

Modèle polygénique pour les caractères mesurables

par

E. DEFRISE-GUSSENHOVEN

1. Introduction

«Pourrait-on déduire de la variance de la stature le nombre de gènes responsables ?». Posée il y a bien longtemps par le Professeur Twiesselmann, cette question, toujours actuelle, éveilla mon intérêt pour la biométrie et la génétique humaine. C'est la même question qui a inspiré les réflexions qui suivent.

On peut espérer que la connaissance accrue des hormones de croissance et la synthèse de certaines d'entre elles, rapproche le moment où une analyse fine des acides aminés révélera des petites variations génétiques, n'entraînant pas des aberrations morphologiques brutales, mais expliquant les légères différences observables dans toute population normale. On sait qu'une partie de la variation de nombreux caractères mesurables est de nature génétique, mais aussi que les modèles mendéliens simples ne suffisent pas à expliquer leur mode de transmission. FISHER (1918), en conciliant la biométrie et les lois de Mendel, a proposé un modèle dans lequel de nombreux locus portent des gènes qui contribuent à l'élaboration du caractère mesurable : c'est le modèle polygénique, dont nous voulons parler.

2. Modèle général expliquant la variation d'un caractère mesurable

Chez l'homme, des études familiales ont prouvé l'héritabilité de caractères quantitatifs très variés comme les dermatoglyphes, les mesures corporelles, les réflectances de la peau, la hauteur de la voix, la quantité d'ions et d'urée dans le sang, les paramètres des électroencéphalogrammes, les tests d'aptitudes et de personnalité, etc. Dans une population homogène, ces caractères présentent souvent des distributions unimodales voire normales, comme si – ainsi que le pensait Quetelet – les déviations à la moyenne étaient comparables à des erreurs de mesure.

En toute généralité, on peut considérer la mesure P d'un caractère quantitatif observé chez un individu comme une fonction

$$P = F(x_1, x_2, \dots, x_p, z_1, z_2, \dots, z_q), \quad (1)$$

x_1, \dots, x_p étant les effets des gènes de p locus (que nous nommons variables génétiques) et z_1, \dots, z_q ceux des facteurs du milieu (et que nous appelons variables mésologiques). Pour évaluer les parts respectives que prennent chez un individu les variables x et z dans l'élaboration du phénotype P, on a cherché des situations dans lesquelles soit les x, soit les z sont constants.

Ainsi JOHANSEN (1903) a obtenu expérimentalement la distribution des poids de haricots d'une lignée pure dans laquelle les différences de poids proviennent uniquement des variables mésologiques, les variables génétiques x restant constantes. Dans d'autres expériences, notamment dans les essais de sélection des animaux domestiques (FALCONER 1960), on tente d'opérer dans un milieu très homogène, les variables mésologiques restant alors quasi constantes, de sorte que les différences observées de P sont dues surtout aux effets des gènes.

Mais il faut bien reconnaître que, même dans un milieu très homogène, les individus n'auront jamais des variables z tout à fait égales : les différences observées entre jumeaux monozygotes élevés ensemble suffiraient à le prouver. On aimerait assez pouvoir classer, même grossièrement, les caractères mesurables suivant l'importance prédominante que prennent les variables génétiques ou mésologiques dans leur détermination. Un tel classement est partiellement réalisable : les anthropologues admettent, par exemple, que la couleur de la peau est principalement d'origine génétique, tandis que d'autres caractères, comme le poids, varient beaucoup avec le milieu. Remarquons que les dissensions actuelles les plus vives sur l'héritabilité concernent surtout des caractères difficiles à mesurer comme le quotient intellectuel et pour lesquels les variables z semblent varier beaucoup d'un individu à l'autre, tout en étant rebelles à toute définition, à toute quantification. Si l'on peut calculer la quantité de protéines absorbées annuellement par un sujet et obtenir ainsi une valeur chiffrée pour l'une des variables z déterminant le poids, il est impossible de mesurer la quantité d'information absorbée par un sujet et qui le rendra plus ou moins apte à réussir un test intellectuel : les variables mésologiques influençant le quotient intellectuel échappent pour une bonne part à tout contrôle.

3. **Modèle simplifié de la variation d'un caractère mesurable**

On peut, sans perdre le caractère très général de la fonction F, décomposer celle-ci de la manière suivante

$$P = G(x_1, \dots, x_p) + E(z_1, \dots, z_q) + f(x_1, \dots, x_p, z_1, \dots, z_q) \quad (2)$$

afin d'en extraire les termes G et E, dépendant uniquement l'un des variables génétiques et l'autre des variables mésologiques. Il reste cependant un noyau dur dans cette expression, la fonction f, peu maniable, mais correspondant souvent à une réalité, car l'interaction continue, au cours du développement, des variables génétiques et mésologiques, peut être considérable. Ainsi, il arrive que le génotype A est supérieur au génotype B dans un milieu donné, mais inférieur à B dans un autre milieu (HALDANE 1946, FALCONER 1960).

Cependant, les généticiens ont décidé de négliger le terme de l'interaction gènes-milieu, pour ne conserver que l'expression

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

qui se prête à l'élaboration de plans expérimentaux et à leur interprétation. Il semble que cette simplification peut se justifier dans le cas de nombreux caractères quantitatifs (CAVALLI-SFORZA et BODMER 1971) mais on peut craindre que pour d'autres c'est justement la fonction f de l'interaction gènes-milieu qui est essentielle (TWIESSLMANN 1947).

4. Modèle $P = G + E$ et normalité de P

Supposons que G et E soient des fonctions simples des variables :

$$G = x_1 + \dots + x_p \text{ et } E = z_1 + \dots + z_q.$$

Alors (3) devient

$$P = x_1 + \dots + x_p + z_1 + \dots + z_q \quad (4)$$

D'après le théorème central limite, P aura, dans une population donnée, une distribution normale à condition que les variables x et z soient

- a) nombreuses,
- b) du même ordre de grandeur,
- c) indépendantes deux à deux.

On admet facilement que les conditions a et b sont remplies, mais la condition c est difficile à contrôler.

Une vérification fort indirecte de l'indépendance des variables a été tentée (DEFRISE 1954) en calculant le coefficient de corrélation entre la taille et son accroissement pendant un an, chez un groupe de 110 garçons de 9 ans fréquentant les écoles de Bruxelles ; on a trouvé $r = -0.02 \pm 0.09$. Cette valeur pratiquement nulle montre l'indépendance de l'accroissement annuel de la taille et de sa valeur initiale à 9 ans. Si donc la taille peut être considérée comme une somme d'accroissements successifs indépendants, on peut prévoir qu'elle sera distribuée normalement, ce qui est en effet le cas.

Remarquons que pour ces mêmes 110 garçons, le coefficient de corrélation entre le poids et son accroissement annuel était significatif : $r = 0.27 \pm 0.08$. L'accroissement annuel du poids dépend donc du poids initial, le théorème central limite ne peut être appliqué ; et, en effet, la distribution du poids est asymétrique comme il est bien connu.

5. Décomposition de la variance de P dans une population donnée

Du modèle (2) $P = G + E + f$, on tire la variance $V(P)$ de P :

$$V(P) = V(G) + V(E) + 2\text{cov}(GE) + V(f) + 2\text{cov}(Gf) + 2\text{cov}(Ef) \quad (5)$$

et du modèle simplifié $P = G + E$ (3), on tire

$$V(P) = V(G) + V(E) + 2\text{cov}(GE) \quad (6),$$

V désignant la variance et "cov" la covariance. Remarquons que dans sa forme (6) simplifiée, comme les valeurs extrêmes de la corrélation entre G et E sont ± 1 , le terme $2\text{cov}(GE)$ ne peut dépasser les limites $\pm \{V(G) + V(E)\}$. FALCONER (1960) donne un exemple de covariance positive entre G et E : on a l'habitude de donner plus de foin aux vaches qui produisent beaucoup de lait ; la corrélation positive qui en résulte entre P et E entraîne une corrélation positive entre G et E. De même, on pourrait s'attendre à une corrélation positive entre la vivacité innée d'un enfant et les stimulants intellectuels qu'il rencontre dans son milieu familial et à l'école. Dans ces exemples, le terme $2\text{cov}(GE)$ dépend de l'effet des variables génétiques x sur les variables mésologiques z. Comme on le voit dans la formule (5), la covariance entre G et E se distingue de $V(f)$, la variance de l'interaction génotype - milieu, discutée au paragraphe 2, et qui peut être très grande, tandis que la valeur de $2\text{cov}(GE)$ ne peut dépasser les limites $\pm \{V(G) + V(E)\}$.

Dans la pratique, on néglige le terme f, sa variance et ses covariances ; de même, dans la formule (6), on néglige la covariance $\text{cov}(GE)$, ou bien on groupe les termes comme suit

$$V(P) = \{V(G) + 2\text{cov}(GE)\} + V(E), \quad (7)$$

le terme entre crochets étant alors considéré comme la partie de la variance de P due aux gènes.

6. Remarques sur l'héritabilité d'un caractère quantitatif

Une des définitions de l'héritabilité d'un caractère mesurable est

$$h^2 = \frac{V(G)}{V(P)} = \frac{V(G)}{V(G) + V(E)} \quad \text{avec } 0 \leq h^2 < 1. \quad (8)$$

Comme $V(P)$ peut se déduire directement des observations, il suffit, pour calculer h^2 , de connaître $V(G)$. Cette variance se déduit des coefficients de corrélation entre sujets apparentés.

On trouve dans SUSANNE (1977) une description des méthodes de calcul de h^2 dans les populations humaines suivie d'un classement des caractères anthropologiques suivant leur degré d'héritabilité. Remarquons que h^2 décrit plus la population dans laquelle on opère que le caractère étudié. En effet, imaginons deux populations A et B dans lesquelles un même caractère (songeons au quotient intellectuel !) est mesuré. Supposons que le milieu dans la population A est très uniforme ; $V(E)$ sera petit et l'essentiel de la variation du phénotype P sera dû aux gènes : h^2 sera proche de 1. Si la population B, au contraire, présente des conditions de milieu très contrastées, la variance de E sera énorme et la plus grande partie de $V(P)$ consistera en $V(E)$: ici, h^2 sera proche de zéro. Le chiffre de 80 % avancé parfois pour l'héritabilité de l'intelligence est au mieux une caractéristique de certaines écoles dans certains pays, mais il n'est pas inéluctablement lié à l'intelligence.

Ajoutons cependant que si, pour un caractère donné, on trouvait toujours $h^2 = 0$, on aurait le droit de conclure que le caractère ne varie qu'avec le milieu ; au contraire, si on obtient parfois $h^2 > 0$, c'est que $V(P)$ contient réellement une partie génétique. C'est donc la variété et la multiplicité des études qui doivent finalement nous éclairer. C'est ce qu'ont senti ERLÉNMEYER-KIMLING et JARVIK (1963) qui ont rassemblé 99 coefficients de corrélation tirés de 52 études sur l'héritabilité de l'intelligence. Pour chaque type de parenté, parent-enfant, frère-frère, faux jumeaux, ces coefficients varient, mais leur moyenne est proche de 0,5, la valeur théorique attendue par la génétique. D'autre part, dans trois études sur les relations entre parents et enfants adoptifs, les coefficients de corrélation s'écartent de la valeur zéro, attendue dans le cas de sujets non apparentés, et ils révèlent ainsi (l'un d'eux atteint 0,40) les effets de l'éducation.

Ces effets se marquent aussi dans les différences de Q.I. observées entre jumeaux monozygotes élevés séparément. Malgré la prudence qui s'impose dans l'interprétation de tels résultats – notamment parce que la majorité des études provient de pays industrialisés peut-être trop peu dissemblables et en tout cas assez riches pour de telles recherches – il nous semble cependant qu'ils montrent que les points obtenus pour un test intellectuel dépendent des effets combinés des gènes et de l'éducation.

7. Modèle dans lequel Fisher analyse la fonction G

Dans un article célèbre de 1918, FISHER, parlant de la corrélation entre les mesures de deux frères, écrit : «an examination of the best available figures for

human measurements shows that there is little or no indication of non-genetic causes» ; cette remarque, bien conforme à l'esprit de l'époque, explique que dans la formule (3) $P = G + E$, Fisher va négliger le terme de l'environnement E. Sa décomposition de $V(G)$ comprend 3 variances : la première est due aux effets additifs des gènes, la deuxième à l'interaction des deux allèles d'un locus appelée dominance et la troisième à l'interaction des allèles de locus différents appelée épistasie.

Dans les paragraphes qui suivent, après une discussion sur le terme E, nous explicitons les termes de G qui conduisent à ces variances et indiquons les conditions nécessaires aux calculs. Nous prenons le cas de deux locus non liés, chacun avec deux allèles codominants, l'un avec les gènes A et a, l'autre avec les gènes B et b et nous supposons que les effets de ces gènes déterminent la partie génétique G du phénotype mesurable P.

8. Le terme E négligé par Fisher dans la formule $P = G(x_1, \dots, x_p) + E(z_1, \dots, z_q)$

La tendance actuelle est de s'interroger particulièrement sur les effets mésologiques puisque c'est sur le milieu que l'homme peut agir. Malheureusement, si Fisher a proposé une fonction G, personne n'a encore imaginé une fonction E, les variables mésologiques étant encore plus difficilement définissables et dénombrables que les gènes. Essayons cependant d'estimer la valeur de G pour un génotype donné, c'est-à-dire pour des valeurs connues de x_1, \dots, x_p , sans supposer le terme E nul pour chaque individu.

Pour tous les sujets de même génotype, donc de variables x_1, \dots, x_p égales et ayant la même valeur G, on peut écrire

$$\mathcal{E}(P) = G + \mathcal{E}(E) \quad (9),$$

\mathcal{E} désignant l'espérance mathématique. Si $\mathcal{E}(E) = 0$ pour chaque groupe d'individus de même génotype, on a

$$\mathcal{E}(P) = G \quad (10),$$

et la moyenne des mesures P des sujets de même génotype donnera leur valeur génotypique G. La condition $\mathcal{E}(E) = 0$ signifie que les sujets de même génotype occupent tous les milieux et que pour l'ensemble des sujets de la population on a $\text{cov}(GE) = 0$. La formule (6) devient

$$V(P) = V(G) + V(E) \quad (11)$$

Les développements exposés dans la suite ne sont légitimes que si on a $\text{cov}(GE) = 0$, condition qui remplace celle plus restrictive de Fisher annulant E pour chaque individu.

9. Fonction G dans le cas de deux allèles A(p) et a(q) d'un même locus (p + q = 1)

La fonction $G(x_1, \dots, x_p)$ ne dépend ici que de la seule variable x_1 puisqu'un seul locus est en jeu. Laissant tomber l'indice inutile, la valeur génotypique est $G(x)$, x prenant les valeurs 2, 1, 0 suivant que le sujet est AA, Aa ou aa. Les valeurs génotypiques resp. $G(x = 2)$, $G(x = 1)$, $G(x = 0)$ sont désignées par g_2 , g_1 , g_0 . Le tableau 1 résume les données dans le cas où la loi de Hardy-Weinberg est vérifiée pour les fréquences génotypiques du locus en question. FISHER (1918) exprime la plus grande partie possible de la variance de g en fonction des effets individuels des gènes A et a. Pour arriver au même résultat, non par l'analyse de la variance de g mais par son analyse directe, nous calculons la régression linéaire de g par rapport à x . Les coefficients b_0 et b_1 de la fonction

$$g = b_0 + b_1(x - \bar{x}) + \text{résidu} \tag{12}$$

sont calculés de manière à rendre minimum la somme des carrés des résidus.

Nous trouvons ainsi (voir Tableau 1)

TABLEAU 1				
Valeurs génotypiques. Un locus, deux allèles				
Génotype	Valeur génotypique	Variable indépendante x	x^2	Fréquence
AA	g_2	2	4	p^2
Aa	g_1	1	1	$2pq$
aa	g_0	0	0	q^2
moyenne	$\mu = p^2 g_2 + 2pq g_1 + q^2 g_0$	$\bar{x} = 2p$	$\overline{x^2} = 2p + 2p^2$	
variance	$\sigma_g^2 = p^2 g_2^2 + 2pq g_1^2 + q^2 g_0^2 - \mu^2$	$s_x^2 = 2pq$		
coef. de régr. $g/x : \alpha = p(g_2 - g_1) + q(g_1 - g_0)$				
résidu = $\delta \{ (1+2p)(x - \bar{x}) - (x^2 - \overline{x^2}) \}$ avec $\delta = g_1 - (g_2 + g_0)/2$				

$b_0 = \mu$ et $b_1 = \alpha$, coefficient de régression linéaire ; le résidu = $\delta \{ (1 + 2p)(x - \bar{x}) - (x^2 - \overline{x^2}) \}$ est une fonction quadratique de x , indépendante de la partie linéaire $x - \bar{x}$. Les trois valeurs génotypiques tirées de (12) s'expriment en fonction de μ , α , δ , p , q et de x (valant resp. 2, 1 ou 0) :

$$\left. \begin{aligned} g_2 &= \mu + 2q\alpha - 2q^2\delta \\ g_1 &= \mu + (1 - 2p)\alpha + 2pq\delta \\ g_0 &= \mu - 2p\alpha - 2p^2\delta \end{aligned} \right\} \tag{13}$$

Dans ces formules, les termes en α et δ sont dits respectivement linéaire ou «additif» et quadratique ou «de dominance». Ce dernier terme disparaît lorsque l'hérédité est strictement intermédiaire : $g_1 = (g_2 + g_0)/2$; $\delta = 0$ est donc une propriété des génotypes et non des fréquences géniques p et q du pays à l'étude.

Le cas de $\alpha = 0$ est différent ; il se présente lorsque

$$\frac{p}{q} = \frac{g_0 - g_1}{g_2 - g_1} \quad (14)$$

Ce cas est rare, non seulement parce que l'hétérozygote doit être à la fois supérieur ou inférieur aux deux homozygotes, mais aussi parce que l'égalité (14) ne peut être vraie dans tous les pays puisqu'elle dépend des valeurs variables de p et q . La propriété $\alpha = 0$ n'est donc pas uniquement de nature génétique. Notons encore que l'on peut avoir en même temps $\alpha = 0$ et $\delta \neq 0$. Quand $\alpha = \delta = 0$, les trois génotypes ont une même valeur.

Remarque

Essayons d'ajuster à g_2, g_1, g_0 une fonction du 2^e degré en x , par exemple

$$g = b_0 + b_1(x - \bar{x}) + b_2(x^2 - \bar{x}^2) + \text{résidu} \quad (15)$$

On trouve $b_0 = \mu$, $b_1 = \alpha + (1 + 2p)\delta$, $b_2 = -\delta$ et un résidu égal à zéro. Les termes en b_1 et b_2 ne sont pas indépendants. Au contraire, si l'on ajuste g_2, g_1, g_0 par la fonction

$$g = b_0 + b_1(x - \bar{x}) + b_2\{(1 + 2p)(x - \bar{x}) - (x^2 - \bar{x}^2)\} + \text{résidu} \quad (16)$$

on trouve $b_0 = \mu$, $b_1 = \alpha$, $b_2 = \delta$ et le résidu est nul. Cette fonction est telle que la partie linéaire $x - \bar{x}$ et la partie quadratique $\{(1 + 2p)(x - \bar{x}) - (x^2 - \bar{x}^2)\}$ peuvent varier indépendamment l'une de l'autre ; leur covariance est nulle ; ce sont des polynômes orthogonaux. Les fonctions (15) et (16) conduisent aux mêmes expressions (13) que la fonction (12).

10. Effets individuels (α_1 et α_2) des gènes A et a. Signification de α

Dans une population panmictique, un gamète contenant le gène A a une probabilité p de s'unir à un gamète contenant le gène A et de donner un génotype AA. Ce même gamète a une probabilité q de produire un génotype Aa. La valeur moyenne des génotypes produits par des gamètes contenant A est par conséquent $pg_2 + qg_1$ et son écart à la moyenne de la population est $pg_2 + qg_1 - \mu = q\alpha$. Cette quantité est souvent notée α_1 ; c'est l'effet individuel du gène A sur la valeur du génotype. On calcule de même $\alpha_2 = -p\alpha$, l'effet individuel du gène a sur la valeur du génotype. Le tableau 2 résume les calculs.

TABEAU 2
Effets individuels α_1 et α_2 des gènes A et a sur les génotypes AA, Aa et aa

Gamète	Génotypes produits. Valeurs et fréquences.			Valeurs moyennes des génotypes produits.	Écarts à la moyenne μ des valeurs moyennes des génotypes produits.	Effets moyens des gènes A et a, dits aussi effets individuels.
	AA g_2	Aa g_1	aa g_0			
A	p	q		$pg_2 + qg_1$	$pg_2 + qg_1 - \mu$	$= qa = \alpha_1$
a		p	q	$pg_1 + qg_0$	$pg_1 + qg_0 - \mu$	$= -pa = \alpha_2$

Rappel : $\mu = p^2g_2 + 2pqg_1 + q^2g_0$; $\alpha = p(g_2 - g_1) + q(g_1 - g_0)$

Décomposition de la variance $V(G)$.

$\sigma_g^2 = V(G) = 2pqa^2 + 4p^2q^2\delta^2$	variance des valeurs génotypiques
$\sigma_a^2 = 2pqa^2 = 2pq(\alpha_1 - \alpha_2)^2$	variance dite additive
$\sigma_d^2 = 4p^2q^2\delta^2$	variance due à la dominance

Remarquons que $\alpha_1 - \alpha_2 = qa - (-pa) = a$, de sorte que a a une signification intéressante : c'est l'effet sur la valeur génotypique de la substitution du gène a par le gène A. Les formules (13) peuvent donc aussi s'écrire :

$$\left. \begin{aligned} g_2 &= \mu + 2\alpha_2 - 2q^2\delta \\ g_1 &= \mu + \alpha_1 + \alpha_2 + 2pq\delta \\ g_0 &= \mu + 2\alpha_1 - 2p^2\delta \end{aligned} \right\} \quad (17)$$

Elles mettent en évidence la partie additive des effets individuels des gènes A et a, le terme en δ concernant leur interaction (la dominance).

On calcule facilement la variance $V(G)$ des valeurs génotypiques en fonction de α et de δ :

$$V(G) = \sigma_g^2 = 2pqa^2 + 4p^2q^2\delta^2. \quad (18)$$

$2pqa^2 = \sigma_a^2$ est la variance dite additive due aux effets individuels des gènes A et a. $4p^2q^2\delta^2 = \sigma_d^2$ est la partie de la variance due à l'effet de dominance ; c'est la variance, rendue minimum, du résidu de la formule (12).

Ajoutons encore des propriétés bien connues de σ_a^2 et de σ_d^2 . Les covariances entre les valeurs génotypiques de différents types de sujets apparentés sont :

$$\left. \begin{aligned} \text{cov}(\text{père, fils}) &= \frac{\sigma_a^2}{2} = pq\alpha^2 \\ \text{cov}(\text{oncle, neveu}) &= \frac{\sigma_a^2}{4} = \frac{1}{2} pq\alpha^2 \\ \text{cov}(\text{frère, frère}) &= \frac{\sigma_a^2}{2} + \frac{\sigma_d^2}{4} = pq\alpha^2 + p^2q^2\delta^2. \end{aligned} \right\} \quad (19)$$

Si l'on suppose donc nul l'effet moyen du milieu, les formules (19) montrent que la covariance frère-frère excède d'autant plus celle de père-fils que l'effet de dominance est plus grand. Les formules (19) et d'autres analogues sont utilisées pour estimer σ_a^2 et σ_d^2 .

11. Polyallélie. Un locus, plus de 2 allèles

La formule (12) dans laquelle x prend les valeurs 2, 1, 0 suivant que le génotype est AA, Aa ou aa ne peut être généralisée telle quelle au cas de plusieurs allèles. Nous indiquons rapidement comment on obtient une estimation des effets additifs $\alpha_1, \alpha_2, \dots, \alpha_k$ dans le cas de k allèles A_1, \dots, A_k pour un locus. Soient p_1, \dots, p_k les fréquences de ces gènes dans la population et désignons par g_{ij} la valeur génotypique d'un sujet $A_i A_j$.

Soit $\mu = \sum_{ij} p_i p_j g_{ij}$ la moyenne des g dans la population et imposons aux α_i la condition $\sum_i p_i \alpha_i = 0$.

Par définition, $\alpha_1, \dots, \alpha_k$, les valeurs individuelles des gènes, sont telles que $\sum_{ij} p_i p_j (g_{ij} - \mu - \alpha_i - \alpha_j)^2$ est minimum.

La solution est $\alpha_i = \sum p_j g_{ij} - \mu$.

La quantité α_i est donc, comme dans le cas de deux allèles, l'écart à la moyenne générale μ de la moyenne des génotypes produits par toutes les gamètes contenant le gène A_i . Un exposé très clair des calculs est donné par EWENS (1979).

12. Deux locus non liés, l'un avec les allèles A(p) et a(q), l'autre avec les allèles B(p₁) et b(q₁)

La fonction $G(x_1, x_2, \dots, x_p)$ du paragraphe 3 ne dépend ici que des deux variables génétiques x_1 et x_2 , que nous désignerons par x et y et qui prennent les valeurs resp. 2, 1, 0 suivant que le génotype contient 2, 1 ou 0 majuscules A ou B. La valeur génotypique $G(x = i, y = j)$ sera notée g_{ij} avec $i, j = 2, 1, 0$. Les fréquences des gènes p, q et p_1, q_1 , avec $p + q = 1$ et $p_1 + q_1 = 1$, donnent les fréquences des génotypes suivant la loi de Hardy-Weinberg. Le tableau 3 montre le détail de ces notations.

Comme g_2, g_1, g_0 sont les moyennes respectives des valeurs des génotypes contenant 2, 1 ou 0 gènes A, les valeurs $\alpha, \alpha_1, \alpha_2$ et δ_A du paragraphe 9 et des tableaux 1 et 2 sont ici :

$$\left. \begin{aligned} \alpha &= p(g_2 - g_1) + q(g_1 - g_0) \\ \alpha_1 &= q\alpha : \alpha_2 = -p\alpha : \alpha_1 - \alpha_2 = \alpha \\ \delta_A &= g_1 - \frac{g_2 + g_0}{2} \end{aligned} \right\} \quad (20)$$

De manière analogue, pour le locus B, on définit :

TABLEAU 3

Génotypes, valeurs et fréquences des génotypes, valeurs de x et y.
En marge, les valeurs génotypiques moyennes et leur fréquence

AABB	AABb	AAbb	
g_{22}	g_{21}	g_{20}	$g_{2.} = p_1^2 g_{22} + 2p_1 q_1 g_{21} + q_1^2 g_{20}$
$p^2 p_1^2$	$p^2 2p_1 q_1$	$p^2 q_1^2$	p^2
$x=2, y=2$	$x=2, y=1$	$x=2, y=0$	
AaBB	AaBb	Aabb	
g_{12}	g_{11}	g_{10}	$g_{1.} = p_1^2 g_{12} + 2p_1 q_1 g_{11} + q_1^2 g_{10}$
$2pq p_1^2$	$2pq 2p_1 q_1$	$2pq q_1^2$	$2pq$
$x=1, y=2$	$x=1, y=1$	$x=1, y=0$	
aaBB	aaBb	aabb	
g_{02}	g_{01}	g_{00}	$g_{0.} = p_1^2 g_{02} + 2p_1 q_1 g_{01} + q_1^2 g_{00}$
$q^2 p_1^2$	$q^2 2p_1 q_1$	$q^2 q_1^2$	q^2
$x=0, y=2$	$x=0, y=1$	$x=0, y=0$	
$g_{.2}$	$g_{.1}$	$g_{.0}$	$g_{..} = \mu = p^2 g_{.2} + 2pq g_{.1} + q^2 g_{.0}$
p_1^2	$2p_1 q_1$	q_1^2	1
$g_{.2} = p^2 g_{22} + 2pq g_{12} + q^2 g_{02}$ $g_{.1} = p^2 g_{21} + 2pq g_{11} + q^2 g_{01}$ $g_{.0} = p^2 g_{20} + 2pq g_{10} + q^2 g_{00}$			$\mu = p_1^2 g_{.2} + 2p_1 q_1 g_{.1} + q_1^2 g_{.0}$

$$\left. \begin{aligned} \beta &= p_1(g_{.2} - g_{.1}) + q_1(g_{.1} - g_{.0}) \\ \beta_1 &= q_1\beta; \beta_2 = -p_1\beta; \beta_1 - \beta_2 = \beta \\ \delta_B &= g_{.1} - \frac{g_{.2} + g_{.0}}{2} \end{aligned} \right\} \quad (21)$$

β_1 et β_2 sont les effets individuels additifs des gènes B et b et β la valeur de substitution de b par B.

Ces notations étant connues, nous calculons la fonction de régression linéaire de g_{ij} par rapport à x et y ,

$$g_{ij} = b_0 + b_1(x - \bar{x}) + b_2(y - \bar{y}) + \text{résidu} \quad (22)$$

en déterminant b_0 , b_1 et b_2 pour rendre minimum la somme des carrés des résidus. Une fois de plus, nous voulons expliquer le plus possible de la valeur g_{ij} par une fonction linéaire du nombre de gènes A et B. Nous utilisons les 9 valeurs g_{ij} pour calculer b_0 , b_1 , b_2 .

On trouve $b_0 = \mu$ (tableau 3)

$$b_1 = \alpha \quad (\text{voir formule 20})$$

$$b_2 = \beta \quad (\text{voir formule 21})$$

$$g_{ij} = \mu + \alpha(x - \bar{x}) + \beta(y - \bar{y}) + \text{résidu} \quad (23)$$

Les termes en α et β et le résidu sont indépendants. Les coefficients de régression sont les valeurs de substitution. Pour inclure les termes de dominance, nous essayons d'ajuster aux g_{ij} une fonction quadratique choisie pour obtenir des termes indépendants comme dans la formule (16) :

$$\begin{aligned} g_{ij} = & b_0 + b_1(x - \bar{x}) + b_2\{(1 + 2p)(x - \bar{x}) - (x^2 - \bar{x}^2)\} \\ & + b_3(y - \bar{y}) + b_4\{(1 + 2p_1)(y - \bar{y}) - (y^2 - \bar{y}^2)\} + \text{résidu} \end{aligned} \quad (24)$$

Cette fois, ce sont 5 inconnues, b_0, b_1, \dots, b_4 , que nous devons tirer des g_{ij} (voir formules (20) et (21)).

En remplaçant les coefficients b par leur valeur, on trouve

$$\begin{aligned} g_{ij} = & \mu + \alpha(x - \bar{x}) + \delta_A\{(1 + 2p)(x - \bar{x}) - (x^2 - \bar{x}^2)\} \\ & + \beta(y - \bar{y}) + \delta_B\{(1 + 2p_1)(y - \bar{y}) - (y^2 - \bar{y}^2)\} + \text{résidu} \end{aligned} \quad (25)$$

La forme (23) s'est compliquée des 2 termes de dominance. Les termes linéaires en α et β et les termes quadratiques en δ_A et δ_B sont indépendants. La somme des carrés des résidus est minimum.

Pour inclure un terme relatif à l'épistasie, c'est-à-dire à l'interaction entre les locus A et B, nous ajoutons à (24) un terme $b_5(x - \bar{x})(y - \bar{y})$. Nous tirons les 6 inconnues b_0, \dots, b_5 des 9 nombres g_{ij} . En remplaçant les coefficients b par leur valeur respective $\mu, \alpha, \delta_A, \beta, \delta_B$ et E , on obtient :

$$\begin{aligned} g_{ij} = & \mu + \alpha(x - \bar{x}) + \delta_A\{(1 + 2p)(x - \bar{x}) - (x^2 - \bar{x}^2)\} \\ & + \beta(y - \bar{y}) + \delta_B\{(1 + 2p_1)(y - \bar{y}) - (y^2 - \bar{y}^2)\} \\ & + E(x - \bar{x})(y - \bar{y}) + \text{résidu} \end{aligned} \quad (26)$$

avec $E = pp_1e_{22} + pq_1e_{21} + qp_1e_{12} + qq_1e_{11}$ (voir tableau 4).

TABLEAU 4
Notations pour les termes d'épistasie

$e_{22} = g_{22} - g_{21} - g_{12} + g_{11}$	}	grandeurs qui ne dépendent que des valeurs génotypiques
$e_{21} = g_{21} - g_{20} - g_{11} + g_{10}$		
$e_{12} = g_{12} - g_{11} - g_{02} + g_{01}$		
$e_{11} = g_{11} - g_{10} - g_{01} + g_{00}$		
$E = pp_1 e_{22} + pq_1 e_{21} + qp_1 e_{12} + qq_1 e_{11}$	}	grandeurs dépendant des valeurs génotypiques et des fréquences géniques
$A = -\{p(e_{22}-e_{21}) + q(e_{12}-e_{11})\}$		
$B = \{p_1(e_{22}-e_{12}) + q_1(e_{21}-e_{11})\}$		
$C = e_{22} - e_{21} - e_{12} + e_{11}$	}	grandeur qui ne dépend que des valeurs génotypiques

E est relatif au terme d'épistasie lin. x lin. (add. x add.),
 A au terme lin. x quadrat. (add. x dom.),
 B au terme quadrat. x lin. (dom. x add.),
 C au terme quadrat. x quadrat. (dom. x dom.).

Le terme en E, relatif à l'épistasie, est appelé parfois terme d'interaction linéaire x linéaire ou additif x additif. Les termes en α , δ_A , β , δ_B et E sont indépendants : ce sont des polynômes orthogonaux en x et y.

Enfin, nous ajoutons à (24), non seulement un terme $b_5(x - \bar{x})(y - \bar{y})$ (lin. x lin.), mais encore les termes

$$\begin{aligned}
 & b_6(x - \bar{x})\{(1 + 2p_1)(y - \bar{y}) - (y^2 - \bar{y}^2)\} && (\text{lin.} \times \text{quadratique}), \\
 & b_7\{(1 + 2p)(x - \bar{x}) - (x^2 - \bar{x}^2)\}(y - \bar{y}) && (\text{quadrat.} \times \text{linéaire}), \\
 \text{et} & b_8\{(1 + 2p)(x - \bar{x}) - (x^2 - \bar{x}^2)\}\{(1 + 2p_1)(y - \bar{y}) - (y^2 - \bar{y}^2)\} && (\text{quadr.} \times \text{quadr.}).
 \end{aligned}$$

Des 9 nombres g_{ij} nous tirons les 9 inconnues b_0, b_1, \dots, b_8 . On trouve $b_0 = \mu$, $b_1 = \alpha$, $b_2 = \delta_A$, $b_3 = \beta$, $b_4 = \delta_B$, $b_5 = E$, $b_6 = A/2$, $b_7 = B/2$, $b_8 = C/4$. Il n'y a pas de résidu, l'ajustement est parfait. Les polynômes sont encore orthogonaux. Le tableau 4 donne le détail des coefficients E, A, B, C.

Finalement, g_{ij} se décompose comme suit, avec $x = i$ et $y = j$ ($i, j = 2, 1, 0$) et $\bar{x} = 2p$, $\bar{y} = 2p_1$, $\bar{x}^2 = 2p + 2p^2$, $\bar{y}^2 = 2p_1 + 2p_1^2$:

$$\begin{aligned}
 g_{ij} = & \mu + \alpha(x - \bar{x}) + \delta_A \{ (1 + 2p)(x - \bar{x}) - (x^2 - \bar{x}^2) \} + \beta(y - \bar{y}) + \\
 & \delta_B \{ (1 + 2p)(y - \bar{y}) - (y^2 - \bar{y}^2) \} + E(x - \bar{x})(y - \bar{y}) \\
 & + \frac{A}{2} (x - \bar{x}) \{ (1 + 2p)(y - \bar{y}) - (y^2 - \bar{y}^2) \} \\
 & + \frac{B}{2} \{ (1 + 2p)(x - \bar{x}) - (x^2 - \bar{x}^2) \} (y - \bar{y}) \\
 & + \frac{C}{4} \{ (1 + 2p)(x - \bar{x}) - (x^2 - \bar{x}^2) \} \{ (1 + 2p)(y - \bar{y}) - (y^2 - \bar{y}^2) \}
 \end{aligned} \tag{27}$$

Les 4 derniers termes sont ceux de l'épistasie.

La variance $\sigma_g^2 = V(G)$ est

$$\begin{aligned}
 \sigma_g^2 = & \underbrace{2pq\alpha^2 + 2p_1q_1\beta^2}_{\sigma_a^2} + \underbrace{4p^2q^2\delta_A^2 + 4p_1^2q_1^2\delta_B^2}_{\sigma_d^2} + \\
 & \underbrace{4pp_1q_1E^2 + 2pp_1^2q_1^2A^2 + 2p^2q^2p_1q_1B^2 + p^2q^2p_1^2q_1^2C^2}_{\sigma_i^2}
 \end{aligned} \tag{28}$$

σ_a^2 , σ_d^2 et σ_i^2 sont respectivement les variances additive, de dominance et d'épistasie (voir paragraphe 7) trouvées par FISHER (1918) et qu'on peut généraliser aux cas de polyallélie et aux cas de plus de deux locus.

Le tableau 5 donne la matrice des variances et des covariances des g_{ij} et des 8 polynômes (27) dont les coefficients sont respectivement α , δ_A , β , δ_B , E , $A/2$, $B/2$, $C/4$. Toutes les covariances sont nulles sauf celles relatives aux g_{ij} : les polynômes sont orthogonaux.

Le tableau 6 contient le détail des valeurs génotypiques g_{ij} analysées d'après la formule (27) : x et y prennent les valeurs i et j ($i, j = 2, 1, 0$) ; α , β , δ_A et δ_B sont calculés d'après (20) et (21), E , A , B , C d'après le tableau 4. Aux fonctions de régression (23), (25), (26), (27) correspondent des résidus de plus en plus petits. L'ajustement (27) est même parfait, le résidu étant nul.

Les raisonnements qui précèdent se généralisent tels quels aux cas de plus de deux locus et on a toujours

$$\sigma_g^2 = \sigma_a^2 + \sigma_d^2 + \sigma_i^2. \tag{29}$$

On peut déduire de (29) la covariance entre frères ou entre jumeaux dizygotes de même sexe :

$$\text{cov (frère, frère)} = \frac{1}{2} \sigma_a^2 + \frac{1}{4} \sigma_d^2 + f \sigma_i^2, \tag{30}$$

f étant un nombre compris entre 0 et 1 ; nous l'avons calculé dans le cas de deux locus non liés, chacun avec deux allèles :

$$f = \frac{1}{2} - \frac{1}{4} (pq + p_1q_1 - pp_1q_1) \tag{31}$$

TABLEAU 5

Variances et covariances des g_{ij} et des 8 polynômes de la régression (27)

Nous notons $x-\bar{x}$: add.A ; $(1+2p)(x-\bar{x}) - (x^2-\bar{x}^2) = \text{dom.A}$
 $y-\bar{y}$: add.B ; $(1+2p_1)(y-\bar{y}) - (y^2-\bar{y}^2) = \text{dom.B}$
 $(x-\bar{x})(y-\bar{y})$: add.A, add.B ; $(x-\bar{x})\{(1+2p_1)(y-\bar{y}) - (y^2-\bar{y}^2)\} = \text{add.A, dom.B.}$
 $\{(1+2p)(x-\bar{x}) - (x^2-\bar{x}^2)\}(y-\bar{y})$: dom.A, add.B
 $\{(1+2p)(x-\bar{x}) - (x^2-\bar{x}^2)\}\{(1+2p_1)(y-\bar{y}) - (y^2-\bar{y}^2)\}$: dom.A, dom.B.

Matrice des variances et des covariances

g_{ij}	add.A	dom.A	add.B	dom.B	add.A, add.B	add.A, dom.B	add.B, dom.A	dom.A, domB
σ_g^2	$2pq\alpha$	$4p^2q^2\delta_A$	$2p_1q_1\beta$	$4p_1^2q_1^2\delta_B$	$4pp_1q_1E$	$4pp_1^2q_1^2A$	$4p^2q^2p_1q_1B$	$4p^2q^2p_1^2q_1^2C$
	$2pq$	0	0	0	0	0	0	0
		$4p^2q^2$	0	0	0	0	0	0
			$2p_1q_1$	0	0	0	0	0
				$4p_1^2q_1^2$	0	0	0	0
					$4pp_1q_1$	0	0	0
						$8pp_1^2q_1^2$	0	0
							$8p^2q^2p_1q_1$	0
								$16p^2q^2p_1^2q_1^2$

TABLEAU 6

Analyse des valeurs génotypiques lorsque 2 loci, chacun avec deux allèles, déterminent le caractère mesuré

$g_{22} = \mu +$	$2q\alpha - 2q^2\delta_A +$	$2q_1\beta - 2q_1^2\delta_B +$	$4qq_1E -$	$2qq_1^2A -$	$2q^2q_1B +$	$q^2q_1^2C$
$g_{21} = \mu +$	$2q\alpha - 2q^2\delta_A + (1-2p_1)\beta + 2p_1q_1\delta_B +$	$2q(1-2p_1)E +$	$2qp_1q_1A - q^2(1-2p_1)B - q^2p_1q_1C$			
$g_{20} = \mu +$	$2q\alpha - 2q^2\delta_A -$	$2p_1\beta - 2p_1^2\delta_B -$	$4qp_1E -$	$2qp_1^2A +$	$2q^2p_1B +$	$q^2p_1^2C$
$g_{12} = \mu + (1-2p)\alpha + 2pq\delta_A +$	$2q_1\beta - 2q_1^2\delta_B +$	$2(1-2p)q_1E -$	$(1-2p)q_1^2A +$	$2ppq_1B -$	ppq_1^2C	
$g_{11} = \mu + (1-2p)\alpha + 2pq\delta_A + (1-2p_1)\beta + 2p_1q_1\delta_B + (1-2p)(1-2p_1)E + (1-2p)p_1q_1A + pq(1-2p_1)B + ppq_1q_1C$						
$g_{10} = \mu + (1-2p)\alpha + 2pq\delta_A -$	$2p_1\beta - 2p_1^2\delta_B -$	$(1-2p)2p_1E -$	$(1-2p)p_1^2A -$	$2ppq_1B -$	ppq_1^2C	
$g_{02} = \mu -$	$2p\alpha - 2p^2\delta_A +$	$2q_1\beta - 2q_1^2\delta_B -$	$4pp_1E +$	$2pp_1^2A -$	$2p^2q_1B +$	$p^2q_1^2C$
$g_{01} = \mu -$	$2p\alpha - 2p^2\delta_A + (1-2p_1)\beta + 2p_1q_1\delta_B -$	$2p(1-2p_1)E -$	$2pp_1q_1A - p^2(1-2p_1)B - p^2p_1q_1C$			
$g_{00} = \mu -$	$2p\alpha - 2p^2\delta_A -$	$2p_1\beta - 2p_1^2\delta_B +$	$4pp_1E +$	$2pp_1^2A +$	$2p^2p_1B +$	$p^2p_1^2C$

Lorsque les termes de dominance et d'épistasie s'annulent, on obtient

$g_{22} = \mu + 2\alpha_1 + 2\beta_1$	$g_{12} = \mu + \alpha_1 + \alpha_2 + 2\beta_1$	$g_{02} = \mu + 2\alpha_2 + 2\beta_1$
$g_{21} = \mu + 2\alpha_1 + \beta_1 + \beta_2$	$g_{11} = \mu + \alpha_1 + \alpha_2 + \beta_1 + \beta_2$	$g_{01} = \mu + 2\alpha_2 + \beta_1 + \beta_2$
$g_{20} = \mu + 2\alpha_1 + 2\beta_2$	$g_{10} = \mu + \alpha_1 + \alpha_2 + 2\beta_2$	$g_{00} = \mu + 2\alpha_2 + 2\beta_2$

Remarquons que $\delta_A = 0$ et $\delta_B = 0$ n'impliquent pas la nullité de E, A, B, C.

CHRISTIAN (1974) a montré que l'analyse d'un caractère quantitatif chez des jumeaux mono- et dizygotes permet, sous certaines conditions, d'estimer

$$G_T = \sigma_g^2 - \text{cov}(\text{frère, frère}) = \frac{1}{2} \sigma_a^2 + \frac{3}{4} \sigma_d^2 + (1-f)\sigma_i^2 \quad (32)$$

L'estimateur est

$$\hat{G}_T = M_{\text{WDZ}} - M_{\text{WMZ}} \quad (33),$$

M_{WDZ} et M_{WMZ} étant les sommes des carrés moyens des différences intrapaires, respectivement chez les jumeaux dizygotes et monozygotes.

13. En guise de conclusion

Dans un cas simple – deux locus non liés, chacun avec deux allèles –, nous avons décomposé g , la valeur génétique d'un individu de génotype donné, en une somme d'éléments, dus soit aux effets individuels des gènes, soit à la dominance, soit aux interactions entre les gènes de locus différents. La variance de g , calculée avec ces éléments, permet de retrouver la décomposition de σ_g^2 proposée par FISHER (1918), COCKERHAM (1954) et bien d'autres. Cette décomposition de g et de σ_g^2 est intéressante parce qu'elle permet d'analyser en termes clairement définissables la variation si complexe observée entre les individus pour chaque caractère supposé polygénique.

Nous rappelons maintenant toutes les conditions que doivent remplir la population et le caractère étudié pour conduire à l'analyse présentée :

1. La population doit être grande, panmictique, fermée. Le modèle est donc déterministe et la loi de Hardy-Weinberg liant les fréquences des génotypes à celles des gènes est satisfaite.
2. Nous supposons nulle la fonction $f(2)$ de l'interaction des gènes et du milieu qui pourrait être importante au cours du développement d'un sujet.
3. Nous supposons que les sujets de génotypes différents ont des chances égales d'occuper un milieu donné, d'où $\text{cov}(GE) = 0$. Pour cela, il suffit que l'espérance mathématique $\mathbf{E}(E)$ étendue à tous les sujets de même génotype, soit une constante dans la population ; sans nuire à la généralité, nous avons pris zéro pour cette constante, donc $\mathbf{E}(E) = 0$.
4. Nous avons choisi d'abord une fonction de régression linéaire pour g/x et $g/x, y$ (formules (12) et (22)), afin d'expliquer la plus grande partie possible de g par les effets individuels des gènes : α_1, α_2 (17) et β_1, β_2 (21). L'effet de la dominance sur le génotype se trouve ainsi réduit au minimum.
5. Dans le cas de deux locus, nous avons utilisé une fonction quadratique de $g/x, y$ de manière à obtenir des termes linéaires et quadratiques orthogonaux et à rendre minimum la somme des carrés des résidus qui concernent l'épistasie. Ce

procédé n'est peut-être pas toujours défendable. En effet, alors qu'il semble logique de rendre minimum l'effet de la dominance, nous pensons que l'épistasie n'a pas nécessairement des effets moins grands que les gènes pris séparément. Les effets de position bien connus chez les drosophiles suffiraient à nous le montrer.

Enfin, remarquons que les coefficients δ_A , δ_B et C de la fonction de régression (27) sont les seuls qui ne dépendent que des valeurs génotypiques. Les autres coefficients sont également des fonctions des g mais ils dépendent aussi des fréquences géniques p et q, donc de la population particulière que l'on étudie. Cela est bien décevant et on trouve dans JACQUARD (1977) des exemples numériques imaginés pour montrer que les valeurs individuelles des gènes varient avec les fréquences géniques de la population.

Les analyses de g et de σ_g^2 que nous venons d'exposer ont donc certaines faiblesses sans pour autant être dénuées de toute valeur. Admirons que l'article de Fisher écrit en 1918 serve encore en 1980 de base au calcul des covariances entre sujets apparentés.

BIBLIOGRAPHIE

- BODMER, W. F. and CAVALLI-SFORZA, L. L.
1976 *Genetics, Evolution, and Man*.
San Francisco, Freeman, 781 p.
- CAVALLI-SFORZA, L. L. and BODMER, W. F.
1971 *The genetics of human populations*.
San Francisco, Freeman, 965 p.
- CHRISTIAN, J. C., KE WON KANG, NORTON, J. A.
1974 Choice of an estimate of genetic variance from twin data.
Am. J. hum. Genet., **26** : 154-161.
- COCKERHAM, C. C.
1954 An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present.
Genetics, **39** : 859-882.
- DEFRISE-GUSSENHOVEN, E.
1954 Croissance et débilité.
Mémoire Inst. roy. Sci. nat. Belgique, **128** : 70 p.
- ERLENMEYER-KIMLING, L. and JARVIK, L. F.
1963 Genetics and intelligence : a review.
Science, **142** : 1477-1479.
- EWENS, W. J.
1979 *Mathematical population genetics*.
Berlin, Springer, 325 p.

FALCONER, D. S.

- 1960 *Introduction to quantitative genetics.*
New York, Ronald Press Company, 365 p.

FISHER, R. A.

- 1918 The correlations between relatives on the supposition of Mendelian inheritance.
Trans. roy. Soc. Edinburgh, **52** : 399-433.

HALDANE, J. B. S.

- 1946 The interaction of nature and nurture.
Ann. Eugen., London, **13** : 197-205.

JACQUARD, A.

- 1977 *Concepts en génétique des populations.*
Paris, Masson, 128 p.

JOHANSEN, W.

- 1903 *Erblichkeit in Populationen und in reinen Linien.*
Jena, G. Fischer.

SUSANNE, C.

- 1977 Heritability of anthropological characters.
Human Biology, **49** : 573-580.

TWIESELMANN, F.

- 1947 *L'hérédité. Buts et méthodes de la génétique humaine.*
Bruxelles, Institut de Sociologie Solvay, 116 p.

Adresse de l'auteur : E. DEFRISE-GUSSENHOVEN
Vrije Universiteit Brussel
Pleinlaan 2
B-1050 Brussel.