

Incidence de la sélection et de l'homogamie sur les paramètres du modèle génétique additif

B Langlois

*Institut National de la Recherche Agronomique, Station de Génétique Quantitative
et Appliquée, Centre de Recherches de Jouy-en-Josas, 78350 Jouy-en-Josas, France*

(Reçu le 17 mai 1988; accepté le 2 novembre 1989)

Résumé – Dans le cadre du modèle génétique additif l'article s'attache à exprimer la variance génétique lorsque la population est soumise à l'action simultanée de la sélection et de l'homogamie. Quelques résultats classiques ont été généralisés et relativisés comme "l'existence" et la valeur de la variance génétique "à l'équilibre" ainsi que l'expression de la variance d'erreur du modèle $A_d = (A_p + A_m)/2 + e$. L'incidence du linkage et d'une intensité de sélection différente dans les 2 sexes a été discutée. L'effet simultané de la consanguinité a été aussi examiné mais en négligeant l'effet du linkage. Cet inventaire des problèmes en suspens laisse apparaître la bonne robustesse des modèles d'estimation de la valeur génétique vis-à-vis des écarts à la panmixie lorsque le nombre de gènes est grand et qu'ils sont peu liés les uns aux autres.

homogamie / sélection / variance génétique / variance intrafamille

Summary – **Effects of selection and assortative mating on parameters of the additive genetic model** – *In the framework of the classical additive genetic model, this paper is concerned with expression of genetic variance when the population is submitted to the simultaneous effects of selection and non-random mating. Classical results have been generalized, and the "existence" and value of the "equilibrium" variance are discussed. The expression of the residual variance of the model $A_0 = (A_s + A_d)/2 + e$ in the case of linkage, and when selection intensity is unequal in the 2 sexes, is also given. The simultaneous effect of inbreeding is also considered, but only on neglecting linkage. This screening of solved and unsolved problems reveals that usual genetic prediction models are robust enough vis-à-vis deviations from panmixia when there is a great number of genes not closely linked.*

homogamy / selection / genetic variance / intrafamily variance

INTRODUCTION

Pour un caractère additif gouverné par un grand nombre de gènes la variance génétique est modifiée par des covariances entre effets de gènes lorsque l'on s'écarte de la panmixie:

– covariances positives entre effets de gènes situés au même locus quand il y a consanguinité (Wright, 1921a);

- covariances positives entre effets de gènes situés en grande majorité à des loci différents quand il y a homogamie (Fisher, 1918; Wright, 1921b, Crow et Felsenstein, 1968);
- covariances négatives entre effets de gènes situés en grande majorité à des loci différents quand il y a de la sélection.

Dans ces 2 derniers cas, il apparaît un déséquilibre de linkage; excès de combinaisons génétiques hautement favorables et défavorables par rapport à un déficit des combinaisons intermédiaires en homogamie; situation inverse dans le cas de la sélection où l'élimination des associations hautement défavorables entraîne un excès des assemblages intermédiaires par rapport aux arrangements hautement favorables.

Nous nous proposons d'aborder l'action conjointe de ces 3 écarts à la panmixie sur la variance génétique et par conséquent sur les paramètres génétiques, en partant de la formulation de ces problèmes proposée par Crow et Felsenstein (1968) et reprise dans l'ouvrage de Crow et Kimura (1970, p 158).

Dans un premier temps nous examinons la situation au niveau des valeurs génétiques ce qui permet d'éviter les problèmes de convergence soulevés par Feldman et Cavalli-Sforza (1979). Le passage au phénotype est abordé dans un second temps.

Le modèle génétique utilisé suppose l'additivité des effets d'un grand nombre de gènes. A l'opposé de l'approche de Lande (1977) la mutation est ignorée. Nous supposons en général malgré la sélection une relative stabilité des fréquences géniques qui peut être admise sur quelques générations lorsque le nombre de gènes impliqués dans le déterminisme du caractère est grand (Crow et Kimura, 1970, p 236).

EXPRESSIONS DE LA VARIANCE GÉNÉTIQUE ADDITIVE

Selon le modèle additif, la valeur génétique A d'un individu est la somme des contributions additives de chacun des allèles dont il se trouve porteur au niveau des n loci qui contrôlent le caractère mesuré:

$$A = \sum_{i=1}^n (a_i + a_i^*) \quad (1)$$

où a_i et a_i^* sont les effets moyens des 2 allèles hérités respectivement du père et de la mère, présents au locus i chez l'individu.

Dans ces conditions

$$\text{var}(A) = \sum_{i,j} \text{cov}(a_i, a_j) + \sum_{i,j} \text{cov}(a_i^*, a_j^*) + \sum_{i,j} \text{cov}(a_i, a_j^*) + \sum_{i,j} \text{cov}(a_i^*, a_j) \quad (2a)$$

$$\text{soit } \text{var}(A) = \text{var}(g) + \text{var}(g^*) + 2\text{cov}(g, g^*) \quad (2b)$$

où g et g^* représentent les valeurs gamétiques paternelles et maternelles.

En reprenant les notations de Crow et Felsenstein (1968) on peut écrire:

$$\text{var}(g) = \sum_{i,j} k_{i,j} \sigma_i \sigma_j \quad (3a)$$

$$\text{var}(g^*) = \sum_{i,j} k_{ij}^* \sigma_i^* \sigma_j^* \quad (3b)$$

$$\text{cov}(g, g^*) = \sum_{i,j} m_{ij} \sigma_i \sigma_j^* \quad (3c)$$

formules dans lesquelles $\sigma_i^2 = \text{var}(a_i)$ et où les k_{ij} et m_{ij} représentent les corrélations des effets des allèles portés aux loci i et j au sein d'un même gamète parental (k_{ij}) ou entre gamètes parentaux (m_{ij}).

On peut rapprocher ces notations de celles de Lande (1977) qui considère les covariances au lieu des corrélations. On note alors:

$$k_{ij} \sigma_i \sigma_j = C_{ij} \text{ et } m_{ij} \sigma_i \sigma_j^* = C'_{ij}$$

si pour tout indice d , p ou m rapportant à un descendant, un père ou une mère on peut supposer:

$$m_{ij,d} \sigma_{i,d} \sigma_{j,d} = m_{ij} \sigma_i \sigma_j^* / 2 + m_{ij} \sigma_i^* \sigma_j / 2 \quad (4a)$$

et

$$k_{ij,d} \sigma_{i,d} \sigma_{j,d} = k_{ij} \sigma_i \sigma_j / 2 + k_{ij}^* \sigma_i^* \sigma_j^* / 2 \quad (5a)$$

Les équations suivantes résultent alors des expressions (2) et (3):

$$\text{var}(A_p) = 2 \sum_{i,j} k_{ij,p} \sigma_i \sigma_j + 2 \sum_{i,j} m_{ij,p} \sigma_i \sigma_j \quad (6)$$

$$\text{var}(A_m) = 2 \sum_{i,j} k_{ij,m} \sigma_i^* \sigma_j^* + 2 \sum_{i,j} m_{ij,m} \sigma_i^* \sigma_j^* \quad (7)$$

$$\text{cov}(A_p, A_m) = 2 \sum_{i,j} m_{ij} \sigma_i \sigma_j^* + 2 \sum_{i,j} m_{ij} \sigma_i^* \sigma_j \quad (8)$$

$$\text{var}(A_d) = 2 \sum_{i,j} k_{ij,d} \sigma_{i,d} \sigma_{j,d} + 2 \sum_{i,j} m_{ij,d} \sigma_{i,d} \sigma_{j,d} \quad (9)$$

sachant

$$\sigma_{i,d} \sigma_{j,d} = (\sigma_i \sigma_j + \sigma_i^* \sigma_j^*) / 2 \quad (10)$$

l'équation (4) suppose:

$$m_{ij,d} = m_{ij} \quad (4b)$$

et

$$\sigma_i \sigma_j + \sigma_i^* \sigma_j^* \approx \sigma_i \sigma_j^* + \sigma_i^* \sigma_j \quad (11)$$

soit $\forall i$

$$\sigma_i \approx \sigma_i^* \quad (12)$$

l'équation (5) implique en plus

$$k_{ij,d} = (k_{ij} + k_{ij}^*) / 2 \quad (5b)$$

si r_{ij} est le pourcentage de recombinaisons des loci i et j , en appliquant l'équation (11) aux parents (indices p et m) on obtient:

$$k_{ij} = (1 - r_{ij})k_{ij,p} + r_{ij}m_{ij,p} \quad (13)$$

$$k_{ij}^* = (1 - r_{ij})k_{ij,m} + r_{ij}m_{ij,m} \quad (14)$$

d'où il résulte d'après (5b)

$$k_{ij,d} = (1 - r_{ij})(k_{ij,p} + k_{ij,m})/2 + r_{ij}(m_{ij,p} + m_{ij,m})/2 \quad (15a)$$

Or les valeurs parentales (indices p et m) découlent par sélection des valeurs à la génération précédente (indice $d - 1$). En supposant la multinormalité des effets alléliques et du phénotype, Lande (1977) a montré que les covariances C_{ij} et C'_{ij} étaient modifiées de façon identique par la sélection. Cette propriété selon Aitken (1934) cité par Verrier (1989, p 29) serait vérifiée aussi dans des conditions plus générales que la multinormalité. Si l'on suppose aussi la stabilité des variances alléliques sur deux générations consécutives on peut écrire:

$$k_{ij,p} = k_{ij,d-1} + d_{ij} \text{ et } m_{ij,p} = m_{ij,d-1} + d_{ij} \quad (16a)$$

ainsi que

$$k_{ij,m} = k_{ij,d-1} + d_{ij}^* \text{ et } m_{ij,m} = m_{ij,d-1} + d_{ij}^* \quad (16b)$$

d'où d'après (15a)

$$k_{ij,d} = (1 - r_{ij})k_{ij,d-1} + r_{ij}m_{ij,d-1} + (d_{ij} + d_{ij}^*)/2 \quad (15b)$$

Si nous nous intéressons maintenant à la variance résiduelle du modèle habituel

$$A_d = (A_p + A_m)/2 + e \quad (17)$$

sachant

$$\text{var}(e) = \text{var}(A_d) - \text{var}(A_p)/4 - \text{var}(A_m)/4 - \text{cov}(A_p, A_m)/2 \quad (18a)$$

on peut déduire des expressions (6) à (9), (11) et (15a)

$$\begin{aligned} \text{var}(e) = & \sum_{i,j} (1 - 2r_{ij})(k_{ij,p} - m_{ij,p})(\sigma_i \sigma_j)/2 \\ & + \sum_{i,j} (1 - 2r_{ij})(k_{ij,m} - m_{ij,m})(\sigma_i \sigma_j)/2 \end{aligned} \quad (18b)$$

Si l'on tient compte des expressions (16) et (10) l'expression (18b) se simplifie encore:

$$\text{var}(e) = \sum_{i,j} (1 - 2r_{ij})(k_{ij,d-1} - m_{ij,d-1})\sigma_i \sigma_j \quad (18c)$$

L'ensemble de ces expressions suppose que l'approximation (11) est robuste vis à vis de différences importantes d'intensité de sélection dans les deux sexes. En posant $\sigma_i = \alpha_i \sigma_i^*$, on doit examiner l'effet des écarts à 1 des coefficients α_i . Pour $|\alpha_i - 1| < 0.10$ l'écart à l'égalité (11) n'est que de 1%, des différences allant jusqu'à 19% entre σ_i^2 et σ_i^{*2} ne génèrent donc qu'une erreur inférieure à 1%. Pour des valeurs de α_i supérieures à 0.80 (36% de différence entre σ_i^2 et σ_i^{*2}) l'erreur reste inférieure à 4%.

Par ailleurs, lorsque la population est grande et que le caractère est gouverné par un grand nombre de gènes les variations de fréquence des allèles à chaque locus

sont trop faibles pour modifier sensiblement la variance sur un cycle de sélection même si elles suffisent à modifier la moyenne. Une analyse de cette question qui était déjà très simplement évoquée par Lush (1945, p 141) est fournie par Crow et Kimura (1970, p 236-239).

L'approximation (11) exclut cependant des situations extrêmes où les différences entre σ_i^2 et σ_i^{*2} excèdent 50%. Dans ce cas si l'on remplace les conditions (4a) et (5a) par celles exprimées par (4b) et (5b) l'ensemble des autres expressions hormis (11) et (12) reste applicable. On peut alors montrer selon (9), (10) et (15a) que,

$$\begin{aligned} \text{var}(A_d) = & \sum_{i,j} (1 - r_{ij})(k_{ij,p} + k_{ij,m})(\sigma_i\sigma_j + \sigma_i^*\sigma_j^*)/2 \\ & + \sum_{i,j} r_{ij}(m_{ij,p} + m_{ij,m})(\sigma_i\sigma_j + \sigma_i^*\sigma_j^*)/2 \\ & + \sum_{i,j} m_{ij}(\sigma_i\sigma_j + \sigma_i^*\sigma_j^*) \end{aligned}$$

Sachant (17) on peut donc déduire selon (6), (7) et (8):

$$\begin{aligned} \text{var}(e) = & \sum_{i,j} (k_{ij,p} - m_{ij,m})[(1 - r_{ij})\sigma_i^*\sigma_j^* - r_{ij}\sigma_i\sigma_j]/2 \\ & + \sum_{i,j} (k_{ij,m} - m_{ij,p})[(1 - r_{ij})\sigma_i\sigma_j - r_{ij}\sigma_i^*\sigma_j^*]/2 \\ & + \sum_{i,j} m_{ij}(\sigma_i\sigma_j + \sigma_i^*\sigma_j^* - \sigma_i\sigma_j^* - \sigma_i^*\sigma_j) \end{aligned} \quad (18d)$$

D'une manière générale sachant $r_{ii} = 0$ et $k_{ii} = 1$, la variance intra famille de pleins frères $\text{var}(e)$ que nous qualifierons de variance d'erreur, ne dépend que de la variance génétique intralocus lorsqu'on peut négliger le linkage ($r_{ij} = 1/2$, $\forall i \neq j$) et les variations de fréquence alléliques entre sexes ($\sigma_i \approx \sigma_i^*$, $\forall i$). Sinon une proportion plus ou moins importante des covariances entre loci intervient également dans son expression selon le degré de linkage des gènes impliqués comme l'ont déjà fait remarquer Keightley et Hill (1987) dans le cas de la sélection et selon les écarts de fréquence alléliques entre les 2 sexes que l'on néglige le plus souvent, mais dont l'effet sur la variance a cependant été discuté par Verrier *et al* (1988a).

ÉVOLUTION DE LA VARIANCE GÉNÉTIQUE ADDITIVE AU COURS DES GÉNÉRATIONS

L'équation (15b) en posant

$$s_{ij,d} = (d_{ij} + d_{ij}^*)/2 \quad (19)$$

permet d'écrire:

$$k_{ij,n+1} = \sum_{d=0}^{d=n} (1 - r_{ij})^{n-d} (s_{ij,d+1} + r_{ij}m_{ij,d}) + (1 - r_{ij})^{n+1} k_{ij,0} \quad (20)$$

Les paramètres $s_{ij,d}$ et $m_{ij,d}$ définissent la structure de la phase parentale produisant la génération d . Leur connaissance au cours des générations permet avec celle de la parenté de décrire l'évolution de la variance additive par rapport à la situation panmictique de référence où $k_{ij,0} = s_{ij,0} = m_{ij,0} = 0$.

La différence $k_{ij,d+1} - k_{ij,d}$ peut alors s'écrire:

$$k_{ij,d+1} - k_{ij,d} = s_{ij,d+1} + r_{ij}m_{ij,d} - \sum_{n=1}^{n=d} r_{ij}(1 - r_{ij})^{d-n}(s_{ij,n} + r_{ij}m_{ij,n-1}) \quad (21)$$

Lorsque $d - n$ est suffisamment grand, la quantité $(1 - r_{ij})^{d-n}$ tend vers 0 exprimant l'influence décroissante avec le temps des états antérieurs sur celui à une génération déterminée.

Parallèlement la somme $\sum_{n=1}^{n=d} r_{ij}(1 - r_{ij})^{d-n}$ tend vers 1.

Ces valeurs peuvent être atteintes assez vite au bout de 6 à 7 générations pour des gènes indépendants ($r_{ij} \approx 0.5$). En revanche elles peuvent être très longues à obtenir pour des gènes très liés. En moyenne si l'on suit Bulmer (1974) en retenant un linkage moyen de 0.40, une dizaine de générations suffit à obtenir le résultat. Il faut donc que les quantités $s_{ij,d+1}$ et $m_{ij,d}$ puissent être supposées à peu près constantes pendant ce nombre de générations pour que la différence $k_{ij,d+1} - k_{ij,d}$ s'annule. Ceci évidemment postule la convergence vers un état d'équilibre dont l'existence contestée par Feldman et Cavalli-Sforza (1979) a pu cependant être démontrée par Bulmer (1980, p 149) pour la sélection et p 127 pour l'homogamie, dans le cas de loci indépendants et d'une variance d'erreur constante.

Plus généralement mais avec un modèle d'hérédité phénotypique qui postule des coefficients de transmission constants de génération en génération Karlin (1979) démontre également l'existence d'un tel équilibre.

Cet équilibre comme le souligne Chevalet (1988) ne peut cependant être envisagé qu'en supposant l'absence d'effet de la sélection sur les variances alléliques. Sinon, sans mutation, la variance génétique tend obligatoirement vers zéro par fixation des allèles. C'est donc un cas d'école qui ne peut être envisagé que sur une échelle de générations relativement limitée, lorsque le progrès génétique croît linéairement. Cela suppose une population suffisamment grande et un caractère gouverné par un grand nombre de gènes.

Dans notre cas il faut supposer que les quantités $s_{ij,d+1}$ et $m_{ij,d}$ convergent sous l'action du processus de sélection et d'accouplements pour qu'un tel équilibre soit atteint. Dans ce cas, l'équation (15b) entraîne l'égalité suivante entre, \widehat{k}_{ij} , \widehat{m}_{ij} , \widehat{s}_{ij} , valeurs de k_{ij} , m_{ij} , s_{ij} à l'équilibre, supposé:

$$\widehat{k}_{ij} - \widehat{m}_{ij} = \widehat{s}_{ij}/r_{ij} \quad (\text{pour } i \neq j) \quad (24)$$

Remarquons aussi que l'expression (24) exprime un équilibre de recombinaison conditionné par la variance

$$V_{0,d} = 2 \sum_{i=1}^n \sigma_{i,d}^2 \quad (25)$$

qui se définit comme la variance génétique qui serait réalisée en panmixie avec les mêmes fréquences alléliques. Comme Verrier (1989, p 20) nous qualifierons cette variance de variance génique pour la différencier de la variance génétique qui inclut l'effet des déséquilibres de linkage. Si $V_{0,d}$ évolue lentement au cours des générations on peut supposer pour des valeurs de d suffisamment grandes

$$k_{ij,d-1} - m_{ij,d-1} \approx s_{ij,d}/r_{ij} \quad (26)$$

S'il paraît difficile de démontrer l'existence d'un état d'équilibre par l'expression de règles de récurrence entre $m_{ij,d}$ et $m_{ij,d-1}$ ainsi qu'entre $s_{ij,d}$ et $s_{ij,d-1}$, il est relativement aisé de le décrire s'il existe comme le font la majorité des auteurs. On peut aussi proposer un suivi expérimental des variances phénotypique et génétique qui permet de vérifier si l'on est en état stable ou pas.

ANALYSE AU NIVEAU PHÉNOTYPIQUE

Si l'héritabilité chez les pères h_p^2 est différente de celle réalisée dans leur population d'origine on peut quand même, compte tenu de la multinormalité des valeurs génétiques et phénotypiques avant sélection, écrire les équations de régression suivantes:

$$A_{d-1} = h_{d-1}^2 P_{d-1} + \varepsilon \quad (27a)$$

qui n'est pas modifiée après sélection des pères, soit

$$A_p = h_{d-1}^2 P_p + \varepsilon \quad (27b)$$

d'où l'on peut déduire comme Bulmer (1971) d'après Pearson (1903)

$$dV_{A_p} = \text{var}(A_p) - \text{var}(A_{d-1}) = h_{d-1}^4 (V_p - V_{d-1}) \quad (28)$$

où P_p et P_{d-1} sont les phénotypes des pères et des descendants à la génération $d-1$; V_p et V_{d-1} les variances phénotypiques correspondantes, ε une erreur aléatoire indépendante des phénotypes et qui n'est donc pas modifiée par la sélection des pères.

On déduit ainsi

$$h_{d-1}^4 (V_p - V_{d-1}) = h_p^2 V_p - h_{d-1}^2 V_{d-1}$$

d'où en posant

$$V_p = (1 - s_p) V_{d-1} \quad (29)$$

$$h_p^2 = h_{d-1}^2 (1 - s_p h_{d-1}^2) / (1 - s_p) \quad (30)$$

L'étude de l'expression (30) pour différentes valeurs de h_{d-1}^2 et de s_p montre que h_p^2 peut dépasser la valeur 1. On remarque aussi que pour des valeurs de h_{d-1}^2 comprises entre 0 et 1 la valeur de h_p^2 est toujours supérieure à celle de h_{d-1}^2 .

L'interprétation que l'on peut donner à ce phénomène est qu'il se crée du fait de la sélection une covariance négative entre les effets génétiques et ceux du milieu, selon les mêmes principes que ceux qui entraînent "l'effet Bulmer" entre effets génétiques. De ce fait la régression de la valeur génétique A_p en son phénotype P_p que l'on peut mesurer selon Langlois (1981) par le double du coefficient de régression du descendant en son père à mère constante ($b_{d/p,m}$) n'est plus égale à l'héritabilité

chez les pères (h_p^2), mais à celle réalisée à la génération $d - 1$ (h_{d-1}^2). Il s'est en effet introduit chez les pères une covariance entre effets génétiques (A_p) et le reste (R_p) que l'on peut évaluer en appliquant les équations de Pearson (1903) à A_p et R_p . On peut montrer que

$$\text{cov}(A_p, R_p) = h_{d-1}^2(1 - h_{d-1}^2)(V_p - V_{d-1}) \quad (31)$$

Le même raisonnement peut bien sûr être conduit pour la voie maternelle et implique les identités suivantes:

$$h_{d-1}^2 = 2b_{d/p.m} = 2b_{d/m.p} \quad (32)$$

en l'absence de corrélations entre les effets du milieu chez les descendants et les phénotypes parentaux.

On dispose donc à tout moment des moyens pour suivre expérimentalement les évolutions de la variance génétique qui se traduisent si les effets de milieu sur les descendants sont indépendants des phénotypes parentaux par la variation des coefficients de régression partiels $b_{d/p.m}$ et $b_{d/m.p}$, cela en opposition semble-t-il, avec le modèle de Karlin (1979).

En revanche, il n'est pas possible de modéliser l'évolution de la variance en fonction des paramètres de la sélection et des accouplements que si l'on connaît la variance d'erreur. En effet dans le cas général la connaissance de h_{d-1}^2 et donc de h_p^2 et de h_m^2 ne permet de déduire celle de h_d^2 que si la variance d'erreur est connue.

L'expression de cette variance sous diverses hypothèses revêt donc un aspect essentiel non seulement pour l'expression des paramètres génétiques mais aussi pour les problèmes d'évaluation des reproducteurs tels qu'ils ont été abordés par Sørensen et Kennedy (1984).

EXPRESSIONS DE LA VARIANCE D'ERREUR

La formule (18d) donne la structure de cette variance dans le cas général qui, en posant $\forall i, \sigma_i = \alpha_i \sigma_i^*$, $k_{ii,p} = k_{ii,m} = 1$ et $r_{ii} = 0$, peut aussi s'écrire:

$$\begin{aligned} \text{var}(e) = & \sum_i (1 - m_{ii,p}) \alpha_i \alpha_j \sigma_i^{*2} / 2 + \sum_i (1 - m_{ii,m}) \sigma_i^{*2} / 2 \\ & + \sum_{i \neq j} (k_{ij,p} - m_{ij,p}) [1 - (1 + \alpha_i \alpha_j) r_{ij}] \sigma_i^* \sigma_j^* / 2 \\ & + \sum_{i \neq j} (k_{ij,m} - m_{ij,m}) [1 - (1 + \alpha_i \alpha_j) r_{ij}] \sigma_i^* \sigma_j^* / 2 \\ & + \sum_{i,j} m_{ij} (1 - \alpha_i) (1 - \alpha_j) \sigma_i^* \sigma_j^* \end{aligned} \quad (35)$$

La première hypothèse qui est faite par quasiment tous les auteurs est de supposer $\alpha_i \approx 1, \forall i$, ce qui permet de faire disparaître le terme en $m_{ij,d}$. Nous avons vu que cette hypothèse était supportable dans la majorité des cas.

La seconde hypothèse consiste à faire disparaître le terme qui dépend du linkage en supposant $r_{ij} = 1/2, \forall i \neq j$. Remarquons que dans le cas où l'homogamie intervient seule sans la sélection (Crow et Felsenstein, 1968) les k_{ij} et m_{ij} tendent

vers une même limite en s'homogénéisant par le jeu des recombinaisons. L'effet du linkage sur la variance d'erreur disparaît alors au bout de quelques générations.

La troisième hypothèse enfin, conduit à négliger les termes en m_{ii} , et permet de simplifier encore l'expression de la variance d'erreur.

- Sous les 2 premières hypothèses: $\sigma_i = \sigma_i^*$, $r_{ij} = 1/2 \forall i \neq j$ et $m_{ii,d-1} = (m_{ii,p} + m_{ii,m})/2$ on obtient:

$$\text{var}(e) = \sum_{i=1}^n \sigma_i^2 - \sum_{i=1}^n m_{ii,d-1} \sigma_i^2 \quad (36)$$

En présence de consanguinité, on peut remarquer que:

$$m_{ii,d-1} = \varphi_{d-1} + (1 - \varphi_{d-1})a_{ii,d-1} \quad (37)$$

où φ_{d-1} est le coefficient de consanguinité et $a_{ii,d-1}$ la corrélation entre effets de gènes homologues au locus i induite par la sélection et/ou l'homogamie indépendamment de la consanguinité.

Des formules (36) et (37) on tire

$$\text{var}(e) = \sum_{i=1}^n \sigma_i^2 - \sum_{i=1}^n [\varphi_{d-1} + (1 - \varphi_{d-1})a_{ii,d-1}] \sigma_i^2$$

soit, en factorisant $(1 - \varphi_{d-1})$

$$\text{var}(e) = (1 - \varphi_{d-1}) \left[\sum_{i=1}^n \sigma_i^2 (1 - a_{ii,d-1}) \right]$$

ou, d'après (25) et après avoir posé

$$f_{a,d-1} = \frac{\sum_{i=1}^n a_{ii,d-1} \sigma_i^2}{\sum_{i=1}^n \sigma_i^2}$$

$$\text{var}(e) = V_0(1 - \varphi_{d-1})(1 - f_{a,d-1})/2 \quad (38)$$

En l'absence de consanguinité on obtient l'expression proposée par Rogers (1983)

$$\text{var}(e) = V_0(1 - f_{a,d-1})/2 \quad (39)$$

De la même façon si la population est consanguine $\varphi_{d-1} = (\varphi_p + \varphi_m)/2$ et en l'absence des effets de l'homogamie et de la sélection on retrouve l'expression proposée par Foulley et Chevalet (1981)

$$\text{var}(e) = V_0(1 - \varphi_{d-1})/2 \quad (40)$$

- L'expression classique de Bulmer (1980, p 125) généralement utilisée ajoute la troisième hypothèse aux 2 précédentes en supposant $f_{a,d-1} = 0$ et $\varphi_{d-1} = 0$ soit:

$$\text{var}(e) = V_0/2 \quad (41)$$

En effet si le nombre de gènes impliqués dans le déterminisme du caractère est important, l'effet intra locus de l'homogamie et/ou de la sélection devient négligeable en regard de l'effet entre locus et $f_{a,d-1}$ est voisin de zéro. Cela confère aux expressions de Bulmer (1980, p 158) et de Foulley et Chevalet (1981) une assez grande généralité (Verrier *et al*, 1988a et b). Ce n'est en revanche pas le

cas si le nombre de gènes impliqués est faible. Remarquons aussi que dans le cas, homogamie et consanguinité interagissent de manière complexe. Wright (1921b), repris par Crow et Felsenstein (1968), montre en effet l'équivalence à la limite de ces 2 écarts à la panmixie lorsqu'un seul gène gouverne le caractère.

– Lorsqu'on ne peut pas faire la seconde hypothèse $r_{ij} = 1/2 \forall i \neq j$ l'expression (35) ne se simplifie pas aussi facilement et il faut alors postuler l'existence d'un état d'équilibre; vrai, si dans (25) $V_{0,d}$ peut être considéré comme stable sur la période considérée; dynamique si l'on peut seulement supposer réalisée l'expression (26) soit l'équilibre des recombinaisons.

Soit A , B et C , les 3 parties de la formule (35) $\text{var}(e) = A + B + C$, les calculs précédents donnent:

$$A = V_0(1 - \varphi_{d-1})(1 - f_{a,d-1})/2 \quad (42)$$

En introduisant la quantité:

$$\rho = \sum_{i \neq j} r_{ij}(k_{ij,d-1} - m_{ij,d-1})\sigma_i\sigma_j / \sum_{i \neq j} (k_{ij,d-1} - m_{ij,d-1})\sigma_i\sigma_j \quad (43)$$

l'expression B s'écrit:

$$B = \left(\frac{1}{2\rho} - 1\right)/2 \sum_{i \neq j} r_{ij}(k_{ij,d-1} - m_{ij,d-1})\sigma_i\sigma_j$$

d'après (26) et (19) on peut supposer $\forall i \neq j$ après installation d'un équilibre de recombinaison

$$r_{ij}(k_{ij,d-1} - m_{ij,d-1}) \approx (d_{ij} + d_{ij}^*)/2$$

d'où sachant selon (16a et b), $d_{ii} = d_{ii}^* = 0$ par définition des k_{ij} :

$$B \approx \left(\frac{1}{2\rho} - 1\right) \sum_{i \neq j} (d_{ij} + d_{ij}^*)\sigma_i\sigma_j$$

D'après (6) et (16a) on peut écrire:

$$\text{var}(A_p) = 2 \sum_{i,j} k_{ij,d-1}\sigma_i\sigma_j + 2 \sum_{i,j} m_{ij,d-1}\sigma_i\sigma_j + 4 \sum_{i,j} d_{ij}\sigma_i\sigma_j$$

qui en supposant $\sigma_i = \beta_i\sigma_{i,d-1} \forall i$ ou plus généralement en posant

$$\beta^2 = \sum_{i,j} (k_{ij,d-1} + m_{ij,d-1})\sigma_i\sigma_j / \sum_{i,j} (k_{ij,d-1} + m_{ij,d-1})\sigma_{i,d-1}\sigma_{j,d-1} \quad (44)$$

donne

$$\text{var}(A_p) = \beta^2 \text{var}(A_{d-1}) + 4 \sum_{i,j} d_{ij}\sigma_i\sigma_j$$

d'où il résulte selon (28) si $\beta^2 \approx 1$

$$4 \sum_{i,j} d_{ij}\sigma_i\sigma_j = dV_{A_p} \quad (45a)$$

de la même façon sous des hypothèses identiques:

$$4 \sum_{i,j} d_{ij}^* \sigma_i^* \sigma_j^* = dV_{A_m} \quad (45b)$$

En posant

$$dV_A = (dV_{A_p} + dV_{A_m})/2 \quad (46)$$

on obtient alors:

$$B \approx \left(\frac{1}{2\rho} - 1 \right) \frac{1}{2} dV_A \quad (47)$$

$$\text{où } dV_A = h_{d-1}^4 [(V_p + V_m)/2 - V_{d-1}] \quad (48)$$

Par ailleurs l'expression

$$C = \sum_{i,j} m_{ij,d-1} (\sigma_i \sigma_j + \sigma_i^* \sigma_j^* - \sigma_i^* \sigma_j - \sigma_i \sigma_j^*)$$

s'annule pour $\alpha_i \approx \alpha_i^*$

Dans le modèle infinitésimal $\alpha_i = \beta_i = 1 \forall i$ et $f_{a,d-1}$ étant voisin de zéro l'expression générale de la variance d'erreur devient:

$$\text{var}(e) = \frac{1}{2} V_0 \left(1 - \frac{\varphi_p + \varphi_m}{2} \right) + \left(\frac{1}{2\rho} - 1 \right) dV_A/2 \quad (49)$$

où V_0 est la variance d'équilibre supposée stable; $(\varphi_p + \varphi_m)/2 = \varphi_{d-1}$ est le coefficient de consanguinité de la génération parentale; ρ est un pourcentage moyen de recombinaison pondéré des différences ($k_{ij,d-1} - m_{ij,d-1}$) pour chaque couple de loci; ρ n'est fixé que si le poids relatif de ces différences pour chaque couple de loci est stable. dV_A exprime la variation de la variance génétique additive réalisée chez les parents du fait de leur sélection. Elle peut être déduite des variations phénotypiques observées. Elle n'est stable que si la population a atteint son niveau d'équilibre.

EXAMEN DES VARIANCES GÉNÉTIQUES "À L'ÉQUILIBRE" DANS LE CAS DU MODÈLE INFINITÉSIMAL

Selon (17) et (28)

$$\text{var}(A_d) = \text{var}(A_{d-1})/2 + (dV_{A_p} + dV_{A_m})/4 + \text{cov}(A_p, A_m)/2 + \text{var}(e)$$

En remplaçant $\text{var}(e)$ par sa valeur selon (49) on obtient après quelques simplifications:

$$\text{var}(A_d) = \frac{1}{2} V_{0,d} \left(1 - \frac{\varphi_p + \varphi_m}{2} \right) + \frac{1}{2} \text{var}(A_{d-1}) + \frac{1}{2} \text{cov}(A_p, A_m) + \frac{1}{4\rho} dV_A$$

à l'équilibre, $\text{var}(A_d) = \text{var}(A_{d-1}) = \text{var}(\hat{A}_d)$ et,

$$\text{var}(\hat{A}_d) = V_0 \left(1 - \frac{\varphi_p + \varphi_m}{2} \right) + \text{cov}(\hat{A}_p, \hat{A}_m) + \frac{1}{2\rho} d\hat{V}_A$$

en posant $\text{cov}(\hat{A}_p, \hat{A}_m) = R_A \text{var}(\hat{A}_d)$ et $d\hat{V}_A = -S_A \text{var}(\hat{A}_d)$ (pour que $d\hat{V}_A$ étant négatif, S_A soit positif) on obtient:

$$\text{var}(\hat{A}_d) = \frac{\left(1 - \frac{\varphi_p + \varphi_m}{2}\right) V_0}{1 + \frac{1}{2\rho} S_A - R_A}$$

expression qui généralise celle de Crow et Felsenstein (1968) qui avait déjà été exprimée par Wright (1921b):

$$\text{var}(\hat{A}_d) = V_0 / (1 - R_A)$$

lorsque l'homogamie est seule à intervenir, $S_A = 0$ et en l'absence de consanguinité $(\varphi_p + \varphi_m)/2 = 0$.

Dans les cas habituels avec $0 < S_A < 1$, $0 < \rho < 0.5$ et $-1 < R_A < +1$ le dénominateur sera tantôt plus petit que 1 si les effets de l'homogamie ($R_A > 0$) l'emportent sur ceux de la sélection ($R_A > S_A/2\rho$), la variance observée à l'équilibre sera alors grandie; elle sera réduite si c'est l'inverse ($S_A/2\rho > R_A$).

Nous voyons qu'il est difficile de prévoir ce que donnera la variance d'équilibre car la réponse dépend en grande partie d'une donnée inconnue qui est le linkage moyen des gènes impliqués. Plus ce dernier est important plus l'effet de la sélection est renforcé par rapport à celui de l'homogamie qui selon le signe de R_A peut agir soit dans le même sens ($R_A < 0$) soit en sens opposé ($R_A > 0$). En effet il apparaît que sélection et linkage sont des phénomènes étroitement associés, que l'expression de la variance d'erreur est peut être en mesure de nous éclairer. Plus le pourcentage moyen de recombinaisons ρ est faible et plus la variance d'erreur est réduite en fonction de la diminution de variance génétique réalisée chez les reproducteurs. Rapportée au niveau d'un individu la sélection entraîne donc un effet de "prépotence" d'autant plus fort que le linkage moyen des gènes impliqués dans la détermination du caractère est élevé. Un reproducteur pour lequel les gènes sélectionnés sont très liés voit donc sa variance d'erreur d'échantillonnage diminuer, il produit avec plus de régularité et son succès reproducteur s'en trouve accru d'autant. On comprend ainsi comment peu à peu les gènes impliqués dans le même processus de sélection en viennent à être liés les uns aux autres.

CONCLUSION

L'action simultanée de la sélection et de l'assortiment non aléatoire des reproducteurs sur les paramètres du modèle génétique additifs, pour courante qu'elle soit dans la pratique, est encore assez peu abordée dans la littérature. L'effet de la sélection ainsi que de l'homogamie ou des croisements consanguins agissant seuls ont en revanche déjà fait l'objet d'études approfondies. La plupart d'entre elles supposent l'absence de linkage. En nous appuyant sur ces travaux nous nous sommes attachés à en proposer ici des généralisations. Elles excluent cependant pour l'instant le cas où le linkage moyen n'étant pas négligeable une certaine consanguinité interfère avec la sélection et l'homogamie. L'effet de consanguinité n'a été envisagé qu'en fonction de son incidence sur la variance intra locus. Cela peut suffire en l'absence

de linkage mais risque en sa présence de compliquer beaucoup les expressions de dV_A et de $\text{cov}(A_p, A_m)$ en fonction de paramètres phénotypiques mesurables.

Néanmoins d'ores et déjà, le résultat de cet examen conduit à considérer différemment la variance génétique additive et les paramètres classiques comme l'héritabilité. En effet, entrent dans la variance génétique et ce paramètre non seulement la variabilité allélique de la population considérée mais aussi une grande part d'associations de gènes induites par la consanguinité, l'homogamie et la sélection.

Ce tour d'horizon a permis de généraliser et de relativiser un certain nombre de résultats classiques comme "l'existence" et la valeur de la variance génétique "à l'équilibre" et l'expression de la variance d'erreur $\text{var}(e)$ en cas de linkage et lorsque l'intensité de sélection est différente dans les 2 sexes.

Cet inventaire des problèmes en suspens devrait permettre un abord plus aisé de ces questions trop souvent passées sous silence et qui ont pourtant un impact non négligeable sur les méthodes de sélection (Dempfle, 1987) et sur les techniques d'évaluation génétique (Sørensen et Kennedy, 1984).

Soulignons toutefois que ces dernières apparaissent bien adaptées lorsque le nombre de gènes est grand et qu'ils sont indépendants. Dans ce contexte qualifié souvent de modèle infinitésimal la nécessité présentée par Karlin (1979) de remplacer le modèle génétique additif classique par des modèles purement statistiques d'hérédité phénotypique n'apparaît pas encore évidente aux zootechniciens.

RÉFÉRENCES

- Bulmer MG (1971) The effect of selection on genetic variability. *Am Naturalist* 105, 201-211
- Bulmer MG (1974) Linkage disequilibrium and genetic variability. *Genet Res Camb* 23, 281-289
- Bulmer MG (1980) *The Mathematical Theory of Quantitative Genetics*. Clarendon Press Oxford, 254 pp
- Chevalet C (1988) Control of genetic drift in selected population. In: *Proceedings of the 2nd International Conference on Quantitative Genetics*, Raleigh, 31 May - 5 June 1987, Sinauer Associates Inc., Sunderland, USA, 379-394
- Crow JF, Felsenstein J (1968) The effect of assortative mating on the genetic composition of a population. *Eugenics Quart* 15, 85-97
- Crow JF, Kimura M (1970) *An Introduction to Population Genetics Theory*. Harper and Row, New York, p 591
- Dempfle L (1987) Statistical aspects of design of animal breeding programs - a comparison among various selection strategies. In: *Advances in Statistical Methods for Genetic Improvement of Livestock* (Gianola D, Hammond K, eds) Int Symp, Armidale, 16-20 February 1987 Animal Genetics and Breeding Unit of New England, Australia, 109-128
- Feldman MW, Cavall-Forza LL (1979) Aspects of variance and covariance analysis with cultural inheritance. *Theor Popul Biol* 15, 276-307

- Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinburgh* 52, 399-433
- Foulley JL, Chevalet C (1981) Méthode de prise en compte de la consanguinité dans un modèle simple de simulation de performances. *Ann Génét Sél Anim* 13, 189-196
- Karlin S (1979) Models of multifactorial inheritance: I. Multivariate formulations and basic convergence results. *Theor Popul Biol* 15, 308-355
- Keightley PD, Hill WG (1987) Directional selection and variation in finite populations. *Genetics* 117, 573-582
- Lande R (1977) The influence of the mating system on the maintenance of genetic variability in polygenic characters. *Genetics* 86, 485-498
- Langlois B (1981) Proposition d'extension du modèle génétique additif classique à des situations non panmictiques. *Ann Génét Sél Anim* 13, 151-164
- Lush JL (1945) *Animal Breeding Plans*. Iowa State University Press, Ames, 3rd edn., p 443
- Pearson K (1903) Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Phil Trans R Soc London, Ser A*, 200, 1-66
- Rogers A (1983) Assortative mating and the segregation variance. *Theor Popul Biol* 23, 110-113
- Sörensen DA, Kennedy BW (1984) Estimation of response to selection using least-squares and mixed model methodology. *J Anim Sci* 58, 1097-1106
- Verrier E (1989) Prédiction de l'évolution de la variance génétique dans les populations animales d'effectif limité soumises à la sélection. *Thèse INAPG*, Paris, 259 pp
- Verrier E, Colleau JJ, Foulley JL (1988a) Effect of mass selection on the within-family genetic variance in finite populations. *Theor Appl Genet* 77, 142-148
- Verrier E, Colleau JJ, Foulley JL, Ducrocq V (1988b) Predicting the additive genetic variance in finite populations undergoing mass selection. (ADSA 83rd Annual Meeting, 26-29 June 1988, Edmonton, Alberta) *J Dairy Sci* 71 (Suppl. 1), P434, 263
- Wright S (1921a) Systems of mating. II. The effect of inbreeding on the genetic composition of a population. *Genetics* 6, 124-143
- Wright S (1921b) Systems of mating. III. Assortative mating based on somatic resemblance. *Genetics* 6, 144-161