

-4 NOV. 1983



O. R. S. T. O. M. Fonds Documentaire

N° : 3634 ex 1

Cote : B

Une analyse génétique chez le petit mil pénicillaire à l'aide de la distance R de HANSON et CASAS

L. MARCHAIS

Office de la Recherche Scientifique et Technique Outre-Mer,
70-74, route d'Aulnay, 93140 Bondy

Résumé

Neuf lignées F_2 de petit mil (*Pennisetum typhoides*) issues d'une même population F_2 ont été étudiées par les deux ensembles parallèles de croisements : Triple et Diallèle.

L'analyse génétique utilise essentiellement les statistiques d'hétérosis et de distance R^2 (HANSON et CASAS), exprimées dans le modèle F_{∞} de Van der VEEN limité à des épistasies entre paires de loci. Les covariances entre loci sont prises en compte.

Onze caractères sur douze manifestent une très forte corrélation entre la distance et l'hétérosis. L'interprétation de ces corrélations conduit à corriger l'estimation des paramètres génétiques de l'analyse classique et à calculer une distance générale entre lignées à savoir le nombre de groupes de gènes pour lequel deux lignées diffèrent.

Le modèle génétique utilisé rend assez bien compte des mesures de distance et d'hétérosis mais explique mal la faible variance des lignées.

L'épistasie apparaît essentiellement de type hétérozygote \times hétérozygote opposé à la dominance.

Il apparaît que les distances entre lignées mesurées sur des croisements lignées \times testeurs permettent de prévoir l'hétérosis entre lignées. Ce résultat peut intéresser le sélectionneur.

Introduction

La distance R entre unités génotypiques définie par HANSON et CASAS (1968) et utilisée par ceux-ci sur le Maïs est apparue dans diverses expériences posséder des propriétés intéressantes (HANSON et MOLL, 1973; CASAS *et al.*, 1968) :

- la distance R était fortement corrélée avec l'hétérosis;
- elle a permis de séparer des populations qui d'après leur phénotype semblaient voisines;
- la validité de la distance R semblait diminuer au-delà d'une certaine divergence génétique entre unités taxonomiques observées à cause de ce que

O. R. S. T. O. M.

Collection de Référence



SSIBAPU

B3634 ex 1

22 DEC. 1978

3634
ex 1
B

ces auteurs ont appelé une chute des relations de dominance suivant en cela le phénomène déjà observé par MOLL *et al.* (1965) sur l'hétérosis.

Malgré ce défaut, cette distance semble avoir plus de signification génétique que bien d'autres distances taxonomiques utilisant aussi des caractères biométriques. Il est, en effet, logique de penser que la différence entre les gamètes de deux unités taxonomiques phénotypiquement voisines s'exprimera si elle existe, dans la différence entre les croisements de ces deux unités avec les mêmes testeurs.

A l'occasion d'un travail d'analyse génétique classique sur des lignées de Petit Mil (*Pennisetum typhoides*), nous avons cherché d'une part à observer les distances entre nos lignées pour mieux les caractériser et à voir si ces distances permettent de prévoir l'hétérosis entre lignées.

D'autre part, nous avons cherché un modèle génétique permettant de reconstituer les hétérosis et distances observées, ceci sur plusieurs caractères. Cette analyse a conduit à corriger l'analyse classique.

Dans la mesure où nous pouvions approcher les mécanismes génétiques de ces caractères, nous avons tenté de construire une distance globale entre lignées réunissant les informations tirées des distances particulières à ces caractères.

Le matériel génétique et les essais

Les données à analyser ont été recueillies sur un croisement diallèle et un croisement triple modifié (JINKS *et al.*, 1969) incluant les mêmes neuf lignées F_5 naines de Petit Mil issues de neuf pieds F_2 provenant d'un même hybride entre une lignée pure indienne naine 1472 et une lignée pure Ouest-africaine 1133, de type traditionnel.

Le croisement diallèle comprenait les autofécondations et les croisements dans un seul sens. Le croisement triple modifié comprenait pour chacune des neuf lignées l'autofécondation, le croisement avec le parent 1472 et le croisement avec l'autre parent 1133.

Les deux essais ont été cultivés à Bambey (Sénégal) pendant l'hivernage 1973, en blocs randomisés complets à deux répétitions de 5 pieds espacés à 1 m \times 1 m, soit sans compétition.

Les neuf lignées ont été sélectionnées au cours des cycles d'autofécondation essentiellement pour le gène de nanisme d_2 et la résistance au *Sclerospora*.

Les observations ont porté sur les 12 caractères suivants :

CY	cycle compté en nombre de jours du semis à l'épiaison,
LF	longueur en cm de la 3 ^e feuille au-dessous de la chandelle principale,
lf	largeur en mm de la 3 ^e feuille,
LT	longueur de la plus haute tige,
PT	périmètre du 3 ^e entre-nœud sous la chandelle principale,
PRO	longueur de la plus grande projection d'une talle sur le sol,
NC	nombre de chandelles du pied,
LC	longueur en cm de la chandelle principale,
PC	périmètre en mm de la chandelle principale,
NE	densité des épillets sur la chandelle,
P500	poids de 500 graines,
PG	poids de grain récolté sur le pied.

Les méthodes

1. — L'analyse classique (supposée connue)

Le croisement diallèle a été analysé suivant les hypothèses et le schéma de HAYMAN (1954), de tout temps critiqué par les généticiens américains, actuellement abandonné par la plupart des généticiens parce qu'il suppose des conditions sans doute jamais satisfaites simultanément : bi-allélisme, distributions indépendantes des gènes et absence d'épistasie. Toutefois à titre de comparaison avec les autres analyses, on a calculé les paramètres D , H_1 , H_2 et le nombre de gènes N même si le test de la regression Wr sur Vr indiquait le non respect des hypothèses. Rappelons que, si les hypothèses sont satisfaites : $D = \Sigma 4uvd^2$, $H_1 = \Sigma 4uvh^2$, $H_2 = \Sigma 16u^2v^2h^2$ d et h sont les effets d'additivité et dominance à un locus donné, u et v sont les fréquences respectives des 2 allèles au locus considéré, dans l'ensemble des parents du diallèle.

Le croisement triple modifié (JINKS *et al.*, 1969) comprend pour chacune des neuf lignées F issues de l'hybride entre les deux lignées initiales P et P' , l'auto-fécondation de valeur Y_f , le croisement avec le parent P de valeur Y_{pf} et le croisement avec le parent P' de valeur $Y_{p'f}$.

Sans hypothèses particulières, ce matériel permet de tester la présence d'épistasie en testant si :

$$Y_{pff} + Y_{p'f} - Y_f = \text{constante pour les 9 lignées}$$

Avec les hypothèses de HAYMAN (1954), les paramètres D et H_1 sont estimés par :

$$D = \text{variance sur } f \text{ de } (Y_{pf} + Y_{p'f}) = \Sigma_f \frac{(Y_{pf} + Y_{p'f})^2}{r} - \frac{[\Sigma_f (Y_{pf} + Y_{p'f})]^2}{r^2}$$

$$H_1 = \text{variance sur } f \text{ de } (Y_{pf} - Y_{p'f}) = \Sigma_f \frac{(Y_{pf} - Y_{p'f})^2}{r} - \frac{[\Sigma_f (Y_{pf} - Y_{p'f})]^2}{r^2}$$

avec $r = 9$.

Signalons que nous avons effectué sur ce croisement triple une analyse de variance en effets fixes et non comme JINKS *et al.* (1969) en modèle mixte, ceci afin de rester dans le même modèle que celui utilisé par HAYMAN (1954), avec le croisement diallèle.

2. — L'analyse des distances et de l'hétérosis

2.1. — Terminologie et formulaire

Précisons d'abord la terminologie employée et les hypothèses retenues.

Le modèle génétique :

Le modèle génétique retenu est le modèle F_{∞} de Van der VEEN (1959) limité à des épistasies entre paires de locus.

La valeur Ym d'un individu M pour un certain caractère s'exprime :

$$\begin{aligned}
 Ym = C + \Sigma_a \theta_{am} da + \Sigma_a (1 - \theta_{am}^2) ha + \Sigma_{a < a'} \theta_{am} \theta_{a'm} iaa' \\
 + \Sigma_{a < a'} \theta_{am} (1 - \theta_{a'm}^2) ja | a' + \Sigma_{a < a'} (1 - \theta_{am}^2) \theta_{a'm} ja' | a \\
 + \Sigma_{a < a'} (1 - \theta_{am}^2) (1 - \theta_{a'm}^2) laa'
 \end{aligned} \quad (1)$$

C désigne l'origine des valeurs génotypiques.

θ_{am} désigne la situation génotypique de l'individu M au locus a : les génotypes AA , Aa et aa sont représentées par les valeurs de θ_{am} 1, 0, -1.

Les effets i désignent les interactions digéniques de type homozygote \times homozygote, les effets j les interactions de type homozygote \times hétérozygote et les effets l les interactions de type hétérozygote \times hétérozygote.

Si le taux de recombinaison des gènes à la formation des gamètes vaut $1/2$ et si les différents gamètes de deux individus M et F s'unissent au hasard, la valeur $Yf \times m$ du croisement $F \times M$ s'écrit grâce à l'indépendance des distributions alléliques aux divers loci :

$$\begin{aligned}
 Yf \times m = C + \Sigma_a \frac{\theta_{af} + \theta_{am}}{2} da + \Sigma_a \frac{(1 - \theta_{af} \theta_{am})}{2} ha \\
 + \Sigma_{a < a'} (\theta_{af} + \theta_{am}) (\theta_{a'f} + \theta_{a'm}) iaa' / 4 \\
 + \Sigma_{a < a'} (\theta_{af} + \theta_{am}) (1 - \theta_{a'm} \theta_{a'f}) ja | a' / 4 \\
 + \Sigma_{a < a'} (1 - \theta_{af} \theta_{am}) (\theta_{a'f} + \theta_{a'm}) ja' | a / 4 \\
 + \Sigma_{a < a'} (1 - \theta_{af} \theta_{am}) (1 - \theta_{a'f} \theta_{a'm}) laa' / 4
 \end{aligned} \quad (2)$$

Dans la suite de cet article, on supposera que les différents effets sont les mêmes à tous les locus à savoir :

$$da = d, ha = h, iaa' = i, ja | a' = ja' | a = j, laa' = l$$

Structure génotypique des neuf lignées et de leurs parents

Les deux parents de la F_1 , source des 9 lignées, sont désignés par les indices p et p' .

Le nombre N de gènes qui les différencient peut s'exprimer sous la forme :

$$N = \Sigma_a \frac{(\theta_{ap} - \theta_{ap'})^2}{4}$$

Les neuf lignées issues de la F_1 par autofécondation sont désignées par l'indice m lorsqu'elles sont testées pour leur hétérosis ou leurs distances mutuelles et par l'indice f lorsqu'elles sont testeurs.

Le nombre n de gènes différents entre 2 lignées M et M' peut s'écrire :

$$n = \Sigma_a \frac{(\theta_{am} - \theta_{am'})^2}{4}$$

Les neuf lignées sont obtenues par des tirages indépendants de la F_1 .

On admet en outre que les différents gènes d'une lignée sont obtenus par tirages indépendants dans le réservoir génique de la F_1 .

Cette hypothèse implique au moment de la formation des gamètes des diverses générations d'autofécondation une liaison absolue de certains locus et une ségré-

gation indépendante de ces groupes de locus liés. A la limite, quand on parle ici de locus, il peut s'agir de chromosomes entiers.

Il en résulte que chaque paramètre θ_{af} peut être vu comme une variable aléatoire indépendante, d'espérance $E = 0$ et de variance $1 - E^2 = 1$.

Pour la même raison, on admettra que dans une lignée M particulière : $\Sigma_a \theta_{am} \neq 0$. Chaque lignée contient un nombre de gènes dominants homozygotes voisin du nombre de gènes homozygotes récessifs. On dira que les locus sont à l'état dispersé.

Les deux parents de la F_1 , par contre, peuvent se situer entre l'état de dispersion totale des locus et l'état d'association totale. Dans ce dernier cas, le parent P contient tous les gènes dominants et, le parent P' tous les gènes récessifs. Pour tout couple de locus a, a' on a la relation :

$$\theta_{ap} \times \theta_{a'p} = \theta_{ap'} \times \theta_{a'p'} = 1$$

Définition et calcul des hétérosis

L'hétérosis entre deux lignées M et M' est défini par :

$$He(m, m') = Ym \times m' - \frac{(Ym \times m + Ym' \times m')}{2}$$

Ici $Ym \times m = Ym$.

L'application des formules (1) et (2) donne :

$$He(m, m') = \Sigma_a (\theta_{am} - \theta_{am'})^2 h / 4 + \Sigma_{a < a'} (\theta_{am} - \theta_{am'}) (\theta_{a'm'} - \theta_{a'm}) i / 4 \\ + \Sigma_{a < a'} [(\theta_{am} + \theta_{am'}) (1 - \theta_{a'm} \theta_{a'm'}) + (1 - \theta_{am} \theta_{am'}) (\theta_{a'm} + \theta_{a'm'})] j / 4 \\ + \Sigma_{a < a'} [(\theta_{am} - \theta_{am'})^2 (\theta_{a'm} - \theta_{a'm'})^2] l / 16$$

Avec l'approximation $\Sigma_a \theta_{am} = 0$, on a les relations :

$$\Sigma_{a' \neq a} (\theta_{a'm'} - \theta_{a'm}) = \theta_{am} - \theta_{am'} \\ \Sigma_a \theta_{am} \theta_{am'} \Sigma_{a' \neq a} \theta_{a'm} = -\Sigma \theta_{am}^2 \theta_{am'} = 0$$

Alors les termes en j disparaissent et les termes en i deviennent

$$\frac{i}{2} \Sigma_a \frac{(\theta_{am} - \theta_{am'})^2}{4} = n \frac{i}{2}$$

En définitive :

$$He(m, m') = n \left(h + \frac{i}{2} + \frac{(n-1)}{2} l \right) \quad (3)$$

L'hétérosis entre P et P' les parents des lignées pourrait s'exprimer sous une forme différente si leurs locus étaient à l'état d'association totale.

Les termes en i deviennent alors :

$$\Sigma_{a < a'} \frac{(\theta_{ap} - \theta_{ap'}) (\theta_{a'p'} - \theta_{a'p})}{4} i = \Sigma_{a < a'} (-i) = -\frac{N(N-1)}{2} i$$

Les termes en j disparaissent parce que $\theta_{ap} + \theta_{ap'} = 0$.

$$\text{Ainsi } He(p, p' \text{ en association}) = N \left(h + \frac{N-1}{2} (l-i) \right) \quad (3 \text{ bis})$$

$$\text{Mais } He(p, p' \text{ en dispersion}) = N \left(h + \frac{i}{2} + \frac{N-1}{2} l \right) \quad (3 \text{ ter})$$

Notons ici qu'avec l'approximation $\Sigma \theta_{am} = 0$ les termes en j disparaissent dans tous les croisements entre lignées.

La formule (3) est approximative pour les termes en i et j .

La formule (3 bis) est exacte.

Définition et calcul des distances

Rappelons la définition de HANSON et CASAS donnée dans le cadre d'un croisement diallèle entre r lignées.

Soit $s_{it} = y_{it} - y_{i.} - y_{.t} + y_{..}$ l'aptitude spécifique à la combinaison.

La distance $R(i, j)$ entre 2 lignées est définie par la relation :

$$R^2(i, j) = \sum_{t=1}^{t=r} (s_{it} - s_{jt})^2$$

Cette relation peut aussi se lire $R^2(i, j) = \Sigma_t [(y_{it} - y_{i.}) - (y_{jt} - y_{j.})]^2$ ce qui est le numérateur de la variance des r différences $y_{it} - y_{jt}$.

Dans le présent travail, la définition de $R^2(i, j)$ a été légèrement modifiée :

$$R^2(i, j) = \text{Variance } (y_{it} - y_{jt}) \text{ sur l'indice variable } t$$

Cette définition se situe en modèle fixe, dans un cadre limité aux lignées i, j à tester et aux lignées t qui servent de testeurs.

Cette définition reste valable quand les lignées à tester sont différentes des lignées testeurs.

On a vu dans la présentation du croisement triple modifié que H_1 était estimé par la variance de $(y_{pf} - y_{p'f})$ pour l'indice f variant de 1 à g .

L'estimation H_1 du croisement triple n'est autre que $R^2(p, p')$ la distance au carré entre les parents des neuf lignées mesurée avec ces lignées comme testeurs.

La distance entre 2 lignées M et M' du diallèle est donnée par $R^2(m, m') = \text{Variance sur } f \text{ de } (Y_{mf} - Y_{m'f})$, F étant aussi une lignée.

Le détail des calculs est donné en annexe. On aboutit à l'étape intermédiaire approximative pour les termes en i et j :

$R^2(m, m') = \text{Variance sur } f \text{ de l'expression } (A + B) \text{ avec :}$

$$A = \Sigma_a \theta_{af} \frac{(\theta_{am'} - \theta_{am})}{2} \left(h + \frac{i}{2} + \frac{N-1}{2} l \right)$$

$$B = \frac{l}{4} \Sigma_{a < a'} (\theta_{am} \theta_{a'm} - \theta_{am'} \theta_{a'm'}) \theta_{af} \theta_{a'f}$$

La deuxième étape de calcul consiste à calculer l'espérance de $R^2(m, m')$ pour deux lignées différentes à n locus et identiques à $N - n$ locus, sachant que les $g \times N$ paramètres θ_{am} ou θ_{af} des neuf lignées sont tels que :

$$E(\theta_{am}) = E(\theta_{af}) = 0 \quad E(\theta_{am} \times \theta_{a'm}) = E(\theta_{am} \times \theta_{a'm'}) = 0$$

Cette structure des neuf lignées n'empêche pas l'effet de covariances dans les calculs de distances, parce que les testés sont aussi les testeurs.

Dans cette étape, on a aussi cherché les valeurs moyennes d'un certain nombre de statistiques définies dans l'échantillon des 9 lignées, à savoir :

$$Va = \Sigma_f \frac{\theta_{af}^2}{r} - \frac{(\Sigma_f \theta_{af})^2}{r^2} \quad r = \text{nombre de lignées}$$

Va est la variance de θ_{af} dans l'échantillon, qui nous intéresse parce qu'elle apparaît dans les calculs de $R^2(m, m')$:

$$Va = 4u_a v_a \text{ de HAYMAN (1954)}$$

Insistons sur le fait que Va correspond à une statistique de l'échantillon nécessaire aux calculs de distances, indépendamment de la structure particulière de nos neuf lignées. Va ne cherche pas à estimer la variance d'ailleurs connue (la valeur 1) d'une des aléatoires θ_{af} définie par tirage au hasard d'un gène dans la F_1 .

Au contraire, on cherche à estimer la valeur moyenne de Va, dans un échantillon de r lignées, parmi tous les échantillons possibles.

De même, on aura besoin d'estimer la valeur moyenne dans notre cas particulier, des statistiques suivantes :

$$C_{aa'} = \Sigma_f \frac{\theta_{af} \theta_{a'f}}{r} - \frac{(\Sigma_f \theta_{af})(\Sigma_f \theta_{a'f})}{r^2}$$

$C_{aa'}$ est la covariance ($\theta_{af}, \theta_{a'f}$) dans l'échantillon.

$C_{aa'}^2$ la covariance au carré.

V_a^2 la variance au carré.

$D_{aa', m} = \theta_{am} \theta_{a'm} C_{aa'}$.

TABLÉAU I

Valeur attendue de quelques paramètres définis dans l'échantillon des r lignées

- V_a : $4u_a v_a$ variance de θ_{af} sur indice f.
- $C_{aa'}$ covariance ($\theta_{af}, \theta_{a'f}$) sur indice f.
- $D_{aa', m} = \theta_{am} \theta_{a'm} C_{aa'}$.
- C_v covariance [$(\theta_{am} - \theta_{am'})^2, (\theta_{a'm} - \theta_{a'm'})^2$] sur les indices m, m'.
- V_n variance de $n = \Sigma_a (\theta_{am} - \theta_{am'})^2 / 4$.

Statistique	Symbole	Espérance
$V_a = 4u_a v_a$	V	$\frac{r-1}{r}$
V_a^2	V_2	$(r-1)(r(r-1) + 2) / r^3$
$C_{aa'}$	C_2	0
$C_{aa'}^2$		$(r-1) / r^2$
$D_{aa', m}$		C_2
C_v		$4C_2$
V_n		$N(zV - V_2 + C_2(N-1)) / 4$

Le tableau 1 donne les espérances de ces expressions. Quelques calculs d'espérances sont donnés en annexe à titre d'exemple.

Remarquons que $E(D_{aa' m}) = \frac{r-1}{r^2}$ alors que $E(D_{aa' p}) = 0$.

En effet, les paramètres des parents ne figurent pas dans $C_{aa'}$ calculé sur les lignées.

Revenons au calcul de l'espérance de $R^2(m, m')$:

$$E(R^2(m, m')) = n(V - C_2) \left[\left(h + \frac{i}{2} + \frac{N-1}{2} l \right)^2 + (N-n) \frac{l^2}{4} \right] + n^2 \left(h + \frac{i}{2} + \frac{n-1}{2} l \right)^2 C_2 \quad (4)$$

Dans cette expression, V est l'espérance de V_a (soit $\frac{r-1}{r}$) et C_2 est l'espérance de $C_{aa'}$ (soit $\frac{r-1}{r^2}$).

Si les deux parents des lignées sont en phase d'association totale, on a approximativement pour les termes en j :

$$R^2(p, p') = \text{Variance de } [h \Sigma_a \theta_{af} + (l-i) \Sigma_{a < a'} \theta_{af}]$$

$$R^2(p, p') \text{ en association} = \left[h + \frac{N-1}{2} (l-i) \right]^2 [\Sigma_a V_a + \Sigma_{a \neq a'} C_{aa'}]$$

et $E(R^2(p, p')) \text{ en association} = NV \left(h + \frac{N-1}{2} (l-i) \right)^2 \quad (4 \text{ bis})$

Si les parents sont en phase de dispersion totale, on a approximativement pour les termes en i et j :

$$R^2(p, p') \text{ en dispersion} = \text{Variance de } \left[\Sigma_a \theta_{ap} \theta_{af} \left(\frac{h+i}{2} + \frac{N-1}{2} l \right) \right]$$

d'où $E(R^2(p, p')) \text{ en dispersion} = NV \left(h + \frac{i}{2} + \frac{N-1}{2} l \right)^2 \quad (4 \text{ ter})$

Expression du test d'épistasie du croisement triple

Ce test calcule la variance de $[Y_{fp} + Y_{fp'} - Y_f] = T$.

Si les deux parents sont en association complète, on a exactement :

$$T = \text{Variance de } \left[\Sigma_{a < a'} \theta_{af} \theta_{a'f} \frac{(l-i)}{2} \right] \quad (5)$$

Dans ce cas, si $i = l$, l'épistasie ne sera pas décelée.

Si les deux parents sont en dispersion, on a approximativement pour les termes en j :

$$T = \text{Variance de } \left[\Sigma_{a < a'} \theta_{af} \theta_{a'f} \frac{(l \theta_{ap} \theta_{a'p} - i)}{2} \right] \quad (5 \text{ bis})$$

Comparaison des estimations de $D = \Sigma V d^2$ par l'essai diallèle et l'essai triple :

D diallèle = Variance (Y_m) = Variance ($\Sigma_a \theta_{am} d + \Sigma_{a < a'} \theta_{am} \theta_{a'm} i$).

Si les deux parents sont en phase d'association complète, on a exactement :
D Triple = Variance ($Y_{fp} + Y_{fp'}$).

$$D \text{ Triple} = \text{Variance} \left[\Sigma_a \theta_{af} d + \Sigma_{a < a'} \theta_{af} \theta_{a'f} \frac{(l + i)}{2} \right] \quad (6 \text{ bis})$$

Si les deux parents sont en phase de dispersion complète, on a approximativement pour les termes en j :

$$D \text{ Triple} = \text{Variance} \left[\Sigma_a \theta_{af} d + \Sigma_{a < a'} \theta_{af} \theta_{a'f} \frac{(l \theta_{ap} \theta_{a'p} + i)}{2} \right] \quad (6 \text{ ter})$$

Ces formules appellent quelques commentaires :

— Les caractères strictement additifs (sans dominance, sans épistasie) engendrent des mesures de distances toujours nulles, donc sans intérêt pour différencier les lignées.

— La limite supérieure du nombre de gènes différents décelables dans les lignées dépend du nombre de locus qui présentent chez les testeurs une variabilité ($V_a \neq 0$).

— Les effets d'épistasie de type j apparaissent négligeables sur les statistiques étudiées.

— Si les deux parents sont en phase d'association complète, et si $i = l$, l'épistasie passe inaperçue au niveau du test d'épistasie triple, des valeurs de l'hétérosis et de la distance entre les 2 parents. Les estimations D diallèle et D Triple sont égales.

— Les formules (3) et (4) montrent bien la relation entre l'hétérosis et la distance tels que définis ici, surtout chez les parents des lignées.

Chez les lignées, la relation est compliquée surtout à cause de l'épistasie de type l et la covariance.

Le terme prépondérant de l'expression (4) est $h + \frac{i}{2} + \frac{N-1}{2} l$ qui correspond au terme $h + \frac{i}{2} + \frac{n-1}{2} l$ de l'expression (3).

Ces deux termes peuvent être vus comme la dominance apparente d'un locus respectivement dans la mesure de distance et dans la mesure d'hétérosis. Ces deux effets de dominance apparente ne sont égaux que pour les parents (P, P') ou les lignées extrêmes. Pour que la corrélation entre hétérosis et distance soit forte, il faut au moins que $h + \frac{i}{2} + \frac{N-1}{2} l$ garde le même signe à tous les locus.

— Montrons que si les unités à tester M sont différentes des unités servant de testeurs F, l'effet des covariances sur les distances est fortement réduit, ceci sans aucune hypothèse sur les structures génotypiques des M et des F.

On suppose ici l'absence d'épistasie :

$$R^2(m, m') = h^2 \Sigma_a \frac{(\theta_{am} - \theta_{am'})^2}{4} V_{af} + h^2 \Sigma_{a \neq a'} \frac{(\theta_{am} - \theta_{am'}) (\theta_{a'm} - \theta_{a'm'})}{4} C_{aa'f}$$

Le terme avec les covariances comprend $n(n-1)$ éléments tels que $\theta_{am'} = -\theta_{am}$.

Dans ce cas, le terme des covariances comprend la somme de $n(n-1)$ éléments $\theta_{am} \times \theta_{a'm} C_{aa'f}$.

Si les M sont distincts des F, ces éléments sont de signe quelconque et la somme est probablement voisine de zéro (c'est le cas des parents testés sur les lignées).

Si $M = F$, dans chaque élément, $\theta_{am}\theta_{a'm}$ est le plus souvent du même signe que $C_{aa',f}$ et la somme des éléments est le plus souvent positive.

Dans le présent travail, avec la structure particulière des lignées, il a été possible de chiffrer la valeur moyenne du terme $D_{aa',m} = \theta_{am}\theta_{a'm}C_{aa',f}$ (voir tabl. 1).

Toujours dans le présent modèle, calculons la moyenne des r^2 distances $\overline{R^2}$:

$$\overline{R^2} = \Sigma_a \frac{1}{2} V_{am} V_{af} h^2 + \Sigma_{a \neq a'} \frac{1}{2} C_{aa',m} C_{aa',f} h^2$$

Sur l'expression de $\overline{R^2}$, on voit également que si $M \neq F$, les termes $C_{aa',m} \times C_{aa',f}$ sont de signe variable et leur somme est sans doute voisine de zéro. Si $M = F$ l'effet des covariances est positif. Ceci sans aucune hypothèse sur la structure du matériel.

2.2. — Utilisation des formules

Retrouver l'analyse classique du diallèle :

L'analyse classique suppose l'absence d'épistasie et $C_{aa'} = 0$ dans notre échantillon particulier.

On a alors les expressions simplifiées :

$$\begin{aligned} He(m, m') &= \Sigma_a \frac{(\theta_{am} - \theta_{am'})^2}{4} h \\ R^2(m, m') &= \Sigma_a \frac{(\theta_{am} - \theta_{am'})^2}{4} V_a h^2 \end{aligned}$$

La moyenne et la variance des r^2 statistiques He sont :

$$\begin{aligned} \overline{He} &= \frac{1}{2} \Sigma_a V_a h \\ VHE &= \Sigma_a \frac{2V_a - V_a^2}{4} h^2 \end{aligned} \quad (7)$$

La moyenne des r^2 distances vaut :

$$\overline{R^2} = \Sigma_a \frac{V_a^2}{2} h^2$$

Sachant que $V_a = 4uava$, on déduit immédiatement de (7) les estimations des paramètres génétiques h_2, H_2, H_1 :

$$\begin{aligned} h_2 &= (\Sigma 2uvh)^2 = (\overline{He})^2 \\ H_2 &= \Sigma 16u^2v^2h^2 = 2\overline{R^2} \\ H_1 &= \Sigma 4uvh^2 = \overline{R^2} + 2VHE \end{aligned} \quad (8)$$

Remarquons que les formules (8) sont une façon sans doute plus élégante de calculer les paramètres h_2, H_1, H_2 mais exactement équivalente à la méthode classique fondée sur les statistiques (V_r, W_r).

En effet $(\overline{He})^2$ est égal au terme $(mL_1 - mL_0)^2$ de HAYMAN (1954).

Exprimons $\overline{R^2}$ et VHE en fonction des termes V_I , V_P , V_r , \overline{V}_r , W_r de MATHER et JINKS (1971) :

$$R^2_{ij} = \Sigma_t (s_{it} - s_{jt})^2 / R \text{ avec } s_{it}, s_{jt} \text{ comme A.S.C.}$$

$$\overline{R^2} = \Sigma_{ij} R^2_{ij} / R^2 = 2 \Sigma_t s^2_{it} / R^2 = 2V_I = \frac{1}{2} H_2$$

$$VHE = \text{Variance} \left(y_{ij} - \frac{y_{ii} + y_{jj}}{2} \right) = 2V_r + V_I + \frac{1}{2} V_P - 2W_r$$

Sachant que $V_I = \overline{V}_r - V_r$, on retrouve bien les expressions de MATHER et JINKS (1971, p. 269) :

$$H_1 = \overline{R^2} + 2VHE = V_P + 4\overline{V}_r - 4W_r$$

Ajustement du modèle

Il s'agit de trouver les paramètres $h + \frac{i}{2}$, l et les 36 n associés aux 36 couples de lignées qui permettent avec les formules (3) et (4) d'approcher au mieux les hétérosis et distances observées, ceci dans le cas où la corrélation observée entre hétérosis et distance est forte.

Une méthode d'estimation rigoureuse n'a pas été trouvée. Dans un premier temps, on a estimé l'ensemble des n et $h + i/2$ en posant $l = 0$.

En pratique, i étant lié à h , on parlera de h à la place de $h + i/2$.

Pour tout couple M, M' on a $n = He(m, m') / h$ qui introduit dans (4) donne une distance calculée R^2_c :

$$R^2_c = (V - C_2)hHe(m, m') + C_2H_e^2(m, m')$$

Un programme informatique permet de trouver la valeur h qui minimise la fonction $\Sigma(R^2 \text{ calculé} - R^2 \text{ observé})^2$ pour les 36 couples.

Avec cette valeur h , on estime alors pour chaque couple M, M' , nhe le nombre de locus différents d'après l'hétérosis et nR^2 le nombre de locus différents d'après la distance.

Le graphique des points (nhe, nR^2) permet d'apprécier si ceux-ci sont alignés sur la bissectrice.

Les formules (3) et (4) doivent permettre d'expliquer le cas échéant les déviations par rapport à la bissectrice par des valeurs l non nulles.

Chaque série de 36 valeurs nhe et nR^2 est utilisée pour estimer le nombre N de locus variables dans les neuf lignées.

$$\text{En effet } n(m, m') = \frac{\Sigma_a (\theta_{am} - \theta_{am'})^2}{4}$$

La moyenne \overline{n} et la variance \overline{V}_n ont les espérances suivantes (voir tabl. 1) :

$$E(n) = NV / 2 \quad (9)$$

$$E(V_n) = N(2V - V_2 + (N - 1)C_2) / 4 \quad (10)$$

Les relations (9) et (10) permettent une estimation de N .

En définitive, si les 4 estimations de N concordent et si le graphique montre un bon alignement des points sur la bissectrice, on regroupe ces diverses estimations de n et N en une seule estimation.

On obtient une série de valeurs n qui sont des mesures de distances indépendantes des effets génétiques et qu'on peut appeler distances standardisées.

La comparaison des séries n correspondant à deux caractères a un sens.

Recherche d'une distance générale

Dans la mesure où plusieurs caractères ont pu être grossièrement ajustés, on dispose pour chaque couple m, m' de plusieurs valeurs n . Il serait intéressant de calculer une sorte de distance générale combinant les distances particulières standardisées « n ». Pour ce faire, on utilise la matrice des covariances entre les différents vecteurs (n) des différents caractères.

On a en espérance :

$$\text{Covariance } (n_i, n_j) = N_{ij}(2V - V_2 + (N_{ij} - 1)C_2) / 4 \quad (11)$$

n_i désigne le vecteur des distances pour le caractère i .

N_{ij} désigne le nombre de locus communs aux caractères i et j .

De (11) on peut calculer N_{ij} .

L'examen des différents N_i et N_{ij} permet de construire une distance générale R_g qui évite de compter le même locus plusieurs fois :

$$R_g^2 = n_1 + \left(1 - \frac{N_{12}}{N_2}\right) n_2 + \left(1 - \frac{N_{13} + N_{23}}{N_3}\right) n_3 + \dots \quad (12)$$

Un nouveau caractère C_i est ajouté dans la mesure où ses locus ne sont pas déjà inclus dans les caractères C_j précédents.

Le nombre de ces locus déjà inclus est estimé par $\sum_j N_{ij}$.

L'assemblage est arrêté quand $\sum_j N_{ij} = N_i$ pour tout nouveau caractère.

Les distances combinées avec la formule (12) sont choisies avec des N_{ij} le plus faible possible et des N_i le plus fort possible.

La vraisemblance de R_g^2 peut être appréciée en calculant les covariances de R_g^2 avec les distances (n_i) . Il faut que covariance $(R_g^2, n_i) \neq$ variance (n_i) .

2.3. — Estimation des distances et des hétéroses

En fait la moyenne génotypique Y_{fm} n'est pas connue. On dispose de la moyenne phénotypique P_{fm} telle que :

$$P_{fm} = Y_{fm} + \varepsilon_{fm}$$

Comme d'habitude, on suppose que les ε_{fm} sont des aléatoires d'espérance zéro et de variance σ_e^2 . On a alors les estimations :

$$\widehat{R}^2(m, m') = \sum_f \frac{(P_{fm} - P_{fm'})^2}{r} - \frac{(\sum_f (P_{fm} - P_{fm'}))^2}{r^2} - \frac{2(r^2 - r + 1)}{r^2} s_e^2$$

$$He(m, m') = P_{mm'} - \frac{P_m + P_{m'}}{2}$$

$$\text{VHE} = \sum_{mm'} \frac{H_e^2(m, m')}{r^2} - \frac{(\sum_{mm'} He(m, m'))^2}{r^4} - \frac{r(r-1)}{r^4} \left[(r^2 - 2)s_e^2 + \frac{r^2 - 2r + 2}{2} s_p^2 \right]$$

s_e^2 est l'estimation de σ_e^2 sur les hybrides.

s_p^2 est l'estimation de la variance erreur sur les lignées (considérée uniquement pour l'hétérosis).

Résultats

Corrélation distance — hétérosis

Le premier résultat à rapporter qui donne de la consistance aux analyses ultérieures est la corrélation entre distance R^2 et hétérosis (on ne s'occupe en fait que du carré de la distance).

Ces corrélations expérimentales sont calculées sur les 36 couples différents observés.

Dans le tableau 2, on voit que sauf pour le caractère NE, les corrélations (R^2 , He) sont très fortes. Par contre, si on remplace l'hétérosis par la valeur de la F_1 , on obtient une forte diminution de la corrélation.

Les formules (3) et (4) expliquent bien ce résultat si on admet que les locus agissent tous dans le même sens.

Ce résultat conduit à se méfier de certaines autres mesures de l'hétérosis utilisées dans la littérature pour établir un lien entre l'hétérosis et la diversité enzymatique ou génétique telles que la valeur de la F_1 ou la différence entre la F_1 et un des deux parents (ZALI et ALLARD, 1976).

D'autre part, ce résultat permet de penser que tout se passe comme si les gènes agissant sur un caractère se classaient en deux catégories : ceux dont l'action est faible et s'exprime sous forme d'un « bruit de fond », et ceux dont l'action est forte.

Pour ces derniers, *grosso modo* la dominance apparente $\left(h + \frac{i}{2} + \frac{N-1}{2}l\right)$ associée à un locus garde un signe constant d'un locus à l'autre.

La très faible corrélation observée chez le caractère NE s'explique au contraire par des effets géniques de signe variable. La figure 1 concernant NE illustre bien cette interprétation : à un niveau donné de R^2 , la moyenne des hétérosis est nulle mais la fluctuation de l'hétérosis augmente avec le niveau de R^2 . Cela signifie que les n gènes différents sont un mélange en proportion variable de gènes à effets positifs et de gènes à effets négatifs. Pour un tel caractère, une analyse génétique plus approfondie n'a pas été tentée.

Paramètres génétiques classiques

Le calcul des paramètres h_2 , H_2 , H_1 à partir des statistiques classiques (V_r , W_r) a permis de vérifier numériquement que ces estimations sont exactement les mêmes que celles obtenues par les formules (8), comme prévu par la théorie.

Analyse génétique approfondie

On a essayé de reconstituer la liaison curvilinéaire observée entre hétérosis et distance à l'aide d'un modèle sans épistasie mais avec les covariances C_2 . Le paramètre h a été ajusté suivant la technique exposée au chapitre Méthodes.

Les figures 2 présentent l'ajustement obtenu, sous la forme suivante : en abscisse, le nombre n de gènes estimé sur R^2 , en ordonnée le nombre n de gènes

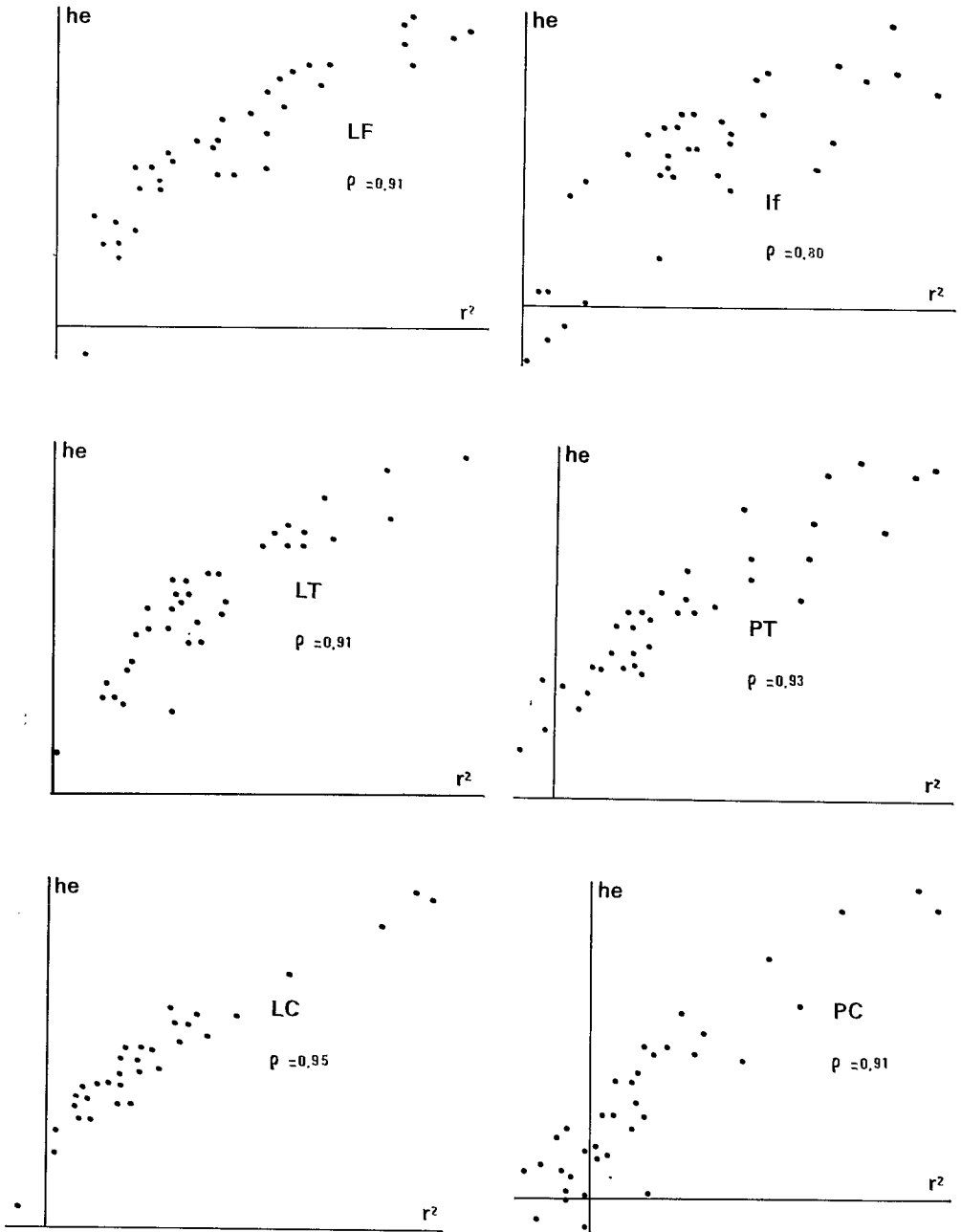


FIG. 1.

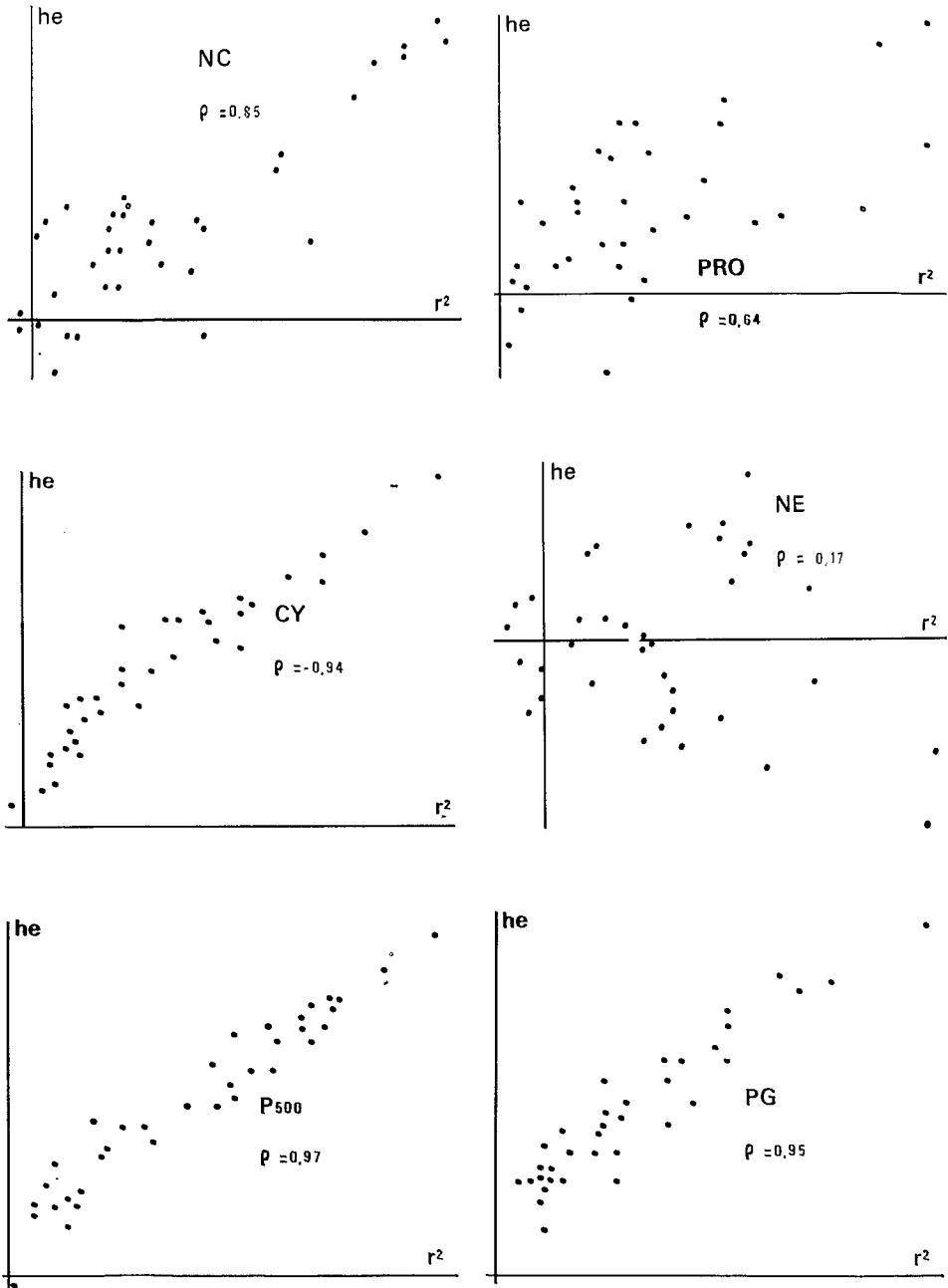


FIG. 1. — Représentation des 36 couples de lignées par leur hétérosis (He) et leur distance (R^2).
 ρ = corrélation entre He et R^2

TABLEAU 2

Estimation de divers paramètres génétiques dans le modèle classique et dans un modèle prenant en compte les covariances entre locus.
 Pour les paramètres D et H_1 , la signification statistique est donnée (* au seuil 5 p. 100, ** au seuil 1 p. 100)

Observations	Caractères											
	CY	LF	h	LT	PT	LC	PC	NC	PRO	NE	P500	PG
Corrélation R^2 , H_c	— 0,94	0,91	0,80	0,91	0,93	0,95	0,90	0,85	0,64	— 0,17	0,97	0,95
Corrélation R^2 , F_1	— 0,43	0,47	0,44	0,87	0,61	0,53	0,50	0,68	0,57	0,008	0,89	0,94
<i>Modèle classique :</i>												
Test d'épistasie Triple T	2,48*	1,18	0,83	2,60*	0,96	1,20	1,89	—	—	3,86**	0,74	—
Diallèle b	0,85	0,88	1,25	0,29*	0,92	0,41*	0,92	0,59*	0,65	0,58*	0,52*	0,10*
$D = \Sigma Vd^2$ Triple	740**	582**	919**	4 550**	335**	1 471**	2 290**	—	—	105**	3,3	—
Diallèle	208**	334**	953**	522**	398**	246**	982**	90**	488**	27,8**	1,3**	2 014**
$H_1 = \Sigma Vh^2$ Triple	1 060**	909**	520**	1 445*	168**	426*	390*	—	—	41**	6,3**	—
Diallèle	429	542	675	2 241	372	627	234	61	519	27,9	11,6	93 134
$H_2 = \Sigma V^2h^2$ Diallèle	366	447	513	2 030	342	550	180	48	456	16,7	8,8	75 084
$N = 4h^2/H_2$ nombre de loci	4,40	5,73	4,37	6,05	6,17	6,68	4,02	2,00	1,8	0,02	4,89	4,83
h/d degré de dominance	1,4	1,3	0,8	2,1	1,0	1,6	0,5	0,8	1,0	1,0	3,0	6,8
<i>Modèle avec covariances :</i>												
N' nouvelle estimation du nombre de loci	5,65	8,59	7,11	8,76	8,91	9,68	6,25	3,27	3,41	—	7,11	6,60
n Max: nombre maximal de gènes différents entre 2 lignées	5,96	6,93	6,05	7,85	7,70	9,30	6,91	3,74	3,33	—	6,17	6,84
h effet moyen de dominance	8,8	7,3	8,6	14,9	6,4	7,8	7,8	4,5	12,3	—	1,2	110
R^2 Max distance maximale entre 2 lignées	665	607	685	2 985	554	1 118	711	80,5	787	—	11,5	129 147
$N'Vh^2$	392	413	471	1 733	329	528	341	58,7	463	—	8,5	71 075
$(h/d)'$ degré de dominance	1,2	1,2	0,7	0,6	0,7	0,5	0,4	0,8	0,97	—	1,4	5,9
Distance générale: N_{gt}/N_t	0,92	1,23	1,24	1,04	0,96	0,97	0,76	1,14	1,49	—	1,11	1,06

estimé sur l'hétérosis. On peut voir en gros les points se répartir le long de la bissectrice avec, cependant, la déformation suivante : près de l'origine, un excès des $n(R^2)$ sur les $n(R^2)$ et l'inverse loin de l'origine.

Cette déformation peut être expliquée en partie par l'inégalité des effets h et l , en partie par une épistasie dupliquée (effet l opposé à la dominance).

En effet, les formules (3) et (4) deviennent en supposant les effets ha et laa' différents suivant les locus :

$$\begin{aligned} He(m, m') &= \Sigma a \Sigma a (ha + i/2) + \Sigma a, a' \Sigma a \Sigma a' laa' / 2 \\ R^2(m, m') \text{ ajusté} &= (R^2(m, m') - C_2 He^2(m, m')) / (V - C_2) \\ &= \Sigma a \Sigma a (ha + i/2 + \frac{N-1}{2} la)^2 + \Sigma aa' \Sigma a (1 - \Sigma a')^2 aa' / 4 \end{aligned}$$

avec $\Sigma a = (\theta am - \theta am')^2 / 4$

$(N-1)l_a = \Sigma a' laa'$.

Une analyse numérique des relations entre He en abscisse et R^2 ajusté en ordonnée a donné les nuages de points suivants :

— Pour des effets ha toujours de même signe mais d'intensité variable, en l'absence d'épistasie, les points se disposent le long d'une courbe en S symétrique par rapport au milieu du nuage.

Avec une épistasie complémentaire (l de même signe que h) la courbe du S à concavité vers le haut disparaît, reste la branche du S à concavité vers le bas.

Avec une épistasie dupliquée (l de signe contraire à h) c'est l'inverse.

— Pour des effets ha de signe variable, avec ou sans épistasie, le nuage devient flou et de forme losangique.

Ces relations entraînent les effets suivants sur les figures 2 :

— A la courbe en S correspond un S retourné (à cause des changements d'axes) que la bissectrice traverse aux deux extrémités et au centre. C'est l'image de CY et P500.

— A la branche du S de l'épistasie dupliquée correspond une figure 2 comme celles de LF, LT, PT, LC, PG : la bissectrice passe par le milieu du nuage mais au-dessus des points extrêmes.

— La branche du S typique de l'épistasie complémentaire n'est pas représentée dans les figures 2 : la bissectrice devrait passer toujours au milieu du nuage mais au-dessous des points extrêmes. La concavité du nuage serait inversée.

— Les autres caractères présentent une figure 2 assez floue caractéristique de la coexistence d'effets h de signe variable : lf, NC, PRO, PC. La présence d'épistasie ne peut alors être décelée sur les graphiques.

L'image de CY n'interdit pas une légère épistasie dupliquée.

Le croisement triple décele l'épistasie dans CY, LT, NE (tabl. 2), ce qui ne contredit pas l'analyse des figures 2 mais n'en décele pas dans les autres caractères en particulier LF, PT et LC. Ce désaccord peut s'expliquer soit par une faible puissance du test d'épistasie du croisement triple modifié, soit par la présence d'effets i égaux à l et des parents P et P' en association totale.

Le test classique du croisement diallèle basé sur la pente b de régression (Wr/Vr) ne décele la présence d'effets de covariances ou d'épistasie que dans 6 caractères, au lieu de tous, ce qui est peu vraisemblable et jette un doute sur la validité de ce test (tabl. 2).

En l'absence de conclusions sûres quant à l'épistasie, le modèle sans épistasie ajusté au mieux (fig. 2) fournit les estimations de la dominance apparente moyenne

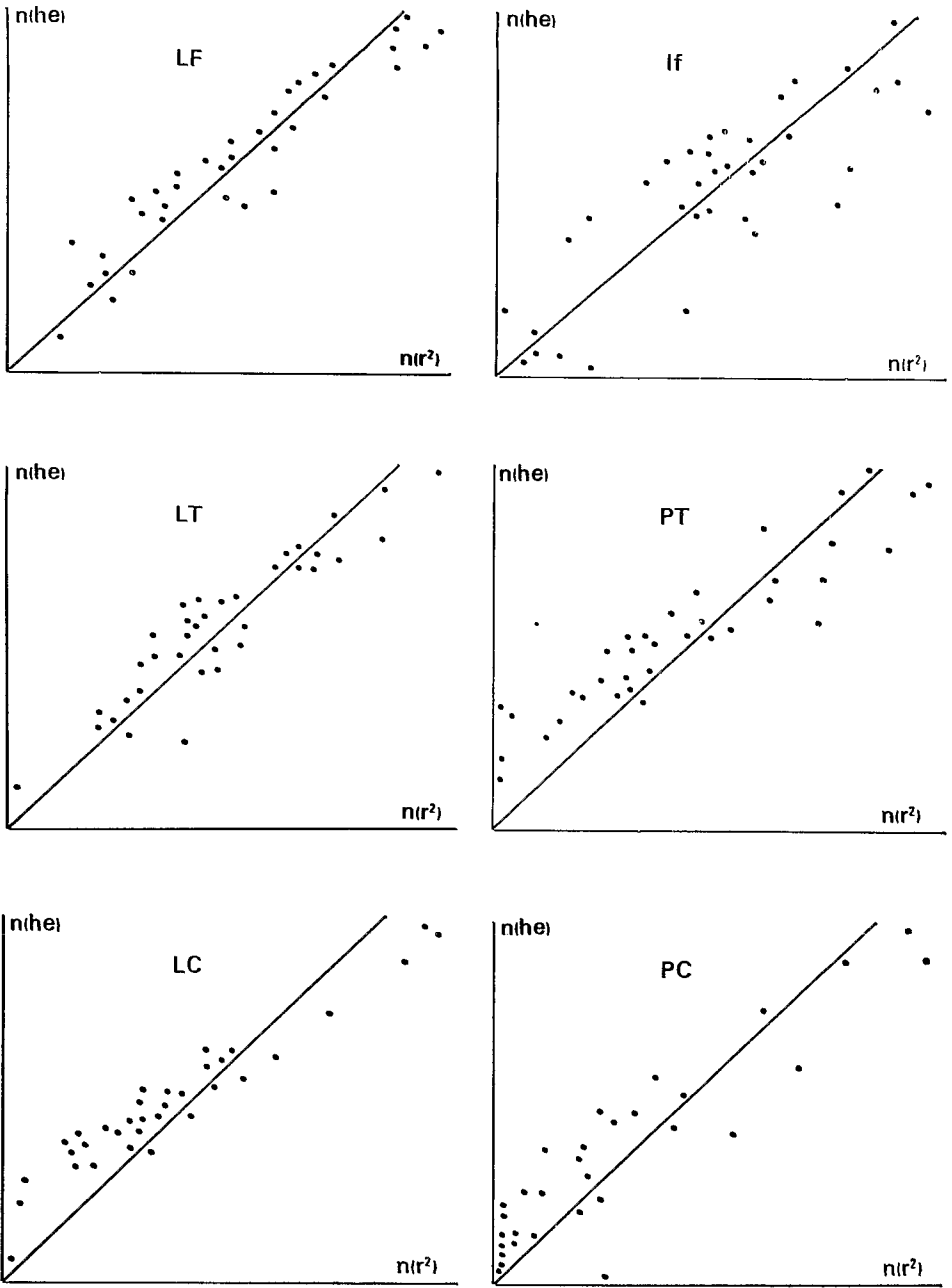


FIG. 2.

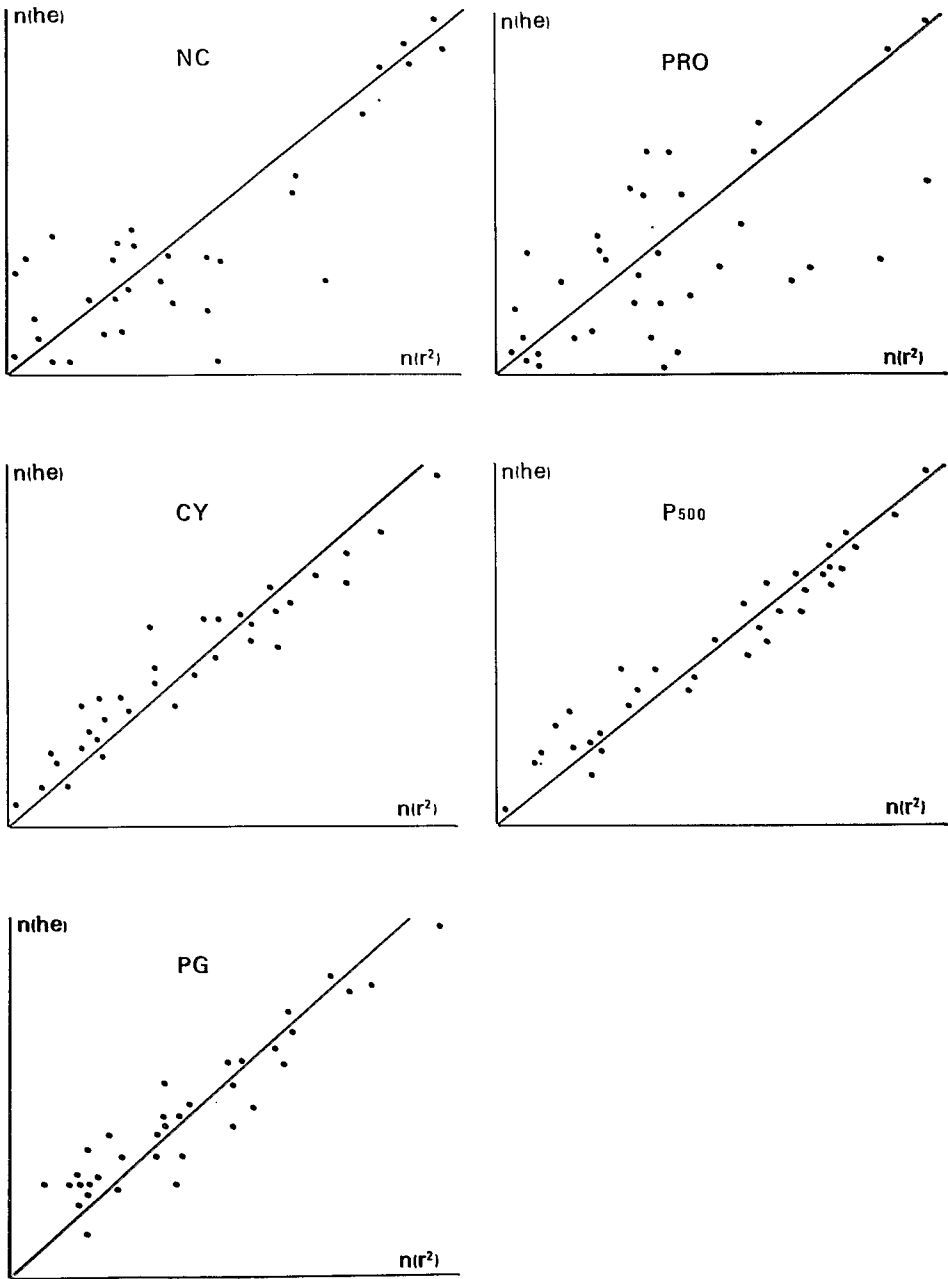


FIG. 2. — Représentation des 36 couples de lignées par $n(he)$ et $n(R^2)$ les estimations sur l'hétérosis et sur la distance du nombre de gènes différents, par ajustement des formules (3) et (4). La droite représente la première bissectrice.

et du nombre total de gènes en cause (N' du tabl. 2). Ce nombre reste voisin de l'estimation N classique. Il vaut environ 7, le nombre de base du génome des mils penicillaires.

Le degré de dominance (h/d) de ce modèle est plus faible que l'estimation classique.

PG et P500 manifestent de la superdominance, CY et LF une légère superdominance et les autres caractères une dominance incomplète.

Ces estimations h et N' , malgré la présence dans certains cas d'épistasie dupliquée, concordent avec H1 Triple, sauf pour CY et LF.

En effet, H1 Triple n'est autre que $R^2(p, p')$. D'après les formules (4), $R^2(p, p') = N'Vh^2$. Le tableau 2 montre que les valeurs $N'Vh^2$ calculées sont proches de H1 Triple.

Pour CY et LF, la forte valeur de H1 Triple pourrait s'expliquer par des parents P, P' en association et une épistasie dupliquée avec $i = l$.

Par contre, les estimations D Triple et D diallèle sont dans l'ensemble très divergentes sans explication évidente.

La présence d'épistasie de type l pourrait d'après les formules (6) expliquer seulement une légère supériorité de D Triple sur D diallèle.

Remarquons que D Triple utilise uniquement des observations sur des croisements alors que D diallèle n'utilise que les lignées autofécondées. Il y aurait sans doute un effet de la consanguinité non pris en compte par notre modèle.

Enfin, le tableau 2 montre que l'estimation H1 diallèle est dans l'ensemble inférieure à R^2 Max, la distance maximale observée. D'après le modèle sans épistasie, H1 diallèle devrait être voisin de R^2 Max.

En effet, d'après (8), H1 diallèle = $\overline{R^2} + 2 \text{ VHE}$

$$\begin{aligned} E(\overline{R^2}) &= Nh^2(V_2 + (N - 1)C_2)/2 \\ E(\text{VHE}) &= Nh^2(2V - V_2 + (N - 1)C_2)/4 \end{aligned}$$

d'où $E(\text{H1 diallèle}) = Nh^2(V + (N - 1)C_2)$ la distance maximale possible.

Ce désaccord pourrait s'expliquer par une moyenne $\overline{R^2}$ trop basse : il y aurait un excès de couples de lignées assez voisines. Les lignées ne seraient pas tirées au hasard de la F_2 , mais auraient subi une sélection dans la même direction.

Recherche d'une distance générale

Le tableau 3 contient la matrice des N_{ij} , le nombre de locus communs aux caractères i et j , obtenue à partir des covariances entre les distances standardisées n_i et n_j par la formule (11). Les N_i sont les éléments diagonaux. On voit qu'aucun N_{ij} n'est supérieur à N_i ou à N_j , comme il est souhaitable.

La figure 3 représente une interprétation de l'occupation du génome par les divers caractères. Il semble y avoir une forte pléiotropie : certains locus semblent communs à tous les caractères. Avec les deux seuls caractères LF et LT, la majeure partie des locus variables semble couverte.

Tout semble se passer comme si les « gènes » visibles à ségrégation indépendante étaient en gros les 7 chromosomes de base du génome des petits mils. Chaque chromosome porterait un ou parfois deux gènes à effet important pour chaque caractère.

TABLEAU 3

Matrice des nombres de locus communs à deux caractères.
La diagonale porte les nombres de locus régissant chacun des caractères

	LT	lf	CY	LC	PC	P500	NC	PRO	PT	LF	PG
LT . . .	9,19	5,60	5,25	5,98	1,41	5,91	3,52	3,82	6,56	5,60	5,29
lf	5,60	7,74	2,84	7,35	5,65	6,12	3,24	3,69	5,97	7,22	5,65
CY	5,25	2,84	6,26	3,78	1,48	2,75	2,16	1,55	3,21	2,62	2,70
LC	5,98	7,35	3,78	11,01	7,93	7,90	1,97	2,77	7,39	7,36	6,39
PC	1,41	5,65	1,48	7,93	8,59	5,67	0,70	1,23	4,05	4,67	3,87
P500 . . .	5,91	6,12	2,75	7,90	5,67	7,88	2,16	2,09	6,95	5,89	6,04
NC	3,52	3,24	2,16	1,97	0,70	2,16	3,82	2,58	1,57	2,75	3,76
PRO . . .	3,82	3,69	1,55	2,77	1,22	2,08	2,58	3,71	1,27	3,97	2,85
PT	6,56	5,97	3,21	7,39	4,05	6,95	1,57	1,27	10,18	6,94	5,37
LF	5,60	7,22	2,62	7,36	4,67	5,89	2,74	3,97	6,94	9,23	9,08
PG	5,29	5,65	2,70	6,39	3,87	6,03	3,76	2,84	5,37	5,08	7,20

locus disposés en ligne

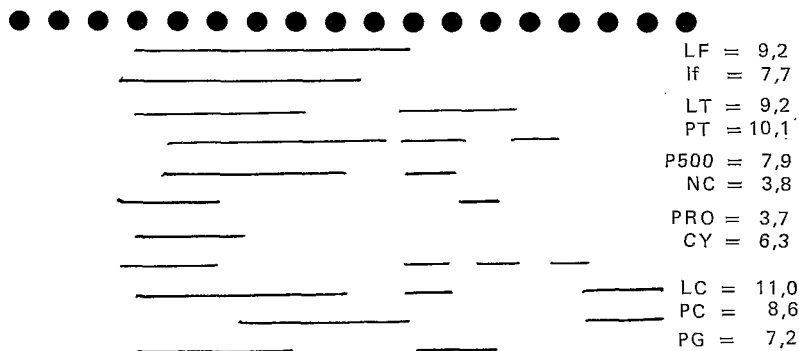


FIG. 3. — Schéma d'occupation du génome par les différents caractères. Le nombre de locus occupés est indiqué pour chaque caractère. La longueur du trait de chaque caractère est proportionnelle au nombre de locus qui le régissent. Le degré de superposition des traits de 2 caractères indique la proportion de locus communs.

La distance générale suivante a été construite :

$$R^2_g = n(LF) + 0,39n(LT) + 0,22n(LC) + 0,16n(CY)$$

Le nombre de gènes N_g couvert par cette distance vaut 13,9 ou 15 suivant les estimations des formules (9) ou (10).

R^2_g traduit donc environ 2 fois plus de locus qu'un caractère particulier.

La dernière ligne du tableau 2 donne le rapport N_{gi}/N_i du nombre de locus communs à la distance générale et au caractère i avec le nombre de locus du caractère i .

Les résultats sont satisfaisants : dans l'ensemble, le rapport est voisin de 1. Donc la distance générale recouvre sans redondance les locus des caractères particuliers.

Enfin, la corrélation entre la distance générale et la distance PG (poids de grain) pour les 36 paires de lignées vaut 0,54. La distance générale n'est donc pas un très bon prédicteur du rendement bien qu'elle couvre tous les locus du rendement.

Conclusion

La confrontation des données observées dans les essais Triple et Diallèle nous a conduit aux conclusions suivantes.

— Les statistiques de distances R^2 et d'hétérosis He permettent l'estimation des paramètres h_2 , H_1 , H_2 dans le cadre de l'analyse classique.

Elles donnent une idée des différences de structure génotypique entre lignées, sans hypothèse particulière.

Elles apparaissent à tout le moins, complémentaires des statistiques (V_r , W_r).

— La corrélation expérimentale entre R^2 et HE est apparue très forte dans 11 cas sur 12.

— Dans le cas de forte corrélation, l'approfondissement de l'analyse classique a conduit à supposer l'existence d'effets épistatiques qui s'opposent aux effets de dominance. Ainsi, une certaine saturation se manifeste dans la croissance de l'hétérosis avec le nombre de locus hétérozygotes, qui fait penser à la loi logistique et au modèle génétique de FORKMANN et SEYFFERT (1977). En pratique, un modèle génétique comportant des covariances entre locus, et une épistasie entre paires de locus de type l négatif a permis une bonne explication des observations, sauf en ce qui concerne la faible valeur de la variance des lignées (D diallèle).

— Malgré ses imperfections certaines, la distance générale représente une tentative intéressante pour le taxinomiste dans la mesure où elle approche le nombre de gènes ou groupes de gènes qui différencient deux unités génotypiques quelconques, et ce sur la base de caractères quantitatifs mesurés sur des croisements. Elle est sans doute aussi utile que bien d'autres distances phénotypiques (RAO, 1952; GOWER, 1971; EDWARDS, 1971). Elle tend à approcher la distance minimale définie par NEI (1975).

— Pour le sélectionneur la distance R devrait permettre de prévoir pour le caractère rendement, l'hétérosis entre lignées à partir de croisements lignées \times testeurs.

Le schéma 2 de la Caroline du Nord (COMSTOCK *et al.*, 1948) croisant entre eux deux ensembles A et B d'unités génotypiques devrait présenter deux avantages: économie d'effort d'environ 50 p. 100 et réduction des effets de covariance sur les mesures de distances.

Considérons, en effet, les croisements de N individus A avec N individus B. La valeur des N^2 hybrides est directement observée. Les distances entre individus A et les distances entre individus B peuvent être calculées et servir d'estimations pour la valeur des hybrides $A \times A$ et $B \times B$. Au total N^2 croisements permettent d'apprécier $N(2N - 1)$ hybrides.

L'économie est d'environ 50 p. 100.

Outre la réduction des effets de covariance (démontrée dans le chapitre des méthodes) l'utilisation de testeurs de distances distincts des unités à tester devrait permettre d'étudier l'évolution de l'hétérosis entre unités avec une croissance continue de la mesure de distance.

Le présent travail devrait donc être repris à partir des données expérimentales d'un schéma 2 de la Caroline du Nord incluant des unités génotypiques très diverses dans chacun des ensembles A et B.

Reçu pour publication en novembre 1977.

Summary

The use of HANSON and CASAS « R » distance to analyse genetically nine inbred lines of Pearl Millet

Nine F_5 inbred lines of pearl millet (*Pennisetum typhoides*) originating from the same F_2 population have been simultaneously studied in a diallel cross and in a modified triple test cross (JINKS *et al.* 1969) for a set of twelve quantitative characters.

Both trials have been analysed by considering the two sets of statistics: heterosis and squared distances R^2 (HANSON, CASAS, 1968).

It is first shown that the classical diallel analysis can be obtained by using these new statistics instead of the (V_r , W_r) statistics.

Among the twelve measured characters, eleven have shown a very strong correlation between heterosis and R^2 squared distances.

It is tempted to interpret those high correlations with a genetic model including covariances between locus states and 2 locus epistasis.

The model can be relatively well adjusted to the observations by considering type 1 epistasis in opposition to dominance.

Consequently, for each character, the number of the main genetic factors differentiating the lines is tentatively estimated.

Furthermore for each pair of characters, the number of genetic factors, common to both characters, is also estimated. Then, a general distance is built based on all the differences measured on all the characters, that hopefully avoids redundancy.

It is proposed to use this general distance in taxonomy.

For the plant breeder, the use of the R distance could reduce by an amount of 50 p. 100 the cost of appreciating the heterosis between each pair of a set of lines.

But the deficiencies of the genetic model concern mainly the explanation of inbred depression. The relation between heterosis and the number of heterozygous locus looks *grosso modo* like the logistic curve.

Références bibliographiques

- CASAS E., HANSON W. D., WELSHAUSEN E. J., 1968. Genetic relationships among collections representing three mexican race composites of *Zea mays* L. *Genetics*, **59**, 299-310.
- COMSTOCK R. E., ROBINSON H. F., 1948. The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. *Biometrics*, **4**, 254-266.
- EDWARDS A. W. F., 1971. Distances between populations on the basis of gene frequencies. *Biometrics*, **27**, 873-882.
- FORKMANN G., SEYFFERT W., 1977. Simulation of quantitative characters by genes with biochemically definable action VI. Modifications of a simple model. *Genetics*, **85**, 557-572.
- GOWER J. C., 1971. A general coefficient of similarity and some of its properties. *Biometrics*, **27**, 857-871.
- HANSON W. D., CASAS E., 1968. Spatial relationships among eight populations of *Zea mays* L. utilizing information from a diallel mating design. *Biometrics*, **24**, 867-880.

- HANSON W. D., MOLL R. H., 1973. Experimental evaluation of relationships among populations resulting from intergradation among cultivars of *Zea mays*. *Genetics*, **74**, 133-138.
- HAYMAN B. I., 1954. The theory and analysis of diallel crosses. *Genetics*, **39**, 789-809.
- JINKS J. L., PERKINS Jean-M., BREESE E. L., 1969. A general method of detecting additive dominance and epistatic variation for metrical traits. II. Application to inbred lines. *Heredity*, **24**, 45-57.
- MATHER K., JINKS J. L., 1971. *Biometrical genetics*. Chapman and Hall, Londres, 382 p.
- MOLL R. H., LONNQUIST J. H., VELEZ FORTUNO J., JOHNSON E. C., 1965. The relationships of heterosis and genetic divergence in maize. *Genetics*, **52**, 139-144.
- NEI M., 1975. *Molecular population genetics and evolution*. North Holland, Amsterdam-Oxford, 288 p.
- RAO C. R., 1952. *Advanced statistical methods in biometric research*. John Wiley-New York.
- Van der VEEN J. H., 1959. Tests of non-allelic interaction and linkage for quantitative characters in generations derived from two diploid pure lines. *Genetica*, **30**, 201-232.
- ZALI A. A., ALLARD R. W., 1976. The effect of level of heterozygosity on the performance of hybrids between isogenic lines of barley. *Genetics*, **84**, 765-775.

Annexe 1

Calcul de l'espérance de $R^2(m, m')$

La distance est mesurée dans un croisement dialléle $r \times r$.
M et M' sont différents à n locus, identiques à $N - n$ locus.

$R^2(m, m')$ = Variance sur indice f de $(Y_{mf} - Y_{m'f})$.

$$\begin{aligned}
 Y_{mf} - Y_{m'f} = & \Sigma_a(\theta_{am} - \theta_{am'})d\alpha/2 + \Sigma_a(\theta_{am'} - \theta_{am})\theta_{af}h\alpha/2. \\
 & + \Sigma_{a < a'}[\theta_{am}\theta_{a'm} - \theta_{am'}\theta_{a'm'} + (\theta_{am} - \theta_{am'})\theta_{a'f} \\
 & \qquad \qquad \qquad + (\theta_{a'm} - \theta_{a'm'})\theta_{af}]i_{aa'}/4. \\
 & + \Sigma_{a < a'}[\theta_{af}\theta_{a'f}(\theta_{a'm'} - \theta_{a'm}) - \theta_{a'm'} + \theta_{am} \\
 & \qquad \qquad \qquad + \theta_{a'f}(\theta_{a'm'}\theta_{a'm'} - \theta_{am}\theta_{a'm})]j_{a'a'}/4. \\
 & + \Sigma_{a < a'}[\theta_{af}\theta_{a'f}(\theta_{a'm} - \theta_{a'm}) - \theta_{a'm'} + \theta_{a'm} \\
 & \qquad \qquad \qquad + \theta_{a'f}(\theta_{a'm'}\theta_{a'm} - \theta_{a'm}\theta_{a'm})]j_{a'f'a}/4. \\
 & + \Sigma_{a < a'}[(\theta_{a'm'} - \theta_{am})\theta_{af} + (\theta_{a'm'} - \theta_{a'm})\theta_{a'f} \\
 & \qquad \qquad \qquad + (\theta_{am}\theta_{a'm} - \theta_{a'm'}\theta_{a'm'})\theta_{af}\theta_{a'f}]l_{aa'}/4.
 \end{aligned}$$

Les termes sans indice f sont des constantes qui n'interviennent pas dans le calcul de $R^2(m, m')$.

Avec les hypothèses :

$$h_a = h, i_{aa'} = i, j_{a'a'} = j_{a'a} = j, l_{aa'} = l \text{ et } \Sigma_a\theta_{af} = \Sigma_a\theta_{am} = 0,$$

les termes variables en i deviennent :

$$i/4 \Sigma_a\theta_{af}\Sigma_{a' \neq a}(\theta_{a'm} - \theta_{a'm'}) = \Sigma_a\theta_{af}(\theta_{a'm'} - \theta_{am})i/4,$$

certaines termes variables en j deviennent :

$$\sum_a \theta_{af} \theta_{am'} \sum_{a' \neq a} \theta_{a'f} = - \sum_a \theta_{am'} = 0,$$

les autres termes variables en j deviennent :

$$\sum_a \theta_{af} \theta_{am'} \sum_{a' \neq a} \theta_{a'm'} = - \sum_a \theta_{af} = 0.$$

Certains termes variables en l sont :

$$l/4 \sum_a \theta_{af} (\theta_{am'} - \theta_{am}) \sum_{a' \neq a} \mathbf{1} = \sum_a \theta_{af} (\theta_{am'} - \theta_{am}) (N - 1) l / 4.$$

Il vient alors en posant :

$$A = \sum_a \theta_{af} \frac{(\theta_{am'} - \theta_{am})}{2} \left(h + \frac{i}{2} + \frac{N-1}{2} l \right)$$

$$B = \frac{l}{4} \sum_{a < a'} (\theta_{am} \theta_{a'm} - \theta_{am'} \theta_{a'm'}) \theta_{af} \theta_{a'f}$$

$R^2(m, m') = \text{Variance}(A + B) = \text{Variance}(A) + \text{Variance}(B) + 2 \text{Covariance}(A, B).$

$$\begin{aligned} \text{Variance}(A) = \left(h + \frac{i}{2} + \frac{N-1}{2} l \right)^2 & \left[\sum_a \left(\frac{\theta_{am'} - \theta_{am}}{2} \right)^2 V_a \right. \\ & \left. + \sum_{a \neq a'} \left(\frac{\theta_{am'} - \theta_{am}}{2} \right) \left(\frac{\theta_{a'm'} - \theta_{a'm}}{2} \right) C_{aa'} \right]. \end{aligned}$$

Pour $N - n$ locus $\theta_{am'} - \theta_{am} = 0$, pour n locus $\frac{\theta_{am'} - \theta_{am}}{2} = \theta_{am'}$.

Variance (A) contient $n(n-1)$ termes :

$$\theta_{am'} \theta_{a'm'} C_{aa'} = \theta_{am'} \theta_{a'm'} \left[\frac{\sum_i \theta_{ai} \theta_{a'i}}{r} - \frac{\sum_{ij} \theta_{ai} \theta_{a'j}}{r^2} \right].$$

Calculons la valeur moyenne d'un de ces termes dans l'ensemble de tous les échantillons de r lignées qu'il est possible de tirer de la F_1 .

Sachant que

$$E(\theta_{ai} \theta_{a'j}) = E(\theta_{ai} \theta_{a'i}) = 0$$

les termes non nuls de $E(\theta_{am'} \theta_{a'm'} C_{aa'})$ correspondent à $i = m'$, $j = m'$. Donc :

$$E(\theta_{am'} \theta_{a'm'} C_{aa'}) = \frac{1}{r} - \frac{1}{r^2} = \frac{r-1}{r^2} = C_2$$

De même, $E(V_a = 1 - \sum_{ij} \theta_{ai} \theta_{a'j} / r_2) = V = \frac{r-1}{r}$, donc :

$$E(\text{Variance}(A)) = \left(h + \frac{i}{2} + \frac{N-1}{2} l \right)^2 (nV + n(n-1)C_2) \quad (\text{I})$$

Dans B, si a et a' appartiennent aux n locus ($a, a' \in n$) ou si a et a' appartiennent aux $N - n$ locus ($a, a' \in N - n$), le terme $\theta_{am} \theta_{a'm} - \theta_{am'} \theta_{a'm'} = 0$.

Seuls sont différents de zéro les $n(N - n)$ termes tels que $a \in n, a' \in (N - n)$ qui sont de la forme $\frac{l}{4} 2\theta_{am}\theta_{a'm}\theta_{af}\theta_{a'f}$ donc $B = \frac{l}{2} \sum_{a \in n} \theta_{am}\theta_{af} \sum_{a' \in N-n} \theta_{a'm}\theta_{a'f}$.

La variance de B contient $n(N - n)$ termes de la forme variance $(\theta_{af}\theta_{a'f}) = 1 - \sum_{i,j} \theta_{ai}\theta_{aj}\theta_{a'i}\theta_{a'j} / r^2$ $E(\theta_{ai}\theta_{aj}\theta_{a'i}\theta_{a'j}) = 1$ si $i = j$, elle est nulle autrement.

Donc :

$$E(\text{Variance } \theta_{af}\theta_{a'f}) = 1 - \frac{1}{r} = \frac{r-1}{r} = V$$

La variance de B contient aussi $n(N - n)[n(N - n) - 1]$ termes de la forme : Covariance $(\theta_{am}\theta_{af}\theta_{a'm}\theta_{a'f}, \theta_{am}\theta_{af}\theta_{a''m}\theta_{a'f})$ ou de la forme :

Covariance $(\theta_{am}\theta_{af}\theta_{a'm}\theta_{a'f}, \theta_{am''}\theta_{a'f}\theta_{a''m}\theta_{a'f})$ qui ont tous pour espérance $C_2 = \frac{r-1}{r^2}$.

Donc :

$$E[\text{Variance (B)}] = \frac{l^2}{4} [n(N - n)V + n(N - n)[n(N - n) - 1]C_2] \quad (2)$$

La covariance (A, B) contient $n \times n(N - n)$ termes non nuls de la forme $\left(h + \frac{i}{2} + \frac{N-1}{2}l\right) l/2$ Covariance $(-\theta_{am}\theta_{af}, \theta_{a'm}\theta_{a'f}\theta_{a''m}\theta_{a'f})$ avec $a, a' \in n$ et $a'' \in N - n$. Toutes ces covariances valent en espérance $-C_2$.

$$E[\text{Covariance (A, B)}] = -\frac{l}{2} \left(h + i/2 + \frac{N-1}{2}l\right) n^2(N - n)C_2 \quad (3)$$

Après assemblage des relations (1), (2) et (3) suivi d'une simplification, on trouve :

$$E[R^2(m, m')] = n(V - C_2) \left[\left(h + i/2 + \frac{N-1}{2}l\right)^2 + (N - n)l^2/4 \right] + n^2C_2 \left(h + i/2 + \frac{n-1}{2}l\right)^2$$

Annexe 2

*Valeurs moyennes des croisements des neuf lignées entre elles (diallèle)
et avec leurs 2 parents P et P' (Triple)*

Chaque tableau présente deux caractères, l'un au-dessus de la diagonale, le second au-dessous de la diagonale.

Lignée		L/T										
		1	2	3	4	5	6	7	8	9	P	P'
PT . . .	1 . . .	82,6 31,6	95,0	80,3	89,6	83,7	79,9	94,0	94,6	101,1	155	93
	2 . . .	33,6	68,5 28,2	73,6	80,8	79,4	78,4	87,8	83,4	93,3	153	75
	3 . . .	38,6	33,5	67,1 36,2	72,3	70,9	78,9	82,9	77,4	81,4	146	78
	4 . . .	35,4	32,7	34,1	64,5 27,1	78,7	75,9	82,8	82,7	80,4	146	91
	5 . . .	40,4	36,4	39,8	39,3	69,0 36,6	75,4	85,4	84,3	82,0	141	68
	6 . . .	39,9	38,8	41,6	41,1	40,1	64,9 39,8	80,1	85,3	88,2	146	71
	7 . . .	40,8	35,1	39,0	33,3	43,2	41,1	73,4 31,4	91,1	91,5	149	77
	8 . . .	36,8	33,4	35,4	31,1	41,7	40,9	35,2	73,4 28,8	86,5	158	72
	9 . . .	38,1	32,1	35,4	33,0	38,3	37,9	38,9	34,2	75,6 28,7	163	96
	P . . .	32	30	34	30	34	36	38	34	30		
P' . . .	30	31	27	30	33	33	36	30	34			

Lignée		L/F										
		1	2	3	4	5	6	7	8	9	P	P'
lf	1 . . .	57,8 44,6	63,8	66,0	67,5	64,9	68,5	68,0	64,9	63,6	67	60
	2 . . .	47,7	53,9 36,0	63,7	66,1	65,1	68,0	64,2	62,7	63,8	66	58
	3 . . .	50,7	42,2	64,4 44,7	70,9	66,3	69,1	69,0	67,2	63,8	72	52
	4 . . .	48,5	45,3	47,7	59,1 37,6	70,7	71,7	67,1	65,1	67,1	62	63
	5 . . .	50,0	50,4	48,6	51,4	63,3 55,1	68,9	67,8	68,1	69,3	69	63
	6 . . .	54,6	54,4	48,8	52,2	52,7	67,4 54,4	68,8	72,4	69,2	69	63
	7 . . .	53,0	45,2	51,0	47,8	52,8	55,3	58,4 40,1	66,2	66,5	73	66
	8 . . .	47,9	45,3	49,3	46,6	52,6	56,7	45,4	60,5 41,2	65,4	75	55
	9 . . .	50,4	45,8	44,3	45,9	49,4	52,6	47,4	47,7	65,2 42,6	71	66
	P . . .	44	37	44	39	43	47	47	48	46		
P' . . .	49	50	44	48	58	53	55	46	57			

Lignée		CY										
		1	2	3	4	5	6	7	8	9	P	P'
NE . . .	1 . . .	49,6 8,8	42,2	49,0	46,4	48,0	48,0	47,3	45,1	45,9	47	45
	2 . . .	10,1	47,1 10,2	43,8	45,7	45,7	46,4	44,9	45,6	42,3	45	60
	3 . . .	7,8	11,0	49,6 9,8	46,0	46,9	47,8	46,0	43,8	46,7	49	57
	4 . . .	7,3	9,9	8,1	51,5 7,5	45,9	49,7	47,5	48,4	47,4	48	44
	5 . . .	7,4	10,7	10,4	9,1	51,9 10,2	49,7	47,7	49,1	46,4	48	52
	6 . . .	8,5	10,7	9,6	10,2	10,9	50,2 9,1	49,3	46,5	48,9	47	50
	7 . . .	6,5	8,0	8,8	6,8	9,5	8,0	55,7 6,9	49,7	53,2	45	48
	8 . . .	7,1	8,8	8,0	7,7	8,8	8,6	6,9	50,5 7,5	48,0	48	61
	9 . . .	8,0	11,2	9,4	7,7	9,8	10,4	7,3	7,4	58,0 9,6	53	49
	P . . .	7,5	7,8	7,6	6,9	7,7	7,0	5,6	6,2	7,1		
P' . . .	9,7	13,4	9,9	7,6	12,1	9,6	8,9	9,6	8,7			

Lignée		PG										
		1	2	3	4	5	6	7	8	9	P	P'
P500 . . .	1 . . .	40,8 3,2	151,2	130,9	120,2	90,7	156,2	148,4	177,8	99,7		
	2 . . .	4,0	28,6 2,6	87,6	64,8	95,3	135,3	99,3	93,4	109,2		
	3 . . .	4,4	3,5	45,5 3,5	83,3	80,1	73,3	108,5	139,5	47,6		
	4 . . .	4,1	3,1	3,5	21,7 3,0	70,8	81,8	54,7	66,2	46,1		
	5 . . .	4,0	3,3	3,7	3,7	39,1 3,3	71,5	98,0	95,3	84,9		
	6 . . .	4,2	3,6	4,0	4,0	3,3	27,2 3,3	108,8	120,3	55,7		
	7 . . .	4,5	3,9	4,4	3,8	3,8	4,2	14,1 2,9	72,2	78,2		
	8 . . .	3,9	3,4	4,2	3,3	3,9	4,1	4,2	26,7 2,9	56,8		
	9 . . .	4,1	3,4	3,5	3,5	3,6	3,4	3,9	3,6	12,7 2,9		
	P . . .	3,9	3,7	4,4	3,7	3,9	4,5	4,3	4,3	2,9		
P' . . .	3,4	2,8	2,9	3,8	2,5	3,3	3,2	2,7	3,3			