The strategy of the genes*. Allen & Unwin, Londres.
- Wright S (1952) The genetics of quantitative variability. In ECR Reeve & CH Waddington eds. : *Quantitative Inheritance*. HMSO, Londres, pp. 5-41.